Hematology Research Unit Helsinki,
Department of Clinical Chemistry and Hematology,
University of Helsinki and Helsinki University Hospital Comprehensive Cancer Center, Helsinki, Finland

Doctoral School in Health Sciences, Doctoral Program in Clinical Research, University of Helsinki

SOMATIC MUTATIONS IN AUTOIMMUNITY

Paula Savola

ACADEMIC DISSERTATION

To be presented, with the permission of the Faculty of Medicine of the University of Helsinki, for public examination in lecture hall 3, Biomedicum Helsinki 1 (Haartmaninkatu 8), on 15 February 2019, at 12 noon.

Helsinki 2018

Supervised by

Professor Satu Mustjoki, MD, PhD

Tiina Kelkka, PhD

Hematology Research Unit Helsinki,
Department of Clinical Chemistry and Hematology,
University of Helsinki and
Helsinki University Hospital Comprehensive Cancer Center,
Helsinki, Finland

Reviewed by

Docent Laura Pirilä, MD, PhD
Department of Rheumatology, Turku University Hospital, Turku, Finland

Docent Arto Orpana, PhD
Laboratory of Genetics, HUSLAB, Helsinki University Hospital, Helsinki, Finland

Opponent

Professor Lambert Busque, MD, FRCPC
Université de Montréal
Research Center and Hematology Division, Hôpital Maisonneuve-Rosemont Montréal, Quebec, Canada

Custos

Professor Satu Mustjoki, MD, PhD
Hematology Research Unit Helsinki, Department of Clinical Chemistry and Hematology,
University of Helsinki and
Helsinki University Hospital Comprehensive Cancer Center,
Helsinki, Finland

ISBN 978-951-51-4807-0 (paperback)

ISBN 978-951-51-4808-7 (PDF)

ISSN 2342-3161 (print)

ISSN 2342-317X (online)

Unigrafia
Helsinki 2018

The Faculty of Medicine uses the Urkund system (plagiarism recognition) to examine all doctoral dissertations.

S.D.G.
-Johann Sebastian Bach

TABLE OF CONTENTS

1 ABBREVIATIONS

AA	Aplastic anemia
ACPA	Anti-citrullinated protein antibody
ACR	American College of Rheumatology
ALPS	Autoimmune lymphoproliferative syndrome
AML	Acute myeloid leukemia
APC	Allophycocyanin
ARCH	Age-related clonal hematopoiesis
ASXL1	ASXL transcriptional regulator 1
BAFFR	B cell-activating factor receptor ; TNF receptor superfamily member 13C
BV	Brialliant violet
CBL	Cbl proto-oncogene
CCR7	C-C motif Chemokine Receptor 7
CD	Cluster of differentiation
CDCP1	CUB domain-containing protein 1
CHIP	Clonal hematopoiesis of indeterminate potential
CLEC	C-type lectin domain containing
CMV	Cytomegalovirus
COSMIC	Catalogue of Somatic Mutations in Cancer
CRP	C-reactive protein
CSF1	Colony-stimulating factor 1
CTLA4	Cytotoxic t-lymphocyte associated protein 4
CVID	Common variable immunodeficiency
CXCL10	C-X-C motif chemokine 10
DAS28	Disease Activity Score 28
DMARD	Disease-modifying anti-rheumatic drug
DMSO	Dimethyl sulfdioxide
DNA	Deoxyribonucleic acid
DNMT3A	DNA methyltransferase 3 alpha
DQB	DQ beta [in the context of HLA II]
DRB	Antigen D-related beta [in the context of HLA II]
EBV	Epstein-Barr virus
EDTA	Ethylenediaminetetraacetic acid
ESID	European Society for Immnodeficiencies
ESR	Erythrocyte sedimentation rate
EULAR	European League Against Rheumatism
FAS	Fas cell surface death receptor

FDR	False discovery rate
FITC	Fluorescein isothiocyanate
FUS	FUS RNA binding protein
FW	Forward
GATK	Genome Analysis Toolkit
GM-CSF	Granulocyte-macrophage colony-stimulating factor
GNAS	GNAS complex locus
GWAS	Genome-wide association study
HAQ	Health-assessment questionnaire
HGF	Hepatocyte growth factor
HLA	Human leukocyte antigen
hMDS	Hypoplastic myelodysplastic syndrome
HPRT	hypoxanthine phosphoribosyltransferase 1
ICON	International consesus document
ICOS	Inducible T cell costimulator
IGV	Integratice Genomics Viewer
IL	Interleukin
IL6R	Interleukin 6 receptor
IKZF1	Ikaros family zinc finger 1
iNKT	Invariant natural-killer T cell
IQR	Interquartile range
ITGA	Integrin alpha subunit
JAK	Janus kinase
LDL	Low-density lipoprotein
LGL	Large granular lymphocyte
LRBA	LPS responsive beige-like anchor protein
MCV	Mean corpuscular volume
MIP-1alpha	Macrophage inflammatory protein 1-alpha
MSH5	MutS Homolog 5
NFKB	Nuclear Factor Kappa B
NGS	Next-generation sequencing
NK	Natural killer
NPX	Normalized protein expression unit
OSM	Oncostatin-M
PADI	Peptidyl arginine deiminase
PAGID	Pan-American Group for Immune Deficiencies
PBMNC	Peripheral-blood mononuclear cell
PCR	Polymerase chain reaction
PD-L1	Programmed death 1 ligand 1

PE	Phycoerythrin
PE-Cy7	Phycoerythrin-Cyanine7
PerCp	Peridinin chlorphyll protein
PIGA	phosphatidylinositol glycan anchor biosynthesis class A
PIK3CD	Phosphatidylinositol-4,5-Bisphosphate 3-Kinase Catalytic Subunit Delta
PPM1D	protein phosphatase, Mg2+/Mn2+ dependent 1D
PTPN	Protein tyrosine phosphatase, non-receptor type
RA	Rheumatoid arthritis
RDW	Red-cell distribution width
RF	Rheumatoid factor
RNA	Ribonucleic acid
RV	Reverse
SE	Shared epitope
SF3B1	Splicing factor 3b subunit 1
SIR	Standardized incidence ratio
STAT	Signal transducer and activator of transcription
TACI	Transmembrane activator and CAML interactor; TNFRSF13B
TBS	Tris-buffered saline
TCR	T-cell receptor
TCRB	T-cell receptor beta
TET2	Tet methylcytosine dioxygenase 2
TNF	Tumor necrosis factor
TNFRSF13B	TNF receptor superfamily member 13B
TP53	Tumor protein p53
Treg	Regulatory-T cell
TRIM	Tripartite motif containing
Tris	Trisaminomethane
TWEAK	TNF superfamily member 12
VAF	Variant allele frequency
ZNF	Zinc finger protein

2 LIST OF ORIGINAL PUBLICATIONS

This thesis is based on the following original publications, which are referred to in the text by their Roman numerals:

I **Savola P*,** Kelkka T*, Rajala H, Kuuliala A, Kuuliala K, Eldfors S, Ellonen P, Lagström S, Lepistö M, Hannunen T, Andersson EI, Kumar Khajuria R, Jaatinen T, Koivuniemi R, Repo H, Saarela J, Porkka K*, Leirisalo-Repo M*, Mustjoki S*. Somatic Mutations in Clonally Expanded Cytotoxic T Lymphocytes in Patients with Newly Diagnosed Rheumatoid Arthritis. Nature Communications 2017, Jun 21;8:15869.

II **Savola P,** Brück O, Olson T, Kelkka T, Kauppi MJ, Kovanen PE, Kytölä S, Sokka-Isler T, Loughran TP, Leirisalo-Repo M, Mustjoki S. Somatic STAT3 mutations in Felty syndrome: an implication for a common pathogenesis with LGL leukemia. Haematologica. 2018 Feb;103(2):304-312.

III **Savola P**, Lungdren S, Keränen MAI, Almusa H, Ellonen P, Leirisalo-Repo M, Kelkka T, Mustjoki S. Clonal hematopoiesis in patients with rheumatoid arthritis. Blood Cancer Journal 2018, 8:69.

IV **Savola P**, Martelius T, Kankainen M, Huuhtanen J, Lundgren S, Koski Y, Eldfors S, Kelkka T, Keränen MAI, Ellonen P, Kovanen P, Kytölä S, Saarela J, Lähdesmäki H, Seppänen M, Mustjoki S. Somatic mutations in CD4+ and CD8+ cells in patients with immunodeficiency. Submitted.

*Indicates equal contribution

3 ABSTRACT

Autoimmune diseases are caused by a dysregulated immune response against self-antigens. They affect more than 5% of the population in the Western countries, but curative therapies do not exist. The complexity of autoimmune reactions is demonstrated in the context of immunodeficiency, in which patients have even higher rates of autoimmune disease than the general population, despite immune system defects. The molecular mechanisms that cause autoimmune diseases are mostly unknown.

In this thesis, we focused on T cells with the aim of characterizing the landscape of somatic mutations in immune-associated genes. We used rheumatoid arthritis (RA) patients and patients suffering from immunodeficiency and autoimmunity as disease models.

In study I, we collected a cohort of 82 newly diagnosed RA patients. We performed deep T-cell receptor (TCR) sequencing from 65 RA patients' CD8+ cells to characterize the TCR repertoire. We also sequenced CD4+ and CD8+ cells of 25 RA patients and 20 healthy controls with a targeted deep sequencing panel covering 986 immune-associated genes; results were complemented with exome sequencing in 3 cases. We discovered 30 novel somatic mutations in the CD8+ cells of 5 RA patients and one mutation in one of the healthy controls. Of the discovered mutations in RA, 30% affected cell proliferation and 20% were related to immune functions. Mutations were restricted to specific, expanded CD8+ memory T-cell populations.

Felty's syndrome (a rare form of RA) and large granular lymphocyte (LGL) leukemia often present with a similar patient phenotype. In study II, we showed that the diseases are unified by somatic, activating *STAT3* mutations in CD8+ cells. Felty's syndrome patients harbor similar *STAT3* mutations, with a prevalence (43%) comparable to previously described LGL leukemia cohorts. In addition, these 2 diseases shared similar cytokine profiles. Since both diseases are treated similarly, we suggest that these entities could be considered parts of the same disease continuum.

Study III investigated the prevalence of somatic mutations in hematopoietic stem cells (clonal hematopoiesis) in RA patients. Targeted deep sequencing from 59 RA patients' whole blood samples revealed

clonal hematopoiesis in 17% of cases. However, clonal hematopoiesis did not associate with clinical parameters in our study setting.

Study IV aimed to investigate whether somatic mutations associate with autoimmunity and lymphoproliferation in the context of immunodeficiency. We sequenced 2533 genes from CD4+ and CD8+ cells of 17 immunodeficiency patients. Immunodeficiency patients harbored 45 somatic mutations in 65% of cases. Mutations in tumor suppressor and oncogenes occurred in 35%, but mutations in genes affecting lymphocyte functions, inflammation, and cell proliferation were also common. Clonal hematopoiesis variants also existed in 24% of patients.

In summary, we have shown that somatic mutations in T cells and clonal hematopoiesis variants are common in RA and immunodeficiency patients. Our work provides a molecular link between autoimmune disease and cancerous processes, introducing a novel concept in the field of immunology.

4 TIIVISTELMÄ

Autoimmuunisairaudet ovat moninainen joukko sairaustiloja, joissa immuunijärjestelmä hyökkää omia kudoksia vastaan. Yli 5%:lla länsimaiden väestöstä on autoimmuunisairaus, mutta parantavia hoitoja näihin sairauksiin ei ole. Immuunijärjestelmän häiriöillä on monet kasvot: vaikka immuunijärjestelmä ei kykenisi torjumaan riittävän tehokkaasti taudinaiheuttajia, kuten immuunipuutostautien yhteydessä, se voi hyökätä omia kudoksia vastaan. Immuunipuutoksiin liittyykin korkeampi autoimmuunisairauksien esiintyvyys kuin väestössä keskimäärin. Autoimmuunisairauksia aiheuttavat molekyylitason mekanismit ovat useimmiten huonosti tunnettuja.

Tässä väitöskirjatyössä pyrimme kuvaamaan somaattisten mutaatioiden kirjoa immuunivastegeeneissä T-soluihin keskittyen. Käytimme nivelreuma-, immuunipuutos- ja autoimmuniteettipotilaita tautimalleina.

Ensimmäisessä osatyössä keräsimme 82 juuri nivelreumadiagnoosin saaneen potilaan kohortin. Sekvensoimme 65 potilaan CD8+ soluja selvittääksemme näiden solujen T-solureseptorien kirjoa. Sekvensoimme myös 25 nivelreumapotilaan ja 20 terveen yksilön CD4+ ja CD8+ -soluja 986 immuunivastegeeniä kattavalla sekvensointipaneelilla. Lisäksi käytimme eksomisekvensointia täydentämään kohdennettua paneelia kolmen potilaan kohdalla. Löysimme yhteensä 30 somaattista mutaatiota viiden reumapotilaan CD8+ -soluista ja yhden mutaation yhdeltä terveeltä yksilöltä. Nivelreumapotilaiden mutaatioista 30% oli geeneissä, jotka vaikuttavat solujen proliferaatioon ja 20% olivat immuunivasteita säätelevissä geeneissä. Mutaatiot esiintyivät ainoastaan spesifisissä, suurissa CD8-positiivisissa muisti-T-soluklooneissa.

Feltyn oireyhtymä on harvinainen nivelreuman alatyyppi, ja Feltyn oireyhtymäpotilaiden sekä suurten granulaaristen lymfosyyttien (LGL) leukemiapotilaiden ilmiasu on usein hyvin samankaltainen. Toisessa osatyössä osoitimme, että näissä molemmissa sairauksissa potilaita yhdistävä tekijä ovat somaattiset *STAT3*-mutaatiot CD8+ -soluissa. Feltyn oireyhtymäpotilailla oli samanlaisia *STAT3*-mutaatioita samantasoisella esiintyvyydellä (43%) kuin LGL-leukemiapotilailla aikaisempiin LGL-leukemiakohortteihin verrattuna. Lisäksi näillä kahdella taudilla oli samankaltaiset plasman sytokiiniprofiilit. Koska

näiden sairauksien hoitolinjatkin ovat samankaltaiset, ehdotamme, että nämä tautitilat ovat osa saman sairauden jatkumoa.

Kolmas osatyö selvitti somaattisten mutaatioiden esiintyvyyttä veren kantasoluissa (klonaalinen hematopoieesi) nivelreumapotilailla. Kohdennettu syväsekvensointi 59 nivelreumapotilaan kokoverinäytteistä osoitti, että klonaalista hematopoieesia oli 17%:lla potilaista. Klonaalinen hematopoieesi ei kuitenkaan liittynyt kliinisiin muuttujiin tässä tutkimusasetelmassa.

Neljäs osatyö tutki autoimmuniteetin, lymfoproliferaation ja somaattisten mutaatioiden yhteyksiä immuunipuutospotilailla. Sekvensoimme 17 immuunipuutospotilaan CD4+ ja CD8+ soluista 2533 geeniä. Immuunipuutospotilailta löytyi yhteensä 45 somaattista mutaatiota, ja 65%:lla potilaista oli mutaatioita. Tunnetuissa kasvunrajoite- ja onkogeeneissa oli mutaatioita 35%:lla potilaista, mutta mutaatiot lymfosyyttien toimintaan, tulehdusreaktioon, sekä soluproliferaatioon liittyvissä geeneissä olivat myös yleisiä. Klonaaliseen hematopoieesiin liittyviä geenimuutoksia esiintyi 24%:lla potilaista.

Tässä väitöskirjatyössä olemme osoittaneet, että sekä somaattiset mutaatiot T-soluissa että klonaalinen hematopoieesi ovat yleisiä nivelreuma- ja immuunipuutospotilailla. Autoimmuunisairauksissa ja syövässä on siis samankaltaisia molekyylitason löydöksiä. Tämä tuo esiin uuden näkökulman immuunisolujen toimintojen sekä immunologisten sairauksien tutkimukseen.

5 INTRODUCTION

The mechanisms that cause autoimmune diseases are incompletely understood. Immune cells that target self-antigens contribute to autoimmune disease pathogenesis, but which factors break immune tolerance and allow for the function of these autoreactive cells? The adaptive immune system is critical for most immune responses. T cells are key players in the adaptive immune response and have well-established roles in autoimmunity.

Rheumatoid arthritis (RA) is a relatively common autoimmune disease that affects approximately 0.5–1% of the population.[1] It is a chronic disease that confers increased mortality. In RA, the immune system attacks joint tissues, leading to joint pain and swelling.[2] The pathogenic roles of B cells and CD4+ T cells are acknowledged in RA,[2,3] but CD8+ T cells are potential disease modulators.[4]

Common variable immunodeficiency (CVID) is the most common immune deficiency disorder in the adult population. It is characterized by recurrent infections and insufficient levels of protective immunoglobulins,[5] but T cells are also implicated in CVID pathogenesis.[6] Current estimates for monogenic germline causes range from 2% to 10% of adult-onset CVID patients,[7,8] although monogenic causes are more common in pediatric patients.[9,10] Despite insufficient infection control, autoimmunity occurs in 30% of patients, and lymphoproliferation is also common.[5,11,12]

Somatic mutations are genomic alterations that are not passed to offspring. Somatic mutations in tumor suppressor genes and activating mutations in oncogenes contribute to tumorigenesis.[13] However, somatic mutations accumulate in multiple tissue types of healthy individuals during the human lifetime.[14] We hypothesized that somatic mutations may have a range of effects on immune cells: mutations in molecular signaling pathways may promote a proinflammatory cell phenotype and/or promote cell survival just enough to allow disruption of the immune balance.

Thus, we aimed to study the landscape and effects of somatic mutations in immune cells in autoimmune disease, with a focus on T cells. We used RA and immunodeficiency patients as models for our investigations. Although recent research has shown that somatic mutations accumulate with age in several cell types, mutational

landscapes and effects in the T cells of autoimmune disease patients are not known. In studies I–IV, we present novel discoveries of somatic mutations in T cells in the contexts of RA and immunodeficiency.

6 REVIEW OF THE LITERATURE

6.1 Rheumatoid arthritis

6.1.1 Epidemiology

Rheumatoid arthritis (RA) is a systemic autoimmune disease in which the immunity targets synovial joints. This results in joint inflammation, swelling, and pain.[1] The disease was described in the medical literature as early as the 17th century,[15] but the term "rheumatoid arthritis" was coined in 1876 by Sir Alfred Garrod.[16] The current prevalence of RA is 0.5–1% in the general population in different countries, including the Finnish population.[17] Incidence rates for RA range from 0.1 to 0.5 per 1,000 adult inhabitants[17] and increase with age.[18,19] RA affects women 3 times more often than it does men.[1]

Genetic factors and the presence of RA-associated autoantibodies are risk factors for RA.[1,2,20] Of environmental factors, smoking increases the risk for RA.[21,22] Research has suggested that periodontitis,[23,24] obesity,[25,26] and short education[26] are also risk factors and that oral contraceptive use[27] and breastfeeding[26] may have protective effects. Although studies on dietary factors often yield controversial results, such studies suggest that alcohol consumption, antioxidant-containing foods, and oily fish may protect from RA, whereas high coffee consumption may be a risk factor.[26]

6.1.2 Genetics and the shared epitope hypothesis

RA shows a clear familial predisposition: twin studies have produced an estimate of 60–70% heritability.[28,29] Of all genetic risks for RA, the human leukocyte antigen (HLA) loci account for 11–37%.[29,30] The HLA-DRB1 (human leukocyte antigen antigen-D related beta 1) locus harbors the most well-acknowledged risk alleles for RA. A subset of DRB1-01 and DRB1-04 alleles, but also DRB1*1001, and DRB1*1402, confer risk for RA.[31–34] The RA risk HLAs accommodate a shared amino acid motif at positions 70–74 of the amino acid binding groove, which allows them to recognize similar peptides. These HLAs are termed shared epitope (SE) HLAs.[35] More recently, Raychaudhuri *et al* showed that the amino acids at positions 11, 71, and 74 in DRB1 but also HLA-B position 9 and DPB1

position 9 amino acid polymorphisms account for most of the risk for RA explained by HLA.[36]

A study combining data from 22 genome-wide association studies (GWAS) identified 100 non-HLA risk loci for RA, and these account for approximately 5% of RA heritability.[37] Other than HLA, the most important risk gene for RA is *PTPN22* (protein tyrosine phosphatase, nonreceptor-type, 22). Other risk genes include *PADI4* (peptidyl arginine deiminase 4), *CD40* (cluster of differentiation 40), *IRF4* (interferon-regulatory factor 4), *IL6R* (interleukin 6 receptor), and more than 300 other genes.[37]

6.1.3 Pathogenesis of seropositive RA

RA is characterized by inflammation in the synovial joints that eventually damages joint cartilage and the underlying bone.[1,2] This inflammation is propagated by both innate and adaptive immune systems. As a hallmark of the adaptive immune response, autoantibodies occur in 50–70% of RA patients at diagnosis.[2] These include the relatively unspecific anti-immunoglobulin Fc region antibodies (rheumatoid factor, RF)[38] and anti-citrullinated protein antibodies (ACPAs), which are highly specific for RA.[38] The antibodies emerge at detectable levels up to 10 years before disease onset, thus demonstrating the long disease process in RA.[39,40] Recent work has also discovered anti-carbamylated peptide antibodies in over 45% of RA patients.[41]

Citrulline is a key component of the antigens which RA patients form antibodies against.[42,43] It is produced as a post-translational modification from the amino acid arginine by peptidylarginase deiminase (PADI) enzymes.[42] RA synovial fluid contains elevated amounts of these enzymes, mostly produced by neutrophils, in addition to a range of citrullinated proteins, including citrullinated vimentin, fibrinogen, collagen II, and alpha-enolase.[42,44] PADI enzymes and citrullinated proteins occur in larger quantities in the lungs of smokers than in those of non-smokers.[45,46] Bacteria that cause periodontal inflammation also promote citrullination.[42]

To elect an autoantibody response, CD4+ T cells have to recognize the target antigen in the context of HLA-II molecules. Some of the HLA-DRB1-04 SE alleles, which predispose to RA, bind citrullinated peptides (collagen II, vimentin, aggrecan, fibrinogen, and others) with higher affinities than native peptides, while others do not show preferential binding.[47] Both RA patients and controls harbor T cells that recognize

citrullinated RA epitopes in peripheral blood in similar frequencies,[48–51] but RA synovial fluid samples show increased frequencies of citrullinated alpha-enolase(326-340)-specific T cells.[48] In addition, the frequency of citrulline-specific T cells correlates with disease activity in RA,[51] and the frequency of memory CD4+ T cells that recognize citrullinated peptides is increased in RA.[52]

Animal models highlight the immunogenic properties of citrulline and the pathogenic role of autoantibodies: rats immunized with citrullinated collagen II present with arthritis earlier than animals immunized with the native protein,[53] and antibodies against citrullinated type II collagen can mediate arthritis in mice.[54] ACPAs also activate macrophages and osteoclasts, which promotes bone remodeling.[1,2]

The citrullination processes described above are epidemiologically linked to RA: smoking is a well-acknowledged and periodontitis a possible risk factor.[21–24] Moreover, smoking, silica, and textile dust are risk factors for ACPA-positive RA in the context of the shared epitope.[45,55,56] Thus, the initial anti-citrulline immune response is triggered in the lungs, mouth, or gut, but additional stimuli are needed to precipitate the immune reaction into the joints.[42,44,57]

In addition to evidence on the recognition of citrulline as an antigen and the discovery of the strong HLA association in RA, the important role of T cells in RA pathogenesis has been demonstrated by genetic studies. Genes conferring RA susceptibility are enriched in B-cell, T-cell, and cytokine signaling pathways.[37] Large datasets comprising 223 murine gene-expression profiles have shown that RA risk genes are enriched in genes expressed in CD4+ effector-memory cell subsets, as well as in specific CD8+ memory Tbet+ cells from the spleen and in one subset of invariant natural killer T (iNKT) cells.[58] SNPs associated with RA overlap with markers of promoter activity (H3K4me3) in CD4+ memory and CD4+ T-regulatory cells.[59] Thus, the genes with variants that confer risk for RA may mediate the risk by altering the function of specific immune cells. Although these “big” datasets do not elucidate the exact mechanisms that trigger RA, they can be used to prioritize variants with most likely biological impact,[60,61] and these genes can be prioritized in functional studies.

Most research in RA pathogenesis has focused on CD4+ and B cells because they are clearly implicated in RA pathogenesis via HLA-DRB1 association and autoantibodies,[1,2] but other immune cells also modulate inflammation in RA. CD8+ T cells may have both proinflammatory and

anti-inflammatory roles in arthritis,[4] as is also suggested by HLA-I RA risk polymorphims.[36,62] In addition to adaptive immune cells (B and T cells), other immune cell subsets—such as platelets and mast cells—may also play roles in RA pathogenesis.[63,64] Macrophages are major mediators of tissue damage and inflammation, as they present antigens and secrete cytokines, chemokines, proteases, vasoactive peptides, and reactive intermediates.[3,65,66] Also, stromal and synovial cells in joints are not bystanders in inflammatory processes. Proinflammatory synovial fibroblasts exist in RA,[67] and fibroblast-like synovial cells remodel the synovium via protease excretion and contribute to the local adaptive immune response via antigen presentation.[3] The immune pathogenesis of RA is extremely complex (Figure 1), and the exact pathogenic mechanisms and triggers are only partially known.

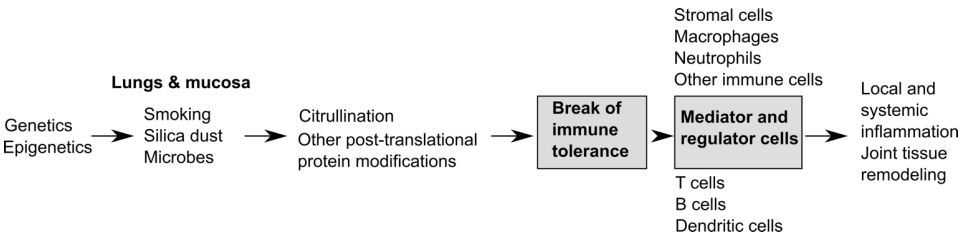

Figure 1 A schematic of the current hypothesis on the pathogenesis of ACPA-positive RA.

6.1.4 Pathogenesis of seronegative RA

Little is known of the pathogenesis of ACPA-negative RA. ACPA-negative disease may not always be autoantibody-negative, as approximately 16–24% of RA patients without ACPA or RF and 45% of patients with ACPA have autoantibodies against carbamylated peptides.[41,68] In addition, because many large genetic study cohorts contain either no or a minority of ACPA-negative subjects, the genetics of ACPA-negative RA is less well understood than that of ACPA-positive RA.[69–71] Despite showing a familial predisposition,[72] genetic studies in seronegative RA have yielded relatively few risk genes. Only the largest GWAS studies have discovered 3 non-HLA risk variants at a genome-wide significance level.[70,73] Fine-mapping of the HLA region has revealed risk factors at the amino acid level in HLA-B (HLA-I) and DRB1 (HLA-II).[62,74] Moreover, ACPA-positive and -negative disease may share some risk genes, although the

effects sizes of these risk genes differ.[75] Seronegative RA likely accommodates a heterogenous group of patients.

6.1.5 Diagnosis

RA diagnosis is aided by using the American College of Rheumatology (ACR) and European League Against Rheumatism (EULAR) 2010 classification criteria (Table 1).[76] The ACR/EULAR criteria emphasize inflammation in several small joints and high titers of ACPA or RF. RA should be suspected if other causes of joint swelling are not likely.[76] Re-evaluation is important if symptoms continue during follow-up, even if the criteria for RA were not fulfilled at first evaluation.[76] Compared to the previous ACR1987 criteria,[77] the new criteria are more sensitive in early RA and accommodate the usage of ACPA status in diagnostic decision-making.[78] Differential diagnostic options for RA include, but are not restricted to, reactive arthritis, psoriatic arthritis, gout, fibromyalgia, systemic lupus erythematosus, infectious arthritis, polymyalgia rheumatica and osteoarthritis.[2,79]

Table 1. *ACR/EULAR classification criteria for RA. Six or more points are required for RA diagnosis. Adapted from Aletaha* et al.[76]

			Score
1.	Joints	1 large joint	0
		2–10 large joints	1
		1–3 small joints	2
		4–10 small joints	3
		>10 joints (at least one small joint)	5
2.	Serology	Negative RF and negative ACPA	0
		Low–positive RF or ACPA	2
		High positive RF or ACPA	3
3.	Inflammation markers	Normal CRP and ESR	0
		Abnormal CRP or ESR	1
4.	Duration of symptoms	Less than 6 weeks	0
		More than 6 weeks	1

Abbreviations: CRP, c-reactive protein; ESR, erythrocyte sedimentation rate.

6.1.6 Treatment

Treatment of RA is based on immunosuppression by pharmaceutical agents. The mechanisms of action of synthetic disease-modifying anti-

rheumatic drugs (DMARDs), such as methotrexate, hydroxychloroquine, sulfasalazine, leflunomide, and parenteral gold, are mostly poorly known.[1,2] Methotrexate, the most commonly used DMARD, inhibits purine synthesis at high doses. Its anti-rheumatic mechanisms are unknown, although it may exert effects on multiple immune-cell subtypes.[80]

More novel drugs based on recombinant protein technology ("biologics") were introduced in the 1990s.[81] The largest group of biologics inhibit tumor necrosis factor (TNF) -alpha,[3] a cytokine produced by T cells, monocytes, and macrophages and a key mediator of inflammation. TNF-alpha promotes proinflammatory cytokine production, upregulates cell adhesion molecules, and modulates cartilage damage and bone resorption.[82] Other biologics target different cytokine pathways and/or immune-cell subsets: treatment strategies include IL-6 blockade, B-cell depletion (anti-CD20), and inhibition of costimulation (CTLA-4 agonist).[3] Novel synthetic DMARDs that inhibit Janus kinase (JAK) signaling have recently been introduced: as many proinflammatory cytokines signal via the JAK-STAT (signal transducer and activator of transcription) pathway, the inhibition of JAKs by tofacitinib (JAK1/3) and baricitinib (JAK1/2) has proven an efficient strategy.[3,83] Tofacitinib and baricitinib, often combined with methotrexate, have shown promise in patients refractory to TNF-alpha inhibitors and even efficacy when compared to methotrexate+adalimumab (a TNF-alpha inhibitor) in clinical trials.[84–87]

According to the current ACR and EULAR RA treatment guidelines, single-DMARD (preferably methotrexate) treatment is recommended as an initial strategy for RA.[88,89] However, the Finnish Medical Society current care guidelines recommend triple combination treatment (methotrexate +sulphasalazine +hydroxychloroquine) for active RA, even as first-line therapy.[90] Importantly, early remission of RA disease activity predicts future remission,[89,91] and thus efficient treatment is critical even in early RA.

6.1.7 Prognosis

Despite the increase in treatment options for RA during the past decades, RA patients have an increased risk for disability and mortality.[1,2] Other complications of RA include cardiovascular disease, vasculitis, interstitial lung disease, secondary amyloidosis and lymphoma.[1,2] The lymphoma risk is related to disease activity,[92] and the cardiovascular risk can be

attenuated by methotrexate and TNF-alpha inhibitors.[93] The mortality in RA associates with high disease activity, and thus treatment results in mortality reduction.[94,95] Early remission in RA activity predicts future remission,[89,91] and aggressive treatment initiation in early RA prevents work disability.[96] These facts underline the importance of efficient treatment to reduce RA-related complications. Indeed, mortality from RA has decreased during the last 50 years.[97]

6.1.8 RA and the risk for hematological malignancy

RA is associated with a 5-10% cancer risk increase when compared to controls.[98] RA treatments may affect the cancer risk, but disease activity often confounds study results.[98] RA confers risk for lymphoma, renal cancer and lung cancer.[98] However, conflicting results on cancer risk exist for many cancer types.

