NGS-Based High-Throughput Screen to Identify MicroRNAs Regulating Growth of B-Cell Lymphoma

Kluiver, Joost; Niu, Fubiao; Yuan, Ye; Kok, Klaas; van den Berg, Anke; Dzikiewicz-Krawczyk, Agnieszka

Published in:
Lymphoma

DOI:
10.1007/978-1-4939-9151-8_12

Publication date:
2019

Methods in Molecular Biology 1956

Springer Protocols

Ralf Küppers *Editor*

Lymphoma

Methods and Protocols

Second Edition

Humana Press

METHODS IN MOLECULAR BIOLOGY

Series Editor
John M. Walker
School of Life and Medical Sciences
University of Hertfordshire
Hatfield, Hertfordshire, AL10 9AB, UK

For further volumes:
http://www.springer.com/series/7651

Lymphoma

Methods and Protocols

Second Edition

Edited by

Ralf Küppers

Institute of Cell Biology (Cancer Research), Medical School, University of Duisburg-Essen, Nordrhein-Westfalen, Essen, Germany

Humana Press

Editor
Ralf Küppers
Institute of Cell Biology (Cancer Research)
Medical School, University of Duisburg-Essen
Nordrhein-Westfalen, Essen, Germany

ISSN 1064-3745 ISSN 1940-6029 (electronic)
Methods in Molecular Biology
ISBN 978-1-4939-9150-1 ISBN 978-1-4939-9151-8 (eBook)
https://doi.org/10.1007/978-1-4939-9151-8

Library of Congress Control Number: 2019934201

This Humana Press imprint is published by the registered company Springer Science+Business Media, LLC, part of Springer Nature.
The registered company address is: 233 Spring Street, New York, NY 10013, U.S.A.

Preface

Since publication of the first edition of this book on methods in lymphoma research in 2013, a variety of further methods have been developed or further refined that can now be used to study the molecular biology of lymphomas. Therefore, this new edition includes several new chapters that describe approaches for single cell RNA sequencing, for the evaluation of whole genome sequencing data, for compound library screens to determine lymphoma cell vulnerabilities and resistance, for CRISPR/Cas9-mediated gene knockout, for analysis of protein interactions by a proximity ligation assay, and for phosphoproteomic analysis of signaling pathways and, finally, an ultrasensitive approach for the detection of circulating cell-free tumor DNA in lymphomas. Besides these new chapters, most of the chapters included in the first edition still address key approaches for the analysis of lymphoid malignancies, so that they remain important for such a collection of methods and have been updated to present the most current version of the methods. As was already the case for the first edition, many of the methods can not only be used for the analysis of lymphoid malignancies but also for other cancers or normal cells of the immune system. Therefore, also this second edition of the book should be of interest not only for hematologists and hematopathologists but also for scientists interested in other fields of cancer research as well as human genetics and immunology.

Nordrhein-Westfalen, Essen, Germany *Ralf Küppers*

Contents

Contributors

Andreas Agathangelidis • *Institute of Applied Biosciences, Centre for Research and Technology Hellas, Thessaloniki, Greece*

Miguel Alcaide • *Department of Molecular Biology and Biochemistry, Simon Fraser University, Burnaby, BC, Canada*

Johannes W. Bagnoli • *Anthropology and Human Genomics, Department of Biology II, Ludwig-Maximilians-University, Martinsried, Germany*

Elke Boone • *Laboratory for Molecular Diagnostics, Department of Laboratory Medicine, AZ Delta Hospital, Roeselare, Belgium*

Sebastian Böttcher • *Division of Internal Medicine, Medical Clinic III—Hematology, Oncology, and Palliative Medicine, Rostock University Medical Center, Rostock, Denmark*

Benedikt Brors • *Applied Bioinformatics, German Cancer Research Center (DKFZ), Heidelberg, Germany; German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ), Heidelberg, Germany*

Monika Brüggemann • *Second Medical Department, University Hospital Schleswig-Holstein, Kiel, Germany*

Christian E. Busse • *Division of B Cell Immunology, German Cancer Research Center (DKFZ), Heidelberg, Germany*

Moez Dawood • *Lymphoid Malignancies Branch, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA; Baylor College of Medicine, Houston, TX, USA*

Agnieszka Dzikiewicz-Krawczyk • *Department of Pathology and Medical Biology, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands; Institute of Human Genetics, Polish Academy of Sciences, Poznan, Poland*

Wolfgang Enard • *Anthropology and Human Genomics, Department of Biology II, Ludwig-Maximilians-University, Martinsried, Germany*

Jonathan R. Fromm • *Department of Laboratory Medicine, University of Washington, Seattle, WA, USA*

Maciej Giefing • *Institute of Human Genetics, Polish Academy of Science, Poznan, Poland*

Emily Glynn • *Department of Laboratory Medicine, University of Washington, Seattle, WA, USA*

Patricia J. T. A. Groenen • *Department of Pathology, Radboud University Medical Center, Nijmegen, The Netherlands*

Martin-Leo Hansmann • *Senckenberg Institute of Pathology, University of Frankfurt, Frankfurt am Main, Germany; Frankfurt Institute of Advanced Studies, Frankfurt am Main, Germany*

Björn Häupl • *Department of Medicine II, Hematology/Oncology, University Hospital Frankfurt, Frankfurt, Germany; German Cancer Research Center (DKFZ) and German Cancer Consortium (DKTK), Heidelberg, Germany*

Kim C. Heezen • *Laboratory Medical Immunology, Department of Immunology, Erasmus MC, Rotterdam, The Netherlands*

Da Wei Huang • *Lymphoid Malignancies Branch, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA*

Daniel Hüebschmann • *Division of Theoretical Bioinformatics (B080), German Cancer Research Center (DKFZ), Heidelberg, Germany; Department of Pediatric Immunology, Hematology and Oncology, University Hospital, Heidelberg, Germany; Division of Stem Cells and Cancer, German Cancer Research Center (DKFZ), Heidelberg, Germany; Heidelberg Institute for Stem Cell Technology and Experimental Medicine (HI-STEM), Heidelberg, Germany*

Charles D. Imbusch • *Applied Bioinformatics, German Cancer Research Center (DKFZ), Heidelberg, Germany; German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ), Heidelberg, Germany; Faculty of Biosciences, Heidelberg University, Heidelberg, Germany*

Aleksandar Janjic • *Anthropology and Human Genomics, Department of Biology II, Ludwig-Maximilians-University, Martinsried, Germany*

Calvin A. Johnson • *Office of Intramural Research, Center for Information Technology, National Institutes of Health, Bethesda, MD, USA*

Joost Kluiver • *Department of Pathology and Medical Biology, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands*

Michael Kneba • *Second Medical Department, University Hospital Schleswig-Holstein, Kiel, Germany*

Klaas Kok • *Department of Genetics, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands*

Ralf Küppers • *Institute of Cell Biology (Cancer Research), Medical School, University of Duisburg-Essen, Nordrhein-Westfalen, Essen, Germany*

Anton W. Langerak • *Laboratory Medical Immunology, Department of Immunology, Erasmus MC, Rotterdam, The Netherlands*

Ryan D. Morin • *Department of Molecular Biology and Biochemistry, Simon Fraser University, Burnaby, BC, Canada*

Fubiao Niu • *Department of Pathology and Medical Biology, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands*

Thomas Oellerich • *Department of Medicine II, Hematology/Oncology, University Hospital Frankfurt, Frankfurt, Germany; German Cancer Research Center (DKFZ) and German Cancer Consortium (DKTK), Heidelberg, Germany*

James D. Phelan • *Lymphoid Malignancies Branch, National Cancer Institutes, National Institutes of Health, Bethesda, MD, USA*

Christoph Plass • *Cancer Epigenomics, German Cancer Research Center (DKFZ), Heidelberg, Germany*

Christiane Pott • *Second Medical Department, University Hospital Schleswig-Holstein, Kiel, Germany*

Fotis Psomopoulos • *Institute of Applied Biosciences, Centre for Research and Technology Hellas, Thessaloniki, Greece*

Matthias Ritgen • *Second Medical Department, University Hospital Schleswig-Holstein, Kiel, Germany*

Sandrine Roulland • *Lymphoid Malignancies Branch, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA; CNRS, INSERM, Centre d'Immunologie de Marseille-Luminy, Aix-Marseille-Luminy, Marseille, France*

Christopher Rushton • *Department of Molecular Biology and Biochemistry, Simon Fraser University, Burnaby, BC, Canada*

Sebastian Scheinost • *Molecular Therapy in Haematology and Oncology, National Center for Tumor Diseases, German Cancer Research Center (DKFZ), Heidelberg, Germany*
Matthias Schlesner • *Bioinformatics and Omics Data Analytics (B240), German Cancer Research Center (DKFZ), Heidelberg, Germany*
Roland Schmitz • *Institute of Pathology, University Hospital Giessen, Justus-Liebig-University Giessen, Giessen, Germany*
Markus Schneider • *Senckenberg Institute of Pathology, University of Frankfurt, Frankfurt am Main, Germany*
René Scholtysik • *Institute of Cell Biology (Cancer Research), Medical School, University of Duisburg-Essen, Nordrhein-Westfalen, Essen, Germany*
Marc Seifert • *Institute of Cell Biology (Cancer Research), Medical School, University of Duisburg-Essen, Nordrhein-Westfalen, Essen, Germany*
Reiner Siebert • *Institute of Human Genetics, Ulm University and Ulm University Medical Center, University Hospital, Ulm, Germany*
Lori Soma • *Department of Laboratory Medicine, University of Washington, Seattle, WA, USA*
Kostas Stamatopoulos • *Institute of Applied Biosciences, Centre for Research and Technology Hellas, Thessaloniki, Greece; Department of Immunology, Genetics and Pathology, Uppsala University, Uppsala, Sweden*
Katarzyna Tomska • *Molecular Therapy in Haematology and Oncology, National Center for Tumor Diseases, German Cancer Research Center (DKFZ), Heidelberg, Germany*
Henning Urlaub • *Bioanalytical Mass Spectrometry Group, Max Planck Institute for Biophysical Chemistry, Goettingen, Germany*
Anke van den Berg • *Department of Pathology and Medical Biology, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands*
Jacques J. M. van Dongen • *Department of Immunohematology and Blood Transfusion, Leiden University Medical Center, Leiden, The Netherlands*
Vincent H. J. van der Velden • *Department of Immunology, Erasmus MC, Erasmus University Medical Center, Rotterdam, The Netherlands*
Menno C. van Zelm • *Department of Immunology and Pathology, Central Clinical School, Monash University and the Alfred Hospital Erasmus MC, Melbourne, VIC, Australia*
Lucas E. Wange • *Anthropology and Human Genomics, Department of Biology II, Ludwig-Maximilians-University, Martinsried, Germany*
Qi Wang • *Applied Bioinformatics, German Cancer Research Center (DKFZ), Heidelberg, Germany; German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ), Heidelberg, Germany*
Hedda Wardemann • *Division of B Cell Immunology, German Cancer Research Center (DKFZ), Heidelberg, Germany*
Daniel E. Webster • *Lymphoid Malignancies Branch, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA*
Dieter Weichenhan • *Cancer Epigenomics, German Cancer Research Center (DKFZ), Heidelberg, Germany*
Brent L. Wood • *Department of Laboratory Medicine, University of Washington, Seattle, WA, USA*
David Wu • *Department of Laboratory Medicine, University of Washington, Seattle, WA, USA*
Ryan M. Young • *Lymphoid Malignancies Branch, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA*

Ye Yuan • *Department of Pathology and Medical Biology, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands; Department of Clinical Pharmacy, The Second Affiliated Hospital of Harbin Medical University, Harbin, China*

Thorsten Zenz • *Department of Haematology, Center for Oncology and Haematology, University Hospital and University of Zurich, Zurich, Switzerland*

Chapter 1

Origin and Pathogenesis of B Cell Lymphomas

Marc Seifert, René Scholtysik, and Ralf Küppers

Abstract

Immunoglobulin (IG) gene remodeling by V(D)J recombination plays a central role in the generation of normal B cells, and somatic hypermutation and class switching of IG genes are key processes during antigen-driven B cell differentiation. However, errors of these processes are involved in the development of B cell lymphomas. IG locus-associated translocations of proto-oncogenes are a hallmark of many B cell malignancies. Additional transforming events include inactivating mutations in various tumor suppressor genes and also latent infection of B cells with viruses, such as Epstein-Barr virus. Many B cell lymphomas require B cell antigen receptor expression, and in several instances, chronic antigenic stimulation plays a role in lymphoma development and/or sustaining tumor growth. Often, survival and proliferation signals provided by other cells in the microenvironment are a further critical factor in lymphoma development and pathophysiology. Many B cell malignancies derive from germinal center B cells, most likely because of the high proliferation rate of these cells and the high activity of mutagenic processes.

Key words B cells, B cell lymphoma, Clonality, Chromosomal translocation, Germinal center, Hodgkin lymphoma, Immunoglobulin genes, V gene recombination, Somatic hypermutation

1 B Cell Development and Differentiation

1.1 Introduction

B cells are lymphocytes that confer efficient and long-lasting adaptive immunity by the generation of high-affinity antibodies against antigens. These cells form an essential part of the humoral immune response and play a central role in immunological memory. Beyond this, B lymphocytes participate in a broad range of immunological functions, including antigen presentation, immune regulation, and provision of a cellular and humoral pre-immune repertoire. Their contribution to the immune system is complex and multilayered.

1.2 B Cell Diversity and Antibody Structure

All mature B cells express a membrane-bound antibody with individual specificity. This immunoglobulin (Ig) is associated with cofactors, and together these molecules form the B cell receptor (BCR). The cofactors immunoglobulin alpha and beta (Igα/Igβ), also termed CD79A and CD79B, participate in signal transduction of this surface receptor. The diversity of immunologically

Ralf Küppers (ed.), *Lymphoma: Methods and Protocols*, Methods in Molecular Biology, vol. 1956,
https://doi.org/10.1007/978-1-4939-9151-8_1,

competent B cells results from the variability of their BCR. This is a consequence of recombination processes during B lymphocyte development in which genes located in the IG loci are joined to give rise to new and individually generated IG genes. Antibodies are composed of four polypeptides, two identical heavy chains (IGH) and two identical light chains (IgL), that are linked by disulfide bonds. The IgL chains are of either κ or λ isotype. All these polypeptides consist of a carboxyterminal constant (C) and an aminoterminal variable (V) fragment. The V region includes four framework regions, each separated by hypervariable regions, the complementarity determining regions 1, 2, and 3 (CDRI to CDRIII). Whereas the V_H region gene is generated by the recombination of three independent genes, the variable (V_H), diversity (D_H), and joining (J_H) genes (also termed *IGHV*, *IGHD*, and *IGHJ*, respectively), the light chain V region genes are composed of only two genes, namely, the V_L and J_L genes [1]. The somatic recombination of these genes is catalyzed by the enzymes recombination-activating genes (RAG)1 and RAG2. These enzymes recognize recombination signal sequences (RSS) flanking the recombining ends of the genes, cut the DNA at these sites, and build hairpin structures at the coding ends [2]. The hairpin structures can be resolved in different ways to generate (palindromic) P elements. Moreover, exonucleases can act arbitrarily to remove nucleotides from the ends of the rearranging genes. The enzyme terminal deoxynucleotidyl transferase (TdT) randomly adds (non-germline-encoded) N nucleotides to the ends of the rearranging genes before they are joined, and DNA repair factors finally complete the recombination process [1].

1.3 B Cell Development and Differentiation

The development of B cells is initiated in the fetal liver and relocated to the bone marrow during maturation of mammalian embryos. Throughout the differentiation processes, the microenvironment of the respective tissues (the microenvironmental niche) plays an essential role in providing nutrition, survival, and developmental stimuli.

Multipotent hematopoietic stem cells give rise to lymphoid precursors that initiate an irreversible differentiation program. The development of B cells from lymphoid precursors is orchestrated by several key transcription factors that determine B cell fate. Early B cell factor 1 (EBF1), E2A, and PAX5 are the three main transcription factors for early B cell development [3]. The production of a functional and unique BCR through V(D)J recombination is the central process for the generation of a mature B cell [4]. Hence, selection processes for appropriate receptor molecules play a key role during B cell development, as nonfunctional or autoreactive B cells have to be eliminated. B cell development is regulated by an ordered rearrangement of antigen receptor genes and can be divided into distinct steps according to the

rearrangement status of the IG loci and phenotypical features. The initial step in B cell development is a D_H to J_H gene rearrangement at the IGH locus on human chromosome 14. In humans, 27 D_H and 6 J_H genes are available for this rearrangement (Fig. 1) [5, 6] that can occur on both alleles. B lymphocyte precursors carrying D_HJ_H joints are called pro-B cells. Subsequently, 1 of about 50 functional V_H genes is rearranged to the D_HJ_H joint (Fig. 1) [7]. The newly generated IG heavy chain is expressed and paired to a surrogate light chain. The so formed pre-BCR is tested for functional competence. If functional, recombination processes of the second allele are suppressed (allelic exclusion), and the B lymphocyte precursor reaches the stage of the pre-B cell [4, 8]. However, often generation of a functional pre-BCR upon V_H to D_HJ_H recombination fails: e.g., 1 of approximately 80 nonfunctional V_H genes encoded in the human genome can be recombined to the D_HJ_H joint [7]. Also, nucleotide insertions or deletions occurring during the rearrangement process can cause frame shifts of the IGH gene (the chance for an out-of-frame rearrangement is 66%), or the expressed heavy chain cannot bind properly to the surrogate light chain and fails to form a stable pre-BCR. In case of a nonfunctional pre-BCR, a rearrangement of the second IGH allele and the potential use of V gene replacement (recombination of a further upstream located V_H gene into the existing $V_HD_HJ_H$ joint, targeting an internal RSS in the rearranged V_H gene) are alternatives for B lymphocyte precursors to generate a functionally competent pre-BCR [9]. In case these escape mechanisms are unsuccessful, the respective B lymphocyte precursor will undergo apoptosis [10]. Only those B cell precursors that survive the selection for a functional pre-BCR start rearranging V_L to J_L light chain genes in order to generate an immunoglobulin light chain. Light chain recombination starts at the κ loci on chromosome 2. In the human, depending on the haplotype, 30–35 functional $V_κ$ genes and 5 $J_κ$ genes are available for recombination [11, 12]. In case of nonfunctional $V_κ$ rearrangements on both alleles, the λ loci on chromosome 22 can be rearranged subsequently with 30–37 functional $V_λ$ and 4 $J_λ$ gene genes available [13, 14]. B cells express either κ or λ light chains, a phenomenon called isotype exclusion. Only in very rare instances (<2% of total B cells), two different light chains (κ and/or λ) are expressed by a single B cell at the same time [4, 15].

When a B cell precursor expresses a functional heavy and light chain that can appropriately pair to form a stable BCR, the stage of immature B cell is reached. At this stage of B cell development, the BCR is expressed exclusively with the IgM isotype of the heavy chain, and the cell is now counterselected for autoreactivity [4]. If the BCR of immature B cells shows reactivity to autoantigens, the corresponding immature B cell has the opportunity to escape counterselection by receptor editing. This includes either the

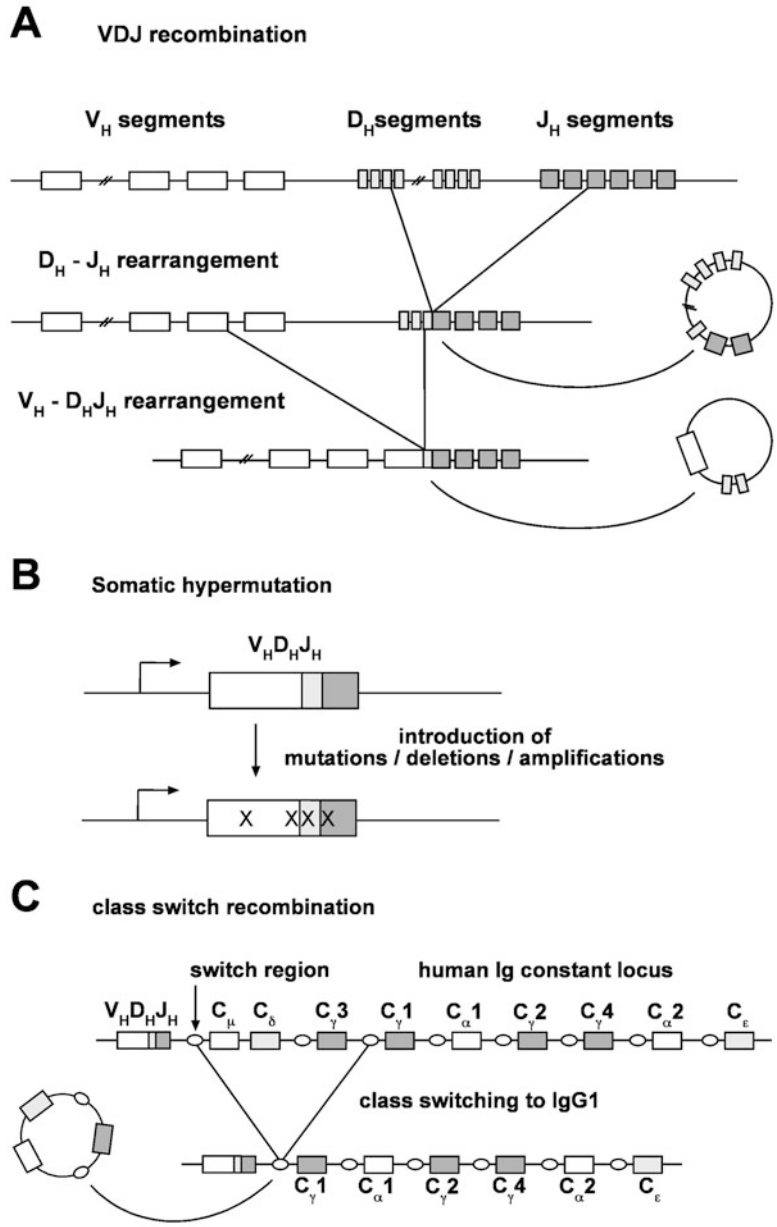


Fig. 1 IG gene remodeling processes in B cells. (**a**) Shown is a schematic presentation of the stepwise rearrangement of V_H, D_H, and J_H genes during B cell ontogeny. Excision circle by-products are depicted to the right. (**b**) The introduction of somatic mutations into a transcribed $V_HD_HJ_H$ gene by the somatic hypermutation machinery active in GC B lymphocytes. Each "X" denotes an independent mutational event. (**c**) Schematic presentation of class switch recombination from Cμ to Cγ1 on the human IGH chain constant region locus. An excised switch circle is shown on the left

usage of so far unrearranged IG loci or potentially the additional rearrangement of upstream V with downstream J genes in the light chain loci [16, 17]. Negatively selected B cells are driven into an anergic state (i.e., immunological unresponsiveness) or are eliminated by apoptosis. Those B cells with positively selected functional BCR are called mature, naive B cells. These cells coexpress IgM and IgD through differential splicing of the V_H region exon to the Cμ and Cδ constant region exons coding for the IgM and IgD isotypes. Mature B cells leave the bone marrow and circulate as small, resting lymphocytes in peripheral blood and secondary lymphoid tissues.

1.4 The Germinal Center Reaction

Upon contact with cognate antigen, mature B cells are activated and migrate into the T cell zones of secondary lymphatic organs. The interaction with activated T helper cells provides further stimulation to the B lymphocytes and induces proliferation, leading to the formation of primary foci. A fraction of the proliferating B cells differentiates into short-lived plasma cells that secrete antibodies of mostly IgM isotype and hence provide an initial wave of low-affinity antibodies. Some B cells, however, migrate together with activated T helper cells into B cell follicles and initiate a germinal center (GC) reaction [4, 18].

In the GC, the highly proliferative GC B cells interact in immunological synapses with follicular T helper cells and follicular dendritic cells (FDCs). These tightly regulated mechanisms of proliferation and cellular interaction lead to a histological structure that is characteristic for all GCs. A loosely associated network of FDCs, follicular T helper cells, and GC B cells (centrocytes) form a so-called light zone, whereas a dense area of quickly dividing B cells (centroblasts) can be recognized as dark zone. The expanding GC displaces the locally residing, resting B lymphocytes and compacts them into a mantle zone surrounding the GC. GC B cells circulate mainly within but also between the two zones [4, 18]. The underlying mechanisms that drive these processes are currently under thorough investigation [19–23].

In the dark zone, the process of somatic hypermutation (SHM) introduces point mutations as well as small deletions, insertions, or duplications into the rearranged IgV genes (Fig. 1) [4, 24, 25]. The SHM process is strictly dependent on transcriptional activity of the affected template and mainly takes place in a regionally defined area of 1–2 kb downstream of the V gene promoters [26]. Moreover, special sequence motifs (SHM hot spots) are preferentially targeted, and nucleotide exchanges have a transition over transversion bias when compared to a random mutation process [26, 27]. The SHM mechanism is essentially dependent on the enzyme activation-induced cytidine deaminase (AID) that converts cytidine into uracil in the affected DNA strand [26]. As uracil is not a normal component of DNA, and as U does not pair with G as does C, these sites are subsequently targeted by error-prone DNA

repair mechanisms [28]. SHM also affects some non-IG genes, although at a much lower rate [29, 30].

The somatic IgV gene mutations may lead to a change in BCR affinity. The centroblasts that acquired mutations migrate into the light zone and compete with other GC B cells for survival signals from FDC and T helper cells. The amount of survival signal correlates with improved or at least retained affinity to the cognate antigen: comparably low affinity is counterselected by induction of apoptosis in the corresponding centrocytes [4, 24, 31]. GC B cells undergo multiple cycles of proliferation, mutation, and selection. This iterative process leads to a stepwise improvement of the affinity of the BCR to its cognate antigen.

Another important DNA recombination process occurs in centrocytes: the constant region of the antibody heavy chain (C_H) may be exchanged by class switch recombination. The C_H region of naive B cells is initially expressed with IgM and IgD isotype due to alternative splicing. During class switching, the Cμ and Cδ genes can be replaced by one of either two Cα, one Cε, or four Cγ genes in humans (Fig. 1). This recombination process is dependent on AID and mediated by DNA double-strand breaks in specialized switch regions upstream of the C_H genes [26]. Upon deletion of Cμ and Cδ, the $V_HD_HJ_H$ exon will now be expressed as part of a heavy chain with the C_H gene that replaced the Cμ gene. Class switching leads to changes in BCR signaling competence and modified effector functions of the antibody [32]. Notably, class switching is not an obligatory feature of GC B cells, as part of the GC B cell progeny—mostly early descendants—leave the GC as non-class-switched lymphocytes [33–35].

After several cycles of proliferation, mutation, and positive selection, GC B cells differentiate into either antibody-secreting plasma cells or resting memory B cells and leave the GC microenvironment [36].

1.5 B Cell Memory

Memory B cells and post-GC plasma cells provide the two most important functions of B cell adaptive immunity: first of all, post-GC plasma cells antagonize the invaded pathogens with a potent wave of high-affinity antibodies. Second, both cell types provide the organism with the potential of enhanced and improved immune responses upon reencounter of antigen [37]. These quick and effective secondary immune responses constitute humoral immunological memory [38]. High-affinity antibodies are secreted over long periods by long-lived post-GC plasma cells that reside in specialized niches in the bone marrow [39]. Quiescent long-lived memory B cells can easily be reactivated to take part in improved secondary responses [37, 38].

1.6 T Cell-Independent Immune Responses

Whereas for the induction of a GC reaction, a B lymphocyte is dependent on the interaction and communication with T cells, the activation of B lymphocytes and their differentiation into plasma cells may also occur without T cell help in T cell-independent (TI) immune responses. There are two types of TI activations of B cells: either antigens (mitogens) that trigger conserved pattern recognition receptors (e.g., Toll-like receptors) provoke a polyclonal B cell response (TI-1) [40], or antigens with highly repetitive structure (e.g., bacterial capsules) activate specific B lymphocytes by intensive BCR triggering (TI-2) [41]. Typically, TI immune responses do not give rise to memory B cells. Plasma cells generated in TI immune responses are short-lived and unmutated [42]. Class switching can also take place during TI immune responses, mainly to the IgG2 isotype in humans.

2 Cellular Origin of Human B Cell Lymphomas

When B cells undergo malignant transformation, they usually retain key features of their cell of origin, including specific characteristics of the particular differentiation stage of the lymphoma precursor [36, 43]. Histological and immunohistochemical studies of lymphomas have hence been very important to classify B cell malignancies and determine the cellular derivation of these tumors. For example, in follicular lymphoma, the tumor cells morphologically resemble GC B cells, they express typical markers of GC B cells, and they grow in follicular structures that resemble GC and that harbor GC T helper cells and FDC networks [44]. Thus, all these features point to a GC B cell derivation of follicular lymphomas.

When it became feasible to comprehensively study IG gene rearrangements by polymerase chain reaction (PCR) and sequencing, the histopathological evaluations of lymphomas were complemented by V gene analyses. Such studies, for example, validated the GC B cell origin of follicular lymphomas by showing that the lymphoma cells carry somatically mutated IG V genes and show intraclonal diversity as a sign of ongoing SHM throughout clonal expansion, further characteristic features of GC B cells [45].

The cellular origin of Burkitt lymphomas could not be clarified based on the histological picture, because this highly aggressive lymphoma shows a disruption of the normal lymph node structure and growth in a diffuse pattern. However, the lymphoma cells morphologically resemble centroblasts, they express key GC B cell markers, and they have somatically mutated IG V genes with ongoing hypermutation in a fraction of cases [46]. Thus, Burkitt lymphomas derive from GC B cells, too.

The development of transcriptome analysis tools enabled a much more comprehensive comparison of the gene expression of human B cell lymphomas to normal B cell subsets than was

previously possible by immunohistochemical stainings for single or few markers. Such comparisons were consequently widely used in recent years to identify the cellular origin of lymphomas. One landmark finding was that the heterogeneous group of diffuse large B cell lymphomas (DLBCL) can be subdivided into several subgroups [47, 48]. The two main subgroups were defined by a high similarity of the lymphoma cells to GC B cells (GCB-DLBCL) or to in vitro-activated B cells (ABC-DLBCL) [47, 48]. As the GCB-DLBCL also often showed ongoing somatic hypermutation, this lymphoma is now considered as a further GC B cell lymphoma [49]. ABC-DLBCL have a highly activated phenotype and carry somatically mutated V genes but lack most specific features of GC B cells. This lymphoma is most similar to post-GC immunoblasts.

A special case is classical Hodgkin lymphoma, because the tumor cells of this malignancy, the Hodgkin and Reed/Sternberg (HRS) cells, express only few B cell markers and express multiple markers of various other hematopoietic cell types [50]. However, the B cell origin of HRS cells from mature B cells was unequivocally shown by the demonstration that these cells carry clonally rearranged and somatically mutated V region genes [51, 52]. Surprisingly, in a quarter of cases, destructive somatic mutations were found in originally productive V gene rearrangements [51, 52]. Based on this finding, it was proposed that HRS cells derive from the pool of pre-apoptotic GC B cells that acquired unfavorable mutations and that normally would have undergone apoptosis [51].

As a main finding of the numerous studies to reveal the cellular origin of human B cell lymphomas, it can be concluded that the majority of these lymphomas is derived from GC or post-GC B cells [36, 46] (Fig. 2). This is remarkable, because B cells reside only for a few days to weeks in a GC, and more than half of the B cell pool is represented by naive B cells. The vigorous proliferation of the GC B cells may represent one critical factor why these cells become transformed, permanently proliferating cells. Moreover, the genetic processes of SHM and class switching are mutagenic processes that strongly increase the risk for a B cell to undergo malignant transformation, as will be discussed in the following paragraph. It should, however, also be stressed that lymphoma development is a multistep process and that genetic lesions may be acquired in a lymphoma precursor over multiple differentiation steps [53, 54]. It is a matter of discussion which of these intermediate steps to define as the cell of origin. For example, in follicular lymphomas, the prototypic GC B cell lymphoma, the t(14;18) *BCL2*/IGH chromosomal translocation found in nearly all cases occurs at the pro-B cell stage of B cell development during misguided V gene recombination. However, this translocation becomes pathogenetically relevant only much later in B cell development in GC B cells when BCL2 is normally downregulated.

To date, a growing number of studies make use of genome-wide profiling strategies to determine the normal counterpart to a given leukemia/lymphoma. Such approaches are also suitable for entities that do not resemble GC B cell phenotypes, including multiple myeloma [55], hairy cell leukemia [56], and chronic lymphocytic leukemia (CLL) [57, 58]. Irrespective of the tumor entity, the key focus of such studies is to determine as precisely as possible what patterns (and functions) among tumor cells are normal, i.e., taken over from the cell type in that the malignant transformation process occurred. Knowing what remained normal in a tumor may not only reveal the initial, most relevant steps in pathogenesis but may also point to tumor-specific features, the potential Achilles' heel for targeted treatment. As tumor development may occur in multiple steps, potentially decades apart, and transformation may have strong, ongoing influence on genome-wide patterns, the integration of multiple screening methods and more than one normal cell type may be mandatory to understand the cellular origin of a tumor (see below).

3 Genetic Lesions in the Pathogenesis of B Cell Lymphomas

To reach full malignancy, several security checks in a B cell have to fail. The current theory of cancer development promulgates that a normal cell needs multiple "hits" that change its normal functions regarding the control of proliferation, apoptosis, and regulation by other cells. These hits are acquired in B cell malignancies both by largely random genetic lesions and by erroneous B cell-specific processes, as discussed below. Subsequently, cells with genetic lesions are selected in an evolutionary process for environmental fitness and survival and have the chance to acquire additional changes. The nature of these hits is diverse, including viral infection, genetic mutations, and epigenetic restructurations (Table 1).

A hallmark of many B cell lymphomas are chromosomal translocations involving one of the IG loci and a proto-oncogene. These translocations happen as by-products during the processes of V(D)J recombination, SHM, and class switching [59]. Apparently, the DNA strand breaks occurring in each of these processes bear an inherent risk of generating translocations. The strand breaks in the loci of the proto-oncogenes may be random, involve some recombination-prone sites, and in some instances are due to off-target activity of SHM in some non-IG genes (e.g., *BCL6*) [60]. Off-target activity of SHM can also cause point mutations in proto-oncogenes and is most frequently seen in DLBCL, where many more genes are targeted by aberrant SHM [60].

The translocation of a proto-oncogene into an IG locus in B cell lymphomas deregulates the oncogene expression through the associated IG enhancers, which are highly active in B cells. For

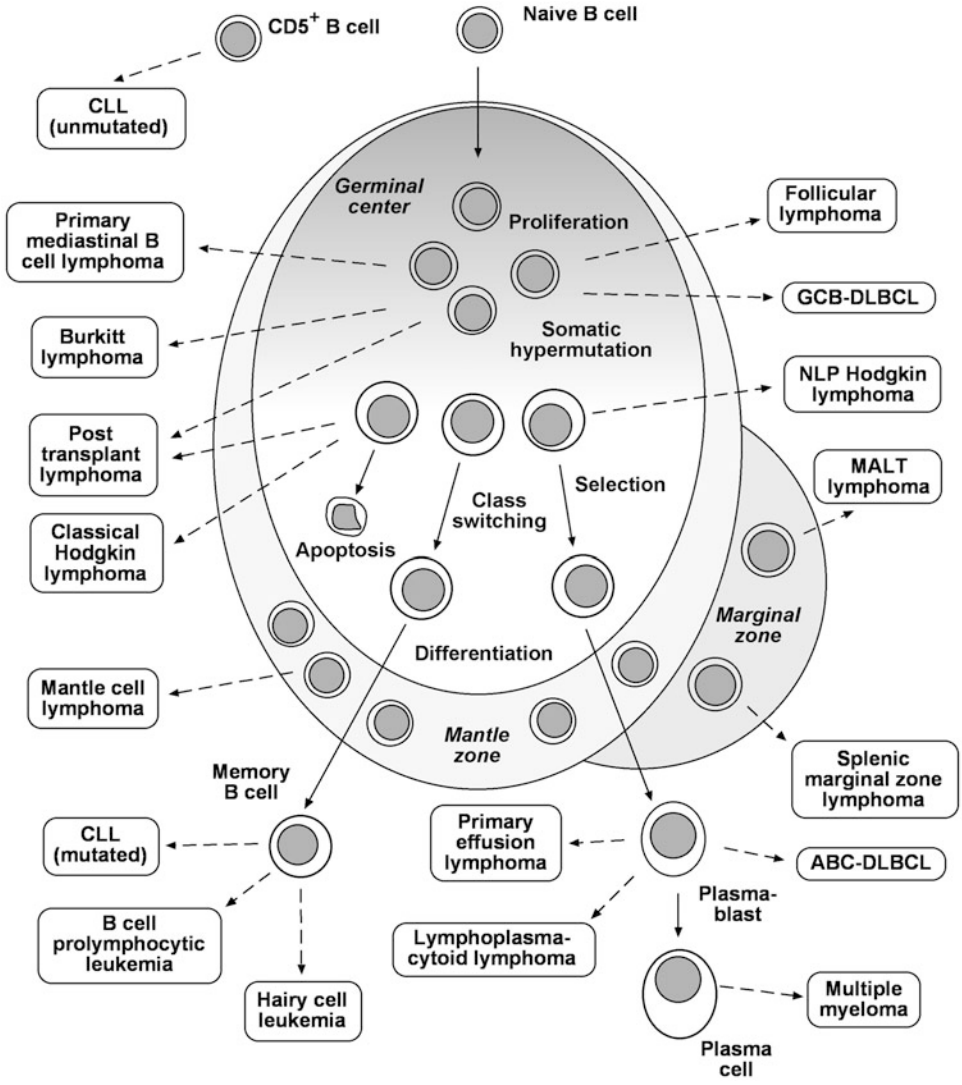


Fig. 2 Germinal center reaction and cellular origin of human B cell lymphomas. Shown are the main steps in mature B cell differentiation in the GC and the presumed cellular origin of human B cell lymphomas. Antigen-activated mature B cells are driven into a GC reaction when T cell help is available. The GC B cells undergo massive clonal expansion in the dark zone of the GC and activate the process of SHM. Mainly in the light zone, mutated B cells are selected for affinity-increasing IgV gene mutations. Positively selected cells will undergo multiple rounds of proliferation, mutation, and selection before they differentiate into memory B cells or plasmablasts and exit the GC. The majority of GC B cells will acquire disadvantageous mutations and undergo apoptosis. Many GC B cells perform class switch recombination in the light zone. Most lymphomas are derived from GC B cells or from post-GC B cells. Also in the latter types of lymphomas, decisive transforming events have presumably occurred in premalignant GC B cell precursors of these lymphomas. CLL with unmutated IgV genes is derived from $CD5^+$ B cells. Mutated CLL is presumably derived from a small subset of $CD5^+$ memory B cells. Most mantle cell lymphomas carry unmutated IgV genes and are presumably derived from ($CD5^+$) mantle zone B cells. However, 20–30% of these lymphomas carry mutated V genes, suggesting a GC experience. Primary mediastinal large B cell lymphoma is likely derived from GC-experienced thymic B cells. A clear marginal zone is present around B cell follicles in the spleen. Marginal zone B cells are likely the

example, in over 90% of follicular lymphomas, *BCL2* is found translocated into the heavy chain locus, sustaining the continuous expression of the *BCL2* oncogene [61, 62]. The situation is similar in Burkitt and mantle cell lymphoma, where *MYC* and *CCND1*, respectively, are translocated into an IG locus in virtually every case [63–65]. MYC and CCND1 both cause uncontrolled proliferation of the cells harboring the respective translocation.

Other lymphoma entities do not show such monotypic translocation patterns as Burkitt lymphoma or mantle cell lymphoma. Although diverse translocations were described in DLBCL, MALT (mucosa-associated lymphatic tissue) lymphoma, and multiple myeloma, none of these is found in the majority of cases (Table 1). This may also reflect the more diverse phenotypes of these lymphomas.

The same processes that mediate a translocation into an IG locus sometimes cause non-IG translocations. For some of them, the postulated effect is deregulation by a similar mechanism as in IG translocation. Here, a regulatory element on the translocation partner overrides the normal expression pattern of a gene. This is, for example, the case in a diverse range of *BCL6* translocations in DLBCL and lymphocyte-predominant Hodgkin lymphoma, leading to a constitutive overexpression of this gene [66–68].

A second class of translocations leads to the formation of fusion genes. These are transcribed as fusion transcripts that span the translocation breakpoint, allowing the translation of proteins that combine parts of two separate genes. In MALT lymphoma, about one third of cases harbors a translocation that leads to the expression of the fusion protein API2-MALT1 [69]. This protein, a combination of the amino terminus of API2 with the carboxy terminus of MALT1, has acquired the function to noncanonically activate the NFκB pathway by cleavage of NFκB inducing kinase MAP3K14 (NIK) into a constitutively active form.

Probably because of the tight surveillance of normal B cell populations, most B cell lymphomas acquire changes during their clonal evolution that allow them to bypass intrinsic or extrinsic signals that normally trigger apoptosis. A central player in the apoptosis network is *TP53*. This gene encodes the tumor suppressor protein p53, a transcription factor influencing the expression of genes that partake directly or indirectly in apoptosis, senescence, and cell cycle progression. Alterations in this single gene have been found in a variety of lymphomas (Table 1) and can have a profound

Fig. 2 (continued) origin of splenic marginal zone B cells, although it is puzzling that normal marginal B cells harbor mutated IgV genes, whereas a considerable fraction of splenic marginal zone B cell lymphomas has unmutated V genes [209]. The tumor cells of classical Hodgkin lymphoma and some posttransplant lymphomas carry destructive IgV gene mutations, indicating a derivation from pre-apoptotic GC B cells

Table 1
Genetic lesions in human B cell lymphomas

Entity	Chromosomal translocations	Tumor suppressor gene mutations/deletions/silencing	Additional alterations
Mantle cell lymphoma	*CCND1*-IGH (95%) [65]	*ATM* (40%) [125, 126] *DLEU2*/miR-15a/16-1 (Del13q14, 50–70%) [74] *CDKN2A* (16–31%) [127] *TP53* (13–45%) [127] *TNFAIP3* (40%) [128] *MEF2B* (31%) [129] *KMT2D* (23%) [129]	*NOTCH1* missense mutations (12%) [130] *CCND1* missense mutations (aberrant somatic hypermutation target, 35%) [131]
Chronic lymphocytic leukemia	–	*ATM* (30%) [132, 133] *TP53* (15%) [134] *DLEU2*/miR-15a/16-1 (Del13q14, 60%) [76]	*NOTCH1* missense mutations (10–20%) [135, 136] *SF3B1* missense mutations (10–20%) [137, 138]
Follicular lymphoma	*BCL2*-IGH (90%) [61, 62]	*KMT2D* (89%) [70, 139] *CREBBP* (33%) [70, 71] *EP300* (10–20%) [71] *TNFRSF14* (30–40%) [140]	*EZH2* missense mutations (20–30%) [141] *ARID1A* missense mutations (15%) [142] *MEF2B* missense mutations (10–20%) [143] *STAT6* missense mutations (10–15%) [144] *RRAGC* missense mutations (15–20%) [145, 146]
Diffuse large B cell lymphoma	*BCL6*-(many loci) (35%) [66, 68] *BCL2*-IGH (15–30%) [147] *MYC*-IGH or *MYC*-IgL (15%) [148]	*CD95* (10–20%) [149] *ATM* (15%) [150] *TP53* (25%) [151, 152] *SOCS1* (25%) [153] *BLIMP1* (20% of ABC-DLBCL) [154] *CREBBP* (20%) [71] *EP300* (10%) [71] *KMT2D* (32%) [70] *CDKN2A* (35%) [155] *TNFAIP3* (38%) [156]	Mutations of proto-oncogenes (by aberrant somatic hypermutation) (50%) [60]

Primary mediastinal B cell lymphoma	*CIITA* (diverse breaks, 38%) [157]	*SOCS1* (40%) [158] *STAT6* (36%) [159] *TNFAIP3* (36%) [160] *NFKBIE* (23%) [161] *CIITA* (53%) [162] *PTPN1* (22%) [163]	Mutation of proto-oncogenes (by aberrant somatic hypermutation) (70%) [164] *XPO1* (E571K, 24%) [165] *IL4R* (activating missense mutations, i.e., I242N, 42%) [166]
Burkitt lymphoma	*MYC*-IGH or *MYC*-IgL (100%) [63, 64]	*TP53* (40%) [134] *RB2* (20–80%) [167] *ID3* [60] [168–170] *DDX3X* (30%) [168, 169] *SMARCA4* (25%) [168–170] *GNA13* (20%) [168–170]	*TCF3* missense mutations (30%) [169] *CCND3* missense mutations (40%) [168, 169]
Posttransplant lymphoma	–	*KMT2D* (35%) [171] *TP53* (35%) [171] *TET2* (14%) [171]	*MYC* aberrations (more frequent in EBV$^+$ cases, average 12%) [171]
Classical Hodgkin lymphoma	*CIITA* (diverse breaks, 15%) [157]	*NFKBIA* (10–20%) [172–174] *NFKBIE* (10%) [161, 175] *TNFAIP3* (44%) [160] *SOCS1* (42%) [176] *B2M* (30–70%) [177, 178] *PTPN1* (20%) [163] *GNA13* (24%) [178] *ITPKB* (16%) [178] *EP300* (12%) [179]	*REL* amplifications (50%) [180] *JAK2* gains (25%) [181] *MAP3K14* gains [182, 183] *STAT6* missense mutations (32%) [178, 179] *XPO1* (E571K, 20%) [165, 178] *BTK* missense mutations (10%) [179] *CSF2RB* missense mutations (12%) [179]
Lymphocyte-predominant Hodgkin lymphoma	*BCL6*-(diverse loci) (48%) [67]	*SOCS1* (50%) [184] *JUNB* (46%) [185]	*SGK1* missense mutations (46%) [185] *DUSP2* missense mutations (54%) [185]
Splenic marginal-zone lymphoma	–	*KMT2D* (15%) [186] *TP53* (14%) [186] *KLF2* (42%) [187]	Del 7q22-36 (40%) [188] *NOTCH2* missense mutations (20%) [186] *MYD88* (L265P, 15%) [189]

(continued)

Table 1
(continued)

Entity	Chromosomal translocations	Tumor suppressor gene mutations/deletions/silencing	Additional alterations
Hairy cell leukemia		*CDKN1B* (15%) [190, 191] *KMT2C* (15%) [190]	*BRAF* (95%) [192]
MALT lymphoma	*API2-MALT1* (30%) [69] *BCL10*-IGH (5%) [193, 194] *MALT1*-IGH (15–20%) [195] *FOXP1*-IGH (10%) [196]	*CD95* (5–80%) [149] *CDKN2A* (60%) [197]	Many alterations specific for site of occurrence (e.g., *TNFAIP3* inactivation in ocular adnexa MALT lymphoma [198], *TET2* inactivation in thyroid MALT lymphoma) [199]
Lymphoplasmacytoid lymphoma	*PAX5*-IGH (50%) [200]	–	*MYD88* (L265P, 91%) [201]
Primary effusion lymphoma	–	–	Lymphoma cells often driven by HHV8 and EBV infection
Multiple myeloma	*CCND1*-IGH (15–20%) [202] *FGFR3*-IGH (10%) [203] *MAF*-IGH [5–10] [204]	*CD95* (10%) [205] *DLEU2*/miR-15a/16-1 (Del13q14, 50%) [75]	*MYC* alterations (40%) [206] *RAS* missense mutations (40%) [207] Gains of 5q (50%) [208]

impact on the cell's reaction to a multitude of stimuli, for example, DNA damage or cell surface receptor signaling.

Another important apoptosis-related pathway that is often blocked is the sensing of cellular stress and DNA damage during cell cycle checkpoints. For example, the checkpoint guardians ATM and CDKN2A are often inactivated by mutations, deletions, or epigenetic silencing in lymphomas (Table 1). This allows cell growth and proliferation despite the accumulation of DNA damage during tumor progression.

The NFκB pathway is important for many cell types during inflammation, including B cells. It is exploited by some tumors to gain resistance to external and internal danger signals. Hodgkin lymphoma, primary mediastinal B cell lymphoma, ABC-DLBCL, MALT lymphoma, and mantle cell lymphoma seem to sustain a gene expression program resembling continuous inflammation by acquiring inactivating mutations in inhibitors of the NFκB complex (e.g., *TNFAIP3*, *NFKBIA*, and *NFKBIE*) and/or activating mutations, e.g., in *REL*, a component of the NFκB complex itself (Table 1).

A cellular function that is not completely understood but nevertheless found to be recurrently deregulated in several types of lymphomas is chromatin remodeling. Follicular lymphoma and DLBCL show recurrent inactivation of KMT2D (MLL2), a histone methyltransferase, and CREBBP, a histone acetyltransferase [70, 71]. Artificial inhibition of these genes in cell lines has shown that their activity controls the chromatin state of a variety of genetic loci. This probably allows the malignant cell to silence tumor suppressor genes and to reactivate epigenetically silenced oncogenes by rewriting histone marks. Indeed, many lymphomas show an abnormal pattern of histone marks and/or DNA methylation when compared to their nonmalignant counterparts [72, 73].

In recent years, with advances in the understanding of microRNAs, it became clear that this class of small noncoding RNAs is another frequent target of deregulation in lymphomas. For example, the microRNA cluster discovered in the minimal deleted region of CLL (comprising *miR-15a* and *miR-16-1*) is also recurrently deleted in mantle cell lymphoma and multiple myeloma [74–76]. Both this cluster and the single protein-coding gene *DLEU2* in this region on chromosome 13 have been shown to act as tumor suppressors [77], providing a selective advantage for tumor clones that have lost one or both copies of this region.

4 Epigenetic Alterations

The detection of numerous genetic lesions in various types of lymphomas that affect epigenetic regulators (see above) indicates that also epigenetic alterations play an important role in the

pathogenesis of lymphomas. Indeed, B cell lymphomas typically show substantial alterations of their DNA methylation pattern (the most widely analyzed epigenetic mark), and a common theme is a general hypomethylation throughout the genome but also hypermethylations in particular regions [78, 79]. In CLL, two independent studies uncovered that three subgroups of CLL can be separated based on their DNA methylation pattern [79, 80]. One is linked to CLL with unmutated IgV genes and one with IgV-mutated CLL, and the third is an intermediate group. These subgroups are also associated with distinct clinical behavior [81]. A detailed comparison of CLL DNA methylation patterns with those of distinct human mature B cell subsets revealed that many differential methylation patterns actually do not represent aberrant patterns as originally thought when comparing CLL to bulk $CD19^+$ B cells but reflect similarities of distinct cases of CLL to distinct differentiation stages of mature B cells [80]. Thus, the extent of tumor-specific methylation alterations in CLL is much lower than implied from earlier studies. A further study revealed that CLL cells show substantial intra-tumor heterogeneity in DNA methylation [82]. This methylation disorder in CLL is linked to transcriptional variation and may affect genetic evolution and clinical outcome [82]. The ability of epigenetic studies to uncover subgroups of a lymphoid malignancy was also seen for mantle cell lymphomas, where based on DNA methylation profiling two subgroups of this lymphoma emerged [83]. One was associated with mainly pre-GC B cells and the other with post-GC B cells, which also fitted to the IgV gene mutation patterns of the subgroups. A major finding of an in-depth epigenetic analysis of Burkitt lymphomas revealed that many pathways which are affected by recurrent genetic lesions in this type of lymphoma are also affected by alterations of the DNA methylation patterns of genes encoding critical factors of these signaling pathways [78]. Hence, genetic and epigenetic alterations cooperate in the pathogenesis of Burkitt lymphoma.

5 Viruses in B Cell Lymphomas

In several types of B cell lymphomas, viruses are implicated in their pathogenesis. In B cell lymphomagenesis, mainly two members of the γ herpesvirus family are involved, i.e., Epstein-Barr virus (EBV) and human herpesvirus 8 (HHV8) [84, 85]. EBV is found in the tumor cells in nearly all cases of endemic Burkitt lymphoma in Africa and about 30% of sporadic Burkitt lymphomas in other parts of the world. EBV is also found in most lymphomas in post-transplant patients (posttransplant lymphoproliferative disease, PTLD) and in a fraction of classical Hodgkin lymphoma. About 30% of classical Hodgkin lymphomas in the Western world show an EBV association, and pediatric cases of this disease in Central

America are EBV-positive in more than 80% of cases. Moreover, EBV is found in some DLBCL (in particular plasmablastic lymphomas) and primary central nervous system lymphomas. EBV establishes a latent infection in B cells, and different forms of latency are seen in the EBV-associated lymphomas [85]. In Burkitt lymphoma, EBV typically shows latency form I, which means that only one EBV-encoded protein is expressed, i.e., EBV nuclear antigen 1 (EBNA1). EBNA1 is essential for the replication of the episomal EBV genome in replicating cells. Whether EBNA1 also has oncogenic functions is still debated. In the HRS cells of Hodgkin lymphoma, besides EBNA1, the two latent membrane proteins 1 and 2a (LMP1 and LMP2a, respectively) of EBV are expressed, a pattern which is called latency II. LMP1 mimics an active CD40 receptor and is a classical oncogene [86]. LMP2a mimics an active BCR. Hence, two main survival signals for B cells in the GC are provided by the virus. Indeed, it has been shown that EBV can rescue BCR-deficient GC B cells from apoptosis, supporting an important role of the virus in the transformation of such B cells [87, 88]. In line with this view, all cases of classical Hodgkin lymphoma in which the HRS cells carried crippling mutations preventing expression of a BCR were found to be EBV-positive [89].

In PTLD, nine viral proteins are usually expressed (latency pattern III). Besides EBNA1, LMP1, and LMP2a, also LMP2b and the EBNAs 2, 3a, 3b, 3c, and LP are expressed [85]. EBNA2 is important to drive the proliferation of the transformed B cells. EBV encodes also multiple noncoding RNAs, including the EBERs (EBV-encoded RNAs) and numerous microRNAs. EBERs are expressed by all EBV-infected B cells, which are often used to detect such cells by in situ hybridization for these transcripts.

HHV8 is found in virtually all cases of primary effusion lymphoma, which is a very rare lymphoma mainly occurring in acquired immune deficiency syndrome (AIDS) patients [84]. Similar to EBV, HHV8 establishes a latent infection in B cells and persists as an episome in the cells. Several viral proteins are implicated in the pathogenetic role of HHV8 in primary effusion lymphomas, including LNA1, which inhibits p53 and Rb, a viral cyclin and a viral FLICE inhibitory protein [84]. Notably, a fraction of primary effusion lymphomas is coinfected with HHV8 and EBV [84].

Hepatitis C virus (HCV) does not establish a latent infection in B cells but is also implicated in B cell lymphomagenesis. In chronic HCV carriers, HCV appears to drive B cell lymphomagenesis by two means. On the one hand, HCV may act as a chronic antigenic stimulus for HCV-specific B cells, which is indicated by the specificity of the BCR of some HCV-associated lymphomas for viral antigens [90]. On the other hand, HCV can upregulate AID, the key factor of SHM, in B cells and also causes the production of reactive oxygen species, two factors that can be mutagenic for B

cells [91, 92]. A general increased mutation load in several candidate genes was, however, not found in mature B cells of HCV-infected individuals [93].

Several additional viruses have been discussed as potential factors in B cell lymphoma pathogenesis [94, 95], but further studies are needed to validate these findings.

6 Microenvironmental Interactions in B Cell Lymphomas and the Role of the BCR

In many B cell lymphomas, interaction of tumor cells with other cells in the microenvironment plays an essential role for survival and proliferation of the tumor. A prototypical example for this is follicular lymphoma; here the transformed GC B cells grow in follicular structures resembling normal GC in close association with FDC and follicular T helper cells. The dependency of follicular lymphoma cells on these normal constituents of a GC is suggested from in vitro studies. The malignant cells survive longer when cocultured with $CD4^+$ T cells or stromal cells and with stimulation of the CD40 receptor, a main survival receptor for GC B cells [96, 97]. As normal GC B cells require BCR triggering for their survival, it was an important finding that the BCRs of follicular lymphomas often acquire replacement mutations in their rearranged IgV genes that allow linking of sugar moieties to the V regions [98]. There is evidence that sugar-binding receptors (lectins) on stromal cells can constantly trigger the BCR of the lymphoma cells and hence replace the original foreign antigen that would normally be only transiently available to stimulate the GC B cells [99, 100].

A similar situation as seen in follicular lymphoma is given in CLL, where proliferation mostly takes place in specialized structures in lymph nodes [101]. In these proliferation centers, CD40-expressing CLL cells are in contact with stromal and $CD4^+$ T cells, the latter partly expressing CD40L [102]. The relevance of these interactions for the survival of CLL cells is indicated from the observation that CLL cells survive only for a short time when cultured alone, but their survival is significantly improved upon coculture with stromal cells or CD40 stimulation [103]. Moreover, there is now evidence that the BCRs of CLL cells are often polyreactive to foreign and autoantigens, including modified components from apoptotic cells [104, 105], or even bind to itself [106]. Whether and how autoantigens are presented in the proliferation centers is currently unclear.

Gastric MALT lymphomas are frequently associated with chronic *Helicobacter pylori* infections. The lymphoma cells themselves usually do not recognize these bacteria via their BCR but often express an autoreactive surface BCR [107]. However, $CD4^+$ T cells in the MALT lymphoma microenvironment are activated

upon interaction with *H. pylori*, and it thus appears that *H. pylori*-activated T cells stimulate the lymphoma B cells [108]. The importance of this interaction is evident from the fact that elimination of *H. pylori* by antibiotic treatment often leads to regression of the lymphoma [109]. Interestingly, also in splenic B cell lymphomas in patients chronically infected with HCV, antiviral treatment resulted in regression of the lymphomas [110], representing a further example for chronic antigenic stimulation as critical survival signal for lymphoma cells. Further evidence for autoreactivity among lymphoma BCRs came from studies on ABC-DLBCL [111], marginal zone lymphomas [112, 113], CLL [114, 115], and follicular lymphoma [116].

Regarding the extent (and possibly the role) of the microenvironment in lymphomas, classical Hodgkin lymphoma is a rather unique lymphoma, because here the tumor cells usually account for only about 1% of cells in the tissue. The vast majority of cells in the lymphoma is composed of various types of nonmalignant cells, including T cells, B cells, eosinophils, neutrophils, plasma cells, mast cells, and others [50]. The normal histological structure of the affected lymph nodes is disturbed. $CD4^+$ T helper and regulatory T cells are usually the most frequent cells in the Hodgkin lymphoma microenvironment, and HRS cells are typically directly surrounded by $CD4^+$ T cells [117, 118]. Many of the infiltrating cells are most likely beneficial for the tumor, and it seems that HRS cells actively attract such cells by secretion of a multitude of chemokines and cytokines [50, 118]. $CD4^+$ T cells may stimulate HRS cells through CD40 ligand–CD40 interaction and CD28–CD80 interaction, as HRS cells express CD40 and CD80 [50]. The regulatory T cells may play a role in the rescue of HRS cells from an attack by cytotoxic T cells and natural killer cells [117]. Other immunosuppressive factors secreted by the HRS cells include TGFβ, galectin-1, and purinergic signaling, and HRS cells also express the T cell-suppressive PD1 ligand 1 (CD274) [119–123]. Eosinophils on the one hand may also contribute to an immunosuppressive microenvironment by secretion of TGFβ and on the other hand may stimulate the CD30-positive HRS cells through interaction with CD30 ligand, which is expressed by the eosinophils [50].

7 Conclusions

B cell lymphoma pathogenesis is a multistep process involving various genetic and epigenetic lesions, viruses, and distorted microenvironmental and antigenic interactions. Often, the specific hit is less important than the affected pathway, because most gene products exert their function not directly but involve a network of relays and switches. In this way, different combinations of hits can lead to phenotypically similar lymphomas, as the resulting

deregulation of cellular processes is largely identical. Numerous hallmark chromosomal translocations in B cell lymphomas occur as errors of the IG gene remodeling processes of V gene recombination, SHM, and class switching. Most B cell lymphomas derive from GC B cells, which is explainable by the highly proliferative activity of these cells and the mutagenic potential of SHM and class switch recombination. A number of mutations in GC B cells appear to function by freezing GC B cells in their proliferative state and preventing their differentiation into resting post-GC memory or plasma cells. *BCL6* translocations and inactivating mutations in *BLIMP1*, encoding the master regulator of the plasma cell program, are prototypical examples for this [124].

The study of rearranged IG V genes and chromosomal translocations in B cell lymphomas has not only been of major relevance to define the cellular origin of these tumors and identify main transforming events, but they are also very important research tools. IG V gene rearrangements and chromosomal translocations have been used as genetic markers to study lymphoma cell dissemination and to identify and quantify minimal residual disease. They are also valuable to search for lymphoma precursor cells, to define the replication history of B cells, and to determine BCR specificity and hence potential antigen triggering of the lymphoma cells, to name several key applications.

Acknowledgments

Own work discussed in this review was supported by grants from the Deutsche Forschungsgemeinschaft (KU1315/8-1, KU1315/9-2, KU1315/10-1, TRR60), the Deutsche Krebshilfe, the Wilhelm Sander Foundation, the Hairy Cell Leukemia Foundation, the SASS Foundation for Medical research, and the Werner Jackstädt-Stiftung. We thank the other members of our group and Martin-Leo Hansmann for many stimulating discussions.

References

1. Tonegawa S (1983) Somatic generation of antibody diversity. Nature 302:575–581
2. van Gent DC, Ramsden DA, Gellert M (1996) The RAG1 and RAG2 proteins establish the 12/23 rule in V(D)J recombination. Cell 85:107–113
3. Medina KL, Singh H (2005) Genetic networks that regulate B lymphopoiesis. Curr Opin Hematol 12:203–209
4. Rajewsky K (1996) Clonal selection and learning in the antibody system. Nature 381:751–758
5. Corbett SJ, Tomlinson IM, Sonnhammer EL, Buck D, Winter G (1997) Sequence of the human immunoglobulin diversity (D) segment locus: a systematic analysis provides no evidence for the use of DIR segments, inverted D segments, "minor" D segments or D–D recombination. J Mol Biol 270:587–597
6. Ravetch JV, Siebenlist U, Korsmeyer S, Waldmann T, Leder P (1981) Structure of the human immunoglobulin mu locus: characterization of embryonic and rearranged J and D genes. Cell 27:583–591

7. Cook GP, Tomlinson IM (1995) The human immunoglobulin VH repertoire. Immunol Today 16:237–242
8. Alt FW, Rathbun G, Oltz E, Taccioli G, Shinkai Y (1992) Function and control of recombination-activating gene activity. Ann N Y Acad Sci 651:277–294
9. Zhang Z (2007) VH replacement in mice and humans. Trends Immunol 28:132–137
10. Tiegs SL, Russell DM, Nemazee D (1993) Receptor editing in self-reactive bone marrow B cells. J Exp Med 177:1009–1020
11. Hieter PA, Maizel JV Jr, Leder P (1982) Evolution of human immunoglobulin kappa J region genes. J Biol Chem 257:1516–1522
12. Schäble KF, Zachau HG (1993) The variable genes of the human immunoglobulin kappa locus. Biol Chem Hoppe Seyler 374:1001–1022
13. Kawasaki K, Minoshima S, Nakato E, Shibuya K, Shintani A, Schmeits JL, Wang J, Shimizu N (1997) One-megabase sequence analysis of the human immunoglobulin lambda gene locus. Genome Res 7:250–261
14. Vasicek TJ, Leder P (1990) Structure and expression of the human immunoglobulin lambda genes. J Exp Med 172:609–620
15. Bräuninger A, Goossens T, Rajewsky K, Küppers R (2001) Regulation of immunoglobulin light chain gene rearrangements during early B cell development in the human. Eur J Immunol 31:3631–3637
16. Nadel B, Tang A, Feeney AJ (1998) V (H) replacement is unlikely to contribute significantly to receptor editing due to an ineffectual embedded recombination signal sequence. Mol Immunol 35:227–232
17. Zhang Z, Zemlin M, Wang YH, Munfus D, Huye LE, Findley HW, Bridges SL, Roth DB, Burrows PD, Cooper MD (2003) Contribution of Vh gene replacement to the primary B cell repertoire. Immunity 19:21–31
18. MacLennan IC (1994) Germinal centers. Annu Rev Immunol 12:117–139
19. Allen CD, Ansel KM, Low C, Lesley R, Tamamura H, Fujii N, Cyster JG (2004) Germinal center dark and light zone organization is mediated by CXCR4 and CXCR5. Nat Immunol 5:943–952
20. Hauser AE, Junt T, Mempel TR, Sneddon MW, Kleinstein SH, Henrickson SE, von Andrian UH, Shlomchik MJ, Haberman AM (2007) Definition of germinal-center B cell migration in vivo reveals predominant intrazonal circulation patterns. Immunity 26:655–667
21. Schwickert TA, Lindquist RL, Shakhar G, Livshits G, Skokos D, Kosco-Vilbois MH, Dustin ML, Nussenzweig MC (2007) In vivo imaging of germinal centres reveals a dynamic open structure. Nature 446:83–87
22. Bannard O, Horton RM, Allen CD, An J, Nagasawa T, Cyster JG (2013) Germinal center centroblasts transition to a centrocyte phenotype according to a timed program and depend on the dark zone for effective selection. Immunity 39:912–924
23. Mayer CT, Gazumyan A, Kara EE, Gitlin AD, Golijanin J, Viant C, Pai J, Oliveira TY, Wang Q, Escolano A, Medina-Ramirez M, Sanders RW, Nussenzweig MC (2017) The microanatomic segregation of selection by apoptosis in the germinal center. Science 358:eaao2602
24. Küppers R, Zhao M, Hansmann ML, Rajewsky K (1993) Tracing B cell development in human germinal centres by molecular analysis of single cells picked from histological sections. EMBO J 12:4955–4967
25. Goossens T, Klein U, Küppers R (1998) Frequent occurrence of deletions and duplications during somatic hypermutation: implications for oncogene translocations and heavy chain disease. Proc Natl Acad Sci U S A 95:2463–2468
26. Pavri R, Nussenzweig MC (2011) AID targeting in antibody diversity. Adv Immunol 110:1–26
27. Neuberger MS (2008) Antibody diversification by somatic mutation: from Burnet onwards. Immunol Cell Biol 86:124–132
28. Di Noia JM, Neuberger MS (2007) Molecular mechanisms of antibody somatic hypermutation. Annu Rev Biochem 76:1–22
29. Pasqualucci L, Migliazza A, Fracchiolla N, William C, Neri A, Baldini L, Chaganti RSK, Klein U, Küppers R, Rajewsky K, Dalla-Favera R (1998) BCL-6 mutations in normal germinal center B cells: Evidence of somatic hypermutation acting outside Ig loci. Proc Natl Acad Sci U S A 95:11816–11821
30. Liu M, Duke JL, Richter DJ, Vinuesa CG, Goodnow CC, Kleinstein SH, Schatz DG (2008) Two levels of protection for the B cell genome during somatic hypermutation. Nature 451:841–845
31. Liu YJ, Joshua DE, Williams GT, Smith CA, Gordon J, MacLennan IC (1989) Mechanism of antigen-driven selection in germinal centres. Nature 342:929–931
32. Manis JP, Tian M, Alt FW (2002) Mechanism and control of class-switch recombination. Trends Immunol 23:31–39

33. Klein U, Rajewsky K, Küppers R (1998) Human immunoglobulin (Ig)M+IgD+ peripheral blood B cells expressing the CD27 cell surface antigen carry somatically mutated variable region genes: CD27 as a general marker for somatically mutated (memory) B cells. J Exp Med 188:1679–1689
34. Seifert M, Küppers R (2009) Molecular footprints of a germinal center derivation of human IgM+(IgD+)CD27+ B cells and the dynamics of memory B cell generation. J Exp Med 206:2659–2669
35. Budeus B, Schweigle de Reynoso S, Przekopowitz M, Hoffmann D, Seifert M, Küppers R (2015) Complexity of the human memory B-cell compartment is determined by the versatility of clonal diversification in germinal centers. Proc Natl Acad Sci U S A 112: E5281–E5289
36. Klein U, Dalla-Favera R (2008) Germinal centres: role in B-cell physiology and malignancy. Nat Rev Immunol 8:22–33
37. Seifert M, Küppers R (2016) Human memory B cells. Leukemia 30:2283–2292
38. McHeyzer-Williams M, Okitsu S, Wang N, McHeyzer-Williams L (2012) Molecular programming of B cell memory. Nat Rev Immunol 12:24–34
39. Manz RA, Hauser AE, Hiepe F, Radbruch A (2005) Maintenance of serum antibody levels. Annu Rev Immunol 23:367–386
40. Han JH, Akira S, Calame K, Beutler B, Selsing E, Imanishi-Kari T (2007) Class switch recombination and somatic hypermutation in early mouse B cells are mediated by B cell and Toll-like receptors. Immunity 27:64–75
41. Mond JJ, Lees A, Snapper CM (1995) T cell-independent antigens type 2. Annu Rev Immunol 13:655–692
42. Toellner KM, Jenkinson WE, Taylor DR, Khan M, Sze DM, Sansom DM, Vinuesa CG, MacLennan IC (2002) Low-level hypermutation in T cell-independent germinal centers compared with high mutation rates associated with T cell-dependent germinal centers. J Exp Med 195:383–389
43. Küppers R (2005) Mechanisms of B-cell lymphoma pathogenesis. Nat Rev Cancer 5:251–262
44. de Jong D (2005) Molecular pathogenesis of follicular lymphoma: a cross talk of genetic and immunologic factors. J Clin Oncol 23:6358–6363
45. Bende RJ, Smit LA, van Noesel CJ (2007) Molecular pathways in follicular lymphoma. Leukemia 21:18–29
46. Küppers R, Klein U, Hansmann M-L, Rajewsky K (1999) Cellular origin of human B-cell lymphomas. N Engl J Med 341:1520–1529
47. Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A, Boldrick JC, Sabet H, Tran T, Yu X, Powell JI, Yang L, Marti GE, Moore T, Hudson J Jr, Lu L, Lewis DB, Tibshirani R, Sherlock G, Chan WC, Greiner TC, Weisenburger DD, Armitage JO, Warnke R, Levy R, Wilson W, Grever MR, Byrd JC, Botstein D, Brown PO, Staudt LM (2000) Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature 403:503–511
48. Rosenwald A, Wright G, Chan WC, Connors JM, Campo E, Fisher RI, Gascoyne RD, Muller-Hermelink HK, Smeland EB, Giltnane JM, Hurt EM, Zhao H, Averett L, Yang L, Wilson WH, Jaffe ES, Simon R, Klausner RD, Powell J, Duffey PL, Longo DL, Greiner TC, Weisenburger DD, Sanger WG, Dave BJ, Lynch JC, Vose J, Armitage JO, Montserrat E, Lopez-Guillermo A, Grogan TM, Miller TP, LeBlanc M, Ott G, Kvaloy S, Delabie J, Holte H, Krajci P, Stokke T, Staudt LM (2002) The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma. N Engl J Med 346:1937–1947
49. Lossos IS, Alizadeh AA, Eisen MB, Chan WC, Brown PO, Botstein D, Staudt LM, Levy R (2000) Ongoing immunoglobulin somatic mutation in germinal center B cell-like but not in activated B cell-like diffuse large cell lymphomas. Proc Natl Acad Sci U S A 97:10209–10213
50. Küppers R (2009) The biology of Hodgkin's lymphoma. Nat Rev Cancer 9:15–27
51. Kanzler H, Küppers R, Hansmann ML, Rajewsky K (1996) Hodgkin and Reed-Sternberg cells in Hodgkin's disease represent the outgrowth of a dominant tumor clone derived from (crippled) germinal center B cells. J Exp Med 184:1495–1505
52. Küppers R, Rajewsky K, Zhao M, Simons G, Laumann R, Fischer R, Hansmann ML (1994) Hodgkin disease: Hodgkin and Reed-Sternberg cells picked from histological sections show clonal immunoglobulin gene rearrangements and appear to be derived from B cells at various stages of development. Proc Natl Acad Sci U S A 91:10962–10966
53. Chung SS, Kim E, Park JH, Chung YR, Lito P, Teruya-Feldstein J, Hu W, Beguelin W, Monette S, Duy C, Rampal R, Telis L, Patel M, Kim MK, Huberman K, Bouvier N, Berger MF, Melnick AM,

Rosen N, Tallman MS, Park CY, Abdel-Wahab O (2014) Hematopoietic stem cell origin of BRAFV600E mutations in hairy cell leukemia. Sci Transl Med 6:238ra271

54. Damm F, Mylonas E, Cosson A, Yoshida K, Della Valle V, Mouly E, Diop M, Scourzic L, Shiraishi Y, Chiba K, Tanaka H, Miyano S, Kikushige Y, Davi F, Lambert J, Gautheret D, Merle-Beral H, Sutton L, Dessen P, Solary E, Akashi K, Vainchenker W, Mercher T, Droin N, Ogawa S, Nguyen-Khac F, Bernard OA (2014) Acquired initiating mutations in early hematopoietic cells of CLL patients. Cancer Discov 4:1088–1101
55. Shaffer AL, Emre NC, Lamy L, Ngo VN, Wright G, Xiao W, Powell J, Dave S, Yu X, Zhao H, Zeng Y, Chen B, Epstein J, Staudt LM (2008) IRF4 addiction in multiple myeloma. Nature 454:226–231
56. Basso K, Liso A, Tiacci E, Benedetti R, Pulsoni A, Foa R, Di Raimondo F, Ambrosetti A, Califano A, Klein U, Dalla Favera R, Falini B (2004) Gene expression profiling of hairy cell leukemia reveals a phenotype related to memory B cells with altered expression of chemokine and adhesion receptors. J Exp Med 199:59–68
57. Seifert M, Sellmann L, Bloehdorn J, Wein F, Stilgenbauer S, Dürig J, Küppers R (2012) Cellular origin and pathophysiology of chronic lymphocytic leukemia. J Exp Med 209:2183–2198
58. Klein U, Tu Y, Stolovitzky GA, Mattioli M, Cattoretti G, Husson H, Freedman A, Inghirami G, Cro L, Baldini L, Neri A, Califano A, Dalla-Favera R (2001) Gene expression profiling of B cell chronic lymphocytic leukemia reveals a homogeneous phenotype related to memory B cells. J Exp Med 194:1625–1638
59. Küppers R, Dalla-Favera R (2001) Mechanisms of chromosomal translocations in B cell lymphomas. Oncogene 20:5580–5594
60. Pasqualucci L, Neumeister P, Goossens T, Nanjangud G, Chaganti RS, Kuppers R, Dalla-Favera R (2001) Hypermutation of multiple proto-oncogenes in B-cell diffuse large-cell lymphomas. Nature 412:341–346
61. Jäger U, Bocskor S, Le T, Mitterbauer G, Bolz I, Chott A, Kneba M, Mannhalter C, Nadel B (2000) Follicular lymphomas' BCL-2/IgH junctions contain templated nucleotide insertions: novel insights into the mechanism of t(14;18) translocation. Blood 95:3520–3529
62. Tsujimoto Y, Gorham J, Cossman J, Jaffe E, Croce CM (1985) The t(14;18) chromosome translocations involved in B-cell neoplasms result from mistakes in VDJ joining. Science 229:1390–1393
63. Dalla-Favera R, Martinotti S, Gallo RC, Erikson J, Croce CM (1983) Translocation and rearrangements of the c-myc oncogene locus in human undifferentiated B-cell lymphomas. Science 219:963–967
64. Taub R, Kirsch I, Morton C, Lenoir G, Swan D, Tronick S, Aaronson S, Leder P (1982) Translocation of the c-myc gene into the immunoglobulin heavy chain locus in human Burkitt lymphoma and murine plasmacytoma cells. Proc Natl Acad Sci U S A 79:7837–7841
65. Vaandrager JW, Schuuring E, Zwikstra E, de Boer CJ, Kleiverda KK, van Krieken JH, Kluin-Nelemans HC, van Ommen GJ, Raap AK, Kluin PM (1996) Direct visualization of dispersed 11q13 chromosomal translocations in mantle cell lymphoma by multicolor DNA fiber fluorescence in situ hybridization. Blood 88:1177–1182
66. Baron BW, Nucifora G, McCabe N, Espinosa R 3rd, Le Beau MM, McKeithan TW (1993) Identification of the gene associated with the recurring chromosomal translocations t(3;14)(q27;q32) and t(3;22)(q27;q11) in B-cell lymphomas. Proc Natl Acad Sci U S A 90:5262–5266
67. Wlodarska I, Nooyen P, Maes B, Martin-Subero JI, Siebert R, Pauwels P, De Wolf-Peeters C, Hagemeijer A (2003) Frequent occurrence of BCL6 rearrangements in nodular lymphocyte predominance Hodgkin lymphoma but not in classical Hodgkin lymphoma. Blood 101:706–710
68. Ye BH, Rao PH, Chaganti RS, Dalla-Favera R (1993) Cloning of bcl-6, the locus involved in chromosome translocations affecting band 3q27 in B-cell lymphoma. Cancer Res 53:2732–2735
69. Dierlamm J, Baens M, Wlodarska I, Stefanova-Ouzounova M, Hernandez JM, Hossfeld DK, De Wolf-Peeters C, Hagemeijer A, Van den Berghe H, Marynen P (1999) The apoptosis inhibitor gene API2 and a novel 18q gene, MLT, are recurrently rearranged in the t(11;18)(q21;q21) associated with mucosa-associated lymphoid tissue lymphomas. Blood 93:3601–3609
70. Morin RD, Mendez-Lago M, Mungall AJ, Goya R, Mungall KL, Corbett RD, Johnson NA, Severson TM, Chiu R, Field M, Jackman S, Krzywinski M, Scott DW, Trinh DL, Tamura-Wells J, Li S, Firme MR, Rogic S, Griffith M, Chan S, Yakovenko O, Meyer IM, Zhao EY, Smailus D, Moksa M,

Chittaranjan S, Rimsza L, Brooks-Wilson A, Spinelli JJ, Ben-Neriah S, Meissner B, Woolcock B, Boyle M, McDonald H, Tam A, Zhao Y, Delaney A, Zeng T, Tse K, Butterfield Y, Birol I, Holt R, Schein J, Horsman DE, Moore R, Jones SJ, Connors JM, Hirst M, Gascoyne RD, Marra MA (2011) Frequent mutation of histone-modifying genes in non-Hodgkin lymphoma. Nature 476:298–303

71. Pasqualucci L, Dominguez-Sola D, Chiarenza A, Fabbri G, Grunn A, Trifonov V, Kasper LH, Lerach S, Tang H, Ma J, Rossi D, Chadburn A, Murty VV, Mullighan CG, Gaidano G, Rabadan R, Brindle PK, Dalla-Favera R (2011) Inactivating mutations of acetyltransferase genes in B-cell lymphoma. Nature 471:189–195
72. Ammerpohl O, Haake A, Pellissery S, Giefing M, Richter J, Balint B, Kulis M, Le J, Bibikova M, Drexler HG, Seifert M, Shaknovic R, Korn B, Küppers R, Martin-Subero JI, Siebert R (2011) Array-based DNA methylation analysis in classical Hodgkin lymphoma reveals new insights into the mechanisms underlying silencing of B cell-specific genes. Leukemia 26:185–188
73. Martin-Subero JI, Kreuz M, Bibikova M, Bentink S, Ammerpohl O, Wickham-Garcia E, Rosolowski M, Richter J, Lopez-Serra L, Ballestar E, Berger H, Agirre X, Bernd HW, Calvanese V, Cogliatti SB, Drexler HG, Fan JB, Fraga MF, Hansmann ML, Hummel M, Klapper W, Korn B, Küppers R, Macleod RA, Möller P, Ott G, Pott C, Prosper F, Rosenwald A, Schwaenen C, Schubeler D, Seifert M, Sturzenhofecker B, Weber M, Wessendorf S, Loeffler M, Trümper L, Stein H, Spang R, Esteller M, Barker D, Hasenclever D, Siebert R (2009) New insights into the biology and origin of mature aggressive B-cell lymphomas by combined epigenomic, genomic, and transcriptional profiling. Blood 113:2488–2497
74. Cuneo A, Bigoni R, Rigolin GM, Roberti MG, Bardi A, Campioni D, Minotto C, Agostini P, Milani R, Bullrich F, Negrini M, Croce C, Castoldi G (1999) 13q14 deletion in non-Hodgkin's lymphoma: correlation with clinicopathologic features. Haematologica 84:589–593
75. Kuehl WM, Bergsagel PL (2002) Multiple myeloma: evolving genetic events and host interactions. Nat Rev Cancer 2:175–187
76. Liu Y, Hermanson M, Grander D, Merup M, Wu X, Heyman M, Rasool O, Juliusson G, Gahrton G, Detlofsson R, Nikiforova N, Buys C, Soderhall S, Yankovsky N, Zabarovsky E, Einhorn S (1995) 13q deletions in lymphoid malignancies. Blood 86:1911–1915
77. Klein U, Lia M, Crespo M, Siegel R, Shen Q, Mo T, Ambesi-Impiombato A, Califano A, Migliazza A, Bhagat G, Dalla-Favera R (2010) The DLEU2/miR-15a/16-1 cluster controls B cell proliferation and its deletion leads to chronic lymphocytic leukemia. Cancer Cell 17:28–40
78. Kretzmer H, Bernhart SH, Wang W, Haake A, Weniger MA, Bergmann AK, Betts MJ, Carrillo-de-Santa-Pau E, Doose G, Gutwein J, Richter J, Hovestadt V, Huang B, Rico D, Juhling F, Kolarova J, Lu Q, Otto C, Wagener R, Arnolds J, Burkhardt B, Claviez A, Drexler HG, Eberth S, Eils R, Flicek P, Haas S, Humme M, Karsch D, Kerstens HHD, Klapper W, Kreuz M, Lawerenz C, Lenzek D, Loeffler M, Lopez C, MacLeod RAF, Martens JHA, Kulis M, Martin-Subero JI, Möller P, Nage I, Picelli S, Vater I, Rohde M, Rosenstiel P, Rosolowski M, Russell RB, Schilhabel M, Schlesner M, Stadler PF, Szczepanowski M, Trümper L, Stunnenberg HG, Küppers R, Ammerpohl O, Lichter P, Siebert R, Hoffmann S, Radlwimmer B (2015) DNA methylome analysis in Burkitt and follicular lymphomas identifies differentially methylated regions linked to somatic mutation and transcriptional control. Nat Genet 47:1316–1325
79. Kulis M, Heath S, Bibikova M, Queiros AC, Navarro A, Clot G, Martinez-Trillos A, Castellano G, Brun-Heath I, Pinyol M, Barberan-Soler S, Papasaikas P, Jares P, Bea S, Rico D, Ecker S, Rubio M, Royo R, Ho V, Klotzle B, Hernandez L, Conde L, Lopez-Guerra M, Colomer D, Villamor N, Aymerich M, Rozman M, Bayes M, Gut M, Gelpi JL, Orozco M, Fan JB, Quesada V, Puente XS, Pisano DG, Valencia A, Lopez-Guillermo A, Gut I, Lopez-Otin C, Campo E, Martin-Subero JI (2012) Epigenomic analysis detects widespread gene-body DNA hypomethylation in chronic lymphocytic leukemia. Nat Genet 44:1236–1242
80. Oakes CC, Seifert M, Assenov Y, Gu L, Przekopowitz M, Ruppert AS, Wang Q, Imbusch CD, Serva A, Koser SD, Brocks D, Lipka DB, Bogatyrova O, Weichenhan D, Brors B, Rassenti L, Kipps TJ, Mertens D, Zapatka M, Lichter P, Döhner H, Küppers R, Zenz T, Stilgenbauer S, Byrd JC, Plass C (2016) DNA methylation dynamics during B cell maturation underlie a continuum of disease phenotypes in chronic lymphocytic leukemia. Nat Genet 48:253–264

81. Queiros AC, Villamor N, Clot G, Martinez-Trillos A, Kulis M, Navarro A, Penas EM, Jayne S, Majid A, Richter J, Bergmann AK, Kolarova J, Royo C, Russinol N, Castellano G, Pinyol M, Bea S, Salaverria I, Lopez-Guerra M, Colomer D, Aymerich M, Rozman M, Delgado J, Gine E, Gonzalez-Diaz M, Puente XS, Siebert R, Dyer MJ, Lopez-Otin C, Rozman C, Campo E, Lopez-Guillermo A, Martin-Subero JI (2015) A B-cell epigenetic signature defines three biologic subgroups of chronic lymphocytic leukemia with clinical impact. Leukemia 29:598–605
82. Landau DA, Clement K, Ziller MJ, Boyle P, Fan J, Gu H, Stevenson K, Sougnez C, Wang L, Li S, Kotliar D, Zhang W, Ghandi M, Garraway L, Fernandes SM, Livak KJ, Gabriel S, Gnirke A, Lander ES, Brown JR, Neuberg D, Kharchenko PV, Hacohen N, Getz G, Meissner A, Wu CJ (2014) Locally disordered methylation forms the basis of intratumor methylome variation in chronic lymphocytic leukemia. Cancer Cell 26:813–825
83. Queiros AC, Beekman R, Vilarrasa-Blasi R, Duran-Ferrer M, Clot G, Merkel A, Raineri E, Russinol N, Castellano G, Bea S, Navarro A, Kulis M, Verdaguer-Dot N, Jares P, Enjuanes A, Calasanz MJ, Bergmann A, Vater I, Salaverria I, van de Werken HJG, Wilson WH, Datta A, Flicek P, Royo R, Martens J, Gine E, Lopez-Guillermo A, Stunnenberg HG, Klapper W, Pott C, Heath S, Gut IG, Siebert R, Campo E, Martin-Subero JI (2016) Decoding the DNA methylome of mantle cell lymphoma in the light of the entire B cell lineage. Cancer Cell 30:806–821
84. Carbone A, Gloghini A (2008) KSHV/HHV8-associated lymphomas. Br J Haematol 140:13–24
85. Küppers R (2003) B cells under influence: transformation of B cells by Epstein-Barr virus. Nat Rev Immunol 3:801–812
86. Kilger E, Kieser A, Baumann M, Hammerschmidt W (1998) Epstein-Barr virus-mediated B-cell proliferation is dependent upon latent membrane protein 1, which simulates an activated CD40 receptor. EMBO J 17:1700–1709
87. Bechtel D, Kurth J, Unkel C, Küppers R (2005) Transformation of BCR-deficient germinal-center B cells by EBV supports a major role of the virus in the pathogenesis of Hodgkin and posttransplantation lymphomas. Blood 106:4345–4350
88. Mancao C, Hammerschmidt W (2007) Epstein-Barr virus latent membrane protein 2A is a B-cell receptor mimic and essential for B-cell survival. Blood 110:3715–3721
89. Bräuninger A, Schmitz R, Bechtel D, Renné C, Hansmann M-L, Küppers R (2006) Molecular biology of Hodgkin and Reed/Sternberg cells in Hodgkin's lymphoma. Int J Cancer 118:1853–1861
90. Quinn ER, Chan CH, Hadlock KG, Foung SK, Flint M, Levy S (2001) The B-cell receptor of a hepatitis C virus (HCV)-associated non-Hodgkin lymphoma binds the viral E2 envelope protein, implicating HCV in lymphomagenesis. Blood 98:3745–3749
91. Machida K, Cheng KT, Sung VM, Lee KJ, Levine AM, Lai MM (2004) Hepatitis C virus infection activates the immunologic (type II) isoform of nitric oxide synthase and thereby enhances DNA damage and mutations of cellular genes. J Virol 78:8835–8843
92. Machida K, Cheng KT, Sung VM, Shimodaira S, Lindsay KL, Levine AM, Lai MY, Lai MM (2004) Hepatitis C virus induces a mutator phenotype: enhanced mutations of immunoglobulin and protooncogenes. Proc Natl Acad Sci U S A 101:4262–4267
93. Tucci FA, Broering R, Johansson P, Schlaak JF, Küppers R (2013) B cells in chronically hepatitis C virus-infected individuals lack a virus-induced mutation signature in the TP53, CTNNB1, and BCL6 genes. J Virol 87:2956–2962
94. Lacroix A, Collot-Teixeira S, Mardivirin L, Jaccard A, Petit B, Piguet C, Sturtz F, Preux PM, Bordessoule D, Ranger-Rogez S (2010) Involvement of human herpesvirus-6 variant B in classic Hodgkin's lymphoma via DR7 oncoprotein. Clin Cancer Res 16:4711–4721
95. Maggio E, Benharroch D, Gopas J, Dittmer U, Hansmann ML, Küppers R (2007) Absence of measles virus genome and transcripts in Hodgkin-Reed/Sternberg cells of a cohort of Hodgkin lymphoma patients. Int J Cancer 121:448–453
96. Johnson PW, Watt SM, Betts DR, Davies D, Jordan S, Norton AJ, Lister TA (1993) Isolated follicular lymphoma cells are resistant to apoptosis and can be grown in vitro in the CD40/stromal cell system. Blood 82:1848–1857
97. Umetsu DT, Esserman L, Donlon TA, DeKruyff RH, Levy R (1990) Induction of proliferation of human follicular (B type) lymphoma cells by cognate interaction with CD4+ T cell clones. J Immunol 144:2550–2557

98. Zhu D, McCarthy H, Ottensmeier CH, Johnson P, Hamblin TJ, Stevenson FK (2002) Acquisition of potential N-glycosylation sites in the immunoglobulin variable region by somatic mutation is a distinctive feature of follicular lymphoma. Blood 99:2562–2568
99. Coelho V, Krysov S, Ghaemmaghami AM, Emara M, Potter KN, Johnson P, Packham G, Martinez-Pomares L, Stevenson FK (2010) Glycosylation of surface Ig creates a functional bridge between human follicular lymphoma and microenvironmental lectins. Proc Natl Acad Sci U S A 107:18587–18592
100. Küppers R, Stevenson FK (2018) Critical influences on the pathogenesis of follicular lymphoma. Blood 131:2297–2306
101. Schmid C, Isaacson PG (1994) Proliferation centres in B-cell malignant lymphoma, lymphocytic (B-CLL): an immunophenotypic study. Histopathol 24:445–451
102. Ghia P, Strola G, Granziero L, Geuna M, Guida G, Sallusto F, Ruffing N, Montagna L, Piccoli P, Chilosi M, Caligaris-Cappio F (2002) Chronic lymphocytic leukemia B cells are endowed with the capacity to attract CD4+, CD40L+ T cells by producing CCL22. Eur J Immunol 32:1403–1413
103. Buske C, Gogowski G, Schreiber K, Rave-Frank M, Hiddemann W, Wormann B (1997) Stimulation of B-chronic lymphocytic leukemia cells by murine fibroblasts, IL-4, anti-CD40 antibodies, and the soluble CD40 ligand. Exp Hematol 25:329–337
104. Chu CC, Catera R, Zhang L, Didier S, Agagnina BM, Damle RN, Kaufman MS, Kolitz JE, Allen SL, Rai KR, Chiorazzi N (2010) Many chronic lymphocytic leukemia antibodies recognize apoptotic cells with exposed nonmuscle myosin heavy chain IIA: implications for patient outcome and cell of origin. Blood 115:3907–3915
105. Herve M, Xu K, Ng YS, Wardemann H, Albesiano E, Messmer BT, Chiorazzi N, Meffre E (2005) Unmutated and mutated chronic lymphocytic leukemias derive from self-reactive B cell precursors despite expressing different antibody reactivity. J Clin Invest 115:1636–1643
106. Dühren-von Minden M, Ubelhart R, Schneider D, Wossning T, Bach MP, Buchner M, Hofmann D, Surova E, Follo M, Kohler F, Wardemann H, Zirlik K, Veelken H, Jumaa H (2012) Chronic lymphocytic leukaemia is driven by antigen-independent cell-autonomous signalling. Nature 489:309–312
107. Bende RJ, Aarts WM, Riedl RG, de Jong D, Pals ST, van Noesel CJ (2005) Among B cell non-Hodgkin's lymphomas, MALT lymphomas express a unique antibody repertoire with frequent rheumatoid factor reactivity. J Exp Med 201:1229–1241
108. Hussel T, Isaacson PG, Crabtree JE, Spencer J (1996) Helicobacter pylori-specific tumour-infiltrating T cells provide contact dependent help for the growth of malignant B cells in low-grade gastric lymphoma of mucosa-associated lymphoid tissue. J Pathol 178:122–127
109. Wotherspoon AC, Doglioni C, Diss TC, Pan L, Moschini A, de Boni M, Isaacson PG (1993) Regression of primary low-grade B-cell gastric lymphoma of mucosa-associated lymphoid tissue after eradication of Helicobacter pylori. Lancet 342:575–577
110. Hermine O, Lefrere F, Bronowicki JP, Mariette X, Jondeau K, Eclache-Saudreau V, Delmas B, Valensi F, Cacoub P, Brechot C, Varet B, Troussard X (2002) Regression of splenic lymphoma with villous lymphocytes after treatment of hepatitis C virus infection. N Engl J Med 347:89–94
111. Young RM, Wu T, Schmitz R, Dawood M, Xiao W, Phelan JD, Xu W, Menard L, Meffre E, Chan WC, Jaffe ES, Gascoyne RD, Campo E, Rosenwald A, Ott G, Delabie J, Rimsza LM, Staudt LM (2015) Survival of human lymphoma cells requires B-cell receptor engagement by self-antigens. Proc Natl Acad Sci U S A 112:13447–13454
112. Dagklis A, Ponzoni M, Govi S, Cangi MG, Pasini E, Charlotte F, Vino A, Doglioni C, Davi F, Lossos IS, Ntountas I, Papadaki T, Dolcetti R, Ferreri AJ, Stamatopoulos K, Ghia P (2012) Immunoglobulin gene repertoire in ocular adnexal lymphomas: hints on the nature of the antigenic stimulation. Leukemia 26:814–821
113. Mahendra A, Gangadharan B, Andre S, Boudjoghra M, Davi F, Lecerf M, Planchais C, Kaveri SV, Lacroix-Desmazes S, Dimitrov JD (2014) Cryptic polyreactivity of IgG expressed by splenic marginal zone B-cell lymphoma. Mol Immunol 60:54–61
114. Chiorazzi N, Hatzi K, Albesiano E (2005) B-cell chronic lymphocytic leukemia, a clonal disease of B lymphocytes with receptors that vary in specificity for (auto)antigens. Ann N Y Acad Sci 1062:1–12
115. Que X, Widhopf GF 2nd, Amir S, Hartvigsen K, Hansen LF, Woelkers D, Tsimikas S, Binder CJ, Kipps TJ, Witztum JL (2013) IGHV1-69-encoded antibodies expressed in chronic lymphocytic leukemia

react with malondialdehyde-acetaldehyde adduct, an immunodominant oxidation-specific epitope. PLoS One 8:e65203

116. Cha SC, Qin H, Kannan S, Rawal S, Watkins LS, Baio FE, Wu W, Ong J, Wei J, Kwak B, Kim S, Popescu MS, Paick DS, Kim K, Luong A, Davis RE, Schroeder HW Jr, Kwak LW, Neelapu SS (2013) Nonstereotyped lymphoma B cell receptors recognize vimentin as a shared autoantigen. J Immunol 190:4887–4898
117. Marshall NA, Christie LE, Munro LR, Culligan DJ, Johnston PW, Barker RN, Vickers MA (2004) Immunosuppressive regulatory T cells are abundant in the reactive lymphocytes of Hodgkin lymphoma. Blood 103:1755–1762
118. Wein F, Küppers R (2016) The role of T cells in the microenvironment of Hodgkin lymphoma. J Leukoc Biol 99:45–50
119. Chemnitz JM, Eggle D, Driesen J, Classen S, Riley JL, Debey-Pascher S, Beyer M, Popov A, Zander T, Schultze JL (2007) RNA fingerprints provide direct evidence for the inhibitory role of TGFbeta and PD-1 on CD4+ T cells in Hodgkin lymphoma. Blood 110:3226–3233
120. Gandhi MK, Moll G, Smith C, Dua U, Lambley E, Ramuz O, Gill D, Marlton P, Seymour JF, Khanna R (2007) Galectin-1 mediated suppression of Epstein-Barr virus specific T-cell immunity in classic Hodgkin lymphoma. Blood 110:1326–1329
121. Juszczynski P, Ouyang J, Monti S, Rodig SJ, Takeyama K, Abramson J, Chen W, Kutok JL, Rabinovich GA, Shipp MA (2007) The AP1-dependent secretion of galectin-1 by Reed Sternberg cells fosters immune privilege in classical Hodgkin lymphoma. Proc Natl Acad Sci U S A 104:13134–13139
122. Yamamoto R, Nishikori M, Kitawaki T, Sakai T, Hishizawa M, Tashima M, Kondo T, Ohmori K, Kurata M, Hayashi T, Uchiyama T (2008) PD-1-PD-1 ligand interaction contributes to immunosuppressive microenvironment of Hodgkin lymphoma. Blood 111:3220–3224
123. Wein F, Weniger MA, Hoing B, Arnolds J, Hüttmann A, Hansmann ML, Hartmann S, Küppers R (2017) Complex immune evasion strategies in classical Hodgkin lymphoma. Cancer Immunol Res 5:1122–1132
124. Shaffer AL, Young RM, Staudt LM (2012) Pathogenesis of human B cell lymphomas. Annu Rev Immunol 30:565–610
125. Camacho E, Hernandez L, Hernandez S, Tort F, Bellosillo B, Bea S, Bosch F, Montserrat E, Cardesa A, Fernandez PL, Campo E (2002) ATM gene inactivation in mantle cell lymphoma mainly occurs by truncating mutations and missense mutations involving the phosphatidylinositol-3 kinase domain and is associated with increasing numbers of chromosomal imbalances. Blood 99:238–244
126. Schaffner C, Idler I, Stilgenbauer S, Dohner H, Lichter P (2000) Mantle cell lymphoma is characterized by inactivation of the ATM gene. Proc Natl Acad Sci U S A 97:2773–2778
127. Jares P, Colomer D, Campo E (2007) Genetic and molecular pathogenesis of mantle cell lymphoma: perspectives for new targeted therapeutics. Nat Rev Cancer 7:750–762
128. Chanudet E, Huang Y, Ichimura K, Dong G, Hamoudi RA, Radford J, Wotherspoon AC, Isaacson PG, Ferry J, Du MQ (2010) A20 is targeted by promoter methylation, deletion and inactivating mutation in MALT lymphoma. Leukemia 24:483–487
129. Wu C, de Miranda NF, Chen L, Wasik AM, Mansouri L, Jurczak W, Galazka K, Dlugosz-Danecka M, Machaczka M, Zhang H, Peng R, Morin RD, Rosenquist R, Sander B, Pan-Hammarstrom Q (2016) Genetic heterogeneity in primary and relapsed mantle cell lymphomas: Impact of recurrent CARD11 mutations. Oncotarget 7:38180–38190
130. Kridel R, Meissner B, Rogic S, Boyle M, Telenius A, Woolcock B, Gunawardana J, Jenkins C, Cochrane C, Ben-Neriah S, Tan K, Morin RD, Opat S, Sehn LH, Connors JM, Marra MA, Weng AP, Steidl C, Gascoyne RD (2012) Whole transcriptome sequencing reveals recurrent NOTCH1 mutations in mantle cell lymphoma. Blood 119:1963–1971
131. Bea S, Valdes-Mas R, Navarro A, Salaverria I, Martin-Garcia D, Jares P, Gine E, Pinyol M, Royo C, Nadeu F, Conde L, Juan M, Clot G, Vizan P, Di Croce L, Puente DA, Lopez-Guerra M, Moros A, Roue G, Aymerich M, Villamor N, Colomo L, Martinez A, Valera A, Martin-Subero JI, Amador V, Hernandez L, Rozman M, Enjuanes A, Forcada P, Muntanola A, Hartmann EM, Calasanz MJ, Rosenwald A, Ott G, Hernandez-Rivas JM, Klapper W, Siebert R, Wiestner A, Wilson WH, Colomer D, Lopez-Guillermo A, Lopez-Otin C, Puente XS, Campo E (2013) Landscape of somatic mutations and clonal evolution in mantle cell lymphoma. Proc Natl Acad Sci U S A 110:18250–18255

132. Schaffner C, Stilgenbauer S, Rappold GA, Döhner H, Lichter P (1999) Somatic ATM mutations indicate a pathogenic role of ATM in B-cell chronic lymphocytic leukemia. Blood 94:748–753
133. Stankovic T, Weber P, Stewart G, Bedenham T, Murray J, Byrd PJ, Moss PA, Taylor AM (1999) Inactivation of ataxia telangiectasia mutated gene in B-cell chronic lymphocytic leukaemia. Lancet 353:26–29
134. Gaidano G, Ballerini P, Gong JZ, Inghirami G, Neri A, Newcomb EW, Magrath IT, Knowles DM, Dalla-Favera R (1991) p53 mutations in human lymphoid malignancies: association with Burkitt lymphoma and chronic lymphocytic leukemia. Proc Natl Acad Sci U S A 88:5413–5417
135. Fabbri G, Rasi S, Rossi D, Trifonov V, Khiabanian H, Ma J, Grunn A, Fangazio M, Capello D, Monti S, Cresta S, Gargiulo E, Forconi F, Guarini A, Arcaini L, Paulli M, Laurenti L, Larocca LM, Marasca R, Gattei V, Oscier D, Bertoni F, Mullighan CG, Foa R, Pasqualucci L, Rabadan R, Dalla-Favera R, Gaidano G (2011) Analysis of the chronic lymphocytic leukemia coding genome: role of NOTCH1 mutational activation. J Exp Med 208:1389–1401
136. Puente XS, Pinyol M, Quesada V, Conde L, Ordonez GR, Villamor N, Escaramis G, Jares P, Bea S, Gonzalez-Diaz M, Bassaganyas L, Baumann T, Juan M, Lopez-Guerra M, Colomer D, Tubio JM, Lopez C, Navarro A, Tornador C, Aymerich M, Rozman M, Hernandez JM, Puente DA, Freije JM, Velasco G, Gutierrez-Fernandez A, Costa D, Carrio A, Guijarro S, Enjuanes A, Hernandez L, Yague J, Nicolas P, Romeo-Casabona CM, Himmelbauer H, Castillo E, Dohm JC, de Sanjose S, Piris MA, de Alava E, San Miguel J, Royo R, Gelpi JL, Torrents D, Orozco M, Pisano DG, Valencia A, Guigo R, Bayes M, Heath S, Gut M, Klatt P, Marshall J, Raine K, Stebbings LA, Futreal PA, Stratton MR, Campbell PJ, Gut I, Lopez-Guillermo A, Estivill X, Montserrat E, Lopez-Otin C, Campo E (2011) Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia. Nature 475:101–105
137. Quesada V, Conde L, Villamor N, Ordonez GR, Jares P, Bassaganyas L, Ramsay AJ, Bea S, Pinyol M, Martinez-Trillos A, Lopez-Guerra M, Colomer D, Navarro A, Baumann T, Aymerich M, Rozman M, Delgado J, Gine E, Hernandez JM, Gonzalez-Diaz M, Puente DA, Velasco G, Freije JM, Tubio JM, Royo R, Gelpi JL, Orozco M, Pisano DG, Zamora J, Vazquez M, Valencia A, Himmelbauer H, Bayes M, Heath S, Gut M, Gut I, Estivill X, Lopez-Guillermo A, Puente XS, Campo E, Lopez-Otin C (2011) Exome sequencing identifies recurrent mutations of the splicing factor SF3B1 gene in chronic lymphocytic leukemia. Nat Genet 44:47–52
138. Rossi D, Bruscaggin A, Spina V, Rasi S, Khiabanian H, Messina M, Fangazio M, Vaisitti T, Monti S, Chiaretti S, Guarini A, Del Giudice I, Cerri M, Cresta S, Deambrogi C, Gargiulo E, Gattei V, Forconi F, Bertoni F, Deaglio S, Rabadan R, Pasqualucci L, Foa R, Dalla-Favera R, Gaidano G (2011) Mutations of the SF3B1 splicing factor in chronic lymphocytic leukemia: association with progression and fludarabine-refractoriness. Blood 118:6904–6908
139. Green MR, Gentles AJ, Nair RV, Irish JM, Kihira S, Liu CL, Kela I, Hopmans ES, Myklebust JH, Ji H, Plevritis SK, Levy R, Alizadeh AA (2013) Hierarchy in somatic mutations arising during genomic evolution and progression of follicular lymphoma. Blood 121:1604–1611
140. Launay E, Pangault C, Bertrand P, Jardin F, Lamy T, Tilly H, Tarte K, Bastard C, Fest T (2012) High rate of TNFRSF14 gene alterations related to 1p36 region in de novo follicular lymphoma and impact on prognosis. Leukemia 26:559–562
141. Morin RD, Johnson NA, Severson TM, Mungall AJ, An J, Goya R, Paul JE, Boyle M, Woolcock BW, Kuchenbauer F, Yap D, Humphries RK, Griffith OL, Shah S, Zhu H, Kimbara M, Shashkin P, Charlot JF, Tcherpakov M, Corbett R, Tam A, Varhol R, Smailus D, Moksa M, Zhao Y, Delaney A, Qian H, Birol I, Schein J, Moore R, Holt R, Horsman DE, Connors JM, Jones S, Aparicio S, Hirst M, Gascoyne RD, Marra MA (2010) Somatic mutations altering EZH2 (Tyr641) in follicular and diffuse large B-cell lymphomas of germinal-center origin. Nat Genet 42:181–185
142. Li H, Kaminski MS, Li Y, Yildiz M, Ouillette P, Jones S, Fox H, Jacobi K, Saiya-Cork K, Bixby D, Lebovic D, Roulston D, Shedden K, Sabel M, Marentette L, Cimmino V, Chang AE, Malek SN (2014) Mutations in linker histone genes HIST1H1 B, C, D, and E; OCT2 (POU2F2); IRF8; and ARID1A underlying the pathogenesis of follicular lymphoma. Blood 123:1487–1498
143. Ying CY, Dominguez-Sola D, Fabi M, Lorenz IC, Hussein S, Bansal M, Califano A, Pasqualucci L, Basso K, Dalla-Favera R (2013) MEF2B mutations lead to

deregulated expression of the oncogene BCL6 in diffuse large B cell lymphoma. Nat Immunol 14:1084–1092

144. Yildiz M, Li H, Bernard D, Amin NA, Ouillette P, Jones S, Saiya-Cork K, Parkin B, Jacobi K, Shedden K, Wang S, Chang AE, Kaminski MS, Malek SN (2015) Activating STAT6 mutations in follicular lymphoma. Blood 125:668–679
145. Krysiak K, Gomez F, White BS, Matlock M, Miller CA, Trani L, Fronick CC, Fulton RS, Kreisel F, Cashen AF, Carson KR, Berrien-Elliott MM, Bartlett NL, Griffith M, Griffith OL, Fehniger TA (2017) Recurrent somatic mutations affecting B-cell receptor signaling pathway genes in follicular lymphoma. Blood 129:473–483
146. Okosun J, Wolfson RL, Wang J, Araf S, Wilkins L, Castellano BM, Escudero-Ibarz L, Al Seraihi AF, Richter J, Bernhart SH, Efeyan A, Iqbal S, Matthews J, Clear A, Guerra-Assuncao JA, Bodor C, Quentmeier H, Mansbridge C, Johnson P, Davies A, Strefford JC, Packham G, Barrans S, Jack A, Du MQ, Calaminici M, Lister TA, Auer R, Montoto S, Gribben JG, Siebert R, Chelala C, Zoncu R, Sabatini DM, Fitzgibbon J (2016) Recurrent mTORC1-activating RRAGC mutations in follicular lymphoma. Nat Genet 48:183–188
147. Weiss LM, Warnke RA, Sklar J, Cleary ML (1987) Molecular analysis of the t(14;18) chromosomal translocation in malignant lymphomas. N Engl J Med 317:1185–1189
148. Ladanyi M, Offit K, Jhanwar SC, Filippa DA, Chaganti RS (1991) MYC rearrangement and translocations involving band 8q24 in diffuse large cell lymphomas. Blood 77:1057–1063
149. Gronbaek K, Straten PT, Ralfkiaer E, Ahrenkiel V, Andersen MK, Hansen NE, Zeuthen J, Hou-Jensen K, Guldberg P (1998) Somatic Fas mutations in non-Hodgkin's lymphoma: association with extranodal disease and autoimmunity. Blood 92:3018–3024
150. Gronbaek K, Worm J, Ralfkiaer E, Ahrenkiel V, Hokland P, Guldberg P (2002) ATM mutations are associated with inactivation of the ARF-TP53 tumor suppressor pathway in diffuse large B-cell lymphoma. Blood 100:1430–1437
151. Koduru PR, Raju K, Vadmal V, Menezes G, Shah S, Susin M, Kolitz J, Broome JD (1997) Correlation between mutation in P53, p53 expression, cytogenetics, histologic type, and survival in patients with B-cell non-Hodgkin's lymphoma. Blood 90:4078–4091
152. Moller MB, Ino Y, Gerdes AM, Skjodt K, Louis DN, Pedersen NT (1999) Aberrations of the p53 pathway components p53, MDM2 and CDKN2A appear independent in diffuse large B cell lymphoma. Leukemia 13:453–459
153. Mottok A, Renne C, Seifert M, Oppermann E, Bechstein W, Hansmann ML, Küppers R, Bräuninger A (2009) Inactivating SOCS1 mutations are caused by aberrant somatic hypermutation and restricted to a subset of B-cell lymphoma entities. Blood 114:4503–4506
154. Pasqualucci L, Compagno M, Houldsworth J, Monti S, Grunn A, Nandula SV, Aster JC, Murty VV, Shipp MA, Dalla-Favera R (2006) Inactivation of the PRDM1/BLIMP1 gene in diffuse large B cell lymphoma. J Exp Med 203:311–317
155. Jardin F, Jais JP, Molina TJ, Parmentier F, Picquenot JM, Ruminy P, Tilly H, Bastard C, Salles GA, Feugier P, Thieblemont C, Gisselbrecht C, de Reynies A, Coiffier B, Haioun C, Leroy K (2010) Diffuse large B-cell lymphomas with CDKN2A deletion have a distinct gene expression signature and a poor prognosis under R-CHOP treatment: a GELA study. Blood 116:1092–1104
156. Honma K, Tsuzuki S, Nakagawa M, Tagawa H, Nakamura S, Morishima Y, Seto M (2009) TNFAIP3/A20 functions as a novel tumor suppressor gene in several subtypes of non-Hodgkin lymphomas. Blood 114:2467–2475
157. Steidl C, Shah SP, Woolcock BW, Rui L, Kawahara M, Farinha P, Johnson NA, Zhao Y, Telenius A, Neriah SB, McPherson A, Meissner B, Okoye UC, Diepstra A, van den Berg A, Sun M, Leung G, Jones SJ, Connors JM, Huntsman DG, Savage KJ, Rimsza LM, Horsman DE, Staudt LM, Steidl U, Marra MA, Gascoyne RD (2011) MHC class II transactivator CIITA is a recurrent gene fusion partner in lymphoid cancers. Nature 471:377–381
158. Melzner I, Bucur AJ, Bruderlein S, Dorsch K, Hasel C, Barth TF, Leithauser F, Möller P (2005) Biallelic mutation of SOCS-1 impairs JAK2 degradation and sustains phospho-JAK2 action in the MedB-1 mediastinal lymphoma line. Blood 105:2535–2542
159. Ritz O, Guiter C, Castellano F, Dorsch K, Melzner J, Jais JP, Dubois G, Gaulard P, Moller P, Leroy K (2009) Recurrent mutations of the STAT6 DNA binding domain in primary mediastinal B-cell lymphoma. Blood 114:1236–1242

160. Schmitz R, Hansmann ML, Bohle V, Martin-Subero JI, Hartmann S, Mechtersheimer G, Klapper W, Vater I, Giefing M, Gesk S, Stanelle J, Siebert R, Küppers R (2009) TNFAIP3 (A20) is a tumor suppressor gene in Hodgkin lymphoma and primary mediastinal B cell lymphoma. J Exp Med 206:981–989

161. Mansouri L, Noerenberg D, Young E, Mylonas E, Abdulla M, Frick M, Asmar F, Ljungstrom V, Schneider M, Yoshida K, Skaftason A, Pandzic T, Gonzalez B, Tasidou A, Waldhueter N, Rivas-Delgado A, Angelopoulou M, Ziepert M, Arends CM, Couronne L, Lenze D, Baldus CD, Bastard C, Okosun J, Fitzgibbon J, Dorken B, Drexler HG, Roos-Weil D, Schmitt CA, Munch-Petersen HD, Zenz T, Hansmann ML, Strefford JC, Enblad G, Bernard OA, Ralfkiaer E, Erlanson M, Korkolopoulou P, Hultdin M, Papadaki T, Gronbaek K, Lopez-Guillermo A, Ogawa S, Küppers R, Stamatopoulos K, Stavroyianni N, Kanellis G, Rosenwald A, Campo E, Amini RM, Ott G, Vassilakopoulos TP, Hummel M, Rosenquist R, Damm F (2016) Frequent NFKBIE deletions are associated with poor outcome in primary mediastinal B-cell lymphoma. Blood 128:2666–2670

162. Mottok A, Woolcock B, Chan FC, Tong KM, Chong L, Farinha P, Telenius A, Chavez E, Ramchandani S, Drake M, Boyle M, Ben-Neriah S, Scott DW, Rimsza LM, Siebert R, Gascoyne RD, Steidl C (2015) Genomic alterations in CIITA are frequent in primary mediastinal large B cell lymphoma and are associated with diminished MHC class II expression. Cell Rep 13:1418–1431

163. Gunawardana J, Chan FC, Telenius A, Woolcock B, Kridel R, Tan KL, Ben-Neriah S, Mottok A, Lim RS, Boyle M, Rogic S, Rimsza LM, Guiter C, Leroy K, Gaulard P, Haioun C, Marra MA, Savage KJ, Connors JM, Shah SP, Gascoyne RD, Steidl C (2014) Recurrent somatic mutations of PTPN1 in primary mediastinal B cell lymphoma and Hodgkin lymphoma. Nat Genet 46:329–335

164. Rossi D, Cerri M, Capello D, Deambrogi C, Berra E, Franceschetti S, Alabiso O, Gloghini A, Paulli M, Carbone A, Pileri SA, Pasqualucci L, Gaidano G (2005) Aberrant somatic hypermutation in primary mediastinal large B-cell lymphoma. Leukemia 19:2363–2366

165. Jardin F, Pujals A, Pelletier L, Bohers E, Camus V, Mareschal S, Dubois S, Sola B, Ochmann M, Lemonnier F, Viailly PJ, Bertrand P, Maingonnat C, Traverse-Glehen-A, Gaulard P, Damotte D, Delarue R, Haioun C, Argueta C, Landesman Y, Salles G, Jais JP, Figeac M, Copie-Bergman C, Molina TJ, Picquenot JM, Cornic M, Fest T, Milpied N, Lemasle E, Stamatoullas A, Moeller P, Dyer MJ, Sundstrom C, Bastard C, Tilly H, Leroy K (2016) Recurrent mutations of the exportin 1 gene (XPO1) and their impact on selective inhibitor of nuclear export compounds sensitivity in primary mediastinal B-cell lymphoma. Am J Hematol 91:923–930

166. Vigano E, Gunawardana J, Mottok A, Van Tol T, Mak K, Chan FC, Chong L, Chavez E, Woolcock B, Takata K, Twa D, Shulha HP, Telenius A, Kutovaya O, Hung SS, Healy S, Ben-Neriah S, Leroy K, Gaulard P, Diepstra A, Kridel R, Savage KJ, Rimsza L, Gascoyne R, Steidl C (2018) Somatic IL4R mutations in primary mediastinal large B-cell lymphoma lead to constitutive JAK-STAT signaling activation. Blood 131:2036–2046

167. Cinti C, Leoncini L, Nyongo A, Ferrari F, Lazzi S, Bellan C, Vatti R, Zamparelli A, Cevenini G, Tosi GM, Claudio PP, Maraldi NM, Tosi P, Giordano A (2000) Genetic alterations of the retinoblastoma-related gene RB2/p130 identify different pathogenetic mechanisms in and among Burkitt's lymphoma subtypes. Am J Pathol 156:751–760

168. Richter J, Schlesner M, Hoffmann S, Kreuz M, Leich E, Burkhardt B, Rosolowski M, Ammerpohl O, Wagener R, Bernhart SH, Lenze D, Szczepanowski M, Paulsen M, Lipinski S, Russell RB, Adam-Klages S, Apic G, Claviez A, Hasenclever D, Hovestadt V, Hornig N, Korbel JO, Kube D, Langenberger D, Lawerenz C, Lisfeld J, Meyer K, Picelli S, Pischimarov J, Radlwimmer B, Rausch T, Rohde M, Schilhabel M, Scholtysik R, Spang R, Trautmann H, Zenz T, Borkhardt A, Drexler HG, Moller P, MacLeod RA, Pott C, Schreiber S, Trümper L, Loeffler M, Stadler PF, Lichter P, Eils R, Küppers R, Hummel M, Klapper W, Rosenstiel P, Rosenwald A, Brors B, Siebert R, Project IM-S (2012) Recurrent mutation of the ID3 gene in Burkitt lymphoma identified by integrated genome, exome and transcriptome sequencing. Nat Genet 44:1316–1320

169. Schmitz R, Young RM, Ceribelli M, Jhavar S, Xiao W, Zhang M, Wright G, Shaffer AL, Hodson DJ, Buras E, Liu X, Powell J, Yang Y, Xu W, Zhao H, Kohlhammer H, Rosenwald A, Kluin P, Muller-Hermelink

HK, Ott G, Gascoyne RD, Connors JM, Rimsza LM, Campo E, Jaffe ES, Delabie J, Smeland EB, Ogwang MD, Reynolds SJ, Fisher RI, Braziel RM, Tubbs RR, Cook JR, Weisenburger DD, Chan WC, Pittaluga S, Wilson W, Waldmann TA, Rowe M, Mbulaiteye SM, Rickinson AB, Staudt LM (2012) Burkitt lymphoma pathogenesis and therapeutic targets from structural and functional genomics. Nature 490:116–120

170. Love C, Sun Z, Jima D, Li G, Zhang J, Miles R, Richards KL, Dunphy CH, Choi WW, Srivastava G, Lugar PL, Rizzieri DA, Lagoo AS, Bernal-Mizrachi L, Mann KP, Flowers CR, Naresh KN, Evens AM, Chadburn A, Gordon LI, Czader MB, Gill JI, Hsi ED, Greenough A, Moffitt AB, McKinney M, Banerjee A, Grubor V, Levy S, Dunson DB, Dave SS (2012) The genetic landscape of mutations in Burkitt lymphoma. Nat Genet 44:1321–1325
171. Menter T, Juskevicius D, Alikian M, Steiger J, Dirnhofer S, Tzankov A, Naresh KN (2017) Mutational landscape of B-cell post-transplant lymphoproliferative disorders. Br J Haematol 178:48–56
172. Cabannes E, Khan G, Aillet F, Jarrett RF, Hay RT (1999) Mutations in the IkBa gene in Hodgkin's disease suggest a tumour suppressor role for IkappaBalpha. Oncogene 18:3063–3070
173. Krappmann D, Emmerich F, Kordes U, Scharschmidt E, Dörken B, Scheidereit C (1999) Molecular mechanisms of constitutive NF-kappaB/Rel activation in Hodgkin/Reed-Sternberg cells. Oncogene 18:943–953
174. Jungnickel B, Staratschek-Jox A, Bräuninger A, Spieker T, Wolf J, Diehl V, Hansmann ML, Rajewsky K, Küppers R (2000) Clonal deleterious mutations in the IkappaBalpha gene in the malignant cells in Hodgkin's lymphoma. J Exp Med 191:395–402
175. Emmerich F, Theurich S, Hummel M, Haeffker A, Vry MS, Döhner K, Bommert K, Stein H, Dörken B (2003) Inactivating I kappa B epsilon mutations in Hodgkin/Reed-Sternberg cells. J Pathol 201:413–420
176. Weniger MA, Melzner I, Menz CK, Wegener S, Bucur AJ, Dorsch K, Mattfeldt T, Barth TF, Möller P (2006) Mutations of the tumor suppressor gene SOCS-1 in classical Hodgkin lymphoma are frequent and associated with nuclear phospho-STAT5 accumulation. Oncogene 25:2679–2684
177. Reichel J, Chadburn A, Rubinstein PG, Giulino-Roth L, Tam W, Liu Y, Gaiolla R, Eng K, Brody J, Inghirami G, Carlo-Stella C, Santoro A, Rahal D, Totonchy J, Elemento O, Cesarman E, Roshal M (2015) Flow-sorting and exome sequencing reveals the oncogenome of primary Hodgkin and Reed-Sternberg cells. Blood 125:1061–1072
178. Tiacci E, Ladewig E, Schiavoni G, Penson A, Fortini E, Pettirossi V, Wang Y, Rosseto A, Venanzi A, Vlasevska S, Pacini R, Piattoni S, Tabarrini A, Pucciarini A, Bigerna B, Santi A, Gianni AM, Viviani S, Cabras A, Ascani S, Crescenzi B, Mecucci C, Pasqualucci L, Rabadan R, Falini B (2018) Pervasive mutations of JAK-STAT pathway genes in classical Hodgkin lymphoma. Blood 131:2454–2465
179. Mata E, Diaz-Lopez A, Martin-Moreno AM, Sanchez-Beato M, Varela I, Mestre MJ, Santonja C, Burgos F, Menarguez J, Estevez M, Provencio M, Sanchez-Espiridion B, Diaz E, Montalban C, Piris MA, Garcia JF (2017) Analysis of the mutational landscape of classic Hodgkin lymphoma identifies disease heterogeneity and potential therapeutic targets. Oncotarget 8:111386–111395
180. Martin-Subero JI, Gesk S, Harder L, Sonoki T, Tucker PW, Schlegelberger B, Grote W, Novo FJ, Calasanz MJ, Hansmann ML, Dyer MJ, Siebert R (2002) Recurrent involvement of the REL and BCL11A loci in classical Hodgkin lymphoma. Blood 99:1474–1477
181. Schmitz R, Stanelle J, Hansmann ML, Küppers R (2009) Pathogenesis of classical and lymphocyte-predominant Hodgkin lymphoma. Annu Rev Pathol 4:151–174
182. Otto C, Giefing M, Massow A, Vater I, Gesk S, Schlesner M, Richter J, Klapper W, Hansmann M-L, Siebert R, Küppers R (2012) Genetic lesions of the TRAF3 and MAP3K14 genes in classical Hodgkin lymphoma. Br J Haematol 157:702–708
183. Steidl C, Telenius A, Shah SP, Farinha P, Barclay L, Boyle M, Connors JM, Horsman DE, Gascoyne RD (2010) Genome-wide copy number analysis of Hodgkin Reed-Sternberg cells identifies recurrent imbalances with correlations to treatment outcome. Blood 116:418–427
184. Mottok A, Renné C, Willenbrock K, Hansmann ML, Bräuninger A (2007) Somatic hypermutation of SOCS1 in lymphocyte-predominant Hodgkin lymphoma is accompanied by high JAK2 expression and activation of STAT6. Blood 110:3387–3390

185. Hartmann S, Schuhmacher B, Rausch T, Fuller L, Döring C, Weniger M, Lollies A, Weiser C, Thurner L, Rengstl B, Brunnberg U, Vornanen M, Pfreundschuh M, Benes V, Küppers R, Newrzela S, Hansmann ML (2016) Highly recurrent mutations of SGK1, DUSP2 and JUNB in nodular lymphocyte predominant Hodgkin lymphoma. Leukemia 30:844–853
186. Rossi D, Trifonov V, Fangazio M, Bruscaggin A, Rasi S, Spina V, Monti S, Vaisitti T, Arruga F, Fama R, Ciardullo C, Greco M, Cresta S, Piranda D, Holmes A, Fabbri G, Messina M, Rinaldi A, Wang J, Agostinelli C, Piccaluga PP, Lucioni M, Tabbo F, Serra R, Franceschetti S, Deambrogi C, Daniele G, Gattei V, Marasca R, Facchetti F, Arcaini L, Inghirami G, Bertoni F, Pileri SA, Deaglio S, Foa R, Dalla-Favera R, Pasqualucci L, Rabadan R, Gaidano G (2012) The coding genome of splenic marginal zone lymphoma: activation of NOTCH2 and other pathways regulating marginal zone development. J Exp Med 209:1537–1551
187. Clipson A, Wang M, de Leval L, Ashton-Key M, Wotherspoon A, Vassiliou G, Bolli N, Grove C, Moody S, Escudero-Ibarz-L, Gundem G, Brugger K, Xue X, Mi E, Bench A, Scott M, Liu H, Follows G, Robles EF, Martinez-Climent JA, Oscier D, Watkins AJ, Du MQ (2015) KLF2 mutation is the most frequent somatic change in splenic marginal zone lymphoma and identifies a subset with distinct genotype. Leukemia 29:1177–1185
188. Mateo M, Mollejo M, Villuendas R, Algara P, Sanchez-Beato M, Martinez P, Piris MA (1999) 7q31-32 allelic loss is a frequent finding in splenic marginal zone lymphoma. Am J Pathol 154:1583–1589
189. Martinez-Lopez A, Curiel-Olmo S, Mollejo M, Cereceda L, Martinez N, Montes-Moreno S, Almaraz C, Revert JB, Piris MA (2015) MYD88 (L265P) somatic mutation in marginal zone B-cell lymphoma. Am J Surg Pathol 39:644–651
190. Durham BH, Getta B, Dietrich S, Taylor J, Won H, Bogenberger JM, Scott S, Kim E, Chung YR, Chung SS, Hullein J, Walther T, Wang L, Lu SX, Oakes CC, Tibes R, Haferlach T, Taylor BS, Tallman MS, Berger MF, Park JH, Zenz T, Abdel-Wahab O (2017) Genomic analysis of hairy cell leukemia identifies novel recurrent genetic alterations. Blood 130:1644–1648
191. Dietrich S, Hüllein J, Lee SC, Hutter B, Gonzalez D, Jayne S, Dyer MJ, Oles M, Else M, Liu X, Slabicki M, Wu B, Troussard X, Dürig J, Andrulis M, Dearden C, von Kalle C, Granzow M, Jauch A, Frohling S, Huber W, Meggendorfer M, Haferlach T, Ho AD, Richter D, Brors B, Glimm H, Matutes E, Abdel Wahab O, Zenz T (2015) Recurrent CDKN1B (p27) mutations in hairy cell leukemia. Blood 126:1005–1008
192. Tiacci E, Trifonov V, Schiavoni G, Holmes A, Kern W, Martelli MP, Pucciarini A, Bigerna B, Pacini R, Wells VA, Sportoletti P, Pettirossi V, Mannucci R, Elliott O, Liso A, Ambrosetti A, Pulsoni A, Forconi F, Trentin L, Semenzato G, Inghirami G, Capponi M, Di Raimondo F, Patti C, Arcaini L, Musto P, Pileri S, Haferlach C, Schnittger S, Pizzolo G, Foa R, Farinelli L, Haferlach T, Pasqualucci L, Rabadan R, Falini B (2011) BRAF mutations in hairy-cell leukemia. N Engl J Med 364:2305–2315
193. Willis TG, Jadayel DM, Du MQ, Peng H, Perry AR, Abdul-Rauf M, Price H, Karran L, Majekodunmi O, Wlodarska I, Pan L, Crook T, Hamoudi R, Isaacson PG, Dyer MJ (1999) Bcl10 is involved in t(1;14)(p22;q32) of MALT B cell lymphoma and mutated in multiple tumor types. Cell 96:35–45
194. Zhang Q, Siebert R, Yan M, Hinzmann B, Cui X, Xue L, Rakestraw KM, Naeve CW, Beckmann G, Weisenburger DD, Sanger WG, Nowotny H, Vesely M, Callet-Bauchu E, Salles G, Dixit VM, Rosenthal A, Schlegelberger B, Morris SW (1999) Inactivating mutations and overexpression of BCL10, a caspase recruitment domain-containing gene, in MALT lymphoma with t(1;14)(p22;q32). Nat Genet 22:63–68
195. Streubel B, Lamprecht A, Dierlamm J, Cerroni L, Stolte M, Ott G, Raderer M, Chott A (2003) T(14;18)(q32;q21) involving IGH and MALT1 is a frequent chromosomal aberration in MALT lymphoma. Blood 101:2335–2339
196. Streubel B, Vinatzer U, Lamprecht A, Raderer M, Chott A (2005) T(3;14)(p14.1;q32) involving IGH and FOXP1 is a novel recurrent chromosomal aberration in MALT lymphoma. Leukemia 19:652–658
197. Takino H, Okabe M, Li C, Ohshima K, Yoshino T, Nakamura S, Ueda R, Eimoto T, Inagaki H (2005) p16/INK4a gene methylation is a frequent finding in pulmonary MALT lymphomas at diagnosis. Mod Pathol 18:1187–1192
198. Johansson P, Klein-Hitpass L, Grabellus F, Arnold G, Klapper W, Pfortner R, Dührsen U, Eckstein A, Dürig J, Küppers R

(2016) Recurrent mutations in NF-kappaB pathway components, KMT2D, and NOTCH1/2 in ocular adnexal MALT-type marginal zone lymphomas. Oncotarget 7:62627–62639

199. Moody S, Thompson JS, Chuang SS, Liu H, Raderer M, Vassiliou G, Wlodarska I, Wu F, Cogliatti S, Robson A, Ashton-Key M, Bi Y, Goodlad J, Du MQ (2018) Novel GPR34 and CCR6 mutation and distinct genetic profiles in MALT lymphomas of different sites. Haematologica 103:1329–1336
200. Iida S, Rao PH, Nallasivam P, Hibshoosh H, Butler M, Louie DC, Dyomin V, Ohno H, Chaganti RS, Dalla-Favera R (1996) The t(9;14)(p13;q32) chromosomal translocation associated with lymphoplasmacytoid lymphoma involves the PAX-5 gene. Blood 88:4110–4117
201. Treon SP, Xu L, Yang G, Zhou Y, Liu X, Cao Y, Sheehy P, Manning RJ, Patterson CJ, Tripsas C, Arcaini L, Pinkus GS, Rodig SJ, Sohani AR, Harris NL, Laramie JM, Skifter DA, Lincoln SE, Hunter ZR (2012) MYD88 L265P somatic mutation in Waldenstrom's macroglobulinemia. N Engl J Med 367:826–833
202. Avet-Loiseau H, Li JY, Facon T, Brigaudeau C, Morineau N, Maloisel F, Rapp MJ, Talmant P, Trimoreau F, Jaccard A, Harousseau JL, Bataille R (1998) High incidence of translocations t(11;14)(q13;q32) and t(4;14)(p16;q32) in patients with plasma cell malignancies. Cancer Res 58:5640–5645
203. Chesi M, Nardini E, Brents LA, Schrock E, Ried T, Kuehl WM, Bergsagel PL (1997) Frequent translocation t(4;14)(p16.3;q32.3) in multiple myeloma is associated with increased expression and activating mutations of fibroblast growth factor receptor 3. Nat Genet 16:260–264
204. Chesi M, Bergsagel PL, Shonukan OO, Martelli ML, Brents LA, Chen T, Schrock E, Ried T, Kuehl WM (1998) Frequent dysregulation of the c-maf proto-oncogene at 16q23 by translocation to an Ig locus in multiple myeloma. Blood 91:4457–4463
205. Landowski TH, Qu N, Buyuksal I, Painter JS, Dalton WS (1997) Mutations in the Fas antigen in patients with multiple myeloma. Blood 90:4266–4270
206. Shou Y, Martelli ML, Gabrea A, Qi Y, Brents LA, Roschke A, Dewald G, Kirsch IR, Bergsagel PL, Kuehl WM (2000) Diverse karyotypic abnormalities of the c-myc locus associated with c-myc dysregulation and tumor progression in multiple myeloma. Proc Natl Acad Sci U S A 97:228–233
207. Liu P, Leong T, Quam L, Billadeau D, Kay NE, Greipp P, Kyle RA, Oken MM, Van Ness B (1996) Activating mutations of N- and K-ras in multiple myeloma show different clinical associations: analysis of the Eastern Cooperative Oncology Group Phase III Trial. Blood 88:2699–2706
208. Munshi NC, Avet-Loiseau H (2011) Genomics in multiple myeloma. Clin Cancer Res 17:1234–1242
209. Dunn-Walters DK, Isaacson PG, Spencer J (1995) Analysis of mutations in immunoglobulin heavy chain variable region genes of microdissected marginal zone (MGZ) B cells suggests that the MGZ of human spleen is a reservoir of memory B cells. J Exp Med 182:559–566

Chapter 2

Flow Cytometry for Non-Hodgkin and Hodgkin Lymphomas

Emily Glynn, Lori Soma, David Wu, Brent L. Wood, and Jonathan R. Fromm

Abstract

Multiparametric flow cytometry is a powerful diagnostic tool that permits rapid assessment of cellular antigen expression to quickly provide immunophenotypic information suitable for disease classification. This chapter describes a general approach for the identification of abnormal lymphoid populations by flow cytometry, including B, T, NK, and Hodgkin lymphoma cells suitable for the clinical and research environment. Knowledge of the common patterns of antigen expression of normal lymphoid cells is critical to permit identification of abnormal populations at disease presentation and for minimal residual disease assessment. We highlight an overview of procedures for processing and immunophenotyping non-Hodgkin B- and T-cell lymphomas and also describe our strategy for the sensitive and specific diagnosis of classical Hodgkin lymphoma and nodular lymphocyte predominant Hodgkin lymphoma.

Key words B cells, B-cell lymphoma, Clonality, Flow cytometry, Hodgkin lymphoma, Light-chain restriction, T cells, T-cell lymphoma, T-cell receptor V-beta repertoire analysis

1 Introduction

Immunophenotypic analysis of both non-Hodgkin lymphomas (NHL) and Hodgkin lymphomas is commonly required for proper diagnostic classification and is currently accomplished by immunohistochemistry and flow cytometry [1]. Multiparameter flow cytometry is an advanced diagnostic technology that is widely used for the clinical evaluation of complex cellular mixtures such as peripheral blood, bone marrow aspirates, body fluids, lymph nodes, and other tissue specimens [2]. The ability of the technique to perform rapid characterization of numerous cellular populations for the simultaneous expression of multiple cell surface or cytoplasmic antigens on individual cells over several orders of magnitude of expression has led flow cytometry to be regarded as a "standard of care" for the evaluation of hematopoietic processes in the clinical laboratory.

In flow cytometry, a suspension of cells is injected into a fluid stream under laminar flow conditions [3] resulting in individual cells being directed through a quartz capillary tube (flow cell).

Ralf Küppers (ed.), *Lymphoma: Methods and Protocols*, Methods in Molecular Biology, vol. 1956,
https://doi.org/10.1007/978-1-4939-9151-8_2,

Following illumination by one or more light sources, typically lasers, multiple cellular properties may be simultaneously assessed, including light scatter and the expression of surface and/or cytoplasmic antigens [3, 4] when cells are labeled with fluorochrome-conjugated antibodies directed against specific antigens [5]. Both liquid (peripheral blood, bone marrow, cerebrospinal fluid, etc.) and tissue specimens are routinely analyzed to provide unique diagnostic information about hematopoietic populations [2]. In general, flow cytometry permits the evaluation of multiple antigens on cells in a given experiment, typically four to six and increasingly in up to ten or more antigens [4]. Here, we discuss practical aspects of lymphoma diagnosis and classification by flow cytometry. As part of our approach, we assume that an abnormal cell population will have an aberrant immunophenotype as compared to background normal or reactive cells [6]. As such, emphasis is placed on identifying aberrant immunophenotypes in a multiparametric analysis. As there have been a numerous articles describing various strategies for flow cytometry, the reader is advised to access other articles in the field to fill deficiencies or intentional omissions herein.

2 Materials

Reagents are generally used as provided from the manufacturer with in-house validation prior to use. However, it is important to titer antibodies for optimal signal-to-noise response under the conditions to be used, and this may result in the use of antibodies at concentrations below that recommended by the manufacturer. The antigens that we target for routine B- and T-cell non-Hodgkin lymphoma and Hodgkin lymphoma analysis are included in Table 1, with specific antibodies and fluorochromes employed using a modified LSRII flow cytometer (Becton-Dickinson).

2.1 Buffers and Cell Staining Reagents

1. PBS-BSA buffer: Dulbecco's phosphate-buffered saline (Gibco®) with 3% bovine serum albumin added. PBS contains 2.67 mM KCl, 1.47 mM KH_2PO_4, 137.9 mM NaCl, and 8.1 mM Na_2HPO_4.
2. RPMI 1640.
3. Lysing/fixation solution: 0.15 mol/L NH_4Cl, pH 7.2 containing 0.25% ultrapure formaldehyde (Polysciences). The solution fixes cells and lyses red blood cells (primarily anucleate forms).
4. Medium A and B for cytoplasmic and nuclear antibody staining (FIX & PERM®, Life Technologies)
5. 40 μm filters.

Table 1
Fluorochrome combinations used for lymphoma immunophenotyping at the University of Washington

Antigen	B cells	B cell add-on for CLL/mantle cell	Add on tube for CLL prognosis	B cell add-on for hairy cell leukemia	B cell add-on for marginal zone lymphoma[a]	T cell	TCR V-beta	cHL (9-color)	cHL (6-color)	NLPHL tube 1	NLPHL tube 2
CD2						FITC					
CD3						PE-Cy7	PE-Cy7		APC-Cy7		
CD4						A594	A594				
CD5	PE-Cy5.5	PE-Cy5.5	PE-Cy5			PE	ECD	ECD		ECD	ECD
CD7						APC	APC				
CD8						BV421	APC-Cy7				
CD10	APC									A488	PE-Cy7
CD11c				APC	APC						
CD15								APC			
CD19	PE-Cy7	PE-CF594	PE-CF594	PE-Cy7	PE-Cy7						
CD20	V450							PE-Cy7	PE-Cy7	PE-Cy7	APC-H7
CD23		PE									
CD25				PE							
CD30						APC-A700		PE	PE		
CD32										PE	
CD34						PE-CF594					
CD38	A594		PE							A594	A594
CD40								PE-Cy5.5	PE-Cy5.5	PE-Cy5.5	PE-Cy5.5
CD45	APC-H7			V450	APC-H7	APC-H7		APC-H7		APC-H7	

(continued)

Table 1
(continued)

Antigen	B cells	B cell add-on for CLL/mantle cell	Add on tube for CLL prognosis	B cell add-on for hairy cell leukemia	B cell add-on for marginal zone lymphoma[a]	T cell	TCR V-beta	cHL (9-color)	cHL (6-color)	NLPHL tube 1	NLPHL tube 2
CD54										APC	APC
CD56						PE-Cy5	PE-Cy5				
CD64								FITC	FITC	FITC	FITC
CD71								APC-A700		APC-A700	APC-A700
CD75										V450	
CD95								PB	APC		
CD103				FITC							
CD200		APC									
HLA-DR							PB				
FMC-7		FITC									
ZAP-70			FITC								
Kappa	FITC				FITC						
Lambda	PE				PE						
V-beta isoform reagents							FITC, PE, and FITC and PE				
BCL-6											PE
DAPI											V450

[a]A second tube for marginal zone lymphoma where CD43-APC (instead of CD11c-APC) is occasional used. Abbreviations: *PB* pacific blue, *FITC* fluorescein isothiocyanate, *PE* phycoerythrin, *ECD/PE-TR* PE-Texas Red, *PECy5* PE-Cyanine-5, *PE-Cy5.5* PE-Cyanine-5.5, *PE-Cy7* PE-Cyanine-7, *A594* AlexaFluor 594, *APC* allophycocyanin, *APC-A700* APC-AlexaFluor 700

3 Methods

3.1 Sample Preparation

Tissue specimens are first disaggregated to create a single-cell suspension:

3.1.1 Disaggregation of Tissue Specimens

1. Mince the tissue with a scalpel in approximately 5 mL of RPMI 1640.
2. Filter the disaggregated cell suspension through a 40 micron filter.
3. Pellet the cells by centrifugation (550 × *g* for 5 min), and decant the remaining RPMI 1640.
4. Resuspend the cells in PBS-BSA or RPMI 1640, pellet the cells again by centrifugation (550 × *g* for 5 min), and resuspend in RPMI 1640 to a cell count of 10,000 cells/μL or less.
5. Add sufficient cell suspension to deliver up to one million cells in a volume of less than 200 μL.

3.1.2 Cell Surface Labeling of Cell Suspensions

1. Add appropriate fluorescently labeled, titered antibodies to cell suspension (from disaggregated tissue, bone marrow, blood, etc.), typically 5–20 μL of each antibody, and mix gently. The antibodies may be cocktailed prior to use for efficient delivery.
2. Incubate the labeled cells for 15 min at room temperature (RT) in the dark.
3. Add 1.5 mL of lysing/fixation solution.
4. Incubate for 15 min at RT in the dark.
5. Centrifuge the cells (550 × *g* for 5 min) and decant the supernatant.
6. Add 3 mL of PBS-BSA, centrifuge (550 × *g* for 5 min), and decant the supernatant.
7. Resuspend the cells in 100 μL of PBS-BSA.
8. Collect 150,000 events (if possible).

3.1.3 Cytoplasmic or Nuclear Labeling of Cell Suspensions (for BCL2, BCL6, and ZAP-70)

1. Add appropriate fluorescently labeled, titered cell surface antibodies to cell suspension, and mix gently.
2. Incubate the labeled cells for 15 min at room temperature (RT) in the dark.
3. After washing the cells twice with PBS-BSA, add 100 μL of medium A and mix well.
4. Incubate for 15 min at room temperature in the dark.
5. Wash the cells twice with PBS-BSA and add 100 μL of medium B and mix well.
6. Add appropriate amount of cytoplasmic or nuclear antibody, mix, and incubate the labeled cells for 30 min (for cytoplasmic

antigens) or 60 min (for nuclear antigens) in the dark at room temperature.

7. Wash the cells twice with PBS-BSA and resuspend the cells in 100 μL of PBS-BSA.
   For classical Hodgkin lymphoma (CHL), nodular lymphocyte predominant Hodgkin lymphoma (NLPHL), and lymphoma minimal residual disease testing:
   The sample processing and immunostaining protocols are the same as those described above with the exception that more events should be collected, preferably ~500,000. The CHL and NLPHL assays have only been validated on lymph nodes and thus are only recommended for tissue specimens.

3.2 Gating Strategies, Data Analysis, and Interpretation

3.2.1 B-Cell Analysis

Compensated data files for B-cell NHL are first gated to exclude so-called "doublet" events using a plot of forward scatter area versus forward scatter height. The doublet events represent coincident cells in the flow cell and need to be excluded as the antigenic profile derived from these events may result in an apparent composite immunophenotype due to the combined antigenic expression of two or more cells. Subsequently, nonviable events are excluded using forward and side light scatter gating. As cells degenerate, initially forward scatter decreases, while side-scatter increases, and later both decrease in intensity. These nonviable events can be readily excluded by selective gating considering these findings.

