Array based genetic profiling of chronic lymphocytic leukemia

Gunnarsson, Rebeqa

2010

Doctoral dissertation

ARRAY BASED GENETIC PROFILING OF CHRONIC LYMPHOCYTIC LEUKEMIA

REBEQA GUNNARSSON

With the approval of Lund University Faculty of Medicine, this thesis will be defended on February 12 2010, at 13.00 in the Segerfalk lecture hall, BMC, Lund.

Supervisor: Professor Gunnar Juliusson
Co-supervisors: Professor Richard Rosenquist,
Associate Professor Ulrike Nuber

Faculty opponent: Assistant Professor Ulf Klein,
Columbia University, New York, US

LUND
UNIVERSITY

Organization
LUND UNIVERSITY

Department of Laboratory Medicine
Stem Cell Center
BMC B10
SE-221 84 Lund

Author(s)
Rebeqa Gunnarsson

Document name
DOCTORAL DISSERTATION

Date of issue
February 12, 2010

Sponsoring organization

Title and subtitle

Array based genetic profiling of chronic lymphocytic leukemia

Abstract

Although no common genetic defect has been described in chronic lymphocytic leukemia (CLL), recurrent genomic aberrations (i.e. deletions of chromosome 11q, 13q, 17p and trisomy 12) are important for prognostication. Deletion of 13q as single aberration is associated with the best prognosis, whereas del(11q) and del(17p) predict a poor outcome. Recent development of high-resolution genomic techniques, i.e. microarrays, provides effective detection of the known recurrent aberrations and simultaneous exploration of the whole genome. Hence, this thesis aimed to map genomic aberrations in CLL through application of high-resolution microarrays. In paper I, we investigated the pros and cons of four differently designed microarray platforms. All platforms readily detected the known recurrent aberrations in CLL as well as other large copy-number aberrations (CNAs). The difference in technical performance and discrepancies in detection of small CNAs were influenced by differential platform density, different reference sets, and platform-specific analysis. In paper II and IV, 250K single nucleotide polymorphism (SNP)-array screening of 203 and 370 samples, respectively, detected CNAs in >90% and the known recurrent aberrations in >70% of patients. Moreover, novel recurrent gains of chromosome 2p in combination with 11q deletion and copy-number neutral loss of heterozygosity on chromosome 13q were revealed. Furthermore, a high genomic complexity was correlated to worse survival, but also closely linked to poor-prognostic markers. In addition, study IV also included follow-up samples (n=43) to investigate clonal evolution, which showed that patients with unmutated immunoglobulin heavy chain variable (IGHV) genes and treated patients with mutated IGHV genes often gained novel aberrations. In paper III, high-density screening revealed a different spectrum of genomic aberrations in CLL patients with 'stereotyped' IGHV3-21 (poor-prognostic) versus IGHV4-34 (good-prognostic) B-cell receptors. IGHV3-21 subset #2 (n=29) showed a high frequency of samples carrying genomic aberrations, and a particularly high prevalence of del(13q) and del(11q), which may correspond to the adverse survival reported for these patients. In contrast, IGHV4-34 subset #4 (n=17) showed a lower frequency and complexity of CNAs, which may reflect an inherent low-proliferative disease, preventing accumulation of genomic alterations. In summary, the studies included in this thesis provided a greater insight of genomic aberrations in newly diagnosed patients, at follow-up and in different subgroups of CLL.

Key words: genomic microarrays, chronic lymphocytic leukemia, copy-number aberrations, copy-number neutral loss of heterozygosity, prognostic markers, clonal evolution, stereotyped B-cell receptors

Classification system and/or index termes (if any):

Supplementary bibliographical information:

Language
English

ISSN and key title:
1652-8220

ISBN
978-91-86443-22-1

Recipient's notes

Number of pages
99

Price

Security classification

ARRAY BASED GENETIC PROFILING OF CHRONIC LYMPHOCYTIC LEUKEMIA

REBEQA GUNNARSSON

LUND
UNIVERSITY

Department of Laboratory Medicine
Faculty of Medicine
Lund University
Sweden
2010

Printed in Sweden
Media Tryck AB, Lund 2010
ISSN 1652-8220
ISBN 978-91-86443-22-1
Lund University, Faculty of Medicine Doctoral Dissertation Series 2010:8

TABLE OF CONTENTS

LIST OF PAPERS

I. Gunnarsson R, Staaf J, Jansson M, Ottesen AM, Göransson H, Liljedahl U, Ralfkiaer U, Mansouri M, Buhl AM, Smedby KE, Hjalgrim H, Syvänen AC, Borg A, Isaksson A, Jurlander J, Juliusson G, Rosenquist R. **Screening of copy-number aberrations and loss-of-heterozygosity in chronic lymphocytic leukemia – a comparative study of four differentially designed microarray platforms.** *Genes Chromosomes Cancer* 2008;47:697-711

II. Gunnarsson R, Isaksson A, Mansouri M, Göransson H, Jansson M, Cahill N, Rasmussen M, Staaf J, Lundin J, Norin S, Buhl AM, Smedby KE, Hjalgrim H, Karlsson K, Jurlander J, Juliusson G, Rosenquist R. **Large but not small copy-number alterations correlate to high-risk genomic aberrations and survival in chronic lymphocytic leukemia: A high resolution genomic screening of newly diagnosed patients.** *Leukemia,* Advanced online publication 10 September 2009

III. Marincevic M*, Cahill N*, Gunnarsson R*, Isaksson A, Göransson H, Rasmussen M, Jansson M, Mansouri M, Ryan F, Karlsson K, Adami HO, Jurlander J, Juliusson G, Davi F, Stamatopolous K, Rosenquist R. **High-density genomic screening reveals a different spectrum of aberrations in chronic leukemia patients with stereotyped IGHV3-21 versus IGHV4-34 B-cell receptors.** Submitted

* Equal contribution

IV. Gunnarsson R, Mansouri M, Isaksson A, Göransson H, Cahill N, Jansson M, Rasmussen M, Lundin J, Norin S, Buhl AM, Smedby KE, Hjalgrim H, Karlsson K, Jurlander J, Juliusson G, Rosenquist R. **Genome-wide array-based screening at diagnosis and follow-up in chronic lymphocytic leukemia**. Manuscript

ABBREVIATIONS

aCGH	Array comparative genomic hybridization
B2M	β-2 microglobulin
BAC	Bacterial artificial chromosome
BASE	Bio-array software environment
BCL2	B-cell CLL/lymphoma 2
BCR	B-cell receptor
CD	Cluster of differentiation
CDR	Complementarity determining region
CGH	Comparative genomic hybridization
CLL	Chronic lymphocytic leukemia
CLLU1	CLL-upregulated gene-1
CNA	Copy-number alteration
CNV	Copy-number variant
CNN-LOH	Copy-number neutral loss-of-heterozygosity
CR	Complete remission
DAPK1	Death associated protein kinase 1
DNA	Deoxyribo-nucleic acid
FISH	Fluorescence in-situ-hybridization
GC	Germinal center
IG	Immunoglobulin
IGH	Immunoglobulin heavy chain
IGHV	Immunoglobuline heavy variable
IGK	Immunoglobulin kappa chain
IGL	Immunoglobulin lambda chain
K	Kilo
Kbp	Kilo base pair
LDT	Lymphocyte doubling time
LPL	Lipoprotein lipase
MCL1	Myeloid cell factor-1
MBL	Monoclonal B-cell lymphocytosis
Mbp	Mega base pair
miR	micro-RNA
PCR	Polymerase chain reaction
RNA	Ribo-nucleic acid
SLL	Small lymphocytic lymphoma
SNP	Single nucleotide polymorphism
STAC	Significant testing for aberrant copy-number

TCL1	T cell leukemia/lymphoma 1
UPD	Uniparental disomy
TK	tymidine kinase
ZAP70	zeta-chain associated protein kinase 70kDa

SAMMANFATTNING PÅ SVENSKA (SUMMARY IN SWEDISH)

Kronisk lymfatisk leukemi (KLL) är en sjukdom som drabbar ca 500 människor i Sverige varje år och räknas därmed som den vanligaste leukemin i västvärlden. Vid KLL ackumuleras B-lymfocyter i blod, benmärg och andra lymfoida organ, vilket på sikt kan leda till symptom och följdsjukdomar. Sjukdomsbilden vid KLL är väldigt heterogen då vissa patienter lever ett relativt symptomfritt liv utan sjukdomsprogression eller behov av behandling, medan andra patienter har snabbt progredierande sjukdom som är svårbehandlad. Denna stora skillnad i sjukdomsförlopp gör att det är mycket angeläget att upptäcka metoder som på ett enkelt sätt kan ställa en säker prognos på KLL-patienter. I dagsläget finns det ett flertal sätt att prognostisera sjukdomen, av vilka en av de med stabila är att analysera mutationsstatus inom immunoglobulingenens tunga kedja. Genom denna metod kan man identifiera ”omuterade” patienter, som har en sämre prognos i jämförelse med de fall som identifieras som ”muterade”. Alternativa metoder fastställer halten av olika molekyler på eller i KLL-cellerna, som t.ex. cellytemarkören CD38 och de intracellulära molekylerna ZAP70, *CLLU1* och *LPL*. Vidare har genetiska studier på KLL visat att det finns specifika, återkommande kromosomala förändringar, som ger en god uppskattning av patienternas sjukdomsförlopp. Patienter med deletion av den långa armen (q) på kromosom 11 och den korta armen (p) av kromosom 17, har en dålig prognos, medan patienter med deletion av kromosom 13q, har den längsta överlevnaden. Denna genetiska information är värdefull vid värdering av patientens status inför insättning av behandling.

Genetiska förändringar kan bland annat detekteras med microarrays, en teknik som möjliggör detektion av avvikelser med hög upplösning. Vid användning av hel-genomiska arrayer kan man dessutom undersöka huvudparten av den genetiska uppsättningen. Denna metod bygger på att man jämför DNA från ett homogent cancercellprov mot normalt DNA från ett stort antal normala individer. Jämförelsen resulterar i detektion av genomiska kopietalsförändringar i cancerprovet. Sådana förändringar kan vara förlust av genomiskt material, vilka uppkommer genom deletioner, eller tillkomst av genomiskt material skapade genom duplicering eller amplifiering. En nyare typ av microarrays, single nucleotide polymorphism (SNP)-arrays, kan förutom detektion av tillkomst eller förlust av DNA, även detektera kopietalsneutrala förändringar. Dessa förändringar

har alltså två intakta kopior av DNA, men består enbart av en av kopiorna från ena föräldern, inte en av varje som är det normala fallet.

Fyra delarbeten ingår i denna avhandling där den övergripande målsättningen har varit att använda microarrayteknologi för att identifiera och utvärdera genetiska avvikelser hos KLL-patienter. Detta gav upphov till en inledande jämförelsestudie, där vi utvärderade fyra olika typer av microarrays genom att jämföra metodernas prestanda och detektionsgrad av genomiska avvikelser i tio KLL-prover (arbete I). I detta första delarbete kunde vi fastställa att det fanns flera för- och nackdelar med de olika metoderna, både gällande den tekniska prestandan och förmågan att detektera stora och små kopietalsförändringar. Vår förvärvade kunskap låg delvis till grund för valet av microarrayteknik för vår fortsatta utvärdering av KLL med den specifika målsättningen att beskriva förekomsten av genomiska avvikelser hos ett stort antal KLL-patienter ur ett populationsbaserat material (203 patienter i arbete II och 370 patienter i arbete IV). Denna undersökning visade att majoriteten av KLL-patienterna bär på genomiska avvikelser, som ofta, men inte uteslutande, representerade de kända genomiska avvikelserna med prognostisk betydelse. Andra kopietalsförändringar sågs återkommande på kromosom 2p, 4p, 8p, 8q, 14q, 18 och 19. Vidare detekterades kopietalsneutrala förändringar på kromosom 13q, vilka ofta inkluderade deletionen på 13q. När de genomiska fynden relaterades till den kliniska informationen, såg vi en korrelation mellan en ökad genomisk komplexitet och kortare överlevnad. Denna korrelation beror till stor del på att de patienter som uppvisade genomisk komplexitet ofta bar på de genomiska avvikelser som indikerar en dålig prognos.

Som ett delmoment i vår mest omfattande microarraystudie undersökte vi tillkomsten av genetiska avvikelser (klonal utveckling) hos ett urval av patienterna (artikel IV). Här jämfördes den genomiska profilen i diagnosprovet med ett uppföljningsprov insamlat 5-7 år efter diagnostillfället för varje patient. Uppföljningsstudien visade att de patienter som bär på omuterade immunoglobulingener hade en större tendens att inkorporera nya avvikelser över tid. Detta kan till viss del bero på att omuterade patienter har en aggressivare sjukdom, där KLL-cellerna utsätts för ett större selektivt tryck.

I en ytterligare microarraystudie jämfördes förekomsten av genomiska avvikelser i olika KLL-undergrupper, där patienterna har blivit grupperade efter graden av homologi av specifika sekvenser inom den mest hypervariabla regionen på tunga immunoglobulingenen, dvs den antigenbindande CDR3 (arbete III). Här fann vi att de patienter som uttrycker IGHV3-21/IGLV3-21 generna (9 aminosyror lång

CDR3) och som uppvisar ett mer aggressivt sjukdomsförlopp, hade en högre genomisk komplexitet och en hög förekomst av deletioner på kromosom 13q och 11q. Vi screenade också patienter som uttrycker IGHV4-34/IGKV2-30 generna (20 aminosyror lång CDR3). Dessa patienter som ofta innefattar unga patienter med ett indolent sjukdomsförlopp uppvisade färre genomiska avvikelser. Detta kan till viss del bero på att denna patientgrupp har en låg-proliferativ sjukdom, som inte ackumulerar genetiska avvikelser i samma takt som en aggressiv form.

Sammanfattningsvis har delarbetena i denna avhandling bidragit till en ökad kunskap om genomiska avvikelser vid KLL. En högre genomisk komplexitet och en större benägenhet att utveckla nya genomiska avvikelser identifierades i patienter med dåliga prognosmarkörer, dvs. hos patienter med deletion 11q och 17p, och hos patienter med omuterade immunoglobulingener. I motsats påvisades en lägre genomisk komplexitet och en mindre benägenhet att utveckla nya genomiska defekter i patienter med goda prognosmarkörer. Denna information visar att de olika biologiska markörerna antingen är beroende av, eller påverkar uppkomsten av kromosomala avvikelser i KLL, vilket på sikt förhoppningsvis kan leda till en ökad förståelse av sjukdomsbilden.

PREFACE

B-cell chronic lymphocytic leukemia (CLL) is a common lymphoid malignancy with a variable clinical course. CLL is an incurable disease, however many patients never require treatment or can be managed with intensive or palliative treatment. Nevertheless, some patients will suffer from a progressive disease, with a fatal outcome despite initiation of treatment. Research in the CLL field has led to identification of molecular markers with prognostic value, such as the mutation status of the immunoglobulin heavy-chain variable (IGHV) genes. Moreover, certain known recurrent aberrations are linked to the survival and response to treatment in CLL patients. Although CLL has been assessed for chromosomal aberrations, there is no commonly detected aberration found in all cases. However, it is unlikely that all important genetic changes have been identified by previously used methods with limited sensitivity, and it is therefore valuable with application of high-resolution techniques for characterization of genomic events. It is also important to correlate such genomic findings to clinical data in order to indentify aberrations of prognostic importance. Microarray screening can also be used to compare the prevalence of genomic aberrations in CLL subsets with specific characteristics, such as patients with very similar, 'stereotyped' B-cell receptors. Such comparisons will hopefully improve the knowledge of the link between genomic aberrations and other biological features in CLL. Furthermore, the detailed information provided by genomic arrays allows straightforward detection of novel aberrations in serial samples, with possible identification of patient risk-groups.

The general aim of this thesis was to characterize genomic aberrations in CLL patients by applying genomic microarrays, in order to gain improved knowledge of genomic aberrations and their possible influence on the disease. This thesis is divided into three different sections; the first part contains a general introduction of the field, including cancer genetics, normal B-cell biology, different aspects of CLL, and the basics of the techniques applied for genomic profiling. The second section specifies the aims of the thesis, summarizes materials and methods and presents the results and conclusions from each study followed by a general discussion. The third part includes the original papers on which this thesis is based.

Lund, January 2010

INTRODUCTION

Cancer genetics

The human genome

The human genome is diploid, represented by 23 chromosomal pairs which are stored in the cell-nucleus of almost all cells in the body. The chromosomes include genetic information and each chromosomal pair enclose analogous genetic information on the two corresponding alleles i.e. the maternal and the paternal alleles. The chromosomes are constructed by long deoxyribo-nucleic acid (DNA) strands with a specific sequence of the chemical base pairs (A, T, G and C). The DNA of a haploid genome include 20000-25000 genes, which are transcribed into messenger RNA (mRNA), a molecule that acts as an intermediate between the genetic and the end product, the protein. Outside the nucleus, the mRNA nucleotide sequence is translated in triplets, called codons, which encodes specific amino acids that function as building blocks for protein synthesis. Moreover, the mRNA codon specifies the order in which the amino acids are assembled. The resulting polypeptide chain will thereafter fold and form a functional protein. The proteins are essential molecules because they participate in almost every process within the cells. Furthermore, these molecules function as important links between cells such as cell signaling and immune responses. When cells proliferate the DNA has to be duplicated in order to preserve the genetic information in the resulting daughter cells. This replication process can sometimes induce errors in the genes which is the first step in carcinogenesis, i.e. the transformation of normal cells into cancer cells. Moreover, external factors such as UV-radiation, chemicals and bacterial and viral infections can induce DNA damage which may lead to cancer development.[1-4] Epigenetic factors, such as the methylation status of the DNA, which influence the transcription of DNA, and microRNAs, which regulate genes on the post-translational level, have also been shown to play a role in carcinogenesis.[5,6]

Genomic alterations in cancer

In the beginning of the 1900s, several researchers such as Boveri, Tyzzler and Whitman postulated that the development of cancer is associated with genomic abnormalities and the work of Boveri has recently been translated and published in English.[7] It is now known that genomic aberrations leading to cancer can arise through several different mechanisms of which mutations are thought to be the most common. Mutations can occur through base substitutions in the DNA, which leads to nonsense or missense amino-acid changes in the protein. Moreover, insertions or deletions of bases in coding sequences or splice sites might cause in-frame or frameshifting alterations. A different form of genomic aberrations is caused by physical rearrangement of genes through translocations of chromosomes. These chromosomal aberrations lead to formation of fusion genes with an altered function, or deregulation of genes caused by their translocation to a novel promoter or enhancer region. Translocations are commonly detected as pathogenic events in leukemia and lymphomas, i.e. cancer affecting the white blood cells. For instance, the translocation occurring between chromosome 9 and 22, i.e. t(9;22), which causes formation of the fusion protein BCR-ABL is the hallmark of chronic myelogenous leukemia (CML).[8] This translocation, which formed the "Philadelphia chromosome"[9] is also found in 25% of adult acute lymphoblastic leukemia (ALL). Other common translocations involve the *MYC* gene on chromosome 8 which is relocated to the immunoglobulin (IG) gene loci on chromosomes 2, 14 and 22 giving rise to t(8;2), t(8;14) and t(8;22) in Burkitt lymphomas.[10-12] Furthermore, unbalanced chromosomal alterations, such as copy-number aberrations (CNAs), are caused by deletions or gains of chromosomal material, respectively (Figure 1). Such events can lead to alteration of the expression-level, i.e. up- or down-regulation of the genes within these regions.

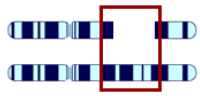
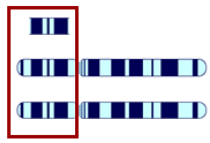
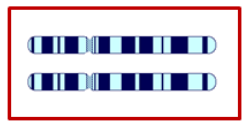
Detection	Copy-number change	Example
Copy-number loss (CNA)	Heterozygous or homozygous deletion of genomic DNA	
Copy-number gain (CNA)	gain or high-copy amplification of genomic DNA	
Copy-number neutral alteration (CNN-LOH)	allelic imbalance without a change in DNA copy-number	

Figure 1. Examples of genomic aberrations in cancer. CNA, copy-number aberration, CNN-LOH, copy-number neutral loss of heterozygosity.

Moreover, combinations of a deleted gene on one allele and a mutation affecting the same gene on the other allele can lead to loss of function of tumor suppressor genes, as shown for *TP53* and *Rb*. Another genetic event detected in cancer is copy-number neutral loss of heterozygosity (CNN-LOH), also referred to as uniparental disomy (UPD) (Figure 1). CNN-LOH is caused by duplication of one allele and a concurrent loss of the other allele. UPDs have been detected in several cancers such as follicular lymphomas, colorectal cancer and CLL.[13-17] However, the link between CNN-LOH and cancer is still elusive. Genomic aberrations such as CNN-LOH, translocations and CNAs can be detected through molecular cytogenetic techniques and microarrays, which will be described in the methodological section.

Carcinogenesis

All dividing cells in the human body are controlled through different mechanisms that prevent carcinogenesis. Importantly, this prevention requires a fine balance between cell division (proliferation) and programmed cell death (apoptosis) to maintain a normal cell-cycle. The breaching of the normal cell-cycle is a critical step in carcinogenesis because it allows accumulation of additional genetic defects and cellular traits which leads to cancer.[18] Carcinogenic defects often impair genes that choreograph the life-cycle of the cells. The two major classes of such genes are tumor suppressor genes and oncogenes.[19] Tumor suppressor genes are important regulators because they prevent cell proliferation upon cellular stress or DNA damage which leads to cell-cycle arrest and/or activation of apoptosis. Genetic defects of tumor suppressor genes, such as mutations or deletions can lead to loss of function in these genes, which in turn can lead to accumulation of damaged DNA. The loss of function in tumor suppressor genes generally follow a two-hit model, which suggests that both allelic copies of the gene have to be affected before an adverse effect can arise.[20] However, recent gene expression studies have suggested a possible gene-dose effect of tumor-suppressor gene function which implies that mono-allelic changes or aberrant imprinting are sufficient predisposing factors in carcinogenesis.[19,21] Examples of tumor suppressor genes involved in carcinogenesis are the *BRCA1*, *RB*, and *TP53* genes. For instance, half of all cancers lack a functional p53 protein due to mutations.[22] In contrast, proto-oncogenes are responsible of promoting cell proliferation in various ways. Mutations or other defects of proto-oncogenes can lead to formation of oncogenes with a gain of function through modification of their expression, function and level of activity. Hence, the gain of function observed in oncogenes can be due to mono-allelic changes. For instance, alterations of *BCL2*, *MYC* and *RAS* produce very potent oncogenes, which will keep the cell in a proliferative state and block apoptosis, thus paving the way for cancer formation.

Carcinogenesis involves a multi-step process in which accumulation of several mutations is needed in order to develop cancer.[20,23] Thus, the tumor development is considered as an evolutionary process, where each genetic change signifies an advantage in cell proliferation, which in combination eventually will convert the normal cells into cancer cells.[18] The finding that only a fraction of the tumor cells are capable of forming colonies *in vitro* or reconstructing a new tumor *in vivo*, suggested the existence of a minor subpopulation within the tumor with stem cell like properties.[24,25] These cancer stem cells are thought to have a self-renewing capacity, which allow maintenance of the tumor, as well as the ability to

differentiate (i.e. form new tumors in transplantation experiments).[24-27] Another explanation to the difference in clonogenecity is suggested by the stochastic model, which states that the low proliferative capacity of tumor cells are of equal low probability in all tumor cells.[28] Genetic changes will occur continuously in the proliferating tumor cells and give rise to clonal evolution and the existence of several genetically distinct cancer cell sub-clones. This feature, which has been documented in CLL, will be described in more detail in the corresponding section.