The risks for both non-Hodgkin lymphoma and Hodgkin lymphoma are increased in RA, with standardized incidence ratios (SIRs) ranging from 1.9 to 24.[99] Of lymphoma subtypes, diffuse large B-cell lymphoma is overrepresented in RA.[98,99] The pathogenesis of RA-associated lymphomas is not clear. Immunosuppressive treatment is not the most likely cause: despite the introduction of aggressive immunosuppressive treatment strategies, lymphoma rates in RA have not increased during the recent decades,[98] and DMARD usage does not increase lymphoma risk.[99] Nor do inherited genetic factors account for the increased lymphoma risk.[98,99] Although Epstein-Barr virus (EBV) can predispose to immunosuppression-associated lymphoproliferative disease, EBV-positive lymphomas form only 12-17% of RA-associated lymphomas.[92,100] However, the lymphoma risk in RA is clearly associated with high disease activity,[92,98,99] and immune stimulation by chronic inflammation in RA may be a key risk factor for lymphomagenesis in RA.[99]

RA also predisposes patients to leukemias. The overall SIR for leukemias is 1.3-1.7 in RA, and the SIR for acute myeloid leukemia (AML) ranges from 2.4 to 4.3.[101] Immune dysregulation in autoimmune disease may disrupt immune-surveillance, and the inflammatory environment in autoimmune disease can promote tumor growth.[101] In addition, chronic inflammation can promote mutagenesis in genomic deoxyribonucleic acids (DNA).[102] Some immunsuppressive treatments promote mutagenesis (mitoxantrone, cyclophosphamide, and azathioprine), but for most immunosuppressants, evidence of mutagenic potential or a predisposition for myeloid malignancy is lacking.[101,103]

6.1.9 RA as a co-manifestation of lymphoproliferative disease: the case of LGL leukemia

Large granular lymphocyte (LGL) leukemia is a rare, chronic hematological disorder characterized by an indolent disease course, LGL lymphocytosis in peripheral blood, and autoimmune manifestations. RA is very common in LGL leukemia, occurring in 11-36% of patients.[104,105] Neutropenia, thrombocytopenia, anemia, splenomegaly, rheumatoid arthritis and monoclonal gammopathy of undetermined significance are the most common manifestations of LGL leukemia.[104–107] A subset of patients are asymptomatic, and treatment indications are the alleviation of autoimmune manifestations and/or cytopenias, especially in the context of recurrent infections.[108] If treatment is needed, the first-line treatment option is immunosuppression, usually with low-dose methotrexate.[108]

The pathogenic cells in LGL leukemia are most commonly CD8+ T cells and, more rarely, CD4+ T cells; in approximately 10-20% of patients, the pathogenic cell clone consists of NK cells.[105,107,108] Interestingly, activating somatic *STAT3* mutations occur in 30-75% of LGL leukemia patients.[107,109–112] These mutations occur in the expanded lymphocyte clones, and patients can harbor multiple *STAT3* mutations.[113] In addition to *STAT3* mutations, *STAT5B* mutations occur in rare cases of LGL leukemia (2%)[114] and are overrepresented in CD4+ T-cell LGL leukemia.[115]

Autoimmune manifestations in LGL leukemia are linked to *STAT3* mutations: patients with *STAT3* mutations have more frequently neutropenia, anemia, and RA than patients with wild-type *STAT3*.[107,109] Patients harboring multiple *STAT3* mutations have even higher prevalence of RA (46%) than do patients without *STAT3* mutations (6%).[113] Although these associations suggest that *STAT3* mutations may play causative roles in autoimmune manifestations in LGL leukemia, the exact pathogenetic mechanisms remain unknown. STAT3 is critical for lymphocyte proliferation[116], it is constitutively activated in LGL leukemia,[110] and STAT3 activation has shown oncogenic potential.[112] Lamy and colleagues suggested that autoimmune manifestations are results of proinflammatory cytokine production, which is mediated by STAT3. Also, pathogenic cells may target joint and marrow cells.[108]

Large T-cell clones are not restricted to LGL leukemia. Healthy, elderly individuals harbor major T-cell clones that target cytomegalovirus,[117] and lymphocytosis may occur in many settings, such

as viral infections, organ transplantations, and autoimmune diseases.[106,108] Interestingly, persistent T-LGL expansions occur in 3.5% of RA patients. This occurrence is associated with TNF-alpha inhibitor usage.[118]

6.1.10 Felty's syndrome

Felty's syndrome, a rare form of RA, is defined by unexplained neutropenia, RA, and splenomegaly. Splenomegaly is not an essential criterion for diagnosis, as patients with the full Felty's syndrome triad and patients with RA and unexplained neutropenia closely resemble each other clinically.[119–121] Felty's syndrome affects approximately 1% of RA patients.[122] It usually develops in patients with longstanding RA, and patients often display severe joint destruction.[119] In addition, Felty's syndrome patients more often have extra-articular manifestations of rheumatoid disease than matched controls with RA, such as weight loss, pyrexia, and vasculitis.[121,123] Interestingly, clinical synovitis may be low in Felty's syndrome patients despite radiographic joint erosions.[119,121]

The Felty's syndrome triad overlaps with the clinical phenotype of many LGL leukemia patients,[124] and thus the differential diagnosis of these 2 diseases may be impossible (Table 2). Moreover, Felty's syndrome and LGL leukemia patients show similar histopathological findings and HLA-DR4 associations, and many Felty's syndrome patients have clonal T-cell LGL populations.[120,125,126] Some authors have suggested that LGL leukemia and Felty's syndrome are actually part of the same disease spectrum.[120,124,126] Indeed, it has been suggested that the first case of Felty's syndrome, described in 1924 by A.R. Felty, actually suffered from LGL leukemia.[104,127]

No standard treatment protocols exist for Felty's syndrome, and most studies on treatment are case reports or small case series. However, splenectomy, parenteral gold, methotrexate, granulocyte-macrophage colony-stimulating factor (GM-CSF) and biologics have been most commonly studied.[128–131] Methotrexate and/or GM-CSF are commonly recommended as first-line therapies for Felty's syndrome.[119,128,132]

The outcomes of Felty's syndrome patients may vary. Studies have reported both markedly high mortality rates (up to 36% in 5 years) as well as similar death rates with controls.[121,133,134] The primary causes of death are sepsis[121,134] and cardiovascular disease.[133] Felty's syndrome also

confers risk for cancers: non-Hodgkin lymphomas, leukemias, lung cancer, and melanoma show increased SIRs.[135]

Table 2. *Comparison of Felty's syndrome and LGL leukemia.*[108,119,136]

	LGL leukemia	Felty's syndrome
Rheumatoid arthritis (RA)	11-36%	Yes
Neutropenia	Up to d80%	Yes
Splenomegaly	20-50%	Often
Duration of RA	Usually not so many years	Usually many years
LGL cells in peripheral blood	Over $0.5*10^9$/l, but not an essential criterion	Not specified
T-cell receptor rearrangement	Yes (in T-cell LGL leukemia)	Not specified
Treatment	Immunosuppression, usually methotrexate	Immunosuppression, usually methotrexate

6.2 Common variable immunodeficiency

6.2.1 Etiology and genetics

Common variable immune deficiency (CVID) is a disease with diverse manifestations and phenotypes.[5,137] It is an umbrella diagnosis for antibody deficiencies with no other known cause, and thus patients diagnosed with CVID show diverse phenotypes.[137] The prevalence of CVID in the Finnish population is 5.5/100,000 inhabitants,[11] which is higher than in other national studies: CVID prevalences range from 0.6 to 3.8 per 100 000 inhabitants in other European countries.[138–144] The age at disease onset and diagnosis seems to have 2 peaks, one in childhood and one at 20-30 years of age,[145–147] but all cohorts have not confirmed this finding.[148] In adult patients, males and females are approximately equally affected,[145,147–152] while children show a male predominance.[145] Males also show an earlier median age at diagnosis.[145,146,149,150]

The cause of CVID is ultimately unknown. Most cases are sporadic, but 5-25% have a positive family history.[5] Estimated rates for monogenic forms of CVID range from 2 to 10%.[7,8] In pediatric patients, monogenic disease variants can be identified in 15-30% of cases.[9,10] In addition to

monogenic, highly pathogenic variants, some genetic variants occur in healthy individuals and confer a smaller risk increase for CVID (Table 3). Currently, CVID is perceived to likely be a multifactorial and/or polygenic disease.[5,7]

Table 3. *Highly pathogenic genetic variants and genes with variants predisposing to CVID. In the "GWAS risk genes" column, variants at genome-wide significance level are shown. The list of monogenic forms of CVID includes only examples, as a myriad of rare variants have been discovered.*

Monogenic forms of CVID/CVID-like disease[5,7]	Predispose to CVID[153–161]	GWAS risk genes[162–164]
PIK3CD	*TNFRSF13B* (*TACI*) variants	HLA (HLA-DQB1*02:01 and *05:03)
LRBA	*MSH5*	*CLEC16A*
ICOS	*BAFFR*	*ITGAM*
CTLA4		*FUS*
TWEAK		*ITGAX*
NFKB1		*TRIM72*
NFKB2		*ZNF646*
IKZF1		
CD19		
CD20		
CD21		
IL21		
IL21R		

6.2.2 Comorbidities of CVID

CVID is associated with various comorbidities, including recurrent infections, autoimmune disease, enteropathy, granulomatous disease, pulmonary disease, and lymphoproliferation.[5,137] Table 4 shows the prevalences of many common CVID comorbidities. Of autoimmune diseases, immune-mediated thrombocytopenic purpura (ITP) and autoimmune haemolytic anemia (AIHA) are the most common,[11,146,149] but other autoimmune manifestations also occur: examples include RA (3.2%),[149] vitiligo (5%),[148] and hypothyroidism (4%).[148] Pulmonary diseases are very common and include bronchiectasis, asthma, and interstitial lung disease.[11,165,166] Bronchiectasis are often due to recurrent infections, and the interstitial lung disease is treated as an autoimmune manifestation with immunosuppression.[5,12] Gastrointestinal

manifestations include entities such as chronic diarrhea,[147,149,150] pernicious anemia,[148] and malabsorption.[142,146,147] Lymphoproliferation is also a common co-morbidity of CVID. LGL lymphoproliferation or elevated LGL counts occurs in 3-41% of CVID patients.[11,167] CVID also confers risk for lymphoma at a SIR of 12.[168,169] The overall SIR for cancer in CVID is 1.8-1.9,[168,169] but up to 20% of CVID patients develop cancer in follow-up.[170]

Table 4. *Prevalences of comorbidities in CVID patients.*[11,142,145–147,149,150,165]

	% of patients
Autoimmune disease	22-30
ITP (immune-mediated thrombocytopenic purpura)	6-20
AIHA (autoimmune hemolytic anemia)	4-10
Granulomatous disease	8-10
Gastrointestinal disease	9-29
Splenomegaly	26-39
Lymphadenopathy	26-37
Bronchiectasis	11-43
Chronic pulmonary disease	27-46
GLILD (granulomatous-lymphocytic interstitial lung disease)	19-20
Solid tumor	5-8

6.2.3 Pathogenesis of CVID

The pathogenesis of CVID is likely as heterogenous as are the patient phenotypes. The disease etiology is apparent in the rare cases of monogenic CVID or CVID-like disease, but the causative mechanisms are unknown in most cases. Hypogammaglobulinemia, caused by insufficient immunoglobulin production by B cells, is a key characteristic of CVID. Isotype-switched memory B cells and bone-marrow plasma cells are reduced in CVID,[171–174] but B cells are absent in only 10% of patients.[12,172]

In addition to B-cell defects, CVID is characterized by the abnormal function of T cells. Studies have demonstrated reduced numbers of naïve CD4+ and CD8+ T cells,[175] regulatory-T cells,[176,177] increased apoptosis,[178]

a decrease in thymic output,[179] and abnormal cytokine production.[5,178] Deep T-cell receptor (TCR) sequencing studies have also demonstrated increased T-cell clonality in CVID,[180] especially in patients with low frequencies of switched-memory B cells and high frequencies of CD21-low cells.[175] This is in line with studies showing that thymic output and naïve CD4+ T cells are decreased in this same subset of patients.[181]

Autoimmune diseases are highly prevalent in CVID, occurring with a prevalence of 20-30%.[5,137,182] Interestingly, although patients produce insufficient amounts of immunoglobulins against pathogens, they retain the ability to produce autoantibodies.[183] Other immunologic defects, such as decreased regulatory-T cells (Tregs), associate with autoimmunity in CVID: Tregs are decreased in CVID patients with granulomas, autoimmunity, or splenomegaly.[177,184] The frequency of Tregs also has an inverse correlation with CD21-low B cells,[177] which may harbor autoreactive B-cell receptors.[185] Constant exposure and ineffective pathogen clearance may promote autoimmune responses via molecular mimicry or bystander-activation.[186,187] Increased survival factor levels in CVID may also disturb lymphocyte homeostasis.[187]

6.2.4 Diagnosis, treatment, and prognosis

Several diagnostic criteria have been proposed for CVID: European Society for Immunodeficiencies (ESID) and Pan American Group for Immune Deficiencies (PAGID) 1999 criteria,[188] Ameratunga criteria,[189] International consensus document (ICON) criteria,[5] and revised ESID registry 2014 criteria.[190] Different criteria have distinct requirements for CVID, but they have common factors: disease onset should be later than 2-4 years of age, immunoglobulin levels should be low, other causes for hypogammaglobulinemia should be ruled out, and patients should show poor response to vaccines. The effort to produce multiple diagnostic criteria reflects the challenge of defining CVID: the diagnosis accommodates heterogenous patient phenotypes, and CVID may actually comprise multiple disease entities.

Treatment of CVID includes subcutaneous or intravenous immunoglobulin replacement therapy and the management and monitoring of comorbidities.[5,137] Although placebo-controlled trials do not exist, historical comparisons have shown that immunoglobulin treatment decreases infections and pneumonias in CVID.[191,192] In addition, monitoring comorbidities is crucial. Comparisons of historical cohorts with more recent data have shown that survival after CVID

diagnosis has improved from a 12-year survival rate of 30% to a 64-67% rate 20 years after diagnosis.[5,146] The most common causes for death are infections, cancer, chronic lung disease, and liver disease.[147–149] Patients with disease manifestations other than infection susceptibility have higher mortality than patients with infections only,[148,149] although not all studies have confirmed this association.[145]

6.3 Somatic mutations in health and disease

6.3.1 Patterns of somatic mutations in healthy tissues

Somatic mutations are well-acknowledged contributors to cancer pathogenesis.[13] However, increasing evidence indicates that healthy individuals' cells acquire somatic mutations over time and that all somatic events do not lead to cancer. The "earliest" timepoint for somatic mutations occurs during embryogenesis: mutations occurring in dividing stem cells are passed to all daughter cells and multiple tissue types.[193,194]

The first estimates of the rate of somatic mutation accumulation were based on studying only small genomic areas, such as the *HPRT* or *PIGA* genes.[195,196] More recent studies using next-generation sequencing (NGS) have estimated that mutation rates differ between tissues, ranging from $3.5^{*}10^{-9}$ (small intestine) to $1.57^{*}10^{-7}$ (skin) mutations per base-pair per cell division.[197] The frequency of somatic mutations increases with age: human stem cells in the colon, small intestine, and liver acquire approximately 36 mutations per year.[198] Blood cells also accumulate mutations during the human lifetime, with cord-blood samples having the smallest numbers of somatic mutations.[199] In addition to these tissues, skin, pancreas, skeletal muscle satellite, brain, and kidney cells accommodate somatic mutations.[200–202]

Somatic mutations are enriched in non-coding areas of the genome in most of the studied tissues: fibroblasts, liver, colon, small intestine, and skeletal muscle satellite cells.[198,202,203] In contrast, one study sequenced single human neurons and reported the enrichment of somatic mutations in exonic areas.[204] Variations in sequenced tissues, mutagenic exposures, and sequencing technologies may explain the differing results.

The somatic mutational spectra and preferentially mutated nucleotides differ between tissue types. Approximately half of all single-nucleotide mutations in blood and intestinal cells are C>T (G>A) alterations.[198,199] Kidney, liver and brain tissue have smaller frequencies of C>T mutations than do blood and intestinal cells.[198–200] C>T mutations

occur more commonly in methylated CpG sites, when a methylated cytosine deaminates into a thymine.[205] As demonstrated by C>T mutations, different mutagenic processes cause distinct mutational profiles –mutational signatures– that depend on the nucleotide context.[206] Aging-associated signatures in healthy tissues are the most prevalent signatures in skin fibroblasts, skeletal muscle satellite cells, liver and gut cells, but all mutations were not attributed to these signatures.[198,202,207,208] Thus, even though aging is a major contributor to somatic mutations in healthy tissues, other mechanisms also likely modulate the process.

Aging-related somatic mutations may have varying effects in different tissue types. In skeletal muscle satellite cells, the number of mutations correlated inversely with differentiation and proliferation capacity. Also, mutations in the elderly individuals were enriched in exons and had predictably more frequent deleterious consequences than in young individuals. This suggests that mutations may affect aging-associated muscle impairment.[202] In contrast, somatic mutations that provide a survival advantage are common in hematopoietic stem cells.[209] Aged hematopoietic stem cells have decreased engraftment capacity, and thus cells with mutations that promote proliferation will be enriched and will produce a larger proportion of blood cells.[210] Indeed, somatic mutations seem to be an inevitable consequence of aging, and they may have varying effects on tissue renewal.

6.3.2 Somatic mutations in non-malignant disease

Recent research has revealed some examples of how somatic mutations can modulate non-malignant disease. The first description of somatic mutations causing non-malignant disease occurred in 2004: somatic *FAS* mutations in hematopoietic precursor cells in autoimmune lymphoproliferative syndrome (ALPS).[211] ALPS patients present with lymphoproliferation, autoimmune blood cytopenias, hypergammaglobulinemia, and an accumulation of double-negative T cells.[212]

Oncogenes and/or tumor suppressor genes seem to influence non-malignant disease in several cases: activating *KRAS* mutations have been observed in arteriovenous malformations in the brain.[213] Also, deeply infiltrating endometriotic lesions harbor exonic somatic mutations in approximately 80% of cases, with mutations in cancer driver genes (such as *ARID1A*, *PIK3CA*, and *KRAS*) in 26% of lesions.[214] Somatic mutations

with unknown biological significance in autoimmune disease have been described in refractory celiac disease and multiple sclerosis patients: Ettersperger *et al* discovered *STAT3* mutations in refractory celiac disease patients' intraepithelial lymphocytes,[215] and CD8+ cells from patients with multiple sclerosis are also shown to harbor somatic mutations.[216] Some studies dating to the pre-NGS era attempted to examine somatic mutations in RA patients' synovial cells and T cells, but due to technical issues and the lack of both germline controls and/or genomic DNA sequencing,[217–221] the full extent and biological meaning of somatic mutations in RA remain to be explored.

6.3.3 The origin and prevalence of clonal hematopoiesis

If a somatic mutation that provides a survival advantage occurs in a hematopoietic stem cell, the cell progeny can be detected in a disproportionaly large frequency in whole blood (Figure 2).[210] The first discoveries of this process, called clonal hematopoiesis, used chromosomal anomalies, copy number variations, and skewed X-chromosome inactivation of genomic aberrations in healthy individuals. Skewed X-chromosome inactivation in blood cells was discovered in the 1980s and 1990s in women with no blood cancer, and the prevalence of skewed X-chromosome inactivation increased with age in healthy female subjects.[222] In addition, the prevalence of copy number variations and chromosome anomalies increases with age in blood cells.[223,224] Subsequent studies using NGS sequencing revealed that the number of somatic small nucleotide variants in blood cells increases with age,[199] and somatic *TET2* mutations were discovered in 5.6% of female subjects with skewed X-chromosome inctivation.[225]

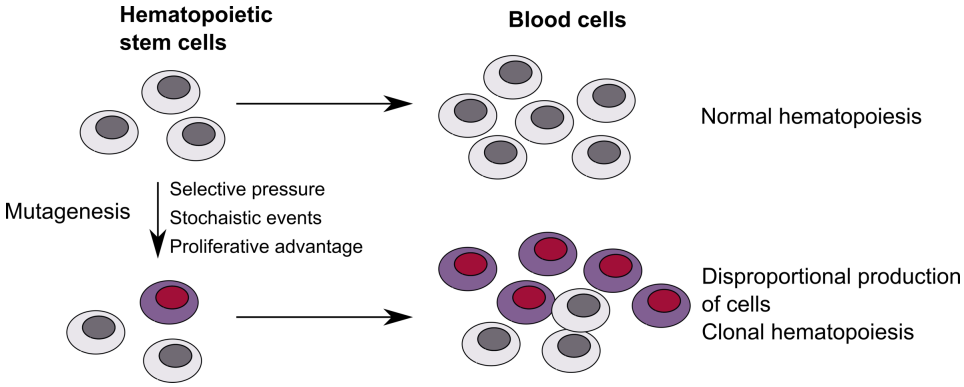

Figure 2 Clonal hematopoiesis is detected in peripheral blood if somatic mutations in a hematopoietic stem cell are passed to an unusually large proportion of peripheral blood cells.

Exome sequencing studies with large cohorts of hematologically unselected patients have reported a clonal hematopoiesis prevalence of approximately 10% in the elderly population (Table 5),[226,227] and subsequent studies using more sensitive targeted deep sequencing panels and whole-genome sequencing have reported higher prevalences (Table 5) that rise to more than 40% in >90-year-olds.[209,228,229] The more sensitive the method, the higher the frequency of clonal hematopoiesis: extremely sensitive sequencing methods have identified mutations associated with clonal hematopoiesis in 95% of 50- to 60-year-old women.[230] In all of these studies, the prevalence of clonal hematopoiesis increased with age.

Most patients harbor only one mutation, but the proportion of patients with more than one mutation increases with age.[209,229] The most common genes with mutations in unselected individuals include *DNMT3A*, *TET2*, *ASXL1*, *TP53*, *PPM1D*, *CBL*, *JAK2, GNAS,* and *SF3B1*,[209,226–229,231,232] where *DNMT3A* and *TET2* mutations are most common.[209,227,229,231,232] These are all candidate cancer driver genes, as they are associated with hematological malignancies.[210,233] The mutations in clonal hematopoiesis can be traced to hematopoietic stem cells, but mutations do not occur in similar frequencies in all consecutive progenitor cells or peripheral blood cells.[209] As an example, *TET2* mutations occur in higher variant allele frequencies (VAFs) in myeloid and NK cells but are nearly absent in T cells.[209,234] *DNMT3A* mutations also have lower VAFs in T cells than in other blood cells, but they occur in higher VAFs in T cells than do *TET2* mutations.[209]

Most studies on clonal hematopoiesis have employed targeted sequencing methods to discover mutations in established driver genes or have focused on candidate driver in exome sequencing data. Although mutations in cancer driver genes such as *DNMT3A* and *TET2* are more likely to provide a proliferative advantage to cells, and are thus more likely to manifest as clonal hematopoiesis, mathematical modeling has suggested that even in the absence of driver mutations, stochastic processes can lead to clonal hematopoiesis.[231] Therefore, clonal

hematopoiesis seems to be an extremely common phenomenon that shapes the aging hematopoietic tissue.[210]

Table 5. *The prevalence of clonal hematopoiesis in different studies. Abbreviations: WES, whole-exome sequencing; WGS, whole-genome sequencing; Targeted, targeted sequencing panel.*

		Cohort age range	Overall prevalence	Prevalence in >65-69 -year-olds	No. of individuals
Genovese *et al* 2014[226]	WES	19-93	4%	10%	>12000
Jaiswal *et al* 2014[227]	WES	19-108	4%	10%	>17000
Zink *et al* 2017[231]	WGS	All ages	12.5%	Over 20%	>11000
Arends *et al* 2018[209]	Targeted	55-98	28%	Over 29%	437
Buscarlet *et al* 2017[229]	Targeted	55-101	14%	15%	2530
Acuna-Hidalgo *et al* 2017[228]	Targeted	20-69	10%	Over 20%	2014

6.3.4 Biological consequences of clonal hematopoiesis

Clonal hematopoiesis predisposes for hematological malignancy, and this association has been replicated in multiple large studies.[226,227,231] In addition, clonal hematopoiesis confers the risk for death and cardiovascular disease. [226,227,235] Other associations between clonal hematopoiesis and clinical parameters are listed in Table 6.

Table 6. *Clonal hematopoiesis and its association with different clinical parameters. Abbreviations: WES, whole-exome sequencing; WGS, whole-genome sequencing; T2D, type 2 diabetes. Clonal hematopoiesis also uniformly associates with age in multiple studies.*

	Association	Method	No. of participants in the study
Jaiswal *et al* 2014[227]	Hematological malignancy; death; cardiovascular disease; increased RDW; no effect on blood-cell counts; T2D	WES	>17000 (3107 with blood cell count data)
Genovese *et al* 2014[226]	Hematological malignancy; death; smoking	WES	>12000
Zink *et a*l 2017[231]	Hematological malignancy; death; smoking; increased blood cell counts; treatment for addiction; psychiatric disease; smoking-related diseases; chronic pulmonary disease	WGS	>11000
Arends *et al* 2018[209]	High neutrophil counts; peripheral artery disease	Targeted sequencing	437
Buscarlet *et al* 2017[229]	Chronic obstructive pulmonary disease – asthma (*TET2* mutations)	Targeted sequencing	2530
Abelson *et al* 2018[236]; Desai *et al* 2018[237]	AML (case-control study); no effect on blood cell counts; AML risk depends on the gene with the mutation and VAF. *TP53* mutations confer the highest risk for AML	Targeted sequencing	95 cases + 414 controls; 212 cases + 212 controls

Many associations of clonal hematopoiesis and clinical parameters lack evidence for a causal relationship. However, the relationship between clonal hematopoiesis and cardiovascular disease risk is backed by mechanistic evidence: deficiency of Tet2, one of the most commonly mutated genes in clonal hematopoiesis, in myeloid cells alone promotes atherosclerosis and proinflammatory cytokine production in a mouse model of atherosclerosis.[238] In humans, clonal hematopoiesis confers a

1.9 odds ratio for coronary heart disease and is associated with increased coronary artery calcification.[235]

As clonal hematopoiesis is often detected with extremely low VAFs, it is not always clear which mutations are likely to have clinical significance. The issue complexity is increased by the fact that mutations in different genes confer differing effects, at least in AML risk.[236,237] The definition of clonal hematopoiesis of indeterminate potential (CHIP) comprises cases in which mutations exist in a hematological malignancy-associated gene with a VAF of more than 2%, but the patient does not fulfill criteria for a hematological disorder.[233] Additional definitions such as age-related clonal hematopoiesis (ARCH) are also used. However, because mutations with even smaller VAFs are detectable by modern methods,[228,230] and because clonal hematopoiesis with unknown drivers also confers risk for hematological malignancy,[226,231] the definitions will likely continue to evolve.

6.4 Next-generation sequencing

Currently, DNA sequencing is based on so-called "next-generation sequencing" (NGS), also called massive parallel sequencing. The first NGS platform sequencer was introduced in 2005,[239] and sequencing technologies have undergone rapid development since then. The first human genome sequence was completed in 2001-2004,[240–242] but it was followed by human genome, exome, and tumor genome resequencing as soon as in 2007-2009.[239,243]

Several NGS platforms exist from different commercial providers, but this thesis comprises data that has been obtained with the Illumina sequencing-by-synthesis technology, sequenced during 2013-2017. Sequencing instrument and sequencing library preparation method development has continued during these years. However, high-resolution sequencing assays are still accompanied by increased costs, and thus these assays often target specific genomic areas of interest and not the entire genome.

In this thesis, we have aimed to detect mutations that occur only in a small proportion of sequenced cells, and we have sequenced only selected genomic areas. Studies I and IV utilized targeted, hybridized capture methods, but the capture probes were updated in the years between these studies. Studies II and III targeted much smaller genomic areas than

studies I and IV, and thus PCR-based methods were used to enrich the genomic areas of interest.

6.4.1 Exome sequencing

The human genome nucleic acid sequence was resolved in 2001–2004.[240–242] Since then, sequencing technologies have evolved rapidly, and sequencing costs have decreased substantially: the cost to sequence an entire human genome in 2017 was less than 1 500 US dollars, while the cost for the first human whole-genome sequence in the Human Genome Project was 500–1000 million US dollars.[244,245] However, sequencing only protein-coding areas (exons) of the genome has been a cost-effective and popular strategy to identify variants that are relevant in disease. Exons span more than 30 million base-pairs, which comprises only approximately 2% of the total genome.[243,246] Exome sequencing applications include the discovery of rare monogenic disease variants and somatic variants in cancer cells,[247] as variants that alter protein structure have more apparent consequences for disease.[248]

6.4.2 Somatic variant calling

Raw sequencing data does not provide useful information on genomic variation. Sequencing data pre-processing includes steps to remove low-quality data and non-genomic sequences, such as adapter sequences. To detect genomic variants, sequencing data is aligned to a reference genome, and variants that differ from the reference sequence are reported. The most commonly used tools for alignment include the Burrows-Wheeler Aligner[249] and Bowtie2.[250] In exome- and other probe-capture technologies, duplicate reads produced by polymerase-chain reaction (PCR) are filtered based on genomic coordinates.[251,252]

Multiple bioinformatic tools exist for somatic variant calling: Genome Analysis ToolKit (GATK)[253] MuTect,[254] MuTect2,[254] VarScan2,[255] Strelka,[256] and others.[257] These tools use different statistical methods to detect somatic variants, and the GATK toolset also performs also local realignment in addition to mutation calling. Tool selection is based on the dataset composition: some tools, such as Seurat and SomaticSniper, are optimized for low-coverage samples, and some, such as VarScan2 and MuTect2, can detect VAFs with very low frequency.[257] Also, although most tools use another sample from the same individual as a background control to filter non-somatic variants, some tools allow variant calling without a paired sample.

Several studies have compared different variant callers. Before major variant filtering steps, the majority of putative variants are detected by only one variant caller.[258,259] However, variant caller parameters and customizable filtering steps significantly affect the rate of true mutation discovery. Some callers, such as Mutect and Strelka, show a higher overlap in results (0.8) than others.[259] Indeed, in a study comparing 4 callers, Strelka and MuTect2 seem to have the highest rate of validated mutations (72-94%).[258] Another study also suggested that MuTect and Strelka show good sensitivity with the smallest false positive rates.[260] Problems arise when aiming to detect mutations with low VAFs: in a study by Sandmann *et al*, when calling mutations with <5% VAF, the positive predictive value for almost all studied callers was nearly zero.[261] Many of these aforementioned comparative studies have yielded differing results because investigators have used differing variant calling parameters in the calling tools.[257]

6.4.3 Sensitivity issues in somatic variant discovery

Somatic variants may not occur in high frequencies, as all somatic variants do not occur in all sequenced cells. Sensitive sequencing methods are often required to detect a variant of interest, especially in the context of samples with polyclonal cells. One way to increase sensitivity in a sequencing experiment is to sequence higher numbers of unique DNA molecules in the initial sequencing assay, which can be called "increasing the sequencing depth." Generally, a variant must be supported by multiple sequencing reads for a mutation call, and thus an increase in sequencing depth can identify variants that occur only in a small proportion of sequenced cells (Figure 3).

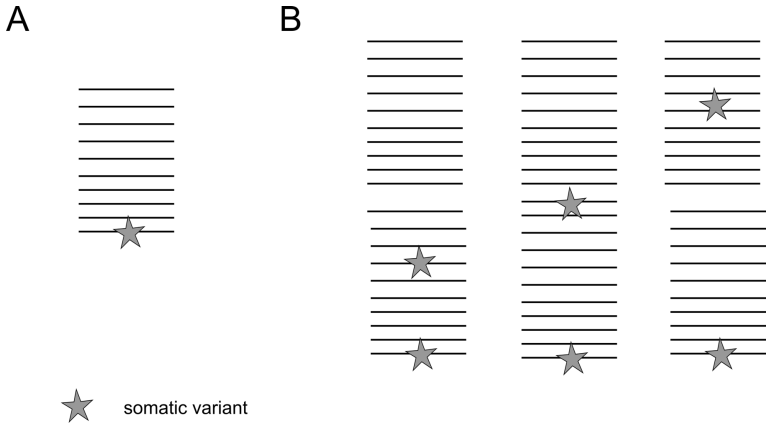

Figure 3 Sequencing depth affects the sensitivity of the sequencing assay. In both A and B, the variant occurs in 10% of sequenced reads. A. With a "depth" of 10, the variant is supported by only one read and thus cannot be called with high confidence. B. With a "depth" of 60, the variant is supported by 6 reads, which increases the confidence level for the variant.

Sequencing costs increase rapidly with increasing depths if the assay targets large genomic areas. Therefore, sequencing approaches that target only specific genomic areas are common when studying variants with very low VAFs. Targeted approaches are often based either on (1) hybridized capture or (2) amplified capture (Figure 4).[262] Hybridized capture methodology is essentially similar to exome capture, with the exception that instead of using exome-specific probes, customized probes that target only the genomic areas of interest are used. Amplified capture utilizes PCR primers to amplify genomic regions of interest.

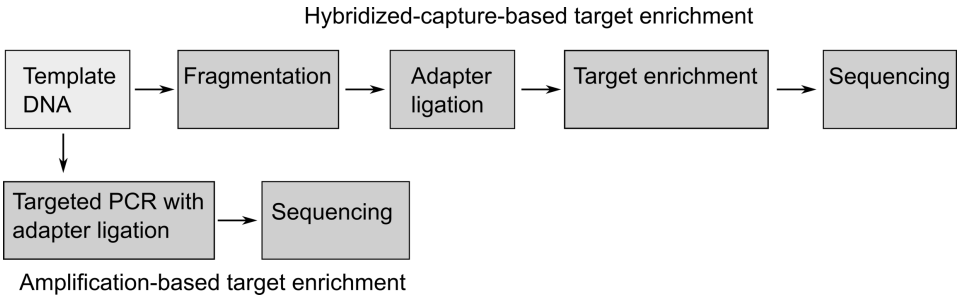

Figure 4 Simplified workflows of hybridized and amplified capture sequencing approaches.

However, when aiming to detect very low-frequency variants, the difficulty of distinguishing true signal from noise in the sequence data exists and cannot be solved simply by increasing sequencing depth. Sequencing instruments are not error free, and they have their own error profiles: Illumina HiSeq 2000-2500 and Genome Analyzer IIx systems have an approximately 10^{-2} to 10^{-3} substitution rate, and a lower error rate for indels,[263–265] but the error rate increases toward the 3' end of reads.[265] The GC content of the sequenced genomic area affects the error rate, with extreme GC contents in particular increasing the indel error rate.[264] GC extremes (GC content <10% or over 75–85%) result in lower relative converage.[264] Sequencing library preparation methods also introduce errors in the sample DNA, especially C>A and G>T alterations,

which may resemble “real” low-frequency variants.[266,267] In addition to errors introduced in library preparation and sequencing, analyzing steps such as genome alignment, may introduce false positive results. Especially alignment of indel-containing sequences is challenging.[262]

Specialized approaches exist to detect variants that occur in approximately 0.1–1% of sequenced cells. For example, unique molecular identifiers (UMIs) are nucleotide sequences that are attached to the sequenced DNA molecules, and each original DNA molecule is labeled with a different UMI. Thus, amplification errors and PCR duplicates can be filtered based on UMIs.[268] A sensitivity of 0.1% may be achieved with UMI technology.[269]

6.4.4 TCR sequencing

T cells are key mediators and regulators of the adaptive immunity, and each T cell has a unique TCR on its cell membrane that recognizes antigens presented by HLA molecules. Most TCRs consist of an alpha and a beta chain, which are encoded by variable (V), diversity (D) and joining (J) gene segments in the TCR alpha or beta chain genomic loci.[270] Each locus contains multiple gene segments of each category.

The minimum number of unique alphabeta T cells in humans is estimated to be $25*10^6$.[271] This great diversity of T cells is not produced by germline-encoded DNA. During T cell maturation in thymus, cells undergo somatic recombination to ensure that they recognize a large spectrum of possible antigens. In the TCR beta locus (TCRB), one of the V-, D-, and J-gene segments are selected and combined by recombining enzymes. Non-templated nucleotides insertions or deletions may occur in gene-segment junctional regions (Figure 5).[270] The TCR alpha locus undergoes a similar procedure, except the alpha locus does not contain D segments.[270]

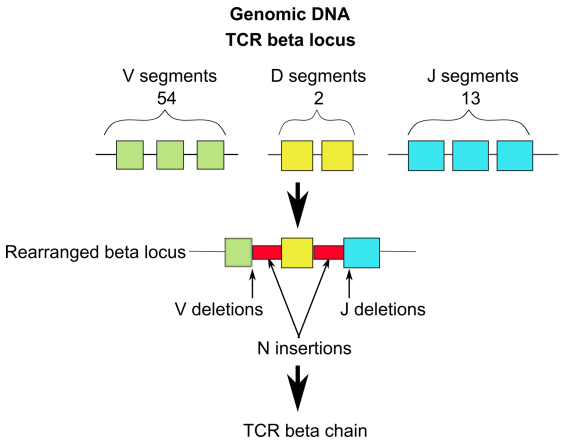

Figure 5 Somatic recombination of the TCR. The TCR beta locus is shown as an example.