The evaluation for B-cell NHL next involves identification of B cells with subsequent evaluation for the presence of surface immunoglobulin light chain-restricted populations with aberrant immunophenotypes (*see* **Note 1**). We currently gate lymphocytes with forward and side light scatter (*see* **Note 2**), computationally subtract the CD5-positive and CD19-negative T cells, and then isolate the B cells on a plot of CD19 versus side scatter. Data is subsequently evaluated by examining various antigens (Table 1) plotted against each other using a gating strategy first focusing on CD19-positive B cells and then evaluating all lymphocytes. Evaluation of lymphocytes based on forward versus side scatter (in addition to CD19-positive B cells) prevents inadvertent exclusion of B-cell populations that have aberrant loss or decreased expression of CD19 (*see* **Note 3**). In general, the normal kappa to lambda ratio of the B cells is approximately 1.4; however, sole emphasis on a skewed light-chain ratio for identifying an abnormal B-cell population is discouraged due to its poor sensitivity and specificity. Rather, an emphasis is placed on identifying abnormalities of other antigens assessed in these studies (CD5, CD10, CD19, CD20, CD38, and CD45) that are often under- or overexpressed relative to any normal population, helping to separate the abnormal and normal populations. Examples of typical examples of chronic lymphocytic leukemia (CLL)/small lymphocytic lymphoma (SLL) (Fig. 1), marginal zone lymphoma (Fig. 2), hairy cell leukemia (Fig. 3),

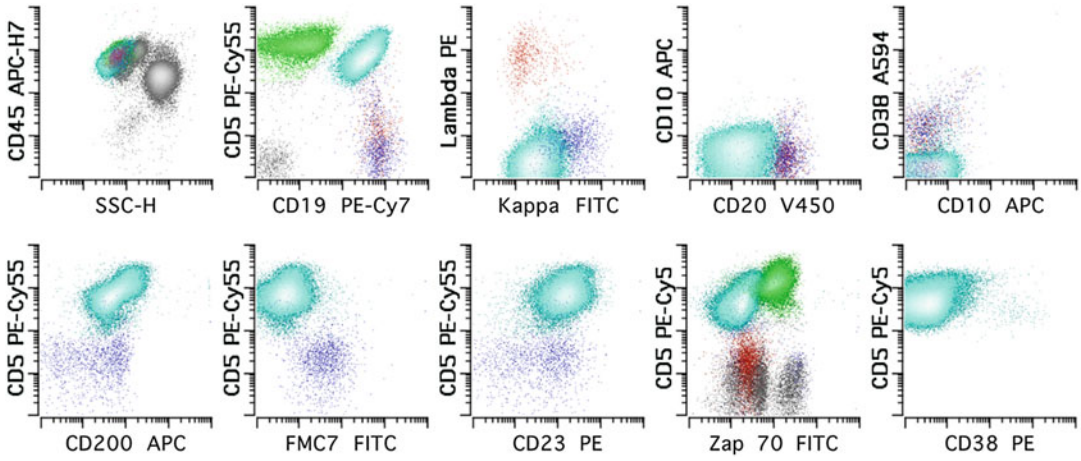


Fig. 1 Flow cytometric characterization of chronic lymphocytic leukemia/small lymphocytic lymphoma (CLL/SLL). The neoplastic population (cyan) demonstrates expression of CD45, CD19, and CD5; low-level kappa surface light chains; decreased CD20, CD23, and CD200; and no expression of FMC7 or CD10. The neoplastic population does not express ZAP-70 (positive cells determined by a discriminator which is negative on the normal B cells and positive on the normal T cells) or CD38, features associated with a more favorable prognosis in CLL/SLL [74, 75]. The first dot plot of the top panel shows all cells, the second shows lymphocytes, and the last three show only B cells. In the bottom panel of dot plots, all lymphocytes are shown in the fourth plot, while the first, second, third, and fifth plots show only B cells. A small population of reactive B cells is identified by blue (kappa expressing) and red (lambda expressing) events. Reactive T cells are colored in green. The color version of this figure is available in the electronic version of this manuscript

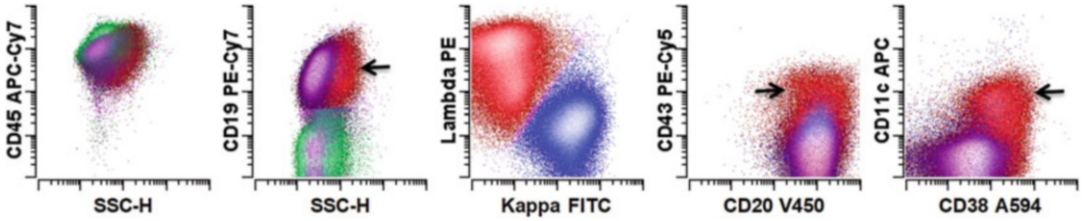


Fig. 2 Characterization of an abnormal B-cell population using CD11c and CD43. The right parotid gland biopsy demonstrates a decreased kappa to lambda ratio (0.83) but no other antigenic abnormalities. Characterization of the B cells using CD11c and CD43 demonstrates a lambda-restricted population with expression of these antigens (arrows), consistent with involvement by a B-cell lymphoproliferative disorder (marginal zone lymphoma by morphology). T cells are green, and kappa- and lambda-expressing B cells are in blue and red, respectively. The first dot plot shows all cells, the second shows all lymphocytes, and the third through fifth shows only B cells. The color version of this figure is available in the electronic version of this manuscript

Burkitt lymphoma (minimal residual disease) (Fig. 4), and follicular lymphoma (Fig. 5) are provided. While individual cases vary, a summary of common immunophenotypes of typical B-cell lymphomas is provided (Table 2). A more comprehensive discussion of immunophenotypes in B-cell NHL can be found in the review by Craig and Foon [2].

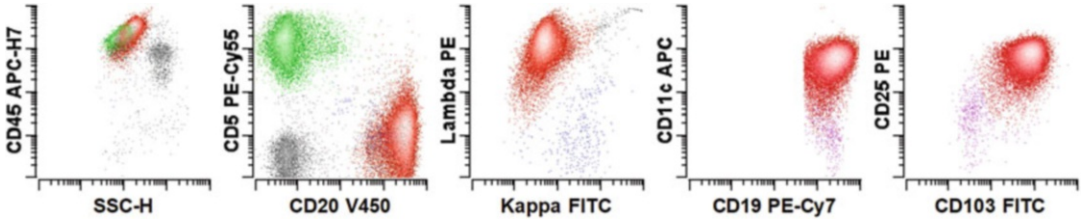


Fig. 3 Immunophenotypic characterization of a case of hairy cell leukemia in the bone marrow. The first dot plot shows all leukocytes, the second dot plot shows all lymphocytes, and the remaining three dot plots show only B cells. T cells are colored in green, kappa-restricted B cells are colored in blue, and the lambda-restricted B cells are colored in red. The neoplastic hairy cell leukemia population demonstrates increased side light scatter (compared to other lymphocytes) and CD45 expression, bright expression of CD20, lambda light-chain expression, and expression of CD11c, CD19, CD25, CD38 (data not shown), and CD103, without CD5 or CD10 (data not shown). The color version of this figure is available in the electronic version of this manuscript

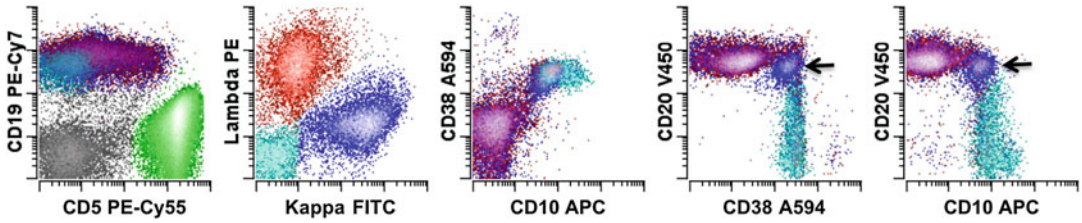


Fig. 4 Minimal residual disease detection of Burkitt lymphoma in the bone marrow. Overall, this bone marrow specimen demonstrates a normal kappa to lambda ratio of 1.5. In addition, multiple dot plots do not demonstrate an abnormal B-cell population. However, a small (1% of leukocytes) kappa-restricted B-cell population is noted (arrows) with relatively bright expression of CD38 and CD10. This immunophenotype is identical to that identified in the prior presenting mesenteric mass. The first dot plot shows all lymphocytes, while the remaining four dot plots show only B cells. T cells are colored green; kappa- and lambda-restricted mature B cells are colored blue and red, respectively; immature, normal B lymphoblasts are colored cyan. The color version of this figure is available in the electronic version of this manuscript

The approach for evaluating specimens for B-cell NHL minimal residual disease (MRD) is identical to that used for evaluating samples at disease presentation, with the obvious difference being that the abnormal population post-therapy can be very small, sometimes less than 0.1% of the viable cells. In practice, a series of sequential gates to selectively include and/or exclude various populations may be required to convincingly isolate the abnormal population and evaluate for definite light-chain restriction (Fig. 4). Identifying these small populations is facilitated by knowledge of the original immunophenotype of the abnormal B-cell population at presentation (when the population is usually present at significantly increased proportion) preferably using the identical reagent combination used to immunophenotype the presentation specimen. Other caveats related to the analysis of B cells are described in **Notes 4–7**.

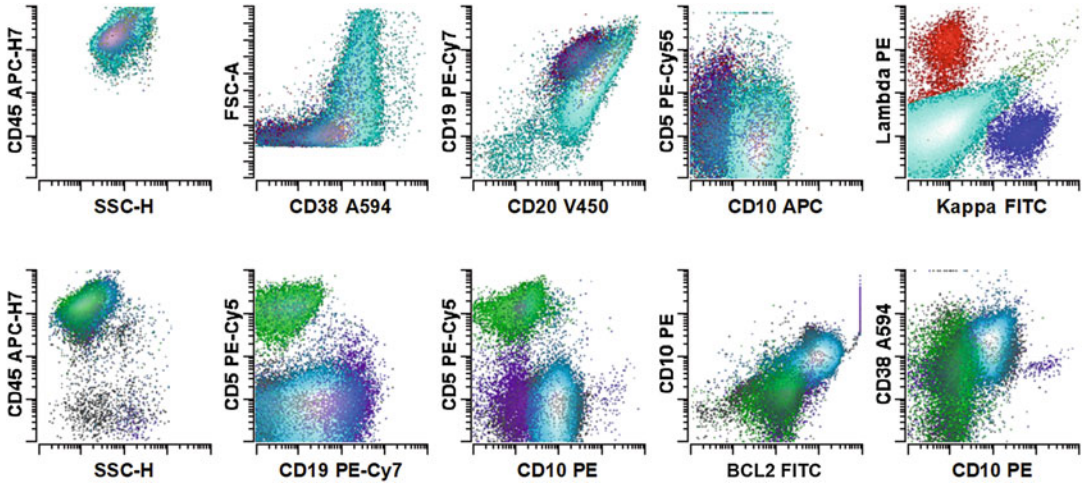


Fig. 5 Immunophenotypic characterization of a case of follicular lymphoma (grade 3). The first dot plots of both the top and bottom panels show all leukocytes, the last four in the top panel show all B cells, and the remaining four plots in the bottom panel show all lymphocytes. T cells are colored in green, kappa-restricted B cells are covered in blue, the lambda-restricted B cells are colored in red, and neoplastic B cells are colored in cyan. The neoplastic cells show expression of CD45, increased forward light scatter, decreased expression of CD19 relatively to the reactive B cells, absent surface light-chain restriction, and expression of CD10, CD20, BCL2 (relative to the reactive T cells), and CD38 (normal to slightly decreased relative to the expression of a normal germinal center population [55]) and no expression of CD5. The color version of this figure is available in the electronic version of this manuscript

Table 2
Immunophenotypes of common B-cell NHL

B-cell lymphoma	Immunophenotype
Diffuse large B-cell lymphoma	Increased forward and side light scatter with variable expression of CD19, CD20, CD45, and monotypic surface light chain. CD10 may or may not be expressed
Follicular lymphoma	CD45, CD19 (decreased), CD20 (decreased), CD10 (increased), CD38 (decreased), monotypic surface light chain, increased BCL2 as compared to background T cells
Chronic lymphocytic leukemia/small lymphocytic lymphoma	CD45, CD5, CD20 (dim), mono- or bitypic dim to absent surface light-chain expression, CD23 and CD200 positive, FMC-7 negative, with variable CD38 and ZAP-70 expression
Mantle-zone lymphoma	CD45, CD5, CD20 (normal), monotypic surface light-chain restriction, CD23 and CD200 negative, FMC-7 positive
Burkitt lymphoma	CD45, CD20, CD10, CD38 (increased), monotypic surface light-chain expression, no overexpression of BCL2
Hairy cell leukemia	CD20 (bright), CD19, monotypic surface light-chain expression, aberrant co-expression of CD11c, CD25, and CD103, without CD5 and CD10
Marginal zone lymphoma	CD20, CD19, negative for CD5 and CD10. Occasional cases have expression of CD43

3.2.2 T- and NK-Cell Analysis

In general, the flow cytometric gating strategy for T cells is similar to that for B cells. Compensated data files are first gated to exclude doublet events and then nonviable events (*see* **Note 8**). T cells are then identified typically by two gating strategies: (1) forward versus side scatter, to identify lymphocytes, or (2) CD3 versus side scatter, to identify T cells. The former approach helps to identify T cells that have aberrant decreased or absent expression of CD3, while the latter identifies T cells using CD3 to permit identification of potentially larger cells that may have increased side scatter (*see* also **Note 9**). While the analysis of B cells rests on identifying an immunophenotypically abnormal B-cell population that is clonal, evaluation of T cells involves the recognition of an immunophenotypically abnormal T-cell population. At some institutions, evaluation of probable clonality of T-cell populations can also be performed using flow cytometry by assessing the T-cell receptor (TCR) V-beta repertoire (*see* below).

As part of the immunophenotypic evaluation of T cells, it is critical to recognize that numerous reactive T-cell populations may be present in any given sample (*see* **Note 10**). These T cells may include memory T cells (decreased CD7 expression), gamma-delta T cells (increased CD3, absent CD4, and absent or partial expression of CD8), and large granular lymphocytes (typically $CD8^+$-/CD5-dim or $CD8^+$-/CD5-negative) and may be identified in any given patient sample, sometimes in increased proportion [2]. Familiarity with this normal spectrum of antigenic variation of reactive T cells is critical to prevent misinterpretation of a reactive population as a malignant clone. On the other hand, it is also possible that a neoplastic, clonal population of T cells may show identical immunophenotypic changes as any of these reactive populations. As such, careful clinical and pathologic correlation is required in every case.

With this spectrum of antigenic variation in mind, it is diagnostically fortuitous that the majority of T-cell lymphoproliferative disorders will demonstrate aberrant antigenic expression, such that the level of expression of T-cell-associated antigens (CD2, CD3, CD4, CD5, CD7, CD8, and CD45) can be either increased, decreased, or completely absent, as compared to the background, normal reactive T cells [7, 8]. Multiparametric identification of these antigenic changes consequently permits one to reliably identify an abnormal T-cell population, which in conjunction with histologic findings or clinical data can permit subsequent definite classification. As an example, flow cytometry of a liver core-needle sample from a 69-year-old man with multiple liver lesions (Fig. 6) identified an abnormal T-cell population comprising 25.5% of total white cells with increased forward and side scatter and expression of CD2 and CD30 (bright) and with aberrant loss of expression of CD3, CD5, and CD7 without CD8, CD34, or CD56. CD4 may be expressed dimly on a small subset. The immunophenotype of the abnormal $CD45^+CD30^+$ cell population was favored to represent a

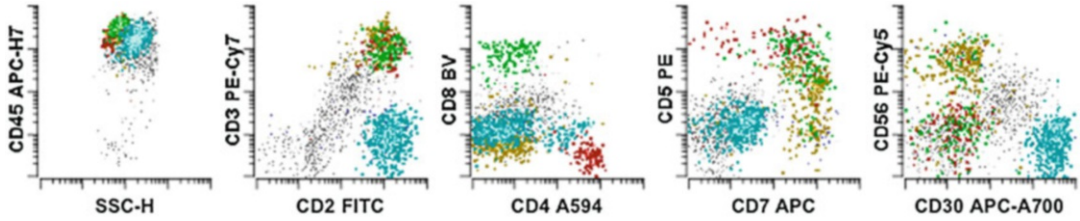


Fig. 6 Example of a CD30$^+$ T-cell lymphoma. An abnormal CD2-positive T-cell population (teal) is identified with strong CD30 expression, CD4 expression on a small subset, and aberrant loss of CD3, CD5, and CD7 without CD8 or CD56. Reactive CD4, CD8, and CD4-/CD8-negative T cells are colored red, green, and orange, respectively. Of note, although the expression of CD2 and possible dim expression of CD4 is not considered lineage-specific, subsequent molecular analysis of T-cell receptor gamma gene rearrangement confirmed the population is clonal. The first dot plots show all leukocytes, and the last four show all lymphocytes. The color version of this figure is available in the electronic version of this manuscript

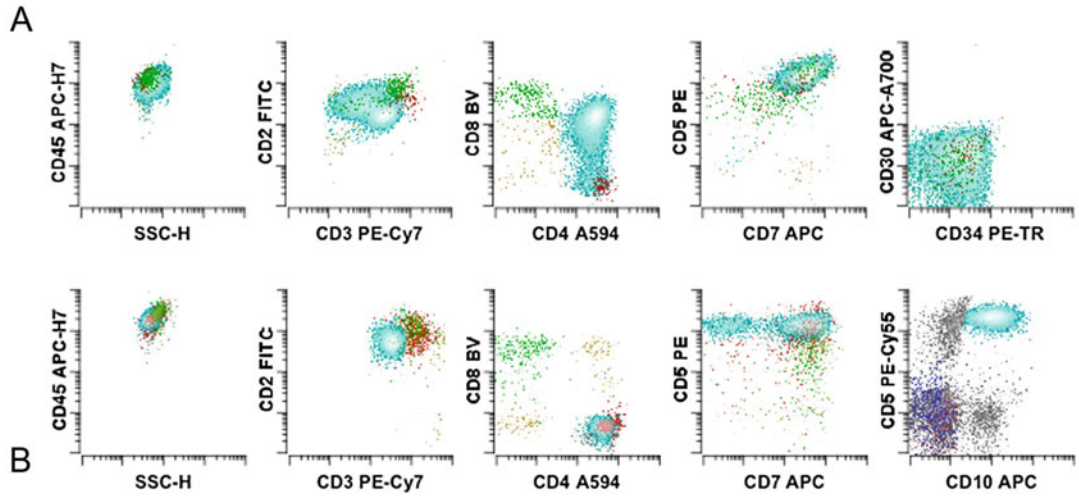


Fig. 7 Examples of other mature T-cell lymphomas. (**a**) T-cell prolymphocytic lymphoma. A population of CD4/CD8+ double-positive T cells (cyan) with decreased expression of CD2, CD3, and CD45 (relative to the other lymphocytes), without evidence of CD34 or TdT (data not shown), is identified in a peripheral blood sample of a 73-year-old male with a white blood cell (WBC) count of 398,000 and prior cytogenetic studies showing involvement of a complex karyotype, including inv(14)(q11.2q32). Background reactive CD4$^+$ (red) and CD8$^+$ (green) T lymphocytes are also shown. All dot plots show all lymphocytes. (**b**) Angioimmunoblastic T-cell lymphoma, a population of CD4$^+$ T cells (cyan) with aberrant expression of CD2 (slightly decreased), CD3 (decreased), and CD10, with variable CD7 and normal expression of CD5 and CD45, is identified in a 91-year-old female peripheral blood sample, comprising 18.6% of total white cells. Background reactive CD4$^+$ (red) and CD8$^+$ (green) T lymphocytes are also shown. The first dot plot shows all leukocytes, and the last four show all lymphocytes. The color version of this figure is available in the electronic version of this manuscript

CD30$^+$ peripheral T-cell lymphoma, and a clonal TCR-gamma gene rearrangement (by PCR) was identified, providing support for T-cell lineage. Other examples of mature T-cell lymphomas, including T-cell prolymphocytic leukemia and angioimmunoblastic T-cell lymphoma, are also shown (Fig. 7). While individual cases

Table 3
Immunophenotypes of common T- and NK-cell NHL

T-cell lymphoma	Immunophenotype
T-PLL	CD4+/−, CD8−/+; variable CD3, CD2, CD5, and CD7+
AITL	CD4+, CD5+, CD10+/−, CD3−/+
Mycosis fungoides/Sezary syndrome	CD4+, variably decreased CD7; CD3, CD5, and CD45+
PTCL-NOS	CD4+/−, often with loss of CD5 and/or CD7
ATLL	CD4+, CD25+, CD3+, decreased CD7
ALCL	CD30+, variable loss of T-cell antigens; TIA1/granzyme B+ and ALK+/− (by immunohistochemistry)
T-LGL	CD3+, CD4−, CD8+, with decreased CD5 and/or CD7
Extranodal NK/T cell lymphoma-nasal type	CD2, CD7, CD56+; CD3 (surface), CD4, CD5, CD8−. EBV+ by in situ hybridization

Abbreviations: *T-PLL* T-cell prolymphocytic leukemia; *AITL* angioimmunoblastic T-cell lymphoma; *PTCL-NOS* peripheral T-cell lymphoma, not otherwise specified; *ATLL* adult T-cell leukemia/lymphoma; *ALCL* anaplastic large-cell lymphoma; *T-LGL* T-cell large granular lymphocyte leukemia; *TIA1* T-cell intracellular antigen 1

vary, a summary of common immunophenotypes of typical T-cell lymphomas is provided (Table 3). A more comprehensive discussion of immunophenotypes in T-cell NHL can be found elsewhere [2]. *See* **Note 11** for an additional caveat regarding the analysis of T cells.

At presentation, a common question clinically is whether a T-cell population of interest represents a clonal process or alternatively is a reactive oligoclonal or polyclonal expansion of an immunophenotypically distinct subset. Flow cytometric analysis of TCR V-beta repertoire is a methodology that uses fluorescently labeled anti-TCR V-beta antibodies to determine if the identified T-cell populations represent a diverse or a clonal process [9–13]. As provided by the manufacturer, the IOTest® Beta Test Mark assay is a set of 24 antibodies that covers approximately 70% of the normal human TCR V-beta repertoire. These V-beta-specific antibodies are each conjugated to one of three fluorochromes: FITC, PE, or PE/FITC dual conjugate. According to the intended original protocol, the assay is run in eight separate tubes, such that three V-beta family-specific antibodies (one labeled with FITC, one labeled with PE, and one labeled with FITC and PE) are present in each tube. Through the use of gating reagents, (e.g., CD3, CD4, or CD8), the T-cell population of interest can be isolated to determine if there is evidence for overrepresentation of a particular TCR V-beta isoform relative to normal, a finding that is strongly suggestive of clonality. Numerous studies have demonstrated the utility of this approach for assessing putative T-cell clonality in both the

clinical and research environment [9–12]. The reader is advised to consult these publications for performing this approach for TCR V-beta analysis.

At our institution, a simple permutation of the original protocol for TCR V-beta repertoire analysis has been developed and used for several years [14]. In this modification, instead of 8-tube analysis as intended by the manufacturer, all of the fluorescently labeled TCR V-beta antibodies are combined into a single tube for combined analysis. This modification permits rapid identification of a putative TCR V-beta restricted T-cell population with emphasis on identifying aberrant antigenic expression through the use of an increased number of T-cell-specific gating reagents (Table 1). As compared to the standard method, this modified approach is fast, can be adopted in any laboratory currently performing routine TCR V-beta analysis, and minimizes the amount of reagent and sample used for analysis, thus permitting application to specimens that are typically hypocellular in nature, such as skin biopsies and cerebrospinal fluid samples. An example of this modified approach is shown in which an expanded large granular lymphocyte population is identified with expression of CD8 and decreased expression of CD5 and CD7 (Fig. 8). Subsequent assessment of TCR V-beta repertoire analysis using a modified approach in which all V-beta

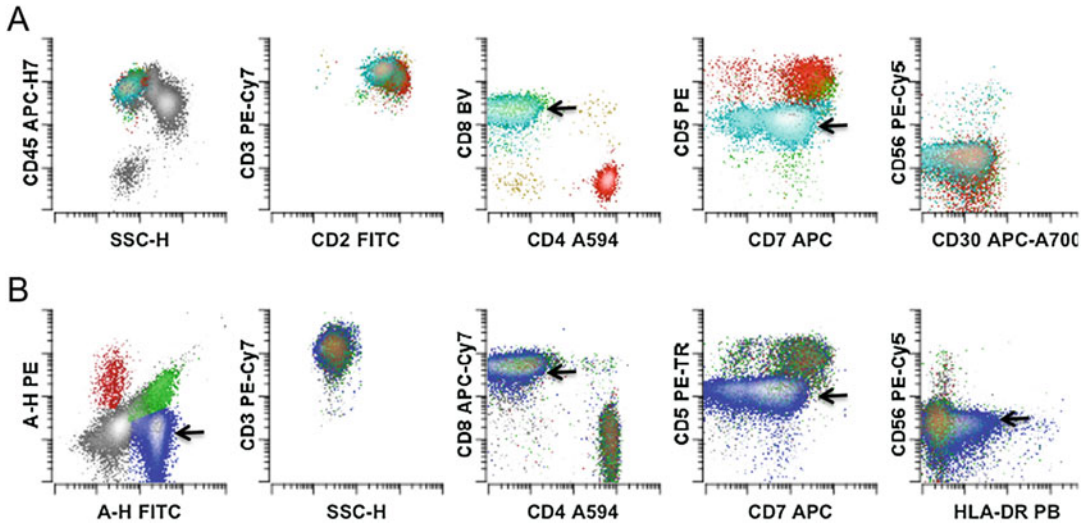


Fig. 8 Example of a combined, modified TCR V-beta assay showing clonality of T-LGL population. (**a**) A $CD8^+$ T-cell population (teal, arrows) is identified with decreased CD5 and CD7 expression as compared to background $CD4^+$ (red) and $CD8^+$ (green) T cells. The first dot plot shows all leukocytes, and the last four show all T cells. (**b**) Subsequent analysis of TCR V-beta expression using the modified assay in which all TCR V-beta antibodies are combined together shows the T-LGL population (blue, arrows) with apparent TCR V-beta restriction limited to a FITC-labeled isoform. By comparison, background, reactive T cells show variable expression of PE, FITC, and PE-FITC and unlabeled TCR V-beta isoforms. All dot plots show only T cells. The color version of this figure is available in the electronic version of this manuscript

antibodies are combined together permits identification of probable FITC-fluorochrome-labeled TCR V-beta isoform. If clinically indicated, molecular analysis of TCR gene rearrangement could be subsequently performed to confirm the presumed clonal nature of this population. In addition, if knowledge of the specific TCR V-beta isoform is important, such as for future minimal residual disease monitoring, the standard TCR V-beta repertoire assay could be performed to permit identification of which V-beta isoform is identified by the FITC-labeled anti-V-beta antibody.

Analysis for NK-cell neoplasms is essentially identical to that for T-cell neoplasms and uses the same flow cytometry combination. Using reagents described in this manuscript, normal and neoplastic NK cells typically express CD2, CD7, CD45, and CD56 without CD3 (surface), CD4, or CD5; CD8 is often expressed in a subset. As with the analysis of T cells, NK-cell neoplasms usually show an aberrant immunophenotype (typically loss of expression of an antigen) [15]. Additionally, reactive NK-cell populations with some immunophenotypic abnormalities can also be seen [16]. Ultimately, correlation with morphologic tissue sections and Epstein-Barr virus (EBV) in situ hybridization studies (NK-cell neoplasms are positive) is required for definitive diagnosis [2].

3.2.3 CHL Analysis

CHL is an unusual type of B-cell lymphoma [17–19] in which the neoplastic Hodgkin and Reed-Sternberg (HRS) cells are rare (less than 1% of the cells in lymph node) [20, 21]; the bulk of the cells in an involved lymph node include reactive lymphocytes, eosinophils, plasma cells, and histiocytes [21]; and the HRS cells bind to non-neoplastic T cells, resulting in HRS cell-T-cell rosettes [22–28]. Traditionally, CHL has been diagnosed by morphology and immunohistochemistry (HRS cells demonstrate expression of CD15 and CD30 but lack expression of CD20, CD3, and CD45 [17, 21, 29–31]). Our studies, however, have demonstrated that HRS cells can be directly identified by flow cytometry with high clinical sensitivity and specificity [28], allowing this neoplasm to be diagnosed by this technique [28, 32]. The interaction of T cells and HRS cells (rosetting) can be directly detected by flow cytometry, as demonstrated by the observation of a composite immunophenotype of the HRS cells and T cells, that is, cells with expression of both HRS cell and T-cell antigens [32, 33]. T-cell-HRS cell rosetting can be disrupted by unlabeled "blocking" antibodies (antibodies that can compete for the binding of the adhesion molecule binding partner [26, 28]), a practice useful for purifying HRS cells [28, 33] but is not necessary for diagnostic flow cytometry [32]. Reagent combinations are proposed for either 9-color [32] or 6-color [34] flow cytometry platforms (Table 1). We have also found the 9-color flow cytometry combination for CHL to be superior to our standard B-NHL flow cytometry combination for immunophenotyping primary mediastinal large B-cell lymphoma [35].

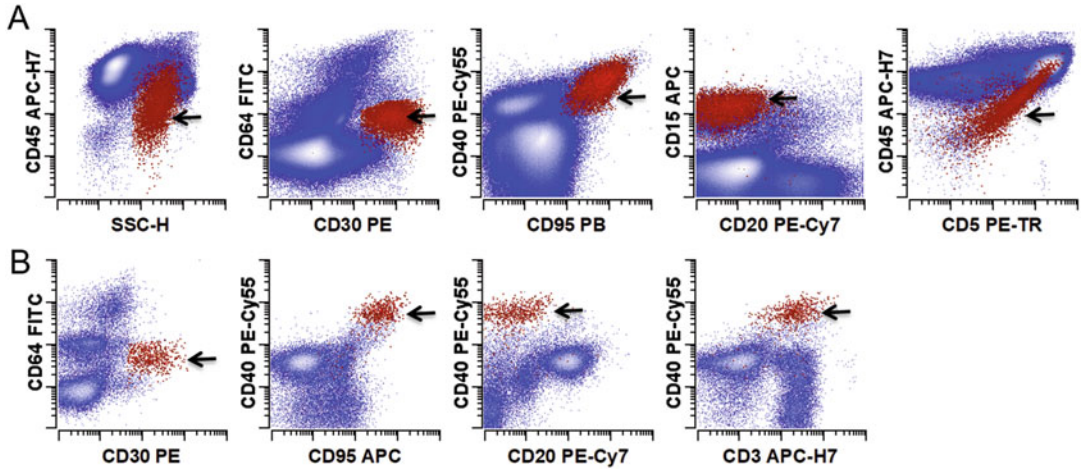


Fig. 9 Representative examples of 9- (**a**) and 6-color (**b**) flow cytometry studies of CHL cases. HRS cells (arrows; also shown in red and emphasized) are identified by their absence of expression of CD64 (position of negative determined by control experiments, not shown); expression of CD30, CD40, and CD95; and increased side light scatter (SSC-H) compared to normal lymphocytes; all remaining viable events are in blue. (**a**) The population of neoplastic HRS cells has expression of intermediate CD15, intermediate to bright CD30, intermediate to bright CD40, variable CD71 (data not shown), and intermediate to bright CD95, without expression of CD64 or CD20. The diagonal relationship between CD45 and CD5 is due to the presence of T cells bound to the HRS cells. (**b**) Neoplastic HRS cells have expression of intermediate CD30, bright CD40, and intermediate CD95, without expression of CD64 or CD20. CD3 is expressed suggesting some degree of HRS−/T-cell rosetting. All dot plots (in **a** and **b**) show all events. The color version of this figure is available in the electronic version of this manuscript

The gating strategy to identify HRS cells of CHL is different from that for B- and T-cell NHL. The first difference is that because of the relatively increased cell size of HRS and rosetted cells, increased side scatter is used to identify these populations (Figs. 9 and 10); HRS cells are then identified by the requisite expression of CD30, CD40, and CD95 with absence of strong CD20 expression [32, 34]. An HRS cell population must be identifiable in all four gates (Fig. 10) having increased side scatter as compared to background lymphocytes, expression of CD30 and increased autofluorescence (as assessed using the FITC channel), absence of or decreased expression of CD20, strong expression of CD40, and at least intermediate expression of CD95. Of note, the increased autofluorescence seen in HRS cells is critical to distinguishing these from $CD30^+$ reactive immunoblasts, and frequently, CD40 expression on these cells is increased as compared to background cells in the tissue preparation. If properly gated, HRS cells may appear as two subset populations, one with apparent T-cell antigen expression (such as CD3 or CD5) and the second with decreased (but not absent) CD45 but no expression of T-cell antigens, due to presence and absence of T-cell rosetting, respectively (*see* **Note 12**).

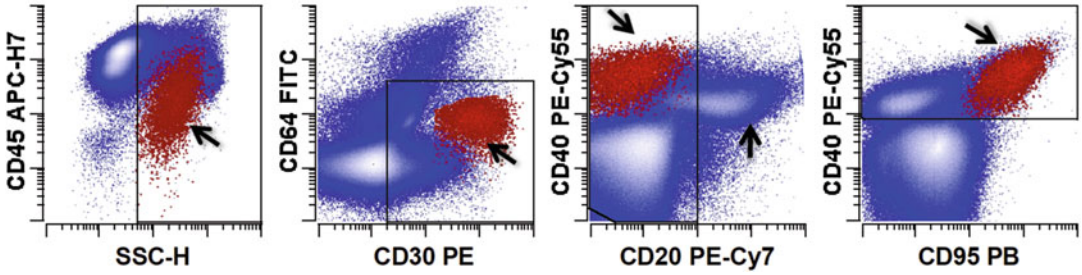


Fig. 10 Gating strategy to identify HRS cells in CHL. HRS cells are in red (also identified by oblique arrows) and are emphasized, while all other events are in blue. The B cells are identified in the vertical arrow ($CD20^+$, $CD40^+$). Putative HRS cell events must fall in all four gates and form a distinct population. Specifically, the cells must have increased side scatter (first plot), be $CD30^+$ and have increased autofluorescence in the FITC channel compared to small lymphocytes (second plot), express no or low CD20 (third plot), and express CD40 (at or greater intensity than a reactive B cells) and CD95 (fourth plot). All dot plots show all events. The color version of this figure is available in the electronic version of this manuscript

Table 4
Criteria for identifying an HRS-cell population. The putative HRS-cell population must meet all of the following criteria

1	Express CD30, CD40, and CD95
2	Have increased light scatter properties (compared to background lymphocytes)
3	Lack moderate to bright expression of CD20
4	Lack expression of CD64
5	Represent a discrete cell population in multidimensional projections of the immunophenotypic data with increased autofluorescence as compared to CD30+ reactive immunoblasts

A summary of the immunophenotypic criteria (*see* also **Notes 13** and **14**) for identifying a putative HRS cell population is provided in Table 4 [28, 32, 34]. *See* **Note 15** for a further caveat regarding the analysis of CHL.

3.2.4 Nodular Lymphocyte Predominant Hodgkin Lymphoma Analysis

Nodular lymphocyte predominant Hodgkin lymphoma (NLPHL) is a rare type of B-cell lymphoma, showing scattered, rare, large neoplastic germinal center B cells (lymphocyte predominant (LP) cells) embedded in a nodular, reactive infiltrate that includes small, nonneoplastic lymphocytes and histiocytes [31, 36, 37]. Immunohistochemical evaluation is critical for making a pathologic diagnosis; the LP cells express PAX5 and CD20 (pan-B-cell markers), CD45 (pan-hematopoietic marker), and BCL6 (marker of germinal center origin) and lack expression of CD3 (pan-T-cell marker), CD15, or CD30 [36–38]. Lack of CD15 and CD30 is particularly important (as noted previously), as HRS cells of CHL express these antigens [21, 29, 39].

The approach to identify LP cells is similar to that for immunophenotyping HRS cells of CHL. Two flow cytometry combinations are used to identify LP cells [40]. Both tubes (Table 1) utilize CD20, CD38, CD40, CD54, and CD71 to identify the LP cells and CD5, CD10, and CD64 to exclude other events. The first tube also employs CD32 and CD75 to evaluate the expression of these antigens on LP cells and CD5 and CD45 (like with CHL 9-color assay) to evaluate for rosette formation. The second tube evaluates DNA content (using DAPI) and nuclear expression of BCL6. The flow cytometry assays allow NLPHL to be immunophenotyped by flow cytometry with high sensitivity and specificity; from all seven cases of NLPHL evaluated, an LP population was identified in both tubes (100% sensitivity). Evaluation of 94 tissues (55 cases of B-cell lymphoma, 30 reactive cases, and 9 cases of CHL) did not identify an LP-like population in 91 cases (96.8% specificity). Perhaps as expected, the misclassified cases were ultimately diagnosed as diffuse large B-cell lymphoma, not otherwise specified (DLBCL, NOS) [40]. The findings from this pilot study need to be confirmed in a larger series of cases.

Criteria to identify an LP cells were developed based on immunophenotyping the NLPHL cell line DEV and published immunohistochemical studies of NLPHL and are as follows: (1) large putative LP cells should have increased forward and side light scatter compared to small lymphocytes; (2) the flow cytometry-derived immunophenotype should be similar to that observed for DEV and in immunohistochemistry for LP cells with expression of CD20 [36–38] and CD40 [41, 42]; (3) LP cells may be immunoglobulin light-chain restricted [43, 44]; (4) LP cells should be of germinal center origin ($BCL6^+$ [31, 36–38] and $CD75^+$ [45]) but with aberrant loss of CD10 [38, 46] and; (5) a putative LP-cell population will form a distinct cluster of events in multidimensional flow cytometric space.

Similar to the immunophenotype described in the medical literature by immunohistochemistry, LP cells by flow cytometry show absence of surface light chains or kappa surface light-chain restriction. LP cells show expression of B-cell markers (CD19 and CD20 (at the level of the reactive B cells), bright expression of CD40 (like CHL)) and BCL6. LP cells do not express CD5, CD10, CD15, or CD64; only minimal CD30 is present [37].

The methods described above readily allow LP cells to be studied by flow cytometry. Our prior studies suggest that LP cells interact with T cells with less avidity than HRS cells interact with T cells; few T-cell-LP cell rosettes are typically observed. However, adhesion molecules CD54 and CD50 likely mediate this interaction; unlike CHL, CD58 likely plays only a minor role in LP cell-T-cell interactions. Little expression of PD-L1 or PD-L2 is seen on LP cells (suggesting the blocking of the PD-1/PD-L1/PD-L2 pathway likely will not be efficacious in NLPHL). Finally, CD32 (like

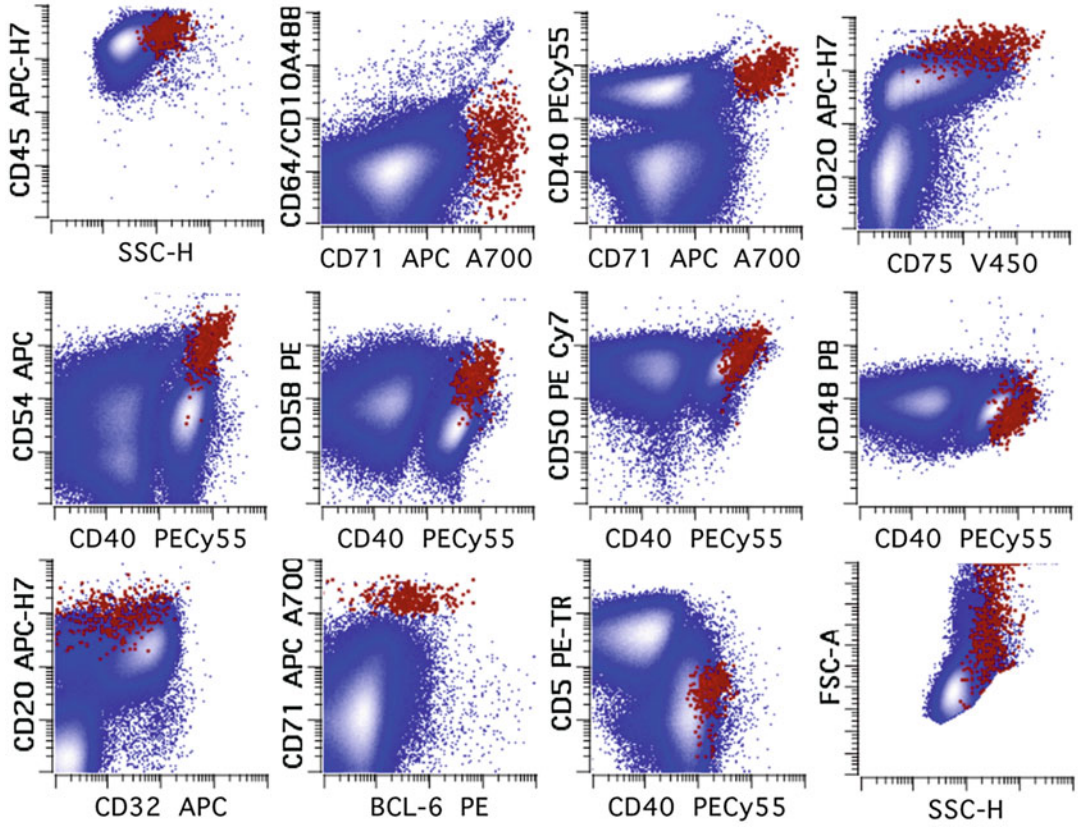


Fig. 11 Example of a lymph node involved with NLPHL characterized by flow cytometry. The large (as determined by increased forward and side light scatter compared to the small lymphocytes) neoplastic LP cells express CD20 (bright), CD40 (intermediate to bright), CD48 (low to absent), CD50 (intermediate to bright), CD54 (intermediate to bright), CD58 (intermediate), CD71 (bright), CD75 (variable), and BCL-6 (low to intermediate) without CD5, CD10, or CD64. The immunophenotype identifies a distinct population in multidimensional space with a germinal center immunophenotype. The putative neoplastic LP population is colored in red; all other events are in blue. All events are shown in all dot plots. Reprinted and modified from [40] with permission from Elsevier

normal germinal center B cells) is weakly or not expressed on LP cells. As CD32 is a low-affinity Fc receptor and its expression provides a mechanism of rituximab resistance [47], our prior flow cytometry studies suggest rituximab will be useful in NLPHL. An example of a case of NLPHL characterized by flow cytometry is shown in Fig. 11.

3.2.5 Contribution of Reactive T Cells to the Diagnosis of CHL, NLPHL, and T-Cell/Histiocyte-Rich Large B-Cell Lymphoma

CHL, NLPHL, and T-cell/histiocyte-rich large B-cell lymphoma (THRLBCL) are related lymphomas where the reactive infiltrate predominates and neoplastic cells are correspondingly rare. The presence of a characteristic, reactive $CD4^+$ T-cell population with increased expression of CD45 and CD7 (Fig. 12, **Note 16**) may suggest the diagnosis of CHL [48, 49] or THRLBCL [50] in the

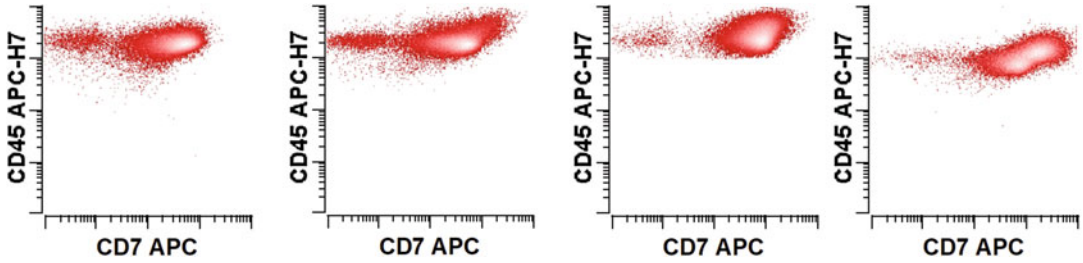


Fig. 12 Reactive T cells in CHL. $CD4^+$ T cells from CHL demonstrate increased expression of CD45 and CD7 when compared to reactive lymph nodes. First dot plot, typical reactive lymph node; second to fourth dot plots, three examples of CHL cases from three different patients. Only $CD4^+$ T cells are shown. The color version of this figure is available in the electronic version of this manuscript

appropriate clinical and histologic context as this population is present in 76.5% and 92.3% or CHL and THRLBCL, respectively, but only in 8.3% NLPHL and 4.1% of reactive cases; therefore, the presence of this population can suggest a diagnosis of CHL or THRLBCL. As T cells predominate and reactive B cells are rare in the infiltrate of THRLBCL, the presence of this population with less than 7–10% T cells, when there is no immunophenotype evidence of a NHL (in the flow cytometry assays), suggests a diagnosis of THRLBCL over CHL [50] (*see* **Note 17**). Other workers have shown that $CD4^+$ T cells that co-express low-level CD8 can suggest a diagnosis of NLPHL [51, 52].

4 Notes

1. Some B-cell populations may demonstrate aberrant loss of surface light chains. Evaluation of cytoplasmic light chains may show monotypic restriction in these cases. Aside from hematogones (normal immature B cells in the marrow and rarely in lymphoid tissues) and plasma cells, absence of surface light-chain expression usually is considered aberrant antigen expression [53]. Note, however, that some normal B-cell populations may exhibit decreased light-chain expression (e.g., germinal center B cells [54]), and consequently this finding should be interpreted cautiously.
2. Some large B-cell lymphomas may be missed if analyzed using the routine gating strategy due to increased light scatter. In every case, it is important to consider flow cytometry detected events that fall outside the normal expected size range for typical small lymphoid cells based on light scatter properties. Evaluating events with increased scatter properties is helpful to ensure complete sampling has been achieved.

3. Some B-cell lymphomas may show absent or decreased expression of CD19 [55–57] and/or CD20 [55] and may therefore be inadvertently excluded from analysis using standard gating strategies. To avoid this pitfall, evaluate more than one B-cell antigen, and examine their expression on all lymphocytes gated by light scatter.
4. Not all clonal germinal center B-cell populations are neoplastic [58]. This not uncommonly occurs in reactive states, particularly Hashimoto's thyroiditis [59]. Correlation with other immunophenotypic abnormalities and histologic studies is required to make a determination of malignancy.
5. Not all neoplastic populations that express CD19 represent B-cell lymphomas. B-lymphoblastic leukemia/lymphoma (B-ALL) and, rarely, myeloid neoplasms (namely, myeloid neoplasms with t(8;21) [60]) may show expression of CD19. Additional studies are required to completely evaluate these neoplasms (discussion beyond the scope of this paper).
6. Occasionally, dying cells and debris may appear to express B-cell antigens and appear clonal. While the differentiation of debris and authentic abnormal B-cell populations can be challenging, as a general rule, debris binds antibodies non-specifically, often resulting in a correlated (diagonal) relationship between antigens in multiple two-dimensional projections of the immunophenotypic data. The exclusion of low forward scatter events, e.g., a viability gate, may significantly reduce apparent non-specific binding and simplify analysis. However, neoplasms with a high proliferative rate may show preferential degeneration with loss of forward scatter, e.g., Burkitt lymphoma, so examination of events prior to exclusion by viability gating is important to detect this occurrence.
7. Small clonal B-cell populations that do not represent B-cell lymphoma are relatively common in normal individuals and have been described in the literature [61–63]. Ultimately, flow cytometric evaluation in the context of clinical findings, peripheral blood counts, and tissue morphology is required for the diagnosis of B-cell lymphoma.
8. As with B cells, debris can mimic an authentic abnormal T-cell population. *See* **Note 6** for B cells for evaluation of these phenomena.
9. Some large T-cell lymphomas may be missed if analyzed using the routine gating strategy due to increased light scatter. In every case, it is important to consider flow cytometric events that fall outside the normal expected size range for typical small lymphoid cells based on light scatter properties. Evaluating events with increased scatter properties is helpful to ensure complete sampling has been achieved.

10. Not all clonal and immunophenotypically aberrant T-cell populations represent T-cell lymphomas [2]. In some situations, clonal populations may arise as part of the normal process of the adaptive immune system, such as in response to cytomegalovirus [64].
11. Some T-cell lymphomas may show aberrant expression of CD19 or CD20 [65, 66]. This should not be inadvertently interpreted as a B-cell neoplasm. Correlation with expression of other T-cell antigens (cytoplasmic CD3, CD7, CD4/CD8) and absence of expression of other B-cell antigens (cytoplasmic CD79a, CD19) is recommended. Careful evaluation of the morphologic section studies may be required. Additionally, B-cell and T-cell clonality studies by PCR may be useful.
12. Putative HRS cell populations should be identifiable using the aforementioned gating strategy. To support this diagnosis, one should attempt to identify the presence of a diagonal relationship on a plot of CD45 versus CD3 or CD5. This diagonal relationship occurs due to the presence of varying numbers of T cells rosetting the HRS cells.
13. There may be slight variance to the typical HRS cell immunophenotype. For instance, while most HRS cells do not express CD20, low-level expression may occasionally be observed. Further, some HRS cell populations may lack expression of CD15, a finding that correlates with reported immunohistochemical studies (~20% of CHL case lack expression of CD15 [67–69]).
14. In clinical samples, there may be populations of cells with an apparent immunophenotype that has some resemblance to HRS cells. However, since the immunophenotype does not meet the four basic criteria outlined in Table 4, these events are almost always shown to be a different type of hematologic neoplasm or cellular debris. For example, anaplastic large T-cell lymphoma cells will commonly express CD30 with increased scatter properties. However, its expression of CD40, if present, is usually at a lower level than reactive, background B cells, as also observed in immunohistochemical studies [41, 42]. DLBCL, not otherwise specified, may occasionally show expression of CD30; however, the level of expression of CD30 in DLBCL (if present) is relatively lower than that seen for a typical HRS cell population. Further, DLBCL cells will have intermediate to bright expression of CD20 (not dim as compared to HRS cells) with a slight increase in side scatter properties (but not as increased in CHL). Lastly and as described above, LP cells of NLPHL express CD20, CD40, and CD45 without CD30 or CD15 [37]. Although these cells cannot yet be reliably identified by the flow cytometry assay for

CHL, the expression of CD20 without CD30 would argue against this population from being a putative HRS cell population. *See* Table 1 and the above discussion for combinations that can be used to immunophenotype NLPHL.

15. It is critical to note that HRS cells may occasionally be identified by flow cytometry in patients with non-Hodgkin lymphomas, such as peripheral T-cell lymphoma [70] and CLL [71–73]. In these instances, a concurrent abnormal non-Hodgkin B- or T-cell population can usually also be identified by flow cytometry. Accordingly, the determination of the presence of a HRS cell population should be considered in the context of all the flow cytometry data. Ultimately, the diagnosis of lymphoma requires the integration of the all the clinical, morphologic, and immunophenotypic data.
16. In addition to the increased expression of CD45 and CD7 on $CD4^+$ T cells (Fig. 12) [48, 49], we have noted increased expression of CD2 and CD5 on the $CD4^+$ T cells, increased CD5 and CD45 on the $CD8^+$ T cells, and decreased expression of CD3 on the $CD4^+$ and $CD8^+$ T cells in CHL.
17. The above $CD4^+$ T-cell population with increased CD45 and CD7 can be seen in some NHL, most notably B-cell lymphomas of germinal center origin and rarely in T-cell lymphomas (mainly PTCL, NOS) [50]. Therefore, this population should only be used to suggest a diagnosis of CHL or THRLBCL if evidence of B- or T-NHL is *not* present on routine NHL immunophenotyping. The presence of this population may also suggest adding the specific combination for CHL, described above.

Acknowledgments

The authors thank the medical technologists in the hematopathology laboratory at the University of Washington for their expert technical assistance.

References

1. Jaffe ES, Campo E, Harris NL, Pileri SA, Stein H, Swerdlow SH (2017) Introduction and overview of the classification of lymphoid neoplasms. In: Swerdlow SH, Campos E, Harris NL, Jaffe ES, Pileri SA, Stein H, Thiele J (eds) WHO classification of tumours of haematopoietic and lymphoid tissues. World Health Organization classification of tumors. IARC Press, Lyon, France, pp 190–198
2. Craig FE, Foon KA (2008) Flow cytometric immunophenotyping for hematologic neoplasms. Blood 111(8):3941–3967. https://doi.org/10.1182/blood-2007-11-120535
3. Wood BL, Borowitz MJ (2017) The flow cytometric evaluation of hematopoietic neoplasia. In: RA MP, Pincus MR (eds) Henry's clinical diagnosis and management by laboratory methods. Elsevier, St. Louis, Missouri, pp 659–679

4. Givan AL (2011) Flow cytometry: an introduction. Methods Mol Biol 699:1–29. https://doi.org/10.1007/978-1-61737-950-5_1
5. Redelman D (2000) Flow cytometric analyses of cell phenotypes. In: Stewart CC, Nicholson, J.K.A. (ed) Immunophenotyping. Wiley-Liss, New York,
6. Keeney M, Hedley BD, Chin-Yee IH (2017) Flow cytometry-recognizing unusual populations in leukemia and lymphoma diagnosis. Int J Lab Hematol 39 Suppl 1:86–92. https://doi.org/10.1111/ijlh.12666
7. Gorczyca W, Weisberger J, Liu Z, Tsang P, Hossein M, Wu CD, Dong H, Wong JY, Tugulea S, Dee S, Melamed MR, Darzynkiewicz Z (2002) An approach to diagnosis of T-cell lymphoproliferative disorders by flow cytometry. Cytometry 50(3):177–190. https://doi.org/10.1002/cyto.10003
8. Jamal S, Picker LJ, Aquino DB, McKenna RW, Dawson DB, Kroft SH (2001) Immunophenotypic analysis of peripheral T-cell neoplasms. A multiparameter flow cytometric approach. Am J Clin Pathol 116(4):512–526. https://doi.org/10.1309/QF6N-VAQW-N74H-4JE2
9. Beck RC, Stahl S, O'Keefe CL, Maciejewski JP, Theil KS, Hsi ED (2003) Detection of mature T-cell leukemias by flow cytometry using anti-T-cell receptor V beta antibodies. Am J Clin Pathol 120(5):785–794. https://doi.org/10.1309/835B-04QX-GNNF-NRJU
10. Langerak AW, van Den Beemd R, Wolvers-Tettero IL, Boor PP, van Lochem EG, Hooijkaas H, van Dongen JJ (2001) Molecular and flow cytometric analysis of the Vbeta repertoire for clonality assessment in mature TCRalphabeta T-cell proliferations. Blood 98 (1):165–173
11. Morice WG, Kimlinger T, Katzmann JA, Lust JA, Heimgartner PJ, Halling KC, Hanson CA (2004) Flow cytometric assessment of TCR-Vbeta expression in the evaluation of peripheral blood involvement by T-cell lymphoproliferative disorders: a comparison with conventional T-cell immunophenotyping and molecular genetic techniques. Am J Clin Pathol 121(3):373–383. https://doi.org/10.1309/3A32-DTVM-H640-M2QA
12. Tembhare P, Yuan CM, Xi L, Morris JC, Liewehr D, Venzon D, Janik JE, Raffeld M, Stetler-Stevenson M (2011) Flow cytometric immunophenotypic assessment of T-cell clonality by Vbeta repertoire analysis: detection of T-cell clonality at diagnosis and monitoring of minimal residual disease following therapy. Am J Clin Pathol 135(6):890–900. https://doi.org/10.1309/AJCPV2D1DDSGJDBW
13. Gibson JF, Huang J, Liu KJ, Carlson KR, Foss F, Choi J, Edelson R, Hussong JW, Mohl R, Hill S, Girardi M (2016) Cutaneous T-cell lymphoma (CTCL): current practices in blood assessment and the utility of T-cell receptor (TCR)-Vbeta chain restriction. J Am Acad Dermatol 74(5):870–877. https://doi.org/10.1016/j.jaad.2015.12.018
14. Wu D, Anderson MM, Othus M, Wood BL (2016) Clinical experience with modified, single-tube T-cell receptor Vbeta flow cytometry analysis for T-cell clonality. Am J Clin Pathol 145(4):467–485. https://doi.org/10.1093/ajcp/aqw015
15. Karube K, Aoki R, Nomura Y, Yamamoto K, Shimizu K, Yoshida S, Komatani H, Sugita Y, Ohshima K (2008) Usefulness of flow cytometry for differential diagnosis of precursor and peripheral T-cell and NK-cell lymphomas: analysis of 490 cases. Pathol Int 58(2):89–97. https://doi.org/10.1111/j.1440-1827.2007.02195.x
16. de Mel S, Li JB, Abid MB, Tang T, Tay HM, Ting WC, Poon LM, Chung TH, Mow B, Tso A, Ong KH, Chng WJ, Liu TC (2018) The utility of flow cytometry in differentiating NK/T cell lymphoma from indolent and reactive NK cell proliferations. Cytometry B Clin Cytom 94(1):159–168. https://doi.org/10.1002/cyto.b.21529
17. Chan WC (2001) The Reed-Sternberg cell in classical Hodgkin's disease. Hematol Oncol 19 (1):1–17
18. Stein H, Hummel M (1999) Cellular origin and clonality of classic Hodgkin's lymphoma: immunophenotypic and molecular studies. Semin Hematol 36(3):233–241
19. Marafioti T, Hummel M, Foss HD, Laumen H, Korbjuhn P, Anagnostopoulos I, Lammert H, Demel G, Theil J, Wirth T, Stein H (2000) Hodgkin and Reed-Sternberg cells represent an expansion of a single clone originating from a germinal center B-cell with functional immunoglobulin gene rearrangements but defective immunoglobulin transcription. Blood 95(4):1443–1450
20. Stein H (1999) Diagnosis of Hodgkin's disease, Hodgkin's like anaplastic large cell lymphoma, and T cell/histiocyte-rich B cell lymphoma. In: Mason DY, Harris NL (eds) Human lymphoma: clinical implications of the REAL classification, vol 52. Springer, London, pp 0–4
21. Stein H, Pileri SA, Weiss LM, Poppema S, Gascoyne RD, Jaffe ES (2017) Hodgkin lymphomas: introduction. In: Swerdlow SH, Campos E, Harris NL, Jaffe ES, Pileri SA, Stein H, Thiele J (eds) WHO classification of tumours of haematopoietic and lymphoid tissues. World Health Organization classification of tumors. IARC Press, Lyon, France, pp 424–430

22. Dorreen MS, Habeshaw JA, Stansfeld AG, Wrigley PF, Lister TA (1984) Characteristics of Sternberg-Reed, and related cells in Hodgkin's disease: an immunohistological study. Br J Cancer 49(4):465–476
23. Kadin ME, Newcom SR, Gold SB, Stites DP (1974) Letter: origin of Hodgkin's cell. Lancet 2(7873):167–168
24. Payne SV, Jones DB, Wright DH (1977) Reed-Sternberg-cell/lymphocyte interaction. Lancet 2(8041):768–769
25. Payne SV, Newell DG, Jones DB, Wright DH (1980) The Reed-Sternberg cell/lymphocyte interaction: ultrastructure and characteristics of binding. Am J Pathol 100(1):7–24
26. Sanders ME, Makgoba MW, Sussman EH, Luce GE, Cossman J, Shaw S (1988) Molecular pathways of adhesion in spontaneous rosetting of T-lymphocytes to the Hodgkin's cell line L428. Cancer Res 48(1):37–40
27. Stuart AE, Williams AR, Habeshaw JA (1977) Rosetting and other reactions of the Reed-Sternberg cell. J Pathol 122(2):81–90
28. Fromm JR, Kussick SJ, Wood BL (2006) Identification and purification of classical Hodgkin cells from lymph nodes by flow cytometry and flow cytometric cell sorting. Am J Clin Pathol 126(5):764–780. https://doi.org/10.1309/7371-XK6F-6P74-74XX
29. Harris NL (1999) Hodgkin's disease: classification and differential diagnosis. Mod Pathol 12 (2):159–175
30. Kuppers R, Engert A, Hansmann ML (2012) Hodgkin lymphoma. J Clin Invest 122 (10):3439–3447. https://doi.org/10.1172/JCI61245
31. Mathas S, Hartmann S, Kuppers R (2016) Hodgkin lymphoma: pathology and biology. Semin Hematol 53(3):139–147. https://doi.org/10.1053/j.seminhematol.2016.05.007
32. Fromm JR, Thomas A, Wood BL (2009) Flow cytometry can diagnose classical Hodgkin lymphoma in lymph nodes with high sensitivity and specificity. Am J Clin Pathol 131 (3):322–332. https://doi.org/10.1309/AJCPW3UN9DYLDSPB
33. Fromm JR, Wood BL (2012) Strategies for immunophenotyping and purifying classical Hodgkin lymphoma cells from lymph nodes by flow cytometry and flow cytometric cell sorting. Methods 57(3):368–375. https://doi.org/10.1016/j.ymeth.2012.03.028
34. Fromm JR, Wood BL (2014) A six-color flow cytometry assay for immunophenotyping classical Hodgkin lymphoma in lymph nodes. Am J Clin Pathol 141(3):388–396. https://doi.org/10.1309/AJCP0Q1SVOXBHMAM
35. Cherian S, Fromm JR (2018) Evaluation of primary mediastinal large B cell lymphoma by flow cytometry. Cytometry B Clin Cytom 94 (3):459–467. https://doi.org/10.1002/cyto.b.21544
36. Schmitz R, Stanelle J, Hansmann ML, Kuppers R (2009) Pathogenesis of classical and lymphocyte-predominant Hodgkin lymphoma. Annu Rev Pathol 4:151–174. https://doi.org/10.1146/annurev.pathol.4.110807.092209
37. Stein H, Swerdlow SH, Gascoyne RD, Poppema S, Jaffe ES, Pileri SA (2017) Nodular lymphocyte predominant Hodgkin lymphoma. In: Swerdlow SH, Campos E, Harris NL, Jaffe ES, Pileri SA, Stein H, Thiele J (eds) WHO classification of tumours of haematopoietic and lymphoid tissues. IARC Press, Lyon, France, pp 431–434
38. Uherova P, Valdez R, Ross CW, Schnitzer B, Finn WG (2003) Nodular lymphocyte predominant Hodgkin lymphoma. An immunophenotypic reappraisal based on a single-institution experience. Am J Clin Pathol 119 (2):192–198. https://doi.org/10.1309/38RK-238F-CDCH-5R22
39. Harris NL (1999) Hodgkin's lymphomas: classification, diagnosis, and grading. Semin Hematol 36(3):220–232
40. Fromm JR, Thomas A, Wood BL (2017) Characterization and purification of neoplastic cells of nodular lymphocyte predominant Hodgkin lymphoma from lymph nodes by flow cytometry and flow cytometric cell sorting. Am J Pathol 187(2):304–317. https://doi.org/10.1016/j.ajpath.2016.10.007
41. Carbone A, Gloghini A, Pinto A (1996) CD40: a sensitive marker of Reed-Sternberg cells. Blood 87(11):4918–4919
42. Carbone A, Gloghini A, Gattei V, Aldinucci D, Degan M, De Paoli P, Zagonel V, Pinto A (1995) Expression of functional CD40 antigen on Reed-Sternberg cells and Hodgkin's disease cell lines. Blood 85(3):780–789
43. Schmid C, Sargent C, Isaacson PG (1991) L and H cells of nodular lymphocyte predominant Hodgkin's disease show immunoglobulin light-chain restriction. Am J Pathol 139 (6):1281–1289
44. Stoler MH, Nichols GE, Symbula M, Weiss LM (1995) Lymphocyte predominance Hodgkin's disease. Evidence for a kappa light chain-restricted monotypic B-cell neoplasm. Am J Pathol 146(4):812–818
45. Torlakovic E, Torlakovic G (2002) B-cell markers in lymphocyte predominance Hodgkin disease. Arch Pathol Lab Med 126

(7):862–863. https://doi.org/10.1043/0003-9985(2002)126<0862:BCMILP>2.0.CO;2

46. Gruss HJ, Kadin ME (1996) Pathophysiology of Hodgkin's disease: functional and molecular aspects. Baillieres Clin Haematol 9 (3):417–446
47. Lim SH, Vaughan AT, Ashton-Key M, Williams EL, Dixon SV, Chan HT, Beers SA, French RR, Cox KL, Davies AJ, Potter KN, Mockridge CI, Oscier DG, Johnson PW, Cragg MS, Glennie MJ (2011) Fc gamma receptor IIb on target B cells promotes rituximab internalization and reduces clinical efficacy. Blood 118(9):2530–2540. https://doi.org/10.1182/blood-2011-01-330357
48. Seegmiller AC, Karandikar NJ, Kroft SH, McKenna RW, Xu Y (2009) Overexpression of CD7 in classical Hodgkin lymphoma-infiltrating T lymphocytes. Cytometry B Clin Cytom 76(3):169–174. https://doi.org/10.1002/cyto.b.20459
49. Fromm JR, Thomas A, Wood BL (2010) Increased expression of T cell antigens on T cells in classical Hodgkin lymphoma. Cytometry B Clin Cytom 78(6):387–388. https://doi.org/10.1002/cyto.b.20535
50. Wu D, Thomas A, Fromm JR (2016) Reactive T cells by flow cytometry distinguish Hodgkin lymphomas from T cell/histiocyte-rich large B cell lymphoma. Cytometry B Clin Cytom 90 (5):424–432. https://doi.org/10.1002/cyto.b.21261
51. Rahemtullah A, Reichard KK, Preffer FI, Harris NL, Hasserjian RP (2006) A double-positive CD4+CD8+ T-cell population is commonly found in nodular lymphocyte predominant Hodgkin lymphoma. Am J Clin Pathol 126(5):805–814. https://doi.org/10.1309/Y8KD-32QG-RYFN-1XQX
52. Rahemtullah A, Harris NL, Dorn ME, Preffer FI, Hasserjian RP (2008) Beyond the lymphocyte predominant cell: CD4+CD8+ T-cells in nodular lymphocyte predominant Hodgkin lymphoma. Leuk Lymphoma 49 (10):1870–1878. https://doi.org/10.1080/10428190802308728
53. Li S, Eshleman JR, Borowitz MJ (2002) Lack of surface immunoglobulin light chain expression by flow cytometric immunophenotyping can help diagnose peripheral B-cell lymphoma. Am J Clin Pathol 118(2):229–234. https://doi.org/10.1309/57G0-1BNF-KB9R-L4HN
54. Jaffe ES, Harris NL, Stein H, Campo E, Pileri SA, Swerdlow SH (2008) Introduction and overview of the classification of lymphoid neoplasms. In: Swerdlow SH, Campos E, Harris NL, Jaffe ES, Pileri SA, Stein H, Thiele J, Vardiman JW (eds) WHO classification of tumours of haematopoietic and lymphoid tissues. World Health Organization classification of Tumors. IARC Press, Lyon, pp 158–166
55. Mantei K, Wood BL (2009) Flow cytometric evaluation of CD38 expression assists in distinguishing follicular hyperplasia from follicular lymphoma. Cytometry B Clin Cytom 76 (5):315–320. https://doi.org/10.1002/cyto.b.20477
56. Yang W, Agrawal N, Patel J, Edinger A, Osei E, Thut D, Powers J, Meyerson H (2005) Diminished expression of CD19 in B-cell lymphomas. Cytometry B Clin Cytom 63(1):28–35. https://doi.org/10.1002/cyto.b.20030
57. Ray S, Craig FE, Swerdlow SH (2005) Abnormal patterns of antigenic expression in follicular lymphoma: a flow cytometric study. Am J Clin Pathol 124(4):576–583. https://doi.org/10.1309/2GFKU23XA1DH38L7
58. Kussick SJ, Kalnoski M, Braziel RM, Wood BL (2004) Prominent clonal B-cell populations identified by flow cytometry in histologically reactive lymphoid proliferations. Am J Clin Pathol 121(4):464–472. https://doi.org/10.1309/4EJ8-T3R2-ERKQ-61WH
59. Chen HI, Akpolat I, Mody DR, Lopez-Terrada D, De Leon AP, Luo Y, Jorgensen J, Schwartz MR, Chang CC (2006) Restricted kappa/lambda light chain ratio by flow cytometry in germinal center B cells in Hashimoto thyroiditis. Am J Clin Pathol 125(1):42–48
60. Hurwitz CA, Raimondi SC, Head D, Krance R, Mirro J Jr, Kalwinsky DK, Ayers GD, Behm FG (1992) Distinctive immunophenotypic features of t(8;21)(q22;q22) acute myeloblastic leukemia in children. Blood 80 (12):3182–3188
61. Rawstron AC, Green MJ, Kuzmicki A, Kennedy B, Fenton JA, Evans PA, O'Connor SJ, Richards SJ, Morgan GJ, Jack AS, Hillmen P (2002) Monoclonal B lymphocytes with the characteristics of "indolent" chronic lymphocytic leukemia are present in 3.5% of adults with normal blood counts. Blood 100 (2):635–639
62. Rawstron AC, Shanafelt T, Lanasa MC, Landgren O, Hanson C, Orfao A, Hillmen P, Ghia P (2010) Different biology and clinical outcome according to the absolute numbers of clonal B-cells in monoclonal B-cell lymphocytosis (MBL). Cytometry B Clin Cytom 78 (Suppl 1):S19–S23. https://doi.org/10.1002/cyto.b.20533
63. Strati P, Shanafelt TD (2015) Monoclonal B-cell lymphocytosis and early-stage chronic lymphocytic leukemia: diagnosis, natural history, and risk stratification. Blood 126

(4):454–462. https://doi.org/10.1182/blood-2015-02-585059

64. Rodriguez-Caballero A, Garcia-Montero AC, Barcena P, Almeida J, Ruiz-Cabello F, Tabernero MD, Garrido P, Munoz-Criado S, Sandberg Y, Langerak AW, Gonzalez M, Balanzategui A, Orfao A (2008) Expanded cells in monoclonal TCR-alphabeta+/CD4+/NKa+/CD8-/+dim T-LGL lymphocytosis recognize hCMV antigens. Blood 112 (12):4609–4616. https://doi.org/10.1182/blood-2008-03-146241
65. Rahemtullah A, Longtine JA, Harris NL, Dorn M, Zembowicz A, Quintanilla-Fend L, Preffer FI, Ferry JA (2008) CD20+ T-cell lymphoma: clinicopathologic analysis of 9 cases and a review of the literature. Am J Surg Pathol 32(11):1593–1607. https://doi.org/10.1097/PAS.0b013e31817d7452
66. Rizzo K, Stetler-Stevenson M, Wilson W, Yuan CM (2009) Novel CD19 expression in a peripheral T cell lymphoma: a flow cytometry case report with morphologic correlation. Cytometry B Clin Cytom 76(2):142–149. https://doi.org/10.1002/cyto.b.20442
67. Hsu SM, Jaffe ES (1984) Leu M1 and peanut agglutinin stain the neoplastic cells of Hodgkin's disease. Am J Clin Pathol 82(1):29–32
68. Stein H, Mason DY, Gerdes J, O'Connor N, Wainscoat J, Pallesen G, Gatter K, Falini B, Delsol G, Lemke H et al (1985) The expression of the Hodgkin's disease associated antigen Ki-1 in reactive and neoplastic lymphoid tissue: evidence that Reed-Sternberg cells and histiocytic malignancies are derived from activated lymphoid cells. Blood 66(4):848–858
69. Stein H, Uchanska-Ziegler B, Gerdes J, Ziegler A, Wernet P (1982) Hodgkin and Sternberg-Reed cells contain antigens specific to late cells of granulopoiesis. Int J Cancer 29 (3):283–290
70. Quintanilla-Martinez L, Fend F, Moguel LR, Spilove L, Beaty MW, Kingma DW, Raffeld M, Jaffe ES (1999) Peripheral T-cell lymphoma with Reed-Sternberg-like cells of B-cell phenotype and genotype associated with Epstein-Barr virus infection. Am J Surg Pathol 23(10):1233–1240
71. Mao Z, Quintanilla-Martinez L, Raffeld M, Richter M, Krugmann J, Burek C, Hartmann E, Rudiger T, Jaffe ES, Muller-Hermelink HK, Ott G, Fend F, Rosenwald A (2007) IgVH mutational status and clonality analysis of Richter's transformation: diffuse large B-cell lymphoma and Hodgkin lymphoma in association with B-cell chronic lymphocytic leukemia (B-CLL) represent 2 different pathways of disease evolution. Am J Surg Pathol 31(10):1605–1614. https://doi.org/10.1097/PAS.0b013e31804bdaf8
72. Momose H, Jaffe ES, Shin SS, Chen YY, Weiss LM (1992) Chronic lymphocytic leukemia/small lymphocytic lymphoma with Reed-Sternberg-like cells and possible transformation to Hodgkin's disease. Mediation by Epstein-Barr virus. Am J Surg Pathol 16 (9):859–867
73. Ohno T, Smir BN, Weisenburger DD, Gascoyne RD, Hinrichs SD, Chan WC (1998) Origin of the Hodgkin/Reed-Sternberg cells in chronic lymphocytic leukemia with "Hodgkin's transformation". Blood 91 (5):1757–1761
74. Crespo M, Bosch F, Villamor N, Bellosillo B, Colomer D, Rozman M, Marce S, Lopez-Guillermo A, Campo E, Montserrat E (2003) ZAP-70 expression as a surrogate for immunoglobulin-variable-region mutations in chronic lymphocytic leukemia. N Engl J Med 348(18):1764–1775. https://doi.org/10.1056/NEJMoa023143
75. Damle RN, Wasil T, Fais F, Ghiotto F, Valetto A, Allen SL, Buchbinder A, Budman D, Dittmar K, Kolitz J, Lichtman SM, Schulman P, Vinciguerra VP, Rai KR, Ferrarini M, Chiorazzi N (1999) Ig V gene mutation status and CD38 expression as novel prognostic indicators in chronic lymphocytic leukemia. Blood 94(6):1840–1847

Chapter 3

Laser-Based Microdissection of Single Cells from Tissue Sections and PCR Analysis of Rearranged Immunoglobulin Genes from Isolated Normal and Malignant Human B Cells

Ralf Küppers, Markus Schneider, and Martin-Leo Hansmann

Abstract

Normal and malignant B cells carry rearranged immunoglobulin (Ig) variable region genes, which due to their practically limitless diversity represent ideal clonal markers for these cells. We describe here an approach to isolate single cells from frozen tissue sections by microdissection using a laser-based method. From the isolated cells, rearranged IgH and Igκ genes are amplified in a semi-nested PCR approach, using a collection of V gene subgroup-specific primers recognizing nearly all V genes together with primers for the J genes. By sequence analysis of V region genes from distinct cells, the clonal relationship of the B lineage cells can unequivocally be determined and related to the histological distribution of the cells. The approach is also useful to determine V, D, and J gene usage. Moreover, the presence and pattern of somatic Ig V gene mutations give valuable insight into the stage of differentiation of the B cells.

Key words B cells, B-cell lymphoma, Clonality, Hodgkin lymphoma, Micromanipulation, V gene recombination, Single-cell PCR, Somatic hypermutation, Taq DNA polymerase errors

1 Introduction

To generate a functional B-cell antigen receptor (BCR), genes coding for the variable region of the immunoglobulin (Ig) heavy and light chains have to be assembled by a somatic recombination process during early B-cell development in the bone marrow. Heavy chain V region genes are composed of three genes, called variable (V_H, IGHV), diversity (D_H, IGHD), and joining (J_H, IGHJ) (Fig. 1), whereas V regions of light chains are composed of V and J genes. There are two types of light chains, κ and λ. The human IGH locus on chromosome 14 contains about 50 functional V_H genes (depending on the haplotype), 27 D_H genes, and 6 J_H genes [1–3]. As each V_H gene can recombine with each D_H and each J_H gene, a large diversity of heavy chain V gene rearrangements is generated. This diversity is further increased by loss of nucleotides at the ends of the rearranging genes, by addition of non-germline-

Ralf Küppers (ed.), *Lymphoma: Methods and Protocols*, Methods in Molecular Biology, vol. 1956,
https://doi.org/10.1007/978-1-4939-9151-8_3,

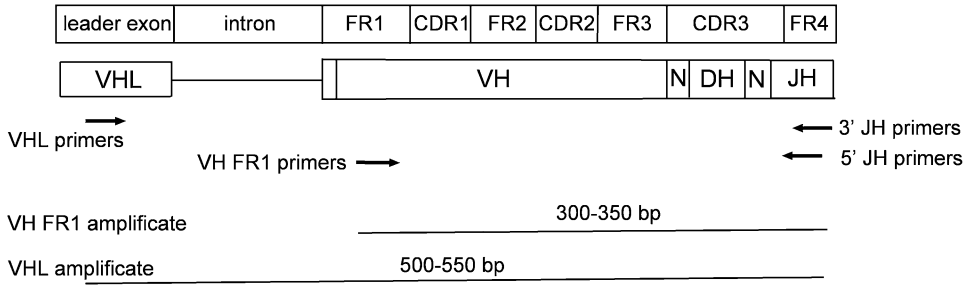


Fig. 1 Structure of the V_H region gene and location of PCR primers. The V_H region gene consists of two exons. The first encodes the most part of the V_H leader peptide; the second exon encodes the rest of the leader exon and the mature V_H region. The latter is composed of the V_H, D_H, and J_H genes. At the joining sites between V_H-D_H and D_H-J_H, one usually finds several bases of non-germline-encoded N nucleotides. The V_H region is subdivided into four relatively conserved framework regions (FR) and three diverse complementarity-determining regions (CDR). CDR1 and CDR2 are encoded by the V_H genes; CDR3 is generated in the rearrangement process at the V-D-J joining sites. The approximate location of the binding sites for V_H FR1, V_H leader, and 5′ and 3′ J_H primer mixes is indicated. With the V_H FR1 and J_H primers, amplificates of about 300–350 bp are obtained; the amplificates with the V_H leader and J_H primers are about 500–550 bp long. The length of the substructures of the V_H region gene as shown here is not to scale

encoded nucleotides (N nucleotides) between the rearranging genes and the variable opening of a hairpin structure generated as a recombination intermediate (P nucleotides). The diversity is more limited for light chain gene rearrangements, as D genes are missing. Nevertheless, a large number of functional V and J genes are also available for V-J recombination: the human Igκ (IGK) locus on chromosome 2 encompassed 30–35 functional V_κ (IGKV) genes and 5 J_κ (IGKJ) genes, and the Igλ (IGL) locus on chromosome 22 consists of 30–37 functional V_λ (IGLV) genes and 4 J_λ (IGLJ) genes [4–7]. Moreover, loss of nucleotides and the addition of N nucleotides and P nucleotides between the V and J genes occur also during light chain gene rearrangements. Hence, a large diversity can also be generated for light chain V region genes. Further diversity can be generated at later stages of B-cell differentiation, when B cells are antigen-activated and become involved in T-cell-dependent immune responses. During such immune responses, B cells establish histological structures called germinal centers (GC) in secondary lymphoid organs (e.g., lymph nodes). In proliferating GC B cells, V region genes acquire somatic mutations at a high rate through a process called somatic hypermutation [8, 9].

Due to the nearly limitless diversity of V region genes, each mature B cell is equipped with a unique BCR, which is inherited to the daughter cells when B cells proliferate. Consequently, V region genes represent ideal markers for members of a B-cell clone. This can be used to study B-cell developmental processes and the clonal relationship of B lineage cells, such as the clonal composition and intraclonal diversification in GC B cells [8]. An analysis of

rearranged Ig V genes can also be used to study malignant B cells. For example, we showed that the malignant Hodgkin and Reed-Sternberg (HRS) cells of Hodgkin lymphoma harbor clonal Ig gene rearrangements [10, 11]. This proved the B-cell origin of these cells and demonstrated that HRS cells represent monoclonal cell populations, a hallmark of tumor cells.

In many situations, PCR amplification of rearranged Ig V genes from cell populations is sufficient to identify clonal B-cell populations and obtain the Ig rearrangements of dominant normal or malignant B-cell clones. Such approaches are often applied to study B-cell non-Hodgkin lymphomas, where usually most of the B cells in the tissue belong to the lymphoma clone [12–14]. However, for a characterization of rare cells, such as the HRS cells in Hodgkin lymphoma, that usually account for only about 1% of cells in the tissue, analysis of single HRS cells is mandatory for a reliable characterization of the cells. Moreover, for a detailed analysis of B cells that are defined by their location in the histological environment, a single-cell isolation and PCR analysis are needed.

We established a PCR protocol to amplify rearranged human Ig V genes from genomic DNA of single B cells using primers binding to the 5′ end of the V genes in framework region (FR)1 combined with primers binding to the J genes in FR4 [8]. As the human V_H and V_L genes are grouped into subgroups (previously called families) based on at least 80% DNA sequence homology for subgroup members, V gene subgroup-specific primers were designed that bind to all or at least most members of a given V gene subgroup. Starting from a single-target molecule in a single cell, one round of PCR is not sufficient to obtain sufficient amplificate for direct sequence analysis. Therefore, the PCR protocol includes a second round of PCR with a small aliquot of the first-round amplificate using the same V gene primers, but J primer mixes that are located 5′ (i.e., internal) of the J primers are used in the first round of PCR (semi-nested PCR) (Fig. 2). With this strategy, PCR products of about 300–350 bp length are obtained that cover most of the V region gene. Hence, besides the determination of the clonal relationship of distinct cells, the sequence analysis of the PCR products also allows one to identify the V, (D), and J genes used in the rearrangements and analyze the sequences for somatic mutations.

We used this single-cell PCR approach for the analysis of single cells isolated from tissue sections by laser-assisted micromanipulation, as described in this chapter. However, the PCR protocol can also be used to amplify rearranged Ig V genes from cells isolated by flow cytometry [15–17].

A

```
                                    FR4                                                                 intron
Cons. GCT GAC TAC TTC GAC TTC TGG GGC CAA GGG ACC CTG GTC ACC GTC TCC TCA GGT
J1     --- --A --- --- --G CA- --- --- --G --C --- --- --- --- --- --- --- ---
J2     TAC TGG --- --- --T C-- --- --- -GT --C --- --- --- --T --- --- --- ---
J3a        --T GCT --T --T G-- --- --- --- --- --A A-- --- --- --- --T --- ---
J3b        --T GCT --T --T A-- --- --- --- --- --A A-- --- --- --- --T --- ---
J4a        --- --- --T --- -A- --- --- --- --A --- --- --- --- --- --- --- ---
J4b        --- --- --T --- -A- --- --- --G --A --- --- --- --- --- --- --- ---
J5a     AC A-- -GG --- --- -C- --- --- --- --A --- --- --- --- --- --- --- ---
J5b     AA A-- -GG --- --- CC- --- --- --G --A --- --- --- --- --- --- --- ---
J6a (TAC)4T-- GGT A-G --- G-- --- --G --- --- --- AC- --- --- --- --- --- ---
J6b (TAC)4T-- GGT A-G --- G-- --- --- --- --- --- AC- --- --- --- --- --- ---
J6c (TAC)4T--     A-G --- G-- --- --- A-- --- --- AC- --- --- --- --- --- ---

5' primers
JH1,4,5                                 --- --G --M --- --- --- --- ---
JH2                                     --- -GT --C --- --- --- --T ---
JH3                                     --- --- --- --A A-- --- --- ---
JH6                                     --- M-- --- --- AC- --- --- ---

3' primers
JH1,2,4,5                                                   --- --- --- --- --- --- --- ---
JH3                                                         --A A-- --- --- --- --T --- ---
JH6                                                         --- AC- --- --- --- --- --- ---


M: A+C
```

B

```
                    FR4                                                                 intron
       J1    G TGG ACG TTC GGC CAA GGG ACC AAG GTG GAA ATC AAA CGT GAGTAG
       J2    - -AC --T --T --- --G --- --- --- C-- --G --- --- --- A----C
       J3    A -TC --T --- --- -CT --- --- --A --- --T --- --- --- A----C
       J4    - CTC --T --- --- GG- --- --- --- --- --G --- --- --- A---GC
       J5    - ATC --C --- --- --- --- --A CGA C-- --G --T --- --- A--C-T

       3' primers
       Jk1,2,4                       -- --- --- S-- --R --- --- --- ----
       Jk3                           -- --- --A --- --T --- --- --- A----C
       Jk5                           -- --A CGA C-- --G --T --- --- A--C


       5' primers
       Jk1,2                 -- --R --- --- --- S-- --R --- --
       Jk3                   -- -CT --- --- --A --- --T --- --
       Jk4                   -- GG- --- --- --- --- --G --- --
       Jk5                   -- --- --- --A CGA C-- --G --T --


       S: G+C; R: A+G
```

Fig. 2 Sequences of human J_H and J_κ genes and of the J primer sequences. Shown are the sequences of the human J_H (**a**) and J_κ (**b**) genes and the location of the first- and second-round PCR J primers. The primer sequences are the reverse complement of the J sequences shown here (*see* Table 1). For J_H3, 4, 5, and 6, several alleles exist (denoted as a, b, c). The beginning of framework region 4 (FR4) and of the intron 3′ of the J genes is indicated. The bases upstream of FR4 belong to CDR3. The numbers in the primer designations indicate the J genes for which the primers were designed. Note, however, that the primers may also amplify other J genes of the respective Ig loci

2 Materials

2.1 Histological Stainings

1. Acetone.
2. Antibodies: for example, CD30 (Ber-H2, mouse antihuman, Agilent Dako, Hamburg, Germany), 1:100 diluted, or CD20 (L26, mouse antihuman, Agilent Dako), 1:100 diluted.
3. SuperSensitive™ Link-Label IHC Detection System, Multi-Link®—alkaline phosphatase (concentrated) (BioGenex, Fremont, CA, USA).
4. Permanent Red Substrate-Chromogen, liquid (Dako).
5. Tris-buffered saline (TBS): dissolve 8 g of NaCl, 0.2 g of KCl, and 3 g of Tris base in ca. 800 mL of H_2O, adjust pH to 7.4 with HCl, and add H_2O to 1 L.
6. TBS/BSA: TBS supplemented with 1 or 5% (w/v) bovine serum albumin.
7. Mayer's hemalaun.
8. Membrane-covered slides (Zeiss, Jena, Germany).

2.2 Laser-Assisted Microdissection

1. PALM MicroBeam (Zeiss).
2. Axiovert 200 microscope (Zeiss).
3. PCR tubes.
4. PCR buffer (without $MgCl_2$) from Expand High Fidelity PCR system (Roche, Mannheim, Germany).

2.3 Proteinase K Digestion

1. Proteinase K, recombinant, PCR grade.
2. 10 mM Tris–HCl, pH 7.6.
3. Mineral oil.

2.4 First Round of PCR

1. Expand High Fidelity PCR system (Roche).
2. dNTPs.
3. V_H/V_κ primer mix: mix same volumes of 2.5 μM working dilutions (V_H1–6, 3′ J_H primers, V_κ1–6, and 3′ J_κ primers) (Table 1).

2.5 Second Round of PCR

2.5.1 V_H PCR

1. Taq DNA polymerase (5 units/μL) with PCR buffer and $MgCl_2$.
2. dNTPs.
3. Mineral oil.
4. V_H primers: see first round of PCR, 2.5 μM working dilution.
5. J_H primer mix: mix the same volumes of 2.5 μM working dilutions of the four 5′ J_H primers (0.625 μM per 5′ J_H primer) (Table 2).

Table 1
Primers for the first round of PCR

Name	Sequence (5′–3′)
VH1	CAG-TCT-GGG-GCT-GAG-GTG-AAG-A
VH2	GTC-CTR-CGC-TGG-TGA-AAC-CCA-CAC-A
VH3	GGG-GTC-CCT-GAG-ACT-CTC-CTG-TGC-AG
VH4	GAC-CCT-GTC-CCT-CAC-CTG-CRC-TGT-C
VH5	AAA-AAG-CCC-GGG-GAG-TCT-CTG-ARG-A
VH6	ACC-TGT-GCC-ATC-TCC-GGG-GAC-AGT-G
3′JH1.2.4.5	ACC-TGA-GGA-GAC-GGT-GAC-CAG-GGT
3′JH3	ACC-TGA-AGA-GAC-GGT-GAC-CAT-TGT
3′JH6	ACC-TGA-GGA-GAC-GGT-GAC-CGT-GGT
Vk1	GAC-ATC-CRG-WTG-ACC-CAG-TCT-CCW-TC
Vk2	CAG-WCT-CCA-CTC-TCC-CTG-YCC-GTC-A
Vk3	TTG-TGW-TGA-CRC-AGT-CTC-CAG-SCA-CC
Vk4	AGA-CTC-CCT-GGC-TGT-GTC-TCT-GGG-C
Vk5	CAG-TCT-CCA-GCA-TTC-ATG-TCA-GCG-A
Vk6	TTT-CAG-TCT-GTG-ACT-CCA-AAG-GAG-AA
3′Jk1.2.4	ACT-CAC-GTT-TGA-TYT-CCA-SCT-TGG-TCC
3′Jk3	GTA-CTT-ACG-TTT-GAT-ATC-CAC-TTT-GGT-CC
3′Jk5	GCT-TAC-GTT-TAA-TCT-CCA-GTC-GTG-TCC

R, A &G; W, A&T; Y, C&T; S, G&C
The numbers in the V primer designations indicate the respective V gene subgroups. The numbers in the J primers indicate that J genes are preferentially amplified by these primers

Table 2
Sequences of J_H and $J_κ$ primers for the second round of PCR

Name	Sequence (5′–3′)
5′JH1.4.5	GAC-GGT-GAC-CAG-GGT-KCC-CTG-GCC
5′JH2	GAC-AGT-GAC-CAG-GGT-GCC-ACG-GCC
5′JH3	GAC-GGT-GAC-CAT-TGT-CCC-TTG-GCC
5′JH6	GAC-GGT-GAC-CGT-GGT-CCC-TTK-GCC
5′Jk1.2	TTG-ATY-TCC-ASC-TTG-GTC-CCY-TGG-C
5′Jk3	TTG-ATA-TCC-ACT-TTG-GTC-CCA-GGG-C
5′Jk4	TTG-ATC-TCC-ACC-TTG-GTC-CCT-CCG-C
5′Jk5	TTA-ATC-TCC-AGT-CGT-GTC-CCT-TGG-C

K, G&T; Y, C&T; S, G&C

2.5.2 V_κ PCR

1. Taq DNA polymerase (5 units/μL) with PCR buffer and $MgCl_2$.
2. dNTPs.
3. Mineral oil.
4. V_κ primers: see first round of PCR, 2.5 μM working dilution.
5. J_κ primer mix: 2.5 μM per 5′ J_κ primer (Table 2).

2.6 PCR Product Purification and Direct Sequence Analysis

1. Agarose.
2. GelRed Nucleic Acid stain, 10,000× in water (Biotium, Hayward, CA, USA).
3. 3 M Na-acetate, pH 5–5.5.
4. Ethanol 100%, 70%.
5. TAE: 40 mM Tris-acetate, 1 mM EDTA.
6. Double pure extraction kit (Analytik Jena, Jena, Germany).
7. DNA mass ladder.
8. BigDye sequencing kit (Thermo Scientific, Darmstadt, Germany).
9. Na-acetate/dextranblue: 3 M Na-acetate (pH 5–5.5), 12.5 mg/mL dextranblue.

3 Methods

3.1 Histological Staining

Dry frozen sections overnight on membrane-covered slides. Fix in acetone for 10 min. Dry sections for 30–60 min. Incubate for 15 min with TBS/3% BSA. Dilute primary antibody in TBS/3% BSA. Incubate sections with primary antibody for 30 min at room temperature. Wash three times with TBS followed by 5 min washing with TBS on a rocker. Dilute MultiLink antibody (a biotinylated secondary bridge antibody) 1:20 in TBS/3% BSA. Incubate for 20 min at room temperature. Wash three times with TBS followed by 5 min washing with TBS on a rocker. Dilute peroxidase-labeled streptavidin 1:20 in TBS/3% BSA. Incubate for 20 min at room temperature. Wash three times with TBS followed by 5 min washing with TBS on a rocker. Incubate for 1–15 min with Permanent Red prepared according to the instructions of the supplier until staining intensity controlled under the microscope is as desired. Stop reaction by washing with distilled water. Counterstain with Mayer's hemalaun. Dry sections for 2 h on a heating plate at 37 °C (*see* **Note 1**).

3.2 Microdissection

1. The stained tissue sections are mounted on the PALM Stage II of the PALM Robot MicroBeam system.
2. Navigation through the tissue section is done with the PALM RoboSoftware which displays the microscope view onto the

monitor screen. A magnification of 40× is used for microdissection.

3. The single cells or group of cells are selected and marked for microdissection by using the pencil tool of the software. With this pencil the cell or area to be microdissected is surrounded and thereby marked (*see* **Note 2**).
4. For microdissection and catapulting of the cells, 20 μL of 1× Expand High Fidelity PCR buffer (without $MgCl_2$) is added into the cap of a thin-walled PCR tube, and this is loaded into the tube holder of the CapMover, which moves the cap into the appropriate position for microdissection right above the tissue section (*see* **Note 3**).
5. Select RoboLPC as cutting mode and start the laser-based microdissection. In RoboLPC the cell is first cut out along the drawn path and then catapulted into the cap of the mounted tube.
6. The PCR tubes are centrifuged at full speed for 1 min to spin down the microdissected cells and are stored at −20 °C or −80 °C.
7. Before and after micromanipulation, a photograph is taken to allow the exact localization of the microdissected cell within the histological microenvironment.
8. In order to monitor possible contamination by DNA from fragmented neighboring cells, every fifth to tenth cell should be followed by one sample which contains only membrane from the slide but no parts of the tissue section (these samples are used as negative controls for microdissection). Microdissection can be carried out for about 24 h without a detectable decrease of PCR efficiency.

3.3 Proteinase K Digestion

Dilute proteinase K 1:1 with 10 mM Tris–HCl, pH 7.6. Add 0.5 μL of this enzyme dilution to the 20 μL PCR buffer containing the single cell, and overlay with mineral oil. Incubate for 2 h at 50 °C. Inactivate the enzyme by incubation at 95 °C for 10 min. Cool down to 4 °C.

3.4 First Round of PCR

Prepare a master mix consisting of 2.5 μL dNTP solution (2 mM), 3 μL 10× Expand High Fidelity PCR system buffer without $MgCl_2$ buffer, 5 μL primer mix, 5 μL 25 mM $MgCl_2$, and 13.5 μL H_2O per reaction. Add 29 μL master mix to the reaction mixture from the proteinase K digestion (20.5 μL). Add 0.5 μL Expand High Fidelity Enzyme Mix after the first denaturation (*see* below), and carry out PCR. The PCR program consists of 95 °C 2 min, 65 °C pause (add enzyme), 72 °C 1 min, 34 × (95 °C 50 s, 59 °C 30 s, 72 °C 60 s), 72 °C 5 min, 10 °C pause (*see* **Notes 4**–**7**).

3.5 Second Round of PCR

3.5.1 V_H PCR

Prepare on ice six master mixes (one for each of the six V_H subgroup-specific primers) each consisting of 5 μL dNTPs (2 mM), 5 μL 10× PCR buffer, 2.5 μL of the respective V_H primer, 2.5 μL 5′ J_H primer mix, 3 μL 25 mM $MgCl_2$ (5 μL for V_H1), 30.7 μL H_2O (28.7 μL for V_H1), and 0.3 μL *Taq* DNA polymerase. Pipette 49 μL master mix in fresh PCR tube. Overlay with mineral oil (if a PCR machine with heated lid is available, it is preferable to use such a machine and omit the oil). To master mix add 1 μL of the first-round reaction. Start PCR program, but put PCR tubes from ice to heating block only after block has reached 95 °C. The program consists of 95 °C 1 min, 45 × (95 °C 50 s, 61 °C 30 s, 72 °C 60 s), 72 °C 5 min, and 10 °C pause.

3.5.2 V_κ PCR

Prepare on ice six master mixes (one for each of the six V_κ subgroup-specific primers) each consisting of 5 μL dNTPs (2 mM), 5 μL 10× PCR buffer, 2.5 μL of the respective V_κ primer, 2.5 μL J_κ primer mix, 5 μL 25 mM $MgCl_2$, 28.7 μL H_2O, and 0.3 μL *Taq* DNA polymerase. Pipette 49 μL master mix in fresh PCR tube. Overlay with mineral oil. To master mix add 1 μL of the first-round reaction. Start PCR program, but put PCR tubes from ice to heating block only after block has reached 95 °C. The program consists of 95 °C 1 min, 45 × (95 °C 50 s, 61 °C 30 s, 72 °C 60 s), 72 °C 5 min, 10 °C pause (*see* **Note 8**).

3.6 PCR Product Purification and Direct Sequence Analysis

Since primer dimers and additional bands are often observed when PCR products are analyzed by agarose gel electrophoresis, it is recommended to purify the PCR products from agarose gels before sequencing. The amount of purified product is estimated from an agarose gel using a DNA mass ladder to estimate the DNA concentration (*see* **Note 9**).

1. PCR products are first analyzed on a 2% analytical agarose gel containing GelRed Nucleic Acid stain and visualized under UV light.
2. For samples with a PCR product, the second round of PCR is repeated with two to three reactions in 50 or 100 μL volume. These reactions are then mixed and precipitated by adding 1/10 volume of 3 M Na-acetate and 2 volumes ethanol. After centrifugation at ca. 12,000 × *g* for 30 min and washing of the pellet with 70% ethanol, the pellet is dissolved in 20 μL TE and loaded on a 2% preparative agarose gel.
3. The PCR product is cut out of the gel under UV light and isolated with the double pure extraction kit, as recommended by the supplier.
4. One tenth of the DNA is run on a 2% agarose gel together with a DNA mass ladder. By comparing the strength of the PCR

product bands to the strength of the DNA fragments of the mass ladder, the DNA concentration is estimated.

5. Sequence reaction: Mix the following components on ice – 10 ng PCR product, 1.5 μL primer (2.5 μM; usually the V primer), 2 μL BigDye reaction mix plus x μL H_2O to obtain a final volume of 10 μL. Overlay with one drop of mineral oil (if a PCR machine with heated lid is available, it is preferable to use such a machine and omit the oil). Run PCR, 96 °C 2 min, 34 times (96 °C 30 s, 50 °C 15 s, 60 °C 4 min), and cool to 10 °C.
6. Precipitation (at room temperature): Add 40 μL H_2O, remove oil, transfer reaction mixture to fresh tube, and add 5 μL 3 mM Na-acetate/dextranblue and 125 μL 100% ethanol. Mix by vortexing. Directly centrifuge at ca. 12,000 × *g* for 30 min. Remove supernatant and wash the pellet with 400 μL 70% ethanol. Centrifuge at ca. 12,000 × *g* for 5 min, remove supernatant, and let the pellet air-dry. Store at 4 °C, if not run on a sequencing gel on the same day.
7. Analyze sequence reaction on an ABI sequencer.

3.7 Sequence Evaluation

V gene sequences can be evaluated for the identity of the genes used in the rearrangement and the presence of somatic mutations by comparing to GenBank entries (www.ncbi.nlm.nih.gov/blast/). However, a more convenient way is to use the IMGT database and the DNA plot software (http://www.imgt.org/IMGT_vquest/vquest) or the IgBLAST tool (http://www.ncbi.nlm.nih.gov/igblast/). With these tools, rearranged human Ig genes are analyzed for sequences homologous to V, D, and J genes, and the most homologous germline gene sequences are given (*see* **Note 10**).

4 Notes

1. While the protocol described here uses alkaline phosphatase for immunostaining, the procedure can also be performed using horseradish peroxidase as enzyme for the color reaction. Moreover, PCR works also efficiently with cells labeled by in situ hybridization, e.g., for EBER transcripts to detect Epstein-Barr virus-infected B cells [18].
2. The software allows the marking/selection of multiple objects before microdissecting them, also on different glass slides mounted on the Stage II in parallel. However, sometimes the slides slip a bit by navigating through the section, and then all drawings are not surrounding the cells anymore and have to be adjusted again. So do not mark too many cells before microdissection.

3. In general, about 200–300 cells can be microdissected in one cap in one turn. However, microdissection should not take too long as the fluid in the cap vaporizes after some time. If you need more cells, it is possible to pick about 300 cells, centrifuge down the fluid, add another 20 μL of buffer into the cap, and microdissect additional 200–300 cells. This can be repeated several times. Microdissection can also be performed dry without adding PCR buffer. Then, adhesive caps should be used. Here up to 400 cells can be picked as a cell conglomerate forms in the cap which prevents proper sticking of the additional cells to the adhesive cap.
4. The protocol described here uses V_H and V_κ subgroup-specific primers binding to the FR1 of the respective genes together with combinations of J gene primers (Fig. 1). We also established PCR strategies for the co-amplification of V_H and V_λ gene rearrangements from single cells [19] and a strategy to co-amplify V_H, V_κ, and V_λ gene rearrangements [19]. Also in these approaches, V gene subgroup-specific primers binding to sequences in the FR are used together with primers for the J genes. For the amplification of V_H gene rearrangements, V_H subgroup-specific primers binding to sequences in the leader exon may be used instead of the FR1 primers [20] (Table 3). With the V_H leader primers, amplificates of about 500–550 bp length are obtained (Fig. 1). There are usually less somatic mutations in the leader exon than in FR1. Therefore, usage of the leader primers is advantageous when highly mutated B cells are studied, because there is less chance for leader primers than for FR1 primers that they cannot bind efficiently to the V genes due to somatic mutations at their binding sites. A further advantage when using V_H leader primers is that the complete V_H gene sequence is obtained. However, when DNA quality is not optimal, the longer amplificates with V_H leader primers

Table 3
Subgroup-specific VH leader exon primers

Name	Sequence (5′–3′)
VH1L	CTC-ACC-ATG-GAC-TGG-ACC-TGG-AG
VH2L	TGC-TCC-ACR-CTC-CTG-CTR-CTG-A
VH3L	ACC-ATG-GAG-TTT-GGG-CTG-AGC-TG
VH3.2L	ACC-ATG-GAA-CTG-GGG-CTC-CGC-TG
VH4L	CTC-CTG-GTG-GCA-GCT-CCC-AGA-T
VH5L	ATC-ATG-GGG-TCA-ACC-GCC-ATC-CT

Two primers are used to cover the members of the V_H3 subgroup. No V_H leader primer was designed for the single-gene subgroup V_H6

may be amplified with lower efficiency. We also established a protocol for the amplification of V_κ gene rearrangements from single cells using V_κ subgroup-specific leader primers [21]. For a detailed characterization of the IGH and IGK loci of human B lineage cells, there are also protocols available to amplify D_HJ_H joints or fragments indicative of germline configuration of the IgH locus [22] and to detect rearrangements of the kappa-deleting element (KDE) [23].

5. Since 60% of human B cells use kappa light chains, the V_H and V_κ primer combination is the preferred one to analyze human B cells for clonality and the presence of somatic mutations. However, one has to keep in mind that most λ light chain expressing B cells carry $V_\kappa J_\kappa$ joints and that these joints are usually inactivated by rearrangements of a KDE that is located downstream of the IGKC gene [23–25]. Rearrangement of the KDE to a sequence in the Jκ-Cκ intron deletes not only the IGKC gene but also the two Igκ enhancers. However, $V_\kappa J_\kappa$ joints remain on the chromosome and can be amplified. As the Ig enhancers are needed for somatic hypermutation activity, such $V_\kappa J_\kappa$ joints remain unmutated when the other Ig loci in the cell acquire mutations [17].
6. The protocol described here allows one to co-amplify V_H and V_κ gene rearrangements from single human cells. By performing first a whole genome amplification with the DNA of a single cell, it is possible to analyze the cells in parallel for multiple genes [26]. Using this strategy, single cells can be analyzed in parallel for IgH, Igκ, and Igλ gene rearrangements and also for other features [27], such as mutations in oncogenes or tumor suppressor genes or infection by viruses [18, 28]. A similar approach to combine whole genome amplification with PCR for rearranged human Ig genes has also been described by Brezinschek and colleagues [29, 30].
7. PCR contamination is a major concern when performing single-cell PCR. Therefore, a number of precautions should be followed to obtain reliable results: pre- and post-PCR work should be performed in separate rooms. Gloves should be worn and frequently changed. Separate pipettes, reaction tubes, aerosol-resistant pipette tips, and other equipment should be dedicated only for the pre-PCR work of the first round of PCR. Tubes containing first-round reaction products should not be opened in the room where the pre-PCR work is done.
8. Since V_κ genes of the $V_\kappa 5$ and $V_\kappa 6$ subgroups are only rarely used, the V_κ analysis may be restricted to subgroups 1–4.
9. The PCR products amplified with the V gene primers binding to the FR1 of human V_H and V_κ genes and the corresponding J

gene primers are usually 300–350 bp long. However, occasionally also longer or shorter PCR products are obtained. These should not be disregarded, as they also may represent V gene rearrangements. On the one hand, one has to keep in mind that the J genes are located close to each other in the IgH and Igκ loci (200–600 bp distances, ref. 3, 4) and that J primer mixes are used that bind to all J_H and J_κ genes. Hence, longer PCR products may represent amplificates ranging from the V primer binding site to a J gene downstream of the one that is actually rearranged. If a somatic mutation in the rearranged J genes prevents successful amplification from this J gene, the product from the next downstream J gene may be the predominant one. On the other hand, somatic hypermutation not only introduces point mutations but also deletions or duplications. These events account for ca. 4% of somatic mutations in nonselected (e.g., out-of-frame) rearrangements [15]. Overall, about 40% of mutated out-of-frame rearrangements from human B cells have been found to carry deletions and/or duplications, which may range in length from single nucleotides to several hundred base pairs [15].

10. While the single-cell approach is technically demanding and laborious, it has several advantages compared with PCR analysis of cell populations: (1) PCR products can be directly sequenced without cloning. This is not only fast but also has the important advantage that errors due to *Taq* DNA polymerase mistakes do not pose a problem for the reliable identification of somatic mutations, as a "consensus" sequence is obtained. Even if a *Taq* DNA polymerase error happens in the first cycle of amplification, it should principally be present only in a quarter of the final molecules of the PCR (assuming that both strands of the target molecule are replicated in that cycle and that the mutated sequence does not have an amplification advantage in the further cycles). (2) PCR hybrid artifacts, which may represent a considerable fraction of sequences if PCR products obtained from DNA of cell populations are cloned before sequencing [31], do not play a role in single-cell PCR. (3) V_H and V_L genes can be co-amplified from a single cell, so that they can be assigned to distinct cells, which is not possible if heavy and light chain genes are amplified from cell populations. (4) The sizes of B-cell clones can be estimated when single cells from defined histological locations are analyzed. If PCR is performed with DNA from cell populations, it is difficult to discern whether identical sequences derive from a single cell or from an expanded clone.

Acknowledgments

This work was supported by the Deutsche Forschungsgemeinschaft and the Deutsche Krebshilfe. We are grateful to all present and previous members of the group and coworkers that were involved in the establishment of the protocols described here.

References

1. Cook GP, Tomlinson IM (1995) The human immunoglobulin VH repertoire. Immunol Today 16:237–242
2. Corbett SJ, Tomlinson IM, Sonnhammer EL et al (1997) Sequence of the human immunoglobulin diversity (D) segment locus: a systematic analysis provides no evidence for the use of DIR segments, inverted D segments, "minor" D segments or D-D recombination. J Mol Biol 270:587–597
3. Ravetch JV, Siebenlist U, Korsmeyer S et al (1981) Structure of the human immunoglobulin mu locus: characterization of embryonic and rearranged J and D genes. Cell 27:583–591
4. Hieter PA, Maizel JV Jr, Leder P (1982) Evolution of human immunoglobulin kappa J region genes. J Biol Chem 257:1516–1522
5. Kawasaki K, Minoshima S, Nakato E et al (1997) One-megabase sequence analysis of the human immunoglobulin lambda gene locus. Genome Res 7:250–261
6. Schäble KF, Zachau HG (1993) The variable genes of the human immunoglobulin kappa locus. Biol Chem Hoppe Seyler 374:1001–1022
7. Vasicek TJ, Leder P (1990) Structure and expression of the human immunoglobulin lambda genes. J Exp Med 172:609–620
8. Küppers R, Zhao M, Hansmann ML et al (1993) Tracing B cell development in human germinal centres by molecular analysis of single cells picked from histological sections. EMBO J 12:4955–4967
9. Rajewsky K (1996) Clonal selection and learning in the antibody system. Nature 381:751–758
10. Kanzler H, Küppers R, Hansmann ML et al (1996) Hodgkin and Reed-Sternberg cells in Hodgkin's disease represent the outgrowth of a dominant tumor clone derived from (crippled) germinal center B cells. J Exp Med 184:1495–1505
11. Küppers R, Rajewsky K, Zhao M et al (1994) Hodgkin disease: Hodgkin and Reed-Sternberg cells picked from histological sections show clonal immunoglobulin gene rearrangements and appear to be derived from B cells at various stages of development. Proc Natl Acad Sci U S A 91:10962–10966
12. Küppers R, Zhao M, Rajewsky K et al (1993) Detection of clonal B cell populations in paraffin-embedded tissues by polymerase chain reaction. Am J Pathol 143:230–239
13. McCarthy KP, Sloane JP, Wiedemann LM (1990) Rapid method for distinguishing clonal from polyclonal B cell populations in surgical biopsy specimens. J Clin Pathol 43:429–432
14. Trainor KJ, Brisco MJ, Wan JH et al (1991) Gene rearrangement in B- and T-lymphoproliferative disease detected by the polymerase chain reaction. Blood 78:192–196
15. Goossens T, Klein U, Küppers R (1998) Frequent occurrence of deletions and duplications during somatic hypermutation: implications for oncogene translocations and heavy chain disease. Proc Natl Acad Sci U S A 95:2463–2468
16. Klein U, Rajewsky K, Küppers R (1998) Human immunoglobulin (Ig)M+IgD+ peripheral blood B cells expressing the CD27 cell surface antigen carry somatically mutated variable region genes: CD27 as a general marker for somatically mutated (memory) B cells. J Exp Med 188:1679–1689
17. Goossens T, Bräuninger A, Klein U et al (2001) Receptor revision plays no major role in shaping the receptor repertoire of human memory B cells after the onset of somatic hypermutation. Eur J Immunol 31:3638–3648
18. Kurth J, Spieker T, Wustrow J et al (2000) EBV-infected B cells in infectious mononucleosis: viral strategies for spreading in the B cell compartment and establishing latency. Immunity 13:485–495
19. Bräuninger A, Küppers R, Spieker T et al (1999) Molecular analysis of single B cells from T cell-rich B-cell lymphoma shows the derivation of the tumor cells from mutating germinal center B cells and exemplifies means

by which immunoglobulin genes are modified in germinal center B cells. Blood 93:2679–2687

20. Braeuninger A, Küppers R, Strickler JG et al (1997) Hodgkin and Reed-Sternberg cells in lymphocyte predominant Hodgkin disease represent clonal populations of germinal center-derived tumor B cells. Proc Natl Acad Sci U S A 94:9337–9342
21. Müschen M, Küppers R, Spieker T et al (2001) Molecular single-cell analysis of Hodgkin- and Reed-Sternberg cells harboring unmutated immunoglobulin variable region genes. Lab Investig 81:289–295
22. Küppers R, Bräuninger A, Müschen M et al (2001) Evidence that Hodgkin and Reed-Sternberg cells in Hodgkin disease do not represent cell fusions. Blood 97:818–821
23. Bräuninger A, Goossens T, Rajewsky K et al (2001) Regulation of immunoglobulin light chain gene rearrangements during early B cell development in the human. Eur J Immunol 31:3631–3637
24. Hieter PA, Korsmeyer SJ, Waldmann TA et al (1981) Human immunoglobulin kappa light-chain genes are deleted or rearranged in lambda-producing B cells. Nature 290:368–372
25. Korsmeyer SJ, Hieter PA, Sharrow SO et al (1982) Normal human B cells display ordered light chain gene rearrangements and deletions. J Exp Med 156:975–985
26. Zhang L, Cui X, Schmitt K et al (1992) Whole genome amplification from a single cell: implications for genetic analysis. Proc Natl Acad Sci U S A 89:5847–5851
27. Kanzler H, Küppers R, Helmes S et al (2000) Hodgkin and Reed-Sternberg-like cells in B-cell chronic lymphocytic leukemia represent the outgrowth of single germinal-center B-cell-derived clones: potential precursors of Hodgkin and Reed-Sternberg cells in Hodgkin's disease. Blood 95:1023–1031
28. Müschen M, Re D, Bräuninger A et al (2000) Somatic mutations of the CD95 gene in Hodgkin and Reed-Sternberg cells. Cancer Res 60:5640–5643
29. Brezinschek HP, Brezinschek RI, Lipsky PE (1995) Analysis of the heavy chain repertoire of human peripheral B cells using single-cell polymerase chain reaction. J Immunol 155:190–202
30. Brezinschek HP, Foster SJ, Dörner T et al (1998) Pairing of variable heavy and variable kappa chains in individual naive and memory B cells. J Immunol 160:4762–4767
31. Küppers R (1997) Ongoing somatic mutation in mantle cell lymphomas questioned. Br J Haematol 97:932–934

Chapter 4

PCR GeneScan and Heteroduplex Analysis of Rearranged Immunoglobulin or T-Cell Receptor Genes for Clonality Diagnostics in Suspect Lymphoproliferations

Elke Boone, Kim C. Heezen, Patricia J. T. A. Groenen, and Anton W. Langerak and On behalf of the EuroClonality Consortium

Abstract

Assessment of the presence of clonal lymphoproliferations via polymerase chain reaction (PCR)-based analysis of rearranged immunoglobulin (IG) or T-cell receptor (TR) genes is a valuable method in the diagnosis of suspect lymphoproliferative disorders. Additionally, this methodology can be used for evaluating dissemination of lymphoma cells and for studying the clonal relationship between multiple (different locations) or consecutive (over time) lymphomas. Here we describe an integrated approach to assess clonality via analysis of Ig heavy chain (IGH), Ig kappa (IGK), TCR beta (TRB), and TCR gamma (TRG) gene rearrangements, based on the standardized multiplex PCRs as originally developed by the European BIOMED-2 consortium. The described protocol covers the pre-analytical phase of DNA isolation (from formalin-fixed paraffin-embedded and fresh tissues, body fluids, peripheral blood, and bone marrow), the analytical phase of PCR GeneScan and heteroduplex analysis, and the post-analytical interpretation of the obtained profiles, following established guidelines.

Key words BIOMED-2, Immunoglobulin, T-cell receptor, Gene rearrangements, Clonality, Heteroduplex analysis, GeneScan analysis, DNA isolation, Polyacrylamide electrophoresis, Capillary electrophoresis

1 Introduction

Cancer cells have the unique feature that they all originate from a single transformed cell and are therefore clonally related. Mature lymphoid B- or T-cells have several unique DNA sequences coding for unique antigen-receptor molecules [1, 2]. This feature is very useful for demonstrating the clonal expansion of B- or T-cells in lymphoid malignancies. B-cells are characterized by unique rearranged DNA sequences in the IGH, IGK, and/or IGL loci, whereas in T-cell rearrangements of the TRD, TRG, TRB, and/or TRA loci can occur. The unique DNA sequences are formed during lymphocyte development by a stepwise

Ralf Küppers (ed.), *Lymphoma: Methods and Protocols*, Methods in Molecular Biology, vol. 1956,
https://doi.org/10.1007/978-1-4939-9151-8_4,

rearrangement of the genes [3]. In case of the IGH, TRB, and TRD loci, a variable (V), diversity (D), and joining (J) genes are coupled to form a functional V(D)J exon [3]. IGK, IGL, TRA, and TRG gene rearrangements are the result of joining of one specific V gene to a J gene [3]. Based on sequence homology, the V and D genes, though each slightly different, can be grouped into families. However, the number of families is strongly dependent on the individual locus; for instance, the 44 functional IGHV genes cluster into 7 families, whereas the 47 functional TRBV genes can only be grouped into 23 families, of which 12 families contain just 1 member. Depending on the locus, a different number of V, D, and J genes are available for the recombination process. Therefore the so-called combinatorial repertoire of these genes can mount up to ~3×10^6 possible rearrangements. Importantly, besides functional genes, also pseudogenes are found, which cannot rearrange or do not result in a functional receptor upon rearrangement. During the joining process, random insertion and/or deletion of nucleotides occurs, resulting in highly diverse (both in size and nucleotide sequence), unique, junctional regions per cell. However, among the different IG and TCR genes, the level of junctional diversity is not identical. Particularly, IGK and IGL rearranged genes show a restricted junctional repertoire as compared to the other loci. Based on the combinatorial and junctional diversity, the total polyclonal repertoire of IG and TR molecules in healthy individuals is estimated to be $>10^{12}$.

During early lymphoid differentiation, the loci are rearranged in a sequential order. Pre-B-cells first rearrange the IGH locus, followed by the IGK locus [4]. In case the rearranged IGK gene is not functional, the gene is deleted and IGL rearrangement occurs [4]. This deletion of the IGK locus is mediated via a special rearrangement in the IGK locus involving the so-called kappa-deleting element (Kde) that can rearrange to any of the IGKV genes or to an isolated recombination signal sequence (RSS) in the IGKJ-IGKC intron [4]. Eventually, B-cells thus express either IgH/κ or IGH/λ receptors. In case of T-cells, either expression of TCRαβ or TCRγδ receptors is found on the cell surface. First the TRD locus rearranges, followed by the TRG locus [5]. In case of nonfunctionality, TRB gene rearrangement will occur followed by deletion of TRD and rearrangement of the TRA locus [5]. In a normal mature B- or T-cell repertoire, both monoallelic and biallelic gene rearrangements can be found. In the latter case, one of the rearrangements resulted in a nonfunctional receptor gene, as lymphoid cells with two functional rearrangements will not fully differentiate into mature cells. Of note, malignant, lymphoid cells sometimes escape differentiation control checkpoints, resulting in unusual rearrangement patterns within one cell.

To discriminate between a polyclonal, reactive, lymphoid repertoire and a monoclonal lymphoproliferation by PCR analysis, the

primers in the PCR reaction should be able to amplify all possible rearranged V(D)J exons and should flank the junctional region. This way, a polyclonal cell population will result in differently sized PCR products, corresponding to the presence of different V(D)J rearrangements which show a Gaussian distribution with respect to the amount of inserted or deleted nucleotides in the junctional region. In case of a monoclonal cell expansion, only PCR products of the same size and sequence will be generated.

In the past, a large number of different PCR protocols were designed, which however often suffered from a limited analytical sensitivity to detect clonality in the suspect lymphoproliferation. Due to the known sequence diversity among the different V, D, and J genes, some of the developed family and/or consensus primers showed insufficient primer annealing, resulting in false-negative data. Furthermore, in case of clonal expansions of germinal center or post-germinal center B-cells, improper annealing might occur due to the presence of mutations that occur through the process of somatic hypermutation (such somatic mutations are mostly single-nucleotide mutations, but also deletions and insertions are seen) in the rearranged IG genes, potentially also resulting in false negativity.

For this reason the BIOMED-2 Concerted Action BMH4 CT98-3936 was initiated to develop standardized PCR protocols to maximally cover all possible rearrangements (including somatically mutated IG rearrangements) for all the lymphoid gene targets. A total of 47 institutes from 7 European countries collaborated in this project and used their expertise on IG and TR rearrangement processes to develop a standardized protocol for each target gene [3], with the exception of TRA due to its high level of complexity. A total of 97 new primers were designed, covering the majority of functional genes and representing 418 single PCR reactions. After initial evaluation of all separate 418 PCR tests, careful evaluation of the combination of primers resulted in only 14 IG/TR multiplex PCR tubes: 3 for complete IGH (V-J) to circumvent possible mismatching due to somatic mutations in the V region, 2 for incomplete IGH (D-J), 2 for IGK (V-J and Kde rearrangements), 1 for IGL (V-J), 2 for complete TRB (V-J), 1 for incomplete TRB (D-J), 2 for TRG (V-J), and 1 for all TRD gene rearrangements [3].

A second major technical pitfall of PCR-based IG/TR clonality assays is the occurrence of false-positive results due to a lack of optimal discrimination between a polyclonal, oligoclonal, and monoclonal PCR product by the techniques used to visualize the obtained DNA molecules. If the obtained PCR products are visualized by normal agarose or acrylamide gel electrophoresis, both polyclonal and monoclonal PCR products will result in a band of apparently the same size and intensity. Therefore either GeneScan fragment analysis or heteroduplex analysis is required to identify the differences in size and/or sequence of the obtained PCR products

[3]. Heteroduplex analysis uses double-stranded PCR products and takes advantage of the sequence diversity in the junctional regions to discriminate between a polyclonal and monoclonal population [3]. After PCR, the double-stranded PCR products are denatured, followed by a very slow renaturation of the single strands at 4 °C. In polyclonal samples renaturation will also occur between PCR products with similar V and J genes, but the different junctional sequences thereby form typical heteroduplexes. In case of clonal PCR products, perfect double-stranded homoduplexes will be present after the denaturation/renaturation process. Heteroduplexes run on a polyacrylamide gel will result in a smear, whereas the double-stranded homoduplexes will migrate faster as a single band and at a predictable, expected product size (Fig. 1).

GeneScan fragment analysis requires PCR products which are fluorescently labeled [3]. Therefore the reverse primers in each BIOMED-2 multiplex PCR contain a 5′ coupled fluorochrome which enables detection of the PCR products with automated sequencing equipment. Following denaturation of PCR products, the single-stranded products are size-separated in denaturing polyacrylamide gels or in a capillary polymer. The labeled PCR products are visualized by automated laser scanning of the migrating molecules and capturing of the fluorescent signals by a CCD camera. Polyclonal proliferations will generate a Gaussian distribution of multiple peaks, corresponding to the differently sized PCR products, whereas monoclonal proliferations will result in a single peak due to the presence of only one type of PCR product in the reaction (Fig. 2).

Heteroduplex analysis is more laborious and less sensitive than the automated GeneScan fragment analysis. Nevertheless, as heteroduplex analysis uses both the size and sequence diversity of the junctional region, this technique is particularly useful for loci with restricted junctional size diversity. On the other hand, due to superior sensitivity of GeneScan fragment analysis and the exact identification of the size of the clonal PCR product, this method can be used for demonstrating the presence of a similar clonal proliferation in case of, e.g., suspected bone marrow invasion by lymphoma cells or for monitoring the presence of the clonal proliferation during follow-up of the patient.

Three large-scale, multicenter BIOMED-2 studies demonstrated that by using these PCR protocols, the false-negative rate for detecting clonal lymphoproliferations in five categories of B-cell malignancies (mantle cell lymphoma, MCL; chronic lymphocytic leukemia, CLL; follicular lymphoma, FL; marginal zone lymphoma, MZL; diffuse large B-cell lymphoma, DLBCL) could be reduced to less than 1% by using a combined strategy of the different IGH and IGK tubes [6]. The PCR for IGL rearrangements did not show additional value and can therefore be omitted in most screening strategies. Likewise, the combination of the three

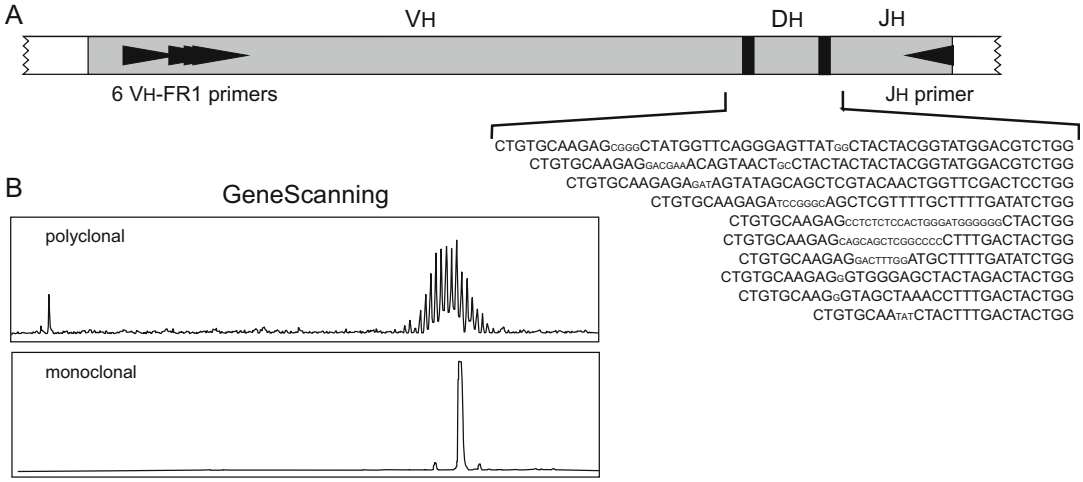


Fig. 1 GeneScan fragment analysis. (**a**) IGH rearrangements showing junctional region diversity, reflecting different levels of deletion and insertion (lower cases) of nucleotides. (**b**) Upon PCR amplification with IGHV and IGHJ primers (*see* **a**), polyclonal PCR products give rise to a normal Gaussian size distribution in GeneScan fragment analysis, whereas in case of monoclonal PCR product, one clear peak of identically sized products is seen

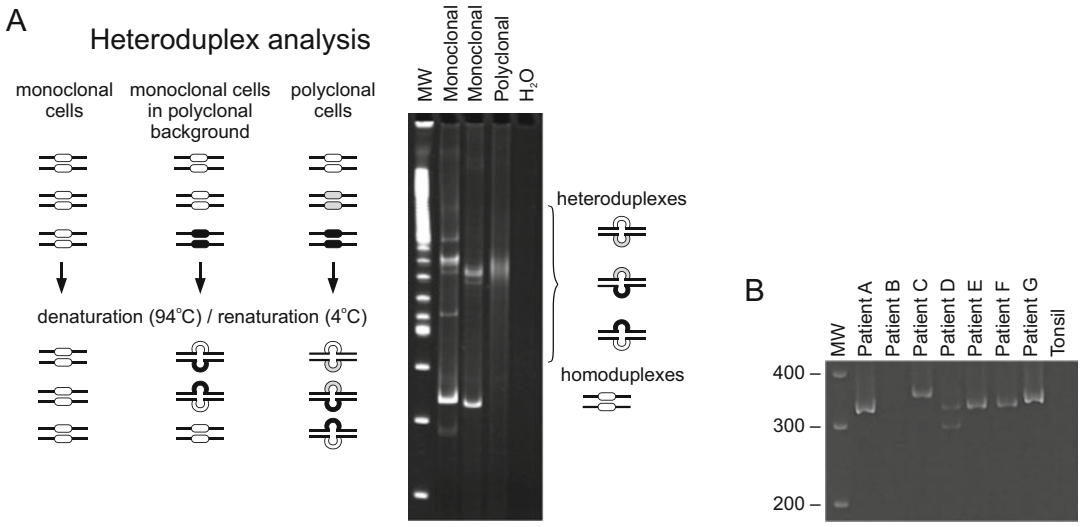


Fig. 2 Heteroduplex analysis. (**a**) Scheme of heteroduplex analysis technique, showing migration effect of homo- and heteroduplexes upon gel electrophoresis. (**b**) Example of polyacrylamide gel showing clonal homoduplexes in patients A, B (biallelic), C, and E–G. In patient B and in tonsil, no clonal PCR products are observed

TRB tubes and two TRG tubes resulted in an overall clonality detection rate of 99% in T-cell prolymphocytic leukemia (T-PLL), T-cell large granular leukemia (T-LGL), peripheral T-cell lymphoma unspecified (PTCL-U), and angioimmunoblastic T-cell lymphoma (AITL) and a rate of 94% when also the category of

anaplastic large cell lymphoma (ALCL) was included [7]. The TRD tube can be omitted, except in cases of suspected TCRγδ or immature T-cells. In the vast majority of reactive samples, IG or TR gene rearrangements appeared to be polyclonal, whereas occasionally monoclonality was seen that prompted further diagnostic work-up [8].

Finally, it is extremely important to realize the limitation of this technique with regard to the assignment of a malignant status and/or a specific cell lineage status to the analyzed sample [9]. Clinical benign clonal TCRαβ$^+$ proliferations can regularly be found in peripheral blood (PB) of elderly individuals. PB or lymph nodes of patients with acute Epstein-Barr virus (EBV) or cytomegalovirus (CMV) infections can show a restricted TCR repertoire, often resulting in an (oligo)clonal peak pattern, especially in immunodeficient patients. Benign monoclonal gammopathies often show a clonal IGH or IGK pattern. Furthermore, TR gene rearrangements occur in 10–20% of B-cell malignancies, and IG gene rearrangements occur in 5–10% of T-cell malignancies, mostly however in a single locus [6, 7]. Therefore, the identification of a clonal pattern in a single locus cannot be used as a marker for B/T-lineage assignment.

In this chapter we present a complete overview of the different steps of clonality analysis in suspected lymphoproliferations. Also the technical interpretation and reporting of the obtained patterns will be addressed. A good knowledge of the immunobiology of IG and TR rearrangements is mandatory to be able to correctly interpret the different molecular patterns. Furthermore, it is of utmost importance that molecular clonality results are eventually interpreted in the context of available clinical, morphological, and immunophenotypic data.

2 Materials

2.1 DNA Extraction

1. Xylene (molecular biology quality grade).
2. 99% ethanol, absolute pro analyze (molecular biology quality grade)
3. Tissue grinder system, e.g., Medimachine (Becton Dickinson, 340588) including CellWASH (Becton Dickinson, 349524) and Medicon 35 μm filter (Becton Dickinson, 340590).
4. Proteinase K (QIAGEN, 19131).
5. ATL buffer (QIAGEN, 19076).
6. Genomic DNA isolation kit, e.g., QIAamp DNA Blood Mini Kit (QIAGEN, 51106) or GenElute mammalian genomic DNA Miniprep Kit (Sigma, G1N70 or G1N350) (*see* **Note 1**); contains all necessary reagents for DNA extraction from PB, bone marrow (BM), or cell suspensions.

7. Centrifuge, e.g., Biofuge Pico (Heraeus) or equivalent equipment.

2.2 PCR Setup

1. PCR-grade H_2O.
2. dNTPs (10 mM): combine equal amounts of 100 mM dATP, dTTP, dCTP, and dGTP (e.g., dNTP sets, Illustra [GE HealthCare] via VWR, 28-4065-51), and dilute this dNTP mix 1/10 in PCR-grade H_2O.
3. Ampli*Taq* Gold 5 U/μL supplied with 25 mM $MgCl_2$ and 10× ABI buffer II or 10× GOLD buffer (Thermo Fisher, N8080245 or 4311820).
4. TE buffer for primers (1 mM Tris–HCl pH 8.0, 0.01 mM EDTA): combine 10 μL of 1 M Tris–HCl pH 8.0, 1 mL of 0.1 mM EDTA, and 8.99 mL PCR-grade H_2O.
5. Highly purified, lyophilized primers to be dissolved in TE buffer for primers at 100 pmol/μL (100 μM), e.g., desalted custom DNA Oligos "Invitrogen Life Sciences" quality (Thermo Fisher).
6. Primer mix for B-cell targets (IGH, IGK): combine 100 μL of each primer (100 μM), and adjust, when necessary, to 800 μL with TE buffer for primers (Table 1); alternatively, IGH and IGK unlabeled (Invivoscribe, 2-101-0010, 2-101-0020, 2-101-0030, 2-101-0040, 2-102-0010, 2-102-0020) or labeled (Invivoscribe, 2-101-0011, 2-101-0101, 2-101-0031, 2-101-0041, 2-102-0011, 2-102-0021) master mixes.
7. Primer mix for TRB targets: combine 100 μL of each primer (100 μM), and adjust, when necessary, to 3200 μL with TE buffer (Table 2); alternatively, TRB unlabeled (Invivoscribe, 2-205-0010, 2-205-0020, 2-205-0030) or labeled (Invivoscribe, 2-205-0011, 2-205-0021, 2-205-0031) master mixes.
8. Primer mix for TRG targets: combine 100 μL of each primer (100 μM) (Table 3) or TRG unlabeled (Invivoscribe, 2-207-0030, 2-207-0040) or labeled (Invivoscribe, 2-207-0031, 2-207-0041) master mixes. An alternative option is to use the novel TRG2.0 1 tube assay (Invivoscribe, 2-207-0091) (manuscript, in preparation).
9. Thermal cyclers, e.g., ABI 2700, 2720, Veriti (Thermo Fisher) or PTC-0200 (MJ Research) or equivalent equipment (*see* **Note 2**).

2.3 GeneScan Fragment Analysis

1. Genetic analyzer, e.g., 3100/3130xl/3700 (Thermo Fisher) or equivalent systems (*see* **Note 3**).
2. Capillary array, 36 cm (Thermo Fisher, 4315931) or equivalent capillary arrays.

Table 1
Composition of multiplex primer mixes for rearranged IG genes

Primer mix	Primer sequence (5′ → 3′)
IGH tube A FR1 V-J	
VH1/7-FR1	GGCCTCAGTGAAGGTCTCCTGCAAG
VH2-FR1	GTCTGGTCCTACGCTGGTGAAACCC
VH3-FR1	CTGGGGGGTCCCTGAGACTCTCCTG
VH4-FR1	CTTCGGAGACCCTGTCCCTCACCTG
VH5-FR1	CGGGGAGTCTCTGAAGATCTCCTGT
VH6-FR1	TCGCAGACCCTCTCACTCACCTGTG
JH consensus (5′ FAM)	CTTACCTGAGGAGACGGTGACC
IGH tube B FR2 V-J	
VH1-FR2	CTGGGTGCGACAGGCCCCTGGACAA
VH2-FR2	TGGATCCGTCAGCCCCCAGGGAAGG
VH3-FR2	GGTCCGCCAGGCTCCAGGGAA
VH4-FR2	TGGATCCGCCAGCCCCCAGGGAAGG
VH5-FR2	GGGTGCGCCAGATGCCCGGGAAAGG
VH6-FR2	TGGATCAGGCAGTCCCCATCGAGAG
VH7-FR2	TTGGGTGCGACAGGCCCCTGGACAA
JH consensus (5′ FAM)	CTTACCTGAGGAGACGGTGACC
IGH tube C FR3 V-J	
VH1-FR3	TGGAGCTGAGCAGCCTGAGATCTGA
VH2-FR3	CAATGACCAACATGGACCCTGTGGA
VH3-FR3	TCTGCAAATGAACAGCCTGAGAGCC
VH4-FR3	GAGCTCTGTGACCGCCGCGGACACG
VH5-FR3	CAGCACCGCCTACCTGCAGTGGAGC
VH6-FR3	GTTCTCCCTGCAGCTGAACTCTGTG
VH7-FR3	CAGCACGGCATATCTGCAGATCAG
JH consensus (5′ FAM)	CTTACCTGAGGAGACGGTGACC
IGH tube D D-J	
DH1	GGCGGAATGTGTGCAGGC
DH2	GCACTGGGCTCAGAGTCCTCT
DH3	GTGGCCCTGGGAATATAAAA
DH4	AGATCCCCAGGACGCAGCA

(continued)

Table 1
(continued)

Primer mix	Primer sequence (5′ → 3′)
DH5	CAGGGGGACACTGTGCATGT
DH6	TGACCCCAGCAAGGGAAGG
JH consensus (5′ FAM)	CTTACCTGAGGAGACGGTGACC
IGK tube A V-J	
Vκ1f/6	TCAAGGTTCAGCGGCAGTGGATCTG
Vκ2f	GGCCTCCATCTCCTGCAGGTCTAGTC
Vκ3f	CCCAGGCTCCTCATCTATGATGCATCC
Vκ4	CAACTGCAAGTCCAGCCAGAGTGTTTT
Vκ5	CCTGCAAAGCCAGCCAAGACATTGAT
Vκ7	GACCGATTTCACCCTCACAATTAATCC
Jκ1-4 (5′ FAM)	CTTACGTTTGATCTCCACCTTGGTCCC
Jκ5 (5′ FAM)	CTTACGTTTAATCTCCAGTCGTGTCCC
IGK tube B V/intron-Kde	
Vκ1f/6	TCAAGGTTCAGCGGCAGTGGATCTG
Vκ2f	GGCCTCCATCTCCTGCAGGTCTAGTC
Vκ3f	CCCAGGCTCCTCATCTATGATGCATCC
Vκ4	CAACTGCAAGTCCAGCCAGAGTGTTTT
Vκ5	CCTGCAAAGCCAGCCAAGACATTGAT
Vκ7	GACCGATTTCACCCTCACAATTAATCC
INTR	CGTGGCACCGCGAGCTGTAGAC
Kde (5′ FAM)	CCTCAGAGGTCAGAGCAGGTTGTCCTA

3. MicroAmp Optical 96-well reaction plate (Thermo Fisher, 4306737) or equivalent plates.
4. MicroAmp Caps (8 caps/strip, Thermo Fisher, N8011535) or equivalent caps/seals.
5. Size standard, e.g., GeneScan 500 ROX size standard (Thermo Fisher, 401734) or equivalent size standard.
6. Hi-Di formamide (e.g., Thermo Fisher, 4311320).
7. Performance Optimized Polymer 7 (POP7, Thermo Fisher, 4352759) or equivalent polymers.
8. PCR-grade H_2O.
9. Centrifuge, e.g., Megafuge 1.0 with plate holder (Heraeus) or equivalent equipment.

Table 2
Composition of multiplex primer mixes for rearranged TRB genes

Primer name	Primer sequence (5′ → 3′)	TRB tube A V-J	TRB tube B V-J	TRB tube C D-J
V2	AACTATGTTTTGGTATCGTCA	+	+	–
Vβ4	CACGATGTTCTGGTACCGTCAGCA	+	+	–
Vβ5/1	CAGTGTGTCCTGGTACCAACAG	+	+	–
Vβ6a/11	AACCCTTTATTGGTACCGACA	+	+	–
Vβ6b/25	ATCCCTTTTTTGGTACCAACAG	+	+	–
Vβ6c	AACCCTTTATTGGTATCAACAG	+	+	–
Vβ7	CGCTATGTATTGGTACAAGCA	+	+	–
Vβ8a	CTCCCGTTTTCTGGTACAGACAGAC	+	+	–
Vβ9	CGCTATGTATTGGTATAAACAG	+	+	–
Vβ10	TTATGTTTACTGGTATCGTAAGAAGC	+	+	–
Vβ11	CAAAATGTACTGGTATCAACAA	+	+	–
Vβ12a/3/13a/15	ATACATGTACTGGTATCGACAAGAC	+	+	–
Vβ13b	GGCCATGTACTGGTATAGACAAG	+	+	–
Vβ13c/12b/14	GTATATGTCCTGGTATCGACAAGA	+	+	–
Vβ16	TAACCTTTATTGGTATCGACGTGT	+	+	–
Vβ17	GGCCATGTACTGGTACCGACA	+	+	–
Vβ18	TCATGTTTACTGGTATCGGCAG	+	+	–
Vβ19	TTATGTTTATTGGTATCAACAGAATCA	+	+	–
Vβ20	CAACCTATACTGGTACCGACA	+	+	–
Vβ21	TACCCTTTACTGGTACCGGCAG	+	+	–
Vβ22	ATACTTCTATTGGTACAGACAAATCT	+	+	–
Vβ23/8b	CACGGTCTACTGGTACCAGCA	+	+	–
Vβ24	CGTCATGTACTGGTACCAGCA	+	+	–
Jβ1.1 (5′ HEX)	CTTACCTACAACTGTGAATCTGGTG	+	–	+
Jβ1.2 (5′ HEX)	CTTACCTACAACGGTTAACCTGGTC	+	–	+
Jβ1.3 (5′ HEX)	CTTACCTACAACAGTGAGCCAACTT	+	–	+
Jβ1.4 (5′ HEX)	CATACCCAAGACAGAGAGCTGGGTTC	+	–	+
Jβ1.5 (5′ HEX)	CTTACCTAGGATGGAGAGTCGAGTC	+	–	+
Jβ1.6 (5′ HEX)	CATACCTGTCACAGTGAGCCTG	+	–	+
Jβ2.1 (5′ FAM)	CCTTCTTACCTAGCACGGTGA	–	+	+

(continued)

Table 2
(continued)

Primer name	Primer sequence (5′ → 3′)	TRB tube A V-J	TRB tube B V-J	TRB tube C D-J
Jβ2.2 (5′ FAM)	CTTACCCAGTACGGTCAGCCT	+	−	+
Jβ2.3 (5′ FAM)	CCCGCTTACCGAGCACTGTCA	−	+	+
Jβ2.4 (5′ FAM)	CCAGCTTACCCAGCACTGAGA	−	+	+
Jβ2.5 (5′ FAM)	CGCGCACACCGAGCAC	−	+	+
Jβ2.6 (5′ FAM)	CTCGCCCAGCACGGTCAGCCT	+	−	+
Jβ2.7 (5′ FAM)	CTTACCTGTAACCGTGAGCCTG	+	−	+
Dβ1	GCCAAACAGCCTTACAAAGAC	−	−	+
Dβ2	TTTCCAAGCCCCACACAGTC	−	−	+

Table 3
Composition of multiplex primer mixes for rearranged TRG genes

Primer mix	Primer sequence (5′ → 3′)
TRG tube A V-J	
Vγfl	GGAAGGCCCCACAGCRTCTT
Vγ10	AGCATGGGTAAGACAAGCAA
Jγ1.1/2.1 (JP1/2) (5′ FAM)	TTACCAGGCGAAGTTACTATGAGC
Jγ1.3/2.3 (J1/2) (5′ HEX)	GTGTTGTTCCACTGCCAAAGAG
TRG tube B V-J	
Vγ9	CGGCACTGTCAGAAAGGAATC
Vγ11	CTTCCACTTCCACTTTGAAA
Jγ1.1/2.1 (JP1/2) (5′ FAM)	TTACCAGGCGAAGTTACTATGAGC
Jγ1.3/2.3 (J1/2) (5′ HEX)	GTGTTGTTCCACTGCCAAAGAG

2.4 Heteroduplex Analysis

1. Vertical polyacrylamide gel electrophoresis (PAGE) system (e.g., Hoefer SE600 Ruby).
2. 30% acrylamide/bisacrylamide (29:1) (e.g., Bio-Rad, 1610156).
3. 10× TBE buffer (e.g., Sigma, T4415).
4. TEMED (e.g., Bio-Rad, 1610800).

5. APS (10% w/v) (e.g., Bio-Rad, 1610700).
6. Ethidium bromide 10 mg/mL (e.g., Bio-Rad, 16110433, or Sigma, E1510).

3 Methods

3.1 Samples and Quality Controls

Clonality analysis can be performed on DNA extracted from any human tissue. The specific extraction procedure for isolating DNA depends on the sample which is received. We here describe the column-based extraction procedure of QIAGEN for formalin-fixed paraffin-embedded (FFPE) tissue, fresh tissues, cell suspensions, PB, and BM; equivalent isolation systems are also possible (*see* **Note 1**).

Quality controls are crucial for correct evaluation of the clonality results.

To monitor PCR reaction performance, three control samples are crucial (*see* **Note 4**):

- No template control (NTC).
- Polyclonal control: DNA extracted from tonsil or mononuclear PB cells.
- Clonal control: DNA extracted from cell lines or clonal patient samples.

A second important level of quality control is the quality of the tissue sample analyzed. It is important to evaluate whether the sample contains representative tissue and/or whether the quality of the obtained DNA is sufficient (*see* **Note 5**).

3.2 DNA Extraction

3.2.1 Extraction from Formalin-Fixed Paraffin-Embedded (FFPE) Tissue

1. Place two to three 10 μm sections of FFPE tissue (approximately 1 cm^2 in size) into a 1.5 mL tube (*see* **Note 5**).
2. Add 1200 μL xylene, and vortex thoroughly for 10 s (work in a protective cabinet).
3. Centrifuge 5 min, 16,200 × *g*.
4. Remove the supernatant carefully and dispense in specific waste containers.
5. Add 1200 μL 99% ethanol.
6. Centrifuge 5 min, 16,200 × *g*.
7. Remove carefully the ethanol.
8. Add again 1200 μL 99% ethanol and mix carefully by inverting the tube.
9. Centrifuge 5 min, 16,200 × *g*.
10. Remove all ethanol.