Techniques for genetic profiling

Many different techniques can be applied for identification of genomic changes, i.e. structural alterations of chromosomal DNA. Examples of such techniques, which have paved the way for cancer genomics, are cytogenetic chromosome banding, fluorescence in situ hybridization (FISH) and conventional comparative genomic hybridization (CGH). These techniques can detect unbalanced aberrations i.e. deletions and gains. Cytogenetic banding techniques and FISH also allow detection of structural (balanced) reorganization of DNA such as inversions and translocations. The aforementioned techniques and the more recently developed genomic microarray techniques will be described in some detail below.

Cytogenetic banding techniques

Karyotyping, which involves characterization of the chromosomes with regards to size, number, and form was improved in 1970 when the first cytogenetic banding technique was developed.[29] Since chromosomal translocations can be detected with this method, several such balanced alterations were identified upon introduction of this technique. For instance, the t(9;22)(q34;q11) translocation resulting in the *BCR-ABL* fusion gene, was identified and further characterized with this technique.[8,30,31] The banding technique allows visualization of chromosomes after destruction of the mitotic spindle in dividing cells, followed by fixation and staining of the DNA. The actual chromosomal resolution resides on the level of DNA condensation in the disrupted cells. The human chromosomes are divided into a short arm (p for petit) and a long arm (q), which are separated by a centromere. The arms are further divided into regions, and the regions can be sectioned into bands. Various banding techniques have been developed for staining of the chromosomal bands such as Q-banding, R-banding and G-banding methods, which simplifies the recognition of the different chromosomes.

Moreover, the bands and regions are useful for description of the chromosomal "location" of genomic aberrations. Although the banding technique is still used for chromosomal characterization, this technique possesses several shortcomings. For instance, this method provides a relatively low resolution which makes small genomic aberrations and cryptic translocations impossible to distinguish. Furthermore, this technique requires dividing cells.

Fluorescence in situ hybridization

Fluorescence in situ hybridization (FISH) was introduced in the late 1980´s and provided a breakthrough in cytogenetic research.[32] This molecular karyotyping technique enables characterization of submicroscopic and/or complex chromosomes. The first two variants of this technique were metaphase FISH, which requires viable cells in metaphase to be performed, and interphase FISH, which can be performed on interphase nuclei, i.e. non-dividing cells. Thus, the latter form allows investigation on material from e.g. paraffin-embedded samples. Currently, several different types of FISH techniques are available, all based on the hybridization of specifically labeled probes with specific labels to a complementary DNA sequence on specific chromosomal locations (Figure 2). Moreover, several FISH probes have been developed for disease specific applications and are used for diagnostics and prognostics in cancers such as CLL, breast cancer and bladder cancer. Moreover, by applying the FISH technique it is possible to detect narrow CNAs, since the detection is based on probes with sizes down to ~100 bp. This method allows detection of homo- and heterozygous deletions through visualization of the probe specific binding on both alleles. Genome-wide screening and detection of novel aberrations are not feasible to perform with conventional FISH, since this technique is directed to specific genomic regions. However, development of spectral karyotyping (SKY) and multiplex FISH (M-FISH) apply multiple hybridization probes that allow simultaneous visualization of all human chromosomes, which can detect complex rearrangements, translocations, deletions and gains.[33,34]

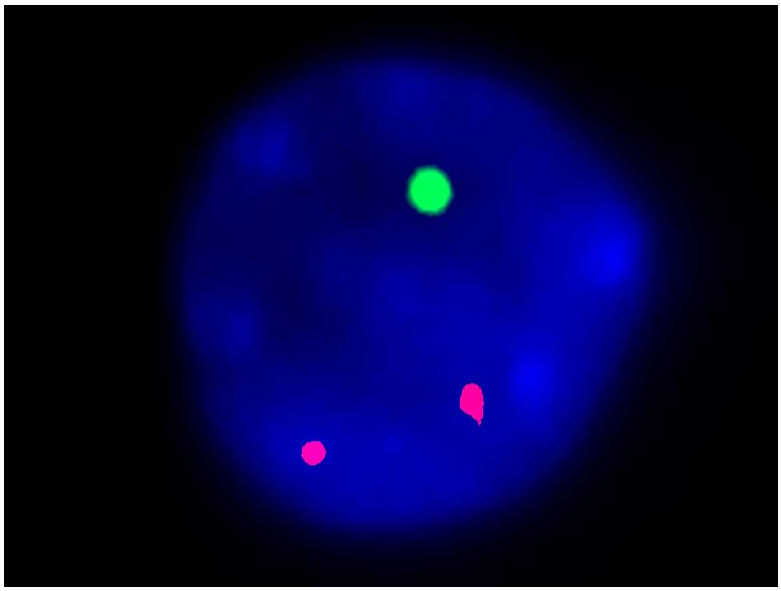


Figure 2. Fluorescence in-situ hybridization (FISH). A cell stained with fluorescent dye with probes directed to *ATM* and *TP53* genes on chromosomes 11q (green) and 17p (red), respectively. This cell has two copies of the *TP53* coding region and one copy of the ATM coding region (i.e corresponding to a heterozygous deletion).

Comparative genomic hybridization

Comparative genomic hybridization (CGH) was first reported in 1992 by Kallioniemi et al.[35] CGH is based on quantitative hybridization of two different samples, i.e. the sample of interest (often derived from a tumor) and a reference (often pooled DNA from a large number of normal individuals).[35] The two differently labeled DNA samples are mixed and hybridized to normal metaphase chromosomes. Chromosomal alterations in the sample are quantifiable due to higher or lower ratio of hybridized sample DNA compared to the reference DNA, and the regional ratio differences are visualized with fluorescence microscopy. CGH is superior to FISH in view of the fact that the whole genome can be investigated for aberrations down to ~10 Mbp in size. Another advantage of CGH

compared to the cytogenetic banding technique and metaphase FISH, is that CGH does not require analyzable metaphases of mitotic cells, which can be an obstacle when investigating cells with a low mitotic index. However, a prerequisite and a major drawback with CGH is that the CNAs have to exist in a high proportion of the cells in a sample in order to be detectable subsequent to hybridization. In comparison, the FISH technique is based on hybridization of probes to individual cell nuclei which enables detection of changes in a low fraction of the tumor sample (i.e. subclones), and can be of prognostic relevance. Moreover, CGH and other newly developed array techniques cannot detect structural changes such as inversions or translocations, although changes in gene dosage have been noted at potential chromosomal breakpoints for these balanced aberrations.

Array comparative genomic hybridization

The first genomic microarray technique for detecting DNA alterations, array comparative genomic hybridization (aCGH), was developed from conventional CGH in the late 1990´s.[36,37] Akin to conventional CGH, aCGH is based on detection of CNAs due to a higher or lower ratio of inbound differently labeled sample DNA compared to a reference DNA (Figure 3). However, instead of metaphase chromosomes acting as targets for sample DNA binding, aCGH utilizes thousands of short DNA probes, each representing a unique part of the genome, arrayed onto for example a glass slide. Fluorescence for each probe is recorded separately for the sample and reference channel using, e.g., a laser scanner, and the formed ratio (sample/reference) can be plotted against the known genomic positions of each probe creating a high-resolution map of DNA CNAs. The advancement of aCGH in terms of resolution has been dependent on the determination of the human genome sequence, which was completed in 2003. With the exception of regions on centromeres and telomeres that consist of highly repetitive sequences i.e. heterochromatin, aCGH are available as whole genome bacterial artificial chromosome (BAC) arrays or oligonucleotide arrays (oaCGH).

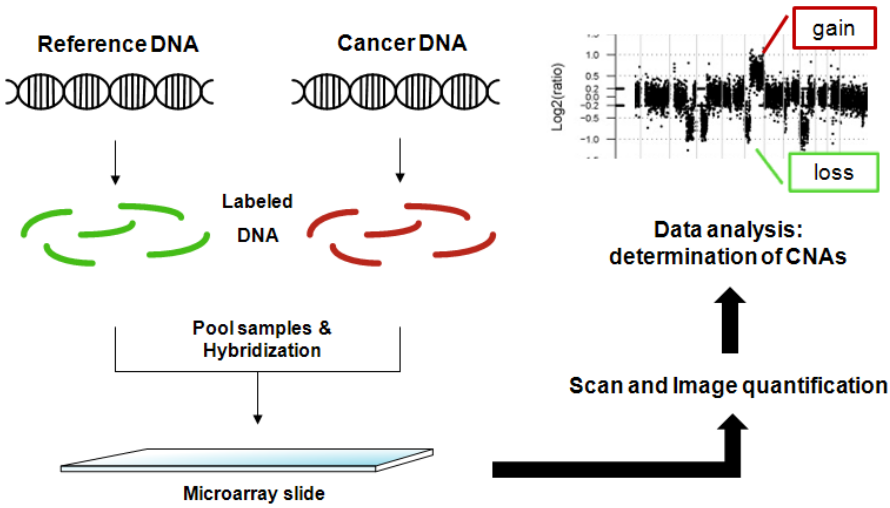


Figure 3. Array comparative genomic hybridization (aCGH). Reference DNA and cancer DNA are differently labeled with fluorescent dyes prior to hybridization to the array. Spotted DNA on the microarray slide represents sequences with a known genomic localization in the human genome. After hybridization of the DNA samples, the array is scanned and the ratio of fluorescent intensity of the inbound cancer DNA, to that of the reference DNA is used to identify copy number losses and gains on the chromosomes.

BAC probes are produced by insertions of DNA sequences to BAC (or sometimes PAC) clones in order to amplify the DNA sequences.[36-38] The DNA is extracted, fragmented and spotted to glass slides at specific positions, where each DNA probe will represent 100-200 kbp of the genome. Although this technique is robust, it has reached its limit in terms of resolution.[39] Furthermore, the fact that different laboratories often use their own BAC aCGH platform makes it harder to compare results conveyed at different study centers. The design of oaCGH is somewhat different from BAC arrays, for instance the probes consist of 60-mer oligonucleotides which are synthesized directly on the arrays. Furthermore, since the oaCGH is constructed with a higher number of probes than the BAC array, the former will permit detection of smaller aberrations with higher resolution.[40]

Single nucleotide polymorphism arrays

The single nucleotide polymorphism (SNP) arrays, are constructed with a slightly different approach than that of aCGH. SNP-arrays are composed of a large number of 25-50mer oligonucletide primers which represents SNPs. The SNPs represent the most common type of genetic variation in the human genome, where almost all known SNPs exist as two common allelic versions. Each SNP position is represented by the two allelic variants (A and B) on SNP-arrays, thus allowing interrogation of allelic imbalances. Hence, the primary application of SNP-arrays has been to identify susceptibility genes through SNP-genotyping of populations or specific groups of interest. In addition, SNP-arrays allow identification of copy neutral loss of heterozygosity (CNN-LOH), which represents alterations without a change in copy-number. Moreover, since the SNP-arrays of today often contain a high number of probes it is custom to align the SNPs in chromosomal order and use them for karyotypic purposes. In contrast to aCGH, only the sample of interest is hybridized to the SNP-array. Therefore, copy number analysis requires the combination of the two corresponding SNP-data points, SNP A and SNP B, in order to generate an estimate of the number of copies for a specific SNP. Moreover, normalization to a reference data set, which consists of samples that have been individually run on the same type of SNP platform, is required in order to generate ratios (sample/reference) similar to those obtained directly in an aCGH experiment for detection of chromosomal aberrations in the investigated sample. Currently, SNP-arrays are available up to ~10^6 probes, which allow high-throughput, genome-wide screening and with detection of cryptic CNAs and CNN-LOH.

Genomic studies of CLL

In CLL, deletions and gains are more commonly detected than translocations, and recurrent deletions of chromosome 11q, 13q and 17p and trisomy 12 were initially shown by applying conventional cytogenetic techniques.[41-45] However, CLL cells have a low proliferative index, which makes it hard to investigate their karyotype with conventional techniques, although, recently novel cell culture protocols have been developed.[46,47] The development of the FISH technique allowed detection of genetic changes in a greater proportion of CLL cases. Where the known recurrent aberrations can be detected in approximately 80% of CLL patients by applying this technique.[41] Today, FISH is routinely performed by applying specific probes directed to the known recurrent aberrations. Furthermore, since the deletion of 13q

is known to affect one or two alleles in CLL cells, FISH has allowed detection of both homo- and heterozygous deletions.[48,49] However, a recent report indicated that atypical aberrations, in this case deletion of 11q, was missed by the FISH-probe directed to del(11q).[50]The different types of microarray platforms are advantageous to older cytogenetic techniques in several aspects, for instance, the arrays allow whole genomic profiling of large sample sets in order to detect novel genomic aberrations and characterize the genomic complexity. Thus, the usage of microarray technology has provided additional knowledge of the known recurrent aberrations as well as identification of novel aberrations in CLL.[15,16,51-54] For instance, the known recurrent aberrations del(17p), which covers the *TP53* gene, has been further delineated and characterized with microarray technology.[55] Also, additional aberrations such as gain of 2p at the *N-MYC* encoding region and alterations of chromosome 8p and 8q have been reported.[14,16,52,56] Recently, a study showed a novel recurrent aberration on chromosome 22q, close to the IGL gene locus.[57] Moreover, by applying a 50K SNP-array a correlation between genomic complexity and a shorter time to initiation of treatment was also revealed.[58] Similar correlations between complex karyotypes and poor survival was established already with cytogenetics in the 1980´s.[59] Furthermore, SNP-array studies in CLL have enabled detection of CNN-LOH, which has been described on chromosome 11p, 13q and 17p.[14-16] Altogether, these findings provided by various types of microarray technologies have led to a greater understanding of genomic changes in CLL.

Development of normal B cells

Normal hematopoesis of B cells

B cells are essential for the immunological defense through identification of antigens on foreign pathogens with the cell membrane bound B-cell receptor (BCR). In the human adult, the generation of B cells occurs in the bone-marrow from hematopoietic stem cells which produce all cells that constitute the blood system.[60] In the early hematopoesis, the stem cells will differentiate into progenitor cells which can give rise to B cells. The stromal cells, which surround the stem cells, secrete different amounts and various types of growth factors that promote the differentiation into B cells. Furthermore, the B cell commitment is dependent on the transcription factor Pax5. The first step in the B-cell commitment is when the progenitor cell differentiates into the earliest distinctive B-linage cell, the progenitor B cell (pro-B cell). At this stage, the cell will start the first

rearrangement step of the immunoglobulin heavy chain (IGH) locus which is a feature that distinguishes B cells from all other cells in the human body. pro-B cells also start expressing the trans-membrane immunoglobulin molecules Igα/Igβ and CD19, which are part of the BCR and the BCR co-receptor, respectively. The next step, represented by pre-B cells, is marked by expression of the pre-B cell receptor and pronounced proliferation and differentiation into immature B cells. At this stage, the immature B cells have rearranged the light chain loci and starts expressing the IgM protein on the cell surface. The immature B cell will undergo a process called negative selection, in which B-cells that recognize and bind to naturally present molecules in the bone marrow, i.e. self-antigens, are eliminated through apoptosis. This mechanism is important for the destruction of self-reactive cells and prevention of auto-immune diseases. The B cells that surpass the negative selection will leave the bone marrow and will most often die in the periphery within a few weeks by apoptosis. However, the B cells that recognize foreign antigens will proceed through antigen-driven activation, proliferation and differentiation which give rise to memory B cells and plasma B cells. These steps will be explained in detail after the description of the recombination of the IG genes.

Immunoglobulin gene rearrangement

The rearrangement of the IG genes occurs during the development of B cells and will ultimately lead to the production of a functional antibody molecule which has two major and essential functions, i.e. antigen recognition and cell signaling through the BCR. Numerous IG gene segments are present in germline DNA, however, these cannot be expressed until they are rearranged into functional genes. The recombination process involves sequential reorganization of gene segments within the IG heavy (IGH), and IG light kappa (IGK) and IG light lambda (IGL) loci on chromosomes 14, 2 and 22, respectively. The IGH locus consists of ~40 variable (V), 23 diversity (D), 6 joining (J) and 9 constant (C) functional genes, that are separated by non-coding sequences.[61] The IGK and IGL loci also include V, J and C genes, however lack D genes. The rearrangement process starts by joining of an IGHD gene to an IGHJ gene, followed by joining of an IGHV gene to the IGHD-J gene complex.[62] The V-D-J joining is dependent on specific DNA sequences termed recombination signal (RS) sequences. The RS sequences are located on the 3´ end of the V genes, the 5´end of the J genes and on both sides of the D genes. Moreover, the sequences flanking the V, J and D genes have different length which ensures that they are appropriately recombined. The RS sequences

are recognized by specific enzymes called recombination-activating-genes RAG-1 and RAG-2 which are essential for the V-D-J recombination process.[62] Upon binding of RAG1/RAG2 to the RS sequences, the enzymes will introduce double-strand breaks of the DNA, which leads to joining of specific D to J and DJ-V gene segments and deletion of all other non-rearranged genes. Subsequently, the enzyme terminal deoxynucleotidyl transferase (TdT) will correct and add additional nucleotides in the junctions between the fused IG gene segments.[63] A successful VDJ rearrangement will downregulate the RAG1/RAG2 expression which prevents further rearrangement on the corresponding allele, and confers the expression of a single IGH gene rearrangement.[64] The IGL and IGK gene rearrangement occur in a similar manner, subsequent to IGH gene recombination. This process starts with recombination of the IGK genes and if both alleles are non-functionally rearranged the process will continue with IGL gene rearrangement. This process culminate in isotypic exclusion since only one type of functionally rearranged IGK or IGL gene will be used. The rearrangement of the IG genes can generate more than 10^6 possible combinations of IG molecules due to the random combinatorial assembly of V, D and J genes. Moreover, the joining of D to J and DJ-V gene segments, are accompanied by nucleotide addition and deletion by TdT and exonuclease, respectively, which result in an even higher diversity. Thus, every single B cell will express a unique antibody molecule with a specific target-antigen. As a result, all B cells that develop will collectively provide a great 'pool' of antibody diversity which allows recognition of a wide variety of antigens. Importantly, the antigen specificity is determined by three different complementarity-determining regions (CDR) on the variable region of the IG molecule. The CDR3 is the most hypervariable region since it is produced through joining of the VDJ genes (Figure 4).[65]

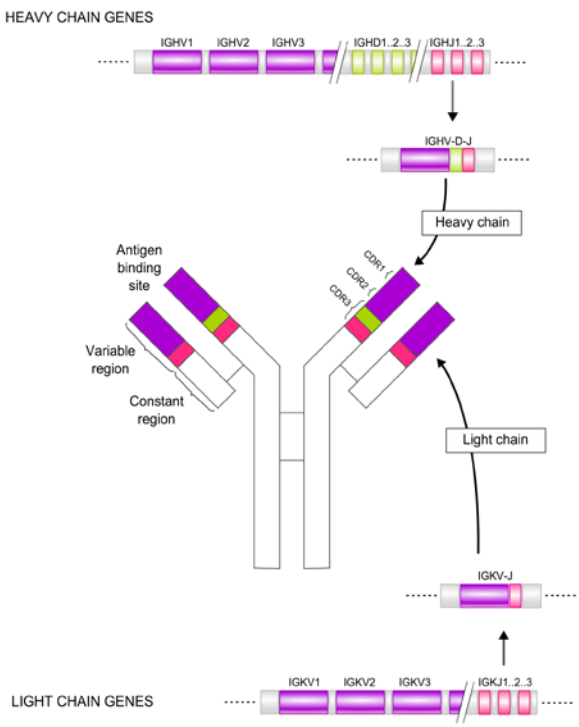


Figure 4. Structure of the antibody molecule. Upon successful IG gene rearrangement, co-expression of the heavy and light chain IG gene will result in the assembly of a functional antibody molecule. The antigen specificity is determined by CDR1, CDR2 and CDR3, where the latter is the most variable region of the IG molecule.

Assembly of the B cell receptor

In the development of B cells, the expression of functionally rearranged IG genes will give rise to the assembly of a functional BCR on the B cell surface. The process starts when pre-B cells have completed the rearrangement of the IGH locus with resulting expression of the heavy-chain Ig. The heavy-chain Ig molecule will form the pre-B cell receptor (pre-BCR) together with Igα/Igβ and a surrogate light chain (V pre-B and λ5).[66] When the subsequent light-chain IG rearrangement has occurred in the immature B-cell, co-expression of the heavy

and light-chain IG will result in assembly and cell surface-expression of a functional antibody molecule (Figure 4). The Igα/Igβ as well as other molecules such as the co-receptors CD19 and CD22 and intracellular mediators such as LYN, SYK and RAS are important for the BCR signaling pathway which is initiated upon antigen recognition.

The germinal center reaction

When a B cell interacts through its BCR with an antigen it may differentiate and give rise to plasma cells and memory cells. Depending on its nature, the antigen will initiate either a T cell dependent or T cell independent B cell maturation. T cell independent antigens are superantigens such as bacterial molecules and highly repetitious molecules which will evoke low-specificity recognition by B cells. These antigens will induce proliferation and antibody-secretion by the B cell, however without further alterations of the IG genes. In contrast, the binding to T cell dependent antigens require additional stimulus from T helper (T_H) cells and will lead to a proliferation process which proceed in germinal centers (GC) in lymphoid tissue (Figure 5).[67] During the initial step of the GC reaction, the B cell will develop into centroblasts, which undergo intense proliferation in the dark zone of the GC. At this point, the IG genes will be altered through a somatic hypermutation process giving rise to mutated B cells. This alteration involves nucleotide substitutions, insertions or deletions which will lead to antibody diversity of the proliferating B cells and transition into centrocytes. The somatic hypermutations are concentrated to the CDRs which are the antigen binding sites of the IG molecule.[68] This targeted event will either increase or decrease the affinity to the antigen which initially provoked the B cell activation. Later, these cells will move to the light zone where they compete for interaction with antigen presenting follicular dendritic cells (FDC). In this step, cells that bind with a higher affinity to the antigen will survive, whereas cells with a low affinity will die by apoptosisis. The positively selected cells receive survival signals from the T_H cells, and differentiate either to short lived antibody secreting plasma cells or long lived memory B-cells which are important for the adaptive immunity by recognizing previously experienced antigens.