Somatic recombination produces stable rearrangements in the genomic DNA, and thus these rearrangements can be detected by sequencing. Any sample that contains T cells also contains TCR rearrangements. However, due to the complexity and high variability in the TCR chain loci, characterizing TCR rearrangements in detail but with accuracy is not trivial. Determining the TCR sequence from sequencing libraries prepared from RNA samples is a common approach, as RNA is abundant and includes only functional rearrangements.[272,273] Also due to RNA splicing, some techniques (5' RACE with template-switch effect) require only one primer pair to amplify all rearrangements.[272] However, this does not allow exact quantification of T cells in the sequenced sample, since different T cells can have differential TCR expression.[273] Sequencing genomic DNA requires multiple primers to picture the TCR diversity, and this approach is prone to amplification biases.[272] However, amplification bias in PCR can be minimized by primer optimization[274] or by using unique molecular identifiers.[275]

7 AIMS OF THE STUDY

The aim of this PhD thesis is to characterize the landscape of somatic mutations in immune-associted genes in autoimmune disease using NGS methods. We use RA and immunodeficiency patients as disease models for our study, and we have studied somatic mutations in mature T cells as well as mutations that originate in hematopoietic progenitor cells.

This entailed 4 projects :

1. Characterization of somatic mutations in mature T cells in newly diagnosed RA patients
2. Elucidation of the common molecular pathogenesis of LGL leukemia and Felty's syndrome by studying somatic *STAT3* mutations
3. Identification and characterization of clonal hematopoiesis in RA patients and its associations with clinical parameters
4. Identification of somatic mutations as possible contributors to immunodeficiency and/or immunodeficiency-related comorbidities

8 PATIENTS AND METHODS

8.1 Patients and ethics

All studies were reviewed and approved by the local institutional review boards. Patients gave written informed consent, and the principles of the Helsinki declaration were followed.

8.1.1 Newly diagnosed RA patients (I)

Newly diagnosed RA patients (n=82) were recruited from the Helsinki University Hospital (Helsinki, Finland) rheumatology outpatient clinic. All patients fulfilled ACR & EULAR 2010 classification criteria for RA, and blood samples were obtained before initiating disease-modifying treatment. In addition, synovial fluid samples were obtained at diagnosis from 2 patients. Laboratory and clinical information were obtained for all patients from their hospital records.

8.1.2 Felty's syndrome patients (II)

We included patients with an established Felty's syndrome diagnosis (n=14) that was stated in patient records. Samples from healthy controls (n=8), RA patients without Felty's syndrome (n=20) and LGL leukemia patients (n=9) were used as controls. The patients had followed up for Felty's syndrome in the Helsinki University Hospital rheumatology outpatient clinic, Päijät-Häme Central Hospital (Lahti, Finland), or in collaborating clinics in the USA. Patient information was extracted retrospectively from patient records. When possible, we collected new blood samples from patients, but we also obtained sample materials that had been collected for diagnostic purposes but were archived thereafter. Examples of these archived samples include whole blood DNA samples and bone marrow cells fixed with methanol and acetic acid. The samples are listed in Table 7. For 3 patients, only archived clinical samples were available.

Table 7. *Sample types available from Felty's syndrome patients.*

	BM MNC	BM cells	PB	PBMNC	PB CD8+
No. of patients	5	5	5	5	6
CD8+ cells available from the same patient?	3	2	3	0	6

Abbreviations: BM MNC, bone marrow mononuclear cells; BM cells, cultured bone marrow cells; PB, whole peripheral blood; PBMNC, peripheral blood mononuclear cells; PB CD8+, peripheral blood CD8+ cells.

8.1.3 Follow-up samples from RA patients (III)

We obtained follow-up whole blood samples from 59 RA patients, of which 53 were included in study I. The samples were obtained 2 to 13 years after diagnosis. Blood cell counts from follow-up timepoints were obtained from the electronic patient record system.

8.1.4 AA- and hMDS patients (III)

Twelve patients diagnosed with acquired aplastic anemia (AA) or hypoplastic myelodysplastic syndrome (hMDS) were recruited from the Helsinki University Hospital hematology clinic for study III. We did not obtain fresh blood samples from these patients but used DNA samples that had been acquired previously for clinical purposes and archived thereafter. The sample types were bone-marrow mononuclear cell DNA for all but for one patient, whose sample type was peripheral blood DNA (MDS1).

8.1.5 Healthy controls (I–IV)

Studies I–IV included healthy controls. The numbers of controls and their characteristics are listed in Table 8. In TCRB sequencing and other DNA sequencing applications, buffy coat samples from Finnish Red Cross Blood Service blood donors were used. As for patient samples, cytokine profiling was performed from peripheral-blood ethylenediaminetetraacetic acid (EDTA) plasma samples.

Table 8. *Information on healthy controls. Abbreviations: RNA, ribonucleic acid; DNA, deoxyribonucleic acid; TCRB, T-cell receptor beta*

	Study I	Study II	Study III	Study IV
TCRB sequencing	20	-	-	27
DNA sequencing	20	14	2	21
Cytokine profiling	-	8	-	-
RNA sequencing	3	-	-	8

8.2 Sample preparation

8.2.1 Cell separations (I, II, IV)

For all studies that investigated T cells, the cells of interest needed to be extracted from peripheral blood samples with cell separation and enrichment methods. Peripheral blood mononuclear cells (PBMNCs) were separated from EDTA blood using Ficoll gradient separation (Ficoll-Paque PLUS, GE Healthcare) according to the manufacturer's instructions. CD4+ and CD8+ cells were selected from PBMNCs using magnetic microbeads (Miltenyi Biotech) and were subjected thereafter to DNA extraction. PBMNCs were also frozen in aliquots suspended in fetal bovine serum with 10% dimethyl-sulfoxide (DMSO).

8.2.2 Flow cytometry (I, II, IV)

Information on the composition of the cell samples of interest is important for sequencing assays. Purities of sorted cell fractions were controlled with flow cytometry using either antibody mix 1 or 2 (Table 9; studies I, II, and IV). Antibodies were incubated with cells in room temperature, in the dark, for 15 minutes, followed by washing.

To screen samples for large T-cell clones, we used variable beta (Vbeta)-specific antibodies which detect TCR variable beta region usage. TCR Vbeta usage was screened using a panel of T-cell receptor β variable chain (TCR Vβ) antibodies (IOTest Beta Mark TCR V kit, Beckman Coulter Immunotech) supplemented with antibody mix 1 (Table 9). The initial antibody screen in study I was performed from whole blood

samples; after staining, erythrocytes were lysed with 1x Lysing solution (Beckton Dickinson). In cell sorting experiments, PBMNCs were used and no erythrocyte lysis was performed. Samples were analyzed with FACSAriaII (Beckton Dickinson) in study I and FACSVerse (Becktron Dickinson) in others.

Table 9. *Antibody mixes used in the studies. The antibody clone and fluorochrome are included in the table columns.*

	Mix 1	Mix 2	Mix 3
Anti-CD45	-	2D1 PerCp	-
Anti-CD3	SK7 APC	SK7 FITC	SK7 APC
Anti-CD4	SK3 PerCP	SK3 APC	SK3 PerCP
Anti-CD8	SK–1 Pe-Cy7	SK1 PE	SK–1 Pe-Cy7
Anti-CCR7			150503 BV421
Anti-CD45RA			HI100 Alexa700

Abbreviations: APC, allophycocyanin; PerCp, Peridinin chlorophyll protein; Pe-Cy7, Phycoerythrin-Cyanine7; FITC Fluorescein isothiocyanate; PE Phycoerythrin; BV, brilliant violet.

In studies I and II, CD8+ T-cell populations using the same Vbeta were sorted using these antibodies with FACSAriaII (Beckton Dickinson). FACS data were analyzed with FACSDiva (Beckton Dickinson) or FlowJo (FlowJo LLC) software. The gating strategy for expanded T-cell clones is shown in Figure 6.

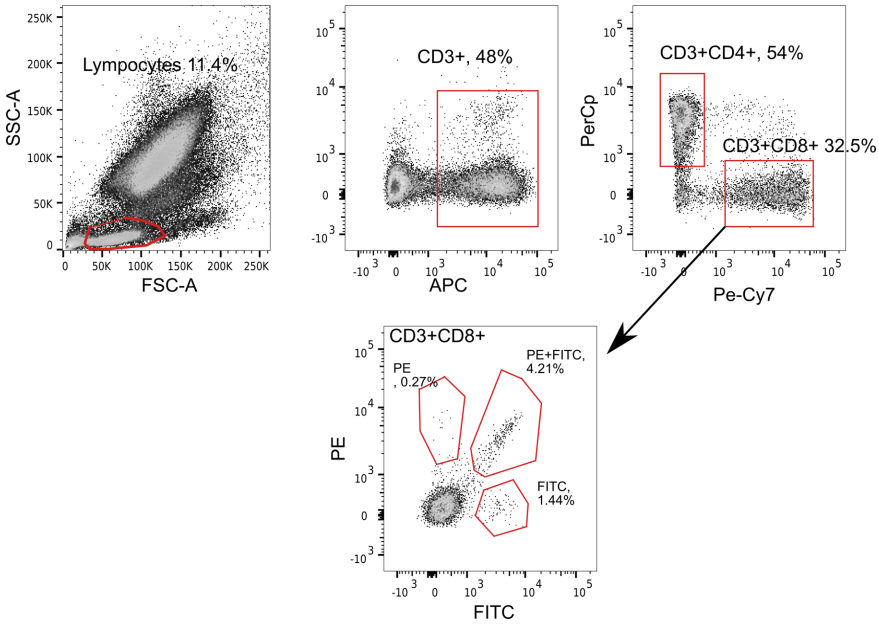

Figure 6 Gating strategy for Vbeta analyses. PE and FITC fluorochromes label the Vbeta-specific antibodies.

For phenotyping analyses in study I, PBMNCs were stained with antibody mix 3 (Table 9) and the appropriate Vβ antibody. Cytomegalovirus (CMV)-specific cells were studied by iTAg major histocompatibility complex pentamer (HLA-A-0201, NLVPMVATV, Proimmune, PE conjugate), and the cells were labeled with antibody mix 1.

STAT3-pathway activation of immune cells was investigated with phosphospecific antibodies in study I. Phosphospecific flow cytometry was performed from whole blood samples stained with anti-CD4 (FITC, SK3) or anti-CD8 (FITC, SK1), incubated for 15 minutes at +37 degrees, and followed by erythrocyte lysis (1x Lyse/Fix Buffer, Becktron Dickinson) and cell permeabilization with 1x PhosFlow Perm/Wash Buffer I (Beckton Dickinson). After permeabilization, anti-pSTAT3 (pTyr705)-PE, anti-pSTAT3(pSer727)-AlexaFluor647, polyclonal rabbit anti-pJAK1(pTyr1022/1023, 1:10 dilution in PBS, Merck), anti-pJAK2(pTyr1007/1008, Santa Cruz Biotechnology), anti-pJAK3(pTyr980, Santa Cruz Biotechnology), or anti-SHP1(pTyr536, 1:10 dilution in PBS, Abcam), was introduced in addition to anti-CD3 (SK7). The phosphospecific conjugated antibodies were incubated for 30 minutes and the unconjugated ones for 60 minutes, in the dark at room temperature. PE-conjugated F(ab')2 donkey anti-rabbit IgG (Beckton

Dickinson) was used as a secondary fluorescent antibody (40-minute incubation) for pJAK1, pJAK2, and pSHP1. The same antibody was used as a PerCP-conjugate for pJAK3. Washing and resuspension in Stain Buffer (Beckton Dickinson) followed.

Stained samples were held on ice and analyzed with FACSCantoII (Beckton Dickinson) within 4 hours of staining. FACSDiva software was used for data analyses. Each week, a healthy control sample was analyzed to determine gates for positive and negative populations (<5% of control cells were positive).

8.2.3 Nucleic acid extraction (I–IV)

DNA extractions for sequencing assays for studies I, II and IV were performed using Nucleospin Tissue DNA or Nucleospin Tissue DNA XS kits (Machery Nagel) according to the manufacturer's instructions. The DNA samples of patient 15 in study IV were extracted using Qiagen DNEasy Blood & Tissue kit (Qiagen). DNA concentrations were measured using the Qubit 2.0 fluorometer (Life Technologies) and the Qubit HS or BR dsDNA Assay Kit (Life Technologies). In study I, RNA was extracted using the Qiagen miRNeasy micro kit (Qiagen). For study III, DNA extractions were performed from whole EDTA blood with a method based on magnetic beads using Chemagic MSM1 (PerkinElmer).

8.3 Nucleic acid sequencing and data analyses

Enriching smaller genomic areas of interest allows higher sequencing depths than traditional exome sequencing, and thus even small cell populations harboring mutations can be detected. The targeted hybridized captures used in studies I and IV represent similar technologies as exome sequencing, but these methods enrich smaller genomic areas than conventional exome sequencing kits (approximately <7 megabases compared to over 30-60 megabases of DNA). Amplified capture technologies and other multiplexed PCR applications are usually used to capture smaller genomic areas than hybridized capture approaches, and the sequencing depths can be even higher, as seen in studies II and III.

8.3.1 Targeted hybridized captures (I, IV)

We aimed to detect mutations with small variant-allele frequencies (VAFs), and as this strategy requires high sequencing depth, we designed

customized probes to enrich smaller genomic areas than the entire exome. Probes were designed for studies I and IV by the Nimblegen SeqCap EZ Developer system ("immunogene panel"). In study I, the panel covered the exons and surrounding untranslated regions of 986 critical immune-related genes. These genes were annotated in the InnateDB database,[276–278] but other targets, such as genes involved in JAK-STAT signaling, were also included. Both CD4+ and CD8+ cells were sequenced for all patients.

In study I, sequencing libraries were prepared by either NEBNext (New England Biolabs) or ThruPLEX DNA-seq (Rubicon Genomics). The NEBNext kit required 1–3 ug fragmented genomic DNA to prepare libraries, and the libraries were prepared according to the manufacturer's instructions (NEBNext DNA Sample Prep Master Mix Set 1 Manual, New England BioLabs) with minor modifications. Target regions were captured according to the NimbleGen SeqCap EZ Exome Library SR User's Guide with minor modifications.

ThruPLEX libraries were prepared from 100ng of genomic DNA fragmented with Episonic Multi-Functional Bioprocessor 1100 (Epigenetik Group Inc) to the mean fragment size of 300 base-pairs, and 50 ng fragmented DNA was processed according to ThruPLEX DNA-seq library preparation kit (Rubicon Genomics) instructions with minor modifications. Captures of genomic regions of interest were performed according to the "Exome Capture of ThruPLEX libraries with Roche NimbleGen SeqCap EZ library" (Rubicon Genomics) and the "NimbleGen SeqCap EZ Exome Library SR User's guide" (Roche). Sequencing was performed using the HiSeq system (Illumina).

In study IV, the custom capture probes covered the exons of 2533 genes. The selected genes are critical for hematopoietic cells, adaptive immunity, and autoimmunity, but we also included genes in the InnateDB database.[276–278] Sequencing libraries were prepared with the ThruPLEX DNA-seq kit or the Kapa Hyper library preparation kit (Roche). CD4+ and CD8+ cells of 9 healthy controls were prepared with ThruPLEX DNA-seq kit. Samples (100ng DNA) were fragmented with the Episonic Multi-Functional Bioprocessor 1100 (Epigenetik Group Inc.) to fragments with a mean length of 200 base-pairs. Sequencing libraries were prepared according to ThruPLEX DNA-seq library preparation kit protocol (Rubicon Genomics). Targets were captured according to the "Exome Capture of ThruPLEX Libraries with Roche NimbleGen SeqCap EZ Library" (Rubicon Genomics) and the "NimbleGen SeqCap EZ Exome

Library SR User's Guide" (Roche). Eight samples (100ng DNA each) were pooled for hybridization and target capture, and post-capture amplification had 10 cycles.

In the remaining of samples in study IV, library preparation was performed with the Kapa Hyper library preparation kit: 300–500ng DNA was fragmented using the Covaris E220 evolution instrument (Covaris) to a mean fragment length of 200 base-pairs. Sample libraries and enrichment were performed according to the "SeqCapEZ HyperCap Workflow User's Guide version 1.0" (Roche Nimblegen). Four samples were multiplexed for capture (1 ug each), and post-capture amplification had 10 cycles. Sequencing of all samples in study IV was performed with the HiSeq2500 (Illumina) with HiSeq high output mode using v4 kits (Illumina)a and 101-length paired-end reads.

8.3.2 Exome sequencing (I)

Somatic mutations in highly clonal populations can be detected with low sequencing depth. We complemented the custom gene panel capture sequencing in study I with exome sequencing from highly clonal, CD8+ T-cell populations. CD8+ T-cell populations expressing the same Vbeta were sorted with FACSAria II from patients 1, 3 and 4 using antibody mix 1 (Table 9) and the appropriate Vbeta antibody. Sequencing libraries were prepared from 50 ng sorted DNA with the Nextera Rapid Exome Kit (Illumina). In addition to the expanded Vbeta clone, CD4+ cells were sequenced to serve as a germline control. The data were analyzed similarly as immunogene gene panel data.

8.3.3 Variant calling with VarSca2 (I)

Somatic variant identification was performed with the VarScan2 algorithm in study I. First, raw Illumina reads were merged with SeqPrep 0.4.5 and paired reads were trimmed of B blocks in the quality scores from read ends. Trimmed reads shorter than 36 base-pairs were omitted. Reads were aligned with Burrows-Wheeler Aligner version 0.6.2 to the GRCh37 reference genome.[249] Multimapping reads were removed. Alignment was refined with GATK Indel Realignment version 2.2-16.[253] PCR duplicates were removed with Picard MarkDuplicates version 1.90.[252]

VarScan2 version 2.3.251 was used to call high-confidence somatic mutations for each patient with the following parameters: strand-filter 1, min-coverage-normal 8, min-coverage-tumor 6, somatic-p-value 1,

normal-purity 1, and min-var-freq 0.05.[255] Mutation annotations were performed using SnpEff version 4.052 with the EnsEMBL v68 annotation database53.[279–281] Variants classified as common population variants in the Single Nucleotide Polymorphisms database (dbSNP, build ID 130) were discarded.[282] VarScan2 analyses were performed in a pairwise manner so that the same patient's CD4+ cells served as germline control for CD8+ cells but also vice versa. Variants with a somatic p-value <0.01 were inspected more closely.

8.3.4 Variant calling with MuTect2 (IV)

In study IV, somatic variants were identified with MuTect2 toolset. Prior to variant calling, the Trimmomatic software was used to pre-process reads shorter than 36 base-pairs, adapter sequences and low-quality bases.[283] Sequencing reads were aligned to the GRCh38 reference genome with BWA-MEM.[284] Alignments were sorted with SortSAM, and PCR duplicate marking was performed with the Picard toolset.[252] The Genome Analysis Toolkit (GATK) 3 was used for variant calling, with the exception that cross-sample contamination level estimation and 8-oxoguanine and deamination artifact filtering were performed by GATK4 CalculateContamination, CollectSequencingArtifactMetrics, and FilterByOrientationBias.[253] ContEst was also used to estimate cross-sample contaminations.[285] CrossMap was used to convert GRCh37 to GRCh38.[286]

Somatic variants were identified with GATK MuTect2 using a panel of 21 healthy controls' samples as a background reference. Variants were called in 2 ways. (1) Paired sample analysis included calling variants by comparing CD4+ to CD8+, and vice versa. This approach identifies variants in mature T cells. (2) To identify variants that originate in hematopoietic progenitor and stem cells, variants were also called from each sample in single-sample mode, as variants occurring in both CD4+ and CD8+ would be discarded in the paired-sample analyses.

Annovar was used for variant annotation.[287] Variants occurring with more than 1% minor allele frequency in established genomic variation databases (gnomAD_exome_ALL, gnomAD_exome_FIN, gnomAD_exome_NFE, esp6500siv2_all, 1000g2015aug_all) were excluded. Intronic, inframe, and synonymous mutations were excluded in all analyses except mutational signature analyses. Variants were required to pass all MuTect2 filters. Variants with <40 quality, as well as positions with coverage <10, were filtered. In addition, variants with >3 strand

orientation bias (SOR) for single-nucleotide alterations and >11 for indels were excluded. Variants with minimum 2% variant allele frequency (VAF) and 7 supporting reads, and with calls in both forward and reverse strands were included in further analyses. Variants with 1–2% VAF were incorporated if they occurred more than 5 times in the Catalogue of Somatic Mutations in Cancer (COSMIC) database in hematopoietic tissue samples, but these variants had to fulfill the other filtering criteria. Indels in repeat regions and at least 5 base-pair homopolymers were filtered. Finally, variants were inspected visually with Integrative Genomics Viewer (IGV).

Paired-sample variant calling analysis eliminates variants that occur in both CD4+ and CD8+ cells. To complement the paired-sample variants with variants that have occurred in hematopoietic progenitors, we also performed variant calling in single-sample mode. The single-sample variants had to occur in both CD4+ and CD8+ cells, and the variant had to occur in a known clonal hematopoiesis-associated gene.[226,227,235] However, the VAF was permitted to be less than 2% in either CD4+ or CD8+ cells.

Mutations identified in paired-sample analyses were included in mutation signature analyses. Signatures were identified with default parameters with deconstructSigs; cancer profiles were downloaded from the COSMIC web site in September 2017.[206,288]

8.3.5 Targeted amplicon sequencing (I, II)

Targeted, deep amplicon sequencing was employed in studies I and II. Amplicon sequencing is a PCR-based method, in which short PCR products are sequenced, yielding sequencing coverage that can be over 10 000–100 000 at the sequenced position and a sensitivity that can be as low as 0.5–1%.[114]

The amplicon primers were designed to cover the genomic area of interest, and primer sequences are listed in Table 10. The sequenced amplicons were produced with either one-step or two-step PCR. In one-step PCR, the locus-specific primers existed in limiting quantities, and the index adapter primers allowed the amplification to continue. In two-step PCR, the first PCR contained locus-specific primers, and the second comprised index primers. The amplified samples were pooled, purified twice with Agencourt AMPure XP beads (Beckman Coulter), quantified with the Agilent 2100 Bioanalyzer (Agilent Genomics), and sequenced

with the Illumina MiSeq system (Illumina) using the Illumina MiSeq Reagent v2 500 cycles kit (Illumina).

Sequencing reads were aligned to the human GRCh37 genome with Bowtie2 and GATK IndelRealigner.[250,253] Both high- and low-quality reads were aligned to the genome, but quality scores were used to filter low-quality variants in further analyses. Variants with >5 supporting reads and VAF >0.5% were called. A frequency ratio for each variant was calculated by dividing the ratio of variant calls/number of all bases by the ratio of variant allele quality sum/quality sum of all bases. Control samples were included in all sequencing runs.

For studies I–II, the frequency ratio was required to be >0.8 and VAF >1%, and the variant was required to be absent in control samples. Study I included validation experiments, and thus the variant was required to occur with a similar VAF as reported in exome- or immunogene panel sequencing.

Table 10. *Amplicon primers for studies I–II. Abbreviations: FW, forward; RV, reverse.*

Study	Mutated gene	FW primer	RV primer
I	*ADCY10*	TTCAGTGGAAAGGGGTGCTG	CCCTCTGGAATCTTGCCAGT
I	*BST1*	TGACCTGGTGATCAGAGCCA	GGGTCCCCCAATTTCATGCA
I	*C16orf89*	CCTGGCCTCCTCTCTCCTTA	TTACTGACAGCACTGCCACC
I	*C5*	TGGGCTCATGAGAAACCGT	AGGGGATCACATGGAATGTTGT
I	*CDK12*	AAGGAGTCCAAGGGTTCACC	TAGTTGGTAGAGGGGGTGGG
I	*CDYL*	GGACCTCTCCCAACAATGCT	GGTCCAGTTTCTCAGGGCTC
I	*CLEC10A*	CCTGGCTTCCAGTTCCTGAG	AAGTCACCACTGCCCTTCTG
I	*CLIP2*	ATGCAGTGGGGTATGTCACC	CTCCTTCCGCAGCAGCTC
I	*CLSPN*	TTCAGGCTGGTCCTGAACTC	GATGAGGAACTGCAGAGTCAAA
I	*CLU*	CGAGCTGTGTCATCCCTCTC	TGGAGCCAGCACAGCTATTC
I	*CNTN2*	GGTTCCCATCTTACGGTGCT	TGAGTCAGAGCTGTGCAGTG
I	*CRYBB2*	AATGGTTGGGAGGCTTCACC	TCACTCTCTGGGAGGTCTGG
I	*EPRS*	TCTGTTTCTTAGGCCCCTGC	GTGTGCTTAGACAATGACTTGAGT
I	*GRM5*	AACCAAGGGTTTCGGTGGTT	GGCCCAAAATTTCCATGGTGT
I	*IRF1*	CGTGAATGTGGCACTTGTGG	GAACTGGAACTGCAGGGTGA
I	*KIF16B*	TTTTCCTTTTGGAGCTGGTC	AAGGAGACAGAAATCGTGCAG
I	*MAST1*	GGTGCTATGGTGCATGTCCA	TGGGGATCTCAGGGATGAGG
I	*MREG*	TGACTGCTGAGTCCTCATGG	GGGTGGGTTTTGTTGTTGTT
I	*PADI4*	GCAAGGGTGAGAGTGAGTGG	GGCAGGCTAAAAGGAGGGAG
I	*PLRG1*	CCATGAGGATCAAAACTTCCA	TTTCTACGTGAATGGGCAAA
I	*PROM1*	GGCAACCTTTCTCTAGAATTTCA	TCATGAGGAATTGTTTTGAAGG
I	*PTPRO*	TGGGATTTCATGCTCTGCTG	ACAGACACACAATGCTGGCA
I	*RP11-766F14.2.1*	CCTGCTCTGGGGTAACACTG	CAGCCGAGCTCAGACTTCAA
I	*SCAF4*	CAAGTCTCTGTGCCTGTCCC	GGGGGTCCTGAAGACAGAGA
I	*SEC14L3*	AACATCTTAACCTAGGCCTGGAC	TCCGCAAGGTGAGACCTATC

I	*SKP1*	AGTTCAAAAAGTGTTCCTTGGTCA	TGGCTCAAGTCAACCTCTAGC
I	*SLAMF6*	CTCCCTCCTCCTGTCTTCCA	CATAGGACTCCCACCCCAGA
I	*SMARCAD1*	GGAACCTACACATTGTGAGGC	AGCAAGTCAGTGTGCCAGAA
I	*STARD5*	GTTTGGGGCTTCTCTCCCTC	TGATTGATGCGTTGGGGTCA
I	*TOPAZ1*	AGAGGAACTGAGCAGAAGAGG	TCAGCCGTCCTGTGTTCAAT
I	*TPTE*	GGAACAATGGTGGTTCAAGG	GGAACTTCAATTTCTTTGGAATG
II	*STAT3-2*	TCCCCATCACCTGTACCCAT	TGACACCTGTGTTGGGCAAT
II	*STAT3-3*	ACACTAACACCCGACTCTGC	TGTATGCGTCGGCTTCAGAG
II	*STAT3-4*	TCCATTCCTCCCAGACCAGG	GCTCTGAAGCCTTTGTTCCG
II	*STAT3-5*	CCGAGGCTTGTAACTTGCAT	TTCCCTTCCTCTTGTGATGG
II	*STAT3-6*	GACCAGGCTCCTTTGAGGAC	CTCTTGGGGATACTGCCTGC
II	*STAT3-7*	CCGATCTAGGCAGATGTTGG	TTCCCTCAGGTCAAGGAGTTT
II	*STAT3-8*	CTGTGGGCCTGCAGTTAAGA	GTTCCTGCTCTGGAGTTGACT
II	*STAT3-9*	AAGAGAAGATGGGCTCACGC	TCCCCTTCTCCATCTCACCT
II	*STAT3-10*	TGGAAAGAATGACCCTGGCC	CACGTGGTAGAGTGAGAGGC
II	*STAT3-11*	AATGCACCCCAAGGCTTTTG	CCTCCCACAGTGCTGAGATT
II	*STAT3-12-14*	CAAGGAAAACACCCCAGTTG	AAATAACAGGTGGTCAAAGTAGGC
II	*STAT3-15*	ATTGCCAGATGGGATGCCAA	CCACACCTGGCCTAAGAGTG
II	*STAT3-16*	GAGGAGAACTGCCAGCTCAG	CCTTTCATTCTGAGCCCCGT
II	*STAT3-17*	AGGGAGAAGGGGTGAAATGC	TGCCCCCTCCTTTTAGTTGG
II	*STAT3-18*	CCTTGCCAGCCATGTTTTCC	AACCTCTTGACCCCAAGCTG
II	*STAT3-19*	GTGCACACTCTGTCCAACCT	GCTTGAAGGCCCTGAACTCT
II	*STAT3-20*	GGAGTCAAGGCCATCTCCAC	TGGATGCCCTGTTAGCAATA
II, I	*STAT3-21*	CCCAAAAATTAAATGCCAGGA	GGTTCCATGATCTTTCCTTCC
II	*STAT3-22*	CTCACCCAGTGTCCCATTCC	GGCAGATGGAGCTTTCCAGA
II	*STAT3-23*	GACCAGCTCTCGGTGTGTAC	TGGAGACCAGAGTTTGATGGC
II	*STAT3-24*	GGCACTTGTCTAAGAACAACAACA	AGTTGCAGAGGGTGGACAAC
II	*STAT5B-16*	TGTTGGGGTTTTAAGATTTCC	CAAATCAGAATGCGAACATTG
II	*STAT5A-17*	TCCTGCTGCTGGTGGATTAT	AGCCCAAGGCTTTGTCTATG

8.3.6 TCRB deep sequencing (I, IV)

NGS has revolutionized the capacity to detect and characterize the T-cell repertoires in biological samples: sequencing and bioinformatics allow obtaining information not only on the receptor sequences but also the abundance of the identified receptor sequences. In study I, we performed deep TCR beta locus (TCRB) sequencing from 65 patients' and 20 controls' CD8 cells. We also sequenced synovial fluid cells from 2 patients for whom synovial fluid samples were available. TCRB sequencing in study IV involved 15 patients and 27 healthy controls. The CD4+ and CD8+ TCRB repertoires of 7 CVID patients and 27 healthy controls were compared. Genomic DNA was used in all cases. Sequencing and data analysis were performed with the ImmunoSEQ human TCRB assay (Adaptive Biotechnologies Corp) at survey level. The methodology employs multiplex PCR in which primer optimization has been used to minimize amplification bias.[274]

ImmunoSEQ Analyzer (Adaptive Biotechnologies) tools were used for clonality metrics and differential abundance calculations. Only productive TCR sequences were included in the analyses. Sample clonality was calculated according to the formula (1), where p_i is the proportional abundance of the rearrangement i and N is the total count of rearrangements.

1. $Clonality = 1 - \frac{-\sum_{i=1}^{N} p_i log_2(p_i)}{log_2(N)}$

Physico-chemical TCR similarity was investigated with a string kernel algorithm that has been designed for small peptides.[289] Unsupervised clustering was performed on the results with Phenograph.[290]

8.3.7 Myeloid panel sequencing (III)

In study III, we aimed to study clonal hematopoiesis in RA patients and compare the results with patients with aplastic anemia (AA) and hypoplastic myelodysplastic syndrome (hMDS). We aimed to detect mutations with low VAF (down to 2%) in coding exons of 34 genes that frequently harbor mutations in clonal hematopoiesis in healthy and AA/hMDS patients (Table 11). Thus, we designed a custom sequencing panel based on the Illumina TruSeq Custom Amplicon platform with Illumina Design Studio software (Illumina). The panel consisted of 583 multiplexed amplicons of approximately 250 base-pairs. Sequencing libraries were prepared from 250ng genomic DNA according to the manufacturer's instructions (Illumina TruSeq Custom Amplicon v1.5 reference guide), except that the last normalization step was omitted to allow for accurate library quantification. The samples were sequenced with the Illumina HiSeq2500 with rapid 150 paired-end reads (Rapid PE150).

Sequencing data were pre-processed similarly to the targeted amplicon sequencing data (described above in Section 7.3.5). However, only high quality bases (Phred >20) were included in analyses, and the frequency ratio was required to be over 0.9. All called variants were required to pass a binomial test threshold p-value $<10^{-5}$, have less than 2% noise, have coverage >500, and have a variant call count >20. The variant also was required to compose more than 80% of all variant bases in the genomic position. Variant calls derived from only one strand were omitted if the genomic area was covered by both forward and reverse strands.

To filter out germline variants, all variants with VAF>35%, variants that had a population frequency >1% in 1000Genomes data,[291] and variants that occurred in >10 individuals were omitted. Variants occurring within 5 base-pairs of a 5 base-pair-long homopolymer were also omitted.

Variants were annotated with the Ensembl Variant Effect Predictor.[292] Because we defined clonal hematopoiesis as mutations that offer likely survival advantage to cells, we discarded non-coding and synonymous mutations, as well as mutations with <2% VAF. The VAF 2% cutoff is also used in the definition of CHIP.[233]

Table 11. *Genomic areas covered by the sequencing panel. The areas included 10 gaps (range: 1–872 base-pairs), which are described in more detail in study III.*

Gene	Sequenced exons	Coding exons in canonical transcript	Gene	Sequenced exons	Coding exons in canonical transcript
DNMT3A	2–23	2–23	*CSMD1*	1–70	1–70
ASXL1	1–13	1–13	*LAMB4*	2–34	2–34
PPM1D	1–2, 4–6	1–6	*WT1*	1–5, 7–10	1–10
SRSF2	1–2	1–2	*CEBPA*	1	1
TET2	3–11	3–11	*JAK2*	12 and 14	3–25
TP53	2–11	2–11	*SF3B1*	8–16	1–25
BCOR	2–15	2–15	*CBL*	8–9, 11	1–16
BCORL1	1–12	1–12	*GNAS*	1*	1–13
RUNX1	1–8	1–8	*IDH2*	4 –5	1–11
PIGA	1–6	2–6	*IDH1*	4–5	3–10
EZH2	2–20	2–20	*NRAS*	2–3	2–5
ETV6	1–8	1–8	*KRAS*	2–3	2–5
PHF6	2–10	2–10	*U2AF1*	2, 5–6, 8	1–8
ZRSR2	1–11	1–11	*SETBP1*	4**	2–6
STAG2	3–5, 7–35	3–35	*MPL*	1–12	1–12
GATA2	2–6	2–6	*STAT3*	12–13, 21	2–24
NPM1	1–11	1–11	*STAT5B*	16	2–19

* chr20: 57429568–57428846 (722 base-pairs)

** chr 18: 42531686–42532297 (611 base-pairs)

8.3.8 RNA sequencing and data analysis (I, IV)

To characterize the expanded CD8+ T-cell clones that harbored somatic mutations in study I, we utilized RNA sequencing from cell samples sorted by flow cytometry. The SMART-Seq v4 Ultra Low Input RNA Kit was used to synthesize complementary DNA libraries from sample RNA. Sequencing libraries were prepared with the Illumina Nextera XT kit. Sequencing was performed as 100 base-pair paired-end reads with HiSeq2000 (Illumina). Sequence data were corrected, trimmed, and filtered for read lengths with Trimmomatics using the following settings: leading 3, trailing 3, sliding window 4:15, and minlen 36.[283] STAR aligner was used to align the final and cleaned paired-end reads to the GRCh38 human reference genome with guidance of EnsEMBL v80 reference gene models.[281,293] Default 2-pass per-sample alignment and indexing settings were used, except that the overhang on each side of a splice junction was set to 99. Picard tools were used to sort and mark duplicates, and genomic features were assigned with SubRead.[252,294] Default parameters were used in the feature summations, with the exception that reads could be assigned to overlapping genome features. RNA-SeQC and FASTQC with default settings were used for quality control analyses. Genome annotation information downloaded from EnsEMBL (v80) was used.[281] Binary Alignment Map files were normalized, and the data were analyzed with the Qlucore Omics Explorer 3.2 (Qlucore AB).