11. Dry the remaining tissue by leaving the tube open, and incubate for 10–15 min at 37 °C in a thermo block.
12. Resuspend the dried tissue in 180 µL ATL buffer + 20 µL proteinase K, and vortex thoroughly for 10 s (*see* **Note 6**).
13. Incubate overnight at 54–56 °C in a thermo block (*see* **Note 6**).
14. Continue the procedure as described in Subheading 3.2.3 from **step 3** onward.

3.2.2 Extraction from Fresh Tissues or Human Fluids

1. Make a cell suspension of the fresh tissue using a Medimachine mixer (*see* **Note 7**).
2. If necessary, first cut the biopsy in small pieces with a sterile blade so that tissue pieces of maximally 10 mm^3, free of fat or necrotizing tissue, are obtained.
3. Place the tissue together with 1 mL of CellWASH in the Medicon holder.
4. Place the Medicon holder in the Medimachine and start the machine.
5. Depending on the tissue type, different run times should be used: 2 × 30 s for firm/solid biopsies and 1 × 45 s for skin biopsies. If necessary, an extra 1 mL of CellWASH can be added to the Medicon holder to obtain a cell suspension and determine the cell concentration.
6. Stop the machine and remove the Medicon holder from the machine.
7. Carefully remove the cover of the Medicon holder, and use a small syringe to recuperate the obtained cell suspension.
8. Determine the cell concentration of the suspension.
9. If the suspension contains less than 625 cells/µL, concentrate the cell suspension by centrifugation for 1 min at 16,200 × *g*, and dissolve the pellet in a smaller volume of CellWASH. Likewise, if the concentration is >40,000 cells/µL, dilute the suspension (*see* **Note 8**).
10. Put 200 µL of the cell suspension in a 1.5 mL tube.
11. Continue the procedure as described in Subheading 3.2.3 from **step 3** onward.

3.2.3 Extraction from Peripheral Blood or Bone Marrow

1. Place 200 µL fresh or frozen EDTA-PB or EDTA-BM in a 1.5 mL tube (*see* **Note 9**).
2. Add 20 µL QIAGEN protease directly into the PB or BM (*see* **Note 10**).
3. Add 200 µL AL buffer.
4. Vortex the tube for at least 15 s at minimally 600 × *g* (*see* **Note 10**).

5. Incubate 10 minutes at 56 °C in a thermo block.
6. Shortly spin the tube before proceeding (*see* **Note 10**).
7. Add 200 μL 99% ethanol.
8. Vortex the tube for at least 15 s at maximum rpm.
9. Shortly spin the tube before proceeding.
10. Label de QIAamp columns and put them in the 2 mL collection tubes.
11. Carefully apply the lysed sample onto the column (*see* **Note 10**).
12. Centrifuge 1 minute, 9600 × *g*.
13. Discard the collection tube with the flow through, and place the column in a new collection tube.
14. Put 500 μL AW1 buffer on the column.
15. Centrifuge 1 min, 9600 × *g*.
16. Discard the collection tube with the flow through, and place the column in a new collection tube.
17. Put 500 μL AW2 buffer on the column.
18. Centrifuge 3 min, 16,200 × *g*.
19. Discard the collection tube with the flow through, and place the column in a new collection tube.
20. Centrifuge 1 min, 16,200 × *g* (*see* **Note 11**).
21. Label new 1.5 mL tubes.
22. Place the column in a labeled 1.5 mL tube.
23. Put 200 μL AE buffer on the middle of the column.
24. Incubate 5 min at room temperature (18–25 °C).
25. Centrifuge 1 min, 9600 × *g*.
26. Discard the column and keep the 1.5 mL tube containing the DNA solution.
27. Determine the DNA concentration, and dilute if necessary to the desired concentration (e.g., 50 ng/μL) with AE buffer (*see* **Note 12**).

3.3 PCR Setup

1. PCR setup will be dependent on the available information accompanying the sample. In most cases either histopathology and/or flow cytometry results will indicate whether the sample is suspicious for the presence of abnormal B- or T-cells. In case the cell origin is unknown or unclear, a combination of the below mentioned targets should be analyzed (Fig. 3) (*see* **Note 13**).
2. Analysis should be performed in duplicate for each target, to be able to distinguish between a true clonal proliferation of cells and an inconsistent clonal pattern due to the presence of few B-

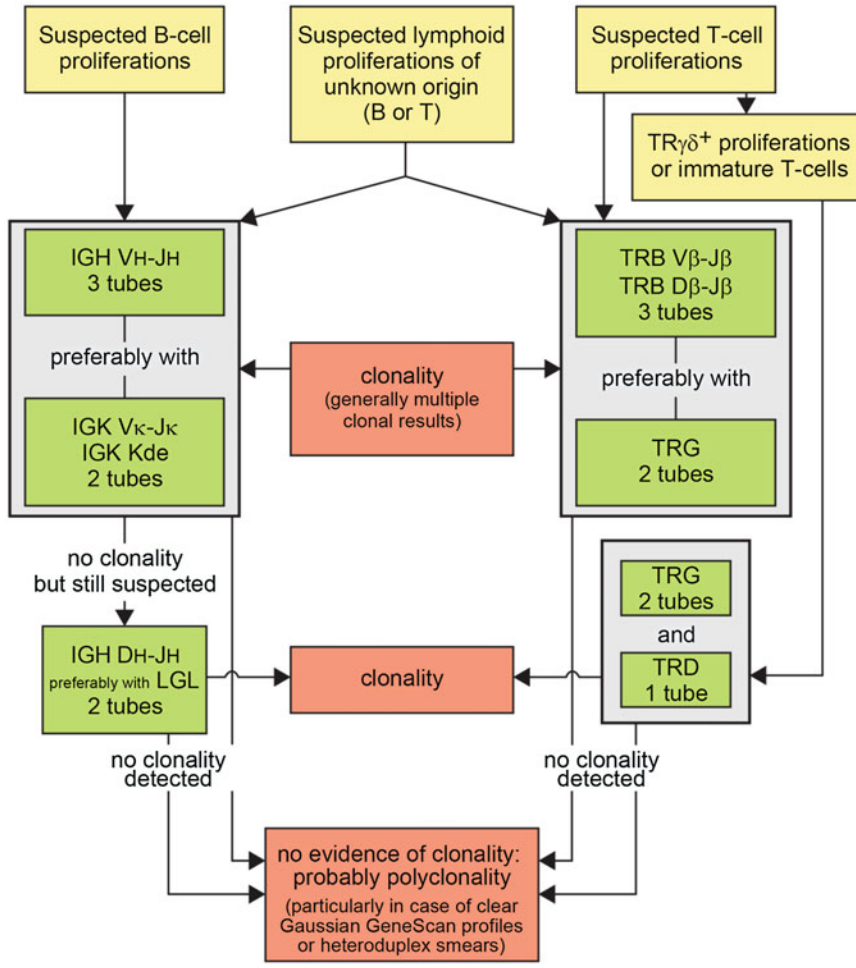


Fig. 3 Proposed IG/TR clonality testing algorithm

or T-lymphocytes or, in case of FFPE tissue, due to degraded DNA (*see* **Note 14**).

3. A standard protocol is used for all PCR reactions: 95 °C 10 min//(94 °C 30 s, 60 °C 30 s, 72 °C 1 min) × 35// 72 °C 10 min//4 °C hold (*see* **Note 2**).
4. We here describe a PCR setup of primer mixes that consist of individually ordered primers. An alternative PCR setup would make use of commercially available master mixes that include primers, dNTPs, buffer, and $MgCl_2$.

3.3.1 Suspect B-Cell Proliferations

1. Make a primer mix for the following targets: IGH tube A VH FR1-JH, IGH tube B VH FR2-JH, IGH tube C VH FR3-JH, IGH tube D DH-JH, IGK tube A Vκ-Jκ, and IGK tube B Vκ/intron-Kde.

2. For each target, combine 32 μL PCR-grade H_2O, 5 μL 10× GOLD buffer, 3 μL 25 mM $MgCl_2$, 4 μL 10 mM dNTPs, and 0.8 μL primer mix (*see* **Note 15**). Vortex thoroughly.
3. Add 0.2 μL Ampli*Taq* Gold and mix by pipetting (do not vortex!).
4. Pipet 45 μL master mix in a PCR tube or PCR plate (*see* **Note 16**).
5. Add 2–5 μL of the 50 ng/μL DNA solution (*see* **Note 12**) or two different amounts (e.g., 50 and 200 ng) for FFPE material.
6. Start the PCR reaction.

3.3.2 Suspect T-Cell Proliferations

1. Make primer mix for the following targets: TRB tube A Vβ-Jβ, TRB tube B Vβ-Jβ, TRB tube C Dβ-Jβ, TRG tube A Vγ-Jγ, and TRG tube B Vγ-Jγ.
2. For TRB tubes A and B, combine 26.4 μL PCR-grade H_2O, 5 μL 10× buffer II, 6 μL 25 mM $MgCl_2$, 4 μL 10 mM dNTPs, and 3.2 μL primer mix. Vortex thoroughly. Add 0.4 μL Ampli-*Taq* Gold and mix by pipetting (do not vortex!).
3. For TRB tube C, combine 29.6 μL PCR-grade H_2O, 5 μL 10× buffer II, 3 μL 25 mM $MgCl_2$, 4 μL 10 mM dNTPs, and 3.2 μL primer mix. Vortex thoroughly. Add 0.2 μL Ampli*Taq* Gold and mix by pipetting (do not vortex!).
4. For TRG tubes A and B, combine 32.4 μL PCR-grade H_2O, 5 μL 10× buffer II, 3 μL 25 mM $MgCl_2$, 4 μL 10 mM dNTPs, and 0.4 μL primer mix (*see* **Note 15**). Vortex thoroughly. Add 0.2 μL Ampli*Taq* Gold and mix by pipetting (do not vortex!).
5. Pipet 45 μL master mix in a PCR tube or PCR plate (*see* **Note 16**).
6. Add 2–5 μL of the 50 ng/μL DNA solution (*see* **Note 12**) or two different amounts (e.g., 50 and 200 ng) for FFPE material.
7. Start the PCR reaction.

3.4 GeneScan Fragment Analysis

We here describe a GeneScan analysis procedure using a Thermo Fisher genetic analyzer.

1. First denature the PCR product in Hi-Di formamide or H_2O together with, e.g., the GeneScan 500 ROX size standard. Add together 9.25 μL Hi-Di formamide (or H_2O), 0.25 μL ROX standard, and 1 μL PCR product in the specific 96-well plate for the genetic analyzer (*see* **Note 17**).
2. Close the wells with the 8-cap strips or seals.
3. Short spin the plate in the centrifuge (e.g., Megafuge 1.0).
4. Put the plate in a thermal cycler, and denature the samples for 2–5 min at 95 °C, followed by a minimal incubation of 2 min at 4 °C.

5. Mount the plate on the genetic analyzer, and analyze on a 36 cm capillary filled with POP7 with the following instrument protocol parameters: Dye Set D; Oven_Temperature 60 °C; Poly_Fill_Vol 6500 steps; Current_Stability 5.0 μA; PreRun_-Voltage 15.0 kV; Pre_Run_Time 180 s; Injection_Voltage 1.2 kV; Injection_Time 23 s; Voltage_Number_Of_Steps 20 nk; Voltage_Step_Interval 15 s; Data_Delay_Time 60 s; Run_Voltage 15.0 kV; and Run_Time 800 s.
6. Analyze the obtained .fsa files with GeneMapper software or freely available Peak Scanner software by using the default microsatellite analysis parameters with one exception: use the advanced peak detection algorithm, and change the baseline window from 51 pts to 251 pts (*see* **Note 18**).
7. Resolve any indicated analysis problems by reviewing all samples for which the software indicated a bad sizing quality. Make sure to set the sample plot settings so that off-scale peaks are identified in the plots. Due to overloading of the CCD camera during data collection, off-scale peaks (broad, capped peaks) appear, which sometimes might give the impression of >1 peak (*see* **Note 17**). Also these off-scale peaks can give problems during size standard analysis as they result in extra peaks in the ROX color due to incomplete color compensation.
8. To display and review the analyzed data, it could be helpful to use a specific sample plot setting for IG or TR genes. The IG targets contain only FAM-labeled PCR products, whereas in the TR targets, both FAM and HEX labels are present (with the exception of TRB tube B). Therefore a sample plot for IG targets should only demonstrate FAM- and ROX-labeled peaks, whereas for TR targets, also the HEX color should be displayed. For the *X*-axis always display the full base pair range (50 bp–500 bp) to be able to identify peaks outside the expected size range. Off-scale peaks should be shown, and the *Y*-axis (corresponding to the signal intensity) is set at 4000 for IG targets and 8000 for TR targets (*see* **Note 19**).
9. Verify whether the control samples give the expected pattern (*see* **Note 20**).

3.4.1 Heteroduplex Analysis

1. Heat 20 μL PCR product for 5 min at 94 °C in a thermal cycler.
2. Immediately cool the products to 4 °C, and incubate for 60 min to reanneal the PCR products.
3. Assemble the glass plates in the holder (use clean plates, appropriate clamps, and spacers).
4. Prepare a 6% non-denaturing polyacrylamide gel (40 mL) by mixing 8 mL 30% acrylamide/bisacrylamide (29:1) and 2 mL

10× TBE; add MilliQ water to a final volume of 40 mL (*see* **Note 21**).

5. Add 240 µL 10% ammonium persulfate (APS).
6. Add 60 µL TEMED and mix well.
7. Pour or pipette the mixture between the glass plates.
8. Place the comb.
9. Let the gel polymerize for approximately 1 h.
10. Remove the comb.
11. Assemble the plates in the vertical PAGE unit.
12. Fill the upper and lower compartments with 0.5× TBE.
13. Add 20 µL (denatured + reannealed) PCR product to one of the wells of a 96-well plate containing 4 µL non-denaturing bromophenol blue loading buffer (ratio 1:6).
14. Immediately load the samples on the 6% non-denaturing polyacrylamide gel (*see* **Note 21**).
15. Run at 40–50 V overnight or alternatively 1 h at 110 V followed by several hours at 180 V (*see* **Note 22**).
16. Disassemble the electrophoresis unit and carefully remove one of the glass plates.
17. Incubate the gel (loosely attached to the other glass plate) for 2–3 min in 250 mL water with 12.5 µL ethidium bromide 10 mg/mL (or use an alternative stain).
18. Discard the staining solution appropriately.
19. Remove access ethidium bromide by washing the gel for 2 min with water.
20. Visualize the DNA fragments by placing the gel on a transilluminator, and take a picture.

3.5 Post-analytical Data Interpretation

The molecular interpretation of the obtained patterns of each sample should be performed in several steps, basically following EuroClonality guidelines [9]:

1. Make a technical interpretation of the obtained pattern per tube. Identify possible non-specific peaks, and assign a technical description to the pattern (*see* **Note 23**). Some examples are given in Fig. 4.
2. Interpret the different tubes of the same IG or TR gene together, and evaluate whether the number of clonal peaks is in agreement with the presence of one clonal population (*see* **Note 24**).
3. Integrate the obtained clonality result with the clinical, morphological, and immunophenotypic data to allow making the final diagnosis and classification (*see* **Note 25**).

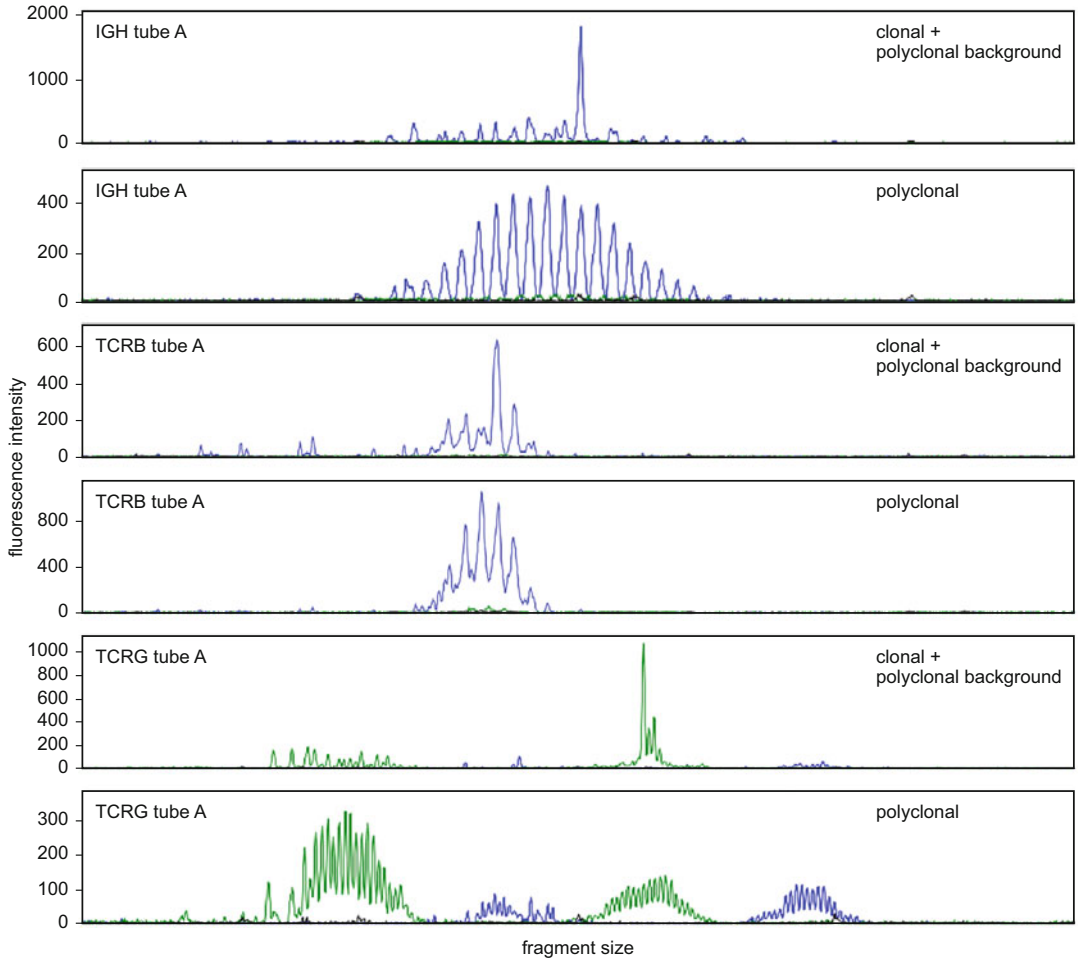


Fig. 4 Examples of IG/TR profiles in GeneScan fragment analysis. From top to bottom, the following profiles are shown: (i) IGH FR1 (tube A) multiplex PCR with clear monoclonal product on a polyclonal background; (ii) IGH FR1 (tube A) multiplex PCR showing normal Gaussian distribution of PCR products reflecting polyclonality; (iii) TRB tube A multiplex PCR with monoclonal product on a polyclonal background; (iv) TRB tube A multiplex PCR showing polyclonality with normal Gaussian distribution of PCR products; (v) TRG tube A multiplex PCR with monoclonal product on a background of polyclonality; (vi) TRG tube A multiplex PCR showing typical polyclonal profile that actually consists of four Gaussian size distributions due to variation in TRGV and TRGJ primer positions

4 Notes

1. Any DNA extraction technique (column-based or other) might be used, as long as the resulting DNA is free of proteins and RNA. Extraction methods which extract both DNA and RNA in parallel are not suitable. The RNA present in the DNA

solution negatively influences the PCR reaction resulting in an abnormal, disturbed polyclonal pattern.

2. The BIOMED-2 PCR protocols were validated by analysis of the same samples in different laboratories using different types of thermal cyclers (e.g., ABI 9700, 2700, or 2720 cyclers). This demonstrates the robustness of the developed protocols with regard to the used thermal cycler and PCR cycling conditions. However, it was shown that with older PCR equipment, better results are obtained with slightly different cycling conditions. Therefore, upon implementation of the technique, it is worthwhile to evaluate different cycling conditions [3]. Also take into account the different ramp rates of different thermal cyclers (e.g., for the Veriti 96-well cycler, a standard ramp rate of 100% is used).
3. Different automated sequencing systems can be used for GeneScan analysis. It is important to mention that the obtained molecular peak profiles are influenced by the polymer used to separate the single-stranded PCR products and by the settings of the software used to identify a peak.
4. A negative control sample (e.g., H_2O or buffer) is used to monitor possible contaminations throughout the entire procedure. Contamination during DNA extraction is sometimes underestimated; especially in samples with few lymphoid cells, a contamination during extraction with a strong clonal sample can cause a false-positive result. To this end buffer (or H_2O) can be used as non-template or negative control. Additionally, an extra control reaction that only contains master mix is useful to monitor contamination during PCR reaction setup. The polyclonal control sample is crucial as it allows to evaluate whether the PCR reaction was successful and to identify the polyclonal range of each tube. To obtain a complete, polyclonal TR pattern, DNA from mononuclear PB cells of healthy individuals is most optimal, although DNA from tonsil can be used as well. For IG polyclonal pattern, tonsil DNA is preferred because of the higher B-cell numbers. Although derived from alleged healthy individuals, it might be, especially for TR targets, that minor clonally activated cell populations are seen. Therefore each new batch of polyclonal control needs to be validated up-front. If needed, the monoclonal control sample might be omitted during routine PCR setup, as it mainly serves to monitor correct base pair calling by the GeneScan software.
5. To assess whether the tissue analyzed is representative, 4 μm hematoxylin-eosin sections just before and after the sections that are used for DNA extraction should be evaluated by an experienced hematopathologist. The exact tissue fixation protocol will have an effect on the quality of the extracted DNA. It

is important that neutral, buffered, formalin is used. Tissues fixated in Bouin or picric acid cannot be used for DNA analysis. Regarding fixation time no standardized protocols are available. However, prolonged fixation has been shown to induce too much cross-linking between DNA and other biomolecules resulting in a bad DNA quality for PCR analysis. Paraffin embedding should optimally be done with undiluted, fresh reagents, and beeswax or high-melt temperature paraffin should be avoided. It is strongly advised to assess the quality of the purified DNA from FFPE samples by a control PCR like the BIOMED-2 control gene primer set (Section 10 in ref. 3).

6. ATL buffer contains sodium dodecyl sulfate (SDS) which at lower temperatures might form a precipitate. If precipitation is seen, first incubate the buffer at 56 °C to dissolve the SDS. If the amount of tissue is too large, 180 μL of ATL buffer will not be sufficient to lyse the tissue entirely. In this case a double volume of buffer and proteinase K can be used. All subsequent volumes of buffers should then also be doubled until application of the obtained lysate on the column. The total lysate can be put on the column in two centrifugation steps. An overnight incubation is sufficient, but longer incubations (e.g., over the weekend) will not hamper the DNA quality.
7. Often fresh tissue samples will also be tested by flow cytometry to determine the presence of clonal B- or T-cell populations. In this case, it is advised to use the same cell suspension for both molecular and flow analysis so that an optimal correlation between the results can be made. The cell suspensions should be made in neutral, buffered solutions such as PBS or CellWASH. All tissue-grinding procedures will do (e.g., simple grinding of the tissue pieces in a Petri plate with the top of a sterile 15 mL Falcon tube might work as well). Flow cytometry fixatives should be avoided or at least tested up-front, as the obtained DNA might no longer be usable for PCR analysis due to irreversible cross-linking (e.g., TransFix from Cytomark). Alternatively, a small piece of tissue (2 mm^3) can be lysed directly in 180 μL ATL + 20 μL proteinase K as done with the FFPE tissues.
8. The here described QIAGEN extraction method is able to extract high-quality DNA from cell suspensions ranging from 625 cells/μL to 40,000 cells/μL as determined by the BIOMED-2 control gene primer set (Section 10 in ref. 3). Suspensions below 625 cells/μL will result in lower DNA concentration (<10 ng/μL), with accordingly weaker PCR amplification results. Therefore it is better to work with concentrated cell suspensions above 625 cells/μL. Concentrations >20,000 cells/μL should be diluted as they might result in bad DNA quality due to the presence of proteins which are not

completely removed. Also, the used columns might get saturated, and therefore no gain in DNA concentration is seen with higher cell numbers.

9. Use EDTA- or citrate-treated PB/BM. Heparin-stabilized PB or BM is not useful for PCR analysis as the heparin will remain in the DNA solution and will inhibit downstream PCR reactions.
10. To obtain high-quality DNA, several points should be taken into consideration. In case the sample volume is <200 μL, add CellWASH before starting the procedure to maintain the buffer proportions in the subsequent steps. It is important to avoid direct contact between QIAGEN protease and AL buffer as this will result in reduced protease activity. Therefore always put the protease into the PB or BM directly. Another crucial step for obtaining high-quality DNA is the 15-s vortexing of each sample with protease and AL buffer at high speed.

    To prevent contamination each vortexing step should be followed by a short spin before opening the tubes.
11. Be sure that all washing buffer is removed before proceeding as this buffer is a PCR inhibitor.
12. In the subsequent PCR reactions, 5 μL of the DNA solution will be used, corresponding to 250 ng DNA input. The original BIOMED-2 PCR protocol was validated using 100 ng of DNA in a 50 μL PCR reaction. However, a DNA input between 100 ng and 500 ng (as measured by NanoDrop) can be used without problem. Higher amounts of input DNA (or contaminating RNA) will lead to inhibition of the PCR reaction, but lower inputs might still be successful to determine the presence of a clonal cell population.
13. In a second series of reports, the here described standardized PCR protocols were used for clonality analysis of various disease entities. From these studies a guideline was developed for implementing clonality analysis in routine lymphoma diagnostics (Fig. 3) [10, 11]. In suspect B-cell proliferations, it is advised to use at least the three IGH VH-JH tubes combined with analysis of the two IGK (Vκ-Jκ and Kde rearrangements) tubes, to avoid delay in the diagnostic process. Subsequently, IGH DH1-6-JH PCR analysis (potentially combined with IGL analyses) can be performed in those cases where the previous PCR analysis failed to demonstrate clonality and no clear polyclonality was demonstrated, as this might indicate a false-negative result due to primer mismatching.

    In suspect T-cell proliferations, it was shown that combining the three TRB tubes with the two TRG tubes is the preferred strategy. TRD analysis should be avoided in routine practice because of complexity of interpretation. In fact, only in case of

suspected TCRγδ$^+$ T-cell proliferations, this tube provides complementary information. In this report we therefore restricted ourselves to describe the protocols for IGH, IGK, TRB, and TRG PCR analysis.

14. When few B- or T-cells are present in a sample, or in case of suboptimal DNA due to fixation cross-linking, the obtained pattern might look clonal as only few amplifiable targets are present, resulting in a peak pattern instead of a polyclonal pattern. By performing the analysis in duplicate, a true clonal pattern will give the same result in both PCR reactions, whereas a "false" clonal pattern will give differently sized peaks. Some labs use different amounts of input DNA to evaluate possible artifacts due to few amplifiable DNA targets (FFPE tissues or PCR inhibitors). Depending on the kind of tissues analyzed (many or few lymphoid cells, FFPE or fresh) and the experience, one might decide not to perform a duplicate analysis up-front but rather in second line for those samples in which no clear results are obtained. The latter strategy should however be used with great caution, as false clonality is an important factor, especially in the more sensitive GeneScan analysis, leading to a wrong interpretation of the results, if not appropriately taken into account.
15. The master mixes can also be obtained commercially, ready for use, in both heteroduplex and GeneScan fragment analysis format via Invivoscribe Technologies (Carlsbad, CA, USA, www.invivoscribe.com). In this way the time-consuming and difficult validation and quality control of the BIOMED-2-multiplex PCR tubes can be avoided. Master mix can be made up-front if stored at <-70 °C. Instead of 50 μL reactions, 25 μL reactions are also feasible, when using 100 ng of input DNA [3]. Verify whether higher amounts of DNA also work before changing the protocol to 25 μL.
16. The described protocols make use of the hot-start AmpliTaq Gold enzyme, allowing PCR setup at room temperature.
17. The optimal amount of PCR product used for GeneScan analysis is dependent on the amount of input DNA, the strength of the PCR reaction, and/or the type of instrument used. It is therefore recommended to validate the GeneScan setup before performing routine analysis. If too much PCR product is present, the CCD camera recording the fluorescent signals might get overloaded (off-scale peaks), leading to difficult pattern interpretation. In this case, a true clonal pattern might look as an oligoclonal pattern due to detection of minor PCR amplification products, which are the result of some normal lymphoid cells still present in the sample. Of note, weak artifact peaks can sometimes be seen at the left side of the stronger truly clonal peak. Such artifact peaks have a typical and

different appearance compared to a true but weak clonal peak, i.e., they are steeply rising and "ski slope" falling.

18. The used analysis parameters of the GeneMapper software influence the obtained peak pattern and the identification of a fluorescent signal as a peak. The newer GeneMapper software uses a slightly different algorithm as compared to the older software used during the BIOMED-2 project. It was observed that when using a baseline window of 51 data points, an irregular polyclonal pattern, due to few lymphoid cells in the samples, might look like an oligoclonal or even clonal pattern. Also the signal intensity of the peaks in the pattern is smaller as more background is subtracted. By using 251 or 501 data points, less background subtraction occurs, and the polyclonal, Gaussian, pattern is more clearly defined. Therefore this setting is preferred to identify true clonal populations (*see* Fig. 3).
19. The intensity of the peaks is dependent on a variety of parameters: lymphoid cell repertoire, quality of the obtained DNA, quality of the primers and/or PCR reaction, dilution factor of the PCR products, and genetic analyzer settings. Therefore the proposed sample plot settings are only a guideline. It might be necessary to review a specific sample with different plot settings to be able to evaluate the obtained pattern correctly.
20. The expected polyclonal patterns and possible non-specific products are listed in Table 4. Some of the mentioned non-specific products will only be seen in samples with few lymphoid cells. Also, in IGH tube D DH-JH, in case of few B-cells, the irregular polyclonal pattern will look like a pattern with multiple peaks instead of a disturbed Gaussian distribution. No amplification products should be seen in the NTC sample, and the peak size of the PCR products of the clonal control sample should be as formerly determined.
21. Percentage polyacrylamide gel and electrophoresis time are dependent on the expected size of the PCR products.
22. For convenient loading use special tips for loading vertical electrophoresis gels, e.g., Prot/Elec tips, Bio-Rad, 223-9915.
23. It is not possible within the frame of this chapter to describe all details and pitfalls of the molecular interpretation of GeneScan patterns. These items have however already been described in several special articles:
    (a) To make a correct interpretation of each PCR tube, an excellent knowledge of the analyzed gene rearrangements per tube is imperative. Details can be found in the different sections of the report of the BIOMED-2 Concerted Action BMH4 CT98-3936 [3].

Table 4
Size ranges, non-specific bands, and method of analysis for IG/TR multiplex PCRs (updated by EuroClonality consortium in 2012; *see* also ref. 9)

Tube	Polyclonal size range (bp)	Non-specific products (bp)	Preferred method of analysis
IGH tube A FR1 V-J	310–360	~85[a]	GS and HD both suitable
IGH tube B FR2 V-J	250–295	~85[a], ~228[a]	
IGH tube C FR3 V-J	100–170	~211[a]	
IGH tube D D-J	110–290 (DH1/2/4/5/6-JH) 390–420 (DH3-JH)	~161[a], ~350[b]	HD slightly preferred over GS (amplicon variation hampers GS)
IGK tube A V-J	120–160 (Vκ1f/6/Vκ7-Jκ) 190–210 (Vκ3f-Jκ) 260–300 (Vκ2f/Vκ4/Vκ5-Jκ)	~217[a]	HD slightly preferred over GS (less junctional size diversity + amplicon variation hampers GS)
IGK tube B V/intron-Kde	210–250 (Vκ1f/6/Vκ7-Kde) 270–300 (Vκ3f/intron-Kde) 350–390 (Vκ2f/Vκ4/Vκ5-Kde)	~404[a]	
TRB tube A V-J	240–285	~213[a,c], ~273[a,c]	GS and HD both suitable
TRB tube B V-J	240–285	~93[a,c],~126[a,c], ~221[a,c]	
TRB tube C V-J	170–210 (Dβ2) 285–325 (Dβ1)	~128[a,c], ~337[a,c]	
TRG tube A V-J	145–175 (Vγ10-Jγ1.3/2.3) 175–195 (Vγ10-Jγ1.1/2.1) 195–230 (Vγfl-Jγ1.3/2.3) 230–255 (Vγfl-Jγ1.1/2.1)	None	GS and HD both suitable
TRG tube B V-J	80–110 (Vγ11-Jγ1.3/2.3) 110–140 (Vγ11-Jγ1.1/2.1) 160–195 (Vγ9-Jγ1.3/2.3) 195–220 (Vγ9-Jγ1.1/2.1)	None	

GS GeneScan analysis; *HD* heteroduplex analysis
[a]Intensity and/or presence of the non-specific product will depend on the lymphoid repertoire present in the sample. Few lymphoid cells will result in stronger non-specific products
[b]Non-specific band as a result of cross-annealing of the DH2 primer to a sequence upstream of IGHJ4. In GeneScan analysis this non-specific band does not comigrate with D-J products
[c]Presence of these non-specific bands also depends on primer quality (primer batch dependent)

(b) Most clonal PCR products will have a number of nucleotides within the appropriate size range. However, occasionally, specific clonal products can be found outside the expected range due to, e.g., insertions or deletions in the rearranged genes. Therefore all isolated peaks, different from the known non-specific peaks, should be taken into account when interpreting the obtained pattern [12].

(c) When no specific PCR products are seen, it is crucial that the DNA quality of the sample is evaluated by the BIOMED-2 control gene primer set (Section 10 in ref. 3). Only in case of optimal DNA quality, the absence of specific PCR products can be attributed to an absence of cells with rearranged IG or TR genes.

(d) To give a technical interpretation of the obtained pattern per tube, a standardized universal description of clonality data has been developed and tested by the EuroClonality group [9]. This scoring system is based on the fact that the type of the obtained pattern is determined by the underlying immunobiological features of the sample.

(e) To avoid a biased interpretation, it is advised to interpret the pattern first without any extra information on the lymphoid cell repertoire in the sample. However, immunophenotypic data from cytomorphology, pathology, and/or flow cytometry are very useful tools to further interpret the overall obtained molecular clonality results.

(f) In case of follow-up samples, the size of the clonal peaks is useful to determine whether the observed peaks correspond to the original (malignant) clone.

(g) In case of request for BM invasion by lymphoma, especially in elderly people, it is important to correlate possible clonal peaks in the bone marrow or peripheral blood, with the original clonal population found in the lymph node. As people tend to live longer, dual pathologies can be seen at the same time, e.g., CLL and lymphoma. Both will result in clonal patterns, however, with differently sized peaks.

24. The number of clonal peaks that can still be compatible with the presence of a single clonal cell expansion is IG/TR target dependent [13].

25. In case clonality results are being reported outside a multidisciplinary setting, a clear remark on the report should be made indicating that the presence of a clonal cell proliferation is not equivalent to the presence of a malignant process. Also, the specific background of the request for clonality analysis should be taken into account before making a final conclusion.

References

1. Tonegawa S (1983) Somatic generation of antibody diversity. Nature 302:575–581
2. Davis MM, Bjorkman PJ (1988) T-cell antigen receptor genes and T-cell recognition. Nature 334:395–402
3. Van Dongen JJM, Langerak AW, Brüggemann M, Evans PA, Hummel M, Lavender FL, Delabesse E, Davi F, Schuuring E, García-Sanz R, Van Krieken JHJM, Droese J, González D, Bastard C, White HE, Spaargaren M, González M, Parreira A, Smith JL, Morgan GJ, Kneba M, Macintyre EA (2003) Design and standardization of PCR primers and protocols for detection of clonal immunoglobulin and T-cell receptor gene recombinations in suspect lymphoproliferations: report of the BIOMED-2 Concerted Action BMH4-CT98-3936. Leukemia 17:2257–2317
4. Van Zelm MC, Van der Burg M, De Ridder D, Barendregt BH, De Haas EF, Reinders MJ, Lankester AC, Révész T, Staal FJT, Van Dongen JJM (2005) Ig gene rearrangement steps are initiated in early human precursor B cell subsets and correlate with specific transcription factor expression. J Immunol 175:5912–5922
5. Dik WA, Pike-Overzet K, Weerkamp F, De Ridder D, De Haas EF, Baert MRM, Van der Spek P, Koster EEL, Reinders MJT, Van Dongen JJM, Langerak AW, Staal FJT (2005) New insights on human T-cell development by quantitative T-cell receptor gene rearrangement studies and gene expression profiling. J Exp Med 201:1715–1723
6. Evans PA, Pott C, Groenen PJ, Salles G, Davi F, Berger F, Garcia JF, Van Krieken JHJM, Pals ST, Kluin PM, Schuuring E, Spaargaren M, Boone E, González D, Martinez B, Villuendas R, Gameiro P, Diss TC, Mills K, Morgan GJ, Carter GI, Milner BJ, Pearson D, Hummel M, Jung W, Ott M, Canioni D, Beldjord K, Bastard C, Delfau-Larue MH, Van Dongen JJM, Molina TJ, Cabeçadas J (2007) Significantly improved PCR-based clonality testing in B-cell malignancies by use of multiple immunoglobulin gene targets. Report of the BIOMED-2 Concerted Action BHM4-CT98-3936. Leukemia 21:207–214
7. Brüggemann M, White H, Gaulard P, Garcia-Sanz R, Gameiro P, Oeschger S, Jasani B, Ott M, Delsol G, Orfao A, Tiemann M, Herbst H, Langerak AW, Spaargaren M, Moreau E, Groenen PJTA, Sambade C, Foroni L, Carter GI, Hummel M, Bastard C, Davi F, Delfau MH, Kneba M, Van Dongen JJM, Macintyre EA, Molina TJ (2007) Powerful strategy for PCR-based clonality assessment in T-cell malignancies. Report of the BIOMED-2 Concerted Action BHM4-CT98-3936. Leukemia 21:215–221
8. Langerak AW, Molina TJ, Lavender FL, Pearson D, Flohr T, Sambade C, Schuuring E, Al Saati T, Van Dongen JJM, Van Krieken JHJM (2007) PCR-based clonality testing in tissue samples with reactive lymphoproliferations: usefulness and pitfalls. A study from the BIOMED-2 Concerted Action BMH4-CT98-3936. Leukemia 21:222–229
9. Langerak AW, Groenen PJTA, Brüggemann M, Beldjord K, Bellan C, Bonello L, Boone E, Carter I, Catherwood M, Davi F, Delfau-Larue MH, Diss T, Evans PAS, Gameiro P, Garcia SR, Gonzalez D, Grand D, Håkansson A, Hummel M, Liu H, Lombardia L, Macintyre EA, Milner B, Montes-Moreno S, Schuuring E, Spaargaren M, Hodges E, Van Dongen JJM (2012) EuroClonality/BIOMED-2 guidelines for interpretation and reporting of Ig/TCR clonality testing in suspected lymphoproliferations. Leukemia 26:2159–2172
10. Diss TC, Molina TJ, Cabeçadas J, Langerak AW (2012) Molecular diagnostics in lymphoma: why, when and how to apply. Diagn Histopathol 18:53–63
11. Van Krieken JHJM, Langerak AW, Macintyre EA, Kneba M, Smith JL, Garcia Sanz R, Morgan GJ, Parreira A, Molina T, Cabeçadas J, Gaulard P, Jasani B, Garcia JF, Ott M, Hannsmann ML, Berger F, Hummel M, Davi F, Brüggemann M, Lavender FL, Schuuring EMD, Evans PAS, White H, Salles G, Groenen PJTA, Gameiro P, Pott C, Van Dongen JJM (2007) Improved reliability of lymphoma diagnostics via PCR-based clonality testing. Report of the BIOMED-2 Concerted Action BHM4-CT98-3936. Leukemia 21:201–206
12. Rothberg PG, Langerak AW, Verhaaf B, Van Dongen JJM, Burack WR, Johnson MD, Slate D, Laughlin TS, Payne K, Figueiredo L, Bandoh BN, Yan Q, Bacon CM, Wright P, Bench A, Du MQ, Liu H (2012) Clonal antigen receptor gene PCR products outside the expected size range. J Hematop 5:57–67
13. Langerak AW, Van Dongen JJM (2012) Multiple clonal Ig/TCR products: implications for interpretation of clonality findings. J Hematop 5:35–43

Chapter 5

Expression Cloning of Antibodies from Single Human B Cells

Hedda Wardemann and Christian E. Busse

Abstract

The majority of lymphomas originate from B cells at the germinal center stage. Preferential selection of B-cell clones by a limited set of antigens has been suggested to drive lymphoma development. While recent studies in chronic lymphocytic leukemia have shown that self-reactive B-cell receptors (BCR) can generate cell-autonomous signaling and proliferation, our knowledge about the role of BCRs for the development or survival of other lymphomas remains limited. Here, we describe a strategy to characterize the antibody reactivity of human B cells. The approach allows the unbiased characterization of the human antibody repertoire at single-cell level through the generation of recombinant monoclonal antibodies from single primary human B cells of defined origin. This protocol offers a detailed description of the method starting from the flow-cytometric isolation of single human B cells to the reverse transcription-polymerase chain reaction (RT-PCR)-based amplification of the expressed immunoglobulin (Ig) transcripts (*IGH*, *IGK*, and *IGL)* and their subsequent cloning into expression vectors for the in vitro production of recombinant monoclonal antibodies. The strategy may be used to obtain information on the clonal evolution of B-cell lymphomas by single-cell sequencing of Ig transcripts and on the antibody reactivity of human lymphoma B cells.

Key words Antibody, Immunoglobulin, Single-cell RT-PCR

1 Introduction

The virtually unlimited diversity of the antibody repertoire efficiently protects from infection. Antibody diversity is primarily established early in lymphocyte development through combinatorial joining of immunoglobulin (Ig) variable (V), diversity (D), and joining (J) gene segments [1]. A side effect of the random Ig gene recombination is the formation of self-reactive immunoglobulins [2]. In healthy individuals, a series of checkpoints purge self-reactive B cells from the repertoire, both centrally in the bone marrow during B-cell development and in peripheral lymphoid tissues [3–5]. Additional diversity in the B-cell repertoire is generated by random somatic hypermutation (SHM) of Ig V regions in antigen-activated B cells primarily during T-cell-dependent immune responses in germinal centers (GC). Somatically mutated

Ralf Küppers (ed.), *Lymphoma: Methods and Protocols*, Methods in Molecular Biology, vol. 1956,
https://doi.org/10.1007/978-1-4939-9151-8_5,

GC B cells with high affinity for their cognate antigen are positively selected and differentiate into memory B cells and humoral effector cells, i.e., short-lived surface Ig-positive plasmablasts and long-lived surface Ig-negative plasma cells [6–8]. The SHM process can again lead to the development of self-reactive antibodies, but how newly arising self-reactive GC B cells are regulated to avoid autoimmunity is still not well understood [5].

The majority of B-cell-associated lymphomas arise from B cells at the GC stage, indicating that malignant B-cell transformation often occurs during GC reactions [9, 10]. The predominance of GC-derived lymphomas may be due to the special microenvironment provided in the GC, where Ig gene modification processes, including DNA breaks due to isotype class switch recombination, and intense cellular proliferation occur [11]. BCR expression and antigen recognition are essential in B-cell lymphoma development and pathogenesis [9, 11, 12]. Stereotyped antigen-binding sites and biased V-segment usage suggest a limited set of antigens as driving force for B-cell activation [13–20]. Over the last years, high-throughput sequencing of adaptive immune receptor repertoires (AIRR-seq) has become a frequently used approach to assess Ig repertoires [21, 22]. While bulk sequencing protocols typically analyze millions of cells and therefore provide a high sensitivity (e.g., for the detection of minimal residual disease), it is still not possible to predict antibody reactivity based on sequence analysis [23]. In addition, due to the clonal nature of malignancies, the diversity of lymphoma cells is rather limited and thus does often not require the analysis of large numbers of cells to identify the relevant Igs. Human monoclonal antibodies can be produced by different methods to assess their antibody-specificity, such as immortalization of B cells with Epstein-Barr virus [24, 25] or the generation of B-cell hybridomas [26]. However, immortalization and fusion efficiencies are often low, largely depend on the maturation status of the B cell, and may require B-cell stimulation. Phage display technology on the other hand relies on the random pairing of human Ig heavy and light chains in vitro; therefore the antibodies that are produced may not mimic the natural situation [27].

Here, we describe an unbiased and highly efficient strategy to characterize the Ig genes and antibody reactivity of primary human B cells on a single-cell level [4, 28]. This protocol offers a detailed guide comprising the flow-cytometric isolation of single-cell B cells; the subsequent unbiased RT-PCR amplification of human *IGH*, *IGK*, and *IGL* transcripts; and Ig gene expression vector cloning for in vitro antibody production. The obtained sequence data provide detailed information on Ig gene rearrangement, V segment somatic mutations, isotype subclass, and clonal relationships of individual cells. Importantly, the recombinant monoclonal antibodies can be used to determine their antigen reactivity and epitope specificity [4, 15, 16, 28–30]. The characterization of

recombinant antibodies from B-cell chronic lymphocytic leukemia (CLL) cells using this approach revealed the basis of the clinically observed difference between cases of *IGHV* mutated CLL and *IGHV* unmutated CLL cases and demonstrated that antigen binding and BCR stimulation can contribute to clonal selection and expansion of neoplastic B-cell clones [19, 20].

2 Materials

2.1 Isolation of Single Human B Cells by Flow Cytometry

1. Mononuclear **cells** (*see* **Notes 1–3**).
2. Staining solution: cold phosphate-buffered saline (PBS) (without Ca^{2+} or Mg^{2+}) with 2% heat-inactivated fetal calf serum (FCS) and antihuman antibodies or antigen at the final concentration (*see* **Notes 4** and **5**).
3. PBS/FCS: cold PBS (without Ca^{2+} or Mg^{2+}) with 2% heat-inactivated FCS, sterile; store at 4 °C.
4. 7-Aminoactinomycin D (7-AAD) solution: prepare a 2 μg/mL solution; store protected from light at 4 °C (*see* **Note 6**).
5. 10× PBS (without Ca^{2+} or Mg^{2+}, molecular biology grade).
6. 100 mM dithiotretiol (DTT) (from Subheading 2.2, **item 4**).
7. Random hexamer primer (RHP) ($pd(N)_6$; Roche Applied Science): prepare a 300 ng/μL solution in PCR water; store in small aliquots at −20 °C (*see* **Note 7**).
8. IGEPAL CA-630, molecular biology grade: prepare a 10% solution in PCR water; store in aliquots at −20 °C.
9. RNAsin (40 U/μL; Promega).
10. 96-well PCR plate, skirted (*see* **Note 8**).
11. Adhesive polyolefine plate seal (Order-# 900 320, HJ-Bioanalytik, Erkelenz, Germany) (*see* **Note 9**).

2.2 RT-PCR and Ig Gene Amplification

1. PCR cycler compatible with skirted 96-well PCR plates.
2. Nuclease-free PCR water.
3. RNAsin (40 U/μL; Promega).
4. SuperScript IV Reverse Transcriptase (200 U/μL; Invitrogen) with 5× RT buffer and 100 mM DTT solution (*see* **Note 10**).
5. dNTP solution, prepare a mix with 25 mM of each nucleotide; store in small aliquots at −20 °C.
6. HotStarTaq DNA polymerase (5 U/μL; Qiagen) with 10× PCR buffer.
7. 5′ primer or primer mix: prepare a stock solution of 50 μM for each primer. For primer mixes, combine equal volumes of

Table 1
Primer sequences

PCR step	Primer name	Primer sequence (5′ → 3′)
IGH first PCR		
Use these nine primers as mix	hIGHV-1/7-066-fw	ACAGGTGCCCACTCCCAGGTGCAG
	hIGHV-1/7-017-fw	ATGGACTGGACCTGGAG
	hIGHV-1/7-041-fw	TCCTCTTTGTGGTGGCAGCAGC
	hIGHV-2-035-fw	TCCACGCTCCTGCTRCTGAC
	hIGHV-3-066-fw	AAGGTGTCCAGTGTGARGTGCAG
	hIGHV-3-057-fw	TAAAAGGTGTCCAGTGT
	hIGHV-4/6-066-fw	CCCAGATGGGTCCTGTCCCAGGTGCAG
	hIGHV-4-022-fw	ATGAAACACCTGTGGTTCTTCC
	hIGHV-5-066-fw	CAAGGAGTCTGTTCCGAGGTGCAG
Use these four primers as mix	hIGHA-111-rv	GTCCGCTTTCGCTCCAGGTCACACT
	hIGHE-140-rv	AAGGTCATAGTTGTCCCGTTGAGG
	hIGHG-137-rv	GGAAGGTGTGCACGCCGCTGGTC
	hIGHM-082-rv	GGAAGGAAGTCCTGTGCGAGGC
IGH second PCR		
	hIGHV-pan-080-fw	AGGTGCAGCTGCTGGAGTCKGG
Use these four primers as mix	hIGHA-076-rv	GGAAGAAGCCCTGGACCAGGC
	hIGHE-070-rv	CCAGGCAGCCCAGAGTCACGG
	hIGHG-074-rv	AGTCCTTGACCAGGCAGCCC
	hIGHM-031-rv	GGGAATTCTCACAGGAGACGA
IGH-specific PCR		
	5′ AgeI VH1	CTGCAACCGGTGTACATTCCCAGGTGCAGC TGGTGCAG
	5′ AgeI VH1-18	CTGCAACCGGTGTACATTCCCAGGTTCAGC TGGTGCAG
	5′ AgeI VH1-24	CTGCAACCGGTGTACATTCCCAGGTCCAGC TGGTACAG
	5′ AgeI VH1/5	CTGCAACCGGTGTACATTCCGAGGTGCAGC TGGTGCAG
	5′ AgeI VH3	CTGCAACCGGTGTACATTCTGAGGTGCAGC TGGTGGAG
	5′ AgeI VH3-9	CTGCAACCGGTGTACATTCTGAAGTGCAGC TGGTGGAG
	5′ AgeI VH3-23	CTGCAACCGGTGTACATTCTGAGGTGCAGCTG TTGGAG
	5′ AgeI VH3-33	CTGCAACCGGTGTACATTCTCAGGTGCAGC TGGTGGAG
	5′ AgeI VH4	CTGCAACCGGTGTACATTCCCAGGTGCAGC TGCAGGAG

(continued)

Table 1
(continued)

PCR step	Primer name	Primer sequence (5′ → 3′)
	5′ AgeI VH4-34	CTGCAACCGGTGTACATTCCCAGGTGCAGC TACAGCAGTG
	5′ AgeI VH4-39	CTGCAACCGGTGTACATTCCCAGCTGCAGC TGCAGGAG
	5′ AgeI VH6-1	CTGCAACCGGTGTACATTCCCAGGTACAGC TGCAGCAG
	5′ AgeI VH7	CTGCAACCGGTGTACATTCTCAGGTGCAGC TGGTGCAATCTGG
	3′ SalI JH1/2/4/5	TGCGAAGTCGACGCTGAGGAGACGGTGACCAG
	3′ SalI JH3	TGCGAAGTCGACGCTGAAGAGACGGTGACCA TTG
	3′ SalI JH6	TGCGAAGTCGACGCTGAGGAGACGGTGACCGTG
IGK first PCR		
Use these three primers as mix	hIGKV-1-060-fw hIGKV-3-049-fw hIGKV-4-049-fw	ATGAGGSTCCCYGCTCAGCTGCTGG CTCTTCCTCCTGCTACTCTGGCTCCCAG ATTTCTCTGTTGCTCTGGATCTCTG
	hIGKC-172-rv	GTTTCTCGTAGTCTGCTTTGCTCA
IGK second PCR		
	hIGKV-pan-fw	ATGACCCAGWCTCCABYCWCCCTG
	hIGKC-032-rv	AACTGCTCATCAGATGGCGGG
IGK-specific PCR		
	5′ AgeI Vκ1-5	CTGCAACCGGTGTACATTCTGACATCCAGA TGACCCAGTC
	5′ AgeI Vκ1-5b	CTGCAACCGGTGTACATTCAGACATCCAG TTGACCCAGTCT
	5′ AgeI Vκ1-6	CTGCAACCGGTGTACATTCTGCCATCCAGA TGACCCAGTC
	5′ AgeI Vκ1-13	CTGCAACCGGTGTACATTCTGCCATCCAG TTGACCCAGTC
	5′ AgeI Vκ1D-43	CTGCAACCGGTGTACATTGTGCCATCCGGA TGACCCAGTC
	5′ AgeI Vκ2-24	CTGCAACCGGTGTACATGGGGATATTGTGA TGACCCAGAC
	5′ AgeI Vκ2-28	CTGCAACCGGTGTACATGGGGATATTGTGA TGACTCAGTC
	5′ AgeI Vκ2-30	CTGCAACCGGTGTACATGGGGATGTTGTGA TGACTCAGTC

(continued)

Table 1
(continued)

PCR step	Primer name	Primer sequence (5′ → 3′)
	5′ AgeI Vκ3-11	CTGCAACCGGTGTACATTCAGAAATTGTG TTGACACAGTC
	5′ AgeI Vκ3-15	CTGCAACCGGTGTACATTCAGAAATAGTGA TGACGCAGTC
	5′ AgeI Vκ3-20	CTGCAACCGGTGTACATTCAGAAATTGTG TTGACGCAGTCT
	5′ AgeI Vκ4-1	CTGCAACCGGTGTACATTCGGACATCGTGA TGACCCAGTC
	3′ BsiWI Jκ1/4	GCCACCGTACGTTTGATYTCCACCTTGGTC
	3′ BsiWI Jκ2	GCCACCGTACGTTTGATCTCCAGCTTGGTC
	3′ BsiWI Jκ3	GCCACCGTACGTTTGATATCCACTTTGGTC
	3′ BsiWI Jκ5	GCCACCGTACGTTTAATCTCCAGTCGTGTC
IGL first PCR		
	hIGLV-1-068-fw	GGTCCTGGGCCCAGTCTGTGCTG
	hIGLV-2-068-fw	GGTCCTGGGCCCAGTCTGCCCTG
	hIGLV-3-068-fw	GCTCTGTGACCTCCTATGAGCTG
	hIGLV-4/5-068-fw	GGTCTCTCTCSCAGCYTGTGCTG
	hIGLV-6-068-fw	GTTCTTGGGCCAATTTTATGCTG
	hIGLV-7-068-fw	GGTCCAATTCYCAGGCTGTGGTG
	hIGLV-8-083-fw	GAGTGGATTCTCAGACTGTGGTG
	hIGLC-057-rv	CACCAGTGTGGCCTTGTTGGCTTG
IGL second PCR		
	hIGLV-pa1-fw	CAGYCTGYSCTGACTCA
	hIGLV-pa2-fw	TCCTATGAGCTGACWCAG
	hIGLC-026-rv	TCAGAGGAGGGYGGGAACAGAGTG
IGL-specific PCR		
	5′ AgeI Vλ1	CTGCTACCGGTTCCTGGGCCCAGTCTGTGC TGACKCAG
	5′ AgeI Vλ2	CTGCTACCGGTTCCTGGGCCCAGTCTGCCC TGACTCAG
	5′ AgeI Vλ3	CTGCTACCGGTTCTGTGACCTCCTATGAGC TGACWCAG
	5′ AgeI Vλ4/5	CTGCTACCGGTTCTCTCTCSCAGCYTGTGC TGACTCA

(continued)

Table 1
(continued)

PCR step	Primer name	Primer sequence (5′ → 3′)
	5′ AgeI Vλ6	CTGCTACCGGTTCTTGGGCCAATTTTATGC TGACTCAG
	5′ AgeI Vλ7/8	CTGCTACCGGTTCCAATTCYCAGRCTGTGG TGACYCAG
	3′ XhoI Cλ	CTCCTCACTCGAGGGYGGGAACAGAGTG
Insert check		
	Ab sense	GCTTCGTTAGAACGCGGCTAC
	IgG internal	GTTCGGGGAAGTAGTCCTTGAC
	hIGKC-172-rv	GTGCTGTCCTTGCTGTCCTGCT
	hIGLC-057-rv	CACCAGTGTGGCCTTGTTGGCTTG

Restriction sites are underlined

50 μM stock solutions of each primer (Table 1); store in small aliquots at −20 °C.

8. 3′ primer or primer mix: prepare a stock solution of 50 μM for each primer. For primer mixes, combine equal volumes of 50 μM stock solutions of each primer (Table 1); store in small aliquots at −20 °C.
9. Adhesive polyolefine plate seal.
10. Agarose (gel electrophoresis grade).
11. TAE buffer: 40 mM Tris-acetate, 1 mM ethylenediaminetetraacetic acid (pH 8.0).
12. 10 mg/mL ethidium bromide solution; dilute 1:2000 in TAE buffer.
13. Loading dye for DNA gels: 60% w/v sucrose, 1 mM Cresol Red.
14. 100 bp DNA ladder.

2.3 Expression Vector Cloning

1. PCR cycler compatible with 96-well PCR plates.
2. 5′ primer: prepare a working dilution of 5 μM for each primer (Table 1, *see* **Note 11**).
3. 3′ primer: prepare a working dilution of 5 μM for each primer (Table 1, *see* **Note 11**).
4. Adhesive polyolefine plate seal.
5. Nuclease-free PCR water.
6. HotStarTaq DNA polymerase (5 U/μL; Qiagen) with 10× PCR buffer.

Table 2
Available expression vectors

Designation	Constant region	Gene product	Restriction site 5′	Restriction site 3′	ENA accession #	AddGene ID #
AbVec2.0-IGHG1	*IGHG1*	Human Ig γ1	AgeI	SalI	LT615368	80795
AbVec2.0-IGHG2	*IGHG2*	Human Ig γ2	AgeI	SalI	n/a	99576
AbVec2.0-IGHG3	*IGHG3*	Human Ig γ3	AgeI	SalI	n/a	99577
AbVec2.0-IGHG4	*IGHG4*	Human Ig γ4	AgeI	SalI	n/a	99578
AbVec1.1-IGKC	*IGKC*	Human Igκ	AgeI	BsiWI	LT615369	80796
AbVec1.1-IGLC2-XhoI	*IGLC2*	Human Igλ2	AgeI	XhoI	n/a	99575

7. dNTP solution, prepare a mix with 25 mM of each nucleotide; store in small aliquots at −20 °C.
8. PCR purification kit, e.g., NucleoSpin 96 PCR Clean-up (Macherey-Nagel).
9. AgeI-HF (20 U/μL; NEB).
10. SalI-HF (20 U/μL; NEB).
11. BsiWI-HF (20 U/μL; NEB).
12. XhoI (20 U/μL; NEB).
13. CutSmart buffer (10×; NEB).
14. Linearized expression vector AbVec2.0-IGHG1 (Table 2) (50 ng/μL; *see* **Note 12**).
15. Linearized expression vector AbVec1.1-IGKC (Table 2) (50 ng/μL; *see* **Note 12**).
16. Linearized expression vector AbVec1.1-IGLC2-XhoI (Table 2) (50 ng/μL; *see* **Note 12**).
17. T4 DNA ligase (400 U/μL) with 10× T4 DNA ligase reaction buffer.
18. Chemical competent *Escherichia coli* (*E. coli*) DH10B cells (Invitrogen).
19. LB medium at room temperature.
20. LB agar plates containing 100 μg/mL ampicillin.
21. Taq DNA polymerase (5 U/μL) with 10× PCR buffer.
22. Primer Ab sense (Table 1): prepare a stock solution of 50 μM; store in small aliquots at −20 °C.
23. Primer IgG internal, hIGKC-172-rv and hIGLC-057-rv (Table 1): prepare stock solutions of 50 μM; store in small aliquots at −20 °C.

24. Terrific broth containing 75 μg/mL ampicillin; pre-warmed to room temperature.
25. Round bottom tube with snap cap, 13 mL.
26. Plasmid purification kit, e.g., NucleoSpin Plasmid (Macherey-Nagel).

2.4 Expression of Recombinant Monoclonal Antibodies

1. Human embryonic kidney (HEK) 293T cells (German Collection of Microorganisms and Cell Cultures (DSMZ), Order-# ACC 635).
2. 150 mm cell culture plates.
3. DMEM growth medium: Dulbecco's Modified Eagle's Medium (DMEM; Gibco) supplemented with 10% heat-inactivated FCS (Invitrogen), 100 μg/mL streptomycin, 100 U/mL penicillin G, and 0.25 μg/mL amphotericin B (all Gibco).
4. DMEM wash medium: DMEM supplemented with 100 μg/mL streptomycin, 100 U/mL penicillin G, and 0.25 μg/mL amphotericin B.
5. DMEM Nutridoma medium: DMEM supplemented with 1% Nutridoma-SP (Roche, Cat.-Nr. 11 011 375 001), 100 μg/mL streptomycin, 100 U/mL penicillin G, and 0.25 μg/mL amphotericin B.
6. 150 mM NaCl, sterile.
7. Vector DNA: equal amounts (10–15 μg) of AbVec2.0-IGHG1 expression vector DNA and AbVec1.1-IGKC or AbVec1.1-IGLC2-XhoI expression vector DNA.
8. Polyethylenimine, branched (PEI; Sigma): prepare a PEI solution of 1 μg/mL in deionized water and filter sterilize. Store at 4 °C for up to 6 months.
9. Sodium azide (NaN_3) solution.
10. Protein G Sepharose (GE Healthcare).
11. PBS, sterile.
12. Chromatography spin columns (Bio-Rad).
13. 0.1 M glycine, pH 3.0.
14. 1 M Tris, pH 8.0, supplemented with 0.5% NaN_3.

3 Methods

3.1 Isolation of Single Human B Cells by Flow Cytometry

Single primary human B cells can be isolated from various sources (e.g., blood, bone marrow, tissue) based on their phenotype or antigen reactivity by flow cytometry (*see* **Note 1**). Throughout all steps, it is important to work under RNAse-free conditions and on ice. When fluorochromes are used, incubation should be carried out in the dark.

1. Prepare Sort/RHP mix with 5.2 μL nuclease-free water, 0.175 μL 10× PBS, 0.35 μL DTT, 0.475 μL IGEPAL CA-630 (10% v/v), 0.475 μL RHP, and 0.325 μL RNAsin (*see* **Note 7**). Dispense 7 μL in each well of a 96-well PCR plate and keep on ice until used for sorting (up to 6 h). Sort/RHP can also be prepared ahead of time and stored at −80 °C for up to 3 months. Frozen plates should be thawed on ice 30 min before use.
2. Incubate purified mononuclear cells with 50–100 μL staining solution for 30 min (*see* **Notes 4–6** and **13–15**).
3. Add 1 mL PBS/FCS, centrifuge for 5 min at 400 × *g*, 4 °C. Carefully discard supernatant.
4. Repeat washing step.
5. Resuspend cell pellet in 100 μL of 7-AAD solution, and incubate for 15 min in the dark.
6. Add 1 mL PBS/FCS, centrifuge cells 5 min at 400 × *g*, 4 °C. Carefully discard supernatant.
7. Repeat washing step.
8. Pre-wet 35 μm cell strainer cap with PBS/FCS (*see* **Note 15**).
9. Resuspend cell pellet in 400 μL PBS/FCS. Filter through strainer cap into round bottom tube.
10. Keep samples on ice until loaded onto the sorter.
11. Set up the flow cytometer for 96-well sorting: sort 50 beads into each corner well of an empty 96-well plate. A single droplet should be visible. Make sure that the bead droplet is located in the center of the bottom of each well (*see* **Notes 8 and 16**).
12. Sort the cells of interest into 96-well PCR plates containing 7 μL Sort/RHP mix per well (*see* **Notes 3** and **17–19**).
13. Immediately seal the sorted plate and put on dry ice before storing at −80 °C (*see* **Notes 9 and 20**).

3.2 RT-PCR and Ig Gene Amplification

3.2.1 cDNA Synthesis

cDNA is synthesized in the original 96-well sort plate. Throughout all steps, it is important to work under RNAse-free conditions and on ice. A master mix is prepared with the indicated surplus, but unless otherwise noted all reaction mixes are specified per well.

1. Thaw a plate of single-sorted cells on ice and spin down (*see* **Notes 7** and **21**).
2. While the plate is thawing, prepare the RT mix with 2.8 μL RT buffer, 2.2 μL nuclease-free water, 1.05 μL DTT, 0.5 μL dNTP solution, 0.2 μL RNAsin, and 0.25 μL SuperScript IV (*see* **Notes 7** and **22**).

3. Incubate sort plate for 1 min at 68 °C, and transfer back onto ice.
4. Add 7 μL RT mix to each well. Mix carefully and seal plate (*see* **Note 19**).
5. Perform reverse transcription at 42 °C for 5 min, 25 °C for 10 min, 50 °C for 60 min, and 94 °C for 5 min (*see* **Note 10**).
6. Store cDNA at −20 °C until further use, or use cDNA plate directly to set up single-cell PCRs.

3.2.2 Ig Gene Amplification

Human *IGH* and *IGK* or *IGL* transcripts from the same cell are amplified separately by nested PCRs. All steps are conducted on ice while wearing gloves. A master mix is prepared with the indicated surplus, but unless otherwise noted all reaction mixes are specified per well.

1. Prepare PCR master mix 1 consisting of 32.16 μL water, 4 μL PCR buffer, 0.13 μL 5′ primer mix, 0.13 μL 3′ primer or primer mix, 0.4 μL dNTP solution, and 0.18 μL HotStarTaq (*see* **Notes 7** and **22**).
2. Transfer PCR master mix 1 into a new 96-well PCR plate (37 μL per well; *see* **Note 19**).
3. Carefully transfer 3 μL cDNA from the cDNA plate into the first PCR plate with PCR master mix 1, mix carefully, and seal plate with PCR film (*see* **Notes 19** and **21**).
4. Run first PCR at 94 °C for 15 min; 50 cycles at 94 °C for 30 s, 58 °C (*IGH* and *IGK*) or 60 °C (*IGL*) for 30 s, and 72 °C for 55 s; and 72 °C for 10 min, hold at 4 °C.
5. Prepare PCR master mix 2 consisting of 31.66 μL water, 4 μL PCR buffer, 0.13 μL 5′ primer or primer mix, 0.13 μL 3′ primer or primer mix, 0.4 μL dNTP solution, and 0.18 μL HotStarTaq (*see* **Notes 7**, **22**, and **23**).
6. Transfer PCR master mix 2 into new 96-well PCR plate (36.5 μL per well; *see* **Note 19**).
7. Add 3.5 μL of the first PCR product, mix carefully, and seal plate (*see* **Notes 19** and **21**).
8. Run second PCR at 94 °C for 15 min; 50 cycles at 94 °C for 30 s, 58 °C (*IGH* and *IGK*) or 60 °C (Igλ) for 30 s, and 72 °C for 45 s; 72 °C for 10 min.
9. Load 3 μL of second PCR product mixed with 3 μL of loading dye onto a 2% analytical agarose gel in TAE buffer containing ethidium bromide and run for 25 min at 120 V. Visualize DNA bands under ultraviolet (UV) light.
10. Perform Sanger sequencing of second PCR products of the expected size (480 bp for *IGH*, 370 bp for *IGK*, and 390 bp for *IGL*) using the respective 3′ primer used for amplification in the second PCR. Purification of the PCR products before

sequencing is recommended, as it substantially improves the sequencing quality.

11. Analyze Ig gene sequences by IgBLAST (https://www.ncbi.nlm.nih.gov/igblast/; *see* **Notes 24–28**).

3.3 Expression Vector Cloning

All steps are conducted on ice while wearing gloves. A master mix is prepared with the indicated surplus, but unless otherwise noted all reaction mixes are specified per well.

3.3.1 Ig Gene-Specific PCR

All primers used in Ig gene-specific PCRs contain restriction sites (AgeI and SalI for *IGH*, AgeI and BsiWI for *IGK*, and AgeI and XhoI for *IGL*), which allow direct cloning into expression vectors (*see* **Notes 11**, **23**, and **29**).

1. Prepare PCR master mix 3 consisting of 29.2 μL water, 4 μL PCR buffer, 0.4 μL dNTP solution, and 0.2 μL HotStarTaq (*see* **Note 22**).
2. Transfer PCR master mix 3 into new 96-well PCR plate (33.8 μL into each well; *see* **Note 19**).
3. Deposit 2 μL each of 5′ and 3′ gene-specific primer into the corresponding well of the 96-well PCR plate containing PCR master mix 3.
4. Add 4.2 μL first PCR product to the corresponding well of the 96-well PCR plate, mix carefully, and seal plate (*see* **Note 21**).
5. Run second PCR program (*see* Subheading 3.2.2, **step 8**).
6. Run second PCR products on a 2% agarose gel to verify amplification.

3.3.2 Directional Cloning into Expression Vectors

1. Purify specific PCR products using a PCR purification kit, and elute with Tris–HCl buffer to a final volume of approx. 40 μL (*see* **Note 30**).
2. Use 30.6 μL purified PCR product, and add 3.4 μL CutSmart Buffer (*see* **Note 31**).

Table 3
Pipetting scheme for restriction digest

	***IGH* V region PCR products**	***IGK* V region PCR products**	***IGL* V region PCR products**
Enzyme mix 1 (*see* **Note 22**)	5.3 μL water	5.3 μL water	5.3 μL water
	0.6 μL CutSmart	0.6 μL CutSmart	0.6 μL CutSmart
	0.05 μL AgeI-HF	0.05 μL AgeI-HF	0.05 μL AgeI-HF
	0.05 μL SalI-HF	0.05 μL BsiWI-HF	0.05 μL XhoI

3. Prepare enzyme mix 1 for Ig V region PCR products according to Table 3. Add 6 µL of enzyme mix 1 to purified PCR product plate, mix carefully, and seal with PCR film. Incubate for 2 h at 37 °C (*see* **Notes 31** and **32**).
4. Purify digested PCR products using the PCR purification kit, elute in 50 µL Tris–HCl (*see* **Note 30**).
5. Ligate 8 µL digested and purified gene-specific PCR product into 0.5 µL expression vector using 1 µL T4 DNA ligase reaction buffer and 0.5 µL T4 DNA ligase (*see* **Note 33**). Incubate for 2 h at room temperature or overnight at 16 °C.
6. Thaw competent bacteria on ice. Add 3 µL ligation product to 5 µL competent DH10B bacteria, and mix carefully. Incubate for 30 min. Perform a heat shock for 45 s at 42 °C, and put bacteria/ligation mix back on ice.
7. Add 200 µL LB medium and incubate at 37 °C for 30 min under moderate shaking.
8. Plate bacteria on ampicillin-containing LB agar plate. Incubate upside down overnight at 37 °C. Label colonies used for insert PCR to screen bacterial colonies for the presence of appropriately sized inserts.
9. Prepare PCR master mix 4 consisting of 19.1 µL water, 2.5 µL PCR buffer, 2.5 µL dNTP solution, 0.2 µL 5′ Ab sense primer, 0.2 µL 3′ primer, and 0.5 µL Taq DNA polymerase (*see* **Notes 22** and **34**).
10. Add PCR master mix 4 to new 96-well PCR plate, place on ice, and inoculate with bacterial colony (*see* **Notes 34** and **35**).
11. Run insert PCR at 94 °C for 15 min; 27 cycles at 94 °C for 30 s, 58 °C for 30 s, and 72 °C for 60 s; and 72 °C for 10 min.
12. Run PCR products on a 2% agarose gel to verify amplification (expected product sizes are 650 bp for AbVec2.0-IGHG1, 700 bp for AbVec1.1-IGKC, and 590 bp for AbVec1.1-IGLC2-XhoI).
13. Perform Sanger sequencing of insert PCR products with the expected size using Ab sense as sequencing primer (*see* **Note 36**).
14. Choose colonies with 100% sequence identity to the second PCR products (Subheading 3.2.2) to inoculate 4 mL Terrific broth with 75 µg/mL ampicillin in round bottom tube. Incubate for 16 h at 37 °C, 200 rpm.
15. Isolate plasmid DNA from 2 mL bacteria culture using a plasmid purification kit (*see* **Note 37**). Elute DNA in 75 µL elution buffer. Determine plasmid DNA concentrations at A_{260} and A_{260}/A_{280} ratios with a spectrophotometer (*see* **Note 38**).

3.4 Expression of Recombinant Monoclonal Antibodies

Using PEI-mediated transfection, antibody concentration in supernatant is on average 40 μg/mL, typically ranging between 10 and 150 μg/mL as determined by ELISA (*see* **Note 39**). Depending on the intended antibody binding assays, the antibody may be purified. For some applications, e.g., ELISA, the antibody-containing supernatant may be sufficient for reactivity testing. Functional Ig production should be confirmed by sodium dodecylsulfate polyacrylamide gel electrophoresis (SDS-PAGE) under nonreducing and reducing conditions [31].

3.4.1 Recombinant Antibody Production

1. Grow HEK 293T in 150 mm plates and 25 mL DMEM growth medium under standard conditions (37 °C, 5% CO_2; *see* **Note 40**). Ensure that cells are evenly spread across the plate.
2. At 80% confluency, remove medium completely and wash with 10 mL DMEM wash medium. Remove wash medium completely, and add 25 mL DMEM Nutridoma medium (*see* **Note 41**). Put plate back into the incubator.
3. Prepare a mix of NaCl, AbVec2.0-IGHG1, and corresponding AbVec1.1-IGKC or AbVec1.1-IGLC2-XhoI vector obtained from Subheading 3.3.2, **step 15** (ratio 150 mM NaCl/total vector DNA [μL/μg] = 50; *see* **Note 42**). Agitate on vortexer for 10 s, add PEI (ratio PEI/total vector DNA [μg/μg] = 3; *see* **Note 42**). Immediately agitate on vortexer for 30 s. Incubate transfection mix for 10 min at room temperature.
4. Gently add PEI mixture dropwise to each plate, and aim for even distribution of the mixture. Cultivate cells for 4 days under standard conditions.
5. Collect culture supernatant, centrifuge at 800 × *g* for 10 min, and transfer to a new reaction tube. Add sodium azide to a final concentration of 0.05%, and store at 4 °C (*see* **Notes 43** and **44**).

3.4.2 Recombinant Antibody Purification

All steps are conducted on ice while wearing gloves.

1. Equilibrate 25 μL of Protein G beads with PBS. Centrifuge for 10 min at 800 × *g*, 4 °C (*see* **Note 45**).
2. Incubate 25 μL Protein G beads with 25 mL culture supernatant overnight at 4 °C under rotation.
3. Equilibrate chromatography spin column with PBS.
4. Centrifuge Protein G beads/culture supernatant for 10 min at 800 × *g*, 4 °C.
5. Carefully remove the supernatant, resuspend beads in residual PBS, and transfer to spin column (*see* **Note 46**).
6. Wash column twice with 1 mL PBS.

7. Elute recombinant antibody in three fractions (200 μL each) with glycine and collect in tubes, each containing 20 μL Tris/NaN_3 (*see* **Notes 44** and **47**).
8. Measure protein concentration in spectrophotometer at A_{280} or in an ELISA [28].

4 Notes

1. Single-cell suspensions can be prepared from virtually any tissue or organ by standard procedures [32]. The appropriate protocol to obtain mononuclear cells depends on the tissue of interest. Mononuclear cells can be isolated from blood or bone marrow using EDTA or citrate blood collection tubes by Ficoll-Paque density gradient centrifugation after 1:2 or 1:10 dilution in medium, respectively. Mononuclear cells from tissue can be isolated by collagenase/DNase digest and subsequent Percoll-based density gradient purification [29]. After obtaining single-cell suspensions, either continue with the protocol or freeze the cells for later processing (*see* **Note 2**). Unless noted otherwise, keep cell suspensions on ice until sorted.
2. For freezing, resuspend cells in a cryotube in FCS + 10% (v/v) DMSO to a maximum concentration of 1×10^7 cells/mL. Freeze immediately at −80 °C using a Biocision CoolCell device, and transfer the tube into liquid nitrogen after 36–48 h. For thawing, first transfer the tube from liquid nitrogen to −80 °C for 2–3 h. Then quickly thaw it for 90 s in a 37 °C water bath, and directly transfer the content of the vial into 50 mL pre-warmed RPMI 1640 medium. Spin down (400 × *g*, 5 min) and wash once with 50 mL cold PBS/FCS. Finally, resuspend in the desired volume of PBS/FCS, and proceed with the protocol. Note that while the freeze/thaw procedure does have a minor impact on cell viability, it offers a substantial increase in workflow flexibility.
3. For successful cell sorting, target cells should not be substantially below 0.01% of total cells. Magnetic cell-enrichment strategies, e.g. by negative depletion for B-cell enrichment are possible prior to isolation by flow cytometry but result in significant cell loss. In addition, high target cell frequencies will result in low sorting efficiency, as repositioning of the PCR plate by the cell sorter becomes the rate-limiting step. As a rule-of-thumb, users should aim for flow rates of one target cell/s.
4. Titrate antibodies for use in flow cytometry. If possible, perform the titration on the tissue of interest.

5. If unlabeled or biotinylated antibody or antigen is used [33], repeat **steps 1–3** accordingly with secondary antibody or streptavidin-conjugate coupled to fluorochrome.
6. 7-AAD is used to exclude dead cells. Alternatively, amine-reactive dyes can be used, which offer a wider selection of fluorescence characteristics [34].
7. Pay attention to work under RNase-free conditions and on ice.
8. For single-cell sorting, use skirted 96-well PCR plates for better fixation on the cell sorter's plate handling device. It is recommended to use plates with a rigid frame to prevent tilting.
9. The seal should cover a temperature range from −80 °C to 105 °C and should be free of DNase, RNase, and DNA.
10. The previously published protocols used SuperScript III, which is still commercially available. However, we found that SuperScript IV has a superior performance for cell populations with low levels of Ig transcripts (e.g. mature naive or GC B cells), resulting in 10–20% increased amplification efficiency.
11. 5′ primers for gene-specific amplification are individually designed based on the V segment germline configuration [28]. Primers contain an AgeI restriction site, the last 3 codons of the leader sequence (not encoded in the vector), and typically the first 18 nt of the FWR1 of the respective V segment (Table 1), preferentially ending with C or G. Gene-specific J-segment primers containing a SalI restriction site for *IGH* or a BsiWI restriction site for *IGK* are listed in Table 1. The 3′ primer for *IGL* binds in the constant region and contains an XhoI site.
12. All expression vectors (Table 2) contain an ampicillin resistance gene and can be propagated in *E. coli* DH10B bacteria [28]. Prepare vectors using a Maxiprep Kit, e.g., NucleoBond Xtra Maxi (Macherey-Nagel), and follow the manufacturer's instructions. After plasmid purification, determine DNA concentration, and sequentially digest 25 μg of circular vector DNA with 50 U of the appropriate restriction enzymes and an extended incubation for 4 h. Load digested expression vectors onto an 0.8% agarose gel, and purify linearized expression vectors by using a gel extraction kit, e.g., NucleoSpin PCR clean-up (Macherey-Nagel). To test for successful linearization, perform ligations with and without a digested, purified V region PCR product as described in Subheading 3.3.2, **step 5**.
13. Depending on the frequency of the target population, aliquot the desired number of cells. Do not exceed 10^7 cells/100 μL.
14. Simultaneously prepare single stains with the same fluorochrome used in the actual staining procedure. Single stains

are important for setting up the instrument and for appropriate compensation. Use compensation beads instead of cells whenever possible.

15. Incubation with antibody or antigen solution may cross-link cells and thus may cause agglutination. DNase may be used to reduce this effect. Use of a cell strainer is essential.
16. This setup ensures that the cell droplet is collected in the 4 μL of lysis buffer. To avoid liquid evaporation, use a cooled stage.
17. Single-cell sorting and subsequent processing are also possible using 384-well plates [35, 36]. The smaller format and the associated lower amounts of reagents typically reduce the costs by a factor of three. Although it is advisable to train the manual handling and liquid transfer before working with actual samples, the contamination risk for experienced users is not increased.
18. Label 96-well plates appropriately ahead of cell sorting. Sort cells using forward vs. side scatter with doublet discrimination, and apply live cell gate.
19. For parallel processing in a 96-well plate, use a multichannel pipette to speed up the process.
20. Alternatively, plates may be used directly for cDNA synthesis after freezing on dry ice.
21. To avoid contaminations, always spin down single-cell plates prior to removal of the seal. Carefully peel off the seal.
22. Prepare a master mix for all reactions. Allow a surplus of 10%.
23. Primer sets for direct cloning of second round PCR products have been described [4, 15, 28, 37]. However, users should be aware that this approach does not allow isotype determination and often introduces amino acid replacement mutations due to primer mismatches.
24. Sequence analysis by IgBLAST (https://www.ncbi.nlm.nih.gov/igblast/) will identify the germline V(D)J segments with the highest identity, the amino acid composition of complementarity determining region 3 (CDR3), and the location of somatic mutations [38]. The analysis of the Ig repertoire and CDR3 characteristics provide information on the clonality of the B-cell population and their differentiation stage. Importantly, certain B-cell tumors including CLL, mantle cell lymphoma, follicular lymphoma, and primary central nervous system lymphoma frequently express characteristic *IGHV* segments with stereotyped CDR3 or show biased *IGKV* segment usage [13–15, 17, 39–45].
25. Determine V- and J-segment usage for gene-specific PCRs and expression vector cloning.

26. Determine CDR3 length as indicated by IgBLAST. Count the amino acid residues following the conserved cysteine at the end of framework region (FWR) 3 up to, but not including, the conserved motif in the J segment (tryptophan-glycine (WGXG) for *IGH*, and phenylalanin-glycine (FGXG) for *IGK* and *IGL*) [46, 47].
27. Cells that have undergone affinity maturation in the GC show clear signs of antigen-mediated selection, as evidenced by higher ratios of amino acid replacement to silent somatic mutations in CDRs than in FWRs. Unmutated V segments have been described in mantle cell lymphoma and a subtype of CLL [15, 40, 48]. Somatically mutated V segments are found in multiple myeloma and most types of lymphomas, e.g., follicular lymphoma, mucosa-associated lymphatic tissue lymphoma, hairy cell leukemia, and a subtype of CLL [49].
28. Determine the Ig isotype using Ensembl and the most recent human genome reference consortium assembly (https://www.ensembl.org/Homo_sapiens/Info/Index).
29. All steps in expression vector cloning can be performed in 96-well plates. If problems occur during cloning, switch to single tubes as this usually results in a better yield.
30. Determine DNA concentration with a spectrophotometer at A_{260}. To ensure for a sufficient yield, at least 300 ng DNA are required for restriction digest. A successful ligation is accomplished using at least 15 ng vector DNA and by aiming for a 3:1 ratio of insert to vector.
31. Include a restriction control, e.g., linearize 100 ng of the respective expression vector.
32. Confirm that V regions of somatically mutated antibodies do not contain restriction sites for the respective restriction enzymes. If so, perform a partial digest with tenfold lower enzyme concentration and only 5–15 min incubation time. In case of multiple internal restriction sites, consider gene synthesis of the respective V region with the restriction sites removed.
33. Include a ligation control if possible, e.g., a purified digested PCR product that has been successfully ligated into the corresponding expression vector.
34. Check three colonies per clone. If none of the inserts have the expected size, repeat cloning from the specific PCR.
35. Prepare backup plates for the colonies of interest on fresh LB agar plates containing ampicillin. Inoculate by touching bacterial colony with small pipette tip, touching a spot on the fresh LB agar plate, and then dipping pipette tip into PCR plate containing master mix 4. Incubate LB agar plates upside down over night at 37 °C.

36. Due to the use of error-prone Taq DNA polymerase and the fixation of these errors by cloning, 10% of all inserts will contain mutations as compared to the nucleotide sequences obtained from the respective second round PCR product [28]. Clones with mutated insert sequences that lead to amino acid exchanges are excluded from further analyses.
37. Use the remaining 2 mL of the bacterial culture as backup for plasmid DNA isolation, and store as cell pellet at −20 °C.
38. Usually, 20–30 μg plasmid DNAs are recovered from a 2 mL bacterial culture.
39. Production of antibodies can also be performed in suspension culture, using adapted HEK293 cells. This approach can double the antibody yields but requires special media and shaking incubators [50].
40. HEK 293T cells should be grown and passaged according to the instructions provided by DSMZ. Be aware that HEK 293T cells require at least 24 h to attach firmly to the culture plate. Ensure that passage number for HEK 293T cells is below 30, as cells may otherwise not produce sufficient amounts of antibody.
41. If cells detach while washing, change plate brand.
42. For example, when working with 10 μg heavy chain vector DNA and 10 μg light chain vector DNA, 1000 μL of 150 mM NaCl and 60 μg of PEI are added.
43. To increase antibody yield, supernatant can be replaced with fresh DMEM Nutridoma medium and harvested again 7 days after transfection.
44. Supernatant and purified antibodies may be stored at 4 °C for several months if sodium azide is added. If sodium azide interferes with the intended antibody reactivity assay, collect the supernatant under sterile conditions. Before use in experiments, spin down supernatant or purified antibody in benchtop centrifuge for 5 min at maximum speed and 4 °C. Determine antibody concentration after long-term storage.
45. If several antibodies are purified in parallel, adjust the required volume of Protein G beads in one tube including 5% surplus.
46. To prevent clogging, cut off pipette tip.
47. Fractions may be pooled after determining the protein concentration.

References

1. Tonegawa S (1983) Somatic generation of antibody diversity. Nature 302:575–581
2. Wardemann H, Nussenzweig MC (2007) B-cell self-tolerance in humans. Adv Immunol 95:83–110

3. Goodnow CC, Cyster JG, Hartley SB et al (1995) Self-tolerance checkpoints in B lymphocyte development. Adv Immunol 59:279–368
4. Wardemann H, Yurasov S, Schaefer A et al (2003) Predominant autoantibody production by early human B cell precursors. Science 301:1374–1377
5. Tiller T, Tsuiji M, Yurasov S et al (2007) Autoreactivity in human IgG+ memory B cells. Immunity 26:205–213
6. Berek C, Milstein C (1988) The dynamic nature of the antibody repertoire. Immunol Rev 105:5–26
7. Klein U, Dalla-Favera R (2008) Germinal centres: role in B-cell physiology and malignancy. Nat Rev Immunol 8:22–33
8. Muramatsu M, Kinoshita K, Fagarasan S et al (2000) Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. Cell 102:553–563
9. Rui L, Schmitz R, Ceribelli M et al (2011) Malignant pirates of the immune system. Nat Immunol 12:933–940
10. Küppers R, Klein U, Hansmann ML et al (1999) Cellular origin of human B-cell lymphomas. N Engl J Med 341:1520–1529
11. Küppers R (2005) Mechanisms of B-cell lymphoma pathogenesis. Nat Rev Cancer 5:251–262
12. Zenz T, Mertens D, Küppers R et al (2010) From pathogenesis to treatment of chronic lymphocytic leukaemia. Nat Rev Cancer 10:37–50
13. Agathangelidis A, Hadzidimitriou A, Rosenquist R et al (2011) Unlocking the secrets of immunoglobulin receptors in mantle cell lymphoma: implications for the origin and selection of the malignant cells. Semin Cancer Biol 21:299–307
14. Messmer BT, Albesiano E, Efremov DG et al (2004) Multiple distinct sets of stereotyped antigen receptors indicate a role for antigen in promoting chronic lymphocytic leukemia. J Exp Med 200:519–525
15. Hervé M, Xu K, Ng Y-S et al (2005) Unmutated and mutated chronic lymphocytic leukemias derive from self-reactive B cell precursors despite expressing different antibody reactivity. J Clin Invest 115:1636–1643
16. Catera R, Silverman GJ, Hatzi K et al (2008) Chronic lymphocytic leukemia cells recognize conserved epitopes associated with apoptosis and oxidation. Mol Med 14:665–674
17. Hadzidimitriou A, Darzentas N, Murray F et al (2009) Evidence for the significant role of immunoglobulin light chains in antigen recognition and selection in chronic lymphocytic leukemia. Blood 113:403–411
18. Bahler DW, Levy R (1992) Clonal evolution of a follicular lymphoma: evidence for antigen selection. Proc Natl Acad Sci U S A 89:6770–6774
19. Dühren-von Minden M, Übelhart R, Schneider D et al (2012) Chronic lymphocytic leukaemia is driven by antigen-independent cell-autonomous signalling. Nature 489:309–312
20. Minici C, Gounari M, Übelhart R et al (2017) Distinct homotypic B-cell receptor interactions shape the outcome of chronic lymphocytic leukaemia. Nat Commun 8:15746
21. Breden F, Luning Prak ET, Peters B et al (2017) Reproducibility and reuse of adaptive immune receptor repertoire data. Front Immunol 8:1418
22. Rubelt F, Busse CE, Bukhari SAC et al (2017) Adaptive immune receptor repertoire community recommendations for sharing immune-repertoire sequencing data. Nat Immunol 18:1274–1278
23. Wardemann H, Busse CE (2017) Novel approaches to analyze immunoglobulin repertoires. Trends Immunol 38:471–482
24. Steinitz M, Klein G, Koskimies S et al (1977) EB virus-induced B lymphocyte cell lines producing specific antibody. Nature 269:420–422
25. Lanzavecchia A, Corti D, Sallusto F (2007) Human monoclonal antibodies by immortalization of memory B cells. Curr Opin Biotechnol 18:523–528
26. Köhler G, Milstein C (1975) Continuous cultures of fused cells secreting antibody of predefined specificity. Nature 256:495–497
27. McCafferty J, Griffiths AD, Winter G et al (1990) Phage antibodies: filamentous phage displaying antibody variable domains. Nature 348:552–554
28. Tiller T, Meffre E, Yurasov S et al (2008) Efficient generation of monoclonal antibodies from single human B cells by single cell RT-PCR and expression vector cloning. J Immunol Methods 329:112–124
29. Benckert J, Schmolka N, Kreschel C et al (2011) The majority of intestinal IgA+ and IgG+ plasmablasts in the human gut are antigen-specific. J Clin Invest 121:1946–1955
30. Binder M, Müller F, Jackst A et al (2011) B-cell receptor epitope recognition correlates with the clinical course of chronic lymphocytic leukemia. Cancer 117:1891–1900
31. Tiller T, Busse CE, Wardemann H (2009) Cloning and expression of murine Ig genes

from single B cells. J Immunol Methods 350:183–193

32. Gu H, Rajewsky K (eds) (2004) B cell protocols. Humana Press, New York
33. Scheid JF, Mouquet H, Feldhahn N et al (2009) A method for identification of HIV gp140 binding memory B cells in human blood. J Immunol Methods 343:65–67
34. Perfetto SP, Chattopadhyay PK, Lamoreaux L et al (2006) Amine reactive dyes: an effective tool to discriminate live and dead cells in polychromatic flow cytometry. J Immunol Methods 313:199–208
35. Busse CE, Czogiel I, Braun P et al (2014) Single-cell based high-throughput sequencing of full-length immunoglobulin heavy and light chain genes. Eur J Immunol 44:597–603
36. Murugan R, Imkeller K, Busse CE et al (2015) Direct high-throughput amplification and sequencing of immunoglobulin genes from single human B cells. Eur J Immunol 45:2698–2700
37. Tsuiji M, Yurasov S, Velinzon K et al (2006) A checkpoint for autoreactivity in human IgM+ memory B cell development. J Exp Med 203:393–400
38. Ye J, Ma N, Madden TL et al (2013) IgBLAST: an immunoglobulin variable domain sequence analysis tool. Nucleic Acids Res 41:34–40
39. Smilevska T, Tsakou E, Hadzidimitriou A et al (2008) Immunoglobulin kappa gene repertoire and somatic hypermutation patterns in follicular lymphoma. Blood Cells Mol Dis 41:215–218
40. Fais F, Ghiotto F, Hashimoto S et al (1998) Chronic lymphocytic leukemia B cells express restricted sets of mutated and unmutated antigen receptors. J Clin Invest 102:1515–1525
41. Thompsett AR, Ellison DW, Stevenson FK et al (1999) VH gene sequences from primary central nervous system lymphomas indicate derivation from highly mutated germinal center B cells with ongoing mutational activity. Blood 94:1738–1746
42. Montesinos-Rongen M, Küppers R, Schlüter D et al (1999) Primary central nervous system lymphomas are derived from germinal-center B cells and show a preferential usage of the V4-34 gene segment. Am J Pathol 155:2077–2086
43. Stamatopoulos K, Kosmas C, Papadaki T et al (1997) Follicular lymphoma immunoglobulin κ light chains are affected by the antigen selection process, but to a lesser degree than their partner heavy chains. Br J Haematol 96:132–146
44. Noppe SM, Heirman C, Bakkus MH et al (1999) The genetic variability of the VH genes in follicular lymphoma: the impact of the hypermutation mechanism. Br J Haematol 107:625–640
45. Aarts WM, Bende RJ, Steenbergen EJ et al (2000) Variable heavy chain gene analysis of follicular lymphomas: correlation between heavy chain isotype expression and somatic mutation load. Blood 95:2922–2929
46. Kabat EA, Wu TT (1991) Identical V region amino acid sequences and segments of sequences in antibodies of different specificities. Relative contributions of VH and VL genes, minigenes, and complementarity-determining regions to binding of antibody-combining sites. J Immunol 147:1709–1719
47. Imkeller K, Arndt PF, Wardemann H et al (2016) sciReptor: analysis of single-cell level immunoglobulin repertoires. BMC Bioinformatics 17:67
48. Hummel M, Tamaru J, Kalvelage B et al (1994) Mantle cell (previously centrocytic) lymphomas express VH genes with no or very little somatic mutations like the physiologic cells of the follicle mantle. Blood 84:403–407
49. Klein U, Goossens T, Fischer M et al (1998) Somatic hypermutation in normal and transformed human B cells. Immunol Rev 162:261–280
50. Lorin V, Mouquet H (2015) Efficient generation of human IgA monoclonal antibodies. J Immunol Methods 422:102–110

Chapter 6

Studying the Replication History of Human B Lymphocytes by Real-Time Quantitative (RQ-)PCR

Menno C. van Zelm

Abstract

The cells of the adaptive immune system, B and T lymphocytes, each generate a unique antigen receptor through V(D)J recombination of their immunoglobulin (Ig) and T-cell receptor (TCR) loci, respectively. Such rearrangements join coding elements to form a coding joint and delete the intervening DNA as circular excision products containing the signal joint. These excision circles are stable structures that cannot replicate and have no function in the cell. Since the coding joint in the genome is replicated with each cell division, the ratio between coding joints and signal joints in a population of B cells can be used as a measure for proliferation. This chapter describes a real-time quantitative (RQ-)PCR-based approach to quantify proliferation through calculating the ratio between coding joints and signal joints of the frequently occurring intronRSS–Kde rearrangements in the *IGK* light chain locus. Besides its use in normal B-cell biology, quantification of B-cell replication can inform on abnormal proliferation in human diseases and in B-cell neogenesis following depletion therapy.

Key words B cell, Immunoglobulin, V(D)J recombination, KREC, Replication history, Proliferation, Real-time quantitative PCR, IntronRSS, Kde, IGK

Abbreviations

BCR	B-cell antigen receptor
D	Diversity
Ig	Immunoglobulin
IgH	Ig heavy chain
Igκ	Ig kappa light chain
Igλ	Ig lambda light chain
J	Joining
Kde	Kappa-deleting element
KREC	Kappa-deleting recombination excision circle
RAG	Recombination activating gene
RQ-PCR	Real-time quantitative PCR
RSS	Recombination signal sequence
V	Variable

Ralf Küppers (ed.), *Lymphoma: Methods and Protocols*, Methods in Molecular Biology, vol. 1956,
https://doi.org/10.1007/978-1-4939-9151-8_6,

1 Introduction

B cells have a unique role in the vertebrate immune system with the production of antibodies (immunoglobulin; Ig) that are capable of neutralizing invading pathogens. Importantly, antibody responses are highly antigen-specific and generate immunological memory: upon secondary encounter, the same pathogen is eliminated much more efficiently. These adaptive capabilities are the result of two independent stages of development [1]. First, each B cell creates a unique antibody during antigen-independent precursor differentiation in bone marrow, which is expressed on the surface membrane: B-cell antigen receptor (BCR). Together, all newly produced B cells have a large repertoire of antigen receptors with the potential to specifically recognize many different pathogens. The second phase of development starts when a B cell actually recognizes antigen in peripheral lymphoid organs. These activated B cells undergo enormous clonal proliferation, thereby generating huge numbers of daughter cells with the potential to recognize the same pathogen. This clonal expansion generates effector cells for a strong response and long-term memory in the form of memory B cells and Ig-producing plasma cells. The host requires a highly dynamic immune system, which maintains a tight balance between the production of a large repertoire of B cells with unique receptors and a strong immune response of groups of cells with an antigen-specific and thereby a more limited (selected) repertoire.

While clonal B-cell proliferation is important to generate a strong immune response, it should be tightly controlled to prevent malignant transformation. Until recently, the extent of properly controlled B-cell expansion following antigen recognition was unknown, due to the lack of techniques to assess in vivo proliferation. In 2007, we introduced a novel real-time quantitative (RQ-) PCR-based method to assess ex vivo the average replication history a population of B cells has undergone in vivo [2]. This method exploits the circular DNA excision products that are generated in precursor B cells that undergo V(D)J recombination.

Ig loci do not contain a functional first exon. Instead, they contain multiple variable (V), diversity (D), and joining (J) genes, which undergo ordered genomic rearrangement during antigen-independent differentiation to form a functional first exon (Fig. 1a) [3, 4]. First, D to J gene rearrangements are initiated in the Ig heavy chain locus (*IGH*), followed by V to DJ gene rearrangements. Upon successful completion, Ig light chain rearrangements are induced in the *IGK* locus, leading to direct V to J coupling (Fig. 1a). When *IGK* light chain rearrangements are not successful, V to J gene rearrangements are induced in the *IGL* light chain locus. Thus, mature B cells can exit the bone marrow expressing an antigen receptor with either an Igκ or an Igλ isotype.

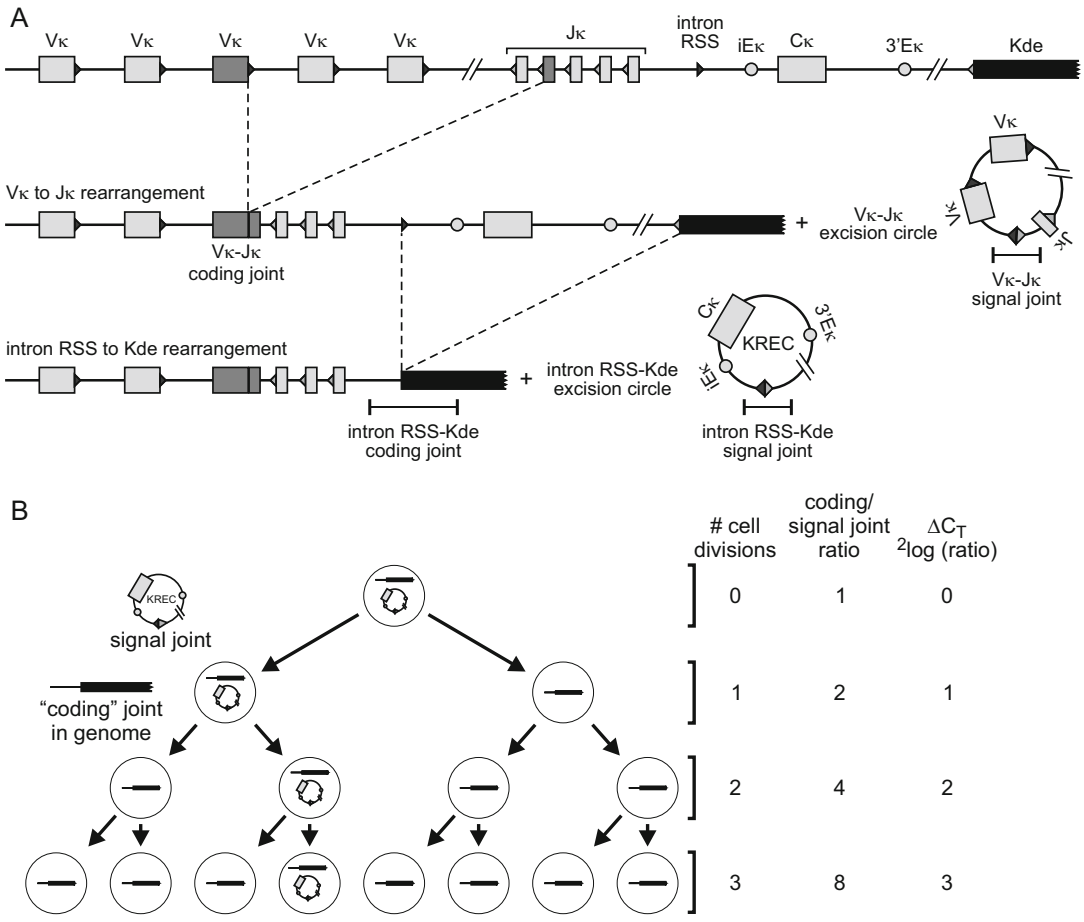


Fig. 1 Detection of intronRSS–Kde coding joints and signal joints to quantify B cell replication history. (**a**) V(D)J recombination on the *IGK* locus results in a VκJκ coding joint. Subsequent rearrangement between the intronRSS and the Kde elements can render the *IGK* allele nonfunctional by deleting the Cκ exons and the enhancers. Consequently, the coding joint precludes any further rearrangements in the *IGK* locus and therefore remains present in the genome, whereas an intronRSS–Kde signal joint is formed on an excision circle (KREC). (**b**) When a B lymphocyte with an intronRSS–Kde rearrangement divides, both daughter cells inherit the intronRSS–Kde coding joint in the genome. However, the signal joint, which is on the episomal KREC, will be inherited by only one of the two daughter cells. Crucially, the ΔC_T of the PCRs detecting the coding joint and the signal joint exactly represent the number of cell divisions a B lymphocyte has undergone, because both processes have an exponential increase with base number 2. Figure panels are reproduced from van Zelm et al. [2] with permission from Rockefeller University Press

V(D)J recombination of Ig loci is initiated by the recombination activating gene products RAG1 and RAG2 that introduce double-stranded DNA breaks on the borders of two genes and their flanking recombination signal sequences (RSS). The resulting coding ends of the genes are ligated together to form a coding joint. The intervening DNA is excised and circularized by ligating the blunt DNA signal ends, thereby forming a signal joint. While multiple of these episomal recombination excision circles are

formed in every developing B cell, these do not appear to have a specific function. Excision circles are stable products, which cannot replicate in the cell, do not duplicate during mitosis, and consequently are diluted during cell division, while the corresponding rearrangement on genomic DNA (coding joint) is inherited by all daughter cells [2, 5, 6].

Since a daughter cell either carries a circle or not, it is not possible to study turnover or proliferation of single cells. In fact, measurement of turnover or proliferation via excision circles is performed at the level of a cell population. By comparing the number of detected signal joints on excision circles with the total number of coding joints in chromosomal DNA, the average proliferation is established. The condition to detect signal joints at the population level requires selection of a common rearrangement. However, in each developing B cell, a unique combination of V, D, and J genes is formed to facilitate generation of a large repertoire of Igs. Therefore, most rearrangements are formed to generate a functional B- or T-cell receptor and will consequently not be useful as targets to study proliferation.

Fortunately, most B cells initiate a single-step gene rearrangement to render a specific locus nonfunctional. This concerns the rearrangement between the intronRSS and the kappa-deleting element (Kde), which fits all the criteria for optimal measurement of proliferation [2]: (a) it is a frequently occurring gene rearrangement that deletes the constant region of 35–40% *IGK* alleles; (b) it is one of the last Ig gene rearrangements to occur in bone marrow (Fig. 1a) [4], ensuring that the corresponding kappa-deletion recombination excision circles (KRECs) are abundantly present in naive B lymphocytes; (c) it is a single-step rearrangement, which allows easy design of RQ-PCR primers and probes for accurate detection of the coding joints and signal joints; and (d) it is an end-stage rearrangement that deletes the enhancers, precluding further rearrangements.

The KREC assay depends on the dilution of signal joints on excision circles: nearly all formed B cells contain a KREC, while only one in many antigen-experienced B cells carries excision circles. Thus, when measuring the replication history of total B cells, the KREC content of newly formed B cells has a dominant effect and will result in a low number of cell divisions. It is therefore recommended to measure the replication history in well-defined, purified B-cell subsets.

In this context, the KREC assay has been applied to study the contribution of proliferation to naive B-cell homeostasis and to antigen-induced B-cell proliferation in memory B-cell subsets from T-cell dependent and T-cell independent origin [2, 7, 8].

Furthermore, abnormal homeostatic proliferation of naive B cells was detected using the KREC assay in patients with primary immunodeficiencies [7, 9, 10], as well as aberrant proliferation of

memory B cells in patients with HIV chronic infection [11], immunodeficient patients [7], Down syndrome [12], and patients with chronic autoinflammatory diseases [13, 14].

Several studies have shown the potential of detecting KRECs in DNA from full blood or mononuclear cells without the intronRSS–Kde coding joint, as a marker for circulating B cells. Although this can be readily applied to newborn screening assays to detect absence of B cells [15–17], it is recommended to perform the coding joint assay when quantification of cell divisions is aimed for or to confirm absence of B cells following a negative KREC result [18].

Analogously, this is recommended in the context of lymphoproliferations: the presence of an enlarged B-cell population with intronRSS–Kde rearrangements and low to absent KRECs is highly suggestive of abnormal proliferation, as has been shown for persistent polyclonal B-cell lymphocytosis (PPBL) [19]. Finally, B-cell reconstitution following therapy, e.g., stem cell transplantation or B-cell depletion, can be monitored with respect to production of new cells versus peripheral expansion through proliferation [20].

2 Materials

2.1 High-Speed Cell Sorting

1. Phosphate-buffered saline (PBS): dissolve 8 g of NaCl, 0.2 g of KCl, 0.12 g of KH_2PO_4, and 1.67 g of $Na_2HPO_4{\cdot}2H_2O$ in approx. 800 mL of water, adjust pH to 7.8 with HCl, and add water to 1 L.
2. PBS/bovine serum albumin (BSA): PBS supplemented with 0.2% BSA and 0.1% azide.
3. Antibodies: *see* Table 1.
4. Fluorescence-activated cell sorter (FACS), e.g., FACSAria (BD Biosciences).
5. Fetal calf serum (FCS).

2.2 DNA Isolation and TaqMan-Based Real-Time Quantitative (RQ-)PCR

1. Column-based miniprep DNA isolation kit, from, e.g., Sigma-Aldrich or Qiagen.
2. Spectrophotometer, e.g., NanoDrop.
3. Universal PCR Master Mix (Applied Biosystems), containing 5 mM $MgCl_2$, 200 μM dNTPs, 0.1 U Taq Gold DNA polymerase, and 10% glycerol.
4. 2% BSA.
5. Primers, in a 30 pmol/μL working dilution (Table 2).
6. Hydrolysis probes with a 5′ fluorescent reporter (e.g., FAM, HEX, VIC) in combination with a 3′ fluorescent quencher

Table 1
Fluorescently labeled antibodies for high-speed cell sorting of human B-cell subsets[a]

Antibody against	Clone	Fluorochrome	Supplier
CD19	J3-119	PC7	Beckman Coulter
CD21	B-ly4	BV711	BD Biosciences
CD27	M-T271	BV421	BD Biosciences
CD38	HB7	APC-H7	BD Biosciences
IgM	MHM-88	BV510	BioLegend
IgD	IA6-2	PE-CF594	BD Biosciences
IgA	IS11-8E10	FITC	Miltenyi Biotec
IgA	IS11-8E10	PE	Miltenyi Biotec
IgG	IS11-3B2.2.3	PE	Miltenyi Biotec
IgE	Goat-anti-human (polyclonal)	FITC	Invitrogen

[a]As used in [22]

Table 2
Oligonucleotide characteristics for intronRSS–Kde coding joint and signal joint and albumin control gene quantification

Target (size)	Oligonucleotide	Sequence (5′–3′)
Albumin[a] (118 bp)	F-Alb R-Alb T-Alb[c]	CCT.GAT.CCT.CTT.GTC.CCA.CAG GGA.TTT.AGA.GTC.TCT.CAG.CTG.GTA.CA TGC.TGA.AAC.ATT.CAC.CTT.CCA.TGC.AGA
IntronRSS–Kde coding joint[b] (144 bp)	Intron-F1-TM Kde-R1-TM T-Kde-cons1[c]	CCC.GAT.TAA.TGC.TGC.CGT.AG CCT.AGG.GAG.CAG.GGA.GGC.TT AGC.TGC.ATT.TTT.GCC.ATA.TCC.ACT.ATT.TGG.AGT
IntronRSS–Kde signal joint[b] (148 bp)	Int-Kde KREC F Int-Kde KREC R T-Kde-RSS_2[c]	TCA.GCG.CCC.ATT.ACG.TTT.CT GTG.AGG.GAC.ACG.CAG.CC CCA.GCT.CTT.ACC.CTA.GAG.TTT.CTG.CAC.GG

[a]Refs. 18, 23
[b]Ref. 2
[c]5′-FAM- and 3′-TAMRA-labeled probes

(e.g., TAMRA) or a 3′ dark quencher (e.g., Iowa Black FQ), in a 5 pmol/μL working dilution (Table 2).

7. Genomic DNA of the control cell line U698-DB01 that contains one intronRSS–Kde coding joint and one signal joint per genome (Invivoscribe; *see* **Note 1**) [2].

8. Thermocycler that is capable to detect FAM signals from TaqMan-based hydrolysis probes, e.g., StepOnePlus or QuantStudio 6 Flex (Applied Biosystems).

3 Methods

3.1 High-Speed Cell Sorting

1. Start with mononuclear cells following Ficoll density separation or with presorted B cells, e.g., via CD19+ enrichment with magnetic beads (Miltenyi Biotec).
2. Resuspend cell pellet after final wash in 500 μL PBS/BSA per 50×10^6 cells.
3. Add appropriate amounts of antibodies of choice (Table 1), vortex shortly, and incubate sample for 10 min in the dark at room temperature.
4. Wash sample once with 10–20× labeling volume of PBS/BSA.
5. Resuspend cells to a final concentration of 25×10^6 cells/mL in PBS/BSA.
6. Define B-cell populations with electronic gating on FACS (Fig. 2) and perform high-speed cell sorting.
7. Collect specified B-cell populations in tubes that contain a small volume (0.1–0.5 mL) of FCS.
8. Upon completion of cell sorting, spin cells down and start DNA isolation procedure.

3.2 DNA Isolation

1. Perform DNA isolation on 30,000 or more cells using a column-based kit according to manufacturer's instructions (*see* **Note 2**).
2. Perform elution of DNA from column with elution buffer preheated to 55 °C. When starting with 30,000–200,000 cells, elute in 50 μL; 200,000–2 million cells in 100 μL; 2–5 million cells in 200 μL. In case of low cell numbers, the eluate can be reloaded on the column for a consecutive elution to increase the DNA yield.
3. Measure DNA quantity and purity on spectrophotometer.
4. If concentration is >15 ng/μL, prepare working dilution of 10 ng/μL in water.

3.3 TaqMan-Based Real-Time Quantitative (RQ-)PCR

1. Prepare three master mixes for independent RQ-PCR amplification of intronRSS–Kde coding joints and signal joints and for the albumin control gene. Perform these three reactions in duplicate on each sample, on the control cell line U698-DB01 and on a non-template (water) control (*see* **Note 3**).
2. The master mix consists of 12.5 μL Universal PCR Master Mix, 0.75 μL of both the forward and the reverse primer (30 pmol/

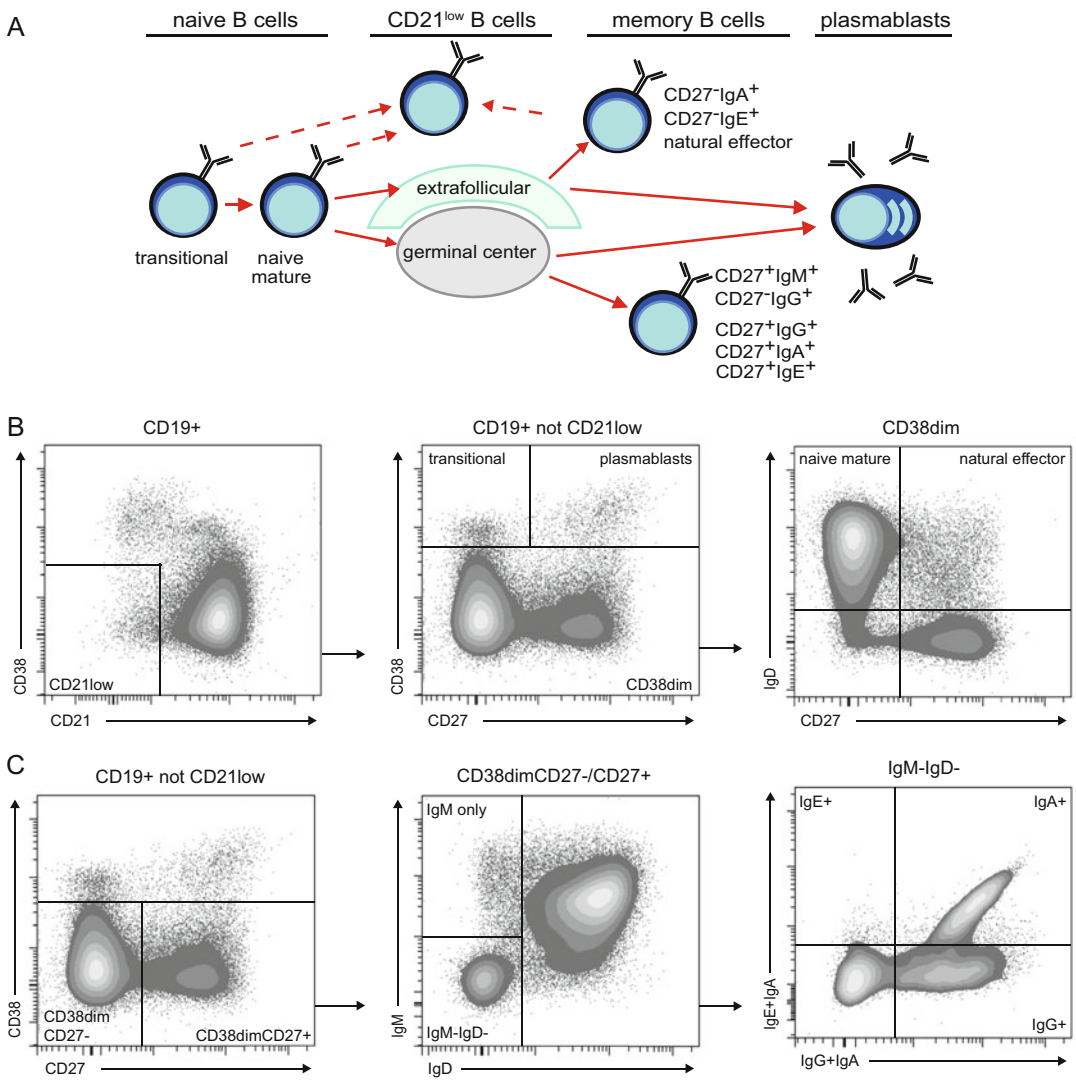


Fig. 2 Definition of human B cell subsets in blood. (**a**) Scheme of human B-cell maturation. Adapted from [12]. (**b**, **c**) Flowcytometric definition of human B-cell subsets utilizing the markers listed in Table 1. Adapted from [22]

μL), 0.5 μL hydrolysis probe (5 pmol/μL), 0.5 μL 2% BSA, and 5 μL water per reaction.

3. Add 5 μL template DNA (working dilution 10 ng/μL) or water per reaction well.
4. Adjust the PCR program for FAM detection and TAMRA quenching. The thermocycler program consists of a 50 °C step for 2 min and 95 °C for 10 s, followed by 50× (95 °C for 15 s and 60 °C for 1 min).
5. Upon completion of RQ-PCR reaction, set the threshold for signal detection such that all samples are in the exponential

amplification phase. For consistency, it is preferred to set the same threshold for all three PCR reactions (intronRSS–Kde coding joint, signal joint, and albumin).

6. Extract the cycle threshold (C_T) value for each reaction, and calculate the average C_T value for duplicate reactions. Duplicates should differ <1 C_T.

3.4 Calculation of Rearranged IntronRSS–Kde Alleles and the B-Cell Replication History

1. The ratio between the amount of genomic albumin and the intronRSS–Kde coding joints is indicative for the frequency of *IGK* alleles with an intronRSS–Kde rearrangement. Since the data are derived from independent PCRs, any technical variation is corrected for using the results on the U698-DB01 cell line which has one intronRSS–Kde coding joint per genome. The frequency of *IGK* alleles containing the intronRSS–Kde coding joint is thus calculated as follows (*see* **Note 4**):

$$2\left[\left(C_{T_{\text{albumin}}}-C_{T_{\text{codingjoint}}}\right)_{\text{sample}}-\left(C_{T_{\text{albumin}}}-C_{T_{\text{codingjoint}}}\right)_{\text{cellline}}\right]\cdot 100\%$$

2. The difference in C_T values for the intronRSS–Kde coding joint and signal joints is directly related to the replication history of the B-cell subset (Fig. 1b). Still, also for this calculation, the results of the U698-DB01 cell line have to be used to correct for differences in PCR efficiencies as follows:

$$\left(C_{T_{\text{signal joint}}}-C_{T_{\text{coding joint}}}\right)_{\text{sample}}-\left(C_{T_{\text{signal joint}}}-C_{T_{\text{coding joint}}}\right)_{\text{cell line}}$$

It is important to note that the increased proliferation will lead to reduced signal joints on KRECs (*see* **Note 5**).

4 Notes

1. The Igκ$^+$ B-cell line U698-M has two Vκ–Jκ rearranged IGK alleles, one of which is functional and has the intronRSS and Kde elements in germline configuration. The second allele has an out-of-frame VκJκ joint and contains an intronRSS–Kde rearrangement. In order to obtain an equal ratio between the intronRSS–Kde coding and signal joints, an intronRSS–Kde signal joint construct was inserted into the genome of the U698-M cell line using retroviral transduction. Individual clones were sorted, and the DB01 clone was found to contain a single insertion per genome, similar to the intronRSS–Kde coding joint [2]. The primer-probe sets for both the coding and signal joint are optimized and show nearly equal efficiencies. Still, minimal variation (<1 C_T) can occur. Therefore, it is

recommended to correct for this technical variation in each experiment.

2. It is recommended to perform the KREC assay on DNA isolated from 30,000 cells or more for reliable signal joint amplification. When low replication histories are expected, it is still possible to obtain reliable signal joint amplification from fewer cells. However, for DNA isolation from low cell numbers, we recommend to use a direct DNA lysis method [21].
3. For optimal calculation of the frequency of rearranged intronRSS–Kde alleles and B-cell replication history, it is recommended to perform the coding joint, signal joint, and albumin control PCRs in the same run. As with all RQ-PCR experiments, we recommend to perform each reaction at least in duplicate. In addition, each run should contain a non-template control and the U698-DB01 control for correction of technical variation (*see* **Note 1**). When expecting low to undetectable signal joint numbers, it is recommended to include serial dilutions of the U698-DB01 for the coding and signal joint PCRs to determine their quantitative ranges.
4. In addition to the intronRSS–Kde coding to signal joint ratio, it is advised to calculate the frequency of rearranged coding joints with the albumin control gene. While this is not crucial for the actual replication history levels, it provides valuable information on how representative this is for the whole population. In general, 35–50% of *IGK* alleles in mature B cells contain an intronRSS–Kde coding joint. While Igλ+ B cells contain higher levels than Igκ+ B cells [4], these rearrangements are independent of the antigen specificity. Thus, substantially high frequencies of intronRSS–Kde coding joints (>35%) reflect a good representation for the total B-cell subset.
5. With each cell division, the V(D)J recombination excision circles are diluted in the total B-cell population. In memory B-cell subsets of healthy controls, the replication history can be up to 10–12 cell divisions. This means that one circle is detected for 2^{10}–2^{12} (i.e., ~1000–4000) coding joints in the genome. Thus, when high replication histories are expected, it is important to start with high amounts of template DNA to maximize the quantitative range.

Acknowledgments

MCvZ is supported by NHMRC Senior Research Fellowship 1117687.

References

1. Berkowska MA, van der Burg M, van Dongen JJM, van Zelm MC (2012) Checkpoints of B cell differentiation: visualizing Ig-centric processes. Ann N Y Acad Sci 1246:11–25
2. van Zelm MC, Szczepanski T, van der Burg M, van Dongen JJ (2007) Replication history of B lymphocytes reveals homeostatic proliferation and extensive antigen-induced B cell expansion. J Exp Med 204:645–655
3. Ghia P, ten Boekel E, Sanz E, de la Hera A, Rolink A, Melchers F (1996) Ordering of human bone marrow B lymphocyte precursors by single-cell polymerase chain reaction analyses of the rearrangement status of the immunoglobulin H and L chain gene loci. J Exp Med 184:2217–2229
4. van Zelm MC, van der Burg M, de Ridder D, Barendregt BH, de Haas EF, Reinders MJ, Lankester AC, Revesz T, Staal FJ, van Dongen JJ (2005) Ig gene rearrangement steps are initiated in early human precursor B cell subsets and correlate with specific transcription factor expression. J Immunol 175:5912–5922
5. Livak F, Schatz DG (1996) T-cell receptor alpha locus V(D)J recombination by-products are abundant in thymocytes and mature T cells. Mol Cell Biol 16:609–618
6. Breit TM, Verschuren MC, Wolvers-Tettero IL, Van Gastel-Mol EJ, Hahlen K, van Dongen JJ (1997) Human T cell leukemias with continuous V(D)J recombinase activity for TCR-delta gene deletion. J Immunol 159:4341–4349
7. Driessen GJ, van Zelm MC, van Hagen PM, Hartwig NG, Trip M, Warris A, de Vries E, Barendregt BH, Pico I, Hop W, van Dongen JJ, van der Burg M (2011) B-cell replication history and somatic hypermutation status identify distinct pathophysiologic backgrounds in common variable immunodeficiency. Blood 118:6814–6823
8. Berkowska MA, Heeringa JJ, Hajdarbegovic E, van der Burg M, Thio HB, van Hagen PM, Boon L, Orfao A, van Dongen JJ, van Zelm MC (2014) Human IgE(+) B cells are derived from T cell-dependent and T cell-independent pathways. J Allergy Clin Immunol 134 (688–697):e686
9. van der Burg M, Pac M, Berkowska MA, Goryluk-Kozakiewicz B, Wakulinska A, Dembowska-Baginska B, Gregorek H, Barendregt BH, Krajewska-Walasek M, Bernatowska E, van Dongen JJ, Chrzanowska KH, Langerak AW (2010) Loss of juxtaposition of RAG-induced immunoglobulin DNA ends is implicated in the precursor B-cell differentiation defect in NBS patients. Blood 115:4770–4777
10. Rakhmanov M, Keller B, Gutenberger S, Foerster C, Hoenig M, Driessen G, van der Burg M, van Dongen JJ, Wiech E, Visentini M, Quinti I, Prasse A, Voelxen N, Salzer U, Goldacker S, Fisch P, Eibel H, Schwarz K, Peter HH, Warnatz K (2009) Circulating CD21low B cells in common variable immunodeficiency resemble tissue homing, innate-like B cells. Proc Natl Acad Sci U S A 106:13451–13456
11. Moir S, Ho J, Malaspina A, Wang W, DiPoto AC, O'Shea MA, Roby G, Kottilil S, Arthos J, Proschan MA, Chun TW, Fauci AS (2008) Evidence for HIV-associated B cell exhaustion in a dysfunctional memory B cell compartment in HIV-infected viremic individuals. J Exp Med 205:1797–1805
12. Verstegen RH, Driessen GJ, Bartol SJ, van Noesel CJ, Boon L, van der Burg M, van Dongen JJ, de Vries E, van Zelm MC (2014) Defective B-cell memory in patients with Down syndrome. J Allergy Clin Immunol 134:1346–1353.e1349
13. Timmermans WM, van Laar JA, van der Houwen TB, Kamphuis LS, Bartol SJ, Lam KH, Ouwendijk RJ, Sparrow MP, Gibson PR, van Hagen PM, van Zelm MC (2016) B-Cell Dysregulation in Crohn's Disease Is Partially Restored with Infliximab Therapy. PLoS One 11:e0160103
14. van der Houwen TB, van Hagen PM, Timmermans WM, Bartol SJ, Lam KH, Kappen JH, van Zelm MC, van Laar JA (2017) Chronic signs of memory B cell activation in patients with Behcet's disease are partially restored by anti-tumour necrosis factor treatment. Rheumatology (Oxford) 56:134–144
15. Nakagawa N, Imai K, Kanegane H, Sato H, Yamada M, Kondoh K, Okada S, Kobayashi M, Agematsu K, Takada H, Mitsuiki N, Oshima K, Ohara O, Suri D, Rawat A, Singh S, Pan-Hammarstrom Q, Hammarstrom L, Reichenbach J, Seger R, Ariga T, Hara T, Miyawaki T, Nonoyama S (2011) Quantification of kappa-deleting recombination excision circles in Guthrie cards for the identification of early B-cell maturation defects. J Allergy Clin Immunol 128 (223–225):e222
16. Borte S, von Dobeln U, Fasth A, Wang N, Janzi M, Winiarski J, Sack U, Pan-Hammarstrom Q, Borte M, Hammarstrom L (2012) Neonatal screening for severe primary immunodeficiency diseases using high-

throughput triplex real-time PCR. Blood 119:2552–2555

17. Kamae C, Nakagawa N, Sato H, Honma K, Mitsuiki N, Ohara O, Kanegane H, Pasic S, Pan-Hammarstrom Q, van Zelm MC, Morio T, Imai K, Nonoyama S (2013) Common variable immunodeficiency classification by quantifying T-cell receptor and immunoglobulin kappa-deleting recombination excision circles. J Allergy Clin Immunol 131:1437–1440.e1435
18. Van Zelm MC, Van Der Burg M, Langerak AW, Van Dongen JJM (2011) PID comes full circle: Applications of V(D)J recombination excision circles in research, diagnostics and newborn screening of primary immunodeficiency disorders. Front Immunol 2:1–9
19. Berkowska MA, Grosserichter-Wagener C, Adriaansen HJ, de Ridder D, Mirani-Oostdijk KP, Agteresch HJ, Bottcher S, Orfao A, van Dongen JJ, van Zelm MC (2014) Persistent polyclonal B-cell lymphocytosis: extensively proliferated CD27+IgM+IgD+ memory B cells with a distinctive immunophenotype. Leukemia 28:1560–1564
20. Skert C, Perucca S, Chiarini M, Giustini V, Sottini A, Ghidini C, Martellos S, Cattina F, Rambaldi B, Cancelli V, Malagola M, Turra A, Polverelli N, Bernardi S, Imberti L, Russo D (2017) Sequential monitoring of lymphocyte subsets and of T-and-B cell neogenesis indexes to identify time-varying immunologic profiles in relation to graft-versus-host disease and relapse after allogeneic stem cell transplantation. PLoS One 12:e0175337
21. van der Burg M, Kreyenberg H, Willasch A, Barendregt BH, Preuner S, Watzinger F, Lion T, Roosnek E, Harvey J, Alcoceba M, Diaz MG, Bader P, van Dongen JJ (2011) Standardization of DNA isolation from low cell numbers for chimerism analysis by PCR of short tandem repeats. Leukemia 25:1467–1470
22. Heeringa JJ, Fieten KB, Bruins FM, van Hoffen E, Knol EF, Pasmans S, van Zelm MC (2018) Treatment for moderate to severe atopic dermatitis in alpine and moderate maritime climates differentially affects helper T cells and memory B cells in children. Clin Exp Allergy 48:679–690
23. Pongers-Willemse MJ, Verhagen OJ, Tibbe GJ, Wijkhuijs AJ, de Haas V, Roovers E, van der Schoot CE, van Dongen JJ (1998) Real-time quantitative PCR for the detection of minimal residual disease in acute lymphoblastic leukemia using junctional region specific TaqMan probes. Leukemia 12:2006–2014

Chapter 7

Stereotyped B Cell Receptor Immunoglobulins in B Cell Lymphomas

Andreas Agathangelidis, Fotis Psomopoulos, and Kostas Stamatopoulos

Abstract

Comprehensive analysis of the clonotypic B cell receptor immunoglobulin (BcR IG) gene rearrangement sequences in patients with mature B cell neoplasms has led to the identification of significant repertoire restrictions, culminating in the discovery of subsets of patients expressing highly similar, stereotyped BcR IG. This finding strongly supports selection by common epitopes or classes of structurally similar epitopes in the ontogeny of these tumors. BcR IG stereotypy was initially described in chronic lymphocytic leukemia (CLL), where the stereotyped fraction of the disease accounts for a remarkable one-third of patients. However, subsequent studies showed that stereotyped BcR IG are also present in other neoplasms of mature B cells, including mantle cell lymphoma (MCL) and splenic marginal zone lymphoma (SMZL). Subsequent cross-entity comparisons led to the conclusion that stereotyped IG are mostly "disease-specific," implicating distinct immunopathogenetic processes. Interestingly, mounting evidence suggests that a molecular subclassification of lymphomas based on BcR IG stereotypy is biologically and clinically relevant. Indeed, particularly in CLL, patients assigned to the same subset due to expressing a particular stereotyped BcR IG display remarkably consistent biological background and clinical course, at least for major and well-studied subsets. Thus, the robust assignment to stereotyped subsets may assist in the identification of mechanisms underlying disease onset and progression, while also refining risk stratification. In this book chapter, we provide an overview of the recent BcR IG stereotypy studies in mature B cell malignancies and outline previous and current methodological approaches used for the identification of stereotyped IG.

Key words Immunoglobulin gene, VH CDR3, Antigen, Pattern, Stereotypy, Subset, Bioinformatics

1 Introduction

1.1 Antigen Selection in Mature B Cell Neoplasms

Studies in the last 3 decades have established a critical role for antigenic interactions in the pathogenesis of mature B cell malignancies. From an immunogenetic perspective, this notion is supported by biases in the expressed repertoires of immunoglobulin (IG) heavy and light variable genes (IGHV and IGKV/IGLV, respectively) [1, 2]. Moreover, the somatic hypermutation (SHM) patterns within the clonotypic IG genes of many lymphomas/leukemias are considered typical of antigen selection [3]. Longitudinal

Ralf Küppers (ed.), *Lymphoma: Methods and Protocols*, Methods in Molecular Biology, vol. 1956,
https://doi.org/10.1007/978-1-4939-9151-8_7,

studies and intraclonal analysis also led to the conclusion that, in some cases, antigen-induced SHM may remain active even after the neoplastic transformation [4]. Strong functional support for the importance of external drive emerged from ex vivo studies revealing that the malignant B cells are not able to survive or proliferate autonomously, indicating that external signals emanating from the tumor microenvironment are crucial for both sustaining the tumor and allowing its further expansion [5]: these signals are probably of different types and sources, yet, at least in some cases, they involve antigens [6]. Overall, it is now widely accepted that antigen selection shapes the IG repertoire of mature B cell malignancies by endowing the clonogenic progenitors and, in some instances, the malignant cells themselves, with pro-survival and proliferation signals perceived through clonotypic B cell receptor immunoglobulin (BcR IG) with distinctive immunogenetic features [3].

More recently, novel agents inhibiting downstream kinases in B cell signaling pathways [ibrutinib for Bruton's tyrosine kinase (BTK) and idelalisib for the phosphatidylinositol 3-kinase (PI3K), respectively] proved remarkably efficacious in a wide range of B cell lymphomas [including, among others, chronic lymphocytic leukemia (CLL), mantle cell lymphoma (MCL), lymphoplasmacytic lymphoma, marginal zone lymphomas, follicular lymphoma, and diffuse large B cell lymphoma] [7, 8]. This in vivo evidence highlights the importance of microenvironmental interactions in lymphomagenesis, prompting further analysis of the implicated immune pathways and processes, inevitably extending to the study of BcR IG structures. Hence, immunogenetics emerged as highly relevant to the study of B cell lymphomas.

1.2 B Cell Receptor Immunoglobulin Stereotypy: It All Started with CLL

The first solid evidence for IGHV gene repertoire restrictions in CLL was published in 1998, when it was also shown that a significant fraction of CLL patients carry somatically hypermutated IGHV genes [9]. Soon thereafter, two independent studies reported a strong correlation between the SHM status of the clonotypic rearranged IGHV genes and clinical course and outcome [10, 11]. In more detail, patients expressing clonotypic IGHV genes with an identity percent of $\geq$98% against the closest germline gene (unmutated CLL, U-CLL) were found to display adverse biological and clinical features, experienced fast progression requiring treatment, and often exhibited resistance to therapy, thus sharply contrasting patients with a higher mutational load (<98% IGHV gene germline identity; mutated CLL, M-CLL) who experienced a far more indolent disease course [10, 11].

These seminal observations were corroborated by all subsequent studies, firmly establishing the SHM status of the clonotypic IGHV genes as one of the most robust prognostic biomarkers in CLL. Growing experience has led to a high level of standardization regarding both the laboratory procedure and the bioinformatics analysis,

hence rendering the determination of SHM status feasible in diagnostic routine [12–14]. Relevant to mention in this respect, the International Workshop on CLL (iwCLL) considers this particular laboratory test as mandatory prior to treatment initiation in all CLL cases, in both general practice and clinical trials [15].

Besides the clinical value of immunogenetic analysis in CLL, the availability of relevant sequence data enabled further important observations to be made concerning the antigen-binding sites of the clonotypic IG. Studies from the mid-1990s first discovered that different, unrelated CLL patients could express highly similar IG receptors with restricted usage of IGHV/IGHD/IGHJ genes [16, 17]. In-depth analysis of IG gene restrictions led to the identification of "prototypic" IG rearrangements in CLL carrying very similar VH complementarity-determining region (CDR) three amino acid sequences: a characteristic example concerns the IGHV1-69/IGHD3-3/IGHJ6 gene rearrangement that encodes a long VH CDR3 enriched in tyrosine residues [18]. Later on, it was demonstrated that virtually half of CLL cases expressing BcR IG encoded by the IGHV3-21 gene had short and highly similar VH CDR3 and also carried lambda light chains utilizing the IGLV3-21 gene displaying VL CDR3 restrictions [19] (Fig. 1). Similar restrictions were soon reported by independent studies revealing the existence of subsets of cases, concerning both M-CLL and U-CLL, with (quasi)identical VH CDR3 (and, often, VL CDR3 as well) [20–24]. Considering that the possibility of this occurring by chance is extremely remote, this finding strongly argues against serendipity and alludes to an antigen-driven path to CLL development.

The term "stereotyped" i.e., repeated without variation, was first coined in 2004 by Chiorazzi's group for describing such (quasi)identical BcR IG [21] and, since then, has been widely adopted by the scientific community.

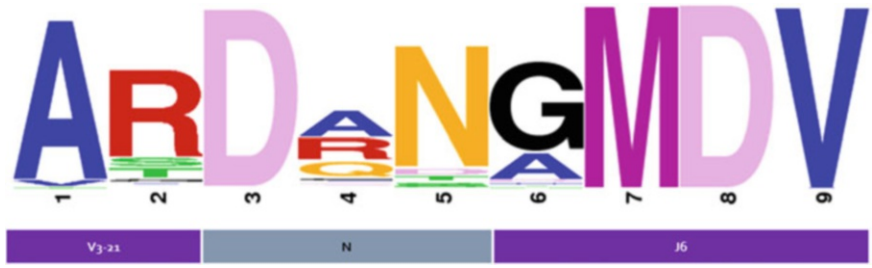


Fig. 1 Restricted VH CDR3s among unrelated CLL cases utilizing the IGHV3-21 gene. In 2003, Tobin et al. [19] showed that almost half of all CLL cases expressing the IGHV3-21 gene carry short and highly similar VH CDR3 leading to their clustering in a subset, nowadays referred to as subset #2. Rearrangements belonging to this subset can be simply identified by a nine amino-acid (aa) long VH CDR3 with an acidic residue (aspartic acid D) at position 107. This sequence logo illustrates the VH CDR3s of all subset #2 cases in our recent study of 20331 IG sequences from patients with CLL [48]. In this logo, each VH CDR3 position is displayed as a stack of upright amino acid symbols. The height of each one-letter amino acid symbol is directly proportional to the relative frequency of that amino acid at a given position among all sequences of the subset

The study of the stereotypy phenomenon in larger cohorts of patients with CLL by us and others has conclusively documented that a large fraction of patients can be assigned to subsets with distinct, stereotyped VH CDR3 sequences [25–28]. Importantly, the previously identified IGHV gene repertoire biases were in essence a property of "stereotyped CLL," raising scientific speculations about a distinct disease trajectory between the two CLL categories, stereotyped and non-stereotyped [27]. These studies also showed that patients assigned to the same stereotyped subset may also share biological and clinical characteristics, suggesting that the structural properties of the clonotypic BcR IG are inextricably linked to CLL clonal behavior [26, 28].

In the first studies on BcR IG stereotypy in CLL, the percent of stereotyped CLL seemed to rise as the cohort size grew, raising questions regarding the actual size of "stereotyped" CLL. In order to address this issue, we performed a large-scale study in 7596 CLL IGHV-IGHD-IGHJ gene rearrangements [29] and found that cases with stereotyped BcR IG accounted approximately for one-third of the cohort. Moreover, we identified 19 subsets, comprising at least 20 sequences each, that we deemed as major, collectively representing 12% of all CLL.

The next challenge concerned how to prove that this novel molecular subclassification of CLL based on BcR IG stereotypy had biological and clinical implications, i.e., reflected fundamental similarities beyond primary amino acid IG sequences. On the available evidence, this challenge appears to be met since major CLL subsets are characterized by unique biological profiles, extending from genetic lesions [30–33] to epigenetic modifications [34, 35]; cell signaling via receptors of both innate (e.g., Toll-like receptors, TLRs) and adaptive immunity (i.e., the BcR) [36–38] and even cell autonomous signaling [39]; as well as clinical presentation and course, including response to treatment [40–46]. Overall, these findings suggest that distinct modes of interplay between cell-intrinsic and cell-extrinsic mechanisms may underlie the ontogeny and evolution of each subset. Therefore, BcR IG stereotypy and, particularly, the identification of major stereotyped CLL subsets have important implications for both advancing our understanding of CLL pathogenesis and improving patient risk stratification toward personalized therapeutic applications [14, 47].

In the course of our immunogenetic studies, we became aware of IG gene sequences that were similar to those belonging in major subsets, yet were set apart due to not fulfilling all adopted criteria. These subset-like cases could, in theory, also share the distinct biological and clinical properties of the respective major subsets. For these reasons, we scanned 20,331 CLL IG sequences for sequence relatives of major IG stereotypes [48]. Overall, 12.1% CLL sequences were assigned to the 19 major subsets. The relative frequencies of all major subsets were highly similar to those

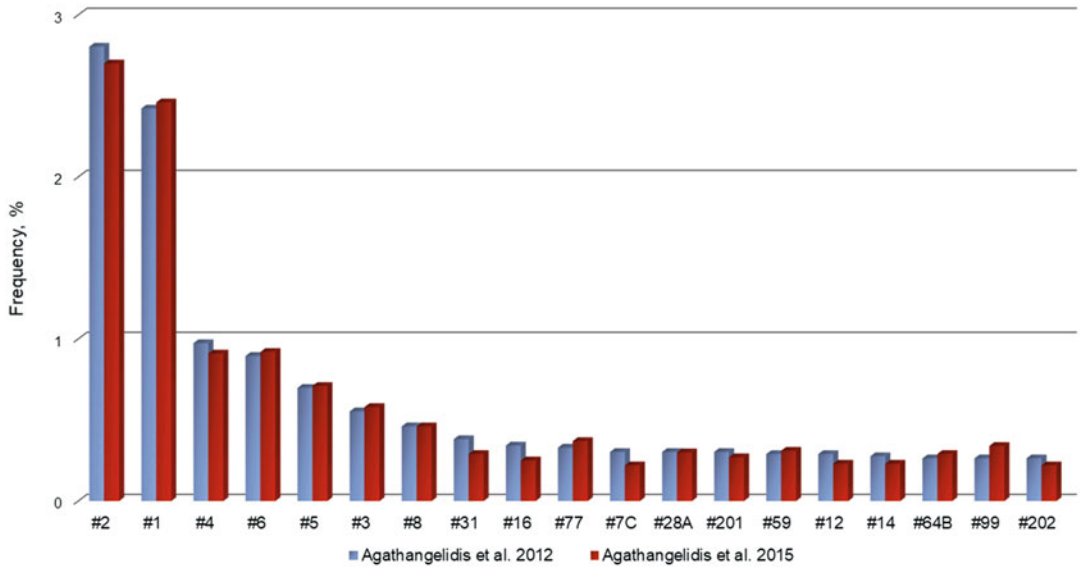


Fig. 2 Frequency of major CLL subsets. The relative size of the 19 major subsets in CLL remained stable in 2 different studies including 7596 [29] and 20331 IG sequences from patients with CLL [48] supporting their definition as major

reported in our previous analysis [29], justifying their consideration as major (Fig. 2). We did not find subset-like sequences for major M-CLL stereotyped subsets, indicating that these subsets have very distinct immunogenetic characteristics. However, we did identify "satellite subsets" of many major U-CLL subsets, which were often quite large (Fig. 3 illustrates the satellite subsets of CLL major subset #8). We interpreted these findings as probably reflecting intrinsic differences between M-CLL and U-CLL subsets in terms of IG gene repertoire and VH CDR3 features. From a biological and clinical perspective, considering that major stereotyped subsets may indeed represent distinct disease variants, our highlighting of higher-order sequence relations may have implications for clinico-biological research aimed at dissecting the heterogeneity of CLL toward identifying distinct profiles that would assist in clinical decision-making.