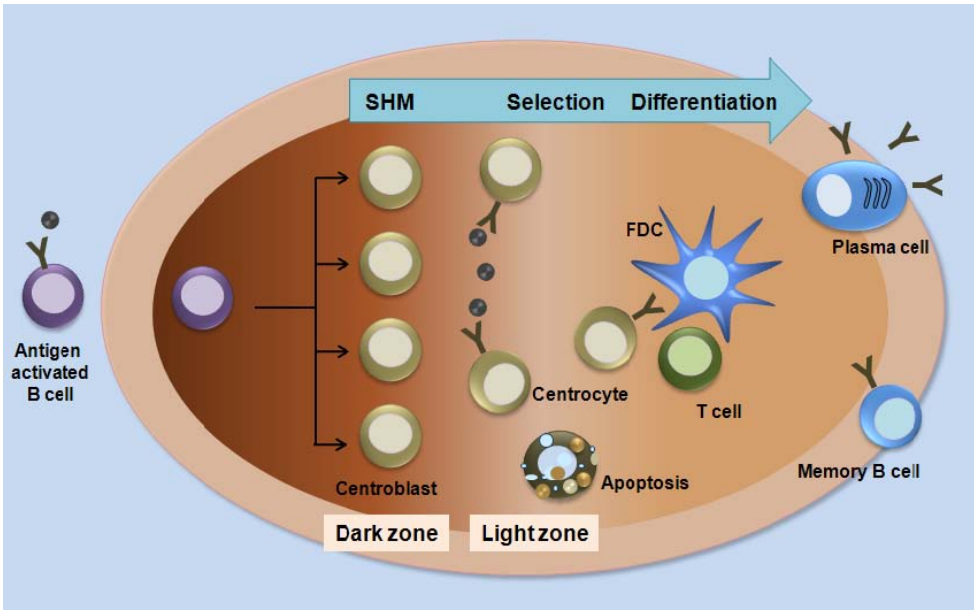


Figure 5. Germinal center reaction. Upon T cell dependent antigen stimulation a proliferation process which proceed in the germinal center. This process will lead to clonal expansion and somatic hypermutation (SHM) in the centroblasts. Positive selection of the centrocytes that have gained the highest affinity to the antigen leads to differentiation into plasma cells and memory B cells.

Immunoglobulin genes as clonal markers

B cell leukemias/lymphomas such as CLL and B cell lymphomas arise from clonal expansion of single B-cells that have passed the pre-B cell stage and have rearranged the IGH and IGL/IGK genes. In view of the fact that the malignant daughter cells will carry the same IG gene rearrangements, these can be used as clonal markers for the malignant cell population. Moreover, the IGHV mutation status in the tumor cell clone will provide useful hints about where the malignant transformation occurred, since IGHV mutated cell clones are thought to have matured through the GC reaction. For instance, several lymphomas such as Hodgkin lymphomas and follicular lymphomas are derived from GC B cells. In contrast, IGHV unmutated cell may have originated from pre-B cells such as for instance in most cases of mantle cell lymphoma.

Chronic lymphocytic leukemia

Epidemiology

CLL is the most common leukemia in the western world with a median age at diagnosis of approximately 70 years.[69] In Sweden, ~500 new patients are being diagnosed annually, and the incidence rate is nearly twice as high in men as in women.[69] It is a heterogeneous disorder and many patients are asymptomatic and follow an indolent disease course, without requirement of treatment, whilst other patients develop an aggressive and fatal disease. Some patients have a family history of CLL, and recently Houlston et. al. identified some predisposing loci.[70-72] CLL is characterized by accumulation of monoclonal B-lymphocytes (i.e. B cell lymphocytosis) with a distinct phenotype in blood, bone marrow and lymphoid organs. The criteria for diagnosis of CLL are i) presence of at least 5×10^9 B-lymphocytes/l in the peripheral blood, ii) a typical immunophenotype with expression of CD5, CD19 and CD23 and a weak to undetectable amount of monoclonal surface immunoglobulins (sIg) on the cell surface, iii) a typical blood and/or bone marrow morphology, with infiltration of small mature looking lymphocytes.[73]

Normally, CLL patients are diagnosed when examined during a routine doctor's visit. These patients may not be in need of treatment, however will be monitored by their physicians during follow-up visits. Notably, one third of all CLL patients will live a normal life with no or mild symptoms and no progression. However, approximately two thirds of all CLL patients will eventually have a progressive disease. These patients will suffer from symptoms such as anemia, thrombocytopenia, hepatosplenomegaly and lympadenopathy or immune dysfunction. The most common causes of death in the aggressive cases are bone marrow failure, immunodeficiency or infections. Moreover, transformation to aggressive lymphomas such as Richter syndrome may affect approximately 5% of patients with CLL.[74]

Treatment of chronic lymphocytic leukemia

The most important indicator for initiation of treatment of CLL is symptomatic disease, which often is reflected by a short lymphocyte doubling time (LDT), with an increase of the blood CLL cell counts with 50% within two months or a doubling within six months. Moreover, other factors such as progression in

Rai/Binet stage, with the development of cytopenia, progressive tumor burden with bulky lymphadenopathy, or symptoms such as weight loss, night sweats or fever indicate need for treatment. Despite frequent responses to initial standard treatment, CLL is considered as an incurable disease since many patients eventually become refractory, i.e. nonresponsive to treatment, or develop disease related or therapy related complications. Patients should be treated with the aim to relieve immediate symptoms, but also to avoid subsequent progression and complications. Many patients that are old benefit more from a low-dose treatment with low toxicity such as chlorambucil instead of receiving fludarabine as first-line treatment.[75] However, modern intensive combination therapy may often provide complete remission from disease, which is characterized by the disappearance of the clonal CLL B cells in blood and lymphoid organs and the restoration of the other hematopoietic cell components in response to the treatment. This is shown to result in a prolonged time to disease progression, and may also lead to a prolonged survival.

The treatment options available are chemotherapy, such as alkylating agents (chlorambucil and cyclophosphamide) and purine analogues (fludarabine), immunotherapy with monoclonal antibodies targeting CD52 (Alemtuzumab) or CD20 (Rituximab) on the CLL cell surface, and hematopoietic stem cell transplantation.[76] The first treatment option for CLL was chlorambucil, introduced in the 1950´s, and is still widely used as first-line treatment for palliative purposes, in particular in elderly patients, since it has a low toxicity. This nitrogen mustard derived alkylating agent adds an alkyl group to DNA, which inhibits the DNA replication of the proliferating CLL cells. Another alkylating agent is cyclophosphamide (C) which is now often used in combination with fludarabin (F), a purine analogue which inhibits DNA synthesis by interfering with enzymes that are essential in the DNA replication and repair processes. Because of the synergistic effects of fludarabin and cyclophosphamide, these drugs have been effectively combined for CLL therapy. Today, FC is considered as standard therapy for first-line treatment since randomized studies have shown that FC have a relatively low toxicity, and improve complete remission and overall response in comparison to fludarabin as monotherapy.[77,78] Furthermore, clinical studies comparing FC, and chemoimmunotherapy, i.e. FC with the addition of rituximab (FCR), have shown the best outcome for patients receiving the latter combination.[73] In particular, patients with high-risk disease carrying del(11q) have shown a good response to the FCR regimen and should therefore be considered for this combination. Patients that are refractory to fludarabin, such as patients with TP53 abnormalities can benefit from monotherapy with alemtuzumab.[79] Lately,

young patients with high risk disease, e.g. patients that are fludarabin-refractory or patients that have TP53 abnormalities and require treatment, may be offered allogenic hematopoietic stem cell transplantation (EBMT guidelines for stem cell transplantation), which in several studies have shown to be a potentially curable treatment option.[80]

Chronic lymphocytic leukemia kinetics

For many years, CLL was regarded as a disease caused by accumulation of non-proliferating monoclonal B cells in the blood and lymphoid tissue due to apoptotic resistance. The birth-rate of the leukemic cell clone was considered to be low and without influence on the disease. However, this view of CLL was recently revised when *in vivo* studies showed that there is a small but variable fraction of the monoclonal CLL cells that actually proliferate, and a fraction of cells that undergo apoptosis.[81] The same study showed that patients with accumulation of B cells showed a higher birth rate and a lower death rate of cells. In comparison, patients with stable B cell numbers displayed equilibrium between proliferation and apoptosis. The CLL cells are thought to divide in proliferation centers (PCs) in lymph nodes and the spleen. In these structures, CLL cells communicate with cells in the microenvironment, such as T cells and stromal cells. The interaction with these cells provide essential signals for survival and expansion of the CLL cells promoted by molecules such as IL-4 and CD40 ligation, whose effect also has been documented *in vitro*.[82,83] Hence, the active PCs are feeding the accumulating CLL compartment in blood and bone marrow by giving birth to new CLL cells.

Monoclonal B cell lymphocytosis

The diagnosis of CLL requires $>5\times10^9$ clonal B-cells with the typical phenotype per liter of blood. However, lower concentrations of such cells have recently been documented among normal people without clinical disease. Monoclonal B cell lymphocytosis (MBL) is defined as the presence of $<5\times10^9$ clonal B-cells per liter of peripheral blood, in the absence of lymphadenopathy or other associated clinical symptoms. There are three immunophenotypes of MBL, i.e. CD5 positive, with dim expression of CD20, which resemble CLL, CD5/CD20 double positive, which bear a resemblance to atypical CLL, and CD5 negative MBL. Recently, it has been shown that MBL is a rather common feature in the general population, affecting 3-5% of the general population aged over 50 years.[84,85] It is not known what causes this accumulation of clonal cells, however, environmental hazard

exposure, genetics and ageing have been suggested to influence development of MBL.[85,86] A relation between MBL and CLL has been proposed, due to the phenotypic and also genotypic similarity of some MBL with CLL, and the documented transitions from MBL to CLL. Additionally, it has been shown that first-degree relatives of CLL patients have a high incidence of MBL.[85,87] Lately, attempts have been made to link the two conditions, for instance, one study investigated the prevalence of MBL in prediagnostic CLL samples.[72] Intriguingly, the results revealed that 98% of the investigated samples had a prediagnostic B-cell clone, hence indicating that the majority of CLL is preceded by MBL. On the other hand, it must be stressed that only a minority (5%) of all MBL cases develop to CLL, since the prevalence of MBL is at least 100 times more common than the incidence of CLL in the population. This is somewhat reflected by the finding that the IG gene repertoire in MBL is different from CLL, and moreover, that MBL in contrast to CLL, seldom share similar HCDR3.[88] The frequent finding of cases with IGHV mutated and/or del(13q) in MBL may reflect that these markers are not only associated with low cell proliferation in CLL, but also in MBL. The infrequent finding of cases with unmutated IGHV genes, poor-prognostic genomic aberrations and CLL-typical IGHV genes in MBL suggest that clones with these features have a relatively fast proliferation rate already before reaching the criteria for CLL, and thus may never be detected as a precursor stage.

Pathogenesis of chronic lymphocytic leukemia

The origin of chronic lymphocytic leukemia

The cellular origin and the true normal counterpart of CLL cells are still elusive. Initially it was proposed that the CLL cells developed from pre-germinal center B cells, which was supported by the shared CD5 expression on CLL cells and naïve B cells. However, the finding that approximately 50% of CLL cases have experienced the somatic hypermutation process challenged this theory.[89] Based on the IGHV mutation status of CLL cells, patients were now divided into two disease entities, with or without somatic mutations, and it was suggested that they developed through two different pathogenetic pathways. It was believed that the mutated CLL cells were activated by antigens and matured through the GC-reaction, thus resembling normal mature B cells. In contrast, the unmutated CLL cells were thought to originate from antigen-inexperienced naïve pre-GC B cells. However, several lines of evidence have contradicted this theory. Important results based on gene expression studies have shown that unmutated and mutated

subgroups share similar gene expression profiles with differential expression of only a restricted number of genes, and both closely resembling memory B cells.[90,91] Furthermore, it has been shown that both unmutated and mutated CLL cells express surface molecules similar to antigen-activated B cells.[92] The finding that unmutated cases also share features with mature B cells have proposed that these cells also have been activated by antigens. Lately, it has been hypothesized that CLL B cells may arise from marginal zone B cells that are stimulated by T cell independent antigens.[93] The reason for this speculation is that these cells share similar characteristics such as expression of IgM antibody molecules, i.e. the cells are not isotype switched.[94-96] Moreover, CLL cells, in particular the unmutated, show a similar autoreactivity and polyreactivity as the natural antibodies produced by marginal zone B cells that become activated by T cell independent antigens.[97-100]

Genetic events in pathogenesis

Since no common genetic defect has been identified in CLL the exact pathogenetic events that lead to development of the disease is still largely unknown. However, during the last decade, extensive research in the field has led to a more detailed knowledge of biological mechanisms that may be of importance for the pathogenesis. For instance, the known recurrent aberrations, which will be further discusses in the section of prognostic markers, are detected in the majority of samples and cover the *ATM* and *TP53* genes and microRNAs *mir-15* and *mir-16*, which have been implicated to play a role in CLL leukemogenesis. Moreover, other genetic/epigenetic defects, for instance involving *MCL1, TCL1* and *DAPK1* genes have also been linked to the disease.

T cell leukemia/lymphoma 1 (*TCL1*) was identified as a specific oncogene in leukemia developing from T cells, but is now also known to be up-regulated in CLL and other B-cell malignancies.[101,102] A recent study of the function of *TCL1* showed that this gene inhibits AP-1 transcriptional activity and activates NFκB.[103,104] AP-1 and NFκB are transcription factors that are important for survival and proliferation of B cells. Thus, the *TCL1* over-expression observed in CLL may enhance these characteristics and promote carcinogenesis. Deregulation of *TCL1* in *TCL1* transgenic mice has been shown to cause a CD5-positive leukemia similar to aggressive human CLL.[105,106] The *TCL1* locus on human chromosome 14q32.1 is activated in T cell leukemias by translocations and inversions that juxtapose the *TCL1* to regulatory elements of T cell receptor genes.

However, since translocations are uncommonly detected in CLL, other regulatory pathways that modulate *TCL1* expression have been suggested for this disease. Recently, expression analysis of micro-RNAs in CLL subgroups has revealed such an alternative regulation of *TCL1* through *miR-29* and *miR-181*.[107] These micro-RNAs were identified as down-regulated in CLL patients that carry del(11q), a CLL subgroup that is known to have a high expression of *TCL1*.[108] Moreover, it was shown that these regulatory RNA molecules are homologous to the untranslated region of *TCL1* and they are able to inhibit expression of TCL1 protein expression.[107] High *TCL1* expression is correlated to unmutated IGHV genes, expression of the IGHV3-21 gene and a shorter overall survival.[109] However, it was not confirmed as an independent prognostic marker in multivariate analysis.

Up-regulation of anti-apoptotic proteins is an important trait in cancer calls because it promotes resistance toward programmed cell death.[18] The anti-apoptotic genes, B cell CLL/Lymphoma 2 (*BCL2*) and myeloid cell factor-1 (*MCL1*) genes are implicated in CLL pathogenesis. High expression of *BCL2* in CLL is thought to be mediated by deregulation of *miR-15* and *miR-16* caused by the 13q deletion.[110] The *MCL1* gene belongs to the BCL2 superfamily and is mapped to chromosome 1q21. Because Mcl-1 up-regulation has been shown to correlate with poor prognostic markers of the disease, such as mutation status of the IGHV genes and high expression of CD38 and ZAP70, this antiapoptotic molecule is thought to play a role in CLL.[111] Moreover, direct down-regulation of *MCL1* by small interfering RNA molecules has been shown to induce apoptosis in CLL cells, which further strengthen its role as an important anti-apoptotic regulator in CLL.[112]

Loss of pro-apoptotic molecules, i.e. molecules that promote induction and execution of cell death is also commonly seen in cancers. An example of a pro-apoptotic molecule is the death-associated protein kinase 1 (DAPK1), which is a positive mediator of interferon-γ induced programmed cell death and a tumor suppressor candidate.[113] In 2007, Raval et. al. showed that *DAPK1* expression is down-regulated in both sporadic and familial CLL.[114] In the familial CLL cases, it was shown that a single nucleotide change in the transcription factor binding site enhanced the binding affinity of the transcription factor HOXB2, which result in down-regulation of *DAPK1*. Hence, down-regulation of *DAPK1* was implicated as a predisposing event in familial CLL pathogenesis. In sporadic CLL cases, the mechanism behind the down-regulation of *DAPK1* was shown to involve promoter methylation, which was higher in the CLL samples as compared to normal cells.

Thus, epigenetic silencing of *DAPK1* may have an important contribution role in CLL pathogenesis.[114]

The overall DNA metylation pattern may be important in CLL pathogenesis since epigenetic changes are commonly detected in cancers i.e. global hypomethylation and regional hypermetylation. Recently, a comparison of the genome-wide epigenetic profiles in mutated and unmutated IGHV CLL patients demonstrated that certain tumor-suppressor genes were methylated, i.e. silenced, in unmutated CLL.[115] Moreover, specific genes involved in proliferation were revealed as unmethylated in IGHV unmutated samples, but metylated in IGHV mutated samples. Thus, CLL pathogenesis seems to be a combination of direct genetic defects, such as the deletion of *ATM* and *TP53,* and regulation of gene-expression by micro-RNAs and methylation.

B cell receptor stereotypy and the role of antigens in pathogenesis

Many CLL research groups have addressed the question of antigen involvement in CLL pathogenesis by investigating the IGHV and IGLV gene rearrangements, because they will make up the antigen binding site of the Ig molecule on the B cell surface. The first report revealed a biased usage of the IGHV1-69 gene usage in CLL patients.[116] Thereafter, a number of studies have confirmed that the IGHV1-69 gene is the most commonly expressed in CLL patients and that the majority of these patients have unmutated IGHV genes and are associated with poor survival.[89,117-121] Moreover, biased IGHV gene usage of IGHV3-21, IGHV4-34 and IGHV3-7 has also been reported in CLL.[89,121,122] In 2002, Rosenquist´s group identified that the second most frequent IGHV gene in their CLL patient cohort was the IGHV3-21 gene.[123] Although this gene was more frequently found in patients with mutated IGHV genes, the IGHV3-21 gene subgroup had a poor prognosis with a short median survival.[121] A geographical difference in the proportion of patients that express IGHV3-21 has been noted since these patients seem to be more common in Scandinavian countries while less common in Mediterranean countries.[122] This discrepancy may reflect different genetic and/or environmental elements with influence on CLL pathogenesis. Regarding the IG light chain gene repertoire usage in CLL, a biased usage of some IGKV genes such as the IGKV3-20, IGKV1-39 and IGKV1-5 genes has been reported.[124] Moreover, frequent IGLV genes have also been detected, which include IGLV3-21, IGLV2-8 and IGLV2-14.[124]

Since the IG light chain genes complement the IGH chain genes in the shaping of BCR specificity, the combination of these gene rearrangements will ultimately form the antigen specificity. Accordingly, CLL patients have recently been categorized into subsets according to the restricted usage of heavy and light chain genes and shared amino acid motifs of the CDR3 (Table 1).[125,126] Actually, up to 30% CLL patients share such closely similar 'quasi-identical' BCRs and currently more than 100 different subsets with 'stereotyped' BCRs have been identified. These subsets have been further investigated for possible antigen selection during CLL pathogenesis but also for their prognostic significance.

The most common subset is subset #1 which is characterized by expression of one of the IGHV1/5/7 clan genes such as IGHV1-2, IGHV1-3 or IGHV5a with strikingly similar CDR3, in combination with usage of IGKV1-39/1D-39.[127] Subset #1 patients show a worse prognosis compare to non-stereotyped subset #1 patients. Intriguingly, it has been shown that Subset #1 bind to an antigen epitope on oxidized LDL which is displayed on apoptotic cells.[128]

Approximately 50% of the CLL patients that use the IGHV3-21 gene show a homologous CDR3 and restricted expression of the IGLV3-21 lambda gene, which make up subset #2.[122,129,130] Patients included in subset #2 show an equally poor overall survival as the non-stereotyped IGHV3-21 patients, irrespective of IGHV mutation status. However, the stereotyped patients have been associated with a shorter time to progression as well as other poor prognostic markers.[127,131]

As many as 8 different subsets using the IGHV1-69 gene but expressing different IGHD, IGHJ and IG light chain genes have been identified and the majority of these subsets are unmutated.[127] IGHV1-69 utilizing patients belonging to different subsets have also been implicated to have a disparate clinical outcome. For instance, subset #3 (IGHV1-69/IGHD2-2/IGHJ6) was associated with an aggressive disease, whereas subset #5 (IGHV1-69/IGHD3-10/IGHJ6) was correlated with a more indolent disease.[127]

Table 1. Restricted usage of heavy and light chain genes and shared amino acid motifs of the HCDR3 in selected subsets.

Subset	IGHV gene	Predominant IGL/IGK gene	HCDR3 amino acid length
Subset #1	IGHV1/5/7	IGKV1-39/1D-39	13-14
Subset #2	IGHV3-21	IGLV3-21	9
Subset #3	IGHV1-69	IGKV1-39/IGKV1D-39 IGKV3-11	20/23
Subset #4	IGHV4-34	IGKV2-30	20
Subset #5	IGHV1-69	IGKV1-33/I1D-33 IGLV3-21	20
Subset #16	IGHV4-34	IGKV3-20	24

Subset #1 patients carry IGHV1-2, IGHV1-3, IGHV1-18, IGHV1-8, IGHV5a or IGHV7-4-1 genes

The IGHV4-34 gene is very frequently used in CLL and has most often undergone somatic hypermutation.[89,122,126,127] This is reflected by the fact that Ig molecules that have rearranged the IGHV4-34 gene most probably have to undergo the somatic hypermutation in order to reverse their inherent autoreactive properties.[132] Several subsets that have been identified among patients with rearranged IGHV4-34 genes such as subset #4 and subset #16.[127] Subset #4 is distinguished by simultaneous usage of the IGHV4-34 and IGKV2-30 genes and with stereotyped CDR3s (20 amino acids long).[127] Moreover, patients belonging to subset #4 have a young median age (43 years) and follow an indolent disease course when compared to non-stereotyped IGHV4-34 patients.[122,127,131] Subset #16 is an uncommon subset which expresses IGHV4-34/IGKV3-20 and displays a CDR3 consisting of 24 amino acids.

The chance that two different B cells would carry identical IG is virtually impossible. Thus, considering the evidence of biased IG gene usage and the existence of CLL subsets with closely similar 'stereotyped' BCRs, strongly implicates a role of antigen selection in the development of CLL, either by promoting pathogenesis in B cells which already carry a certain genetic lesion, or by activating and promoting the B cell to incur genetic defects that can lead to CLL.[133] Moreover, both the nature of the antigen and the response that follow upon antigen recognition may be important for the clinical outcome of CLL patients, since as discussed above, there are differences between patients expressing certain IGHV genes or belonging to specific subsets. However, it is still elusive how these clinical differences arise.

The preferential IGHV gene usage and subsets with restricted BCRs have evoked questions regarding the type of antigens which may drive leukemogenesis. Several studies have therefore investigated the antibody specificity of the BCR on CLL cells. An important finding was made in 2005, when Herve et.al. evaluated the binding properties of recombinant antibodies from both IGHV mutated and unmutated CLL B cells. This study revealed that the IGHV unmutated antibodies were highly polyreactive because they cross-reacted with intracellular structures such as DNA, insulin and LPS.[134] Moreover, it was shown that the IGHV mutated cases expressed non-polyreactive antibodies, however which gained polyreactive features upon reversion to germline. This finding suggests that at least a fraction of the IGHV mutated CLL may arise from naïve polyreactive B cells as well. Further investigations of possible targets for CLL antibodies have revealed that molecular motifs on cytoskeletal proteins, bacterial polysaccharides and oxidized lipoproteins evoke antigen-recognition.[128] Corroborating evidence that CLL cells bind to molecular motifs associated with apoptosis and oxidation has recently been provided.[135] Thus, activation and clonal expansion of CLL cells may occur under influence of bacterial infections or apoptotic cells by recognition of B cells that produce natural antibodies. Finally, a striking observation of persistent infections of Epstein-Barr virus (EBV) and cytomegalovirus (CMV) in CLL patients with stereotyped IGHV4-34 further strengthens the notion that antigens are involved in CLL leukemogenesis.[136] Taken all these observations together, it is likely that CLL arise from cells that are persistently activated by a wide range of self-antigens or pathogens that contribute to the pathogenesis of the disease.