In study IV, RNA sequencing was used to document the expression of genes that harboured mutations in healthy CD8+ and CD4+ cells. RNA was extracted with RNA was extracted with miRNeasy kit including DNAse I digestion (Qiagen) or NucloSpin RNA II kit (Machery-Nagel) according to manufacturer's instructions. After depletion of ribosomal RNA (Ribo-Zero, Epicentre) and RNA purification, RNA was transcribed to double-stranded cDNA (SuperScript Double-Stranded cDNA Synthesis Kit, Life Technologies or NEBNext mRNA Library Prep Master Mix Set for Illumina). Random hexamers were used to prime the first strand synthesis (New England Biolabs). cDNA was purified with SPRI beads (Agencourt AMPure XP, Beckman Coulter). Illumina Nextera technology with 60ng of double-stranded cDNA was used to prepare RNA sequencing libraries. The tagmentation reaction was followed with SPRI bead purification. Illumina-specific bridge-PCR sites and library enrichment were performed via limited-cycle (5 cycles) PCR according to the Nextera system instructions. However, 50 X Nextera Adaptor 2 was replaced with Nextera Bar Codes kit adapters (Illumina). PCR products

were purified with SPRI beads, and the sequencing library qualities were evaluated with Agilent Bioanalyzer (Agilent Technologies). Samples were sequenced with Illumina HiSeq2000 or Illumina Genome Analyzer II with 100 base-pair long paired-end reads.

In study IV, sequencing reads were trimmed with Trimmomtics[283] and mapped to the GRCh38 reference genome using STAR65 aligner with the default 2-pass per-sample mapping settings.[293] The Picard toolset was used to sort reads and mark duplicates.[252] Sequencing quality was evaluated with FASTQC and RNA-SeQC. FeatureCounts was used to assign the mapped reads to gene features.[294] Multimapping and assignment to more than one overlapping feature were allowed. Read count normalization with the Trimmed Mean of M-values (TMM) method was performed with edgeR with default parameters,[295] and results were transformed to log2 values for statistical analyses.

8.3.9 HLA typing (I)

All HLA typing was performed by the histocompatibility-testing laboratory of the Finnish Red Cross Blood Service, which has been accredited by the European Federation for Immunogenetics. Results are reported based on the World Health Organization (WHO) HLA nomenclature. HLA alleles were determined by using Luminex bead-array technology with sequence-specific oligonucleotide probes (Commercial LabType kits RSSO1A, RSSO1B, RSSO1C, RSSO2B1, One Lambda). The results were analyzed with HLA Fusion v3.2 software (One Lambda) according to the manufacturer's instructions. A subset of samples was typed with Sanger sequencing for higher resolution (Commercial AlleleSEQR kits 08K60-06, 08K61-06, 08K62-06, 08K63-06, GenDx), and sequencing data were analyzed using the SBTengine software 3.9.0.2563 (GenDx).

8.4 Biomarker panel analysis (II)

Cytokine profiling of 92 cytokines was performed in study II using the Proseek Multiplex Inflammation I (Olink Biosciences) to compare plasma cytokine profiles of healthy controls (n=8), RA patients (n=9), Felty's syndrome patients (n=7), and LGL leukemia patients (n=8). The assay is based on oligonucleotide-labeled antibodies, and the target-antibody complexes are detected with real-time PCR. The PCR was performed with a 96.96 dynamic array on the HD (Fluidigm) system.

The assay results are reported as normalized protein-expression units (NPX), and the levels of the same biomarker are comparable with different patients. NPX values were obtained from Cq values after normalizing for extension and interplate controls and correction factors.

(1) $dCq_{analyte} = Cq_{analyte} - Cq_{Ext\ Ctrl}$

(2) $ddCq_{analyte} = dCq_{analyte} - dCq_{Interplate\ Ctrl}$

(3) $NPX = Correction\ factor - ddCq_{analyte}$

NPX values were reported in log2 scale and were analyzed with Qlucore Omics Explorer (Qlucore). Biomarkers with a false discovery rate (FDR) <0.1 (Benjamini-Hochberg method) were reported and Dunn's multiple comparison tests were used in post-hoc testing.

8.5 Immunohistochemistry (II)

Phosphorylation of STAT3 was investigated in study II in formalin-fixed, paraffin-embedded bone marrow samples of 7 patients and 14 control subjects. All samples had been obtained for clinical purposes and were archived afterwards. Sections (3.5um) were cut on Superfrost objective slides (Kindler O Gmbh) with a Microm 355S microtome (Thermo Scientific). Glass slides were dried overnight at +37°C and maintained in +4C for short-term storage.

Slides were deparaffinized in xylene and rehydrated in graded ethanol series and H_2O. Heat-induced epitope retrieval was performed in 10 mM trisaminomethane (Tris)-hydrochloride 1 mM EDTA buffer (pH 9) at +99°C for 20 minutes (PT Module, Thermo Fisher Scientific). Peroxide activity was blocked in a 0.9% hydrogen peroxide solution for 15 minutes, followed by 15 minutes of 10% normal goat serum (in Tris-buffered saline, pH 7.4, TBS) application. For washing, 0.1% Tween-20 (Thermo Scientific) diluted in 10 mM trisaminomethane hydrochloride (pH 7.4, TBS) was used. Washing was performed 3 times after peroxide blocking, antibody incubations, and chromogen reactions. A humid chamber was used for protein blocking and antibody incubations.

Antibodies (CD57, clone VC1.1, Sigma-Aldrich and STAT3 Tyr705, Cell Signaling) were diluted 1:100 and 1:250, respectively, in protein block solution and were incubated for 90 minutes. Secondary antibodies (1:1 mix) were incubated for 45 minutes. Secondary antibodies comprised

horseradish peroxidase (HRP)-conjugated anti-mouse and alkaline phosphatase (AP)-conjugated anti-rabbit (Immunologic).

Antibody complexes were detected with Liquid Permanent Red (Dako) and VinaGreen (Biocare Medical) chromogens and washed with H_2O 30 seconds after each reaction. Counterstaining was performed using Mayer´s hematoxylin diluted 1:10 in H_2O for 1 minute and 30 seconds. Digitalization of stained tissue slides was performed with 0.33 µm/pixel resolution using a Pannoramic P250 Flash II tissue slide scanner (3DHistech, Hungary) with a Zeiss Plan-Apochromat 20x objective (NA 0.8).

8.6 Statistical methods (I-IV)

In all studies, p-values <0.05 (corrected for multiple comparisons if necessary) were considered as statistically significant. Statistical software used in the studies are shown in Table 12. Normality of the data was inspected graphically and tested with the Shapiro-Wilk test. Nonparametric tests were used if the sample groups were small or if studied groups did not follow normal distribution.

Table 12. *Statistical software used in studies I-IV*

Software	Study	Provider
GraphPad Prism version 6	I–IV	GraphPad Software
SPSS Statistics version 23	III	IBM
R version 3.3.1	I	R project, www.r-project.org
R version 3.4.2	IV	R project, www.r-project.org
Qlucore Omics Explorer	II	Qlucore
immunoSEQ Analyzer	I, IV	Adaptive Biotechnologies

Statistical tests included the T-test, the Mann Whitney test, Fisher's exact test, linear regression, the Kruskall-Wallis test, Dunn's multiple comparison tests, and Spearman rank correlation. Study II used Qlucore Omics Explorer to produce heatmaps and calculate FDRs. Study III analyzed longitudinal data in a natural logarithm-transformed format with a linear mixed model using unstructured covariance. Also, in study III, multiple comparisons were corrected with the Sidak method. In study IV, Benjamini-Hochberg corrections for multiple testing were used in V-

family and J-gene usage analyses. TCRB sequencing result analyses also employed the immunoSEQ Analyzer platform (Adaptive Biotechnologies) to calculate TCR clonality indices and to perform public TCR queries.

9 RESULTS

9.1 Major T-cell clones in health and disease (I and IV)

9.1.1 Flow cytometry versus TCRB sequencing in RA (I)

We screened 82 newly diagnosed RA patients' blood samples with a flow cytometry-based method based on a commercial panel of monoclonal antibodies that recognize specific V-beta families. We used this method to study whether large T-cell expansions exist in either CD8+ or CD4+ cells. The flow-cytometric method showed that CD8+ T cells were more clonal than CD4+ T cells (Figure 7): a larger percentage of patients had major (> 10%) Vbeta-family restricted expansions in CD8+ than in CD4+ T cells.

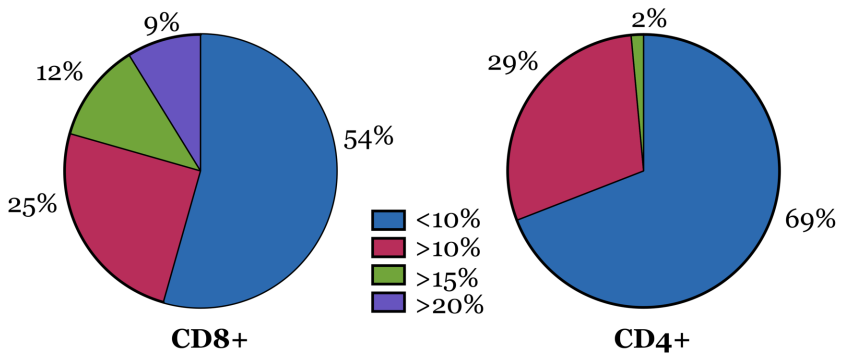

Figure 7 The percentages of patients with a T-cell clone forming more of the TCR repertoire than the denoted thresholds (10%, 15%, 20%), as measured with flow-cytometric Vbeta-specific antibodies.

Because CD8+ cells were more clonal than CD4+ cells, we performed deep TCRB sequencing to confirm the flow-cytometric results and further characterize the identified clones. The V-gene usage identified with flow cytometry and sequencing correlated well (Spearman $p<0.0001$, Figure 8A). T-cell clonality calculated from the TCRB sequencing results showed that CD8 clonality increases with age in RA patients ($p<0.0001$, Figure 8B). However, when clonality scores were compared between patients and controls, the clonality scores did not differ significantly (Figure 8C).

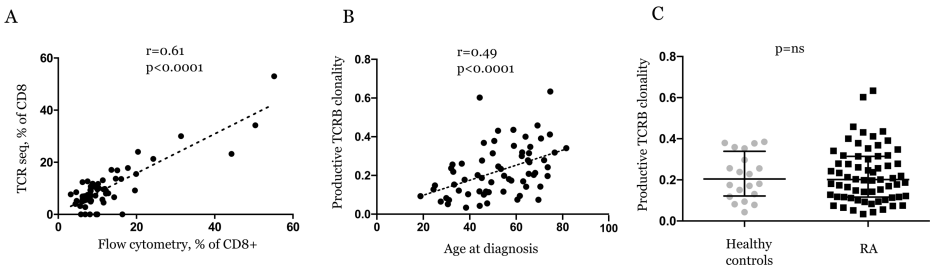

Figure 8 Clonality analysis in RA patients. In A and B, correlation was calculated with Spearman correlation.

TCRB sequencing identified 8 (of 65) patients in whom the largest CD8 clone composed more than 20% of the total CD8+ cells. In the age-matched healthy controls, 2 of 20 patients had clones that made up more than 20% of the CD8+ repertoire by TCRB sequencing.

9.1.2 Clinical associations with T-cell clonality (I)

CD8+ clonality scores were correlated with different clinical parameters in newly diagnosed RA patients: different blood cell counts, disease activity measures, age, and CD4/CD8 ratio. Only age (Figure 8B), CD4/CD8-ratio and hemoglobin level correlated significantly with CD8 clonality. No differences were detected in clonality in patients when they were grouped according to radiographic erosions, seropositivity, and "shared epitope" HLA usage.

9.1.3 TCRB repertoire convergence in CVID (IV)

We compared the TCRB repertoires of CD8+ and CD4+ cells in 7 CVID patients and 27 healthy controls. Both groups showed higher clonality in CD8+ cells than in CD4+ cells, and both also harbored individuals with extremely large CD8+ clones, similar to the RA cohort in study I. As in study I, median clonality scores between CVID and healthy controls did not differ. CD4+ clonality showed an inversed correlation with the frequency of switched-memory B cells, and a positive correlation with CD8+ clonality.

In addition, no preferential V family usage was observed when CD4+ and CD8+ repertoires of CVID patients were compared to controls. Nor did differences exist between controls and CVID patients in the frequency or abundance of public, pathogen-specific TCRs documented in the McPas database.[296] However, CVID patients' CD4+ and CD8+ repertoires

harbored higher frequencies of TCRs that shared an identical amino-acid sequence, but had distinct nucleotide sequences (TCR convergence). These convergent TCRs accounted for a median 7.0% of the CVID CD8+ repertoire and 3.2% of the healthy controls' CD8+ cells (Mann-Whitney test p=0.0215). In CD4+ cells, convergent TCRs accounted for 4.5% of the CVID CD4+ repertoire and 3.5% of the healthy controls' repertoires (p=0.0123).

9.2 *STAT3* mutation screening in RA (I–II)

9.2.1 STAT3-mutation screening in newly diagnosed RA patients (I)

Because of the close link of *STAT3* mutations and RA in LGL leukemia, we investigated whether RA patients harbored small cell clones with *STAT3* mutations. Newly diagnosed RA patients' CD4, CD8, or MNC samples were sequenced for the STAT3 exon 21 mutational hotpot with deep amplicon sequencing. Despite a median sequencing depth of 20000, no mutations were detected.

9.2.2 *STAT3* mutations in Felty's syndrome (II)

Felty's syndrome is a subtype of RA in which patients present with RA, neutropenia, and often splenomegaly. Because the clinical presentation resembles closely LGL leukemia, we collected cell samples and clinical information from these patients and sequenced all coding exons of *STAT3* and *STAT5A/B* exon 17/16 with amplicon sequencing. *STAT3* mutations occurred in 43% (6/14) of patients. All of the mutations occurred in the mutational hotspot in exon 21 of *STAT3*, and all mutations have been previously described in LGL leukemia (Table 13). One patient had 2 mutations (N647I and Y640F).

Table 13. *Mutations discovered in 6 Felty's syndrome patients. One patient harbored 2 different STAT3 mutations (N647I and Y640F), others only 1.*

STAT3 mutation	No. of mutations	VAF in sequenced cells
Y640F	3	1.1–8.8%
N647I	1	1.6–2.4%
D661Y	1	2.9–8.9%
D661V	2	1–2%

We had access to CD8+ cell samples for 6 patients, and *STAT3* mutations occurred in 4 of them (Figure 9). In most cases, *STAT3* mutations persisted in follow-up: of the 5 *STAT3*-mutated patients with sequential samples, 3 cases harbored mutations in both timepoints (Figure 9). Also, flow cytometry-mediated sorting and sequencing experiments from 2 patient samples showed that *STAT3* mutations do not occur polyclonally in all CD8+ T cells but exist in more specific cell subsets. In the case of patient 6, the major CD8+ T-cell clone was the only cell population to harbor the Y640F mutation, while patient 4 did not harbor the Y640F mutation in the largest clone but did in other CD8+ T cells. Moreover, Felty's syndrome patients showed increased STAT3 phosphorylation in bone marrow biopsy samples, regardless of *STAT3* mutation status or mutation VAF.

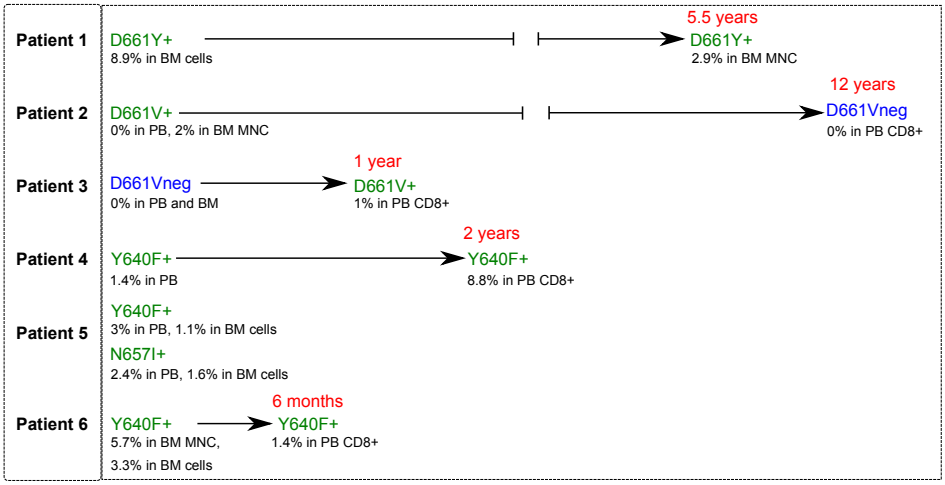

Figure 9 *STAT3* mutations in Felty's syndrome patients. Abbreviations: BM cells, cultured bone marrow cells; BM MNC, bone marrow mononuclear cells; PB, peripheral blood; PB CD8+, peripheral blood CD8+ cells.

We also compared the plasma cytokine profiles of Felty's syndrome with those of LGL leukemia patients, RA patients, and healthy controls. Of the 92 assayed cytokines, 10 were increased in both Felty's syndrome and LGL leukemia when compared to healthy controls. All 3 immune-mediated diseases showed elevated CDCP1, IL-6, and HGF when compared to healthy controls. CSF1, CXCL10, IL-15RA, MIP-1alpha, OSM, TNFRSF9, and PD-L1 were elevated in both Felty's syndrome and LGL leukemia patients, but not in RA patients.

9.3 Somatic mutations in mature T cells (I)

To identify somatic mutations in mature T cells in RA patients, we used hybridized target enrichment technologies to capture coding exons and related untranslated regions of 986 immune-related genes. This "immunogene panel" sequencing was performed from rhe CD4+ and CD8+ cells of 25 newly diagnosed RA patients' and 20 healthy controls. Exome sequencing was performed for 3 patients: these patients harbored CD8+ T-cell clones that comprised approximately 50% of all CD8+ T cells in flow cytometry, thus allowing sufficient sample material collection for exome sequencing. A schematic of the study design is shown in Figure 10.

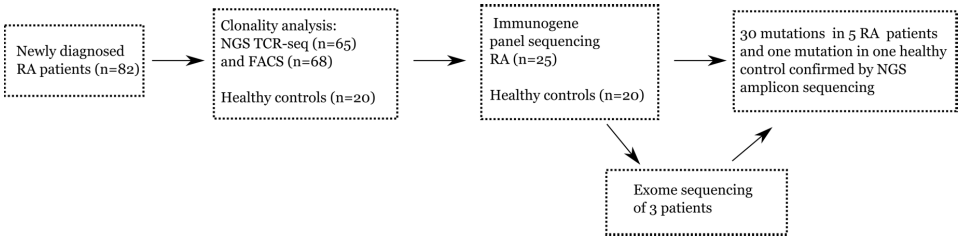

Figure 10 Study design and sample collection.

In variant calling, we compared CD4+ and CD8+ cells from each patient to discard germline variants. Variant calling with VarScan2 was performed in 2 ways: using CD4+ as a germline control and CD8+ cells as a "tumor" sample, and vice versa. The average sequencing coverage median was 397 (interquartile range, IQR 223-435) in RA patients' CD4+ cells and 386 (IQR 284-441) in CD8+ cells. In healthy controls, the median coverage was 465 (IQR 322-1293) for CD4+ amnd 437 (IQR 254-1329) for CD8+.

9.3.1 Immunogene panel and exome sequencing in RA (I)

To confirm that mutations reported by VarScan2 were true somatic mutations, variants were inspected visually with IGV and confirmed with another sequencing method, amplicon sequencing, to occur in the patient's cells. With this stringent approach, immunogene panel sequencing identified 10 mutations that occurred in CD8+ cells but not in CD4+ cells. One of the identified mutations occurred in one of the healthy controls. Exome sequencing detected 21 additional mutations in flow-sorted CD8+ T-cell clones. All mutations identified by immunogene panel and exome sequencing are shown in Figure 11.

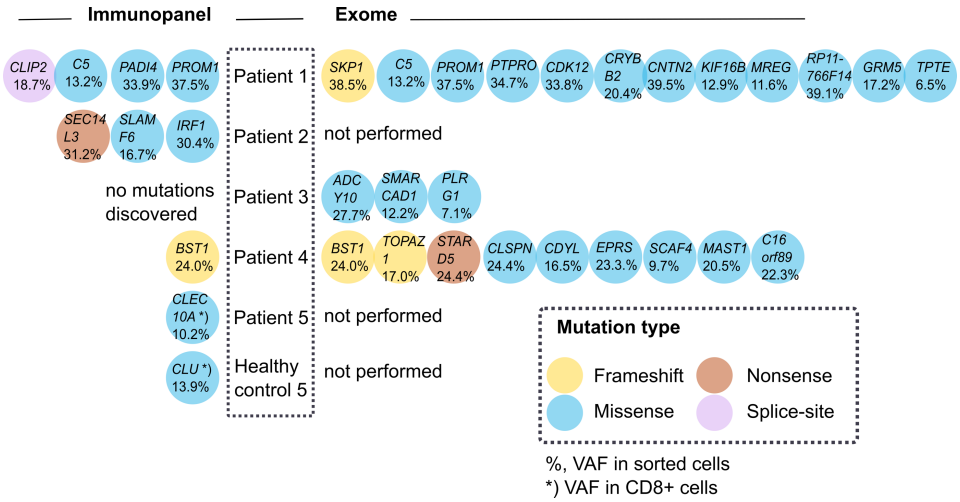

Figure 11 Somatic mutations identified in the study. Abbreviations: VAF, variant allele frequency; Immunopanel, immunogene panel.

Missense mutations comprised 81%, frameshift mutations 10%, nonsense mutations 6%, and splice-site mutations 3%. Of the mutations identified by immunogene panel sequencing, 8/10 were in genes associated with either immune functions or cell proliferation. Exome sequencing identified a larger proportion of mutations that were not associated with cell proliferation or immune functions (14/24). PolyPhen2[297] and SIFT[298] prediction algorithms predicted 32% (8/25) of missense mutations to be damaging.

9.3.2 Somatic mutations are restricted to specific CD8+ memory T cells in RA (I)

The mutations that occurred in RA patients 1–3 in study I were screened with amplicon sequencing from different cell fractions: CD4+ cells, the expanded CD8+ population, and other CD8+ cells. The mutations were absent in CD4+ cells and in other CD8+ cells than the expanded population (Figure 12).

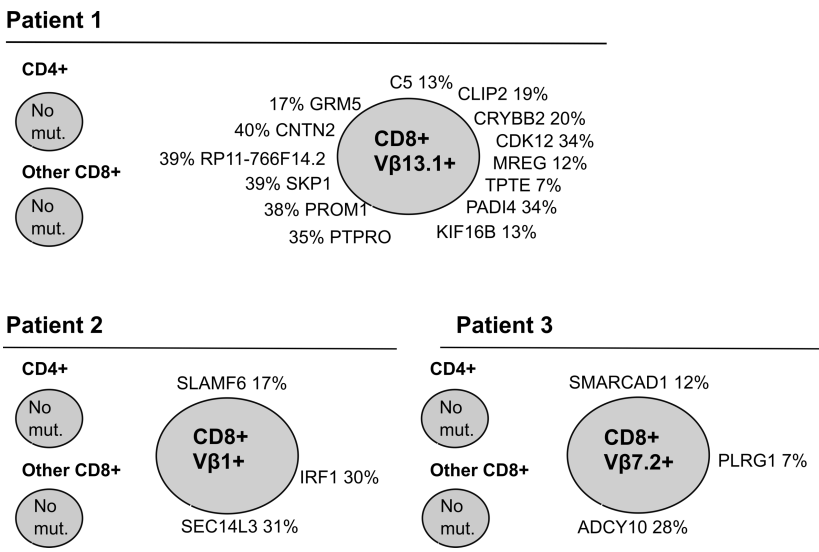

Figure 12 Somatic mutations occur only in the expanded CD8+ T cells, not in other cells. Abbreviations: mut, mutations.

Flow cyometry revealed that the mutation-harboring CD8+ T-cell clones (as determined by flow cytometry antibody staining) had a memory phenotype. Expanded clones of patients 1 and 3 had an effector-memory phenotype with decreased C-C chemokine receptor type 7 (CCR7) and CD45 expression, but the mutation-harboring clone in patient 2 comprised terminal effector-memory cells. The CD8+ TCRB repertoires of these 3 patients were also followed with TCRB sequencing, and the largest clone sizes (as % of all CD8) remained relatively stable during 0.6–3 years of follow-up (Figure 13). RNA sequencing results suggested that the expanded, mutation-harboring cell clones had a survival-associated gene expression signature.

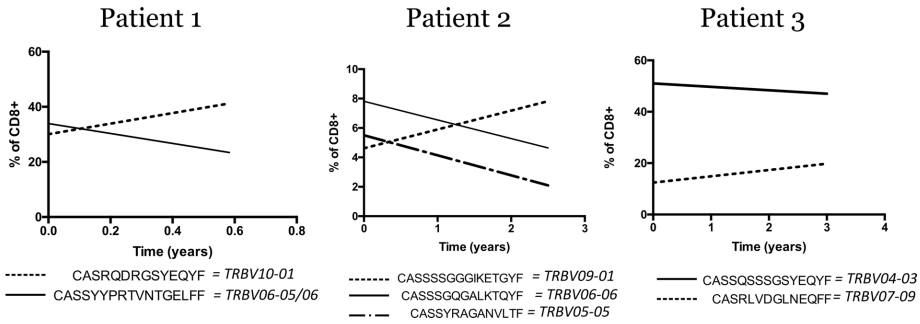

Figure 13 T-cell clones persist in follow-up.

Healthy controls harbor large T-cell clones that target CMV,[117] and thus we investigated the possibility that the major, mutation-harboring CD8+ clones in RA were CMV-specific. However, only a minority of CD8+ T cells in patients 1 and 2 were specific for the CMV NLV-pentamer. Also, previously published CMV-specific TCR sequences were queried against TCRB sequences in 65 RA patients, but major clones were not included in the public CMV-specific CMV-specific TCR sequences.

9.3.3 Immunogene panel sequencing in immunodeficiency patients (IV)

The more comprehensive sequencing panel covering 2533 genes identified 45 somatic mutations in 11 of 17 (65%) immunodeficiency patients. CVID patients harbored somatic mutations in 6 of 8 (75%) of cases. Figure 14A shows the proportion and numbers of patients and healthy controls with mutations in CD4+ and CD8+ cells. Most mutations occurred in CD8+ cells, as shown in Figure 14B. The identified mutations comprised of 38 missense, 5 nonsense, and 2 frameshift mutations. Of the 38 missense mutations, 47% were predicted to be damaging by both PolyPhen2 and SIFT algorithms.

Mutations in genes associated with hematological malignancies comprised 20% of mutations in immunodeficiency, with 9 mutations in 6 of 17 (35%) patients. Gene Ontology Consortium annotations revealed that mutations in genes affecting lymphocyte functions, inflammation, or cell proliferation were also common (Figure 14C). RNA sequencing showed that most mutated genes were expressed in CD4+ and CD8+ cells of healthy controls. Interesting highlights of the data included 2 patients with *STAT5B* mutations, 2 with *C5AR1* mutations, and one with a *KRAS* mutation.

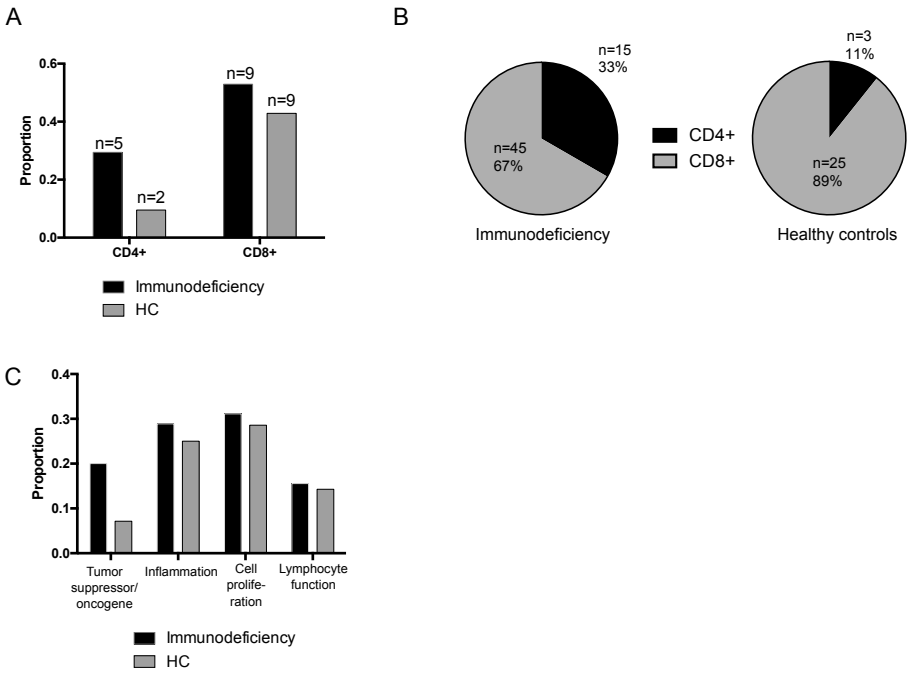

Figure 14 A. The proportion of immunodeficiency patients and healthy controls harboring mutations in CD4+ or CD8+ cells. The absolute numbers of individuals with mutations are also shown. B. The absolute numbers and proportions of mutations occurring in either CD4+ or CD8+ cells. C. The proportion of mutations occurring in tumor suppressors/oncogenes or affecting different cell functions. Abbreviations: HC, healthy control.

However, healthy controls also harboured mutations in 10 of 21 (48%) cases, with no statistically significant difference when compared to immunodeficiency patients. Altogether, 28 mutations existed in healthy control samples, and 89% of them occurred in CD8+ cells (Figure 14B). Healthy controls harbored 3 splicing and 25 missense mutations, and 25% of the missense mutations were predicted to be damaging by both PolyPhen2 and SIFT, which occurred in lower proportion than in immunodeficiency patients (Fisher's exact test p=0.034). Of note, one healthy control harbored a *NRAS* hotspot G13D mutation.

9.4 Somatic mutations in hematopoietic progenitor cells (III)

9.4.1 Clonal hematopoiesis in RA and bone marrow failure (III)

We utilized a targeted amplicon-based deep sequencing panel to identify clonal hematopoiesis in RA patients and used AA/hMDS patients as comparators for the mutational spectrum. Somatic mutations occurred in 17% (10/59) of studied RA patients, where 14% (8/59) had only one mutation and 3% (2/59) had 2 mutations. Because these mutations had a VAF of more than 2% and occurred in genes that have been previously associated with hematological malignancies, they fulfilled the criteria for CHIP. *DNMT3A* mutations were the most common, occurring in 6.8% of patients (Figure 15). *TET2* mutations were the second most common, occurring in 5.1%. None of the *DNMT3A* mutations occurred in the R882 hotspot that is commonly mutated in AML.

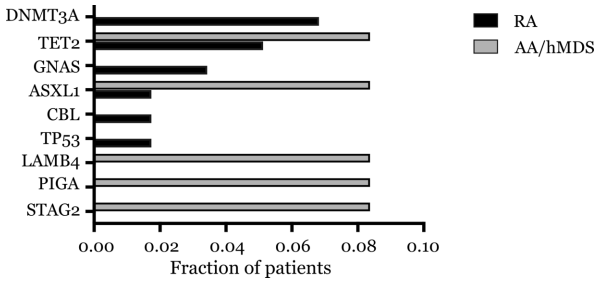

Figure 15 Frequencies of mutations in different genes in RA and AA/hMDS patients.

Nonsense and frameshift mutations were the most common mutation types, as these mutations composed 67% of all observed mutations in RA (Figure 16A). These mutations truncate the affected protein and are thus clearly damaging. C>T transitions comprised 50% of single-base alterations. In RA and AA/hMDS patients, the truncating mutations had higher median VAFs than missense mutations (Figure 16B, Mann-Whitney test p-value 0.019).

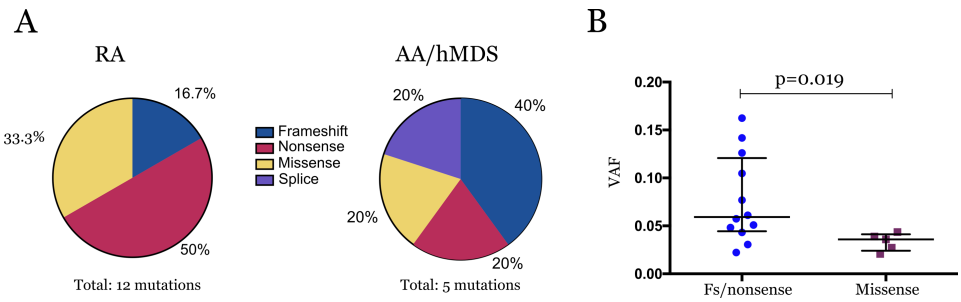

Figure 16 A. Most mutations were frameshift or nonsense mutations in RA and AA/hMDS. B. The median VAFs and interquartile ranges of mutations with frameshift/nonsense or missense consequences are shown. Statistical significance was determined by the Mann-Whitney test. Abbreviations: Splice, splice-site mutations with harmful prediction; Fs, frameshift.