1.3 B Cell Receptor Immunoglobulin Stereotypes in Other Mature B Cell Neoplasms

1.3.1 Mantle Cell Lymphoma

Despite the fact that the initial immunogenetic studies in MCL concerned small patient cohorts, however, they consistently reported IGHV gene restrictions [49–54]. In 2011 we published a large-scale IG gene repertoire study of 807 patients with MCL and reported that only four genes (namely, IGHV3-21, IGHV4-34, IGHV1-8, and IGHV3-23) accounted for virtually half of the total repertoire [55]. On this evidence, we hypothesized that antigenic and/or superantigenic selection may be important for disease pathogenesis also in MCL [56], a claim corroborated by subsequent studies [57, 58]. Importantly, this hypothesis soon

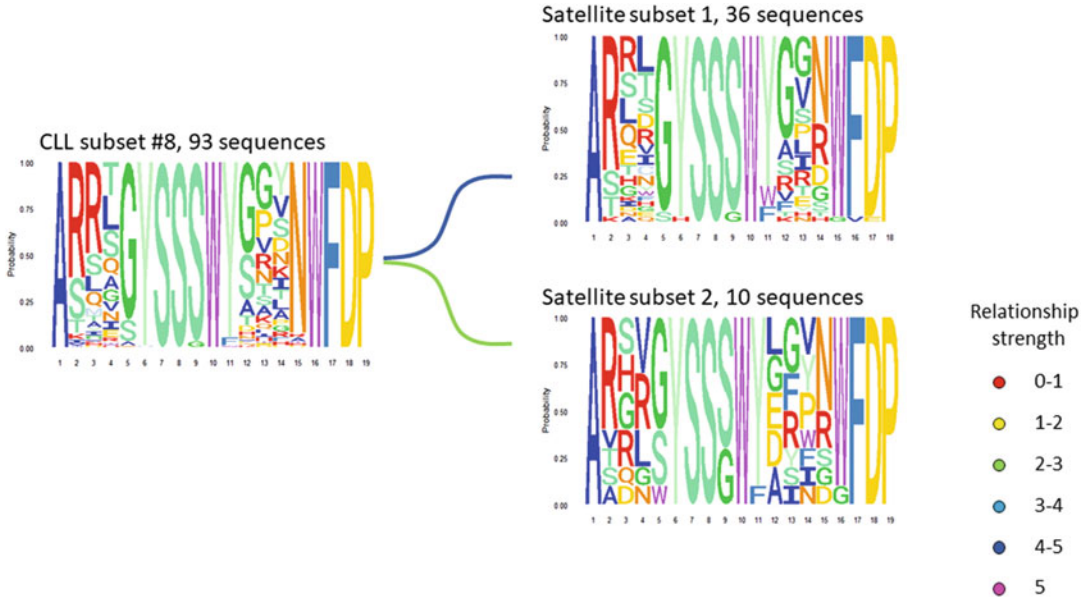


Fig. 3 Satellite subsets of major CLL subset #8. The search for satellite subsets of all 19 major CLL stereotypes resulted in vastly different results for M-CLL and U-CLL subsets. In the case of M-CLL, we did not identify large satellite subsets. For major U-CLL subsets, certain subsets remained "unique," whereas others had satellite subsets. In particular, in the case of subset #8 (IGHV4-39), we identified 2 "satellite" subsets that consisted of 36 and 10 cases and displayed relative sizes of 37.6% and 10.8% (compared to subset #8), respectively [48]

gained strong clinical support when ibrutinib was found to be an effective therapeutic agent for MCL [59].

Back to the immunogenetics, we also identified certain biased associations between IGHV genes and certain IGHD and/or IGHJ genes [55]. These biased associations between IGHV, IGHD, and IGHJ genes underlay the formation of restricted (stereotyped) VH CDR3 motifs, especially for cases expressing the IGHV3-21 and IGHV4-34 genes [55]. Comparing MCL stereotypes from cases utilizing the IGHV3-21 and IGHV4-34 genes against stereotypes expressing the same IGHV genes previously reported in CLL, we documented that stereotypes were indeed disease-specific. This result further supports the distinct types of postulated progenitors and distinct immune mechanisms implicated in the pathogenesis of MCL and CLL [55, 56].

1.3.2 Splenic Marginal Zone Lymphoma

Splenic marginal zone lymphoma (SMZL) is characterized by a remarkably biased IGHV gene repertoire with greater than 30% of patients utilizing a single IGHV gene allele, namely, IGHV1-2*04 [60]. IGHV1-2*04 gene rearrangements in SMZL carry few somatic mutations and very often express the IGHD3-3 or IGHD3-10 genes, mainly in reading frame 3, forming long and electropositive VH CDR3. However, despite this restricted

association, only few IGHV1-2*04 gene rearrangements in SMZL fulfilled the criteria for stereotypy previously applied to CLL and MCL, mainly due to variable IGHJ gene usage and heterogeneous junctional residues. However, rare "SMZL-specific" stereotypes with shared IGHD-derived amino acid motifs do actually exist [60, 61]. These results might be taken as evidence for a "SMZL-biased" type of IG stereotypy, defined by restrictions in VH CDR3 length along with few "landmark" (i.e., conserved) amino acids. Again, as in MCL, cross-entity comparison of stereotyped VH CDR3s revealed those to be disease-specific [60], with very few noteworthy exceptions [62, 63].

2 Identification of BcR IG Stereotypes: Methodologies and Criteria

The study of BcR IG stereotypy has focused on the IG heavy chains and, more specifically, the VH CDR3, not least because this is the most variable part of the IG and plays the most critical role in antigen recognition, at least in the pre-immune repertoire [64]. Indeed, evidence suggests that the more similar the primary VH CDR3 sequences of two different IG, the more similar their folding and, in turn, their specificities [65].

The analysis of stereotypy requires the application of comparisons between primary IG amino acid sequences and, thus, is largely dependent on alignment of the nucleotide sequence of the rearranged IGHV gene against that of its closest germline counterpart. In other words, detailed and accurate IG sequence annotation is a critical prerequisite for BcR IG stereotypy analysis: it is reliably provided by IMGT®, the international ImMunoGeneTics information system (http://www.imgt.org), the global reference in immunogenetics and immunoinformatics. Indeed, IMGT/V-QUEST [66], the analytical tool for rearranged IG sequences of IMGT, enables robust identification of the rearranged V and J genes and alleles and incorporates the results of the IMGT/JunctionAnalysis for the detailed analysis of the VH CDR3 properties and the identification of the D gene and its reading frame.

In the following section, we describe the evolution of methodologies and criteria for the identification of BcR IG stereotypes in CLL and other B cell lymphomas.

2.1 Messmer et al. (2004)

The first set of criteria for the identification of stereotyped subsets was proposed in 2004 by Messmer et al. [21]. In detail, the criteria were the following:

1. Utilization of the same IGHV, IGHD, and IGHJ genes.
2. The IGHD gene reading frame should be identical.

3. VH CDR3 amino acid identity between members of the same stereotyped subset should be ≥60%, a threshold decided based on established bioinformatics concepts such as amino acid substitution matrices and in particular BLOSUM62 [67]. Substitution scoring systems such as BLOSUM (blocks substitution matrix) have been used for the study of nonidentical residues in short, highly conserved sequence blocks: in particular, BLOSUM62 was generated from sequences that were at least 62% identical between them. While similar tables have been generated at the 50% or 80% thresholds, BLOSUM62 has been shown to be the most reliable scoring system over a wide range of protein families thus also initially applicable to the study of IG receptors.

2.2 Stamatopoulos et al. (2007)

A significant proportion of cases with stereotyped IG were identified using the first set of criteria. However, it was evident that the first criterion concerning the usage of IGHV genes could not be met in all cases. Even in the publication by Messmer et al. [21], there were cases assigned to the same subset that utilized different IGHV genes. In a subsequent study by our group [28], we took this into consideration while also incorporating the concept of amino acid functional similarity, thereby refining the criteria as follows:

1. Different IGHV genes are allowed.
2. VH CDR3 length difference should be ≤3 amino acids.
3. VH CDR3 amino acid sequence identity of ≥60%.
4. Functional conservation of amino acids (based on the BLOSUM concept outlined above).

Following these revised criteria, we identified subsets of cases expressing different IGHV genes, e.g., subsets #1 or #77 (Fig. 4): interestingly, the IGHV genes utilized by subset #1 or subset #77 cases are members of the same phylogenetic clan and, thus, exhibit a high degree of similarity.

2.3 Darzentas et al. (2010)

The need to better understand the phenomenon of IG stereotypy prompted the study of IG gene sequence data from thousands of patients. This, in turn, meant that automated approaches were a prerequisite for ensuring robustness and sensitivity. To this end, we designed and developed a purpose-built, pattern-discovery bioinformatics approach based on the TEIRESIAS algorithm from IBM Research [68]. This systematic approach enabled the robust identification of VH CDR3 sequence similarities, whereas it was also able to describe more distant sequence relationships in a tree-like fashion by grouping together related first (ground)-level clusters into larger high-level clusters [27].

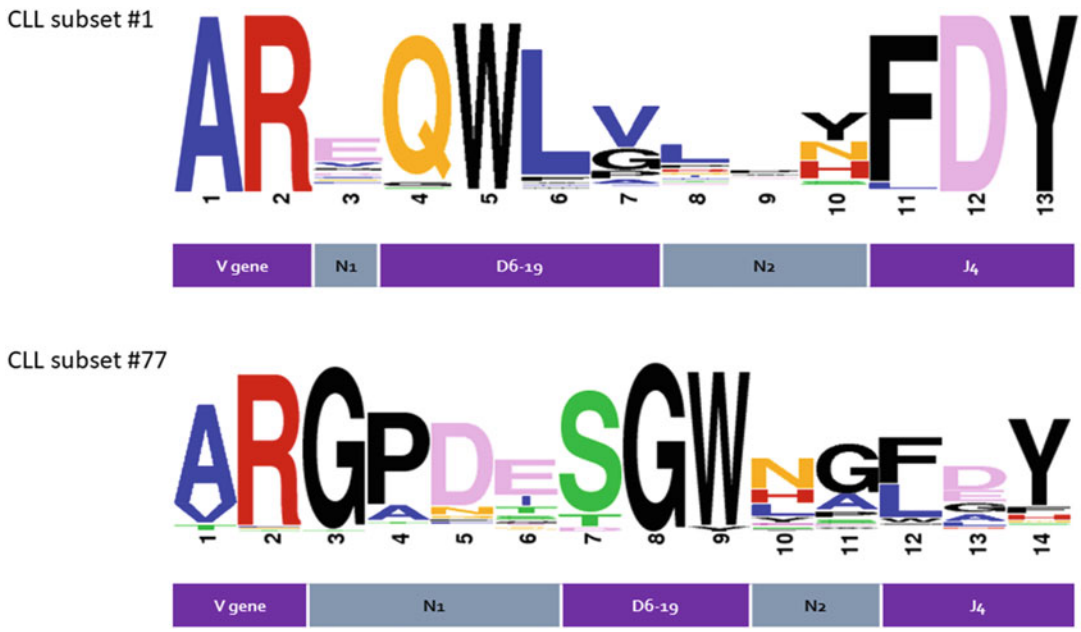


Fig. 4 Restricted VH CDR3 in stereotyped subsets #1 and #77 CLL cases. These subsets are characterized by usage of different IGHV genes. Subset #1 cases express the IGHV1-2, IGHV1-3, IGHV1-18, IGHV1-8, IGHV5-a, and IGHV7-4-1 genes, all members of phylogenetic clan I, while subset #77 includes cases carrying the IGHV4-4 and IGHV4-59 genes, both belonging to the phylogenetic clan II. In both subsets the IGHV gene is expressed in association with the IGHD6-19 and IGHJ4 genes. The sequence logos presented here depict the VH CDR3s of all cases assigned to subsets #1 and #77 in our recent study of 20,331 IG sequences from patients with CLL [48]

In summary, the algorithm builds the largest possible shared amino acid motifs between different VH CDR3 and clusters the respective IG sequences accordingly at a first (ground) level. IG sequences assigned to the same ground-level cluster satisfy the following criteria:

1. VH CDR3 amino acid identity of $\geq$50% and amino acid similarity of $\geq$70%.
2. VH CDR3 lengths should be $\leq$2 amino acids different.
3. The offset of the shared sequence pattern should be $\leq$2.

A critical characteristic of the algorithm is that sequences can carry more than one shared pattern and, thus, can be members of more than one ground-level cluster, enabling their grouping into clusters at progressively higher levels of hierarchy. This leads to the formation of clusters characterized by certain critically positioned residues in the VH CDR3 sequence. High-level clusters are usually much larger in size and describe more distant sequence relationships in the form of more widely shared, though more "degenerate," sequence patterns [27].

Table 1
Evolution of the criteria for the identification of IG stereotypy in CLL

	Messmer et al. (2004)	Stamatopoulos et al. (2007)	Darzentas et al. (2010)	Agathangelidis et al. (2012)	Satellite subsets (2015)
VH CDR3 amino acid identity (%)	≥60	≥60	≥50	≥50	Presence of the "subset-specific" amino acid pattern
VH CDR3 amino acid similarity (%)	n/a	n/a	≥70	≥70	
IGHV genes	Same	Any	Any	IGHV phylogenetic clan	Same
VH CDR3 length difference	?	≤3	≤2	=0	≤2
Offset of the pattern	n/a	n/a	≤2	=0	n/a
IGHD gene reading frame	Same	n/a	n/a	n/a	n/a

n/a non-applicable

2.4 Agathangelidis et al. (2012)

Subsequent studies in different contexts provided the basis for an update in the criteria proposed by Darzentas et al. [29] (Table 1). The most important change concerned the criterion for IGHV gene usage. Considering the role of the germline-encoded specificities in antigen recognition and also realizing that distinct yet phylogenetically related IGHV genes are represented in the gene repertoire of subsets, we deemed that only sequences carrying IGHV genes of the same clan could be assigned to the same cluster. In particular, three ancestral phylogenetic clans were imported in the pipeline: IGHV1/5/7 subgroup genes from clan I, IGHV2/4/6 from clan II, and finally, IGHV3 genes from clan III [69]. Additionally, we applied more stringent criteria including the requirement for identical VH CDR3 lengths and identical pattern offsets (i.e., exact locations within the VH CDR3 region) between connected sequences (Table 1).

2.5 ARResT/AssignSubsets (2015)

As mentioned above, 1 in 8 patients with CLL belongs to 1 of 19 different major subsets [29], and this subclassification has important biological and clinical implications [47]. Thus, the systematic assignment of CLL cases to major subsets appears of high value for both research and prognostication.

On these grounds, we designed a novel application, called ARResT/AssignSubsets, which performs automated assignment

of IG sequences from CLL patients to the 19 major stereotyped subsets [70]. At the core of the tool lies a set of rules captured in a probabilistic model, a Bayes classifier implemented in R. ARResT/AssignSubsets evaluates IG sequence characteristics divided into two categories: core and secondary. Core characteristics are based on the last set of stringent criteria for subset assignment [29]. If any of the values of core characteristics differs from the one shared by the members of a subset, then the sequence will not be assigned to that particular subset. The three core features are (1) VH CDR3 length, which is critical for the formation of the antigen-binding site [65]; (2) IGHV gene phylogenetic clan that takes into account IGHV gene ancestry and amino acid similarity [69]; and (3) SHM status of the IG sequence, an important prognosticator in CLL [10, 11]. Secondary characteristics are less strict, i.e., different values are accepted (but scored negatively)—and include (1) IGHV and IGHJ genes and (2) the VH CDR3 amino acid sequence composition through the calculation of amino acid frequencies for all given sequence positions.

Assignment to major subsets is performed after the evaluation of all core and secondary characteristics. The tool report includes the quality of the submitted IG sequence, the assignment to major subsets and its confidence (Fig. 5), and, when possible, heat maps of core and secondary features with their significance.

assignment frequencies table

CLL#2	CLL#1	CLL#4	CLL#6	CLL#5	CLL#3	CLL#8	CLL#31	CLL#16	CLL#77
2.8%	2.4%	1.0%	0.9%	0.7%	0.6%	0.5%	0.4%	0.3%	0.3%

1

CLL#7H	CLL#28A	CLL#201	CLL#12	CLL#59	CLL#14	CLL#64B	CLL#99	CLL#202
0.3%	0.3%	0.3%	0.3%	0.3%	0.3%	0.3%	0.3%	0.3%

assignment report table

label [+ heat map, if appl.]	SeqCure	subset	confidence	score
IGsequence-6	warning	CLL#2	extreme	78.26
IGsequence-1	warning	unassigned	extreme	-8.56
IGsequence-2	warning	unassigned	extreme	-Inf
IGsequence-3	warning	unassigned	extreme	-Inf
IGsequence-4	warning	unassigned	extreme	-Inf
IGsequence-5	warning	unassigned	extreme	-Inf
IGsequence-7	note	unassigned	extreme	-Inf
IGsequence-8	note	unassigned	extreme	-Inf
IGsequence-9	note	unassigned	extreme	-Inf
IGsequence-10	note	unassigned	extreme	-Inf

hosted at the Bioinformatics Analysis Team / BAT

Fig. 5 ARResT/AssignSubsets report table for assignment to major subsets. IG sequences from ten CLL patients in fasta format (IG sequences 1–10) were submitted to the ARResT/AssignSubsets tool. Case IG sequence 6 was assigned to CLL stereotyped subset #2 with extreme confidence. All other cases/sequences were reported as unassigned. A comprehensive explanation of the table headings and the tool is available at http:/tools.bat.infspire.org/arrest/assignsubsets/

2.6 Major Subset Satellites (2015)

The adoption of stringent criteria for subset assignment enabled to reliably identify cases with high IG sequence relatedness. However, it also led to the exclusion of sequences that were similar to those assigned to major subsets, yet did not satisfy the entire set of the applied criteria. In order to identify such IG sequences, IG sequences from CLL (on any other B cell malignancy) are clustered if displaying the following set of features:

1. Utilization of the same IGHV gene.
2. VH CDR3 length difference of $\leq$2 amino acids.
3. Presence of the "subset-specific" VH CDR3 sequence motif.

In detail, the algorithm uses the output of the aforementioned pattern-discovery clustering tool in order to identify all major subsets. Furthermore, a configuration file is created that contains all the criteria that will be applied in the analysis (*see* above). The input file is the Summary table output file provided by IMGT/V-QUEST. The algorithm identifies specific data columns and searches for specific values that satisfy the applied criteria while parsing the file line by line. After the identification of all sequences that display the required properties, the tool creates subsets of sequences by performing clustering based on VH CDR3 features. Finally, an output file for each major subset is created containing all satellite subsets. Furthermore, a report file is created that includes relevant information regarding the number of sequences grouped to at least one satellite subset and the number of satellite subsets for each major subset. A free online version of the tool is currently in preparation.

3 Future Developments and Perspective

In the future, we plan to upscale the analysis by expanding our dataset of CLL IG sequences contained within the IMGT/CLL-DB (https://www.imgt.org/CLLDBInterface/bylaws), developed by IMGT® and the IgCLL group. The primary aim of this initiative concerns the maintenance and expansion of a database with immunogenetic data from patients with CLL attended at academic institutions, to foster high-quality collaborative research on CLL and related disorders. The main outlet is a purpose-built interactive website platform providing all data, information, and knowledge to members in a user-friendly yet robust way.

Furthermore, capitalizing on the availability of sequence datasets from different B cell lymphoma subtypes consolidated within our wide network of collaborating institutions, we will also perform extensive cross-entity comparisons extending from CLL, MCL, and SMZL to other marginal zone lymphomas, hairy cell leukemia, and diffuse large B cell lymphoma, holding the potential to assist in

dissecting their immunopathogenesis. We are also already working toward developing workflows enabling the construction of robust 3D IG models followed by their automated clustering aimed at gaining deeper insight into IG structure diversity in B cell malignancies [71]. This initiative, also supported by published [39] and ongoing actual crystallographic analysis of the clonotypic BcR IG in various lymphomas, is anticipated to refine the concept of BcR IG stereotypy phenomenon with implications for the development of novel treatment strategies targeting the clonotypic IG.

Finally, in this era of next-generation, high-throughput immunogenetics, an important prerequisite for performing research on IG stereotypy concerns the development of a validated chain of processes from sample collection to laboratory methodologies, to IG data curation and, finally, systematic bioinformatics analysis. Many of the applications and methods described here are the results of efforts in this direction, including the IMGT-based tools and databases (such as the IMGT/CLL-DB initiative); the formulation of recommendations and guidelines for IG analysis [13, 14]; the establishment of an expert online forum to discuss and provide guidance on IG gene sequence interpretation in CLL by the European Research Initiative on CLL (ERIC; www.ericll.org); and, finally, the standardization of the methodology for IG analysis in CLL using novel NGS techniques jointly undertaken by ERIC and the EuroClonality-NGS Consortium [72] (https://euroclonalityngs.org/usr/pub/pub.php).

Acknowledgments

This work was supported in part by H2020 "AEGLE, An analytics framework for integrated and personalized healthcare services in Europe", by the EU; "MEDGENET, Medical Genomics and Epigenomics Network" (No. 692298) by the EU; "TRANSCAN-179" NOVEL JTC 2016; and "Odysseus" implemented under the Action "Reinforcement of the Research and Innovation Infrastructure," funded by the Operational Programme "Competitiveness, Entrepreneurship and Innovation" (NSRF 2014–2020) and co-financed by Greece and the European Union (European Regional Development Fund). Andreas Agathangelidis is a recipient of a Fellowship for Postgraduate Research by the Hellenic Foundation for Research and Innovation.

References

1. Chiorazzi N, Ferrarini M (2003) B cell chronic lymphocytic leukemia: lessons learned from studies of the B cell antigen receptor. Annu Rev Immunol 21:841–894. https://doi.org/10.1146/annurev.immunol.21.120601.141018

2. Stevenson FK, Sahota SS, Ottensmeier CH et al (2001) The occurrence and significance of V gene mutations in B cell-derived human malignancy. Adv Cancer Res 83:81–116

3. Sutton LA, Agathangelidis A, Belessi C et al (2013) Antigen selection in B-cell lymphomas—tracing the evidence. Semin Cancer Biol 23(6):399–409. https://doi.org/10.1016/j.semcancer.2013.07.006
4. Dunn-Walters D, Thiede C, Alpen B et al (2001) Somatic hypermutation and B-cell lymphoma. Philos Trans R Soc Lond Ser B Biol Sci 356(1405):73–82. https://doi.org/10.1098/rstb.2000.0751
5. Ghia P, Granziero L, Chilosi M et al (2002) Chronic B cell malignancies and bone marrow microenvironment. Semin Cancer Biol 12 (2):149–155. https://doi.org/10.1006/scbi.2001.0423
6. Scott DW, Gascoyne RD (2014) The tumour microenvironment in B cell lymphomas. Nat Rev Cancer 14(8):517–534. https://doi.org/10.1038/nrc3774
7. Burger JA, Chiorazzi N (2013) B cell receptor signaling in chronic lymphocytic leukemia. Trends Immunol 34(12):592–601. https://doi.org/10.1016/j.it.2013.07.002
8. Jerkeman M, Hallek M, Dreyling M et al (2017) Targeting of B-cell receptor signalling in B-cell malignancies. J Intern Med 282 (5):415–428. https://doi.org/10.1111/joim.12600
9. Fais F, Ghiotto F, Hashimoto S et al (1998) Chronic lymphocytic leukemia B cells express restricted sets of mutated and unmutated antigen receptors. J Clin Invest 102 (8):1515–1525. https://doi.org/10.1172/JCI3009
10. Damle RN, Wasil T, Fais F et al (1999) Ig V gene mutation status and CD38 expression as novel prognostic indicators in chronic lymphocytic leukemia. Blood 94(6):1840–1847
11. Hamblin TJ, Davis Z, Gardiner A et al (1999) Unmutated Ig V(H) genes are associated with a more aggressive form of chronic lymphocytic leukemia. Blood 94(6):1848–1854
12. Ghia P, Stamatopoulos K, Belessi C et al (2007) ERIC recommendations on IGHV gene mutational status analysis in chronic lymphocytic leukemia. Leukemia 21(1):1–3. https://doi.org/10.1038/sj.leu.2404457
13. Langerak AW, Davi F, Ghia P et al (2011) Immunoglobulin sequence analysis and prognostication in CLL: guidelines from the ERIC review board for reliable interpretation of problematic cases. Leukemia 25(6):979–984. https://doi.org/10.1038/leu.2011.49
14. Rosenquist R, Ghia P, Hadzidimitriou A et al (2017) Immunoglobulin gene sequence analysis in chronic lymphocytic leukemia: updated ERIC recommendations. Leukemia 31 (7):1477–1481. https://doi.org/10.1038/leu.2017.125
15. Hallek M, Cheson BD, Catovsky D et al (2018) Guidelines for diagnosis, indications for treatment, response assessment and supportive management of chronic lymphocytic leukemia. Blood 131:2745–2760. https://doi.org/10.1182/blood-2017-09-806398
16. Efremov DG, Ivanovski M, Siljanovski N et al (1996) Restricted immunoglobulin VH region repertoire in chronic lymphocytic leukemia patients with autoimmune hemolytic anemia. Blood 87(9):3869–3876
17. Johnson TA, Rassenti LZ, Kipps TJ (1997) Ig VH1 genes expressed in B cell chronic lymphocytic leukemia exhibit distinctive molecular features. J Immunol 158(1):235–246
18. Widhopf GF 2nd, Kipps TJ (2001) Normal B cells express 51p1-encoded Ig heavy chains that are distinct from those expressed by chronic lymphocytic leukemia B cells. J Immunol 166(1):95–102
19. Tobin G, Thunberg U, Johnson A et al (2003) Chronic lymphocytic leukemias utilizing the VH3-21 gene display highly restricted Vlambda2-14 gene use and homologous CDR3s: implicating recognition of a common antigen epitope. Blood 101(12):4952–4957. https://doi.org/10.1182/blood-2002-11-3485
20. Ghiotto F, Fais F, Valetto A et al (2004) Remarkably similar antigen receptors among a subset of patients with chronic lymphocytic leukemia. J Clin Invest 113(7):1008–1016. https://doi.org/10.1172/JCI19399
21. Messmer BT, Albesiano E, Efremov DG et al (2004) Multiple distinct sets of stereotyped antigen receptors indicate a role for antigen in promoting chronic lymphocytic leukemia. J Exp Med 200(4):519–525. https://doi.org/10.1084/jem.20040544
22. Tobin G, Thunberg U, Karlsson K et al (2004) Subsets with restricted immunoglobulin gene rearrangement features indicate a role for antigen selection in the development of chronic lymphocytic leukemia. Blood 104 (9):2879–2885. https://doi.org/10.1182/blood-2004-01-0132
23. Widhopf GF 2nd, Rassenti LZ, Toy TL et al (2004) Chronic lymphocytic leukemia B cells of more than 1% of patients express virtually identical immunoglobulins. Blood 104 (8):2499–2504. https://doi.org/10.1182/blood-2004-03-0818
24. Stamatopoulos K, Belessi C, Hadzidimitriou A et al (2005) Immunoglobulin light chain repertoire in chronic lymphocytic leukemia. Blood

106(10):3575–3583. https://doi.org/10.1182/blood-2005-04-1511

25. Murray F, Darzentas N, Hadzidimitriou A et al (2008) Stereotyped patterns of somatic hypermutation in subsets of patients with chronic lymphocytic leukemia: implications for the role of antigen selection in leukemogenesis. Blood 111(3):1524–1533. https://doi.org/10.1182/blood-2007-07-099564
26. Bomben R, Dal Bo M, Capello D et al (2009) Molecular and clinical features of chronic lymphocytic leukaemia with stereotyped B cell receptors: results from an Italian multicentre study. Br J Haematol 144(4):492–506. https://doi.org/10.1111/j.1365-2141.2008.07469.x
27. Darzentas N, Hadzidimitriou A, Murray F et al (2010) A different ontogenesis for chronic lymphocytic leukemia cases carrying stereotyped antigen receptors: molecular and computational evidence. Leukemia 24(1):125–132. https://doi.org/10.1038/leu.2009.186
28. Stamatopoulos K, Belessi C, Moreno C et al (2007) Over 20% of patients with chronic lymphocytic leukemia carry stereotyped receptors: pathogenetic implications and clinical correlations. Blood 109(1):259–270. https://doi.org/10.1182/blood-2006-03-012948
29. Agathangelidis A, Darzentas N, Hadzidimitriou A et al (2012) Stereotyped B-cell receptors in one-third of chronic lymphocytic leukemia: a molecular classification with implications for targeted therapies. Blood 119(19):4467–4475. https://doi.org/10.1182/blood-2011-11-393694
30. Malcikova J, Stalika E, Davis Z et al (2014) The frequency of TP53 gene defects differs between chronic lymphocytic leukaemia subgroups harbouring distinct antigen receptors. Br J Haematol 166(4):621–625. https://doi.org/10.1111/bjh.12893
31. Navrkalova V, Young E, Baliakas P et al (2016) ATM mutations in major stereotyped subsets of chronic lymphocytic leukemia: enrichment in subset #2 is associated with markedly short telomeres. Haematologica 101(9):e369–e373. https://doi.org/10.3324/haematol.2016.142968
32. Strefford JC, Sutton LA, Baliakas P et al (2013) Distinct patterns of novel gene mutations in poor-prognostic stereotyped subsets of chronic lymphocytic leukemia: the case of SF3B1 and subset #2. Leukemia 27(11):2196–2199. https://doi.org/10.1038/leu.2013.98
33. Sutton LA, Young E, Baliakas P et al (2016) Different spectra of recurrent gene mutations in subsets of chronic lymphocytic leukemia harboring stereotyped B-cell receptors. Haematologica 101(8):959–967. https://doi.org/10.3324/haematol.2016.141812
34. Papakonstantinou N, Ntoufa S, Chartomatsidou E et al (2013) Differential microRNA profiles and their functional implications in different immunogenetic subsets of chronic lymphocytic leukemia. Mol Med 19:115–123. https://doi.org/10.2119/molmed.2013.00005
35. Kanduri M, Marincevic M, Halldorsdottir AM et al (2012) Distinct transcriptional control in major immunogenetic subsets of chronic lymphocytic leukemia exhibiting subset-biased global DNA methylation profiles. Epigenetics 7(12):1435–1442. https://doi.org/10.4161/epi.22901
36. Arvaniti E, Ntoufa S, Papakonstantinou N et al (2011) Toll-like receptor signaling pathway in chronic lymphocytic leukemia: distinct gene expression profiles of potential pathogenic significance in specific subsets of patients. Haematologica 96(11):1644–1652. https://doi.org/10.3324/haematol.2011.044792
37. Ntoufa S, Vardi A, Papakonstantinou N et al (2012) Distinct innate immunity pathways to activation and tolerance in subgroups of chronic lymphocytic leukemia with distinct immunoglobulin receptors. Mol Med 18:1281–1291. https://doi.org/10.2119/molmed.2011.00480
38. Ntoufa S, Papakonstantinou N, Apollonio B et al (2016) B cell anergy modulated by TLR1/2 and the miR-17 approximately 92 cluster underlies the indolent clinical course of chronic lymphocytic leukemia stereotyped subset #4. J Immunol 196(10):4410–4417. https://doi.org/10.4049/jimmunol.1502297
39. Minici C, Gounari M, Ubelhart R et al (2017) Distinct homotypic B-cell receptor interactions shape the outcome of chronic lymphocytic leukaemia. Nat Commun 8:15746. https://doi.org/10.1038/ncomms15746
40. Baliakas P, Agathangelidis A, Hadzidimitriou A et al (2015) Not all IGHV3-21 chronic lymphocytic leukemias are equal: prognostic considerations. Blood 125(5):856–859. https://doi.org/10.1182/blood-2014-09-600874
41. Baliakas P, Hadzidimitriou A, Sutton LA et al (2014) Clinical effect of stereotyped B-cell receptor immunoglobulins in chronic lymphocytic leukaemia: a retrospective multicentre study. Lancet Haematol 1(2):e74–e84. https://doi.org/10.1016/S2352-3026(14)00005-2
42. Del Giudice I, Chiaretti S, Santangelo S et al (2014) Stereotyped subset #1 chronic lymphocytic leukemia: a direct link between B-cell

receptor structure, function, and patients' prognosis. Am J Hematol 89(1):74–82. https://doi.org/10.1002/ajh.23591

43. Maura F, Cutrona G, Fabris S et al (2011) Relevance of stereotyped B-cell receptors in the context of the molecular, cytogenetic and clinical features of chronic lymphocytic leukemia. PLoS One 6(8):e24313. https://doi.org/10.1371/journal.pone.0024313
44. Rossi D, Spina V, Cerri M et al (2009) Stereotyped B-cell receptor is an independent risk factor of chronic lymphocytic leukemia transformation to Richter syndrome. Clin Cancer Res 15(13):4415–4422. https://doi.org/10.1158/1078-0432.CCR-08-3266
45. Xochelli A, Baliakas P, Kavakiotis I et al (2017) Chronic lymphocytic leukemia with mutated IGHV4-34 receptors: shared and distinct immunogenetic features and clinical outcomes. Clin Cancer Res 23(17):5292–5301. https://doi.org/10.1158/1078-0432.CCR-16-3100
46. Baliakas P, Mattsson M, Hadzidimitriou A et al (2018) No improvement in long-term survival over time for chronic lymphocytic leukemia patients in stereotyped subsets #1 and #2 treated with chemo(immuno)therapy. Haematologica 103(4):e158–e161. https://doi.org/10.3324/haematol.2017.182634
47. Stamatopoulos K, Agathangelidis A, Rosenquist R et al (2017) Antigen receptor stereotypy in chronic lymphocytic leukemia. Leukemia 31(2):282–291. https://doi.org/10.1038/leu.2016.322
48. Agathangelidis A, Bystry V, Hadzidimitriou A et al (2015) Higher-order immunoglobulin sequence relations for major subsets of chronic lymphocytic leukemia: uniqueness versus equivalence. In: Paper presented at the 20th congress of the European Hematology Association (EHA), Vienna
49. Camacho FI, Algara P, Rodriguez A et al (2003) Molecular heterogeneity in MCL defined by the use of specific VH genes and the frequency of somatic mutations. Blood 101(10):4042–4046. https://doi.org/10.1182/blood-2002-11-3456
50. Kienle D, Krober A, Katzenberger T et al (2003) VH mutation status and VDJ rearrangement structure in mantle cell lymphoma: correlation with genomic aberrations, clinical characteristics, and outcome. Blood 102(8):3003–3009. https://doi.org/10.1182/blood-2003-05-1383
51. Orchard J, Garand R, Davis Z et al (2003) A subset of t(11;14) lymphoma with mantle cell features displays mutated IgVH genes and includes patients with good prognosis, nonnodal disease. Blood 101(12):4975–4981. https://doi.org/10.1182/blood-2002-06-1864
52. Schraders M, Oeschger S, Kluin PM et al (2009) Hypermutation in mantle cell lymphoma does not indicate a clinical or biological subentity. Mod Pathol 22(3):416–425. https://doi.org/10.1038/modpathol.2008.199
53. Thorselius M, Walsh S, Eriksson I et al (2002) Somatic hypermutation and V(H) gene usage in mantle cell lymphoma. Eur J Haematol 68(4):217–224
54. Walsh SH, Thorselius M, Johnson A et al (2003) Mutated VH genes and preferential VH3-21 use define new subsets of mantle cell lymphoma. Blood 101(10):4047–4054. https://doi.org/10.1182/blood-2002-11-3479
55. Hadzidimitriou A, Agathangelidis A, Darzentas N et al (2011) Is there a role for antigen selection in mantle cell lymphoma? Immunogenetic support from a series of 807 cases. Blood 118(11):3088–3095. https://doi.org/10.1182/blood-2011-03-343434
56. Agathangelidis A, Hadzidimitriou A, Rosenquist R et al (2011) Unlocking the secrets of immunoglobulin receptors in mantle cell lymphoma: implications for the origin and selection of the malignant cells. Semin Cancer Biol 21(5):299–307. https://doi.org/10.1016/j.semcancer.2011.09.009
57. Xochelli A, Sutton LA, Agathangelidis A et al (2015) Molecular evidence for antigen drive in the natural history of mantle cell lymphoma. Am J Pathol 185(6):1740–1748. https://doi.org/10.1016/j.ajpath.2015.02.006
58. Pouliou E, Xochelli A, Kanellis G et al (2017) Numerous ontogenetic roads to mantle cell lymphoma: immunogenetic and immunohistochemical evidence. Am J Pathol 187(7):1454–1458. https://doi.org/10.1016/j.ajpath.2017.02.017
59. Wang ML, Rule S, Martin P et al (2013) Targeting BTK with ibrutinib in relapsed or refractory mantle-cell lymphoma. N Engl J Med 369(6):507–516. https://doi.org/10.1056/NEJMoa1306220
60. Bikos V, Darzentas N, Hadzidimitriou A et al (2012) Over 30% of patients with splenic marginal zone lymphoma express the same immunoglobulin heavy variable gene: ontogenetic implications. Leukemia 26(7):1638–1646. https://doi.org/10.1038/leu.2012.3
61. Zibellini S, Capello D, Forconi F et al (2010) Stereotyped patterns of B-cell receptor in splenic marginal zone lymphoma.

Haematologica 95(10):1792–1796. https://doi.org/10.3324/haematol.2010.025437

62. Kostareli E, Gounari M, Janus A et al (2012) Antigen receptor stereotypy across B-cell lymphoproliferations: the case of IGHV4-59/IGKV3-20 receptors with rheumatoid factor activity. Leukemia 26(5):1127–1131. https://doi.org/10.1038/leu.2011.311
63. Xochelli A, Bikos V, Polychronidou E et al (2015) Unique versus common: disease-biased immunoglobulin gene repertoires along with public antigen receptor stereotypes in marginal zone B-cell lymphoproliferations. In: 57th ASH annual meeting, Orlando, FL
64. Xu JL, Davis MM (2000) Diversity in the CDR3 region of V(H) is sufficient for most antibody specificities. Immunity 13(1):37–45
65. Barrios Y, Jirholt P, Ohlin M (2004) Length of the antibody heavy chain complementarity determining region 3 as a specificity-determining factor. J Mol Recognit 17 (4):332–338. https://doi.org/10.1002/jmr.679
66. Brochet X, Lefranc MP, Giudicelli V (2008) IMGT/V-QUEST: the highly customized and integrated system for IG and TR standardized V-J and V-D-J sequence analysis. Nucleic Acids Res 36(Web Server):W503–W508. https://doi.org/10.1093/nar/gkn316
67. Henikoff S, Henikoff JG (1993) Performance evaluation of amino acid substitution matrices. Proteins 17(1):49–61. https://doi.org/10.1002/prot.340170108
68. Rigoutsos I, Floratos A (1998) Combinatorial pattern discovery in biological sequences: the TEIRESIAS algorithm. Bioinformatics 14 (1):55–67
69. Kirkham PM, Mortari F, Newton JA et al (1992) Immunoglobulin VH clan and family identity predicts variable domain structure and may influence antigen binding. EMBO J 11 (2):603–609
70. Bystry V, Agathangelidis A, Bikos V et al (2015) ARResT/AssignSubsets: a novel application for robust subclassification of chronic lymphocytic leukemia based on B cell receptor IG stereotypy. Bioinformatics 31 (23):3844–3846. https://doi.org/10.1093/bioinformatics/btv456
71. Polychronidou E, Kalamaras I, Agathangelidis A et al (2018) Automated shape-based clustering of 3D immunoglobulin protein structures in chronic lymphocytic leukemia. BMC Bioinformatics 19:414
72. Langerak AW, Bruggemann M, Davi F et al (2017) High-throughput immunogenetics for clinical and research applications in immunohematology: potential and challenges. J Immunol 198(10):3765–3774. https://doi.org/10.4049/jimmunol.1602050

Chapter 8

Flow Cytometric MRD Detection in Selected Mature B-Cell Malignancies

Sebastian Böttcher

Abstract

The quantification of submicroscopic minimal residual disease (MRD) after therapy proved to have independent prognostic significance in many mature B-cell malignancies. With the advent of routine benchtop cytometers capable of simultaneously analyzing ≥4 colors and with improved standardization, flow cytometry has become the method of choice for MRD assessments in some lymphoma entities. Herein we describe general aspects of flow cytometric standardization. Chronic lymphocytic leukemia and multiple myeloma (MM) are used as examples to explain the technical standardization of flow cytometry for MRD detection according to EuroFlow strategies. MRD data acquisition and detailed analysis using a newly developed approach (so-called next generation flow, NGF) in MM is a particular focus of this chapter.

Key words Minimal residual disease, Chronic lymphocytic leukemia, Multiple Myeloma, Flow cytometry, Next generation flow

1 Introduction

During the last decade, the research community witnessed an unprecedented rate at which novel, highly effective drugs and drug combinations were introduced into the treatment of mature B-cell malignancies. Many of the innovative treatment modalities lead to complete remissions without morphological evidence of disease. The improved reduction in tumor load that becomes achievable using those new agents translates into median progression-free survival (PFS) of more than 5 years in certain entities [1, 2]. Nevertheless, except for patients with diffuse large B-cell and Burkitt lymphomas, there is currently no evidence for definite cure in patients with mature B-lineage non-Hodgkin's lymphoma (B-NHL), and relapses can occur even after years. This observation suggests the persistence of minimal residual disease (MRD) at submicroscopic levels.

The long PFS that can be achieved with current state-of-the-art treatment regimens is increasingly hampering clinical research in

Ralf Küppers (ed.), *Lymphoma: Methods and Protocols*, Methods in Molecular Biology, vol. 1956,
https://doi.org/10.1007/978-1-4939-9151-8_8,

many incurable but relatively indolent mature B-cell malignancies. PFS data often only become mature at a time when likely the next generation of antileukemic drugs will have already become available. As submicroscopic disease levels were shown to be correlated to PFS, MRD gained interest as a surrogate marker for treatment efficacy in clinical trials [3] and additionally as a means to tailor intensity and duration of treatment in individual patients [4, 5].

MRD can be assessed in mature B-NHL using either flow cytometry (MRD flow) or molecular methods. The first techniques for MRD measurements were PCR-based, targeting either recurrent chromosomal translocations or the rearranged immunoglobulin genes of individual patients. The sensitivity of more simple PCR methods (e.g., consensus primer immunoglobulin heavy chain PCR) is dependent on the number of polyclonal benign B-cells in the sample as well as on the structure of the rearranged immunoglobulin gene of the individual patient [6, 7]. This disadvantage was overcome by the introduction of real-time quantitative (RQ)-PCR technology that allowed the assessment of quantitative disease levels. Allele-specific oligo (ASO) primer immunoglobulin heavy chain (*IGH*) RQ-PCR targets the hypervariable region of the rearranged immunoglobulin genes of individual patients. The method is well-standardized [8], reproducibly achieves sensitivities of 1 malignant cell in 50,000 normal nucleated cells (2×10^{-5}), and typically yields quantitative results down to 1 lymphoma cell in 10,000 benign leukocytes (10^{-4}). RQ-PCR utilizes DNA, so that sample stability is maintained for several days. The disadvantage of ASO primer *IGH* RQ-PCR is the considerable effort required for the design and evaluation of individual MRD assays for single patients, thereby increasing cost and turnaround time. Next generation sequencing (NGS), the most recently introduced molecular method in the field, utilizes massively parallel sequencing for MRD detection. Similarly to ASO primer *IGH* RQ-PCR, that technique targets the rearranged immunoglobulin genes of the individual patient. The method reportedly can achieve even higher sensitivities (up to 10^{-6}) provided sufficient numbers of cells are tested. NGS utilizes the same set of reagents for all patients but requires the sequence information from a heavily infiltrated sample for reliable results. Moreover, the method currently lacks internationally accepted standardization, in particular with respect to bioinformatics. Molecular methods that target the rearranged immunoglobulin genes of individual patients in general show similar sensitivities and specificities across different mature B-NHL entities. Very similar reagents and experimental approaches are applicable for different diseases from this group. However, the applicability of ASO *IGH* RQ-PCR might be occasionally hampered by mutations within the primer binding sites that particularly occur in multiple myeloma (MM) [9].

MRD flow relies on the identification of expression patterns of antigens that are not present in benign tissues (aberrant immunophenotypes). MRD flow approaches were pioneered in MM and chronic lymphocytic leukemia (CLL) [10, 11], as in those mature B-cell neoplasms well-defined normal counterparts exist and many markers to distinguish malignant from benign have already been identified a decade ago [11–13]. Flow cytometric MRD assessments have to incorporate markers to define the population of interest as well as markers to detect aberrations within that population. Both the population of interest (the immunophenotypic normal counterpart within the B-cell lineage) and the marker combinations to detect aberrations differ by entity, so that MRD flow assays have to be developed separately by disease. However, the assays published to date can be applied for MRD assessments in all patients suffering from a given entity. Sensitive MRD assessments require markers with a large difference in staining intensity between malignant and normal cells. If such clear immunophenotypic aberrations exist in a given disease, the sensitivity of MRD flow primarily depends on the total number of acquired leukocytes for analysis [7, 14, 15]. Depending on the respective B-NHL entities and assays, sensitivities of 10^{-4} to 10^{-5} were reported [11, 14, 16, 17]. Advantages of flow cytometric MRD assessments are applicability without requiring the assessments of a heavily infiltrated sample, very short turnaround time (less than 6 h), relatively low costs, and broad availability of the equipment. Centralized MRD diagnostics in large international randomized trials might be hampered by the requirement that viable cells are to be analyzed within 48 (at most 60) h from sampling.

Data on the clinical significance and the utility of different techniques to quantify MRD are available for several B-NHL entities, in particular for CLL, MM, follicular lymphoma (FL), and mantle cell lymphoma (MCL).

Chemoimmunotherapy [1, 18–20] as well as BCL2- [21–23] and B-cell-receptor pathway inhibitors [24–27] are the mainstay of modern CLL treatment. The former two therapeutic approaches typically cause a profound reduction of the tumor burden [1, 18–20, 22, 28], the extent of which, when determined by MRD, proved to be prognostic for PFS [20, 29–41]. Data from two randomized trials suggest that patients from the same study who achieve the same MRD levels experience similar PFS regardless of the treatment arm [30, 36]. Based on those observations, the European Medicines Agency (EMA) has accepted MRD as a surrogate efficacy endpoint for licensure in randomized clinical trials in this disease [3]. In keeping with current International Workshop on CLL (iwCLL) guidelines [42], EMA defines that MRD negativity equals less than 1 CLL cell in 10,000 benign nucleated cells (10^{-4}). MRD flow in CLL reliably reaches this sensitivity [11, 43] and can reportedly even achieve sensitivities of 10^{-5} [16]. Disease levels

assessed by flow cytometry and by ASO *IGH* RQ-PCR correlate very well within the quantifiable ranges of the two methods [7, 11, 14]. Both techniques are equally suited to detect MRD down to the threshold of 10^{-4}, even after anti-CD20 antibody treatment [14]. EMA recommends either ASO *IGH* RQ-PCR or flow cytometry as standardized methods for the evaluation of MRD [3]. To date, information on the prognostic significance of MRD in CLL measured by NGS is scarce [44].

Clinical development in MM proved at least as fast as in CLL, again driven by the advent of new drugs and their combinations. Novel proteasome inhibitors [45–48], monoclonal antibodies [49–52], and immunomodulatory drugs [53, 54] all contributed to prolonged PFS by inducing high-quality remissions. MRD was shown to predict for PFS in different clinical situations, including transplant-eligible and transplant-ineligible patient cohorts [9, 15, 45, 49, 50, 55–63]. Frequent mutations within the primer binding sites are considered causative for reduced applicability of ASO *IGH* RQ-PCR in MM [64]. These technical pitfalls likely limited the number of studies investigating the correlation of PCR-based MRD to PFS [62, 63, 65]. NGS was successfully applied for MRD assessments in MM [49, 50, 58] and repeatedly showed a good correlation to PFS. There is evidence that the applicability of NGS is higher than what was formerly observed with ASO *IGH* RQ-PCR, but quantitative differences between both methods were observed and have not been fully explained yet [64]. Substantial data sets proving the correlation between MRD flow and PFS have accumulated, as the flow cytometric method has been applied in large prospective clinical trials [15, 55–57, 59]. Moreover, since MRD flow data are available on samples taken already years ago, they can often be correlated to very mature PFS follow-up [60]. Based upon international general consensus guidelines on MRD flow in MM [10, 66], EuroFlow technical standardization [67], and comprehensive assessment of candidate markers available from immunophenotypic sample libraries [68], the EuroFlow consortium developed a fully standardized 2-tube 8-color MRD assay [17]. That novel method (so-called next generation flow, NGF) demonstrated improved sensitivity and quantitative range (at least 10^{-5}) compared to previous MRD flow methods, a very good correlation to NGS as well as superior prediction of clinical outcome. Quality control rounds using NGF in MM have been performed within the EuroFlow consortium and are expected to become publically available in 2019. The most recent consensus criteria of the International Myeloma Working Group (IMWG) accepted both NGF and NGS as equally suited technologies for MRD monitoring in MM [69].

There is a growing body of evidence demonstrating the clinical significance of MRD in MCL at different time points during the course of treatment and in different age groups [2, 70–73]. The

prognostic significance of MRD so far has been exclusively shown using RQ-PCR approaches, while flow cytometry is of proven utility to detect low-level disease in blood and bone marrow at presentation [6]. The application of NGS for MRD detection in MCL is still in its infancy [74]. Similarly to older flow cytometric approaches [6], the clinical impact of a novel MRD flow method applied after induction treatment could not be shown, while even very low MRD levels detected by molecular methods impacted on outcome [75]. Reduced sensitivity of MRD flow (10^{-3} to 10^{-4}) compared to RQ-PCR has also very recently been corroborated [76]. In summary, available data show that clinically useful MRD assessments require sensitivities beyond 10^{-4}, which currently cannot be obtained using flow cytometry.

The prognostic significance of MRD using RQ-PCR for PFS in FL was demonstrated by several groups [77–80], while this observation could not be reproduced in the EORTC 20981 trial [81]. Most of the investigators used assays that targeted *BCL2/IGH* rearrangements. There are no published MRD flow assays that achieve sensitivities comparable to molecular techniques. Given the lack of clinical applicability, this chapter will cover neither MRD flow in MCL nor in FL.

MRD flow in MM and CLL are the focus of this book chapter, as the clinical significance of the method is most robustly established for those two entities. Most of the immunophenotypic aberrations in CLL and MM are quantitative, i.e., malignant cells are identified by over- or underexpression of antigens relative to the most similar normal B-cell population. Nevertheless, the detection of quantitative differences in antigen expression has for a long time relied on operator experience and was not unequivocally defined [11]. To allow reproducible quantification of fluorescence intensities, the international EuroFlow consortium developed and tested optimized flow cytometer settings and immunophenotyping protocols, covering all technical aspects of flow cytometry in hematological malignancies (www.euroflow.org; *see* "protocols"). The application of these EuroFlow standard operation procedures (SOPs) now allows to directly compare expression levels measured in different laboratories and within one laboratory over time [67, 82]. These SOPs form the basis of MRD flow in CLL and MM described herein. The adherence to EuroFlow SOPs can be monitored using dedicated biannual quality assessment (QA) rounds.

This chapter describes the implementation of the MRD flow approach in CLL proposed by the ERIC-led consortium in 2007 [11] in combination with EuroFlow technical standardization [67]. The author considers this combination the technology affording the highest level of standardization for MRD flow in this disease. MRD flow in CLL as described herein reproducibly achieves a sensitivity of 10^{-4}. The method has been validated

against RQ-PCR in 530 samples [14] and was applied to detect MRD in at least 1008 patients from large clinical series and trials [20, 30, 31, 33, 35, 39]. The assay combines CD19 and CD5 to detect $CD19^{+}CD5^{+}$ lymphocytes as a whole with CD20, CD38, CD22, CD81, CD79b, and CD43 for the discrimination of malignant from benign lymphocytes [11]. This selection of monoclonal antibodies (mABs) has been maintained over consecutive versions of the updated ERIC recommendations [16, 43]. Whereas the initial 2007 publication provided exact recommendations on clones and gating strategies, subsequent ERIC publications lacked these important specifications [16]. The development of a fully standardized MRD flow assay aiming at reproducible sensitivities of at least 10^{-5} is currently underway.

The design and application of the NGF method in MM is based on EuroFlow technical standards [67] as well. Among other improvements the novel assay features the identification of total plasma cells using scatter, CD138 and CD38, whereas comprehensive unsupervised analyses revealed CD19, CD56, CD27, CD81, CD45, and CD117 in combination with intracellular light chain restriction as most powerful markers for the distinction of aberrant from benign plasma cells [17]. MRD flow in CLL and MM as described herein require the same instrument setup to achieve reproducible staining intensities. However, the two methods described in this chapter utilize two different fully standardized staining protocols. In order to be able to stain sufficient numbers of leukocytes in a cost-efficient way, a bulk lysis protocol was optimized for the novel NGF assay in MM [17], while MRD flow in CLL is still performed without bulk lysis step. The novel bulk lysis protocol will likely form the basis for future developments of highly sensitive flow cytometric MRD assays in other diseases. This chapter focuses in particular on the detailed description of the data analysis using NGF in MM.

2 Materials

2.1 Staining for MRD Flow

1. Washing buffer (PBS + 0.2% BSA + 0.09% NaN_3): Dissolve 4 phosphate-buffered saline (PBS) tablets (Life Technologies), 4 g bovine serum albumin (BSA, AppliChem), and 1.8 g sodium azide (NaN_3, Merck Millipore) at room temperature (RT) in 2 L of distilled water. Store at 4 °C for a maximum of 14 days.
2. Acquisition buffer (PBS + 0.2% BSA): Dissolve 1 PBS tablet (Life Technologies) and 1 g BSA at RT in 500 mL of distilled water. Store at 4 °C for a maximum of 14 days.

3. Premix dilution buffer (PBS + 0.09% NaN_3): Dissolve 1 PBS tablet and 0.45 g NaN_3 at RT in 500 mL of distilled water. Store at RT for a maximum of 28 days.
4. PBS: Dissolve 4 PBS tablets in 2 L of distilled water. Store at RT for a maximum of 14 days.
5. FACS Lysing Solution (Becton Dickinson, BD): Prepare working solution by diluting 100 mL reagent with 900 mL distilled water. Store at RT for a maximum of 28 days.
6. Pharm Lyse (commercially available ammonium chloride (NH_4Cl)-based lysing solution from BD). Prepare fresh working solution daily by diluting 5 mL reagent with 45 mL distilled water.
7. Fix & Perm Lysing Solution for detection of intracellular antigens (Dianova). Ready-to-use solutions. Observe expiry dates provided by manufacturer.
8. Tube 5 mL (polystyrene, 12 × 75 mm, VWR).
9. Blue cap tube 50 mL (polypropylene, Greiner Bio-One).
10. Blue cap tube 15 mL (polypropylene, Greiner Bio-One).
11. Tube 1.5 mL (polypropylene, Eppendorf, Hamburg).
12. Antibody premixes for CLL MRD: Calculate reagents necessary for mAB premixes by multiplying volumes of individual mABs and PBS + NaN_3 from Table 1 by the number of required tests. Combine mABs (clones, labels, vendors outlined in Table 2), and premix dilution buffer for each premix in a capped Eppendorf tube. Store at 4 °C protected from light for a maximum of 14 days.
13. Antibody premixes for MM MRD: Calculate reagents necessary for mAB premixes by multiplying volumes of individual mABs and PBS + NaN_3 from Table 3 by the number of required tests. Combine mABs (clones, labels, vendors outlined in Table 4), and premix dilution buffer for each premix in a capped Eppendorf tube. Store at 4 °C protected from light for a maximum of 14 days.
14. MABs additionally required for MM MRD but not included into the premixes are listed in Table 5. Clones, labels, and vendors are outlined in Table 4. These mABs are not included into premixes. For expiry dates observe manufacturer's recommendation.

2.2 Photomultiplier Tube Voltage Setup

1. Cytometer Setup and Tracking Beads (CS&T beads, BD). Dilute 1 drop in 0.5 mL BD FACS Flow solution.
2. 8-peak rainbow bead calibration particles. Obtain EuroFlow tested lots in Europe from Cytognos, Salamanca, Spain (www.cytognos.com). Dilute 1 drop in 750 μL of distilled water.

Table 1
Reagents and sample preparation for MRD flow in CLL

Tube	Susp. vol. [μL]	mAB [μL]				PBS NaN_3 [μL]	Lysing sol. [μL]	Final vol. [μL]	File # for analysis
		FITC	PE	PerCP Cy5.5	APC				
1	50	CD8 [10]	CD56 [2.5]	CD4 [5]	CD3 [0.5]	7	1000	250	1
2	50	κ/λ light chains [5]		CD19 [5]	CD5 [5]	10	1000	250	5
4	200	CD20 [20]	CD38 [40]	CD19 [20]	CD5 [20]	0	4000	500	2
6	200	CD81 [10]	CD22 [40]	CD19 [20]	CD5 [20]	10	4000	500	3
8	200	CD43 [40]	CD79b [20]	CD19 [20]	CD5 [20]	0	4000	500	4

All volumes given correspond to a single test. If mABs and PBS + NaN_3 are used in premixes, volumes have to be multiplied by test number
Susp. vol. suspension volume of patient sample after adjustment of cell concentration, *FITC* fluorescein isothiocyanate, *PE* phycoerythrin, *PerCP Cy5.5* peridinin chlorophyll protein-cyanin5.5, *APC* allophycocyanin, *lysing sol.* FACS Lysing Solution, *Final vol.* volume in which the suspension is acquired by the cytometer

Table 2
mABs for MRD flow in CLL

Antibody	Label	Clone	Distributor
CD8	FITC	SK1	BD
κ/λ light chains	FITC/PE	TB 28-2/1-155-2	BD
CD20	FITC	L27	BD
CD81	FITC	JS-81	BD
CD43	FITC	1G10	BD
CD56	PE	B159	BD
CD38	PE	HB-7	BD
CD22	PE	HIB22	BioLegend
CD79b	PE	SN8	BD
CD4	PerCP Cy5.5	SK3	BD
CD19	PerCP Cy5.5	SJ25C1	BD
CD3	APC	SK7	BD
CD5	APC	UTCH2	BD

Table 3
mABs for premixes and sample requirements for MRD flow in MM

Tube	Susp. vol. [μL]	mAB [μL] FITC	PE	PerCP Cy5.5	*PECy7*[a]	BV421	BV510	PBS + NaN_3 [μL]
1 + 3	210	CD38 [12]	CD56 [4]	CD45 [20]	*CD19 [10]*	CD138 [4]	CD27 [20]	10

All volumes given correspond to a single sample test. If mABs and PBS + NaN_3 are used in premixes, volumes have to be multiplied by test number
Susp. vol. suspension volume of patient sample after bulk lysis and adjustment of cell concentration to 100,000/μL, *FITC* fluorescein isothiocyanate, *PE* phycoerythrin, *PerCP Cy5.5* peridinin chlorophyll protein-cyanin5.5, *PECy7* phycoerythrin cyanine7, *APC* allophycocyanin, *BV421* Brilliant Violet 421, *BV510* Brilliant Violet 510
[a]CD19 PECy7 is not included into the premix but is added to the cell suspension after the premix

Table 4
mABs for MRD flow in MM

Antibody	Label	Clone	Distributor
CD138	BV421	MI15	BD
CD27	BV510	O323	BioLegend
CD38	FITC	Multi-epitope	Cytognos
CD56	PE	C5.9	Cytognos
CD45	PerCP Cy5.5	HI30	BioLegend
CD117	APC	104D2	BD
CD19	PE Cy7	J3-119	Beckman Coulter
CD81	APC C750	M38	Cytognos
Kappa	APC	Polyclonal	Dako
Lambda	APC C750	Polyclonal	Cytognos

Table 5
Sample preparation for the two separate tubes required for NGF in MM

Tube	Susp. vol. [μL]	mAB [μL] APC	APC C750	PBS NaN_3 BSA [μL]	File # for analysis
1	140	CD117 [5]	CD81 [10]	45	1
3	140	Kappa [5]	Lambda [5]	50	2

All volumes given correspond to a single sample test
Susp. vol. suspension volume obtained after combining backbone markers with cell suspensions according to Table 3, *APC* allophycocyanin, *APC C750* allophycocyanin cyanine 750, *PBS NaN_3 BSA* washing buffer to be added to surface staining reaction

2.3 Light Scatter and Compensation Setup

1. Washing buffer: *see* Subheading 2.1, **item 1**.
2. FACS Lysing Solution (BD): *see* Subheading 2.1, **item 4**.
3. Acquisition buffer: *see* Subheading 2.1, **item 2**.
4. 2 mL peripheral blood from a healthy donor.
5. UltraComp eBeads (Invitrogen).
6. Mouse anti-goat antibody, polyclonal (Invitrogen).
7. MABs as described in Table 6.

2.4 Data Analyses for MRD Flow in MM and CLL

Infinicyt software v2.0 (Cytognos, Salamanca, Spain).

3 Methods

3.1 Staining for MRD Flow in CLL

An initial washing step removes serum including possible soluble antigens as well as cell-membrane-bound immunoglobulins followed by a stain-lyse-wash protocol (*see* **Note 1**). Take necessary biohazard precautions, and dispose of all NaN_3 containing waste according to EU and national regulations.

1. This method has been validated using blood or bone marrow samples with heparin as anticoagulant. Determine the WBC using a 100 μL sample aliquot and an automatic cell counter (e.g., Sysmex XP-300, Sysmex Corporation, Kobe, Japan). Pipet the sample volume corresponding to 15,000,000 leukocytes into a 50 mL tube.
2. Add approximately 45 mL washing buffer, mix well, and centrifuge for 5 min at 540 × *g*.
3. Remove the supernatant using a vacuum system or a Pasteur pipette carefully avoiding disturbing the pellet.
4. Add approximately 10 mL washing buffer, mix well, transfer the suspension to a 15 mL tube, and centrifuge for 5 min at 540 × *g*.
5. Remove the supernatant carefully avoiding disturbing the pellet. In order to achieve a high leukocyte concentration, the supernatant should be removed completely at this stage. Vortex thoroughly.
6. Determine the WBC using a 100 μL aliquot of the pellet. Dilute the pellet with washing buffer if necessary to obtain a cell suspension with a WBC of 10,000 leukocytes/μL.
7. Label eight 5 mL tubes (suitable for sample acquisition for your flow cytometer) with numbers 1–8 and with the laboratory code of the sample. Add the appropriate volumes of the cell

Table 6
Reagents and sample preparation for compensation setup

Tube	Susp. vol. [μL]	mAB [μL]									PBS NaN_3 [μL]	Lysing sol. [μL]	Final vol. [μL]
		BV421	BV510	FITC	PE	PerCP Cy5.5	PE Cy7	APC	APC C750	APC C750			
1	50	–	–	–	–	–	–	–	–	–	50	1000	500
2	50	CD8 [5]	–	–	–	–		–	–	–	45	1000	500
3	50	–	CD27 [5]	–	–	–	–	–	–	–	45	1000	500
4	50	–	–	CD8 [1]	–	–	–	–	–	–	49	1000	500
5	50	–	–	–	CD8 [5]	–	–	–	–	–	45	1000	500
6	50	–	–	–	–	CD5 [15]	–	–	–	–	35	1000	500
7	50	–	–	–	–	–	CD19 [5]	–	–	–	45	1000	500
8	50	–	–	–	–	–	–	CD8 [5]	–	–	45	1000	500
9	50	–	–	–	–	–	–	–	CD81 [3]	–	47	1000	500
10	One[a] drop	–	–	–	–	–	–	–	–	Lambda[a] [5]	–	–	500

[a]Please note that Lambda APC C750 has to be compensated using UltraComp eBeads

suspensions to tubes 1, 2, 4, 6, and 8 (Table 1) while putting aside tubes 3, 5, and 7.

8. Add 25 μL of the appropriate mAB premixes according to Table 1 to tubes 1 and 2. Add 100 μL of the appropriate mAB premixes according to Table 1 to tubes 4, 6, and 8 (*see* **Note 2**).
9. Mix well. Incubate with mABs for 30 min in the dark at RT.
10. Add FACS Lysing Solution to tubes 1, 2, 4, 6, and 8 using the volumes specified in Table 1. Carefully mix the suspension while adding FACS Lysing Solution (*see* **Note 3**). Incubate for 10 min at RT protected from light.
11. Centrifuge at 540 × *g* for 5 min. Decant the supernatant by turning the tubes upside-down. Remove drops of the supernatant that adhere to the rim of the plastic tubes by gently touching a paper towel.
12. Add 4 mL washing buffer to each of the tubes. Vortex thoroughly to detach the pellet from the bottom of the tube and resuspend the pellet.
13. Centrifuge again at 540 × *g* for 5 min. Decant the supernatant by turning the tubes upside-down. Remove drops of the supernatant that adhere to the rim of the plastic tubes by gently touching a paper towel.
14. Add the appropriate volume acquisition buffer to each of the tubes (*see* Table 1, final volume). Vortex thoroughly to detach the pellet from the bottom of the tube and suspend the pellet.
15. Fill tubes 3, 5, and 7 with 2 mL of PBS. These tubes serve to decrease the risk of contamination caused by carry-over between subsequent tubes during acquisition.
16. Acquire data on a flow cytometer within 60 min (store at 4 °C in the dark unless acquired immediately after staining).

3.2 Staining for MRD Flow in MM

An initial washing step is not necessary, as immunoglobulin stains target intracellular antigens and several washing steps are performed for the bulk lysis procedure. The protocol described herein is the most updated version of a protocol initially optimized for NGF in MM [17]. For future updates visit www.euroflow.org ("protocols").

Use common biohazard precautions, and dispose of all NaN_3 containing waste according to EU and national regulations.

1. This method has been validated using bone marrow samples with ethylenediaminetetraacetate (EDTA) as anticoagulant (*see* **Note 4**). Determine the WBC using a 100 μL sample aliquot and an automatic cell counter (e.g., Sysmex XP-300, Sysmex Corporation).

2. Pipet the sample volume corresponding to 20,000,000 leukocytes into a 50 mL tube. Do not use more than 2 mL sample per 50 mL tube. Run multiple 50 mL tubes in parallel if necessary.
3. Add Pharm Lyse solution until reaching 50 mL volume; mix well by using a sample-shaker device for 15 min.
4. Centrifuge for 10 min at 800 × *g*.
5. Remove the supernatant using a vacuum system or a Pasteur pipette carefully avoiding disturbing the pellet.
6. Add 2 mL washing buffer and resuspend the cell pellet thoroughly, preferably by vortexing.
7. Refill with washing buffer until reaching a total volume of 50 mL per tube.
8. Mix well by inverting the tube several times.
9. Centrifuge for 5 min at 800 × *g*.
10. Remove the supernatant using a vacuum system or a Pasteur pipette carefully avoiding disturbing the pellet.
11. Add 2 mL washing buffer and resuspend the cell pellet thoroughly again.
12. Transfer the volume to a 5 mL polystyrene round-bottom tube.
13. Add 2 mL of additional washing buffer into the 50 mL tube to recover remaining cells, and transfer entire volume to the 5 mL tube as well.
14. In case of using several tubes, the cell suspensions from the same sample should be combined at this point.
15. Centrifuge for 5 min at 540 × *g*. Decant the supernatant by turning the tubes upside-down. Remove drops of the supernatant that adhere to the rim of the plastic tubes by gently touching a paper towel. Determine the WBC using an aliquot of 10 μL of the pellet diluted in 290 μL PBS. Dilute the cell suspension pellet with washing buffer if necessary to obtain a cell suspension with a WBC of 100,000 leukocytes/μL.
17. Label three 5 mL tubes (suitable for sample acquisition for your flow cytometer) with numbers 1–3 and with the laboratory code of the sample.
18. Prepare the backbone tube by combining 210 μL cell suspension, 70 μL of the premix, and 10 μL CD19 PECy7 (compare Table 3). Vortex.
19. Pipet 140 μL of backbone mix each into tubes 1 and 3 (Table 5) while putting aside tube 2**.** Add surface markers labelled with APC and APC C750, respectively, to tube 1. Use

washing buffer according to Table 5 to reach a final staining volume of 200 µL for each of the tubes.

20. Mix well by vortexing. Incubate for 30 min in the dark at RT.
21. **Steps 22–25** are applicable to tube 1 (surface marker staining only).
22. Add 2 mL FACS Lysing Solution to tube 1; incubate for 10 min in the dark at RT (*see* **Note 3**).
23. Centrifuge for 5 min at 540 × *g*. Decant the supernatant by turning the tubes upside-down. Remove drops of the supernatant that adhere to the rim of the plastic tubes by gently touching a paper towel.
24. Add 4 mL washing buffer, centrifuge for 5 min at 540 × *g*, and decant the supernatant.
25. Add 500 µL acquisition buffer.
26. **Steps 27–35** are applicable to tube 3 (combined surface marker and intracellular staining).
27. Add 4 mL washing buffer to tube 3; centrifuge for 5 min at 540 × *g*.
28. Carefully decant the supernatant by turning the tube upside-down. At this stage 100 µL or less suspension should remain at the bottom of the tube.
29. Add 100 µL reagent A (fixative from Fix & Perm kit); incubate for 15 min in the dark at RT.
30. Add 4 mL washing buffer to tube 3; centrifuge for 5 min at 540 × *g*.
31. Decant the supernatant by turning the tube upside-down. Remove drops of the supernatant that adhere to the rim of the plastic tube by gently touching a paper towel.
32. Add Kappa-APC and Lambda APC C750 according to Table 5 as well as 100 µL reagent B (permeabilization solution from Fix & Perm kit) to the tube. Incubate for 15 min in the dark at RT (*see* **Note 5**).
33. Add 4 mL washing buffer to tube 3; centrifuge for 5 min at 540 × *g*.
34. Decant the supernatant by turning the tube upside-down. Remove drops of the supernatant that adhere to the rim of the plastic tube by gently touching a paper towel.
35. Add 500 µL acquisition buffer.
36. Prepare tube 2 (safeguard against carry-over) by adding 500 µL PBS.
37. Acquire data on a flow cytometer within 60 min from staining (store at 4 °C in the dark unless acquired immediately after staining).

3.3 Standardized Instrument PMT Voltage Setup

Standardized instrument setup and staining allows for direct comparison of fluorescence intensities of benign and malignant cells over time and between different flow cytometers, thus increasing specificity of MRD flow. For a detailed description of the rationale for the instrument setup and the technical variation determined using this SOP, refer to the manuscripts by the EuroFlow consortium [67, 82]; for additional information and regular updates on the setup, *see* "protocols" at www.euroflow.org (*see* **Note 6**). Target values for rainbow beads are available for BD Canto II as well as for Beckman Coulter Navios instruments for four-color setups. Target values for NGF eight-color setup have been extensively evaluated on BD FACSCanto II instruments. These values for NGF for Beckman Coulter Navios instruments have been established and will become publically available in 2018. The respective protocols are currently being developed for the BD FACSLyric cytometer and are also expected to be released in 2018.

1. Set up the flow cytometer according to manufacturer's recommendations. Allow for sufficient laser warm-up time (at least 30 min). Perform necessary procedures to degas and clean the flow cell and to set the time delay in multi-laser instruments according to manufacturer's instructions (e.g., use the CS&T module of the BD FACSCanto instrument) (*see* **Note 7**).
2. Create a forward scatter-area (FSC-A) vs. side scatter-area (SSC-A) bivariate dot plot. Create a gate ("singlets") within this dot plot. Create a FITC vs. PE dot plot showing all events of region "singlets." Create a gate "7-peak" within that dot plot, so that "7-peak" becomes a subpopulation of "singlets." Create eight histograms each showing one of the fluorescence intensities in channels for Pacific Orange (PacO), Pacific Blue (PacB), fluorescein isothiocyanate (FITC), phycoerythrin (PE), peridinin chlorophyll protein-cyanine 5.5 (PerCP Cy5.5), phycoerythrin cyanine 7 (PE Cy7), allophycocyanin (APC), and allophycocyanin-hilite7 (APC H7) for events within the gate "singlets." For BD FACSCanto instruments, PacB and Brilliant Violet 421 (BV421), PacO and BV510 as well as APC H7 and APC C750 require identical rainbow beads target values. For a discussion of target values on additional instruments, please *see* **Note 6**. Display mean fluorescence and coefficient of variation (CV) for events in "7-peak" for each of the eight fluorescence channels of your instrument.
3. Run diluted rainbow beads at low speed (flow rate: approximately 10 μL/min) at your flow cytometer. Adjust FSC and SSC photomultiplier tube (PMT) voltages to see rainbow beads on scale. Move the "singlets" gate to designate the major bead population showing low FSC and SSC values, thus excluding debris (very low FSC/SSC) and doublets (higher FSC/SSC).

4. Assign the second brightest bead population in the FITC vs. PE dot plot to the "7-peak" population. Check that FITC and PE match target values. Change PMT voltages for the two channels and the position of the "7-peak" gate as necessary.
5. Adjust the PMT voltages for the remaining six fluorescence channels so that the fluorescence intensities of the second brightest peak of the rainbow beads match the targets. EuroFlow provides reference values per master lot at www.euroflow.org.
6. Record the information for 10,000 beads; check again for match with target values. Adjust PMT voltages as necessary in case of deviation from target values. Check for CVs of the second brightest peaks. For optimal instrument performance, the CVs of the seventh peak should be below 4% for FITC and PE channels, below 5% for PerCP Cy5.5, PacO, and PacB channels, and below 8% for APC and APC H7 channels and not exceed 10% for the PE Cy7 channel. Failure to reach those specifications is most commonly caused by problems of the fluidics system of the instrument. Additional cleaning cycles are therefore the recommended first steps for problem correction.
7. PMT values from all eight fluorescence channels resulting in optimal match of the second brightest (seventh) peak fluorescence of rainbow beads with EuroFlow targets are recorded and used for all subsequent compensation experiments and MRD assessments (*see* **Note 8**).

3.4 Standardized Light Scatter and Compensation Setup

Light scatter is standardized using lymphocytes from healthy donor peripheral blood as reference. Compensation is automatically calculated by dedicated software tools (e.g., integrated into BD FACSDiva, BD FACSuite, or Beckman Coulter Navios software) using lymphocytes stained in separate tubes for a brightly expressed antigen for each of the fluorochromes. The compensation setup for Lambda APC C750 requires a bead-based approach, as too few normal donor's cells are positive for that antigen. The procedure described herein allows for the assessment of up to eight fluorochromes and includes tube-specific labels for all antigens included into MM MRD flow. The resulting compensation matrix is also suitable for four-color CLL MRD flow.

1. Label ten 5 mL tubes, add 50 μL of healthy donor blood to tubes 1–9, and incubate with mABs specified in Table 6. Perform the stain-lyse-wash procedure as described in Subheading 3.1.
2. **Steps 3–11** are applicable to tube 10 only.
3. Mix UltraComp eBeads well by inverting the bottle several times.

4. Add one drop to 5 mL tube number 10.
5. Add 5 μL mouse anti-goat mAB and mix thoroughly.
6. Incubate in the dark at 4 °C for 15 min.
7. Add 4 mL washing buffer and centrifuge at 540 × *g* for 5 min. Decant the supernatant by turning the tube upside-down. Remove drops of the supernatant that adhere to the rim of the plastic tube by gently touching a paper towel.
8. Add Lambda APC C750 mAB and mix thoroughly.
9. Incubate in the dark at 4 °C for 30 min.
10. Add 4 mL washing buffer while vortexing and centrifuge at 540 × *g* for 5 min. Decant the supernatant by turning the tube upside-down. Remove drops of the supernatant that adhere to the rim of the plastic tube by gently touching a paper towel.
11. Add 500 μL acquisition buffer.
12. Acquire tube number 1.
13. Gate lymphocytes and display FSC-A and SSC-A mean values for the lymphocyte population. Adjust FSC and SSC PMT voltages so that mean of lymphocyte populations of a healthy donor achieves values of 55,000 (FSC-A) and 13,000 (SSC-A), respectively.
14. Record the PMT values in the FSC-A and SSC-A channels necessary to obtain those values.
15. Run tubes 1–9 using FSC-A and SSC-A voltages determined in **step 11** and fluorescence detector PMT voltages as determined in Subheading 3.3.
16. Acquire 50,000 cellular events per tube in uncompensated mode (i.e., with compensation turned off) at medium speed (flow rate: approximately 60 μL/min).
17. Adjust FSC and SSC to move the UltraComp eBeads into the visible part of the light scatter dot plot. Acquire 10,000 beads.
18. Identify lymphocytes and beads in a light scatter dot plot, gate for positive events in each of the fluorescence channels, and apply the automatic compensation algorithm of your instrument (e.g., "Calculate compensation" of FACSDiva software; *see* **Note 9**).

3.5 Data Acquisition

1. Start your flow cytometer according to manufacturer's recommendations. Use standard procedures to clean the fluidics system and to remove air bubbles from the cytometer.
2. Allow at least 30 min for the lasers to stabilize.
3. Adjust time delay between lasers according to manufacturer's recommendations (e.g., use the CS&T module of the FACSDiva software on FACSCanto II flow cytometers).

4. Set PMT voltages for scatter and fluorescence channels as determined in Subheadings 3.3 and 3.4. Verify daily that the fluorescence intensities of the rainbow beads match the specified EuroFlow ranges for that lot (*see* **Note 8**).
5. Run MRD samples which were stained as described in Subheadings 3.1 and 3.2. Set the flow rate to medium (approximately 60 μL/min).
6. For CLL MRD flow: Record 50,000 cellular events in tubes 1 and 2. For tubes 4, 6, and 8, acquire as many cellular events as possible. For those tubes a time limit is applied so that sample acquisition is automatically stopped before the tube is empty. Tubes 3, 5, and 7 are run in a time-dependent mode for 120 s (tubes contain PBS only, serve to reduce possible sample carry-over between the tubes).
7. For MM MRD flow: For tubes 1 and 3, acquire as many cellular events as possible. For those tubes a time limit is applied so that sample acquisition is automatically stopped before the tube is empty. Tube 2 is run in a time-dependent mode for 120 s (tube contains PBS only, serves to reduce possible sample carry-over between the tubes).
8. Export the acquired data after compensation using Flow Cytometry Standard (.fcs) format.

3.6 Data Analysis for MRD Flow in CLL

A gating strategy using Infinicyt software is described herein (*see* **Note 10**). While Infinicyt features high speed of analysis, innovative statistics in flow cytometry (e.g., principal component analysis, PCA), and advanced tools for simultaneous analysis of multi-tube panels, other software solutions from other manufacturers can be adjusted to serve the same purpose. It should be noted that Infinicyt software assigns events always only to one defined population of the lowest hierarchical level. Moreover, a gating step affects only events which are visible at the time point of gating. Therefore, the result of gating can depend on the temporal order of gating steps. For reproducible analyses the application of predefined profiles and gating strategies can be advantageous. The Infinicyt profile (.inp), reference image, and gating strategy applied here are available upon request from the author (sebastian.boettcher@med.uni-rostock.de). The following instructions by convention quote the *x*-axis parameter followed by the *y*-axis parameter when describing bivariate plots. Tube numbers refer to numbers in the very right column of Table 1. Populations of the population tree to which cells are assigned are highlighted by underscore. Figures to illustrate the gating strategy in several clinical situations (presentation sample, blood and bone marrow assessments, samples taken after CD20 treatment) have been published recently [83].

1. Open fcs files 1, 2, 4, 6, and 8 corresponding to a single MRD assessment (Table 1). Make sure to follow the file order outlined in the very right column of Table 1 when positioning the files in the File List window of the Infinicyt software. Use drag-drop as necessary. The software will automatically identify parameters common to all tubes and designate them with a "C." Disable the common parameter "Time."
2. Create or load a population tree that meets the gate structure outlined in Table 7. Load a reference image representing the autofluorescence of normal lymphocytes from approximately ten normal donors.
3. The following gates are simultaneously applied to files 1–5.
   (a) Assign doublets from a forward scatter-height (FSC-H) vs. FSC-A dot plot to the population Doublets. Doublets characteristically show disproportionally high FSC-A when compared to FSC-H (Fig. 1f).
   (b) Identify nucleated cells in a FSC-A vs. SSC-A density plot that does not depict Doublets (this can, e.g., be accomplished by disabling the visibility of Doublets in the population tree). Include lymphocytes, neutrophils, monocytes, eosinophils, as well as precursors of the erythroid (bottom left) and myeloid lineages. Exclude debris on the left. Assign to the population NUCLEATED CELLS in the hierarchy tree. NUCLEATED CELLS are the denominator for calculating MRD levels, other subpopulations of lymphocytes (e.g., T-cells, NK-cells, etc.), as well as the sensitivity of the MRD assessments per tube.
   (c) Gate Lymphocytes in a FSC-A vs. SSC-A density plot which visualizes NUCLEATED CELLS as small cells (low FSC) with little or no granulation (low SSC). It is important to carefully exclude monocytes (upper right border of this gate) from this analysis gate. Nucleated erythroid precursors (typically present in bone marrow aspirates only) are often hard to distinguish from lymphocytes based on scatter features only and should therefore be included in the Lymphocytes population.
4. Within the Lymphocytes population of tube 1, CD3+ T-cells; their CD4+CD8−, CD4−CD8+, CD4+CD8+, and CD4−CD8− subpopulations; as well as the CD56+CD3− NK-cells are identified using standard approaches (not described in detail; compare Table 7).
5. A CD19/CD5 dot plot and a CD19 histogram depict Lymphocytes from all tubes that contain these parameters (files #2–5 from Table 1). The CD19 histogram additionally shows the autofluorescence of normal lymphocytes from the reference image. CD19−CD5+ T-cells are designated in tube

Table 7
Gating strategy and populations for MRD flow in CLL

Region name	Subpopulation of	Defined in	File # for analysis
Doublets	EVENTS	FSC-H vs. FSC-A dot plot	1–5
NUCLEATED CELLS	EVENTS after exclusion of doublets	FSC-A vs. SSC-A dot plot	1–5
Lymphocytes	NUCLEATED CELLS	FSC-A vs. SSC-A dot plot	1–5
CD3+	Lymphocytes	CD3 APC histogram	1
CD4+CD8−	CD3+	CD4 PerCP Cy5.5 histogram, CD8 FITC histogram	1
CD4−CD8+	CD3+	CD4 PerCP Cy5.5 histogram, CD8 FITC histogram	1
CD4+CD8+	CD3+	CD4 PerCP Cy5.5 histogram, CD8 FITC histogram	1
CD4−CD8−	CD3+	CD4 PerCP Cy5.5 histogram, CD8 FITC histogram	1
CD56+CD3−	Lymphocytes after exclusion of CD3+	CD56 PE histogram	1
CD19−CD5+	Lymphocytes	CD19 PerCP Cy5.5 vs. CD5 APC dot plot	5
B-cells	Lymphocytes	CD19 PerCP Cy5.5 histogram	2–5
CD19CD5_kl_BG	B-cells	CD19 PerCP Cy5.5 vs. CD5 APC dot plot	5
Kappa	CD19CD5_kl_BG	Lambda PE vs. kappa FITC dot plot	5
Lambda	CD19CD5_kl_BG	Lambda PE vs. kappa FITC dot plot	5
CD20_CD38_CD5_BG	B-cells	Boolean "AND" connected gates in CD20 FITC vs. CD5 APC and CD20 FITC vs. CD38 PE dot plots	2
CD81_CD22_CD5_BG	B-cells	Boolean "AND" connected gates in CD22 PE vs. CD5 APC and CD81 FITC vs. CD22 PE dot plots	3
CD43_CD79b_CD5_BG	B-cells	Boolean "AND" connected gates in CD79b PE vs. CD5 APC and CD43 FITC vs. CD22 PE dot plots	4

5. The proportion of T-cells identified as CD19−CD5+ lymphocytes in tube 5 should roughly equal the T-cell fraction determined as CD3+ lymphocytes in tube 1 (column "Frequency" in Infinicyt Population Tree). T-cells identified as

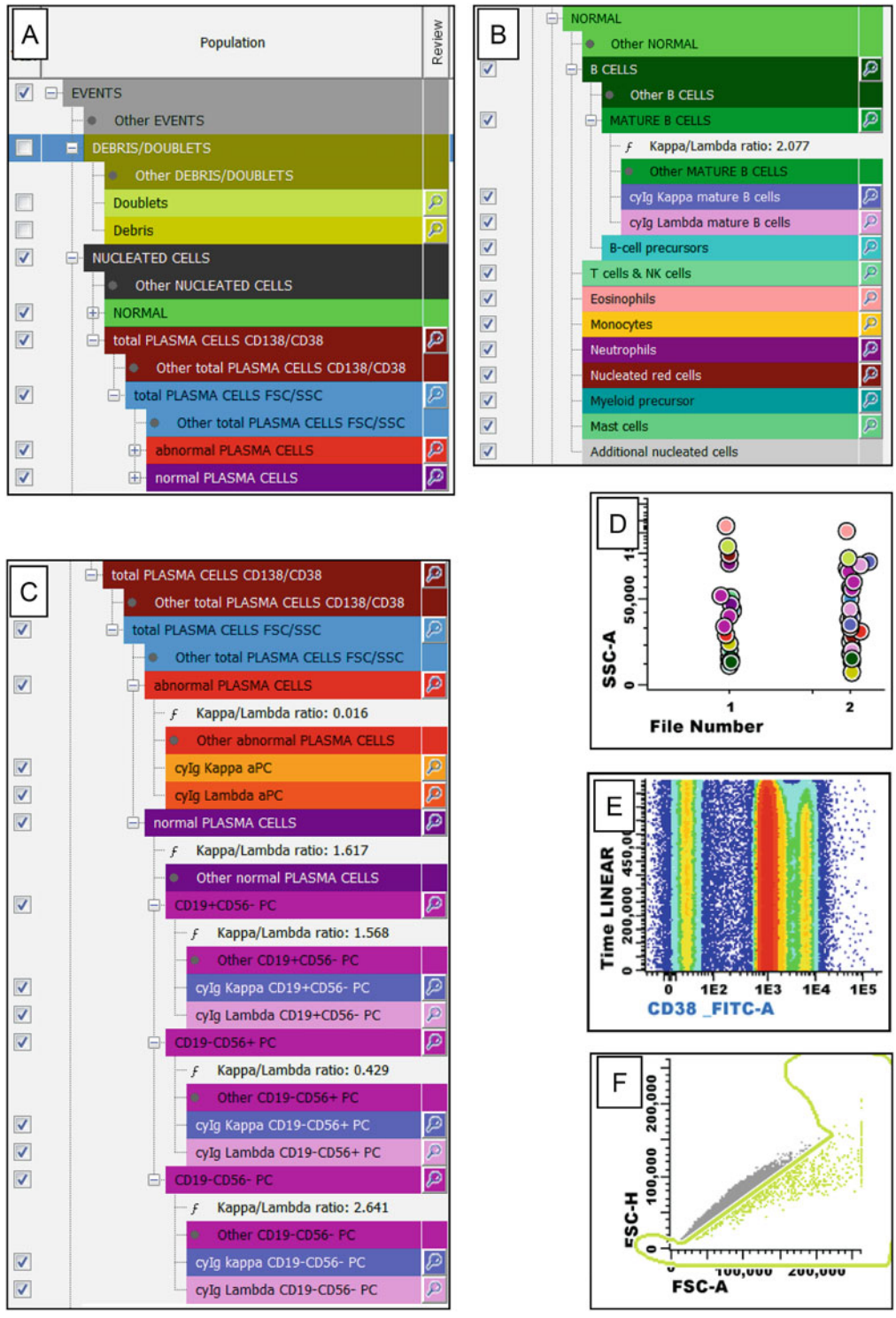


Fig. 1 Population hierarchy and fundamental dot plots for data analysis using MRD flow in MM. (**a**) Main populations. Subpopulations of NORMAL, abnormal PLASMA CELLS, and normal PLASMA CELLS are

CD19−CD5+ lymphocytes serve as negative control for light chain assessment.

6. B-cells are identified as CD19 expressing lymphocytes using a region within the CD19 histogram. The left border of that region is adjusted so that all B-cells are included, whereas the vast majority of T-cells ($CD19^{-}CD5^{+}$) and NK-cells/erythroblast ($CD19^{-}CD5^{-}$) are excluded. This is best assessed using the CD19/CD5 dot plot. The presence of B-cells with low-level CD19 expression occasionally necessitates the inclusion of minor T- and/or NK-cell subpopulations (*see* **Note 11**). The CD19 region must not overlap the autofluorescence of lymphocytes (depicted by the reference image from ten normal donors) in the PerCP Cy5.5 channel. Provided CD19-positive and CD19-negative populations are visible in the CD19 histogram, the left border of the CD19 region should be set at the nadir of the CD19 distribution.
7. Putative CLL cells are designated in a CD19/CD5 dot plot using population CD19CD5_kl_BG within the B-cells acquired in tube 5. The borders of this population are optimized so that the largest population with pathological light chain restriction is assigned. CLL cells typically show CD5 expression; however, occasionally the lower border of CD5 is lower than the upper border of CD5 expression in NK-cells or erythroblasts, i.e., subpopulations of the CLL might show no CD5 expression when compared to background. The kappa and lambda light chains are identified using the lack of light chain expression of the T-cells as reference. A B-cell subpopulation is considered light chain restricted provided the kappa/lambda ratio is below or exceeds the range of 0.38–6 (*see* **Note 12**).
8. CLL cells are identified in tube 2 using two regions in CD20/CD5 and CD20/CD38 dot plots, respectively, and assigned to population CD20_CD38_CD5_BG. The regions target $CD5^{+}CD20^{low}$ and $CD20^{low}CD38^{low}$ cells (*see* **Note 13**). Both gates are combined using a Boolean "AND" operator for final identification of CLL cells. The resulting population CD20_CD38_CD5_BG is shown in a CD38/CD5 dot plot as well as a plot of principal component 1/principal component 2 of the CLL cells' total immunophenotype analyzed by PCA (so-called automated population separator, APS). The CLL cells typically form homogeneous distributions in both

Fig. 1 (continued) collapsed. (**b**) Population tree showing all subpopulations of NORMAL. (**c**) Expanded view of both all subpopulations of abnormal PLASMA CELLS and all subpopulations of normal PLASMA CELLS. (**d**) Medians of all populations separated by file number. This plot simplifies gating strategies separated by file. (**e**) CD38/time density plot to check the stability of acquisition over time. (**f**) FSC-A/FSC-H dot plot of all events. Doublets (pastel green) are simultaneously assigned for both files by virtue of disproportionally high FSC-A

types of plots. However, there are occasional CLL cases with bimodal CD38 distribution.

In bone marrow and blood samples taken more than 9 months after the last CD20 mAB treatment, tube 2 is usually the most informative one. Samples obtained within a 9-month period from the last anti-CD20 treatment lack assessable CD20. Nevertheless, CLL cells can often be identified by lower CD5 than detected in T-cells and low-level CD38. Anti-CD20 mAB therapy leads to profound depletion of mature benign B-cells from the blood, thus allowing the identification of samples from patients who received this treatment within the last 9 months.

9. An analogous gating strategy is applied on CD22/CD5 and CD81/CD22 dot plots to identify CLL cells in tube 3, which are assigned to the CD81_CD22_CD5_BG population. CLL cells can be identified based on the characteristic $CD81^{low}CD22^{low}CD5^{+}$ immunophenotype. The result is verified in a CD81/CD5 plot. The exact position of all gates is best determined by comparison with a pre-treatment sample from the same patient, since the extent of underexpression of both antigens differs between patients. Tube 3 is particularly useful after anti-CD20 mAB therapy.
10. Tube 4 is analyzed in a similar fashion. CLL cells are primarily identified using CD79b/CD5 and CD43/CD5 dot plots with results verified in CD43/CD79b and APS diagrams. $CD79b^{-/low}CD5^{+}$ B-cells are identified first followed by adjustment of the CLL regions within the CD43 vs. CD5 dot plot. CD79b, CD5, and CD43 are typically homogeneous within the CLL clone. Most often, tube 4 provides the second best discrimination of all three CLL MRD tubes.
11. The proportions of CLL events from all nucleated cells are provisionally calculated. Relative differences in CLL MRD levels between the three MRD tubes 2–4 rarely exceed 50%. In case of larger frequency differences between tubes, reasons should be actively looked for and gating strategies scrutinized. Consider only tubes as MRD positive in which more than 20 total events with typical CLL immunophenotype can be identified. Those events should form a cluster in all six dot plots described under **steps 8–10**. A sample is considered MRD positive if at least two out of three tubes are MRD positive (*see* **Note 14**). The MRD level for a sample is the mean of the MRD levels of all positive tubes. The sensitivity of the MRD assessment per tube is calculated as 20 divided by number of NUCLEATED CELLS of that tube. The sensitivity of a negative MRD test equals the second highest sensitivity of the three MRD tubes.

3.7 Data Analysis for MRD Flow in MM

A gating strategy using Infinicyt software is described herein (*see* **Note 15**). The Infinicyt profile (.inp) and gating strategy applied here are available upon request from the author (sebastian.boettcher@med.uni-rostock.de). Naming conventions, features of Infinicyt software, and applicability of different software solutions are explained under Subheading 3.6.

1. Open fcs files 1 and 2 corresponding to a single MRD assessment (Table 5). Make sure to follow the file order outlined in the very right column of Table 5 when positioning the files in the File List window of the Infinicyt software. Use drag-drop if necessary. The software will automatically identify parameters common to both tubes and designate them with a "C."
2. Create or load a population tree as shown in Fig. 1a–c.
3. Display the medians of all populations in a File Number/SSC-A plot (Fig. 1d). Gates in this diagram control whether a particular gating step is applicable to either or both files acquired for a sample (*see* **Note 16**).
4. Check the stability of the laminar flow during sample acquisition using the common parameter CD38 plotted vs. time (Fig. 1e). Disturbances of the sample flow (e.g., caused by air bubbles) are visualized by abrupt shifts of the leukocyte populations.
5. Use a FSC-A/FSC-H dot plot to define the population Doublets for both files simultaneously (Fig. 1f). Doublets characteristically show disproportionally high FSC-A when compared to FSC-H. Scatter characteristics of aberrant plasma cells cannot be predicted. Therefore the Doublets gate should be adjusted so that neither events with high nor with low FSC are erroneously included. Doublets are excluded from all further analyses. Fig. 2a shows the CD45/SSC-A characteristics of the remaining events.
6. The following gates are separately applied to files 1 and 2. The single exception are immunoglobulin light chains (**step c**) that can be identified in file 2 only.
   (a) B CELLS including B-cell precursors are identified as CD19-expressing leukocytes with intermediate to high CD45 expression and low SSC (Fig. 2b/c). Lymphocytes are characterized in the CD45/SSC-A density plot by their low SSC and high CD45 expression. B-cell precursors show a lower CD45 expression (comparable to that of neutrophils or slightly lower than neutrophils) and should be included in the B CELLS gate. The CD19 histogram (Fig. 2c) visualizes the CD19-positive B-cells and CD19-negative T- & NK-lymphocytes. The left border of the CD19 region is adjusted so that all B-cells are included,

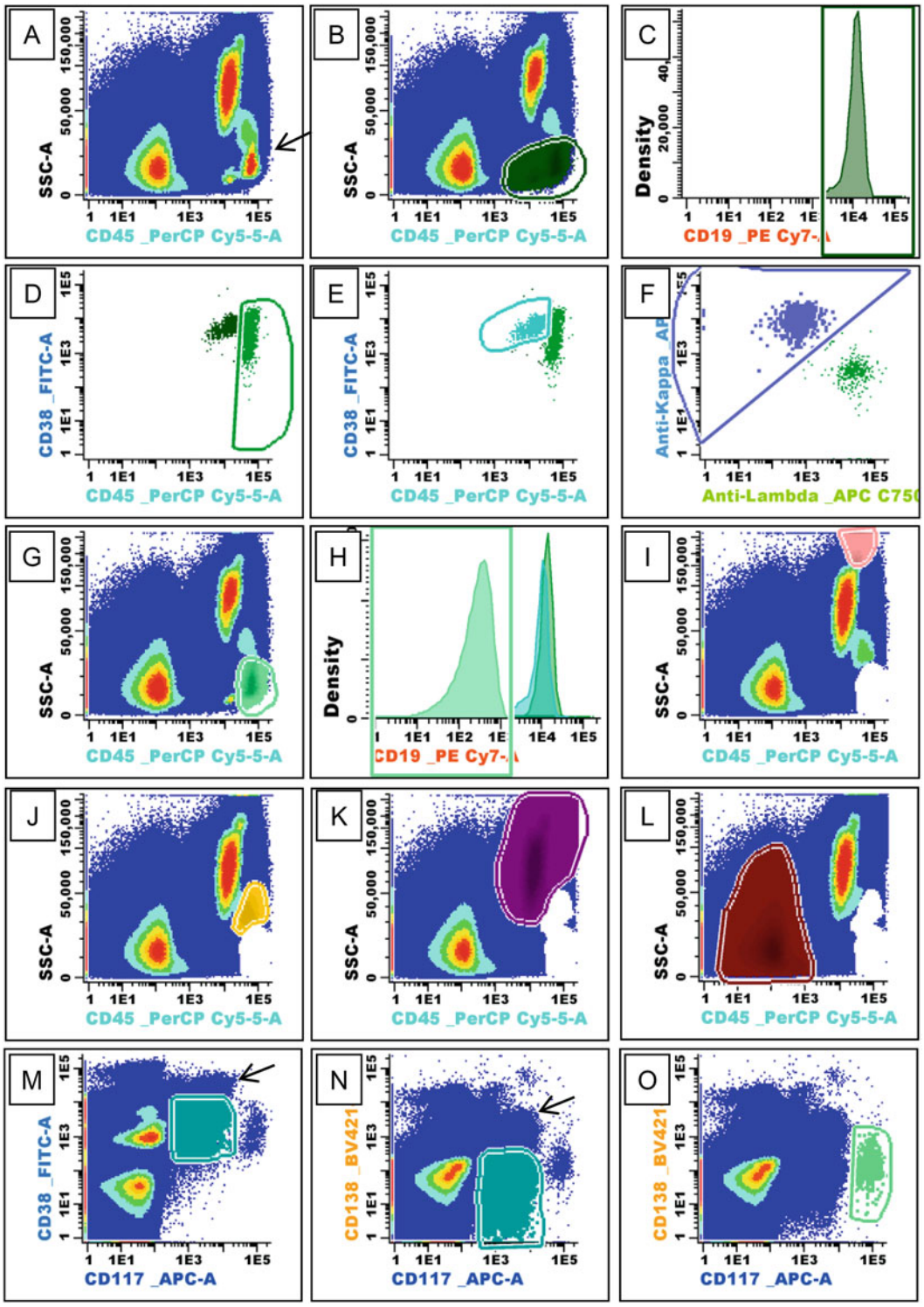


Fig. 2 MRD flow for MM: Identification of major bone marrow cell populations except for plasma cells. (**a**) CD45/SSC density plot of all EVENTS excluding Doublets. Eosinophils, neutrophils, monocytes, lymphocytes

whereas the T- and NK-cells are excluded. It should be set at the nadir of the CD19 distribution. With standard EuroFlow setup (262,144 channel resolution), the lower limit of CD19 positivity is typically found between channels 1000 and 2000.

(b) Within the B CELLS mature B-cells are distinguished from B-cell precursors by their characteristic expression pattern of CD45 and CD38. MATURE B CELLS comprise the lymphocyte subpopulation from B CELLS with the highest expression level of CD45, while CD38 ranges from intermediate to high (Fig. 2d). Compared to MATURE B CELLS B-cell precursors show lower CD45 expression combined with very high levels of CD38 (Fig. 2e). Occasionally plasma cells can contaminate the B CELLS region at this stage of analysis. Normal plasma cells form a cluster with even higher CD38 expression and similar CD45 expression in comparison to B-cell precursors. This plasma cell population will disappear from the B CELLS population upon adjustment of the plasma cell gates at subsequent (CD138/scatter based) steps of the analysis.

(c) Check the correct identification of mature B-cells in file 2 on the basis of their kappa/lambda expression in a dot plot depicting all mature B-cells (Fig. 2f). MATURE B CELLS should express either kappa or lambda. If kappa/lambda double-negative B-cells are visible, either T-/NK-cells or B-cell precursors have been erroneously included into that gate. Assign MATURE B CELLS to cyIg kappa mature B cells and cyIg lambda mature B cells, respectively. The kappa/lambda ratio is calculated automatically

Fig. 2 (continued) (arrow), lymphoid precursors/degranulated basophils, and erythroid precursors are discernible. (**b**/**c**) Identification of total B CELLS including B-cell precursors (dark green) using combined gates in a CD45/SSC dot plot and a CD19 histogram. (**d**/**e**) Identification of MATURE B CELLS (green, **d**) and B-cell precursors (turquoise, **e**) as subpopulations of total B CELLS. (**f**) Kappa-positive MATURE B CELLS (blue) are assigned to the population cyIg kappa mature B cells (Lambda-positive MATURE B CELLS are analogously assigned to cyIg lambda mature B cells, not shown.) (**g**/**h**) Identification of T cells & NK cells (light green) using combined gates in a CD45/SSC dot plot and a CD19 histogram. (**i**/**j**) Eosinophils (pink, **i**) and Monocytes (yellow, **j**) are identified in a CD45/SSC density plot showing all EVENTS but Doublets, B CELLS, and T cells, and NK cells. (**k**/**l**) Neutrophils (dark violet, **k**) and Nucleated red cells (brown, **l**) are gated in a plot that is identical to the density plot shown in (**j**) but shows neither Monocytes nor Eosinophils. (**m**/**n**) Two Boolean "AND" connected gates in CD117/CD38 and CD117/CD138 density plots are used to distinguish Myeloid precursors (green) from all other cellular EVENTS. The density plots do not depict Doublets, B CELLS, and Eosinophils. Arrows indicate the position of CD117-positive (aberrant) plasma cells in this example case ($CD138^{+}CD38^{++}$). (**o**) Mast cells (green) feature very high CD117 expression as shown in a CD117/CD138 density plot of identical populations as in (**n**). The gates shown in (**b**)–(**e**) and in (**g**)–(**l**) are adjusted separately for files 1 and 2. Gate from (**f**) is applied to file 2 only, whereas gates shown in (**m**)–(**o**) are applicable to file 1 data

(normal range 0.38–6, *see* **Note 12**). A deviation from that normal range indicates the presence of an additional mature B-cell non-Hodgkin's lymphoma or that myeloma cells are included into the gate.

(d) Identify T cells & NK cells using a CD45/SSC-A density plot as well as a CD19 histogram. Besides their CD19 negativity (compare **step a**), T-/NK-cells show a comparable or slightly higher CD45 expression than MATURE B CELLS and a similar SSC (Fig. 2g/h). Since the panel lacks T- and NK-cell markers, the T cells & NK cells gate might occasionally contain degranulated basophils.

(e) Define Eosinophils in a CD45/SSC density plot that neither depicts B CELLS nor T cells & NK cells (Fig. 2i). Eosinophils feature very high SSC and autofluorescence. Gate Monocytes within the same density plot (Fig. 2j) using the characteristic high CD45 expression together with intermediate SSC.

(f) Assign Neutrophils (Fig. 2k) and Nucleated red cells (Fig. 2l) to their respective populations in the hierarchy tree using a CD45/SSC density plot that does not display B CELLS, T cells & NK cells, Eosinophils, or Monocytes.

7. Identify Myeloid precursors in file 1 using two "AND" connected populations in CD117/CD38 and CD117/CD138 density plots (Fig. 2m/n). In both plots the populations Eosinophils and B CELLS are invisible to improve the separation of the remaining populations. Myeloid precursors are defined as $CD117^{+}CD38^{dim}CD138^{-}$ cell population. Using EuroFlow standard settings at a 262,144 channel resolution, the lower border of CD117 expression typically ranges between 1500 and 4000 channels, while the upper limit is usually situated between 10,000 and 20,000. Dimmer CD38 expression and lack of CD138 are the distinguishing features of those myeloid precursors vs. CD117-positive MM cases.
8. Define Mast cells in file 1 using a CD117/CD138 density plot not depicting Eosinophils and B CELLS as $CD117^{high}$ population (Fig. 2o). Mast cells represent the leukocyte population in bone marrow with the highest CD117 expression (lower gate limits typically higher than 20,000 channels). Due to autofluorescence mast cells appear as $CD138^{dim}$. The percentage of mast cells serves as internal control for the quality of bone marrow aspirates. Mast cells counts below 0.002% of all nucleated cells were not observed in normal bone marrow aspirates and might suggest hemodilution of the sample [17].
9. Identify Debris using a FSC-A/SSC-A dot plot depicting all populations (except Doublets). This step has to be performed separately by file. The 2SD curves (curves that represent the

second standard deviation of a population) of B CELLS, Eosinophils, Monocytes, and Neutrophils are shown for visual orientation. The Debris gate should include events with low to high SSC-A and low FSC-A typically sparing events included into the 2SD curves of the mentioned major leukocyte populations (Fig. 3a).

10. Plasma cells as a whole (including normal and aberrant subpopulations) are identified in a two-step procedure separately per file (Fig. 3b–h).
    (a) In order to set the boundaries of the gate total PLASMA CELLS CD138/CD38 for the identification of $CD138^{high}CD38^{high}$ events as precisely as possible, CD138/CD38 will be presented as both dot and density plots (Fig. 3b–e). Moreover, for better discrimination, two of those plots do not visualize Debris, Doublets, Eosinophils, Monocytes, or Neutrophils, i.e., populations with high autofluorescence that can in particular interfere with CD138 staining. However, the population total PLASMA CELLS CD138/CD38 is assigned to the population hierarchy tree using a dot plot that shows all EVENTS but Debris, Doublets, and Eosinophils, thus guaranteeing high sensitivity for plasma cells. Particular attention should be directed at the identification of abnormal plasma cell populations which might present with reduced CD38 and CD138 expression in comparison to normal plasma cells.
    (b) Total PLASMA CELLS CD138/CD38 are displayed in a FSC-A/SSC-A dot plot. The total plasma cells are defined more precisely as events with intermediate FSC and SSC and assigned to the population total PLASMA CELLS FSC/SSC. This region should be set large enough to take into account the variable size of normal and abnormal plasma cells, while in particular excluding $FSC^{high}SSC^{high}$ events (usually granulocytes) and discrete clusters with very low SSC and FSC (debris) (Fig. 3f). For orientation, two additional density plots show the CD138/CD38 properties of the events that are included and excluded by the FSC/SSC gate on plasma cells, respectively (Fig. 3g/h). Excluded events usually should form a cluster continuously extending from the lower left border of the plasma cell population as a whole.
11. The events from the total PLASMA CELLS FSC/SSC population are divided into abnormal PLASMA CELLS and normal PLASMA CELLS using six dot plots each for files 1 (Fig. 4a–f) and 2 (Fig. 4g–l).

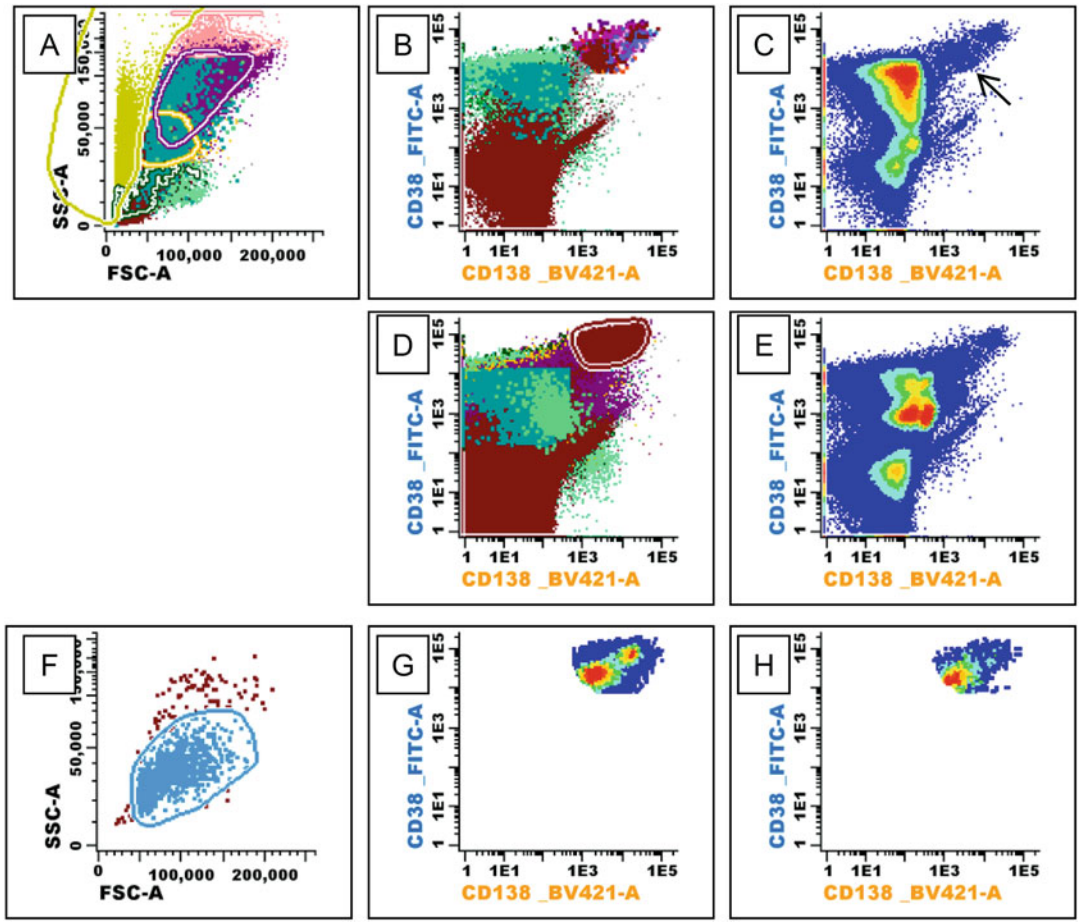


Fig. 3 MRD flow for MM: Exclusion of debris and delineation of total plasma cells. The gating strategy is illustrated using file 1 as an example. However, all gates shown have to be adjusted separately for files 1 and 2. (**a**) Identification of Debris as events with low FSC in a FSC/SSC dot plot. Two standard deviations of B CELLS (green), Monocytes (yellow), Neutrophils (dark violet), and Eosinophils (pink) serve as guidance for the right border of the Debris gate. All EVENTS but Doublets are depicted. (**b**–**e**) Identification of total PLASMA CELLS CD138/CD38. CD138/CD38 dot plots (**b**, **d**) and density plots (**c**, **e**) are used to visualize all EVENTS but Debris, Doublets, Eosinophils, Monocytes, and Neutrophils (**b**, **c**) or all EVENTS but Debris, Doublets, and Eosinophils (**d**, **e**). Please note the presence of a plasma cell subpopulation with low-level CD138 and CD38 (arrow, **c**). The gate is adjusted in a dot plot as shown in (**d**) in order to include even plasma cells that might have been erroneously assigned to previously defined cell populations. Figures 3**b**, **c**, and **e** serves to visualize the exact borders of the plasma cells in a particular case. (**f**) The FSC/SSC dot plot depicts total PLASMA CELLS CD138/CD38. Contaminating cellular events form clusters outside of the population total PLASMA CELLS FSC/SSC. (**g**/**h**) Density plots of total PLASMA CELLS CD138/CD38 that are included into (**g**) or are excluded by (**h**) the border of total PLASMA CELLS FSC/SSC. If the population total PLASMA CELLS FSC/SSC is adjusted correctly, (**g**) shows one to two discrete clusters, while (**h**) shows one cluster adjacent to the lower left border of the total population

(a) Abnormal PLASMA CELLS and normal PLASMA CELLS each are defined using three gates that are connected by Boolean "AND." There are separate gates for files 1 and 2 that are adjusted separately. The gates

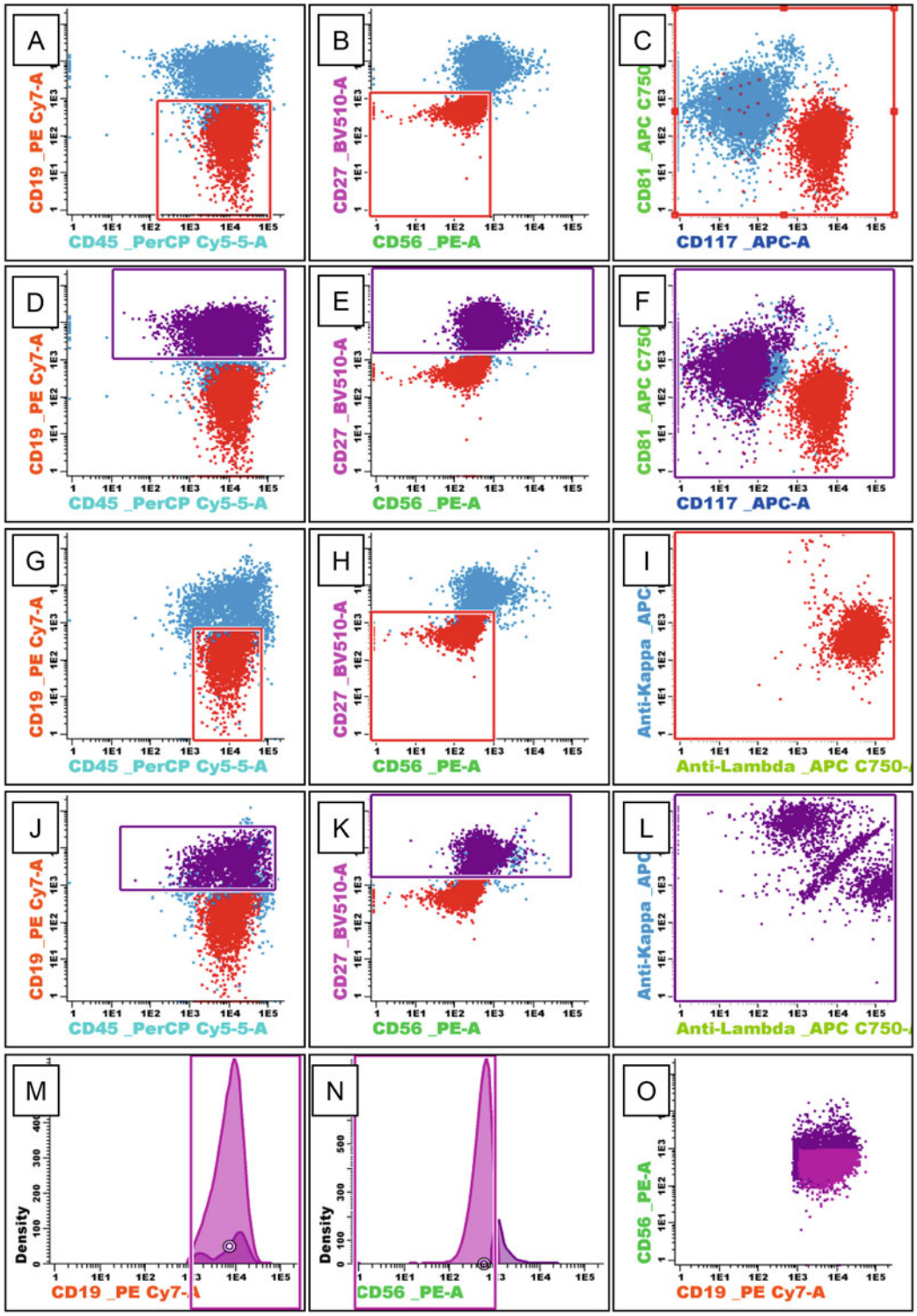


Fig. 4 Distinction between aberrant and normal plasma cells using MRD flow for MM. (**a**–**f**) Dot plots of file 1. (**g**–**l**) Dot plots of file 2. Cellular events from the population total PLASMA CELLS FSC/SSC only are shown.

define any combination of immunophenotypes discernible within the CD45/CD19, CD56/CD27, and CD117/CD81 dot plots from file 1 as abnormal and normal, respectively, whereas this distinction is made based on CD45/CD19, CD56/CD27, and lambda/kappa dot plots for file 2.

(b) The most universal and unequivocal sign of clonality is a pathological light chain restriction (kappa/lambda ratio $\leq$0.38 or $\geq$6, *see* **Note 12**). Starting with file 2 gates within the CD45/CD19 and CD56/CD27 dot plots are optimized so that the largest population showing light chain restriction is defined (Fig. 4g/h). The vast majority of MM patients are expected to show at least one of the following immunophenotypic aberrations: $CD19^{-}$, $CD45^{low}$, $CD56^{+}$, or $CD27^{low}$ (*see* **Note 17**). Use as few parameters for the identification of the aberrant population as possible, thus leaving the other markers for unsupervised confirmation of aberrancies. Once a light chain-restricted plasma cell population was detected in file 2 and assigned to abnormal PLASMA CELLS, the same population should be identified in file 1 (Fig. 4a/b) and additionally assigned to abnormal PLASMA CELLS. File 1 can be used for further confirmation of an aberrant immunophenotype by additionally revealing $CD117^{+}$ and/or $CD81^{low}$ plasma cells.

(c) Populations of normal and abnormal plasma cells are mutually exclusive. For identification of normal plasma cells, the remaining (not as aberrant assigned) plasma cells should be assigned to normal PLASMA CELLS, again primarily using CD45/CD19 and CD56/CD27 dot plots in both files 1 and 2 (Fig. 4d/e, j/k).

(d) Normal plasma cells are divided into three subpopulations (CD19+CD56− PC, CD19−CD56+ PC, CD19−CD56− PC) according to their CD19 and CD56 expression in two respective histograms (Fig. 4m/n). The additional CD19/

Fig. 4 (continued) Three gates each are connected by Boolean "AND" to delineate abnormal PLASMA CELLS (red) and normal PLASMA CELLS (violet) in files 1 and 2, respectively. If possible, the distinction should be based upon CD45/CD19 and CD56/CD27 expression levels, thus leaving CD117, CD81, and light chain expression as markers to confirm the aberration of abnormal PLASMA CELLS. Gating using CD45/CD19 and CD56/CD27 facilitates the identification of similar plasma cell subpopulations in files 1 and 2. (**m**–**o**) Boolean "AND" combinations of regions set in CD19 and CD56 histograms allow for the definition of CD19+CD56− PC, CD19−CD56+ PC, and CD19−CD56− PC as subpopulations of normal PLASMA CELLS in file 2, which are further visualized using a CD19/CD56 dot plot. All three subpopulations of normal PLASMA CELLS are individually scrutinized for light chain restriction

CD56 dot plot serves as visual control for possible discrete populations (Fig. 4o).

12. Assign the kappa and lambda light chains to the respective subpopulations of abnormal PLASMA CELLS, CD19+CD56− PC, CD19−CD56+ PC, and CD19−CD56− PC defined in file 2 (figures not shown). The software will automatically calculate kappa/lambda ratios. Confirm that all subpopulations of normal PLASMA CELLS show kappa/lambda ratios within the normal range.
13. Check for possible aberrant expression ($CD38^{low}$, $CD45^{low}$, $CD27^{low}$, $CD81^{low}$, $CD117^{+}$) within the populations CD19+CD56− PC, CD19−CD56+ PC, and CD19−CD56− PC defined in file 1 to rule out the erroneous inclusion of myeloma cells into these populations of normal plasma cells.
14. Calculate results based upon counts from file 1 only. A sample is considered MRD positive provided that at least 20 tumor plasma cells are assigned to abnormal PLASMA CELLS from file 1 and is quantifiable provided that number equals at least 50. Exact MRD levels should be given in case of a quantifiable result and calculated as abnormal PLASMA CELLS divided by numbers of NUCLEATED CELLS. The limit of detection (LOD) is calculated as 20/numbers of NUCLEATED CELLS. The lower limit of quantification (LLOQ) is calculated as 50/numbers of NUCLEATED CELLS. In MRD-negative samples, the LOD has to be provided. For MRD-positive samples below the LLOQ, the result is reported as "positive, between LOD and LLOQ." The immunophenotype of the myeloma cells has to be described. The report should be flagged as possibly hemodiluted in cases with mast cell percentages below 0.002% of all nucleated cells.

4 Notes

1. Neither the 2007 international harmonization approach for MRD flow in CLL [11] nor later versions of the ERIC guidelines [16, 43] required a particular staining or instrument setup protocol, which in turn precluded the direct comparison of staining intensities between different laboratories. To improve standardization the author has been using the EuroFlow SOP for staining and instrument setup in all our experiments, including MRD flow, for the last 11 years (www.euroflow.org).

   Different sample preparation methods have been investigated by the EuroFlow consortium, and the technique presented herein was found to be most reliable with respect to practicability, scatter properties, and preservation of fluorescence

intensities [67]. The addition of a protein source (we recommend BSA) to the washing solution is essential for reduced cell loss. Very recently the consortium established that 0.2% BSA suffices for that purpose, while previous EuroFlow publications [17, 67] recommended 0.5% BSA. NaN_3 prevents capping of mABs bound to vital cells and reduces the risk of bacterial contamination in protein-containing solutions. As it is poisonous, the final resuspension step omits that reagent (resulting in so-called acquisition buffer) so that the flow cytometer is not contaminated. Washing of cell suspensions prior to incubation with mABs is strictly speaking essential only for staining with anti-immunoglobulin mABs. We use a bulk-wash approach for the sample as a whole in order to improve the comparability of the staining patterns in the individual tubes.

2. With the exception of CD5, all mABs used for MRD flow in CLL were selected from the reagents recommended by the international standardization project [11]. The reagents specified in this chapter show low lot-to-lot variation (relative differences between consecutive lots <30%) when checked in our laboratory. We additionally recommend careful monitoring of lot-to-lot variability between consecutive lots of mABs [84].
3. FACS Lysing Solution contains a fixative that might increase doublet formation when added to cells that adhere to each other. We therefore add this solution mL by mL while vortexing the cell suspension.
4. While for CLL MRD flow heparin is considered the preferred anticoagulant (assumed to improve viability of leukocytes), this molecule interferes with CD138 staining patterns so that EDTA is the anticoagulant of choice for NGF in MM.
5. Residual erythrocytes are expected to be completely lysed after completion of this incubation step only.
6. PMT voltages are adjusted using rainbow beads that emit fluorescence light of stable intensity in all channels of modern flow cytometers. Voltages are tuned so that the fluorescence of those beads match target ranges on each of the fluorescence channels, resulting in highly similar optical output of all instruments for identical molecule numbers per cell. This is a prerequisite for direct comparison of fluorescence intensities measured at different instruments and over time at the same cytometer. The target values of bead fluorescence emissions were initially created by analyzing many different flow cytometers at all EuroFlow consortium laboratories. These standard values were chosen to minimize electronic noise and at the same time allow the on-scale analysis of all antigens used in clinical flow cytometry. For subsequent master lots of beads, these reference values are cross-calibrated at a EuroFlow

reference laboratory. Target values for the seventh peak of the rainbow beads (i.e., the second brightest bead population) for individual master lots are available at www.euroflow.org for FACSCanto II (BD), BD LSR II (BD), and CyAn ADP (Dako/Beckman Coulter) cytometers for the fluorochromes Pacific Blue, Pacific Orange, FITC, PE, PerCP Cy5.5, PECy7, APC, and APC H7. On these reference instruments, targets for BV421 equal targets for Pacific Blue, targets for BV510 equal targets for Pacific Orange, and targets for APC H7 equal the APC C750 targets. Target values for Beckman Coulter Navios instruments are available for Pacific Blue, OC515, FITC, PE, PerCP Cy5.5, PECy7, APC, and APC C750, whereas target values for BV421 and BV510 are scheduled to be released later in 2018. The consortium is currently establishing target values and SOPs for the BD FACSLyric flow cytometer.

7. EuroFlow setup and compensation SOPs have been extensively tested on FACSCanto II (BD), BD LSR II (BD), and CyAn ADP (Dako/Beckman Coulter) cytometers by the consortium [67]. The personal experience of the author is restricted to FACSCanto II and FACSLyric machines.
8. Bead emissions have to match optimal target values ±15%. Set your instrument once a month at optimal values. Check daily for adherence to target values before acquiring samples. In case of a sudden deterioration of CV or loss in fluorescence intensity, most often problems in the fluidics system are causative. This can usually be overcome by vigorous cleaning of the instrument and removal of air bubbles. Gradual decline in instrument performance usually requires the help of a service engineer from the manufacturer of your cytometer. The author with the help of service engineers most often discovered laser problems or other problems of the optical system (e.g., de-adjustment of lenses, gradual disconnection of fiber-optic light guides) as underlying reason for such issues.
9. Compensation and light scatter values can be used for a month as long as the PMT values for the rainbow beads are within target values. If an adjustment of PMT values of the fluorescence channels becomes necessary or maintenance at fluidics or optical systems of your instrument is required, a new setup of compensation and light scatter is strongly recommended (refers specifically to BD FACSCanto II instruments).
10. The gating approach follows the general principles of the international standardization project on MRD flow in CLL [11]. Staining for CD45/CD14/CD19/CD3 is recommended by this consortium to identify B-T-cell-doublets, assess the contamination of the B-cell gate with T-cells, and refer CLL numbers to $CD45^+$ leukocytes. While the

discrimination against T-cells is crucial for MRD flow, we found a high specificity when comparing flow cytometry to ASO primer PCR without using this test [14]. Moreover, identifying nucleated cells by light scatter allowed us to demonstrate a high concordance of MRD levels measured by flow and PCR. Consequently, we prefer to omit this CD45/CD14/CD19/CD3 test in order to save mABs. According to our own [7, 14] and the experience of an international consortium [11], the most important factor for the sensitivity of MRD flow in CLL is the total number of acquired cells. We therefore use the total cell number to calculate sensitivity of a test [14] and recommend not to analyze samples in which less than 200,000 nucleated cells are measured per tube [30]. Data from our group demonstrated that CD20 cannot be assessed up to 9 months after the last treatment with one of the CD20 mABs, e.g., rituximab [14], obinutuzumab, or ofatumumab. Mature benign B-cells are more sensitive to rituximab treatment than CLL cells and absent from PB for similar periods of time [14]. Therefore, B-cells presenting with the immunophenotype $CD19^{+}CD5^{+}CD22^{low}CD81^{low}CD79b^{low\ to\ -}CD43^{+}$ can be considered CLL if no CD20-expressing lymphocytes are discernible. The specificity of MRD flow is greatly enhanced by comparison with the expression levels of all antigens at presentation, made possible by standardized sample preparation and instrument setup. Individual CLL patients show underexpression of the B-cell antigens CD22, CD81, and CD79b to a different extent. Knowledge of the exact initial immunophenotype is particularly helpful for interpretation of CD81 vs. CD22 dot plots. In our experience all antigens included into the MRD flow panel are generally stable in the MRD situation relative to the initial presentation. The only exception is CD20 that is not assessable up to 9 months after anti-CD20 treatment. In our own work in 530 samples from CLL patients, a sensitivity of at least 10^{-4} and an excellent specificity of this MRD flow approach were demonstrated by comparison to ASO *IGH* RQ-PCR [14].

11. The exclusion of T-cells using this gate is important, while all B-cells have to be included. In case of doubt, set the gate to deliberately include some T-cells; identify the position of T-cells ($CD20^{-}CD5^{high}CD81^{high}CD22^{-}$) in CD20/CD5, CD81/CD22, and CD22/CD5 dot plots; and exclude T-cells using, e.g., CD22. The CD5 expression of T-cells is usually higher than on CLL and benign B-cells.
12. There is no generally accepted definition for light chain restriction. The definition provided was established in our work on 4-color flow in MCL [6].

13. Mature benign B-cells can show considerable upregulation of CD5 after therapy, so that up to 90% of benign B-cells express this marker [11, 85]. A similar effect was observed for CD43, while B-cell antigens (CD22, CD79b) and CD81 showed stable expression levels on benign regenerating B-cells after treatment.
14. Using comparative analysis to ASO *IGH* RQ-PCR, we found a minimal requirement of more than 20 cells in at least 2 tubes sufficient evidence of MRD [14], while the international consortium set this limit at 50 cells per tube being detectable in at least 2 tubes [11]. Our minimum specification of 40 positive events in total for a quantifiable result is in the same range as the requirement of a total 50 positive events per quantifiable positive test defined by the consensus guidelines on MRD flow in MM [66].
15. The gating strategy proposed herein features sequential gating. Doublet discrimination is followed by the identification of the major leukocyte subpopulations in bone marrow, which in turn guide the exclusion of debris based on scatter properties. A particularly important gating step is the unequivocal assignment of the plasma cell population as a whole. To achieve this with high accuracy, leukocyte subpopulations with high autofluorescence are excluded prior to application of CD138/CD38 gating. CD138 is a marker particularly susceptible to background signal caused by autofluorescence [86]. Scatter is used to further define the plasma cell populations. After the identification of all plasma cells in a sample, these cells are divided into benign and aberrant subpopulations. The EuroFlow consortium has recently proposed a fully automated gating strategy using databases of normal bone marrow samples [17]. This strategy is not described herein as it currently still lacks validation on instruments different from BD FACSCanto II.
16. Scatter, but to a lesser extent also certain fluorescence markers, is affected by the different lysing procedures (surface and combined surface plus intracellular stains, respectively) so that with the exception of doublet discrimination, all gates have to be adjusted separately for both tubes.
17. The marker with the highest frequency of aberrations is CD19 (>90% of MM show underexpression [10, 13, 17]). However, it should be noted that CD19-negative subpopulations of plasma cells occur in healthy donors and that these subpopulations can expand, e.g., after intensive treatment. Loss of CD45 and expression of CD56 are the next most frequent aberrations. However, in particular CD56 overexpressing benign plasma cells can be pronounced in posttreatment samples.

According to the observations of the author in more than 1000 analyzed samples, CD27 underexpression and the aberrant expression of CD117 are specific for MM. With the exception of CD81, all markers included in the NGF assay are stable from presentation to MRD. While knowledge of the immunophenotype of a MM patient at presentation is no prerequisite for successful NGF, this information can be helpful in selected cases.

Acknowledgments

The author is grateful to E. Harbst, L. Falck, J. Hanani, D. Paape, B. Wiebeck, S. Lange, E. Koppitz, and L. Henseleit for excellent technical support with establishing the protocols and thanks the members of the EuroFlow and ERIC consortiums for long-standing collaborations. B. Wiebeck, S. Lange, PhD, R. Engelmann, PhD, and D. Paape are acknowledged for critical reading of the manuscript. Parts of this manuscript are based on a previously published chapter in *Methods in Molecular Biology* by the author, Matthias Ritgen, and Michael Kneba [87].

References

1. Fischer K, Bahlo J, Fink AM et al (2016) Long-term remissions after FCR chemoimmunotherapy in previously untreated patients with CLL: updated results of the CLL8 trial. Blood 127:208–215
2. Hermine O, Hoster E, Walewski J et al (2016) Addition of high-dose cytarabine to immunochemotherapy before autologous stem-cell transplantation in patients aged 65 years or younger with mantle cell lymphoma (MCL Younger): a randomised, open-label, phase 3 trial of the European Mantle Cell Lymphoma Network. Lancet 388:565–575
3. European Medicines Agency Committee for Medicinal Products for Human Use (2016) Appendix 4 to the guideline on the evaluation of anticancer medicinal products in man. http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2016/02/WC500201945.pdf, Accessed 15 Mar 2018
4. Fink AM, Bottcher S, Ritgen M et al (2013) Prediction of poor outcome in CLL patients following first-line treatment with fludarabine, cyclophosphamide and rituximab. Leukemia 27:1949–1952
5. Cramer P, von Tresckow J, Bahlo J et al (2018) CLL2-BXX Phase II trials: sequential, targeted treatment for eradication of minimal residual disease in chronic lymphocytic leukemia. Future Oncol (London, England) 14:499–513
6. Bottcher S, Ritgen M, Buske S et al (2008) Minimal residual disease detection in mantle cell lymphoma: methods and significance of four-color flow cytometry compared to consensus IGH-polymerase chain reaction at initial staging and for follow-up examinations. Haematologica 93:551–559
7. Bottcher S, Ritgen M, Pott C et al (2004) Comparative analysis of minimal residual disease detection using four-color flow cytometry, consensus IgH-PCR, and quantitative IgH PCR in CLL after allogeneic and autologous stem cell transplantation. Leukemia 18:1637–1645
8. van der Velden VH, Cazzaniga G, Schrauder A et al (2007) Analysis of minimal residual disease by Ig/TCR gene rearrangements: guidelines for interpretation of real-time quantitative PCR data. Leukemia 21:604–611
9. Puig N, Sarasquete ME, Balanzategui A et al (2014) Critical evaluation of ASO RQ-PCR for minimal residual disease evaluation in multiple myeloma. A comparative analysis with flow cytometry. Leukemia 28:391–397

10. Rawstron AC, Orfao A, Beksac M et al (2008) Report of the European Myeloma Network on multiparametric flow cytometry in multiple myeloma and related disorders. Haematologica 93:431–438
11. Rawstron AC, Villamor N, Ritgen M et al (2007) International standardized approach for flow cytometric residual disease monitoring in chronic lymphocytic leukaemia. Leukemia 21:956–964
12. Rawstron AC, de Tute R, Jack AS et al (2006) Flow cytometric protein expression profiling as a systematic approach for developing disease-specific assays: identification of a chronic lymphocytic leukaemia-specific assay for use in rituximab-containing regimens. Leukemia 20:2102–2110
13. Mateo G, Montalban MA, Vidriales MB et al (2008) Prognostic value of immunophenotyping in multiple myeloma: a study by the PETHEMA/GEM cooperative study groups on patients uniformly treated with high-dose therapy. J Clin Oncol 26:2737–2744
14. Bottcher S, Stilgenbauer S, Busch R et al (2009) Standardized MRD flow and ASO IGH RQ-PCR for MRD quantification in CLL patients after rituximab-containing immunochemotherapy: a comparative analysis. Leukemia 23:2007–2017
15. Paiva B, Cedena MT, Puig N et al (2016) Minimal residual disease monitoring and immune profiling in multiple myeloma in elderly patients. Blood 127:3165–3174
16. Rawstron AC, Fazi C, Agathangelidis A et al (2016) A complementary role of multiparameter flow cytometry and high-throughput sequencing for minimal residual disease detection in chronic lymphocytic leukemia: an European Research Initiative on CLL study. Leukemia 30:929–936
17. Flores-Montero J, Sanoja-Flores L, Paiva B et al (2017) Next Generation Flow for highly sensitive and standardized detection of minimal residual disease in multiple myeloma. Leukemia 31:2094–2103
18. Goede V, Fischer K, Busch R et al (2014) Obinutuzumab plus chlorambucil in patients with CLL and coexisting conditions. N Engl J Med 370:1101–1110
19. Hallek M, Fischer K, Fingerle-Rowson G et al (2010) Addition of rituximab to fludarabine and cyclophosphamide in patients with chronic lymphocytic leukaemia: a randomised, open-label, phase 3 trial. Lancet 376:1164–1174
20. Stilgenbauer S, Leblond V, Foa R et al (2018) Obinutuzumab plus bendamustine in previously untreated patients with CLL: a subgroup analysis of the GREEN study. Leukemia 32:1778–1786
21. Roberts AW, Davids MS, Pagel JM et al (2016) Targeting BCL2 with Venetoclax in relapsed chronic lymphocytic leukemia. N Engl J Med 374:311–322
22. Seymour JF, Ma S, Brander DM et al (2017) Venetoclax plus rituximab in relapsed or refractory chronic lymphocytic leukaemia: a phase 1b study. Lancet Oncol 18:230–240
23. Stilgenbauer S, Eichhorst B, Schetelig J et al (2016) Venetoclax in relapsed or refractory chronic lymphocytic leukaemia with 17p deletion: a multicentre, open-label, phase 2 study. Lancet Oncol 17:768–778
24. Byrd JC, Brown JR, O'Brien S et al (2014) Ibrutinib versus ofatumumab in previously treated chronic lymphoid leukemia. N Engl J Med 371:213–223
25. Furman RR, Sharman JP, Coutre SE et al (2014) Idelalisib and rituximab in relapsed chronic lymphocytic leukemia. N Engl J Med 370:997–1007
26. Ahn IE, Farooqui MZH, Tian X et al (2018) Depth and durability of response to ibrutinib in CLL: 5-year follow-up of a phase II study. Blood 131:2357–2366
27. Burger JA, Tedeschi A, Barr PM et al (2015) Ibrutinib as initial therapy for patients with chronic lymphocytic leukemia. N Engl J Med 373:2425–2437
28. Fischer K, Al-Sawaf O, Fink AM et al (2017) Venetoclax and obinutuzumab in chronic lymphocytic leukemia. Blood 129:2702–2705
29. Moreno C, Villamor N, Colomer D et al (2006) Clinical significance of minimal residual disease, as assessed by different techniques, after stem cell transplantation for chronic lymphocytic leukemia. Blood 107:4563–4569
30. Bottcher S, Ritgen M, Fischer K et al (2012) Minimal residual disease quantification is an independent predictor of progression free and overall survival in chronic lymphocytic leukemia. A multivariate analysis from the randomized GCLLSG CLL8 trial. J Clin Oncol 30:980–988
31. Fischer K, Cramer P, Busch R et al (2012) Bendamustine in combination with rituximab for previously untreated patients with chronic lymphocytic leukemia: a multicenter phase II trial of the German Chronic Lymphocytic Leukemia Study Group. J Clin Oncol 30:3209–3216
32. Pettitt AR, Jackson R, Carruthers S et al (2012) Alemtuzumab in combination with methylprednisolone is a highly effective induction regimen for patients with chronic lymphocytic

leukemia and deletion of TP53: final results of the national cancer research institute CLL206 trial. J Clin Oncol 30:1647–1655

33. Bouvet E, Borel C, Oberic L et al (2013) Impact of dose intensity on outcome of fludarabine, cyclophosphamide, and rituximab regimen given in the first-line therapy for chronic lymphocytic leukemia. Haematologica 98:65–70
34. Santacruz R, Villamor N, Aymerich M et al (2014) The prognostic impact of minimal residual disease in patients with chronic lymphocytic leukemia requiring first-line therapy. Haematologica 99:873–880
35. Strati P, Keating MJ, O'Brien SM et al (2014) Eradication of bone marrow minimal residual disease may prompt early treatment discontinuation in CLL. Blood 123:3727–3732
36. Ritgen M, Langerak A, Goede V et al (2016) Quantitative MRD is prognostic for progression free & overall survival in elderly patients receiving chlorambucil alone or with obinutuzumab/rituximab: a prospective analysis of the GCLLSG CLL11 study. Haematologica 101:149–150 [abstract]
37. Howard DR, Munir T, McParland L et al (2017) Results of the randomized phase IIB ARCTIC trial of low-dose rituximab in previously untreated CLL. Leukemia 31:2416–2425
38. Kwok M, Rawstron AC, Varghese A et al (2016) Minimal residual disease is an independent predictor for 10-year survival in CLL. Blood 128:2770–2773
39. Munir T, Howard DR, McParland L et al (2017) Results of the randomized phase IIB ADMIRE trial of FCR with or without mitoxantrone in previously untreated CLL. Leukemia 31:2085–2093
40. Feugier P, Aurran T, Mahe B et al (2018) Long-term follow-up of the CLL2007FMP trial evaluating fludarabine and cyclophosphamide in combination with either rituximab or alemtuzumab in previously untreated patients with chronic lymphocytic leukemia. Haematologica 103:e304–e306
41. Seymour JF, Kipps TJ, Eichhorst B et al (2018) Venetoclax-rituximab in relapsed or refractory chronic lymphocytic leukemia. N Engl J Med 378:1107–1120
42. Hallek M, Cheson BD, Catovsky D et al (2018) Guidelines for diagnosis, indications for treatment, response assessment and supportive management of chronic lymphocytic leukemia. Blood 131:2745–2760
43. Rawstron AC, Bottcher S, Letestu R et al (2013) Improving efficiency and sensitivity: European Research Initiative in CLL (ERIC) update on the international harmonised approach for flow cytometric residual disease monitoring in CLL. Leukemia 27:142–149
44. Logan AC, Zhang B, Narasimhan B et al (2013) Minimal residual disease quantification using consensus primers and high-throughput IGH sequencing predicts post-transplant relapse in chronic lymphocytic leukemia. Leukemia 27:1659–1665
45. Roussel M, Lauwers-Cances V, Robillard N et al (2014) Front-line transplantation program with lenalidomide, bortezomib, and dexamethasone combination as induction and consolidation followed by lenalidomide maintenance in patients with multiple myeloma: a phase II study by the Intergroupe Francophone du Myelome. J Clin Oncol 32:2712–2717
46. San Miguel JF, Schlag R, Khuageva NK et al (2008) Bortezomib plus melphalan and prednisone for initial treatment of multiple myeloma. N Engl J Med 359:906–917
47. Stewart AK, Rajkumar SV, Dimopoulos MA et al (2015) Carfilzomib, lenalidomide, and dexamethasone for relapsed multiple myeloma. N Engl J Med 372:142–152
48. Moreau P, Masszi T, Grzasko N et al (2016) Oral ixazomib, lenalidomide, and dexamethasone for multiple myeloma. N Engl J Med 374:1621–1634
49. Dimopoulos MA, Oriol A, Nahi H et al (2016) Daratumumab, lenalidomide, and dexamethasone for multiple myeloma. N Engl J Med 375:1319–1331
50. Mateos MV, Dimopoulos MA, Cavo M et al (2018) Daratumumab plus bortezomib, melphalan, and prednisone for untreated myeloma. N Engl J Med 378:518–528
51. Palumbo A, Chanan-Khan A, Weisel K et al (2016) Daratumumab, bortezomib, and dexamethasone for multiple myeloma. N Engl J Med 375:754–766
52. Lonial S, Dimopoulos M, Palumbo A et al (2015) Elotuzumab therapy for relapsed or refractory multiple myeloma. N Engl J Med 373:621–631
53. Chari A, Suvannasankha A, Fay JW et al (2017) Daratumumab plus pomalidomide and dexamethasone in relapsed and/or refractory multiple myeloma. Blood 130:974–981
54. San Miguel J, Weisel K, Moreau P et al (2013) Pomalidomide plus low-dose dexamethasone versus high-dose dexamethasone alone for patients with relapsed and refractory multiple myeloma (MM-003): a randomised, open-

label, phase 3 trial. Lancet Oncol 14:1055–1066

55. Paiva B, Vidriales MB, Cervero J et al (2008) Multiparameter flow cytometric remission is the most relevant prognostic factor for multiple myeloma patients who undergo autologous stem cell transplantation. Blood 112:4017–4023
56. Paiva B, Gutierrez NC, Rosinol L et al (2012) High-risk cytogenetics and persistent minimal residual disease by multiparameter flow cytometry predict unsustained complete response after autologous stem cell transplantation in multiple myeloma. Blood 119:687–691
57. Paiva B, Martinez-Lopez J, Vidriales MB et al (2011) Comparison of immunofixation, serum free light chain, and immunophenotyping for response evaluation and prognostication in multiple myeloma. J Clin Oncol 29:1627–1633
58. Martinez-Lopez J, Lahuerta JJ, Pepin F et al (2014) Prognostic value of deep sequencing method for minimal residual disease detection in multiple myeloma. Blood 123:3073–3079
59. Rawstron AC, Child JA, de Tute RM et al (2013) Minimal residual disease assessed by multiparameter flow cytometry in multiple myeloma: impact on outcome in the Medical Research Council Myeloma IX Study. J Clin Oncol 31:2540–2547
60. Lahuerta JJ, Paiva B, Vidriales MB et al (2017) Depth of response in multiple myeloma: a pooled analysis of three PETHEMA/GEM clinical trials. J Clin Oncol 35:2900–2910
61. Korde N, Roschewski M, Zingone A et al (2015) Treatment with carfilzomib-lenalidomide-dexamethasone with lenalidomide extension in patients with smoldering or newly diagnosed multiple myeloma. JAMA Oncol 1:746–754
62. Ferrero S, Ladetto M, Drandi D et al (2015) Long-term results of the GIMEMA VEL-03-096 trial in MM patients receiving VTD consolidation after ASCT: MRD kinetics' impact on survival. Leukemia 29:689–695
63. Ladetto M, Pagliano G, Ferrero S et al (2010) Major tumor shrinking and persistent molecular remissions after consolidation with bortezomib, thalidomide, and dexamethasone in patients with autografted myeloma. J Clin Oncol 28:2077–2084
64. Ladetto M, Bruggemann M, Monitillo L et al (2014) Next-generation sequencing and real-time quantitative PCR for minimal residual disease detection in B-cell disorders. Leukemia 28:1299–1307
65. Martinez-Lopez J, Paiva B, Lopez-Anglada L et al (2015) Critical analysis of the stringent complete response in multiple myeloma: contribution of sFLC and bone marrow clonality. Blood 126:858–862
66. Arroz M, Came N, Lin P et al (2016) Consensus guidelines on plasma cell myeloma minimal residual disease analysis and reporting. Cytometry B Clin Cytom 90:31–39
67. Kalina T, Flores-Montero J, van der Velden VH et al (2012) EuroFlow standardization of flow cytometer instrument settings and immunophenotyping protocols. Leukemia 26:1986–2010
68. van Dongen JJ, Lhermitte L, Bottcher S et al (2012) EuroFlow antibody panels for standardized n-dimensional flow cytometric immunophenotyping of normal, reactive and malignant leukocytes. Leukemia 26:1908–1975
69. Kumar S, Paiva B, Anderson KC et al (2016) International Myeloma Working Group consensus criteria for response and minimal residual disease assessment in multiple myeloma. Lancet Oncol 17:e328–e346
70. Pott C, Hoster E, Delfau-Larue MH et al (2010) Molecular remission is an independent predictor of clinical outcome in patients with mantle cell lymphoma after combined immunochemotherapy: a European MCL intergroup study. Blood 115:3215–3223
71. Pott C, Schrader C, Gesk S et al (2006) Quantitative assessment of molecular remission after high-dose therapy with autologous stem cell transplantation predicts long-term remission in mantle cell lymphoma. Blood 107:2271–2278
72. Kolstad A, Laurell A, Jerkeman M et al (2014) Nordic MCL3 study: 90Y-ibritumomab-tiuxetan added to BEAM/C in non-CR patients before transplant in mantle cell lymphoma. Blood 123:2953–2959
73. Kolstad A, Pedersen LB, Eskelund CW et al (2017) Molecular monitoring after autologous stem cell transplantation and preemptive rituximab treatment of molecular relapse; results from the nordic mantle cell lymphoma studies (MCL2 and MCL3) with median follow-up of 8.5 years. Biol Blood Marrow Transplant 23:428–435
74. Armand P, Redd R, Bsat J et al (2016) A phase 2 study of Rituximab-Bendamustine and Rituximab-Cytarabine for transplant-eligible

patients with mantle cell lymphoma. Br J Haematol 173:89–95

75. Cheminant M, Derrieux C, Touzart A et al (2016) Minimal residual disease monitoring by 8-color flow cytometry in mantle cell lymphoma: an EU-MCL and LYSA study. Haematologica 101:336–345
76. Tam CS, Anderson MA, Pott C et al (2018) Ibrutinib plus venetoclax for the treatment of mantle-cell lymphoma. N Engl J Med 378:1211–1223
77. Ladetto M, Lobetti-Bodoni C, Mantoan B et al (2013) Persistence of minimal residual disease in bone marrow predicts outcome in follicular lymphomas treated with a rituximab-intensive program. Blood 122:3759–3766
78. Ladetto M, De Marco F, Benedetti F et al (2008) Prospective, multicenter randomized GITMO/IIL trial comparing intensive (R-HDS) versus conventional (CHOP-R) chemoimmunotherapy in high-risk follicular lymphoma at diagnosis: the superior disease control of R-HDS does not translate into an overall survival advantage. Blood 111:4004–4013
79. Galimberti S, Luminari S, Ciabatti E et al (2014) Minimal residual disease after conventional treatment significantly impacts on progression-free survival of patients with follicular lymphoma: the FIL FOLL05 trial. Clin Cancer Res 20:6398–6405
80. Zohren F, Bruns I, Pechtel S et al (2015) Prognostic value of circulating Bcl-2/IgH levels in patients with follicular lymphoma receiving first-line immunochemotherapy. Blood 126:1407–1414
81. van Oers MH, Tonnissen E, Van Glabbeke M et al (2010) BCL-2/IgH polymerase chain reaction status at the end of induction treatment is not predictive for progression-free survival in relapsed/resistant follicular lymphoma: results of a prospective randomized EORTC 20981 phase III intergroup study. J Clin Oncol 28:2246–2252
82. Kalina T, Flores-Montero J, Lecrevisse Q et al (2015) Quality assessment program for EuroFlow protocols: summary results of four-year (2010–2013) quality assurance rounds. Cytometry A 87:145–156
83. Böttcher S (2019) Minimal residual disease quantification in chronic lymphocytic leukemia: clinical significance and flow cytometric methods. Methods Mol Biol 1881:211–238. https://doi.org/10.1007/978-1-4939-8876-1_17
84. Böttcher S, van der Velden VHJ, Villamor N, Ritgen M, Flores-Montero J, Murua Escobar H, Kalina T, Brüggemann M, Grigore G, Martin-Ayuso M, Lecrevisse Q, Pedreira CE, van Dongen JJM, Orfao A (2017) Lot-to-lot stability of antibody reagents for flow cytometry. J Immunol Methods. pii: S0022-1759(17)30075-3. https://doi.org/10.1016/j.jim.2017.03.018. [Epub ahead of print]
85. Bomberger C, Singh-Jairam M, Rodey G et al (1998) Lymphoid reconstitution after autologous PBSC transplantation with FACS-sorted CD34+ hematopoietic progenitors. Blood 91:2588–2600
86. Flores-Montero J, de Tute R, Paiva B et al (2016) Immunophenotype of normal vs. myeloma plasma cells: toward antibody panel specifications for MRD detection in multiple myeloma. Cytometry B Clin Cytom 90:61–72
87. Bottcher S, Ritgen M, Kneba M (2013) Flow cytometric MRD detection in selected mature B-cell malignancies. Methods Mol Biol 971:149–174. https://doi.org/10.1007/978-1-62703-269-8_9

Chapter 9

MRD Detection in B-Cell Non-Hodgkin Lymphomas Using Ig Gene Rearrangements and Chromosomal Translocations as Targets for Real-Time Quantitative PCR

Christiane Pott, Monika Brüggemann, Matthias Ritgen, Vincent H. J. van der Velden, Jacques J. M. van Dongen, and Michael Kneba

Abstract

Minimal residual disease (MRD) diagnostics is of high clinical relevance in patients with indolent B-cell non-Hodgkin lymphomas (B-NHL) and serves as a surrogate parameter to evaluate treatment effectiveness and long-term prognosis. MRD diagnostics performed by real-time quantitative PCR (RQ-PCR) is still the gold standard and currently the most sensitive and the most broadly applied method in follicular lymphoma (FL) and mantle cell lymphoma (MCL). Alternatively, droplet digital PCR (ddPCR) can be used for MRD monitoring in multiple myeloma, mantle cell lymphoma, and follicular lymphoma with comparable sensitivity, accuracy, and reproducibility.