Prognostic markers in chronic lymphocytic leukemia

The value of prognostic markers

Several factors have been associated to CLL prognosis and suggested to have a clinical implication. However, it is of high importance that these prognostic factors are thoroughly examined for their prognostic value before application in the clinical setting. It is also important to evaluate if these prognostic markers could and should be used to direct the treatment, such as directing the choice and intensity of treatment or implying transplantation. At present, only the prognostic marker del(17p), which represent ~5% of all CLL patients, is used for directing treatment, whereas the remaining prognostic markers show no clear difference in response to treatment, even though they predict overall survival.[137] Furthermore, prognostic markers are valuable when describing the distribution of patients in clinical trials such as the proportion of IGHV mutated and unmutated patients. Finally, the prognostic markers are important for our understanding of the underlying biology of the disease, and may eventually lead to targeted treatment of specific prognostic subgroups in CLL. Since the prognostic markers are valuable from many different aspects, a selection of these markers will be described in some detail below.

Clinical prognostic factors

Clinical staging systems

The Rai and Binet CLL staging systems have been applied for CLL prognostication for over 25 years.[138,139] These systems are based on physical examination of the number of enlarged lymph nodes and/or enlarged liver or spleen (i.e. lymphadenopathy, hepatomegaly and splenomegaly) and standard laboratory blood tests for evaluating anemia or thrombocytopenia. The staging systems are well established and straightforward protocols used for assessment of CLL patients into different risk groups, i.e. 0-IV for Rai and A-C for Binet. The low stages include patients with a long median survival and an indolent disease, whereas the high stages include patients with high-risk CLL. The median survival for low-risk patients (Binet stage A or Rai Stage 0) is often more than 10 years whereas intermediate-risk patients (Binet B and Rai I/II) have a median survival of 5-7 years. Patients in the high-risk group (Binet C and Rai III/IV) have a short median survival less than three years.[140] Rai and Binet staging rely on simple

clinical evaluation and blood tests that are readily available. However, these staging systems have limitations for prognostication of patients in the majority of patients that belong to the low-risk groups.[141,142]

Lymphocyte doubling time

Lymphocyte doubling time (LDT) is determined by the number of months it takes to double the absolute lymphocyte count. This marker has been proven to have an independent prognostic significance, where a shorter LDT relates to a higher proliferation rate and a more aggressive disease. Accordingly, LDT does also correlate to clinical stage and the level of marrow infiltration.[143] According to the NCI guidelines, a LDT of less than 6 months might support the initiation of treatment. Since LDT require at least two sequential measurements, it cannot be assessed at diagnosis. On the other hand, only patients with severe and obvious CLL-related manifestations should receive treatment immediately after diagnosis.

Serum markers

Lactic dehydrogenase (LD) is an enzyme that catalyzes the conversion of pyruvate to lactate. An elevated level of LD is detected in several disorders such as cancers, meningitis and HIV. The readily available laboratory test for measuring LD level indicates cell proliferation and is an essential part in the clinical staging systems for lymphoma, such as the in the International Prognostic Index.[144] LD has also shown to be useful in CLL, although not very specific. Another serum marker is thymidine kinase (TK), an enzyme that is active in dividing cells and involved in a salvage pathway for DNA synthesis. The active form of the enzyme is found in dividing cells, which makes it a marker for proliferation. In CLL, a high serum TK level has been shown to correlate with a higher proliferative activity and an advanced Rai stage, progression of disease, unmutated IGHV genes mutation status and shorter LDT.[145-147] Moreover, the TK level has also been shown to have an independent prognostic value in multivariate analysis.[148] Another molecule of prognostic value in CLL is β2 microglobulin (B2M).[149] B2M is an extracellular protein component of MHC class 1 molecules, which are present on all nucleated cells. High levels of B2M have been detected in CLL patients and in patients with monoclonal gammopathies, although the reason for high B2M concentration in serum is not known.[149,150] Elevated serum level of B2M correlates to a higher clinical stage and tumor burden in CLL.[148,150] serum markers are not specific for

CLL, and standardized methods for CLL are not readily available, but they still provide useful information in the routine management.

Genetic abnormalities

Genomic aberrations

Several genomic aberrations have been identified in CLL and some, such as trisomy 12 and complex karyotypes were established as prognostic markers in CLL already in the 1980´s.[41,59,151] By applying FISH, deletions of 11q, 13q, 17p, and trisomy 12 can be detected in approximately 80% of the patients.

The deletion of chromosome 11q is detected in ~18% of cases.[41] This aberration is correlated to a younger age and an inferior outcome.[41] Moreover, patients with del(11q) often display a bulky disease with lympadenopathy, a short LDT and a rapid disease progression.[42] The most frequently deleted region spans between chromosome bands q22 and q23, which encode several tumor suppressor genes e.g. *FDX*, *ATM*, *MLL* and *RDX*, involved in a number of hematological diseases.[42,152] The *ATM* gene, which is mutated in an autosomal recessive genetic syndrome called ataxia telangiectacia, has been implicated in the pathogenesis of CLL since it is included in the commonly deleted segment. Moreover, it has been revealed that approximately one third of CLL patients carry germline or somatic mutations of *ATM*.[153,154] In addition, some patients that display del(11q) have been shown to have a mutation on the corresponding *ATM* allele, thus corroborating the involvement of this gene in CLL pathogenesis.[155] However, it has recently been demonstrated that some cases with del(11q) do not encompass *ATM*, thus involvement of other or additional genes in CLL pathogenesis of patients carrying the deletion of 11q cannot be ruled out.[50]

Table 2. The known recurrent aberrations as prognostic markers

CNA:	del(6q)	del(11q)	del(13q)	trisomy 12	del(17p)
Candidate gene	no candidate	*ATM*	*miR- 15/16*	-	*TP53*
Prognosis	progression	poor	good	intermediate	poor

Copy-number loss of chromosome 13q is the most common aberration in CLL, affecting ~50% of patients.[41] The patients with sole deletion of 13q, which often display mutated IGHV genes, have the best prognosis.[41,151] Recently, a study on patients with homozygous and heterozygous deletions on 13q revealed that patients with either one or two allelic allelic losses had a similar outcome in terms of time to initiation of treatment and overall survival. However, this report revealed that patients with higher fraction (subclone) of cells carrying 13q deletions had an inferior outcome to patients with lower del(13q) cell fractions. Moreover, this finding was independent of mutation status or expression of CD38 and ZAP70.[156] The size of del(13q) varies between patients, as detected in sizes from <1 Mbp on the 13q14 region to involvement of the entire q arm. Moreover, both heterozygous and homozygous 13q deletions are frequently observed, the latter occurring in ~15% of the patients.[41] Initially the recurrent deletions of 13q14 suggested a role for the *RB* gene, a well established tumor suppressor gene. Further investigations of 13q14 identified *DLEU1* and *DLEU2*, but their roles have remained puzzling.[48,157,158] However, subsequent findings have revealed a role for microRNAs *miRNA-15a* and *miR-16-1* located between exon 2 and 5 of *DLEU2*, in the pathophysiology of CLL. Specifically, it was discovered that the *miRNA-15a* and *miR-16-1*, which normally downregulate the anti-apoptotic BCL2, are underexpressed in CLL.[110,159,160] Hence, the antiapoptotic feature of CLL cells in del(13q) patients may in part be caused by a reduced post-transcriptional BCL2 repression by *miRNA-15a* and *miR-16*.

Trisomy 12 is correlated to an intermediate prognosis and is marginally associated with unmutated IGHV genes.[47,161] This abnormality was reported as the first and most common recurrent aberration in CLL, through detection with conventional cytogenetic banding techniques, however is rarely seen in other leukemias.[59,151] Expression analyses of genes and proteins on chromosome 12 have been

performed in order to find candidate genes that could be involved in CLL pathogenesis, however, with inconclusive results.[162,163] More recently, the *CLLU1* gene located on chromosome 12q22 was discovered to be deregulated in CLL.[164] Hence, it was suggested as a possible candidate gene for involvement in pathogenesis in trisomy 12 patients, however, was shown to be regulated independently of trisomy 12.[164]

The deletion of 17p is the least common among the known recurrent aberrations in CLL and rarely present at diagnosis.[41] This aberration is the strongest marker of progression, especially of resistance to therapy among the recurrent aberrations, and thus confers the worst prognosis in CLL. Patients with del(17p) commonly present other poor prognostic markers such as unmutated IGHV genes and high expression of CD38 and ZAP70.[47,161] The disease aggressiveness in patients with del(17p) is reflected by several features. For instance, it is shown that these patients have an instable genome, with a higher genomic complexity than patients with other or no recurrent aberrations.[14,46,47,165] Delineation of the deletion of 17p demonstrated inclusion of the tumor suppressor gene *TP53* in the commonly deleted segment. This tumor suppressor gene is an important inducer of cell cycle arrest and apoptosis. Further investigation of this gene led to detection of *TP53* mutations, predominantly in patients del(17p) but also in a fraction of CLL patients without the deletion.[166,167] Notably, the mutations of *TP53* are more common in advanced stages of the disease, and are detected in more than 30% of patients that have developed Richter transformation.[168,169] Recent studies have shown an equally poor survival for patients with combination of *TP53* mutation and deletion, or sole mutation or deletion.[46,167,169,170] Moreover, a similar level of genomic complexity was identified in patients that only showed the TP53 mutation and patients that had del(17p).[171] Importantly, patients that carry del(17p) or have mutations in *TP53* are highly resistant to chemotherapy, such as alkalyting agents and purine analogues, and require other treatment types such as alemtuzumab or steroids.[172] Consequently, it is important that CLL patients are investigated for genomic aberrations, but also for *TP53* mutation prior to initiation of treatment.[166]

In addition to the known recurrent aberrations, the del(6q) has been identified as a progression marker in CLL as it was recurrently detected in a study of follow-up samples. This study showed that patients that acquired a deletion on chromosome 6q had a shorter interval to clinical progression compared to patients without secondary 6q abnormalities.[173] Furthermore, in a previous study del(6q) was correlated to an advanced disease, however was detected in a low number of cases.[174] In an attempt to map the deletion of 6q it was found that most deletions

mapped to 6q21, whereas a minority mapped to 6q27, thus no commonly deleted region was found for this loss.[175] At present, no candidate gene of pathogenic significance has been identified on 6q.

Telomeres

CLL patients also diverge in terms of telomere length and telomerase activity.[176-178] Telomeres are repetitive DNA sequences on the chromosome ends (TTAGGG) and are important during cell division since they prevent pairing of unmatched chromosomes and because they prevent DNA degradation leading to cell senescence. Telomerase is an enzyme that elongates the telomere sequence in dividing cells, thus preventing cell ageing and death. Telomerase activity is detected in stem cell compartments and cancer cells but generally not in adult somatic cells. Short telomeres and high telomerase activity are detected in CLL patients with a worse clinical outcome, i.e. patients with unmutated IGHV genes, high levels of CD38 and ZAP70 and in patients with del(11q) or del(17p).[177-179] Moreover, it has been shown that patients with unmutated IGHV genes and a low telomerase levels had a similar overall survival as patients with mutated IGHV genes and high telomerase levels.[179] The fact that patients with mutated IGHV genes show longer telomeres may be reflected by that normal activated B cells in the GC activate their telomerase. This will prevent cell senescence during the high cell proliferation seen when B cells undergo somatic hypermutations.[180] Hence, the difference in telomere length between mutated and unmutated patients may be due differences incurred already before the CLL transformation. Moreover, the short telomeres in unmutated CLL and telomerase activity in these cells may be due to the fact that these cells have a higher proliferation rate after transformation

Molecular and biological markers

Immunoglobulin gene mutation status

In 1998, it was shown that CLL patients can be divided into two subgroups according to the mutation status of the IGHV genes.[89] The two different groups were defined based on the homology to normal germline DNA sequences, where mutated cases show <98% identity to the germline, whereas unmutated cases have an IGHV identity that exceeds >98%. The prognostic significance of the mutation status was published back-to-back by Hamblin et al and Damle et al in 1999.[117,181]

Both groups showed that patients with mutated IGHV genes had a better outcome than patients with unmutated IGHV genes. Actually, the median overall survival was more than double in mutated patients, ~25 years, compared to ~8 years in the unmutated patients.[117] Patients with unmutated IGHV genes displayed a more aggressive disease, a higher clinical stage, the poor prognostic genomic aberrations del(11q) and del(17p) and a higher risk of relapse after transplantation.[117,181] Determination of the mutation status is a rather complicated procedure, which requires highly standardized methods for analysis of the data. Moreover, it has been revealed that some mutated patients with borderline (97-98%) germline identity, will develop a more aggressive disease than highly mutated patients.[182] Furthermore, additional factors such as usage of IGHV3-21 predict for an inferior outcome regardless of mutation status.[123] However, despite these drawbacks, the IG gene analysis is still one of the most reliable prognostic markers in CLL, widely used in laboratories and at enrollment in clinical trials.

CD38

The cellmembrane glycoprotein CD38 is expressed on B-cells among various other cell types and function as a surface receptor by interacting with CD31.[183] This surface molecule is active in cell adhesion but also in BCR signaling.[184] The expression of this molecule in over 30% of CLL cells has been shown to correlate with mutation status of the IGHV genes.[181] A high expression of CD38 was initially shown in patients with unmutated IGHV genes, and this molecule was therefore proposed as a surrogate marker for the more technically complex determination of mutational status. However, further investigation of the prognostic value of this molecule presented several complications. Both mutated and unmutated patients can have low expression of CD38, which makes it a poor surrogate marker, the expression of CD38 may change over time, and the level of CD38 expression may be differential between CLL cells in a sample. The differential CD38 expression necessitates the introduction of a cut-off for determination of CD38 positivity, which is not yet fully established.[161,185,186] Nevertheless, the change of CD38 expression in CLL may be important in a clinical point of view, since this alteration is closely correlated to aggressiveness.[185-187] This observation is thought to mirror that a higher CD38 expression and signaling promotes CLL cell proliferation and survival.[185-189]

ZAP70

The tyrosine kinase ZAP70 regulates mechanisms such as activation, cell migration and apoptosis in T-cells and NK-cells.[190] This molecule was discovered to be differentially expressed between IGHV unmutated and mutated CLL in a gene expression study performed by Rosenwald and colleagues.[91] They showed that a high ZAP70 expression level was associated with IGHV unmutated patients, and thus implicated this molecule as a surrogate marker for IGHV mutation status. This finding was further substantiated by a report which showed that ZAP70 was the best gene to use for characterization of IGHV mutation status.[191] However, discordance of ZAP70 level and mutation status have been reported in up to 25% of cases, mostly involving patients that display poor prognostic markers such as del(11q), del(17p) and usage of IGHV3-21.[192] ZAP70 is an intracellular protein that in B cells play part in BCR signaling, and it is thought that expression level of this molecule and its downstream targets can reflect the level of BCR signaling in response to antigenic stimulation, i.e. the activation status of the cell.[193] Regarding the clinical relevance of this molecule, it has been shown that the level of ZAP70 protein or mRNA expression correlates to a shorter progression free survival and an inferior outcome.[194] A favorable property of ZAP70 over CD38 is that the former have a more stable expression pattern during the course of the disease, and moreover, that ZAP70 is homogenously expressed in the CLL cells.[194,195] However, this prognostic marker also has its drawbacks, such as the lack of standardized methods, a debatable cut-off level and a simultaneous expression by T-cells. Moreover, ZAP70 has been shown to be expressed by normal B cells in different maturational stages.[196] The routine technique for measuring ZAP70 is flow-cytometry, however, since the tyrosine kinase is an intracellular protein it had been difficult to produce a standardized protocol for ZAP-70 evaluation.

RNA based markers

CLLU1

Expression of CLL upregulated gene 1 (*CLLU1*) has been shown to correlate well to prognosis in CLL. *CLLU1* was discovered by differential display screening, which showed that this gene was highly upregulated in patients with unmutated IGHV genes compared to mutated patients.[164] Subsequently, it was shown that high *CLLU1* expression correlates to a shorter time to initiation of treatment (TTT) and to high expression of CD38 and ZAP70.[197] However, this correlation is more

valid if investigated in patients younger than 70 years. Moreover, this gene has cloned to chromosome 12q22 and shown to be exclusively expressed in CLL cells, although so far, no functional protein has been discovered. Recently, it was shown that *CLLU1* mRNA exits in different splice variants, which may be differently expressed by low- and high-risk patients. The same study also revealed that *CLLU1* is equally expressed in all CLL cells of a patient, and that its expression is stable over time.[198] Since this gene appears to be disease specific, i.e. not expressed in any other tissue type or hematological diseases, it has a potential for CLL risk prediction.

LPL

Lipoprotein lipase (*LPL*) was also shown to be differentially expressed in mutated and unmutated CLL by gene expression analysis, with higher levels detected in unmutated patients.[90,91] The natural role of *LPL* is to hydrolyze circulating triacylglycerol and it plays a central role in lipid metabolism. Moreover, it has been suggested that LPL plays a role in cell adhesion, which may be a potential function in CLL.[199] A high *LPL* level correlates to a poor outcome, but also to prognostic markers such as ZAP70, CD38, LDT and genomic aberrations.[200,201] When *LPL* mRNA expression was investigated in relation to mutation status and usage of the IGHV3-21, it emerged as the strongest predictor among the investigated prognostic factors. However, it was noted that *LPL* mRNA expression level was equally high in mutated patients regardless of IGHV3-21 expression.[109] Hence, the poor prognostic, mutated IGHV3-21 patients showed the same expression level as the remainder mutated patients. Recently, an investigation of mRNA levels of *LPL*, *ZAP70*, *CLLU1*, *TCL1* and *MCL1* in CLL reported that *LPL* was strong independent markers for TTT and the best predictor for survival in multivariate analysis (unpublished data). Additionally, the LPL mRNA level can be measured directly on peripheral blood without any CLL cell separation. Taken together, *LPL* appears to be a promising prognostic marker in CLL, although its expression should be sequentially investigated in CLL patients for confirmation of the stability of mRNA expression over time.

Clonal evolution in CLL

In 1976, Novell and colleges proposed that the original cancer cell clone will acquire additional genomic defects resulting in specific cancer genotypes between patients and a mosaic of different cells within a single tumor.[202] Nowadays, clonal evolution is thought upon as a steady ongoing Darwinian process which involves natural selection in the cancer cell clone. This implies that the evolving clone often is more aggressive than the founder clone, which is exemplified by the fact that many cancers such as leukemia will develop drug resistance for therapies such as metotrexate and Imatinib, which provoke a selective pressure in the cancer cells.[203-205] Moreover, it has been demonstrated that therapies such as isotopomerase II-inhibitors, e.g., anthracycline can cause secondary AML with changes on 11q23 (*MLL* gene).

In CLL, clonal evolution has been investigated by analyzing samples derived from different time-points during the course of disease. This type of follow-up studies, with relatively long time between first and second samplings, is possible in CLL since most patients have a slow disease progression. The early studies of clonal evolution included rather few samples and applied cytogenetic banding techniques. The results pointed to that acquisition of novel aberrations is a rare feature in CLL, affecting less than a fifth of the investigated patients.[206-209] These studies also revealed that the chromosomal aberrations detected at baseline only in rare cases disappeared, and that the patients that acquired novel aberrations often progressed to a more aggressive disease. On the other hand, recent FISH studies of CLL have shown that clonal evolution is more common than originally believed.[173,210-212] This increase is probably mostly due to the switch to a more efficient technique, however the discrepancy in clonal evolution may also be affected by inclusion of different patient groups with a more or less aggressive disease. In addition, the time between the first and second sampling of the patients and prior treatment is also critical for the results detected at follow-up. Current follow-up studies have also aimed to address whether there is a difference in clonal evolution between different CLL subgroups in respect to mutation status, expression of ZAP70 and other prognostic markers.[173,210-212] A recent long term-follow up study revealed that 42% of ZAP70 positive patients developed novel aberrations, compared to the ZAP70 negative group, in which clonal evolution occurred in only 10%.[211] Another study, which included 64 patients showed that a higher incidence of clonal evolution was shown in patients with unmutated IGHV genes and underscored that these patients often acquired the poor prognostic genomic aberrations.[212] In contrast, a FISH-study on 97 CLL patients performed

by Berkova and colleagues contradicted a correlation between clonal evolution and single negative prognostic factors such as unmutated IGHV genes, CD38 positivity or high ZAP70 expression.[210] Moreover, studies of clonal evolution have detected clonal acquisition of del(6q) in follow-up samples, suggesting that this is a progression marker in CLL.[173] Interestingly, a recent microarray study applying representational oligonucleotide microarray analysis (ROMA) compared CD38 positive and CD38 negative cell fractions from individual patients and showed that there were differences between the two fractions.[213] This result suggests that cells which show differential expression of CD38 expression have acquired novel aberrations subsequent to the divergence of these cell populations.

SPECIFIC AIMS OF THE STUDY

The general aim of this thesis was to characterize genetic aberrations in patients with CLL by applying high-resolution microarray techniques to improve the knowledge of genome-wide alterations in this disease. More specifically the aims were as follows;

-To investigate the different microarray technologies available for screening of genomic aberrations and compare the pros and cons of these techniques (Paper I).

-To screen and identify genetic aberrations in a large number of newly diagnosed CLL patients, with samples from a population-based study and correlate the array findings to clinical data (Paper II and IV).

-To evaluate and compare genomic aberrations in CLL patients with stereotyped and non-stereotyped IGHV3-21 and IGHV4-34 BCRs (Paper III).

-To study clonal evolution by comparing sequential samples taken at diagnosis and at follow-up in patients from different prognostic subsets (Paper IV).

MATERIALS AND METHODS

Patient material

In papers I, II and IV, all CLL samples were obtained from a Scandinavian population based case-control study called SCALE (Scandinavian Lymphoma Etiology).[214] In paper I, 10 CLL samples, five with mutated and five with unmutated IGHV genes, were selected for inclusion in the comparison of microarray platforms. In Papers II and IV, 203 and 370 peripheral blood samples from newly diagnosed CLL patients, with a tumor cell content >70%, were selected for microarray analysis. Additionally, in Paper IV, 43 follow-up samples were included for investigation of clonal evolution. These samples were collected with a median of 6.7 years after diagnosis and all samples were collected from peripheral blood. In paper IV, 101 IGHV3-21 and IGHV4-34 CLL samples were obtained from collaborating institutes in France (n=16), Greece (n=25), Denmark (n=4) and Sweden (n=57). All samples showed a typical immunophenotype and were diagnosed according to the recently revised criteria for CLL.[73] Clinical data was obtained from medical records.

Immunoglobulin gene analysis

In papers I-IV, DNA samples were subjected to sequence analysis in order to establish the mutational status of the IGHV genes. In all papers, samples with less than 98% homology to the germline IGHV sequences were considered as mutated. Additionally, in paper III, subsets were defined according to Stamatopolous et.al. and Murray et.al.[126,127]

Microarrays

In paper I, DNA samples were subjected to four microarray platforms; a 32K BAC- (bacterial artificial chromosome) array produced in-house, a 185K oligonucleotide array from Agilent and two differently designed SNP-(single nucleotide polymorphism) arrays containing 250K and 317K SNPs from Affymetrix and Illumina, respectively. The array-experiments were performed according to the manufacturers' recommendations. For BAC and Agilent arrays, which are based on comparative genomic hybridization (CGH) the same genomic reference DNA was applied. For the SNP-arrays, different normalization methods

were applied, i.e. CNAT 4.1.0 with 96 CEU samples from the HapMap project for Affymetrix, and by applying BeadStudio Software (Illumina) using 120 normal samples. Throughout papers II-IV, the Affymetrix 250K Nsp SNP-array was applied. Array experiments were performed according to the standard protocols for Affymetrix GeneChip® Mapping 250K arrays (Gene Chip Mapping 500K Assay Manual (P/N 701930 Rev2.), Affymetrix Inc., Santa Clara, CA, USA). Quality control (QC), genotype calling and probe level normalization were performed in the Affymetrix GeneChip® Genotyping Analysis Software (GTYPE) 4.1. The Dynamic Model (DM) algorithm was used to perform single sample QC and genotype calls were made using BRLMM.[215] The applied QC specifies a neighbor score, which is an average of the euclidean distances between $\log_2$ratio of 5 consecutive SNPs along chromosomes. Low neighbor scores indicate a low noise level, thus a neighbor score of less than 0.4 was applied as cut-off for inclusion of samples. Copy number normalization, was performed in order to produce $\log_2$ratios, using the Copy Number Analysis Tool (CNAT) 4.0.1. Eighty-two normal samples analyzed at the Uppsala Array Platform were used as the reference set.