Although the prevalence of clonal hematopoiesis seems to increase with age in the RA patient cohort, reaching 25% in 70-79-year-olds, patients with clonal hematopoiesis were not statistically significantly older than patients without clonal hematopoiesis (Mann-Whitney test p-value 0.31).

Acquired bone marrow failure patients (AA and hMDS) showed a slightly different mutational landscape. The prevalence of clonal hematopoiesis was 33%, and 8% (1/12) of patients had only one mutation. AA and hMDS patients were characterized by *TET2*, *ASXL1*, *LAMB4*, *PIGA* and *STAG2* mutations (Figure 15).

9.4.2 Clonal hematopoiesis and its association with clinical parameters in RA (III)

We explored whether clonal hematopoiesis was associated with disease activity or other clinical parameters in the 59 RA patients included in study III. However, clonal hematopoiesis did not associate with studied clinical parameters (Table 14; Fisher's exact test for categorical and Mann-Whitney test for numerical variables).

We obtained follow-up data up to 4 years after diagnosis on blood cell counts and indices of the RA patients (Table 14). Differences between patients with/without CH were analyzed on natural-logarithm-transformed data with a linear mixed model, with multiple comparison adjustments in post-hoc testing. The model included time, smoking status with two options (never-smoked or ex-smokers+smokers), and clonal hematopoiesis status (yes/no). In this model, only mean corpuscular volume (MCV) and lymphocyte count showed statistically

significant changes during follow-up. Time, clonal hematopoiesis, and smoking also had a statistically significant interaction effect on lymphocyte counts. However, when the patients were split into groups according to clonal hematopoiesis and smoking status, no statistically significant differences were observed.

Table 14. *Clinical parameters included in study III.*

Numerical parameters at diagnosis	Categorical parameters	Follow-up data
DAS28	Sex	MCV
ESR	Smoking status	Leukocyte count
CRP	Serostatus	Lymphocyte count
Duration of symptoms	Autoimmune disease	Neutrophil count
No. of swollen joints	Previous cancer	Monocyte count
No. of tender joints	Diabetes	Thrombocyte count
HAQ score	Astma	Hemoglobin level
Leukocyte count	Allergy	RDW
Hemoglobin level	COPD	
Lymphocyte count	Hypertension	
Monocyte count	Elevated LDL	
Thrombocyte count	Atherosclerotic disease	
Neutrophil count		
MCV		
RDW		

Abbreviations: DAS28, disease activity score 28; HAQ, health assessment questionnaire; MCV, mean corpuscular volume; RDW, red cell distribution width; LDL, low-density lipoprotein.

9.4.3 Clonal hematopoiesis variants in immunodeficiency patients (IV)

Our paired-sample variant calling strategy that uses CD4+ and CD8+ cells excludes somatic variants that occur in both CD4+ and CD8+ cells. Variants originating with hematopoietic progenitors may be left undetected if they occur in both CD4+ and CD8+ lineages. Thus, we aimed to identify variants associated with clonal hematopoiesis in immunodeficiency patients' CD4+ and CD8+ samples with a single-sample variant calling strategy.

Variants in genes associated with clonal hematopoiesis and occurring in both CD4+ and CD8+ cells in the same individual were identified in 4 of 17 immunodeficiency patients (23.5%) and in none of the healthy controls. Of the 7 discovered variants, 3 occurred in *DNMT3A*, one in *PRPF8*, and 3 in *TET2*. Curiously, all of the *TET2* mutations occurred in the same patient, and 3 were truncating mutations (frameshift or nonsense).

10 DISCUSSION

10.1 Discovery of novel and recurrent somatic mutations in T cells (I, IV)

In this study, we showed that somatic mutations in T cells are common findings in RA patients (study I) and immunodeficiency patients with lymphoproliferation and autoimmunity (study IV), but also in healthy controls (studies I and IV). Recent research has revealed a landscape of somatic mutations in other non-malignant tissues.[14] For example, somatic mutations accumulate in skeletal muscle satellite cells, stem cells in the gut, and liver, fibroblasts, and hematopoietic stem cells with age.[198,199,202,203]

Using a custom capture panel covering 986 immune-related genes, combined with exome sequencing in selected cases, study I identified somatic mutations in mature CD8+ cells in 20% of studied RA patients. In the variant calling procedure, CD8+ cells were compared to CD4+ cells, and vice versa. Thus, germline variants and somatic variants occurring in both cell fractions were excluded. Mutations were discovered only in CD8+ cells, and sorting experiments confirmed that these mutations existed exclusively in specific, expanded CD8+ T cell populations. Only one mutation was discovered in one of the 20 healthy controls.

Study I discovered that 32% of genes with mutations were associated with cell proliferation, and 19% of mutations were in immunogenes. Due to lacking functional assays, we do not know the exact effects of these mutations on the cell phenotype or disease activity. However, some of the mutations could contribute to altered immune cell functions. For example, *IRF1* regulates the type 1 interferon response, which promotes activation of adaptive immune cells.[299] An interferon signature can also be detected in RA patients.[299] Another example highlights *PTPRO,* a protein phosphatase, tumor suppressor, and a negative regulator of *STAT3,*[300] which is also an anti-apoptotic signal tranductor.[112]

Study IV employed a custom capture panel that targeted 2,533 genes. The panel covered mostly the same genes as in study I, but it was appended with additional genes associated with autoimmune disease, adaptive immunity, and hematopoietic malignancies. We applied this panel to sequence CD4+ and CD8+ cells of 21 healthy controls and 17 patients with immunodeficiency. Of the immunodeficiency patients, 8

had CVID, 3 had germline *STAT3* gain-of-function mutations, one had germline ADA2 deficiency, one had Jacobsen syndrome, and 4 had some other immune deficiency and/or severe autoimmunity. Lymphoproliferation occurred in 65% and autoimmunity in 94% of cases.

We discovered 45 somatic mutations in mature CD4+ and CD8+ cells in 65% of immunodeficiency patients, of which many occurred in genes that regulate cell proliferation, inflammation, and lymphocyte functions or are known tumor suppressors or oncogenes. Interesting highlights of discovered mutations are 2 patients with *STAT5B* mutations, one with a *KRAS* mutation, and 2 with *C5AR1* mutations. Overactive *STAT5B* and *KRAS* associate with both autoimmunity and malignancies,[115,212,301–304] whereas *C5AR1* promotes inflammatory responses in CD4+ cells.[305] We hypothesize that mutations in mature T cells can operate together or with clonal hematopoiesis variants to provide a slight survival advantage or proinflammatory properties for T cells, which can contribute to autoimmunity and/or lymphoproliferation.

Somatic mutations are not restricted to diseases such as immunodeficiency or RA. Studies I and IV revealed somatic mutations in healthy controls: study I identified only one mutation in one of 20 healthy controls, whereas study IV identified 28 mutations in 10 of 21 of individuals. This is likely due to a broader panel of sequenced genes in study IV and a lower VAF threshold in study IV: study I had a VAF cutoff at 5% and study IV at 2%. However, it is not unexpected that healthy individuals harbor some somatic mutations, even if they occurred in genes implicated for malignancy, such as *DNMT3A* and *NRAS,* because apparently healthy individuals can have somatic variants in hematopoietic cells that predispose to hematological malignancies.[226,227]

All mutations in RA patients and nearly all mutations in healthy controls occurred in CD8+ cells in studies I and IV, which is in line with findings in a study reporting somatic mutations in multiple sclerosis patients' samples.[216] The majority of mutations in immunodeficiency patients also occurred in CD8+ cells, but they harbored a slightly larger proportion of mutations in CD4+ cells than in healthy controls. The higher frequency of mutations in CD8+ cells than in CD4+ cells is likely due to at least 2 reasons. First, CD8+ cells form larger T-cell clones than CD4+ cells, which makes the detection of mutations easier with sequencing methods. Second, CD8+ proliferate highly to eliminate threats such as virus-infected or cancerous cells, and numerous cell

divisions may result in DNA replication error accumulation in highly expanding CD8+ T-cell clones.

Due to the targeted approaches used in studies I and IV, the total mutational spectrum and mutation burden in RA and immunodeficiency patients cannot be determined. However, we demonstrated that somatic mutations are a common phenomenon in newly diagnosed RA, immunodeficiency (especially late-onset CVID), and healthy controls. We also highlight mutations in genes that are implicated in autoimmunity and malignancies. Although healthy controls had some mutations with pathogenic implications, we suggest that some of the mutations in RA and immunodeficiency (such as *STAT5B* and *KRAS* mutations) could play modulative roles in disease pathology, although they are not likely to be the only causative factors for disease.

10.2 *STAT3* mutations, RA, and autoimmunity (II)

The clinical presentation of Felty' syndrome is often nearly indistinguishable from T-LGL leukemia with RA as a comanifestation. Due to the similar clinical presentations, HLA-DR4 prevalences, similar treatment choices, histopathological splenectomy findings, and LGL expansions in both diseases, many authors have suggested that these diseases are part of the same disease spectrum.[120,124,306] Study II provided molecular evidence for this hypothesis: a reclassification of Felty's syndrome and LGL leukemia would be needed, as both conditions are characterized by similar somatic *STAT3* mutations.

Activating somatic *STAT3* mutations occur in 30–75% of LGL leukemia patients' lymphocytes.[112] Study II comprised 14 patients with clinical Felty's syndrome and showed that 43% of these patients harbored somatic *STAT3* mutations. The mutations occurred in Y640F, D661, and N647I mutational hotspots in the *STAT3* SH2 domain. Although the VAFs of the observed mutations were low in T cells (1–8.8%), they were comparable with results in LGL leukemia patients: some LGL leukemia patients have also *STAT3* mutations with very small (0.8%) VAFs.[113]

Do the activating *STAT3* mutations cause the autoimmune manifestations in Felty's syndrome? Evidence for causative mechanisms on the subject is scarce, but STAT3 activation does link with autoimmune manifestations. In LGL leukemia, patients with multiple *STAT3* mutations have RA (43%) more often than do patients with no *STAT3* mutations (6%).[113] Neutropenia and anemia in LGL leukemia are also

associated with *STAT3* mutations.[107,109,110] In addition, activating germline *STAT3* mutations associate with multi-organ autoimmunity and cytopenias.[307–309] Lamy, Moignet, and Loughran proposed a mechanism for disease manifestations: STAT3 activation causes proinflammatory cytokine production, but LGLs may also target bone marrow progenitor and joint cells.[108] The exact mechanism remains unknown, however. Several possible explanations for the link between *STAT3* mutations and autoimmunity exist. Chronic inflammation may predispose patients for mutagenesis (in *STAT*3 or other genes). Alternatively, patients with detectable *STAT3* mutations may be prone to develop autoimmune manifestations. Selective pressure –mediated by inflammatory factors or antigen stimulation– on immune cells may enrich *STAT3*-mutated clones.

By determining 92 biomarkers in patients' plasma samples, study II showed that LGL leukemia patients and Felty's syndrome patients share similar cytokine profiles, although the inflammatory cytokine profile in Felty's syndrome and LGL leukemia was not entirely specific to these diseases. RA patients shared some similarities with both diseases, but since Felty's syndrome patients always have RA and LGL leukemia patients often have RA, the finding is not surprising.

However, *STAT3* mutations seem to distinguish the lymphoproliferative syndrome of LGL leukemia from "conventional" RA. In study I, we screened 82 newly diagnosed RA patients for *STAT3* hotspot mutations with sensitive methods but did not detect any mutations. *STAT3* mutations occur in multiple diseases, such as inflammatory hepatocellular adenomas and T-cell lymphomas.[112] A few reports of somatic *STAT3* mutations exist in the context of autoimmune disease. Aplastic anemia patients harbor *STAT3* mutations.[310,311] Refractory celiac disease patients' gut intraepithelial lymphocytes harbor *STAT3* D661 mutations,[215] and one multiple sclerosis patient (of 16) harbored a *STAT3* D661Y mutation.[216] Thus, activating *STAT3* mutations may exist in multiple autoimmune disease individuals, but the VAF and immune homeostasis may modulate the pathogenic potential of the mutations.

10.3 Clonal hematopoiesis as a possible regulator of immune-mediated disease (III, IV)

In addition to existing in mature T cells, somatic mutations are common in RA patients' hematopoietic stem cells. This was highlighted in study

III, in which we showed that clonal hematopoiesis occurred in 17% of RA patients. This is similar to frequencies in studies using sensitive, targeted sequencing assays in unselected individuals.[228,229] The most commonly mutated genes in RA patients were *DNTM3A* and *TET2*, and this is concordant with previous publications.[226–229] All of the mutations discovered in our study fulfilled the criteria for clonal hematopoiesis of indeterminate potential (CHIP).[233]

We hypothesized that clonal hematopoiesis could be linked with disease activity in RA. RA is a disease with longstanding and chronic inflammation,[2] and chronic inflammation may predispose to DNA damage and mutagenesis.[102] However, we did not observe any associations between inflammatory markers or disease activity and clonal hematopoiesis, which is likely due to lacking statistical power. Previous studies have not investigated clonal hematopoiesis specifically in the context of autoimmune disease. One previous study included data on the self-reported history of arthritis,[229] and a study with 361 participants with data on autoimmune disease did not identify increased rates of autoimmune disease in patients with CHIP.[209]

Data on larger patient cohorts could address the issue, as published data suggest that inflammation and clonal hematopoiesis are linked: loss of function of Tet2 in myeloid cells led to increased atherosclerosis and proinflammatory cytokine production by macrophages in mice.[238] Intriguingly, similar to clonal hematopoiesis, RA itself predisposes to both cardiovascular disease and AML.[2,101] Also, both clonal hematopoiesis and RA are associated with shortened telomeres.[231,312]

Study IV addressed clonal hematopoiesis variants in CD4+ and CD8+ cells. In immunodeficiency patients, clonal hematopoiesis variants occurred in CD4+ and CD8+ cells in 4/17 (24%). No clonal hematopoiesis variants that were shared between CD4+ and CD8+ cells were discovered in healthy controls. Although clonal hematopoiesis variants are often enriched in myeloid cells, clonal hematopoiesis occurs in several cell lineages. Whereas *DNMT3A* mutations occur in both myeloid and T cells,[209,234] *TET2* mutations seem to exert a myeloid proliferation bias: they are almost nonexistent in T cells, but they occur in higher frequencies in myeloid cells, B cells, and NK cells.[209,234] *TET2* mutations do occur in T-cell malignancies,[313–315] and Tet2 deficiency affects T- and B-cell differentiation in mice.[316] Moreover, a hypomorphic *TET2* mutation in therapeutic CAR-T cells caused major clonal expansion and persistence of the mutation-harboring cell clone.[317] We hypothesize that

clonal hematopoiesis in T cells, in combination with the other mutations, could contribute to immune dysfunction and lymphoproliferation in immunodeficiency.

10.4 T-cell clonality and somatic mutations (I, II, IV)

The human TCR repertoires are extremely complex, harboring more than $2.5*10^7$ different alphabeta TCRs.[271] High-throughput methods to characterize the TCR repertoire are relatively novel, and previously used methods include flow cytometry and spectratyping. Flow-cytometric methods in study I indicated that although flow cytometry does not distinguish unique TCR clones, it yields a relatively robust estimate of the V-gene usage of the TCR repertoire. Flow cytometry-mediated cell sorting with Vbeta-specific antibodies allows enrichment of the clone of interest. Sequencing these specific cell populations yields information on mutations in T-cell clones: in studies I and II, sorting experiments showed that the mutations occur in a "clone" as defined by flow cytometry, and not in other T cells. This finding raises questions about the origin of the identified mutations: have the mutations occurred at a late stage of T-cell development, or are they derived from progenitor cells despite being nonexistent in other cell types (Figure 17)?

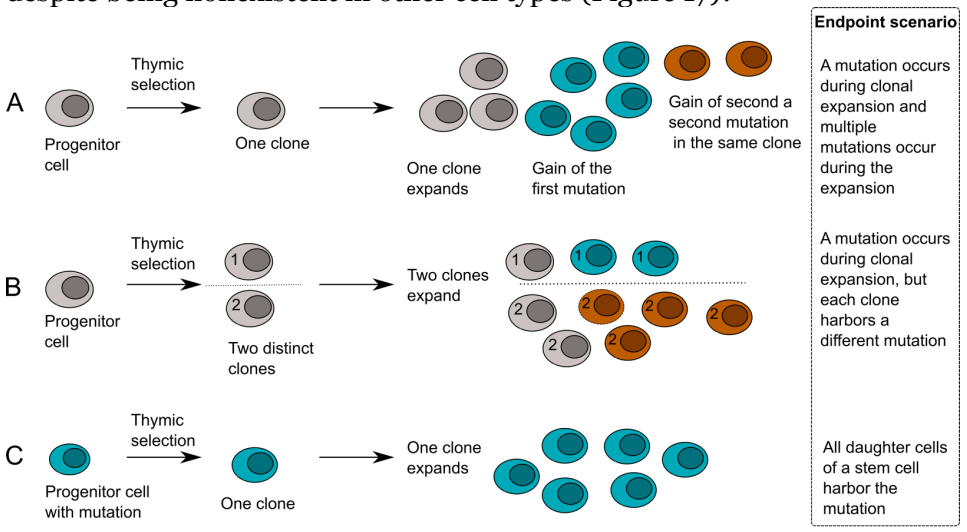

Figure 17 Different scenarios for the origin of somatic mutations in mature T cells. A. A mutation occurs in a mature T cell during clonal expansion, and another mutation occurs in the same cell clone later during the expansion. Thus, the 2 mutations have different VAFs in the same clone. B. Two distinct clones acquire two different somatic mutations during the clonal expansion process, but they harbor homologous TCRs and are thus detected as one by flow cytometry. Single-cell sequencing technologies could help to differentiate scenarios A and B. C. A mutation occurs already in a progenitor cell and is passed on to all daughter T cells. This scenario can mimic scenario A or B, because if only one T-cell clone with the mutation expands and the others do not, sequencing assays will not detect the other mutation-harboring clones.

In studies I and IV, with TCRB sequencing data, no differences in median TCRB repertoire clonalities were observed between patients and controls. Since the median ages of patients and controls were similar, and since elderly individuals harbor large T-cell clones and decreased TCR receptor repertoire diversity,[318,319] the finding is not surprising. Consistent with these observations, we identified a correlation between CD8+ clonality and increasing age in newly diagnosed RA patients. Although we described highly expanded major CD8+ clones in peripheral blood, synovial tissue in early RA is more enriched with expanded T-cell clones when compared to peripheral blood.[320]

A previous study in CVID patients by Ramesh *et al* described increased TCR clonality in CVID patients,[180] but the age of healthy controls and the heterogenous nature of CVID patients may explain the conflicting results with study IV. Taken together, the aging TCR repertoire in RA, CVID, and healthy individuals is relatively clonal, which is likely due to the physiological clonal expansions and decreased diversity in both naïve and memory T cells.[318] Greater differences in TCR clonality between diseased and healthy individuals would be observed in lymphoproliferative clonal disorders, such as LGL leukemia.

Although very large (20–50% of all CD8+) T-cell clones occur in some RA patients, immunodeficiency patients, and healthy controls, the target antigen of these T cells is unknown. Public, previously reported TCR sequences that target pathogens such as CMV and EBV were not presented in these major clones. Studying the TCR target antigen *ex vivo* is not trivial, and with no prior knowledge on the potential antigen target, high-throughput screening methods are needed to recognize potential targets.[321] Moreover, TCRs show significant plasticity, and because a single TCR can recognize multiple different peptides,[322] discovering one target does not exclude others.

Although no clear target antigen for the major T-cell clones was identified, the TCRB repertoires of CVID patients showed an overrepresentation of TCRs that had identical amino acid sequences but distinct nucleotide sequences (=convergent TCRs). Convergent TCRs are more often shared between individuals in mice experiments,[323,324] and they often have larger clone sizes.[325] An excess of TCR convergence may be a marker for selective pressure: either convergent TCRs harbor a structure that provides stronger survival signals during thymic development, or a post-thymic antigen exposure propagates the expansion of multiple, similar clones. Regardless of cause, convergence seems to be a more common phenomenon in CVID than in healthy controls.

11 CONCLUSIONS

Our aim was to characterize biologically relevant somatic mutations in the leukocytes of autoimmune disease patients using NGS and focusing on T cells. As disease models, we used RA, Felty's syndrome, and immunodeficiency co-manifested with autoimmunity.

Study I made the novel discovery of somatic mutations in immune-related genes in mature T cells in newly diagnosed RA patients. The mutations occurred exclusively in specific, expanded CD8+ T-cell populations, and many mutated genes were associated with autoimmunity or cell survival. Study III identified clonal hematopoiesis in RA patients' whole blood samples at a prevalence similar to previously published rates, although this was not associated with inflammatory markers or disease activity in this cohort.

Study II discovered somatic *STAT3* mutations in Felty's syndrome. The mutations showed that Felty's syndrome shares molecular markers with LGL leukemia: in Felty's syndrome, the mutations were located in the *STAT3* mutational hotspot, occurred at a similar prevalence as in LGL leukemia, and existed in CD8+ cells. Thus, study II provided molecular evidence for the hypothesis that Felty's syndrome and LGL leukemia are actually the same disease.

Study IV discovered clonal hematopoiesis variants and other somatic mutations in immunodeficiency patients' CD4+ and CD8+ cells. A significant proportion of the discovered mutations associate with malignancies, inflammation, lymphocyte functions, and cell proliferation. Some of the mutations had implications for lymphoproliferation and autoimmunity. Also, the TCR repertoires of CVID patients showed slight selective pressure.

These discoveries are just the beginning of our understanding of the roles somatic mutations play in autoimmune disease and immunodeficiencies. The stage is set: the knowledge of somatic mutations in non-malignant cells is growing, but the physiological effects of these mutations require more study. Future studies, with access to deep full exome and genome sequencing, will probably assess the full mutational landscape, signature, and functional effects of somatic mutations in autoimmune disease.

12 FUTURE PROSPECTS

The research community has gained small glimpses of the spectrum of somatic mutations in non-malignant cell types. Previously published works have assessed the accumulation of somatic mutations in hematopoietic, gut, liver, fibroblast, skin, and muscle cells. Although a few examples exist of somatic mutations altering non-malignant cell function and modulating non-malignant disease, the full effect of somatic mutations in human disease is far from resolved.

Autoimmune diseases and immunodeficiencies are mostly incurable with current treatment options. RA and CVID, for example, represent heterogenous groups of patients, and similar treatments may not be beneficial for all. Personalized treatment for autoimmune manifestations may be facilitated by methods such as biomarker analysis, RNA expression analysis, and identification of somatic mutations. As in cancer, targeted drugs could be used to target disease-specific processes, although more research evidence on the subject is needed.

Large-scale, collaborative efforts will be needed to elucidate the roles somatic mutations play in health and disease. Leukocytes are key mediators of autoimmune disease, and they are relatively easily accessible. Thus, they could be some of the first cells to be investigated in detail. Some possible experimental setups are listed below.

- Knowledge on the full landscape of somatic mutations in healthy individuals' cells is critical for understanding the effects of somatic mutations in disease. To catalog somatic mutations in healthy cells, exome and genome sequencing of different cell types from healthy individuals of varying ages is crucial. This approach can assess whether specific mutations enrich into different genomic areas in different cell types. Some of these data already exist, but data re-analysis is likely needed for future experimental designs.
- Cataloging somatic mutations in healthy individuals and patients with disease will allow comparisons between healthy and diseased cells' mutational burdens in different genomic areas and the assessment of possible enrichment or depletion of somatic alterations in immune-mediated diseases.

- Functional studies are needed to fully understand the impact of selected, specific mutations associated with disease. Genes may have different functions in different tissues, and even different mutations within the same gene may not have similar effects. The number of cells harboring a mutation can also affect the disease phenotype. Thus, functional experiments will need to account for huge complexity, which will likely require high-throughput experimental setups.
- Nucleic acid transitions, transversions, indels, and structural variations in DNA are not the only genetic modifications that affect cell functions. Epigenetic modifications should eventually be characterized in addition to somatic mutations.

These are merely examples of research issues to address, and when more data accumulate, experimental setups will change accordingly. Novel technologies, such as single-cell sequencing techniques, will provide many opportunities to gain insights on the effects of somatic mutations on disease pathogenesis. Although seeking answers gives rise to even more questions, as a research community, we can learn more than we can imagine in the process.

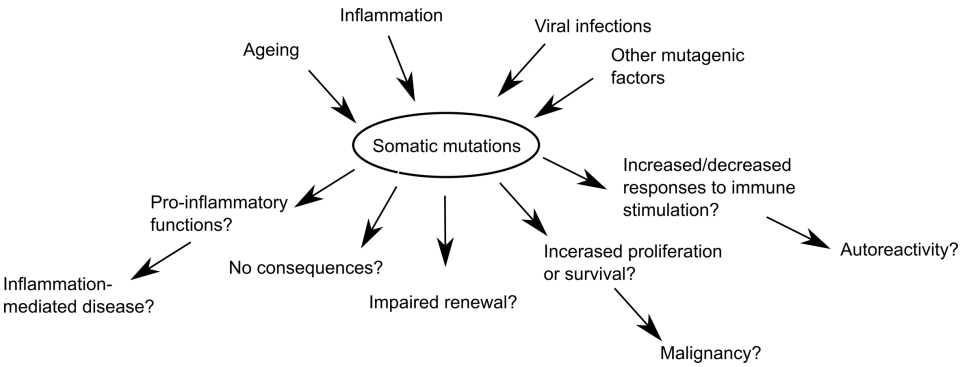

Figure 18 Some hypothetical causes and consequences of somatic mutations in immune cells.

13 ACKNOWLEDGMENTS

It is a time to say thank you. Without the efforts of multiple people, this work would not have been possible. With so much so say, but having only so few words, I hope to convey my thanks to all people involved in this project: co-authors, colleagues, helping hands, funders, but also to other people who have supported me during these years.

This work was performed in Hematology Research Unit Helsinki (HRUH), the Department of Hematology, University of Helsinki, and Helsinki University Hospital Comprehensive Cancer Center. This work was supported financially by the European Research Council (M-IMM), the Academy of Finland, the Finnish special government subsidy for health sciences, research, and training, the Sigrid Juselius Foundation, the Finnish Cancer Societies, the Finnish Cancer Institute, the Finnish Medical Association, the Instrumentarium Research Foundation, the Biomedicum Helsinki Foundation, the Orion Research Foundation, the Paulo Foundation, the K. Albin Johansson Foundation, and the Juhani Aho Foundation. I would like to thank the Doctoral School in Health Sciences and the Doctoral Programme in Clinical Research for support. I would like to give special thanks for the Helsinki University MD/PhD Programme for support, financial and otherwise, when combining both MD and PhD studies.

I wish to thank my supervisors, Satu and Tiina, for mentoring. Satu, thank you for guidance, help, and support during all these years. Thank you for caring not only for research issues but also for the people doing the work. Tiina, we have had great scientific brainstorms, technical issue troubleshootings, and discussions of life in general. I am grateful for your help and support, but, most of all, your friendship.

I wish to thank Prof. Lambert Busque for accepting the invitation to be my opponent for the thesis dissertation. I am also grateful for the thesis reviewers, docent Arto Orpana and docent Laura Pirilä, for constructive feedback on the thesis. I wish also to thank docent Päivi Saavalainen and Professor Kari Eklund for being on my thesis committee and for their constructive feedback.

I am also grateful to have worked with experts from different fields. Marjatta Leirisalo-Repo has been invaluable for patient cohort collections. Marjatta, I am grateful for your support for me and help on patient-related matters. Clinical immunology experts, Mikko Seppänen

and Timi Martelius, thank for your work with study patients and for discussions on the projects. I wish to also thank Kimmo Porkka for support and scientific feedback on my projects. Special thanks also go to our other collaborators who have provided valuable patient samples: Riitta Koivuniemi, Tuulikki Sokka-Isler, Thomas Loughran, Thomas Olson, and Markku Kauppi.

It has been a privilege to work with so many great people at the HRUH. It has been great to brainstorm how to cure autoimmunity, cancer, world poverty, climate change, or any other pressing matter. Special thanks to my fellow colleagues pursuing the MD/PhD track—Olli, Sofie, and Jani—but also to Oscar, Helena and Shady for sharing both the often laborious MD/PhD path but also for the science and fun moments. I am equally grateful to have worked with and have had a great time with other students in the lab: Heikki, Henna, Yrjö, Antti, Moon, Dipa, Daehong, and Mikko T.

Mikko K. and Hanna R., thank you for being role models as researcher-clinicians, but also for the scientific discussion, help, and support. Samuli, Matti, and Hena: without you, this work would not have been possible. I am very grateful for your hard work, and I deeply respect your expertise. Hanna L, Tiina Ka, Jay—thank you for your help with technical issues and patient samples. Saara and Minna: I am grateful for your help with samples and for your friendly support. Special thanks also belong to the Seqlab staff for troubleshooting, expertise, and help during the years, and to all of my co-authors for their contributions in the projects.

My friends have been an invaluable asset. Thank you for sharing life, happy and sad times, for giving me lots of opportunities to think of stuff other than medicine or science, and for letting me learn from you. Marjo, Paula R., Alli, Elsa, Sara, Eveliina, Katja, Jenni, Eerik, Janne, Iiku, Anu, Ellu, Annika, Tiia, Karri, Erika, Tuomas, and others, I am grateful for your friendship.

Finally, I would like to thank my family. Mom and Dad, thank you for your never-ending love, support, and help in all sorts of matters in life. My dear little sister, Pihla, thank you for supporting me in the good and bad days, but especially for the inside jokes we share. My other siblings, Päivi, Pekka, Pauli, Pia, Petri, Tessa, and Pasi: I am grateful to have you all and to have grown up with you. Thank you for the help, fun, and banter during the past 26 years and for the years to come.

I am sincerely grateful to have shared the road with colleagues, friends, and a family like mine. Even though all days are not sugar and rainbows, I would not change this time for another. With you, this has been the best time of my life.

Espoo, 14.12.2018

Paula Savola

14 REFERENCES

1. Scott, D. L., Wolfe, F. & Huizinga, T. W. J. Rheumatoid arthritis. *Lancet* **376,** 1094–1108 (2010).
2. Smolen, J. S., Aletaha, D. & McInnes, I. B. Rheumatoid arthritis. *Lancet* **388,** 2023–2038 (2016).
3. Firestein, G. S. & McInnes, I. B. Immunopathogenesis of Rheumatoid Arthritis. *Immunity* **46,** 183–196 (2017).
4. Petrelli, A. & van Wijk, F. CD8+ T cells in human autoimmune arthritis: the unusual suspects. *Nat. Rev. Rheumatol.* **12,** 421–428 (2016).
5. Bonilla, F. A., Barlan, I., Chapel, H., Carvalho, B. T. C.-, Cunningham-Rundles, C., Morena, M. T. de la, Espinosa-Rosales, F. J., Hammarström, L., Nonoyama, S., Quinti, I., Routes, J. M., Tang, M. L. K. & Warnatz, K. International Consensus Document (ICON): Common Variable Immunodeficiency Disorders. *J Allergy Clin Immunol Pr.* **4,** 38–59 (2016).
6. Wong, G. K. & Huissoon, A. P. T-cell abnormalities in common variable immunodeficiency: The hidden defect. *J. Clin. Pathol.* **69,** 672–676 (2016).
7. Bogaert, D. J. A., Dullaers, M., Lambrecht, B. N., Vermaelen, K. Y., De Baere, E. & Haerynck, F. Genes associated with common variable immunodeficiency: one diagnosis to rule them all? *J. Med. Genet.* **53,** 575–590 (2016).
8. van Schouwenburg, P. A., Davenport, E. E., Kienzler, A. K., Marwah, I., Wright, B., Lucas, M., Malinauskas, T., Martin, H. C., Lockstone, H. E., Cazier, J. B., Chapel, H. M., Knight, J. C. & Patel, S. Y. Application of whole genome and RNA sequencing to investigate the genomic landscape of common variable immunodeficiency disorders. *Clin. Immunol.* **160,** 301–314 (2015).
9. de Valles-Ibáñez, G., Esteve-Solé, A., Piquer, M., Azucena González-Navarro, E., Hernandez-Rodriguez, J., Laayouni, H., González-Roca, E., Plaza-Martin, A. M., Deyà-Martínez, Á., Martín-Nalda, A., Martínez-Gallo, M., García-Prat, M., del Pino-Molina, L., Cuscó, I., Codina-Solà, M., Batlle-Masó, L., Solís-Moruno, M., Marquès-Bonet, T., Bosch, E., *et al.* Evaluating the genetics of common variable immunodeficiency: Monogenetic model and beyond. *Front. Immunol.* **9,** 636 (2018).
10. Maffucci, P., Filion, C. A., Boisson, B., Itan, Y., Shang, L., Casanova, J.-L. & Cunningham-Rundles, C. Genetic Diagnosis Using Whole Exome Sequencing in Common Variable Immunodeficiency. *Front. Immunol.* **7,** (2016).
11. Selenius, J. S., Martelius, T., Pikkarainen, S., Siitonen, S., Mattila, E., Pietikäinen, R., Suomalainen, P., Aalto, A. H., Saarela, J., Einarsdottir, E., Järvinen, A., Färkkilä, M., Kere, J. & Seppänen, M. Unexpectedly high prevalence of common variable immunodeficiency in Finland. *Front. Immunol.* **8,** 1190 (2017).

12. Salzer, U., Warnatz, K. & Peter, H. H. Common variable immunodeficiency - an update. *Arthritis Res. Ther.* **14,** 223 (2012).

13. Hanahan, D. & Weinberg, R. A. The hallmarks of cancer. *Cell* **100,** 57–70 (2000).

14. Martincorena, I. & Campbell, P. J. Somatic mutation in cancer and normal cells. *Science (80-.).* **349,** 1483–1489 (2015).

15. Storey, G. O., Comer, M., Scott, D. L. & Corner, M. Chronic Arthritis before 1876:Early British Cases Suggesting Rheumatoid Arthritis. *Ann. Rheum. Dis.* **53,** 557–560 (1994).

16. Garrod, A. A treatise on gout and rheumatic gout (rheumatoid arthritis). *London: Longman Green* (1876).

17. Alamanos, Y., Voulgari, P. V. & Drosos, A. A. Incidence and Prevalence of Rheumatoid Arthritis, Based on the 1987 American College of Rheumatology Criteria: A Systematic Review. *Semin. Arthritis Rheum.* **36,** 182–188 (2006).