The most broadly applicable MRD target in B-NHL is the junctional regions of the rearranged immunoglobulin heavy chain gene (IGHV). Chromosomal translocations like the t(14;18) translocation in FL and t(11;14) translocation in MCL can be used as MRD target in selected lymphoma subtypes. In patients with B-cell chronic lymphocytic leukemia, both flow-cytometry and RQ-PCR are equally suited for MRD assessment as long as a sensitivity of 10^{-4} shall be achieved.

MRD diagnostics targeting the IGHV gene is complex and requires extensive knowledge and experience because the junctional regions of each lymphoma have to be identified before the patient-specific RQ-PCR assays can be designed for MRD monitoring. In addition, the presence and load of somatic hypermutation (SHM) within the rearranged IG heavy variable (IGHV) gene occurring as during B-cell development of germinal center and post-germinal center lymphomas may hamper appropriate primer binding leading to false-negative results. The translocations mentioned above have the advantage that consensus forward primers and probes, both placed in the breakpoint regions of chromosome 18 in FL and chromosome 11 in MCL, can be used in combination with a reverse primer placed in the IGH joining region of chromosome 14. RQ-PCR-based methods can reach a good sensitivity ($\leq 10^{-4}$). This chapter provides all relevant background information and technical aspects for the complete laboratory process from detection of the clonal IGHV gene rearrangement and the chromosomal translocations at diagnosis to the actual MRD measurements in clinical follow-up samples of B-NHL. However, it should be noted that MRD diagnostics for clinical treatment protocols has to be accompanied by regular international quality control rounds to ensure the reproducibility and reliability of the MRD results.

Ralf Küppers (ed.), *Lymphoma: Methods and Protocols*, Methods in Molecular Biology, vol. 1956,
https://doi.org/10.1007/978-1-4939-9151-8_9,

Key words Minimal residual disease (MRD), B-cell non-Hodgkin lymphomas (B-NHL), Chronic lymphocytic leukemia (CLL), Immunoglobulin heavy chain gene (IGH), t(14;18) translocation, t(11;14) translocation, Follicular lymphoma (FL), Mantle cell lymphoma (MCL), Real-time quantitative PCR (RQ-PCR), ddPCR, Quantitative range, Sensitivity

1 Introduction

Several studies have shown that detection of minimal residual disease (MRD) in indolent B-cell non-Hodgkin lymphomas (B-NHL) is clinically relevant [1–12]. Based on these results, MRD diagnostics is currently implemented in clinical treatment protocols, and future studies plan to stratify patients according to MRD levels in bone marrow (BM) or peripheral blood (PB) samples obtained after induction therapy. In most of these MRD-based stratification studies, MRD diagnostics is performed by real-time quantitative PCR (RQ-PCR) analysis of immunoglobulin heavy chain gene rearrangements (IGHV) or chromosomal translocations in B-NHL subtypes (e.g., in follicular lymphoma (FL) and mantle cell lymphoma (MCL)) [10, 11, 13–17]. Also clonal IG light chain rearrangements are principally applicable as MRD targets [18].

Droplet digital PCR (ddPCR) is a method for performing digital PCR and is based on water-oil emulsion droplet technology. ddPCR has several theoretical advantages compared with qPCR, most notably allowing for absolute quantification of target DNA molecules and avoiding the need for a reference standard curve. The utility of ddPCR as an MRD monitoring tool has been demonstrated by a comparative analysis in different types of lymphoma including follicular and mantle cell lymphomas [19, 20].

The most broadly applicable MRD target in B-NHL is the junctional regions of the rearranged immunoglobulin heavy chain gene (IGHV) that serves as reliable MRD target in preferably low or unmutated disease entities like MCL. After rearrangement of the V, (D), and J gene segments of the Ig gene complexes in early B-cell differentiation, each lymphocyte obtains a specific combination of V-(D-)J segments that codes for the variable domains of Ig molecules. The random insertion and deletion of nucleotides at the junction sites of V, (D), and J gene segments make the junctional regions of Ig genes into “fingerprint-like” sequences, which are most probably different in each lymphocyte and thus also in each lymphoid malignancy. Therefore, junctional regions of malignant lymphoma cells can be used as tumor-specific targets for MRD-PCR analysis. Such targets can be identified by PCR analysis at initial diagnosis in ~80% of B-NHL [21, 22]. Subsequently, the precise nucleotide sequence of the junctional regions can be determined. This sequence information allows the design of junctional region-specific oligonucleotides (either probes or primers), which

like in acute lymphoblastic leukemias (ALL) can be used for sensitive detection of low frequencies of malignant cells, down to one malignant cell in 10^4 to 10^5 normal cells (10^{-4} to 10^{-5}) [23].

However, the presence and load of somatic hypermutation (SHM) within the rearranged IG heavy variable (IGHV) gene occurring mostly in germinal center and post-germinal center lymphomas may hamper appropriate primer binding leading to false-negative PCR results [21, 24].

Alternatively, in certain histologic subtypes of B-NHL recurrent cytogenetic aberrations as the chromosomal translocations t(14;18) in follicular lymphoma (FL) and t(11;14) in MCL can serve as MRD markers. In both translocations the *IGH* locus at 14q32.3 is involved, and the *BCL2* gene from 18q21 [25–27] or the *BCL1* region from 11q13 is translocated into the *IGH* locus [28]. Both translocations involve the process of VDJ recombination, and one of the six germline JH gene segments is closely opposed to the *BCL2* gene or the *BCL1* region. As a consequence of the translocation process during VDJ recombination, random insertion of nucleotides at the junction sites of *BCL2* and *BCL1* and (D) J gene segments takes place and makes the junctional regions of the translocations into "fingerprint-like" sequences.

However, usage of chromosomal translocations as MRD markers is not only restricted to selected subtypes of B-NHL but is also detectable only in a subset of patients, about 60% of FL and 30% of MCL [29–31]. This is mainly due to scattered breakpoint regions on the derivative chromosomal regions involved in the translocations of chromosomes 18 or 11 that are not accessible by PCR in 100% of translocations [30, 32]. A further considerable confounder of using the chromosomal translocations t(14;18) as MRD marker in follicular lymphoma is the occurrence of nonmalignant t(14;18) positive cells that are detectable by consensus PCR, especially when nested PCR is used [33].

In this chapter, we describe RQ-PCR-based MRD detection using IGH gene rearrangements and the chromosomal translocations t(14;18) and t(11;14) as MRD-PCR targets in B-NHL patients. Several of the methods for MRD detection using Ig gene rearrangements have already been published by van der Velden and van Dongen [34].

Because there is a relevant methodical overlap of MRD analysis in acute leukemias and B-NHL, several of the methods described here, including clonality assessment and sample processing, have been adopted or modified for the application in B-NHL. The approaches described here identify MRD-PCR targets in >80% of B-NHL patients and allow MRD monitoring in the majority of B-NHL patients.

2 Materials

2.1 Bone Marrow and Peripheral Blood Sample Processing at Diagnosis

1. Biocoll separating solution (Biochrom AG). Store in the dark at room temperature.
2. Cell storage medium: 10% dimethyl sulfoxide (DMSO), 20% fetal calf serum (FCS), 70% RPMI-1640 medium.

2.2 Detection of Clonal IG Targets or Chromosomal Translocations at Diagnosis

2.2.1 Agarose Gel Electrophoresis and Gene Scan Analysis

1. 1× TBE buffer: 90 mM Tris–HCl, 90 mM H_3BO_3, and 2.5 mM EDTA, in Milli-Q water.
2. Bromophenol Blue (BPB) gel loading buffer: 30% (v/v) glycerol, 0.25% (w/v) bromophenol blue (electrophoresis purity reagent), 0.25% (w/v) Xylene Cyanol FF (electrophoresis purity reagent), in Milli-Q water.
3. GelRed 1%.
4. $MgCl_2$ (25 mM).
5. Deoxynucleotide (dNTP) mix (10 mM each).
6. DNA mass ladder.
7. EagleTaq DNA polymerase (5 U/μL) (Roche).
8. PCR buffer w/o $MgCl_2$ (Roche).
9. Agarose gel (2%).
10. Forward and reverse primers for IGH, t(14;18) or t(11;14) translocation PCR, reverse primers labeled with 5′ FAM 100 μM each (Tables 1 and 2).
11. Electrophoresis units.
12. 3500 Genetic Analyzer with POP-7 polymer (Applied Biosystems).
13. HIDI-Formamide (ABI P/N 4311320) for 3500 Genetic Analyzer.
14. ROX500 size standard according to manufacturer's instructions (ABI); from our experience, a 1:2 dilution is sufficient.
15. Cathode and anode buffer for the respective Genetic Analyzer.
16. DNA isolated from PB mononuclear cells (MNC) obtained from 5–10 healthy individuals (100 ng/μL).
17. DNA from cell lines known to be positive for IGH gene rearrangements or t(14;18) or t(11;14) translocation (100 ng/μL).

2.2.2 PCR-Heteroduplex Analysis

1. 1× TBE buffer: 90 mM Tris–HCl, 90 mM H_3BO_3, 2.5 mM EDTA, in Milli-Q water.
2. Bromophenol blue (BPB) loading buffer: 30% (v/v) glycerol, 0.25% (w/v) Bromophenol blue (electrophoresis purity

Table 1
Primer combinations for detection of IGH rearrangement, t(14;18) and t(11;14) translocation

Target	Forward primer	Reverse primer	Positive control	Product size (bp)	Protocol[a]	Reference
IGH	VH1/7	JH	Nalm6	300–400	A	[22]
	VH2	JH	Patient DNA	300–400	A	[22]
	VH3	JH	REH	300–400	A	[22]
	VH4/6	JH	ROS16	300–400	A	[22]
	VH5	JH	VH5-TL	300–400	A	[22]
t(14;18)						
MBR	MBR	JH	DOHH2	150–450	B	
3′MBR	3′MBR	JH	K231	130–250	B	[22]
mcr	mcr	JH	OZ	150–300	B	[22]
5′mcr	5′mcr	JH	SC1	260–490	B	[22]
t(11;14)						
	BCL1 MTC forward primer G	JH	JVM-2	100–300	C	[30]
	BCL1 MTC forward primer B	JH	JVM-2	100–300	C	[30]

[a] *See* Table 4 for detailed information about PCR mixture per protocol

Table 2
Primer sequences for t(14;18), t(11;14), and IGH PCR

Locus	Primer	Sequence
IGH	FR1-VH1-2	5′-GGCCTCAGTGAAGGTCTCCTGCAAG-3′
	FR1-VH2nw	5′-GTCTGGTCCTACGCTGGTGAAACCC-3′
	FR1-VH3-7	5′-CTGGGGGGTCCCTGAGACTCTCCTG-3′
	FR1-VH4-4	5′-CTTCGGAGACCCTGTCCCTCACCTG-3′
	FR1-VH5-51	5′-CGGGGAGTCTCTGAAGATCTCCTGT-3′
	FR1-VH6	5′-TCGCAGACCCTCTCACTCACCTGTG-3′
	FR2-VH1-2	5′-CTGGGTGCGACAGGCCCCTGGACAA-3′
	FR2-VH2-5	5′-TGGATCCGTCAGCCCCCAGGAAGG-3′
	FR2-VH3-7	5′-GGTCCGCCAGGCTCCAGGGAA-3′
	FR2-VH4-4	5′-TGGATCCGCCAGCCCCCAGGGAAGG-3′
	FR2-VH5-51	5′-GGGTGCGCCACATGCCCGGGAAAGG-3′
	FR2-VH6	5′-TGGATCAGGCAGTCCCCATCGAGAG-3′
	FR3-VH1-2	5′-TGGAGCTGAGCAGCCTGAGATCTGA-3′
	FR3-VH2-5	5′-CAATGACCAAACATGGACCCTGTGGA-3′
	FR3-VH3	5′-TCTGCAAATGAACAGCCTGAGAGCC-3′
	FR3-VH4-4	5′-GAGCTCTGTGACCGCCGCGGACACG-3′
	FR3-VH5-51	5′-CAGCACCGCCTACCTGCAGTGGAGC-3′
	FR3-VH6	5′-GTTCTCCCTGCAGCTGAACTCTGTG-3′
	FR3-VH7	5′-CAGCACGGCATATCTGCAGATCAG-3′
Chr 18 *BCL2*	5′MBR	5′-GCAATTCCGCATTTAATTCATGGTATTCAGGAT-3′
	3′MBR1 681	5′-GCACCTGCTGGATACAACACTG-3′
	3′MBR2 1413	5′-GGTGACAGAGCAAAACATGAACA-3′
	3′MBR3 1745	5′-GTAATGACTGGGGAGCAAATCTT-3′
	3′MBR4 2676	5′-ACTGGTTGGCGTGGTTTAGAGA-3′
	5′mcr	5′-CCTTCTGAAAGAAACGAAAGCA-3′
	mcr 1	5′-TAGAGCAAGCGCCCAATAAATA-3′
	mcr 2	5′-TGAATGCCATCTCAAATCCAA-3′
Chr 11 *BCL1*	BCL1—G	5′-GGAGCATAATTGCTGCACTGC-3′
	Bcl-1—B	5′-TTCGGTTAGACTGTGATTAGC-3′
IGH	JH consensus[a]	5′-CTTACCTGAGGAGACGGTGACC-3′

[a]For gene scanning this primer is labeled with 5′FAM

reagent), 0.25% (w/v) Xylene Cyanol FF (electrophoresis purity reagent), in Milli-Q water.

3. GelRed 1%.
4. $MgCl_2$ (25 mM).
5. Deoxynucleotide (dNTP) mix (10 mM each).
6. Bovine serum albumin (BSA; 20 mg/mL): 20% (w/v) BSA (Fraction V) in Milli-Q water.
7. Orange G: 20% Ficoll, 10 mM Tris–HCl pH 7.6, 1 mg/mL Orange G, in Milli-Q water.
8. DNA ladders.
9. EagleTaq DNA Polymerase (5 U/μL) (Roche).

10. PCR buffer w/o $MgCl_2$ (Roche).
11. Agarose gel (2%).
12. Forward and reverse primers for the various Ig gene rearrangements or translocations.
13. Electrophoresis units.
14. DNA isolated from PB MNC obtained from 5–10 healthy individuals (100 ng/μL).
15. DNA from cell lines or patients known to be positive for certain Ig gene rearrangements or translocations (100 ng/μL).

2.2.3 Sequencing of Clonal Rearrangements

1. MinElute 96 UF Plates [11] (Qiagen).
2. Vacuum manifold (Millipore).
3. Sephadex G-50, 6% (w/v) fine DNA grade in Milli-Q water on a Millipore MultiScreen HV plate.
4. 5× buffer with EDTA (Applied Biosystems).
5. Elution buffer: 0.5 M ammonium acetate, 10 mM magnesium acetate, 1 mM EDTA, 0.1% SDS, in Milli-Q water.
6. Big Dye® Terminator cycle sequencing ready reaction kit version 3.1 (Applied Biosystems).
7. Forward and reverse primers for the various Ig gene rearrangements or chromosomal translocations.
8. 3500 Genetic Analyzer with POP-7 polymer (Applied Biosystems).

2.3 RQ-PCR Sensitivity Testing

1. 1× TE buffer: 10 mM Tris–HCl pH 7.6, 1 mM EDTA, in Milli-Q water.
2. Spectrophotometer (NanoDrop, Wilmington, USA).

2.3.1 Design of Allele-Specific Oligonucleotide Primers

3. Oligo Primer Analysis Software v. 7 (Molecular Biology Insights, Inc., Cascade, USA) or any other primer design software.

2.3.2 RQ-PCR Analysis of Dilutions of Diagnostic Sample

1. BSA (5%).
2. Universal Master Mix (2×) (Applied Biosystems).
3. Forward primer (20 pmol/μL).
4. Reverse primer (20 pmol/μL).
5. TaqMan probe (20 pmol/μL).
6. DNA isolated from PB MNC obtained from 5 to 10 healthy individuals (60 and 100 ng/μL).

2.4 ddPCR for MRD Quantification

1. 2× ddPCR supermix for probes (no dUTP) (BIORAD).
2. ddPCR Droplet Generator Oil for Probes (BIORAD).
3. ddPCR Droplet Reader Oil (BIORAD).

4. Forward primer (20 pmol/μL).
5. Reverse primer (20 pmol/μL).
6. TaqMan probe (20 pmol/μL).
7. DG8 cartridges and gaskets for QX200 Droplet Generator (BIORAD).
8. QX200 Droplet Generator (BIORAD).
9. Twin.tec PCR Plate 96 (Eppendorf).
10. Pierceable Foil Heat Seal (BIORAD).
11. QX200 Droplet Reader (BIORAD).
12. PX1 PCR Plate Sealer (BIORAD).

2.5 Bone Marrow and Peripheral Blood Sample Processing During Follow-Up

See Subheading 2.1 for materials needed.

2.6 MRD Analysis of Follow-Up Samples

See Subheadings 2.3.2 and 2.5 for materials needed.

3 Methods

3.1 Bone Marrow and Peripheral Blood Sample Processing at Diagnosis

1. Perform Ficoll density centrifugation of the BM or PB sample obtained at diagnosis to obtain MNC.
2. Assess the percentage of FL, MCL, and CLL cells in the MNC fraction by ddPCR or flow cytometric immunophenotyping because this information is needed for correction of the tumor load in the standard curve (*see* Subheading 3.3) (*see* **Note 1**). For reasons of standardization and prediction of DNA recovery, use a fixed number of 10×10^6 MNC for DNA extraction (*see* **Note 2**).
3. Determine the concentration of the DNA stock solution by analyzing the optical density at 260 and 280 nm in duplicate on the NanoDrop (*see* **Note 3**).
4. Prepare 150 μL of a DNA working solution of 100 ng/μL in TE buffer; this working dilution will be used for the PCR analyses and RQ-PCR analyses (see below).
5. Check the DNA concentration (NanoDrop).
6. Store the stock solution and working dilution at 4 °C (or − 20 °C) for short periods of time (<3 months) or at −20 °C for long-term storage.
7. Store remaining MNC at diagnosis in cell storage medium for later studies; store in liquid nitrogen or at −80 °C.

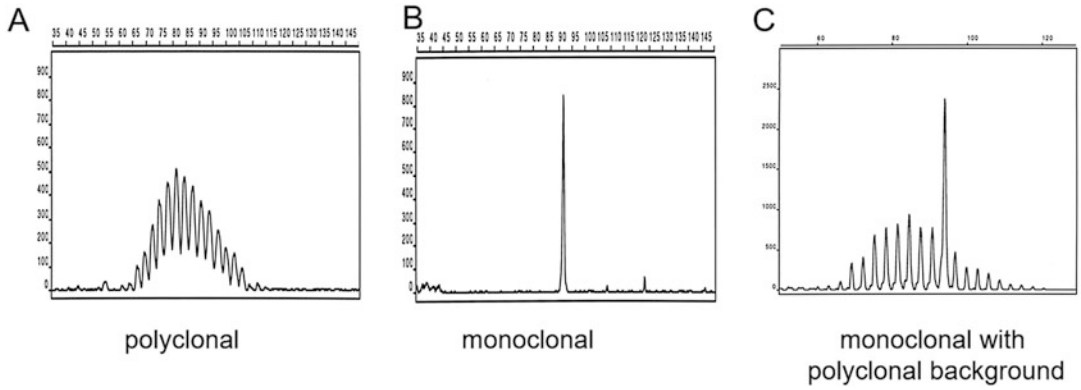


Fig. 1 GeneScan analysis of IG gene rearrangements. (**a**) Mononuclear cells from healthy donors show a polyclonal pattern, (**b**) monoclonal peak in case of a patient with CLL, (**c**) monoclonal peak with polyclonal B-cell background

3.2 Detection of Clonal IG Gene Rearrangements and Determination of Chromosomal Translocations at Diagnosis

For MRD marker screening, IGH consensus PCR is performed in all B-NHL subtypes including CLL. According to histologic subtype, additionally t(14;18) PCR in FL and t(11;14) PCR in MCL are performed. To determine clonality of IGH gene rearrangements, PCR products have to be further processed by either PCR heteroduplex analysis or gene scanning. This is especially important in malignant lymphomas, where besides the clonal lymphoma cell population polyclonal B cells are mostly present and are amplified by IGH PCR as well. By applying gene scan analysis or PCR heteroduplex analysis with appropriate primer sets, clonal IGH gene rearrangements can be visualized and distinguished from a polyclonal B-cell background. In gene scanning single-stranded PCR products are separated in a high-resolution polymer according to their length only. The PCR primers need to be labeled with a fluorochrome to allow detection of the PCR products with automated sequencing equipment. Figure 1 demonstrates gene scan results of IGH PCR in different settings: (a) a polyclonal genescan pattern when only polyclonal B cells are present, (b) a monoclonal gene scan pattern in case of a CLL patient, and (c) a monoclonal IGH rearrangement with a polyclonal B-cell background. The latter genescan pattern is mostly found in nodal FL where besides the lymphoma cell population also polyclonal B cells are present in BM or PB.

In PCR heteroduplex analysis, double-stranded PCR products are denatured at high temperatures followed by rapid random renaturation at low temperatures, preferably +4 °C for 1 h [35]. This results in case of clonal PCR products in homoduplexes with identical rapid migration speed compared to polyclonal PCR products forming many heteroduplexes with different migration

speed when electrophoresis is performed on a 6% polyacrylamide gel.

For the detection of the chromosomal translocations, visualization of the PCR product in an agarose gel is sufficient to detect a t(14;18) or t(11;14) translocation. Here, gene scanning is used to precisely distinguish PCR products from different patients by exact size determination of respective PCR products.

3.2.1 Gene Scan Analysis

1. Perform respective PCR reaction with 100–500 ng diagnostic DNA (*see* Table 1). The reaction mixes are shown in Table 1, the primer sequences are shown in Table 2, and reaction conditions are shown in Tables 3 and 4. Use Milli-Q water as a negative control, normal MNC DNA as a polyclonal control, and an appropriate positive control (*see* Table 1).
2. Analyze PCR products in a 2% analytical agarose gel containing GelRed.
3. For each of the PCR products, positive and negative controls, and the DNA ladder, mix 7 μL of the product with 2 μL gel loading buffer, and load these samples on the agarose gel; run for 30 min at 130 V. PCR products are visualized under ultraviolet (UV) light. The positive controls should give a clear band of the expected size, whereas the negative controls should be negative. If this is not the case, repeat the PCR assays with insufficient results.

Table 3
Composition of PCR mixture per protocol

	PCR protocol		
Reaction mix per 50 μL reaction	**A**	**B**	**C**
PCR Puffer w/o $MgCl_2$	5	5	5
$MgCl_2$ (25 mM)	3	4	4
Milli-Q water	Ad 45 μL	Ad 45 μL	Ad 45 μL[a]
dNTP (10 mM)	1	1	1
BSA (20 mg/mL)	1	1	1
Forward primer (diluted to 100 pmol/μL each)	0.1	0.1	0.1
Reverse primer (diluted to 100 pmol/μL)	0.1	0.1	0.1
EagleTaq (5 U/μL)	0.2	0.2	0.4
DNA (100 ng/μL)	5	5	5[a]

[a]For the second PCR in t(11;14) seminested PCR, 1.5 μL PCR product from the first PCR round is used so 48.4 μL Milli-Q water and no DNA has to be added

Table 4
PCR reaction conditions per protocol

	PCR protocol					
Cycler protocols	**A**		**B[a]**		**C[b]**	
Initial denaturation	94 °C	10 min	94 °C	10 min	94 °C	10 min
Denaturation	94 °C	60 s	94 °C	30 s	94 °C	60 s
Annealing	60 °C	60 s	66 °C −0.5 °C ramp	30 s	60 °C	60 s
Extension	72 °C	30 s	–	–	–	–
Final extension	72 °C	30 min	–	–	72 °C	5 min
Number of cycles	35		23		35	
Initial denaturation			–	–	94 °C	10 min
Denaturation			94 °C	30 s	94 °C	60 s
Annealing			55 °C	30 s	60 °C	60 s
Extension			72 °C	30 s	–	–
Final extension			72 °C	10 min	72 °C	10 min
Number of cycles			25		35	

[a]The touchdown protocol does not require another pipetting step, and the second cycler protocol is run directly after the first
[b]The seminested PCR protocol requires a second pipetting step where 1.5 μL PCR product of PCR 1 is used as template

4. For gene scanning select samples and put 1 μL PCR product with 0.5 μL ROX standard and 12 μL formamide (HIDI-formamide if a 3500 genetic analyzer is used) in a 96-well plate. Centrifuge the plate shortly, and heat the covered plate (rubber seal) at 92 °C for 2 min in a PCR machine.
5. Place samples in the respective genetic analyzer and start gene scan run.
6. Analyze data according to manufacturer's instructions with the gene mapper program, and identify clonal rearrangements.
7. Interpret the results (*see* Fig. 1). One peak of expected size indicates one clonal rearrangement; one dominant peak with an additional polyclonal background pattern indicates the presence of polyclonal B cells, and a polyclonal B-cell pattern indicates polyclonal rearrangements only.

3.2.2 PCR-Heteroduplex Analysis

1. Perform PCR reaction with 100–500 ng diagnostic DNA (*see* Table 1). The reaction mixes and primers are shown in Tables 1 and 2. For PCR heteroduplex analysis, use unlabeled primers only. Use Milli-Q water as a negative control, normal MNC

DNA as a polyclonal control, and an appropriate positive control (*see* Table 1).

2. For each of the positive and negative controls and the DNA ladder, mix 7 μL of the control with 2 μL gel loading buffer, and load these samples on the agarose gel; run for 30 min at 15 V/cm. The positive controls should give a clear band of the expected size, whereas the negative controls should be negative. If this is not the case, repeat the PCR assays with insufficient results.
3. Heat the PCR products of the patient samples tested for 5 min at 94 °C in a PCR machine.
4. Cool the PCR products for 60 min at 4 °C in a PCR machine.
5. Use criterion precast gels (*see* **Note 4**); remove the white strip at the bottom of the cassette, and place the criterion precast gel in the electrophoresis cell.
6. Fill the buffer chamber of the gel cassette with 60 mL TBE buffer (1×), and fill both sides of the electrophoresis cell with 1× TBE buffer till the indicated bar (approximately 800 mL).
7. Remove the comb.
8. Mix 10 μL denaturated and renaturated PCR product with 2 μL loading buffer in a 96 well plate.
9. Pipet the mixture immediately in one of the gel slots.
10. Load the DNA ladder on the gel.
11. Cover the electrophoresis cell, place the electrodes, and run the gel for 12 h at 3 V/cm (the exact time depends on the size of the PCR products that are loaded on the gel).
12. After turning off the power, remove the cover and carefully take out the gel cassette.
13. Open the gel cassette and rinse the gel of the cassette using 100 mL Milli-Q water.
14. Incubate the gel for 2–5 min in 100 mL Milli-Q water with 10 μL GelRed on a rocker platform.
15. Wash the gel for 2–5 min in 100 mL Milli-Q water on a rocker platform.
16. Analyze the gel using UV light; make a photograph of the gel.
17. Interpret the gel. One band of expected size indicates one clonal rearrangement; two homoduplex bands plus two heteroduplex bands indicate two clonal rearrangements (biallelic or bi-clonal); >4 bands ("ladder pattern") indicate oligoclonal rearrangements; a smear indicates polyclonal rearrangements. Weak or very weak bands may indicate subclonal rearrangements.

3.2.3 PCR Product Purification and Sanger Sequencing of Clonal Rearrangements

After identification of clonal IGH gene rearrangements or the chromosomal translocations t(11;14) in the DNA sample of the patient at diagnosis, the precise sequences have to be determined. By comparing the obtained sequences with the sequences of the germline gene segments, the exact composition of the patient-specific junctional regions can be obtained. Clonal t(14;18) rearrangements can be quantified by RQ-PCR with a consensus primer-probe system without prior sequencing.

Monoallelic rearrangements and PCR products of t(14;18) and t(11;14) rearrangements can generally be sequenced directly from the PCR product. Use 1 μL PCR product for the sequencing reaction.

1. Prior to sequencing, PCR products have to be purified to remove primer dimers, excess primers, and dNTPs.
2. PCR products are purified by putting the 50 μL PCR products each in a MinElute 96 UF Plates [11] (Qiagen). Put the plate on a vacuum device (Millipore), and apply vacuum for max. 10 min until the membranes are dry.
3. Put 25 μL Milli-Q water in each well and shake for 5–10 min.
4. Transfer the solution in fresh Eppendorf cups.
5. To estimate the concentration of DNA, 1/10 of the DNA is run on a 1% analytical agarose gel containing ethidium bromide together with a DNA mass ladder which allows the estimation of the DNA concentration by comparing the strength of the PCR product band with the strength of the DNA fragments of the mass ladder.
6. Perform sequencing using the BigDye® Terminator (BDT) v3.1 Cycle Sequencing Kit in a total reaction volume of 20 μL. The reaction contains template DNA (PCR product; *see* **step 4**), 1 μL BDT, 7 μL 5× sequencing buffer, 2 μL primer (5 pmol/μL), and Milli-Q water (*see* **Note 5**). In case of multiplex sequencing reactions, use 0.5 μL of each primer (2 pmol/μL).
7. Use the same primers as used for the PCR reactions, but consider that primers should be unlabeled when used for sequencing (Table 2). In case of multiple primers in the first PCR reaction, use all forward or reverse primers for sequencing.
8. The amount of template DNA should be singleplex PCR, strong band, 1 μL; weak band, 2 μL; and very weak band, 3 μL, and eluted homo- or heteroduplex band, strong, 2 μL; weak, 3 μL; and very weak, 4 μL.
9. The PCR conditions are 96 °C, 5 min; 50 cycles of 96 °C 10 s, 55 °C 5 s, and 60 °C 4 min.

10. After the sequencing reaction, purify the samples, e.g., by using Sephadex G-50 on a Millipore Multiscreen-HV-96-well plate, and centrifugate the plate 5 min at 860 g.
11. Load and run the purified samples on the sequencing instrument, e.g., a 3500 Genetic Analyzer.
12. In order to obtain a reliable junctional region sequence, sequence the clonal PCR product always from both directions (i.e., using both a forward and a reverse primer). In case of doubt, a second (independent) clonal PCR product should be sequenced.
13. Analyze the sequence data using sequence navigator or comparable software programs.

In case of biallelic rearrangements or polyclonal background, heteroduplex analysis is used to isolate the clonal bands for sequencing. For this purpose follow the instructions in Subheading 3.2.2 for PCR heteroduplex analysis. Instead of 10 μL PCR product as recommended in 11. Mix 3 × 10 μL denaturated and renaturated PCR product with 2 μL BFB loading buffer each in a 96-well plate, and pipet the mixture immediately in three lanes of 10 μL each of the gel slots. Then continue with the procedure as in Subheading 3.2.2, **step 10**.

1. Interpret the gel. In case of biallelic rearrangement or a clonal PCR product with polyclonal background, the clonal band has to be excised.
   (a) In case of biallelic rearrangements, the relevant homoduplex bands need to be isolated. If the two homoduplexes are not well separated, the relevant heteroduplex bands should be isolated. The bands of interest are cut from the gel (on an UV lamp), and each put in a microcentrifuge tube. 100 μL aqua dest. is added to each tube, and the tubes are incubated overnight at 37 °C. The eluate is subsequently transferred to a new tube, and 10 μL NaAc and 200 μL ice-cold 96% EtOH are added. After overnight incubation at −20 °C (or 2× 45 min on dry ice), the tube is centrifuged for 20 min at 10,000 rpm. The pellet is washed with 200 μL 70% EtOH, followed by centrifugation for 20 min at 13,000 rpm (Eppendorf centrifuge 16000 × *g*). After removal of the supernatant, the pellet is air-dried and dissolved in 10 μL Milli-Q water. Generally, 2 μL PCR product can be used for the sequencing reaction. If the recovery of the excised DNA is low, a higher volume may however be necessary.
   (b) Alternatively, 50 μL aqua dest. is added to each tube. After vortexing and short centrifugation, the tubes are incubated overnight at 4 °C. The next day the tubes are

vortexed again, and 7 μL of the solution is directly used for sequencing.

(c) Exceptionally, if the above approaches are not successful, one might decide to clone the PCR products, followed by sequencing of 10–15 clones.

3.2.4 Sequence Interpretation

The junctional regions of rearranged IGH genes can be considered as "DNA fingerprints" of the lymphoma cells which are used as MRD-PCR targets. Also chromosomal translocations t(14;18) and t(11;14) involving the IGH gene on chromosome 14 bear a unique junctional region. Therefore, it is crucial to correctly assess the precise composition of the junctional region. To this end, the IGH sequence obtained should be aligned with the germline V, D, and/or J gene segments, so that the exact number of deleted and inserted nucleotides can be determined. Alignment can be done using several databases freely available on the Internet or using hard copies. Since some variations can be seen between different databases, it is advised to check the alignment of the sequence in at least two ways.

1. Identify sequences using IMGT (http://imgt.cines.fr) for human Ig and TCR genes. Sequences can be confirmed by using BLAST sequence similarity searching tool (National Center for Biotechnology Information: http://www.ncbi.nlm.nih.gov/BLAST/ or http://www.ncbi.nlm.nih.gov/igblast/).
2. First, try to assign the appropriate V, D, and/or J gene segments to the sequence. For recognition of D segments, a minimum length of 1/3 of the complete germline segment is required, with a minimum length of five base pairs (*see* also **Note 6**).
3. Analysis of D and/or J gene segment should be done by IMGT (http://imgt.cines.fr) for human Ig and TCR genes.
4. After having identified the appropriate V, D, and/or J gene segments of a clonal IGH rearrangement, determine the number of deletions at both sides of all joining sites, that is, the number of nucleotides present in the germline V, D, and/or J gene segments that are lost in the patient's sequence. The remaining parts of the germline V, D, and/or J gene segments should give a 100% alignment with the patient's sequence. Lack of complete alignment may be due to somatic mutations or due to polymorphisms. Since such variations may impact RQ-PCR primer and/or probe binding, these variations should clearly be marked in the patient's sequence.
5. Subsequently, determine the number of insertions in the joining site(s).

6. For chromosomal translocations: try to assign the chromosomal breakpoints on chromosome 18 or 11 by using published sequences. The identification of the junctional region with D and J gene segments can be done manually from published sequences or by using electronic data bases like IMGT (http://imgt.cines.fr) for the exact identification of human Ig and TCR genes requires some modification. Because a minimum sequence length (about 150 bp) is required and the junctional region and the JH gene segment involved in the translocation region are too short, an "artificial "VH-germline sequence can be added manually 5′ of the junctional N-DH-N-JH sequence of the respective rearrangement instead of the chromosome 18 or 11 sequence. With this artificial extension sequence, proper identification of DH and JH gene segments involved in the translocation can be performed.

3.2.5 Selection of MRD-PCR Targets

In order to limit the risk of false-negative MRD results, preferably two MRD-PCR targets should be used for each lymphoma patient if available. These MRD-PCR targets should be selected based on expected sensitivity. A sensitivity of at least 10^{-4} should be obtained. When using IGH gene rearrangements, somatic mutations may hamper RQ-PCR sensitivity especially in germinal center/post-germinal center B-cell malignancies (e.g., FL) since RQ-PCR primer and/or probe binding may be constrained. Therefore, these variations should be carefully marked in the patient's sequence and should be considered for the selection of allele-specific primer. In patients with chromosomal translocations as t(14;18) translocation or t(11;14), these translocations can serve as the preferred MRD target because they are highly stable and the JH-region involved in the translocation is generally unmutated.

The sensitivity of MRD analysis of junctional regions is dependent on several factors, including the size of the junctional region, the "background" of normal lymphoid cells with comparable polyclonal IGH gene rearrangements and the frequency of somatic IGH mutations [36]. If a relatively high proportion of lymphoma cells and a low background of polyclonal B cells are present as, e.g., in CLL at diagnosis, the MRD level can generally reliably be quantified from the diagnostic sample. In case of low MRD levels in the peripheral blood and bone marrow, e.g., in primary nodal lymphomas as FL or MCL, the assay becomes less accurate. A prerequisite for quantification from diagnostic material is therefore the assessment of the percentage of lymphoma cells in the MNC fraction by flow cytometric immunophenotyping. In cases with infiltration of less than 1% or an unfavorable standard curve, cloning of the junctional regions into plasmids and plasmid-based quantification of patient-specific IGH VDJ targets might be a good alternative [37].

Within the EURO-MRD group (*see* **Note 7**), a subdivision of the European Scientific foundation for Laboratory Hemato Oncology (ESLHO) (http://www.eslho.org) technical parameters was defined for the *quantitative range* and the *sensitivity* for RQ-PCR [23]. The "quantitative range" reflects the part of the standard curve in which the MRD levels can be quantified reproducibly and accurately, whereas the "sensitivity" reflects the lowest MRD level that still can be detected, although not reproducibly and accurately. These definitions are also used for the quantification of MRD levels in malignant lymphomas.

Generally, IGH gene-based MRD assessment results in higher sensitivities in lymphoma entities with a lower frequency of somatic mutations, e.g., unmutated or low-mutated MCL and CLL. In germinal center/post-germinal center B-cell malignancies with a high mutation load, a completely sequence-based design of primers and probes might be necessary to achieve the required sensitivity of 10^{-4}. Alternatively, light chain rearrangements of rearranged Ig kappa (IgK) are less frequently mutated and may serve as alternative MRD targets [18].

3.3 RQ-PCR Sensitivity Testing

3.3.1 Design of Allele-Specific Oligonucleotide Primers

The following instructions have been developed within the EURO-MRD network for the design of allele-specific oligonucleotide (ASO) primers in general and have been successfully applied for MRD assessment in ALL, CLL, and MCL for RQ-PCR. For the application in FL, the same guidelines can be followed. In principle, the same allele-specific oligonucleotide primers and consensus probes can be used for ddPCR.

For RQ-PCR analysis generally the TaqMan (hydrolysis) probe approach is being used (*see* **Note 8**). Several primer-probe sets for RQ-PCR-based detection of tumor-specific *IGH* have been designed and successfully applied in patients with ALL [38, 39]. Although both an ASO probe approach and ASO primer approach can be used, the ASO primer approach has the advantage that it is much cheaper since no patient-specific probes need to be ordered. However, in cases with frequent somatic mutations of IGH gene rearrangements, a completely sequence-based design of primers and probes is recommended for RQ-PCR analysis.

The germline primer and probe sets should be used in combination with a patient-specific ASO primer. This ASO primer should be positioned within and around the junctional region in such a way that it provides maximal specificity.

1. Enter the sequence of the IGH rearrangement or the t(11;14) translocation in Primer Express software or any other software that is available.
2. Select an appropriate germline primer and probe.

3. Mark the position of the germline primer and probe in the sequence.
4. Design an allele-specific oligonucleotide (ASO) primer using any available software according to the following guidelines.
   (a) Place the 3′ end of the primer in or just over the junctional region; maximum overlap with germline sequence: 2–4 nucleotides. In case of *IGH* gene rearrangements, the design of an ASO primer within the D_H–J_H junctional region has the advantage of high specificity. In case of chromosomal translocations, the ASO primer can be placed either completely in the D_H–J_H junctional region or overlapping the breakpoint region on the respective chromosome and the D_H–J_H junctional region.
   (b) Melting temperature (T_m) 57–60 °C, according to the nearest neighbor algorithm [40].
   (c) The GC content should be in the range of 20–80%.
   (d) Check in Oligo Primer Analysis Software v. 7 or in your primer design program:
      - Internal stability graph (if possible no more than two G/Cs in the last five bases on 3′ end according to Primer Express): strong binding in the middle part of primer and lower binding at 3′ end ($\Delta G > -10$).
      - 3′-dimers forward-forward and forward-reverse primer (<5 kcal/mol).
      - Dimers forward-forward and forward-reverse primer (preferably <10 kcal/mol).
      - Hairpin stems 3′ ends (no hairpin with T_m within 10 °C from T_m primer).
5. Order the primers in 1× TE, concentration 100 pmol/μL.
6. Prepare 300 μL of a primer working dilution in 1× TE buffer with a concentration of 20 pmol/μL. Store the stock and working solution at −20 °C.

3.3.2 Application of Consensus Primers and Probes for t(14;18) RQ-PCR

For RQ-PCR and ddPCR analysis of the t(14;18) translocation, forward primer and probe are placed in chromosome 18 in combination with a reverse primer placed in the consensus JH region on chromosome 14. Because only cells bearing the translocation can be amplified, the assay is specific without the necessity to use an ASO primer. With this approach the same primers and probes can be universally used for different patients making this approach much cheaper and faster since no ASO primers have to be designed. A further advantage of this approach is the possible use of t(14;18) bearing cell lines as diagnostic samples. This is especially important in FL which frequently doesn't have a leukemic presentation and

the degree of infiltration in the BM is unknown. Beyond this, the usage of a cell line as diagnostic sample provides the advantage that different patients can be quantified with the same cell line as diagnostic samples. Primers and probes for this approach are listed in Table 2.

However, there are several restrictions for the usage of t(14;18) RQ-PCR assays. The t(14;18) breakpoints cluster in at least four different regions: the major breakpoint region (MBR), the minor clustering region (mcr), the 3′-MBR, and the 5′-mcr [25, 26, 41]. A number of variant translocations have been described that cannot be usefully assessed by PCR [41]. Chromosome 14 breakpoints are almost exclusively located within one of the six germline JH segments allowing PCR amplification of the translocation. Only MBR, and to a lesser extent mcr, have been extensively exploited for MRD analysis. Thus, only a subset of FL patients, ranging from 50% to 65%, can be assessed using the BCL2-IGH PCR [21].

It is part of the current activities of the EURO-MRD group to develop and validate RQ-PCR assays for other breakpoints than MBR. Basically, the procedures described here for the t(14;18) MBR region may be extrapolated to other (t(14;18) breakpoints. However, the size of the respective RQ-PCR product is of great importance since PCR products excessing 200 bp in size are not reliably quantified due to a lower amplification efficiency and to a possibly lower amplification efficiency in comparison with the diagnostic cell line.

3.3.3 RQ-PCR Analysis of Dilution Series of Diagnostic Sample

In order to determine the sensitivity of the RQ-PCR assay and the specificity of a designed ASO primer, serial dilutions of the diagnostic sample should be used.

1. Based on the lymphoma cell percentage (PB and BM) in the diagnostic sample determined immunophenotypically or by digital PCR, the diagnostic DNA sample (60 ng/μL) should be diluted in normal control MNC DNA (60 ng/μL) to a lymphoma cell percentage of 10% (*see* **Notes 9** and **10**). Prepare 150 μL of this 10^{-1} dilution (*see* **Note 11**). If the lymphoma cell infiltration in diagnostic samples is really low (<1%), one might consider plasmid-based quantification of patient-specific IGH VDJ targets as published [37].
2. Prepare a 10^{-2} dilution by mixing 15 μL of the 10^{-1} dilution with 135 μL normal control MNC DNA (60 ng/μL). Comparably, prepare 150 μL of a 10^{-3}, 10^{-4}, and 10^{-5} dilution. Prepare a 5×10^{-4} dilution by mixing 7 μL of the 10^{-2} dilution with 133 μL normal control MNC DNA (60 ng/μL).
3. Perform a RQ-PCR reaction using the 10^{-1} to 10^{-5} dilutions in duplicate as well as the normal control MNC DNA in sixfold (*see* **Note 12**) and Milli-Q water in duplicate. 600 ng DNA

Table 5
RQ-PCR TaqMan reaction mixtures

Reaction mix per 25 μL reaction	Final concentration	μL/well
Universal Master mix (2×)	1×	12.5
Forward primer (20 pmol/μL)	300 nM	0.38
Reverse primer (20 pmol/μL)	300 nM	0.38
TaqMan probe (20 pmol/μL)	100 nM	0.13
BSA (5%)	0.04%	0.2
Milli-Q water		5.4
DNA (100 ng/μL)		6

should be used per well (equivalent to approximately 90,000 cells, required to obtain a theoretical sensitivity of 10^{-5}). Reaction mixes are shown in Table 5.

4. Thermal cycler conditions are 50 °C, 2 min; 95 °C, 10 min; 50 cycles of 95 °C, 30 s; and 60 °C, 1 min.
5. Analyze the RQ-PCR data using appropriate software.
6. The threshold of the RQ-PCR assay should always be set in the region of exponential amplification across all amplification plots. This region is depicted in the log view of amplification plots as the portion of the curve that is linear. Often the threshold automatically determined by the instrument software can be used. However, if the threshold appears to be positioned outside the linear part of (some of) the amplification curves, adjustments may be made [24, 25].
7. By plotting the logarithmic value of the dilution series against the cycle threshold (C_T) obtained, a standard curve can be obtained.

3.3.4 RQ-PCR Data Interpretation: Quantitative Range and Sensitivity

For defining the sensitivity, several criteria (including reproducibility of the measurement, the difference between specific and non-specific amplification, slope and correlation coefficient of the standard curve) should be taken into account [25]. Furthermore, in order to compare data between different studies and/or different laboratories, it is essential to have uniform guidelines for RQ-PCR data interpretation. For Ig/TCR-based MRD data interpretation in ALL, such guidelines have been developed within the ESG-MRD-ALL (now EURO-MRD consortium) and have been applied for MRD assessment in malignant lymphoma by the same group (*see* **Notes 13** and **14**).

3.4 Bone Marrow and Peripheral Blood Sample Processing During Follow-Up

1. Obtain a total of 2.5–3 mL BM in one first aspirate or use 10 mL PB.
2. Perform Ficoll density centrifugation for obtaining MNC.
3. Extract DNA from 10×10^6 MNC using a DNA extraction column or any other DNA extraction method, generally resulting in 30–40 μg of DNA. This is sufficient for standard RQ-PCR analysis (applying two MRD-PCR targets and one control gene) (*see* **Note 15**).
4. Make an RQ-PCR solution of 100 ng/μL and store at 4 °C (or −20 °C).
5. Store remaining MNC of the BM follow-up samples in cell storage medium for later studies; store in liquid nitrogen or as a dry pellet at −80 °C.

3.5 MRD Analysis of Follow-Up Samples

The actual MRD analysis includes three main steps: RQ-PCR analysis of a control gene for checking the quality and quantity of the DNA, RQ-PCR analysis of the IGH targets or chromosomal translocations, and finally interpretation of the data obtained.

3.5.1 Control Gene RQ-PCR Analysis

In order to check for the quantity and quality of the DNA, RQ-PCR analysis of a control gene is required. For this purpose, the albumin (*ALB*) gene on chromosome 4 can be used, and *ALB* quantification has been highly standardized [42].

1. Make dilutions of normal control MNC DNA (100 ng/μL) in Milli-Q water, so that solutions of 50, 10, and 1 ng/μL are obtained.
2. Analyze the four normal control DNA samples (i.e., 100, 50, 10, and 1 ng/μL), the (undiluted) diagnostic DNA (100 ng/μL) and the follow-up samples (100 ng/μL) in duplicate. Use Milli-Q water as a negative control (in duplicate). Use the reaction mix as indicated in Table 5.
3. Thermal cycler conditions (for LC480; other platforms may require different optimal temperatures and times): 40 °C, 10 min; 95 °C, 10 min; 95 °C, 15 s; 58 °C, 30 s; and 72 °C, 1 s (50×).
4. Analyze the RQ-PCR data using appropriate software. It is advised to use a common threshold setting for all control gene RQ-PCR reactions, since this allows an easier comparison of data obtained in different RQ-PCR runs. The C_t values obtained for the standard curve wells can be used to monitor the performance of the RQ-PCR reaction over time.
5. Construct a standard curve by plotting the C_t values of the four normal control MNC DNA against their C_t values.

6. Determine the DNA content of the diagnostic and follow-up DNA samples using their Ct value and the obtained standard curve (*see* **Note 16**).

3.5.2 MRD-PCR Target RQ-PCR Analysis

For the actual analysis of MRD in the follow-up samples, RQ-PCR analysis is performed for the selected targets.

1. For each MRD-PCR target, analyze the dilutions series (10^{-1} to 10^{-5}; *see* Subheading 3.3.2) of the diagnostic DNA (60 ng/μL) in duplicate, the follow-up samples (100 ng/μL) in triplicate, and normal control MNC DNA in sixfold. Use Milli-Q water as a negative control (in triplicate). Use the reaction mix as indicated in Table 5.
2. Of importance, the control gene PCR and MRD-PCR target PCRs should be performed using exactly the same DNA sample in identical reaction volumes and using identical amounts of DNA.
3. Thermal cycler conditions (for LC480; other platforms may require different optimal temperatures and times): 40 °C, 10 min; 95 °C, 10 min; 95 °C, 15 s; 62 °C, 30 s; and 72 °C, 1 s (50×) (*see* **Note 17**).
4. Analyze the RQ-PCR data using appropriate software.
5. The threshold of the RQ-PCR assay should always be set in the region of exponential amplification across all amplification plots. This region is depicted in the log view of amplification plots as the portion of the curve that is linear. Often the threshold automatically determined by the instrument software can be used. However, if the threshold appears to be positioned outside the linear part of (some of) the amplification curves, adjustments may be made [24, 25].

3.5.3 RQ-PCR MRD Data Interpretation

For interpretation of the MRD data, the guidelines of the ESG-MRD-ALL [43] should be used (*see* **Note 18**).

For MRD status interpretation, the MRD-PCR target with the highest MRD level should be regarded as the most reliable target. So, always the highest MRD level per time point should be reported. The optimal marker should have a quantitative range of at least 10^{-4} and a sensitivity of 10^{-5}, less sensitive markers with a quantitative range of $>10^{-3}$ and a sensitivity at least 10^{-4} are acceptable. The acceptance of a MRD marker is dependent on the protocol and purpose of a clinical trial.

3.6 ddPCR

All analyses were performed on a QX100 Droplet Digital PCR system (Bio-Rad Laboratories, Hercules, CA).

3.6.1 Target ddPCR Analysis

For ddPCR the allele-specific oligonucleotide primers and consensus probes like in RQ-PCR can be used.

Table 6
ddPCR reaction mixtures

Reaction mix per 20 μL reaction	Final concentration	μL/well
2× ddPCR Supermix for Probes	1×	10
Forward primer (20 pmol/μL)	300 nM	0.3
Reverse primer (20 pmol/μL)	300 nM	0.3
TaqMan probe (20 pmol/μL)	100 nM	0.1
Restrictions endonuclease (2 U/μL)[a]	2 U	1
Milli-Q water		Ad 20 μL
DNA (100 ng/μL)		5

[a]Optional

1. Setup for every ddPCR-run (replicates): positive control (2×), sample (3×), NTC (2×), and normal control MNC DNA (2×). Reaction mixes as indicated in Table 6.
2. Loading 20 μL of reaction mix and 70 μL of droplet generation oil into the DG8 cartridge, covered with the DG8 gasket and finally placed into the droplet generator.
3. Transfer the 40 μL of droplets carefully into a Twin.tec 96 well semi-skirted PCR plate.
4. Close the PCR plate with a pierceable foil heat seal in the PX1 PCR plate sealer.
5. PCR cond.: 95 °C, 10 min; 40 cycles of 94 °C, 30 s; 60 °C 1, min, and final 98 °C, 10 min.
6. Run the samples on the QX200 Droplet Reader.

3.6.2 ddPCR Data Interpretation

Using the following guidelines for analysis and interpretation as positive (pos.), negative (neg.), or positive not quantifiable (PNQ):

1. Only replicates ≥ 9.000 droplets should be considered for the analysis.
2. The threshold must be set manually. The threshold should be set close to the positive control droplets to exclude the unspecific background signals.
3. In case of positive events in NTC or BC, verify the consistency of the amplification signal by checking the presence of positive droplets in channel 2 (ch2). If signals in ch2 (green) are present, this represents an unspecific amplification (false-positive signal) and should be excluded from the analysis.
4. Interpretation and definition of results:
   (a) *MRD positive*: samples with three positive replicates and a merge of ≥ 3 events.

(b) *MRD negative*: samples with all three replicates negative (with no events) *or* replicates with a merge of ≤ 2 events.[1]

(c) *MRD PNQ*: samples with < 3 positive replicates and a merge of > 2 events.[1]

3.7 Concluding Remarks

In contrast to MRD diagnostics in ALL the clinical implications of PCR-based MRD diagnostics in malignant lymphoma has only evolved recently. As in ALL patients, it is also clear that the PCR-based MRD diagnostics via IG genes and chromosomal translocations is complex and requires extensive knowledge about gene rearrangement processes as well as the consideration of specific characteristics of the heterogeneous histological subtypes. Therefore, usage of standardized assays and internationally accepted guidelines for interpretation of RQ-PCR results as well as regular internal and external quality control rounds is a prerequisite for MRD-based treatment modification in lymphoma.

4 Notes

1. Ideally the diagnostic material for generating the standard curves contains 70–100% tumor cells. However, in the majority of MCL and especially nodal FL as well as in a subset of CLL patients, infiltration of BM or PB is less than 50%. Therefore, flow cytometric immunophenotyping or ddPCR allows the determination of the percentage of tumor cells and with this the correction of the tumor load in the diagnostic sample. While in MCL and CLL the percentage of lymphoma cells is almost equal in PB and BM [11], in FL circulating lymphoma cell level is about one log higher in BM compared to PB [44, 45]. Therefore in FL it is useful in many cases to use BM for the standard curve. The use of samples with low percentages of lymphoma cells (<10%) may hamper the reliable quantification.
2. For DNA extraction different column based and other methods give comparable good results. The usage of fixed numbers of MNC facilitates reproducibility of PCR analysis because extraction of DNA from 10×10^6 MNC will generally result in 30–40 μg of DNA, which is sufficient for all relevant PCR analyses at diagnosis including the preparation of a standard curve.

[1] ddPCR should be repeated on more replicates (≥ 6, based on the sample availability) to verify if the results are reproducible.

3. For assessment of DNA concentration, the NanoDrop method is preferred over analysis using another type of spectrophotometer, since the results are more robust and accurate.
4. Instead of commercially available gels, also home-made 6% PAGE gels can be used.
5. Alternatively, comparable commercially available sequencing kits can be used.
6. The recognition of (second) D segments is often not done by software. Therefore, if the N region consists of >10 bp, check for D segment by hand (using hard copies), or use only the sequence around the N region for analysis in databases.
7. The EuroMRD group under the umbrella of ESHLO (European Scientific Foundation of Laboratory Hemato-Oncology) is a consortium of international MRD-PCR laboratories, coordinated by J.J.M. van Dongen. The main aims of the group are the organization of a quality control program twice per year, the collaborative development and clinical evaluation of new MRD strategies and techniques, and the development of guidelines for the interpretation of RQ-PCR-based MRD data. Within EuroMRD, a group focusing on MRD in mature lymphoid malignancies has developed standards for MRD assessment in lymphoma and is performing regular quality controls twice per year.
8. For RQ-PCR analysis, other probe formats (e.g., hybridization probes) can also be used. The use of non-specific dyes, like SYBR Green I, should be avoided, since this limits the sensitivity.
9. MNC obtained from PB can be used. For Ig gene rearrangements, a pool of at least five healthy individuals should be used. When performing t(14;18) PCR, normal control MNC DNA should be tested for the presence of t(14;18) translocation because t(14;18)-positive cells can be present in healthy donors without lymphoma. Alternatively, DNA from cell lines without the t(14;18) translocation can be used.
10. In many cases the degree of infiltration is less than 10% in the diagnostic material. To our experience, it is possible to generate reliable standard curves from diagnostic samples with a minimum infiltration of 1%. This implies that the standard curve starts at 10^{-2} dilution as the lowest dilution of the standard curve.
11. If only a limited amount of diagnostic DNA is available, one might decide to skip the 10^{-1} dilution and to use the 10^{-2} dilution as the lowest dilution of the standard curve.
12. Since non-specific amplification is generally only detected at low levels and outside the reproducible range of the RQ-PCR,

non-specific amplification controls should be run at least in sixfold in each RQ-PCR analysis for each Ig or TCR marker.

13. The background of the RQ-PCR assay, i.e., the non-specific amplification of comparable Ig gene rearrangements present in normal cells or t(14;18)-positive cells not belonging to the lymphoma, is defined by the lowest C_T value obtained in the non-specific amplification controls (normal MNC DNA).
14. If the quantitative range and/or sensitivity of the RQ-PCR test is not sufficient, i.e., has non-specific amplification, the annealing temperature can be increased (up to 70 °C) in order to improve specificity. If this is not successful, one can decide to design an alternative ASO primer.
15. At least 2×10^6 MNC should be used for DNA extraction since at least 5–8 μg of DNA should be available for RQ-PCR analysis.
16. If the RQ-PCR of the control gene shows a lower amount of template than expected based on physical measurements (e.g., NanoDrop measurement), special caution is needed since this lower value can be the result of inhibition, which can be found in a substantial number of BM or PB samples (5–10%). Addition of BSA prevents inhibition, and the EURO-MRD therefore recommends the inclusion of 0.04% BSA in all RQ-PCR reactions [25]. Since the addition of less template will result in loss of sensitivity, the control gene values of all samples need to be within predefined ranges, that is, between 250 and 1000 ng/reaction. If the DNA amount appears to be higher, the follow-up sample should be diluted and re-tested because too high DNA concentrations may also result in PCR inhibition. If the DNA amount is lower, correct interpretation may not be possible. Consequently, reanalysis is recommended after adaptation of the amount of DNA [25].
17. The annealing temperature can be different, dependent on the results of the sensitivity testing.
18. Logically, very low MRD levels (below the "quantitative range") should always be judged with caution, especially if only one well of the three replicates is positive. In such case, reanalysis of the doubtful sample(s) may be performed, but one should be aware that by definition the results will often not be reproducible.

Acknowledgment

We gratefully acknowledge Anne-Wiebke Kruse for her excellent technical support.

References

1. Brown JR, Feng Y, Gribben JG, Neuberg D, Fisher DC, Mauch P, Nadler LM, Freedman AS (2007) Long-term survival after autologous bone marrow transplantation for follicular lymphoma in first remission. Biol Blood Marrow Transplant 13:1057–1065
2. Goff L, Summers K, Iqbal S, Kuhlmann J, Kunz M, Louton T, Hagenbeek A, Morschhauser F, Putz B, Lister A, Rohatiner A (2009) Quantitative PCR analysis for Bcl-2/IgH in a phase III study of yttrium-90 ibritumomab tiuxetan as consolidation of first remission in patients with follicular lymphoma. J Clin Oncol 27:6094–6100
3. Hirt C, Schuler F, Kiefer T, Schwenke C, Haas A, Niederwieser D, Neser S, Assmann M, Srock S, Rohrberg R, Dachselt K, Leithauser M, Rabkin CS, Herold M, Dolken G (2008) Rapid and sustained clearance of circulating lymphoma cells after chemotherapy plus rituximab: clinical significance of quantitative t(14;18) PCR monitoring in advanced stage follicular lymphoma patients. BrJ Haematol 141:631–640
4. Galimberti S, Luminari S, Ciabatti E, Grassi S, Guerrini F, Dondi A, Marcheselli L, Ladetto M, Piccaluga PP, Gazzola A, Mannu C, Monitillo L, Mantoan B, Del Giudice I, Della Starza I, Cavalli M, Arcaini L, Tucci A, Palumbo GA, Rigacci L, Pulsoni A, Vitolo U, Boccomini C, Vallisa D, Bertoldero G, Gaidano G, Musto P, Petrini M, Federico M (2014) Minimal residual disease after conventional treatment significantly impacts on progression-free survival of patients with follicular lymphoma: the FIL FOLL05 trial. Clin Cancer Res 20:6398–6405
5. Ladetto M, Lobetti-Bodoni C, Mantoan B, Ceccarelli M, Boccomini C, Genuardi E, Chiappella A, Baldini L, Rossi G, Pulsoni A, Di Raimondo F, Rigacci L, Pinto A, Galimberti S, Bari A, Rota-Scalabrini D, Ferrari A, Zaja F, Gallamini A, Specchia G, Musto P, Rossi FG, Gamba E, Evangelista A, Vitolo U (2013) Persistence of minimal residual disease in bone marrow predicts outcome in follicular lymphomas treated with a rituximab-intensive program. Blood 122:3759–3766
6. Ladetto M, De Marco F, Benedetti F, Vitolo U, Parti C, Rambaldi A, Pulsoni A, Musso M, Liberati AM, Olivieri A, Gallamini A, Pogliani E, Scalabrini DR, Callea V, Di Raimondo F, Pavone V, Tucci A, Cortelazzo S, Levis A, Boccadoro M, Majolino I, Pileri A, Gianni AM, Passera R, Corradini P, Tarella C (2008) Prospective, multicenter randomized GITMO/IIL trial comparing intensive (R-HDS) versus conventional (CHOP-R) chemoimmunotherapy in high-risk follicular lymphoma at diagnosis: the superior disease control of R-HDS does not translate into an overall surviva. Blood 111:4004–4013
7. Rambaldi A, Carlotti E, Oldani E, Della Starza I, Baccarani M, Cortelazzo S, Lauria F, Arcaini L, Morra E, Pulsoni A, Rigacci L, Rupolo M, Zaja F, Zinzani PL, Barbui T, Foa R (2005) Quantitative PCR of bone marrow BCL2/IgH positive cells at diagnosis predicts treatment response and long-term outcome in Follicular non Hodgkin's Lymphoma. Blood 105:3428–3433
8. Geisler CH, Kolstad A, Laurell A, Jerkeman M, Räty R, Andersen NS, Pedersen LB, Eriksson M, Nordström M, Kimby E, Bentzen H, Kuittinen O, Lauritzsen GF, Nilsson-Ehle H, Ralfkiaer E, Ehinger M, Sundström C, Delabie J, Karjalainen-Lindsberg ML, Brown P, Elonen E, Nordic Lymphoma Group (2012) Nordic MCL2 trial update: six-year follow-up after intensive immunochemotherapy for untreated mantle cell lymphoma followed by BEAM or BEAC + autologous stem-cell support: still very long survival but late relapses do occur. Br J Haematol 158:355–362
9. Hermine O, Hoster E, Walewski J, Bosly A, Stilgenbauer S, Thieblemont C, Szymczyk M, Bouabdallah R, Kneba M, Hallek M, Salles G, Feugier P, Ribrag V, Birkmann J, Forstpointner R, Haioun C, Hänel M, Casasnovas RO, Finke J, Peter N, Bouabdallah K, Sebban C, Fischer T, Dührsen U, Metzner B, Maschmeyer G, Kanz L, Schmidt C, Delarue R, Brousse N, Klapper W, Macintyre E, Delfau-Larue M-H, Pott C, Hiddemann W, Unterhalt M, Dreyling M (2016) Addition of high-dose cytarabine to immunochemotherapy before autologous stem-cell transplantation in patients aged 65 years or younger with mantle cell lymphoma (MCL Younger): a randomised, open-label, phase 3 trial of the European Mantle Cell Lymphoma N. Lancet 388:565–575
10. Pott C, Schrader C, Gesk S, Harder L, Tiemann M, Raff T, Bruggemann M, Ritgen M, Gahn B, Unterhalt M, Dreyling M, Hiddemann W, Siebert R, Dreger P, Kneba M (2006) Quantitative assessment of molecular remission after high-dose therapy with autologous stem cell transplantation predicts long-

term remission in mantle cell lymphoma. Blood 107:2271–2278

11. Pott C, Hoster E, Delfau-Larue MH, Beldjord K, Bottcher S, Asnafi V, Plonquet A, Siebert R, Callet-Bauchu E, Andersen N, van Dongen JJ, Klapper W, Berger F, Ribrag V, van Hoof AL, Trneny M, Walewski J, Dreger P, Unterhalt M, Hiddemann W, Kneba M, Kluin-Nelemans HC, Hermine O, Macintyre E, Dreyling M (2010) Molecular remission is an independent predictor of clinical outcome in patients with mantle cell lymphoma after combined immunochemotherapy: a European MCL intergroup study. Blood 115:3215–3223
12. Hoster E, Klapper W, Hermine O, Kluin-Nelemans HC, Walewski J, Van HA, Trneny M, Geisler CH, Di RF, Szymczyk M, Stilgenbauer S, Thieblemont C, Hallek M, Forstpointner R, Pott C, Ribrag V, Doorduijn J, Hiddemann W, Dreyling MH, Unterhalt M (2014) Confirmation of the mantle-cell lymphoma International Prognostic Index in randomized trials of the European Mantle-Cell Lymphoma Network. JClinOncol 32:1338–1346
13. Rambaldi A, Carlotti E, Oldani E, Della Starza I, Baccarani M, Cortelazzo S, Lauria F, Arcaini L, Morra E, Pulsoni A, Rigacci L, Rupolo M, Zaja F, Zinzani PL, Barbui T, Foa R (2005) Quantitative PCR of bone marrow BCL2/IgH+ cells at diagnosis predicts treatment response and long-term outcome in follicular non-Hodgkin lymphoma. Blood 105:3428–3433
14. Voena C, Locatelli G, Castellino C, Omede P, Ladetto M, Zappone E, Milani R, Perfetti V, Boccadoro M, Pileri A, Lusso P, Villa C, Malnati M, Corradini P (2002) Qualitative and quantitative polymerase chain reaction detection of the residual myeloma cell contamination after positive selection of CD34+ cells with small- and large-scale Miltenyi cell sorting system. Br J Haematol 117:642–645
15. Ladetto M, Corradini P, Vallet S, Benedetti F, Vitolo U, Martelli M, Brugiatelli M, Coser P, Perrotti A, Majolino I, Fioritoni G, Morandi S, Musso M, Zambello R, Chisesi T, Di Renzo N, Vivaldi P, De Crescenzo A, Gallamini A, Salvi F, Santini G, Boccomini C, Sorio M, Astolfi M, Drandi D, Pileri A, Tarella C (2002) High rate of clinical and molecular remissions in follicular lymphoma patients receiving high-dose sequential chemotherapy and autografting at diagnosis: a multicenter, prospective study by the Gruppo Italiano Trapianto Midollo Osseo (GITMO). Blood 100:1559–1565
16. Andersen NS, Donovan JW, Borus JS, Poor CM, Neuberg D, Aster JC, Nadler LM, Freedman AS, Gribben JG (1997) Failure of immunologic purging in mantle cell lymphoma assessed by polymerase chain reaction detection of minimal residual disease. Blood 90:4212–4221
17. van Krieken JHJM, Langerak AW, Macintyre EA, Kneba M, Hodges E, Sanz RG, Morgan GJ, Parreira A, Molina TJ, Cabeçadas J, Gaulard P, Jasani B, Garcia JF, Ott M, Hannsmann ML, Berger F, Hummel M, Davi F, Brüggemann M, Lavender FL, Schuuring E, Evans PAS, White H, Salles G, Groenen PJTA, Gameiro P, Pott C, Van Dongen JJM (2007) Improved reliability of lymphoma diagnostics via PCR-based clonality testing: report of the BIOMED-2 Concerted Action BHM4-CT98-3936. Leukemia 21:201–206
18. Payne K, Wright P, Grant JW, Huang Y, Hamoudi R, Bacon CM, Du MQ, Liu H (2011) BIOMED-2 PCR assays for IGK gene rearrangements are essential for B-cell clonality analysis in follicular lymphoma. Br J Haematol 155:84–92
19. Cavalli M, De Novi LA, Della I, Cappelli LV, Nunes V, Pulsoni A, Del Giudice I, Guarini A, Fo R (2017) Comparative analysis between RQ-PCR and digital droplet PCR of BCL2/IGH gene rearrangement in the peripheral blood and bone marrow of early stage follicular lymphoma. Br J Haematol 177:588–596
20. Drandi D, Kubiczkova-besse L, Ferrero S, Dani N, Passera R, Mantoan B, Gambella M, Monitillo L, Saraci E, Ghione P, Genuardi E, Barbero D, Omedè P, Barberio D, Hajek R, Vitolo U, Palumbo A, Cortelazzo S, Boccadoro M, Inghirami G, Ladetto M (2015) Minimal residual disease detection by droplet digital PCR in multiple myeloma, mantle cell lymphoma, and follicular lymphoma: a comparison with real-time PCR. J Mol Diagn 17:652–660
21. Evans PA, Pott C, Groenen PJ, Salles G, Davi F, Berger F, Garcia JF, van Krieken JH, Pals S, Kluin P, Schuuring E, Spaargaren M, Boone E, Gonzalez D, Martinez B, Villuendas R, Gameiro P, Diss TC, Mills K, Morgan GJ, Carter GI, Milner BJ, Pearson D, Hummel M, Jung W, Ott M, Canioni D, Beldjord K, Bastard C, Delfau-Larue MH, van Dongen JJ, Molina TJ, Cabecadas J (2007) Significantly improved PCR-based clonality testing in B-cell malignancies by use of multiple immunoglobulin gene targets. Report of the BIOMED-2 Concerted Action BHM4-CT98-3936. Leukemia 21:207–214

22. van Dongen JJ, Langerak AW, Brüggemann M, Evans PA, Hummel M, Lavender FL, Delabesse E, Davi F, Schuuring E, García-Sanz R, van Krieken JH, Droese J, González D, Bastard C, White HE, Spaargaren M, González M, Parreira A, Smith JL, Morgan GJ, Kneba M, Macintyre EA (2003) Design and standardization of PCR primers and protocols for detection of clonal immunoglobulin and T-cell receptor gene recombinations in suspect lymphoproliferations: report of the BIOMED-2 Concerted Action BMH4-CT98-3936. Leukemia 17:2257–2317
23. Van Der Velden V, Cazzaniga G, Schrauder A, Hancock J, Bader P, Panzer-Grumayer E, Flohr T, Sutton R (2007) Analysis of minimal residual disease by Ig/TCR gene rearrangements: guidelines for interpretation of real-time quantitative PCR data. Leukemia 21:604–611
24. Stamatopoulos K, Kosmas C, Belessi C, Stavroyianni N, Kyriazopoulos P, Papadaki T (2000) Molecular insights into the immunopathogenesis of follicular lymphoma. Immunol Today 21:298–305
25. Akasaka T, Yonetani N, Ohno H, Yamabe H, Fukuhara S, Okuma MAH (1998) Refinement of the BCL2/immunoglobulin heavy chain fusion gene in t(14;18)(q32;q21) by polymerase chain reaction amplification for long targets. Genes Chromosomes Cancer 21:17–29
26. Buchonnet G, Lenain P, Ruminy P, Lepretre S, Stamatoullas A, Parmentier F, Jardin F, Duval C, Tilly H, Bastard C (2000) Characterisation of BCL2-JH rearrangements in follicular lymphoma: PCR detection of 3′ BCL2 breakpoints and evidence of a new cluster. Leukemia 14:1563–1569
27. Buchonnet G, Jardin F, Jean N, Bertrand P, Parmentier F, Tison S, Lepretre S, Contentin N, Lenain P, Stamatoullas-Bastard-A, Tilly H, Bastard C (2002) Distribution of BCL2 breakpoints in follicular lymphoma and correlation with clinical features: specific subtypes or same disease? Leukemia 16:1852–1856
28. Rosenberg CL, Wong E, Petty EM, Bale AE, Tsujimoto Y, Harris NL, Arnold A (1991) PRAD1, a candidate BCL1 oncogene: mapping and expression in centrocytic lymphoma. Proc Natl Acad Sci USA 88:9638–9642
29. Kneba M, Eick S, Herbst H, Willigeroth S, Pott C, Bolz I, Bergholz M, Neumann C, Stein H, Krieger G (1991) Frequency and structure of t(14;18) major breakpoint regions in non-Hodgkin's lymphomas typed according to the Kiel classification: analysis by direct DNA sequencing. Cancer Res 51:3243–3250
30. Pott C, Tiemann M, Linke B, Ott MM, von Hofen M, Bolz I, Hiddemann W, Parwaresch R, Kneba M (1998) Structure of Bcl-1 and IgH-CDR3 rearrangements as clonal markers in mantle cell lymphomas. Leukemia 12:1630–1637
31. Williams ME, Meeker TC, Swerdlow SH (1991) Rearrangement of the chromosome 11 bcl-1 locus in centrocytic lymphoma: analysis with multiple breakpoint probes. Blood 78:493–498
32. Williams ME, Swerdlow SH, Rosenberg CL, Arnold A (1992) Characterization of chromosome 11 translocation breakpoints at the bcl-1 and PRAD1 loci in centrocytic lymphoma. Cancer Res 52:5541s–5544s
33. Ladetto M, Mantoan B, De Marco F, Drandi D, Aguzzi C, Astolfi M, Vallet S, Ricca I, Aquila MD, Pagliano G, Monitillo L, Pollio B, Santo L, Cristiano C, Rocci A, Francese R, Bodoni CL, Borchiellini A, Schinco P, Boccadoro M, Tarella C (2006) Cells carrying nonlymphoma-associated bcl-2/IgH rearrangements (NLABR) are phenotypically related to follicular lymphoma and can establish as long-term persisting clonal populations. Exp Hematol 34:1680–1686
34. van der Velden VHJ, van Dongen JJ (2009) MRD detection in acute lymphoblastic leukemia patients using Ig/TCR gene rearrangements as targets for real-time quantitative PCR In: Walker J (ed) Methods in Molecular Biology. Humana Press A Part of Springer Science and Media, New York, pp 115–150
35. van Dongen JJ, Langerak AW, Bruggemann M, Evans PA, Hummel M, Lavender FL, Delabesse E, Davi F, Schuuring E, Garcia-Sanz R, van Krieken JH, Droese J, Gonzalez D, Bastard C, White HE, Spaargaren M, Gonzalez M, Parreira A, Smith JL, Morgan GJ, Kneba M, Macintyre EA (2003) Design and standardization of PCR primers and protocols for detection of clonal immunoglobulin and T-cell receptor gene recombinations in suspect lymphoproliferations: report of the BIOMED-2 Concerted Action BMH4-CT98-3936. Leukemia 17:2257–2317
36. Pott C, Schrader C, Trautmann H, Irmer S, Desgranges C, Ritgen M, Siebert R, Harder S, Unterhalt M, Hiddemann W, Hansmann ML, Parwaresch MR, Kneba M (2003) VH mutational status and VH gene usage define molecular heterogeneity in mantle cell lymphoma. Onkologie 26:162

37. Gimenez E, Chauvet M, Rabin L, Puteaud I, Duley S, Hamaidia S, Bruder J, Rolland-Neyret V, Le Gouill S, Tournilhac O, Voog E, Maisonneuve H, Jacob MC, Leroux D, Béné MC, Formisano-Tréziny C, Gabert J, Gressin R, Callanan MB (2012) Cloned IGH VDJ targets as tools for personalized minimal residual disease monitoring in mature lymphoid malignancies; a feasibility study in mantle cell lymphoma by the Groupe Ouest Est d'Etude des Leucémies et Autres Maladies du Sang. Br J Haematol 158:186–197
38. Brüggemann M, Schrauder A, Raff T, Pfeifer H, Dworzak M, Ottmann OG, Asnafi V, Baruchel A, Bassan R, Benoit Y, Biondi A, Cavé H, Dombret H, Fielding AK, Foà R, Gökbuget N, Goldstone AH, Goulden N, Henze G, Hoelzer D, Janka-Schaub GE, Macintyre EA, Pieters R, Rambaldi A, Ribera J-MM, Schmiegelow K, Spinelli O, Stary J, von Stackelberg A, Kneba M, Schrappe M, van Dongen JJM, Bruggemann M, Schrauder A, Raff T, Pfeifer H, Dworzak M, Ottmann OG, Asnafi V, Baruchel A, Bassan R, Benoit Y, Biondi A, Cave H, Dombret H, Fielding AK, Foa R, Gokbuget N, Goldstone AH, Goulden N, Henze G, Hoelzer D, Janka-Schaub GE, Macintyre EA, Pieters R, Rambaldi A, Ribera J-MM, Schmiegelow K, Spinelli O, Stary J, Von SA, Kneba M, Schrappe M, van Dongen JJM (2010) Standardized MRD quantification in European ALL trials: proceedings of the second international symposium on MRD assessment in Kiel, Germany, 18–20 September 2008. Leukemia 24:521–535
39. Bruggemann M, Raff T, Flohr T, Gokbuget N, Nakao M, Droese J, Luschen S, Pott C, Ritgen M, Scheuring U, Horst HA, Thiel E, Hoelzer D, Bartram CR, Kneba M (2006) Clinical significance of minimal residual disease quantification in adult patients with standard-risk acute lymphoblastic leukemia. Blood 107:1116–1123
40. SantaLucia J (1998) A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. Proc Natl Acad Sci USA 95:1460–1465
41. Albinger-Hegyi A, Hochreutener B, Abdou M-TT, Hegyi I, Dours-Zimmermann MT, Kurrer MO, Heitz PU, Zimmermann DR (2002) High frequency of t(14;18)-translocation breakpoints outside of major breakpoint and minor cluster regions in follicular lymphomas: improved polymerase chain reaction protocols for their detection. Am J Pathol 160:823–832
42. Pongers-Willemse MJ, Verhagen OJ, Tibbe GJ, Wijkhuijs AJ, de Haas V, Roovers E, van der Schoot CE, van Dongen JJ (1998) Real-time quantitative PCR for the detection of minimal residual disease in acute lymphoblastic leukemia using junctional region specific TaqMan probes. Leukemia 12:2006–2014
43. van der Velden VHJ, Cazzaniga G, Schrauder A, Hancock J, Bader P, Panzer-Grumayer ER, Flohr T, Sutton R, Cave H, Madsen HO, Cayuela JM, Trka J, Eckert C, Foroni L, Zur Stadt U, Beldjord K, Raff T, van der Schoot CE, van Dongen JJM (2007) Analysis of minimal residual disease by Ig//TCR gene rearrangements: guidelines for interpretation of real-time quantitative PCR data. Leukemia 21:604–611
44. Summers KE, Davies AJ, Matthews J, Jenner MJ, Cornelius V, Amess JA, Norton AJ, Rohatiner AZ, Fitzgibbon J, Lister TA, Goff LK (2002) The relative role of peripheral blood and bone marrow for monitoring molecular evidence of disease in follicular lymphoma by quantitative real-time polymerase chain reaction. Br J Haematol 118:563–566
45. Pott C, Hoster E, Kehden B, Unterhalt M, Herold M, van der Jagt R, Janssens A, Kneba M, Mayer M, Pocock C, Danesi N, Fingerle-Rowson G, Harbron C, Mundt K, Marcus RE, Hiddemannnn W (2016) Minimal residual disease in patients with follicular lymphoma treated with Obinutuzumab or Rituximab as first-line induction immunochemotherapy and maintenance in the phase 3 GALLIUM study. Blood 128: abstract 623

Chapter 10

Generation of Whole Genome Bisulfite Sequencing Libraries from Very Low DNA Input

Dieter Weichenhan, Charles D. Imbusch, Qi Wang, Benedikt Brors, and Christoph Plass

Abstract

DNA methylation changes are dynamic processes which occur at cytosines of CpG dinucleotides and contribute to normal development but also to diseases. DNA methylation changes are most effective in promoters and enhancers, the former frequently being CpG-rich and the latter, in contrast, CpG-poor. Many genome-wide methods for DNA methylation analysis interrogate predominantly CpG-rich regions and, hence, spare enhancers and other potentially important genomic regions. Whole genome bisulfite sequencing (WGBS), in contrast, analyzes the DNA methylome in its entirety. Standard tagmentation-based whole genome bisulfite sequencing (TWGBS) is a Tn*5* transposon-based method which requires only 30 ng of human input DNA and, hence, is particularly suited for precious biological samples like cells sorted by flow cytometry or laser capture microdissected tissue specimens. In the standard version, tagmentation generates DNA fragments flanked by uniform sequencing adapters. In a subsequent step, the non-covalently bound adapter oligonucleotide needs to be replaced by a novel oligonucleotide to provide the proper adapter sequence for the reverse strand in paired-end sequencing. The presented protocol describes an improved, simplified version of TWGBS where the inefficient oligo-replacement is circumvented by usage of a sequencing-compatible transposase-adapter complex. Consequently, genomic DNA of only a few hundred human cells is required to interrogate the complete human DNA methylome.

Key words Epigenome, Whole genome bisulfite sequencing, DNA methylation, Tagmentation, Tn*5* transposase

1 Introduction

1.1 Global Determination of Genomic Methylation in Precious Biological Specimens

In human and other mammals, genomic DNA is largely methylated by reversible, post-replicative methylation of cytosines in CpG dinucleotides [1, 2]. DNA methylation plays important roles in embryonic development, differentiation, imprinting, X-chromosome inactivation, aging, and malignant diseases [3]. As a consequence of global DNA methylation changes, each developmental stage, age, tissue type, or tumor has its own specific DNA methylation pattern. Hence, knowing the specific correlation between methylation pattern and tissue or cell type provides

Ralf Küppers (ed.), *Lymphoma: Methods and Protocols*, Methods in Molecular Biology, vol. 1956,
https://doi.org/10.1007/978-1-4939-9151-8_10,

valuable information about the (patho)physiological state of the biological source. Historically, determination of DNA methylation was largely confined to only a small fraction of the DNA methylome, usually to regulatory regions like gene promoters which are frequently associated with the CpG-rich CpG islands. This confinement was not only due to the central role attributed to promoters in regulatory processes but also to the technical and financial limitations early investigators of DNA methylation were facing. The advent of affordable high-throughput technologies such as next-generation sequencing (NGS) and microarray analysis allowed the extension of a rather limited number of genomic target sequences to the complete genome and of rather small to large sample numbers, respectively. While promoters are still considered important in the regulation of the associated genes, further genomic elements, e.g., enhancers, super enhancers, locus control regions, and silencers, which are defined by chromatin configurations and their association with modified histones and transcription factors, were recognized as central in the complex regulation of global cellular processes taking place from early embryogenesis to late adulthood. Mutations but also aberrant methylation of such elements lead to their dysfunction and to malignancies [4]. As a conclusion, methods which interrogate only a minor portion of the DNA methylome may miss genomic regions essential for the physiological state of the cells or tissue of interest.

Many approaches to analyze the DNA methylome include a bisulfite conversion step during which unmethylated cytosines are converted to uracil, while methylated cytosines remain as cytosines. Subsequent PCR manifests unmethylated cytosines as thymines, thus, allowing discrimination between the unmethylated and methylated cytosines. The gold standard to address the complete DNA methylome is whole genome bisulfite sequencing (WGBS) [5]. For sequencing library preparation, WGBS comprises diverse steps of genomic DNA manipulation, including sonication, end repair, A-tailing, adapter ligation, size selection, bisulfite conversion, and PCR. Together with intermediate purification steps, these manipulations are cumbersome and prone to DNA loss. As a consequence, WGBS of a human specimen usually requires at least two μg of input DNA which is equivalent to about 660,000 cells assuming a yield of 50% in DNA isolation. During the past few years, several alternative approaches requiring much less input DNA for WGBS were introduced. Post-bisulfite adapter tagging (PBAT) works with only about 100 ng human DNA and uses 5′-tagged random primers to linearly amplify bisulfite-treated genomic DNA in its initial step, followed by a limited number of exponential PCR cycles to generate the sequencing library [6]. Bisulfite treatment and random priming generate sufficiently small fragments as substrates for library PCR. PBAT even proved suitable to analyze the DNA methylome of single cells [7]. Both the PBAT and a similar

approach which starts with bisulfite treatment as well are offered as commercial kits for the generation of WGBS libraries from a small amount of human input DNA in the 100 ng range (https://www.zymoresearch.com/downloads/dl/file/id/628/d5455i.pdf; https://swiftbiosci.com/products/accel-ngs-methyl-seq-dna-library-kit/accel-ngs-methyl-seq-dna-library-kit/#exp-workflow). Standard tagmentation-based WGBS (TWGBS) uses a transposome, which consists of a hyperactive Tn*5* transposase complexed with a uniform, partially methylated, double-stranded oligonucleotide adapter, for initial fragmentation and simultaneous adapter tagging of as low as 30 ng human input DNA [8, 9] (Fig. 1a). Upon tagmentation, only the methylated adapter oligonucleotide is covalently linked to the genomic DNA. The unlinked, unmethylated oligonucleotide is separated from the genomic DNA by a gap of nine bases and needs to be replaced by a second methylated oligonucleotide which is covalently linked to the fragmented DNA by gap repair and ligation. By this means, the DNA fragments are flanked by appropriate adapters for paired-end sequencing. Subsequent bisulfite treatment is finally followed by library PCR under real-time conditions which enables monitoring of the amplification process, hence, avoiding under- or overamplification which would lead to insufficient DNA for sequencing or high PCR duplication frequencies, respectively. The PCR primers match the adapter sequences and are compatible with sequencing primers from Illumina (TruSeq Dual Index Sequencing Primer Box, Paired End). PCR-generated libraries are analyzed by 125 bp paired-end NGS, e.g., on an Illumina HiSeq 2000 platform. The improved version of TWGBS described here is even more efficient and reduces the required input DNA amount to that from only a few hundred cells.

1.2 Applications of Tagmentation and Principle of TWGBS for Very Low-Input Amounts

At the latest with the first presentation of the assay for transposase-accessible chromatin sequencing (ATAC-seq) [10], the principle of tagmentation became very popular as a means to readily and efficiently prepare libraries for NGS. While ATAC-seq is used to map open regions of the native chromatin, ChIPmentation, another variant of the tagmentation tool box, uses cross-linked chromatin in combination with antibodies to map the binding sites of chromatin-associated proteins like histone variants or transcription factors [11]. Initially, however, tagmentation with a transposome, consisting of a hyperactive Tn*5* transposase complexed with double-stranded oligonucleotide adapters, had been developed for NGS of low genomic DNA inputs [12]. Subsequent variants applied tagmentation even in single-cell RNA and DNA sequencing [13, 14].

The need for whole genome methylation analysis of precious DNA samples prompted the development of TWGBS [8]. TWGBS was readily established in our lab and soon proved indispensable in diverse epigenetic studies of rare cells sorted by flow cytometry

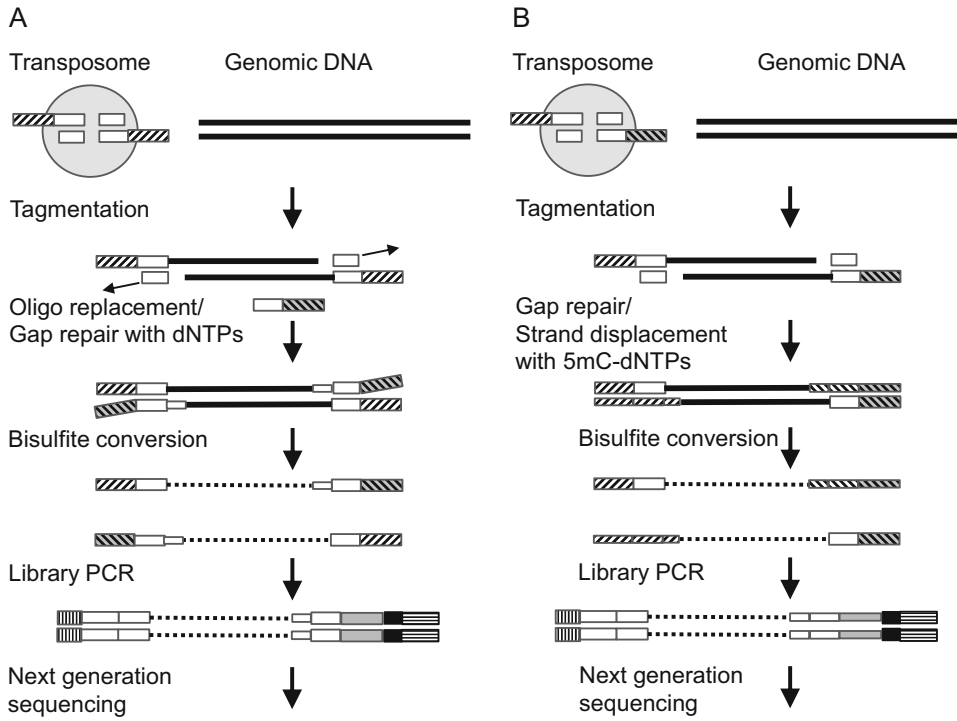


Fig. 1 Standard tagmentation-based (**a**) and improved tagmentation-based (**b**) library preparation for whole genome bisulfite sequencing. (**a**) Standard TWGBS uses a transposome (gray circle) generated from a Tn*5* transposase dimer complexed with two identical load adapters. These consist of two partially methylated oligonucleotide adapters containing the sequencing primer 1 (white fasciated box) and the transposase recognition sequence (white box). After tagmentation, the unmethylated oligonucleotides are replaced with methylated oligonucleotides which are at their 3′ end (gray fasciated box) complementary to the sequencing primer 2 but not complementary to the covalently bound top strand of the load adapter; the remaining gap is repaired with dNTPs. (**b**) For very low-input library preparation, a transposome with two different load adapters is used which already contain the sequence information for sequencing primer 1 and 2 (white and gray fasciated box, respectively). After tagmentation, the gaps are repaired and the unmethylated oligonucleotides removed by strand displacement using 5mC-dNTPs. In both (**a**) and (**b**), the tagmented and repaired DNA fragments (black lines) are flanked by methylated adapters. The final two steps, bisulfite conversion and library PCR, are identical in both approaches. Bisulfite treatment leads to conversion of unmethylated cytosines to uracil in the genomic sequence (interrupted lines) and in the repaired gaps in (**a**) but not in (**b**). Sequencing libraries are generated under real-time PCR conditions using a primer bearing the flow cell primer 1 (left vertically striated box) and the sequencing primer 1 sequences and a second primer bearing the flow cell primer 2 (right horizontally striated box), a barcode (black box), and the sequencing primer 2 sequences

[15–17]. However, although TWGBS routinely works well with the DNA from about 5000 human cells, equivalent to 30 ng, it did not work efficiently enough in our and the hands of others [18] with input DNA from less than 2000 cells. The intricate step of combined oligonucleotide replacement, gap repair, and ligation in the standard TWGBS protocol (*see* Fig. 1a) proved a bottleneck, in addition to the considerable loss of input DNA through the bisulfite treatment [9, 18]. The improved TWGBS protocol presented here avoids oligonucleotide replacement after tagmentation by

usage of a transposome containing two different partially methylated, double-stranded adapters instead of two identical adapters [18] (Fig. 1b). Efficient end filling by strand extension and displacement reduces the required DNA amount to that from only a few hundred human cells.

2 Materials (*See* Note 1)

2.1 Preparation of the Load Adapters and Assembly of the Transposome

1. Oligonucleotides Tn5mC-Apt1, Tn5mC-Apt2, and Tn5mC1.1-a1block (*see* Table 1 and **Note 2**).
2. Sterile deionized H_2O and sterile 100% glycerol.
3. 1.5 mL reaction tube and eight-well strip with lid.
4. 96-well gradient thermocycler.
5. Heater block.
6. EZ-Tn*5* transposase (Lucigen).

2.2 Tagmentation

1. 2–10 ng genomic DNA (*see* **Note 3**).
2. 5× buffer for tagmentation (5×TMA-DMF; *see* **Note 4**).
3. Unmethylated phage λ DNA.
4. 1.5 mL reaction tube and eight-well strip with lid.
5. 96-well thermocycler.

Table 1
Oligonucleotides

Name	Purpose	Sequence (5′ to 3′)	Remark
Tn5mC-Apt1[a]	Load adapter 1	TcGTcGGcAGcGTcAGATGTGTAT-	c: 5C-methylated
		AAGAGAcAG	
Tn5mC-Apt2[a]	Load adapter 2	GTcTcGTGGGcTcGGAGATGTGT-	c: 5C-methylated
		ATAAGAGAcAG	
Tn5mC1.1A1block	Load adapter	pCTGTCTCTTATACAddC	p, phosphate;
			dd, dideoxy
Tn5mCP1	Library PCR	AATGATACGGCGACCACCGAG-	
		ATCTACACTCGTCGGCAGCGTC	
Tn5mCBar[b]	Library PCR	CAAGCAGAAGACGGCATACGA-	Barcode of 8 or 9
		GAT(8-9N)GTCTCGTGGGCTCGG	Bases

[a]The transposon end recognition sequence is underlined
[b]Barcode sequences: (1) GGATGTTCT, (2) CTTATCCAG, (3) GTAAGTCAC, (4) TTCAGTGAG, (5) CTCGTAATG, (6) CATGTCTCA, (7) AATCGTGGA, (8) GTATCAGTC, (9) TCGCCTTA, (10) CTAGTACG, (11) TTCTGCCT, (12) GCTCAGGA, (13) AGGAGTCC, (14) CATGCCTA, (15) GTAGAGAG, (16) CCTCTCTG, (17) AGCGTAGC, (18) AGCCTCG, (19) TGCCTCTT, (20) TCCTCTAC

2.3 Post Tagmentation DNA Purification

1. DNA purification magnetic beads.
2. 5 M guanidinium thiocyanate; prepare with sterile deionized H_2O and store aliquoted at −20 °C.
3. AMPure buffer (*see* **Note 5**).
4. Sterile deionized H_2O and 80% ethanol.
5. Eight-well strip with lid.
6. Magnetic separator for eight-well strips.

2.4 Gap Repair and Strand Displacement

1. 5mC-deoxynucleotide mix (100 mM of each dNTP). Dilute 1:40 in H_2O to have a final concentration of 2.5 mM of each dNTP.
2. BstI DNA polymerase (large fragment) and 10× reaction buffer.
3. 96-well thermocycler.

2.5 Bisulfite Conversion and Purification of DNA

1. EZ DNA Methylation™ Kit (ZYMO Research; *see* **Note 6**).
2. LoBind reaction tubes and eight-well strip with lid.
3. 96-well thermocycler.
4. Benchtop centrifuge.

2.6 Preparative Real-Time PCR

1. Oligonucleotide primers Tn5mCP1 and Tn5mCBar (*see* Table 1 and **Note 2**).
2. KAPA 2G Robust Hot Start Ready Mix.
3. 10,000× SYBR® Green I nucleic acid gel stain. 100× SYBR-Green: dilute 1 μL 10,000× SYBR® Green I nucleic acid gel stain with 99 μL H_2O. The 100× solution can be stored protected from light at −20 °C for at least 6 months.
4. 96-well real-time PCR plate.
5. Real-time thermocycler.

2.7 Purification of TWGBS Library and Quality Control

1. Refer to Subheading 2.3 for material.
2. EB elution buffer (Qiagen).
3. Qubit dsDNA HS assay kit.
4. Qubit 2.0 fluorometer.
5. Qubit assay tubes.
6. Quant-iT dsDNA HS Assay Kit.
7. Bioanalyzer 2100.

2.8 Next-Generation Sequencing of the TWGBS Libraries

1. TruSeq Dual Index Sequencing Primer Box, Paired End (Illumina).
2. Illumina HiSeq2000 Sequencer.

3 Methods

3.1 Preparation of Load Adapters 1 and 2 and Assembly of the Transposome

1. For load adapter 1, mix in a well of an eight-well strip 10 μL of oligonucleotide Tn5mC1.1-A1block with 10 μL of Tn5mC-Apt1; for load adapter 2, mix in a second well 10 μL of Tn5mC1.1-A1block with 10 μL of Tn5mC-Apt2 (initial concentration of each oligonucleotide is 200 μM; *see* Table 1). To each mix, add 20 μL H_2O. The final concentration of each oligonucleotide is 50 μM.
2. Carry out adapter assembly in a thermocycler with the following conditions:

Cycle number	Denature	Anneal	Ramp to 26 °C	Hold
1	95 °C, 3 min	70 °C, 3 min		
2–46		70 °C, 30 s	−1 °C/cycle, 30 s	
47				25 °C, infinite

3. Heat about 150 μL 100% glycerol to 90 °C in a heater block to enable exact pipetting of the hot glycerol.
4. Transfer 100 μL hot glycerol to a 1.5 mL reaction tube, and cool down to room temperature (RT).
5. Add 20 μL H_2O and each 40 μL of load adapter 1 and load adapter 2 from **step 2**, and mix by repeated pipetting. This adapter mix is designated Tn5ME-Apt1 + Apt2 (stable at −20 °C for at least 6 months).
6. To assemble the transposome, transfer 10 μL Tn5ME-Apt1 + Apt2 to 10 μL Ez-Tn*5* transposase, and mix by repeated pipetting; avoid air bubbles.
7. Maintain the mixture at RT for 30 min during which the transposome forms. The transposome mixture is sufficient for about 13 tagmentation reactions and stable at −20 °C for at least 6 months.

3.2 Tagmentation (See Note 7 and Fig. 2)

1. For a single tagmentation reaction, mix on ice in an eight-well strip 4 μL 5× TMA-DMF buffer and 14.5 μL DNA solution containing 2–10 ng genomic DNA and 1–5 pg unmethylated phage λ DNA (5 pg λ DNA/10 ng genomic DNA). Add 1.5 μL transposome (from Subheading 3.1, **step 7**), and mix by repeated pipetting; avoid air bubbles (*see* **Note 8**).
2. Perform the tagmentation reaction in a thermocycler as follows:

Cycle number	Tagmentation	Cool
1	55 °C, 8 min	4 °C, infinite

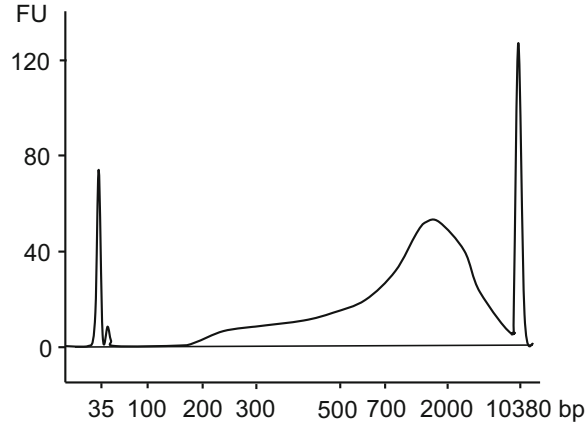


Fig. 2 Transposome quality control by fragment size analysis after tagmentation. Two microliter transposome and 30 ng genomic DNA were used in a 20 μL tagmentation test reaction. The size profile ranges from about 200 bp, an indication for good transposome quality, to smaller than 10 kb. The extent of the profile beyond 1 kb is of minor relevance for the assessment of transposome quality. A size profile starting at 500 bp or larger would indicate an inappropriately active transposome. The sharp peaks at 35 bp and 10,380 bp indicate lower and upper size markers, respectively, which were added to the sample. The small peak close to the right of the lower size marker is derived from unused adapters. *FU* fluorescence units

3.3 Post Tagmentation DNA Purification

1. Add to each tagmentation reaction mix in an eight-well strip 15 μL 5 M guanidinium thiocyanate (DNA solution has then a total volume of 35 μL), 10 μL DNA purification magnetic beads, and 36 μL AMPure buffer (total volume of the bead solution is then 46 μL; bead/DNA volume ratio is $46/35 \approx 1.3$), and homogenize by repeated pipetting.
2. Keep the eight-well strip at RT for 10 min.
3. Transfer the strip to the magnetic separator, and wait until the bead suspensions have been completely cleared (about 1 min). Remove the clear supernatants thoroughly without disturbing the bead pellets, and discard the supernatants.
4. Still keeping the strip in the magnetic separator, add 75 μL of 80% ethanol to each bead pellet. Wash the beads with the 80% ethanol by repeated (five to ten times) pipetting without touching the beads with the pipette tips, then remove and discard the wash solution.
5. Keep the strip without lid in the magnetic separator for about 10 min at RT until the beads are completely dry.
6. Remove the strip from the magnet, and resuspend each bead pellet thoroughly in 16 μL sterile deionized H_2O.

7. Transfer the strip back to the magnetic separator, and wait until the bead suspensions have been completely cleared (about 1 min).
8. Transfer the clear supernatants containing the purified tagmented DNA carefully to a new eight-well strip; avoid bead carryover.
9. For quality control of tagmentation reactions with 10 ng input DNA, analyze 1 μL of the supernatant using the Bioanalyzer 2100.
10. If the DNA has been efficiently fragmented (*see* Fig. 2), proceed to the next step or freeze the DNA at −20 °C.

3.4 Gap Repair and Strand Displacement

1. To each well containing the 15–16 μL supernatant with the tagmented DNA, add 2 μL of 5mC-deoxynucleotide mix dNTP (each dNTP 2.5 mM), 2 μL of 10× BstI reaction buffer, and 1 μL of BstI DNA polymerase (large fragment; *see* **Note 9**).
2. Incubate in a thermocycler.

Cycle number	Gap repair/strand displacement	Hold
1	65 °C, 20 min	4 °C, infinite

3.5 DNA Purification After Gap Repair and Strand Displacement

1. Add to each reaction mix 10 μL DNA purification magnetic beads and 26 μL AMPure buffer (bead/DNA volume ratio is 36/20 = 1.8), and homogenize by repeated pipetting.
2. Repeat **steps 2–8** from Subheading 3.3, but resuspend the beads in 50 μL (instead of 16 μL) of sterile deionized H_2O.
3. Transfer 5 μL of each 50 μL supernatant to a second, new eight-well strip, and keep at −20 °C for later control in the final PCR (*see* **Note 10**).
4. Either proceed to the bisulfite treatment or store the supernatants at −20 °C.

3.6 Bisulfite Conversion and Purification of the DNA

1. Add to the rest of each supernatant (45 μL) 5 μL M-dilution buffer, and mix by repeated pipetting.
2. Incubate in a thermocycler.

Cycle number	Gap repair	Hold
1	37 °C, 15 min	25 °C, infinite

3. Add to each reaction mixture 100 μL CT conversion reagent and mix by repeated pipetting; distribute mixtures in 75 μL aliquots to each two wells of an eight-well strip and incubate overnight in a thermocycler.

Cycle number	Denature	Conversion	Hold
1–16	95 °C, 15 s	50 °C, 1 h	
17			4 °C, infinite

4. For each sample, transfer 400 μL M-binding buffer to a Zymo-Spin™ IC Column inserted in a collection tube and load the 2 × 75 μL of each sample into the Zymo-Spin™ IC Column containing the M-binding buffer. Close the column lid, and invert the column-tube assembly several times. Centrifuge the assembly with 11,000 × *g* for 30 s, and discard the flow-through.
5. Add 100 μL M-wash buffer to each column, and centrifuge with 13,000 × *g* for 30 s; discard the flow-through.
6. Add 200 μL M-desulfonation buffer to each column, and keep the assembly at RT for 15–20 min. Subsequently centrifuge with 13,000 × *g* for 30 s, and discard the flow-through.
7. Add 200 μL M-wash buffer to each column, and centrifuge with 13,000 × *g* for 30 s; discard the flow-through. Repeat washing once, but extend centrifugation with 13,000 × *g* to 3 min.
8. Transfer the column to a LoBind reaction tube and pipette 11.5 μL M-elution buffer onto each column matrix. Centrifuge with 13,000 × *g* for 30 s to elute the converted DNA. Repeat this elution step (*see* **Note 11**).
9. Proceed to the preparative real-time PCR or store the eluate at −80 °C.

3.7 Preparative Real-Time PCR

1. For each tagmented, bisulfite converted or unconverted sample, prepare in a 96-well real-time PCR plate the following mixture on ice (*see* **Note 12**):

Component	Amount (μL)	Final
Kapa 2G Robust HotStart ReadyMix (2×)	12.5	1×
Primer Tn5mCP1n (10 μM; *see* Table 1)	0.75	0.3 μM
Primer Tn5mCBar (10 μM; *see* Table 1)	0.75	0.3 μM
100× SYBRGreen	0.25	1×
DNA	10.75	
Total	25	

2. Incubate in a real-time thermocycler using the following conditions:

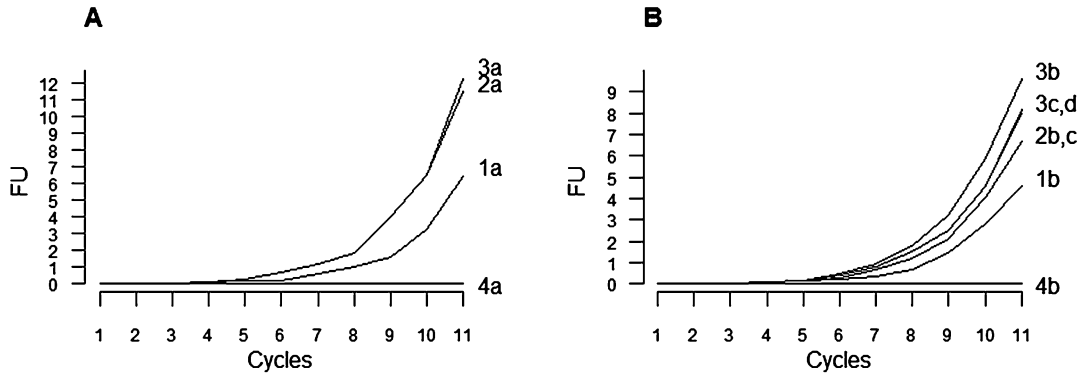


Fig. 3 PCR amplification curves of TWGBS libraries generated from 3, 10, and 30 ng tagmented human input DNA (1, 2, and 3, respectively) and a tagmented mock sample 4. (**a**) Generation of the first libraries, "a," in a series of up to 4, was stopped at cycle 11. The curves of samples 2 and 3 with 10 ng and 30 ng input, respectively, indicate almost identical amplification efficiency. Appropriate amplification is indicated by an increase in the fluorescence intensity between cycles 4 and 8. (**b**) Generation of the additional libraries "b," "c," and "d" was also stopped at cycle 11. From sample 1 with 3 ng input DNA, only one more library was generated, from sample 2 with 10 ng input DNA, two more libraries, and from sample 3 with 30 ng input DNA, three more libraries were generated. Libraries 2b and 2c and libraries 3c and 3d showed identical amplification, respectively. *FU* fluorescence units

Cycle number	Denature	Annealing	Read	Extension
1	95 °C, 3 min			
1–20	95 °C, 20 s	62 °C, 15 s	72, 30 s	72, 10 s

3. Stop the PCR at the latest at cycle 15 to avoid inconveniently high PCR duplication frequencies (Fig. 3a, b).
4. Prepare up to three additional TWGBS libraries, depending on the input DNA amount (*see* **Notes 11** and **12**).

3.8 Library Purification and Quality Control (See Note 13)

1. Transfer the PCR reaction mixes to an eight-well strip, and add to each mix 10 μL DNA purification magnetic beads plus 35 μL AMPure buffer (bead/DNA volume ratio is 45/25 = 1.8); homogenize by repeated pipetting.
2. Repeat **steps 2–7** from Subheading 3.3, but resuspend the beads in 12 μL EB elution buffer.
3. Transfer the clear supernatants corresponding to the TWGBS libraries carefully to 1.5 mL reaction tubes; avoid bead carryover.
4. Transfer 1 μL of each library to a LoBind reaction tube, and dissolve with 3 μL H_2O (1:4) for the subsequent determination of the fragment size distributions with the Bioanalyzer and of the DNA concentrations by Qubit fluorimetry (Fig. 4a, b and Table 2; *see* **Note 14**).

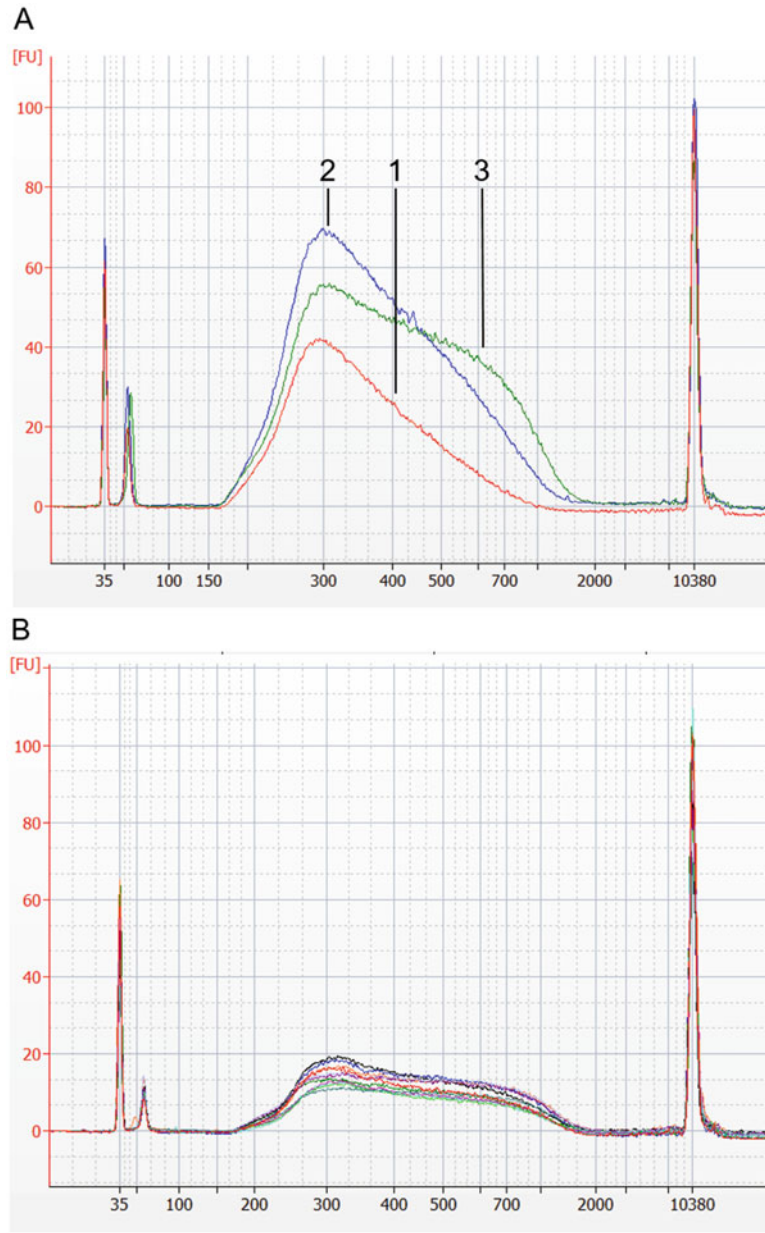


Fig. 4 Fragment size ranges of TWGBS libraries. (**a**) Fragment size ranges of libraries "a" generated from sample 1, 2, and 3 corresponding to 3 ng, 10 ng, and 30 ng of human input DNA, respectively (refer to Fig. 3a). All libraries peak at around 300 bp. While the sample 1 library ranges from 200 to 700 bp, those of samples 2 and 3 both range from 200 to about 1000 bp. Additional libraries from the same sample and tagmentation reaction have identical fragment size characteristics. (**b**) Fragment size ranges of exemplary TWGBS libraries from 11 different DNA samples isolated from flow cytometry-sorted human cells. Each tagmentation reaction was performed with about 10 ng input DNA and 1 μL

3.9 Next-Generation Sequencing of the TWGBS Libraries and Bioinformatics Analysis

1. Sequence the pool of up to four differently barcoded libraries of a tagmented sample in two lanes of a 125 bp paired-end Illumina HiSeq2000 sequencing run, using the appropriate sequencing primers (TruSeq Dual Index Sequencing Primer Box, Paired End) compatible with the TWGBS library primers Tn5mCP1 and Tn5mCBar (*see* Table 1 and **Note 15**).
2. Bioinformatics analysis
   (a) Preprocessing: (i) Trim adaptor sequences from the reads using SeqPrep (https://github.com/jstjohn/SeqPrep). (ii) In order to process nondirectional data with aligners that require directional data (i.e., only reads from the original strand), initially determine whether read pairs originated from the original strand in the genome: through bisulfite conversion, cytosine (C) converted to thymine (T) in read 1 (R1) and guanine (G) converted to

Table 2
TWGBS library characteristics

Library	ng/μL	ng total[a]	Peak bp[b]	Range bp	nM[c]	fmol	fmol total	FU[d]
1a	3.4	34	300	200–700	17.4	174		6.6
1b	2.9	29	300	200–700	14.9	149	323	5
2a	5.8	58	300	200–1000	29.7	297		11.4
2b	3.8	38	300	200–1000	19.5	195		6.6
2c	4.2	42	300	200–1000	21.5	215	707	6.6
3a	5.2	52	300	200–1000	26.7	267		11.4
3b	4.9	49	300	200–1000	25.1	251		9.6
3c	4.1	41	300	200–1000	21.0	210		8.2
3d	4.5	45	300	200–1000	23.1	231	959	8.2

[a]Final volume of each library is 10 μL
[b]Refer to Fig. 4a
[c]Fragment concentration calculated from the concentration c in ng/μL and the fragment size peak fs in bp and the average molecular weight 650 of a base pair by $c/(fs \times 650 \times 10^{-6})$
[d]Refer to Fig. 3a, b

Fig. 4 (continued) transposome. Per sample and tagmentation reaction, four differently barcoded TWGBS libraries, compatible to be analyzed as a pool on a single HiSeq2000 lane, were generated with 12 PCR cycles. All libraries range from 200 to 1000 bp, and most show a slight peak at around 300 bp. The sharp peaks at 35 bp and 10,380 bp in (**a**) and (**b**) indicate lower and upper size markers, respectively, which were added to the samples. The additional peaks close to the right of the lower size marker are derived from unincorporated PCR primers. *FU* fluorescence units

adenine (A) in read 2 (R2), or the reverse complementary, namely, G converted to A in R1 and C converted to T in R2. Since most Cs in non-CpG context are unmethylated, thus converted to Ts upon bisulfite treatment, expect more Ts than Cs on the original strand compared to the reverse complementary strand, which should have more As than Gs. Accordingly, do the following before the alignment: first mask CpG, TpG, GpC, and GpT dinucleotides, because they are potential CpG sites with mainly methylated Cs which are not converted by bisulfite treatment and, thus, uninformative. For each read in a read pair, calculate the base ratios T/C and A/G, denoted as R1T/C, R1A/G and R2T/C, R2A/G. Because of the gap repair step (*see* **Note 16**), the first and the last nine base pairs of both R1 and R2 are excluded, when calculating the base ratios T/C and A/G. Then compare these ratios between the first and second read within a read pair to determine if the read pair should be assigned the original R1–R2 order or the reverted R2–R1 order. The following rules are applied: for read pairs for which the condition R1T/C > R2T/C and R2A/G > R1A/G holds true, assume the read pair comes from the original strand and keep the original read order (R1–R2); for read pairs for which the conditions R1T/C < R2T/C and R2A/G < R1A/G are met, assume the read pair comes from the reverse complementary strand and the reverse read labels R2–R1 are assigned. Remove from downstream analysis ambiguous read pairs that meet neither of these conditions. An example implementation of part (ii) of the preprocessing can be found here (https://github.com/cimbusch/TWGBS).

(b) Alignment: Bisulfite-convert the reference genome in silico to create two reference genomes. One genome has all Cs converted to Ts, and the other has all Gs converted to As. Combine the two converted genomes into one reference genome. Then bisulfite convert the reads in silico: in each read pair, C to T in R1 and G to A in R2. Align the converted reads to the combined reference genome using BWA (bwa-0.6.2-tpx) [19] with default parameters, but disable the quality threshold for read trimming (-q) of 20 and the Smith-Waterman for the unmapped mate (-s). Reconvert the reads after the alignment to their original sequences and remove R1s that are mapped to the G to A converted genome and R2s that are mapped to the C to T converted genome. Mark duplicate reads and determine library complexity using Picard MarkDuplicates (http://broadinstitute.github.io/picard/). Filter out reads with

alignment scores of less than 1, i.e., reads which mapped to multiple locations.

(c) Methylation calling: To determine the methylation state of each C position in the reads, consider a C that is consistent with the reference genome as methylated and a C converted to T as unmethylated, taking only bases with a Phred-scaled quality score of $\geq$20 into account, while excluding the first nine bases of R2 and the last nine bases before the adapter of R1 using the assigned R1/R2 labels from the preprocessing step (*see* **Note 16**). In addition, when R1 and R2 in a read pair overlap, only information from R1 is considered in the overlapping segment in order to avoid scoring the same segment twice.

4 Notes

1. Use only filter pipetting tips throughout the whole procedure of library generation. As a contamination control, use a DNA-free H_2O sample in parallel to the DNA samples throughout the whole process. Multichannel pipettes are recommended, if several samples are processed in parallel.
2. All oligonucleotides (*see* Table 1) were custom-made. Some oligonucleotides contain methylated carbon 5 atoms in the cytosine bases to prevent cytosine to uracil conversion through the bisulfite treatment. The 5′ phosphate of the load adapter oligonucleotide Tn5mC1.1-A1block may not be required; its 3′ dideoxy cytidylate prevents unwanted extension during the PCR if it has not been completely removed before. The TWGBS library primers Tn5mCP1 and Tn5mCBar are compatible with the sequencing primers of the TruSeq Dual Index Sequencing Primer Box, Paired End from Illumina.
3. Recommended is the use of a commercial kit, e.g., the QIAmp DNA mini or micro kit (Qiagen) for the isolation of genomic DNA. If the expected quantity allows, concentration determination should be performed with a Qubit fluorometer and the Quant-iT dsDNA HS Assay Kit.
4. 5× tagmentation buffer (5× TMA-DMF) is 50 mM Tris (hydroxymethyl)aminomethane, 25 mM $MgCl_2$, 50% (vol/-vol) dimethylformamide [20]. Adjust the pH to 7.6 with 100% acetic acid before the addition of DMF. The buffer is filter sterilized and can be stored at −20 °C for at least 6 months.
5. AMPure buffer is 2.5 M NaCl, 20% polyethylene glycol 8000. The solution is filter sterilized and can be stored at 4 °C for at least 6 months.

6. The EZ DNA Methylation™ Kit (ZYMO Research) contains the CT conversion reagent as a solid substance. Immediately before use of the reagent, add 750 µL of H_2O and 210 µL of M-dilution buffer, also a component of the EZ DNA methylation kit, and mix by frequent vortexing for 10 min. The dissolved CT reagent can be stored protected from light at −20 °C for up to 1 month. As a further component of the EZ DNA methylation kit, the M-wash buffer is supplied without ethanol and must be completed by the addition of 96 mL of 100% ethanol to 24 mL of buffer. The buffer mixture can be stored at RT for at least 1 month.
7. Since the activity of a transposome varies from batch to batch, each newly prepared transposome must be quality checked under standard conditions, prior to tagmentation of precious DNA samples. Test tagmentation in a 20 µL reaction mix (*see* Subheading 3.2) is done with 2 µL transposome and an amount of genomic test DNA, e.g., 30 ng, suitable for size assessment after tagmentation. The DNA is subsequently purified (*see* Subheading 3.3), and 1 µL of the eluate is subjected to Bioanalyzer analysis (Fig. 2). Efficient fragmentation by a transposome of good quality is indicated by a size profile ranging from about 200 bp to less than 10 kb. Inappropriate transposome quality is indicated by a profile starting at 500 bp or larger.
8. The reaction mix with 1.5 µL transposome described in Subheading 3.2 is suitable for an input amount of 10 ng human DNA. To avoid too small fragments with DNA input amounts as low as 2 ng, the volume of the transposome may be lowered to 1 µL or even less. If the volume of the DNA solution exceeds 14.5 µL, the tagmentation reaction volume can be adjusted accordingly, e.g., to 40 µL, without changing the volume of the transposome. In the subsequent purification step (*see* Subheading 3.3), the volume of 5 M guanidinium thiocyanate may be increased to 20 µL, and the volume of the bead/AMPure buffer mix adjusted to a final ratio of ~1.3, for example, to 80 µL, if the volume of the tagmentation mix plus guanidinium thiocyanate sums up to 60 µL. Unmethylated λ DNA is used to control the conversion frequency of unmethylated cytosines to uracils in the bioinformatics analysis (*see* Subheading 3.9 and **Note 16**). Using multichannel pipettes in subsequent purification steps, up to eight samples (seven DNA samples and a H_2O control) can readily be processed in parallel.
9. If multiple samples are processed in parallel, the gap repair/strand displacement step can be simplified by preparation of a reaction master mix. For example, if eight samples are processed at a time, a master mix of each 17 µL of 5mC-dNTPs and 10× ThermoPol buffer plus 8.5 µL BstI DNA polymerase

(large fragment) can be prepared and distributed in 5 μL aliquots to each sample.

10. The set-aside 5 μL aliquots of the 50 μL supernatants serve as a positive control in the final library PCR to monitor if the bisulfite treatment of the DNA in the remaining 45 μL may cause severe DNA loss and, hence, failure in PCR amplification. In addition, the control aliquots may also be used to generate sequencing libraries for genomic profiling such as screening for copy number and single nucleotide variants [21].
11. The total volume used for elution may range from 23 μL to 46 μL depending on the initial amount of input DNA for tagmentation. A 10 ng DNA input allows for the generation of four independent PCR libraries requiring 44 μL eluate in total, while an input of 2 ng DNA may be good for two libraries requiring 22 μL eluate.
12. Preparation of a PCR master mix without DNA templates and barcode primers Tn5mCBar is recommended. Different barcode primers may be used for different template DNAs and for additional libraries from the same template DNA. Non-tagmented genomic DNA and the mock sample without DNA processed in parallel to the DNA samples serve as negative control, the tagmented but not bisulfite-treated aliquot serves as positive control (*see* **Note 10**). If more than a single library is generated per sample, it is recommended to generate the PCR libraries in at least two different PCR runs to allow compensation of possible failure of the first run due to experimental error. As long as the bisulfite-treated samples exhibit amplification curves as shown in Fig. 3a, b, the loss of DNA by the bisulfite treatment should not compromise successful TWGBS library generation. However, if no or substantially delayed amplification of the bisulfite-treated sample is observed, the time of treatment may be shortened.
13. To avoid contamination of any equipment and solutions used for tagmentation and subsequent steps by the final PCR amplification products, post-PCR purification and subsequent quality and quantity control should be performed at a different working place with extra pipettes, DNA purification magnetic beads, magnetic separator, etc.
14. TWGBS library size range and concentration are used to assess proper composition of pools from differently barcoded libraries (Table 2). Ideally, library fragments in pools should be present in equimolar quantities such that each library is equally represented. For example, if a pool with a final volume of 30 μL consists of four libraries and should have a fragment concentration of 10 nM, each library should contribute 75 fmol fragments which sum up to 300 fmol in 30 μL which is equivalent to 10 nM. Since paired-end sequencing on a

HiSeq2000 platform usually requires less library DNA, a pool concentration of 2 nM might even be sufficient.

15. If four TWGBS libraries are analyzed per lane of a HiSeq2000 paired-end run, barcode combinations 6, 16, 19, and 20 or 8, 11, 17, and 18 (*see* Table 1) are recommended. Many other barcode combinations are possible with the barcode primers presented in Table 1, depending on the number of differently barcoded samples in a library pool. For more information on barcode combinations, refer to [22]. To enable an in depth statistical methylation analysis at the single base level, a 10- to 20-fold genomic coverage (5- to 10-fold per each Watson and Crick strand) is recommended [23]. To acquire such a sequencing depth, analysis of the pools on two HiSeq2000 lanes is usually required. High frequencies of sequencing read duplicates (>20%) within the libraries of a tagmented sample may be prohibitive to reach this goal and, hence, may require generation of additional TWGBS libraries. Too high duplication frequencies can be avoided by efficient tagmentation (*see* Fig. 2), proper bisulfite treatment and an appropriately low number of PCR cycles (*see* Fig. 3a, b). In case of reproducible library fragment ranges and peaks, FU values monitored during the real-time library PCR run may enable assessment at which cycle the PCR has generated enough fragments for sequencing (Fig. 5). In the given example (*see* Table 2 and

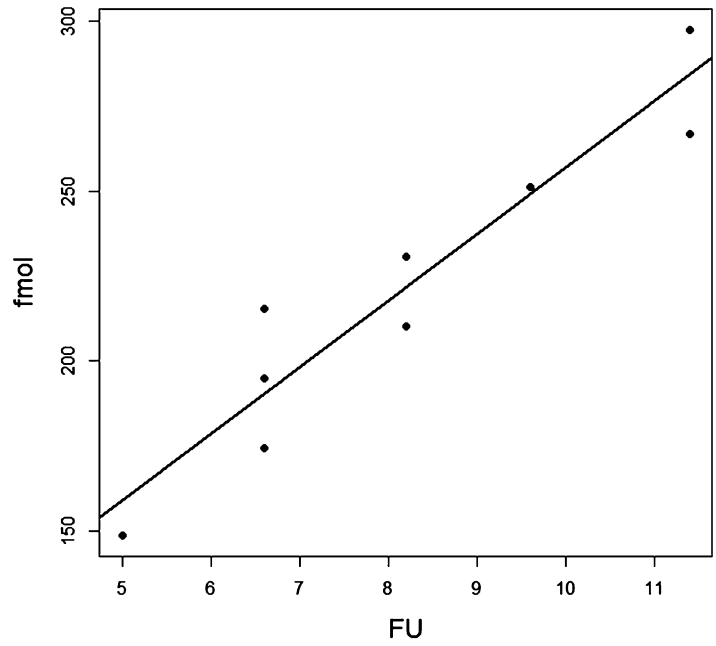


Fig. 5 Correlation between fragment numbers, given in fmol, in TWGBS libraries and corresponding fluorescence units (FU) indicated by real-time PCR. Values were obtained from Table 2; $R^2 = 0.9$

Fig. 5), an FU value of 5 corresponds to about 150 fmol fragments in the TWGBS library (10 μL volume). Lower FU values than 5, however, may lead to false assumptions of fragment numbers, as the nearly linear correlation may be valid only in a limited range of values.

16. 5mC-dNTP nucleotides are incorporated in vitro during the gap repair step into the nine bases gaps which remain after tagmentation [8] (*see* Fig. 1a, b). Consequently, the first nine bases of the second read and the last 9 bases before the adapter of the first read must be excluded from the methylation analysis of the genomic DNA. Appropriate bisulfite treatment determined with spiked-in phage λ DNA is indicated by a conversion frequency of ≥99.5%.

Acknowledgments

We gratefully acknowledge excellent technical support by Marion Bähr. We also acknowledge the excellent support from the sequencing core facility at the DKFZ. Work in the Plass laboratory was supported by the Helmholtz Foundation and the German Federal Ministry of Education and Science (BMBF) in the program for medical genome research (FKZ: 01KU1001A). B.B. and C.P. also receive funding within the DFG Research Consortium FOR 2674. B.B. has been further supported by a grant from BMBF in the framework of the German Epigenome Project DEEP (FKZ: 01KU1216B) and by a grant for the de.NBI-epi project in the context of the German National Network for Bioinformatics Infrastructures (FKZ: 031L0101A). Parts of the bioinformatics workflow are based on methylCtools, which has been originally developed by Volker Hovestadt.

Author Contributions: *D.W., C.D.I. and Q.W. conceived the study. B.B. and C.P. contributed materials. D.W. did the experiments and analyzed data. D.W., C.D.I. and C.P. wrote the manuscript.*

References

1. Fazzari MJ, Greally JM (2004) Epigenomics: beyond CpG islands. Nat Rev Genet 5:446–455
2. Lister R, Ecker JR (2009) Finding the fifth base: genome-wide sequencing of cytosine methylation. Genome Res 19:959–966
3. http://ihec-epigenomes.org/about/why-epigenomics/
4. Ko JY, Oh S, Yoo KH (2017) Functional enhancers as master regulators of tissue-specific gene regulation and cancer development. Mol Cells 40:169–177
5. Lister R, Pelizzola M, Dowen RH et al (2009) Human DNA methylomes at base resolution show widespread epigenomic differences. Nature 462:315–322
6. Miura F, Enomoto Y, Dairiki R et al (2012) Amplification-free whole-genome bisulfite sequencing by post-bisulfite adaptor tagging. Nucleic Acids Res 40:e136

7. Smallwood SA, Lee HJ, Angermueller C et al (2014) Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. Nat Methods 11:817–820
8. Adey A, Shendure J (2012) Ultra-low-input, tagmentation-based whole-genome bisulfite sequencing. Genome Res 22:1139–1143
9. Wang Q, Gu L, Adey A et al (2013) Tagmentation-based whole-genome bisulfite sequencing. Nat Protoc 8:2022–2032
10. Buenrostro JD, Giresi PG, Zaba LC et al (2013) Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat Methods 10:1213–1218
11. Schmidl C, Rendeiro AF, Sheffield NC et al (2015) ChIPmentation: fast, robust, low-input ChIP-seq for histones and transcription factors. Nat Methods 12:963–965
12. Adey A, Morrison HG, Asan et al (2010) Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density in vitro transposition. Genome Biol 11:R119
13. Amini S, Pushkarev D, Christiansen L et al (2014) Haplotype-resolved whole-genome sequencing by contiguity-preserving transposition and combinatorial indexing. Nat Genet 46:1343–1349
14. Picelli S, Bjorklund AK, Reinius B et al (2014) Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. Genome Res 24:2033–2040
15. Cabezas-Wallscheid N, Klimmeck D, Hansson J et al (2014) Identification of regulatory networks in HSCs and their immediate progeny via integrated proteome, transcriptome, and DNA methylome analysis. Cell Stem Cell 15:507–522
16. Lipka DB, Wang Q, Cabezas-Wallscheid N et al (2014) Identification of DNA methylation changes at cis-regulatory elements during early steps of HSC differentiation using tagmentation-based whole genome bisulfite sequencing. Cell Cycle 13:3476–3487
17. Oakes CC, Seifert M, Assenov Y et al (2016) DNA methylation dynamics during B cell maturation underlie a continuum of disease phenotypes in chronic lymphocytic leukemia. Nat Genet 48:253–264
18. Lu H, Yuan Z, Tan T et al (2015) Improved tagmentation-based whole-genome bisulfite sequencing for input DNA from less than 100 mammalian cells. Epigenomics 7:47–56
19. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754–1760
20. Grunenwald HL, Caruccio N, Jendrisak J et al (2010) Transposon end compositions and methods for modifying nucleic acids. USA Patent US20100120098A1
21. Jager N, Schlesner M, Jones DT et al (2013) Hypermutation of the inactive X chromosome is a frequent event in cancer. Cell 155:567–581
22. http://res.illumina.com/documents/products%5Ctechnotes%5Ctechnote_nextera_low_plex pooling_guidelines.pdf
23. Ziller MJ, Hansen KD, Meissner A et al (2015) Coverage recommendations for methylation analysis by whole-genome bisulfite sequencing. Nat Methods 12:230–232

Chapter 11

FISH and FICTION in Lymphoma Research

Maciej Giefing and Reiner Siebert

Abstract

Fluorescence in situ hybridization (FISH) is a powerful and robust technique allowing the visualization of target sequences like genes in interphase nuclei. It is widely used in routine diagnostics to identify cancer-specific aberrations including lymphoma-associated translocations or gene copy number changes in single tumor cells. By combining FISH with immunophenotyping—a technique called fluorescence immunophenotyping and interphase cytogenetic as a tool for investigation of neoplasia (FICTION)—it is moreover possible to identify a cell population of interest. Here we describe standard protocols for FISH and FICTION as used in our laboratories in diagnosis and research.

Key words Fluorescent in situ hybridization (FISH), Fluorescence immunophenotyping and interphase cytogenetics as a tool for investigation of neoplasia (FICTION), Lymphoma research

1 Introduction

In situ hybridization (ISH) is based on the ability of complementary nucleic acids to anneal and form stable DNA-RNA or DNA-DNA structures. It allows the precise localization of a genomic target sequence with the complementary probe—historically an RNA sequence labeled with a radio isotope. In the late 1970s and 1980s of the twentieth century, the potentially harmful radio isotopes have been replaced by fluorophores. The detection of the fluorescently labeled hybrids by means of fluorescent microscopy resulted in the first fluorescent in situ hybridization (FISH) experiments [1–5]. The introduction of DNA labeling procedures like nick translation or random priming that are based on the direct incorporation of labeled nucleotides into the nucleic acid further facilitated the preparation of FISH probes. Together with the development of chromosome banding techniques FISH allowed for the first time gene mapping with relatively high resolution and resulted in a burst of new applications in cancer research and other fields. Detailed information on the historical development of ISH and FISH is excellently reviewed elsewhere [5–8].

Ralf Küppers (ed.), *Lymphoma: Methods and Protocols*, Methods in Molecular Biology, vol. 1956,
https://doi.org/10.1007/978-1-4939-9151-8_11,

A major advantage of FISH is its ability to visualize target sequences in interphase nuclei without prior cell culture and laborious metaphase preparations. It is of special importance in cancer diagnostics and research where it is difficult to obtain high-quality metaphase spreads from tumor biopsies. Thus, FISH became a method of choice for fast and reliable identification of chromosomal alterations including copy number abnormalities, genomic breakpoints, and the ensuing translocations or oncogenic fusion genes.

With the discovery of recurrent chromosomal aberrations in lymphomas like t(8;14) in Burkitt lymphoma, t(14;18) in follicular lymphoma, or t(11;14) in mantle cell lymphoma, FISH became widely used in routine diagnostics of lymphatic neoplasms [9]. Despite the unprecedented advances of molecular biology techniques FISH is still regarded as a fast and reliable tool for the identification of chromosomal aberrations in diagnostic procedures [10]. This has been recently highlighted in the fourth edition of the *WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues* published in 2017 [11, 12].

Today, the availability of bacterial artificial chromosomes (BAC) or P1 artificial chromosome (PAC) libraries—wherein the entire human genome is stored in fragments of 50–300 kb per bacterial clone—allows fast probe preparation for the desired genomic region. Using differently labeled BACs/PACs, it is possible to design multicolor FISH assays for any combination of genes that aided the identification of secondary aberrations in lymphomas and provided insight into the complex genetic background of lymphomagenesis (Table 1) [9, 13, 14].

It is to be expected that with the rapid development of synthetic biology, the laborious procedure of BAC-/PAC-based probe preparation will be eagerly replaced by novel approaches for probe development. These will include the already available commercially synthesized repeat free oligonucleotides used for RNA-FISH [15] as well as custom synthesis of fluorescently labeled long DNA probes. The availability of the latter is hampered mainly by technical limitations of chemical synthesis.

In the field of lymphoma research, an important modification of FISH was developed in need for analyses of Hodgkin lymphoma. The application of this technique was initially hampered mainly for the following reasons: (1) Hodgkin lymphoma is characteristic for its scarcity of the neoplastic Hodgkin and Reed-Sternberg (HRS) cells that usually do not exceed 1% of cell content in the tumor, (2) the necessity to widely remove the cytoplasm during the isolation of interphase nuclei for FISH what leads to the loss of cellular phenotype and makes the HRS cell nuclei difficult to distinguish from other cell populations, and (3) the difficulties in visual identification of HRS cells in primary tissue sections. This was overcome by the

Table 1
Commercially available FISH probes for the detection of chromosomal translocations recurrently identified in lymphomas

Lymphoid malignancy	Chromosomal translocation	Genes involved	Available commercial probes[a]
B-cell derived			
Burkitt lymphoma	t(8;14)(q24;q32)	*MYC/IGH*	MYC/IGH Dual Fusion
	t(2;8)(p12;q24)	*IGK/MYC*	MYC Break Apart; IGK Break Apart
	t(8;22)(q24;q11)	*MYC/IGL*	MYC Break Apart; IGL Break Apart
Diffuse large B-cell lymphoma	t(3q27)	*BCL6*	BCL6 Break Apart
	t(14q32)	*IGH*	IGH Break Apart
	t(14;18)(q32;q21)	*IGH/BCL2*	IGH Break Apart; BCL2 Break Apart; IGH/BCL2 Dual Fusion
Follicular lymphoma	t(14;18)(q32;q21)	*IGH/BCL2*	IGH Break Apart; BCL2 Break Apart; IGH/BCL2 Dual Fusion
	t(3;14)(q27;q32)	*BCL6/IGH*	BCL6 Break Apart; IGH Break Apart
MALT lymphoma	t(11;18)(q21;q21)	*API2/MALT1*	BIRC3 (API2)/MALT1 Dual Fusion; MALT1 Break Apart
	t(14;18)(q32;q21)	*IGH/MALT1*	IGH/MALT1 Dual Fusion; MALT1 Break Apart; IGH Break Apart
	t(1;14)(p22;q32)	*BCL10/IGH*	IGH Break Apart
Mantle cell lymphoma	t(11;14)(q13;q32)	*CCND1/IGH*	IGH/CCND1 Dual Fusion; CCND1 Break Apart; IGH Break Apart
T-cell derived			
Anaplastic large cell lymphoma	t(2;5)(p23;q35)	*NPM/ALK*	ALK Break Apart
Precursor T-cell lymphoblastic lymphoma	14q11.2; 7q35; 7p14-15/10q24 or 5q35	*αTCR* and *δTCR; βTCR; γTCR/TLX1* or *TLX3*	TCRAD BAP; TLX1 Break Apart; TLX3 Break Apart

The table was prepared referring to refs. 9, 14, 19
MALT mucosa-associated lymphatic tissue
[a]Suppliers like Abbott Molecular, Cytocell or Kreatech offer a large collection of FISH probes relevant for hematological diseases

development of the fluorescence immunophenotyping and interphase cytogenetics as a tool for investigation of neoplasia (FICTION) technique that combines fluorescence immunophenotyping and fluorescence in situ hybridization [16–18]. The advantage of FICTION over FISH in analyzing Hodgkin lymphoma is that it uses

the phenomenon that HRS cells, in contrast to the wide majority of other cell populations infiltrating the tumor of Hodgkin lymphoma, strongly express the CD30 protein. For FICTION, tissue sections or imprints are immunophenotyped with primary antibodies directed against CD30 and visualized either by conjugation of the primary or the secondary antibody with a fluorescent dye. Thereafter, fluorescent in situ hybridization is performed with the desired combination of probes. Using different filters in the fluorescence microscope, FICTION allows visualization of the neoplastic HRS cells in the section by the immunophenotype and simultaneous analysis of the fluorescent signals in the nuclei.