Detection of copy number alterations

In paper I, copy number analysis was performed by applying circular binary segmentation (CBS)[216] on normalized $\log_2$ratio data sets. However, special parameters were used for the different platforms (e.g. p-value = 0.01 for BAC and p-value = 0.05 for the remaining platforms. Log_2ratio thresholds for copy-number aberrations were ±0.2 for all platforms and segments >0.2 and <-0.2 were transformed to 1 or -1 for comparison reasons. A filter step was performed to remove unreliably detected segments according to the functional resolution of the different platforms (<25 kbp in the high density platforms and <300 kbp in the BAC platform. In order to facilitate the cross-platform comparison, copy number data was transformed into a virtual probe set consisting of ~300K probes spaced at every 10 kbp throughout the genome. Next, the defined copy-number of each element was associated to the closest virtual probe, for each sample and platform. In paper II-IV, copy number analysis was performed in Nexus Copy Number 3.0 Software (BioDiscovery, El Segundo, CA, USA) by applying a faster variant of CBS, called Rank segmentation. The copy-number analysis was performed using a p-value of $1\text{x}10^{-6}$, a $\log_2$ratio cutoff at 0.15 and 0.2 for regions >500 kbp and 500 kbp, respectively. These settings were established through validation experiments of copy-number regions in Paper II.

Detection of loss of heterozygosity

In paper I, analysis of LOH was performed applying the HMM algorithm in dChip, which is applicable for both Illumina and Affymetrix data. Sixty CEPH parents from dChip and 32 CEU parents from Illumina were used as reference data sets for Affymetrix and Illumina, respectively. Regions with LOH in >10% of the reference data were removed. Calling of LOH data were set as -1 (undefined), 0 (normal), and 1 (LOH) for both platforms. The LOH data was transformed into a virtual probe set as described in the section for detection of CNAs. In papers II-IV, we applied a recently developed method for identifying CNN-LOH which detects LOH-regions that do not result of copy-number changes.[217]

Validation of genomic aberrations and removal of copy-number variations

In paper I and II, quantitative real-time PCR was performed for validation of CNAs. In paper I, 11 genomic regions which were detected by all platforms (n = 4) or by 2-3 platforms (n = 7) were selected for validation. In paper II, 27 genomic regions were selected for validation in order to establish the optimal settings for copy-number analysis. The copy-number analysis in papers II-IV was adjusted according to the validation experiments performed in paper II results. Additionally, to exclude the regions that contain copy-number variations (CNVs), i.e. polymorphic regions which exist as genomic variants in the population and are non-cancer specific events, the data was matched to CNV regions reported by Redon et.al. and MacCarroll et.al.[218-220] CNAs that overlapped >50% with these CNVs were removed from further analysis.

Statistical analysis

In paper I, autocorrelation, which is a measure of noise level, and mean $\log_2$ratio ± SD were calculated on selected regions. In papers II-IV, significance testing for aberrant copy-number (STAC), a method for testing the significance of small CNAs across multiple array experiments, was performed in Nexus Copy Number 3.0 Software (BioDiscovery, El Segundo, CA, USA). In paper II Mann-Whitney U test was used to calculate differences of clinical variables in IGHV mutated and unmutated groups. In paper III, a Chi-square test was applied to determine any statistical differences in the frequency of the known recurrent aberrations between

studied groups. One way ANOVA test, T-test, was applied to determine any statistical differences in the number and size of CNAs between patient groups. Kaplan-Meier survival analysis and the log-rank test were applied to calculate differences in OS or TTT between different patient groups, as defined by various molecular markers. The clinical endpoints were defined as the time from diagnosis until date of last follow-up or death (OS) or as time from diagnosis to initiation of treatment (TTT). Statistical analysis was performed using the Statistica Software 8.0 (Stat Soft inc., Tulsa, OK, USA).

RESULTS AND DISCUSSION

Paper I – A comparison study of microarray platforms

The recent development of different types of microarray technologies inspired us to evaluate differences and similarities of four differently designed platforms. This rather technical approach was performed since our definitive goal was to screen a large series of CLL samples with microarrays.

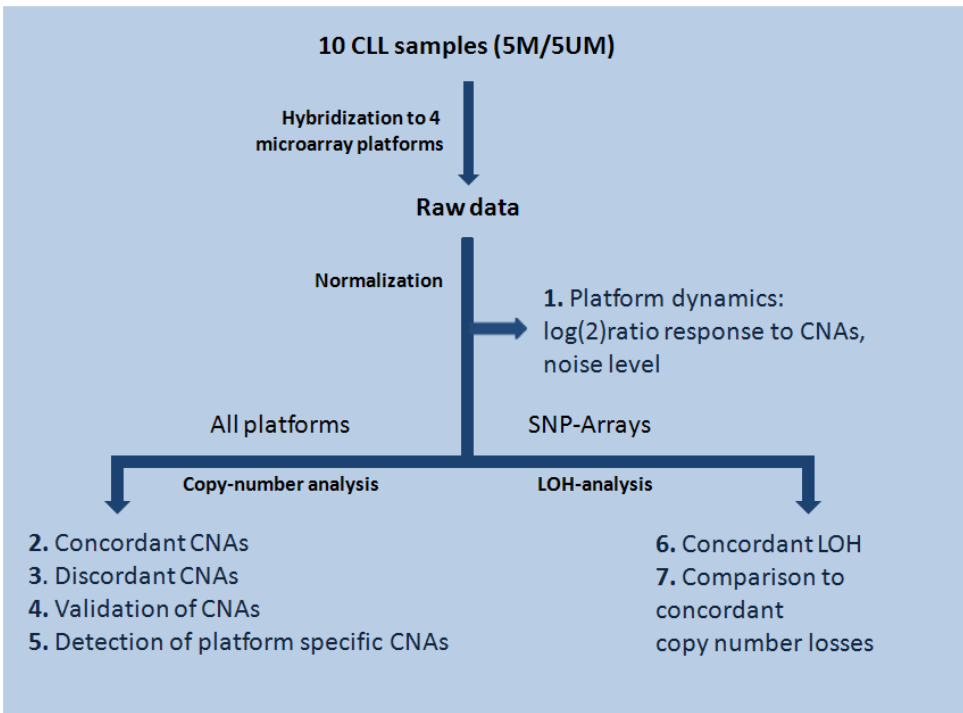


Figure 6. Outline for comparison of the differently designed microarray platforms. After hybridization of the 10 CLL samples and subsequent data-management, the technical performance was evaluated by comparing platform dynamics (1). Detection of copy-number aberrations were evaluated for all platforms (2-5) and the detection of LOH was evaluated for the SNP-arrays (6-7).

Ten CLL samples were selected for inclusion in this study, whereof five had mutated IGHV genes and the remaining five were unmutated. These 10 samples were subjected to four different microarray platforms i.e. a 32K BAC-array produced in-house, a 185K oligonucleotide-array from Agilent, and two SNP-arrays from Affymetrix (250K) and Illumina (317K), respectively (Figure 6). The raw data from the different platforms was normalized with different methods that were selected according to adequacy. In the segmentation analysis, i.e. the method for identifying chromosomal aberrations, the settings were adjusted to apt each specific platform.[221] The segmented data was transformed to a virtual probe sets in order to facilitate cross-platform comparison, thus representing four different versions of the 10 samples.

To determine the difference and similarities between the platforms, we investigated the data from different aspects, i.e. technical performance, detection of genomic aberrations, and for the SNP-arrays, detection of loss of heterozygosity. The technical performance was evaluated in three different ways. First, the average $\log_2$ratio ±SD was computed over chromosome 1 for all samples and platforms. This evaluation demonstrated a lower SD for BAC<Agilent<Affymetrix<Illumina. In addition, the technical performance was assessed by calculating the platform specific $\log_2$ratio response to FISH-validated known recurrent aberrations, i.e. deletions of 11q, 13q, 17p and trisomy 12, which were concordantly detected by the arrays. This analysis showed that Agilent had the best response to single copy deletions and gains, while the BAC platform had the best response to homozygous del(13q) (Figure 7A). Furthermore, the noise level of the platforms was investigated by calculating the autocorrelation for each platform and sample.[40] A low noise level indicates tightly correlated data, whereas a high noise level signifies a lower correlation between consecutive probes. As could be expected, the SNP-arrays showed a higher noise level than the BAC- and oligonucleotide-arrays Figure 7B). Similar results have been reported in other comparison studies.[222,223] Notably, despite the large number of probes on the oligonucleotide-array, this platform showed the lowest noise level.

A three-step analysis was undertaken to investigate genomic aberrations detected by the four platforms. Firstly, concordant aberrations, i.e. CNAs overlapping between all four platforms were identified and evaluated. Secondly, discordant aberrations that were overlapping between 2-3 platforms were assessed, and finally, platform-specific CNAs were evaluated. The evaluation of concordant aberrations showed 29 overlapping regions between the four platforms; 10 gains and 19 losses. These aberrations were mostly large and had an average size of 11.8

Mbp, trisomies excluded. Hence, the concordantly detected regions were represented by regions that, in the majority of cases, would be easily detected in all platforms by segmentation analysis. Accordingly, of the total eleven known recurrent aberrations carried by the 10 CLL patients, nine were detected by all platforms. The remainders were two heterozygous deletions of 13q14, carried by 20% and 26% of cells as detected by FISH. Notably, the deletion of 13q identified in 26% of cells was detected by the BAC-array, which has a dense coverage in this specific region. Excluding the known recurrent aberrations, concordantly detected CNAs included a trisomy 7, gain of 8q and deletions of 1p and 9q, the latter being the most narrow among the overlapping alterations, with an overlapping region of 140 kbp.

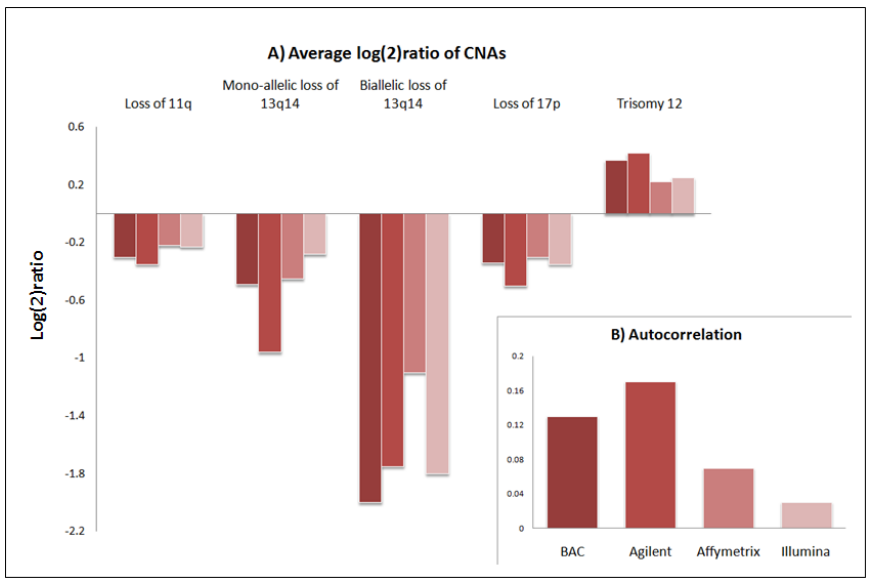


Figure 7. Platform dynamics. The platform dynamics is shown as A) log2ratio response to known recurrent aberrations and B) autocorrelation of the four different microarrays. The different platforms are indicated in dark red to light red BAC>Agilent>Affymetrix>Illumia.

We continued by investigating the aberrations detected by two or three microarrays. The 32 genomic regions that were detected by these criteria showed an average size of 0.49 Mbp, distinctly smaller than the concordantly detected regions. When each platform was evaluated it was shown that Agilent, Affymetrix and BAC detected ≥50% of these 32 CNAs, individually, however, that the Illumina platform only detected 22%. A high number of the discordantly detected regions were noted to contain genomic sequences that were either involved in IG gene rearrangements, i.e. regions on 2p11.2 and 14q32.33, or regions known to contain CNVs. Concerning the IG loci, the reason for discordance between platforms could be that the IG gene sequences are not equally covered by the different arrays. Furthermore, the differential detection of CNV containing genomic regions is affected by the coverage of arrayed elements, and more importantly, the type of reference used. The aberrations that were identified in these regions are most likely due to CNVs in the patient. The regions covering IG gene sequences and CNVs where repeatedly detected by the BAC, Agilent and Affymetrix arrays in several samples. Conversely, these regions were not even detected once by the Illumina platform, which gives an explanation to why this platform detected a lower fraction of the 32 discordant regions. Furthermore, in keeping with other comparative reports, we found that 16/32 discordant regions were not detected by the BAC-array.[40,223,224] This is probably due to the fact that the Agilent- and the two SNP-arrays have a higher probe density.

When we investigated platform specific aberrations, we found a large variation in number of platform specific aberrations. The oaCGH array detected as many as 393 specific aberrations, followed by BAC, which detected 250 CNAs. The SNP-arrays from Affymetrix and Illumina identified 76 and 22 aberrations, respectively. In all platforms, between 50-99% of these aberrations were smaller than 0.5 Mbp. Actually, in all platforms except the BAC-array, between 33-60% of the non-overlapping CNAs were smaller than 0.1 Mbp. The finding that the majority of platform-specific aberrations were small led us to speculate that these regions were either detected in the segmentation analysis due to technical variation in the data, or were detected in individual platforms due to different coverage between the arrays.

The final comparison in this study was focused to CNN-LOH in the SNP-arrays. Here we applied another method in order to detect genomic aberrations due to losses, gains or CNN-LOH. Evaluation of the results from the two SNP-arrays showed that 86 LOH regions were concordantly detected. Some of these regions were CNN-LOH, thus we confirmed the gaining of additional information by

application of SNP-arrays, as reported in other comparative array studies.[225,226] Moreover, when these regions were compared to all LOH regions detected, it was noted that only 10% of the regions smaller than 1 Mbp were concordant. In contrast, >70% of the regions larger than 4 Mbp were concordant. This is in line with our finding in the copy-number analysis, which also showed a higher concordance of large aberrations. Furthermore, we compared the regions detected in the LOH analysis with the concordant CNAs from the copy-number analysis. This comparison revealed that both SNP-platforms failed to detect several of the small aberrations, and that the Illumina platform had a higher detection rate of some of the large CNAs.

When summarizing the findings from this study our goal was to present the pros and cons for each specific platform. This was a rather difficult task since there are so many aspects of the arrays to be considered, and since some array-properties will eventuate in both positive and negative qualities. Nevertheless, we concluded that all platforms are effective tools for screening of copy-number aberrations. The BAC and Agilent arrays had a higher technical performance compared to the SNP-arrays. However, the high density of SNPs on the Affymetrix and Illumina arrays compensated for the lower technical performance. The tiling BAC-array was sensitive and robust, but did not offer a reliable detection of small alterations. The Agilent array had a good coverage and could therefore detect many non-overlapping CNAs. However, this platform did not offer LOH-analysis. The Affymetrix platform detected a higher degree of CNAs, compared to Illumina, whereas the latter showed a better performance in the LOH-analysis. Finally, we concluded that SNP-arrays are advantageous since they can provide detection of CNN-LOH.

Paper II – High density array-screening of 203 newly diagnosed CLL patients

Several CLL studies have applied microarrays for detection of CNAs, and have provided additional information on genome-wide alterations in this disease.[15,16,51-53,165,213] Our aim was to investigate genomic aberrations in newly diagnosed CLL from a Scandinavian population based case control study (SCALE). Here, we applied high density 250K SNP-array screening and performed an interim analysis of 203 CLL samples. The study was divided into four main parts; in the first part we focused on mapping the known recurrent aberrations, which are important prognostic markers in CLL.[41] In the second part we excluded the known recurrent aberrations and evaluated the remaining small and large CNAs. In the third part, we aimed to correlate the known recurrent aberrations and the genomic complexity to clinical data from the patients, while in the last part we evaluated CNN-LOH.

The copy-number analysis revealed that 90% of the samples carried CNAs. Moreover, 74% of the samples showed the known recurrent aberrations, and these aberrations constituted one-third of the total aberrations in this cohort. The deletion of 13q was the most common, detected in 54% of the patients, which is in close agreement to other CLL studies.[16,41,227] The second most common aberration was del(11q), followed by trisomy 12, detected in 13% and 11%, respectively. The deletion on 17p was only detected in 4.4%. The detection rate of the intermediate/poor-prognostic markers was lower than reported in most other studies, which reflect the population-based nature of this study.[41] Moreover, we did not detect the progression marker del(6q) in this study, probably due to the fact that our samples were taken at diagnosis.[173]

Apart from the known recurrent aberrations, additional alterations which constituted two-thirds of the total number of CNAs were detected and evaluated in the second part of this paper. These aberrations were divided into groups of <1 Mbp and >1 Mbp. The large aberrations had an average size of 20 Mbp, and the majority was non-overlapping. However, some overlapping regions were identified, such as deletions of 4p and 8p, gains of 8q and trisomy 18 and 19. Notably, the latter trisomies existed exclusively in patients with trisomy 12. In keeping with this finding, a previous report has proposed a specific mechanism for patients with trisomy 12 together with trisomy 18/19.[45] Interestingly, large recurrent gains were identified on chromosome 2p, which was always detected in combination of del(11q). Several recent studies have reported a gain on chromosome 2p, and the *MYCN* gene has been reported as the gene of interest in

this region, and this gene was recently demonstrated to be up-regulated in CLL patients with 2p gain.[16,53,56]

In CLL, the existence of small recurrent aberrations, which may contain genes associated to the pathogenesis of the disease still remain to be determined. However, in this study, with the level of resolution offered by the 250K SNP-array, we found that the majority of the small aberrations were non-overlapping. Moreover, these small aberrations were equally common in all prognostic groups, based on the known recurrent aberrations. These results indicate that the small aberrations detected in this study occur as secondary events in CLL, which may include genes with influential effect on the disease.

In the third part of this study, overall survival was investigated in relation to the presence or absence of genomic aberrations. In line with previous reports, which demonstrated the prognostic significance of the known recurrent aberrations based on survival analysis, we found that patients with del(13q) had the best survival, followed by patients with no recurrent aberration. However, one exception to the aforementioned studies was that the patients carrying the poor prognostic del(11q) had an intermediate disease course, similar to patients carrying trisomy 12. Finally, and as expected, the patients with del(17p) showed that worst survival.

In a recent CLL study performed with 50K SNP-arrays, genomic complexity (≥3 CNAs) was associated to a shorter time to first or second treatment.[58] Inspired by this concept, we investigated the time to initiation of treatment but also, more importantly, the overall survival in relation to the level of genomic complexity. Hence, the samples were grouped according to number of aberrations, by employing four different size cutoffs for alterations, i.e. all CNAs, aberrations <1 Mbp, between 1 and 5 Mbp, or >5 Mbp. By arranging the alterations accordingly, we were able to investigate the impact of increasing complexity on survival. This analysis showed that a higher number of large (>5 Mbp) aberrations correlated to a shorter time to initiation of treatment and a shorter overall survival (Figure 8). However, patients with a high complexity of large aberrations were shown to carry the poor-prognostic markers, del(17p) and del(11q). These aberrations cover the *ATM* and *TP53* genes, respectively, which have essential functions in the cell-cycle control and DNA-repair mechanisms. Impairment of their functionality on the protein level may therefore lead to accumulation of genomic abnormalities.[228,229] Consequently, the significant association between genomic complexity and poor survival was lost when patients carrying these prognostic markers were removed from the survival analysis. On the contrary, patients

without genomic complexity (<3 CNAs) often belonged to patients with del(13q) or patients without known recurrent aberrations.

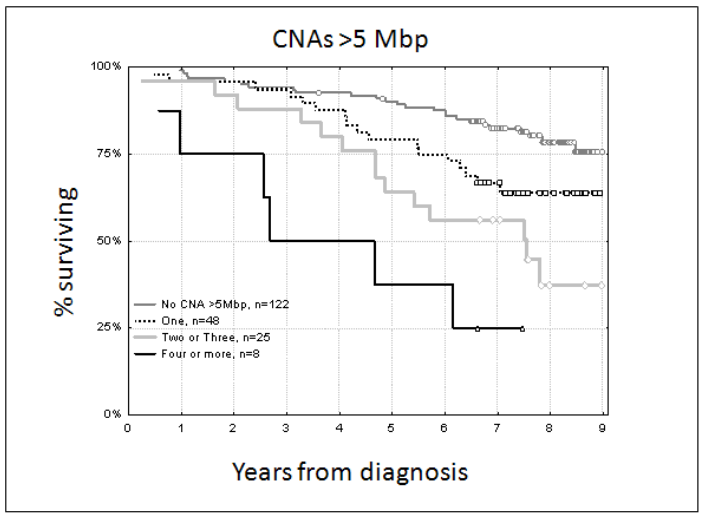


Figure 8. Overall survival in relation to number of aberrations > 5 Mbp. All 203 CLL patients were included in the analysis. P=0.00002

CNN-LOH represents genomic events without a change in copy-number. The identification of CNN-LOH was performed according to a newly developed method, taking into account the fraction of normal cells obtained from the flow-cytometry data, as described by Göransson et. al.[217] Interestingly, we found that eight patients had CNN-LOH of large parts of chromosome 13q concomitant with homozygous del(13q) (Figure 9). This combination has been reported by others, however not in such high frequency as seen in this study.[15,16] Our finding that all copy-neutral regions included a homozygous del(13q) supports the hypothesis that the loss of 13q14 is the first genetic incident in the sequence of events leading to CNN-LOH in these patients. The following event would be a proximal 13q14 double-stand break on the corresponding allele, which is repaired by the allele with the 13q deletion through mitotic recombination. This process would allow the allele which harbors the deletion to be duplicated and form CNN-LOH with an interstitial homozygous deletion. Thus, this finding is an example of how CNN-LOH can affect chromosomal regions with heterozygous deletions and

"transform" these to homozygous events, which might be an important and relatively common mechanism in other cancers as well as in CLL.

In conclusion, this SNP-array study revealed a high frequency of patients carrying CNAs, including the known recurrent aberrations. Moreover, novel recurrent combinations of aberrations were identified, i.e. the gain of chromosome 2p with concomitant del(11q) and the CNN-LOH on chromosome 13q in concurrence with homozygous del(13q). In the survival analysis we found that a high genomic complexity was correlated to worse outcome, although noted that this association was strongly linked to poor-prognostic markers.