18. Humphreys, J. H., Verstappen, S. M. M., Hyrich, K. L., Chipping, J. R., Marshall, T. & Symmons, D. P. M. The incidence of rheumatoid arthritis in the UK: Comparisons using the 2010 ACR/EULAR classification criteria and the 1987 ACR classification criteria. Results from the Norfolk Arthritis Register. *Ann. Rheum. Dis.* **72,** 1315–1320 (2013).

19. Eriksson, J. K., Neovius, M., Ernestam, S., Lindblad, S., Simard, J. F. & Askling, J. Incidence of rheumatoid arthritis in Sweden: A Nationwide population-based assessment of incidence, its determinants, and treatment penetration. *Arthritis Care Res.* **65,** 870–878 (2013).

20. Deane, K. D. & Deane, K. D. Preclinical Rheumatoid Arthritis and Rheumatoid Arthritis Prevention. *Curr. Rheumatol. Rep.* **20,** 50 (2018).

21. Heliovaara, M., Aho, K., Aromaa, A., Knekt, P. & Reunanen, A. Smoking and risk of rheumatoid arthritis. *J. Rheumatol.* **20,** 1830–1835 (1993).

22. Källberg, H., Padyukov, L., Plenge, R. M., Rönnelid, J., Gregersen, P. K., van der Helm-van Mil, A. H. M., Toes, R. E. M., Huizinga, T. W., Klareskog, L. & Alfredsson, L. Gene-Gene and Gene-Environment Interactions Involving HLA-DRB1, PTPN22, and Smoking in Two Subsets of Rheumatoid Arthritis. *Am. J. Hum. Genet.* **80,** 867–875 (2007).

23. Fuggle, N. R., Smith, T. O., Kaul, A. & Sofat, N. Hand to mouth: A systematic review and meta-analysis of the association between rheumatoid arthritis and periodontitis. *Front. Immunol.* **7,** 80 (2016).

24. Potempa, J., Mydel, P. & Koziel, J. The case for periodontitis in the pathogenesis of rheumatoid arthritis. *Nat. Rev. Rheumatol.* **13,** 606–620 (2017).

25. Qin, B., Yang, M., Fu, H., Ma, N., Wei, T., Tang, Q., Hu, Z., Liang, Y., Yang, Z. & Zhong, R. Body mass index and the risk of rheumatoid arthritis: a systematic review and dose-response meta-analysis. *Arthritis Res. Ther.* **17,** 86 (2015).

26. Lahiri, M., Morgan, C., Symmons, D. P. M. & Bruce, I. N. Modifiable risk factors for RA: Prevention, better than cure? *Rheumatology* **51,** 499–512 (2012).

27. Orellana, C., Saevarsdottir, S., Klareskog, L., Karlson, E. W., Alfredsson, L. & Bengtsson, C. Oral contraceptives, breastfeeding and the risk of developing rheumatoid arthritis: Results from the Swedish EIRA study. *Ann. Rheum. Dis.* **76,** 1845–1852 (2017).

28. MacGregor, A. J., Snieder, H., Rigby, A. S., Koskenvuo, M., Kaprio, J., Aho, K. & Silman, A. J. Characterizing the quantitative genetic contribution to rheumatoid arthritis using data from twins. *Arthritis Rheum.* **43,** 30–37 (2000).

29. Van Der Woude, D., Houwing-Duistermaat, J. J., Toes, R. E. M., Huizinga, T. W. J., Thomson, W., Worthington, J., Van Der Helm-Van Mil, A. H. M. & De Vries, R. R. P. Quantitative heritability of anti-citrullinated protein antibody-positive and anti-citrullinated protein antibody-negative rheumatoid arthritis. *Arthritis Rheum.* **60,** 916–923 (2009).

30. Deighton, C. M., Walker, D. J., Griffiths, I. D. & Roberts, D. F. The contribution of HLA to rheumatoid arthritis. *Clin. Genet.* **36,** 178–182 (1989).

31. Reveille, J. D. The genetic contribution to the pathogenesis of rheumatoid arthritis. *Curr. Opin. Rheumatol.* **10,** 187–200 (1998).

32. Van Der Woude, D., Lie, B. A., Lundström, E., Balsa, A., Feitsma, A. L., Houwing-Duistermaat, J. J., Verduijn, W., Nordang, G. B. N., Alfredsson, L., Klareskog, L., Pascual-Salcedo, D., Gonzalez-Gay, M. A., Lopez-Nevot, M. A., Valero, F., Roep, B. O., Huizinga, T. W. J., Kvien, T. K., Martín, J., Padyukov, L., *et al.* Protection against anti-citrullinated protein antibody-positive rheumatoid arthritis is predominantly associated with HLA-DRB1*1301: A meta-analysis of HLA-DRB1 associations with anti-citrullinated protein antibody-positive and anti-citrullinated protein . *Arthritis Rheum.* **62,** 1236–1245 (2010).

33. Huizinga, T. W. J., Amos, C. I., Van Der Helm-Van Mil, A. H. M., Chen, W., Van Gaalen, F. A., Jawaheer, D., Schreuder, G. M. T., Wener, M., Breedveld, F. C., Ahmad, N., Lum, R. F., De Vries, R. R. P., Gregersen, P. K., Toes, R. E. M. & Criswell, L. A. Refining the complex rheumatoid arthritis phenotype based on specificity of the HLA-DRB1 shared epitope for antibodies to citrullinated proteins. *Arthritis Rheum.* **52,** 3433–3438 (2005).

34. Stastny, P. Association of the B-cell alloantigen DRw4 with rheumatoid arthritis. *N. Engl. J. Med.* **298,** 869–871 (1978).

35. Gregersen, P. K., Silver, J. & Winchester, R. J. The Shared Epitope Hypothesis. *Arthritis Rheum.* **30,** 1205–1212 (1987).

36. Raychaudhuri, S., Sandor, C., Stahl, E. A., Freudenberg, J., Lee, H. S., Jia, X., Alfredsson, L., Padyukov, L., Klareskog, L., Worthington, J., Siminovitch, K. A., Bae, S. C., Plenge, R. M., Gregersen, P. K. & De Bakker, P. I. W. Five amino acids in three HLA proteins explain most of the association between MHC and seropositive rheumatoid arthritis. *Nat. Genet.* **44,** 291–296 (2012).

37. Okada, Y., Wu, D., Trynka, G., Raj, T., Terao, C., Ikari, K., Kochi, Y., Ohmura, K., Suzuki, A., Yoshida, S., Graham, R. R., Manoharan, A.,

Ortmann, W., Bhangale, T., Denny, J. C., Carroll, R. J., Eyler, A. E., Greenberg, J. D., Kremer, J. M., *et al.* Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature* **506,** 376–381 (2014).

38. Nishimura, K., Sugiyama, D., Kogata, Y., Tsuji, G. & Nakazawa, T. Annals of Internal Medicine Review Meta-analysis : Diagnostic Accuracy of Anti – Cyclic Citrullinated Peptide Antibody and Rheumatoid Factor for Rheumatoid Arthritis. *Ann. Intern. Med.* **146,** 797–808 (2007).

39. Rantapää-Dahlqvist, S., De Jong, B. A. W., Berglin, E., Hallmans, G., Wadell, G., Stenlund, H., Sundin, U. & Van Venrooij, W. J. Antibodies Against Cyclic Citrullinated Peptide and IgA Rheumatoid Factor Predict the Development of Rheumatoid Arthritis. *Arthritis Rheum.* **48,** 2741–2749 (2003).

40. Nielen, M. M. J., Van Schaardenburg, D., Reesink, H. W., Van De Stadt, R. J., Van Der Horst-Bruinsma, I. E., De Koning, M. H. M. T., Habibuw, M. R., Vandenbroucke, J. P. & Dijkmans, B. A. C. Specific Autoantibodies Precede the Symptoms of Rheumatoid Arthritis: A Study of Serial Measurements in Blood Donors. *Arthritis Rheum.* **50,** 380–386 (2004).

41. Shi, J., Knevel, R., Suwannalai, P., van der Linden, M. P., Janssen, G. M. C., van Veelen, P. A., Levarht, N. E. W., van der Helm-van Mil, A. H. M., Cerami, A., Huizinga, T. W. J., Toes, R. E. M. & Trouw, L. A. Autoantibodies recognizing carbamylated proteins are present in sera of patients with rheumatoid arthritis and predict joint damage. *Proc. Natl. Acad. Sci.* **108,** 17372–17377 (2011).

42. Darrah, E. & Andrade, F. Rheumatoid arthritis and citrullination. *Curr. Opin. Rheumatol.* **30,** 72–78 (2018).

43. Schellekens, G. A., De Jong, B. A. W., van den Hoogen, F. H. J., van de Putte, L. B. A. & van Venrooij, W. J. Citrulline is an Essential Constituent of Antigenic Determinants Recognized by Rheumatoid Arthritis–specific. *J. Clin. Invest.* **101,** 273–281 (1998).

44. Klareskog, L., Malmström, V., Lundberg, K., Padyukov, L. & Alfredsson, L. Smoking, citrullination and genetic variability in the immunopathogenesis of rheumatoid arthritis. *Semin. Immunol.* **23,** 92–98 (2011).

45. Klareskog, L., Stolt, P., Lundberg, K., Källberg, H., Bengtsson, C., Grunewald, J., Rönnelid, J., Harris, H. E., Ulfgren, A. K., Rantapää-Dahlqvist, S., Eklund, A., Padyukov, L. & Alfredsson, L. A new model for an etiology of rheumatoid arthritis: Smoking may trigger HLA-DR (shared epitope)-restricted immune reactions to autoantigens modified by citrullination. *Arthritis Rheum.* **54,** 38–46 (2006).

46. Makrygiannakis, D., Hermansson, M., Ulfgren, A. K., Nicholas, A. P., Zendman, A. J. W., Eklund, A., Gmnewald, J., Skold, C. M., Klareskog, L. & Catrina, A. I. Smoking increases peptidylarginine deiminase 2 enzyme expression in human lungs and increases citrullination in BAL cells. *Ann. Rheum. Dis.* **67,** 1488–1492 (2008).

47. Ting, Y. T., Petersen, J., Ramarathinam, S. H., Scally, S. W., Loh, K. L., Thomas, R., Suri, A., Baker, D. G., Purcell, A. W., Reid, H. H. & Rossjohn, J. The interplay between citrullination and HLA-DRB1 polymorphism in

shaping peptide binding hierarchies in rheumatoid arthritis. *J. Biol. Chem.* **293,** 3236–3251 (2018).

48. Pieper, J., Dubnovitsky, A., Gerstner, C., James, E. A., Rieck, M., Kozhukh, G., Tandre, K., Pellegrino, S., Gebe, J. A., Rönnblom, L., Sandalova, T., Kwok, W. W., Klareskog, L., Buckner, J. H., Achour, A. & Malmström, V. Memory T cells specific to citrullinated α -enolase are enriched in the rheumatic joint. *J. Autoimmun.* 1–10 (2018). doi:10.1016/j.jaut.2018.04.004

49. Gerstner, C., Dubnovitsky, A., Sandin, C., Kozhukh, G., Uchtenhagen, H., James, E. A., Rönnelid, J., Ytterberg, A. J., Pieper, J., Reed, E., Tandre, C., Rieck, M., Zubarev, R. A., Rönnblom, L., Sandalova, T., Buckner, J. H., Achour, A. & Malmström, V. Functional and Structural Characterization of a Novel HLA-DRB1*04:01-Restricted α -Enolase T Cell Epitope in Rheumatoid Arthritis. *Front. Immunol.* **7,** 494 (2016).

50. Chemin, K., Pollastro, S., James, E., Ge, C., Albrecht, I., Herrath, J., Gerstner, C., Tandre, K., Sampaio Rizzi, T., Rönnblom, L., Catrina, A., Holmdahl, R., Klareskog, L., De Vries, N. & Malmström, V. A Novel HLA-DRB1*10:01-Restricted T Cell Epitope from Citrullinated Type II Collagen Relevant to Rheumatoid Arthritis. *Arthritis Rheumatol.* **68,** 1124–1135 (2016).

51. Scally, S. W., Petersen, J., Law, S. C., Dudek, N. L., Nel, H. J., Loh, K. L., Wijeyewickrema, L. C., Eckle, S. B. G., van Heemst, J., Pike, R. N., McCluskey, J., Toes, R. E., La Gruta, N. L., Purcell, A. W., Reid, H. H., Thomas, R. & Rossjohn, J. A molecular basis for the association of the *HLA-DRB1* locus, citrullination, and rheumatoid arthritis. *J. Exp. Med.* **210,** 2569–2582 (2013).

52. James, E. A., Rieck, M., Pieper, J., Gebe, J. A., Yue, B. B., Tatum, M., Peda, M., Sandin, C., Klareskog, L., Malmström, V. & Buckner, J. H. Citrulline-specific Th1 cells are increased in rheumatoid arthritis and their frequency is influenced by disease duration and therapy. *Arthritis Rheumatol.* **66,** 1712–1722 (2014).

53. Lundberg, K., Nijenhuis, S., Vossenaar, E. R., Palmblad, K., van Venrooij, W. J., Klareskog, L., Zendman, A. J. W. & Harris, H. E. Citrullinated proteins have increased immunogenicity and arthritogenicity and their presence in arthritic joints correlates with disease severity. *Arthritis Res. Ther.* **7,** R458-67 (2005).

54. Uysal, H., Bockermann, R., Nandakumar, K. S., Sehnert, B., Bajtner, E., Engström, Å., Serre, G., Burkhardt, H., Thunnissen, M. M. G. M. & Holmdahl, R. Structure and pathogenicity of antibodies specific for citrullinated collagen type II in experimental arthritis. *J. Exp. Med.* **206,** 449–462 (2009).

55. Stolt, P., Yahya, A., Bengtsson, C., Källberg, H., Rönnelid, J., Lundberg, I., Klareskog, L., Alfredsson, L., Lindahl, G., Sverdrup, B., Hellström, H., Weitoft, T., Lindell, B., Nordmark, B., Bratt, J., Hafström, I., Leden, I., Löfström, B., Bengtsson, A., *et al.* Silica exposure among male current smokers is associated with a high risk of developing ACPA-positive rheumatoid arthritis. *Ann. Rheum. Dis.* **69,** 1072–1076 (2010).

56. Too, C. L., Muhamad, N. A., Ilar, A., Padyukov, L., Alfredsson, L., Klareskog, L., Murad, S. & Bengtsson, C. Occupational exposure to textile dust increases the risk of rheumatoid arthritis: Results from a Malaysian population-based case-control study. *Ann. Rheum. Dis.* **75,** 997–1002 (2016).

57. Catrina, A. I., Jimmy Ytterberg, A., Reynisdottir, G., Malmström, V. & Klareskog, L. Lungs, joints and immunity against citrullinated proteins in rheumatoid arthritis. *Nat. Rev. Rheumatol.* **10,** 645–653 (2014).

58. Hu, X., Kim, H., Stahl, E., Plenge, R., Daly, M. & Raychaudhuri, S. Integrating autoimmune risk loci with gene-expression data identifies specific pathogenic immune cell subsets. *Am. J. Hum. Genet.* **89,** 496–506 (2011).

59. Trynka, G., Sandor, C., Han, B., Xu, H., Stranger, B. E., Liu, X. S. & Raychaudhuri, S. Chromatin marks identify critical cell types for fine mapping complex trait variants. *Nat. Genet.* **45,** 124–130 (2013).

60. Zhu, Z., Zhang, F., Hu, H., Bakshi, A., Robinson, M. R., Powell, J. E., Montgomery, G. W., Goddard, M. E., Wray, N. R., Visscher, P. M. & Yang, J. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat. Genet.* **48,** 481–487 (2016).

61. Thalayasingam, N., Nair, N., Skelton, A. J., Massey, J., Anderson, A. E., Clark, A. D., Diboll, J., Lendrem, D. W., Reynard, L. N., Cordell, H. J., Eyre, S., Isaacs, J. D., Barton, A. & Pratt, A. G. CD4+ and B Lymphocyte Expression Quantitative Traits at Rheumatoid Arthritis Risk Loci in Patients With Untreated Early Arthritis: Implications for Causal Gene Identification. *Arthritis Rheumatol.* **70,** 361–370 (2018).

62. Han, B., Diogo, D., Eyre, S., Kallberg, H., Zhernakova, A., Bowes, J., Padyukov, L., Okada, Y., González-Gay, M. A., Rantapää-Dahlqvist, S., Martin, J., Huizinga, T. W. J., Plenge, R. M., Worthington, J., Gregersen, P. K., Klareskog, L., De Bakker, P. I. W. & Raychaudhuri, S. Fine mapping seronegative and seropositive rheumatoid arthritis to shared and distinct HLA alleles by adjusting for the effects of heterogeneity. *Am. J. Hum. Genet.* **94,** 522–532 (2014).

63. Boilard, E., Nigrovic, P. A., Larabee, K., Watts, G. F. M., Coblyn, J. S., Weinblatt, M. E., Massarotti, E. M., Remold-O 'donnell, E., Farndale, R. W., Ware, J. & Lee, D. M. Platelets Amplify Inflammation in Arthritis via Collagen- Dependent Microparticle Production. *Science (80-.).* **327,** 580–583 (2010).

64. Rivellese, F., Nerviani, A., Rossi, F. W., Marone, G., Matucci-Cerinic, M., de Paulis, A. & Pitzalis, C. Mast cells in rheumatoid arthritis: friends or foes? *Autoimmun. Rev.* **16,** 557–563 (2017).

65. Falconer, J., Murphy, A. N., Young, S., Clark, A. R., Tiziani, S., Guma, M. & Buckley, C. D. Synovial cell metabolism and chronic inflammation in rheumatoid arthritis. *Arthritis Rheumatol.* **70,** 984–999 (2018).

66. Udalova, I. A., Mantovani, A. & Feldmann, M. Macrophage heterogeneity in the context of rheumatoid arthritis. *Nat. Rev. Rheumatol.* **12,** 472–485 (2016).

67. Mizoguchi, F., Slowikowski, K., Wei, K., Marshall, J. L., Rao, D. A., Chang, S. K., Nguyen, H. N., Noss, E. H., Turner, J. D., Earp, B. E., Blazar, P. E., Wright, J., Simmons, B. P., Donlin, L. T., Kalliolias, G. D., Goodman, S. M., Bykerk, V. P., Ivashkiv, L. B., Lederer, J. A., *et al.* Functionally distinct disease-associated fibroblast subsets in rheumatoid arthritis. *Nat. Commun.* **9,** 789 (2018).

68. Truchetet, M. E., Dublanc, S., Barnetche, T., Vittecoq, O., Mariette, X., Richez, C., Blanco, P., Mahler, M., Contin-Bordes, C. & Schaeverbeke, T. Association of the Presence of Anti–Carbamylated Protein Antibodies in Early Arthritis With a Poorer Clinical and Radiologic Outcome: Data From the French ESPOIR Cohort. *Arthritis Rheumatol.* **69,** 2292–2302 (2017).

69. Stahl, E. A., Raychaudhuri, S., Remmers, E. F., Xie, G., Eyre, S., Thomson, B. P., Li, Y., Kurreeman, F. A. S., Zhernakova, A., Hinks, A., Guiducci, C., Chen, R., Alfredsson, L., Amos, C. I., Ardlie, K. G., Barton, A., Bowes, J., Brouwer, E., Burtt, N. P., *et al.* Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci. *Nat. Genet.* **42,** 508–514 (2010).

70. Eyre, S., Bowes, J., Diogo, D., Lee, A., Barton, A., Martin, P., Zhernakova, A., Stahl, E., Viatte, S., McAllister, K., Amos, C. I., Padyukov, L., Toes, R. E. M., Huizinga, T. W. J., Wijmenga, C., Trynka, G., Franke, L., Westra, H. J., Alfredsson, L., *et al.* High-density genetic mapping identifies new susceptibility loci for rheumatoid arthritis. *Nat. Genet.* **44,** 1336–1340 (2012).

71. Okada, Y., Terao, C., Ikari, K., Kochi, Y., Ohmura, K., Suzuki, A., Kawaguchi, T., Stahl, E. A., Kurreeman, F. A. S., Nishida, N., Ohmiya, H., Myouzen, K., Takahashi, M., Sawada, T., Nishioka, Y., Yukioka, M., Matsubara, T., Wakitani, S., Teshima, R., *et al.* Meta-analysis identifies nine new loci associated with rheumatoid arthritis in the Japanese population. *Nat. Genet.* **44,** 511–516 (2012).

72. Frisell, T., Hellgren, K., Alfredsson, L., Raychaudhuri, S., Klareskog, L. & Askling, J. Familial aggregation of arthritis-related diseases in seropositive and seronegative rheumatoid arthritis: A register-based case-control study in Sweden. *Ann. Rheum. Dis.* **75,** 183–189 (2016).

73. Bossini-Castillo, L., De Kovel, C., Kallberg, H., Van't Slot, R., Italiaander, A., Coenen, M., Tak, P. P., Posthumus, M. D., Wijmenga, C., Huizinga, T., Van Der Helm-Van Mil, A. H. M., Stoeken-Rijsbergen, G., Rodriguez-Rodriguez, L., Balsa, A., González-Álvaro, I., González-Gay, M. Á., Gómez-Vaquero, C., Franke, B., Vermeulen, S., *et al.* A genome-wide association study of rheumatoid arthritis without antibodies against citrullinated peptides. *Ann. Rheum. Dis.* **74,** (2015).

74. Okada, Y., Suzuki, A., Ikari, K., Terao, C., Kochi, Y., Ohmura, K., Higasa, K., Akiyama, M., Ashikawa, K., Kanai, M., Hirata, J., Suita, N., Teo, Y. Y., Xu, H., Bae, S. C., Takahashi, A., Momozawa, Y., Matsuda, K., Momohara, S., *et al.* Contribution of a Non-classical HLA Gene, HLA-DOA, to the Risk of Rheumatoid Arthritis. *Am. J. Hum. Genet.* **99,** 366–374 (2016).

75. Viatte, S., Plant, D., Bowes, J., Lunt, M., Eyre, S., Barton, A. & Worthington, J. Genetic markers of rheumatoid arthritis susceptibility in anti-citrullinated peptide antibody negative patients. *Ann. Rheum. Dis.* **71,** 1984–1990 (2012).

76. Aletaha, D., Neogi, T., Silman, A. J., Funovits, J., Felson, D. T., Bingham, C. O., Birnbaum, N. S., Burmester, G. R., Bykerk, V. P., Cohen, M. D., Combe, B., Costenbader, K. H., Dougados, M., Emery, P., Ferraccioli, G., Hazes, J. M. W., Hobbs, K., Huizinga, T. W. J., Kavanaugh, A., *et al.* 2010 Rheumatoid arthritis classification criteria: An American College of Rheumatology/European League Against Rheumatism collaborative initiative. *Arthritis Rheum.* **62,** 2569–2581 (2010).

77. Arnett, F. C., Edworthy, S. M., Bloch, D. A., Mcshane, D. J., Fries, J. F., Cooper, N. S., Healey, L. A., Kaplan, S. R., Liang, M. H., Luthra, H. S., Medsger, T. A., Mitchell, D. M., Neustadt, D. H., Pinals, R. S., Schaller, J. G., Sharp, J. T., Wilder, R. L. & Hunder, G. G. The american rheumatism association 1987 revised criteria for the classification of rheumatoid arthritis. *Arthritis & Rheumatism* **31,** 315–324 (1988).

78. Radner, H., Neogi, T., Smolen, J. S. & Aletaha, D. Performance of the 2010 ACR/EULAR classification criteria for rheumatoid arthritis: A systematic literature review. *Ann. Rheum. Dis.* **73,** 114–123 (2014).

79. Wasserman, A. M. Diagnosis and management of rheumatoid arthritis. *Am. Fam. Physician* **84,** 1245–1252 (2011).

80. Brown, P. M., Pratt, A. G. & Isaacs, J. D. Mechanism of action of methotrexate in rheumatoid arthritis, and the search for biomarkers. *Nat. Rev. Rheumatol.* **12,** 731–742 (2016).

81. O'Dell, J. R. Anticytokine Therapy - A New Era in the Treatment of Rheumatoid Arthritis? *N. Engl. J. Med.* **340,** 310–312 (1999).

82. McInnes, I. B. & Schett, G. Cytokines in the pathogenesis of rheumatoid arthritis. *Nat. Rev. Immunol.* **7,** 429–442 (2007).

83. McInnes, I. B. & Schett, G. Pathogenetic insights from the treatment of rheumatoid arthritis. *Lancet* **389,** 2328–2337 (2017).

84. Burmester, G. R., Blanco, R., Charles-Schoeman, C., Wollenhaupt, J., Zerbini, C., Benda, B., Gruben, D., Wallenstein, G., Krishnaswami, S., Zwillich, S. H., Koncz, T., Soma, K., Bradley, J. & Mebus, C. Tofacitinib (CP-690,550) in combination with methotrexate in patients with active rheumatoid arthritis with an inadequate response to tumour necrosis factor inhibitors: A randomised phase 3 trial. *Lancet* **381,** 451–460 (2013).

85. Genovese, M. C., Kremer, J., Zamani, O., Ludivico, C., Krogulec, M., Xie, L., Beattie, S. D., Koch, A. E., Cardillo, T. E., Rooney, T. P., Macias, W. L., Bono, S. De, Schlichting, D. E. & Smolen, J. S. Baricitinib in patients with refractory rheumatoid arthritis. *N. Engl. J. Med.* **374,** 1243–1252 (2016).

86. Taylor, P. C., Keystone, E. C., van der Heijde, D., Weinblatt, M. E., del Carmen Morales, L., Reyes Gonzaga, J., Yakushin, S., Ishii, T., Emoto, K., Beattie, S., Arora, V., Gaich, C., Rooney, T., Schlichting, D., Macias, W. L., de Bono, S. & Tanaka, Y. Baricitinib versus Placebo or Adalimumab in

Rheumatoid Arthritis. *N. Engl. J. Med.* **376,** 652–662 (2017).

87. van Vollenhoven, R. F., Fleischmann, R., Cohen, S., Lee, E. B., García Meijide, J. A., Wagner, S., Forejtova, S., Zwillich, S. H., Gruben, D., Koncz, T., Wallenstein, G. V., Krishnaswami, S., Bradley, J. D. & Wilkinson, B. Tofacitinib or Adalimumab versus Placebo in Rheumatoid Arthritis. *N. Engl. J. Med.* **367,** 508–519 (2012).

88. Singh, J. A., Saag, K. G., Bridges, S. L., Akl, E. A., Bannuru, R. R., Sullivan, M. C., Vaysbrot, E., McNaughton, C., Osani, M., Shmerling, R. H., Curtis, J. R., Furst, D. E., Parks, D., Kavanaugh, A., O'Dell, J., King, C., Leong, A., Matteson, E. L., Schousboe, J. T., *et al.* 2015 American College of Rheumatology Guideline for the Treatment of Rheumatoid Arthritis. *Arthritis Care Res. (Hoboken).* **68,** 1–25 (2016).

89. Smolen, J. S., Landewé, R., Bijlsma, J., Burmester, G., Chatzidionysiou, K., Dougados, M., Nam, J., Ramiro, S., Voshaar, M., Van Vollenhoven, R., Aletaha, D., Aringer, M., Boers, M., Buckley, C. D., Buttgereit, F., Bykerk, V., Cardiel, M., Combe, B., Cutolo, M., *et al.* EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs: 2016 update. *Ann. Rheum. Dis.* **76,** 960–977 (2017).

90. Finnish Medical Society Duodecim. Rheumatoid arthritis: current care guidelines. (2017). Available at: http://www.kaypahoito.fi/web/english/guidelines/guideline?id=ccg00005#suositus. (Accessed: 19th July 2018)

91. Rantalaiho, V., Puolakka, K., Korpela, M., Hannonen, P. & Möttönen, T. Long-term results of the FIN-RACo trial; treatment with a combination of traditional disease-modifying anti-rheumatic drugs is an excellent option in early rheumatoid arthritis. *Clin. Exp. Rheumatol.* **30,** (2012).

92. Baecklund, E., Iliadou, A., Askling, J., Ekbom, A., Backlin, C., Granath, F., Catrina, A. I., Rosenquist, R., Feltelius, N., Sundström, C. & Klareskog, L. Association of chronic inflammation, not its treatment, with increased lymphoma risk in rheumatoid arthritis. *Arthritis Rheum.* **54,** 692–701 (2006).

93. Lauper, K. & Gabay, C. Cardiovascular risk in patients with rheumatoid arthritis. *Semin. Immunopathol.* **39,** 447–459 (2017).

94. Listing, J., Kekow, J., Manger, B., Burmester, G. R., Pattloch, D., Zink, A. & Strangfeld, A. Mortality in rheumatoid arthritis: The impact of disease activity, treatment with glucocorticoids, TNFα inhibitors and rituximab. *Ann. Rheum. Dis.* **74,** 415–421 (2015).

95. Choi, H. K., Hernán, M. A., Seeger, J. D., Robins, J. M. & Wolfe, F. Methotrexate and mortality in patients with rheumatoid arthritis: a prospective study. *Lancet* **359,** 1173–77 (2002).

96. Puolakka, K., Kautiainen, H., Möttönen, T., Hannonen, P., Korpela, M., Hakala, M., Järvinen, P., Ahonen, J., Forsberg, S. & Leirisalo-Repo, M. Early suppression of disease activity is essential for maintenance of work capacity in patients with recent-onset rheumatoid arthritis: Five-year experience from the FIN-RACo trial. *Arthritis Rheum.* **52,** 36–41 (2005).

97. Dadoun, S., Zeboulon-Ktorza, N., Combescure, C., Elhai, M., Rozenberg, S., Gossec, L. & Fautrel, B. Mortality in rheumatoid arthritis over the last fifty years: Systematic review and meta-analysis. *Jt. Bone Spine* **80,** 29–33 (2013).

98. Wilton, K. M. & Matteson, E. L. Malignancy Incidence, Management, and Prevention in Patients with Rheumatoid Arthritis. *Rheumatol. Ther.* **4,** 333–347 (2017).

99. Klein, A., Polliack, A. & Gafter-Gvili, A. Rheumatoid arthritis and lymphoma: Incidence, pathogenesis, biology, and outcome. *Hematol. Oncol.* 1–7 (2018). doi:10.1002/hon.2525

100. Baecklund, E., Sundström, C., Ekbom, A., Catrina, A. I., Biberfeld, P., Feltelius, N. & Klareskog, L. Lymphoma subtypes in patients with rheumatoid arthritis: Increased proportion of diffuse large B cell lymphoma. *Arthritis Rheum.* **48,** 1543–1550 (2003).

101. Ramadan, S. M., Fouad, T. M., Summa, V., Hasan, S. K. H. & Lo-Coco, F. Acute myeloid leukemia developing in patients with autoimmune diseases. *Haematologica* **97,** 805–817 (2012).

102. Nakad, R. & Schumacher, B. DNA damage response and immune defense: Links and mechanisms. *Front. Genet.* **7,** 147 (2016).

103. Ertz-Archambault, N., Kosiorek, H., Taylor, G. E., Kelemen, K., Dueck, A., Castro, J., Marino, R., Gauthier, S., Finn, L., Sproat, L. Z., Palmer, J., Mesa, R. A., Al-Kali, A., Foran, J. & Tibes, R. Association of therapy for autoimmune disease with myelodysplastic syndromes and acute myeloid leukemia. *JAMA Oncol.* **3,** 936–943 (2017).

104. Bockorny, B. & Dasanu, C. A. Autoimmune manifestations in large granular lymphocyte leukemia. *Clin. Lymphoma, Myeloma Leuk.* **12,** 400–405 (2012).

105. Bareau, B., Rey, J., Hamidou, M., Donadieu, J., Morcet, J., Reman, O., Schleinitz, N., Tournilhac, O., Roussel, M., Fest, T. & Lamy, T. Analysis of a French cohort of patients with large granular lymphocyte leukemia: A report on 229 cases. *Haematologica* **95,** 1534–1541 (2010).

106. Sokol, L. & Loughran, T. P. Large Granular Lymphocyte Leukemia. 263–273 (2006). doi:10.1634/theoncologist.11-3-263

107. Sanikommu, S. R., Clemente, M. J., Chomczynski, P., Afable, M. G., Jerez, A., Thota, S., Patel, B., Hirsch, C., Nazha, A., Desamito, J., Lichtin, A., Pohlman, B., Sekeres, M. A., Radivoyevitch, T. & Maciejewski, J. P. Clinical features and treatment outcomes in large granular lymphocytic leukemia (LGLL). *Leuk. Lymphoma* **59,** 416–422 (2018).

108. Lamy, T., Moignet, A. & Loughran, T. P. LGL leukemia: From pathogenesis to treatment. *Blood* **129,** 1082–1094 (2017).

109. Koskela, H. L. M., Eldfors, S., Ellonen, P., Van Adrichem, A., Kuusanmäki, H., Andersson, E. I., Lagström, S., Clemente, M. J., Olson, T., Jalkanen, S. E., Majumder, M. M., Almusa, H., Edgren, H., Lepistö, M., Mattila, P., Guinta, K., Koistinen, P., Kuittinen, T., Penttinen, K., *et al.* Somatic STAT3 Mutations in Large Granular Lymphocytic Leukemia. *N. Engl. J. Med.* **366,** 1905–1913 (2012).

110. Jerez, A., Clemente, M. J., Makishima, H., Koskela, H., Leblanc, F., Ng, K. P., Olson, T., Przychodzen, B., Afable, M., Gomez-Segui, I., Guinta, K., Durkin, L., Hsi, E. D., McGraw, K., Zhang, D., Wlodarski, M. W., Porkka, K., Sekeres, M. a, List, A., *et al.* STAT3 mutations unify the pathogenesis of chronic lymphoproliferative disorders of NK cells and T cell large granular lymphocyte leukemia. *Blood* **120,** 3048–3058 (2012).

111. Andersson, E., Kuusanmäki, H., Bortoluzzi, S., Lagström, S., Parsons, A., Rajala, H., Van Adrichem, A., Eldfors, S., Olson, T., Clemente, M. J., Laasonen, A., Ellonen, P., Heckman, C., Loughran, T. P., Maciejewski, J. P. & Mustjoki, S. Activating somatic mutations outside the SH2-domain of STAT3 in LGL leukemia. *Leukemia* **30,** 1204–1208 (2016).

112. Rajala, H. L. M., Porkka, K., Maciejewski, J. P., Loughran, T. P. & Mustjoki, S. Uncovering the pathogenesis of large granular lymphocytic leukemia—novel *STAT3* and *STAT5b* mutations. *Ann. Med.* **46,** 114–122 (2014).