In this chapter we describe the standard protocols for FISH and FICTION used in our laboratories for routine diagnostics of lymphatic neoplasms and cancer research.

2 Materials

Disclaimer: The reagents and kits used in the described procedures in our laboratories can be replaced by alternative products offered by various companies.

2.1 Preparation of Locus-Specific FISH Probes

2.1.1 DNA Preparation from BAC/PAC Clones

Instruments/Website for Probe Selection

1. http://genome.ucsc.edu/cgi-bin/hgTracks (FISH clones, BAC end pairs, fosmid end pairs).
2. Centrifuge capable of 15,000 × *g* at 2–8 °C equipped with racks for 15 mL, 50 mL tubes or alternative for large-scale preparations.
3. Microcentrifuge for 1.5 mL and 2 mL tubes.
4. Incubation oven with shake option.
5. Water bath.
6. NanoDrop or alternative for DNA concentration and quality measurement.

Reagents and Buffers

1. BAC/PAC clones harboring the DNA of interest.
2. Selection antibiotic (*see* **Note 1**).
3. Glycerine 50%.
4. LB medium: dissolve 20 g LB Broth Base in 1000 mL distilled water. Autoclave and store at 4 °C for maximum 1 year.
5. PhasePrep™ BAC DNA Kit (Sigma-Aldrich Chemie GmbH, Munich, Germany).

Equipment

1. Vented tubes for culturing BACs/PACs.
2. 250 mL conical flasks.
3. 1.5 mL tubes.

2.1.2 DNA Labeling by Random Priming

Instruments

1. Water bath (0–99 °C).
2. Centrifuge capable of 15,000 × *g*.

Reagents and Buffers

1. Bioprime DNA Labeling System (ThermoFisher Scientific).
2. Deoxynucleoside Triphosphate Set (dilute unlabeled dNTPs each 1:10 with distilled water to final concentration of 10 mM).
3. Orange-dUTP or Green-dUTP (Enzo Life Sciences GmbH, Lörrach Germany): 50 nmol of dUTP, 50 µL distilled water.
4. dUTPs labeling mixture**:** 2.5 µL 10 mM dATP, 2.5 µL 10 mM dCTP, 2.5 µL 10 mM dGTP, 1.25 µL 10 mM dTTP, 12.5 µL 1 mM dUTP (orange or green), 3.75 µL distilled water.

Equipment

1. Amicon Ultra-0.5, Ultracel-30 Membrane, 30 kDa (Merck Millipore).
2. 1.5 mL tubes.

2.1.3 Probe Precipitation

Instruments

1. Centrifuge capable of 15,000 × *g*.

Reagents and Buffers

1. Cot1 DNA.
2. 3 M sodium acetate, pH 5.0.
3. Ethanol abs.
4. 20× SSC: 3 M NaCl, 0.3 M sodium citrate, dissolve in 800 mL distilled water and adjust to 1000 mL, pH 7.0.
5. Mastermix: 2.5 mL formamide, deionized, 1.25 mL 40% dextran sulfate, 0.5 mL 20× SSC, 0.25 mL distilled water, stored at −20 °C. We recommend using a face mask during the preparation of mastermix as dextran sulfate is toxic.
6. Control chromosome enumeration probes (CEP probes) if appropriate (Abbott, Kreatech, Metasystems or alternative).

Equipment

1. 1.5 mL tubes.
2. Face mask.

2.2 Sample Preparation

2.2.1 Preparation of Cell Suspensions from Cultured Tumor Cells from Bone Marrow or Blood

Instruments

1. Sterile bench.
2. Cell culture incubator with CO_2.
3. Centrifuge with racks for 15 mL tubes.
4. Vacuum pump.
5. Sterile bench.

Reagents and Buffers

1. RPMI 1640 medium with 25 mM Hepes, without L-glutamine.
2. Fetal calf serum (FCS).
3. L-Glutamine.
4. Penicillin/streptomycin
5. Conditioned medium:
   (a) Collect peripheral blood from volunteers using heparin-monovettes.
   (b) Incubate at 37 °C in 5% CO_2 for 2–5 h till sediment is formed.
   (c) Prepare in a sterile flask: 500 mL RPMI 1640, 125 mL FCS, 6.25 mL L-glutamine, 6.25 mL penicillin/streptomycin, 6.25 mL *phytohemagglutinin*.
   (d) Aliquot each 9 mL of the medium in 25 cm^2 surface area culture flasks.
   (e) Collect the supernatant from the blood samples (in case a leukocytes ring formed mix it with the supernatant), and add 1 mL of the supernatant to the 9 mL of medium.
   (f) Incubate for 72 h in 37 °C with 5% CO_2.
   (g) Transfer the cultures in 15 mL tubes and centrifuge at 200 × *g* for 10 min.
   (h) Collect the supernatants in 50 mL tubes and filter the medium using Steriflips with a vacuum pump.
   (i) Remove the Steriflips sterile under the bench, und close the tubes with a sterile caps.
   (j) Store at −20 °C.
6. Phytohemagglutinin.
7. 0.075 M KCl.
8. Ice-cold fixative: 50 mL acetic acid, 150 mL absolute methanol.

Equipment

1. Heparin-monovettes.
2. 500 mL flask.

3. 15 mL and 50 mL tubes.
4. Steriflips (Merck Millipore).
5. Twist top vials.

2.2.2 Preparation of FISH Slides from Cell Suspensions

Instruments

1. Fume hood.
2. Light microscope.

Reagents and Buffers

1. Protease 1 solution: 250 mg protease 1, 500 mL VP2000 Protease Buffer, 0.01 N HCl (Abbott or alternative).
2. Paraformaldehyde.
3. Ethanol abs.

Equipment

1. Pasteur pipette.
2. Glass slides.
3. Diamond pen.
4. Coplin-jars for glass slides.

2.3 Hybridization

2.3.1 Instruments

1. Water bath.
2. Hybridization oven.

2.3.2 Equipment

1. Cover glasses Ø 10 mm.
2. Fixogum rubber cement (Marabu GmbH & Co. KG, Tamm, Germany).
3. Hybridization metal box.

2.4 Washing

2.4.1 Instruments

1. Water bath.

2.4.2 Reagents and Buffers

1. 2× SSC.
2. Wash buffer 1: 2 mL 20× SCC, 300 µL Tween20, adjust to 100 mL with distilled water.
3. Wash buffer 2: 10 mL 20× SCC, 100 µL Tween20, adjust to 100 mL with distilled water.
4. DAPI stock solution: 0.2 mg DAPI (4′,6-diamidino-2-phenylindole·2HCl)/mL distilled water, aliquot in 1.5 mL tubes, and store at 4 °C in the dark.
5. DAPI working solutions: 100 µL DAPI stock solution, 100 mL 2× SSC. The solution can be used up to 1 month.
6. Vectashield mounting medium H-1200 (Novus Biologicals or Vector Laboratories).

2.4.3 Equipment

1. Coplin-jars for glass slides.
2. Cover glasses 124 × 60 mm

2.5 Evaluation

1. Fluorescent microscope with appropriate filter set (*see* **Note 2**).
2. Documentation software, for example, Isis-Workstation Metasystems software (Metasystems GmbH, Altlussheim, Germany).

2.6 FISH to Paraffin-Embedded Tissue Sections

2.6.1 Instruments

1. Fume hood.
2. Pressure cooker.

2.6.2 Reagents and Buffers

1. Roti-Histol (Carl Roth GmbH, Karlsruhe, Germany).
2. Ethanol abs.
3. 0.01 M citrate buffer: 2.3 g 99.5% citric acid (dehydrated), 800 mL distilled water, 5.4 mL 5 N NaOH, adjust with distilled water to 1000 mL, pH 6.0, prepare daily new solution.
4. Protease 1 solution: 250 mg protease 1, 500 mL VP2000 Protease Buffer, 0.01 N HCl (Abbott or alternative).
5. Paraformaldehyde.

2.6.3 Equipment

1. Coplin-jars for glass slides.
2. Steel rack for glass slides.

2.7 Fiction

2.7.1 FICTION on Cryo-Sections

Reagents and Buffers

1. Acetone.
2. PN buffer: 13.8 g sodium hydrogen phosphate ($NaH_2PO_4 \times H_2O$), 14.2 g disodium hydrogen phosphate (Na_2HPO_4), adjust to 1000 mL with distilled water, pH 8.0, store at room temperature up to 1 year.
3. PNM buffer: 50 g skimmed powdered milk, 0.2 g sodium azide (NaN_3), adjust to 1000 mL with PN buffer, warm up, and stir the buffer for 1–2 h, centrifuge the buffer to pellet the debris for 15 min at 850 × *g*, aliquot the buffer in 50 mL tubes, and store at 4 °C up to 1 year.
4. Ice-cold fixative: 50 mL acetic acid, 150 mL absolute methanol.
5. Paraformaldehyde.
6. Ethanol abs.

Equipment

1. Shandon cover plates and racks (Thermo Fisher Scientific).
2. Coplin-jars for glass slides.

3 Methods

3.1 Preparation of Locus-Specific FISH Probe

3.1.1 DNA Preparation from BAC/PAC Clones

1. Select appropriate BAC/PAC clones (*see* **Note 3**).
2. Transfer 50 μL of the medium in which the clones were shipped to 3 mL LB medium, and add 1.5 μL of the appropriate selective antibiotic [200 mg/mL] (*see* **Note 4**).
3. Incubate overnight at 37 °C, shaking.
4. Prepare several glycerine stocks by adding 500 μL of 50% glycerine to each 500 μL of the overnight culture.
5. In a 250 mL conical flask prepare 100 mL LB medium, add 50 μL [200 mg/mL] of the appropriate antibiotic and 10 μL of the glycerine stock (*see* **Note 5**).
6. Incubate overnight at 37 °C in an incubation oven, shaking.
7. Isolate BAC/PAC DNA using the PhasePrep™ BAC DNA Kit (Sigma-Aldrich Chemie GmbH, Munich, Germany) according to the instructions of the supplier (*see* **Note 6**).
8. Dissolve the isolated DNA for more than 1 h at 4 °C. Thereafter thoroughly vortex and measure the concentration of DNA. DNA of sufficient quality should show the value of 1.8–2.0 for the absorption ratio A_{260}/A_{280}. Store the DNA at −20 °C.

3.1.2 DNA Labeling by Random Priming

1. In a 1.5 mL tube with rubber gasket, dissolve 1 μg DNA in a final volume of 24 μL distilled water.
2. Add 20 μL of 2.5× random primer solution.
3. Denature the mixture in boiling water for 5 min.
4. Immediately thereafter cool the mixture down on ice for 5 min.
5. Add 5 μL of diluted dUTPs labeling mixture. From now on try to avoid exposing the probe to light.
6. Add 1 μL Klenow fragment.
7. Vortex carefully and centrifuge. Avoid harsh vortexing and centrifugation that may damage the sensitive Klenow polymerase.
8. Incubate overnight at 37 °C.
9. Add 5 μL Stop Buffer, vortex and centrifuge.
10. Label the Amicon centrifugation filter, and insert it into the 1.5 mL collection tube.
11. Add 300 μL distilled water to the labeling mixture, and transfer it into the centrifugation filter.
12. Centrifuge at 13,000 × *g* for 10 min.
13. Discard the flowthrough.

14. Add 300 μL distilled water into the centrifugation filter, and centrifuge at 13,000 × *g* for 10 min.
15. Add 50 μL distilled water into the centrifugation filter, and insert the filter upside down into a new collection tube.
16. Centrifuge at 1700 × *g* for 5 min.
17. Transfer the purified labeled DNA into the original 1.5 mL tube with rubber gasket. Store at −20 °C.

3.1.3 Probe Precipitation

1. Prepare in a 1.5 mL tube the following: 5 μL Cot1 DNA.
2. Add 10 μL of the labeled DNA (larger volumes of the probe, or two or more differently labeled BAC/PAC clones may be jointly used without increasing the volume of Cot1 DNA).
3. Add 1/10 volume 3 M sodium acetate, pH 5.0.
4. Add × 2.5 volume abs. ethanol.
5. Centrifuge at 15,000 × *g* or higher for 30 min at room temperature.
6. Carefully remove the liquid with a pipette tip not disturbing the pellet.
7. Let the pellet dry in the dark for 15 min (leave the tube open) (alternatively use a SpeedVac for 10 min).
8. Resuspend the pellet in 10 μL 50% mastermix (or larger volumes if more labeled DNA was used in point 1), and add 1 μL commercially available chromosome enumeration probe (CEP) for each 10 μL of the mastermix.
9. Shake the probe in 37 °C for at least 1 h.
10. Store at −20 °C.

3.2 Sample Preparation

*(For **steps 1–5** work under a sterile bench)*

3.2.1 Preparation of Cell Suspensions from Cultured Tumor Cells, from Bone Marrow, or from Blood

1. Transfer 1–9 mL the bone marrow or blood sample into a 15 mL tube, and adjust with RPMI 1640 up to 14 mL. Mix gently.
2. Centrifuge at 200 × *g* for 10 min.
3. Remove the supernatant.
4. Resuspend the cells in 10 mL RPMI and transfer them to a culture flask.
5. Add 1 mL FCS, 1 mL conditioned culture medium, 100 μL L-glutamine.
6. Incubate overnight at 37 °C in 5% CO_2.
7. Transfer the cell culture into a 15 mL tube and centrifuge at 200 × *g* for 10 min.
8. Remove the supernatant (we recommend using a vacuum pump).

9. Add 37 °C warm 0.075 M KCl to the cell pellet up to 9 mL and mix.
10. Incubate at 37 °C for 30 min.
11. Add few drops of ice-cold fixative and mix immediately.
12. Centrifuge at 200 × *g* for 10 min.
13. Remove the supernatant and add ice-cold fixative up to 10 mL.
14. Repeat **steps 13** and **14** at least three times until the pellet is white.
15. Resuspend the pellet in 2 mL ice-cold fixative and transfer it to twist top vials.

3.2.2 Preparation of FISH Slides from Cell Suspensions

1. Mix the cell suspension.
2. Using a Pasteur pipette drop from about 20 cm height one droplet of the cell suspension onto a glass slide (work under a fume hood).
3. Let the slides dry in room temperature for 20 min.
4. Analyze the slides under a light microscope, and mark the region of highest cell density using a diamond pen.
5. Digest the slides in protease 1 solution for 5 min at 37 °C (*see* **Note 7**).
6. Wash the slides in a Coplin-jar with distilled water for 2 min at room temperature.
7. Fix the slides in a Coplin-jar in 1% paraformaldehyde for 2 min at room temperature.
8. Wash the slides in a Coplin-jar with distilled water for 1 min at room temperature.
9. Dehydrate the slides in 70%, 85%, and abs. ethanol each for 2 min at room temperature.
10. Air-dry the glass slides for 10 min at room temperature.

3.3 Hybridization

1. Apply 1.3 μL probe for one hybridization. Pipette the probe onto the region previously marked with the diamond pen having high cell density (avoid exposition to light).
2. Cover the hybridization region with Ø 10 mm cover glass). Avoid formation of air bubbles.
3. Cover the entire hybridization region with Fixogum rubber cement.
4. Transfer the glass slides into a hybridization metal box with lid, and denature 7 min at 75 °C in a water bath.
5. Hybridize for 12–72 h at 37 °C in a hybridization oven.

3.4 Washing (See Note 8)

1. Place a Coplin-jar with wash buffer 1 in 72 °C water bath.
2. Using a pincette carefully remove the Fixogum rubber cement and the cover glass.
3. Wash the glass slides in a Coplin-jar with 2× SSC for 1–10 min.
4. Place the slides in 72 °C warm wash buffer 1 for 2 min.
5. Transfer the glass slides into wash buffer 2 for 1 min in room temperature.
6. Transfer the glass slides to 2× SSC for 2–10 min in room temperature.
7. Stain the slides in DAPI for 5 min in room temperature.
8. Wash once in 2× SSC for 3 min at room temperature.
9. Put two drops Vectashield Mounting Medium onto a wet glass slide, and cover it with 124 × 60 mm cover glass.
10. Store in the dark at room temperature.

3.5 Evaluation

1. Hybridize your probe to 5 normal male controls, and analyze the signal quality in at least 100 nuclei each and 10 metaphases each to test for potential cross hybridizations or hybridization to a another chromosome. Only probes giving intensive and not diffused signals at the correct chromosomal position are appropriate for further analysis (*see* **Note 9**).
2. Calculate the detection threshold of a probe taking, for example, the mean % of false-positive signals from the 5 controls (100 interphase nuclei each), and add three times the standard deviation calculated for the 5 controls. Evaluate regions with bright signals and non-overlapping nuclei. For break-apart or dual-fusion probes detect a "merge" pattern if the distance between the signals is less than twice the signal diameter (*see* **Note 10**).
3. For evaluation of your samples analyze at least 100 nuclei each and 10 metaphases if present. In case of tissue sections, analyze cells across the whole hybridization area to detect possible subpopulations of cells. Any signal pattern distinct from the expected should be regarded as abnormal and be interpreted. Each sample should be evaluated independently by two observers. Further considerations on FISH probe evaluation may be found in [19].

3.6 FISH to Paraffin-Embedded Tissue Sections

(For deparaffinization work under a fume hood)

1. Incubate the slides with the paraffin sections 3 × 5 min in Roti-Histol in room temperature.
2. Hydrate the slides in abs, 80% und 70% ethanol for 5 min each at room temperature.

3. Place the slides in a steel rack and put it in a Coplin-jar with distilled water.
4. Boil 1 L 0.01 M citrate buffer in a pressure cooker and when boiling place the steel rack with the slides into the cooker; close the lid.
5. Warm up the pressure cooker until it reaches the maximal pressure, and let it cook for 160 s.
6. Cool down the pressure cooker under cold tap water and release gradually the steam.
7. Open the lid and carefully pour cold tap water inside. Do not pour the water directly onto the steel rack.
8. Place the steel rack with the slides into a Coplin-jar with distilled water.
9. Digest the slides in protease 1 solution for 30 min in 37 °C.
10. Wash the slides in a Coplin-jar with distilled water for 2 min at room temperature.
11. Fix the slides in a Coplin-jar in 1% paraformaldehyde for 2 min at room temperature.
12. Wash the slides in a Coplin-jar with distilled water for 1 min at room temperature.
13. Dehydrate the slides in 70%, 85%, and abs. ethanol each for 2 min at room temperature.
14. Air-dry the slides for 5 min at room temperature.
15. Proceed from Subheading 3.3 onward.

3.7 FICTION on Cryo-Sections

1. Air-dry the slides for 30 min.
2. Fix the slides for 10 min in acetone.
3. Pour one drop of the PN buffer on the slides and cover it with the plastic Shandon cover plates and attach them in the Shandon racks and pour 100 μL of the PN buffer into the funnel (Fig. 1).
4. After the PN buffer passes through (position 2), add 100 μL of the primary antibody solution for immunohistochemical staining of the section, and incubate for 30 min; cover the rack with the attached lid to protect the slides from light (*see* **Note 11**).
5. Wash the slides with 100 μL of PN buffer.
6. After the PN buffer passes through (position 2), add 100 μL of the secondary antibody solution, and incubate for 30 min; cover the rack with the attached lid to protect the slides from light (*see* **Note 12**).
7. Remove the slides from the Shandon racks and place them in a Coplin-jar containing PN buffer in room temperature, and protect the slides from light.

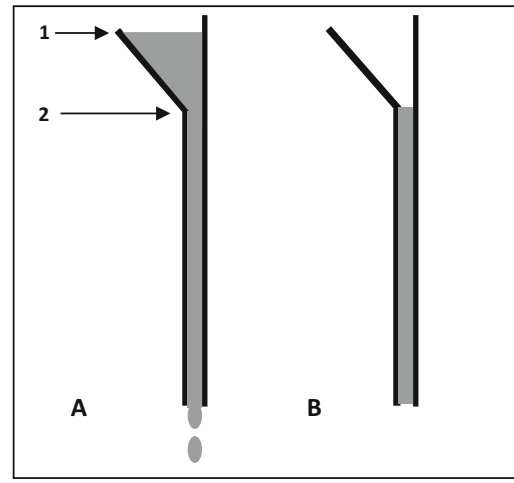


Fig. 1 Shandon cover plates; (**a**) Antibody solution is added to the funnel (position 1) created between the Shandon cover plate and the slide. (**b**) The antibody solution passes between the Shandon cover plate and the slide until the funnel is empty (position 2); the slide remains continuously covered by the antibody solution thanks to capillarity [20]

8. Fix the slides in ice-cold fixative for 10 min.
9. Wash the slides in distilled water for 1 min.
10. Fix the slides in a Coplin-jar in 1% paraformaldehyde for 2 min at room temperature.
11. Wash the slides in a Coplin-jar with distilled water for 1 min at room temperature.
12. Dehydrate the slides in 70%, 85%, and abs. ethanol each for 2 min at room temperature.
13. Air-dry the slides for 10 min at room temperature.
14. Proceed from Subheading 3.3 onward.

3.8 FICTION "Rescue"

In case no immunophenotype is observed after the hybridization and washing steps, the experiment can be rescued by applying the following procedure (work in dark):

1. Remove the coverslips carefully.
2. Wash the slides twice in 2× SSC for 10 min in room temperature, and shake several times manually.
3. Wash the slides twice in PN buffer for 10 min in room temperature, and shake several times manually.
4. Repeat immunophenotyping Subheading 3.7, **steps 3**–7.
5. Place the slides in a Coplin-jar containing PN buffer at room temperature for 5 min.
6. Air-dry the slides shortly.

7. Put two drops Vectashield Mounting Medium onto the wet glass slide, and cover it with 124 × 60 mm cover glass.
8. Store in the dark at room temperature.

3.9 FICTION on Paraffin-Embedded Tissue

1. Proceed with deparaffinization as described in Subheading 3.6, **steps 1–8**, and continue with Subheading 3.7, **steps 1–14** (*see* **Note 13**).

4 Notes

1. Detailed information on the antibiotic resistance of the used BAC/PAC clone can be found on the supplier info sheet.
2. The exciter and emitter ranges of the filter should cover the absorption and emission maximums of the fluorophore used, respectively. For spectrum green the absorption maximum is 496 nm, and the emission maximum is 520 nm; for spectrum orange the absorption maximum is 552 nm, and the emission maximum is 576 nm.
3. The BACs/PACs can be conveniently chosen using one of the genome browsers UCSC *Genome Browser—BAC end pairs* or Ensembl Human—Tilepath).

   The design of probes differs depending on the intended application:

   (a) To identify copy number changes of the region of interest (e.g., a gene) select clones spanning the region (*spanning probe*). Besides, you will need a reference probe against which the number of fluorescent signals will be evaluated. Commercially available chromosome enumeration probes (CEP) for the same chromosome as the analyzed region can be used although the disadvantage of such combination is that commercial probes usually have a higher hybridization efficiency resulting in stronger signal intensities than locus-specific probes. This may lead to false-positive results. The locus-specific clone and the reference probe must be labeled with different fluorophores for easy identification of the signals (Fig. 2).

   (b) To identify translocation breakpoints, select two clones flanking the potential breakpoint (*break-apart probe BAP*). For example, the clones may be located outside the 5′ and 3′ regions of a putative translocation-involved gene. Label the two clones with different fluorophores. The genomic distance between clones shall normally be less than 1 Mb so they appear as a co-localized signal. We recommend choosing green and red appearing fluorophores that usually merge into yellow signal indicating

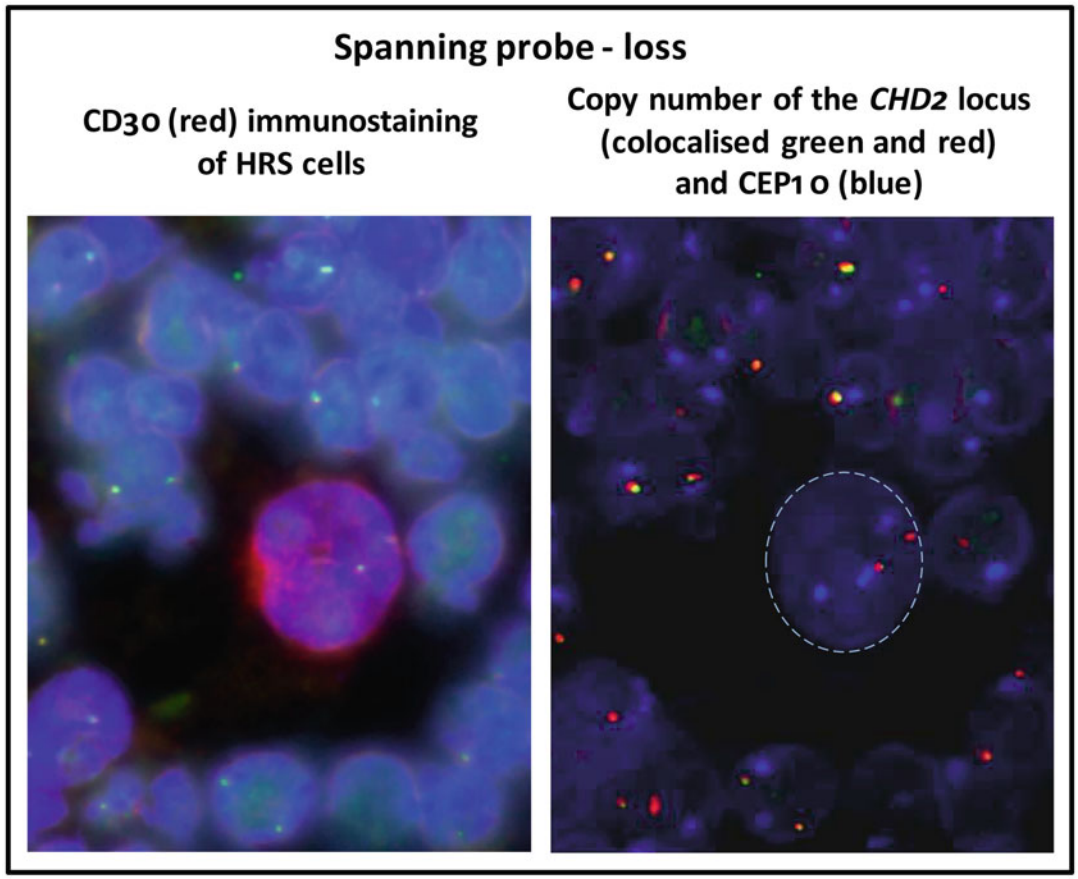


Fig. 2 Copy number analysis of Hodgkin and Reed-Sternberg cells in a classical Hodgkin lymphoma with tetraploid chromosome number. The FICTION technique was used combining immunofluorescent staining against the surface marker CD30 (left image, red) and fluorescent probes and counterstained nuclei (blue, DAPI). Loss of *CHD2* locus (colocalized green and red signals) compared to a CEP10 control (blue signals) and the ploidy level (right image). Not all signals in focus plain; false color display

no break in the investigated region. Using this probe a clearly distinct green and red signal indicates a breakpoint within the flanked region (Fig. 3a).

(c) To identify fusion genes, select two clones each spanning one gene putatively involved in the fusion (*dual-fusion probe*). Label the two clones with different fluorophores. Distinct green and red signals will indicate no fusion of the spanned genes. Any merge signal (that may appear in yellow) in a significant proportion of cells indicates gene fusion (Fig. 3b).

4. We recommend using vented culture tubes. Alternatively use standard 15 mL tubes but ensure not to close the cap tightly to allow gas exchange.

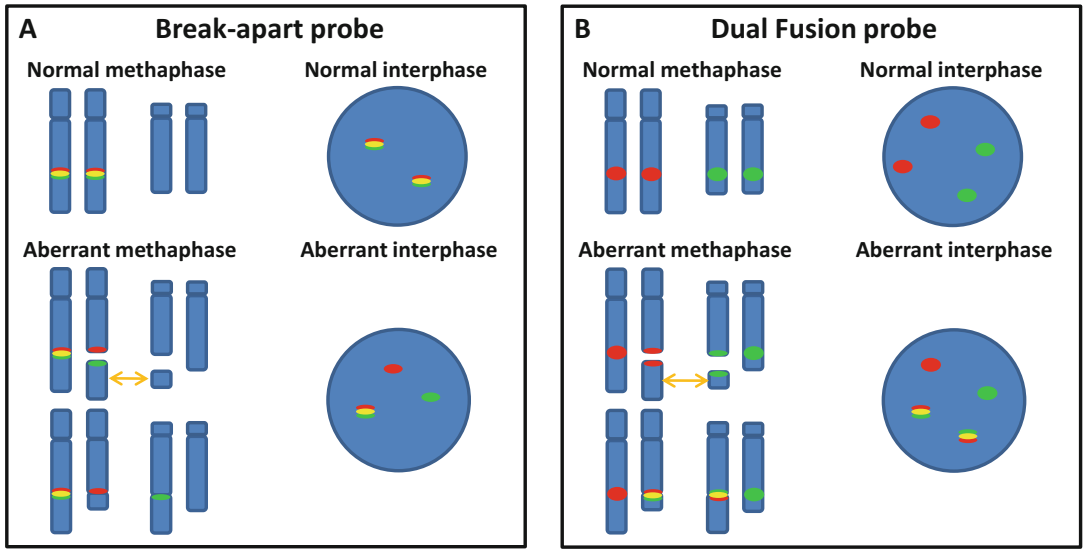


Fig. 3 Signal patterns obtained with different probe combinations. (**a**) Signal pattern expected for a break-apart probe in normal and aberrant metaphase or nucleus. (**b**) Signal pattern expected for a dual fusion probe in normal and aberrant metaphase or nucleus [19]

5. The given volumes may be scaled down to a mini or microscale preparation according to the instructions of the PhasePrep™ BAC DNA Kit supplier and depending on the quantity of BAC DNA that is needed for FISH/FICTION experiments. In our experience a mini-scale preparation yields at least 20 µg of DNA. In case of using mini-scale preparations, all subsequent centrifugation steps can be performed in 50 mL tubes. For microscale preparations 15 mL tubes can be used.
6. In contrast to the protocol of the supplier, we recommend dissolving the final DNA precipitate in 40–100 µL of distilled water.
7. You can alternatively digest the slides in 99 mL distilled water with 5 mg pepsin and 1 mL 1 M HCl warmed to 37 °C for 5 min.
8. The washing protocol can be automated using the VP 2000 Processor (Abbott GmbH & Co. KG, Wiesbaden-Delkenheim, Germany) or an alternative.
9. Use Isis-Workstation Metasystems software or an alternative for documentation.
10. As the preparations of sections may truncate cells resulting in false-positive results, the threshold may be up to 15% especially for complex probes. In research projects aimed at detecting copy number losses in tumor sections, to avoid false-positive results, the authors sometimes set the threshold arbitrary to 30%.

11. For the analysis of classical Hodgkin lymphoma to stain the HRS cells, we use the anti-CD30 (*BER-H2*) monoclonal antibody (dilution 1:20–1:50 the exact dilution must be determined experimentally) as the primary antibody. As secondary antibody we use the Alexa 594-conjugated rabbit anti-mouse antibody (Invitrogen, GmbH, Darmstadt, Germany) (dilution 1:50). The antibodies are diluted with PNM buffer.
12. Further antibodies may be used if necessary. These should not cross-react with other antibodies used. Always wash the slide with 100 μL of the PN buffer prior the incubation with a new antibody.
13. Avoid any type of digestion after the deparaffinization as it destroys the cells' phenotype, and thus no immunophenotyping will be possible.

Acknowledgments

The work of the authors on FISH/FICTION has been supported by the Polish Ministry of Science and Higher Education (grant IP2012 012072) to M.G. and from the Deutsche Krebshilfe, the Kinderkrebsinitiative Buchholz/Holm-Seppensen (KKI), and the BMBF to R.S.

References

1. Cheung SW, Tishler PV, Atkins L et al (1977) Gene mapping by fluorescent in situ hybridization. Cell Biol Int Rep 1:255–262
2. Rudkin GT, Stollar BD (1977) High resolution detection of DNA-RNA hybrids in situ by indirect immunofluorescence. Nature 3:472–473
3. Bauman JG, Wiegant J, Borst P et al (1980) A new method for fluorescence microscopical localization of specific DNA sequences by in situ hybridization of fluorochromelabelled RNA. Exp Cell Res 128:485–490
4. Langer-Safer PR, Levine M, Ward DC (1982) Immunological method for mapping genes on Drosophila polytene chromosomes. Proc Natl Acad Sci U S A 79:4381–4385
5. Van Prooijen-Knegt AC, Van Hoek JF, Bauman JG et al (1982) In situ hybridization of DNA sequences in human metaphase chromosomes visualized by an indirect fluorescent immunocytochemical procedure. Exp Cell Res 141:397–407
6. Trask BJ (2002) Human cytogenetics: 46 chromosomes, 46 years and counting. Nat Rev Genet 3:769–778
7. Levsky JM, Singer RH (2003) Fluorescence in situ hybridization: past, present and future. J Cell Sci 116:2833–2838
8. Wan TS, Ma ES (2012) Molecular cytogenetics: an indispensable tool for cancer diagnosis. Chang Gung Med J 35:96–110
9. Siebert R, Weber-Matthiesen K (1997) Fluorescence in situ hybridization as a diagnostic tool in malignant lymphomas. Histochem Cell Biol 108:391–402
10. Baranger L, Cuccuini W, Lefebvre C et al (2016) Cytogenetics in the management of children and adult acute lymphoblastic leukemia (ALL): an update by the Groupe francophone de cytogénétique hématologique (GFCH). Ann Biol Clin (Paris) 74:547–560
11. WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues, Revised 4th Edition, Volume 2, edited by Swerdlow SH, Campo E, Harris NL, et al. ISBN-13 (Print Book) 9789283244943
12. Swerdlow SH, Campo E, Pileri SA et al (2016) The 2016 revision of the World Health Organization classification of lymphoid neoplasms. Blood 127:2375–2390

13. Schröck E, Padilla-Nash H (2000) Spectral karyotyping and multicolor fluorescence in situ hybridization reveal new tumor-specific chromosomal aberrations. Semin Hematol 37:334–347
14. Oscier DG, Gardiner AC (2001) Lymphoid neoplasms. Best Pract Res Clin Haematol 14:609–630
15. Bradley S, Zamechek L, Aurich-Costa J (2009) Oligonucleotide FISH Probes. In: Liehr T (ed) Fluorescence in situ hybridization (FISH)—application guide. Springer, Berlin, Heidelberg, pp 67–73
16. Weber-Matthiesen K, Winkemann M, Müller-Hermelink A et al (1992) Simultaneous fluorescence immunophenotyping and interphase cytogenetics: a contribution to the characterization of tumor cells. J Histochem Cytochem 40:171–175
17. Weber-Matthiesen K, Deerberg J, Müller-Hermelink A et al (1993) Rapid immunophenotypic characterization of chromosomally aberrant cells by the new FICTION method. Cytogenet Cell Genet 63:123–125
18. Weber-Matthiesen K, Pressl S, Schlegelberger B et al (1993) Combined immunophenotyping and interphase cytogenetics on cryostat sections by the new FICTION method. Leukemia 7:646–649
19. Ventura RA, Martin-Subero JI, Jones M et al (2006) FISH analysis for the detection of lymphoma-associated chromosomal abnormalities in routine paraffin-embedded tissue. J Mol Diagn 8:41–51
20. Martín-Subero JI (2001) Development of the multicolor FICTION technique. Applications to the diagnosis and research of malignant lymphomas. Doctoral thesis. Christian-Albrechts University, Kiel, Germany; currently: University of Barcelona, Barcelona, Spain

Chapter 12

NGS-Based High-Throughput Screen to Identify MicroRNAs Regulating Growth of B-Cell Lymphoma

Joost Kluiver, Fubiao Niu, Ye Yuan, Klaas Kok, Anke van den Berg, and Agnieszka Dzikiewicz-Krawczyk

Abstract

MicroRNAs (miRNAs) play important roles in development, differentiation, and homeostasis by regulating protein translation. In B-cell lymphoma, many miRNAs have altered expression levels, and for a limited subset of them, experimental data supports their functional relevance in lymphoma pathogenesis. This chapter describes an unbiased next-generation sequencing (NGS)-based high-throughput screening approach to identify miRNAs that are involved in the control of cell growth. First, we provide a protocol for performing high-throughput screening for miRNA inhibition and overexpression. Second, we describe the procedure for next-generation sequencing library preparation. Third, we provide a workflow for data analysis.

Key words B-cell lymphoma, MicroRNA, High-throughput screen, Cell growth, MiRNA inhibition, MiRNA overexpression

1 Introduction

MicroRNAs (miRNAs) belong to the family of small noncoding (nc) RNAs. They are transcribed as longer primary transcripts from the genome and processed to 19–22-nucleotide-long mature miRNAs based on the presence of a stem-loop-like structure. In general one of the two strands is incorporated into the RNA-induced silencing complex (RISC) [1], whereas the other strand is degraded. The mature miRNA guides the RISC to homologous regions in their target gene transcripts resulting in inhibition of protein synthesis. About 30–50% of all genes can be regulated by miRNAs indicating the widespread involvement of miRNAs in the regulation of gene expression [2].

MiRNAs are involved in almost all known biological processes including cell cycle, apoptosis, and other major signaling pathways. Multiple miRNAs play fundamental roles in hematopoiesis [3] and are important determinants of B-cell development and maturation

Ralf Küppers (ed.), *Lymphoma: Methods and Protocols*, Methods in Molecular Biology, vol. 1956,
https://doi.org/10.1007/978-1-4939-9151-8_12,

[4, 5]. In B-cell lymphoma, oncogenic or tumor suppressor roles have been described for several miRNAs. A broadly studied oncogenic miRNA is miR-155. This miRNA is highly expressed in Hodgkin lymphoma, primary mediastinal B-cell lymphoma and diffuse large B-cell lymphoma [6, 7]. Eμ-miR-155 transgenic mice show polyclonal B-cell proliferations that can progress into B-cell leukemia or lymphoma [8]. The locus containing the oncogenic miR-17~92 cluster consisting of six miRNAs is amplified in part of the diffuse large B-cell, mantle cell, and Burkitt lymphoma cases [9–11]. MiR-17~92 transgenic mice develop lymphoproliferative disease [12], and miR-17~92 overexpression accelerates B-cell lymphomagenesis in the Eμ-MYC model [13–15]. Conditional overexpression of miR-21 in a mouse model resulted in pre-B-cell malignancies that were addicted to miR-21 expression [16]. Examples of miRNAs with tumor suppressor activity include miR-15a/miR-16-1 [17]. Mice with a deletion of the corresponding chromosomal region developed chronic lymphocytic leukemia-like disease [18].

Next to these examples of functionally characterized miRNAs, many more are differentially expressed in B-cell lymphoma. Selection of the most relevant candidates for functional follow-up studies is based on current knowledge and therefore biased at least to a certain extent. High-throughput gain- or loss-of-function approaches allow efficient and unbiased identification of miRNAs that induce specific phenotypes in lymphoma cells. These phenotypes can include cell growth by studying changes in the abundance of specific constructs over time or focus on other characteristics using readouts such as resistance to certain drugs by selective pressure or changes in expression of selected genes by flow cytometry analysis of individual cells [19]. For miRNAs, several high-throughput approaches have been established (reviewed in ref. [20]). The earliest study with a few hundreds of miRNA constructs was published in 2006 [21]. A marked increase in the number of high-throughput screening studies can be observed in the last 8 years. For miRNA gain-of-function studies, miRNA precursors or mimic oligos and retro- or lentiviral overexpression vectors are most commonly used [22–24]. For loss-of-function approaches, miRNA inhibitor oligos and lentiviral inhibitor or tough-decoy vectors are most popular [25, 26]. In a very recent paper, a high-throughput screen was performed for almost 1300 miRNAs in 8 cancer cell lines to identify miRNAs that either positively or negatively affect growth [27]. Recently, the CRISPR/Cas9 system has offered a new approach to knock down miRNAs that can also be easily used for high-throughput screening [28].

In this chapter, we present a protocol to identify miRNAs that affect growth of B-cell lymphoma in an unbiased high-throughput approach using the sequence of the lentiviral miRNA overexpression or inhibition constructs as a molecular barcode. Cells are

transduced with lentiviral pools of miRNA inhibition or overexpression constructs. At several time points during culturing, cell aliquots are taken for DNA isolation. Changes in the abundance of individual constructs over time are then determined by next-generation sequencing (NGS). Statistical analysis identifies outliers whose abundance in the cell pool changes significantly. The identified miRNAs can be then further functionally characterized in follow-up studies to determine, e.g., via which target genes the phenotype is influenced.

2 Materials

2.1 High-Throughput Loss- and Gain-of-Function MiRNA Screen

2.1.1 Confirmation of the Expression of MiRNA Inhibition and Overexpression Constructs (Optional)

1. miRNeasy micro or mini kit (#217084 & #217004, Qiagen, Venlo, the Netherlands).
2. Nanodrop Spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA).
3. NEXTflex™ Small RNA-Sequencing Kit v3 (NOVA-5132-06, Bioo Scientific, Austin, TX, USA).
4. MiSeq™ (Illumina, San Diego, CA, USA).
5. miRDeep 2.0 [29].
6. miRBase release 22 (http://www.mirbase.org/).
7. Phosphate-buffered saline (PBS): 140 mM NaCl, 9 mM Na_2HPO_4, 1.3 mM NaH_2PO_4.

2.1.2 Performing the Screen

1. Lymphoma cell lines (ATCC, Middlesex, UK/DSMZ, Braunschweig, Germany).
2. Culture media appropriate for the cell lines.
3. miRZip™ Pooled Anti-miRNA Virus Library (SBI, Palo Alto, CA, MZIPPLVA-1).
4. Lentiviral library of miRNA overexpression constructs prepared according to the protocol in [30] (*see* **Notes 1** and **2**).
5. 4 μg/μL polybrene (Sigma, Seelze, Germany, hexadimethrine bromide): dissolve 0.4 g polybrene in approx. 80 mL water and add water to 100 mL. Filter-sterilize with a 0.2 μM filter.
6. Flow cytometer.
7. Cell sorter.

2.2 DNA Isolation and Library Preparation and Next-Generation Sequencing

2.2.1 DNA Isolation

1. Harvested cell pellets.
2. SE buffer: 75 mM NaCl, 25 mM EDTA.
3. 10% SDS
4. Proteinase K (20 mg/mL) (#25530049, ThermoFisher Scientific).
5. Saturated NaCl (6 M).
6. Chloroform.
7. GlycoBlue coprecipitant 15 μg/μL (#AM9515, Thermo Fisher Scientific).
8. Isopropanol.
9. 70% ethanol
10. TE^{-4} buffer: 10 mM Tris–HCl and 0.1 mM EDTA, pH 8.0.

2.2.2 Library Preparation and Next-Generation Sequencing

1. Vector-specific primers (Table 1).
2. NEBNext High-Fidelity polymerase with 2× PCR Master Mix (M0541S, New England Biolabs, Ipswich, MA) or another high-fidelity polymerase.
3. PCR machine.
4. DNA Clean & Concentrator™-5 kit (Zymo Research, Irvine, CA, USA).
5. NEBNext Multiplex oligos for Illumina (#E7335, New England Biolabs).
6. MiSeq™ (Illumina).

2.3 Next-Generation Sequencing Data Analysis

1. BWA (version 0.7.12; https://github.com/lh3/bwa).
2. SAM tools (version 1.3; http://www.htslib.org/).
3. MATLAB (version 6.1, The MathWorks Inc., Natick, MA, USA).

Table 1
Primer sequences

Vector	Forward (5′ → 3′)[a]	Reverse (5′ → 3′)
pmiRZIP	CTGGGAAATCACCATAAACG	R1: TAACCAGAGAGACCCAGTAG R2: CTAACCAGAGAGACCCAGTAG R3: TCTAACCAGAGAGACCCAGTAG R4: GTCTAACCAGAGAGACCCAGTAG
pCDH-EF1α-miR-IRES-GFP	GTGACCGGCGCCTACTCTAG	R1: AAGCGGCTTCGGCCAGTAACGTT R2: CAAGCGGCTTCGGCCAGTAACGTT R3: CCAAGCGGCTTCGGCCAGTAACGTT R4: TCCAAGCGGCTTCGGCCAGTAACGTT

[a]Forward primers should be combined with a barcode sequence. *See also* Subheading 3.2.2 and **Note 7**

3 Methods

3.1 High-Throughput Loss- and Gain-of-Function MiRNA Screen

3.1.1 Confirmation of the Expression of MiRNA Inhibition and Overexpression Constructs (Optional)

For miRNA overexpression, both strands of the miRNA stem-loop are present in the overexpression construct, and it is important to realize that both strands (3p and 5p) can be incorporated in the RISC and can induce a phenotype. For miRNA inhibition constructs, the design is made in such a way that it favors overexpression of the desired 3p antisense arm. However, the unwanted 5p sense arm can also be generated and induce off-target effects. Testing the abundance of the 5p and 3p strands in cells infected with the miRNA overexpression or inhibition pool is recommended as a quality control. It is important to realize that in the below described validation experiment, each construct will be present in only a proportion of the cells. Thus, the amount of overexpression observed will be a substantial underestimation of the actual levels.

1. Infect the cells of interest with the miRNA inhibition or overexpression library at a high multiplicity of infection (MOI), and collect at least 1×10^6 GFP^+ cells. For the overexpression library, cells should be infected in parallel with an empty vector control to determine the endogenous levels of the miRNAs. In case of low GFP^+ cell percentages, sort cells to a purity of at least >90%. Wash cells once with PBS and homogenize in Qiazol reagent.
2. Isolate RNA using the miRNeasy mini kit according to the manufacturer's protocol.
3. Check quantity and quality of RNA by Nanodrop and gel electrophoresis.
4. Prepare small RNA libraries using the NEXTflex™ kit according to manufacturer's protocol and sequence on MiSeq (Illumina). Process the reads and analyze the data as described previously [31]. For the alignment of the miRNA inhibition library, the insert sequences are used, while for the miRNA overexpression library, this is done using miRDeep 2.0 in combination with the latest miRBase annotation.
5. For the miRNA inhibition library, analyze the 5p and the 3p arm normalized read counts. The 3p arm is preferably more prevalent than the 5p arm. An example of such an analysis is shown in Fig. 1a. For the miRNA overexpression library, plot the expression levels of the miRNAs (plot 5p and 3p if both are detected) in the conditions with and without miRNA overexpression library (Fig. 1b). Twofold-higher miRNA levels in miRNA library-transduced cells indicate successful overexpression.

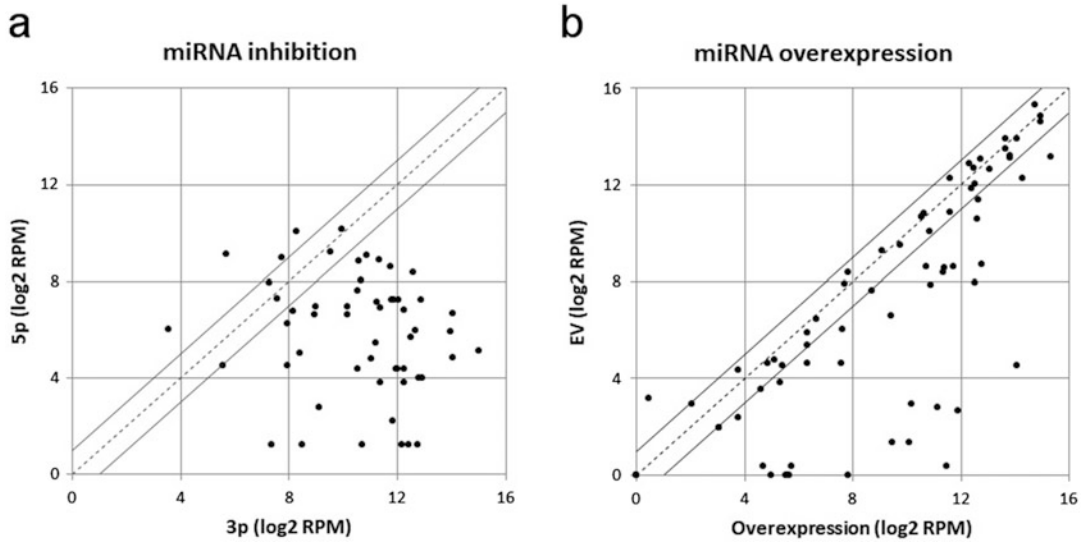


Fig. 1 Quality control of miRNA inhibition and overexpression pools. (**a**) Scatterplot depicting the abundance of the sense (5p) and antisense (3p, functional strand) strand of miRZip constructs. In this example 49 out of 55 constructs show a higher abundance of the functional 3p strand as compared to the 5p strand. (**b**) Scatterplot depicting expression of mature miRNAs in cells transduced with the miRNA overexpression library compared to empty vector (EV)-transduced cells. In this example approximately 70% of the constructs show at least for one of the two arms of the miRNA a twofold increase in expression level, indicating successful miRNA overexpression. Dashed line represents equal expression levels, and solid lines represent fold change of 2. *RPM* reads per million reads

3.1.2 Experimental Setup of the High-Throughput Screen

To ensure sufficient and uniform representation of all constructs from the library among the transduced cells, it is recommended to aim for a minimum of 1000 cells per construct. Thus, for a library of 1000 constructs (used as an example in the steps below), at least $1000 \times 1000 = 1 \times 10^6$ cells should be infected. To minimize the number of cells transduced with multiple constructs, the initial infection efficiency should not be more than 15–20%. This means that for a library of 1000 constructs, the number of cells to be used for the transduction should be at least 7×10^6. It is recommended to perform the infection in duplicate to allow selection of consistent phenotypes.

The first time point to harvest cells should be as soon as possible, but not before GFP is stably expressed. Typically this is at day 4 or 5 postinfection. The final time point for harvesting depends on the strength of the expected phenotypes and on the doubling time of the cells. In our experience, 15–20 population doublings are a good number to be able to identify clear changes in construct abundance. Including additional intermediate harvesting time points may be considered to obtain information on the kinetics of the phenotype(s).

3.1.3 Performing the Screen

1. To determine the amount of virus that results in 15–20% transduced cells, each batch of virus needs to be titrated for each cell line. Mix 0.5×10^6 cells in the logarithmic phase of growth in a volume of 1 mL with different amounts of virus supernatant in a 6-well plate (*see* **Note 3**), add polybrene to a final concentration of 4 μg/mL, mix, and incubate overnight (*see* **Note 4**).
2. Wash cells three times with PBS after 24 h, and culture under standard conditions.
3. On day 4–5 after infection, take an aliquot of the cells to determine the percentage of GFP^+ cells by flow cytometry. Select the optimal amount of virus resulting in 15–20% of transduced cells for the actual screen.
4. Mix 7×10^6 cells of choice in the logarithmic phase of growth with the appropriate volume of virus supernatant, and add polybrene to a final concentration of 4 μg/mL. Infection conditions (virus volume per amount of cells) should be exactly the same as used for virus titration. Perform the infection in duplicate for each cell line.
5. Wash cells three times with PBS 24 h after infection, and continue the culture using all cells under normal conditions.
6. Sort at least 2×10^6 GFP^+ cells on a cell sorter at day 4 or 5 after infection. Use 1×10^6 GFP^+ cells to continue the culture, and harvest 1×10^6 cells for DNA isolation. Collect the sorted cells in a tube with FBS. For subsequent culture, resuspend the cells in normal growth medium immediately after sorting (*see* **Notes 5** and **6**). For DNA isolation, wash cells with PBS, and freeze the pellets or proceed directly with DNA isolation.
7. Continue culturing cells till the final time point. At each passage keep at least 1×10^6 cells in culture to ensure good representation of all initially infected cells. Harvest at least 1×10^6 cells at each intermediate time point and at the final time point for DNA isolation.

3.2 DNA Isolation and Library Preparation and Next-Generation Sequencing

3.2.1 DNA Isolation

Volumes are for up to 1×10^6 cells; increase volumes accordingly with increased cell numbers.

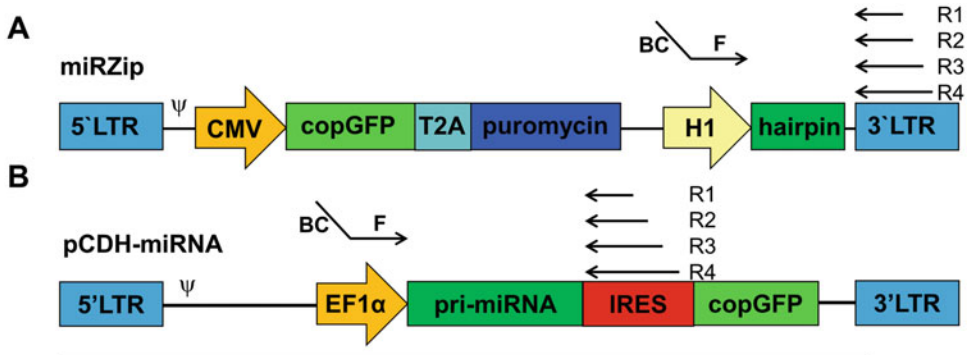
1. Add 0.5 mL SE buffer to the cell pellet and mix gently.
2. Add 50 μL 10% SDS and 2.5 μL proteinase K and incubate overnight at 55 °C.
3. Add 0.33 volume (180 μL) pre-warmed 6 M NaCl and 1 volume (700 μL) chloroform, and mix for 1 h on a top-over-top rotor at room temperature.
4. Centrifuge for 1 h at 3200 × *g* at 4 °C, and transfer the upper phase to a clean tube.

5. Optionally, add 1 μL GlycoBlue coprecipitant to better visualize the DNA pellet in the precipitation step below.
6. Add 1 volume (~700 μL) isopropanol and mix gently by inverting. If a DNA precipitate (cloud) is visible, carefully remove the isopropanol. If no precipitate is visible, centrifuge for 30 min at 14,000 × *g* at 4 °C to obtain a pellet.
7. Wash precipitate/pellet with 0.5 mL 70% ethanol, and carefully remove all ethanol.
8. Dry pellet at room temperature for 5–10 min.
9. Add 30–50 μL TE^{-4} and incubate overnight on a roller bank.
10. Measure DNA concentration by Nanodrop, and check quality on a 0.8% agarose gel.
11. Store DNA at 4 °C.

3.2.2 Library Preparation and Next-Generation Sequencing

Use barcoded forward primers in combination with general reverse primers flanking the inserts to amplify the miRZip and miRNA overexpression constructs (Fig. 2). The use of barcoded forward primers allows pooling of multiple samples in a single sequencing reaction. To avoid sequencing errors due to lack of heterogeneity at the initial nucleotide positions, it is essential to introduce small changes in the length of the primers. This is achieved by adding one to three nucleotides to the 5′ end of the reverse primer and by using a constant forward primer with 5′ barcodes that not only vary in sequence but also in length (eight to ten nucleotides). To limit variation in representation of constructs due to differences in amplification efficiency within each infection over time, it is advised to keep the reverse primer constant and use forward primers with similar-size barcodes. Primers are listed in Table 1. For barcode considerations, *see* **Note 7**.

1. The optimal total amount of input DNA for the amplification is dependent on the complexity of the library. For a library including 1000 constructs and an initial infection of 1000 cells per construct, the required input to have a good representation of the initially infected cells is ~6 μg genomic DNA (assuming 6 pg genomic DNA per cell and a GFP^+ cell percentage close to 100%). It is recommended to prepare six independent PCR reactions per DNA sample with 1 μg of genomic DNA as input in each tube. *See also* Fig. 2c for further explanation.
2. Prepare required number of PCR reactions, each tube containing 15 μL 2× NEBNext High-Fidelity PCR Master Mix, 1 μL of 10 μM forward primer, 1 μL of 10 μM reverse primer, and 1 μg of genomic DNA, and add water to a total volume of 30 μL.



C

Cell line	Day	Triplo	BC	Forward	Reverse	# PCR tubes
A	4	1	BC1	→	← R1	2
		2	BC2	→	← R1	2
		3	BC3	→	← R1	2
A	40	1	BC4	→	← R1	2
		2	BC5	→	← R1	2
		3	BC6	→	← R1	2
B	4	1	BC7	→	← R2	2
		2	BC8	→	← R2	2
		3	BC9	→	← R2	2
B	40	1	BC10	→	← R2	2
		2	BC11	→	← R2	2
		3	BC12	→	← R2	2

Fig. 2 Schematic overview of the miRZip and pCDH vectors and the locations of the primers used to amplify the inserts. (**a**) miRZip vector used for miRNA inhibition. (**b**) pCDH vector used for miRNA overexpression. (**c**) Scheme that shows how the barcoded (BC) forward (F) primers should be combined with the reverse (R1–R4) primers. Within one cell line/infection, the reverse primer should be constant. Per sample six PCR reactions with each an input of 1 μg of genomic DNA is performed divided over three different barcoded forward primers (so two PCR reactions per barcoded forward primer). This allows comparison of the consistency of the results after the NGS. Together the six PCR reactions ensure a 1000× coverage complexity. *IRES* internal ribosomal entry site, *LTR* long terminal repeat, *T2A Thosea asigna* virus 2A

3. Amplify the inserts using a PCR program consisting of 94 °C for 5 min, 26–30 cycles of 30 s at 94 °C, 45 s at 59 °C (miRZip) or 55 °C (pCDH-miRNA), 45 s at 72 °C (miRZip), or 2 min 30 s at 68 °C (pCDH-miRNA), followed by a final extension step at 72 °C of 7 min.
4. Combine the duplo PCR reactions (three per DNA sample), and check a small aliquot of the pooled PCR product (~5 μL) by agarose gel electrophoresis to confirm the size and estimate the PCR yield.
5. Prepare a mix of the PCR products based on the band intensities.
6. Purify the pooled PCR products using the DNA clean & ConcentratorTM-5 kit following the manufacturer's protocol.
7. Ligate adapters using the NEBNext Multiplex (#E7335) oligos for Illumina following the manufacturer's protocol.

8. Perform paired-end next-generation sequencing on MiSeq or similar sequencing machine, and aim at obtaining a number of reads that is the double of the number of initially infected cells (much more reads than the number of infected cells does not add to the quality of your data).

3.3 Next-Generation Sequencing Data Analysis

3.3.1 Reads Processing, Quality Control, and Read Quantification

1. Demultiplex reads based on the sample-specific barcodes.
2. Process reads and align to the insert sequences of the miRZip and miRNA overexpression libraries allowing for one mismatch using BWA [32].
3. Aligned reads are further processed with SAM tools [33], including counting the number of reads per sample and per construct. In case the average coverage of a sample or a construct is outside of the normal range of the experiment, it may be better to exclude that sample or construct from further analysis.
4. Normalize your read counts to the approximate minimum coverage that was maintained throughout all steps × the number of constructs in the library. In our example this was 1000 constructs for each 1000 infected cells resulting in an optimal normalization value of 1×10^6 reads (RPM).

3.3.2 Data Analysis and Outlier Test

1. Calculate fold changes relative to the first time point for each measurement.
2. Calculate the trend line that is forced to start from the first time point (typically day 4 or 5 postinfection), and register the slope (Fig. 3). *See* **Note 8** for how this process can be automated.
3. Use the slopes of the trend lines to identify outliers in the population using Tukey's interquartile range (IQR) approach [34]. If the slope is below the lower cutoff (Q1 − [1 × IQR]) or above the upper cutoff (Q3 + [1 × IQR]), it is considered an outlier. Note that different cutoffs can be applied: Using a cutoff +/− 1 × IQR results in approximately 5% of outliers, while using a cutoff of +/− 1.5 × IQR will give you 1% of outliers.
4. The constructs that are indicated as outliers in both independent infections represent the constructs that consistently result in an increased or decreased abundance over time and can be selected as candidates for further studies. *See also* **Note 9** on candidate selection.

4 Notes

1. Several vectors with different promoters are available (e.g., EF1α, CMV, MSCV) that may be suitable to drive miRNA

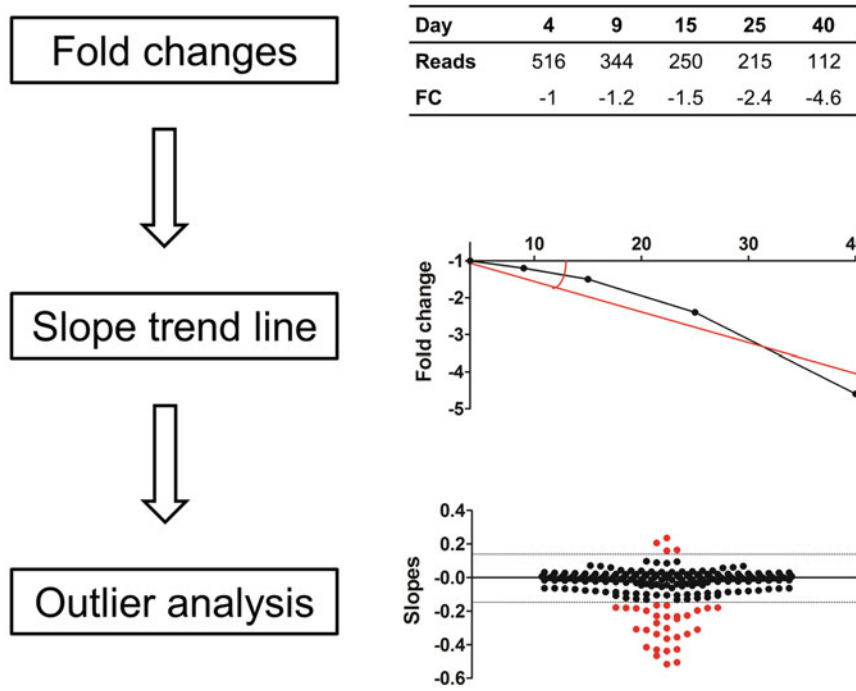


Fig. 3 Flowchart illustrating how candidates can be defined from the NGS data. First fold changes in construct abundance relative to the 1st day of harvesting (day 4 or 5) are calculated. Next a trend line is calculated, and the slope of each trend line is used in an outlier analysis to determine which constructs behave significantly different from the population

expression. In our hands the EF1α promoter (#CD530A-2, SBI = pCDH-miRNA) gave the highest miRNA expression levels in lymphoid cell lines. Alternatively, a microRNA precursor pooled virus library can be purchased from SBI (PMIRHPLVA-1 or PMIRHPLVAHT-1), which contains a CMV promoter.

2. To avoid bias in representation caused by differences in amplification efficiency due to variation in insert size, it is important to use constructs with similar insert sizes.
3. We observed that virus titers for the miRZip vectors are much higher than for miRNA overexpression vectors. Thus, achieving 20% infection efficiency for miRZip vector generally may require as little as 15–100 μL of virus supernatant, while for miRNA overexpression volumes of >100 μL are often required.
4. For some cell lines, spin infection can significantly increase the efficiency. After adding viral supernatant and polybrene, cells are spun down in a plate for 2 h at 1000 × *g*, 33 °C.
5. Some cell lines may be very sensitive to sorting and have poor survival rates after sorting. For these more sensitive cells, it

might be better to culture the original mixture of infected and non-infected cells for the entire screen. To obtain cells for DNA isolation, a proportion of the cell mixture can be used for flow sorting, followed by direct harvesting for DNA. So on each time point, a new sort is required. For the pCDH-miRNA vector, this can give problems as the GFP intensity is not that high and subtle sorting gate differences may affect representation of the constructs at different time points.

6. The miRZip vector carries a puromycin resistance gene. As an alternative approach to GFP sorting, selection with puromycin can be used starting at day 1 after infection. The minimal effective dose of puromycin should be determined earlier by performing a killing curve. Non-transduced cells should be treated with puromycin in parallel with transduced cells, and selection should be continued until no surviving cells are present in the control non-transduced cells (~4–7 days).
7. Barcodes should be attached to the 5′ end of the forward primer. Considerations for the minimal length and sequence distance between barcodes and barcode lists have been published [35, 36].
8. To automate the process of slope calculation of the trend lines, programming software can be used such as MATLAB.
9. For miRNA inhibition experiments, it is important to realize that one or more seed family members might be present in the cell type of interest and that these might be inhibited simultaneously with the miRNA of interest. So depending on the degree of homology between members of the same family, the effect of miRNA inhibition may come from the combined inhibition of multiple members of that family.

Acknowledgments

This work was supported by grants from the National Science Centre, Poland (grant no. 2016/23/D/NZ1/01611 to A.D.-K.), and the Pediatric Oncology Foundation Groningen, the Netherlands (SKOG 11-001 to J.K. and A.v.d.B.).

References

1. Chendrimada TP, Gregory RI, Kumaraswamy E et al (2005) TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing. Nature 436:740–744
2. Lewis BP, Burge CB, Bartel DP (2005) Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are MicroRNA targets. Cell 120 (1):15–20
3. Vasilatou D, Papageorgiou S, Pappa V et al (2009) The role of microRNAs in normal and malignant hematopoiesis. Eur J Haematol 84 (1):1–16

4. Koralov SB, Muljo SA, Galler GR et al (2008) Dicer ablation affects antibody diversity and cell survival in the B lymphocyte lineage. Cell 132(5):860–874
5. de Yébenes VG, Bartolomé-Izquierdo N, Ramiro AR (2013) Regulation of B-cell development and function by microRNAs. Immunol Rev 253(1):25–39
6. Eis PS, Tam W, Sun L et al (2005) Accumulation of miR-155 and *BIC* RNA in human B cell lymphomas. Proc Natl Acad Sci U S A 102 (10):3627–3632
7. Kluiver J, Haralambieva E, de Jong D et al (2006) Lack of BIC and microRNA miR-155 expression in primary cases of Burkitt lymphoma. Genes Chromosom Cancer 45 (2):147–153
8. Costinean S, Zanesi N, Pekarsky Y et al (2006) Pre-B cell proliferation and lymphoblastic leukemia/high-grade lymphoma in Eμ-miR155 transgenic mice. Proc Natl Acad Sci U S A 103(18):7024–7029
9. Ota A, Tagawa H, Karnan S et al (2004) Identification and characterization of a novel gene, *C13orf25*, as a target for 13q31-q32 amplification in malignant lymphoma. Cancer Res 64 (9):3087–3095
10. Robertus JL, Kluiver J, Weggemans C et al (2010) MiRNA profiling in B non-Hodgkin lymphoma: a MYC-related miRNA profile characterizes Burkitt lymphoma. Br J Haematol 149(6):896–899
11. Scholtysik R, Kreuz M, Klapper W et al (2010) Detection of genomic aberrations in molecularly defined Burkitt's lymphoma by array-based, high resolution, single nucleotide polymorphism analysis. Haematologica 95 (12):2047–2055
12. Xiao C, Srinivasan L, Calado DP et al (2008) Lymphoproliferative disease and autoimmunity in mice with increased miR-17-92 expression in lymphocytes. Nat Immunol 9:405–414
13. He L, Thomson JM, Hemann MT et al (2005) A microRNA polycistron as a potential human oncogene. Nature 435:828–833
14. Mu P, Han Y, Betel D et al (2009) Genetic dissection of the miR-17~92 cluster of microRNAs in Myc-induced B-cell lymphomas. Genes Dev 23(24):2806–2811
15. Olive V, Bennett MJ, Walker JC et al (2009) miR-19 is a key oncogenic component of mir-17-92. Genes Dev 23(24):2839–2849
16. Medina PP, Nolde M, Slack FJ (2010) OncomiR addiction in an in vivo model of microRNA-21-induced pre-B-cell lymphoma. Nature 467:86–90
17. Calin GA, Dumitru CD, Shimizu M et al (2002) Frequent deletions and down-regulation of micro- RNA genes *miR15* and *miR16* at 13q14 in chronic lymphocytic leukemia. Proc Natl Acad Sci U S A 99 (24):15524–15529
18. Klein U, Lia M, Crespo M et al (2010) The *DLEU2/miR-15a/16-1* cluster controls B cell proliferation and its deletion leads to chronic lymphocytic leukemia. Cancer Cell 17 (1):28–40
19. Shang W, Wang F, Fan G et al (2017) Key elements for designing and performing a CRISPR/Cas9-based genetic screen. J Genet Genomics 44(9):439–449
20. Eulalio A, Mano M (2015) MicroRNA screening and the quest for biologically relevant targets. J Biomol Screen 20(8):1003–1017
21. Voorhoeve PM, le Sage C, Schrier M et al (2006) A genetic screen implicates miRNA-372 and miRNA-373 as oncogenes in testicular germ cell tumors. Cell 124(6):1169–1181
22. Choi Y, Yoon S, Byun Y et al (2015) MicroRNA library screening identifies growth-suppressive microRNAs that regulate genes involved in cell cycle progression and apoptosis. Exp Cell Res 339(2):320–332
23. Maudet C, Mano M, Sunkavalli U et al (2014) Functional high-throughput screening identifies the miR-15 microRNA family as cellular restriction factors for Salmonella infection. Nat Commun 5:4718
24. Morris VA, Cummings C, Meshinchi S et al (2014) Functional miRNA expression library screen identifies miRNAs that alter proliferation and differentiation in acute myeloid leukemia. Blood 124(21):3541–3541
25. Du L, Borkowski R, Zhao Z et al (2013) A high-throughput screen identifies miRNA inhibitors regulating lung cancer cell survival and response to paclitaxel. RNA Biol 10 (11):1700–1713
26. Mullokandov G, Baccarini A, Ruzo A et al (2012) High-throughput assessment of microRNA activity and function using microRNA sensor and decoy libraries. Nat Methods 9:840–846
27. Nikolic I, Elsworth B, Dodson E et al (2017) Discovering cancer vulnerabilities using high-throughput micro-RNA screening. Nucleic Acids Res 45(22):12657–12670
28. Chang H, Yi B, Ma R et al (2016) CRISPR/cas9, a novel genomic tool to knock down microRNA in vitro and in vivo. Sci Rep 6:22312
29. Friedländer MR, Chen W, Adamidi C et al (2008) Discovering microRNAs from deep

sequencing data using miRDeep. Nat Biotechnol 26:407–415

30. Kluiver J, Slezak-Prochazka I, van den Berg A (2013) Studying microRNAs in lymphoma. Methods Mol Biol 971:265–276
31. Yuan Y, Kluiver J, Koerts J et al (2017) miR-24-3p is overexpressed in Hodgkin lymphoma and protects Hodgkin and Reed-Sternberg cells from apoptosis. Am J Pathol 187(6):1343–1355
32. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25(14):1754–1760
33. Li H, Handsaker B, Wysoker A et al (2009) The sequence alignment/map format and SAMtools. Bioinformatics 25(16):2078–2079
34. Tukey J (1977) Exploratory data analysis. Pearson, London, UK
35. Buschmann T, Bystrykh LV (2013) Levenshtein error-correcting barcodes for multiplexed DNA sequencing. BMC Bioinformatics 14 (1):272
36. Bystrykh LV (2012) Generalized DNA barcode design based on hamming codes. PLoS One 7(5):e36852

Chapter 13

RNA Sequencing in B-Cell Lymphomas

Da Wei Huang, Moez Dawood, Calvin A. Johnson, and Roland Schmitz

Abstract

High-throughput mRNA sequencing (RNA-Seq) provides both qualitative and quantitative evaluation of the transcriptome. This method uses complementary DNA (cDNA) to generate several millions of short sequence reads that are aligned to a reference genome allowing the comprehensive characterization of the transcripts in a cell. RNA-Seq has a wide variety of applications which lead to a pervasive adoption of this method well beyond the genomics community and a deployment of this technique as a standard part of the toolkit applied in life sciences. This chapter describes a protocol to perform mRNA sequencing using the Illumina NextSeq or MiSeq platforms, presents sequencing data quality metrics, and outlines a bioinformatic pipeline for sequence alignment, digital gene expression, identification of gene fusions, detection of transcript isoforms, description and annotation of genetic variants, and de novo immunoglobulin gene assembly.

Key words High-throughput sequencing, RNA-Seq, Transcriptome, Gene expression, B cell, B-cell lymphoma, Immunoglobulin genes, VDJ, Mutation

1 Introduction

In the last decade, high-throughput sequencing has become an integral part of cancer research. With this method, sequencing libraries are prepared from virtually any nucleic acid sample, amplified to produce clonal clusters, and sequenced using massively parallel synthesis. The rapid advancement of this technique has transformed numerous applications of structural and functional genomics leading to immense progress in our understanding of biology and pathogenesis of B-cell lymphomas [1–4]. These studies included analyses of whole genomes, exomes (the coding sequences of the genome), mRNA, and a combination of these approaches. High-throughput sequencing of mRNA represents an elegant way to study the structure and quantity of the transcripts in a cell. Compared with alternative high-throughput gene expression technologies, such as microarrays, RNA-Seq achieves a base pair-level resolution, a higher dynamic range of expression levels, and a lower background level [5]. Owing to the fact that RNA-Seq does not

Ralf Küppers (ed.), *Lymphoma: Methods and Protocols*, Methods in Molecular Biology, vol. 1956,
https://doi.org/10.1007/978-1-4939-9151-8_13,

rely upon a predetermined set of probe sequences, it also provides information on alternative spliced isoforms, novel transcripts, and gene fusions. RNA-Seq can detect rare transcripts with an abundance of one to ten RNA molecules per cell [6] facilitating the detection of mutations in oncogenes and tumor suppressor genes even though these transcripts might be present at low abundance [7].

This chapter contains the relevant protocols for performing RNA-Seq on Illumina MiSeq and NextSeq Systems using kits and reagents from commercial sources. Following RNA sample quantification and quality control, a double-stranded cDNA library is generated, controlled for size range, and quantified. The denatured cDNA library is annealed to oligos that are covalently bound to the surface of the flow cell. These oligos prime the synthesis of a complementary DNA strand creating a double-stranded DNA template of which one strand is immobilized on the flow cell. Following denaturation, the free ends of these covalently attached DNA molecules are bound by neighboring oligos on the surface of the flow cell, which in turn prime the synthesis of a complementary DNA strand that is consequently also covalently bound to the flow cell. This process is repeated to create local clusters of identical DNA molecules. The sequencing reaction in the Illumina instrument is performed using a sequencing primer that is complementary to the adapter sequence attached to each template strand. The four different bases (A, G, T, C) are labeled with different colored fluorophores (MiSeq) or a mix of fluorophores (NextSeq). Following incorporation of the labeled bases, the camera of the instrument records the individual colors for each cluster, and the software modules convert the dye information into DNA sequence data with associated intensity and quality scores. The subsequent bioinformatic analysis of RNA-Seq is done in several stages. In the first step, the sequencing reads are mapped. The correct alignment of short RNA sequences is a crucial, albeit complex, task. Depending on the experimental setup, RNA-Seq generates tens of millions of sequence reads corresponding to complex transcriptomes. Several short-read assembly programs have been developed to tackle these problems applying different mathematical algorithms and scoring schemes [8]. The alignment algorithm used in the outlined RNA-Seq pipeline (STAR2) is as a gapped aligner that splits intron-exon spanning reads at exon boundaries during alignment to the genome [9]. After the data has been mapped to the genome, an expression score for every mapped read is computed. The genome alignment is parsed to identify genetic variants that are further annotated based on their predicted function. Furthermore, alternative alignment algorithms are used to identify and enumerate transcript variants including gene fusions and alternatively spliced isoforms. In a final step, de novo assembly of expressed

immunoglobulin genes is computed, and the resulting contigs are annotated based on the conserved cysteine-104 and tryptophan-118 [10].

2 Materials

2.1 Sample Quality Control

1. Agilent RNA 6000 Nano Kit (Agilent Technologies).
2. Rat Brain Total RNA (Life Technologies).
3. Agilent 2100 Bioanalyzer (Agilent Technologies).
4. RNaseZAP (Ambion).
5. DNAse-/RNAse-free water (DEPC-treated).
6. Bioanalyzer Chip Vortexer (IKA MS 3).

2.2 Library Preparation

1. TruSeq RNA Sample Preparation v2 (Illumina).
2. 0.2 mL clear thin-wall PCR strip tubes.
3. DynaMag™-96 Side (Invitrogen).
4. SuperScript II Mix (Invitrogen).
5. Agencourt AMPure XP 60 mL (Beckman Coulter).
6. Ethanol Absolute.
7. Tris–Cl 10 mM, pH 8.5 with 0.1% Tween 20.

2.3 Library Quality Control

Agilent DNA 1000 Kit (Agilent Technologies).

2.4 qPCR

1. Illumina Eco Real-Time PCR System (Illumina, Inc.).
2. SYBR FAST Master Mix (KAPA Biosystems).
3. Illumina GA Primer Premix (KAPA Biosystems).
4. 6× Illumina GA DNA Standards (KAPA Biosystems).
5. DNAse-/RNAse-free water.
6. 10 m*M* Tris–HCl + 0.05% Tween-20 (DNase-/RNase-free).
7. Eco Plates (Illumina).
8. Eco Adhesive Seals (Illumina).

2.5 Sequencing

1. MiSeq Reagent Kit v3 or NextSeq 500/550 v2 Kit.
2. MiSeq or NextSeq Systems.

2.6 RNA-Seq Data Quality Control and Assessment

Software:

Trimmomatic: http://www.usadellab.org/cms/?page=trimmomatic

FastaQC: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

2.7 Alignment of Sequences to Reference

Software:

STAR2: https://code.google.com/archive/p/rna-star/

2.8 Digital Gene Expression

Software:

HTseq-Count: http://htseq.readthedocs.io/en/master/count.html

DESeq2: https://bioconductor.org/packages/release/bioc/html/DESeq2.html

GSEA: http://software.broadinstitute.org/gsea/index.jsp

Cluster: http://bonsai.hgc.jp/~mdehoon/software/cluster/manual/

TreeView: http://jtreeview.sourceforge.net/

2.9 Gene Fusion Detection

Software:

FusionCatcher: https://github.com/ndaniel/fusioncatcher

IGV: http://software.broadinstitute.org/software/igv/

2.10 Alternative Splicing Discovery

Software:

Cufflinks: http://cole-trapnell-lab.github.io/cufflinks/manual/

IGV: http://software.broadinstitute.org/software/igv/

2.11 Identification and Annotation of Putative Single-Nucleotide Variants

1. Software:
   SAMtools: http://samtools.sourceforge.net/
   VarScan2: http://varscan.sourceforge.net/
   SnpEff: http://snpeff.sourceforge.net/
   ANNOVAR: http://annovar.openbioinformatics.org/en/latest/

2.12 Immunoglobulin Gene Assembly

1. Software:
   BLAST: http://blast.ncbi.nlm.nih.gov
   Trinity: https://github.com/trinityrnaseq
   IMGT High V-QUEST: https://www.imgt.org/HighV-QUEST/login.action
   Download reference sequences:
   Immunoglobulin gene segments: http://www.imgt.org/genedb

3 Methods

3.1 Sample Quality Control

High-throughput mRNA sequencing relies on high-quality RNA for successful library preparation. Degraded RNA can result in low yield, over presentation of the 5′ ends of RNA, or failure of library

preparation. Sequencing libraries using low-quality RNA, including RNA from formalin-fixed, paraffin-embedded (FFPE) tissues, should be prepared with kits designed for capturing coding RNA and specific noncoding RNA.

The RNA quality is measured with the Agilent Bioanalyzer which detects the fluorescence intensity of 18S and 28S ribosomal RNA and calculates the RNA integrity number (RIN).

1. Prepare the gel by pipetting 550 μL of RNA-6000 nanogel matrix into a spin filter. Centrifuge at 1500 × *g* for 10 min at room temperature. Aliquot 65 μL into RNase-free microfuge tubes (*see* **Note 1**).
2. Vortex the RNA 6000 Nano dye concentration for 10 s, spin down, and add 1 μL of dye into a 65 μL aliquot of filtered gel. Vortex and spin at 13,000 × *g* for 10 min (*see* **Note 2**).
3. Heat denature RNA ladder for 2 min at 70 °C and cool on ice for 1 min (*see* **Note 3**).
4. Put an RNA 6000 Nano chip on the chip priming station. Pipette 9 μL of gel-dye mix in the well marked G. Position the plunger at 1 mL and close the chip priming station. Make sure the priming station is latched in by pushing the silver button. Press the plunger until it is held by the clip. Wait for exactly 30 s and release the clip. Wait for additional 5 s. Pull plunger slowly back to 1 mL position. Open chip priming station and pipette 9 μL of gel-dye mix in the wells marked G.
5. Pipette 5 μL of RNA 6000 Nano Marker in all 12 sample wells and in the well marked with the ladder symbol.
6. Pipette 1 μL of prepared ladder in the well marked with the ladder symbol. Pipette 1 μL of sample in each of the 11 sample wells and Rat Brain Total RNA control in well 12 (*see* **Note 4**). Pipette 1 μL of RNA 6000 Nano Marker in each unused sample well. Put the chip in the adapter of the chip vortexer and vortex for 1 min at 2400 *rpm*. Run the chip in the Agilent 2100 Bioanalyzer within 5 min.

3.2 Planning Sequencing Runs

The number of samples per sequencing run depends on the desired read-depth and downstream application. For gene expression analyses, a minimum of 10–20 million reads can be sufficient, whereas the detection of genetic variants including gene fusions can require at least 50–100 million reads. Note that pooling of sequencing libraries can result in variation of the yield of individual libraries despite accurate quantification. The choice of the read length as well as single- or paired-end runs is also influenced by the analytical method. While gene expression analyses can be done with single-end reads of 75 bp, single-nucleotide variants and gene fusion detection require longer, paired-end reads.

3.2.1 Setup of Sequencing Run

1. Use the Illumina Experiment Manager to create a sample plate, select the appropriate library kit, and enter a unique plate name.
2. Select 1 for single-indexed libraries in the index reads field.
3. Use the table or plate view to enter a unique sample ID and index adapter used, and then save the sample plate file.
4. Use the Illumina Experiment Manager to create a sample sheet and select NextSeq or MiSeq.
5. Enter the reagent kit ID from the label of box 1 or box 2 of the SBS kit in the reagent kit barcode field.
6. Select the appropriate library prep workflow and index adapters. Select 1 for single-indexed libraries in the index reads field.
7. Select the read type (single or paired end), and enter the number of cycles adding 1 to the number of cycles (e.g., for a 75-cycle read, enter 76).
8. Select sample plate created in **step 1**, choose samples to include, and select Finish to save sample sheet file.

3.3 Library Preparation

0.1–1 μg of total RNA is required for cDNA library construction (*see* **Note 5**).

3.3.1 Purification and Fragmentation of mRNA

1. Dilute the total RNA with DNAse-/RNAse-free water to a final volume of 50 μL.
2. Add 50 μL of RNA purification beads to each well of the RBP plate using a multichannel pipette to bind mRNA to oligo-dT magnetic beads (*see* **Note 6**). Mix by gently pipetting up and down about six times.
3. Seal plate and denature RNA in a thermal cycler (65 ° C for 5 min, 4 °C hold). Remove from the thermal cycler.
4. Incubate at room temperature for 5 min to allow RNA to bind to the beads.
5. Place plate on magnetic stand at room temperature for 5 min to separate beads from the solution (*see* **Note 7**).
6. Carefully remove and discard the supernatant using a multichannel pipette. Remove plate from the magnetic stand.
7. Wash beads by adding 200 μL of bead wash buffer to remove unbound RNA (*see* **Note 8**).
8. Place plate on magnetic stand at room temperature for 5 min.
9. Carefully remove and discard supernatant. Remove plate from magnetic stand.
10. Add 50 μL of mRNA elution buffer. Mix gently by pipetting.
11. Seal plate and elute mRNA in a thermal cycler (80 °C for 2 min, 25 °C hold) from the beads.

12. Remove plate from thermal cycler, add 50 μL of bead-binding buffer, and incubate at room temperature for 5 min.
13. Place plate on magnetic stand at room temperature for 5 min (*see* **Note 7**).
14. Carefully remove and discard entire supernatant.
15. Remove the plate from magnetic stand, and wash beads by adding 200 μL of bead wash buffer to remove unbound RNA.
16. Place plate on magnetic stand at room temperature for 5 min.
17. Carefully remove and discard supernatant. Remove plate from magnetic stand.
18. Add 19.5 μL of Elute, Fragment Mix. Mix gently by pipetting (*see* **Note 9**).
19. Place the sealed plate in a thermal cycler and elute fragment, and prime the RNA using the program: 94 °C for 8 min and 4 °C hold.
20. Remove from thermal cycler and spin briefly. Proceed immediately to synthesize first-strand cDNA.

3.3.2 Synthesize First-Strand cDNA

1. Place plate on magnetic stand at room temperature for 5 min.
2. Transfer 17 μL of supernatant (fragmented and primed mRNA) to a new plate.
3. Add 8 μL of First-Strand Master Mix and SuperScript II mix. Mix by pipetting.
4. Incubate plate in a thermal cycler: 25 °C for 10 min, 42 °C for 50 min, 70 °C for 15 min, and 4 °C hold.
5. Remove from thermal cycler, and proceed immediately to synthesize second-strand cDNA.

3.3.3 Synthesize Second-Strand cDNA

1. Add 25 μL of Second-Strand Master Mix. Mix by pipetting.
2. Incubate in a thermal cycler, at 16 °C for 1 h.
3. Remove the plate from thermal cycler. Bring the reaction mixture to room temperature.

3.3.4 Ampure XP Cleanup

1. Vortex the Ampure XP beads until they are well dispersed, and add 90 μL Ampure XP beads to reaction mixture. Mix by vortexing.
2. Incubate at room temperature for 15 min.
3. Place the plate on magnetic stand at room temperature for 5 min (*see* **Note 7**).
4. Carefully remove and discard 135 μL of supernatant. Some liquid may remain in the wells (*see* **Note 10**).

5. With the plate remaining on the magnetic stand, carefully add 200 μL of freshly prepared 80% EtOH.
6. Incubate at room temperature for 30 s, and then carefully remove and discard the entire supernatant. Repeat **steps 5** and **6** once more for a total of two 80% EtOH washes.
7. Dry beads at room temperature for 15 min and remove plate from magnetic stand.
8. Add 52.5 μL resuspension buffer. Mix by pipetting.
9. Incubate at room temperature for 2 min.
10. Place the plate on magnetic stand at room temperature for 5 min (*see* **Note 7**).
11. Transfer 50 μL of supernatant (double-stranded cDNA) to a new 0.2 mL PCR plate. Library preparation can be interrupted at this step. cDNA can be stored up to 7 days at −20 °C.

3.3.5 End Repair

1. Add 10 μL of resuspension buffer.
2. Add 40 μL of end repair mix. Mix by pipetting.
3. Incubate in a thermal cycler at 30 °C for 30 min. Library preparation can be interrupted at this step. cDNA can be stored up to 7 days at −20 °C.

3.3.6 Ampure XP Cleanup

1. Add 160 μL of mixed Ampure XP beads. Mix by pipetting.
2. Incubate at room temperature for 15 min.
3. Place the plate on magnetic stand at room temperature for 5 min (*see* **Note 7**).
4. Carefully remove and discard 127.5 μL of supernatant. Some liquid may remain in the wells (*see* **Note 11**).
5. Wash as described in **steps 5** and **6** of Sect. 3.3.4.
6. Resuspend the dried pellet with 25 μL resuspension buffer. Mix well by pipetting.
7. Incubate at room temperature for 2 min.
8. Place the plate on magnetic stand at room temperature for at least 5 min (*see* **Note 7**).
9. Transfer 17.5 μL of supernatant to a new 0.2 mL PCR plate. Library preparation can be interrupted at this step. cDNA can be stored up to 7 days at −20 °C.

3.3.7 Adenylate 3′ Ends

1. Add 12.5 μL of A-tailing mix. Mix by pipetting.
2. Incubate at 37 °C for 30 min, 70 °C for 5 min, and 4 °C hold.
3. Proceed immediately to ligate adapters.

3.3.8 Ligate Adapters

1. Add 2.5 μL of DNA ligase mix, 2.5 μL of resuspension buffer, and 2.5 μL of each RNA Adapter Index. Mix by pipetting.
2. Incubate in a thermal cycler at 30 °C for 10 min.
3. Remove the plate from the thermal cycler.
4. Inactivate ligation by adding 5 μL of Stop Ligase Mix. Mix by pipetting.

3.3.9 Ampure XP Cleanup

1. Add 42 μL of mixed Ampure XP beads. Mix by pipetting.
2. Incubate at room temperature for 15 min.
3. Place the plate on the magnetic stand at room temperature for at least 5 min (*see* **Note 7**).
4. Carefully remove and discard supernatant (*see* **Note 10**).
5. Wash as in **steps 32–34**.
6. Resuspend the dried pellet with 52.5 μL resuspension buffer. Mix by pipetting.
7. Incubate at room temperature for 2 min.
8. Place plate on magnetic stand at room temperature for at least 5 min (*see* **Note 7**).
9. Transfer 50 μL of the supernatant into a new plate.
10. Add 50 μL of mixed Ampure XP beads. Mix by pipetting.
11. Incubate at room temperature for 15 min.
12. Place the plate on the magnetic stand at room temperature for at least 5 min (*see* **Note 7**).
13. Carefully discard 95 μL of the supernatant. Some liquid may remain in each well (*see* **Note 10**).
14. Wash as in **steps 32–34**.
15. Add 22.5 μL resuspension buffer. Mix by pipetting.
16. Incubate at room temperature for 2 min.
17. Place plate on magnetic stand at room temperature for at least 5 min (*see* **Note 7**).
18. Transfer 20 μL of the supernatant to the corresponding wells of the plate of **step 10**. Library preparation can be interrupted at this step. cDNA can be stored up to 7 days at −20 °C.

3.3.10 Enrich DNA Fragments

1. Add 5 μL of PCR Primer Cocktail.
2. Add 25 μL of PCR Master Mix. Mix by pipetting.
3. Amplify cDNA library using the PCR program: 98 °C for 30 s, 15 × (98 °C for 10 s, 60 °C for 30 s, 72 °C for 30 s), 72 °C for 5 min, and hold at 4 °C.
4. Add 50 μL of mixed Ampure XP beads. Mix by pipetting.
5. Incubate at room temperature for 15 min.

6. Place plate on magnetic stand at room temperature for at least 5 min (*see* **Note 7**).
7. Carefully discard 95 μL of the supernatant. Some liquid may remain in each well (*see* **Note 11**).
8. Wash as in **steps 32–34**.
9. Resuspend dried pellet with 32.5 μL resuspension buffer. Mix by pipetting.
10. Incubate at room temperature for 2 min.
11. Place plate on magnetic stand at room temperature for at least 5 min (*see* **Note 7**).
12. Carefully transfer 30 μL of the supernatant to a new plate.

3.4 Library Quality Control

The cDNA library is analyzed using an Agilent DNA 1000 Bioanalyzer chip to validate quality and size range of the library. The final product for paired-end libraries should be a band that peaks at approximately 280 bp with an average size of generally 330 bp.

1. Bring DNA dye concentrate and DNA gel matrix to room temperature. Vortex DNA dye concentrate for 10 s and spin down (*see* **Note 11**).
2. Pipette 25 μL of dye concentrate into DNA gel matrix vial.
3. Vortex for 10 s. Check proper mixing of gel and dye.
4. Transfer the gel-dye mix to the top receptacle of a spin filter. Place the spin filter in a microcentrifuge, and spin for 15 min at room temperature at 2240 × *g*.
5. Place DNA chip on the chip priming station.
6. Pipette 9 μL of the gel-dye mix at the bottom of the well marked G (*see* **Note 12**). Position plunger at 1 mL and close the chip priming station. Make sure the priming station is latched in by pushing the silver button.
7. Press the plunger of the syringe down until it is held by the clip. Wait for exactly 60 s and then release the plunger with the clip release mechanism.
8. Wait for 5 s, and then slowly pull back the plunger to the 1 mL position. Open the chip priming station.
9. Pipette 9 μL of gel-dye mix in each of the wells marked.
10. Pipette 5 μL of DNA marker into the well marked with the ladder symbol and into each of the 12 sample wells (*see* **Note 13**).
11. Pipette 1 μL of the DNA ladder in the well marked with the ladder symbol. In each of the 12 sample wells, pipette 1 μL of sample for used wells or 1 μL of deionized water for unused wells.

Table 1
qPCR reaction mix

[μL]	Reagent
5	2× SYBR Master Mix
1	10× Illumina GA Primer Premix
2	Water
2	DNA template, GA Standard, or water

12. Place the chip horizontally in the adapter of the chip vortexer. Vortex for 60 s at 2400 *rpm*. Run within 5 min.

3.5 qPCR

Accurate quantification of the cDNA library is critical for successful cluster generation. For libraries of a narrow insert size range, this technique is a very effective method for quantification as it measures only functional library fragments rather than all DNA species within a sample (e.g., primer dimers, free nucleotides, library fragments).

1. Dilute cDNA libraries 1:10,000 (*see* **Note 14**).
2. Pipette qPCR reaction mix (Table 1).
3. Place a 48-well qPCR plate on the dock assembly, and adjust the angle of the dock as desired.
4. Pipette 8 μL of the qPCR reaction mix into in the sample plate.
5. Add 2 μL of either DNA, Illumina DNA Standards (1–6), or water into sample wells.
6. Seal the 48-well microplate with adhesive film.
7. Vortex and spin plate. No air bubbles should be visible.
8. Load the sample plate into the qPCR system (*see* **Note 15**).
9. Normalize the concentration of sample libraries to 10 nM.

3.6 Performing Sequencing Run

1. Combine 5 μL of 4 nM sample library and 5 μL of freshly prepared 0.2 N NaOH in a microcentrifuge tube. Vortex briefly and incubate at room temperature for 5 min.
2. Add 5 μL 200 mM Tris–HCl, pH 7. Vortex briefly and incubate at room temperature for 1 min.
3. Add 985 μL prechilled HT1 to the denatured library. Vortex briefly and centrifuge at 280 × *g* for 1 min. Place the 20 pM libraries on ice.
4. For MiSeq, load 600 μL of 20 pM library into the reagent cartridge. For NextSeq, dilute 20 pM sample library to 1.8 pM using chilled HT1, and load 1.3 mL into the reagent cartridge.

5. Clean and load flow cell. Empty waste container, load buffer container, and load reagent cartridge.
6. Review run parameters and start sequencing run.

3.7 RNA-Seq Data Quality Control and Assessment

Quality control (QC) for RNA-Seq data should be performed before implementing any of the downstream data analysis. Sequencing runs might need to be repeated with improved conditions if any of the major QC criteria failed. In lieu of satisfying QC, downstream analysis of RNA-Seq data could run into a typical situation of "garbage in and garbage out" if forcing poor quality into an ill-fitting pipeline.

A variety of publicly available software tools exist for RNA-Seq QC and analysis. This section focuses on data quality control and assessment based on Illumina software tools.

3.7.1 Instrument Run Monitoring and Data Evaluation

During the Illumina instrument run for the MiSeq and the NextSeq platforms, the data quality control is accessible on BaseSpace or viewable from an off-instrument (remote) computer using the Sequencing Analysis Viewer (SAV). Table 2 lists acceptable run performance metrics based on control sample (e.g., Phix). Performance may vary for a given mRNA sample based on sample quality and other experimental factors. The density of clusters on a flow cell is a critically important metric that influences run quality, including reads passing filter, Q30 scores, and total yield. Underclustering usually results in high data quality but also produces low data yield. In contrast, overclustering creates image analysis problems that can lead to poor run performance such as lower Q30 scores and lower total data output. Causes for under- and overclustering include library quality issues (e.g., contamination with adapter dimers or partial library constructs) and inaccurate library quantification.

3.7.2 Post-sequencing Data Processing

1. Download FASTQ files and SAV files from BaseSpace by selecting download of the appropriate run in run tab.
2. Perform adapter clipping and low-quality trimming using the Java-based Trimmomatic software.
3. Generate quality reports using the FastQC software (Table 3).

3.8 Alignment of Sequences to Reference

Essentially, there are two assembly strategies for RNA-Seq data. The sequence reads can be directly aligned to a reference such as the genome or transcriptome or assembled de novo and subsequently aligned. While de novo assembly constructs a whole genome transcriptome map allowing for discovery of previously unknown transcripts, de novo assembly requires greater computational resources and is less accurate at mapping low-abundance transcripts. Thus, except for immunoglobulin gene assembly (Subheading 3.13), this chapter will focus on reference-based mapping strategies. For the reference-based mapping, a major challenge is the alignment of

Table 2
Run performance metrics

Matrix	Description	Optimal value range
Cluster density (K/mm^2)	Density of clusters detected by image analysis	150–250 K/mm^{2a} 1200–1400 K/mm^{2b}
Cluster PF (%)	Percentage of clusters passing intensity filter criteria	>80%
Q30+ (%)	Percentage of bases with a Phred score of 30 or higher	≥80%
Estimated yield	The projected number of bases expected to be sequenced at the end of the run	See specifications for sequencing kits (www.illumina.com)

[a]NextSeq
[b]MiSeq v3

Table 3
Quality assessment using FASTQ

Matrix	Fail QC value range
Per base sequence quality score	If the lower quartile for any base is less than 5 or if the median for any base is less than 20
Per sequence quality score	If the most frequently observed mean quality for sequences is below 20
Per base sequence content	If the difference between A and T or G and C is greater than 20% in any position
Sequence GC content	If the sum of the deviations from the normal distribution represents more than 30% of the reads
Sequence duplication levels	If nonunique sequences make up more than 50% of the total
Overrepresented sequences	If any sequence is found to represent more than 1% of the total

RNA-Seq reads that span across intron-exon boundaries. STAR2 is as a gapped aligner that, in contrast to the linear alignment algorithms BWA or BOWTIE, can effectively split these reads at the intron-exon junction during alignment to the genome [9, 11, 12]. In addition, the two-pass method by STAR2 aligns each read group separately and then merges the resulting alignments to improve alignment accuracy. Other parameters that control mapping have been outlined by the NCI Genomic Data Commons (GDC) platform facilitating the standardization of analytical RNA-Seq pipelines (*see* **Note 16**). The output BAM file from STAR2 alignment is compatible with most other analysis tools such as visualization of aligned target sequences in the Integrated Genomics Viewer (IGV) [13].

3.9 Digital Gene Expression

Digital gene expression (DGE) analysis is one of the most common applications of RNA-Seq [14]. In comparison with the hybridization-based microarray methods, DGE gives a broader dynamic range, provides better resolution, and enhances the fold change determination, particularly of transcripts expressed at a very high or low level. Similar to the microarray technology, the first key step of DGE analysis is to effectively normalize the sample set within a given study to make the data cross-comparable. FPKM (fragment per kilobase per million) is perhaps the most popular choice (preferable in principle to RPKM (read per kilobase per million)), since FPKM normalization avoids redundant counts on paired-end reads. The HTseq-count package can be used directly on the STAR2 BAM file for FPKM calculation [15]. The fold change of each gene between biological conditions can be estimated based on their FPKM values. DESeq2 is a popular software package that calculates, which is based on an advanced negative binomial distribution model rather than simple division calculation. In addition, DESeq2 offers a number of enhancements for comparative RNA-Seq analysis including "shrinkage" estimators to better estimate dispersion and fold change [16]. The biological context of the generated gene lists of differentially expressed genes by DESeq2 can be interpreted using pathway enrichment analysis or a gene ontology enrichment tools, such as GSEA [17, 18]. DGE profiles can be further analyzed by Cluster software, followed by TreeView heat map view for the interactive visualization [19].

3.10 Gene Fusion Detection

RNA-Seq allows for gene fusion discovery [20]. This analytical method is based on the detection of fusion events represented by chimeric reads or read pairs that consist of sequences from two different locations in the genome. Though RNA-Seq has been successfully used to accurately detect gene fusion events in lymphomas, the methodology can generate numerous artifacts [3]. These stem mainly from reads that are too short to be uniquely identified in the genome, reads that align to repetitive regions, mapping of reads in genomic regions of great homology, or imperfection in the library preparation. FusionCatcher is a gene fusion detection software that utilizes FASTQ files to produce gene fusion models including an assembled junction gene [21]. Fusion events that are supported by a low number of junction reads should be filtered out. Focused analysis of gene fusion of interest using BLAST searches with the supporting junction assemblies is essential to manually confirm identified fusion events. Further validation using RT PCR approaches is recommended.

RNA-Seq data offers the opportunity to examine alternative splicing (AS) events on a genome-wide scale [22]. The study of AS can be divided into two subclasses of analyses: first, the differential gene

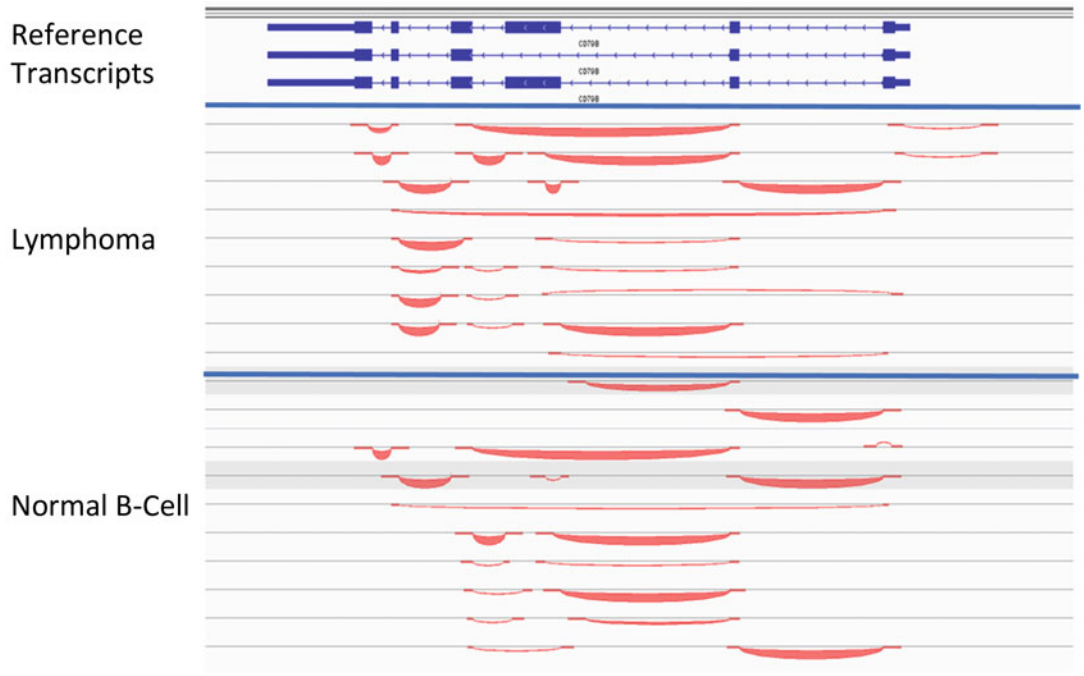


Fig. 1 Visualization of genome alignment of RNA-Seq reads used for alternative splicing discovery. Junction reads, reported from Cufflinks software, are displayed by IGV. The thickness of the red line is proportional to the observed number of junction spanning reads. The black rectangle indicates irregular alternative splicing events between the two exons in lymphoma and normal B-cell control

3.11 Alternative Splicing Discovery and Visualization

expression of known transcript isoforms and, second, the detection of novel transcripts and structural aberrations of known isoforms. Data mining of AS is arguably the most challenging analysis among the various RNA-Seq applications. This is due to difficulties in distinguishing reads that can be shared by multiple transcripts, potential contamination of genomic DNA, and limitation of the ability to detect large structural alterations using short reads. Cufflinks is a broadly established software for AS analysis, utilizing FASTQ files to produce results for both FPKM on isoform level and transcript structure as a general feature format (GTF) file. To analyze differentially expressed isoforms, fold change calculations on isoform FPKM can be done following the methodology described in Subheading 3.9. Moreover, using the junction bed file from Cufflinks, structural aberration of isoforms can be visualized with IGV (Fig. 1).

3.12 Genetic Variation Identification and Annotation

Identification of single-nucleotide variants (SNVs) and other types of structural variants (e.g., insertions and deletions) is achieved mainly through the Exome-Seq or Genome-Seq techniques. However, matched RNA-Seq experiments can serve as validation data sets of genetic variants obtained by high-throughput

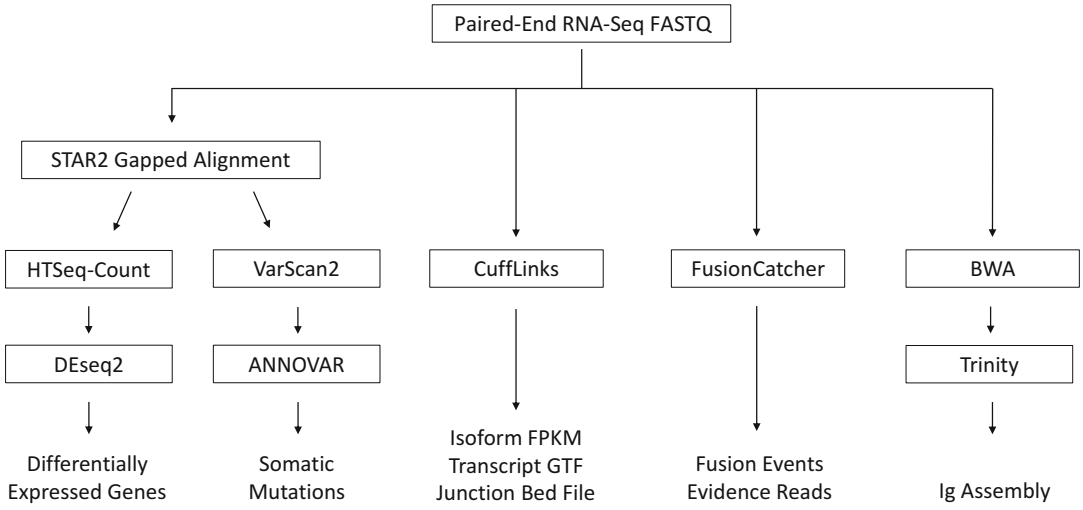


Fig. 2 Schematic workflow of RNA-Seq pipeline

DNA-sequencing approaches [23]. Compared to Exome-Seq, variant calling using RNA-Seq data can result in a greater level of false positives. These artifacts result primarily from mismapping of intron-exon spanning reads using a gapped alignment algorithm. VarScan2 processes BAM alignment via SAMtools to produce a variant calling file (*VCF*) (Fig. 2). To understand the biological effects of a SNV, the annotation of the observed variant is paramount. While the majority of SNVs discovered by RNA-Seq are single-nucleotide polymorphisms (SNPs), the main focus in cancer biology lies on the identification of non-SNP mutations with potential to activate or inactivate gene functions. Hence, in the first step of the annotation process, SNVs that corresponded to known SNPs are identified based on the genomic coordinates. The interpretation of potentially somatic, non-synonymous SNVs can be facilitated using annotation databases, such as ANNOVAR or SnpEff, that offer dozens of annotation categories and are available online or as locally run scripts [24, 25].

3.13 Immunoglobulin Gene Assembly

Most B-cell non-Hodgkin lymphomas (NHL) express a B-cell receptor (BCR) [26]. The BCR is a membrane-bound antibody composed of two identical heavy chain and two identical light chain immunoglobulin (Ig) polypeptides, non-covalently bound to the signaling components CD79A and CD79B. Immunoglobulins consist of variable (V) regions and constant (C) regions. The heavy chain V regions are composed of three (V, D, and J) gene segments, whereas the light chains consist of two gene segments (V and J). During B-cell development, these gene segments are assembled by somatic DNA rearrangement to encode a functional Ig. This Ig gene segment assembly produces non-germline-

encoded immunoglobulin molecules that are strongly expressed in most mature B cells. Consequently, most B-cell NHL express unique Ig heavy chain and Ig light chain molecules allowing de novo assembly, mapping, and annotation of these transcripts in 80–90% of NHL [27].

3.13.1 De Novo Assembly of Ig Genes

1. To map and assemble Ig transcripts from RNA-Seq, Ig germline segments including V, D, and J and constant regions of heavy and light chain are downloaded from the Immunogenetics (IMGT) database (*see* **Note 17**) [28].
2. FASTQ files are aligned to RefSeq database (Build 37) using the Burrows-Wheeler Aligner (BWA) tool [12].
3. Reads that do not map to RefSeq are aligned to the Ig reference sequences using the BLAST algorithm allowing for three mismatches per read. The BLAST alignments are parsed using PERL, to determine the number and sequence of reads mapping to the various Ig gene segments.
4. Mapped reads are assembled de novo using Trinity (Version July 17, 2014) with default parameters and standard instructions on the "Running Trinity" webpage [29].
5. Using the assemblies from Trinity (contigs), set up a FASTA file containing sequence IDs from **step 1** with unique sample names.
6. For each lymphoma sample, the reads utilized to generate the Trinity assemblies are remapped to these contigs using BLAST. The percentage of reads that map to each contig is parsed out for abundance statistic used for data filtering.

3.13.2 Data Filtering and Annotation of Ig Gene Contigs

Although the reads that entered de novo assembly were stringently filtered, the method creates several artifacts, including several hundred contigs with un-rearranged or incomplete *V* genes. Therefore, the resulting contigs are filtered based on the following assumptions: An expressed Ig transcript should include a productive VDJ rearrangement without premature stop codons. The contigs supported by the greatest number of raw reads are the best possible description of the expressed Ig. Detailed annotation of the contigs representing potential Ig transcripts by IMGT High V-Quest is used for filtering the data. The following steps below are written for Microsoft Excel for simplicity but can be accomplished in other programming language as well:

1. Process the FASTA file with IMGT High V-Quest (*see* **Note 18**) [10].
2. Using Excel, open the resulting text file found in the output of the IMGT High V-Quest analysis.
3. Filter out light chains for analysis at a later time.

4. Using a pivot table, filter out every entry except *Productive*, *Productive (See Comment)*, and *Unknown (See Comment)* in *Functionality* column.
5. In the *Functionality Comment* column, filter out *No rearrangement found* and *No rearrangement found stop codons*. This ensures all remaining data corresponds to a functional VDJ recombination.
6. Split the *FR-IMGT lengths* column into four separate columns, one for each framework region denoting the number of amino acids in each framework region (*see* **Note 19**).
7. To reduce the number of incomplete Ig assemblies, filter all sequences to have more than 23 amino acids listed for framework 1. No restrictions are placed on the other frameworks, to prevent instances of insertion/deletions or recombined V genes from being parsed out. The IMGT output also annotates insertions or deletions in V genes (*see* **Note 20**).
8. Filter out all sequences with an amino acid junction out of frame. Now all productive VDJ rearrangements with complete (or near complete) coverage of the V gene should be isolated (*see* **Note 21**).
9. Sort the remaining sequences from top to bottom based on the percentage of RNA-Seq reads used to compile each contig (*see* above). Sort further based on this scheme:
   (a) Sample name, A to Z.
   (b) V gene abundance statistic, largest to smallest
   (c) C gene abundance statistic, largest to smallest
   (d) J gene abundance statistic, largest to smallest
   (e) Sequence length, largest to smallest.
10. At this point, all the productive VDJ rearrangements observed in a certain sample are sorted in the order from strongest to weakest evidence (*see* **Note 22**).

4 Notes

1. Kit should be used at room temperature and protected from light for the whole procedure.
2. Use prepared gel-dye mix within 1 day.
3. Aliquots can be stored at −70 °C. Thaw aliquots on ice.
4. It is recommended to heat denature all RNA samples and RNA ladder before use.
5. The RIN of total RNA used for library preparation should be greater or equal to 8.

6. Vortex the thawed RNA purification beads' tube vigorously to completely resuspend the oligo-dT beads.
7. All beads should be attached to the side of the wells. The liquid should appear clear.
8. Remove the entire ethanol from the bottom of the wells. Residual ethanol can hamper downstream enzymatic reactions and may contain contaminants.
9. The Elute, Prime, Fragment Mix contains reaction buffer and random hexamers for priming of reverse transcription.
10. Leave the plate on the magnetic stand while performing the following 80% EtOH washes.
11. The gel-dye mix can be stored for 4 weeks at 4 °C.
12. When pipetting the gel-dye mix, do not draw up particles that may sit at the bottom of the gel-dye mix vial. Insert the tip of the pipette to the bottom of the chip well when dispensing. This prevents a large air bubble forming under the gel-dye mix. Placing the pipette at the edge of the well may lead to poor results.
13. Empty wells may cause improper running of the chip. Add 5 μL of DNA marker plus 1 μL of deionized water to each unused sample well.
14. Optimal quantification of cDNA libraries is achieved using an input DNA amount of 1 pM.
15. If the concentration for any sample is greater than 50 nM, the sample should be diluted and rerun in the qPCR.
16. A Linux computing environment is highly recommended to process the RNA-Seq data. It is better to have a computer system with >60 GB memory and >8 CPUs. If there are more than ten samples, a clustered computer arrangement is highly recommended to run the analysis in parallel.
17. To download of the reference sequences of V, D, and J or constant region gene segments, submit IMGT/GENE-DB query selecting the respective identification and localization (species, molecular component, and locus/main locus). Select all resulting gene segments, and choose display of IMGT/GENE-DB allele reference sequences in FASTA format.
18. Use the following parameters: species, human; receptor type or locus, IG; multiple individuals; include individual result files; all options checked in detailed view; and all options checked in CSV view.
19. In "FR-IMGT lengths" IMGT has data about the lengths (in terms of the number of amino acids) of each framework region of the V gene in the form of [Framework 1. Framework 2. Framework 3. Framework 4]. According to IMGT

references, complete lengths for framework regions 1–4 are [25.17.38.11]. However, the length of framework 4 can vary.

20. The pivotal factor in framework analysis is usually the length of framework 1. If a sequence has a complete framework 1, then it almost always has complete frameworks 2 and 3 of 17 and 38 amino acids, etc. However, there are exceptions due to insertions/deletions and potentially recombined V genes that may vary the amino acid count in frameworks 2, 3, etc.
21. Do not filter out sequences based on parameters in complementarity-determining regions (CDRs) as there is a range of lengths recorded for all three CDRs including sequences that according to IMGT have missing CDRs. Do not filter out sequences that have no D gene called as some D genes may be mutated enough to be unrecognizable.
22. The rearrangements found less frequently in a sample may either result from a subclone or represent a contamination of VDJ rearrangements from non-tumor B cells.

Acknowledgments

This work was supported by the Dr. Mildred Scheel Stiftung für Krebsforschung (Deutsche Krebshilfe). We are grateful to Bao Tran, Yuliya Kriga, Jyoti Shetty, and Yongmei Zhao who were instrumental in establishing the protocols described here.

References

1. Pasqualucci L, Dalla-Favera R (2018) Genetics of diffuse large B-cell lymphoma. Blood 131 (21):2307–2319
2. Chapuy B, Stewart C, Dunford AJ et al (2018) Author correction: molecular subtypes of diffuse large B cell lymphoma are associated with distinct pathogenic mechanisms and outcomes. Nat Med 24(8):1290–1291
3. Schmitz R, Wright GW, Huang DW et al (2018) Genetics and pathogenesis of diffuse large B-cell lymphoma. N Engl J Med 378 (15):1396–1407
4. Rosenquist R, Bea S, Du MQ et al (2017) Genetic landscape and deregulated pathways in B-cell lymphoid malignancies. J Intern Med 282(5):371–394
5. Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet 10(1):57–63
6. Mortazavi A, Williams BA, McCue K et al (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods 5 (7):621–628
7. Frischmeyer PA, Dietz HC (1999) Nonsense-mediated mRNA decay in health and disease. Hum Mol Genet 8(10):1893–1900
8. Garber M, Grabherr MG, Guttman M et al (2011) Computational methods for transcriptome annotation and quantification using RNA-seq. Nat Methods 8(6):469–477
9. Dobin A, Gingeras TR (2015) Mapping RNA-seq reads with STAR. Curr Protoc Bioinformatics 51:11.14.1–11.1419
10. Li S, Lefranc MP, Miles JJ et al (2013) IMGT/HighV QUEST paradigm for T cell receptor IMGT clonotype diversity and next generation repertoire immunoprofiling. Nat Commun 4:2333
11. Langmead B, Trapnell C, Pop M et al (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10(3):R25

12. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25(14):1754–1760
13. Robinson JT, Thorvaldsdottir H, Winckler W et al (2011) Integrative genomics viewer. Nat Biotechnol 29(1):24–26
14. Costa-Silva J, Domingues D, Lopes FM (2017) RNA-Seq differential expression analysis: an extended review and a software tool. PLoS One 12(12):e0190152
15. Anders S, Pyl PT, Huber W (2015) HTSeq—a Python framework to work with high-throughput sequencing data. Bioinformatics 31(2):166–169
16. Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15(12):550
17. Huang da W, Sherman BT, Lempicki RA (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 4(1):44–57
18. Subramanian A, Tamayo P, Mootha VK et al (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 102(43):15545–15550
19. de Hoon MJ, Imoto S, Nolan J et al (2004) Open source clustering software. Bioinformatics 20(9):1453–1454
20. Kumar S, Vo AD, Qin F et al (2016) Comparative assessment of methods for the fusion transcripts detection from RNA-Seq data. Sci Rep 6:21597
21. D. Nicorici Satalan M, H. Edgren, S. Kangaspeska, A. Murumagi, O. Kallioniemi, S. Virtanen, O. Kilkku (2014) FusionCatcher—a tool for finding somatic fusion genes in paired-end RNA-sequencing data. bioRxiv. https://doi.org/10.1101/011650
22. Ding L, Rath E, Bai Y (2017) Comparison of alternative splicing junction detection tools using RNA-Seq data. Curr Genomics 18 (3):268–277
23. Sheng Q, Zhao S, Li CI et al (2016) Practicability of detecting somatic point mutation from RNA high throughput sequencing data. Genomics 107(5):163–169
24. Cingolani P, Platts A, Wang le L et al (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin) 6(2):80–92
25. Wang K, Li M, Hakonarson H (2010) ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res 38(16):e164
26. Young RM, Shaffer AL 3rd, Phelan JD et al (2015) B-cell receptor signaling in diffuse large B-cell lymphoma. Semin Hematol 52 (2):77–85
27. Young RM, Wu T, Schmitz R et al (2015) Survival of human lymphoma cells requires B-cell receptor engagement by self-antigens. Proc Natl Acad Sci U S A 112 (44):13447–13454
28. Giudicelli V, Chaume D, Lefranc MP (2005) IMGT/GENE-DB: a comprehensive database for human and mouse immunoglobulin and T cell receptor genes. Nucleic Acids Res 33(Database issue):D256–D261
29. Haas BJ, Papanicolaou A, Yassour M et al (2013) De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat Protoc 8(8):1494–1512

Chapter 14

Studying Cancer Heterogeneity by Single-Cell RNA Sequencing

Johannes W. Bagnoli, Lucas E. Wange, Aleksandar Janjic, and Wolfgang Enard

Abstract

A major hurdle for the treatment of cancer is the incomplete understanding of its evolution through the course of its emergence, dispersal, and relapse. Genetic and epigenetic changes in combination with external cues and selective forces are the driving factors behind tumor heterogeneity. Understanding this variability within and across patients may partly explain the unpredictable outcomes of cancer treatments. Measuring the variation of gene expression levels within cells of the same tumor is a crucial part of this endeavor. Hence, the recently developed single-cell RNA-sequencing (scRNA-seq) technologies have become a valuable tool for cancer research. In practice, however, this is still challenging, especially for clinical samples. Here, we describe mcSCRB-seq (molecular crowding single-cell RNA barcoding and sequencing), a highly sensitive and powerful plate-based scRNA-seq method, which shows great capability to generate transcriptome data for cancer cells. mcSCRB-seq is not only characterized by high sensitivity due to molecular crowding and the use of unique molecular identifiers (UMIs) but also features an easy workflow and a low per-cell cost and does not require specialized equipment.

Key words Single-cell RNA-seq, Tumor heterogeneity, Transcriptomics, mcSCRB-seq, Cancer, FACS

1 Introduction

While the evolutionary nature of cancer and hematopoietic neoplasms is well established, it is challenging to generate and analyze data to quantify the dynamics of this Darwinian process [1–6]. This is a clinical problem, for example, when therapy-resistant cancer cells relapse after an initially successful treatment. Hence, patients often die because the tumor is able to evolve. Analogous to the importance of understanding bacterial evolution to prevent antibiotic resistance, it is crucial to understand and measure the clonal evolution of cancer cell populations to prevent therapy resistance. Due to technological advances, genetic heterogeneity and its

Johannes W. Bagnoli and Lucas E. Wange contributed equally to this work.

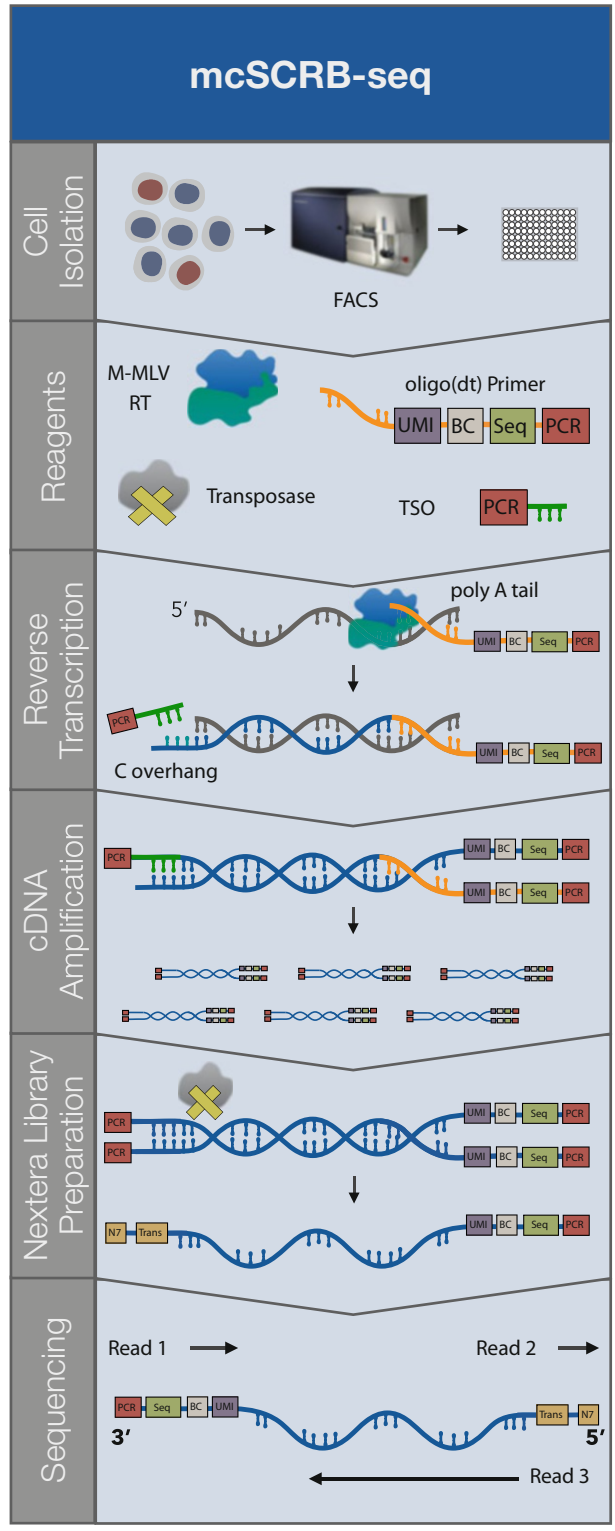
Ralf Küppers (ed.), *Lymphoma: Methods and Protocols*, Methods in Molecular Biology, vol. 1956,
https://doi.org/10.1007/978-1-4939-9151-8_14,

change during tumorigenesis can increasingly be measured (e.g., [7]). Therefore more reliable estimates for mutation rates as well as selection pressures for genetic changes can be obtained (e.g., [8]). In contrast, the epigenetic heterogeneity and its direct output, transcriptional heterogeneity, are more difficult to infer as the large quantitative range of transcriptional levels needs to be measured at single-cell resolution [9].

In recent years, many single-cell RNA-sequencing (scRNA-seq) protocols have been developed to measure transcriptomes of single cells [10, 11] and have been applied in studying cancer heterogeneity (e.g., [12, 13]). While the molecular biology of scRNA-seq protocols is similar in that after cellular lysis, mRNA gets reverse transcribe and second-strand cDNA is generated from which sequenceable libraries are created [11], isolation methods differ between protocols. It can be very challenging to isolate intact cells for analysis, in particular for clinical samples and/or samples that are difficult to access, which is why methods that include an enrichment for intact cells using, e.g., FACS are preferable.

The majority of RNA molecules within a cell are tRNAs and rRNAs, while mRNAs make up only 1–5% of the total cellular RNA. Therefore, polyA-targeting oligo-dT primers are commonly used to specifically reverse transcribe mRNAs [14, 15]. The resulting cDNA must be amplified and further transformed into sequenceable libraries by generating small fragments that bind to the sequencing flow cell using introduced adapter ends [16]. Due to the low amount of starting material, scRNA-seq experiments rely on heavy amplification, which might introduce PCR-based biases in the resulting expression data [17]. This bias can be removed by using unique molecular identifiers (UMIs), which are random DNA sequences that are linked to each mRNA molecule before amplification. Duplicated reads from the same starting mRNA molecule can be easily identified via the UMI sequence and removed computationally. This increases the power of scRNA-seq experiments considerably. The analysis of the resulting data, including preprocessing, normalization, visualization, and interpretation, requires bioinformatic and statistical expertise [11, 18, 19].

In this protocol we focus on the molecular biology aspect of generating such high-throughput transcriptome data. We describe the generation of high-quality libraries from cancer cells using molecular crowding single-cell RNA barcoding and sequencing (mcSCRB-seq) [20]. This method relies on the isolation of single cells using fluorescence-activated cell sorting (FACS), highly sensitive reverse transcription in the presence of a molecular crowding reagent (polyethylene glycol, PEG), and less biased amplification due to the incorporation of UMIs and the usage of Terra polymerase (Fig. 1). Although other methods have higher throughput (e.g., 10× Genomics or Drop-seq), mcSCRB-seq, being a plate-based method, is highly advantageous when low-quality samples


mcSCRB-seq
Cell Isolation
FACS
Reagents
M-MLV RT
oligo(dt) Primer
UMI
BC
Seq
PCR
Transposase
TSO
PCR
Reverse Transcription
5′
poly A tail
C overhang
cDNA Amplification
Nextera Library Preparation
N7
Trans
Sequencing
Read 1
Read 2
3′
5′
Read 3

such as frozen tumor samples from biopsies are used [21, 22]. Especially if samples only comprise a small number of intact cells, plate-based methods in combination with FACS show a much higher success rate.

2 Materials

2.1 General Equipment

1. Thermal cycler (96 or 384 block, with heated lid).
2. Fluorescence-activated cell sorter (Sony/BD).
3. Capillary gel electrophoresis system (Agilent Bioanalyzer/AATI Fragment Analyzer).
4. Fluorescence plate reader.
5. Precast agarose gel system (Thermo Fisher).
6. SPRI bead magnet racks (384-/96-well format/2 mL tube format).
7. Filter tips.
8. Micropipettes (single and multichannel).
9. Plate centrifuge.
10. 96-/384-well DNA low-binding plates.
11. 0.5/1.5/2 mL DNA low-binding tubes.
12. Aluminum cold storage seals.
13. Adhesive PCR film seals.

Fig. 1 Schematic overview of mcSCRB-seq. *Cell isolation*: Isolation of single cells is usually achieved by sorting single cells into microplates (96/384 well) using FACS. *Reverse transcription*: mRNA molecules are targeted by oligo-dT primers incorporating cell barcodes, UMIs, PCR anchor, and sequencing anchors. Reverse transcription in the presence of molecular crowding agent (PEG 8000) is highly efficient. MMLV-derived reverse transcriptase allows for template switching using poly C overhangs and TSO primers. *cDNA amplification*: cDNA amplification is performed using PCR anchors introduced by oligo-dT and TSO primers in a one-primer PCR. Terra polymerase allows amplification of the library with minimal bias. UMIs can later on be used to account for PCR duplicates. *Nextera library preparation:* Full-length cDNA fragments are tagmented using Tn5 transposase (Nextera). Fragments derived from the 3′ end are enriched in the subsequent PCR using a primer binding to the sequencing anchor introduced in the reverse transcription. *Sequencing:* Sequencing is performed with three read starts. The first read (Read 1, 16 bp) detects the cell barcode and UMI for each fragment. The second read (Read 2, 8 bp) targets the Nextera P7 index which allows for multiplexing of several libraries with the same cell barcode set. Lastly, Read 3 (50 bp) covers the cDNA sequence for each fragment

2.2 SPRI Beads

1. Magnetic carboxylate modified particles (Thermo Fisher).
2. DNase-/RNase-free distilled water.
3. TE buffer: 10 mM Tris–HCl, pH 8.0, 1 mM EDTA.
4. Bead-binding buffer: 11 g PEG 8000, 10 mL 5 M NaCl, 500 μL 1 M Tris–HCl pH 8.0, 100 μL EDTA, 50 μL 10% IGEPAL, 250 μL 10% sodium azide, DNase-/RNase-free distilled H_2O to 49 mL or 50 mL.

2.3 Lysis and RNA Isolation

1. 5 M guanidine hydrochloride buffer for 96-well plate (110×): 343.75 μL 8 M guanidine hydrochloride, 1.1 μL 5× Phusion HF buffer (NEB), 5.5 μL 2-mercaptoethanol, 199.7 μL DNase-/RNase-free distilled H_2O (for 384 use 410×)
2. SPRI beads (*see* Subheading 3.1).
3. Molecular biology-grade ethanol.
4. DNase-/RNase-free distilled water.

2.4 Reverse Transcription

1. PEG 8000 (50%) (m/v): Do not add the total amount of H_2O, but adjust the total volume only after PEG is completely solubilized. Incubate at 40 °C and vortex regularly until PEG is completely dissolved.
2. MMLV Reverse Transcriptase (Maxima H-) with supplied buffer (Thermo Fisher).
3. dNTPs.
4. Barcoded oligo-dT primer with UMI (E3V6NEXT) (IDT, TruGrade) (*see* Table 1).
5. Template-switching oligo (TSO, E5V6NEXT) (IDT, HPLC) (*see* Table 1).
6. DNase-/RNase-free distilled water.

Table 1
Primer sequences for mcSCRB-seq

Primer	Sequence (5′–3′)
E3V6NEXT TruGrade	Biotin-ACACTCTTTCCCTACACGACGCTCTTCCGATCT[BC6][UMI10][T30]VN
E5V6NEXT HPLC	ACACTCTTTCCCTACACGACGCrGrGrG
SINGV6	Biotin-ACACTCTTTCCCTACACGACGC
P5NEXTPT5	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCG*A*T*C*T
i7 Index Primer TruGrade	CAAGCAGAAGACGGCATACGAGAT[i7]GTCTCGTGGGCTCGG

V, A, G, C; *, phosphorothioated DNA base; [i7], 8 bp barcode sequence (*see* Illumina sequence letter); r, ribonucleotide; [BC6], 6 bp barcode; [UMI10], random 10 bp (N)

7. Bead-binding buffer (*see* Subheading 2.2).
8. Molecular biology-grade ethanol.

2.5 Exol Digest and cDNA Preamplification

1. Exonuclease I with supplied buffer.
2. Terra polymerase with supplied buffer (Takara).
3. Amplification primer (SINGV6) (IDT) (*see* Table 1).
4. SPRI beads (*see* Subheading 3.1).
5. Molecular biology-grade ethanol.
6. DNase-/RNase-free distilled water.
7. Quant-iT PicoGreen dsDNA Assay Kit (Thermo Fisher).
8. High Sensitivity DNA Analysis Kit for Agilent Bioanalyzer 2100 (Agilent) or High Sensitivity NGS Kit for Fragment Analyzer (AATI).

2.6 Tagmentation, Indexing, and Library Amplification

1. Nextera XT DNA Library Preparation Kit (Nextera).
2. i7 Index Primer (IDT, TruGrade) (*see* Table 1),
3. P5NEXTPT5 Primer (IDT) (*see* Table 1).
4. SPRI beads (*see* Subheading 3.1).
5. Molecular biology-grade ethanol.
6. DNase-/RNase-free distilled water.
7. Precast agarose gels (2%) (Thermo Fisher).
8. MinElute Gel Extraction Kit (Qiagen).
9. Quant-iT PicoGreen dsDNA Assay Kit (Thermo Fisher).
10. High Sensitivity DNA Analysis Kit for Agilent Bioanalyzer 2100 (Agilent) or NGS High Sensitivity.
11. Kit for Fragment Analyzer (AATI).

3 Methods

3.1 SPRI Bead Preparation

1. Add all reagents of bead-binding buffer into a 50 mL tube. Do not add the total amount of H_2O, but adjust the total volume only after PEG is completely solubilized.
2. Incubate at 40 °C and vortex regularly until PEG is completely dissolved.
3. Resuspend bead stock carefully (Sera-Mag SpeedBeads).
4. Withdraw 1000 μL of bead suspension into a 1.5 mL tube. Place on magnet stand. Remove supernatant.
5. Add 1000 μL TE, and resuspend beads. Place on magnet stand. Remove supernatant.

6. Add 1000 μL TE, and resuspend beads. Place on magnet stand. Remove supernatant.
7. Add 900 μL TE, and resuspend beads. Add to PEG solution and mix well.

3.2 Sorting

Prepare lysis buffer, distribute 5 μL to each well of the multiwell plate, and seal the plates until sorting. Adjust FACS (*see* **Note 1**), spin down lysis plate briefly before sorting, sort one cell in each well, and immediately seal the plate with an aluminum cold storage seal. Centrifuge plate at max. speed and place on dry ice immediately. Transfer to −80 °C freezer.

3.3 RNA Cleanup

1. Thaw plates briefly and spin down (*see* **Note 2**).
2. Add 10 μL of SPRI beads (ratio 2:1) to each well. Seal plate immediately, briefly vortex to mix, and briefly spin down (max. 200 × *g*).
3. Incubate for 5 min at room temperature.
4. Place the plate on a magnetic rack until the solution is clear.
5. Discard the supernatant.
6. Carefully add 50 μL freshly made 80% ethanol to each well.
7. Discard the ethanol and repeat the washing step.
8. Remove any residual ethanol and incubate up to 2 min until the beads are dry.
9. Take the plate off the magnet and add 4 μL ddH_2O to each well (*see* **Note 3**).
10. Immediately continue with reverse transcription.

3.4 Reverse Transcription

1. Prepare reverse transcription mix consisting of 2 μL Maxima RT buffer (5×), 0.4 μL 25 mM dNTPs, 0.2 μL 100 μM TSO E5V6NEXT, DNase-/RNase-free distilled H_2O per well prior to cleanup (*see* **Note 4**).
2. Add 1.5 μL 50% PEG and 0.1 μL Maxima H- Reverse Transcriptase directly before reverse transcription, and mix thoroughly using a 1 mL pipette (*see* **Note 5**).
3. Add 5 μL of the complete RT mix to each well containing cleaned RNA.
4. Transfer 1 μL of barcoded E3V6 oligo-dT primer (2 μM) to each well, maintaining the well position from the source plate.
5. Seal plate immediately and mix thoroughly by vortexing, making sure the bead pellets are completely resuspended.
6. Spin down briefly (max. 200 × *g*), and incubate for 1.5 h at 42 °C in a thermal cycler with heated lid.

3.5 Pooling

1. Spin down plate after reverse transcription (max 200 × *g*).
2. Combine all wells including the beads into a 2 mL LoBind tube.
3. Add 960 μL bead-binding buffer and mix well.
4. Incubate for 5 min at room temperature. Place on appropriate magnetic rack until clear.
5. Discard supernatant and add 2 mL of freshly prepared 80% ethanol.
6. Discard ethanol and repeat washing step. Remove any residual ethanol and dry the beads for up to 10 min.
7. Elute first-strand cDNA by adding 17 μL ddH_2O.
8. Completely resuspend the beads and incubate for 5 min at room temperature. Place on appropriate magnetic rack until the solution is clear.
9. Transfer elution to a new well/tube.

3.6 Exonuclease I Digest

Add 1 μL exonuclease I (20 units) as well as 2 μL exonuclease I 10× buffer to each cDNA pool. Mix well and incubate for 20 min at 37 °C, followed by 10 min heat inactivation at 80 °C, in a thermocycler with heated lid.

3.7 Preamplification of cDNA

1. Mix 1 μL Terra polymerase, 25 μL Terra direct buffer 2×, 1 μL 10 μM SINGV6 primer, and 3 μL of DNase-/RNase-free distilled water per cDNA pool for preamplification PCR.
2. Add 30 μL of this mix directly to exonuclease I digest.
3. Mix well and incubate using the following program, 98 °C for 3 min, 13–21 × (98 °C 15 s, 65 °C 30 s, 68 °C 4 min), 72 °C 10 min, and 8 °C for storage until further processing. The number of PCR cycles for preamplification is highly dependent on the RNA content of the cells, which is correlated to its size, but also dependent on sample quality.

3.8 Purification of cDNA Libraries

1. Add 40 μL (ratio 0.8:1) SPRI beads to each amplification reaction. Mix well and incubate for 5 min at room temperature.
2. Place on appropriate magnetic rack until clear.
3. Discard supernatant and add 100 μL freshly prepared 80% ethanol.
4. Discard ethanol and repeat washing step.
5. Remove any residual ethanol and dry the beads for up to 5 min.
6. Elute by adding 10 μL ddH_2O. Completely resuspend the beads and incubate for 5 min at room temperature.
7. Place on appropriate magnetic rack until the solution is clear.
8. Transfer elution to a new well/tube.

3.9 Quantification of cDNA Libraries

Quantify the cDNA using the Quant-iT PicoGreen dsDNA Assay Kit. Follow the protocol as provided by the manufacturer, and use 1 μL of the amplified cDNA for quantification. Concentration should be at least 1 ng/μL but will be variable depending on cell type and cycle numbers.

3.10 Quality Control of cDNA Libraries

Analyze the quality of the obtained cDNA using an automated capillary gel electrophoresis system such as the 2100 Bioanalyzer (Agilent) or Fragment Analyzer (AATI). Perform the assay as described in the manufacturer's protocol. Good-quality cDNA should show a distribution from ~300 to 6000 bp, with the distribution peak above 1000 bp and an average size greater than 700 bp (Fig. 2a). If you should encounter a pronounced peak around 150 bp stemming from primer dimers (Fig. 2b), an additional cleanup using SPRI beads (1/0.7) is recommended (*see* **Note 4**).

3.11 Sequencing Library Construction

In order to minimize batch effects as well as to reduce costs, it is recommended to pool equal amounts of cDNAs with different barcode sets. Several of these pools can be combined after Nextera library preparation using different indices. This ensures that no additional sequencing batch effect is generated and all plates or conditions of one experiment can be sequenced on the same flow cell and lane.

3.12 Tagmentation

1. Dilute amplified cDNA with DNase-/RNase-free water to 0.8 ng/μL. For each cDNA library, it is recommended to perform several replicates (3–5) to minimize sampling biases.
2. Prepare tagmentation mix containing 10 μL 2× Tagment DNA Buffer, 5 μL Amplicon Tagment Mix, and 4 μL DNase-/RNase-free distilled water, per replicate.
3. Dispense 19 μL to a new 96-well plate, and add 1 μL of the corresponding cDNA (0.8 ng/μL) to each reaction.
4. Incubate for 10 min at 55 °C (thermocycler with heated lid).
5. Add 5 μL NT buffer to each reaction and mix. Incubate for 5 min at room temperature to strip Tn5 off the DNA.

3.13 Library Index PCR

1. Prepare Nextera Index PCR Mix containing 15 μL NPM PCR Mix, 0.5 μL 5 μM P5NEXTPT5, and 8.5 μL DNase-/RNase-free distilled water per reaction.
2. Add 24.5 μL to each tagmentation reaction.
3. Add 0.5 μL of i7 Index Primer (5 μM). Mix well and incubate using the following program: 72 °C for 3 min, 98 °C for 30 s, 13 × (95 °C 10 s, 55 °C 30 s, 72 °C 1 min), 72 °C 10 min, and 8 °C for storage until further processing.

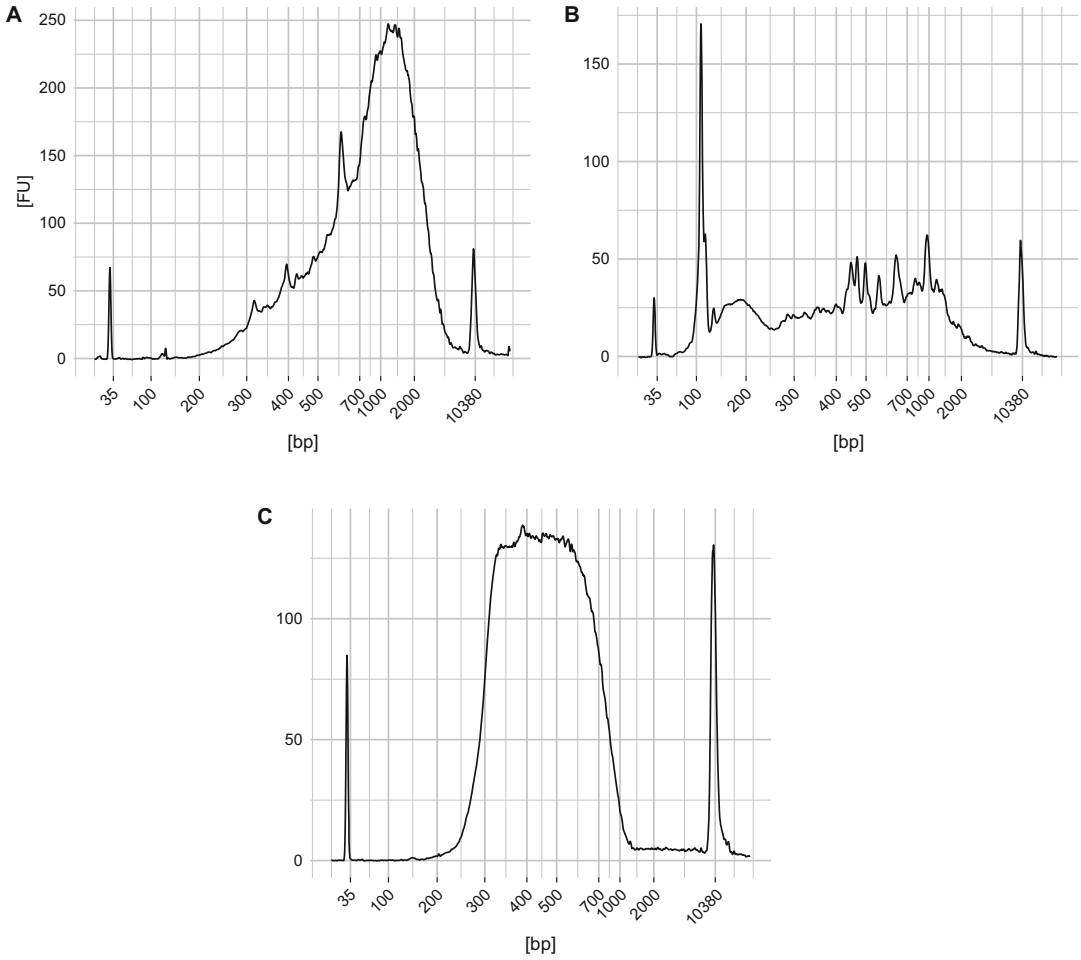


Fig. 2 Examples for high-quality full-length cDNA and Nextera library. (**a**) Capillary gel electrophoresis of high-quality full-length cDNA. High-quality cDNA should have a size peak above 1000 bp with no detectable fragments below 200 bp. (**b**) Capillary gel electrophoresis of low-quality full-length cDNA with substantial amounts of degradation. A high amount of fragments below 200 bp as well as a shift of the distribution to a lower mean size fragment length suggests RNA degradation prior to reverse transcription. (**c**) Capillary gel electrophoresis of a high-quality sequencing library. Fragments between 250 bp and 1000 bp cluster efficiently to the sequencing flow cell. Therefore, the size distribution for the finished library should lay within these borders

3.14 Library Purification and Size Selection

1. Pool all library replicates.
2. Add 50 μL (ratio 1:1) SPRI beads (*see* attached protocol), per each amplification reaction (replicate). Mix well and incubate for 5 min at room temperature.
3. Place on appropriate magnetic rack until clear.
4. Discard supernatant and add 100 μL freshly prepared 80% ethanol.
5. Discard ethanol and repeat washing step.