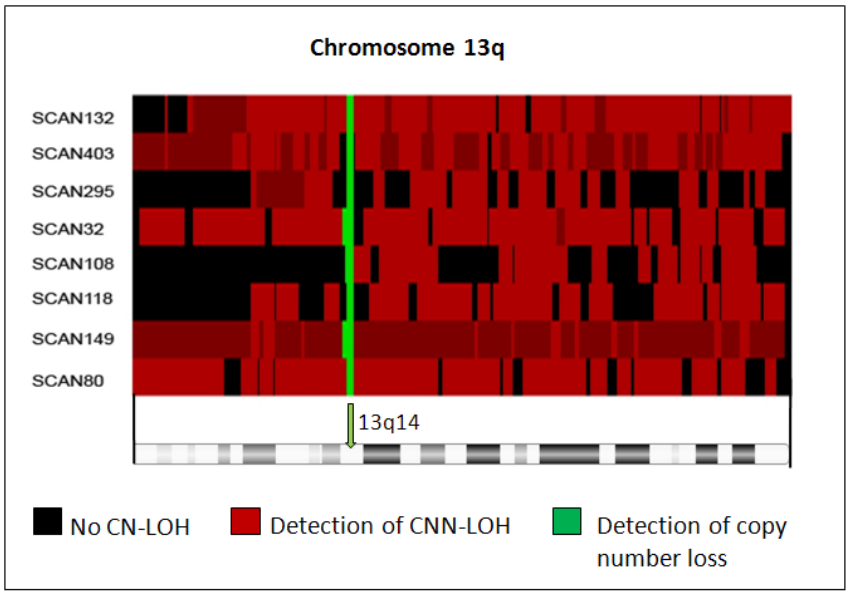


Figure 9. CNN-LOH on chromosome 13q. Eight patients, indicated with SCAN numbers, had a CNN-LOH of chromosome 13q and had a concurrent homozygous loss at 13q14.

Paper III – Genomic profiling of stereotyped and non-stereotyped CLL patients

Previous studies have shown that CLL patients display a remarkably biased IGHV gene repertoire with overrepresentation of a limited number of IGHV genes, such as IGHV1-69, IGHV4-34, IGHV3-23 and IGHV3-21.[122,123,126,127] Moreover, up to a third of the patients carry almost identical stereotyped BCRs. This stereotypy, which particularly involves the antigen-binding heavy-chain CDR3 (HCDR3), allows further identification of specific patient subsets. For instance, patients in subset #2 express the IGHV3-21 and the IGLV3-21 genes and have a 9 amino acid long CDR3.[130] Subset #2 and IGHV3-21 non-subset #2 are shown to display an equally poor survival, independent on mutation status,[129,130] although subset #2 patients appear to have a shorter time to progression.[122] Moreover stereotypic subset #4 and subset #16 are identified in patients expressing the IGHV4-34 gene. Subset #4 is distinguished by simultaneous usage of IGHV4-34 and IGKV2-30 genes with a stereotyped heavy chain HCDR3 (20 amino acids). Patients in this subset show a low median age at diagnosis and a better outcome than non-stereotyped IGHV4-34 patients.[127] Subset #16 is a rare subset which express the IGHV4-34 and IGKV3-20 genes and have a 24 amino acid long CDR3. However, little is known regarding the spectrum of genomic abnormalities in these subsets. Thus, in paper III we compared the genomic aberrations in IGHV3-21 subset #2, and IGHV4-34 subset #4 and subset #16 patients by applying 250K SNP-arrays. Non-stereotyped IGHV3-21 and IGHV4-34 patients were also included for comparison to the subset samples. This comparative study will be summarized in four sections, one describing the evaluation of the known recurrent aberrations, and the following section focusing on the level of genomic complexity, regarding the number and size of genomic alterations. The third section will bring up the analysis of CNN-LOH, and the final part will discuss the results based on the biological and clinical differences between subsets/non-subsets.

Copy-number analysis of IGHV3-21 samples showed a high prevalence of CNAs, i.e. 97% of subset #2 and 92% of non-subset #2 samples (Table 3). In comparison, 75% of subset #16, 76.5% of subset #4 and 88% of non-subset #4/16 patients displayed genomic aberrations. Moreover, when we specifically investigated the known recurrent aberrations, these were detected in as many as 90% of subset #2, however, only in 54% of non-subset #2 patients. Furthermore, these aberrations were identified in 59% of non-subset #4/16 cases, and in as little as 35% of subset #4, and 25% of subset #16. In comparison, the known recurrent aberrations were identified in 74% of the general CLL cohort presented in paper II.

Table 3. Genomic aberrations detected in the subsets in paper III and the general CLL cohort in paper II.

Subgroup	IGHV3-21		IGHV4-34			General CLL cohort
Subset (number)	Subset#2 (n=29)	Non-subset#2 (n=13)	Subset# 4 (n=17)	Subset #16 (n=8)	Non-subset #4/16 (n=34)	All (n=203)
Samples with CNAs	97%	92%	76%	75%	88%	90%
del(11q)	31%	23%	0	13%	12%	13%
del(13q)	79%	31%	35%	0	44%	54%
trisomy 12	0	23%	0	13%	6%	11%
del(17p)	0	7.7%	0	0	3%	4%

Total CNAs and known recurrent aberrations in IGHV3-21 and IGHV4-34 subsets compared to the CLL cohort investigated in paper II.

We continued by investigating the individual known recurrent aberrations, i.e. del(11q), del(13q), del(17p) and trisomy 12. Interestingly, subset #2 had the highest frequency of del(13q), which was detected in 79% in this subset compared to 31% in non-subset #2 samples. Also, the frequency of del(13q) in subset#2 was also higher compared to the frequency observed in our CLL cohort studied in paper II. Moreover, del(11q) was more common in both subset #2 and non-subset #2 samples compared to the general cohort, detected in 31%, 23% and 13% respectively. Notably, of the subset #2 samples carrying del(11q), 75% (9/12) also displayed a deletion on 13q. Furthermore, non-subset #2 patients also carried del(17p) and trisomy 12 (Table 3). In the evaluation of IGHV4-34 samples, the good-prognostic marker deletion of 13q was detected as the exclusive known recurrent aberration in subset #4 (35%). This CNA was also observed as the most prevalent aberration in non-subset #4/16 (44%), it was however absent in subset

#16 samples. Instead, subset #16 samples exclusively carried the poor-prognostic markers. Thus, the most interesting result from the two first parts of paper III is that subset #2 and non-subset #2 patients, who are associated with a poor prognosis, display a higher prevalence of aberrations on chromosome 11 and 13, and that most of these patients carry CNAs.

When we evaluated the level of genomic complexity, subset #2 and non-subset #2 showed a high prevalence of samples with ≥3 CNAs (48% and 39%, respectively) compared to subset #4 and subset#16 (18% and 13%, respectively) (Figure 10). For subset #2, this feature was partly explained by the fact that many samples carried a combination of del(13q) and del(11q). However, for the non-subset #2 samples the genomic complexity was mostly due to other CNAs than the known recurrent aberrations. In subset #4, the majority of patients carried 1 aberration (53%), in subset #16 the greater part carried 2 aberrations (50%), whereas in the non-subset #4/16 patients, there was a more even distribution of samples with 1, 2 or ≥3 CNAs (35%, 23%, 29%). The size distribution was also investigated among the subsets, which showed that subset #4 was the only patient group that did not carry aberrations >5 Mbp. This trait is partly accounted for by the fact that subset #4 patients did not carry trisomy 12 or the deletions of 11q and 17p, which often span regions larger than 5 Mbp. In addition, the average size of the del(13q) in this subset was smaller than seen in all other groups. Furthermore, subset #4, subset #16 and non-subset #4/16 had a higher prevalence of small aberrations (<1 Mbp) compared to subset #2 and non-subset #2, which showed a higher frequency of CNAs larger than >1 Mbp.

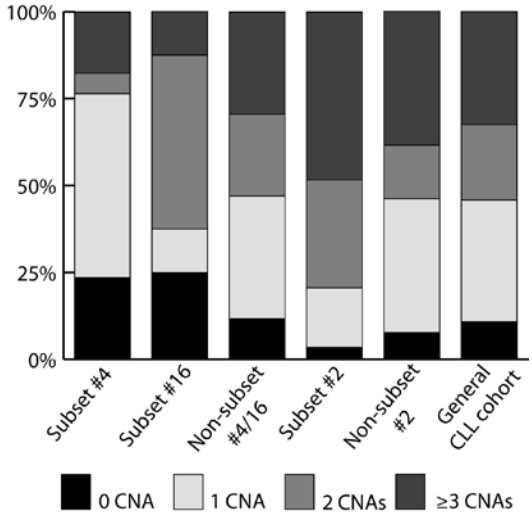


Figure 10. Level of genomic complexity. Subset #4 (n=17), subset #16 (n=8), non-subset #4/16 (n=34), subset #2 (n=29) and non-subset #4/16 (n=13).

In this array-based screening of genomic aberrations, we aimed to determine the presence of specific CNAs within the subsets. However, evaluation of the aberrations excluding the known recurrent aberrations revealed few overlapping subset-specific aberrations. For instance, two subset #2 samples with concurrent deletions of 11q and 13q displayed an overlapping gain of 2q32.1. Moreover, two additional subset #2 samples with 13q deletion demonstrated an overlapping loss of 3p21.31-3p21.1. No recurring subset-specific CNAs were observed in IGHV4-34 samples, although two samples in the non-subset #4/16 group showed a gain on 7q34. Correspondingly, additional non-recurring aberrations were identified in all subsets and non-subset groups, which may be secondary events that contribute to CLL pathogenesis.

In the evaluation of copy-neutral events, we identified five cases with CNN-LOH on chromosome 13q, which has been recently described in CLL.[14-16] In detail, large recurrent regions with CNN-LOH with sizes ranging between 68-95 Mbp were detected on the long chromosomal arm in two subset #2, one subset #4 and two non-subset #4/16 samples. In keeping with the association of del(13q) and CNN-LOH on 13q described in paper II, all five samples with this CNN-LOH

carried a homozygous loss of 13q. However, since this recurrent event was detected in both IGHV3-21 and IGHV4-34 CLL we excluded a bias towards any subset/non-subset group. Conclusively, this LOH might be of specific importance in CLL biology although not specifically in relation to stereotypy.

In summary, this study revealed a different spectrum of genomic aberrations in patients with stereotyped IGHV3-21 and IGHV4-34 BCRs. The patients belonging to subset #4, which express the IGHV4-34/IGKV2-30 genes, show a low median age at diagnosis and a more indolent disease course than non-stereotyped IGHV4-34 patients. Subset #4 showed a lower incidence of CNAs and absence of poor-prognostic aberrations, which may reflect an inherent low-proliferative disease, thus preventing accumulation of genomic aberrations. Conversely, the IGHV3-21 subset #2 and non-subset #2, which share an equally poor overall survival, showed a higher prevalence of genomic aberrations than subset #4. The frequencies of del(11q) and del(13q) were especially high in subset #2, thus these deletions may represent important genetic events during development of IGHV3-21 CLL. Furthermore, the particularly high frequency of del(11q) in subset #2 may coincide with the shorter time to progression reported for these patients.

Paper IV – SNP-array screening of diagnostic and follow-up samples in CLL

In this study, we have extended the analysis of CLL patients in paper II to include more cases from the population-based Scandinavian CLL cohort. Moreover, we have included follow-up samples for investigation of clonal evolution in CLL, i.e. the acquisition of novel aberrations over time. To date, studies on clonal evolution have only been performed by applying FISH or conventional cytogenetic techniques, with most focus on the known recurrent aberrations. Thus, application of 250K arrays here enabled evaluation of deletions, gains and copy-number neutral events in sequential samples in a detailed and genome-wide fashion. This study will be discussed in relation to paper II, with a comparison of the findings made in these related studies. Subsequently, the acquisition of novel aberrations in follow-up samples will be presented and discussed.

The principal findings from paper IV and paper II, which included 370 and 203 CLL samples, respectively, are presented in Table 4. In line with paper II, the majority of patients carried CNAs whereas 10% presented a normal karyotype. Over 70% of the samples carried between 1 and 3 CNAs, whereas only 2% were highly complex, displaying 10 or more aberrations. Thus, while genomic aberrations seem to be common in CLL, the majority of patients display a modest genomic complexity. In agreement to paper II, the majority of patients with a high genomic complexity carried the poor-prognostic markers del(11q) and del(17p). Losses were more commonly detected than gains, representing two thirds of all CNAs, partially explained by the fact that many del(13q) were detected in this cohort. Moreover, a large proportion of CNAs (42%) were smaller than 1 Mbp, whereas CNAs ranging between 1-5 Mbp and CNAs >5 Mbp constituted 29%, respectively.

Table 4. Genomic aberrations detected in paper II and IV.

Variable	Paper II (n=203)	Paper IV (n=370)
Patients with CNAs	90%	90%
Patients with known recurrent aberrations	74%	70%
del(11q)	13%	10%
del(13q)	54%	55%
del(17p)	4%	3.5%
Trisomy 12	11%	10.5%
Gain 2p	2.5 %	2.2 %
CNN-LOH on 13q	3.9%	3.5%

The known recurrent aberrations were detected in 70% of samples, which is similar to our findings in paper II. The most common aberration was deletion of 13q, carried by 203 patients (55%). Heterozygous deletion in this region was more commonly detected than homozygous deletion, 76% vs. 24%, respectively. The homozygous deletions were centered at the 13q14 region and were considerably smaller than the heterozygous losses (Figure 11). Accordingly, a high frequency of both heterozygous and homozygous deletions of 13q covered the *miR-15/16* locus, whereas the heterozygous more often also covered the *Rb* encoding region compared to the homozygous deletions (46% vs. 8%). In keeping with this present study, a recent microarray study which focused on the characterization of deletion of 13q showed that only 40% of all del(13q) covered the *Rb* gene, thus this gene may not be of great importance in pathogenesis of CLL. However, it cannot be ruled out that impairment of this gene leads to additional cellular defects in the CLL cell.[230]In this study, trisomy 12 was the second most common aberration, identified in 10.5%, closely followed by del(11q), which was detected in 10%. Deletion of 17p was noted in 4%. The deletions of 11q and 17p encompassed the *ATM* and the *TP53* encoding regions, respectively, in all cases. Loss of 6q, which is considered as a progression marker in CLL, was only detected in 2 samples

(0.05%). The prevalence of the known recurrent aberrations observed in paper II and in this extended study is reflected by the population-based nature of this study, which may present a more accurate frequency of these aberrations at early stages of the disease than previous reports which mainly contained studies from referral centers.[41,47,170,231]

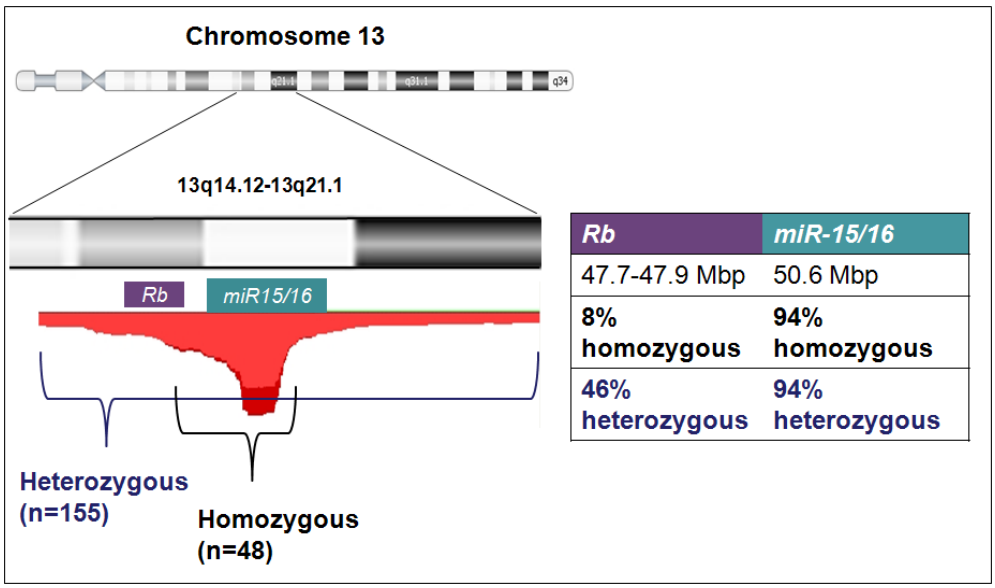


Figure 11. Distribution of homozygous and heterozygous deletions of 13q. The table indicates coverage of the Retinoblastoma gene (*Rb*) and the micro-RNAs miR-15 and miR-16 by homozygous and heterozygous deletions.

In addition to the known recurrent aberrations several other overlapping recurring regions were detected in this cohort. For instance, large gains of chromosome 2p were identified in 7 samples, on chromosome 8q in 6 samples and trisomy 18 and 19 in 3 and 5 samples, respectively. Additional deletions included loss of large part of chromosome 4p in 5 samples, loss of large parts of chromosome 8p in 6 samples and large losses of chromosome 14q in 6 samples. Although we almost doubled the number of CLL patients in this study compared to paper II, the many small aberrations that were detected were consistently non-overlapping. Evaluation of CNN-LOH in the 370 patients revealed that 13 samples had

recurrent CNN-LOH covering large parts of chromosome 13q. Of these samples, 11/13 carried a homozygous deletion of 13q14.

Furthermore, we investigated overall survival and time to treatment in relation to various prognostic markers. For instance, survival in IGHV mutated versus unmutated patient-groups was compared, confirming that unmutated patients had a significantly worse clinical outcome compared to the mutated patients. Next, survival analysis was performed according to the hierarchal model of known recurrent alterations.[41] As expected, deletion of 13q corresponded to a better survival and deletion of 17p with the worst prognosis. Additionally, in keeping with the survival analysis in paper II, del(11q) and trisomy 12 appeared to have a similar intermediate survival. Investigation of time to initiation of treatment showed that patients carrying the poor-prognostic markers received treatment at an earlier stage than patients carrying del(13q) or patients with no recurrent aberrations. Finally, to validate the impact of genomic complexity on overall survival observed in paper II, various cut-off levels for size of alterations (all, <1 Mbp, ≥1-5 Mbp and >5 Mbp) were applied to assign patients into particular groups. This analysis showed a significant association of CNAs 1-5 Mbp and a strong correlation of CNAs larger than 5 Mbp and a worse overall survival. Moreover, the investigation of time to treatment, in relation to an increasing complexity of large aberrations (>5 Mbp), showed that patients with a higher genomic complexity experienced a shorter time from diagnosis to initiation of treatment. Thus, genomic complexity was not only linked to a shorter survival, but also to a shorter time to initiation of treatment. As shown in our previous study, patients with del(11q) and del(17p) showed a high incidence of large aberrations Hence, we argue that the genomic complexity is indeed a poor prognostic marker, but that patients presenting with a complex genome most often also carry del(17p) or del(11q). This is reflected by the fact that deletions of 11q and 17p are often found in patients with higher proliferation of the CLL cells, thus more likely to acquire additional aberrations compared to the patients with an indolent disease.

Our investigation of clonal evolution included 43 samples from the population based cohort, obtained at diagnosis, and 43 samples taken at follow-up between 5-8 years after the first sample. The samples were divided into different groups depending on mutation status (18 unmutated and 25 mutated), and the mutated group was further subdivided according to treatment status (10 untreated and 15 treated). At diagnosis, 75% of the IGHV unmutated patients carried CNAs, followed by 60% and 53% of the untreated and treated IGHV mutated patients, respectively. At follow-up, 44% of the IGHV unmutated patients had developed

novel aberrations, while 27% of the IGHV mutated and treated patients showed novel aberrations. This finding suggests that clonal evolution is common in patient groups that are associated with a more aggressive disease such as unmutated IGHV genes or are in need of treatment. In contrast, none of the untreated patients in the IGHV mutated group acquired new aberrations. This is probably due to the fact that these patients have a low-proliferative disease with a low selective pressure in the proliferating CLL cells.

Deletion of 13q was the most common secondary event, acquired in 4/18 IGHV unmutated (22%) and 2/15 (13%) IGHV mutated/treated patients, either as single aberration or in combination with other CNAs. Clonal evolution did also appear as acquirement of the known recurrent aberrations del(11q) (n=2) and del(17p) (n=2). However, the acquisition of novel aberrations where not restricted to the known recurrent aberrations, in fact, clonal evolution was revealed on several other chromosomes, such as chromosomes 6, 8, 9 and 20. For instance, deletion of chromosome 6q, which is widely accepted as a progression marker in CLL, was detected in one IGHV unmutated case, which had developed a complex genome at follow-up (Figure 12).

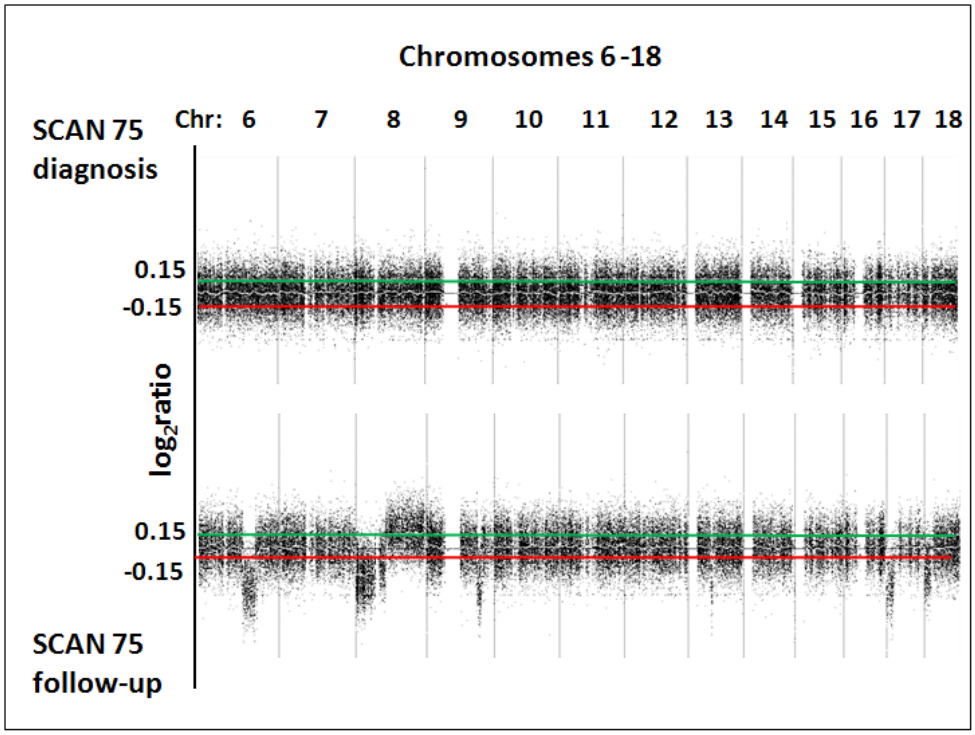


Figure 12. Genomic complexity developed over time. The upper panel shows the patient at diagnosis and the lower panel shows the same patient at follow-up.

In summary, whole-genome screening with SNP-arrays revealed a high frequency of known recurrent alterations as well as additional small and large CNAs in newly diagnosed CLL patients. The fact that the majority of alterations <1 Mbp were non-overlapping indicate that these CNAs occur as random events in early CLL cells. Future studies applying whole genome sequencing may reveal novel genetic events that are recurrent and involved in CLL pathogenesis. The follow-up study showed that clonal evolution is a common feature in CLL, particularly in the patients which had received treatment. This finding implies that the CLL cell clones are under a higher selective pressure in patients that display an aggressive disease, which leads to evolution and expansion of subclones that have acquired genetic defects.