113. Rajala, H. L. M., Olson, T., Clemente, M. J., Lagström, S., Ellonen, P., Lundan, T., Hamm, D. E., Arshi Uz Zaman, S., Lopez Marti, J. M., Andersson, E. I., Jerez, A., Porkka, K., Maciejewski, J. P., Loughran, T. P. & Mustjoki, S. The analysis of clonal diversity and therapy responses using STAT3 mutations as a molecular marker in large granular lymphocytic leukemia. *Haematologica* **100,** 91–99 (2015).

114. Rajala, H. L. M., Eldfors, S., Kuusanmäki, H., Van Adrichem, A., Olson, T., Lagström, S., Andersson, E. I., Jerez, A., Clemente, M. J., Yiyi, Y., Zhang, D., Awwad, A., Ellonen, P., Kallioniemi, O., Wennerberg, K., Porkka, K., Maciejewski, J. P., Loughran, T. P., Heckman, C., *et al.* Discovery of somatic STAT5b mutations in large granular lymphocytic leukemia. *Blood* **121,** 4541–4551 (2014).

115. Andersson, E. I., Tanahashi, T., Sekiguchi, N., Gasparini, V. R., Bortoluzzi, S., Kawakami, T., Matsuda, K., Mitsui, T., Eldfors, S., Bortoluzzi, S. & Coppe, A. High incidence of activating STAT5B mutations in CD4-positive T-cell large granular lymphocyte leukemia. *Blood* **128,** 2465–2468 (2016).

116. Akira, S. Roles of STAT3 defined by tissue-specific gene targeting. *Oncogene* **19,** 2607–2611 (2000).

117. Khan, N., Shariff, N., Cobbold, M., Bruton, R., Ainsworth, J. A., Sinclair, A. J., Nayak, L. & Moss, P. A. H. Cytomegalovirus seropositivity drives the CD8 T cell repertoire toward greater clonality in healthy elderly individuals. *J. Immunol.* **169,** 1984–1992 (2002).

118. Schwaneck, E. C., Renner, R., Junker, L., Einsele, H., Gadeholt, O., Geissinger, E., Kleinert, S., Gernert, M., Tony, H.-P. & Schmalzing, M. Prevalence and Characteristics of Persistent Clonal T-LGL Expansions in Rheumatoid Arthritis: a Comprehensive Analysis of 529 Patients. *Arthritis Rheumatol. (Hoboken, N.J.)* **79,** 1914–1922 (2018).

119. Balint, G. P. & Balint, P. V. Felty's syndrome. *Best Pract. Res. Clin. Rheumatol.* **18,** 631–645 (2004).

120. Bowman, S. J., Sivakumaran, M., Snowdendegs, N., Bhavnani, M., Hall, M. A., Panayi, G. S. & Lanchbury, J. S. The large granular lymphocyte

syndrome with rheumatoid arthritis. Immunogenetic evidence for a broader definition of Felty's syndrome. *Arthritis Rheum.* **37,** 1326–1330 (1994).

121. Campion, G., Maddison, P. J., Goulding, N., James, I., Ahern, M. J., Watt, I. & Sansom, D. The Felty Syndrome: A Case-Matched Study of Clinical Manifestations and Outcome, Serologic Features, and Immunogenetic Associations. *Med.* **69,** 69–80 (1990).

122. Saway, P. A., Prasthofer, E. F. & Barton, J. C. Prevalence of granular lymphocyte proliferation in patients with rheumatoid arthritis and neutropenia. *Am. J. Med.* **86,** 303–307 (1989).

123. Sienknecht, C. W., Urowitz, M. B., Pruzanski, W. & Stein, H. B. Felty's syndrome. Clinical and serological analysis of 34 cases. *Ann. Rheum. Dis.* **36,** 500–507 (1977).

124. Liu, X. & Loughran, T. P. The spectrum of large granular lymphocyte leukemia and Felty's syndrome. *Curr. Opin. Hematol.* **18,** 254–259 (2011).

125. Bowman, S. J., Bhavnani, M., Geddes, G. C., Corrigall, V., Boylston, a W., Panayi, G. S. & Lanchbury, J. S. Large granular lymphocyte expansions in patients with Felty's syndrome: analysis using anti-T cell receptor V beta-specific monoclonal antibodies. *Clin. Exp. Immunol.* **101,** 18–24 (1995).

126. Starkebaum, G., Loughran, T. P., Gaur, L. K., Davis, P. & Nepom, B. S. Immunogenetic similarities between patients with Felty's syndrome and those with clonal expansions of large granular lymphocytes in rheumatoid arthritis. *Arthritis Rheum.* **40,** 624–626 (1997).

127. Felty, A. Chronic arthritis in the adult associated with splenomegaly and leucopenia. *Bull. Johns Hopkins Hosp.* **35,** 16–20 (1924).

128. Rashba, E. J., Rowe, J. M. & Packman, C. H. Treatment of the neutropenia of Felty syndrome. *Blood Rev.* **10,** 177–184 (1996).

129. Dillon, A. M., Luthra, H. S., Conn, D. L. & Ferguson, R. H. Parenteral gold therapy in the Felty syndrome. Experience with 20 patients. *Medicine (Baltimore).* **65,** 107–112 (1986).

130. Wassenberg, S., Herborn, G. & Rau, R. Methotrexate treatment in Felty's syndrome. *Br. J. Rheumatol.* **37,** 908–911 (1998).

131. Narváez, J., Domingo-Domenech, E., Gómez-Vaquero, C., López-Vives, L., Estrada, P., Aparicio, M., Martín-Esteve, I. & Nolla, J. M. Biological Agents in the Management of Felty's Syndrome: A Systematic Review. *Semin. Arthritis Rheum.* **41,** 658–668 (2012).

132. Owlia, M. B., Newman, K. & Akhtari, M. Felty's syndrome, insights and updates. *Open Rheumatol. J.* **8,** 129–136 (2014).

133. Sibley, J. T., Haga, M., Visram, D. A. & Mitchell, D. M. The clinical course of Felty's syndrome compared to matched controls. *J. Rheumatol.* **18,** 1163–1167 (1991).

134. Thorne, C. & Urowitz, M. B. Long-term outcome in Felty ' s syndrome. *Ann. Rheum. Dis.* **41,** 486–489 (1982).

135. Gridley, G., Klippel, J. H., Hoover, R. N. & Fraumeni, J. F. Incidence of

cancer among men with the Felty syndrome. *Ann. Intern. Med.* **120,** 35–39 (1994).

136. Jr, X. L. and T. P. L. The spectrum of LGL and Felty's Syndrome. *Curr Opin Hematol* **18,** 254–259 (2011).

137. Abbott, J. K. & Gelfand, E. W. Common Variable Immunodeficiency. Diagnosis, Management, and Treatment. *Immunol. Allergy Clin. North Am.* **35,** 637–658 (2015).

138. CEREDIH. The French national registry of primary immunodeficiency diseases. *Clin. Immunol.* **135,** 264–272 (2010).

139. Edgar, J. D. M., Buckland, M., Guzman, D., Conlon, N. P., Knerr, V., Bangs, C., Reiser, V., Panahloo, Z., Workman, S., Slatter, M., Gennery, A. R., Davies, E. G., Allwood, Z., Arkwright, P. D., Helbert, M., Longhurst, H. J., Grigoriadou, S., Devlin, L. A., Huissoon, A., *et al.* The United Kingdom Primary Immune Deficiency (UKPID) Registry: report of the first 4 years' activity 2008-2012. *Clin. Exp. Immunol.* **175,** 68–78 (2014).

140. Ludviksson, B. R., Sigurdardottir, S. T., Johannsson, J. H., Haraldsson, A. & Hardarson, T. O. Epidemiology of Primary Immunodeficiency in Iceland. *J. Clin. Immunol.* **35,** 75–79 (2015).

141. Marschall, K., Hoernes, M., Bitzenhofer-Grüber, M., Jandus, P., Duppenthaler, A., Wuillemin, W. A., Rischewski, J., Boyman, O., Heininger, U., Hauser, T., Steiner, U., Posfay-Barbe, K., Seebach, J., Recher, M., Hess, C., Helbling, A., Reichenbach, J., Angelini, F., Bonhoeffer, J., *et al.* The Swiss National Registry for Primary Immunodeficiencies: report on the first 6 years' activity from 2008 to 2014. *Clin. Exp. Immunol.* **182,** 45–50 (2015).

142. Westh, L., Mogensen, T. H., Dalgaard, L. S., Bernth Jensen, J. M., Katzenstein, T., Hansen, A. B. E., Larsen, O. D., Terpling, S., Nielsen, T. L. & Larsen, C. S. Identification and Characterization of a Nationwide Danish Adult Common Variable Immunodeficiency Cohort. *Scand. J. Immunol.* **85,** 450–461 (2017).

143. Matamoros Flori, N., Mila Llambi, J., Espanol Boren, T., Raga Borja, S. & Fontan Casariego, G. Primary immunodeficiency syndrome in Spain: first report of the National Registry in Children and Adults. *J. Clin. Immunol.* **17,** 333–339 (1997).

144. Stray-Pedersen, A., Abrahamsen, T. G. & Froland, S. S. Primary immunodeficiency diseases in Norway. *J. Clin. Immunol.* **20,** 477–485 (2000).

145. Gathmann, B., Mahlaoui, N., Gerard, L., Oksenhandler, E., Warnatz, K., Schulse, I. & Kindle, G. Clinical picture and treatment of 2212 patients with common variable immunodeficienc. *J. Allergy Clin. Immunol.* **134,** 116–126 (2014).

146. Cunningham-Rundles, C. & Bodian, C. Common variable immunodeficiency: Clinical and immunological features of 248 patients. *Clin. Immunol.* **92,** 34–48 (1999).

147. Quinti, I., Soresina, A., Spadaro, G., Martino, S., Donnanno, S., Agostini, C., Claudio, P., Franco, D., Pesce, A. M., Borghese, F., Guerra, A.,

Rondelli, R., Plebani, A., De Mattia, D., Martire, B., Cossu, F., Schirilló, G., Castagnola, E., Pietrogrande, M. C., *et al.* Long-term follow-up and outcome of a large cohort of patients with common variable immunodeficiency. *J. Clin. Immunol.* **27,** 308–316 (2007).

148. Chapel, H., Lucas, M., Lee, M., Bjorkander, J., Webster, D., Grimbacher, B., Fieschi, C., Thon, V., Abedi, M. R. & Hammarstrom, L. Common variable immunodeficiency disorders : division into distinct clinical phenotypes. *Blood* **112,** 277–286 (2008).

149. Resnick, E. S., Moshier, E. L., Godbold, J. H. & Cunningham-Rundles, C. Morbidity and mortality in common variable immune deficiency over 4 decades. *Blood* **119,** 1650–1657 (2012).

150. Oksenhendler, E., Gérard, L., Fieschi, C., Malphettes, M., Mouillot, G., Jaussaud, R., Viallard, J., Gardembas, M., Galicier, L., Schleinitz, N., Suarez, F., Soulas-Sprauel, P., Hachulla, E., Jaccard, A., Gardeur, A., Théodorou, I., Rabian, C. & Debré, P. Infections in 252 Patients with Common Variable Immunodeficiency. *Clin. Infect. Dis.* **46,** 1547–1554 (2008).

151. Aghamohammadi, A., Abolhassani, H., Latif, A. H., Tabassomi, F., Shokuhfar, T., Torabi Sagvand, B., Shahinpour, S., Mirminachi, B., Parvaneh, N., Movahedi, M., Gharagozlou, M., Sherkat, R., Amin, R., Aleyasin, S., Faridhosseini, R., Jabbari-Azad, F., Cheraghi, T., Eslamian, M. H., Khalili, A., *et al.* Long-term evaluation of a historical cohort of Iranian common variable immunodeficiency patients. *Expert Rev. Clin. Immunol.* **10,** 1405–1417 (2014).

152. Ramírez-Vargas, N., Arablin-Oropeza, S. E., Mojica-Martínez, D., Yamazaki-Nakashimada, M. A., De la Luz García-Cruz, M., Terán-Juárez, L. M., Cortés-Grimaldo, R. M., Torres-Lozano, C., Madrigal-Beas, I., Ortega-Cisneros, M., Vargas-Camaño, M. E., Staines-Boone, T., Pietropaolo-Cienfuegos, D., Berrón-Ruiz, L., Espinosa-Rosales, F. J., Guevara-Cruz, M. & Blancas-Galicia, L. Clinical and immunological features of common variable immunodeficiency in Mexican patients. *Allergol Immunopathol (Madr)* **42,** 235–240 (2014).

153. Salzer, U., Chapel, H. M., Webster, A. D. B., Pan-Hammarström, Q., Schmitt-Graeff, A., Schlesier, M., Peter, H. H., Rockstroh, J. K., Schneider, P., Schäffer, A. A., Hammarström, L. & Grimbacher, B. Mutations in TNFRSF13B encoding TACI are associated with common variable immunodeficiency in humans. *Nat. Genet.* **37,** 820–828 (2005).

154. Zhang, L., Radigan, L., Salzer, U., Behrens, T. W., Grimbacher, B., Diaz, G., Bussel, J. & Cunningham-Rundles, C. Transmembrane activator and calcium-modulating cyclophilin ligand interactor mutations in common variable immunodeficiency: Clinical and immunologic outcomes in heterozygotes. *J. Allergy Clin. Immunol.* **120,** 1178–1185 (2007).

155. Castigli, E., Wilson, S. A., Garibyan, L., Rachid, R., Bonilla, F., Schneider, L. & Geha, R. S. TACI is mutant in common variable immunodeficiency and IgA deficiency. *Nat. Genet.* **37,** 829–834 (2005).

156. Pan-Hammarström, Q., Salzer, U., Du, L., Björkander, J., Cunningham-Rundles, C., Nelson, D. L., Bacchelli, C., Gaspar, H. B., Offer, S., Behrens,

T. W., Grimbacher, B. & Hammarström, L. Reexamining the role of TACI coding variants in common variable immunodeficiency and selective IgA To the Editor : Reexamining the role of TACI coding variants in common variable immunodeficiency and selective IgA deficiency To the Editor : *Nat. Genet.* **39,** 429–430 (2007).

157. Sekine, H., Ferreira, R. C., Pan-Hammarstrom, Q., Graham, R. R., Ziemba, B., de Vries, S. S., Liu, J., Hippen, K., Koeuth, T., Ortmann, W., Iwahori, A., Elliott, M. K., Offer, S., Skon, C., Du, L., Novitzke, J., Lee, A. T., Zhao, N., Tompkins, J. D., *et al.* Role for Msh5 in the regulation of Ig class switch recombination. *Proc. Natl. Acad. Sci.* **104,** 7193–7198 (2007).

158. Offer, S. M., Pan-Hammarström, Q., Hammarström, L. & Harris, R. S. Unique DNA repair gene variations and potential associations with the primary antibody deficiency syndromes igAD and CVID. *PLoS One* **5,** e12260 (2010).

159. Losi, C. G., Silini, A., Fiorini, C., Soresina, A., Meini, A., Ferrari, S., Notarangelo, L. D., Lougaris, V. & Plebani, A. Mutational analysis of human BAFF receptor TNFRSF13C (BAFF-R) in patients with common variable immunodeficiency. *J. Clin. Immunol.* **25,** 496–502 (2005).

160. Pieper, K., Rizzi, M., Speletas, M., Smulski, C. R., Sic, H., Kraus, H., Salzer, U., Fiala, G. J., Schamel, W. W., Lougaris, V., Plebani, A., Hammarstrom, L., Recher, M., Germenis, A. E., Grimbacher, B., Warnatz, K., Rolink, A. G., Schneider, P., Notarangelo, L. D., *et al.* A common single nucleotide polymorphism impairs B-cell activating factor receptor's multimerization, contributing to common variable immunodeficiency. *J. Allergy Clin. Immunol.* **133,** 1222–1225 (2014).

161. Warnatz, K., Salzer, U., Rizzi, M., Fischer, B., Gutenberger, S., Böhm, J., Kienzler, A.-K., Pan-Hammarström, Q., Hammarström, L., Rakhmanov, M., Schlesier, M., Grimbacher, B., Peter, H.-H. & Eibel, H. B-cell activating factor receptor deficiency is associated with an adult-onset antibody deficiency syndrome in humans. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 13945–50 (2009).

162. Orange, J. S., Glessner, J. T., Resnick, E., Sullivan, K. E., Lucas, M., Ferry, B., Kim, C. E., Hou, C., Wang, F., Chiavacci, R., Kugathasan, S., Sleasman, J. W., Baldassano, R., Perez, E. E., Chapel, H., Cunningham-Rundles, C. & Hakonarson, H. Genome-wide association identifies diverse causes of common variable immunodeficiency. *J. Allergy Clin. Immunol.* **127,** 1360–1367.e6 (2011).

163. Li, J., Jørgensen, S. F., Maggadottir, S. M., Bakay, M., Warnatz, K., Glessner, J., Pandey, R., Salzer, U., Schmidt, R. E., Perez, E., Resnick, E., Goldacker, S., Buchta, M., Witte, T., Padyukov, L., Videm, V., Folseraas, T., Atschekzei, F., Elder, J. T., *et al.* Association of CLEC16A with human common variable immunodeficiency disorder and role in murine B cells. *Nat. Commun.* **6,** (2015).

164. Maggadottir, S. M., Li, J., Glessner, J. T., Li, Y. R., Wei, Z., Chang, X., Mentch, F. D., Thomas, K. A., Kim, C. E., Zhao, Y., Hou, C., Wang, F., Jørgensen, S. F., Perez, E. E., Sullivan, K. E., Orange, J. S., Karlsen, T. H.,

Chapel, H., Cunningham-Rundles, C., *et al.* Rare variants at 16p11.2 are associated with common variable immunodeficiency. *J. Allergy Clin. Immunol.* **135,** 1569–1577 (2015).

165. Bates, C. A., Ellison, M. C., Lynch, D. A., Cool, C. D., Brown, K. K. & Routes, J. M. Granulomatous-lymphocytic lung disease shortens survival in common variable immunodeficiency. *J. Allergy Clin. Immunol.* **114,** 415–421 (2004).

166. Prasse, A., Kayser, G. & Warnatz, K. Common variable immunodeficiency-associated granulomatous and interstitial lung disease. *Curr. Opin. Pulm. Med.* **19,** 503–509 (2013).

167. Holm, A. M., Tjonnfjord, G., Yndestad, A., Beiske, K., Müller, F., Aukrust, P. & Froland, S. S. Polyclonal expansion of large granular lymphocytes in common variable immunodeficiency - Association with neutropenia. *Clin. Exp. Immunol.* **144,** 418–424 (2006).

168. Mellemkjær, L., Hammarström, L., Andersen, V., Yuen, J., Heilmann, C. & Barington, T. Cancer risk among patients with IgA deficiency or common variable immunodeficiency and their relatives : a combined Danish and Swedish study. *Clin. Exp. Immunol.* **130,** 495–500 (2002).

169. Vajdic, C. M., Mao, L., Van Leeuwen, M. T., Kirkpatrick, P., Grulich, A. E. & Riminton, S. Are antibody deficiency disorders associated with a narrower range of cancers than other forms of immunodeficiency? *Blood* **116,** 1228–1234 (2010).

170. Quinti, I., Agostini, C., Tabolli, S., Brunetti, G., Cinetto, F., Pecoraro, A. & Spadaro, G. Malignancies are the major cause of death in patients with adult onset common variable immunodeficiency. *Blood* **120,** 1953–1954 (2012).

171. Ochtrop, M. L. G., Goldacker, S., May, A. M., Rizzi, M., Draeger, R., Hauschke, D., Stehfest, C., Warnatz, K., Goebel, H., Technau-ihling, K., Werner, M., Salzer, U., Eibel, H., Schlesier, M. & Peter, H. H. T and B lymphocyte abnormalities in bone marrow biopsies of common variable immunodeficiency. *Blood* **118,** 309–319 (2011).

172. Wehr, C., Kivioja, T., Schmitt, C., Ferry, B., Witte, T., Eren, E., Hernandez, M., Detkova, D., Bos, P. R., Poerksen, G., Bernuth, H. Von, Baumann, U., Goldacker, S., Gutenberger, S., Schlesier, M., Bergeron-van, F., Cruyssen, D., Garff, M. Le, Debré, P., *et al.* The EUROclass trial : defining subgroups in common variable immunodeficiency. *Blood* **111,** 77–85 (2008).

173. Warnatz, K., Denz, A., Dra, R., Braun, M., Groth, C., Wolff-vorbeck, G., Eibel, H., Schlesier, M. & Peter, H. H. Severe deficiency of switched memory B cells (CD27+1 IgM- IgD-) in subgroups of patients with common variable immunodeficiency : a new approach to classify a heterogeneous disease. *Blood* **99,** 1544–1552 (2002).

174. Piqueras, B., Lavenu-Bombled, C., Galicier, L., Bergeron-van der Cruyssen, F., Mouthon, L., Chevret, S., Debre, P., Schmitt, C. & Oksenhendler, E. Common variable immunodeficiency patient classification based on impaired B cell memory differentiation correlates with clinical aspects. *J. Clin. Immunol.* **23,** 385–400 (2003).

175. Wong, G. K., Millar, D., Penny, S., Heather, J. M., Mistry, P., Buettner, N., Bryon, J., Huissoon, A. P. & Cobbold, M. Accelerated Loss of TCR Repertoire Diversity in Common Variable Immunodeficiency. *J. Immunol.* **197,** 1642–9 (2016).

176. Melo, K. M., Carvalho, K. I., Bruno, F. R., Ndhlovu, L. C., Ballan, W. M., Nixon, D. F., Kallas, E. G. & Costa-Carvalho, B. T. A decreased frequency of regulatory T cells in patients with common variable immunodeficiency. *PLoS One* **4,** 3–8 (2009).

177. Arumugakani, G., Wood, P. M. D. & Carter, C. R. D. Frequency of treg cells is reduced in CVID patients with autoimmunity and splenomegaly and is associated with expanded CD21lo B lymphocytes. *J. Clin. Immunol.* **30,** 292–300 (2010).

178. Giovannetti, A., Pierdominici, M., Mazzetta, F., Marziali, M., Renzi, C., Mileo, A. M., De Felice, M., Mora, B., Esposito, A., Carello, R., Pizzuti, A., Paggi, M. G., Paganelli, R., Malorni, W. & Aiuti, F. Unravelling the Complexity of T Cell Abnormalities in Common Variable Immunodeficiency. *J. Immunol.* **178,** 3932–3943 (2007).

179. De Vera, M. J., Al-Harthi, L. & Gewurz, A. T. Assessing thymopoiesis in patients with common variable immunodeficiency as measured by T-cell receptor excision circles. *Ann. Allergy, Asthma Immunol.* **93,** 478–484 (2004).

180. Ramesh, M., Hamm, D., Simchoni, N. & Cunningham-Rundles, C. Clonal and constricted T cell repertoire in Common Variable Immune Deficiency. *Clin. Immunol.* **178,** 1–9 (2017).

181. Moratto, D., Gulino, A. V., Fontana, S., Mori, L., Pirovano, S., Soresina, A., Meini, A., Imberti, L., Notarangelo, L. D., Plebani, A. & Badolato, R. Combined decrease of defined B and T cell subsets in a group of common variable immunodeficiency patients. *Clin. Immunol.* **121,** 203–214 (2006).

182. Tam, J. S. & Routes, J. M. Common variable immunodeficiency. *Am. J. Rhinol. Allergy* **27,** 260–265 (2013).

183. Michel, M., Chanet, V., Galicier, L., Ruivard, M., Levy, Y., Hermine, O., Oksenhendler, E., Schaeffer, A., Bierling, P. & Godeau, B. Autoimmune thrombocytopenic purpura and common variable immunodeficiency: Analysis of 21 cases and review of the literature. *Med.* **83,** 254–263 (2004).

184. Horn, J., Manguiat, A., Berglund, L. J., Knerr, V., Tahami, F., Grimbacher, B. & Fulcher, D. A. Decrease in phenotypic regulatory T cells in subsets of patients with common variable immunodeficiency. *Clin. Exp. Immunol.* **156,** 446–454 (2009).

185. Patuzzo, G., Barbieri, A., Tinazzi, E., Veneri, D., Argentino, G., Moretta, F., Puccetti, A. & Lunardi, C. Autoimmunity and infection in common variable immunodeficiency (CVID). *Autoimmun. Rev.* **15,** 877–882 (2016).

186. Brandt, D. & Gershwin, M. E. Common variable immune deficiency and autoimmunity. *Autoimmun. Rev.* **5,** 465–470 (2006).

187. Van De Ven, A. A. J. M. & Warnatz, K. The autoimmune conundrum in common variable immunodeficiency disorders. *Curr. Opin. Allergy Clin. Immunol.* **15,** 514–524 (2015).

188. Conley, M. E., Notarangelo, L. D. & Etzioni, A. Diagnostic Criteria for Primary Immunodeficiencies. *Clin. Immunol.* **93,** 190–197 (1999).

189. Ameratunga, R., Woon, S. T., Gillis, D., Koopmans, W. & Steele, R. New diagnostic criteria for common variable immune deficiency (CVID), which may assist with decisions to treat with intravenous or subcutaneous immunoglobulin. *Clin. Exp. Immunol.* **174,** 203–211 (2013).

190. European Society for Immunodeficiencies Registry Working Party. Available at: https://esid.org/Education/Diagnostic-Criteria-PID. (Accessed: 10th September 2018)

191. Busse, P. J., Razvi, S. & Cunningham-Rundles, C. Efficacy of intravenous immunoglobulin in the prevention of pneumonia in patients with common variable immunodeficiency. *J. Allergy Clin. Immunol.* **109,** 1001–1004 (2002).

192. Cunningham-Rundles, C., Siegal, F. P. & Smithwick, E. M. Efficacy of intravenous immunoglobulin in primary humoral immunodeficiency disease. *Ann. Intern. Med.* **101,** 435–439 (1984).

193. Behjati, S., Huch, M., Van Boxtel, R., Karthaus, W., Wedge, D. C., Tamuri, A. U., Martincorena, I., Petljak, M., Alexandrov, L. B., Gundem, G., Tarpey, P. S., Roerink, S., Blokker, J., Maddison, M., Mudie, L., Robinson, B., Nik-Zainal, S., Campbell, P., Goldman, N., *et al.* Genome sequencing of normal cells reveals developmental lineages and mutational processes. *Nature* **513,** 422–425 (2014).

194. Bae, T., Tomasini, L., Mariani, J., Zhou, B., Roychowdhury, T., Franjic, D., Pletikos, M., Pattni, R., Chen, B.-J., Venturini, E., Riley-Gillis, B., Sestan, N., Urban, A. E., Abyzov, A. & Vaccarino, F. M. Different mutational rates and mechanisms in human cells at pregastrulation and neurogenesis. *Science (80-.).* **359,** 550–555 (2017).

195. Araten, D. J., Golde, D. W., Zhang, R. H., Thaler, H. T., Gargiulo, L., Notaro, R. & Luzzatto, L. A quantitative measurement of the human somatic mutation rate. *Cancer Res.* **65,** 8111–8117 (2005).

196. Lichtenauer-Kaligis, E. G., Thijssen, J., den Dulk, H., van de Putte, P., Tasseron-de Jong, J. G. & Giphart-Gassler, M. Comparison of spontaneous hprt mutation spectra at the nucleotide sequence level in the endogenous hprt gene and five other genomic positions. *Mutat. Res.* **351,** 147–155 (1996).

197. Werner, B. & Sottoriva, A. Variation of mutational burden in healthy human tissues suggests non-random strand segregation and allows measuring somatic mutation rates. *PLos Comput. Biol.* **14,** e1006233 (2018).

198. Blokzijl, F., De Ligt, J., Jager, M., Sasselli, V., Roerink, S., Sasaki, N., Huch, M., Boymans, S., Kuijk, E., Prins, P., Nijman, I. J., Martincorena, I., Mokry, M., Wiegerinck, C. L., Middendorp, S., Sato, T., Schwank, G.,

Nieuwenhuis, E. E. S., Verstegen, M. M. A., *et al.* Tissue-specific mutation accumulation in human adult stem cells during life. *Nature* **538,** 260–264 (2016).

199. Welch, J. S., Ley, T. J., Link, D. C., Miller, C. A., Larson, D. E., Koboldt, D. C., Wartman, L. D., Lamprecht, T. L., Liu, F., Xia, J., Kandoth, C., Fulton, R. S., McLellan, M. D., Dooling, D. J., Wallis, J. W., Chen, K., Harris, C. C., Schmidt, H. K., Kalicki-Veizer, J. M., *et al.* The origin and evolution of mutations in acute myeloid leukemia. *Cell* **150,** 264–278 (2012).

200. Hoang, M. L., Kinde, I., Tomasetti, C., McMahon, K. W., Rosenquist, T. A., Grollman, A. P., Kinzler, K. W., Vogelstein, B. & Papadopoulos, N. Genome-wide quantification of rare somatic mutations in normal human tissues using massively parallel sequencing. *Proc. Natl. Acad. Sci.* **113,** 9846–9851 (2016).

201. Martincorena, I., Roshan, A., Gerstung, M., Ellis, P., Van Loo, P., McLaren, S., Wedge, D. C., Fullam, A., Alexandrov, L. B., Tubio, J. M., Stebbings, L., Menzies, A., Widaa, S., Stratton, M. R., Jones, P. H. & Campbell, P. J. High burden and pervasive positive selection of somatic mutations in normal human skin. *Science (80-.).* **348,** 880–886 (2015).

202. Franco, I., Johansson, A., Olsson, K., Vrtačnik, P., Lundin, P., Helgadottir, H. T., Larsson, M., Revêchon, G., Bosia, C., Pagnani, A., Provero, P., Gustafsson, T., Fischer, H. & Eriksson, M. Somatic mutagenesis in satellite cells associates with human skeletal muscle aging. *Nat. Commun.* **9,** 800 (2018).

203. Milholland, B., Dong, X., Zhang, L., Hao, X., Suh, Y. & Vijg, J. Differences between germline and somatic mutation rates in humans and mice. *Nat. Commun.* **8,** 15183 (2017).

204. Lodato, M. A., Woodworth, M. B., Lee, S., Evrony, G. D., Mehta, B. K., Karger, A., Lee, S., Chittenden, T. W., D'Gama, A. M., Cai, X., Luquette, L. J., Lee, E., Park, P. J. & Walsh, C. A. Somatic mutation in single human neurons tracks developmental and transcriptional history. *Science (80-.).* **350,** 94–98 (2015).

205. Pfeifer, G. P. Mutagenesis at methylated CpG sequences. *Curr. Top. Microbiol. Immunol.* **301,** 259–281 (2006).

206. Alexandrov, L. B., Nik-Zainal, S., Wedge, D. C., Aparicio, S. A. J. R., Behjati, S., Biankin, A. V., Bignell, G. R., Bolli, N., Borg, A., Børresen-Dale, A. L., Boyault, S., Burkhardt, B., Butler, A. P., Caldas, C., Davies, H. R., Desmedt, C., Eils, R., Eyfjörd, J. E., Foekens, J. A., *et al.* Signatures of mutational processes in human cancer. *Nature* **500,** 415–421 (2013).

207. Alexandrov, L. B., Jones, P. H., Wedge, D. C., Sale, J. E., Campbell, P. J., Nik-Zainal, S. & Stratton, M. R. Clock-like mutational processes in human somatic cells. *Nat. Genet.* **47,** 1402–1407 (2015).

208. Abyzov, A., Tomasini, L., Zhou, B., Vasmatzis, N., Coppola, G., Amenduni, M., Pattni, R., Wilson, M., Gerstein, M., Weissman, S., Urban, A. E. & Vaccarino, F. M. One thousand somatic SNVs per skin fibroblast cell set baseline of mosaic mutational load with patterns that suggest proliferative origin. *Genome Res.* **27,** 512–523 (2017).

209. Arends, C. M., Galan-Sousa, J., Hoyer, K., Chan, W., Jäger, M., Yoshida, K., Seemann, R., Noerenberg, D., Waldhueter, N., Fleischer-Notter, H., Christen, F., Schmitt, C. A., Dörken, B., Pelzer, U., Sinn, M., Zemojtel, T., Ogawa, S., Märdian, S., Schreiber, A., *et al.* Hematopoietic lineage distribution and evolutionary dynamics of clonal hematopoiesis. *Leukemia* **32,** 1908–1919 (2018).

210. Bowman, R. L., Busque, L. & Levine, R. L. Clonal Hematopoiesis and Evolution to Hematopoietic Malignancies. *Cell Stem Cell* **22,** 157–170 (2018).

211. Holzelova, E., Vonarbourg, C., Stolzenberg, M.-C., Arkwright, P. D., Selz, F., Prieur, A.-M., Blanche, S., Bartunkova, J., Vilmer, E., Fischer, A., Le Deist, F. & Rieux-Laucat, F. Autoimmune Lymphoproliferative Syndrome with Somatic *Fas* Mutations. *N. Engl. J. Med.* **351,** 1409–1418 (2004).

212. Rieux-Laucat, F. What's up in the ALPS. *Curr. Opin. Immunol.* **49,** 79–86 (2017).

213. Nikolaev, S. I., Vetiska, S., Bonilla, X., Boudreau, E., Jauhiainen, S., Rezai Jahromi, B., Khyzha, N., DiStefano, P. V., Suutarinen, S., Kiehl, T.-R., Mendes Pereira, V., Herman, A. M., Krings, T., Andrade-Barazarte, H., Tung, T., Valiante, T., Zadeh, G., Tymianski, M., Rauramaa, T., *et al.* Somatic Activating *KRAS* Mutations in Arteriovenous Malformations of the Brain. *N. Engl. J. Med.* **378,** 250–261 (2018).

214. Anglesio, M. S., Papadopoulos, N., Ayhan, A., Nazeran, T. M., Noë, M., Horlings, H. M., Lum, A., Jones, S., Senz, J., Seckin, T., Ho, J., Wu, R.-C., Lac, V., Ogawa, H., Tessier-Cloutier, B., Alhassan, R., Wang, A., Wang, Y., Cohen, J. D., *et al.* Cancer-Associated Mutations in Endometriosis without Cancer. *N. Engl. J. Med.* **376,** 1835–1848 (2017).

215. Ettersperger, J., Montcuquet, N., Malamut, G., Guegan, N., Lopez-Lastra, S., Gayraud, S., Reimann, C., Vidal, E., Cagnard, N., Villarese, P., Andre-Schmutz, I., Gomes Domingues, R., Godinho-Silva, C., Veiga-Fernandes, H., Lhermitte, L., Asnafi, V., Macintyre, E., Cellier, C., Beldjord, K., *et al.* Interleukin-15-Dependent T-Cell-like Innate Intraepithelial Lymphocytes Develop in the Intestine and Transform into Lymphomas in Celiac Disease. *Immunity* **45,** 610–625 (2016).