6. Remove any residual ethanol and dry the beads for up to 5 min.
7. Elute by adding 20 μL ddH_2O. Resuspend the beads well and incubate for 5 min at room temperature.
8. Place on appropriate magnetic rack until the solution is clear.
9. Transfer elution to a new well/tube.
10. Run the eluted libraries on a precast agarose gel (2%) with an appropriate DNA ladder (e.g., 100–3000 bp).
11. Open the precast gel and excise the library between 300 and 800 bp.
12. Isolate the DNA with the Qiagen MinElute Gel Extraction Kit as recommended by the manufacturer.
13. Elute the library from the silica membrane with 20 μL DNase-/RNase-free distilled water.

3.15 Library Quantification

Quantify the library using the Quant-iT PicoGreen dsDNA Assay Kit. Follow the protocol as provided by the manufacturer, and use 1 μL of the amplified library for quantification. Concentration should be >2 ng/μL, depending on the number of replicates.

Quality check the library using an Agilent 2100 Bioanalyzer or Fragment Analyzer. Perform the assay as described in the manufacturer's protocol. A good library should show a distribution from ~300 to 800 bp, with the distribution peak around 500 bp (Fig. 2c). In addition, quantify the molarity of the library between 200 and 1000 bp using the analysis software of the capillary gel electrophoresis instrument. Accurate quantification is important to ensure high clustering efficiency on the sequencer.

3.16 Sequencing

Make equimolar pools of all libraries with different indices corresponding to the same experiment, and combine them for sequencing. mcSCRB-seq requires an unbalanced paired-end design. With the first read, 16 bases are sequenced to obtain cellular and molecular barcodes, and 50 bases are sequenced in the second read into the cDNA fragment. When several libraries are multiplexed on sequencing lanes, an additional 8-base i7 barcode read needs to be performed. We recommend sequencing to a nominal depth of 2.5×10^5–1×10^6 reads per cell.

3.17 Primary Data Analysis

Primary data analysis includes filtering of sequencing reads based on Phred scores, mapping to the reference genome, assigning features, and finally counting of unique molecular identifiers per gene. We recommend using zUMIs [23], a fast and flexible pipeline, to perform these steps. The major output of zUMIs is digital gene expression (dge) matrices with samples (barcodes) as columns and genes as rows. Each value describes the number of UMIs that were detected for a certain gene in a given cell.

3.18 Analysis of Heterogeneity

After the primary processing of the data, we recommend using R [24] and RStudio [25] to continue analyzing the resulting dge matrices. As there is no standard way of analyzing such datasets, we will only mention crucial steps in this section. First of all, doublets and broken cells need to be filtered out by excluding cells with markedly fewer or more UMIs as compared to the majority of the cells. Next, it is recommended to filter out genes with only sporadic detection within the dataset. In order to perform any further analysis, the data needs to be normalized. There are several methods published to normalize noisy single-cell data, which also deal with the problem of zero inflation. As mentioned before, ERCC spike-ins often help to normalize the data as one can account for technical variation between the cells. We recommend using scran or SCnorm for normalization [26, 27].

In order to identify expression heterogeneity, the underlying variance structure of the normalized count data can be visualized via dimension reduction analysis like principal component analysis or t-distributed stochastic neighbor embedding (tSNE). Using these methods it is possible to assign portions of cells to clusters and thereby identify subpopulations having different expression profiles. Between these newly assigned clusters, differential gene expression analysis can be performed. These are just a few of the possible analyses that can be performed on scRNA-seq data.

4 Notes

1. Fluorescence-activated cell sorting is primarily used to enrich for living cells as well as specific cell types not only by the size and granularity but also via staining of specific markers. It is equally important to sort for cancerous cells (if other cell types are present) as well as selecting living cells which are in good condition. This can be achieved by using live/dead staining such as Zombie Yellow, propidium iodide, or DAPI as well as apoptosis stainings such as antibodies for annexin V. Additionally, it might help to use known surface markers common to all cells of interest to distinguish cancerous cells from normal cells. Additionally, when using a plate-based scRNA-seq method, it is of high importance that plate adjustments are performed correctly. This is especially true for 384-well plates, as it is necessary to adjust the droplet so it hits the well accurately. In cases where it is not possible to achieve a single-cell solution from given tumor samples, it may be necessary to use methods like laser microdissection to isolate single cells, which can then be lysed in the same lysis buffer as sorted cells. However, since large-scale transcriptome analysis often requires a large sample size (>10,000 cells), sorting of cells is the more practical approach.

2. Compared to DNA, RNA is not very stable due to its autohydrolysis and the ubiquitous occurrence of RNases. There are also dedicated cellular mechanisms to degrade RNA during stress and apoptosis [28], and hence the rate of decay depends strongly on the type of cell and its isolation and storage conditions. Therefore, scRNA-seq can be thought of as a race between reverse transcription and RNA degradation. As RNases are also commonly used in biological laboratories (e.g., plasmid isolation or DNA extraction), it is important to separate scRNA-seq work from lab work that uses RNases. Using a decontamination reagent such as RNase Away to clean lab surfaces, instruments, and gloves is highly recommended, in order to remove RNases and DNA contamination. In addition, the use of a UV-decontaminated PCR hood is recommended. As scRNA-seq relies on very low input amounts, it is also very helpful to only use plasticware with low nucleic acid-binding properties.
3. In contrast to column-based purification methods of nucleic acids, solid-phase reversible immobilization (SPRI) beads have the advantage of high recovery efficiency (above 90%) and a much lower cost, especially with self-made buffers as described above [29]. Although the workflow is very similar to column-based methods, there are some key differences. The volume ratio between the nucleic acid solution and the SPRI beads defines the length cutoff for the cleanup. The fewer SPRI beads used, the higher the cutoff, meaning smaller fragments are less likely to bind. The exact cutoffs for the different ratios highly depend on the buffer conditions of the SPRI bead solution.

   We highly recommend not overdrying the beads after the last washing step. Not only does this decrease the recovery efficiency immensely, especially for longer DNA or RNA fragments, but in most cases, small ethanol carryovers do not influence downstream reactions. Additionally, it is important that all beads stick to the wall of the well/tube when removing any supernatants in order to minimize loss. During washing, the plate or tube should remain on the magnet, and the beads should not be resuspended. In contrast thereto, it is important to mix the beads very well with the nucleic acid solution as well as the water when eluting. When cleaning up nucleic acids eluted in water, we recommend adding an equal volume of buffer EB to the sample prior to purification.
4. PEG is a polyether which can be used to induce molecular crowding in enzymatic reactions. Due to its high viscosity, it is very important to mix the reverse transcription mix thoroughly after adding PEG. It is not sufficient to vortex the tube. Mixing by pipetting up and down is more efficient and less

harmful to the enzyme. Make sure the mix does not show any smears of PEG after mixing.

5. We recommend preparing 110× RT mix for a 96-well plate, for each additional plate 100× RT mix. We highly recommend using ERCC (External RNA Controls Consortium) spike-ins in the RT master mix (substitute a part of H_2O) in order to account for technical variation in the downstream computational analysis. The exact amount of ERCC molecules per reaction/cell depends highly on the RNA content of the cells; however, a range between 5000 and 50,000 molecules is thought to work in most cases. Additionally, we recommend using 100 pg of UHRR as a positive control for the RT reaction in a separate plate or in a dedicated well on the sample plate.

References

1. Greaves M, Maley CC (2012) Clonal evolution in cancer. Nature 481:306–313
2. Maley CC, Aktipis A, Graham TA et al (2017) Classifying the evolutionary and ecological features of neoplasms. Nat Rev Cancer 17:605–619
3. Wu C-I, Wang H-Y, Ling S et al (2016) The ecology and evolution of cancer: the ultra-microevolutionary process. Annu Rev Genet 50:347–369
4. Podlaha O, Riester M, De S et al (2012) Evolution of the cancer genome. Trends Genet 28:155–163
5. Lipinski KA, Barber LJ, Davies MN et al (2016) Cancer evolution and the limits of predictability in precision cancer medicine. Trends Cancer Res 2:49–63
6. McGranahan N, Swanton C (2017) Clonal heterogeneity and tumor evolution: past, present, and the future. Cell 168:613–628
7. Turajlic S, Xu H, Litchfield K et al (2018) Deterministic evolutionary trajectories influence primary tumor growth: TRACERx renal. Cell 173:595–610.e11
8. Martincorena I, Raine KM, Gerstung M et al (2017) Universal patterns of selection in cancer and somatic tissues. Cell 171:1029–1041.e21
9. Tanay A, Regev A (2017) Scaling single-cell genomics from phenomenology to mechanism. Nature 541:331–338
10. Kolodziejczyk AA, Kim JK, Svensson V et al (2015) The technology and biology of single-cell RNA sequencing. Mol Cell 58:610–620
11. Ziegenhain C, Vieth B, Parekh S et al (2018) Quantitative single-cell transcriptomics. Brief Funct Genomics 17:220–232
12. Tirosh I, Izar B, Prakadan SM et al (2016) Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. Science 352:189–196
13. Ebinger S, Özdemir EZ, Ziegenhain C et al (2016) Characterization of rare, dormant, and therapy-resistant cells in acute lymphoblastic leukemia. Cancer Cell 30:849–862
14. Picelli S (2017) Single-cell RNA-sequencing: the future of genome biology is now. RNA Biol 14:637–650
15. Darnell JE Jr (1968) Ribonucleic acids from animal cells. Bacteriol Rev 32:262–290
16. Ziegenhain C, Vieth B, Parekh S et al (2017) Comparative analysis of single-cell RNA sequencing methods. Mol Cell 65:631–643.e4
17. Parekh S, Ziegenhain C, Vieth B et al (2016) The impact of amplification on differential expression analyses by RNA-seq. Sci Rep 6:25533
18. Stegle O, Teichmann SA, Marioni JC (2015) Computational and analytical challenges in single-cell transcriptomics. Nat Rev Genet 16:133–145
19. Vallejos CA, Risso D, Scialdone A et al (2017) Normalizing single-cell RNA sequencing data: challenges and opportunities. Nat Methods 14:565–571
20. Bagnoli JW, Ziegenhain C, Janjic A, et al (2018) Sensitive and powerful single-cell RNA sequencing using mcSCRB-seq. Nat Commun 9:2937
21. Macosko EZ, Basu A, Satija R et al (2015) Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. Cell 161:1202–1214

22. Zheng GXY, Terry JM, Belgrader P et al (2017) Massively parallel digital transcriptional profiling of single cells. Nat Commun 8:14049
23. Parekh S, Ziegenhain C, Vieth B et al (2018) zUMIs – A fast and flexible pipeline to process RNA sequencing data with UMIs. Gigascience 7
24. R Development Core Team (2008) R: a language and environment for statistical computing. http://www.R-project.org
25. RStudio Team (2015) RStudio: integrated development environment for R. http://www.rstudio.com/
26. Lun ATL, McCarthy DJ, Marioni JC (2016) A step-by-step workflow for low-level analysis of single-cell RNA-seq data with bioconductor. F1000Res 5:2122
27. Bacher R, Chu L-F, Leng N et al (2017) SCnorm: robust normalization of single-cell RNA-seq data. Nat Methods 14:584–586
28. Thomas MP, Lieberman J (2013) Live or let die: posttranscriptional gene regulation in cell stress and cell death. Immunol Rev 253:237–252
29. DeAngelis MM, Wang DG, Hawkins TL (1995) Solid-phase reversible immobilization for the isolation of PCR products. Nucleic Acids Res 23:4742–4743

Chapter 15

Evaluation of Whole Genome Sequencing Data

Daniel Hübschmann and Matthias Schlesner

Abstract

Whole genome sequencing (WGS) can provide comprehensive insights into the genetic makeup of lymphomas. Here we describe a selection of methods for the analysis of WGS data, including alignment, identification of different classes of genomic variants, the identification of driver mutations, and the identification of mutational signatures. We further outline design considerations for WGS studies and provide a variety of quality control measures to detect common quality problems in the data.

Key words Next-generation sequencing, Quality control, Variant calling, Mutational signatures

1 Introduction

Compared to traditional sequencing techniques, next-generation sequencing (NGS) offers a tremendous increase in throughput combined with a drastic decrease in cost per sequenced base. This makes it possible to perform whole genome sequencing (WGS) for large cohorts of samples and also to use it as a diagnostic tool in clinical settings.

Especially in cancer research, genome sequencing has been quickly adopted as genome alterations play a central role in malignant transformation [1]. Since the first publication of a cancer genome in 2008 [2], NGS formed the basis for substantial advancements in the understanding of the mechanisms underlying initiation and progression of several cancers including lymphomas.

Cancer genomes can either be sequenced in total (WGS), or sequencing is restricted to protein-coding regions (whole exome sequencing, WES) or selected genes or regions (panel sequencing). Only WGS covers the entire genome, including all exons, introns, and intergenic regions (but excluding highly repetitive parts which are not accessible with current short-read sequencing techniques), reveals all classes of alterations, and thus provides a comprehensive characterization of the cancer genome.

Ralf Küppers (ed.), *Lymphoma: Methods and Protocols*, Methods in Molecular Biology, vol. 1956,
https://doi.org/10.1007/978-1-4939-9151-8_15,

Four classes of variants can be identified from WGS data: single-nucleotide variants (SNVs), small insertions and deletions (indels), structural variants (SVs), and copy number aberrations (CNAs) [3]. In cancer sequencing studies, it is important to distinguish between germline variants, i.e., variants which the patient has inherited or which arose de novo in the very early stages of embryonic development and which are hence present in (almost) all cells of the patient, and somatic variants, which were acquired later during lifetime. Since a tumor stems from clonal expansion of a single cell, all variants which this cell has acquired before or in the early stages of tumor development appear as somatic variants in all cells of the tumor. Genomic variants which are acquired later during tumor evolution are only present in a subset of tumor cells and are commonly referred to as subclonal variants.

Although numerous computational tools dedicated to specific aspects of NGS data analysis have been developed in the past few years, most have project-specific features, and their functionality and parameterization are complicated. Furthermore, even when analyzing the same input data, the use of different processing pipelines results in different results [4]. In this chapter, we will introduce selected methods for processing and analysis of cancer WGS data. The current de facto standard for human genome sequencing experiments is paired-end sequencing on the Illumina platform. The analysis methods presented here assume short-read (100–150 bp) paired-end WGS data generated on Illumina sequencers. To enable reliable detection of genomic variants and to a certain extent also subclonal variants, the average depth of coverage should be at least 30×. The presented methods require availability of a Unix or Linux computing system with sufficient memory and storage (a BAM file of a human genome sequenced at 30× coverage is almost 100 GByte in size; the minimum RAM requirement of some of the processing steps is 50 GByte). It is recommended to perform WGS data analysis on a high-performance compute cluster or in a cloud environment to enable parallel processing and reduce analysis time. It should be noted that human WGS data cannot be de-identified without drastically reducing its utility and hence always bears the risk that the sample donor can be identified. Therefore, all infrastructure involved in storage, processing, and transfer of such data has to be adequately secured to minimize the risk of a privacy breach. The exact regulations differ between countries, and the patient's informed consent might impose additional restrictions on the allowed usage of the data. It is thus necessary to check for each individual scenario that the planned data generation, transfer, storage, and sharing are in agreement with all relevant regulations.

2 Methods

2.1 Experimental Design Considerations

2.1.1 Use of Matched Normal Controls

In addition to the tumor tissue, sequencing of matched normal tissue from the same patient is necessary to discriminate between germline and somatic variants (*see* **Note 1**). Any non-clonal tissue from the same patient can be used as matched normal tissue; due to good accessibility, most often white blood cells are chosen. Care has to be taken that the matched normal sample is not contaminated with tumor cells, which would lead to a misclassification of somatic variants as germline.

2.1.2 Whole Genome, Whole Exome, or Panel Sequencing

Only WGS enables the comprehensive identification of variants in the nonprotein-coding part of the genome. However, many other downstream analyses also benefit tremendously from the availability of broader data as delivered by WGS. These include the identification of structural variants (SVs) and copy number aberrations (CNAs) as well as the determination of tumor cell content and ploidy. Furthermore, WGS information results in a much higher power to identify genomic patterns including mutational signatures. The identification of mutational signatures and similar patterns might even have therapeutic implications, for example, in the case of "BRCAness" which indicates potential sensitivity of the tumor to PARP inhibitors [5]. A disadvantage of WGS is the higher cost (for data generation and for data analysis), especially if a high sequencing coverage is needed to detect variants with low allele frequencies (e.g., to analyze samples with low tumor cell content). In such settings exome sequencing or gene panel sequencing might be a more suitable choice.

2.1.3 Coverage

In the first large-scale cancer WGS projects as, for example, in the International Cancer Genome Consortium (ICGC) [6], the required minimum coverage was 30×. This depth should be considered as absolute minimum for cancer WGS analyses. If a study aims at deciphering the subclonal composition of a tumor or analyzes tumors with low tumor cell content (*see* below), considerably higher coverage might be required. It is discussed controversially whether the matched control tissue should be sequenced to the same coverage depth as the tumor or whether a fixed coverage of 30× in the matched normal control is sufficient even if the tumor is sequenced much deeper. In our experience, the analysis of samples with higher coverage in the tumor than in the matched normal requires additional filtering steps like filtering against a panel of normals to remove artifacts from the somatic variant calls [4].

2.1.4 Tumor Cell Content

Low tumor cell content of a sample causes reduced sensitivity in somatic variant identification [4] and especially impairs the identification of subclonal variants. The ICGC requires a tumor cell

content of at least 60%. It should be noted that the tumor cell content as determined from histological analysis is often considerably higher than the tumor cell content as determined from sequencing data analysis, so that samples which appear to be "still good" in histopathology might result in hardly usable data. To a certain extent, a lower tumor cell content can be compensated by higher sequencing depth and the use of improved bioinformatic methods to detect variants with low mutant allele fraction (MAF), but the discrimination between sequencing and library errors and true variants becomes more challenging.

2.1.5 Sample Requirements

WGS should be done on DNA extracted from fresh frozen tissue. Blood samples, for example, for use as matched normal, should be collected in EDTA-coated tubes; heparin-containing tubes should not be used because heparin inhibits PCR reactions and hence interferes with sequencing library preparation. Sequencing of DNA from FFPE tissue can, with some limitations, be used for small variant and CNA identification [7], but the identification of SVs is usually not possible. The depth of coverage in FFPE-derived data often shows strong fluctuations, and considerable parts of the genome might not be sufficiently covered for reliable variant calling. In many cases FFPE-induced DNA damage causes drastically increased sequencing error rates, and hence additional filtering steps are needed to reduce the number of false-positive variant calls.

2.2 Alignment

2.2.1 Choice of the Reference Genome

Next-generation sequencers produce millions of short sequence reads, which are stored in FASTQ files. The first step of WGS data analysis is to align the sequence reads to a reference genome. Two builds of the human reference genome are concurrently used: GRCh37 (hg19) and GRCh38. For both builds several different versions exist, which differ in the patches applied by the Genome Reference Consortium but also in the inclusion (or exclusion) of additional sequences like the "unlocalized sequences," "unplaced sequences," "decoy sequences," and in the case of GRCh38 also "alternate loci" (ALT contigs). For GRCh37, the reference sequence of the 1000 Genomes Project Phase II, called hs37d5 (ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/technical/reference/phase2_reference_assembly_sequence), is commonly used. This version differs from the standard build in the sequence of mitochondrial DNA, in masking of the PAR regions in chromosome Y and in the inclusion of decoy sequences, which reduce artifacts in variant calling. For GRCh38, the inclusion of ALT contigs, which represent common variation (e.g., the HLA loci) and multi-placed regions, makes this step of the analysis more challenging. Processing with standard pipelines which are not adapted to handle these new features properly will lead to a loss of variant calls in all regions affected by these issues. Many recent aligners (e.g., the current version of BWA-MEM) are "ALT-

aware" and can hence be used with a reference sequence containing ALT contigs, as, for example, included in the current GATK bundle (https://software.broadinstitute.org/gatk/download/bundle). For other pipelines, versions without ALT contigs have to be used (e.g., ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000/001/405/GCA_000001405.15_GRCh38/seqs_for_alignment_pipelines.ucsc_ids/README_analysis_sets.txt).

2.2.2 Preprocessing

Most sequencing facilities remove sequencing barcodes already during generation of the FASTQ files; otherwise they should be removed in the data analysis prior to alignment, for example, using TRIMMOMATIC [8]. Trimming of low-quality parts of the sequencing reads improves alignment quality if methods which require reads to be mapped in full length are used (e.g., BWA-backtrack [9]), but it is not required for methods based on local alignments like BWA-MEM [10].

2.2.3 Alignment Algorithms

While a plethora of algorithms to align short low-divergent sequences to a large reference genome has been developed, only few of them are commonly used for human WGS data, including BWA [9], Bowtie 2 [11], and GEM [12]. Several papers have reviewed the differences between them [13, 14], and we will not review these tools here but focus on the most commonly used algorithm, BWA. BWA is based on backward search with the Burrows-Wheeler transform [15], which enables high alignment accuracy while keeping a small memory footprint also for large genomes [9]. The first two algorithms of the BWA family, BWA-backtrack and BWA-SW [16], have later been complemented by a third algorithm BWA-MEM [10], which is suited to align sequences in the range from 70 bp to a few megabases against reference genomes. Hence BWA-MEM is usually the algorithm of choice when aligning WGS data generated on current Illumina instruments.

2.2.4 BAM File Processing

After the alignment, several postprocessing steps are required to generate BAM files for further analyses. It is recommended to combine several steps with Unix pipes instead of writing every output file to disk in order to reduce disk space requirements and enable fast processing. Several tools are available for these processing steps, including samtools [17], Picard [18], biobambam [19], and sambamba [20]. Essential postprocessing steps after alignment are coordinate sorting of the reads, merging of data (if a sample has been sequenced on more than one lane), and marking or removal of PCR duplicates (*see* **Note 2**). To reduce processing time, it is recommended to choose a tool that can perform merging and duplicate marking in one step (e.g., Picard or biobambam). In addition to these essential steps, other protocols [21, 22] recommend to perform base quality score recalibration and indel

realignment to improve BAM file quality, but in our setting these steps did not result in detectable improvements in the downstream analyses.

2.3 Variant Calling

In cancer sequencing projects, variant calling methods need to distinguish between germline and somatic variants. For somatic variant calling, it is crucial to use dedicated tools. Germline variant callers usually make assumptions about allele frequencies when they determine the most likely genotype as a measure to reduce the number of false-positive calls. For somatic variants, however, these assumptions cannot be made, because due to variable tumor cell content, aneuploidy, copy number aberrations, and the presence of different subclones, virtually all allele frequencies can be present.

2.3.1 Single-Nucleotide Variants (SNVs)

SNVs are the most abundant variant type in most cancer genomes. In general, SNV detection relies on single-nucleotide mismatches in the alignment after mapping the reads to the reference genome. Almost all SNV calling algorithms compute "pileups" (i.e., a section of all bases aligned to the respective position and their quality values through a stack of all reads overlapping this position), which are then used to determine the most likely genotype (germline variants) given the observed data or to determine the presence of a variant and its variant allele frequency (somatic variants). Prior to computation of the pileups, additional steps like local realignment or adjustment of quality scores might be performed.

Widely used tools for germline SNV calling include GATK HaplotypeCaller [23] and Platypus [24]. Dedicated somatic SNV callers are, for example, Mutect2 and our own pipeline based on samtools and a chain of empirical filters (https://dockstore.org/containers/quay.io/pancancer/pcawg-dkfz-workflow). FreeBayes [25] and Strelka2 [26] can be applied for both germline and somatic variant calling. A current review of somatic SNV calling algorithms is, for example, provided by [27].

2.3.2 Insertions and Deletions (Indels)

Like SNVs, small indels (usually up to 15–20 bp) can be directly detected from the alignment. Similar to SNV callers, many tools for small indel identification perform local realignment around indel candidate sites. For this reason, many recent tools combine SNV and small indel detection, like Mutect2 [28], Strelka2 [26], Platypus [24], and others. As described above, Mutect2 is designed to identify somatic variants, while Strelka2 and Platypus can be used for both germline and somatic variant calling. In our hands, Platypus works especially well for somatic indel calling if additional filtering rules are added which rescue variants that did not pass all Platypus internal filters to allow the detection of low allele frequency variants (https://dockstore.org/containers/quay.io/pancancer/pcawg-dkfz-workflow).

2.3.3 Structural Variants (SVs)

Structural variants comprise all types of genomic alterations except SNVs and small indels (*see* **Note 3**). This includes insertions, deletions, duplications, inversions, and translocations. Copy number aberrations (CNAs) are also a class of structural variants, but since different methods are used for CNA identification, CNAs are discussed separately in Subheading 2.3.4. Current tools for SV detection typically rely on one or several of the following signals: (a) discordant read pairs in the mapping of paired-end data, (b) split-read mapping, and (c) depth of coverage. Some tools as, for example, Manta [29], novoBreak [30], and SvABA [31] further employ assembly-based sequence reconstruction after initial candidate identification. Other popular tools for SV detection include DELLY [32] and LUMPY [33]. As the abovementioned signals (a–c) are independent from the assignment of tumor-normal pairs, these tools can be used for both germline and somatic variant calling.

2.3.4 Copy Number Aberrations (CNAs)

The detection of CNAs and allelic imbalances relies on changes in the depth of coverage and the B-allele frequencies of heterozygous SNPs. Before WGS became widely available, analogous information (with slightly lower resolution) could be derived from high-density SNP arrays, and indeed several CNA callers for WGS data are based on methods which have initially been developed for SNP array analysis. Most tools employ change-point detection methods to partition the genome into segments of equal coverage and allelic balance.

The depth of coverage in WGS data is often affected by various biases. The most common bias is GC bias, i.e., the dependence of the coverage on the GC content of the respective genomic window [34]. Another coverage bias, which is particularly prominent in fast-replicating tumor cells (e.g., Burkitt lymphomas), is replication timing bias [35, 36]. Samples affected by this bias have a higher coverage in early-replicating regions of the genome than in late-replicating regions. Many CNA callers correct for GC bias [36–38] and some also for replication timing bias [36] prior to genome segmentation to prevent the erroneous introduction of segment borders and thus over-segmentation. The inclusion of previously determined breakpoints (from SV calls) into the segmentation can additionally improve genome segmentation [36]. The segment borders indicate (relative) changes in the total or allele-specific copy numbers, but the absolute copy numbers of the segments cannot be directly inferred from depth of coverage and B-allele frequencies. Absolute allele-specific copy numbers can only be calculated if the tumor ploidy and tumor cell content (TCC) are known. Since this information is usually not known a priori, tools to determine absolute copy numbers are equipped with methods to estimate TCC and ploidy from the WGS data [36, 39, 40] (*see* **Note 4**).

2.4 Quality Control

At various stages of the analysis, thorough quality control (QC) should be performed.

The first QC can already be done at the level of FASTQ files so that samples with severe quality problems are not even processed further. A popular tool for this is FastQC [41]. Issues detected at this stage include problems originating from sequencing but also certain problems originating from library preparation like adapter contamination. The FastQC homepage [41] contains extensive documentation and also example reports for good and bad data.

Next, several parameters which are acquired during BAM file postprocessing and at BAM file level (e.g., through samtools flagstat) should be assessed. These include:

- The duplication rate. For WGS libraries the duplication rate is usually below 15%. A high duplication rate indicates that the library complexity is rather low. However, as long as the desired coverage (without counting duplicates) is still reached, downstream analyses are usually not affected.
- Median insert size and insert size distribution. The median insert size should be considerably larger than two times the read length for paired-end sequencing, as otherwise for many fragments the middle part would be sequenced twice. Such overlap between the mates can result in artifacts during variant calling (if the used tools do not handle the overlap adequately) and reduces the effective coverage. The insert size distribution should have only one mode and no long tail to either side. Libraries with bimodal insert size distribution can lead to drastically increased false-positive SV calls when choosing SV callers which use the paired-end insert size as detection criterion.
- Read pairs mapping to different chromosomes. Usually this value is below 5%; higher values indicate problems originating from library preparation. Samples with slightly increased values are usually not problematic for further analyses, but samples with drastically increased values (>15–20%) can cause errors especially during SV calling. Note that even in highly rearranged tumor genomes and in genomes with chromothripsis, this parameter is only slightly elevated.
- It should be verified that the required coverage has been reached. To determine the usable coverage, PCR duplicates should not be counted. Especially in libraries with small insert size (considerably smaller than two times the read length), also overlapping parts of both mates should only be counted once to avoid overestimation of the effectively usable coverage.

Finally, different variant calling methods themselves provide valuable QC information:

- Number and fraction of somatic SNVs which match an entry in dbSNP [42]. A very high number (greater than one million) of somatic SNVs in dbSNP indicate that tumor and matched normal are not from the same genetic background. This context is even more likely if the peak of the distribution of MAF values is close to 0.5. Such a situation can either be due to a sample swap or when the patient has received an allogenic stem cell transplant. An increased number and fraction of somatic SNVs in dbSNP with a low mutant allele frequency can indicate contamination of the tumor sample with DNA from another individual.
- Fraction of synonymous SNVs among all SNVs in coding regions. This fraction is increased when the tumor sample is contaminated with DNA from another species (e.g., mouse) as a result of cross contamination or if a xenograft is sequenced. Good samples usually have a synonymous fraction <0.35, while contaminated samples have often >0.5. Samples with values between 0.35 and 0.5 should be carefully checked. Samples which are contaminated with DNA from another species can be rescued by alignment against a combined reference genome containing both the human genome sequence and the genome of the respective species.
- Number of "intron deletions." A high number of deletions with breakpoints exactly at exon boundaries can indicate contamination with RNA. In addition, RNA-contaminated samples show an increased fraction of coding SNVs among all SNVs.
- Tumor cell content. Most CNA calling algorithms provide estimates for the tumor cell content, and it should be checked that the tumor cell content is high enough to enable somatic variant calling with adequate sensitivity.
- GC bias. Many sequencing libraries show a coverage bias which depends on the GC content of the sequenced fragment. CNA callers like ACEseq [36] correct the GC bias prior to copy number estimation to prevent false-positive copy number variant calls and can provide a quantitative estimate of the GC bias. Note that while GC bias can be corrected for CNA calling, strong GC bias will still negatively affect the identification of other types of variants due to the reduced coverage in parts of the genome.
- Other coverage fluctuations. Various problems during library preparation or with the template DNA can result in GC-independent coverage fluctuations. Such coverage fluctuations can lead to difficulties in the detection of CNAs and, like GC bias, will lead to reduced power for the detection of other variant types in genomic regions with low coverage.

2.5 Variant Annotation

To increase usability of variant information in biological analyses, the variants need to be annotated with functional information. This includes gene annotation to identify whether the variant affects, e.g., the protein-coding sequence of a gene; variant database information to disclose if a variant is, e.g., a known SNP or a known somatic cancer mutation; and potentially other information tracks, e.g., about sequence conservation or regulatory elements. Commonly used tools for variant annotation include ANNOVAR [43], SnpEff [44], variant effect predictor [45], and Rbbt [46]. Note that both the choice of the variant annotation software and the choice of the gene and transcript database can have a large impact on variant annotation and variant consequence prediction [47–50]. It is not possible to make a general recommendation which gene annotation should be used, but the choice should be made taking into consideration the aim of the study. For example, it should be evaluated whether a more comprehensive annotation is of higher importance or rather a more simple and reliable annotation. Furthermore, when it comes to the detection of high-impact variants, it may be beneficial to assess *a priori* the respective importance of sensitivity and specificity.

2.6 Identification of Driver Mutations

Driver genes are genes whose deregulation confers a selective advantage for the tumor. Mutations which cause such deregulation are called driver mutations. In a typical tumor genome, the vast majority of mutations are passenger mutations, i.e., mutations which have not been selected and have not conferred a clonal growth advantage. To understand the mechanisms of tumor development, it is important to delineate the driver mutations in a tumor genome. Methods to identify driver mutations search for signals of selection. This can be a higher mutation rate in a gene than expected by chance (e.g., MuSiC [51] or MutsigCV [52]), a bias toward high functional impact of mutations (e.g., Oncodrive-fm [53]), or clustering of mutations in certain parts of a protein (OncodriveCLUST [54]). Recent approaches have generalized these techniques to also enable the detection of driver mutations in noncoding regions of the genome; examples include LARVA [55] and OncodriveFML [56].

2.7 Mutational Signature Analysis

The set of mutations in a cancer genome (including both driver and passenger mutations) is the imprint of the activity of multiple mutational processes. Several mutational processes have specific preferences with respect to the caused mutations. For SNVs, this means that a specific mutational process has a certain probability to introduce each of the possible transitions or transversions. Furthermore, this specificity does not only result in different nucleotide exchanges but also extends over the nucleotides flanking the mutated position. Each mutational process thus leaves a certain footprint in the genome, which is referred to as a mutational

signature [57]. An analysis of mutational signatures can therefore provide insights into the mutational processes that have been active during the life history of a tumor and its precursor cell.

2.7.1 Unsupervised Analysis of Mutational Signatures

A mutational catalogue contains the frequencies of SNVs in their trinucleotide context per sample. It is possible to run a de novo extraction of mutational signatures from the mutational catalogue of a sufficiently large number of cancer genomes. To this end, the mutational catalogue is decomposed in a mutational signature matrix and an exposure matrix (which contains the activity of each identified signature in each genome) using nonnegative matrix factorization (NMF). In a large-scale analysis of more than 500 WGS and more than 6,500 exome samples from 30 different tumor entities, Alexandrov et al. identified 21 mutational signatures, of which around one-half could be associated to known mutational processes [57]. Later this analysis has been extended to more than 1,000 WGS samples and more than 10,000 exomes from 40 entities, and 30 validated mutational signatures have been identified. This set of signatures is available from the COSMIC database [58].

Different frameworks are available to perform unsupervised analysis of mutational signatures. The original framework by Alexandrov et al. [57] is implemented in MATLAB. As alternative, an R package for mutational signature analysis is available from the Bioconductor Project [59]. Finally, the R package Bratwurst [60] can be used for different types of NMF analyses, including unsupervised mutational signature analysis. It provides wrapper functions for NMF solvers on graphical processing units (GPUs) using the Compute Unified Device Architecture (CUDA) 8 framework and the CUDAMat library [61] and therefore enables much faster NMF calculations due to massive parallelization.

2.7.2 Supervised Analysis of Mutational Signatures

Unsupervised mutational signature analysis enables the detection of novel signatures. However, unsupervised signature analysis requires the availability of large cohorts of samples. Especially for smaller studies, a supervised analysis of mutational signatures is therefore the better option. In a supervised analysis, the contributions of known mutational signatures to a given mutational catalogue are determined. This approach requires much less statistical power and can hence be applied to small cohorts or even to single cases. Supervised mutational signature analysis can be performed, e.g., with the R packages deconstructSigs [62] or YAPSA [63].

3 Notes

1. Analysis of samples without matched control may be performed if databases (e.g., dbSNP [42] or more recent resources

like ExAC or gnomAD [64]) are used to remove common variants. However, with such an approach, a complete removal of germline variants from the somatic set is not possible since every individual has private germline variants. It should be noted that certain databases like dbSNP do also contain well-studied somatic variants, and filtering against these databases without additional selection criteria should thus be avoided. In case of deep sequencing (usually >100×) of tumors of relatively low tumor cell content (<70%), the mutant allele fraction can be employed for a better discrimination between germline and somatic variants. Here the allele-specific copy numbers of the genomic segment where the variant is located have to be taken into account, and it is recommended to filter against a panel of independent normal control samples processed with the same workflows as the tumor samples to remove pipeline-specific artifacts. Some dedicated tools for the analysis of tumor samples without matched controls integrate various sources of information to predict somatic mutations [65–67].

2. Although current analyses typically make no use of duplicate reads, we recommend to just mark them instead of removing them from the BAM file to keep the full information in the BAM file. With this strategy, the FASTQ files can be reconstructed from the BAM files and might be discarded to save disk space. If read trimming algorithms were applied before alignment, however, loss of information might occur, and FASTQ files should be kept regardless of whether duplicates were only marked or removed.
3. Many SV callers have a reduced sensitivity for the identification of SVs of short length (20–300 bp) [29, 31], and hence there might be a gap between the events reported by small indel callers and SV callers. If SVs in this size range are in the focus of the analysis, a careful choice of tools and possibly the use of multiple tools is recommended.
4. The determination of TCC and ploidy from WGS data is nontrivial, and often multiple combinations of TCC and ploidy lead to good fits to the data. Furthermore, all existing methods make assumptions about the tumor sample (e.g., that the majority of aberrations is present in all tumor cells or that there is only a low number of subclones in the sample) which are not necessarily true for the analyzed samples. Certain conditions like the presence of tumor cells of different ploidy (i.e., diploid and tetraploid tumor cells after genome duplication as late event in tumor evolution) are to our experience not handled well by any existing tool. It might therefore be necessary to employ complementary techniques like FISH or karyotyping to determine the tumor cell ploidy and reliably estimate absolute copy numbers.

Acknowledgments

This work has been supported by the German Ministry of Science and Education (BMBF) in the framework of the ICGC MMML-Seq (01KU1002A-J) and the ICGC DE-MINING (01KU1505E) projects and the Heidelberg Center for Human Bioinformatics (HD-HuB) within the German Network for Bioinformatics Infrastructure (de.NBI) (#031A537A, #031A537C). We are grateful to all present and previous members of the Division of Theoretical Bioinformatics, the DKFZ-HIPO bioinformatics team, the Omics IT and Data Management Core Facility, and the Bioinformatics and Omics Data Analytics group of the German Cancer Research Center (DKFZ, Heidelberg) as well as coworkers in the ICGC MMML-seq and PedBrain projects who were involved in the establishment of the procedures described here.

References

1. Stratton MR, Campbell PJ, Futreal PA (2009) The cancer genome. Nature 458:719–724. https://doi.org/10.1038/nature07943
2. Ley TJ, Mardis ER, Ding L et al (2008) DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. Nature 456:66–72. https://doi.org/10.1038/nature07485
3. Meyerson M, Gabriel S, Getz G (2010) Advances in understanding cancer genomes through second-generation sequencing. Nat Rev Genet 11:685–696
4. Alioto TS, Buchhalter I, Derdak S et al (2015) A comprehensive assessment of somatic mutation detection in cancer using whole-genome sequencing. Nat Commun 6:10001. https://doi.org/10.1038/ncomms10001
5. Davies H, Glodzik D, Morganella S et al (2017) HRDetect is a predictor of BRCA1 and BRCA2 deficiency based on mutational signatures. Nat Med 23:517–525. https://doi.org/10.1038/nm.4292
6. Hudson TJ, Anderson W, Aretz A et al (2010) International network of cancer genome projects. Nature 464:993–998. https://doi.org/10.1038/nature08987
7. Robbe P, Popitsch N, Knight SJL et al (2018) Clinical whole-genome sequencing from routine formalin-fixed, paraffin-embedded specimens: pilot study for the 100,000 Genomes Project. Genet Med 20(10):1196–1205. https://doi.org/10.1038/gim.2017.241
8. Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170
9. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754–1760. https://doi.org/10.1093/bioinformatics/btp324
10. Li H (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. https://www.arxiv.org/abs/1303.3997
11. Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359. https://doi.org/10.1038/nmeth.1923
12. Marco-Sola S, Sammeth M, Guigó R, Ribeca P (2012) The GEM mapper: fast, accurate and versatile alignment by filtration. Nat Methods 9:1185–1188. https://doi.org/10.1038/nmeth.2221
13. Li H, Homer N (2010) A survey of sequence alignment algorithms for next-generation sequencing. Brief Bioinform 11:473–483
14. Treangen TJ, Salzberg SL (2011) Repetitive DNA and next-generation sequencing: computational challenges and solutions. Nat Rev Genet 13:36–46. https://doi.org/10.1038/nrg3117
15. Lippert RA (2005) Space-efficient whole genome comparisons with burrows–wheeler transforms. J Comput Biol 12:407–415. https://doi.org/10.1089/cmb.2005.12.407
16. Li H, Durbin R (2010) Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 26:589–595. https://doi.org/10.1093/bioinformatics/btp698

17. Li H, Handsaker B, Wysoker A et al (2009) The sequence alignment/map format and SAMtools. Bioinformatics 25:2078–2079. https://doi.org/10.1093/bioinformatics/btp352
18. BroadInstitute (2016) Picard Tools—By Broad Institute. http://broadinstitute.github.io/picard/. Accessed 6 May 2018
19. Tischler G, Leonard S (2014) Biobambam: tools for read pair collation based algorithms on BAM files. Source Code Biol Med 9:13
20. Tarasov A, Vilella AJ, Cuppen E et al (2015) Sambamba: fast processing of NGS alignment formats. Bioinformatics 31:2032–2034. https://doi.org/10.1093/bioinformatics/btv098
21. Van der Auwera GA, Carneiro MO, Hartl C et al (2013) From fastQ data to high-confidence variant calls: the genome analysis toolkit best practices pipeline. Curr Protoc Bioinformatics 43:11.10.1–11.1033. https://doi.org/10.1002/0471250953.bi1110s43
22. Depristo MA, Banks E, Poplin R et al (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet 43:491–501. https://doi.org/10.1038/ng.806
23. Poplin R, Ruano-Rubio V, DePristo MA, et al (2017) Scaling accurate genetic variant discovery to tens of thousands of samples. bioRxiv 201178. https://doi.org/10.1101/201178
24. Rimmer A, Phan H, Mathieson I et al (2014) Integrating mapping-, assembly- and haplotype-based approaches for calling variants in clinical sequencing applications. Nat Genet 46:912–918. https://doi.org/10.1038/ng.3036
25. Garrison E, Marth G (2012) Haplotype-based variant detection from short-read sequencing. arXiv:1207.3907. https://arxiv.org/abs/1207.3907
26. Kim S, Scheffler K, Halpern AL, et al (2017) Strelka2: Fast and accurate variant calling for clinical sequencing applications. bioRxiv 192872. https://doi.org/10.1101/192872
27. Xu C (2018) A review of somatic single nucleotide variant calling algorithms for next-generation sequencing data. Comput Struct Biotechnol J 16:15–24
28. Cibulskis K, Lawrence MS, Carter SL et al (2013) Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nat Biotechnol 31:213–219. https://doi.org/10.1038/nbt.2514
29. Chen X, Schulz-Trieglaff O, Shaw R et al (2016) Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. Bioinformatics 32:1220–1222. https://doi.org/10.1093/bioinformatics/btv710
30. Chong Z, Ruan J, Gao M et al (2017) novoBreak: local assembly for breakpoint detection in cancer genomes. Nat Methods 14:65–67. https://doi.org/10.1038/nmeth.4084
31. Wala JA, Bandopadhayay P, Greenwald NF et al (2018) SvABA: genome-wide detection of structural variants and indels by local assembly. Genome Res 28:581–591. https://doi.org/10.1101/gr.221028.117
32. Rausch T, Zichner T, Schlattl A et al (2012) DELLY: structural variant discovery by integrated paired-end and split-read analysis. Bioinformatics 28:i333–i339. https://doi.org/10.1093/bioinformatics/bts378
33. Layer RM, Chiang C, Quinlan AR, Hall IM (2014) LUMPY: a probabilistic framework for structural variant discovery. Genome Biol 15:R84. https://doi.org/10.1186/gb-2014-15-6-r84
34. Benjamini Y, Speed TP (2012) Summarizing and correcting the GC content bias in high-throughput sequencing. Nucleic Acids Res 40:e72. https://doi.org/10.1093/nar/gks001
35. Koren A, Handsaker RE, Kamitaki N et al (2014) Genetic variation in human DNA replication timing. Cell 159:1015–1026. https://doi.org/10.1016/j.cell.2014.10.025
36. Kleinheinz K, Bludau I, Huebschmann D, et al (2017) ACEseq—allele specific copy number estimation from whole genome sequencing. bioRxiv 210807. https://doi.org/10.1101/210807
37. Boeva V, Popova T, Bleakley K et al (2012) Control-FREEC: a tool for assessing copy number and allelic content using next-generation sequencing data. Bioinformatics 28:423–425. https://doi.org/10.1093/bioinformatics/btr670
38. Favero F, Joshi T, Marquard AM et al (2015) Sequenza: allele-specific copy number and mutation profiles from tumor sequencing data. Ann Oncol 26:64–70. https://doi.org/10.1093/annonc/mdu479
39. Van Loo P, Nordgard SH, Lingjærde OC et al (2010) Allele-specific copy number analysis of tumors. Proc Natl Acad Sci U S A 107:16910–16915. https://doi.org/10.1073/pnas.1009843107
40. Carter SL, Cibulskis K, Helman E et al (2012) Absolute quantification of somatic DNA alterations in human cancer. Nat Biotechnol 30:413–421. https://doi.org/10.1038/nbt.2203

41. Simon A (2010) FastQC: a quality control tool for high throughput sequence data. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/
42. Sherry ST, Ward MH, Kholodov M et al (2001) dbSNP: the NCBI database of genetic variation. Nucleic Acids Res 29:308–311
43. Wang K, Li M, Hakonarson H (2010) ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res 38:e164–e164. https://doi.org/10.1093/nar/gkq603
44. Cingolani P, Platts A, Wang LL et al (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin) 6:80–92. https://doi.org/10.4161/fly.19695
45. McLaren W, Gil L, Hunt SE et al (2016) The Ensembl variant effect predictor. Genome Biol 17:122. https://doi.org/10.1186/s13059-016-0974-4
46. Vazquez M, Nogales R, Carmona P et al (2010) Rbbt: a framework for fast bioinformatics development with ruby. Springer, Berlin, Heidelberg
47. McCarthy DJ, Humburg P, Kanapin A et al (2014) Choice of transcripts and software has a large effect on variant annotation. Genome Med 6:26. https://doi.org/10.1186/gm543
48. Frankish A, Uszczynska B, Ritchie GR et al (2015) Comparison of GENCODE and RefSeq gene annotation and the impact of reference geneset on variant effect prediction. BMC Genomics 16:S2. https://doi.org/10.1186/1471-2164-16-S8-S2
49. Wu PY, Phan JH, Wang MD (2013) Assessing the impact of human genome annotation choice on RNA-seq expression estimates. BMC Bioinformatics 14(Suppl 1):S8. https://doi.org/10.1186/1471-2105-14-S11-S8
50. Zhao S, Zhang B (2015) A comprehensive evaluation of ensembl, RefSeq, and UCSC annotations in the context of RNA-seq read mapping and gene quantification. BMC Genomics 16:97. https://doi.org/10.1186/s12864-015-1308-8
51. Dees ND, Zhang Q, Kandoth C et al (2012) MuSiC: Identifying mutational significance in cancer genomes. Genome Res 22:1589–1598. https://doi.org/10.1101/gr.134635.111
52. Lawrence MS, Stojanov P, Polak P et al (2013) Mutational heterogeneity in cancer and the search for new cancer-associated genes. Nature 499:214–218. https://doi.org/10.1038/nature12213
53. Gonzalez-Perez A, Lopez-Bigas N (2012) Functional impact bias reveals cancer drivers. Nucleic Acids Res 40:e169. https://doi.org/10.1093/nar/gks743
54. Tamborero D, Gonzalez-Perez A, Lopez-Bigas N (2013) OncodriveCLUST: exploiting the positional clustering of somatic mutations to identify cancer genes. Bioinformatics 29:2238–2244. https://doi.org/10.1093/bioinformatics/btt395
55. Lochovsky L, Zhang J, Fu Y et al (2015) LARVA: an integrative framework for large-scale analysis of recurrent variants in noncoding annotations. Nucleic Acids Res 43:8123–8134. https://doi.org/10.1093/nar/gkv803
56. Mularoni L, Sabarinathan R, Deu-Pons J et al (2016) OncodriveFML: a general framework to identify coding and non-coding regions with cancer driver mutations. Genome Biol 17:128. https://doi.org/10.1186/s13059-016-0994-0
57. Alexandrov LB, Nik-Zainal S, Wedge DC et al (2013) Signatures of mutational processes in human cancer. Nature. https://doi.org/10.1038/nature12477
58. COSMIC—signatures of mutational processes in human cancer. https://cancer.sanger.ac.uk/cosmic/signatures. Accessed 9 May 2018
59. Gehring JS, Fischer B, Lawrence M, Huber W (2015) SomaticSignatures: inferring mutational signatures from single-nucleotide variants. Bioinformatics 31:3673–3675. https://doi.org/10.1093/bioinformatics/btv408
60. Huebschmann D, Kurzawa N, Steinhauser S, et al (2017) Deciphering programs of transcriptional regulation by combined deconvolution of multiple omics layers. bioRxiv 199547. https://doi.org/10.1101/199547
61. Mnih V (2009) CUDAMat: a CUDA-based matrix class for Python. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.232.4776&rep=rep1&type=pdf
62. Rosenthal R, McGranahan N, Herrero J et al (2016) deconstructSigs: delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. Genome Biol 17:31. https://doi.org/10.1186/s13059-016-0893-4
63. Huebschmann D, Gu Z, Schlesner M (2015) YAPSA: yet another package for signature analysis. R package. http://bioconductor.org/packages/release/bioc/html/YAPSA.html
64. Lek M, Karczewski KJ, Minikel EV et al (2016) Analysis of protein-coding genetic variation in 60,706 humans. Nature 536:285–291. https://doi.org/10.1038/nature19057

65. Kalatskaya I, Trinh QM, Spears M et al (2017) ISOWN: accurate somatic mutation identification in the absence of normal tissue controls. Genome Med 9:59. https://doi.org/10.1186/s13073-017-0446-9
66. Smith KS, Yadav VK, Pei S et al (2016) SomVarIUS: somatic variant identification from unpaired tissue samples. Bioinformatics 32:808–813. https://doi.org/10.1093/bioinformatics/btv685
67. Madubata CJ, Roshan-Ghias A, Chu T et al (2017) Identification of potentially oncogenic alterations from tumor-only samples reveals Fanconi anemia pathway mutations in bladder carcinomas. NPJ Genomic Med 2:29. https://doi.org/10.1038/s41525-017-0032-5

Chapter 16

Protocols for CRISPR-Cas9 Screening in Lymphoma Cell Lines

Daniel E. Webster, Sandrine Roulland, and James D. Phelan

Abstract

Genome-wide screens are a powerful technique to dissect the complex network of genes regulating diverse cellular phenotypes. The recent adaptation of the CRISPR-Cas9 system for genome engineering has revolutionized functional genomic screening. Here, we present protocols used to introduce Cas9 into human lymphoma cell lines, produce high-titer lentivirus of a genome-wide sgRNA library, transduce and culture cells during the screen, isolate genomic DNA, and prepare a custom library for next-generation sequencing. These protocols were tailored for loss-of-function CRISPR screens in human lymphoma cell lines but are highly amenable for other experimental purposes.

Key words CRISPR, CRISPR-Cas9, Lymphoma, DLBCL, Functional genomics, High-throughput screen

1 Introduction

The field of functional genomics emerged more than two decades ago [1] and has rapidly evolved by incorporating technologies to interrogate specific DNA sequences. The aim of this field is to identify specific genomic elements responsible for a given phenotype, and until recently, this has largely been achieved through the implementation of screening technologies using siRNA or shRNA to reduce the level of mRNA and inhibit gene function. However, the recent discovery of an ancient adaptive immune system in prokaryotes, known as CRISPR, has allowed for large-scale genetic deletion screens that have transformed functional genomics into a tool available for most biomedical research laboratories. The following chapter describes the methods we used to uncover a multiprotein complex present in cancer cells that respond to targeted therapies [2].

The clustered regularly interspaced short palindromic repeats (CRISPR)-associated proteins (Cas) are a family of adaptive immune genes identified in bacteria that function to remove viral

Ralf Küppers (ed.), *Lymphoma: Methods and Protocols*, Methods in Molecular Biology, vol. 1956,
https://doi.org/10.1007/978-1-4939-9151-8_16,

associated DNA from their host [3]. The elegant system has evolved many Cas variants that can function with two RNA guides. The CRISPR-RNA (crRNA) can bind target sequences in the genome through complementary base pairing, and the tracrRNA (trans-activating crRNA) can bind both the crRNA and Cas9 [4, 5]. When bound in a complex on DNA, the CRISPR-Cas system homes an endonuclease to a sequence in the genome. If a species-specific sequence, known as the protospacer adjacent motif (PAM), is directly adjacent to the crRNA complementary sequence, the Cas endonuclease can enzymatically cleave both strands of DNA.

The discovery of this system was quickly recognized as an engineering opportunity to repurpose a fine-tunable delivery system for genome engineering in mammalian cells. Indeed, the Cas9 protein from *S. pyogenes* was combined with a single guide RNA (sgRNA, or gRNA in some literature,) containing elements of both the tracrRNA and the crRNA. When Cas9 was combined with an sgRNA in mammalian cells, sequencing revealed disrupted nucleotide sequences at the loci targeted by the sgRNAs [6–8]. Cas9-sgRNA complexes can induce double-strand DNA breaks that are frequently repaired through the error-prone process of nonhomologous end joining (NHEJ) resulting in insertions and deletions at the target site. This process frequently leads to the introduction of frameshifts and premature stop codons resulting in a genetic knockout. Subsequent studies coupled lentiviral delivery of Cas9 with large-scale libraries of sgRNAs targeting nearly every protein-coding gene in the genome. Thus, genome-wide knockout screens were proven to be effective in mammalian cells [9, 10].

CRISPR-Cas9 screens have revolutionized the dissection of regulatory pathways in mammalian cells in general, and cancer cells specifically, revealing mechanisms of cell proliferation, survival, and drug resistance [9–14]. In contrast to previous screening methods such as shRNA or siRNA, CRISPR-Cas9 screens have shown fewer off-target effects, better reproducibility, and greater versatility [15]. In fact, Cas9 has been adapted to serve as a delivery device for fusion proteins optimized for imaging, gene activation, gene silencing, epigenetic modification, and DNA mutagenesis [16–21]. These customizable DNA-binding proteins allow for the screening of phenotypes previously unimagined.

Here, we describe the steps necessary to perform unbiased genome-wide, loss-of-function CRISPR-Cas9 genetic screens to elucidate tumor subtype-specific drivers of oncogenesis. In general, the procedures described below include the generation of Cas9-expressing cell lines, the production of high-titer pooled sgRNA library virus, transduction and cell culture procedures for screening, and the preparation of an Illumina-compatible library for next-generation sequencing (Fig. 1). Using these procedures in cell line models of the two major subtypes of diffuse large B cell

Fig. 1 General screening overview. First, obtain Cas9 plasmids and engineer efficient cutting cell lines. Next, produce, titer, and transduce sgRNA library into target cells of interest at a low MOI. Then, select transduced cells, and prepare pellets for DNA purification at day 0 and day 21 of doxycycline treatment. After DNA isolation, use a nested PCR approach and size selection to prepare a NextGen library containing the sgRNA sequences. Finally, sequence the library, and analyze relative sgRNA fractions between experimental (day 21) and control (day 0) populations

lymphoma (DLBCL), we uncovered fundamentally different modes of B cell receptor signaling [2] that revealed the molecular basis of exceptional responses to BTK-targeted therapies observed in clinical trials [22]. However, these procedures are highly adaptable to other CRISPR-based systems and phenotypic screens.

2 Materials

2.1 Plasmids

pCW-Cas9-blast (addgene.org).

pLKO.1-puro U6 sgRNA vector (addgene.org).

sgRNA library, Brunello (addgene.org).

Lentiviral packaging construct, psPAX2 (addgene.org).

Lentiviral envelope construct, pMD2.G (addgene.org).

EndoFree Plasmid Maxi Kit (QIAGEN).

2.2 Generation of Cas9-Expressing Cell Lines

293FT cells (Thermo Fisher).

DMEM (Thermo Fisher).

Fetal bovine serum (Atlanta Biologicals).

Penicillin-streptomycin-glutamine (100×) (Thermo Fisher).

TransIT-293 transfection reagent (Mirus).

Lenti-X concentrator (Takara/Clontech)

Phosphate-buffered saline (Thermo Fisher).

0.45 μm low protein-binding syringe-drive filter (Millipore)

12 mL syringe (Thermo Fisher).

Blasticidin (Thermo Fisher).

sgRNA vector targeting nonessential surface markers: for example, CD20 (5′ GATCATCAGAAGACCCCCCA 3′) or ICAM1 (5′ ACCCCGGGCCAACCTCACCG 3′).

Antibodies: for example, CD20 (anti-human CD20, clone 2H7, Biolegend) or ICAM1 (anti-human CD54, clone HCD54).

Flow cytometer (BD Biosciences).

2.3 Library Virus Preparation and Titration

Puromycin (Thermo Fisher).

Stericup filter units, 0.45 μm, PVDF (Millipore).

2.4 Transduction, Cell Culture, and Pellet Collection

Doxycycline (Sigma).

T225 cell culture flasks, filter cap, sterile (Thermo Fisher).

Cell culture medium appropriate for target cell lines, e.g., RPMI 1640.

2.5 Library Preparation and Detection of sgRNA Abundance

QIAamp DNA Blood Maxi Kit (QIAGEN).

Heat block (70 °C).

NanoDrop spectrophotometer (NanoDrop).

PCR-grade water (Quality Biological).

Takara ExTaq (Takara/Clontech).

Primers U6-F1 and Tracer-R1 (Table 1).

Primers D501-D508, D701-D712 (Table 2).

96-well PCR plates (Thermo Fisher).

PCR plate covers (Thermo Fisher).

Thermocycler.

1.5 mL Eppendorf tubes (Thermo Fisher).

E-Gel Electrophoresis System (Invitrogen).

E-Gel SizeSelect II Agarose Gels, 2% (Invitrogen).

E-Gel ladder (Invitrogen).

Qubit dsDNA HS Assay Kit (Thermo Fisher).

Qubit assay tubes (Thermo Fisher).

Access to an Illumina NextSeq500.

Table 1
Primary PCR primer sequences U6-F1 and Tracer-R1

Primer	Sequence (5′ to 3′)
U6-F1	AATGGACTATCATATGCTTACCGTAACTTGAAAGTATTTCG
Tracer-R1	GTAATTCTTTAGTTTGTATGTCTGTTGCTATTATG

Table 2
Secondary PCR primer sequences D501–D508, D701–D712

Primer	Sequence (5′ to 3′)
D501	AATGATACGGCGACCACCGAGATCTACACTATAGCCTACACTCTTTCCCTACACGACGCTC TTCCGATCTATGCATGCTCTTGTGGAAAGGACGAAA CACCG
D502	AATGATACGGCGACCACCGAGATCTACACATAGAGGCACACTCTTTCCCTACACGACGCTC TTCCGATCTTGCATGCAGTCTTGTGGAAAGGACGAA ACACCG
D503	AATGATACGGCGACCACCGAGATCTACACCCTATCCTACACTCTTTCCCTACACGACGCTC TTCCGATCTGCATGCATCGTCTTGTGGAAAGGACGA AACACCG
D504	AATGATACGGCGACCACCGAGATCTACACGGCTCTGAACACTCTTTCCCTACACGACGCTC TTCCGATCTCATGCATGACGTCTTGTGGAAAGGACG AAACACCG
D505	AATGATACGGCGACCACCGAGATCTACACAGGCGAAGACACTCTTTCCCTACACGACGCTC TTCCGATCTCGTACGTATACATCTTGTGGAAAGGAC GAAACACCG
D506	AATGATACGGCGACCACCGAGATCTACACTAATCTTAACACTCTTTCCCTACACGACGCTC TTCCGATCTACGTACGTGTATATCTTGTGGAAAGGA CGAAACACCG
D507	AATGATACGGCGACCACCGAGATCTACACCAGGACGTACACTCTTTCCCTACACGACGCTC TTCCGATCTTACGTACGCGTGTATCTTGTGGAAAGG ACGAAACACCG
D508	AATGATACGGCGACCACCGAGATCTACACGTACTGACACACTCTTTCCCTACACGACGCTC TTCCGATCTGTACGTACACCCGTATCTTGTGGAAAG GACGAAACACCG
D701	CAAGCAGAAGACGGCATACGAGATCGAGTAATGTGACTGGAGTTCAGACGTGTGCTC TTCCGATCtctactattctttcccctgcactgt
D702	CAAGCAGAAGACGGCATACGAGATTCTCCGGAGTGACTGGAGTTCA GACGTGTGCTCTTCCGATCtctactattctttcccctgcactgt

(continued)

Table 2
(continued)

Primer	Sequence (5′ to 3′)
D703	CAAGCAGAAGACGGCATACGAGATAATGAGCGGTGACTGGAGTTCAGACGTGTGC TCTTCCGATCtctactattctttcccctgcactgt
D704	CAAGCAGAAGACGGCATACGAGATGGAATCTCGTGACTGGAGTTCAGACGTGTG CTCTTCCGATCtctactattctttcccctgcactgt
D705	CAAGCAGAAGACGGCATACGAGATTTCTGAATGTGACTGGAGTTC AGACGTGTGCTCTTCCGATCtctactattctttcccctgcactgt
D706	CAAGCAGAAGACGGCATACGAGATACGAATTCGTGACTGGAGTTCAGACGTG TGCTCTTCCGATCtctactattctttcccctgcactgt
D707	CAAGCAGAAGACGGCATACGAGATAGCTTCAGGTGACTGGAGTTCAGACGT GTGCTCTTCCGATCtctactattctttcccctgcactgt
D708	CAAGCAGAAGACGGCATACGAGATGCGCATTAGTGACTGGAGTTCAGACG TGTGCTCTTCCGATCtctactattctttcccctgcactgt
D709	CAAGCAGAAGACGGCATACGAGATCATAGCCGGTGACTGGAGTTCAGAC GTGTGCTCTTCCGATCtctactattctttcccctgcactgt
D710	CAAGCAGAAGACGGCATACGAGATTTCGCGGAGTGACTGGAGTTCAGA CGTGTGCTCTTCCGATCtctactattctttcccctgcactgt
D711	CAAGCAGAAGACGGCATACGAGATGCGCGAGAGTGACTGGAGTTCAG ACGTGTGCTCTTCCGATCtctactattctttcccctgcactgt
D712	CAAGCAGAAGACGGCATACGAGATCTATCGCTGTGACTGGAGTTCAG ACGTGTGCTCTTCCGATCtctactattctttcccctgcactgt

3 Methods

3.1 Obtain Plasmids

Prepare EndoFree Maxiprep DNA stocks of plasmids for Cas9 and lentiviral envelope and packaging constructs. We replaced the puromycin fusion gene from pCW-Cas9-Puro (Addgene.org #50661) with a blasticidin resistance gene (pcW-Cas9-Blast); however, the general procedure described below can be adapted for any Cas-like genome-editing system.

3.2 Generate Cas9 Efficient Clonal Cell Lines

We have found that single-cell cloning a transduced Cas9-expressing population of cells and functionally testing for gene inactivation are excellent methods to ensure inducible, complete inactivation of a target gene. All-in-one-vector systems, containing both Cas9 and a sgRNA, have been reported [10, 23]; however, we find two vector systems to be more efficient for genome editing in DLBCL cell lines. While Cas9 protein expression can be measured with standard antibody-based techniques, we find measuring Cas9 cutting activity a better predictor of the functional efficiency of a given cell line. We therefore use FACS to quickly test the cutting efficiency of many single-cell clones expressing a sgRNA targeting a nonessential surface marker expressed on the target population. The following describes an approach using the HEK293 FT cell line optimized for lentiviral production using Mirus TransIT-293T; however, it can be modified for optimal viral titers using other viral constructs for other target cells.

1. Plate approximately 3×10^6 293FT cells in 10 cm plate 1 day before transfection in 9 mL of fresh DMEM supplemented with 10% FBS, 1% penicillin-streptomycin, and 1% Gluta-MAX.
2. The following day, combine 6 μg of Cas9 plasmid, 4.5 μg of psPAX2, and 1.5 μg of pMD2.G in 1 mL of Opti-MEM. Vortex briefly.
3. Add 36 μL of TransIT-293 transfection reagent to the DNA-media mixture. Vortex briefly and incubate the mixture at room temperature for 15 min.
4. Gently pipet the transfection mixture on a semi-confluent plate of 293FT (70–80%). Swirl the plate to thoroughly mix before returning the cells to a 37 °C CO_2 incubator.
5. Viral particles can be harvested by collecting the supernatants of the 293FT cells at 24, 48, and 72 h post transfection. Carefully remove 10 mL of supernatant, and replace with fresh DMEM media supplemented with 10% FBS, 1% penicillin-streptomycin, and 1% Gluta-MAX.
6. Centrifuge supernatants for 5 min at 500 × *g* to pellet any cells. Pass supernatants through a 0.45 μM low protein-binding filter using a 12 mL syringe (or Stericup filter units for larger volumes), and combine with Lenti-X concentrator at a 3:1 ratio.

7. After all viral collections have been made, combine and centrifuge at 1500 × *g* for 45 min at 4 °C. Decant and resuspend pelleted virus at desired concentration in sterile-filtered PBS. Combine all resuspended aliquots to homogenize viral titers between various viral supernatant harvests. Store aliquots of virus at −80 °C.
8. Transduce 0.5×10^6 target cells in a 24-well plate with 50 μL of 40× concentrated Cas9 virus. Repeating the infection 24 h later may increase Cas9 expression and subsequent clone selection efficiency. Incubate cells with concentrated virus for 3 days to allow for maximal viral integration and protein expression of antibiotic resistance genes.
9. Select transduced cells with a range of blasticidin doses (1–50 μg/mL). Untransduced cells should be used as an appropriate control to find the optimal selection dose. Selection can take up to 7 days, and additional blasticidin should only be added to fresh media to maintain a 1× concentration.
10. Transduce a fraction of the selected pCW-Cas9-Blast cells with a GFP-marked sgRNA vector targeting a nonessential surface marker such as ICAM1, CD20, CD10, or others. Three days after transduction, split the transduced cells into two wells, and add doxycycline-containing media to one of the two wells.
11. Maintain the cultures for 2 weeks after dox induction, and measure the surface expression of the targeted gene by FACS (*see* **Notes 1** and **2**).
12. Establish single-cell clones from the functionally validated pool of Cas9-expressing cells using limiting dilution or FACS sorting. Expand clones, and repeat the FACS assay to test Cas9 function on each clone in the presence and absence of dox.

3.3 Library Virus Preparation and Titration

1. Select a suitable library of sgRNAs (either purchased through addgene.org or custom designed), and generate concentrated lentivirus as described in Subheading 3.2 (*see* **Note 3**).
2. Titrate sgRNA library virus on each cell line to be screened. Using 1×10^6 cells per virus volume, combine 100, 50, 25, 12.5, 6.25, or 0 μL of concentrated virus in a 6-well plate. Incubate cells and virus together at 37 °C for 3 days.
3. Split each condition to a new plate, such that two identical 6-well plates exist. Add puromycin at appropriate concentration to one of the two plates. Add media as needed to maintain optimal growth conditions. Incubate at 37 °C until untransduced, puromycin-treated cells have died (*see* **Note 4**).
4. Enumerate live cells in each well.
5. Calculate the transduction percentage by dividing the number of live cells in the puromycin-treated sample by the number of

live cells in the untreated sample for each volume of virus. Calculate a linear regression for the data points in the linear range of the curve using the volume of virus and transduction percentage as data points.

6. Estimate the volume of virus needed to achieve 33% transduction (multiplicity of infection of ~0.3) from the linear regression, and multiply it by the total cell number needed for the desired coverage of the library (*see* **Note 5**).

3.4 Transduction, Cell Culture, and Pellet Collection

1. Calculate cell number required for desired library coverage, and multiply by 3 (as only 33% of cells will be transduced and survive selection).
2. Combine virus volume calculated in Subheading 3.3, **step 6** with cell number calculated in Subheading 3.4, **step 1** for each replicate of the screen. We find two independently transduced replicates sufficient for genome-wide screening, while three are preferred for smaller focused libraries. Incubate virus and cells at 37 °C for 3 days.
3. Treat each replicate with appropriate concentration of puromycin, and add fresh media as needed (*see* **Note 6**).
4. When puromycin selection is complete, expand transduced cells until 2× library coverage is achieved. Pellet 1× library coverage of cells by centrifugation at 300 × *g* for 5 min at room temperature. Wash cells twice in PBS, combining all cell pellets for a given biological replicate, and freeze the decanted pellet as a day 0 reference control.
5. Replace media with doxycycline-containing media. Continue passaging library cultures for 3 weeks at optimal growth conditions, counting and splitting every 2–3 days. At each passage, enumerate total cell numbers so population doublings can be calculated and library coverage maintained. After 3 weeks of culture, pellet 1× library coverage of cells by centrifugation at 300 × *g* for 5 min at room temperature. Wash cells twice in PBS, combining all cell pellets for a given biological replicate, and freeze the decanted pellet as a day 21 experimental sample.

3.5 Library Preparation and Detection of sgRNA Abundance

On average, each cell will contain a single sgRNA that can serve as a unique barcode sequence to be counted by next-generation sequencing. By comparing the relative frequency between the day 0 and day 21 samples, the selective pressure of a given sgRNA can be inferred. To do so, the genomic DNA from each cell must be isolated, the sgRNA sequence amplified, and sequencing adapters appended before the sequences can be read by next-generation sequencing (Fig. 2). The protocol below details how to add Illumina-compatible sequencing adapters, which can be used on a NextSeq500.

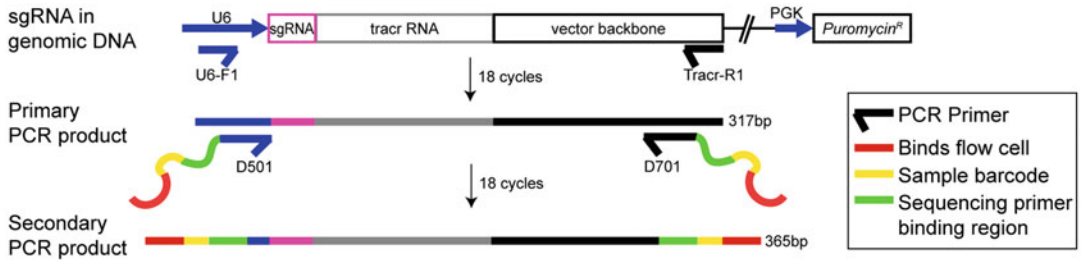


Fig. 2 Library preparation requires a two-step PCR amplification. The sgRNA sequences are represented within an integrated loci of the genome. Using PCR primers U6-F1 and Tracr-R1, the sgRNA sequences can be amplified from genomic DNA (top). The expected PCR product is 317 bp. For each biological replicate, pooled primary PCRs are used as a template to attach sequencing adapters compatible with the Illumina NextSeq500. Variable length products are expected depending on the primer pair used. Sequence-specific regions are outlined in the legend

1. Isolate genomic DNA from each cell pellet using a QIAamp DNA Blood Maxi Kit with the following adjustments. Resuspend cell pellets in 1 mL of QIAGEN Protease from the kit, and add 10 mL of PBS in a 50 mL conical tube. Incubate at 70 °C for 30 min, vortexing frequently. Add 12 mL of buffer AL; mix and incubate at 70 °C for 30 min. Proceed with the instructions from the manufacturer. Elute DNA with 700 μL of PCR-grade H_2O (*see* **Notes 7** and **8**).
2. Determine genomic DNA concentration and total yield with a NanoDrop spectrophotometer.
3. Set up PCR reactions with 10 μg of genomic DNA in 100 μL final volume using 1.5 μL of ExTaq, 10 μL of 10× buffer, 8 μL of dNTPs, and 5 μL each of 5 μM primary PCR primers (Table 1).
4. Perform as many PCR replicates per sample as needed to maintain desired library coverage using the estimate of 6.6 μg DNA/1 × 10^6 cells. For the Brunello genome-wide CRISPR library maintained at 500× coverage, this results in 27 PCR replicates per sample.
5. The following thermocycler conditions should be used: 95 °C for 5 min, 18× (94 °C for 30 s, 65 °C for 30 s, 72 °C for 20 s), 72 °C for 5 min.
6. Pool all primary PCR products for each sample.
7. Set up secondary PCR reactions using 100 μL final volume, 1.5 μL of ExTaq, 10 μL of 10× buffer, 8 μL of dNTPs, 5 μL of the pooled primary PCR product, and 5 μL of 5 μM secondary PCR primers (Table 2). *See* **Notes 9** and **10**.
8. The following thermocycler conditions should be used: 95 °C for 5 min, 18× (94 °C for 30 s, 65 °C for 30 s, 72 °C for 20 s), 72 °C for 5 min.

9. Load 20 μL of each sample on a 2% size-select E-Gel, and isolate the secondary PCR product at ~370 bp. Remove H_2O from the collection wells before the sample enters, and replace with fresh H_2O to reduce contamination from smaller PCR products and primers.
10. Measure the concentration of each sample using a high-sensitivity method such as Qubit HS (*see* **Note 11**).
11. Proceed with NextGen sequencing on an Illumina NextSeq500 sequencer using a single-read flow cell. The primers above will need to be modified for other sequencing platforms.
12. Analysis of CRISPR-screening sequencing results may be achieved by a number of open-source software packages (2) [13, 24–27].

4 Notes

1. Various cell lines, sgRNAs, and surface proteins can display different kinetics with respect to decreased surface expression. The GFP^- fraction of untransduced cells serves as a positive control for surface marker staining, and the non-dox-treated, sgRNA-transduced cell populations serve as a control for Cas9 inducibility.
2. Pools of Cas9-transduced cells may show as little as 5–10% of the GFP^+ fraction of cells that have lost surface expression of the targeted gene. This is sufficient validation of Cas9 function to move forward with single-cell cloning.
3. Virus production will need to be scaled up to achieve sufficient screen coverage. T225 flasks can be used for virus production and have approximately three times the surface area of a 10 cm tissue culture dish. We find using 48 μg of plasmid DNA in a 4:3:1 ratio of library, packing, and envelope constructs to be optimal, but these ratios and total DNA should be optimized for a given system.
4. The optimal puromycin dose should be predetermined by generating a dose response killing curve on each cell line of interest. The optimal puromycin dose will achieve near 100% killing within 72 h of treatment on untransduced cells.
5. Library coverage is equal to the library size multiplied by the desired independent integration events for a single sgRNA. For instance, 500× coverage for a library with 80,000 sgRNAs requires $40e^6$-transduced cells. Thus, each biological replicate will require $120e^6$ cells to begin the screen assuming a 33% transduction efficiency.
6. Transduction rates of the screen cultures can be estimated by removing transduced cells before and after puromycin

treatment to a 6-well plate and enumerated 3 days later as in Subheading 3.3.

7. Centrifugation at 3000 × *g* for twice the recommended time can be substituted if a centrifuge cannot reach the recommended 4500 × *g*.
8. Elutions in H_2O prevent salt contamination of subsequent PCR reactions.
9. Unique pairs of secondary PCR primers should be used to amplify each sample.
10. Primers in Table 2 should be ordered as "Ultramers" from IDT due to their length. D501–D508 differ in length by 8 bp to improve sequence diversity.
11. Additional quantification methods, such as Agilent Bioanalyzer or KAPA Library Quantification Kits, can also be used.

Acknowledgments

We thank all current and former colleagues who helped establish these protocols. This research was supported by the Intramural Research Program of the NIH, CCR, NCI. D.E.W. is a Damon Runyon Fellow (DRG-2208-14). S.R. is a H2020 Marie Sklodowska-Curie global fellow (#661066).

References

1. Hieter P, Boguski M (1997) Functional genomics: it's all how you read it. Science 278 (5338):601–602
2. Phelan JD, Young RM, Webster DE, Roulland S, Wright GW, Kasbekar M, Shaffer AL 3rd, Ceribelli M, Wang JQ, Schmitz R, Nakagawa M, Bachy E, Huang DW, Ji Y, Chen L, Yang Y, Zhao H, Yu X, Xu W, Palisoc MM, Valadez RR, Davies-Hill T, Wilson WH, Chan WC, Jaffe ES, Gascoyne RD, Campo E, Rosenwald A, Ott G, Delabie J, Rimsza LM, Rodriguez FJ, Estephan F, Holdhoff M, Kruhlak MJ, Hewitt SM, Thomas CJ, Pittaluga S, Oellerich T, Staudt LM (2018) A multiprotein supercomplex controlling oncogenic signalling in lymphoma. Nature 560(7718):387–391. https://doi.org/10.1038/s41586-018-0290-0
3. Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, Romero DA, Horvath P (2007) CRISPR provides acquired resistance against viruses in prokaryotes. Science 315(5819):1709–1712. https://doi.org/10.1126/science.1138140
4. Marraffini LA, Sontheimer EJ (2008) CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. Science 322 (5909):1843–1845. https://doi.org/10.1126/science.1165771
5. Deltcheva E, Chylinski K, Sharma CM, Gonzales K, Chao Y, Pirzada ZA, Eckert MR, Vogel J, Charpentier E (2011) CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. Nature 471 (7340):602–607. https://doi.org/10.1038/nature09886
6. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science 337 (6096):816–821. https://doi.org/10.1126/science.1225829
7. Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA, Zhang F (2013) Multiplex genome engineering using CRISPR/Cas systems. Science 339(6121):819–823. https://doi.org/10.1126/science.1231143

8. Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, Norville JE, Church GM (2013) RNA-guided human genome engineering via Cas9. Science 339(6121):823–826. https://doi.org/10.1126/science.1232033
9. Wang T, Wei JJ, Sabatini DM, Lander ES (2014) Genetic screens in human cells using the CRISPR-Cas9 system. Science 343 (6166):80–84. https://doi.org/10.1126/science.1246981
10. Shalem O, Sanjana NE, Hartenian E, Shi X, Scott DA, Mikkelsen TS, Heckl D, Ebert BL, Root DE, Doench JG, Zhang F (2014) Genome-scale CRISPR-Cas9 knockout screening in human cells. Science 343(6166):84–87. https://doi.org/10.1126/science.1247005
11. Hart T, Chandrashekhar M, Aregger M, Steinhart Z, Brown KR, MacLeod G, Mis M, Zimmermann M, Fradet-Turcotte A, Sun S, Mero P, Dirks P, Sidhu S, Roth FP, Rissland OS, Durocher D, Angers S, Moffat J (2015) High-resolution CRISPR screens reveal fitness genes and genotype-specific cancer liabilities. Cell 163(6):1515–1526. https://doi.org/10.1016/j.cell.2015.11.015
12. Wang T, Birsoy K, Hughes NW, Krupczak KM, Post Y, Wei JJ, Lander ES, Sabatini DM (2015) Identification and characterization of essential genes in the human genome. Science 350 (6264):1096–1101. https://doi.org/10.1126/science.aac7041
13. Hart T, Moffat J (2016) BAGEL: a computational framework for identifying essential genes from pooled library screens. BMC Bioinformatics 17:164. https://doi.org/10.1186/s12859-016-1015-8
14. Gilbert LA, Horlbeck MA, Adamson B, Villalta JE, Chen Y, Whitehead EH, Guimaraes C, Panning B, Ploegh HL, Bassik MC, Qi LS, Kampmann M, Weissman JS (2014) Genome-scale CRISPR-mediated control of gene repression and activation. Cell 159(3):647–661. https://doi.org/10.1016/j.cell.2014.09.029
15. Hart T, Brown KR, Sircoulomb F, Rottapel R, Moffat J (2014) Measuring error rates in genomic perturbation screens: gold standards for human functional genomics. Mol Syst Biol 10:733. https://doi.org/10.15252/msb.20145216
16. Tanenbaum ME, Gilbert LA, Qi LS, Weissman JS, Vale RD (2014) A protein-tagging system for signal amplification in gene expression and fluorescence imaging. Cell 159(3):635–646. https://doi.org/10.1016/j.cell.2014.09.039
17. Konermann S, Brigham MD, Trevino AE, Joung J, Abudayyeh OO, Barcena C, Hsu PD, Habib N, Gootenberg JS, Nishimasu H, Nureki O, Zhang F (2015) Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. Nature 517 (7536):583–588. https://doi.org/10.1038/nature14136
18. Gilbert LA, Larson MH, Morsut L, Liu Z, Brar GA, Torres SE, Stern-Ginossar N, Brandman O, Whitehead EH, Doudna JA, Lim WA, Weissman JS, Qi LS (2013) CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. Cell 154 (2):442–451. https://doi.org/10.1016/j.cell.2013.06.044
19. Liu XS, Wu H, Ji X, Stelzer Y, Wu X, Czauderna S, Shu J, Dadon D, Young RA, Jaenisch R (2016) Editing DNA methylation in the mammalian genome. Cell 167 (1):233–247. e217. https://doi.org/10.1016/j.cell.2016.08.056
20. Hilton IB, D'Ippolito AM, Vockley CM, Thakore PI, Crawford GE, Reddy TE, Gersbach CA (2015) Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. Nat Biotechnol 33(5):510–517. https://doi.org/10.1038/nbt.3199
21. Komor AC, Kim YB, Packer MS, Zuris JA, Liu DR (2016) Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. Nature 533 (7603):420–424. https://doi.org/10.1038/nature17946
22. Wilson WH, Young RM, Schmitz R, Yang Y, Pittaluga S, Wright G, Lih CJ, Williams PM, Shaffer AL, Gerecitano J, de Vos S, Goy A, Kenkre VP, Barr PM, Blum KA, Shustov A, Advani R, Fowler NH, Vose JM, Elstrom RL, Habermann TM, Barrientos JC, McGreivy J, Fardis M, Chang BY, Clow F, Munneke B, Moussa D, Beaupre DM, Staudt LM (2015) Targeting B cell receptor signaling with ibrutinib in diffuse large B cell lymphoma. Nat Med 21(8):922–926. https://doi.org/10.1038/nm.3884
23. Sanjana NE, Shalem O, Zhang F (2014) Improved vectors and genome-wide libraries for CRISPR screening. Nat Methods 11 (8):783–784. https://doi.org/10.1038/nmeth.3047
24. Li W, Xu H, Xiao T, Cong L, Love MI, Zhang F, Irizarry RA, Liu JS, Brown M, Liu XS (2014) MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. Genome Biol 15(12):554. https://doi.org/10.1186/s13059-014-0554-4
25. Li W, Koster J, Xu H, Chen CH, Xiao T, Liu JS, Brown M, Liu XS (2015) Quality control, modeling, and visualization of CRISPR screens with MAGeCK-VISPR. Genome Biol 16:281. https://doi.org/10.1186/s13059-015-0843-6

26. Winter J, Breinig M, Heigwer F, Brugemann D, Leible S, Pelz O, Zhan T, Boutros M (2016) caRpools: an R package for exploratory data analysis and documentation of pooled CRISPR/Cas9 screens. Bioinformatics 32(4):632–634. https://doi.org/10.1093/bioinformatics/btv617

27. Jan Winter, Marc Schwering, Oliver Pelz, Benedikt Rauscher, Tianzuo Zhan, Florian Heigwer, Michael Boutros (2017) CRISPRAnalyzeR: interactive analysis, annotation and documentation of pooled CRISPR screens. bioRxiv. https://doi.org/10.1101/109967

Chapter 17

Lymphoma and Leukemia Cell Vulnerabilities and Resistance Identified by Compound Library Screens

Katarzyna Tomska, Sebastian Scheinost, and Thorsten Zenz

Abstract

Response to anticancer agents is often restricted to subsets of patients. The recognition of factors underlying this heterogeneity and the identification of biomarkers associated with response to drugs would greatly improve the efficacy of drug treatment. Platforms that can comprehensively map drug response in high-throughput ex vivo provide a unique tool to identify associated biomarkers and provide hypotheses for mechanisms underlying variable response. Such screens can be performed on cell lines and short-term cultures of primary cells to take advantage of the respective models' strength, which include, e.g., the ability to silence genes in cell lines and the "indefinite" supply of primary cells where clonal selection can be avoided. Cohorts of such samples represent the natural diversity of cancers, including rarer mutations and combinatorial patterns of mutations.

We here summarize a simple and scalable method for the measurement of viability after drug exposure based on ATP measurements as a surrogate for viability, which we use to measure and understand drug response in cell lines and primary cells.

Key words Leukemia, Lymphoma, Drug response profiling, Assay plates, Drug library, Genomics of drug response

1 Introduction

Response to anticancer agents is often restricted to subsets of patients. To identify factors underlying this heterogeneity, there is a need for assays that can map drug responses, identify associated biomarkers, and provide hypotheses for mechanisms underlying variable response.

Determinants of drug response have been profiled in immortalized cancer cell lines (http://www.cancerRxgene.org [1]; http://www.broadinstitute.org/ccle [2]; http://www.broadinstitute.org/ctrp [3]), and recent technology improvements allow increased throughput [4]. However, suitable cell lines are not available for

Katarzyna Tomska and Sebastian Scheinost are shared first authors.

Ralf Küppers (ed.), *Lymphoma: Methods and Protocols*, Methods in Molecular Biology, vol. 1956,
https://doi.org/10.1007/978-1-4939-9151-8_17,

all cancers, including chronic lymphocytic leukemia (CLL), or may not represent the full disease spectrum. Short-term cultures of primary cells have advantages over immortalized cells, as clonal selection is less likely to occur. Cohorts of patient-derived samples represent the natural diversity of cancers. Responses of primary cells to panels of inhibitors ex vivo have recently been used to derive individualized therapeutic options for the donating patients [5–7]. Molecular characterization of individual samples yielded novel genetic markers and led to drug repurposing opportunities [8–10].

We and others recently reported a large-scale study of drug sensitivities of primary leukemia and lymphoma, which systematically linked drug response to genotypes and molecular processes involved in disease pathogenesis [11–13].

While it is tempting to propagate more complex ways of measuring response to drugs, there is a trade-off between assay complexity and cost, reproducibility, and current uncertainty regarding the "perfect" method, which may be expected to be drug specific.

We continue to use ATP-based measurements of drug response, which has been a main method for the large cell line drug response encyclopedias (http://www.cancerRxgene.org [1]; http://www.broadinstitute.org/ccle [2]; http://www.broadinstitute.org/ctrp [3]). With the advent of CRISPR/Cas9 modification of cell lines, the use of isogenic cell lines will further leverage the power of drug-based screening.

While the range of technical optimizations including 3D cultures, spheroid cultures for solid tumors, and the use of xenografts is long, a major milestone is the demonstration of proof of concept to identify known biomarkers, which may be considered as prerequisite for any platform.

2 Materials

2.1 Cell Culture

1. Cell lines or viably cryopreserved primary cells (primary mononuclear cells (PMCs), protocol for tumor cells in liquid culture).
2. Supplemented RPMI 1640 medium (10–20% fetal bovine serum (FBS) or human serum, 1% glutamine, 1% penicillin-streptomycin) or other medium compatible with the cell lines used.
3. Supplemented RPMI 1640 medium (10% heat-inactivated AB-type human serum, 1% glutamine, 1% penicillin-streptomycin) for PMCs.

2.2 Inhibitor Library

1. Compounds of choice dissolved in dimethyl sulfoxide (DMSO) at 10 mM.

2. DMSO and phosphate-buffered saline (PBS) for controls.
3. 96-well polypropylene plates.
4. 12-channel multifunction pipette.

2.3 Single Agent Screen

1. Inhibitor library plates.
2. Cell aliquots with the desired cell density (7.5×10^4–1.75×10^5 cells per mL for cell lines and 6.25×10^5–1.0×10^6 cells per mL for PMCs).
3. Culture medium.
4. 12-channel multifunction pipette or 96-channel multifunction pipette (Viaflo, Integra Biosciences).
5. 384-well non-binding plates.
6. Sterile plate lids.
7. Centrifuge.
8. V-shaped reservoirs or 96-well reservoirs.
9. Cell strainer (40 μm).
10. Counting device: microscope or automated cell counter (Scepter Cell Counter, MerckMillipore).

2.4 Combination Screen

1. Inhibitor library plates.
2. Compounds of choice to combine with the inhibitor library.
3. Cell aliquots with the desired cell density (1.0×10^5–2.33×10^5 cells per mL for cell lines and 8.33×10^5–1.0×10^6 cells per mL for PMCs).
4. DMSO.
5. Culture medium.
6. 12-channel multifunction pipette.
7. Sterile 384-well non-binding plates.
8. Sterile plate lids.
9. V-shaped reservoirs.

2.5 Luciferase Assay

1. Screen plates incubated for 48 h.
2. Luciferase assay reagent, i.e., CellTiter-Glo (Promega) or ATPlite.
3. 12-channel multifunction pipette or rapid reagent dispenser (Viafill, Integra Biosciences).
4. Multiplate reader (e.g., Infinite M200 Tecan Group Ltd. or EnVision Plate Reader, PerkinElmer).

3 Methods

3.1 Cell Culture

1. Determine the optimal cell density for each cell line (*see* **Note 1**), i.e., 3000 cells/well for fast growing cell lines (corresponds to 3000 cells per 25–50 μL) and 7000 cells/well (corresponds to 7000 cells per 25–50 μL) for slow growing lines, as well as 20,000–40,000 cells/well (corresponds to 20,000–40,000 cells per 25–50 μL) for PMCs.
2. Culture cell lines in proper culture media at 37 °C and 5% CO_2 with humidification for 1 week before conducting the drug screen.
3. Thaw PMCs directly at the drug screen day; dissolve them in preheated (37 °C) medium (cell line medium with 10% FBS) in 50 mL plastic centrifugation tubes; centrifuge them at 400 × *g* for 5 min; remove supernatant; resuspend the cell pellet with preheated, fresh medium (now PMC medium with 10% human serum); and equilibrate tubes on a roller mixer at room temperature for 3 h to remove remaining DMSO (*see* **Note 2**).

3.2 Inhibitor Library

1. Define compounds of interest, range of compound concentrations, sample throughput per day, and number of drug screen days ("experiments"; *see* **Notes 3** and **4**). Depending on the planned throughput and number of experiments, decide on the drug volume per well on 96-well plates ("master plates", ideally the master plates are prepared once for the entire screen; *see* **Note 5**).
2. Dissolve all inhibitors in DMSO at starting concentration, e.g., 10 mM. The plates should also include DMSO and PBS controls as a reference to determine percentage of living cells and an internal control to assess DMSO toxicity (*see* **Note 6**).
3. Conduct a serial dilution of the compound stock (e.g., 10 mM) in DMSO with 1:3–1:5 steps (depending on the required concentration range) with a 12-channel multifunction pipette on the master plate. For 4 master plates, plate, e.g., 32 compounds on 96-well plates with 8 compounds per plate and 10 concentrations each (Fig. 1). Alternatively, choose 64 compounds and use only 5 concentrations per drug (16 compounds per plate) or any other format needed for the respective experiment. If multiple concentrations are used, we avoid replicates. It is important to adapt the concentration range to the toxicity of the drug (toxicity as found in literature or seen in previous screens).
4. Transfer 3 μL onto daughter plates using a 12-channel multifunction pipette (*see* **Note 7**). If there are four different master plates, there are also four different copies (daughter plates).

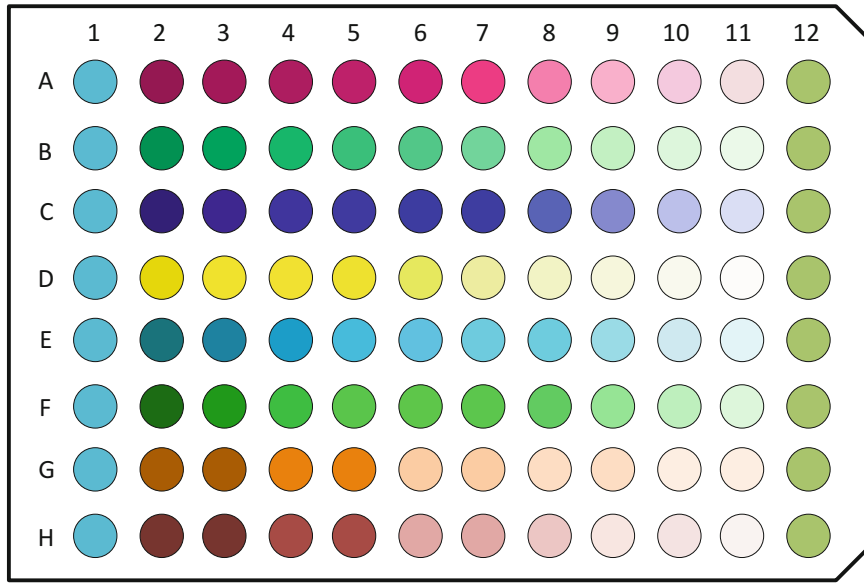


Fig. 1 Distribution of drug library on a 96-well plate for eight drugs in ten concentrations; columns 1 and 12 contain DMSO and PBS controls (format may be modified)

The number of daughter plates per master plate should be sufficient for the estimated number of experiments.

5. Seal plates, e.g., with aluminum foil and store them at −20 °C until needed for drug screen.

3.3 Single Agent Screen

1. Thaw a set of four 96-well daughter plates containing 32 or 64 inhibitors before every single experiment. After thawing, centrifuge the plates manually for 5–10 s to ensure that the drugs are on the bottom of the well.
2. Add 197 µL medium to each well using dispensing mode with multichannel device resulting in a dilution of 3:200.
3. Use mixing mode on the multichannel device to mix the contents of each well at least five to ten times with around half of the well volume (here 100 µL). Bubble formation in the wells should be avoided by submerging sufficiently with the pipette tips into the diluted compound.
4. Then use the dispensing mode, and pipette 5–10 µL of the contents of one row of a 96-well plate onto 384-well plates (depending on your final volume on the 384-well plates). Alternatively, use a 96-channel multifunction pipette to transfer the drugs and controls of the whole 96-well plate onto the 384-well plates at once. To increase accuracy, the first dispense should be discarded. For both pipettes, compounds from each well of the 96-well plate are transferred on every second well of the 384-well plate.

5. Change tips after each row to prevent contamination with a drug from a row above. If using a 96-channel multifunction pipette, change tips after every 96-well plate.
6. After finishing the drug transfer from all four daughter plates, centrifuge the plates manually for 5–10 s to ensure that the drugs are on the bottom of the well.
7. Transfer cell line suspension from culture flasks into 50 mL tubes and centrifuge at 400 × *g* for 5 min. PMCs do not need to be centrifuged again, but taken from the roller mixer and strained through a 40 μm filter into new tubes.
8. After removing supernatant, resuspend cell line pellets in 2–5 mL fresh medium, and count the cells under light microscope. Count PMCs using a small aliquot with 1:100 dilution in PBS with an automated cell counter. Prepare cell suspension of 7.5×10^4 cells/mL or 1.75×10^5 cells/mL for cell lines or 6.25×10^5–1.0×10^6 cells/mL for PMCs.
9. Seed out mixed cells at 20–40 μL per well onto the treated 384-well plate (one sample per 384-well plate; *see* **Notes 7** and **8**).
10. Repeat the same procedure for each cell line or primary sample.
11. After completing cell seeding, centrifuge the plates manually for 5–10 s to ensure that the cells are on the bottom of the well.
12. Incubate for 48 h (or alternative duration as 24 h, 72 h) at 37 °C, at 5% CO_2 with humidification (*see* **Note 9**). Many groups use 72 h time point, but the general trade-off is noise and effect size. We continue to use 48 h.

3.4 Combination Screen

1. Thaw a set of four 96-well plates containing 32 inhibitors before every single experiment.
2. Add 197 μL medium to each well using dispensing mode with multichannel device resulting in a dilution of 3:200.
3. Use mixing mode on the multichannel device to mix the contents of each well at least ten times.
4. Then use the dispensing mode, and pipette 10 μL of the contents of one row of a 96-well plate onto 384-well test plates. Use a total of four 384-well plates per cell line.
5. Dilute three compounds chosen as combination partners to a desired concentration (*see* **Note 10**) as well as DMSO and PBS for controls. First, thaw and vortex 10 mM stocks. For a drug end concentration of 200 nM, prepare a 1:500 dilution in two steps: mix 3 μL stock 297 μL medium in a 1.5 mL tube; then mix 200 μL of the resulting solution with 800 μL medium in a second 1.5 mL tube. Pipette 750 μL of the 1:500 dilution onto a reservoir, and fill it with 14.25 mL medium. Mix using a multichannel pipette.

6. Pipette 10 μL DMSO onto the first of four pretreated plates per cell line (to determine the effect of single drug); add 10 μL combination drugs at a constant concentration onto the remaining plates pretreated with the drug library leaving control wells empty. Pipette DMSO and PBS control wells separately also using the multichannel device.
7. Count cells under light microscope, and prepare cell suspension of 1.0×10^5 cells/mL or 2.33×10^5 cells/mL (8.33×10^5–1.0×10^6 for PMCs).
8. Seed out cells at 30 μL per well. Pipette each cell line onto four 384-well plates, DMSO control plate, and the combination plates with three different drugs.
9. Seal the plates using breathable sealing foils to avoid evaporation, and incubate for 48 h at 37 °C, at 5% CO_2 with humidification.

3.5 Luciferase Assay

1. Measure cell viability 48 h (or 72 h) after cell seeding by adding 7–12 μL CellTiter-Glo reagent per well. As an alternative to a 12-channel multifunction pipette, a rapid reagent dispenser can be used. Here, the reagent is automatically dispensed on 20 or more plates. Different speed settings allow synchronizing dispensing velocity to readout velocity of the multiplate reader, so the incubation time of every 384-well plate is identical.
2. Measure luminescence signal after 20 min incubation time using a multiplate reader (Tecan Microplate Reader with an integration time of 200 ms or EnVision Plate Reader with a measurement time of 100 ms).
3. Calculate relative viability in percent by normalizing the data to the mean of DMSO controls (for further details *see* **Notes 11–14**).

4 Notes

1. *Initial experiments:* Many factors including cell line selection, cell culture, cross-contamination, optimal cell density, plating cells, storage of compounds, as well as their concentration range, type of cell viability assay used, or statistical model implemented may influence reproducibility of drug sensitivity data [14]. Consider designing initial experiments to determine optimal cell density for the cell lines used. It is known that passage number of cell lines may contribute to the reproducibility of drug sensitivity data [14], therefore screen cell lines after 1 week in culture, which corresponds to three or four passages. For PMCs, availability of material is a crucial factor. Here one should aim for the lowest possible cell density

providing reproducible results. This may be different between blood cancer entities and should be assessed before starting a large-scale drug screen.

2. *Sample viability:* While using primary material, the fraction of viable cells within the sample might be an issue. Consider running a quality control of the samples using FACS and Annexin V staining. Samples with low viability may provide unreliable results in the screen. We perform a day 0 measurement of ATP levels to identify samples with severe drop in ATP after incubation. Overall ATP levels are very stable over 48 h in samples with good viability.
3. *Drug library design:* The choice of the drug library is extremely important. It should cover a range of drugs targeting various oncogenic pathways in order to yield useful results. We also recommend choosing two drugs with a similar target and mechanism of action for some pathways for internal controls.
4. *Drug concentrations:* Consider the highest drug concentration after all dilution steps (3:200 on the library plate and 1:5 on the 384-well plate). If the drug is very toxic, an additional dilution step using DMSO as solvent might be needed before plating the library.
5. *Consideration of total volume per well:* The end volume on the library plate determines the amount of samples that can be tested during a single experiment, i.e., with 200 μL one would be able to treat a maximum of 18 384-well plates if 10 μL of drug is needed per plate. Different volumes may be chosen; however keep in mind the maximum volume per well on the 96-well library plate and the end DMSO concentration.
6. *DMSO toxicity:* Since the drugs are dissolved in DMSO, calculate the end concentration of DMSO on the 384-well plate. Make sure it does not exceed 0.3%, or consider conducting an additional experiment to determine DMSO toxicity within the cell lines and PMCs used.
7. *Choice of the multifunction device:* Using a 12-channel multifunction pipette is quite demanding and might be time-consuming if a larger number of cell lines are tested. The implementation of a semiautomatic pipette with a 96-channel head provides an alternative, and liquid handling stations/robotics are emerging to increase throughput. Moreover, the establishment of automated steps helps to minimize noise in the experiment. As an example, for cell seeding, the 96-channel multifunction pipette has to (a) move deep enough into the wells to safely dispense the cells and (b) stop with sufficient distance to the drug on the bottom to avoid cross-contamination. Moreover, at low volumes, suspension droplets stick to the tip's peak and do not fall into the well. This can be

resolved by moving the pipette 1–2 mm to the left or right to enable contact of the tip with the wall of the well.

8. *Seeding out cells:* With the 12-channel multifunction pipette, pour cell suspension in a V-shaped reservoir, and mix it manually using a 10 mL pipette and then at least five times with 200 μL volume using a multichannel device. Cell dispensing should start in PBS control and move from low to high drug concentrations. To avoid cross-contamination, tips should not touch deeper than the upper third of the well. Tips should be changed after half of the plate, and cells in the reservoir should be mixed again before continuing with seeding. Last change of tips and mix of cells should be performed before adding cells to the DMSO control. With the 96-channel multifunction pipette, pour cells into a 96-well reservoir, and mix them at least four times with 50 μL volume. Tips do not need to be changed within one cell line or primary sample (see **Note 7**).
9. *Sealing of 384-well plates with breathable foil:* Initially, we have sealed the 384-well plates using a breathable foil to avoid evaporation during 48 h of incubation. In high-throughput screens, this step is time-consuming and expensive. We have tried out drug screening without the breathable foil, but only with a lid on the plate during 48 h incubation, and results are reproducible. However, in the wells on the edge of the plate, decreased viability is observed ("edge effect"). This can be corrected in the analysis and may need to be accepted. Make sure crucial drugs are not arranged at the edge. Alternatively, the wells on the edge could be filled with water or PBS only.
10. *Combination drug concentration:* For the combination drug screening, the drug library is tested in combination with a chosen drug at a set concentration. Hence, a pre-screen determining the optimal concentration of the combination drug is required and critical. The concentration of the combination drug should not be too toxic, in order not to mask any cooperative interactions. We prefer to use low nM concentrations for kinase inhibitors, and toxicity should not exceed 80% viability. For higher resolution of drug concentration and their interaction, a 10 × 10 matrix experiment can be designed, where two drugs are used in ten concentrations.
11. *Quality assessment and control:* We use multiple steps of data quality assessment. First, we assess the sensitivity of the platform by asking whether it can detect known gene-drug associations and expected correlations between similar drugs. To assess the robustness of the platform and safeguard against "batch effects" that have the potential to confound high-throughput experiments [15], we repeat the analyses taking the time point of experiments as potential batch confounders

into account. In addition, we assess the reproducibility between different batch time points and compare the same samples from cell lines/patients in independent experiments. The data should be highly reproducible for all samples (Pearson correlation coefficients: >0.8).

12. *Drug screen data analysis:* To quantify the response of a primary sample to a drug at a given concentration in the drug sensitivity assays, we use viability relative to control; this value is the CellTiter-Glo luminescence readout of the respective well divided by the mean of luminescence readouts of the DMSO control wells on the same plate. We use the R/Bioconductor package cellHTS2 version 2.26.0 [16] for processing the raw files obtained from the plate scanner.
13. *Batch effects:* Screens are often performed in groups of batches over a time period; to control for confounding by different batch groups, repeat the drug-feature association tests using batch group as a blocking factor and a two-way ANOVA test. Individual drugs may show discrepant *p*-values, e.g., as drugs lose its activity during storage.
14. *Use of effect surrogates as AUC, IC50:* Much of the drug response literature relies on IC50, which is often used to summarize drug effects. We usually use effects at all individual concentrations and averaged drug toxicity (AUC), as IC50 may not be useful for kinase drugs with limited viability effects.

References

1. Garnett MJ, Edelman EJ, Heidorn SJ, Greenman CD, Dastur A, Lau KW, Greninger P, Thompson IR, Luo X, Soares J, Liu Q, Iorio F, Surdez D, Chen L, Milano RJ, Bignell GR, Tam AT, Davies H, Stevenson JA, Barthorpe S, Lutz SR, Kogera F, Lawrence K, McLaren-Douglas A, Mitropoulos X, Mironenko T, Thi H, Richardson L, Zhou W, Jewitt F, Zhang T, O'Brien P, Boisvert JL, Price S, Hur W, Yang W, Deng X, Butler A, Choi HG, Chang JW, Baselga J, Stamenkovic I, Engelman JA, Sharma SV, Delattre O, Saez-Rodriguez J, Gray NS, Settleman J, Futreal PA, Haber DA, Stratton MR, Ramaswamy S, McDermott U, Benes CH (2012) Systematic identification of genomic markers of drug sensitivity in cancer cells. Nature 483 (7391):570–575. https://doi.org/10.1038/nature11005
2. Barretina J, Caponigro G, Stransky N, Venkatesan K, Margolin AA, Kim S, Wilson CJ, Lehar J, Kryukov GV, Sonkin D, Reddy A, Liu M, Murray L, Berger MF, Monahan JE, Morais P, Meltzer J, Korejwa A, Jane-Valbuena J, Mapa FA, Thibault J, Bric-Furlong E, Raman P, Shipway A, Engels IH, Cheng J, Yu GK, Yu J, Aspesi P Jr, de Silva M, Jagtap K, Jones MD, Wang L, Hatton C, Palescandolo E, Gupta S, Mahan S, Sougnez C, Onofrio RC, Liefeld T, MacConaill L, Winckler W, Reich M, Li N, Mesirov JP, Gabriel SB, Getz G, Ardlie K, Chan V, Myer VE, Weber BL, Porter J, Warmuth M, Finan P, Harris JL, Meyerson M, Golub TR, Morrissey MP, Sellers WR, Schlegel R, Garraway LA (2012) The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. Nature 483(7391):603–607. https://doi.org/10.1038/nature11003
3. Basu A, Bodycombe NE, Cheah JH, Price EV, Liu K, Schaefer GI, Ebright RY, Stewart ML, Ito D, Wang S, Bracha AL, Liefeld T, Wawer M, Gilbert JC, Wilson AJ, Stransky N, Kryukov GV, Dancik V, Barretina J, Garraway LA, Hon CS, Munoz B, Bittker JA, Stockwell BR, Khabele D, Stern AM, Clemons PA, Shamji AF, Schreiber SL (2013) An interactive

resource to identify cancer genetic and lineage dependencies targeted by small molecules. Cell 154(5):1151–1161. https://doi.org/10.1016/j.cell.2013.08.003

4. Yu C, Mannan AM, Yvone GM, Ross KN, Zhang YL, Marton MA, Taylor BR, Crenshaw A, Gould JZ, Tamayo P, Weir BA, Tsherniak A, Wong B, Garraway LA, Shamji AF, Palmer MA, Foley MA, Winckler W, Schreiber SL, Kung AL, Golub TR (2016) High-throughput identification of genotype-specific cancer vulnerabilities in mixtures of barcoded tumor cell lines. Nat Biotechnol 34 (4):419–423. https://doi.org/10.1038/nbt.3460
5. Tyner JW, Yang WF, Bankhead A, Fan G, Fletcher LB, Bryant J, Glover JM, Chang BH, Spurgeon SE, Fleming WH, Kovacsovics T, Gotlib JR, Oh ST, Deininger MW, Zwaan CM, Den Boer ML, van den Heuvel-Eibrink MM, O'Hare T, Druker BJ, Loriaux MM (2013) Kinase pathway dependence in primary human leukemias determined by rapid inhibitor screening. Cancer Res 73:285–296. https://doi.org/10.1158/0008-5472.CAN-12-1906
6. Pemovska T, Kontro M, Yadav B, Edgren H, Eldfors S, Szwajda A, Almusa H, Bespalov MM, Ellonen P, Elonen E, Gjertsen BT, Karjalainen R, Kulesskiy E, Lagström S, Lehto A, Lepistö M, Lundán T, Majumder MM, Marti JML, Mattila P, Murumägi A, Mustjoki S, Palva A, Parsons A, Pirttinen T, Rämet ME, Suvela M, Turunen L, Västrik I, Wolf M, Knowles J, Aittokallio T, Heckman CA, Porkka K, Kallioniemi O, Wennerberg K (2013) Individualized systems medicine strategy to tailor treatments for patients with chemorefractory acute myeloid leukemia. Cancer Discov 3:1416–1429. https://doi.org/10.1158/2159-8290.CD-13-0350
7. Crystal AS, Shaw AT, Sequist LV, Friboulet L, Niederst MJ, Lockerman EL, Frias RL, Gainor JF, Amzallag A, Greninger P, Lee D, Kalsy A, Gomez-Caraballo M, Elamine L, Howe E, Hur W, Lifshits E, Robinson HE, Katayama R, Faber AC, Awad MM, Ramaswamy S, Mino-Kenudson M, Iafrate AJ, Benes CH, Engelman JA (2014) Patient-derived models of acquired resistance can identify effective drug combinations for cancer. Science 346(6216):1480–1486. https://doi.org/10.1126/science.1254721
8. Maxson JE, Gotlib J, Pollyea DA, Fleischman AG, Agarwal A, Eide CA, Bottomly D, Wilmot B, McWeeney SK, Tognon CE, Pond JB, Collins RH, Goueli B, Oh ST, Deininger MW, Chang BH, Loriaux MM, Druker BJ, Tyner JW (2013) Oncogenic CSF3R mutations in chronic neutrophilic leukemia and atypical CML. N Engl J Med 368:1781–1790. https://doi.org/10.1056/NEJMoa1214514
9. Pemovska T, Johnson E, Kontro M, Repasky GA, Chen J, Wells P, Cronin CN, McTigue M, Kallioniemi O, Porkka K, Murray BW, Wennerberg K (2015) Axitinib effectively inhibits BCR-ABL1(T315I) with a distinct binding conformation. Nature 519:102–105. https://doi.org/10.1038/nature14119
10. Maxson JE, Abel ML, Wang J, Deng X, Reckel S, Luty SB, Sun H, Gorenstein J, Hughes SB, Bottomly D, Wilmot B, Mcweeney SK, Radich J, Hantschel O, Middleton RE, Gray NS, Druker BJ, Tyner JW (2016) Identification and characterization of tyrosine kinase nonreceptor 2 mutations in leukemia through integration of kinase inhibitor screening and genomic analysis. Cancer Res 76:127–138. https://doi.org/10.1158/0008-5472.CAN-15-0817
11. Dietrich S, Oles M, Lu J, Sellner L, Anders S, Velten B, Wu B, Hullein J, da Silva Liberio M, Walther T, Wagner L, Rabe S, Ghidelli-Disse S, Bantscheff M, Oles AK, Slabicki M, Mock A, Oakes CC, Wang S, Oppermann S, Lukas M, Kim V, Sill M, Benner A, Jauch A, Sutton LA, Young E, Rosenquist R, Liu X, Jethwa A, Lee KS, Lewis J, Putzker K, Lutz C, Rossi D, Mokhir A, Oellerich T, Zirlik K, Herling M, Nguyen-Khac F, Plass C, Andersson E, Mustjoki S, von Kalle C, Ho AD, Hensel M, Durig J, Ringshausen I, Zapatka M, Huber W, Zenz T (2018) Drug-perturbation-based stratification of blood cancer. J Clin Invest 128 (1):427–445. https://doi.org/10.1172/JCI93801
12. Walter R, Pan KT, Doebele C, Comoglio F, Tomska K, Bohnenberger H, Young RM, Jacobs L, Keller U, Bonig H, Engelke M, Rosenwald A, Urlaub H, Staudt LM, Serve H, Zenz T, Oellerich T (2017) HSP90 promotes Burkitt lymphoma cell survival by maintaining tonic B-cell receptor signaling. Blood 129(5):598–608. https://doi.org/10.1182/blood-2016-06-721423
13. Tomska K, Kurilov R, Lee KS, Hullein J, Lukas M, Sellner L, Walther T, Wagner L, Oles M, Brors B, Huber W, Zenz T (2018) Drug-based perturbation screen uncovers synergistic drug combinations in Burkitt lymphoma. Sci Rep 8(1):12046. https://doi.org/10.1038/s41598-018-30509-3
14. Hatzis C, Bedard PL, Birkbak NJ, Beck AH, Aerts HJ, Stem DF, Shi L, Clarke R, Quackenbush J, Haibe-Kains B (2014)

Enhancing reproducibility in cancer drug screening: how do we move forward? Cancer Res 74(15):4016–4023. https://doi.org/10.1158/0008-5472.CAN-14-0725

15. Leek JT, Scharpf RB, Bravo HC, Simcha D, Langmead B, Johnson WE, Geman D, Baggerly K, Irizarry RA (2010) Tackling the widespread and critical impact of batch effects in high-throughput data. Nat Rev Genet 11 (10):733–739. https://doi.org/10.1038/nrg2825
16. Boutros M, Bras LP, Huber W (2006) Analysis of cell-based RNAi screens. Genome Biol 7(7): R66. https://doi.org/10.1186/gb-2006-7-7-R66

Chapter 18

Proximity Ligation Assay

Ryan M. Young

Abstract

Many methods exist to study oncogenic signaling in lymphoma cells, each with certain strengths and weaknesses. The proximity ligation assay (PLA) is a technique used to examine protein-protein interactions in situ. PLA offers many advantages over traditional biochemical approaches to signaling in that it can detect both transient interactions, which are hard to capture otherwise, as well as protein interactions within large supramolecular signaling platforms that are often intractable to traditional biochemical methodologies. In addition, PLA can be adapted for use in lymphoma biopsy samples to translate findings from the lab. Implementation of PLA can complement many types of lymphoma research, and the methods presented herein can be adapted to a number of experimental settings.

Key words Proximity ligation assay, Biochemistry, Lymphoma, Clinical research, Protein-protein interactions

1 Introduction

The study of oncogenic signaling pathways in lymphoma cells has yielded many insights into lymphoma biology and guided the development of novel precision medicine strategies, including B-cell receptor (BCR) pathway inhibitors [1]. Numerous experimental methods are available to interrogate signaling networks. Commonly, these include analysis of protein phosphorylation or enzyme activity in cells to examine changes in signaling following a perturbation, such as stimulation of a receptor or inhibition of a kinase. Likewise, protein-protein interactions can be studied using co-immunoprecipitation of endogenous protein or by overexpression of tagged proteins to observe binding to known interactors. These methods are typically used to reveal changes on a global scale within large populations of cells and are most useful in cell line models of lymphoma. Alternatively, immunohistochemistry studies can examine the broad distribution of protein in individual cells, and co-staining with additional antibodies may reveal overlap between two or more proteins. However, all of these methods

Ralf Küppers (ed.), *Lymphoma: Methods and Protocols*, Methods in Molecular Biology, vol. 1956,
https://doi.org/10.1007/978-1-4939-9151-8_18,

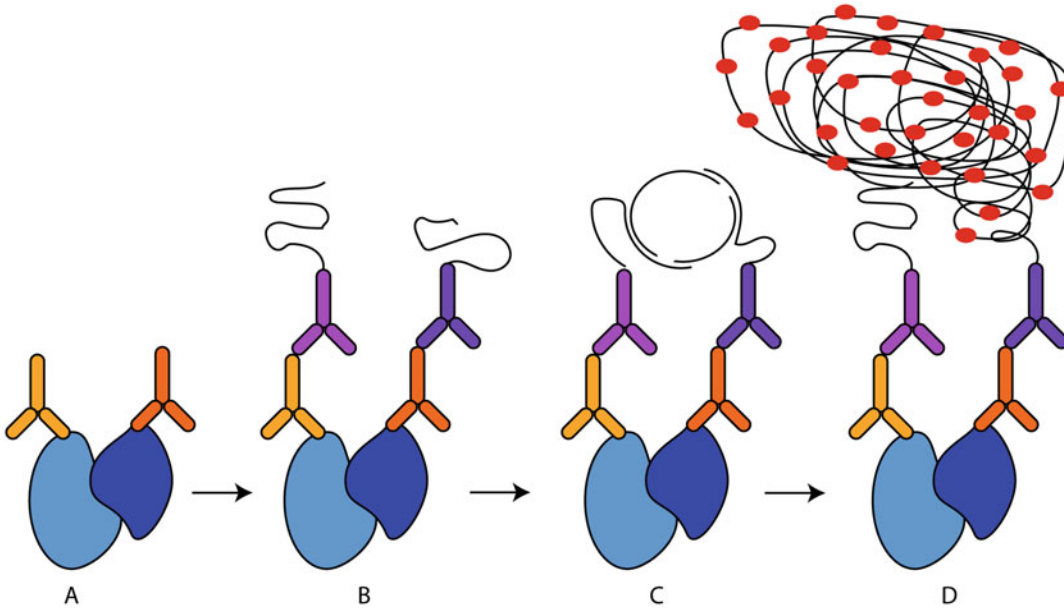


Fig. 1 Graphical workflow for PLA. (**a**) Two primary antibodies raised in separate species are bound to target proteins. (**b**) Species-specific secondary antibodies conjugated to complementary oligonucleotides are bound to primary antibodies. (**c**) The antibody-conjugated oligonucleotides are ligated to connector oligonucleotides to generate a circular single-stranded piece of DNA. (**d**) A rolling circle amplification is performed to amplify the DNA, which is then labeled with fluorescently tagged probes

have limitations when studying exceptionally transient signaling, spatially localized signaling, or signaling in stable supramolecular complexes that may only utilize a small fraction of the total protein in the cell [2].

The proximity ligation assay (PLA) is a method to visualize protein-protein interactions within the cell in situ [3, 4]. The basic workflow is depicted in Fig. 1. This method detects protein-protein interactions by co-staining with antibodies conjugated to complementary oligonucleotides that, when in close enough proximity, can be ligated together and their signal magnified by a rolling circle amplification [5]. Oligonucleotide probes conjugated to either fluorophores or horseradish peroxidase (HRP) are used to detect the amplified DNA and are visualized as distinct puncta (Figs. 1d and 2). This amplification step allows for the sensitive detection of interactions between discrete proteins and has the advantage of detecting relatively rare protein-protein interactions that are still critical to oncogenic signaling and are otherwise difficult to detect using techniques such as fluorescence resonance energy transfer (FRET). However, PLA may not be as quantitative as FRET [6]. PLA can also be used to study particularly large or stable protein complexes that are often insoluble and difficult to study using traditional biochemical techniques [7].

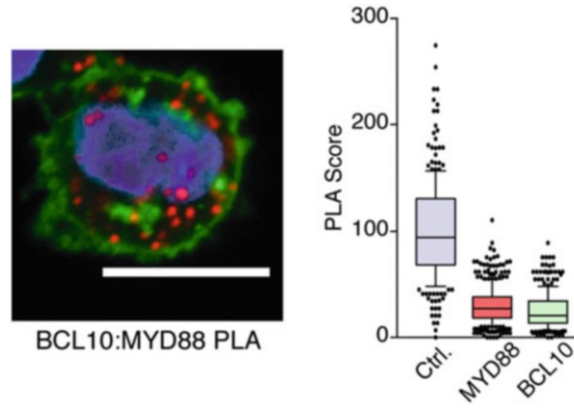


Fig. 2 Sample data in the ABC-DLBCL cell line HBL1 of PLA between BCL10 and MYD88 (left). Data quantitation is shown on the right. The number of PLA puncta per cell was normalized to 100 in cells expressing control shRNA. Knockdown of either MYD88 or BCL10 reduced the number of puncta per cell, demonstrating the specificity of this PLA pair. Red is BCL10:MYD88 PLA and blue is DAPI staining of the nucleus. Cellular membranes were stained in green with wheat germ agglutinin labeled with AlexaFluor 488

An additional advantage of PLA is that it can be applied to formalin-fixed paraffin-embedded (FFPE) clinical biopsy lymphoma samples to readily translate findings in cell line models to clinical samples. In this application, PLA may be used as a marker of oncogenic signaling to be targeted by specific precision medicine strategies. Measuring oncogenic signaling in clinical biopsy samples has been difficult because protein phosphorylation marks often decay in the preparation of FFPE biopsy samples, due to prolonged amounts of time at room temperature when the samples are processed [8]. However, many protein-protein interactions are more stable and are retained in FFPE samples. For example, PLA can detect interactions between the IgM-BCR and TLR9 in FFPE biopsy samples from activated B-cell-like diffuse large B-cell lymphoma (ABC-DLBCL) patients that responded to the BTK inhibitor, ibrutinib [7].

The following is a protocol for PLA on lymphoma cell lines. The workflow starts with preparing cells for imaging by allowing them to attach to a treated optical surface, followed by fixation and permeabilization in preparation for intracellular antibody staining. Cells are then simultaneously stained with two primary antibodies—one for each target protein or protein modification. These primary antibodies must be raised in different host species. We frequently use either a mouse and rabbit or a goat and rabbit antibody combination. Next, cells are stained with anti-species-specific secondary antibodies conjugated to complementary oligonucleotides, followed by a ligation step and an amplification step. Oligonucleotide probes labeled with fluorescent markers are

incorporated in the amplification step. Finally, cells are imaged and data is analyzed.

2 Materials

2.1 Cell Preparation and Fixation

1. Chamber-well slides treated to allow tissue culture cells to adhere: 15 μ-slide angiogenesis slide (Ibidi).
2. Phosphate-buffered saline solution (PBS).
3. 16% paraformaldehyde ampules (PFA) diluted to 4% in PBS immediately before use.
4. Methanol, stored at −20 °C before use.
5. Blocking solution: Duolink blocking solution (Sigma) or 5% bovine serum albumin (BSA) (w/v) diluted in PBS.

2.2 Antibody Staining

1. Two primary antibodies raised from different species, preferentially rabbit and mouse.
2. Antibody dilution buffer: Duolink antibody dilution buffer (Sigma) or 1% BSA (w/v) in PBS.
3. Duolink secondary antibody conjugated to a PLUS oligonucleotide: Duolink in situ PLA probe anti-Rabbit PLUS (Sigma).
4. Duolink secondary antibody conjugated to a MINUS oligonucleotide: Duolink in situ PLA probe anti-Mouse MINUS (Sigma) or Duolink in situ PLA probe anti-Goat MINUS (Sigma).

2.3 Proximity Ligation Assay

1. Duolink In Situ Detection Reagents kit in Orange (Cy3) (Sigma). Other options/fluorophores are available. The kit contains 5× ligation stock, 1× ligase, 5× amplification stock, 1× polymerase.
2. Wash Buffer A: Tris-buffered saline solution with 0.05% Tween 20, pH 7.4. Dissolve 8.8 g NaCl, 1.2 g Tris base, and 0.5 mL Tween 20 in 1 L of H_2O, adjust pH to 7.4 using HCl, and sterile filter before use.
3. Wash Buffer B: Dissolve 5.84 g NaCl, 4.24 g Tris base, and 26.0 g Tris–HCl in 1 L of H_2O, adjust pH to 7.5 using HCl, and sterile filter before use.
4. Mounting media with DAPI: Fluoroshield Mounting Medium with DAPI (Abcam).

3 Methods

3.1 Cell Preparation and Fixation

1. Plate 50 μL of lymphoma cells resuspended at a concentration of $0.5–1.0 \times 10^6$ cells/mL into a well of the cell culture-treated chamber slide (*see* **Note 1**).
2. Allow cells to attach to the treated surface by incubating in a tissue culture incubator at 37 °C and 5% CO_2 for 30 min.
3. Aspirate off media and fix cells with 4% PFA in PBS for 20 min at room temperature (*see* **Note 2**).
4. Aspirate off PFA and wash with PBS.
5. Aspirate off PBS and add 50 μL of cold methanol to cells. Immediately place at −20 °C for 15–30 min to permeabilize.
   (a) Other permeabilization methods can be used, such as detergents like saponin or TX-100.
6. Wash cells twice with PBS.
7. Block cells in blocking solution for 30 min at room temperature.

3.2 Antibody Staining

1. Dilute primary antibodies in antibody dilution buffer and stain overnight at 4 °C (*see* **Notes 3** and **4**).
2. Wash twice for 10 min in large volume of 1% BSA in PBS. Use a coplin jar to wash slides in large volumes.
3. Meanwhile, dilute secondary antibodies (12 μL total per reaction):

   2.4 μL PLUS antibody

   2.4 μL MINUS antibody.

   7.2 μL Duolink antibody dilution buffer

   Mix and let stand for 20 min at room temperature.
4. Add 12 μL of secondary antibody mix to each sample and incubate in 37 °C incubator for 1 h.

3.3 Ligation and Amplification

1. Gently aspirate off secondary antibody mix and wash slides in buffer A two times for 5 min each. Use a coplin jar to wash slides in large volumes.
2. Prepare ligation mix (12 μL total per reaction): Add 2.4 μL 5× ligation stock to 9.3 μL pure water, mix, and add 0.3 μL ligase.
3. Add 12 μL of ligation mix to each sample and incubate for 30 min in 37 °C incubator.
4. Wash slides twice for 2 min each in buffer A. Use a coplin jar to wash slides in large volumes.

5. Prepare amplification mix (12 μL total per reaction): Add 2.4 μL of 5× amplification stock to 9.45 μL of pure water, mix, and add 0.15 μL of polymerase.
6. Add 12 μL of amplification mix to each sample and incubate for 100 min in 37 °C incubator (*see* **Note 5**).
7. Aspirate/blot off amplification mix and wash twice for 10 min in large volume of 1× buffer B. Use a coplin jar to wash slides in large volumes.
8. Dilute 1 mL of buffer B into 99 mL pure water to make 0.01× buffer B.
9. Wash slides in 0.01× buffer B for 1 min.
10. Aspirate/blot off buffer B and then add a drop of mounting medium with DAPI to each well.

3.4 Imaging and Data Analysis

1. Image cells for both DAPI and PLA puncta (detection settings for Cy3) with a confocal microscope, collecting either flat images or z-stacks (*see* **Note 6**).
2. Data analysis can be done using a variety of software packages (*see* **Note 7**).
3. To analyze PLA signal, both the number of spots and their intensity can be measured (*see* **Notes 8** and **9**).

4 Notes

1. Any similar polylysine-treated chamber well slides can be used, but volumes may have to be adjusted to cover total surface area.
2. Work in small batches, and do not completely aspirate all liquid out of the well to avoid letting cells dry out.
3. Finding a pair of antibodies that will give reliable data about protein-protein interactions using PLA requires careful dilution of primary antibodies. Typically, less antibody is used for PLA staining than when staining for direct immunofluorescence.
4. Further validation of PLA antibody pairs should be done by genetically disrupting protein expression through either shRNAs or CRISPR in cell lines. Genetic disruption should disrupt any PLA signal to validate a given antibody pair.
5. These methods can be adapted for use in formalin-fixed paraffin-embedded (FFPE) clinical biopsy samples by increasing the volume of reagents to cover the area of a given biopsy sample and by increasing the length of time for the amplification step to 2 h.

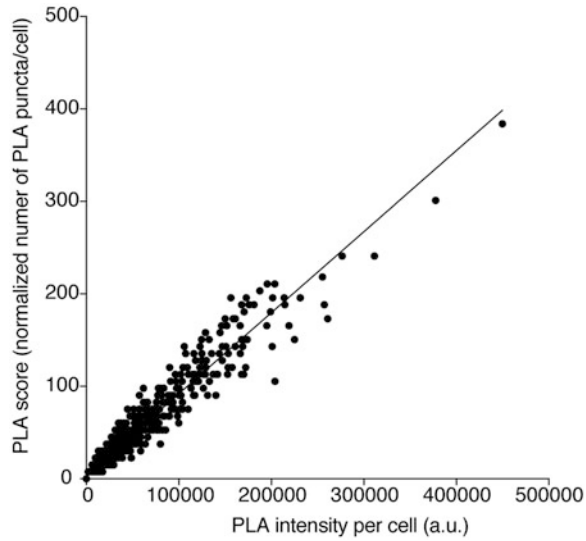


Fig. 3 Quantitation of PLA puncta per cell versus PLA intensity per cell shows that these measurements result in similar data interpretation. BCL10:MYD88 PLA performed in HBL1 cells

6. Collecting flat images allows for more experimental conditions to be measured, but z-stack data will be a more complete measurement of each cell. We have found that they are comparable and generally opt for flat images.
7. For data analysis, we have used BlobFinder [9]: (http://www.cb.uu.se/~amin/BlobFinder/).
8. We typically have counted the number of PLA puncta per cell. When pooling biological replicates of PLA experiments, we have normalized the number of PLA puncta to a standard control and called this the PLA score (Fig. 2, left side).
9. The number of PLA puncta per cell and the intensity of the PLA signal per cell are highly correlated (Fig. 3).

References

1. Young RM, Staudt LM (2013) Targeting pathological B cell receptor signalling in lymphoid malignancies. Nat Rev Drug Discov 12:229–243
2. Wu H, Fuxreiter M (2016) The structure and dynamics of higher-order assemblies: amyloids, signalosomes, and granules. Cell 165:1055–1066
3. Soderberg O, Gullberg M, Jarvius M, Ridderstrale K, Leuchowius KJ, Jarvius J, Wester K, Hydbring P, Bahram F, Larsson LG, Landegren U (2006) Direct observation of individual endogenous protein complexes in situ by proximity ligation. Nat Methods 3:995–1000
4. Weibrecht I, Leuchowius KJ, Clausson CM, Conze T, Jarvius M, Howell WM, Kamali-Moghaddam M, Soderberg O (2010) Proximity ligation assays: a recent addition to the proteomics toolbox. Expert Rev Proteomics 7:401–409
5. Soderberg O, Leuchowius KJ, Gullberg M, Jarvius M, Weibrecht I, Larsson LG, Landegren U (2008) Characterizing proteins and their

interactions in cells and tissues using the in situ proximity ligation assay. Methods 45:227–232

6. Mocanu MM, Varadi T, Szollosi J, Nagy P (2011) Comparative analysis of fluorescence resonance energy transfer (FRET) and proximity ligation assay (PLA). Proteomics 11:2063–2070
7. Phelan JD, Young RM, Webster DE, Roulland S, Wright GW, Kasbekar M, Shaffer AL 3rd, Ceribelli M, Wang JQ, Schmitz R, Nakagawa M, Bachy E, Huang DW, Ji Y, Chen L, Yang Y, Zhao H, Yu X, Xu W, Palisoc MM, Valadez RR, Davies-Hill T, Wilson WH, Chan WC, Jaffe ES, Gascoyne RD, Campo E, Rosenwald A, Ott G, Delabie J, Rimsza LM, Rodriguez FJ, Estephan F, Holdhoff M, Kruhlak MJ, Hewitt SM, Thomas CJ, Pittaluga S, Oellerich T, Staudt LM (2018) A multiprotein supercomplex controlling oncogenic signalling in lymphoma. Nature 560:387–391
8. O'Hurley G, Sjostedt E, Rahman A, Li B, Kampf C, Ponten F, Gallagher WM, Lindskog C (2014) Garbage in, garbage out: a critical evaluation of strategies used for validation of immunohistochemical biomarkers. Mol Oncol 8:783–798
9. Allalou A, Wahlby C (2009) BlobFinder, a tool for fluorescence microscopy image cytometry. Comput Methods Prog Biomed 94:58–65

Chapter 19

Phosphoproteomic Analysis of Signaling Pathways in Lymphomas

Björn Häupl, Henning Urlaub, and Thomas Oellerich

Abstract

Cell fate decisions are controlled by complex signal transduction processes that transmit information via posttranslational protein modifications such as phosphorylation. In lymphoma, as in other cancer types, these signaling networks are often dysregulated and thus contribute to malignant transformation and tumor maintenance. For example, B-cell antigen receptor signals are rewired in certain lymphoma types, such as diffuse large B-cell lymphomas, to promote cell growth and survival of the malignant cell clones. Hence, global elucidation of such intricate signaling networks is important for an improved understanding of the biology of these tumors and the identification of target proteins for therapeutic purposes.

We describe here a mass spectrometry-based phosphoproteomic approach for characterization of intracellular signaling events and their dynamics. This integrated phosphoproteomic technology combines phosphopeptide enrichment and fractionation with liquid-chromatography-coupled mass spectrometry for the site-specific mapping and quantification of thousands of phosphorylation events in a given cell type. Such global signaling analyses provide valuable insights into oncogenic signaling networks and can inform drug development efforts.

Key words Phosphoproteomics, Signaling, Mass spectrometry, B-cell lymphomas, Antigen receptors

1 Introduction

The development and optimum function of the lymphoid lineage rely on antigen receptor signaling, which controls several key processes including differentiation and proliferation responses as well as immune-cell effector functions [1]. However, dysregulation of these signals, for example, by acquisition of mutations in signaling effectors, can lead to malignant transformation [2]. For example, certain B-cell malignancies such as chronic lymphocytic leukemia and activated B-cell-like diffuse large B-cell lymphomas depend on "chronic active" B-cell receptor (BCR) signaling [3, 4], which is caused by (i) auto-antigen binding by their BCRs [5] and (ii) activatory mutations in BCR signaling effectors [6]. Other malignancies, such as Burkitt lymphomas, rely on tonic BCR

Ralf Küppers (ed.), *Lymphoma: Methods and Protocols*, Methods in Molecular Biology, vol. 1956,
https://doi.org/10.1007/978-1-4939-9151-8_19,

signaling, which provides an antigen-independent baseline survival signal [7, 8]. Chronic active and tonic BCR signaling differ in their phosphoproteomic and transcriptional output, a fact that demonstrates how a given receptor can activate distinct pathways by different signaling modes [8]. The complexity of such signaling networks is further increased by cross-talk between active signaling pathways that are integrated to define the ultimate signaling output. Importantly, oncogenic signaling has therapeutic implications. For example, targeting of aberrant BCR signaling has proven to be a promising therapeutic strategy in BCR-dependent malignancies; this finding is reflected by the frequent use and therapeutic success of Bruton's tyrosine kinase (BTK) and PI3-kinase inhibitors [9]. All in all, this underscores the relevance of signal transduction processes for lymphoma pathogenesis as well as for therapeutic approaches.

Despite the importance of signal transduction for inter- and intracellular communication, signaling networks have so far mostly been studied at the level of individual signaling molecules or complexes. Initially, the incorporation of radioactive phosphate into substrate proteins or peptides was used to study phosphorylation events. This changed when phosphosite-specific antibodies became available, and these are still frequently used to investigate phosphorylation events by various techniques including Western blotting, immunofluorescence approaches, enzyme-linked immunosorbent assays (ELISA), and flow as well as mass cytometry. However, the main limitations of these targeted conventional techniques are the restricted availability of reliable phosphosite-specific antibodies and their focus on known rather than undescribed signaling events.

For global signaling analyses, especially those focused upon discovery, mass spectrometric techniques have evolved as the state-of-the-art approach. They have made possible site-specific identification and quantification of more than 30,000 p-sites in a single experiment, making mass spectrometry (MS) the most powerful tool for comprehensive signaling analysis [10]. The analytical procedure for such MS-based phosphoproteomic workflows consists of three main parts which are described here in the Methods section: (1) sample preparation including phosphopeptide enrichment and fractionation, (2) the MS analysis, and (3) computational data processing.

A key aspect of the sample preparation process is the enrichment of phosphopeptides before MS analysis. Various enrichment strategies have been developed that exploit the physicochemical properties of the phosphate group. One possibility for phosphopeptide enrichment is ion-exchange chromatography using strong cation exchange (SCX) [11]. This technique is efficient, but not very specific, as it enriches peptides with acidic modifications other than phosphorylations as well. An alternative is hydrophilic

interaction chromatography (HILIC), which separates peptides on the basis of their hydrophilicity [12]. More recent studies have combined chromatographic workflows to achieve increased phosphoproteome coverage. For example, metal oxide affinity chromatography is an option for such combined workflows where metal oxides such as TiO_2 are used for further enrichment of phosphopeptides [13]. Notably, tyrosine phosphorylation is a comparatively rare protein modification, as it represents only about 1% of the phosphoproteome. Antibody-based workflows aimed at specific enrichment of tyrosine-phosphorylated peptides have been established to increase the coverage of this relatively rare modification [8].

The key principle of MS-based phosphoproteomics is the fragmentation of phosphopeptide ions and their subsequent mass-based analysis and identification. Discovery-mode (also called shotgun) experiments for global phosphoproteome profiling are usually performed in the data-dependent acquisition mode, where a certain number of detected peptide ions are selected for fragmentation and subsequent mass analysis. Importantly, various strategies have been developed for quantitative phosphoproteomic analyses. These are either label-free strategies or workflows based on stable isotope labeling [14]. Various MS techniques and instrument settings have been described for phosphoproteome analyses and are reviewed elsewhere [15].

In this book chapter, we describe a robust quantitative phosphoproteomic workflow that is useful for global signaling analyses.

2 Materials

2.1 Sample Preparation (See Note 1)

1. Phosphate-buffered saline (PBS).
2. Lysis buffer: 50 mM Tris–HCl, pH 7.5, 150 mM NaCl, 1 mM EDTA, 1% NP-40, 0.1% deoxycholate sodium salt. Add the following components freshly before use of lysis buffer: protease inhibitor cocktail (e.g., Roche complete), 5 mM β-glycerophosphate, 5 mM NaF, 1 mM sodium orthovanadate, 10 mM *N*-ethylmaleimide.
3. Sonication device.
4. Protein assay (e.g., Thermo Scientific Pierce 660 nm Protein Assay kit, use according to manufacturer's instructions).

2.2 Protein Precipitation and In-Solution Digestion

1. Acetone.
2. Denaturation buffer: 10 mM HEPES-NaOH, pH 8.0, 6 M urea.
3. Dithiothreitol (DTT) (e.g., 10 mM stock solution).
4. Iodoacetamide (IAA) (e.g., 55 mM stock solution).

5. Lys-C (e.g., Promega MS grade Lys-C, reconstitute according to manufacturer's instructions).
6. 50 mM ammonium bicarbonate.
7. Trypsin (e.g., Promega MS grade Trypsin, reconstitute according to manufacturer's instructions).
8. 20% trifluoroacetic acid (TFA).

2.3 Peptide Purification

1. Reversed-phase C18 columns (e.g., Waters Sep-Pak C18 Classic Cartridge, 360 mg silica sorbent, 55–105 μm particle size, 125 Å pore size).
2. 5 and 10 mL syringes.
3. Acetonitrile (ACN).
4. 0.1% TFA.
5. Elution buffer: 50% ACN, 0.1% TFA.
6. Vacuum concentrator.

2.4 Enrichment of Phosphopeptides

1. TiO_2 beads (e.g., MZ Analysentechnik Mainz).
2. TFA.
3. Binding buffer: 50% ACN, 6% TFA.
4. Wash buffer: 50% ACN, 0.1% TFA.
5. Stage tips packed with C8 material (e.g., 3 M Empore C8 extraction disk).
6. Elution solutions: 5% aqueous ammonia and 10% aqueous ammonia/25% ACN.
7. Vacuum concentrator.

2.5 Sample Fractionation

1. Stage tips packed with SCX material (e.g., 3 M Empore SCX extraction disk).
2. Methanol.
3. Wash buffer: 40% ACN, 0.1% TFA.
4. Elution buffer stock solutions: 40 mM acetic acid, 40 mM boric acid, 40 mM phosphoric acid. Adjust to stock buffers (pH 3.0/3.5/4.0/4.5/5.5/6.0/6.5/8.5) with NaOH.
5. Allocate elution buffer with desired pH by preparing solutions with 40% ACN and 60% stock buffer (pH 3.0/3.5/4.0/4.5/5.5/6.0/6.5/8.5).
6. Vacuum concentrator.
7. 0.1% TFA.
8. Equipment for C18 stage tipping (e.g., 3 M Empore C18 extraction disk).

2.6 Mass Spectrometry Analysis and Data Processing

1. Mass spectrometer with capability for tandem mass spectrometry (MS/MS), coupled online to a nano-HPLC system for reversed-phase separation of peptides, e.g., Orbitrap-based instruments with high resolution and mass accuracy such as Q Exactive HF interfaced with Dionex 3000 Ultimate LC systems (both from Thermo Fisher Scientific).
2. Software to process MS raw data for protein identification and phosphorylation site assignment by matching MS/MS spectra to protein databases (e.g., Uniprot), e.g., the MaxQuant framework including the Andromeda peptide search engine [16, 17].
3. Downstream analysis software for data filtering, processing of quantitative MS data, statistical evaluation, and display of results, e.g., the Perseus application (compatible with MaxQuant) [18].

3 Methods

Here we describe an MS workflow for global phosphoproteomic analysis of oncogenic signaling (Fig. 1, *see* **Note 2**). The first step is the isolation of proteins from cells followed by enzymatic digestion of typically 3 mg (range 0.5–5 mg) of total protein. After peptide purification, phosphopeptides are enriched by using TiO_2 beads (*see* **Note 3**). In the next step, the resulting peptide mixture is separated into eight individual fractions by strong cation-exchange solid-phase extraction. These fractions are analyzed by nano-HPLC/nano-ESI tandem mass spectrometry. For protein identification and assignment of phosphorylation sites, the MS raw data is processed with suitable software tools comprising peptide search engines, e.g., the MaxQuant framework.

3.1 Sample Preparation

1. Harvest the cells and wash them twice with ice-cold PBS.
2. Resuspend the cell pellet in 500 µL of lysis buffer per 10^8 cells and incubate on ice for 30 min.
3. Sonicate the cell lysate three times for 10 s at 40–50% amplitude (*see* **Note 4**). Cool the sample on ice between sonication cycles.
4. Centrifuge the samples at 15,000 × *g* and 4 °C for 15 min, transfer the supernatant to a fresh sample tube, and determine protein concentration.

3.2 Protein Precipitation and In-Solution Digestion

1. Withdraw the sample volume equivalent to 3 mg of total protein content, add four volumes of ice-cold acetone, and incubate at −20 °C overnight.