CONCLUDING REMARKS

In this thesis we focused on characterizing genomic aberrations in CLL by applying microarrays. When investigating chromosomal aberrations from tumor samples, the detection of genomic alteration may vary depending on the technique applied. Therefore, we began our CLL microarray studies by comparing four differently designed microarray platforms by analyzing 10 CLL samples. Through this study we learned that all high-resolution techniques that we investigated readily detected the known recurrent aberrations in CLL, as well as other large aberrations. However, we also noted that the platforms detected non-overlapping, predominantly small chromosomal alterations. This discrepancy is due to many factors such as differences in technical performance, e.g. technical noise level and the ratio response to copy number, between the arrays. Moreover, the resolution and density of probes were not the same for the investigated platforms, which in some cases affected the result. Additionally, other important aspects to consider are normal references and methods for normalization and copy number data analysis, which can vary within and between microarray platforms. Accordingly, selection of reference set and fine tuning of the normalization and copy-number analysis is probably crucial for the final results when performing microarray analysis. Finally, CLL samples, although generally containing a high percentage of tumor cells, may contain subclones which are too small to be detected by the microarray technology. Thus, we concluded that all platforms were effective for screening of genomic aberrations in CLL, but that there are several features that have to be considered before selecting the optimal method.

In paper II and IV, we investigated the genomes of 203 and 370 newly diagnosed CLL patients by applying high-density SNP-arrays. The samples were collected from a population-based case control study, thus resembling an unselected cohort of CLL patients. We found that many of the CLL patients carried the known recurrent aberrations which further corroborate that these alterations are early events in CLL pathogenesis. Moreover, 90% of all cases carried copy-number aberrations, and most cases carried between 1-3 aberrations in the genome. Thus, there appears to be other genetic defects than the known recurrent aberration that are also involved in the disease development of CLL. However, apart from the known recurrent aberrations, we mostly identified genomic alterations that were non-overlapping between samples which indicate that these CNAs are probably secondary events which could contribute to the CLL pathogenesis. Furthermore, we noted that genomic complexity was a more common feature in patients with poor-prognostic markers such as deletion of 11q and 17p. In addition, it appeared

that a high genomic complexity of large aberrations predicted a poor outcome. However, since the genomic complexity and poor-prognostic markers were strongly associated, FISH directed to the known recurrent aberrations will detect most patients with genomic complexity.

In study III, we aimed to investigate genomic aberrations in CLL patient subsets based on their IGHV gene usage and the stereotypy of the BCR. In particular, we were interested in the spectrum of genomic aberrations in IGHV4-34 subset #4 patients, which show a lower median age at diagnosis and a more indolent disease than non-subset #4 patients. Moreover we investigated the level of genomic aberrations in IGHV3-21 subset #2 compared to non-subset #2, which share a similar poor prognosis. The microarray based screening showed that subset 4# cases had a lower prevalence of genome wide aberrations as well as known recurrent aberration compared to the IGHV3-21 samples and to the general CLL cohort presented in paper II. An interesting finding was that the subset #4 exclusively carried the good prognostic marker del(13q) of the known recurrent aberrations. In contrast, IGHV3-21 subset#2 and non-subset#2 showed a high proportion of patients with copy number aberrations. Furthermore, the prevalence of known recurrent aberrations was notably high in subset #2 patients, where a third of patients carried the poor prognostic deletion of 11q and 79% had a 13q deletion. Additionally, subset #2 and non-subset 2 patients showed a higher level of genomic complexity in a higher number of samples. Taken together, the microarray screening revealed a different spectrum of genomic aberrations in patients with stereotyped IGHV3-21 and IGHV4-34 BCR, which may reflect the diverging disease courses noted in these subsets.

In study IV, we investigated clonal evolution in CLL by applying microarray screening of diagnostic and follow-up samples. Acquisition of novel aberrations over time would probably occur in CLL cells which has the capacity to give rise to new CLL subclones. Moreover, in order for the new cell subclones to expand, the acquired aberration should add a proliferative advantage to the original CLL cell clone from which it derived. Thus, in our study, we investigated the development of novel CLL subclones by comparing the genomic profiles of the diagnostic and follow-up samples. With the purpose of elucidating whether patients with different mutation status of the IGHV gene also had different incidence of clonal evolution, we selected samples that had mutated IGHV genes and compared to patients with unmutated IGHV genes. We found that 44% of the IGHV unmutated patients and 27% of IGHV mutated/treated patients developed novel clones. In contrast, none of the untreated patients in the IGHV mutated group acquired new aberrations.

Taken together, our follow-up study suggests that clonal evolution is common in the patient groups that display unmutated IGHV genes, which is associated with an inferior outcome. Furthermore, those patients with mutated IGHV genes are likely to acquire novel aberrations if they progress and require treatment. Finally, patients with mutated IGHV genes who do not require treatment have a more indolent disease which is less likely to provoke clonal evolution.

In conclusion, our studies of genomic aberrations in CLL have led to a better understanding of genomic aberrations in this disease. CNAs were detected in the majority of patients where a higher number of genomic aberrations (genomic complexity) was observed in patients with poor-prognostic markers, such as the deletions of 11q and 17p. In comparison, a lower genomic complexity was noted in patients with good prognostic markers such del(13q) or patients with no recurrent aberration. Furthermore, a high prevalence of genomic aberrations was seen in poor-prognostic IGHV3-21 subset #2 patients, whereas a low occurrence of CNAs was noted for good-prognostic subset #4 patients. These results implicate that the overall genomic complexity is not only dependent on specific deleterious genetic events, such as deletions of *ATM* and *TP53* genes, but also due to other factors such as stereotypy of the BCRs and the level of antigen stimuli. Studies that specifically address the link between genomic alterations and other biological factors are needed to acquire an improved understanding of how these different features are connected and how they contribute to the disease.

ACKNOWLEDGEMENTS

I would like to express my sincere appreciation to all co-authors and collaborators that have contributed to the work included in this thesis.

A special thanks goes to:

My supervisor **Gunnar Juliusson** for letting me join the group and for encouraging me to take part in the Swedish-Danish SCAN-CLL research group. Your clinical expertise has been very helpful during my PhD training. I have truly learned a lot during the years.

Richard Rosenquist my co-supervisor, thank you for your knowledge, encouragement and your never-ending enthusiasm! For always having time to discuss the projects. Most of all, thank you for making me feel like a part of your research group and for giving me valuable input on the manuscripts and on my thesis-kappa.

Ulrike Nuber, my co-supervisor, thank you for giving me the opportunity to take part in your lab work, and for arranging the stem cell seminar series.

The Rudbeck laboratory crew: **Larry**, **Mattias**, **Nikki**, **Mi**, **Fiona**, **Meena**, **Marie**, **Lesley** and **Arifin**. Thank you for being extremely helpful, kind and for always welcoming me during my Uppsala visits. You have been a great company during the CLL conferences and meetings!

Past and present people at B10, in particular **Ariane, Jan**, **Teia**, **Ing-Britt**, **Stuart**, **Falk** and **Cosima** for making B10 enjoyable and for being a good company inside and outside the lab!

Anders Isaksson, **Hanna Göransson** and **Markus Rasmussen** at the Department of Medical Sciences, Uppsala for you support, help and expertise in the SNP-array field.

Our Danish collaborators: **Jesper Jurlander** and **Anne-Mette Buhl.**

Åke Borg and the people at Department of Oncology/Sci Blu genomics. **Johan Staaf**, thank you for sharing your great expertise in the microarray-field, for

giving me advice and for always being so helpful. **Göran**, **Johan V**, **Johan E**, **Eleonor** and **Jeanette**, thank you for being very friendly and helpful in during my first year of training in the Swegene lab

I also want to thank **Anders Malmström** and **Jens Forsberg,** in the Stem Cell Center Research Committee, for giving me the opportunity to act as a PhD representative in the Stem Cell Center Research School planning group. It has been very fun and instructive to be a part of the group and to arrange the career days and the retreat. Thank you for your trust! **Carmen** for being the best "PhD representative college"! You are a brilliant person and a very good friend! (No more surprises…;)

Kicki, **Märta**, **Stina** and **Mats** at B10 for being very kind and helpful!

Johanna W, **Katja** och **Charlotte**, för alla trevliga tjejfikor under årens lopp!

Johanna R, thank you for giving me valuable feedback on my thesis-kappa!

Mina älskade föräldrar, **Pia** och **Christer,** thank you for your endless love and for always encouraging me and trusting in my capability. Your support during these years has meant everything! You are the best parents one could ever whish for! My brothers **David** and **Johannes**, (aka Babbis and Tussen) you always cheer me up during the times we get to spend together! You are the "bästa bröderna" a sister can have! Mina morföräldrar **Haga** och **Åke**. Tack för er omtanke, uppmuntran och för att ni har intresserat er i min forskning, ni är höjdpunkten för mina hemvändarresor!

Erik First of all, thank you for taking me out on the most interesting "first date" ever and for all the moments we have spent together since then! You have been tremendously understanding, supportive and helpful during my time as a PhD student…and most important of all… you always make me laugh regardless of my emotional state! Tack för att DU finns i mitt liv!

REFERENCES

1. Colin P, Koenig P, Ouzzane A, et al. Environmental factors involved in carcinogenesis of urothelial cell carcinomas of the upper urinary tract. BJU Int. 2009;104:1436-1440.
2. Parsonnet J, Friedman GD, Vandersteen DP, et al. Helicobacter pylori infection and the risk of gastric carcinoma. N Engl J Med. 1991;325:1127-1131.
3. Rass K, Reichrath J. UV damage and DNA repair in malignant melanoma and nonmelanoma skin cancer. Adv Exp Med Biol. 2008;624:162-178.
4. zur Hausen H. Viruses in human cancers. Science. 1991;254:1167-1173.
5. Eden A, Gaudet F, Waghmare A, Jaenisch R. Chromosomal instability and tumors promoted by DNA hypomethylation. Science. 2003;300:455.
6. Iorio MV, Croce CM. MicroRNAs in cancer: small molecules with a huge impact. J Clin Oncol. 2009;27:5848-5856.
7. Boveri T. Concerning the origin of malignant tumours by Theodor Boveri. Translated and annotated by Henry Harris. J Cell Sci. 2008;121 Suppl 1:1-84.
8. Rowley JD. Letter: A new consistent chromosomal abnormality in chronic myelogenous leukaemia identified by quinacrine fluorescence and Giemsa staining. Nature. 1973;243:290-293.
9. Nowell PC, Hungerford DA. Chromosome studies on normal and leukemic human leukocytes. J Natl Cancer Inst. 1960;25:85-109.
10. Berger R, Bernheim A, Weh HJ, et al. A new translocation in Burkitt's tumor cells. Hum Genet. 1979;53:111-112.
11. Miyoshi I, Hiraki S, Kimura I, Miyamoto K, Sato J. 2/8 translocation in a Japanese Burkitt's lymphoma. Experientia. 1979;35:742-743.
12. Zech L, Haglund U, Nilsson K, Klein G. Characteristic chromosomal abnormalities in biopsies and lymphoid-cell lines from patients with Burkitt and non-Burkitt lymphomas. Int J Cancer. 1976;17:47-56.
13. Andersen CL, Wiuf C, Kruhoffer M, Korsgaard M, Laurberg S, Orntoft TF. Frequent occurrence of uniparental disomy in colorectal cancer. Carcinogenesis. 2007;28:38-48.
14. Gunnarsson R, Isaksson A, Mansouri M, et al. Large but not small copy-number alterations correlate to high-risk genomic aberrations and survival in chronic lymphocytic leukemia: a high-resolution genomic screening of newly diagnosed patients. Leukemia. 2009. Advanced online publication 10 September 2009.

15. Lehmann S, Ogawa S, Raynaud SD, et al. Molecular allelokaryotyping of early-stage, untreated chronic lymphocytic leukemia. Cancer. 2008;112:1296-1305.
16. Pfeifer D, Pantic M, Skatulla I, et al. Genome-wide analysis of DNA copy number changes and LOH in CLL using high-density SNP arrays. Blood. 2007;109:1202-1210.
17. Ross CW, Ouillette PD, Saddler CM, Shedden KA, Malek SN. Comprehensive analysis of copy number and allele status identifies multiple chromosome defects underlying follicular lymphoma pathogenesis. Clin Cancer Res. 2007;13:4777-4785.
18. Hanahan D, Weinberg RA. The hallmarks of cancer. Cell. 2000;100:57-70.
19. Liu E, Weissman B. Oncogenes and tumor suppressor genes. Cancer Treat Res. 1992;63:1-13.
20. Knudson AG, Jr. Mutation and cancer: statistical study of retinoblastoma. Proc Natl Acad Sci U S A. 1971;68:820-823.
21. Xu YQ, Grundy P, Polychronakos C. Aberrant imprinting of the insulin-like growth factor II receptor gene in Wilms' tumor. Oncogene. 1997;14:1041-1046.
22. Harris CC. p53 tumor suppressor gene: from the basic research laboratory to the clinic--an abridged historical perspective. Carcinogenesis. 1996;17:1187-1198.
23. Nordling CO. A new theory on cancer-inducing mechanism. Br J Cancer. 1953;7:68-72.
24. Hamburger AW, Salmon SE. Primary bioassay of human tumor stem cells. Science. 1977;197:461-463.
25. Park CH, Bergsagel DE, McCulloch EA. Mouse myeloma tumor stem cells: a primary cell culture assay. J Natl Cancer Inst. 1971;46:411-422.
26. Bergsagel DE, Valeriote FA. Growth characteristics of a mouse plasma cell tumor. Cancer Res. 1968;28:2187-2196.
27. McCulloch EA. Stem cells in normal and leukemic hemopoiesis (Henry Stratton Lecture, 1982). Blood. 1983;62:1-13.
28. Reya T, Morrison SJ, Clarke MF, Weissman IL. Stem cells, cancer, and cancer stem cells. Nature. 2001;414:105-111.
29. Caspersson T, Zech L, Johansson C. Differential binding of alkylating fluorochromes in human chromosomes. Exp Cell Res. 1970;60:315-319.
30. Heisterkamp N, Stam K, Groffen J, de Klein A, Grosveld G. Structural organization of the bcr gene and its role in the Ph' translocation. Nature. 1985;315:758-761.

31. Shtivelman E, Lifshitz B, Gale RP, Canaani E. Fused transcript of abl and bcr genes in chronic myelogenous leukaemia. Nature. 1985;315:550-554.
32. Pinkel D, Landegent J, Collins C, et al. Fluorescence in situ hybridization with human chromosome-specific libraries: detection of trisomy 21 and translocations of chromosome 4. Proc Natl Acad Sci U S A. 1988;85:9138-9142.
33. Schrock E, du Manoir S, Veldman T, et al. Multicolor spectral karyotyping of human chromosomes. Science. 1996;273:494-497.
34. Speicher MR, Gwyn Ballard S, Ward DC. Karyotyping human chromosomes by combinatorial multi-fluor FISH. Nat Genet. 1996;12:368-375.
35. Kallioniemi A, Kallioniemi OP, Sudar D, et al. Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors. Science. 1992;258:818-821.
36. Pinkel D, Segraves R, Sudar D, et al. High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. Nat Genet. 1998;20:207-211.
37. Solinas-Toldo S, Lampel S, Stilgenbauer S, et al. Matrix-based comparative genomic hybridization: biochips to screen for genomic imbalances. Genes Chromosomes Cancer. 1997;20:399-407.
38. Pollack JR, Perou CM, Alizadeh AA, et al. Genome-wide analysis of DNA copy-number changes using cDNA microarrays. Nat Genet. 1999;23:41-46.
39. Ylstra B, van den Ijssel P, Carvalho B, Brakenhoff RH, Meijer GA. BAC to the future! or oligonucleotides: a perspective for micro array comparative genomic hybridization (array CGH). Nucleic Acids Res. 2006;34:445-450.
40. Wicker N, Carles A, Mills IG, et al. A new look towards BAC-based array CGH through a comprehensive comparison with oligo-based array CGH. BMC Genomics. 2007;8:84.
41. Dohner H, Stilgenbauer S, Benner A, et al. Genomic aberrations and survival in chronic lymphocytic leukemia. N Engl J Med. 2000;343:1910-1916.
42. Dohner H, Stilgenbauer S, James MR, et al. 11q deletions identify a new subset of B-cell chronic lymphocytic leukemia characterized by extensive nodal involvement and inferior prognosis. Blood. 1997;89:2516-2522.
43. Fitchett M, Griffiths MJ, Oscier DG, Johnson S, Seabright M. Chromosome abnormalities involving band 13q14 in hematologic malignancies. Cancer Genet Cytogenet. 1987;24:143-150.

44. Gahrton G, Robert KH, Friberg K, Zech L, Bird AG. Nonrandom chromosomal aberrations in chronic lymphocytic leukemia revealed by polyclonal B-cell-mitogen stimulation. Blood. 1980;56:640-647.
45. Juliusson G, Merup M. Cytogenetics in chronic lymphocytic leukemia. Semin Oncol. 1998;25:19-26.
46. Dicker F, Herholz H, Schnittger S, et al. The detection of TP53 mutations in chronic lymphocytic leukemia independently predicts rapid disease progression and is highly correlated with a complex aberrant karyotype. Leukemia. 2009;23:117-124.
47. Haferlach C, Dicker F, Schnittger S, Kern W, Haferlach T. Comprehensive genetic characterization of CLL: a study on 506 cases analysed with chromosome banding analysis, interphase FISH, IgV(H) status and immunophenotyping. Leukemia. 2007;21:2442-2451.
48. Liu Y, Corcoran M, Rasool O, et al. Cloning of two candidate tumor suppressor genes within a 10 kb region on chromosome 13q14, frequently deleted in chronic lymphocytic leukemia. Oncogene. 1997;15:2463-2473.
49. Merup M, Jansson M, Corcoran M, et al. A FISH cosmid 'cocktail' for detection of 13q deletions in chronic lymphocytic leukaemia--comparison with cytogenetics and Southern hybridization. Leukemia. 1998;12:705-709.
50. Gunn SR, Hibbard MK, Ismail SH, et al. Atypical 11q deletions identified by array CGH may be missed by FISH panels for prognostic markers in chronic lymphocytic leukemia. Leukemia. 2009;23:1011-1017.
51. Gunn SR, Bolla AR, Barron LL, et al. Array CGH analysis of chronic lymphocytic leukemia reveals frequent cryptic monoallelic and biallelic deletions of chromosome 22q11 that include the PRAME gene. Leuk Res. 2008.
52. Patel A, Kang SH, Lennon PA, et al. Validation of a targeted DNA microarray for the clinical evaluation of recurrent abnormalities in chronic lymphocytic leukemia. Am J Hematol. 2008;83:540-546.
53. Schwaenen C, Nessling M, Wessendorf S, et al. Automated array-based genomic profiling in chronic lymphocytic leukemia: development of a clinical tool and discovery of recurrent genomic alterations. Proc Natl Acad Sci U S A. 2004;101:1039-1044.
54. Tyybakinoja A, Vilpo J, Knuutila S. High-resolution oligonucleotide array-CGH pinpoints genes involved in cryptic losses in chronic lymphocytic leukemia. Cytogenet Genome Res. 2007;118:8-12.

55. Fabris S, Mosca L, Todoerti K, et al. Molecular and transcriptional characterization of 17p loss in B-cell chronic lymphocytic leukemia. Genes Chromosomes Cancer. 2008;47:781-793.
56. Chapiro E, Leporrier N, Radford-Weiss I, et al. Gain of the short arm of chromosome 2 (2p) is a frequent recurring chromosome aberration in untreated chronic lymphocytic leukemia (CLL) at advanced stages. Leuk Res. 2009.
57. Gunn SR, Bolla AR, Barron LL, et al. Array CGH analysis of chronic lymphocytic leukemia reveals frequent cryptic monoallelic and biallelic deletions of chromosome 22q11 that include the PRAME gene. Leuk Res. 2009;33:1276-1281.
58. Kujawski L, Ouillette P, Erba H, et al. Genomic complexity identifies patients with aggressive chronic lymphocytic leukemia. Blood. 2008;112:1993-2003.
59. Juliusson G, Robert KH, Ost A, et al. Prognostic information from cytogenetic analysis in chronic B-lymphocytic leukemia and leukemic immunocytoma. Blood. 1985;65:134-141.
60. Ceredig R, Rolink AG, Brown G. Models of haematopoiesis: seeing the wood for the trees. Nat Rev Immunol. 2009;9:293-300.
61. Lefranc MP, Lefranc, G. The Immunoglobulin Facts Book. London UK. Academic Press. 2001.
62. Tonegawa S. Somatic generation of antibody diversity. Nature. 1983;302:575-581.
63. Desiderio SV, Yancopoulos GD, Paskind M, et al. Insertion of N regions into heavy-chain genes is correlated with expression of terminal deoxytransferase in B cells. Nature. 1984;311:752-755.
64. Grawunder U, Leu TM, Schatz DG, et al. Down-regulation of RAG1 and RAG2 gene expression in preB cells after functional immunoglobulin heavy chain rearrangement. Immunity. 1995;3:601-608.
65. VanDyk L, Meek K. Assembly of IgH CDR3: mechanism, regulation, and influence on antibody diversity. Int Rev Immunol. 1992;8:123-133.
66. Burrows PD, Stephan RP, Wang YH, Lassoued K, Zhang Z, Cooper MD. The transient expression of pre-B cell receptors governs B cell development. Semin Immunol. 2002;14:343-349.
67. Kelsoe G. The germinal center: a crucible for lymphocyte selection. Semin Immunol. 1996;8:179-184.
68. Papavasiliou FN, Schatz DG. Somatic hypermutation of immunoglobulin genes: merging mechanisms for genetic diversity. Cell. 2002;109 Suppl:S35-44.

69. Swedish Lymphoma Registry, Report for 2000-2006 URL: http://www.ocsyd.se/VP-verksamhet/Kvalitetsreg/SvenskalymfomregistretRapport2000-2006.pdf. 2000-2006:[in Swedish].
70. Goldin LR, Bjorkholm M, Kristinsson SY, Turesson I, Landgren O. Elevated risk of chronic lymphocytic leukemia and other indolent non-Hodgkin's lymphomas among relatives of patients with chronic lymphocytic leukemia. Haematologica. 2009;94:647-653.
71. Houlston RS, Sellick G, Yuille M, Matutes E, Catovsky D. Causation of chronic lymphocytic leukemia--insights from familial disease. Leuk Res. 2003;27:871-876.
72. Landgren O, Albitar M, Ma W, et al. B-cell clones as early markers for chronic lymphocytic leukemia. N Engl J Med. 2009;360:659-667.
73. Hallek M, Cheson BD, Catovsky D, et al. Guidelines for the diagnosis and treatment of chronic lymphocytic leukemia: a report from the International Workshop on Chronic Lymphocytic Leukemia updating the National Cancer Institute-Working Group 1996 guidelines. Blood. 2008;111:5446-5456.
74. Tsimberidou AM, Keating MJ. Richter syndrome: biology, incidence, and therapeutic strategies. Cancer. 2005;103:216-228.
75. Eichhorst BF, Busch R, Stilgenbauer S, et al. First-line therapy with fludarabine compared with chlorambucil does not result in a major benefit for elderly patients with advanced chronic lymphocytic leukemia. Blood. 2009;114:3382-3391.
76. Onkologiskt centrum, Södra sjukhusregionen. 2008-2010. National guidelines for investigation and treatment of chronic lymphocytic leukemia
URL:http://www.oc.gu.se/pdf/KLL_Nationella_riktlinjer_20060601.pdf
[in Swedish].
77. Catovsky D, Richards S, Matutes E, et al. Assessment of fludarabine plus cyclophosphamide for patients with chronic lymphocytic leukaemia (the LRF CLL4 Trial): a randomised controlled trial. Lancet. 2007;370:230-239.
78. Flinn IW, Neuberg DS, Grever MR, et al. Phase III trial of fludarabine plus cyclophosphamide compared with fludarabine for patients with previously untreated chronic lymphocytic leukemia: US Intergroup Trial E2997. J Clin Oncol. 2007;25:793-798.