216. Valori, M., Jansson, L., Kiviharju, A., Ellonen, P., Rajala, H., Awad, S. A., Mustjoki, S. & Tienari, P. J. A novel class of somatic mutations in blood detected preferentially in CD8 + cells. *Clin. Immunol.* **175,** 75–81 (2017).

217. Bang, H., Egerer, K., Gauliard, A., Lüthke, K., Rudolph, P. E., Fredenhagen, G., Berg, W., Feist, E. & Burmester, G. R. Mutation and citrullination modifies vimentin to a novel autoantigen for rheumatoid arthritis. *Arthritis Rheum.* **56,** 2503–2511 (2007).

218. Da Sylva, T. R., Connor, A., Mburu, Y., Keystone, E. & Wu, G. E. Somatic mutations in the mitochondria of rheumatoid arthritis synoviocytes. *Arthritis Res. Ther.* **7,** R844-51 (2005).

219. Cannons, J. L., Karsh, J., Chaim Birnboim, H. & Goldstein, R. hprt-Mutant T cells in the peripheral blood and synovial tissue of patients with rheumatoid arthritis. *Arthritis Rheum.* **41,** 1772–1782 (1998).

220. Firestein, G. S., Echeverri, F., Yeo, M., Zvaifler, N. J. & Green, D. R. Somatic mutations in the p53 tumor suppressor gene in rheumatoid arthritis synovium. *Proc. Natl. Acad. Sci. U. S. A.* **94,** 10895–900 (1997).

221. Kullmann, F., Judex, M., Neudecker, I., Lechner, S., Justen, H. P., Green, D. R., Wessinghage, D., Firestein, G. S., Gay, S., Scholmerich, J. & Muller-Ladner, U. Analysis of the p53 tumor suppressor gene in rheumatoid arthritis synovial fibroblasts. *Arthritis Rheum.* **42,** 1594–1600 (1999).

222. Busque, L., Mio, R., Mattioli, J., Brais, E., Blais, N., Lalonde, Y., Maragh, M. & Gilliland, D. G. Nonrandom X-inactivation patterns in normal females: lyonization ratios vary with age. *Blood* **88,** 59–65 (1996).

223. Forsberg, L. A., Rasi, C., Razzaghian, H. R., Pakalapati, G., Waite, L., Thilbeault, K. S., Ronowicz, A., Wineinger, N. E., Tiwari, H. K., Boomsma, D., Westerman, M. P., Harris, J. R., Lyle, R., Essand, M., Eriksson, F., Assimes, T. L., Iribarren, C., Strachan, E., O'Hanlon, T. P., *et al.* Age-related somatic structural changes in the nuclear genome of human blood cells. *Am. J. Hum. Genet.* **90,** 217–228 (2012).

224. Laurie, C. C., Laurie, C. A., Rice, K., Doheny, K. F., Zelnick, L. R., McHugh, C. P., Ling, H., Hetrick, K. N., Pugh, E. W., Amos, C., Wei, Q., Wang, L. E., Lee, J. E., Barnes, K. C., Hansel, N. N., Mathias, R., Daley, D., Beaty, T. H., Scott, A. F., *et al.* Detectable clonal mosaicism from birth to old age and its relationship to cancer. *Nat. Genet.* **44,** 642–650 (2012).

225. Busque, L., Patel, J. P., Figueroa, M. E., Vasanthakumar, A., Provost, S., Hamilou, Z., Mollica, L., Li, J., Viale, A., Heguy, A., Hassimi, M., Socci, N., Bhatt, P. K., Gonen, M., Mason, C. E., Melnick, A., Godley, L. A., Brennan, C. W., Abdel-Wahab, O., *et al.* Recurrent somatic TET2 mutations in normal elderly individuals with clonal hematopoiesis. *Nat. Genet.* **44,** 1179–81 (2012).

226. Genovese, G., Kähler, A. K., Handsaker, R. E., Lindberg, J., Rose, S. a., Bakhoum, S. F., Chambert, K., Mick, E., Neale, B. M., Fromer, M., Purcell, S. M., Svantesson, O., Landén, M., Höglund, M., Lehmann, S., Gabriel, S. B., Moran, J. L., Lander, E. S., Sullivan, P. F., *et al.* Clonal Hematopoiesis and Blood-Cancer Risk Inferred from Blood DNA Sequence. *N. Engl. J. Med.* **371,** 2477–2487 (2014).

227. Jaiswal, S., Fontanillas, P., Flannick, J., Manning, A., Grauman, P. V, Mar, B. G., Lindsley, R. C., Mermel, C. H., Burtt, N., Chavez, A., Higgins, J. M., Moltchanov, V., Kuo, F. C., Kluk, M. J., Henderson, B., Kinnunen, L., Koistinen, H. a, Ladenvall, C., Getz, G., *et al.* Age-related clonal hematopoiesis associated with adverse outcomes. *N. Engl. J. Med.* **371,** 2488–98 (2014).

228. Acuna-Hidalgo, R., Sengul, H., Steehouwer, M., van de Vorst, M., Vermeulen, S. H., Kiemeney, L. A. L. M., Veltman, J. A., Gilissen, C. & Hoischen, A. Ultra-sensitive Sequencing Identifies High Prevalence of Clonal Hematopoiesis-Associated Mutations throughout Adult Life. *Am. J. Hum. Genet.* **101,** 50–64 (2017).

229. Buscarlet, M., Provost, S., Zada, Y. F., Barhdadi, A., Bourgoin, V., Lépine, G., Mollica, L., Szuber, N., Dubé, M. P. & Busque, L. DNMT3A and TET2

dominate clonal hematopoiesis and demonstrate benign phenotypes and different genetic predispositions. *Blood* **130,** 753–762 (2017).

230. Young, A. L., Challen, G. A., Birmann, B. M. & Druley, T. E. Clonal haematopoiesis harbouring AML-associated mutations is ubiquitous in healthy adults. *Nat. Commun.* **7,** 12484 (2016).

231. Zink, F., Stacey, S. N., Norddahl, G. L., Frigge, M. L., Magnusson, O. T., Jonsdottir, I., Thorgeirsson, T. E., Sigurdsson, A., Gudjonsson, S. A., Gudmonsson, J., Jonasson, J. G., Tryggvadottir, L., Jonsson, T., Helgason, A., Gylfason, A., Sulem, P., Rafnar, T., Thorsteinsdottir, U., Gudbjartsson, D. F., *et al.* Clonal Hematopoiesis , With and Without Candidate Driver Mutations , is Common in the Elderly. *Blood* **130,** 742–753 (2017).

232. Xie, M., Lu, C., Wang, J., McLellan, M. D., Johnson, K. J., Wendl, M. C., McMichael, J. F., Schmidt, H. K., Yellapantula, V., Miller, C. A., Ozenberger, B. A., Welch, J. S., Link, D. C., Walter, M. J., Mardis, E. R., Dipersio, J. F., Chen, F., Wilson, R. K., Ley, T. J., *et al.* Age-related mutations associated with clonal hematopoietic expansion and malignancies. *Nat. Med.* **20,** 1472–8 (2014).

233. Steensma, D. P., Bejar, R., Jaiswal, S., Lindsley, R. C., Sekeres, M. A., Hasserjian, R. P. & Ebert, B. L. Clonal hematopoiesis of indeterminate potential and its distinction from myelodysplastic syndromes. *Blood* **126,** 9–16 (2015).

234. Buscarlet, M., Provost, S., Feroz Zada, Y., Bourgoin, V., Mollica, L., Dube, M.-P. & Busque, L. Lineage restriction analyses in CHIP indicate myeloid bias for TET2 and multipotent stem cell origin for DNMT3A. *Blood* **132,** 277–280 (2018).

235. Jaiswal, S., Natarajan, P., Silver, A. J., Gibson, C. J., Bick, A. G., Shvartz, E., McConkey, M., Gupta, N., Gabriel, S., Ardissino, D., Baber, U., Mehran, R., Fuster, V., Danesh, J., Frossard, P., Saleheen, D., Melander, O., Sukhova, G. K., Neuberg, D., *et al.* Clonal Hematopoiesis and Risk of Atherosclerotic Cardiovascular Disease. *N. Engl. J. Med.* **377,** 111–121 (2017).

236. Abelson, S., Collord, G., Stanley, W., Ng, K., Weissbrod, O., Cohen, N. M., Niemeyer, E., Barda, N., Zuzarte, P. C., Heisler, L., Sundaravadanam, Y., Luben, R., Hayat, S., Wang, T. T., Zhao, Z., Cirlan, I., Pugh, T. J., Soave, D., Ng, K., *et al.* Prediction of acute myeloid leukaemia risk in healthy individuals. *Nature* **559,** 400–404 (2018).

237. Desai, P., Mencia-Trinchant, N., Savenkov, O., Simon, M. S., Cheang, G., Lee, S., Samuel, M., Ritchie, E. K., Guzman, M. L., Ballman, K. V., Roboz, G. J. & Hassane, D. C. Somatic mutations precede acute myeloid leukemia years before diagnosis. *Nat. Med.* **24,** 1015–1023 (2018).

238. Fuster, J. J., MacLauchlan, S., Zuriaga, M. A., Polackal, M. N., Chakraborty, R., Wu, C.-L., Sano, S., Muralidharan, S., Rius, C., Vuong, J., Jacob, S., Muralidhar, V., Robertson, A. A. B., Cooper, M. A., Andrés, V., Hirschi, K. K., Martin, K. K. & Walsh, K. Clonal hematopoiesis associated with Tet2 deficiency accelerates atherosclerosis development in mice. *Science (80-.).* **355,** 842–847 (2017).

239. Shendure, J., Balasubramanian, S., Church, G. M., Gilbert, W., Rogers, J., Schloss, J. A. & Waterston, R. H. DNA sequencing at 40: past, present and future. *Nature* **550,** 345–353 (2017).

240. Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., *et al.* Initial sequencing and analysis of the human genome. *Nature* **409,** 860–921 (2001).

241. Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., Evans, C. A., Holt, R. A., Gocayne, J. D., Amanatides, P., Ballew, R. M., Huson, D. H., Wortman, J. R., Zhang, Q., Kodira, C. D., Zheng, X. H., Chen, L., *et al.* The sequence of the human genome. *Science* **291,** 1304–1351 (2001).

242. International Human Genome Sequencing Consortium, . Finishing the euchromatic sequence of the human genome. *Nature* **431,** 931–945 (2004).

243. Ng, S. B., Turner, E. H., Robertson, P. D., Flygare, S. D., Bigham, A. W., Lee, C., Shaffer, T., Wong, M., Bhattacharjee, A., Eichler, E. E., Bamshad, M., Nickerson, D. A. & Shendure, J. Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* **461,** 272–276 (2009).

244. The Cost of Sequencing a Human Genome. Available at: https://www.genome.gov/sequencingcostsdata/. (Accessed: 10th September 2018)

245. The Cost of Sequencing a Human Genome. Available at: https://www.genome.gov/27565109/the-cost-of-sequencing-a-human-genome/. (Accessed: 17th September 2018)

246. Caspar, S. M., Dubacher, N., Kopps, A. M., Meienberg, J., Henggeler, C. & Matyas, G. Clinical sequencing: From raw data to diagnosis with lifetime value. *Clin. Genet.* **93,** 508–519 (2018).

247. Koboldt, D. C., Steinberg, K. M., Larson, D. E., Wilson, R. K. & Mardis, E. R. The next-generation sequencing revolution and its impact on genomics. *Cell* **155,** 27–38 (2013).

248. Nakagawa, H. & Fujita, M. Whole genome sequencing analysis for cancer genomics and precision medicine. *Cancer Sci.* **109,** 513–522 (2018).

249. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25,** 1754–1760 (2009).

250. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9,** 357–359 (2012).

251. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G. & Durbin, R. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25,** 2078–2079 (2009).

252. Picard.

253. McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M. & DePristo, M. A. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20,**

1297–1303 (2010).

254. Cibulskis, K., Lawrence, M. S., Carter, S. L., Sivachenko, A., Jaffe, D., Sougnez, C., Gabriel, S., Meyerson, M., Lander, E. S. & Getz, G. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nature biotechnology* **31,** 213–219 (2013).

255. Koboldt, D. C., Zhang, Q., Larson, D. E., Shen, D., McLellan, M. D., Lin, L., Miller, C. A., Mardis, E. R., Ding, L. & Wilson, R. K. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res.* **22,** 568–576 (2012).

256. Saunders, C. T., Wong, W. S. W., Swamy, S., Becq, J., Murray, L. J. & Cheetham, R. K. Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. *Bioinformatics* **28,** 1811–1817 (2012).

257. Xu, C. A review of somatic single nucleotide variant calling algorithms for next-generation sequencing data. *Comput. Struct. Biotechnol. J.* **16,** 15–24 (2018).

258. Cai, L., Yuan, W., Zhang, Z., He, L. & Chou, K. C. In-depth comparison of somatic point mutation callers based on different tumor next-generation sequencing depth data. *Sci. Rep.* **6,** 36540 (2016).

259. Krøigård, A. B., Thomassen, M., Lænkholm, A. V., Kruse, T. A. & Larsen, M. J. Evaluation of nine somatic variant callers for detection of somatic mutations in exome and targeted deep sequencing data. *PLoS One* **11,** 1–15 (2016).

260. Xu, H., DiCarlo, J., Satya, R. V., Peng, Q. & Wang, Y. Comparison of somatic mutation calling methods in amplicon and whole exome sequence data. *BMC Genomics* **15,** 1–10 (2014).

261. Sandmann, S., De Graaf, A. O., Karimi, M., Van Der Reijden, B. A., Hellström-Lindberg, E., Jansen, J. H. & Dugas, M. Evaluating Variant Calling Tools for Non-Matched Next-Generation Sequencing Data. *Sci. Rep.* **7,** 43169 (2017).

262. Jennings, L. J., Arcila, M. E., Corless, C., Kamel-Reid, S., Lubin, I. M., Pfeifer, J., Temple-Smolkin, R. L., Voelkerding, K. V. & Nikiforova, M. N. Guidelines for Validation of Next-Generation Sequencing–Based Oncology Panels: A Joint Consensus Recommendation of the Association for Molecular Pathology and College of American Pathologists. *J. Mol. Diagnostics* **19,** 341–365 (2017).

263. Fox, E. J., Reid-Bayliss, K. S., Emond, M. J. & Loeb, L. A. Accuracy of Next Generation Sequencing Platforms. *J. Next Gener. Seq. Appl.* **1,** (2014).

264. Ross, M. G., Russ, C., Costello, M., Hollinger, A., Lennon, N. J., Hegarty, R., Nusbaum, C. & Jaffe, D. B. Characterizing and measuring bias in sequence data. *Genome Biol.* **14,** R51 (2013).

265. Minoche, A. E., Dohm, J. C. & Himmelbauer, H. Evaluation of genomic high-throughput sequencing data generated on Illumina HiSeq and Genome Analyzer systems. *Genome Biol.* **12,** R112 (2011).

266. Costello, M., Pugh, T. J., Fennell, T. J., Stewart, C., Lichtenstein, L.,

Meldrim, J. C., Fostel, J. L., Friedrich, D. C., Perrin, D., Dionne, D., Kim, S., Gabriel, S. B., Lander, E. S., Fisher, S. & Getz, G. Discovery and characterization of artifactual mutations in deep coverage targeted capture sequencing data due to oxidative DNA damage during sample preparation. *Nucleic Acids Res.* **41,** e67 (2013).

267. Chen, L., Liu, P., Evans, T. C. & Ettwiller, L. M. DNA damage is a pervasive cause of sequencing errors, directly confounding variant identification. *Science (80-.).* **355,** 752–756 (2017).

268. Kivioja, T., Vähärautio, A., Karlsson, K., Bonke, M., Enge, M., Linnarsson, S. & Taipale, J. Counting absolute numbers of molecules using unique molecular identifiers. *Nat. Methods* **9,** 72–74 (2012).

269. Phallen, J., Sausen, M., Adleff, V., Leal, A., Hruban, C., White, J., Anagnostou, V., Fiksel, J., Cristiano, S., Papp, E., Speir, S., Reinert, T. & Etc. Direct detection of early-stage cancers using circulating tumor DNA. *Sci. Transl. Med.* **9,** eaan2415 (2017).

270. Janeway, C. J., Travers, P., Walport, M. & Shlomchik, M. J. The Immune System in Health and Disease. 5th edition. T-cell receptor gene rearrangement. (2001). Available at: https://www.ncbi.nlm.nih.gov/books/NBK27145/. (Accessed: 1st August 2018)

271. Arstila, T. P., Casrouge, A., Baron, V., Even, J. & Kourilsky, P. A Direct Estimate of the Human αβ T Cell Receptor Diversity. *Science (80-.).* **286,** 958–961 (1999).

272. Rosati, E., Dowds, C. M., Liaskou, E., Henriksen, E. K. K., Karlsen, T. H. & Franke, A. Overview of methodologies for T-cell receptor repertoire analysis. *BMC Biotechnol.* **17,** 61 (2017).

273. Hou, X.-L., Wang, L., Ding, Y.-L., Xie, Q. & Diao, H.-Y. Current status and recent advances of next generation sequencing techniques in immunological repertoire. *Genes Immun.* **17,** 153–164 (2016).

274. Carlson, C. S., Emerson, R. O., Sherwood, A. M., Desmarais, C., Chung, M. W., Parsons, J. M., Steen, M. S., LaMadrid-Herrmannsfeldt, M. A., Williamson, D. W., Livingston, R. J., Wu, D., Wood, B. L., Rieder, M. J. & Robins, H. Using synthetic templates to design an unbiased multiplex PCR assay. *Nat. Commun.* **4,** 2680 (2013).

275. Shugay, M., Britanova, O. V., Merzlyak, E. M., Turchaninova, M. A., Mamedov, I. Z., Tuganbaev, T. R., Bolotin, D. A., Staroverov, D. B., Putintseva, E. V., Plevova, K., Linnemann, C., Shagin, D., Pospisilova, S., Lukyanov, S., Schumacher, T. N. & Chudakov, D. M. Towards error-free profiling of immune repertoires. *Nat. Methods* **11,** 653–655 (2014).

276. Breuer, K., Foroushani, A. K., Laird, M. R., Chen, C., Sribnaia, A., Lo, R., Winsor, G. L., Hancock, R. E. W., Brinkman, F. S. L. & Lynn, D. J. InnateDB: Systems biology of innate immunity and beyond - Recent updates and continuing curation. *Nucleic Acids Res.* **41,** D1228–D1233 (2013).

277. Lynn, D. J., Winsor, G. L., Chan, C., Richard, N., Laird, M. R., Barsky, A., Gardy, J. L., Roche, F. M., Chan, T. H. W., Shah, N., Lo, R., Naseer, M.,

Que, J., Yau, M., Acab, M., Tulpan, D., Whiteside, M. D., Chikatamarla, A., Mah, B., *et al.* InnateDB: Facilitating systems-level analyses of the mammalian innate immune response. *Mol. Syst. Biol.* **4,** 218 (2008).

278. Lynn, D. J., Chan, C., Naseer, M., Yau, M., Lo, R., Sribnaia, A., Ring, G., Que, J., Wee, K., Winsor, G. L., Laird, M. R., Breuer, K., Foroushani, A. K., Brinkman, F. S. L. & Hancock, R. E. W. Curating the innate immunity interactome. *BMC Syst. Biol.* **4,** 117 (2010).

279. Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., Land, S. J., Ruden, D. M. & Lu, X. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly* **6,** 80–92 (2012).

280. Hubbard, T., Barker, D., Birney, E., Cameron, G., Chen, Y., Clark, L., Cox, T., Cuff, J., Curwen, V., Down, T., Durbin, R., Eyras, E., Gilbert, J., Hammond, M., Huminiecki, L., Kasprzyk, A., Lehvaslaiho, H., Lijnzaad, P., Melsopp, C., *et al.* The Ensembl genome database project. *Nucleic Acids Res.* **30,** 38–41 (2002).

281. Yates, A., Akanni, W., Amode, M. R., Barrell, D., Billis, K., Carvalho-Silva, D., Cummins, C., Clapham, P., Fitzgerald, S., Gil, L., Giron, C. G., Gordon, L., Hourlier, T., Hunt, S. E., Janacek, S. H., Johnson, N., Juettemann, T., Keenan, S., Lavidas, I., *et al.* Ensembl 2016. *Nucleic Acids Res.* **44,** D710-6 (2016).

282. Sherry, S., Ward, M., Kholodov, M., Baker, J., Phan, L., Smigielski, E. & Sirotkin, K. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* **29,** 308–11 (2001).

283. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30,** 2114–2120 (2014).

284. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *Prepr.* (2013). doi:arXiv:1303.3997 [q-bio.GN]

285. Cibulskis, K., McKenna, A., Fennell, T., Banks, E., DePristo, M. & Getz, G. ContEst: estimating cross-contamination of human samples in next-generation sequencing data. *Bioinformatics* **27,** 2601–2602 (2011).

286. Zhao, H., Sun, Z., Wang, J., Huang, H., Kocher, J.-P. & Wang, L. CrossMap: a versatile tool for coordinate conversion between genome assemblies. *Bioinformatics* **30,** 1006–1007 (2014).

287. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38,** e164 (2010).

288. Rosenthal, R., McGranahan, N., Herrero, J., Taylor, B. S. & Swanton, C. deconstructSigs: delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. *Genome Biol.* **17,** 31 (2016).

289. Giguère, S., Marchand, M., Laviolette, F., Drouin, A. & Corbeil, J. Learning a peptide-protein binding affinity predictor with kernel ridge regression. *BMC Bioinformatics* **14,** 82 (2013).

290. Levine, J. H., Simonds, E. F., Bendall, S. C., Davis, K. L., Amir, E. D., Tadmor, M. D., Litvin, O., Fienberg, H. G., Jager, A., Zunder, E. R., Finck,

R., Gedman, A. L., Radtke, I., Downing, J. R., Pe'er, D. & Nolan, G. P. Data-Driven Phenotypic Dissection of AML Reveals Progenitor-like Cells that Correlate with Prognosis. *Cell* **162,** 184–197 (2015).

291. Auton, A., Abecasis, G. R., Altshuler, D. M., Durbin, R. M., Bentley, D. R., Chakravarti, A., Clark, A. G., Donnelly, P., Eichler, E. E., Flicek, P., Gabriel, S. B., Gibbs, R. A., Green, E. D., Hurles, M. E., Knoppers, B. M., Korbel, J. O., Lander, E. S., Lee, C., Lehrach, H., *et al.* A global reference for human genetic variation. *Nature* **526,** 68–74 (2015).

292. McLaren, W., Gil, L., Hunt, S. E., Riat, H. S., Ritchie, G. R. S., Thormann, A., Flicek, P. & Cunningham, F. The Ensembl Variant Effect Predictor. *Genome Biol.* **17,** 1–14 (2016).

293. Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M. & Gingeras, T. R. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29,** 15–21 (2013).

294. Liao, Y., Smyth, G. K. & Shi, W. The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Res.* **41,** e108 (2013).

295. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26,** 139–140 (2010).

296. Tickotsky, N., Sagiv, T., Prilusky, J., Shifrut, E. & Friedman, N. McPAS-TCR: A manually curated catalogue of pathology-associated T cell receptor sequences. *Bioinformatics* **33,** 2924–2929 (2017).

297. Adzhubei, I. A., Schmidt, S., Peshkin, L., Ramensky, V. E., Gerasimova, A., Bork, P., Kondrashov, A. S. & Sunyaev, S. R. A method and server for predicting damaging missense mutations. *Nature methods* **7,** 248–249 (2010).

298. Sim, N.-L., Kumar, P., Hu, J., Henikoff, S., Schneider, G. & Ng, P. C. SIFT web server: predicting effects of amino acid substitutions on proteins. *Nucleic Acids Res.* **40,** W452–W457 (2012).

299. Rönnblom, L. The importance of the type I interferon system in autoimmunity. *Clin. Exp. Rheumatol.* **34,** S21–S24 (2016).

300. Hou, J., Xu, J., Jiang, R., Wang, Y., Chen, C., Deng, L., Huang, X., Wang, X. & Sun, B. Estrogen-sensitive PTPRO expression represses hepatocellular carcinoma progression by control of STAT3. *Hepatology* **57,** 678–688 (2013).

301. Kanai, T., Jenks, J. & Nadeau, K. C. The STAT5b pathway defect and autoimmunity. *Front. Immunol.* **3,** 234 (2012).

302. Kiel, M. J., Velusamy, T., Rolland, D., Sahasrabuddhe, A. A., Chung, F., Bailey, N. G., Schrader, A., Li, B., Li, J. Z., Ozel, A. B., Betz, B. L., Miranda, R. N., Medeiros, L. J., Zhao, L., Herling, M., Lim, M. S. & Elenitoba-johnson, K. S. J. Integrated genomic sequencing reveals mutational landscape of T-cell prolymphocytic leukemia. *Blood* **124,** 1460–1473 (2014).

303. Niemela, J. E., Lu, L., Fleisher, T. a, Davis, J., Caminha, I., Natter, M., Beer, A., Dowdell, K. C., Pittaluga, S., Raffeld, M., Rao, V. K., Oliveira, J.

B. & Beer, L. a. Somatic KRAS mutations associated with a human nonmalignant syndrome of autoimmunity and abnormal leukocyte homeostasis. *Blood* **117,** 2883–2886 (2012).

304. Hobbs, G. A., Der, C. J. & Rossman, K. L. RAS isoforms and mutations in cancer at a glance. *J. Cell Sci.* **129,** 1287–1292 (2016).

305. Arbore, G., West, E. E., Spolski, R., Robertson, A. A. B., Klos, A., Rheinheimer, C., Dutow, P., Woodruff, T. M., Yu, Z. X., O'Neill, L. A., Coll, R. C., Sher, A., Leonard, W. J., Köhl, J., Monk, P., Cooper, M. A., Arno, M., Afzali, B., Lachmann, H. J., *et al.* T helper 1 immunity requires complement-driven NLRP3 inflammasome activity in CD4+ T cells. *Science (80-.).* **352,** aad1210 (2016).

306. Starkebaum, G. Leukemia of large granular lymphocytes and rheumatoid arthritis. *The American journal of medicine* **108,** 744–745 (2000).

307. Flanagan, S. E., Haapaniemi, E., Russell, M. A., Caswell, R., Allen, H. L., De Franco, E., Mcdonald, T. J., Rajala, H., Ramelius, A., Barton, J., Heiskanen, K., Heiskanen-Kosma, T., Kajosaari, M., Murphy, N. P., Milenkovic, T., Seppänen, M., Lernmark, Å., Mustjoki, S., Otonkoski, T., *et al.* Activating germline mutations in STAT3 cause early-onset multi-organ autoimmune disease. *Nat. Genet.* **46,** 812–814 (2014).

308. Haapaniemi, E. M., Kaustio, M., Rajala, H. L. M., van Adrichem, A. J., Kainulainen, L., Glumoff, V., Doffinger, R., Kuusanmäki, H., Heiskanen-Kosma, T., Trotta, L., Chiang, S., Kulmala, P., Eldfors, S., Katainen, R., Siitonen, S., Karjalainen-Lindsberg, M.-L., Kovanen, P. E., Otonkoski, Ti., Porkka, K., *et al.* Autoimmunity, hypogammaglobulinemia, lymphoproliferation, and mycobacterial disease in patients with activating mutations in STAT3. *Blood* **125,** 639–649 (2015).

309. Milner, J. D., Vogel, T. P., Forbes, L., Ma, C. A., Stray-Pedersen, A., Niemela, J. E., Lyons, J. J., Engelhardt, K. R., Zhang, Y. & Al, E. Early-onset lymphoproliferation and autoimmunity caused by germline STAT3 gain-of-function mutations. *Blood* **125,** 591–599 (2015).

310. Jerez, A., Clemente, M. J., Makishima, H., Rajala, H., Gómez-Seguí, I., Olson, T., McGraw, K., Przychodzen, B., Kulasekararaj, A., Afable, M., Husseinzadeh, H. D., Hosono, N., Leblanc, F., Lagström, S., Zhang, D., Ellonen, P., Tichelli, A., Nissen, C., Lichtin, A. E., *et al.* STAT3-mutations indicate the presence of subclinical T cell clones in a subset of aplastic anemia and myelodysplastic syndrome patients. *Blood* **122,** 2453–2459 (2013).

311. Yoshizato, T., Dumitriu, B., Hosokawa, K., Makishima, H., Yoshida, K., Townsley, D., Sato-Otsubo, A., Sato, Y., Liu, D., Suzuki, H., Wu, C. O., Shiraishi, Y., Clemente, M. J., Kataoka, K., Shiozawa, Y., Okuno, Y., Chiba, K., Tanaka, H., Nagata, Y., *et al.* Somatic Mutations and Clonal Hematopoiesis in Aplastic Anemia. *N. Engl. J. Med.* **373,** 35–47 (2015).

312. Colmegna, I. & Weyand, C. M. Haematopoietic stem and progenitor cells in rheumatoid arthritis. *Rheumatology* **50,** 252–260 (2011).

313. Kataoka, K., Nagata, Y., Kitanaka, A., Shiraishi, Y., Shimamura, T., Yasunaga, J. I., Totoki, Y., Chiba, K., Sato-Otsubo, A., Nagae, G., Ishii, R., Muto, S., Kotani, S., Watatani, Y., Takeda, J., Sanada, M., Tanaka, H.,

Suzuki, H., Sato, Y., *et al.* Integrated molecular analysis of adult T cell leukemia/lymphoma. *Nat. Genet.* **47,** 1304–1315 (2015).

314. Lemonnier, F., Couronne, L., Parrens, M., Jais, J. P., Travert, M., Lamant, L., Tournillac, O., Rousset, T., Fabiani, B., Cairns, R. A., Mak, T., Bastard, C., Bernard, O. A., de Leval, L., Gaulard, P., Jaïs, J., Travert, M., Lamant, L., Tournillac, O., *et al.* Recurrent TET2 mutations in peripheral T-cell lymphomas correlate with TFH -like features and adverse clinical parameters. *Blood* **120,** 1466–1470 (2012).

315. López, C., Bergmann, A. K., Paul, U., Murga Penas, E. M., Nagel, I., Betts, M. J., Johansson, P., Ritgen, M., Baumann, T., Aymerich, M., Jayne, S., Russell, R. B., Campo, E., Dyer, M. J., Dürig, J. & Siebert, R. Genes encoding members of the JAK-STAT pathway or epigenetic regulators are recurrently mutated in T-cell prolymphocytic leukaemia. *Br. J. Haematol.* **173,** 265–273 (2016).

316. Quivoron, C., Couronné, L., Della Valle, V., Lopez, C. K., Plo, I., Wagner-Ballon, O., Do Cruzeiro, M., Delhommeau, F., Arnulf, B., Stern, M. H., Godley, L., Opolon, P., Tilly, H., Solary, E., Duffourd, Y., Dessen, P., Merle-Beral, H., Nguyen-Khac, F., Fontenay, M., *et al.* TET2 Inactivation Results in Pleiotropic Hematopoietic Abnormalities in Mouse and Is a Recurrent Event during Human Lymphomagenesis. *Cancer Cell* **20,** 25–38 (2011).

317. Fraietta, J. A., Nobles, C. L., Sammons, M. A., Lundh, S., Carty, S. A., Reich, T. J., Cogdill, A. P., Morrissette, J. J. D., DeNizio, J. E., Reddy, S., Hwang, Y., Gohil, M., Kulikovskaya, I., Nazimuddin, F., Gupta, M., Chen, F., Everett, J. K., Alexander, K. A., Lin-Shiao, E., *et al.* Disruption of TET2 promotes the therapeutic efficacy of CD19-targeted T cells. *Nature* **558,** 307–312 (2018).

318. Qi, Q., Liu, Y., Cheng, Y., Glanville, J., Zhang, D., Lee, J.-Y., Olshen, R. A., Weyand, C. M., Boyd, S. D. & Goronzy, J. J. Diversity and clonal selection in the human T-cell repertoire. *Proc. Natl. Acad. Sci.* **111,** 13139–13144 (2014).

319. Blackman, M. A. & Woodland, D. L. The narrowing of the CD8 T cell repertoire in old age. *Curr. Opin. Immunol.* **23,** 537–542 (2011).

320. Klarenbeek, P. L., De Hair, M. J. H., Doorenspleet, M. E., Van Schaik, B. D. C., Esveldt, R. E. E., Van De Sande, M. G. H., Cantaert, T., Gerlag, D. M., Baeten, D., Van Kampen, A. H. C., Baas, F., Tak, P. P. & De Vries, N. Inflamed target tissue provides a specific niche for highly expanded T-cell clones in early human autoimmune disease. *Ann. Rheum. Dis.* **71,** 1088–1093 (2012).

321. Birnbaum, M. E., Dong, S. & Garcia, K. C. Diversity-oriented approaches for interrogating T-cell receptor repertoire, ligand recognition, and function. *Immunol. Rev.* **250,** 82–101 (2012).

322. Wooldridge, L., Ekeruche-Makinde, J., Van Den Berg, H. A., Skowera, A., Miles, J. J., Tan, M. P., Dolton, G., Clement, M., Llewellyn-Lacey, S., Price, D. A., Peakman, M. & Sewell, A. K. A single autoimmune T cell receptor recognizes more than a million different peptides. *J. Biol. Chem.* **287,** 1168–1177 (2012).

323. Madi, A., Shifrut, E., Reich-Zeliger, S., Gal, H., Best, K., Ndifon, W., Chain, B., Cohen, I. R. & Friedman, N. T-cell receptor repertoires share a restricted set of public and abundant CDR3 sequences that are associated with self-related immunity. *Genome Res.* **24,** 1603–1612 (2014).

324. Quigley, M. F., Greenaway, H. Y., Venturi, V., Lindsay, R., Quinn, K. M., Seder, R. A., Douek, D. C., Davenport, M. P. & Price, D. A. Convergent recombination shapes the clonotypic landscape of the naive T-cell repertoire. *Proc. Natl. Acad. Sci.* **107,** 19414–19419 (2010).

325. Venturi, V., Quigley, M. F., Greenaway, H. Y., Ng, P. C., Ende, Z. S., McIntosh, T., Asher, T. E., Almeida, J. R., Levy, S., Price, D. A., Davenport, M. P. & Douek, D. C. A Mechanism for TCR Sharing between T Cell Subsets and Individuals Revealed by Pyrosequencing. *J. Immunol.* **186,** 4285–4294 (2011).

ORIGINAL PUBLICATIONS