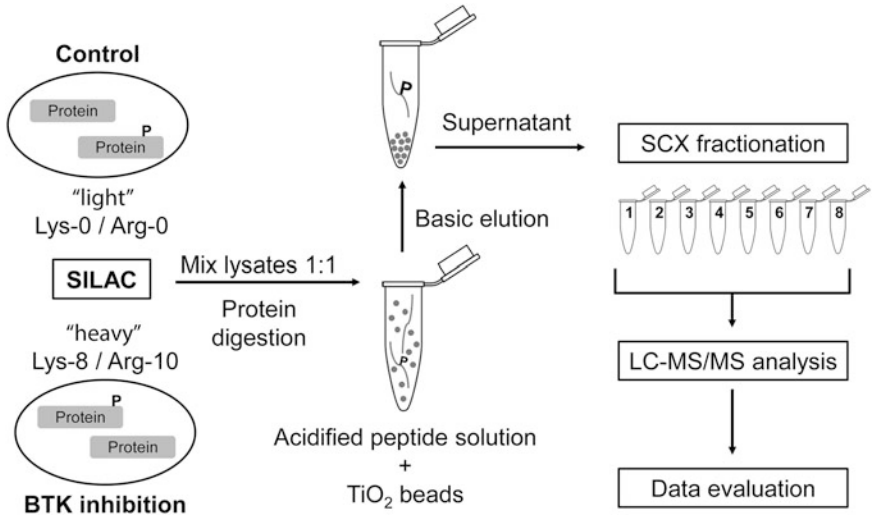


Fig. 1 Quantitative MS workflow for global phosphoproteomic analysis of oncogenic signaling. The influence of BTK inhibition on cellular phosphorylation events is determined by global phosphopeptide enrichment combined with SILAC-based relative quantification. First, total cell lysates of differentially labeled samples from control experiments ("light") and BTK inhibition treatments ("heavy") are prepared and mixed in a 1:1 ratio according to protein concentration. After protein precipitation and in-solution digestion, the peptide mixture is subjected to phosphopeptide enrichment using TiO_2 beads. Next, the beads are washed, and the phosphopeptides are eluted under basic conditions. The phosphopeptide-enriched supernatant is further fractionated into eight fractions by SCX solid-phase extraction, and the individual fractions are analyzed by LC-MS/MS. Finally, MS raw data is processed for protein identification, phosphorylation site assignment, and quantitative analysis of differences in protein phosphorylation

2. Centrifuge the sample for 10 min at 15,000 × *g* and 4 °C to sediment precipitated protein, decant the supernatant, and allow evaporation of residual acetone.
3. Resuspend the precipitate in 3 mL of denaturation buffer, add DTT to a final concentration of 1 mM, and incubate for 1 h at 37 °C.
4. Add IAA to a final concentration of 5.5 mM and incubate for 1 h at 37 °C protected from light.
5. Add Lys-C in an enzyme-to-protein ratio of 1:100 (w/w) and incubate at 37 °C for 4 h.
6. Dilute the sample with three volumes of 50 mM ammonium bicarbonate, add trypsin in an enzyme-to-protein ratio of 1:100 (w/w), and incubate at 37 °C overnight (*see* **Note 5**).
7. Acidify the sample by adding 150 µL of 20% TFA and incubate on ice for 30 min.
8. Centrifuge the resulting peptide solution at 1800 × *g* for 15 min at 4 °C and keep the supernatant.

3.3 Peptide Purification

1. Connect a C18 cartridge to a 10 mL syringe, and wash the column with 5 mL of ACN.
2. Wash the column three times with 5 mL of 0.1% TFA, and load the peptide solution. Repeat the loading step with the sample flow-through.
3. Wash the column three times with 10 mL of 0.1% TFA.
4. Change to a fresh syringe with a volume of 5 mL, and elute the peptides with 3 mL of elution buffer. Repeat the elution step twice.
5. Reduce the sample volume to approximately 2 mL using a vacuum concentrator.

3.4 Enrichment of Phosphopeptides

1. Use 3 mg of TiO_2 beads per 1 mg of total protein (adjust buffer volumes for the following steps according to the amount of protein and beads used). Wash the beads with 200 μL of binding buffer for 5 min on a rotating tube shaker, centrifuge at 2000 × *g* for 30 s, and discard the supernatant.
2. Adjust the peptide solution to 6% TFA, add to the equilibrated TiO_2 beads, and incubate for 2 h on a rotating tube shaker at room temperature.
3. Wash the beads twice with binding buffer.
4. Wash the beads twice with wash buffer.
5. Resuspend the beads in 200 μL of wash buffer, transfer them to a prepared C8 stage tip, and centrifuge at 5000 × *g* for 30 s (*see* **Note 6**).
6. Elute the phosphopeptides with 100 μL of 5% aqueous ammonia, and centrifuge at 5000 × *g* for 30 s. Repeat the elution step with 10% aqueous ammonia/25% ACN.
7. Reduce the sample volume to approximately 10 μL by using a vacuum concentrator, and acidify the concentrate with 100 μL of 0.1% TFA.

3.5 SCX Sample Fractionation

1. Wash the prepared SCX tip with 100 μL of methanol (centrifuge at 5000 × *g* for 30 s).
2. Wash the tip with 50 μL of pH 3.0 buffer and then with 50 μL of pH 8.5 buffer.
3. Wash tip with 100 μL of wash buffer.
4. Load the sample of enriched phosphopeptides.
5. Elute with 100 μL of each elution buffer in turn into separate fresh sample tubes, starting with the pH 3.0 buffer.
6. Reduce the volume of the individual fractions to approximately 10 μL each by using a vacuum concentrator, and mix with 100 μL of 0.1% TFA.
7. Perform phosphopeptide purification by C18 stage tipping.

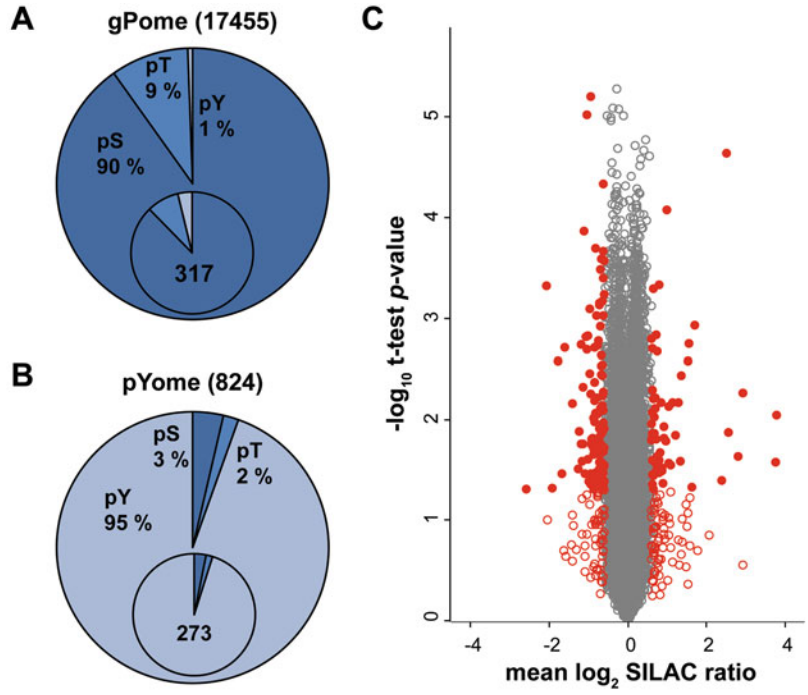


Fig. 2 Quantitative analysis of protein phosphorylation after BTK inhibition. (**a**) On the basis of the strategy presented here for global phosphopeptide enrichment (gPome) combined with SILAC-based quantification, ~17,500 class I phosphorylation sites (localization probability >75%) were assigned. The largest fraction (90%) represented serine phosphorylation (pS), followed by pT (9%) and to a minor extend pY (1%). Statistical analysis of the quantification data (SILAC ratios) showed 317 sites with significant differences in abundance, depending on BTK inhibition (inset). (**b**) According to **Note 3**, the global enrichment strategy was augmented by antibody-based specific enrichment of tyrosine-phosphorylated peptides (pYome). (**c**) The visualization of the statistical analysis of the quantification data (SILAC ratios) assigned to each identified phosphorylation site revealed a large fraction of unregulated sites (gray circles) as well as up- and downregulated sites. These showed significance, either in an outlier analysis of the mean SILAC ratios (red circles and dots) or, additionally, in a *t*-test analysis with the individual replicate ratios for each site (red dots). Analysis was conducted in two biological replicate experiments with two technical replicates each. At least three of four valid SILAC ratio values were required for data evaluation

3.6 Mass Spectrometry Analysis and Data Processing

Tandem MS analysis of enriched and purified phosphopeptides may be based on a standard experimental setup comprising a high-performance liquid chromatography system interfaced with a mass spectrometer featuring both high resolution and high mass accuracy. Examples of results obtained using the protocol described above are shown for the quantitative analysis of differential protein phosphorylation after the inhibition of BTK in the lymphoma cell line U-2932 with ibrutinib (Fig. 2). As described in detail elsewhere [8], LC-MS/MS analysis was performed on a Q Exactive HF mass spectrometer coupled to a Dionex Ultimate 3000 RSLCnano

system via a nano-electrospray ionization (ESI) interface. In brief, analysis was carried out using a two-column LC setup. Phosphopeptides were first trapped on a precolumn and then separated on an analytical column by using a linear solvent gradient. For MS measurements, a data-dependent acquisition scheme was used. On the basis of full MS survey scans, precursor ions were selected and isolated for collision-induced dissociation, followed by mass analysis of the product ions (MS/MS).

MS raw data was processed with the MaxQuant software, and MS/MS spectra were searched against the human reference proteome obtained from the UniProt database. For downstream analysis of phosphorylation sites assigned to identified proteins, the Perseus software was employed. After removal of potential contaminants, decoy hits, and phosphorylation sites identified with a localization probability below 75%, statistical evaluation was carried out in order to reveal whatever differences in protein phosphorylation were present between the samples under investigation.

4 Notes

1. All reagents should be of the highest purity available. Aqueous solutions should be prepared with ultrapure water.
2. The steps described in the Methods section refer to the basic procedures of sample preparation and phosphopeptide enrichment for MS analysis. Typically, these may be combined with quantitative approaches in order to elucidate differential protein phosphorylation in a given set of experiments. Here, one useful strategy is the application of stable isotope labeling with amino acids in cell culture (SILAC) [14, 19], which allows relative quantification, comparing different experimental conditions. For this purpose, differentially labeled samples would be mixed in equivalent ratios after the preparation of total cell lysates (*see* Fig. 1).
3. The protocol described here makes possible the global enrichment of phosphorylated peptides, which leads to strong representation of amino-acid-specific phosphorylations according to their natural occurrence (*see* Fig. 2a). Thus, serine phosphorylation contributes the largest fraction of identified phosphorylation sites (~85–90%), followed by threonine phosphorylation (~10%) and tyrosine phosphorylation with less than 1%. However, it may be of interest to gain further insights into rare tyrosine phosphorylation events on (likewise) low-abundance proteins. For that purpose, this protocol may be augmented by enrichment using antibodies specific for tyrosine-phosphorylated peptides (*see* Fig. 2b).

4. Sonication intensity may depend on the sonication device used. Settings that result in optimum cell lysis and shearing of released DNA should be used.
5. The combination of Lys-C that retains proteolytic activity already under denaturing conditions (6 M urea) and trypsin treatment is advantageous for achieving a high degree of protein digestion, which is crucial for a high coverage of phosphopeptides.
6. For stage tipping either commercially available or self-made, filter tips may be used. For the latter, cut disks from the sorbent material using a 17 gauge Hamilton syringe, and place them into a 200 µL pipet tip. For centrifugation the filter tips may be plugged into the lid of a sample tube, which was punctured beforehand.

References

1. Kurosaki T, Shinohara H, Baba Y (2010) B cell signaling and fate decision. Annu Rev Immunol 28:21–55. https://doi.org/10.1146/annurev.immunol.021908.132541
2. Shaffer AL 3rd, Young RM, Staudt LM (2012) Pathogenesis of human B cell lymphomas. Annu Rev Immunol 30:565–610. https://doi.org/10.1146/annurev-immunol-020711-075027
3. Duhren-von Minden M, Ubelhart R, Schneider D et al (2012) Chronic lymphocytic leukaemia is driven by antigen-independent cell-autonomous signalling. Nature 489 (7415):309–312. https://doi.org/10.1038/nature11309
4. Davis RE, Ngo VN, Lenz G et al (2010) Chronic active B-cell-receptor signalling in diffuse large B-cell lymphoma. Nature 463 (7277):88–92. https://doi.org/10.1038/nature08638
5. Young RM, Wu T, Schmitz R et al (2015) Survival of human lymphoma cells requires B-cell receptor engagement by self-antigens. Proc Natl Acad Sci U S A 112 (44):13447–13454. https://doi.org/10.1073/pnas.1514944112
6. Schmitz R, Wright GW, Huang DW et al (2018) Genetics and pathogenesis of diffuse large B-cell lymphoma. N Engl J Med 378 (15):1396–1407. https://doi.org/10.1056/NEJMoa1801445
7. Schmitz R, Young RM, Ceribelli M et al (2012) Burkitt lymphoma pathogenesis and therapeutic targets from structural and functional genomics. Nature 490(7418):116–120. https://doi.org/10.1038/nature11378
8. Corso J, Pan KT, Walter R et al (2016) Elucidation of tonic and activated B-cell receptor signaling in Burkitt's lymphoma provides insights into regulation of cell survival. Proc Natl Acad Sci U S A 113(20):5688–5693. https://doi.org/10.1073/pnas.1601053113
9. Young RM, Staudt LM (2013) Targeting pathological B cell receptor signalling in lymphoid malignancies. Nat Rev Drug Discov 12 (3):229–243. https://doi.org/10.1038/nrd3937
10. Sacco F, Humphrey SJ, Cox J et al (2016) Glucose-regulated and drug-perturbed phosphoproteome reveals molecular mechanisms controlling insulin secretion. Nat Commun 7:13250. https://doi.org/10.1038/ncomms13250
11. Beausoleil SA, Jedrychowski M, Schwartz D et al (2004) Large-scale characterization of HeLa cell nuclear phosphoproteins. Proc Natl Acad Sci U S A 101(33):12130–12135. https://doi.org/10.1073/pnas.0404720101
12. Singer D, Kuhlmann J, Muschket M et al (2010) Separation of multiphosphorylated peptide isomers by hydrophilic interaction chromatography on an aminopropyl phase. Anal Chem 82(15):6409–6414. https://doi.org/10.1021/ac100473k
13. Nilsson CL (2012) Advances in quantitative phosphoproteomics. Anal Chem 84 (2):735–746. https://doi.org/10.1021/ac202877y
14. Bantscheff M, Lemeer S, Savitski MM et al (2012) Quantitative mass spectrometry in proteomics: critical review update from 2007 to the present. Anal Bionanal Chem 404

(4):939–965. https://doi.org/10.1007/s00216-012-6203-4

15. Junger MA, Aebersold R (2014) Mass spectrometry-driven phosphoproteomics: patterning the systems biology mosaic. Wiley Interdiscip Rev Dev Biol 3(1):83–112. https://doi.org/10.1002/wdev.121
16. Cox J, Mann M (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol 26(12):1367–1372. https://doi.org/10.1038/nbt.1511
17. Cox J, Neuhauser N, Michalski A et al (2011) Andromeda: a peptide search engine integrated into the MaxQuant environment. J Proteome Res 10(4):1794–1805. https://doi.org/10.1021/pr101065j
18. Tyanova S, Temu T, Sinitcyn P et al (2016) The Perseus computational platform for comprehensive analysis of (prote)omics data. Nat Methods 13(9):731–740. https://doi.org/10.1038/nmeth.3901
19. Ong S-E, Blagoev B, Kratchmarova I et al (2002) Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. Mol Cell Proteomics 1(5):376–386. https://doi.org/10.1074/mcp.M200025-MCP200

Chapter 20

Ultrasensitive Detection of Circulating Tumor DNA in Lymphoma via Targeted Hybridization Capture and Deep Sequencing of Barcoded Libraries

Miguel Alcaide, Christopher Rushton, and Ryan D. Morin

Abstract

Liquid biopsies are rapidly emerging as powerful tools for the early detection of cancer, noninvasive genomic profiling of localized or metastatic tumors, prompt detection of treatment resistance-associated mutations, and monitoring of therapeutic response and minimal residual disease in patients during clinical follow-up. Growing evidence strongly supports the utility of circulating tumor DNA (ctDNA) as a biomarker for the stratification and clinical management of lymphoma patients. However, ctDNA is diluted by variable amounts of cell-free DNA (cfDNA) shed by nonneoplastic cells causing a background signal of wild-type DNA that limits the sensitivity of methods that rely on DNA sequencing. Here, we describe an error suppression method for single-molecule counting that relies on targeted sequencing of cfDNA libraries constructed with semi-degenerate barcode adapters. Custom pools of biotinylated DNA baits for target enrichment can be designed to specifically track somatic mutations in one patient, survey mutation hotspots with diagnostic and prognostic value or be comprised of comprehensive gene panels with broad patient coverage in lymphoma. Such methods are amenable to track ctDNA levels during longitudinal liquid biopsy testing with high specificity and sensitivity and characterize, in real time, the genetic profiles of tumors without the need of standard invasive biopsies. The analysis of ultra-deep sequencing data according to the bioinformatics pipelines also described in this chapter affords to harness lower limits of detection for ctDNA below 0.1%.

Key words Liquid biopsy, Noninvasive genetic profiling, Mutation detection, Cell-free DNA, Minimal residual disease, Therapeutic response, Tumor burden, Targeted enrichment, Duplex sequencing, DNA damage

1 Introduction

Growing evidence from a substantial number of proof-of-principle studies indicates that liquid biopsies hold promise for the early detection and noninvasive repetitive monitoring of solid tumors, patient stratification, prompt discovery of treatment-resistance mutations, and comprehensive genomic profiling of complex cancer abnormalities showing spatiotemporal heterogeneity and/or multiple metastasis [1–6]. Cancerous and nonneoplastic cells can

Ralf Küppers (ed.), *Lymphoma: Methods and Protocols*, Methods in Molecular Biology, vol. 1956,
https://doi.org/10.1007/978-1-4939-9151-8_20,

shed fragments of their genome into the surrounding body fluids through processes such us apoptosis, necrosis, or active cellular secretion. This easy-to-extract cell-free DNA (cfDNA) most commonly circulates within body fluids wrapped around the histones that comprise the chromatosome (nucleosome + linker histone H1), with a size distribution peak around 167 bp or multiples thereof [7, 8]. Such fragmentation pattern is consistent with a caspase-dependent cleavage, and hence the presence of cfDNA in body fluids such as blood, urine, or cerebrospinal fluid (CSF) is typically ascribed to apoptosis. Longer tumor-derived DNA fragments (>10 kb) have been suggested to be derived from necrotic cells [9, 10] or found to travel inside secreted microvesicles such as exosomes [11]. Interestingly, studies conducted in pancreatic cancer patients hypothesize that exosome-derived DNA (exoDNA) might be more abundant during early stages of the disease than apoptotic or necrotic cfDNA [12]. Although each particular release mechanism is expected to have a drastic effect on the size distribution and isolation of tumor-derived DNA fragments, there is still controversy regarding the primary mechanistic source of cfDNA release in vivo and in vitro [7, 13–16]. Notably, physiological factors or disorders such as pregnancy, cancer, intensive exercise, trauma, postradiation cellular damage, inflammation, or cardiovascular injury may also occasionally contribute to a great extent to the total pool of cfDNA [7, 8].

The presence of somatic mutations not occurring in the genome of nonmalignant cells differentiates circulating tumor DNA (ctDNA) from the rest of cfDNA, and therefore it can be specifically targeted and quantified by numerous analytical methods [7, 8, 17, 18]. Various reports suggest that ctDNA is, on average, shorter than the cfDNA shed by nonmalignant cells, and it has been hypothesized that this could be related to differences in the patterns of nucleosome positioning that regulate chromatic structure and gene expression [19, 20]. It has been suggested that the relative abundance of DNases in a given body fluid can also influence the size profile of cfDNA [15, 21]. Overall, cfDNA is quickly and steadily removed through degradation or filtration by the liver and the kidney in just a few hours, which turns ctDNA into an ideal biomarker to detect "real-time" changes in tumor composition and prevalence [7, 8, 22, 23]. One of the major hurdles in liquid biopsy research, however, is the dilution of ctDNA in a background of variable quantities of cfDNA that derives from nonmalignant lymphoid and myeloid cells undergoing regular apoptotic cell death [8, 15]. The limited sensitivity for ctDNA detection, exacerbated by this phenomenon, is particularly problematic in samples from patients with early stages of the disease, immediately after surgical removal of tumors, or is a general limitation associated with certain cancer diagnoses of primary sites [24–26]. One classic example involves the difficulty to detect

ctDNA in the blood of patients diagnosed with tumors affecting the central nervous system owing to the effect of the brain-blood barrier on cfDNA dissemination. Recent work has nonetheless corroborated that liquid biopsies relying on CSF samples are more appropriate for the genetic characterization and monitoring of these malignancies [27, 28]. In general, absolute ctDNA levels are expected to vary according to tumor type, burden, and stage but other factors such as the tumor microenvironment, treatment modality and timing of liquid biopsy, cellular turnover rates, or passive release of ctDNA by necrotic cells could also affect this [8, 29–31].

Given the enormous potential of ctDNA but the frequently low prevalence of tumor-derived aberrations in liquid biopsies, comprehensive research efforts have been allocated into the design of ultrasensitive methods for the accurate detection and quantification of ctDNA. A plethora of such methods has become available to the scientific and clinical community (reviewed in refs. [8, 17, 18, 32–34]). Methods utilize a variety of molecular biology techniques such as co-amplification at lower denaturation temperature-PCR (COLD-PCR), synchronous coefficient of drag alteration (SCODA), and differential strand separation at critical temperature (DISSECT); allele-specific or blocking techniques such as the amplification-refractory mutation system (ARMS) or those relying on modified synthetic oligonucleotides incorporating locked nucleotide acid (LNA) or peptide nucleic acid (PNA) residues; digital PCR and beaming for the compartmentalization and absolute quantification of mutant alleles; and a plethora of NGS-based strategies such as Safe-SeqS, Circle-Seq, Cypher-Seq, iDES, Pro-Seq, duplex sequencing, or TAm-seq. Other advanced detection systems for ctDNA detection such as those relying on matrix-assisted laser desorption/ionization (MALDI), surface-enhanced Raman spectroscopy (SERS), electrochemical chips, or fluorescently coded microparticles are reviewed elsewhere [18].

Growing evidence supports the clinical pertinence of ctDNA in establishing diagnostic and prognostic stratification of lymphoma patients and its potential to evaluate clinical response, predict relapsed or refractory cases, monitor minimal residual disease (MRD), and detect acquired resistance-related mutations during follow-up [35–41]. Strategies for detecting ctDNA in lymphoid malignancies were devised early on owing to the convenience of VDJ recombination leaving a clonal genetic marker in all cancer cells [42]. However, unlike strategies used in other cancers, these approaches do not target driver mutations and, owing to various technical limitations, cannot be applied to all patients [43]. In this chapter, we focus on a more general strategy for ctDNA detection and quantification that relies on the error-suppression duplex sequencing approach [44, 45] and entails cfDNA library construction using semi-degenerate barcoded adapters and target

enrichment by means of custom pools of biotinylated DNA baits [46, 47]. This approach benefits from cost-effective synthesis of single-stranded barcoded oligonucleotides that can be readily hybridized to yield adapter sets, which are amenable to multiplexing through the incorporation of hundreds or even thousands of sample-specific indexes. We also guide the reader through some data analysis pipelines that leverage the unique library properties to perform error correction, single-molecule counting, and mutation detection. Simply put, only the variants that are independently supported by the consensus sequences of the two independent strands derived from the same double-stranded cfDNA duplex are considered as high-confidence variants. Other cutting-edge and high-throughput ultrasensitive methods applied to the investigation, management, and monitoring of lymphoma patients, such as digital PCR or alternative NGS-based methods, have been reviewed or described elsewhere [36, 41, 48–50].

2 Materials

1. Cell-free DNA BCT© blood collection tubes (Streck, La Vista, NE, USA) or EDTA blood collection tubes (*see* **Note 1**).
2. Centrifuge with a rotor adapted for blood collection tubes (*see* **Note 2**).
3. Biosafety cabinet (*see* **Note 3**).
4. 2 mL Cryovial tubes and caps for plasma storage.
5. Cryoboxes.
6. Falcon tubes, 15 mL.
7. Sterile disposable transfer pipettes.
8. Fetal bovine serum, FBS.
9. Dimethyl sulfoxide, DMSO.
10. RPMI 1640 medium, no glutamine.
11. Red blood cell (RBC) lysis solution (e.g., Qiagen, Hilden, Germany).
12. Phosphate-buffered saline (PBS): 2.67 mM KCl, 1.47 mM KH_2PO_4, 137.9 mM NaCl, 8.1 mM Na_2HPO_4, pH = 7.4.
13. Cell-free DNA isolation kits, such as the QIAamp Circulating Nucleic Acid Kit (Qiagen, Hilden, Germany) or the MagMAX™ Cell-Free DNA Isolation Kit (Applied Biosystems, Foster City, CA, USA) (*see* **Note 4**).
14. Accessories for specific cfDNA isolation kits, such as the QIAvac Connecting System, vacuum pump, VacValves, and VacConnectors for the QIAamp Circulating Nucleic Acid Kit, or magnetic stands such as those belonging to the DynaMag™ Magnet series (ThermoFisher Scientific, Waltham, MA, USA)

for the MagMAX™ Cell-Free DNA Isolation Kit. Thermo Fisher Scientific offers the possibility to automate the cfDNA purification procedure (up to 96 extractions in one single run) through the KingFisher Flex Magnetic Particle Processor.

15. Proteinase K solution.
16. 20% SDS solution.
17. Agilent 2100 Bioanalyzer instrument and High Sensitivity DNA Analysis kits (Agilent, Santa Clara, CA, USA) or similar to infer DNA fragment size distribution.
18. Digital heaters with modular blocks to incubate small tubes.
19. Temperature-controlled water bath.
20. Benchtop microcentrifuge for 1.5–1.7 mL tubes.
21. Benchtop microcentrifuge for the quick spin of PCR tubes and strips.
22. Benchtop microcentrifuge for the quick spin of PCR plates.
23. Vortex mixer.
24. Low-binding DNA microcentrifuge tubes to minimize cfDNA sample loss.
25. Qubit fluorometric DNA quantitation system (Thermo Fisher Scientific, Waltham, MA, USA): It must be accompanied by Qubit Assay Tubes (set of 500) and Qubit dsDNA HS Assay kits (*see* **Note 5**).
26. NEBNext Ultra II DNA Library Prep Kit for Illumina (New England Biolabs, Ipswich, MA, USA) or similar to build cfDNA libraries (*see* **Note 6**).
27. Multichannel, adjustable pipettes with multi-dispense function (optional but extremely useful when handling large number of samples in the manual mode).
28. Disposable pipetting/reagent reservoirs (optional but extremely useful when handling large number of samples in the manual mode).
29. PCR machine/thermal cycler.
30. Adhesive PCR plate foils (e.g., Thermo Fisher Scientific, Waltham, MA, USA).
31. MAXYMum Recovery PCR tubes or low-binding DNA PCR strips or tubes, 0.2 mL, and flat caps (e.g., VWR, Radnor, PA, USA).
32. 96-Well semi-skirted PCR plates.
33. Custom, semi-degenerate barcoded adapters for single-molecule tagging and error correction purposes [46] (*see* **Note 7**). Oligonucleotides carrying semi-degenerate DNA sequences for barcoded adapter generation can be ordered from Integrated DNA Technologies (Skokie, IL, USA) or

other manufacturers permitting 3′-phosphorothioate bonds, 5′-phosphate modifications, and HPLC purification. The DNA sequences of these oligonucleotides are provided in Table 1. Instructions concerning how to anneal pairs of oligos

Table 1
Sequences of the 64 single-stranded DNA oligonucleotides required to generate 96 unique semi-degenerate barcode adapters

Well Position	Name	Sequence (5'-3')	Well Position	Name	Sequence (5'-3')
A01	1FA	[SEQF]AGTWSMRWSYWKMWW*T	A05	5FE	[SEQF]GTCWSMRWSYRKMKW*T
B01	2FA	[SEQF]CTGWSMRWSYWKMWW*T	B05	6FE	[SEQF]TGAWSMRWSYRKMKW*T
C01	3FA	[SEQF]TCAWSMRWSYWKMWW*T	C05	7FE	[SEQF]ACTWSMRWSYRKMKW*T
D01	4FA	[SEQF]GACWSMRWSYWKMWW*T	D05	8FE	[SEQF]CAGWSMRWSYRKMKW*T
E01	1RA	/5Phos/WWKMWRSWYKSWGTC[SEQR]	E05	5RE	/5Phos/WMKMYRSWYKSWCTG[SEQR]
F01	2RA	/5Phos/WWKMWRSWYKSWTGA[SEQR]	F05	6RE	/5Phos/WMKMYRSWYKSWAGT[SEQR]
G01	3RA	/5Phos/WWKMWRSWYKSWACT[SEQR]	G05	7RE	/5Phos/WMKMYRSWYKSWTCA[SEQR]
H01	4RA	/5Phos/WWKMWRSWYKSWCAG[SEQR]	H05	8RE	/5Phos/WMKMYRSWYKSWGAC[SEQR]
A02	1FB	[SEQF]AGTSWMRSWYKMKYW*T	A06	5FF	[SEQF]GTCSWMRSWYYMKWY*T
B02	2FB	[SEQF]CTGSWMRSWYKMKYW*T	B06	6FF	[SEQF]TGASWMRSWYYMKWY*T
C02	3FB	[SEQF]TCASWMRSWYKMKYW*T	C06	7FF	[SEQF]ACTSWMRSWYYMKWY*T
D02	4FB	[SEQF]GACSWMRSWYKMKYW*T	D06	8FF	[SEQF]CAGSWMRSWYYMKWY*T
E02	1RB	/5Phos/WRMKMRWSYKWSGTC[SEQR]	E06	5RF	/5Phos/RWMKRRWSYKWSCTG[SEQR]
F02	2RB	/5Phos/WRMKMRWSYKWSTGA[SEQR]	F06	6RF	/5Phos/RWMKRRWSYKWSAGT[SEQR]
G02	3RB	/5Phos/WRMKMRWSYKWSACT[SEQR]	G06	7RF	/5Phos/RWMKRRWSYKWSTCA[SEQR]
H02	4RB	/5Phos/WRMKMRWSYKWSCAG[SEQR]	H06	7RF	/5Phos/RWMKRRWSYKWSGAC[SEQR]
A03	1FC	[SEQF]AGTWSKYSWRMKMRW*T	A07	5FG	[SEQF]GTCWSKYSWRWKMWR*T
B03	2FC	[SEQF]CTGWSKYSWRMKMRW*T	B07	6FG	[SEQF]TGAWSKYSWRWKMWR*T
C03	3FC	[SEQF]TCAWSKYSWRMKMRW*T	C07	7FG	[SEQF]ACTWSKYSWRWKMWR*T
D03	4FC	[SEQF]GACWSKYSWRMKMRW*T	D07	8FG	[SEQF]CAGWSKYSWRWKMWR*T
E03	1RC	/5Phos/WYKMKYWSRMSWGTC[SEQR]	E07	5RG	/5Phos/YWKMWYWSRMSWCTG[SEQR]
F03	2RC	/5Phos/WYKMKYWSRMSWTGA[SEQR]	F07	6RG	/5Phos/YWKMWYWSRMSWAGT[SEQR]
G03	3RC	/5Phos/WYKMKYWSRMSWACT[SEQR]	G07	7RG	/5Phos/YWKMWYWSRMSWTCA[SEQR]
H03	4RC	/5Phos/WYKMKYWSRMSWCAG[SEQR]	H07	8RG	/5Phos/YWKMWYWSRMSWGAC[SEQR]
A04	1FD	[SEQF]AGTSWKYWSRWMKMW*T	A08	5FH	[SEQF]GTCSWKYWSRWMKWM*T
B04	2FD	[SEQF]CTGSWKYWSRWMKMW*T	B08	6FH	[SEQF]TGASWKYWSRWMKWM*T
C04	3FD	[SEQF]TCASWKYWSRWMKMW*T	C08	7FH	[SEQF]ACTSWKYWSRWMKWM*T
D04	4FD	[SEQF]GACSWKYWSRWMKMW*T	D08	8FH	[SEQF]CAGSWKYWSRWMKWM*T
E04	1RD	/5Phos/WKMKWYSWRMWSGTC[SEQR]	E08	5RH	/5Phos/KWMKWYSWRMWSCTG[SEQR]
F04	2RD	/5Phos/WKMKWYSWRMWSTGA[SEQR]	F08	6RH	/5Phos/KWMKWYSWRMWSAGT[SEQR]
G04	3RD	/5Phos/WKMKWYSWRMWSACT[SEQR]	G08	7RH	/5Phos/KWMKWYSWRMWSTCA[SEQR]
H04	4RD	/5Phos/WKMKWYSWRMWSCAG[SEQR]	H08	8RH	/5Phos/KWMKWYSWRMWSGAC[SEQR]

Strand-specific tags are colored in orange and red; 12-nucleotide semi-degenerate barcode regions are colored in blue and green

to generate ready-to-use barcoded adapters are provided in Table 2. Low numbers of oligos (i.e., 8, one column) can be ordered in individual tubes. We recommend ordering the 64 oligos needed to generate 96 unique semi-degenerate barcoded adapters in a 96-well plate.

34. TE buffer, 10 mM Tris–HCl (pH = 8.0), 0.1 mM EDTA.
35. Agencourt® AMPure® XP-PCR Purification beads for DNA size selection (Beckman Coulter) (*see* **Note 8**).
36. High-speed mixer for the small and efficient mixing of small volumes.
37. 75% and 80% ethanol.
38. 96-Well ALPAQUA® magnetic plate or similar (e.g., 96S Super Magnet Plate, Alpaqua, Beverly, MA, USA).
39. 10 mM Tris–HCl, pH = 8.0.
40. Dual-indexed PCR primers for Illumina sequencing (*see* **Note 9**).
41. Human Cot-1 DNA (*see* **Note 10**).
42. xGen® Universal Blockers: TS Mix, 96 rxn (Integrated DNA Technologies, Skokie, IL, USA) (*see* **Note 11**).
43. SpeedVac concentrator.
44. Mineral oil, molecular biology grade.
45. Dynabeads™ M-270 Streptavidin (Thermo Fisher Scientific, Waltham, MA, USA).
46. xGen® Hybridization and Wash Kit, 96 rxn (Integrated DNA Technologies, Skokie, IL, USA).
47. Biotinylated DNA baits such as xGen® Lockdown probes and/or predesigned xGen®capture pools or panels (Integrated DNA Technologies, Skokie, IL, USA; http://www.idtdna.com/pages/products/next-generation-sequencing/hybridization-capture/custom-probes-panels/xgen-gene-capture-pools).
48. Premixed P5/P7 Illumina primers (10 μM final concentrations each, *see* **Note 12**).
49. KAPA HiFi HotStart ReadyMix PCR kits (Roche, Basel, Switzerland).
50. Nuclease-free water.
51. M13-tailed primers for the custom generation of biotinylated baits (optional).
52. Human genomic DNA for the custom generation of biotinylated DNA baits (optional).

Table 2
Annealing scheme of pairs of oligonucleotides for the generation of 96 semi-degenerate barcode adapters and how they can be distributed in a 96-well plate

	1	2	3	4	5	6	7	8	9	10	11	12
A	A1-E1	B1-F1	C1-G1	D1-H1	A1-F1	A1-H1	B1-E1	B1-G1	C1-E1	C1-H1	D1-F1	D1-G1
B	B2-F2	A2-E2	D2-H2	C2-G2	B2-E2	B2-G2	A2-F2	A2-H2	D2-F2	D2-G2	C2-E2	C2-H2
C	C3-G3	D3-H3	A3-E3	B3-F3	C3-H3	C3-E3	D3-G3	D3-F3	A3-H3	A3-F3	B3-G3	B3-E3
D	D4-H4	C4-G4	B4-F4	A4-E4	D4-G4	D4-F4	C4-H4	C4-E4	B4-G4	B4-E4	A4-H4	A4-F4
E	A5-E5	B5-F5	C5-G5	D5-H5	A5-F5	A5-G5	B5-E5	B5-H5	C5-E5	C5-H5	D5-F5	D5-G5
F	B6-F6	A6-E6	D6-H6	C6-G6	B6-E6	B6-H6	C6-H6	C6-E6	D6-F6	D6-G6	A6-G6	A6-F6
G	C7-G7	D7-H7	A7-E7	B7-F7	C7-H7	C7-E7	D7-G7	D7-F7	A7-G7	A7-F7	B7-E7	B7-H7
H	D8-H8	C8-G8	B8-F8	A8-E8	D8-G8	D8-F8	A8-F8	A8-G8	B8-H8	B8-E8	C8-H8	C8-E8

For each individual well, the position of each oligo comprising a given barcoded adapter refers to the well positions shown in Table 1

53. Universal primers for the custom generation of biotinylated DNA baits (optional; 5′-biotin-GTTTTCCCAGTCACGAC-3′ and 5′-P-CAGGAAACAGCTATGAC-3′).
54. Lambda exonuclease and Lambda exonuclease reaction buffer for the custom generation of biotinylated DNA baits (optional).
55. Illumina automated sequencers (i.e., Miseq, Hiseq, NextSeq, MiniSeq; Illumina, San Diego, CA, USA).
56. Geneious genomics workbench (Biomatters, Ltd.; https://www.geneious.com/) (optional) for the analysis of NGS data.
57. Dellingr (https://github.com/morinlab/Dellingr) (optional), as a more computationally efficient alternative to Geneious for the analysis of NGS data.

3 Methods

3.1 Blood Processing and Plasma Collection

1. Draw blood directly into cell-free DNA BCT© (Streck) or EDTA blood collection tubes. Label tubes appropriately and store them according to the type of tube used (*see* **Note 1**).
2. In an area designated for the manipulation of biohazardous samples, centrifuge the blood collection tubes at 820–2800 × *g* for 15 min at room temperature (*see* **Notes 2** and **13**).
3. Using a sterile disposable transfer pipette, aspirate the upper phase (plasma) and aliquot it into pre-labeled cryovial tubes. Discard transfer pipette into the biohazard waste. Avoid disturbing the buffy coat during the process.
4. Place cryovials inside cryoboxes and store them at −80 °C until cfDNA extraction.

3.2 Buffy Coat Processing (Optional, See Note 14)

1. Aliquot 8 mL of RBC lysis buffer into pre-labeled 15 mL Falcon tubes.
2. Use a clean disposable serological pipette to transfer the buffy coat layer from a couple of blood collection tubes drawn from the same patient into the 15 mL Falcon tubes containing RBC lysis buffer. Mix the sample by inverting the Falcon tube several times.
3. Incubate the samples during 10 min at room temperature.
4. Centrifuge the Falcon tubes at 820–2800 × *g* for 10 min.
5. Remove and discard the supernatant into a plastic container with 10% bleach.
6. Add 1 mL of RBC lysis buffer using a disposable transfer pipette.

7. Pipette up and down to mix and transfer the mixture to a labeled 1.5 mL Eppendorf tube.
8. Centrifuge samples for 1 min at 12,000–16,000 × *g*.
9. Remove and discard the supernatant into a plastic container with 10% bleach.
10. Add 1 mL of PBS and pipette up and down to mix.
11. Centrifuge samples for 1 min at 12,000–16,000 × *g*.
12. Remove and discard the supernatant into a plastic container with 10% bleach.
13. Add 1 mL of freezing media to each sample (*see* **Note 15**).
14. Pipette up and down to mix and transfer the sample into a pre-labeled cryovial tube.
15. Place cryovials inside cryoboxes and store them at −80 °C.

3.3 cfDNA Extraction and QC

1. Thaw cryovials on ice (*see* **Note 16**).
2. Centrifuge cryovials at 12,000–16,000 × *g* during 10 min at 4 °C (optional; *see* **Note 2**) and transfer the supernatant to a clean 1.5 mL Eppendorf tube or 15 mL Falcon tube (depending on the plasma volume to be processed).
3. Follow the manufacturer's protocol associated with each particular cfDNA isolation kit (*see* **Notes 4** and **17**).
4. Elute cfDNA using the appropriate elution buffer provided by the manufacturer (*see* **Note 18**) and keep the cfDNA extract at 4 °C for short-term storage (i.e., 24 h) or at −20 °C for long-term storage (*see* **Note 19**).
5. Measure cfDNA concentration by replicate using a Qubit dsDNA High Sensitivity assay kit (*see* **Note 20**).
6. Run an Agilent High Sensitivity DNA chip (optional) to inspect cfDNA size distribution (*see* **Note 21**).

3.4 Preparation of Semi-Degenerate Barcoded Adapters

1. We recommend ordering batches of at least sixteen oligonucleotides for barcoded adapter generation as shown in Table 2 (i.e., A01 to H01, one column). The rationale is to provide enough sequence diversity for the strand-specific tags and the semi-degenerate barcode sequences. For example, sequencing libraries constructed from a single adapter construct alone in a flow cell can cause low sequence diversity and low-cluster passing filter, both of which can negatively impact the quality and quantity of NGS reads.
2. Resuspend oligos to 100 μM using 10 mM Tris–HCl and 0.1 mM EDTA, pH = 8.0.
3. Oligonucleotides should be paired as detailed in Table 2. One column of 8 oligonucleotides as shown in Table 1 (i.e. column

Table 3
Annealing of two semi-degenerate oligos carrying strand-specific tags in a final volume of 25 μL to generate barcoded adapters

Reagent	Volume (μL)
Oligo 1 (100 μM stock)	10
Oligo 2 (100 μM stock)	10
10× stringent wash buffer*	2.5
Nuclease-free water	2.5

*Provided as part of the xGen® Hybridization and Wash Kit.

Table 4
Thermal cycling program for the annealing of oligonucleotides carrying strand-specific tags and semi-generate barcode sequences to generate ready-to-use barcoded adapters

Cycle step	*T*	Time (min)	Cycles
Initial denaturation	98 °C	5	1
Annealing	−0.1 °C/s		1
Final annealing	25 °C	60	1
Hold	4 °C		

A01 to H01) can generate 12 unique semi-degenerate barcoded adapters using various combinations of individual oligos (*see* **Note 22**). Ordering 64 unique oligonucleotides (*see* Table 1) permits generating 96 unique semi-degenerate barcoded adapters and these can be dispensed in individual wells of a 96-well plate (Table 2).

4. Dispense pairs of oligos in low-binding 0.2 mL PCR tubes/strips or in a 96-well plate according to the instructions provided in Table 2 and following the volumes and concentrations shown in Table 3.
5. Anneal each pair of oligos by starting the program shown in Table 4 in a thermal cycler with the lid preheated at 105 °C. Semi-degenerate barcode adapters are ready to use at this point and can be stored at −20 °C for at least 6 months (*see* **Note 23**).
6. Adapter QC: The concentration of annealed adapters should be estimated by performing a Qubit™ High Sensitivity dsDNA assay. Annotate adapter concentration and annealing date accordingly. The quality of annealed adapters can be

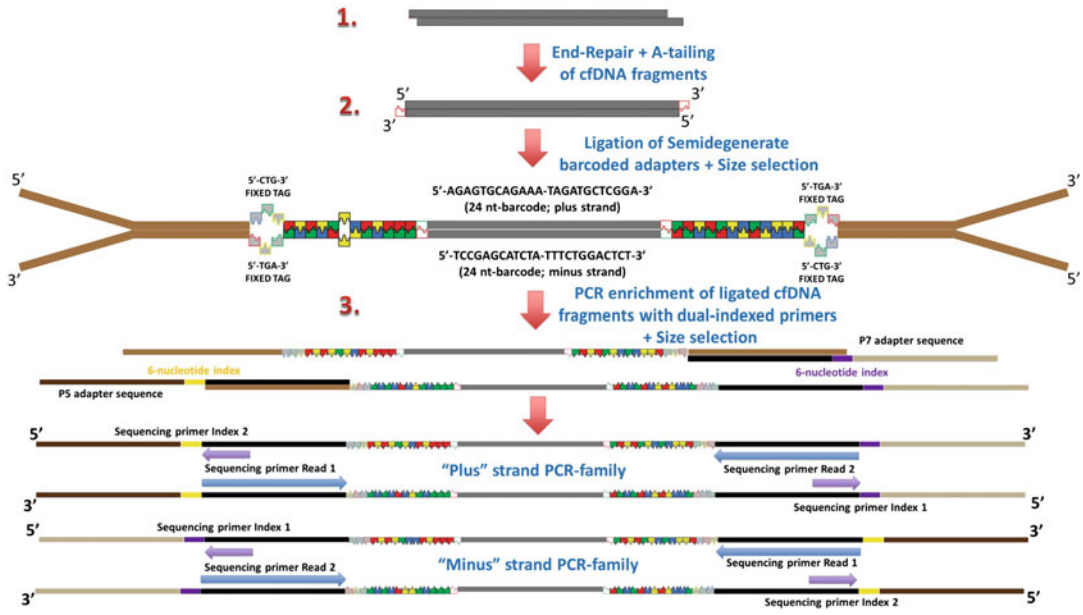


Fig. 1 cfDNA library construction workflow. First, the ends of cfDNA fragments are enzymatically treated with a cocktail of exonucleases and DNA polymerases to facilitate adapter ligation (*1*). Semi-degenerate barcoded adapters are subsequently ligated into end-repaired and A-tailed cfDNA fragments (*2*). The unique molecule identifiers (UMIs) and strand-specific motifs contained within our adapter constructs permit tagging each of the two strands of a given double-stranded cfDNA molecule. Ligated cfDNA fragments are then denatured and replicated using dual-indexed PCR primers. The incorporation of the DNA sequences provided by our dual-indexed PCR primers permits tracking the individual samples associated with a certain population of cfDNA fragments and their subsequent analysis and de-multiplexing in Illumina sequencers (*3*)

investigated by running an Agilent High-Sensitivity DNA chip on the 2100 Bioanalyzer instrument or similar (*see* Fig. 1).

3.5 Library Construction with the NEBNext Ultra™ II DNA Library Prep Kit for Illumina (See Fig. 2)

1. End-repair and A-tailing of cfDNA fragments: In low-binding 0.2 mL PCR tubes/strips or 96-well PCR plates, combine the components listed in Table 5 (*see* **Note 24**).
2. Pipette up and down at least ten times with the pipette set at 50 μL to mix thoroughly. Collect all liquid from the sides of the tubes by performing a quick spin using a microcentrifuge.
3. Place tubes/plates in a thermal cycler with the lid preheated at 75 °C and start the program shown in Table 6.
4. Proceed to the adapter ligation step (*see* **Note 25**). To each tube/well, add the components indicated in Table 7. The final volume of the reaction should be now 98 μL (60 μL + 38 μL). Add first the barcoded adapters (use a multichannel pipette if incorporating multiple adapters) and then dispense the ligation master mix + ligation enhancer premix (*see* **Notes 23** and **26**).

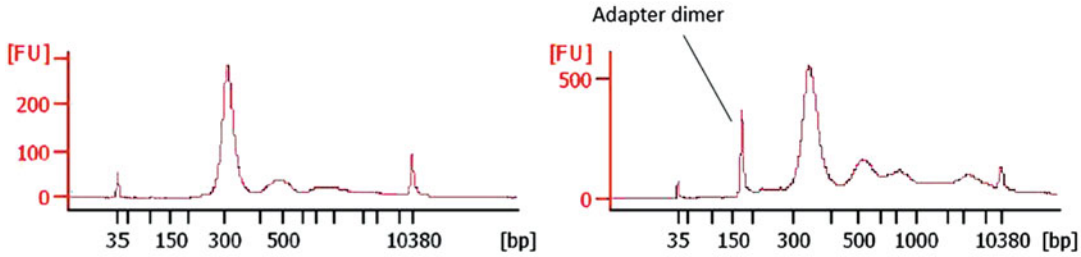


Fig. 2 BioAnalyzer 2100 profile of an optimal cfDNA library (left) and a suboptimal cfDNA library (right) showing a considerable adapter dimer peak that requires further size selection. Sequencing experiments revealed very high ctDNA levels (>50%) in the left library but very low ctDNA levels in the right library (<1%). It is common to observe that cfDNA library peak size is smaller in cases with higher abundance of ctDNA

Table 5
Volumes and composition of the end-repair/A-tailing reaction to prepare cfDNA fragments for adapter ligation

Reagent	Volume (μL)
NEBNext Ultra II End Prep Reaction Buffer	7
NEBNext Ultra II End Prep Enzyme Mix	3
cfDNA extract (*see* **Note 18**)	50
	60

Table 6
End-repair plus A-tailing thermocycling program

Cycle step	*T* (°C)	Time (min)	Cycles
End repair	20	30	1
A-tailing	65	30	1
Hold	4		

Table 7
Reagents and volumes for the ligation reaction that incorporates barcoded adapters into individual cfDNA fragments

Reagent	Volume (μL)
NEBNext Ultra II Ligation Master Mix	32
NEBNext Ligation Enhancer	1
Semi-degenerate barcoded adapters	5
	38

Table 8
Thermocycling program for the ligation reaction performed during library construction

Cycle step	*T* (°C)	Time (min)	Cycles
Ligation	20	15	1
Ligation	16	≥45	1
Hold	4		

5. Pipette up and down at least ten times with the pipette set at 80 μL to mix thoroughly and collect all liquid from the sides of the tubes/plates by performing a quick spin using a microcentrifuge (*see* **Note 27**).
6. Place tubes/plates in a thermal cycler with the preheated lid off and start the program shown in Table 8 (*see* **Note 28**).
7. Proceed to the size selection step and cleanup of cfDNA-ligated fragments (*see* **Note 29**). Add 78 μL of AMPure XP magnetic beads (~0.8× volumes, *see* **Note 8**) to each sample and mix vigorously by pipetting up and down.
8. Incubate in a high-speed mixer (1800 rpm) for 5–10 min at room temperature (*see* **Note 30**) and collect all liquid from the sides of the tubes by performing a quick spin using a microcentrifuge.
9. Place tubes or plates on an ALPAQUA® magnetic plate or on an appropriate magnetic stand and allow beads to bind to the walls of the tubes (~2 min).
10. Using a pipette, carefully remove the clear supernatant without disturbing/removing the beads.
11. Fully resuspend the beads in 75% ethanol by pipetting up and down for at least 30 s and place tubes/plates back in the magnetic separator after a quick spin to collect all liquid from the sides of the tubes (*see* **Note 31**).
12. Using a pipette, carefully remove the clear supernatant without disturbing/removing the beads.
13. Fully resuspend the beads in 75% ethanol by pipetting up and down for at least 30 s and place tubes/plates back in the magnetic separator after a quick spin to collect all liquid from the sides of the tubes (*see* **Note 31**).
14. Using a pipette, carefully remove the clear supernatant and small remaining ethanol drops. Let the samples dry (caps/plate open) at room temperature (*see* **Note 32**).

15. Resuspend the beads in 25 μL of 10 mM Tris–HCl, pH = 8.0, and incubate in the high-speed mixer at 2000 rpm for at least 2 min to elute the DNA. Collect all liquid from the sides of the tubes by performing a quick spin using a microcentrifuge (*see* **Note 33**).
16. Place the samples in an appropriate magnetic separator and transfer the clean supernatant into new, pre-labeled low-binding 0.2 mL PCR tubes/strips or 96-well PCR plates (*see* **Note 34**).
17. To each individual sample, add 5 μL of premixed dual-indexed PCR primers (Tables 16, 17, and 18) and 30 μL of the NEBNext Ultra II Q5 Master Mix (*see* **Note 35**). The final volume of the PCR reaction should be 60 μL. Keep track of the dual-index combination used for each sample for de-multiplexing purposes later on.
18. Mix vigorously by pipetting up and down at least ten times with the pipette set at 50 μL. Spin down liquids in a microcentrifuge.
19. Run the program shown in Table 9 using a thermocycler with the lid preheated at 105 °C (*see* **Notes 36** and **37**).
20. Proceed to the size selection and cleanup of amplified library fragments. Add 48 μL of AMPure magnetic beads (~0.8× volumes) to each sample and mix vigorously by pipetting up and down (*see* **Note 38**).
21. Incubate in a high-speed mixer for 5–10 min at room temperature. Collect all liquid from the sides of the tubes by performing a quick spin using a microcentrifuge.
22. Place tubes or plates on an ALPAQUA® magnetic plate or on an appropriate magnetic stand and allow beads to bind to the walls of the tube (~ 2 min).
23. Using a pipette, carefully remove the clear supernatant without disturbing/removing the beads (*see* **Note 34**).

Table 9
Thermocycling program for the PCR enrichment of ligated cfDNA fragments

Cycle step	*T* (°C)	Time	Cycles
Initial denaturation	98	30 s	1
Denaturation	98	10 s	8 (*see* **Note 37**)
Annealing/extension	65	90 s	
Final extension	65	5 min	1
Hold	4		

24. Fully resuspend the beads in 75% ethanol by pipetting up and down for at least 30 s and place tubes/plates back in the magnetic separator after a quick spin to collect all liquid from the sides of the tubes.
25. Using a pipette, carefully remove the clear supernatant without disturbing/removing the beads (*see* **Note 34**).
26. Fully resuspend the beads in 75% ethanol pipetting up and down for at least 30 s and place tubes/plates back in the magnetic separator after a quick spin to collect all liquid from the sides of the tubes.
27. Using a pipette, carefully remove the clear supernatant and small remaining ethanol drops. Let the samples dry (caps/plate open) at room temperature (*see* **Note 32**).
28. Resuspend the beads in 30 μL of 10 mM Tris–HCl pH = 8.0 and incubate in the high-speed mixer at 2000 rpm for at least 2 min to elute the DNA. Collect all liquid from the sides of the tubes by performing a quick spin using a microcentrifuge (*see* **Note 33**).
29. Place the tubes/plates in an appropriate magnetic separator and transfer the clean supernatant into new, pre-labeled low-binding 0.2 mL PCR tubes/strips or 96-well PCR plates (*see* **Note 34**).
30. Measure cfDNA library yields by performing Qubit High-Sensitivity dsDNA assays (*see* **Note 39**).
31. (Optional) Run an Agilent High Sensitivity DNA chip or similar to visually inspect library quality. Carefully check for size distribution and presence and quantity of adapter dimers (*see* **Note 40**; *see also* Fig. 3).

3.6 Target Enrichment via Hybridization Capture (See Note 41; Fig. 4)

1. Set up custom pools of biotinylated DNA baits, predesigned gene capture pools, or panels before starting the experiment (*see* **Note 42** for the handling of commercial XGen® Lockdown® biotinylated DNA baits and **Note 43** for the in-house generation of biotinylated DNA baits). Here, we can use small personalized pools of biotinylated probes to track ctDNA levels in a given patient [46, 47] or design custom pools of probes targeting genomic regions recurrently mutated in lymphoma [40, 41, 51, 52].
2. Add the components listed in Table 10 in a MAXYMum Recovery® 1.7 mL Microtube or similar. Equimolar amounts of different libraries should be added into library pools (at least 250 ng of DNA per library) or use at least 250 ng if using one single cfDNA library (*see* **Note 44**).
3. Completely dry down the mixture using a SpeedVac concentrator (*see* **Note 45**).

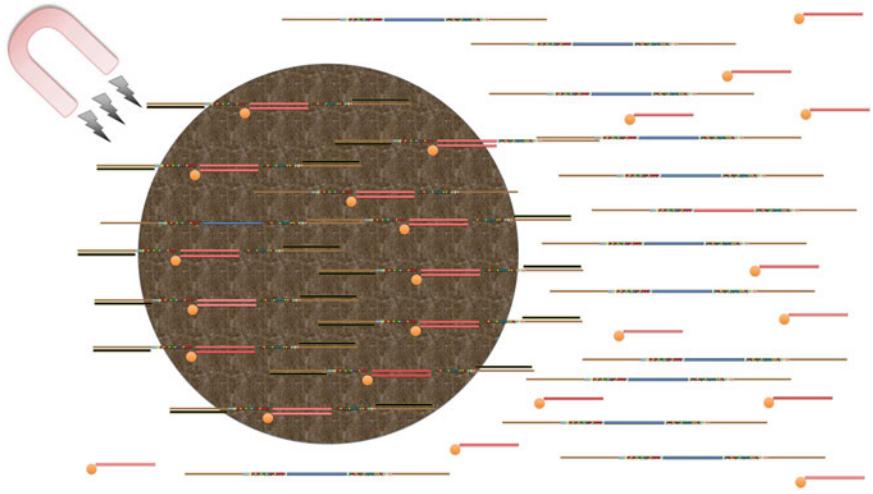


Fig. 3 Fundamentals of targeted enrichment using biotinylated DNA baits. Library fragments are denatured and then hybridized with short, single-stranded biotinylated DNA fragments (red bars with orange circle) complementary to the sequences of interest. These probes can form heteroduplexes with fully and semi-complementary library fragments and these hybrids are physically separated from the rest of the library using magnetic streptavidin beads (large brown circle) that strongly bind to the biotin residues incorporated in the 5′-end of the DNA bait. Library fragments carrying or lacking DNA sequences complementary to those contained within the DNA baits are colored in red or blue, respectively. Blocker oligonucleotides that minimize nonspecific hybridization between the complementary ends of adapters are represented by black rectangles

4. Resuspend the DNA in the hybridization master mix described in Table 11 (*see* **Notes 46** and **47**).
5. Incubate the resuspended samples in the high-speed mixer (2000 rpm) for at least 10 min at room temperature. Pipette up and down the mix several times and spin down the contents of the tubes/plate using a microcentrifuge (*see* **Note 48**).
6. Transfer the resuspended DNA into a new, pre-labeled low-binding 0.2 mL PCR tube.
7. Place the tube inside a thermal cycler with the lid preheated at 75 °C and execute the program listed in Table 12 (*see* **Note 49**).
8. Add 4 μL of custom pools of xGen® Lockdown probes (0.75 pmol/μL) or xGen® predesigned gene capture pools or panels immediately after the initial denaturation step (*see* **Notes 42** and **50**).
9. Briefly vortex and spin down the tube/s and put them back inside the thermal cycler. Close the thermocycler lid and incubate at 65 °C during at least 4 h (*see* **Note 51**).
10. Dilute the buffers supplied as part of the xGen® Hybridization and Wash Kit to prepare 1× stock solutions according to the instructions provided in Table 13 (*see* **Note 52**).

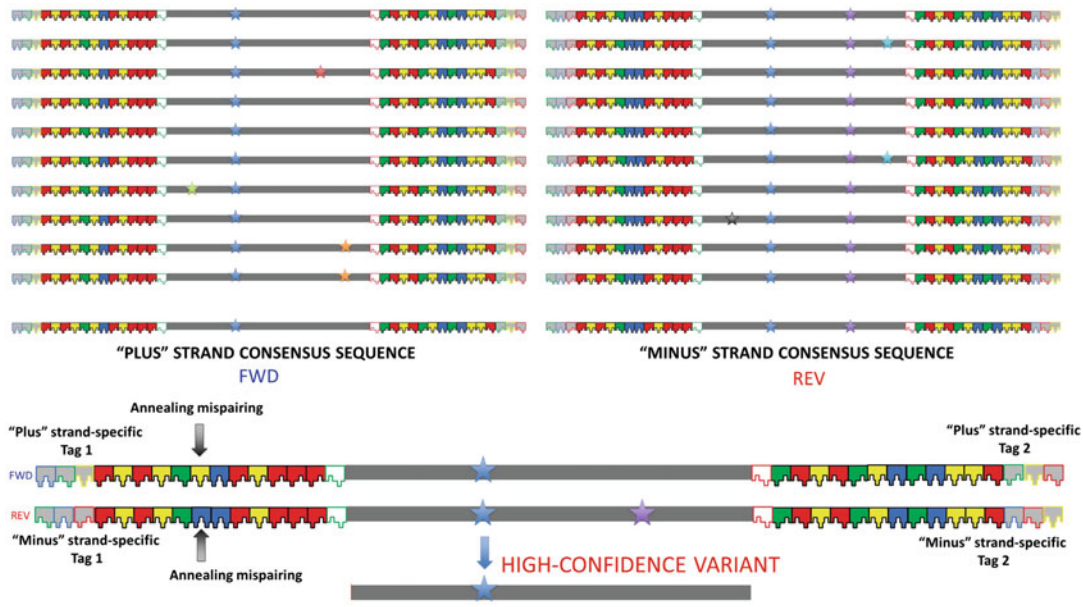


Fig. 4 Fundamentals of error correction in data from deeply sequenced barcoded libraries. Independent PCR families for the two strands of every single cfDNA molecule are generated during the process of library preparation. The members of every single PCR family, which can be tracked owing to the incorporation of unique molecule identifiers, are collapsed into single consensus sequences. Two given strands are considered to be derived from the same cfDNA molecule if their consensus sequences share the same mapping coordinates, exhibit alternate strand-specific tags, and do not differ in more than three mismatches in the two 6-nucleotide semi-degenerate barcode that flank the ligation site. Only those variants independently supported by the consensus sequence of the two strands are considered as high-confidence variants

Table 10
Libraries built from cfDNA samples are pooled with Human Cot-1 DNA and xGen® Universal Blockers during the first step of the targeted hybridization capture experiment

Reagent	Volume (μL)
Human Cot-1 DNA	5
xGen® Universal Blockers: TS Mix	2
cfDNA library/libraries	Variable

11. Start preparing the M-270 streptavidin beads at least 1 h before the end of the hybridization reaction. Streptavidin beads need to be equilibrated at room temperature during at least 30 min prior to use. Pipette 75 μL of streptavidin beads (per capture) and transfer them to a new, 0.2 mL PCR tube (*see* **Note 53**).
12. Place the tube/s on the ALPAQUA® magnetic plate or on an appropriate magnetic stand and allow the beads to bind to the walls of the tube.

Table 11
Composition of the hybridization buffer master mix in which cfDNA libraries, Human Cot-1 DNA, and xGen® Universal Blockers are resuspended before DNA denaturation and hybridization with biotinylated DNA baits

Reagent	Volume (μL)
xGen 2× hybridization buffer	8.5
xGen hybridization buffer enhancer	2.7
Nuclease-free water	1.8
	13

Table 12
Denaturation and hybridization times and temperatures during library enrichment experiments

Cycle step	*T* (°C)	Time
Initial denaturation	95	10 min
Hybridization	65	At least 4 h

Table 13
Dilution scheme for the preparation of working solutions of the different types of wash buffers used during hybridization capture experiments

Component	Buffer (μL)	Nuclease-free water (μL)	Total (μL)	Storage
xGen 2× bead wash buffer	750	750	1500	Room T
xGen 10× wash buffer I	120	1080	1200	Preheat 100 μL/capture at 65 °C
				Keep remaining solution at room T
xGen 10× wash buffer II	160	1440	1600	Room T
xGen 10× wash buffer III	160	1440	1600	Room T
xGen 10× stringent wash buffer	160	1440	1600	Preheat at 65 °C

13. Using a pipette, carefully remove the clear supernatant.
14. Fully resuspend the beads in 200 μL of 1× bead wash buffer by pipetting up and down and put the tube/s back on the magnetic separator. Repeat **steps 13** and **14**.
15. Using a pipette, carefully remove the clear supernatant.

16. Fully resuspend the beads in 100 μL of 1× bead wash buffer and transfer the mixture to a new, low-binding 0.2 mL PCR tube.
17. Place the tube on the ALPAQUA® magnetic plate or on an appropriate magnetic stand and allow the beads to bind to the walls of the tube.
18. Using a pipette, carefully remove the clear supernatant.
19. Immediately transfer the entire volume of the incubating hybridization reaction (17 μL) into the tube containing the dried streptavidin beads (*see* **Note 54**).
20. Vortex thoroughly to allow the total resuspension of the beads and perform a quick spin to collect all liquid from the walls of the tube.
21. Incubate the mixture at 65 °C during 45 min in a thermal cycler with the lid preheated at 75 °C, thoroughly mixing and spinning down the beads every 15 min.
22. Add 100 μL of preheated wash buffer I (*see* **Note 55**) and mix thoroughly by pipetting up and down several times while the tube is still inside the thermocycler. Minimize the formation of air bubbles when possible. Perform a quick spin to collect all the liquid from the walls of the tube.
23. Quickly place tube/s on an ALPAQUA® magnetic plate or on an appropriate magnetic stand and allow beads to bind to the walls of the tube.
24. Using a pipette, carefully remove the clear supernatant.
25. Fully resuspend the beads in 200 μL of preheated 1× stringent wash buffer, being careful not to introduce air bubbles.
26. Incubate resuspended beads at 65 °C for 5 min in a digital heater or water bath.
27. After performing a quick spin, place tubes on an ALPAQUA® magnetic plate or on an appropriate magnetic stand and allow beads to bind to the walls of the tube.
28. Using a pipette, carefully remove the clear supernatant.
29. Repeat **steps 25–28**.
30. Fully resuspend the beads in 200 μL of 1× wash buffer I equilibrated at room temperature.
31. Thoroughly mix beads and 1× wash buffer I in a high-speed mixer (2000 rpm) during 2 min.
32. After performing a quick spin, place tubes on an ALPAQUA® magnetic plate or on an appropriate magnetic stand and allow beads to bind to the walls of the tube.
33. Using a pipette, carefully remove the clear supernatant.

34. Fully resuspend the beads in 200 μL of 1× wash buffer II equilibrated at room temperature.
35. Thoroughly mix beads and 1× wash buffer II in a high-speed mixer (2000 rpm) during 1 min.
36. After performing a quick spin, place tubes on an ALPAQUA® magnetic plate or on an appropriate magnetic stand and allow beads to bind to the walls of the tube.
37. Using a pipette, carefully remove the clear supernatant.
38. Fully resuspend the beads in 200 μL of 1× wash buffer III equilibrated at room temperature.
39. Thoroughly mix beads and 1× wash buffer III in a high-speed mixer (2000 rpm) during 30 s.
40. After performing a quick spin, place tubes on an ALPAQUA® magnetic plate or on an appropriate magnetic stand and allow beads to bind to the walls of the tube.
41. Using a pipette, carefully remove the clear supernatant.
42. Thoroughly resuspend the beads in the solution described in Table 14 (do not discard beads).
43. Perform a quick spin and place the tube containing the beads inside a thermal cycler to start the program listed in Table 15 (*see* **Note 56**).
44. After completion, add 56 μL (~0.8× volumes) of AMPure XP beads equilibrated at room temperature to each post-capture PCR reaction and mix thoroughly by pipetting up and down several times.
45. Incubate samples in a high-speed mixer (1800 rpm) during 5–10 min.
46. Place tubes on an ALPAQUA® magnetic plate or on an appropriate magnetic stand and allow beads to bind the beads to the walls of the tube.
47. Using a pipette, carefully remove the clear supernatant.

Table 14
Dynabeads M-270 carrying biotinylated DNA bait-library fragment hybrids need to be resuspended in a solution containing the following reagents and PCR primers

Reagent	Volume (μL)
Ultrapure nuclease-free water	28
P5/P7 Illumina primers (10 μM)	7
KAPA HiFi HotStart ReadyMix	35
	70

Table 15
Thermocycling program corresponding to the post-capture PCR enrichment of cfDNA libraries

Cycle step	*T* (°C)	Time	Cycles
Initial denaturation	98	45 s	1
Denaturation	98	15 s	12 (*see* **Note 56**)
Annealing	60	30 s	
Extension	72	2 min	
Final extension	72	4 min	1
Hold	4		

48. Fully resuspend the beads in 75% ethanol and place tubes back in the magnetic separator.
49. Using a pipette, carefully remove the clear supernatant.
50. Fully resuspend the beads in 75% ethanol and place tubes back in the magnetic separator.
51. Using a pipette, carefully remove the clear supernatant and remaining ethanol drops. Let the samples dry (caps open) at room temperature. Avoid overdrying the beads.
52. Resuspend the beads in 30 μL of 10 mM Tris–HCl pH = 8.0 and incubate in the high-speed mixer at 2000 rpm for at least 2 min to elute the DNA.
53. Place the tube/s in an appropriate magnetic separator and transfer the clean supernatant into a new, clean, and pre-labeled low-binding 1.5–1.7 mL PCR tube. Discard beads at this point.
54. Measure enriched cfDNA library yields by performing Qubit High-Sensitivity dsDNA assays.
55. (Optional) Run an Agilent High Sensitivity DNA chip or similar to visually inspect library quality. Carry out a second size selection using AMPure XP beads if a substantial amount of adapter dimers is observed.
56. Proceed with the sequencing of enriched cfDNA libraries or perform a second round of hybridization capture over the same library pool (repeat **steps 1–55**; *see* **Note 57**).

3.7 Sequencing of cfDNA Libraries (See Fig. 5)

1. Follow Illumina's instructions for the sequencing of enriched cfDNA libraries using the MiSeq, MiniSeq, HiSeq, NextSeq, or similar automated instruments.
2. Provide sample sheet with the appropriate dual-index information for sample de-multiplexing (Table 16).

Fig. 5 Alignment of Illumina NGS reads derived from targeted enrichment experiments reveals significant spikes in coverage in the exons (gray annotations) of targeted genes (green annotations). The annotation of these regions (**a**, yellow annotations) and their further extraction and concatenation generate simplified references to facilitate the analysis and storage of Illumina sequencing data in Geneious (**b**)

Table 16
Index sequence information for sample de-multiplexing in Illumina sequencers

	i7	i5 (Miseq, HiSeq2500)	i5 (R) (NextSeq)
25	GCTATA	TATAGC	GCTATA
26	CTCTAT	ATAGAG	CTCTAT
27	GATAGG	CCTATC	GATAGG
28	AGAGCC	GGCTCT	AGAGCC
29	TCGCCT	AGGCGA	TCGCCT
30	AGATTA	TAATCT	AGATTA
31	GTCCTG	CAGGAC	GTCCTG
32	CAGTAC	GTACTG	CAGTAC
33	AGTAAT	ATTACT	AGTAAT
34	TCCGGA	TCCGGA	TCCGGA
35	TGAGCG	CGCTCA	TGAGCG
36	AATCTC	GAGATT	AATCTC
37	CTGAAT	ATTCAG	CTGAAT
38	GAATTC	GAATTC	GAATTC
39	CTTCAG	CTGAAG	CTTCAG
40	GCATTA	TAATGC	GCATTA

3.8 Bioinformatics Analysis of NGS Reads Using Geneious (Optional; See Note 58)

1. Import the read 1 and read 2 FastQ files associated to each sample and generated during Subheading 3.7 into Geneious (Go to File > Import > From File).
2. Set the two sequence lists (_R1 and _R2) associated to each sample as paired reads. Go to Sequence > Set Paired Reads. Set expected distance/insert size to 300 bp.
3. Set up reference (*see* **Note 59**) by mapping the paired list of Illumina reads to the human genome reference. Go to Align/Assemble > Map to Reference and follow the parameter settings shown in **Note 60**.
4. Annotate regions of high coverage in the human genome by selecting Annotate & Predict > Find Low/High Coverage according to the parameter settings shown in **Note 60**.
5. Extract and concatenate within each chromosome the regions of high coverage by following the instructions displayed in **Note 60**. Go to Tools > Extract Annotations (*see* Fig. 6).
6. Download and install (just drag files into Geneious) the workflows from the supplemental material associated with our prior publication (https://www.nature.com/articles/s41598-017-10269-2).

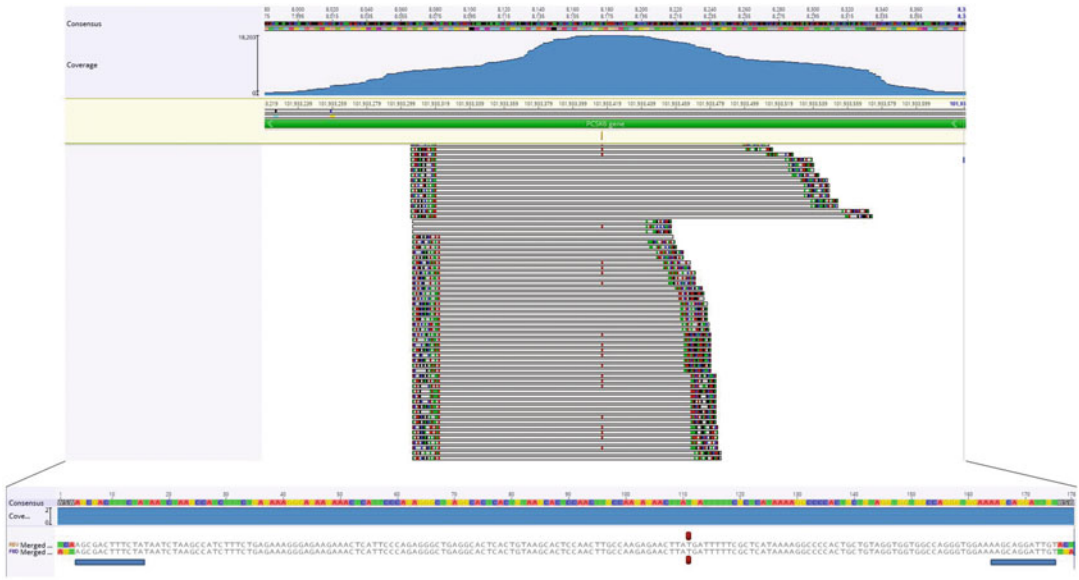


Fig. 6 Alignment of the consensus sequences of individual DNA strands against reference sequence in Geneious. This figure shows consensus sequences with colored molecular barcode sequences in the ends of each molecule (for paired reads with overlapping ends) and strong support for ctDNA (in red) for the position annotated in yellow in the reference sequence. Multiple individual DNA molecules support that variant call, and many of them can be reconstructed into duplexes (as shown in the example duplex at the bottom). The consensus sequences of these two strands differ in their strand-specific strand but share an identical 24 nt molecular barcode (blue bars) and start/end mapping coordinates

Fig. 7 Mapping coordinate collision of 12 unique DNA duplexes. This figure shows the DNA sequences of the consensus sequences derived from 26 individual DNA strands. Twenty-four of these strands can be reconstructed into 12 unique DNA duplexes sharing the same start/end mapping coordinates. The strand-specific tag and the first six nucleotides of each semi-degenerate barcode sequences have been soft-trimmed. Within each duplex, the 12-nucleotide semi-degenerate barcode preceding the ligation site in both ends of each strand can differ due to annealing mismatches, polymerase and sequencing errors but this does not exceed one nucleotide difference in this example

7. Select your list of paired sequences and reference sequence and execute the cfDNA-GenWkf1 workflow by clicking on the Workflow menu. This workflow is well suited for the analysis of small, personalized captures targeting a small number of genomic positions. Alternatively, execute the cfDNA-GenWkf1lp to analyze cfDNA libraries enriched with large panels with broad patient coverage (*see* **Note 61**).
8. After completing the cfDNA-GenWkf1 workflow, select the two new lists containing consensus sequences and your reference and execute the workflow cfDNA-GenWkf3; after completing the cfDNA-GenWkf1lp workflow, select the two new lists containing consensus sequences and your reference and execute the workflow cfDNA-GenWkf3lp (*see* **Note 62**).
9. Inspect the variant calls contained within the Variants annotation track in the resulting assemblies and search for variants with duplex support (*see* **Note 63**, Figs. 7 and 8).
10. Annotate filtered variants (optional) using Ensembl Variant Effect Predictor (e.g., https://uswest.ensembl.org/info/docs/tools/vep/index.html) or Oncotator (http://portals.broadinstitute.org/oncotator/) (optional) (*see* **Note 64**).

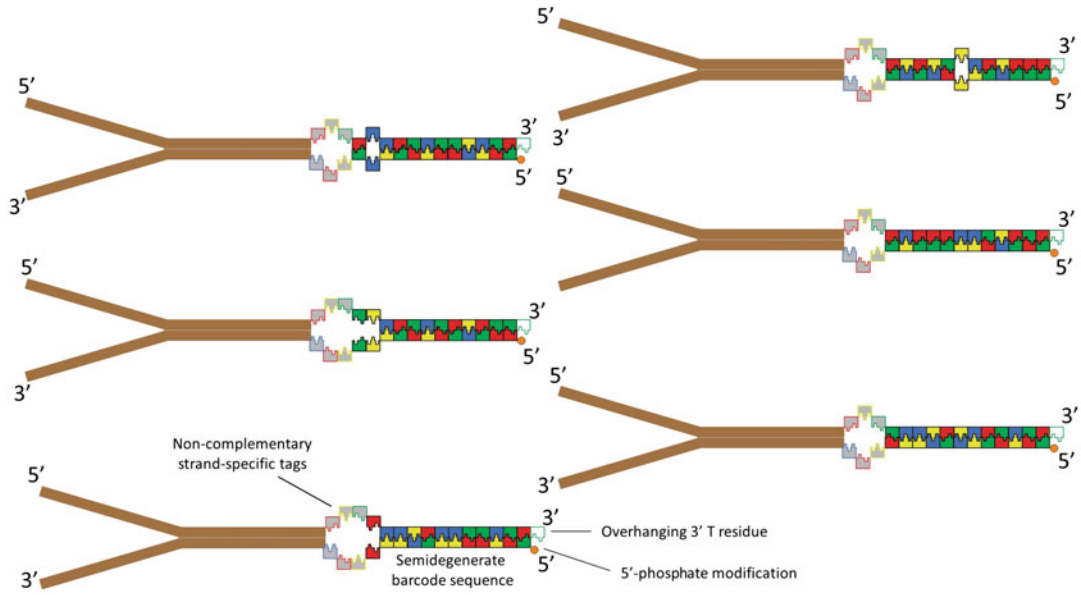


Fig. 8 Structure of semi-degenerate barcoded adapters. Two oligos are annealed to generate Y-shape adapters with overhanging 3′ thymine residues and 5′ phosphate moieties. The two oligoes carry noncomplementary strand-specific tags that form an unpaired "bubble" in the adapter construct adjacent to a stretch of semi-degenerate barcode sequences with potential complementarity. Each color in this figure represents a different nucleotide base (red = adenine; blue = cytosine; yellow = guanine; green = thymine). Two oligos can anneal with perfect complementarity during annealing but annealing mismatches (indicated by the presence of the same base in the same position of the two adapter strands) are more common. We recently demonstrated that these annealing mismatches are less common within the 6-nucleotide residues preceding the ligation site [46]. Our barcode constructs allow $2^{24} = 16{,}777{,}216$ combinations to tag individual DNA strands. We encourage to restrict the barcode information to the last six nucleotide residues preceding the ligation site, which still allows $2^{12} = 4096$ combinations, in order to match up the two strands derived from a given dsDNA molecule [46]

3.9 Bioinformatic Analysis of NGS Reads Using Dellingr (Optional)

1. The following analysis assumes the user has administrator access to a UNIX/Linux-based system that meets Dellingr's minimum system requirements (64-bit OS, 8 GB RAM), and is familiar with the bash command line. Windows users will require a Linux subsystem or virtual box (*see* **Note 65**).
2. Download and install all Dellingr dependencies (*see* **Note 66**).
3. Download Dellingr from Github (https://github.com/morinlab/Dellingr).
4. Navigate into the Dellingr directory, and install both Dellingr and its python packages using the setup script provided:

```
$python3 setup.py install
```

5. Ensure that Dellingr was installed successfully by typing `delligr` into the terminal. A list of Dellingr commands should appear (*see* **Note 67**).

6. Copy the FASTQ files generated by the sequencing run performed in Subheading 3.7 over to a working directory of your choice.
7. Create a copy of the Dellinger configuration file located in the Dellingr directory:

```
etc/dellingr_config.ini
```

8. Open the configuration file in a text editor, and update the arguments to reflect the sample characteristics (*see* **Note 68** for an in-depth description of each parameter).
9. Create a copy of the Dellingr sample configuration file, also located in the Dellingr directory:

```
etc/sample_config.ini
```

   Open the sample configuration file in a text editor, and specify all samples that you wish to analyze, along with any custom parameters (*see* **Note 69** for more details).
10. Run Dellingr using the configuration files that were created in Subheading 3.9, **steps 7** and **9**:

```
$dellingr run_dellingr -c config.ini
```

   (*see* **Notes 70–72**).
11. Examine the final variant calls after analysis is complete. These can be found in the `results` directory corresponding to each sample (*see* **Note 73**).

4 Notes

1. Minimizing contamination of plasma with genomic DNA from lysed white blood cells is critical for PCR-based determinations and other methods for ctDNA quantification because it may greatly dilute the relative abundance of tumor-derived DNA in a sample and cause a misleading underestimate of the relative amount of ctDNA. Blood collection tubes containing the anticoagulant EDTA may suffice when processing blood samples and storing plasma at −80 °C within 4–6 h. If rapid processing is not an option, collection tubes containing stabilizing agents such as cell-free DNA BCT© tubes (Streck) should be used. The manufacturer indicates cell and cfDNA stability for up to 14 days at 6–37 °C. However, we caution that some studies have detected increased genomic DNA levels in plasma only 7 days after storage outside room temperature ranges

[53, 54]. It is also recommended to minimize the presence of air bubbles and moving/shaking tubes during transportation. The construction of cfDNA libraries subjected to NGS is not as significantly impacted by genomic DNA contamination because high molecular weight DNA does not efficiently engage in the adapter ligation step, PCR enrichment, or clustering in Illumina sequencers. Nonetheless, it is preferable to avoid this contamination.

2. A series of two rounds of centrifugation is the accepted standard procedure to isolate plasma. Serum is reported to be more affected by high-molecular-weight DNA contamination than plasma [55]. The first centrifugation usually takes place at 800–2500 × *g* for 10–20 min and it is intended to isolate the plasma fraction. The second spin is carried out at faster speeds (2000–16,000 × *g*) and it is intended to further remove cells and cell debris, although its efficacy is debatable. Temperature during centrifugation does not seem to influence the quality and quantity of cfDNA [8]. Centrifugation should be carried out inside a biosafety room to minimize the dispersion of aerosols. Opening the centrifuge lid a few minutes after centrifugation may also help to settle aerosols.
3. The opening of blood collection tubes after centrifugation and the transfer of plasma to other tubes must be performed inside a biosafety cabinet to minimize exposure to potentially infectious agents. The samples can be manipulated outside the biosafety cabinet after treatment with proteinase K and SDS.
4. There are several commercial kits available for researchers and clinicians to isolate high-quality cfDNA samples. In essence, cfDNA isolation kits leverage silica gel membrane-based columns (e.g., QIAamp Circulating Nucleic Acid Kit, Qiagen), magnetic beads (e.g., MagMAX™ Cell-Free DNA Isolation Kit, Applied Biosystems), or polymer-mediated enrichment techniques (e.g., PME free-circulating DNA extraction kit, Analytik Jena). A helpful review of the currently available commercial kits for cfDNA extraction is available [8] and an important consideration of some kits is the simultaneous purification of circulating DNA, RNA, and miRNA molecules (e.g., MagMAX Cell-Free Total Nucleic Acid Kit, Applied Biosystems). One of the major advantages of the cfDNA purification methods involving magnetic beads is that they are highly amenable to automation (e.g., KingFisher™ Flex Purification System, ThermoFisher Scientific).
5. Qubit fluorescence quantitation is known to be more sensitive, accurate, and reproducible than traditional methods measuring absorbance at 260 nm (e.g., Nanodrop™ spectrophotometers, Thermo Fisher Scientific). Total cfDNA yields may greatly vary

according to different parameters, including extraction method, elution volume, and different physiological conditions or disorders. It is widely accepted that the amount of cfDNA circulating in the blood of cancer patients and pregnant women is significantly higher than in the rest of the population. However, within cancer patients, cfDNA yields can vary from just 2 ng/mL to more than 100 ng/mL of plasma and the level of cfDNA itself is not a robust metric for ctDNA abundance [7, 22].

6. The NEBNext Ultra™ II DNA Library Prep Kit for Illumina includes all reagents necessary for the straightforward and efficient construction of libraries from minute amounts of cfDNA (as low as 0.5 ng). It includes an enzymatic module for end-repair and A-tailing plus modules for ligation (barcoded adapters not included) and PCR enrichment. This kit performs very well with the adapters described in this chapter but we cannot comment on the performance of our molecular barcoding method with other similar kits such as the KAPA HyperPrep kits (Roche, Basel, Switzerland) or the QIAseq Ultralow Input Library Kit (Qiagen, Hilden, Germany). We have demonstrated the compatibility of these adapters in previous versions of the KAPA library preparation chemistries (i.e., Roche, Basel, Switzerland). Any ligation-based chemistry that makes use of Y-shaped adaptors should theoretically be suitable, though ligation efficiency may vary. Most of the cfDNA library prep kits are amenable for automation (e.g., PerkinElmer Automated NGS Library Prep Workstations, http://www.perkinelmer.com/es/automated-liquid-handling/automated-library-prep/).
7. Our custom barcoded adapters are Y-shaped adapters that incorporate 2 noncomplementary trinucleotide strand-specific tags and a 12 semi-degenerate and potentially complementary sequence flanking the ligation site (*see* details in Fig. 9). Ready-to-use adapters are generated via the annealing of pairs of oligonucleotides with potentially complementary semi-degenerate barcode sequences. These oligonucleotides incorporate a phosphorothioate bond in their 3′-end to increase oligo stability and a 5′-phosphate group to enable adapter-cfDNA ligation. Oligonucleotides can be ordered, for example, from Integrated DNA Technologies (Skokie, IL, USA) using HPLC purification and resuspended in TE buffer to a final 100 μM concentration. Owing to the flexibility in terms of the different trinucleotide and semi-degenerate barcode sequences that can be incorporated, as many as 96 unique adapters can be generated from a batch of 64 unique oligonucleotides (*see* Tables 1 and 2). Oligonucleotides can be ordered individually or in plates. Pairs of oligonucleotides should be

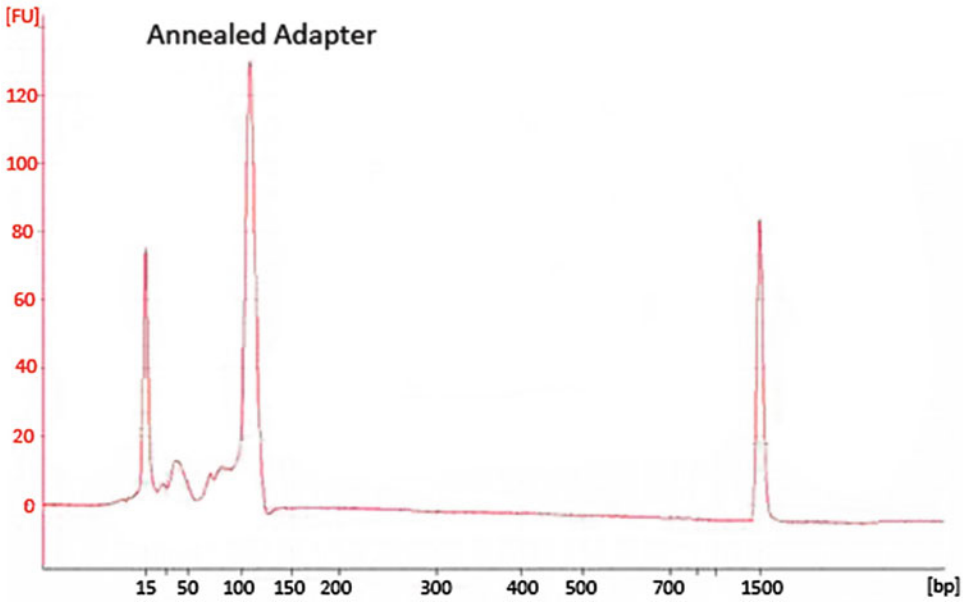


Fig. 9 Agilent Bioanalyzer trace of a pair of oligonucleotides annealed to generate Y-shaped semi-degenerate barcode adapters. Partially complementary annealed adapters migrate as 120 bp but they are actually ~65 bp in length

annealed as indicated in Table 2. Individual oligos can be stored at −20 °C for several years. Other groups have developed different sets of barcoded adapters amenable for duplex sequencing, which are described in detail elsewhere [44, 45, 56, 57].

8. Ensure to let AMPure XP magnetic beads equilibrate at room temperature at least 30 min before use. Cold beads have been related to suboptimal DNA binding efficiencies. Alternative magnetic beads for the purification and size selection of DNA fragments are available from other manufacturers but have not been fully tested by our group.
9. Our dual-indexed PCR primers provide 6-nucleotide tags for sample de-multiplexing and incorporate the DNA sequences needed for cluster generation and sequencing-by-synthesis in Illumina platforms. Previous studies have demonstrated that dual-indexed sequencing strategies successfully decrease cross-talk rates (i.e., the probability of a NGS read to be incorrectly assigned to a different sample and cause a false positive, [57]). Besides dual-index primers, our approach provides up to 96 unique adapter constructs. Thus, only those variants found in molecules presenting the right adapter sequence and assigned to the right dual-index sequence will be considered as true variants. This "double" safety tags not only minimize false-positive rates related to sample cross talk during

sequencing but also enable researchers to detect and control for cross-contamination occurring during or after library prep. Pairs of premixed PCR primers were ordered from Integrated DNA Technologies (Skokie, IL, USA) using HPLC purification in 96-well plates and diluted to a final concentration of 10 μM in TE buffer. Our design is comprised of 256 unique dual-indexed combinations (*see* Table 17) and primer premixes can be stored at −20 °C for several years. Higher order multiplexing is possible by simply increasing the number of primers carrying different indexes. The DNA sequences of individual oligonucleotides are displayed in Table 18. The index sequence information needed for de-multiplexing NGS reads is shown in Table 16.

10. Cot-1 Human DNA increases the specificity of hybridization capture experiments because it blocks the carryover of repetitive DNA elements. If ctDNA from other species rather than human is being investigated, we strongly recommend using a mixture of DNAs that better represents the repetitive fraction of the genome of that particular species.
11. xGen® blocking oligos increase the efficiency of targeted enrichment experiments because they hamper the daisy-chaining of library fragments caused by complementary regions in their ends. Our custom adapters are fully compatible with these blocking oligonucleotides.
12. The DNA sequences of P5 and P7 primers are as follows: P5-5′-AAT GAT ACG GCG ACC ACC GA-3′ and P7-5′-CAA GCA GAA GAC GGC ATA CGA GAT-3′. They can be ordered from any standard oligonucleotide manufacturer as HPLC-purified oligos and premixed at a final 10 μM concentration (each).
13. Three layers should be visible after centrifugation of blood collection tubes. The upper phase (pale yellow-reddish, depending on the amount of hemolysis) corresponds to the plasma fraction (≈55% of total blood); the thin grayish white interface corresponds to the buffy coat and contains white blood cells and platelets (<1% of total blood); the bottom phase (red) contains red blood cells (≈45% of total blood).
14. Buffy coats are a valuable source of genomic DNA from non-malignant cells that help discriminating between germline and somatic mutations. Buffy coats can also be used for the isolation and manipulation of circulating tumor cells.
15. For the preparation of freezing media, mix 35 mL of RPMI 1640 medium, 10 mL of fetal bovine serum (FBS), and 5 mL of DMSO.
16. Most of the commercial kits for liquid biopsy research recommend to extract cfDNA from 1 to 5 mL of biological fluids,

Table 17
Distribution of 256 premixed dual-indexed PCR primers for sample multiplexing

	1	2	3	4	5	6	7	8	9	10	11	12	
A	25F25R	25F26R	25F27R	25F28R	25F29R	25F30R	25F31R	25F32R	25F33R	25F34R	25F35R	25F36R	**PLATE 1**
B	26F25R	26F26R	26F27R	26F28R	26F29R	26F30R	26F31R	26F32R	26F33R	26F34R	26F35R	26F36R	
C	27F25R	27F26R	27F27R	27F28R	27F29R	27F30R	27F31R	27F32R	27F33R	27F34R	27F35R	27F36R	
D	28F25R	28F26R	28F27R	28F28R	28F29R	28F30R	28F31R	28F32R	28F33R	28F34R	28F35R	28F36R	
E	29F25R	29F26R	29F27R	29F28R	29F29R	29F30R	29F31R	29F32R	29F33R	29F34R	29F35R	29F36R	
F	30F25R	30F26R	30F27R	30F28R	30F29R	30F30R	30F31R	30F32R	30F33R	30F34R	30F35R	30F36R	
G	31F25R	31F26R	31F27R	31F28R	31F29R	31F30R	31F31R	31F32R	31F33R	31F34R	31F35R	31F36R	
H	32F25R	32F26R	32F27R	32F28R	32F29R	32F30R	32F31R	32F32R	32F33R	32F34R	32F35R	32F36R	
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	
A	33F25R	33F26R	33F27R	33F28R	33F29R	33F30R	33F31R	33F32R	33F33R	33F34R	33F35R	33F36R	**PLATE 2**
B	34F25R	34F26R	34F27R	34F28R	34F29R	34F30R	34F31R	34F32R	34F33R	34F34R	34F35R	34F36R	
C	35F25R	35F26R	35F27R	35F28R	35F29R	35F30R	35F31R	35F32R	35F33R	35F34R	35F35R	35F36R	
D	36F25R	36F26R	36F27R	36F28R	36F29R	36F30R	36F31R	36F32R	36F33R	36F34R	36F35R	36F36R	
E	37F25R	37F26R	37F27R	37F28R	37F29R	37F30R	37F31R	37F32R	37F33R	37F34R	37F35R	37F36R	
F	38F25R	38F26R	38F27R	38F28R	38F29R	38F30R	38F31R	38F32R	38F33R	38F34R	38F35R	38F36R	
G	39F25R	39F26R	39F27R	39F28R	39F29R	39F30R	39F31R	39F32R	39F33R	39F34R	39F35R	39F36R	
H	40F25R	40F26R	40F27R	40F28R	40F29R	40F30R	40F31R	40F32R	40F33R	40F34R	40F35R	40F36R	

	1	2	3	4	
A	25F37R	25F38R	25F39R	25F40R	**PLATE 3**
B	26F37R	26F38R	26F39R	26F40R	
C	27F37R	27F38R	27F39R	27F40R	
D	28F37R	28F38R	28F39R	28F40R	
E	29F37R	29F38R	29F39R	29F40R	
F	30F37R	30F38R	30F39R	30F40R	
G	31F37R	31F38R	31F39R	31F40R	
H	32F37R	32F38R	32F39R	32F40R	
	1	**2**	**3**	**4**	
A	33F37R	33F38R	33F39R	33F40R	**PLATE 4**
B	34F37R	34F38R	34F39R	34F40R	
C	35F37R	35F38R	35F39R	35F40R	
D	36F37R	36F38R	36F39R	36F40R	
E	37F37R	37F38R	37F39R	37F40R	
F	38F37R	38F38R	38F39R	38F40R	
G	39F37R	39F38R	39F39R	39F40R	
H	40F37R	40F38R	40F39R	40F40R	

The DNA sequences of each premixed PCR primer are provided in Table 18

Table 18
DNA sequences of the dual-indexed PCR primers needed to multiplex hundreds of samples in one single sequencing experiment

Oligo ID	DNA sequence 5′-3′
Indx PCR F	CAA GCA GAA GAC GGC ATA CGA GAT *NNN NNN* GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC* T
Indx PCR R	AAT GAT ACG GCG ACC ACC GAG ATC TAC AC*N NNN NN*A CAC TCT TTC CCT ACA CGA CGC TCT TCC GAT C*T

Each PCR primer carries a set of 6-nucleotide indexes (indicated by *NNNNNN*). For example, a suite of 16 unique forward primers (F) can be combined with 16 unique reverse primers (R) to tag 256 unique cfDNA libraries (see more details in Tables 16 and 17)

mostly plasma. Plasma volume will dictate the capacity of the magnetic rack separators used during bead-based extraction methods (i.e., MagMAX™ Cell-Free DNA Isolation Kit, Applied Biosystems). Plasma is the main biological fluid used for ctDNA determinations in lymphoma patients [40, 41, 50]. cfDNA can be isolated from CSF specimens using the same commercial kits [27, 58]. Some manufactures have released specific protocols for the handling of urine samples (e.g., MagMAX™ Cell-Free DNA Isolation Kit User Bulletin (urine samples) (Pub. no. MAN0015628)).

17. Protocols differ depending on the specific cfDNA extraction kit and the type of blood collection tubes used. For example, Applied Biosystems recommends pretreating plasma samples preserved in Cell-free DNA BCT© (Streck) tubes with proteinase K during 20 min at 60 °C when using the MagMAX™ Cell-Free DNA Isolation Kit. For bead-based purification protocols, cfDNA is specifically bound to magnetic particles subjected to successive washing rounds with buffers supplied by the manufacturer and 80% ethanol. cfDNA is finally eluted using elution buffers provided by the manufacturer. This particular kit offers the possibility to extract cfDNA manually or automatize the process in the KingFisher™ Flex Magnetic Particle Processor. For the popular QIAamp Circulating Nucleic Acid Kit, samples are pretreated with proteinase K and the resulting lysates are filtered through silica gel membrane-based columns using vacuum force. The membrane specifically binds DNA and it is successively washed with buffers provided by the manufacturer and 80% ethanol. Finally, cfDNA is eluted by the addition of an elution buffer provided by the manufacturer after centrifugation of the column and the collection of the cfDNA extract in a clean 1.5 mL Eppendorf tube.

18. The elution volume may affect both the recovery efficiency and total concentration of cfDNA fragments. We normally elute cfDNA extracts in 75–100 μL of elution buffer and use 50 μL of the cfDNA extract for library preparation, unless the concentration of cfDNA is extraordinarily large (e.g., >300 ng cfDNA inputs allocated for library prep). If smaller cfDNA volumes are used during library preparation, adjust the volume of the end-repair plus A-tailing reaction with ultrapure water. Total cfDNA yields are highly variable depending on the patient, tumor type, and stage of the disease, but we find cfDNA yields ranging from 4 to 80 ng/mL of plasma to be the most commonly observed (*see also*, for example, [59–61]).
19. Some studies have documented a 30% degradation rate of cfDNA fragments when stored at −20 °C for 1 year [62]. Thus, prompt use of cfDNA extracts may maximize assay sensitivity.
20. Extraordinarily high cfDNA yields (i.e., >150 ng/mL plasma) are not rarely observed in cancer patients with a very advanced stage of the disease but can also be caused by large genomic DNA contamination associated with poor blood collection and storage or other processes such as extensive necrosis or post-radiation cellular damage.
21. The typical size distribution of cfDNA fragments exhibits a prominent peak at ≈170 bp and two additional peaks at 340 bp and 510 bp showing progressive decay of detected fluorescence intensity [7]. Aberrant cfDNA patterns show a large abundance of high-molecular-weight DNA fragments that can be attributed to poor blood handling and storage, extensive necrosis, or postradiation cellular damage (among other possible causes).
22. Only oligos with perfectly complementary strand-specific tags and potentially complementary semi-degenerate barcode sequences are not annealed together. As explained in the manuscript detailing this approach [46], noncomplementary strand-specific tags minimize annealing mismatches across the semi-degenerate barcode sequence.
23. Old adapter stocks subjected to multiple freezing-thaw cycles may exhibit lower ligation efficiencies than freshly made adapters, although we have not fully tested the extent to which this factor may affect library diversity.
24. A master mix of the first two reagents can be prepared immediately before use and dispensed with the aid of an electronic pipette with multi-dispense function and a reagent reservoir. Thoroughly vortex the master mix and perform a quick spin before pouring the master mix inside the reservoir. The use of adjustable multichannel pipettes is recommended to transfer

cfDNA extracts into individual tubes/wells containing the end-repair/A-tailing master mix in order to minimize accidental sample mix-ups during the handling of large numbers of samples.

25. The manufacturer recommends proceeding immediately with adapter ligation since a small drop in cfDNA recovery efficiency (~20%) may be observed if samples are kept at −20 °C until further processing.
26. We and others recommend adding barcoded adapters in at least a 100-fold excess with respect to cfDNA input. This strategy is aimed at maximizing recovery efficiencies of cfDNA fragments, which have been documented to exceed 50% in some studies [30, 46, 56]. We do not recommend premixing the adapters, ligation enhancer, and ligation master mix in one single tube because of the high probabilities to generate undesired adapter dimers in advance. It is fine to premix the ligation master mix and the ligation enhancer prior to use. Large master mixes of the ligation enhancer and the ligation master mix can be poured into a reagent reservoir to facilitate multiple dispense using an adjustable multichannel pipette.
27. Incomplete mixing may translate into reduced ligation efficiency. A small number of bubbles will not interfere with the ligation step.
28. Longer or overnight ligations at 16 °C may increase recovery efficiency of cfDNA fragments [30] but we have observed highly diverse libraries by just following the fast protocol described above.
29. Size selection using magnetic beads is intended to remove adapter dimers formed during the adapter ligation reaction. Adapters are added to the ligation reaction in great excess, and as a result adapter dimer formation is common. Adapter dimers can be particularly abundant when cfDNA inputs are low (i.e., <10 ng). For those cases, two successive rounds of size selection using 0.8× volumes of magnetic beads are recommended to remove adapter dimers prior to library amplification. Smaller amounts of adapters can also minimize adapter dimer formation for these particular cases. Large input cfDNA amounts considerably decreases the impact of adapter dimers in final libraries. Small traces of adapter dimers can be removed at later steps (i.e., during size selection of post-PCR-enriched library fragments or during size selection of post-capture library fragments). Pouring AMPure XP beads in a pipetting/reagent reservoir (ensure to thoroughly mix the bead solution before this) and use of a multichannel pipette may help to process large batches of samples.

30. Make sure that plates are well sealed and avoid excessive agitation when using the high-speed mixer to avoid cross-contamination between wells. Never place the strips/tubes in a standard vortex.
31. 75% Ethanol may be more efficient in removing adapter dimers than 80% ethanol. Regardless of the final ethanol concentration, ensure that fresh solutions are prepared regularly.
32. Avoid the presence of ethanol traces, which may inhibit enzymatic reactions during the next steps, but also to overdry magnetic beads because this may cause suboptimal recovery efficiency of cfDNA fragments. Elute the beads when they are still dark brown and glossy looking rather than when they are lighter brown and/or cracked. Remove ethanol first with a pipette suitable for large volumes (i.e., 200 μL) and use a pipette set for smaller volumes (i.e., 10 μL) to remove remaining drops. Tubes and plates can be dried faster in a thermal cycler with the lid open and set at 37 °C, but at a higher risk of overdrying the beads.
33. Ensure that beads are back in solution and not still bound to the walls of the tubes. Overdried beads usually bind strongly to plastic tubes and are more difficult to get them back in suspension.
34. Avoid disturbing and pipetting the ring of beads. Leaning individual PCR tubes and strips help concentrate the beads on one side of the tubes. Introduce the pipette tips toward the opposite side of the tubes to avoid disturbing/pipetting the beads. Tubes or plates carrying magnetic beads can be discarded at this point after transferring the supernatant to new tubes/plates.
35. Use a multichannel pipette to transfer multiple dual-indexed PCR primers at once to minimize accidental mix-ups when dealing with large numbers of samples. Add first the dual-indexed PCR primers and then use a pipetting/reagent reservoir and a multichannel pipette to dispense the NEBNext Ultra II Q5 Master Mix.
36. The extension time during the cycling protocol has been slightly increased with respect to the manufacturer's protocol to further minimize amplification biases and the formation of chimeric sequences related to in vitro PCR-mediated recombination.
37. cfDNA inputs >50 ng may benefit from smaller numbers of PCR amplification cycles to avoid library overamplification. Library overamplification manifests by a substantial increase in high-molecular-weight artifactual fragments and also exacerbates GC-related amplification biases. The number of PCR cycles is also known to have a significant effect on the relative

size of PCR families and on the total fraction of PCR duplicates during NGS experiments. For example, we have occasionally built two parallel libraries from high cfDNA inputs (>100 ng) in plasma samples with high ctDNA content to conduct whole-genome sequencing. Here, we only perform four PCR cycles per library and then pool both libraries prior to sequencing to minimize the extent of PCR duplicates in the final sequencing data. cfDNA inputs <10 ng may benefit from nine PCR cycles instead.

38. This size selection step is more aggressive than most of the cleanup step recommended by manufacturers and it is intended to minimize the impact of adapter dimers. We rather carry out a single size selection step after adapter ligation but a more stringent size selection step after PCR enrichment. In other words, we prefer to carry over a small fraction of adapter dimers and lose PCR duplicates rather than miss unique cfDNA fragments prior to amplification.
39. Normal cfDNA library yields range from 5 to 50 ng/μL when libraries are eluted in 25–30 μL.
40. An additional size selection using 0.8× volumes of AMPure XP magnetic beads and 75% ethanol washes can help removing adapter dimers if their amount is higher than desired.
41. We follow here the methods and reagents supplied by the xGen® hybridization capture technology (Integrated DNA Technologies; Skokie, IL, USA). Other hybridization-based target enrichment technologies, such as SeqCap products (Roche, Basel, Switzerland), have not been evaluated in the present protocol. The detailed xGen® hybridization capture manual can be downloaded from IDT's webpage (https://www.idtdna.com/pages/products/next-generation-sequencing/hybridization-capture/hybridization-reagents/hybridization-and-wash-kit) and it is amenable for the handling of both individual tubes (one single hybridization capture) or 96-well plates (multiple simultaneous hybridization captures).
42. Individual xGen® Lockdown® probes can be ordered at different synthesis scales in 96-well plates from Integrated DNA Technologies (Skokie, IL, USA). The minimum order for individual xGen® Lockdown® probes is 96 and their cost is ~30 USD per probe. Biotinylated DNA ultramers are also suitable for hybridization capture but are more expensive (>100 USD per probe). Custom pools of small numbers of xGen® Lockdown probes are highly amenable for personalized assays to track ctDNA [46, 47] or for the survey of mutation hotspots [46]. The standard synthesis scale (24 pmol/probe) is usually enough for most experiments. xGen® Lockdown® probes are biotinylated single-stranded DNA oligonucleotides of ~120 nt

in length. We recommend designing these probes in a way that the targeted position of interest is situated in the middle of the probe to ensure maximum recovery efficiency of cfDNA fragments spanning that particular position [46]. We have also observed that flanking the position of interest with partially overlapping probes in the opposite direction (5 bp at both the 3′- and 5′-ends of the first probe) significantly increases the recovery efficiency of cfDNA fragments spanning the mutations of interest. We commonly apply this probe design strategy during the investigation of mutation hotspot panels. It is possible to order individual probes to cover all the exons from one or multiple genes. Equimolar amounts of each probe can be then pooled to generate custom panels, which should be diluted to a final concentration of 0.75 pmoles of total probes/μL, and 4 μL of probes should be used during the hybridization reaction. IDT also offers the possibility to order xGen® predesigned gene capture pools. These pools contain probes that tile the entire coding region of a given gene and need to be ordered in batches of at least ten genes. The manufacturer can pre-pool different gene pools in one single tube or ship the probes pertaining to a given gene in individual wells of a 96-well plate, which allows more flexibility. Custom multi-gene pools can be prepared by the user by mixing equimolar amounts of different genes. Multigene pools should be dried down in a SpeedVac concentrator to at least two times the volume corresponding to that of a single gene used in the pool. Four microliter of multigene pools should be used for the hybridization reaction. Predesigned xGen® Lockdown panels are also available from Integrated DNA Technologies and should be resuspended according to the instructions provided by the manufacturer and used in volumes of 4 μL per hybridization reaction. Examples of xGen® Lockdown panels with interest in cancer research include xGen® Exome Research, Pan Cancer or AML Panels.

43. We recently described a protocol for the custom generation of biotinylated DNA baits [46]. This protocol may be particularly useful when a whole order of 96 xGen Lockdown probes cannot be fulfilled. These biotinylated baits can be spiked into xGen® Lockdown pools to capture regions poorly covered, for example, by predesigned gene capture pools or panels (i.e., 5′-UTR and 3′-UTR regions or regions commonly involved in recurrent genomic rearrangements).
    (a) Multiplex PCR primers (3–4 amplicons per reaction) with M13 tails (i.e., 5′-biotin-GTTTTCCCAGTCACGAC-locus-specific primer-3′ and 5′-P-CAGGAAACAGCTATGAC-locus-specific primer-3′) in the following

Table 19
PCR recipe for the first round of PCR amplification aimed at the generation of homemade biotinylated DNA baits

Reagent	Volume (μL)
KAPA HiFi HotStart ReadyMix	5
0.2 pmol of each locus-specific primer	Variable
3 ng of human WT DNA	Variable
	10

Table 20
PCR thermal cycling protocol for the generation of M13-tailed PCR amplicons later subjected to a second round of amplification using primers conjugated with biotin and phosphate modifications

Cycle step	*T* (°C)	Time	Cycles
Initial denaturation	95	3 min	1
Denaturation	98	15 s	15
Annealing	58	30 s	
Extension	72	2 min	
Final extension	72	4 min	1
Hold	4		

pre-amplification reaction and according to the PCR recipe and cycling protocol specified in Tables 19 and 20.

(b) Carry out a double-size selection using magnetic AMPure XP beads. First, 0.5× volumes of AMPure beads (5 μL) should be added to the PCR reaction to specifically bind high-molecular-weight DNA. Collect the supernatant in a new 0.2 mL PCR tube. Next, add 1.5× volumes of AMPure XP beads to the collected supernatant in order to purify PCR amplicons.

(c) Subject PCR amplicons to a second round of amplification according to the conditions and cycling protocol described in Tables 21 and 22.

(d) Purify the new batch of PCR amplicons using 1.5× volumes of AMPure XP magnetic beads and elute the DNA in 25 μL of 10 mM Tris–HCl pH = 8.0. Measure DNA concentration in a Qubit fluorometer and prepare amplicon stocks at ~15 ng/μL.

(e) Proceed with the specific degradation of the 5′-phosphorylated DNA strand of each PCR amplicon by incubating

Table 21
PCR recipe for the second round of PCR amplification aimed at the generation of homemade biotinylated DNA baits

Reagent	Volume (μL)
KAPA HiFi HotStart ReadyMix	25
5′-Biotinylated universal primer (10 μM)	2
5′-Phosporylated universal primer (10 μM)	2
Eluted PCR amplicons	21
	50

Table 22
PCR thermal cycling protocol for the generation of PCR amplicons carrying 5′-biotin and 5′-phosphate groups

Cycle step	*T* (°C)	Time	Cycles
Initial denaturation	95	3 min	1
Denaturation	98	15 s	30
Annealing	60	30 s	
Extension	72	2 min	
Final extension	72	4 min	1
Hold	4		

with lambda exonuclease as described in Tables 23 and 24.

(f) Purify treated PCR amplicons with 1.5× volumes of AMPure XP beads. Elute the final biotinylated baits in 25 μL of 10 mM Tris–HCl and 0.1 mM EDTA, pH = 8.0.

(g) Use 4 μL of biotinylated DNA baits per hybridization reaction.

44. The most recently released version of the xGen® hybridization capture of DNA libraries protocol recommends using 500 ng of DNA per library, regardless whether one or multiple libraries are subjected to targeted enrichment. We have nonetheless observed good results when using lower amounts of cfDNA libraries (~150 ng of individual cfDNA libraries in pools). Higher amounts of cfDNA libraries may nevertheless translate into higher diversity of final enriched libraries. IDT does not recommend increasing the amount of Human Cot-1 DNA or blockers when handling large library pools.

Table 23
Incubation of PCR amplicons with lambda exonuclease leads to the generation of single-stranded DNA fragments carrying 5′-biotin modifications

Reagent	Volume (μL)
Lambda exonuclease (5 U/μL)	5
10× lambda exonuclease reaction buffer	5
Eluted PCR amplicons (15 ng/μL)	20
Nuclease-free water	20

Table 24
Thermal cycling protocol for the incubation of PCR amplicons with lambda exonuclease

Cycle step	*T* (°C)	Time (min)	Cycles
Degradation of 5′-phosphorylated strand	37	30	1
Heat enzyme inactivation	75	10	1
Hold	4		

45. The amount of time to completely dry down the mixture of cfDNA libraries, Human Cot-1 DNA, and blockers will depend on the volume of the mixture, the efficiency of the SpeedVac concentrator, and the temperature used during the process. SpeedVac concentrations usually allow selecting between three different temperatures, indicated as low (slower), medium, or high (faster). We typically use the highest temperature to accelerate the process. Sealed tubes/plates containing dried DNA samples can be kept at room temperature overnight or at −20 °C for longer periods of time.
46. Thaw components at room temperature and warm hybridization buffer stock at 65 °C if crystal precipitates are observed.
47. The most up-to-date version of the xGen® hybridization capture protocol recommends to add the biotinylated DNA baits directly into the hybridization master mix instead of after denaturing the library pool, Human Cot-1 DNA, and blockers. We have not fully evaluated this new protocol before writing this chapter and therefore we follow previous versions.
48. Place the samples in an appropriate magnetic separator if magnetic beads are still present due to carryover during previous steps. Transfer the clean supernatant into a new low-binding 0.2 mL PCR tube.
49. The most up-to-date version of the xGen® hybridization capture protocol has decreased the denaturation time to only 30 s and recommends to preheat the lid at 100 °C.

50. Certain pools of biotinylated DNA baits may generate better results at alternative final concentrations than those recommended by the manufacturer. In such situations, titration experiments may point toward the concentration of probes that generates the highest fraction of on-target reads. In most of the occasions, significantly lower concentrations of baits translate into enhanced enrichment efficiencies, particularly when small pools/targeted regions are involved.
51. Ensure covering the hybridization reaction with 10 μL of mineral oil and seal tubes and plates firmly to prevent evaporation during overnight or longer hybridizations. Long hybridizations are predicted to increase enrichment efficiency, particularly for small capture pools, but are also expected to increase the extent of DNA damage [45, 56].
52. Heat the 10× wash buffer I solution at 65 °C if cloudy until it shows complete solubilization of precipitates. The 1× stock working solutions are stable at room temperature for at least 4 weeks.
53. Integrated DNA Technologies has recently updated the protocol regarding the manipulation of M-270 streptavidin beads. We refer here to previous versions of the protocol. The newest protocol version can be downloaded from https://www.idtdna.com/pages/products/next-generation-sequencing/hybridization-capture/hybridization-reagents/hybridization-and-wash-kit.
54. Carrying over of small traces of mineral oil, if used during long hybridizations, do not seem to negatively impact the performance of the enrichment experiment.
55. It is critical that wash buffers I and stringent wash buffers have reached 65 °C before starting the heated washes.
56. The number of PCR cycles should be optimized for each individual pool of biotinylated probes. It is always recommended to carry out as few PCR cycles as possible to obtain optimal sequencing results. Final concentrations of enriched libraries around 1 ng/μL are sufficient for sequencing purposes.
57. Two rounds of hybridization capture may be necessary for certain pools, particularly for small pools of biotinylated baits because they usually generate low percentages of reads on-target [46, 63]. Add half the amount of blockers, Human Cot-1 DNA, and biotinylated baits during additional rounds of hybridization capture. Hybridization and incubation times during heated washes can be shortened in half (e.g., 2-h hybridization at 65 °C instead of 4 h). Perform the smaller number of post-capture PCR cycles possible at the end of the process.

58. The user-friendly Geneious genomics workbench offers a straightforward analysis of cfDNA libraries built with semi-degenerate barcode adapters for researchers and clinicians without a strong background in bioinformatics or poor programming skills. Geneious is particularly amenable for the analysis of small personalized captures or relatively small gene panels. However, this genomics workbench is not as computationally efficient as standard bioinformatics tools for the analysis of complex experiments such as whole-exome sequencing or for the processing of large numbers of samples. Please note that cfDNA libraries built with other type of barcoded adapters should be analyzed through alternative bioinformatics workflows [44, 45, 56, 64].
59. Targeted sequencing experiments may benefit from the use of simplified references in Geneious to minimize computational time and data storage requirements. In any case, be sure to download an annotated version of the human reference genome and import it into Geneious (e.g., http://support.illumina.com/sequencing/sequencing_software/igenome.html).
60. Setting parameters for the annealing of Illumina paired reads against the human genome reference using the Geneious aligner (Panel a) and settings to find, extract, and merge regions of high coverage in the human genome reference (Panel b) are shown in Fig. 10. The minimum number of sequences to differentiate between on-target (specific enrichment) and off-target (nonspecific enrichment) would depend on the efficiency of the enrichment experiment and the amount of sequencing data allocated for a certain library pool. In this particular example, Geneious will incorporate "High-Coverage" annotations to all those regions with coverage >20×. Geneious can generate a concatenated sequence of high-coverage regions for each individual chromosome while keeping the original reference coordinates (Fig. 10, panel c). For personalized assays tracking a small number of somatic mutations, all chromosomes can be concatenated in one single reference sequence if desired by selecting Tools > Concatenate Sequences or Alignments and leaving 500 bp spacers between each individual sequence. Genomic positions mutated in primary tumors can be annotated by adding, for example, "somatic mutations" annotations. This can help Geneious to specifically look for variants at those sites later on during the workflow and ignore the rest of nucleotide positions if desired. The same reference can be used for multiple samples as long as the targeted space remains the same.
61. In essence, these workflows generate consensus sequences from either merged or unmerged reads that do not differ in

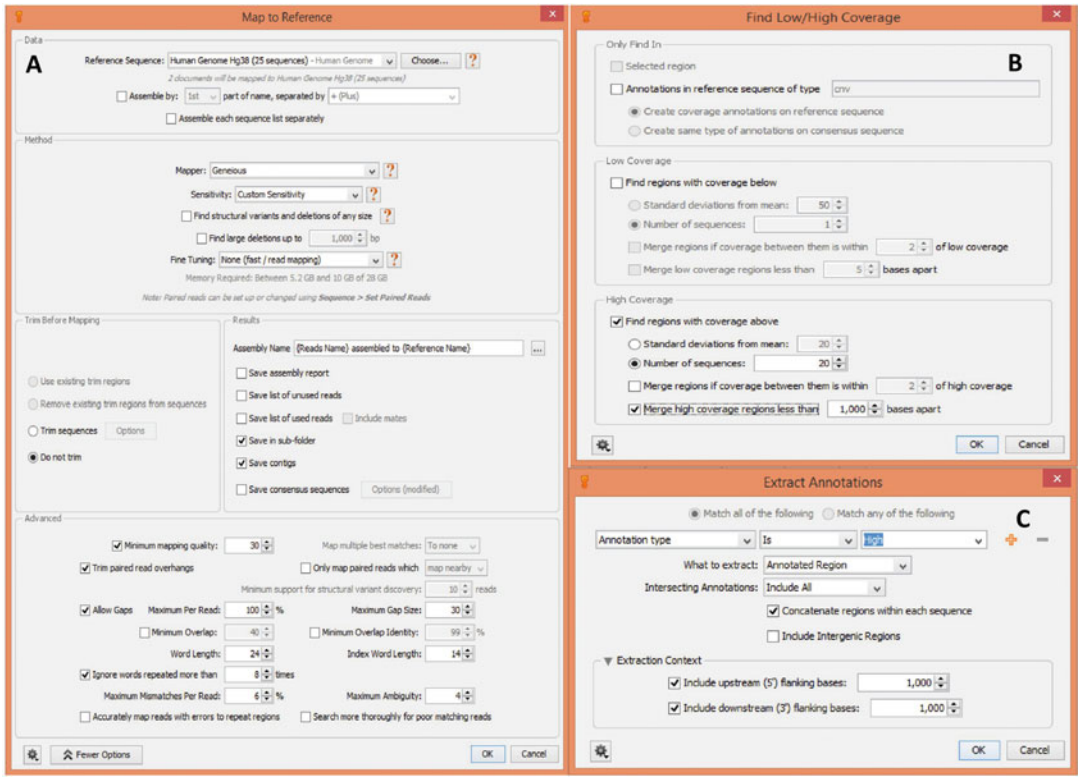


Fig. 10 Settings to map NGS reads derived from the sequencing of cfDNA libraries against the human reference genome (Panel **a**), the annotation of high-coverage regions (Panel **b**), and optional concatenation of several high-coverage annotations into a single sequence (Panel **c**)

more than 1% mismatches. The cfDNA-GenWkf1 workflow only keeps consensus sequences generated from at least three independent reads while the cfDNA-GenWkf1lp workflow keeps both any type of consensus sequences and singletons. The use of the cfDNA-GenWkf1 or the cfDNA-GenWkf1lp workflow depends on sequencing depth at each targeted site. For example, for libraries enriched with small panels of baits and deeply sequenced we recommend executing the cfDNA-GenWkf1 because the probabilities to generate large PCR families are high. Libraries enriched with large gene panels and only shallowly sequenced may benefit, on the other hand, by executing the cfDNA-GenWkf1lp instead. We recommend to allocate approximately two million reads for hybridization capture experiments relying on small personalized panels of baits (i.e., <10 baits). More details concerning each specific workflow can be found in the supplemental information section of our publication detailing this technology [46].

62. The cfDNA-GenWkf3 workflow specifically searches for variants at those genomic positions specifically indicated

(by means of annotations) to contain somatic mutations in primary tumors; the cfDNA-GenWkf3lp workflow searches for any variant across coding regions and splice sites. These by-default parameters can be modified by manually editing the workflows in Geneious [46]. Geneious generates a variant annotation track for each sample in the original reference sequence that summarizes useful information such as the relative frequency of a given variant, the number of raw reads or consensus sequences supporting that variant, start and end coordinates, strand-bias values or average Illumina Phred scores, among other useful parameters.

63. We recommend using the variant annotation table to remove SNV calls with Phred scores <30 or those variants only supported by one single read/consensus sequence. Since the DNA sequences of individual strands are independently aligned against the reference sequence in this analysis pipeline, a minimum of two reads/consensus sequences are needed to obtain duplex support.
64. Refer to the instructions provided for each variant annotation tool concerning the input format of your variants called in Geneious. The mutation annotation files (MAF) generated for each variant annotation tool can be merged with the .csv files exported from Geneious. These exported .csv files summarize important information such as variant allele frequencies, number of raw reads/consensus sequences supporting a certain variant call, strand-bias values, or average Illumina's Phred scores. We found particularly useful the implementation and design of Macros in Excel to summarize and display the most relevant information associated to each variant.
65. Windows users can install Dellingr within the Linux subsystem for Windows, by following the instructions outlined at https://docs.microsoft.com/en-us/windows/wsl/install-win10. Please note that cfDNA libraries built with other type of barcoded adapters should be analyzed through alternative bioinformatics workflows [44, 45, 56, 64]. It is highly recommended that all packages are updated after installing the Linux subsystem:

```
$sudo apt update
$sudo apt upgrade
```

66. The dependencies listed in Table 25 are required prior to installing or running Dellingr.
67. Numerous different commands are available when running Dellingr (Table 26). Further documentation can be found at http://dellingr.readthedocs.io/en/latest/.

Table 25
Summary of the dependencies required prior to installing or running Dellingr

Software	Minimum version	Available from	Notes
python	3.4	https://www.python.org/downloads/	
pip	8.00	https://pip.pypa.io/en/stable/installing/	Ensure that your version of pip is compatible with python version 3+
Burrows-Wheeler Aligner	0.7.0	https://github.com/lh3/bwa	
Samtools	1.3.1	http://www.htslib.org/download/	
Git	1.9.0	https://git-scm.com/downloads	

Table 26
Summary of Dellingr commands

Command	Description
run_dellingr	Runs the entire Dellingr pipeline on each sample. This will be the primary command used
resume_dellingr	If `run_dellingr` is terminated, this command can be used to restart the pipeline
trim	Trims off the degenerate barcodes from raw FASTQ files
collapse	Identifies reads that originate from the same parental molecule, and merges them into a consensus
clip	If the reads within a read pair overlap, a consensus is obtained from the overlapping bases, and the redundant overlap is removed
call	Identifies variants
update_config	Updates an older config file
adapter_predict	Identifies the degenerate barcode used in given sample
train	Retrains the variant calling filter

68. The Dellingr configuration file lists all parameters that are mandatory to run the full Dellingr pipeline (Table 27).
69. The sample configuration file is an easy method to analyze multiple samples simultaneously. The format of the sample configuration file is relatively straightforward. For example, if running a sample called "Thor," with FASTQ files located in /

Table 27
Mandatory parameters required to run the full Dellingr pipeline are listed in the configuration file

Parameter	Description	Notes
barcode_sequence	The sequence of the degenerate barcode used in this sample, represented in IUPAC bases	If removed from the configuration file, Dellingr will attempt to guess the barcode used in each sample
barcode_position	Which positions in the barcode to examine when trimming barcodes from the FASTQ files	If using 12 nt degenerate barcodes (*see* **Note 7**), this can be left at default. Used with max_mismatch
max_mismatch	When trimming the barcode from a read pair, if more than this many positions fall outside of the specified barcode_sequence, that read pair is discarded	Can usually be left at the default value
family_mask	Which positions in the barcode to examine when determining if two read pairs originated from the same parental molecule (termed "family")	If using 12 nt degenerate barcodes (*see* **Note 7**), this can be left at default
family_mismatch	How many barcode sequence mismatches are permitted before two read pairs are considered members of different families	Can usually be left at the default value. Increasing this will lead to more lenient family identification
duplex_mask	Which positions in the barcode to examine when determining if two families originated from different strands of the same parental molecule (termed a "duplex")	If using 12 nt degenerate barcodes (*see* **Note 7**), this can be left at default. The first six bases of the barcode are usually excluded, as they are more error prone
duplex_mismatch	How many barcode sequence mismatches are permitted before two families are not considered to be in duplex	If using 12 nt degenerate barcodes (*see* **Note 7**), this can be left at default
Reference	Reference genome FASTA file	Can be downloaded from http://hgdownload.cse.ucsc.edu/downloads.html
sample_config	Sample configuration file (*see* **Note 69**)	

home/user/sequence_data/thor.R1.fastq.gz and /home/user/sequence_data/thor.R2.fastq.gz, that sample would be specified as follows:

```
    [Thor]
fastqs=/home/user/sequence_data/thor.R1.fastq.gz,/home/
user/sequence_data/thor.fastq.gz
```

Multiple samples can be specified in the same manner:

```
    [Thor]
fastqs=/home/user/sequence_data/thor.R1.fastq.gz,/home/
user/sequence_data/thor.R2.fastq.gz
[Kratos]
fastqs=/home/user/sequence_data/kratos.R1.fastq.gz,/home/
user/sequence_data/kratos.R2.fastq.gz
[Odin]
fastqs=/home/anotheruser/sequence_data/lightning.R1.
fastq.gz,/home/anotheruser/sequence_data/lightning.R2.
fastq.gz
```

Finally, if one or more samples should be run with slightly different parameters than other samples (most commonly, a barcode sequence), the relevant parameter can be provided here:

```
    [Thor]
fastqs=/home/user/sequence_data/thor.R1.fastq.gz,/home/
user/sequence_data/thor.R2.fastq.gz
barcode_sequence=NNNSWMRWSYWKMSST
[Kratos]
fastqs=/home/user/sequence_data/kratos.R1.fastq.gz,/home/
user/sequence_data/kratos.R2.fastq.gz
max_mismatch=4
[Odin]
fastqs=/home/anotheruser/sequence_data/lightning.R1.
fastq.gz,/home/anotheruser/sequence_data/lightning.R2.
fastq.gz
```

70. Dellingr can process multiple samples at once, by adding the `--jobs` flag.
71. The Dellingr pipeline is composed of numerous discrete steps that are preformed sequentially on each sample. This includes sample preprocessing (trim, read alignment using bwa-mem), error correction (collapse), and variant calling (call).
    (a) Trim barcodes (using `dellingr trim`): During read mapping, reads may become misaligned as a result of the degenerate barcode mapping to an arbitrary region of the reference. To prevent this from occurring, the barcode is trimmed from each read, and stored for downstream use, prior to read mapping. Additionally, any read pair in which the degenerate barcode falls significantly outside the specified barcode range is discarded, which removes almost all contaminating sequences.
    (b) Map reads against a reference genome:

```
$bwa mem -C /path/refGenome.fa /path/r1.fastq.gz
/path/r2.fastq.gz | samtools view -b | samtools sort
> outFile.bam
```

To determine if two read pairs originate from the same parental DNA molecule (members of the same "family"), the barcode sequences must be compared. To significantly reduce the number of comparisons that must be performed, Dellingr first maps all reads against a reference. As reads that originate from the same starting molecule should share an almost identical sequence, all family members should map to the same position in the reference. Thus, by comparing barcode sequences only between read pairs which map to the same location, performance is significantly improved.

(c) Identify read families, perform error correction, and identify duplexes (using `dellingr collapse`): Examine all read pairs that map to a given location in the reference, and compare the barcode sequences between them. If the barcodes are very similar, the read pairs are grouped, and merged into a consensus. If there is disagreement between the sequences within each family, the base with the highest number of family members supporting it is used as the consensus. Following family collapsing, barcodes are compared between families to determine if any families might originate from different strands (i.e., + or − strand) of the same parental duplex. These duplexes are flagged for downstream use.

(d) Variant calling (using `dellingr call`): After flagging all positions which support a variant allele, candidate variants are filtered to remove noisy variants. This filter considers features that are both incorporated into other variant callers, such as base quality, mapping quality, and homopolymer repeats, but also features that are specific to the barcoded adapters, such as family size, and duplex support. If any positions overlap within a read pair, the base with the highest quality score is used as a consensus for variant calling.

72. If the Dellingr pipeline fails, analysis can be restarted using `dellingr resume_dellingr` after resolving any problems that resulted in pipeline failure.

73. Once the full Dellingr pipeline has been completed, numerous options are available for subsequent analysis. The identification of copy number alterations, as well as additional simple somatic variant calling, can be performed using the final BAM file

found in the "results" folder that is generated for each sample. This BAM file represents the full error suppression potential of the Dellingr pipeline.

References

1. Crowley E, Di Nicolantonio F, Loupakis F, Bardelli A (2013) Liquid biopsy: monitoring cancer-genetics in the blood. Nat Rev Clin Oncol 10:472–484
2. Heitzer E, Ulz P, Geigl JB (2015) Circulating tumor DNA as a liquid biopsy for cancer. Clin Chem 61:112–123. https://doi.org/10.1373/clinchem.2014.222679
3. Wan JCM, Massie C, Garcia-Corbacho J et al (2017) Liquid biopsies come of age: towards implementation of circulating tumour DNA. Nat Rev Cancer 17:223–238. https://doi.org/10.1038/nrc.2017.7
4. Siravegna G, Marsoni S, Siena S, Bardelli A (2017) Integrating liquid biopsies into the management of cancer. Nat Rev Clin Oncol 14:531–548. https://doi.org/10.1038/nrclinonc.2017.14
5. Dominguez-Vigil IG, Moreno-Martinez AK, Wang JY et al (2018) The dawn of the liquid biopsy in the fight against cancer. Oncotarget 9:2912–2922. https://doi.org/10.18632/oncotarget.23131
6. Stewart CM, Kothari PD, Mouliere F et al (2018) The value of cell-free DNA for molecular pathology. J Pathol 244:616–627. https://doi.org/10.1002/path.5048
7. Volik S, Alcaide M, Morin RD, Collins CC (2016) Cell-free DNA (cfDNA): clinical significance and utility in cancer shaped by emerging technologies. Mol Cancer Res 14:898–908. https://doi.org/10.1158/1541-7786.MCR-16-0044
8. Volckmar A-L, Sultmann H, Riediger A et al (2018) A field guide for cancer diagnostics using cell-free DNA: from principles to practice and clinical applications. Genes Chromosomes Cancer 57:123–139. https://doi.org/10.1002/gcc.22517
9. Jahr S, Hentze H, Englisch S et al (2001) DNA fragments in the blood plasma of cancer patients: quantitations and evidence for their origin from apoptotic and necrotic cells. Cancer Res 61:1659–1665
10. Schwarzenbach H (2013) Circulating nucleic acids as biomarkers in breast cancer. Breast Cancer Res 15:211. https://doi.org/10.1186/bcr3446
11. Kahlert C, Melo SA, Protopopov A et al (2014) Identification of double-stranded genomic DNA spanning all chromosomes with mutated KRAS and p53 DNA in the serum exosomes of patients with pancreatic cancer. J Biol Chem 289:3869–3875. https://doi.org/10.1074/jbc.C113.532267
12. Allenson K, Castillo J, San Lucas FA et al (2017) High prevalence of mutant KRAS in circulating exosome-derived DNA from early-stage pancreatic cancer patients. Ann Oncol 28:741–747. https://doi.org/10.1093/annonc/mdx004
13. Wang W, Kong P, Ma G et al (2017) Characterization of the release and biological significance of cell-free DNA from breast cancer cell lines. Oncotarget 8:43180–43191. https://doi.org/10.18632/oncotarget.17858
14. Bronkhorst AJ, Wentzel JF, Aucamp J et al (2016) Characterization of the cell-free DNA released by cultured cancer cells. Biochim Biophys Acta 1863:157–165. https://doi.org/10.1016/j.bbamcr.2015.10.022
15. Snyder MW, Kircher M, Hill AJ et al (2016) Cell-free DNA comprises an in vivo nucleosome footprint that informs its tissues-of-origin. Cell 164:57–68. https://doi.org/10.1016/j.cell.2015.11.050
16. Zhivotosky B, Orrenius S (2001) Assessment of apoptosis and necrosis by DNA fragmentation and morphological criteria. Curr Protoc Cell Biol Chapter 18:Unit 18.3. doi: https://doi.org/10.1002/0471143030.cb1803s12
17. Campos CDM, Jackson JM, Witek MA, Soper SA (2018) Molecular profiling of liquid biopsy samples for precision medicine. Cancer J 24:93–103. https://doi.org/10.1097/PPO.0000000000000311
18. Rodda AE, Parker BJ, Spencer A, Corrie SR (2018) Extending circulating tumor DNA analysis to ultralow abundance mutations: techniques and challenges. ACS Sensors 3:540–560. https://doi.org/10.1021/acssensors.7b00953
19. Mouliere F, Rosenfeld N (2015) Circulating tumor-derived DNA is shorter than somatic DNA in plasma. Proc Natl Acad Sci U S A 112:3178–3179. https://doi.org/10.1073/pnas.1501321112
20. Underhill HR, Kitzman JO, Hellwig S et al (2016) Fragment length of circulating tumor DNA. PLoS Genet 12:e1006162. https://doi.org/10.1371/journal.pgen.1006162

21. Peng M, Chen C, Hulbert A et al (2017) Non-blood circulating tumor DNA detection in cancer. Oncotarget 8:69162–69173. https://doi.org/10.18632/oncotarget.19942
22. Fleischhacker M, Schmidt B (2007) Circulating nucleic acids (CNAs) and cancer—a survey. Biochim Biophys Acta 1775:181–232
23. Lo YM, Zhang J, Leung TN et al (1999) Rapid clearance of fetal DNA from maternal plasma. Am J Hum Genet 64:218–224
24. Bettegowda C, Sausen M, Leary RJ et al (2014) Detection of circulating tumor DNA in early- and late-stage human malignancies. Sci Transl Med 6:224ra24. https://doi.org/10.1126/scitranslmed.3007094
25. Phallen J, Sausen M, Adleff V et al (2017) Direct detection of early-stage cancers using circulating tumor DNA. Sci Transl Med 9. https://doi.org/10.1126/scitranslmed.aan2415
26. Burgener JM, Rostami A, De Carvalho DD, Bratman SV (2017) Cell-free DNA as a post-treatment surveillance strategy: current status. Semin Oncol 44:330–346. https://doi.org/10.1053/j.seminoncol.2018.01.009
27. De Mattos-Arruda L, Mayor R, Ng CKY et al (2015) Cerebrospinal fluid-derived circulating tumour DNA better represents the genomic alterations of brain tumours than plasma. Nat Commun 6:8839
28. Wang Y, Springer S, Zhang M et al (2015) Detection of tumor-derived DNA in cerebrospinal fluid of patients with primary tumors of the brain and spinal cord. Proc Natl Acad Sci 112:9704 LP–9709709
29. Abbosh C, Birkbak NJ, Wilson GA et al (2017) Phylogenetic ctDNA analysis depicts early-stage lung cancer evolution. Nature 545:446–451. https://doi.org/10.1038/nature22364
30. Newman AM, Bratman SV, To J et al (2014) An ultrasensitive method for quantitating circulating tumor DNA with broad patient coverage. Nat Med 20:548–554. https://doi.org/10.1038/nm.3519
31. Riediger AL, Dietz S, Schirmer U et al (2016) Mutation analysis of circulating plasma DNA to determine response to EGFR tyrosine kinase inhibitor therapy of lung adenocarcinoma patients. Sci Rep 6:33505. https://doi.org/10.1038/srep33505
32. Sloan DB, Broz AK, Sharbrough J, Wu Z (2018) Detecting rare mutations and DNA damage with sequencing-based methods. Trends Biotechnol 36:729–740. https://doi.org/10.1016/j.tibtech.2018.02.009
33. Salk JJ, Schmitt MW, Loeb LA (2018) Enhancing the accuracy of next-generation sequencing for detecting rare and subclonal mutations. Nat Rev Genet 19:269–285. https://doi.org/10.1038/nrg.2017.117
34. Olmedillas-López S, García-Arranz M, García-Olmo D (2017) Current and emerging applications of droplet digital PCR in oncology. Mol Diagn Ther 21:493–510. https://doi.org/10.1007/s40291-017-0278-8
35. Camus V, Bohers E, Dubois S et al (2018) Circulating tumor DNA: an important tool in precision medicine for lymphoma. Expert Rev Precis Med Drug Dev 3:11–21. https://doi.org/10.1080/23808993.2018.1412798
36. Scherer F, Kurtz DM, Diehn M, Alizadeh AA (2017) High-throughput sequencing for non-invasive disease detection in hematologic malignancies. Blood 130:440–452. https://doi.org/10.1182/blood-2017-03-735639
37. Hiemcke-Jiwa LS, Minnema MC, Radersma-van Loon JH et al (2017) The use of droplet digital PCR in liquid biopsies: a highly sensitive technique for MYD88 p.(L265P) detection in cerebrospinal fluid. Hematol Oncol 36:429–435. https://doi.org/10.1002/hon.2489
38. Chase ML, Armand P (2018) Minimal residual disease in non-Hodgkin lymphoma—current applications and future directions. Br J Haematol 180:177–188. https://doi.org/10.1111/bjh.14996
39. Camus V, Jardin F, Tilly H (2017) The value of liquid biopsy in diagnosis and monitoring of diffuse large b-cell lymphoma: recent developments and future potential. Expert Rev Mol Diagn 17:557–566. https://doi.org/10.1080/14737159.2017.1319765
40. Assouline SE, Nielsen TH, Yu S et al (2016) Phase 2 study of panobinostat with or without rituximab in relapsed diffuse large B-cell lymphoma. Blood 128:185–194
41. Scherer F, Kurtz DM, Newman AM et al (2016) Distinct biological subtypes and patterns of genome evolution in lymphoma revealed by circulating tumor DNA. Sci Transl Med 8:364ra155. https://doi.org/10.1126/scitranslmed.aai8545
42. Kurtz DM, Green MR, Bratman SV et al (2015) Noninvasive monitoring of diffuse large B-cell lymphoma by immunoglobulin high-throughput sequencing. Blood 125:3679–3687
43. Roschewski M, Staudt LM, Wilson WH (2016) Dynamic monitoring of circulating tumor DNA in non-Hodgkin lymphoma. Blood 127:3127–3132. https://doi.org/10.1182/blood-2016-03-635219
44. Schmitt MW, Kennedy SR, Salk JJ et al (2012) Detection of ultra-rare mutations by next-

generation sequencing. Proc Natl Acad Sci U S A 109:14508–14513. https://doi.org/10.1073/pnas.1208715109

45. Kennedy SR, Schmitt MW, Fox EJ et al (2014) Detecting ultralow-frequency mutations by Duplex Sequencing. Nat Protoc 9:2586–2606. https://doi.org/10.1038/nprot.2014.170
46. Alcaide M, Yu S, Davidson J et al (2017) Targeted error-suppressed quantification of circulating tumor DNA using semi-degenerate barcoded adapters and biotinylated baits. Sci Rep 7:10574. https://doi.org/10.1038/s41598-017-10269-2
47. Lavoie J-M, Alcaide M, Fisher RA et al (2017) Targeted error-suppressed detection of circulating paternal DNA to establish a diagnosis of gestational trophoblastic neoplasm. JCO Precis Oncol:1–6. https://doi.org/10.1200/PO.17.00154
48. Alcaide M, Yu S, Bushell K et al (2016) Multiplex droplet digital PCR quantification of recurrent somatic mutations in diffuse large B-cell and follicular lymphoma. Clin Chem 62:1238–1247. https://doi.org/10.1373/clinchem.2016.255315
49. Gordon LI (2016) Precision monitoring by next-generation sequencing in lymphoma: circulating tumor DNA as a new biomarker. Oncology (Williston Park) 30:745–746
50. Spina V, Bruscaggin A, Cuccaro A et al (2018) Circulating tumor DNA reveals genetics, clonal evolution and residual disease in classical Hodgkin lymphoma. Blood 131:2413–2425. https://doi.org/10.1182/blood-2017-11-812073
51. Arthur SE, Jiang A, Grande BM et al (2018) Genome-wide discovery of somatic regulatory variants in diffuse large B-cell lymphoma. Nat Commun 9:4001. https://doi.org/10.1038/s41467-018-06354-3
52. Hung SS, Meissner B, Chavez EA et al (2018) Assessment of capture and amplicon-based approaches for the development of a targeted next-generation sequencing pipeline to personalize lymphoma management. J Mol Diagn 20:203–214. https://doi.org/10.1016/j.jmoldx.2017.11.010
53. Norton SE, Luna KK, Lechner JM et al (2013) A new blood collection device minimizes cellular DNA release during sample storage and shipping when compared to a standard device. J Clin Lab Anal 27:305–311. https://doi.org/10.1002/jcla.21603
54. Medina Diaz I, Nocon A, Mehnert DH et al (2016) Performance of streck cfDNA blood collection tubes for liquid biopsy testing. PLoS One 11:e0166354. https://doi.org/10.1371/journal.pone.0166354
55. Board RE, Williams VS, Knight L et al (2008) Isolation and extraction of circulating tumor DNA from patients with small cell lung cancer. Ann N Y Acad Sci 1137:98–107. https://doi.org/10.1196/annals.1448.020
56. Newman AM, Lovejoy AF, Klass DM et al (2016) Integrated digital error suppression for improved detection of circulating tumor DNA. Nat Biotechnol 34:547–555
57. MacConaill LE, Burns RT, Nag A et al (2018) Unique, dual-indexed sequencing adapters with UMIs effectively eliminate index cross-talk and significantly improve sensitivity of massively parallel sequencing. BMC Genomics 19:30. https://doi.org/10.1186/s12864-017-4428-5
58. Li YS, Jiang BY, Yang JJ et al (2018) Unique genetic profiles from cerebrospinal fluid cell-free DNA in leptomeningeal metastases of EGFR-mutant non-small cell lung cancer: a new medium of liquid biopsy. Ann Oncol 29:945–952. https://doi.org/10.1093/annonc/mdy009
59. Maggi EC, Gravina S, Cheng H et al (2018) Development of a method to implement whole-genome bisulfite sequencing of cfDNA from cancer patients and a mouse tumor model. Front Genet 9:6. https://doi.org/10.3389/fgene.2018.00006
60. Mussolin L, Burnelli R, Pillon M et al (2013) Plasma cell-free DNA in paediatric lymphomas. J Cancer 4:323–329. https://doi.org/10.7150/jca.6226
61. Primerano S, Burnelli R, Carraro E et al (2016) Kinetics of circulating plasma cell-free DNA in paediatric classical Hodgkin lymphoma. J Cancer 7:364–366. https://doi.org/10.7150/jca.13593
62. Sozzi G, Roz L, Conte D et al (2005) Effects of prolonged storage of whole plasma or isolated plasma DNA on the results of circulating DNA quantification assays. J Natl Cancer Inst 97:1848–1850. https://doi.org/10.1093/jnci/dji432
63. Schmitt MW, Fox EJ, Prindle MJ et al (2015) Sequencing small genomic targets with high efficiency and extreme accuracy. Nat Methods 12:423–425. https://doi.org/10.1038/nmeth.3351
64. Stoler N, Arbeithuber B, Guiblet W et al (2016) Streamlined analysis of duplex sequencing data with Du Novo. Genome Biol 17:180. https://doi.org/10.1186/s13059-016-1039-4

INDEX

Ralf Küppers (ed.), *Lymphoma: Methods and Protocols*, Methods in Molecular Biology, vol. 1956,
https://doi.org/10.1007/978-1-4939-9151-8,

S

T

U

V

W