79. Stilgenbauer S, Dohner H. Campath-1H-induced complete remission of chronic lymphocytic leukemia despite p53 gene mutation and resistance to chemotherapy. N Engl J Med. 2002;347:452-453.
80. Dreger P, Corradini P, Kimby E, et al. Indications for allogeneic stem cell transplantation in chronic lymphocytic leukemia: the EBMT transplant consensus. Leukemia. 2007;21:12-17.
81. Messmer BT, Messmer D, Allen SL, et al. In vivo measurements document the dynamic cellular kinetics of chronic lymphocytic leukemia B cells. J Clin Invest. 2005;115:755-764.
82. Buske C, Gogowski G, Schreiber K, Rave-Frank M, Hiddemann W, Wormann B. Stimulation of B-chronic lymphocytic leukemia cells by murine fibroblasts, IL-4, anti-CD40 antibodies, and the soluble CD40 ligand. Exp Hematol. 1997;25:329-337.
83. Granziero L, Ghia P, Circosta P, et al. Survivin is expressed on CD40 stimulation and interfaces proliferation and apoptosis in B-cell chronic lymphocytic leukemia. Blood. 2001;97:2777-2783.
84. Rawstron AC, Bennett FL, O'Connor SJ, et al. Monoclonal B-cell lymphocytosis and chronic lymphocytic leukemia. N Engl J Med. 2008;359:575-583.
85. Rawstron AC, Yuille MR, Fuller J, et al. Inherited predisposition to CLL is detectable as subclinical monoclonal B-lymphocyte expansion. Blood. 2002;100:2289-2290.
86. Rawstron AC. Prevalence and characteristics of monoclonal B-cell lymphocytosis (MBL) in healthy individuals and the relationship with clinical disease. J Biol Regul Homeost Agents. 2004;18:155-160.
87. Marti GE, Carter P, Abbasi F, et al. B-cell monoclonal lymphocytosis and B-cell abnormalities in the setting of familial B-cell chronic lymphocytic leukemia. Cytometry B Clin Cytom. 2003;52:1-12.
88. Dagklis A, Fazi C, Sala C, et al. The immunoglobulin gene repertoire of low-count chronic lymphocytic leukemia (CLL)-like monoclonal B lymphocytosis is different from CLL: diagnostic implications for clinical monitoring. Blood. 2009;114:26-32.
89. Fais F, Ghiotto F, Hashimoto S, et al. Chronic lymphocytic leukemia B cells express restricted sets of mutated and unmutated antigen receptors. J Clin Invest. 1998;102:1515-1525.
90. Klein U, Tu Y, Stolovitzky GA, et al. Gene expression profiling of B cell chronic lymphocytic leukemia reveals a homogeneous phenotype related to memory B cells. J Exp Med. 2001;194:1625-1638.

91. Rosenwald A, Alizadeh AA, Widhopf G, et al. Relation of gene expression phenotype to immunoglobulin mutation genotype in B cell chronic lymphocytic leukemia. J Exp Med. 2001;194:1639-1647.
92. Damle RN, Ghiotto F, Valetto A, et al. B-cell chronic lymphocytic leukemia cells express a surface membrane phenotype of activated, antigen-experienced B lymphocytes. Blood. 2002;99:4087-4093.
93. Chiorazzi N, Ferrarini M. B cell chronic lymphocytic leukemia: lessons learned from studies of the B cell antigen receptor. Annu Rev Immunol. 2003;21:841-894.
94. Kantor AB, Stall AM, Adams S, Herzenberg LA. Differential development of progenitor activity for three B-cell lineages. Proc Natl Acad Sci U S A. 1992;89:3320-3324.
95. Stall AM, Adams S, Herzenberg LA, Kantor AB. Characteristics and development of the murine B-1b (Ly-1 B sister) cell population. Ann N Y Acad Sci. 1992;651:33-43.
96. Tarlinton DM, McLean M, Nossal GJ. B1 and B2 cells differ in their potential to switch immunoglobulin isotype. Eur J Immunol. 1995;25:3388-3393.
97. Borche L, Lim A, Binet JL, Dighiero G. Evidence that chronic lymphocytic leukemia B lymphocytes are frequently committed to production of natural autoantibodies. Blood. 1990;76:562-569.
98. Broker BM, Klajman A, Youinou P, et al. Chronic lymphocytic leukemic (CLL) cells secrete multispecific autoantibodies. J Autoimmun. 1988;1:469-481.
99. Martin F, Kearney JF. Marginal-zone B cells. Nat Rev Immunol. 2002;2:323-335.
100. Sthoeger ZM, Wakai M, Tse DB, et al. Production of autoantibodies by CD5-expressing B lymphocytes from patients with chronic lymphocytic leukemia. J Exp Med. 1989;169:255-268.
101. Roos J, Hennig I, Schwaller J, et al. Expression of TCL1 in hematologic disorders. Pathobiology. 2001;69:59-66.
102. Russo G, Isobe M, Gatti R, et al. Molecular analysis of a t(14;14) translocation in leukemic T-cells of an ataxia telangiectasia patient. Proc Natl Acad Sci U S A. 1989;86:602-606.
103. Pekarsky Y, Koval A, Hallas C, et al. Tcl1 enhances Akt kinase activity and mediates its nuclear translocation. Proc Natl Acad Sci U S A. 2000;97:3028-3033.

104. Pekarsky Y, Palamarchuk A, Maximov V, et al. Tcl1 functions as a transcriptional regulator and is directly involved in the pathogenesis of CLL. Proc Natl Acad Sci U S A. 2008;105:19643-19648.
105. Bichi R, Shinton SA, Martin ES, et al. Human chronic lymphocytic leukemia modeled in mouse by targeted TCL1 expression. Proc Natl Acad Sci U S A. 2002;99:6955-6960.
106. Hoyer KK, French SW, Turner DE, et al. Dysregulated TCL1 promotes multiple classes of mature B cell lymphoma. Proc Natl Acad Sci U S A. 2002;99:14392-14397.
107. Pekarsky Y, Santanam U, Cimmino A, et al. Tcl1 expression in chronic lymphocytic leukemia is regulated by miR-29 and miR-181. Cancer Res. 2006;66:11590-11593.
108. Herling M, Patel KA, Khalili J, et al. TCL1 shows a regulated expression pattern in chronic lymphocytic leukemia that correlates with molecular subtypes and proliferative state. Leukemia. 2006;20:280-285.
109. Mansouri M, Sevov M, Fahlgren E, et al. Lipoprotein lipase is differentially expressed in prognostic subsets of chronic lymphocytic leukemia but displays invariably low catalytical activity. Leuk Res. 2009.
110. Cimmino A, Calin GA, Fabbri M, et al. miR-15 and miR-16 induce apoptosis by targeting BCL2. Proc Natl Acad Sci U S A. 2005;102:13944-13949.
111. Pepper C, Lin TT, Pratt G, et al. Mcl-1 expression has in vitro and in vivo significance in chronic lymphocytic leukemia and is associated with other poor prognostic markers. Blood. 2008;112:3807-3817.
112. Hussain SR, Cheney CM, Johnson AJ, et al. Mcl-1 is a relevant therapeutic target in acute and chronic lymphoid malignancies: down-regulation enhances rituximab-mediated apoptosis and complement-dependent cytotoxicity. Clin Cancer Res. 2007;13:2144-2150.
113. Inbal B, Shani G, Cohen O, Kissil JL, Kimchi A. Death-associated protein kinase-related protein 1, a novel serine/threonine kinase involved in apoptosis. Mol Cell Biol. 2000;20:1044-1054.
114. Raval A, Tanner SM, Byrd JC, et al. Downregulation of death-associated protein kinase 1 (DAPK1) in chronic lymphocytic leukemia. Cell. 2007;129:879-890.
115. Kanduri M, Cahill N, Goransson H, et al. Differential genome-wide array-based methylation profiles in prognostic subsets of chronic lymphocytic leukemia. Blood. 2009.
116. Kipps TJ, Tomhave E, Pratt LF, Duffy S, Chen PP, Carson DA. Developmentally restricted immunoglobulin heavy chain variable region

gene expressed at high frequency in chronic lymphocytic leukemia. Proc Natl Acad Sci U S A. 1989;86:5913-5917.
117. Hamblin TJ, Davis Z, Gardiner A, Oscier DG, Stevenson FK. Unmutated Ig V(H) genes are associated with a more aggressive form of chronic lymphocytic leukemia. Blood. 1999;94:1848-1854.
118. Johnson TA, Rassenti LZ, Kipps TJ. Ig VH1 genes expressed in B cell chronic lymphocytic leukemia exhibit distinctive molecular features. J Immunol. 1997;158:235-246.
119. Potter KN, Orchard J, Critchley E, Mockridge CI, Jose A, Stevenson FK. Features of the overexpressed V1-69 genes in the unmutated subset of chronic lymphocytic leukemia are distinct from those in the healthy elderly repertoire. Blood. 2003;101:3082-3084.
120. Schroeder HW, Jr., Dighiero G. The pathogenesis of chronic lymphocytic leukemia: analysis of the antibody repertoire. Immunol Today. 1994;15:288-294.
121. Tobin G, Soderberg O, Thunberg U, Rosenquist R. V(H)3-21 gene usage in chronic lymphocytic leukemia--characterization of a new subgroup with distinct molecular features and poor survival. Leuk Lymphoma. 2004;45:221-228.
122. Ghia P, Stamatopoulos K, Belessi C, et al. Geographic patterns and pathogenetic implications of IGHV gene usage in chronic lymphocytic leukemia: the lesson of the IGHV3-21 gene. Blood. 2005;105:1678-1685.
123. Tobin G, Thunberg U, Johnson A, et al. Somatically mutated Ig V(H)3-21 genes characterize a new subset of chronic lymphocytic leukemia. Blood. 2002;99:2262-2264.
124. Stamatopoulos K, Belessi C, Hadzidimitriou A, et al. Immunoglobulin light chain repertoire in chronic lymphocytic leukemia. Blood. 2005;106:3575-3583.
125. Ghiotto F, Fais F, Valetto A, et al. Remarkably similar antigen receptors among a subset of patients with chronic lymphocytic leukemia. J Clin Invest. 2004;113:1008-1016.
126. Murray F, Darzentas N, Hadzidimitriou A, et al. Stereotyped patterns of somatic hypermutation in subsets of patients with chronic lymphocytic leukemia: implications for the role of antigen selection in leukemogenesis. Blood. 2008;111:1524-1533.
127. Stamatopoulos K, Belessi C, Moreno C, et al. Over 20% of patients with chronic lymphocytic leukemia carry stereotyped receptors: Pathogenetic implications and clinical correlations. Blood. 2007;109:259-270.

128. Lanemo Myhrinder A, Hellqvist E, Sidorova E, et al. A new perspective: molecular motifs on oxidized LDL, apoptotic cells, and bacteria are targets for chronic lymphocytic leukemia antibodies. Blood. 2008;111:3838-3848.
129. Thorselius M, Krober A, Murray F, et al. Strikingly homologous immunoglobulin gene rearrangements and poor outcome in VH3-21-using chronic lymphocytic leukemia patients independent of geographic origin and mutational status. Blood. 2006;107:2889-2894.
130. Tobin G, Thunberg U, Johnson A, et al. Chronic lymphocytic leukemias utilizing the VH3-21 gene display highly restricted Vlambda2-14 gene use and homologous CDR3s: implicating recognition of a common antigen epitope. Blood. 2003;101:4952-4957.
131. Bomben R, Dal Bo M, Capello D, et al. Molecular and clinical features of chronic lymphocytic leukaemia with stereotyped B cell receptors: results from an Italian multicentre study. Br J Haematol. 2009;144:492-506.
132. Pugh-Bernard AE, Silverman GJ, Cappione AJ, et al. Regulation of inherently autoreactive VH4-34 B cells in the maintenance of human B cell tolerance. J Clin Invest. 2001;108:1061-1070.
133. Messmer BT, Albesiano E, Messmer D, Chiorazzi N. The pattern and distribution of immunoglobulin VH gene mutations in chronic lymphocytic leukemia B cells are consistent with the canonical somatic hypermutation process. Blood. 2004;103:3490-3495.
134. Herve M, Xu K, Ng YS, et al. Unmutated and mutated chronic lymphocytic leukemias derive from self-reactive B cell precursors despite expressing different antibody reactivity. J Clin Invest. 2005;115:1636-1643.
135. Catera R, Silverman GJ, Hatzi K, et al. Chronic lymphocytic leukemia cells recognize conserved epitopes associated with apoptosis and oxidation. Mol Med. 2008;14:665-674.
136. Kostareli E, Hadzidimitriou A, Stavroyianni N, et al. Molecular evidence for EBV and CMV persistence in a subset of patients with chronic lymphocytic leukemia expressing stereotyped IGHV4-34 B-cell receptors. Leukemia. 2009;23:919-924.
137. Byrd JC, Gribben JG, Peterson BL, et al. Select high-risk genetic features predict earlier progression following chemoimmunotherapy with fludarabine and rituximab in chronic lymphocytic leukemia: justification for risk-adapted therapy. J Clin Oncol. 2006;24:437-443.
138. Binet JL, Lepoprier M, Dighiero G, et al. A clinical staging system for chronic lymphocytic leukemia: prognostic significance. Cancer. 1977;40:855-864.

139. Rai KR, Sawitsky A, Cronkite EP, Chanana AD, Levy RN, Pasternack BS. Clinical staging of chronic lymphocytic leukemia. Blood. 1975;46:219-234.
140. Hamblin TJ. Prognostic markers in chronic lymphocytic leukaemia. Best Pract Res Clin Haematol. 2007;20:455-468.
141. Molica S, Levato D, Dattilo A. Natural history of early chronic lymphocytic leukemia. A single institution study with emphasis on the impact of disease-progression on overall survival. Haematologica. 1999;84:1094-1099.
142. Thurmes P, Call T, Slager S, et al. Comorbid conditions and survival in unselected, newly diagnosed patients with chronic lymphocytic leukemia. Leuk Lymphoma. 2008;49:49-56.
143. Montserrat E, Sanchez-Bisono J, Vinolas N, Rozman C. Lymphocyte doubling time in chronic lymphocytic leukaemia: analysis of its prognostic significance. Br J Haematol. 1986;62:567-575.
144. A predictive model for aggressive non-Hodgkin's lymphoma. The International Non-Hodgkin's Lymphoma Prognostic Factors Project. N Engl J Med. 1993;329:987-994.
145. Di Raimondo F, Giustolisi R, Lerner S, et al. Retrospective study of the prognostic role of serum thymidine kinase level in CLL patients with active disease treated with fludarabine. Ann Oncol. 2001;12:621-625.
146. Kallander CF, Simonsson B, Hagberg H, Gronowitz JS. Serum deoxythymidine kinase gives prognostic information in chronic lymphocytic leukemia. Cancer. 1984;54:2450-2455.
147. Matthews C, Catherwood MA, Morris TC, et al. Serum TK levels in CLL identify Binet stage A patients within biologically defined prognostic subgroups most likely to undergo disease progression. Eur J Haematol. 2006;77:309-317.
148. Hallek M, Wanders L, Ostwald M, et al. Serum beta(2)-microglobulin and serum thymidine kinase are independent predictors of progression-free survival in chronic lymphocytic leukemia and immunocytoma. Leuk Lymphoma. 1996;22:439-447.
149. Di Giovanni S, Valentini G, Carducci P, Giallonardo P. Beta-2-microglobulin is a reliable tumor marker in chronic lymphocytic leukemia. Acta Haematol. 1989;81:181-185.
150. Di Giovanni S, Valentini G, Ravazzolo E, Carducci P, Giallonardo P, Maschio C. Serum beta 2-microglobulin in patients with monoclonal gammopathies. Int J Biol Markers. 1987;2:169-172.

151. Juliusson G, Oscier DG, Fitchett M, et al. Prognostic subgroups in B-cell chronic lymphocytic leukemia defined by specific chromosomal abnormalities. N Engl J Med. 1990;323:720-724.
152. Stilgenbauer S, Liebisch P, James MR, et al. Molecular cytogenetic delineation of a novel critical genomic region in chromosome bands 11q22.3-923.1 in lymphoproliferative disorders. Proc Natl Acad Sci U S A. 1996;93:11837-11841.
153. Austen B, Skowronska A, Baker C, et al. Mutation status of the residual ATM allele is an important determinant of the cellular response to chemotherapy and survival in patients with chronic lymphocytic leukemia containing an 11q deletion. J Clin Oncol. 2007;25:5448-5457.
154. Bullrich F, Rasio D, Kitada S, et al. ATM mutations in B-cell chronic lymphocytic leukemia. Cancer Res. 1999;59:24-27.
155. Schaffner C, Stilgenbauer S, Rappold GA, Dohner H, Lichter P. Somatic ATM mutations indicate a pathogenic role of ATM in B-cell chronic lymphocytic leukemia. Blood. 1999;94:748-753.
156. Van Dyke DL, Shanafelt TD, Call TG, et al. A comprehensive evaluation of the prognostic significance of 13q deletions in patients with B-chronic lymphocytic leukaemia. Br J Haematol. 2009.
157. Liu Y, Grander D, Soderhall S, Juliusson G, Gahrton G, Einhorn S. Retinoblastoma gene deletions in B-cell chronic lymphocytic leukemia. Genes Chromosomes Cancer. 1992;4:250-256.
158. Liu Y, Szekely L, Grander D, et al. Chronic lymphocytic leukemia cells with allelic deletions at 13q14 commonly have one intact RB1 gene: evidence for a role of an adjacent locus. Proc Natl Acad Sci U S A. 1993;90:8697-8701.
159. Calin GA, Dumitru CD, Shimizu M, et al. Frequent deletions and down-regulation of micro- RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. Proc Natl Acad Sci U S A. 2002;99:15524-15529.
160. Kitada S, Andersen J, Akar S, et al. Expression of apoptosis-regulating proteins in chronic lymphocytic leukemia: correlations with In vitro and In vivo chemoresponses. Blood. 1998;91:3379-3389.
161. Krober A, Seiler T, Benner A, et al. V(H) mutation status, CD38 expression level, genomic aberrations, and survival in chronic lymphocytic leukemia. Blood. 2002;100:1410-1416.
162. Haslinger C, Schweifer N, Stilgenbauer S, et al. Microarray gene expression profiling of B-cell chronic lymphocytic leukemia subgroups defined by genomic aberrations and VH mutation status. J Clin Oncol. 2004;22:3937-3949.

163. Winkler D, Schneider C, Krober A, et al. Protein expression analysis of chromosome 12 candidate genes in chronic lymphocytic leukemia (CLL). Leukemia. 2005;19:1211-1215.
164. Buhl AM, Jurlander J, Jorgensen FS, et al. Identification of a gene on chromosome 12q22 uniquely overexpressed in chronic lymphocytic leukemia. Blood. 2006;107:2904-2911.
165. Forconi F, Rinaldi A, Kwee I, et al. Genome-wide DNA analysis identifies recurrent imbalances predicting outcome in chronic lymphocytic leukaemia with 17p deletion. Br J Haematol. 2008;143:532-536.
166. Malcikova J, Smardova J, Rocnova L, et al. Monoallelic and biallelic inactivation of TP53 gene in chronic lymphocytic leukemia: selection, impact on survival and response to DNA damage. Blood. 2009.
167. Zenz T, Krober A, Scherer K, et al. Monoallelic TP53 inactivation is associated with poor prognosis in chronic lymphocytic leukemia: results from a detailed genetic characterization with long-term follow-up. Blood. 2008;112:3322-3329.
168. Dohner H, Fischer K, Bentz M, et al. p53 gene deletion predicts for poor survival and non-response to therapy with purine analogs in chronic B-cell leukemias. Blood. 1995;85:1580-1589.
169. Fenaux P, Preudhomme C, Lai JL, et al. Mutations of the p53 gene in B-cell chronic lymphocytic leukemia: a report on 39 cases with cytogenetic analysis. Leukemia. 1992;6:246-250.
170. Malcikova J, Smardova J, Rocnova L, et al. Monoallelic and biallelic inactivation of TP53 gene in chronic lymphocytic leukemia: selection, impact on survival, and response to DNA damage. Blood. 2009;114:5307-5314.
171. Zenz T, Benner A, Dohner H, Stilgenbauer S. Chronic lymphocytic leukemia and treatment resistance in cancer: the role of the p53 pathway. Cell Cycle. 2008;7:3810-3814.
172. Boyd K, Dearden CE. Alemtuzumab in the treatment of chronic lymphocytic lymphoma. Expert Rev Anticancer Ther. 2008;8:525-533.
173. Finn WG, Kay NE, Kroft SH, Church S, Peterson LC. Secondary abnormalities of chromosome 6q in B-cell chronic lymphocytic leukemia: a sequential study of karyotypic instability in 51 patients. Am J Hematol. 1998;59:223-229.
174. Philip P, Geisler C, Hansen MM, et al. Aberrations of chromosome 6 in 193 newly diagnosed untreated cases of chronic lymphocytic leukemia. Cancer Genet Cytogenet. 1991;53:35-43.

175. Stilgenbauer S, Bullinger L, Benner A, et al. Incidence and clinical significance of 6q deletions in B cell chronic lymphocytic leukemia. Leukemia. 1999;13:1331-1334.
176. Bechter OE, Eisterer W, Pall G, Hilbe W, Kuhr T, Thaler J. Telomere length and telomerase activity predict survival in patients with B cell chronic lymphocytic leukemia. Cancer Res. 1998;58:4918-4922.
177. Damle RN, Batliwalla FM, Ghiotto F, et al. Telomere length and telomerase activity delineate distinctive replicative features of the B-CLL subgroups defined by immunoglobulin V gene mutations. Blood. 2004;103:375-382.
178. Roos G, Krober A, Grabowski P, et al. Short telomeres are associated with genetic complexity, high-risk genomic aberrations, and short survival in chronic lymphocytic leukemia. Blood. 2008;111:2246-2252.
179. Terrin L, Trentin L, Degan M, et al. Telomerase expression in B-cell chronic lymphocytic leukemia predicts survival and delineates subgroups of patients with the same igVH mutation status and different outcome. Leukemia. 2007;21:965-972.
180. Weng NP, Granger L, Hodes RJ. Telomere lengthening and telomerase activation during human B cell differentiation. Proc Natl Acad Sci U S A. 1997;94:10827-10832.
181. Damle RN, Wasil T, Fais F, et al. Ig V gene mutation status and CD38 expression as novel prognostic indicators in chronic lymphocytic leukemia. Blood. 1999;94:1840-1847.
182. Tobin G, Thunberg U, Laurell A, et al. Patients with chronic lymphocytic leukemia with mutated VH genes presenting with Binet stage B or C form a subgroup with a poor outcome. Haematologica. 2005;90:465-469.
183. Deaglio S, Aydin S, Vaisitti T, Bergui L, Malavasi F. CD38 at the junction between prognostic marker and therapeutic target. Trends Mol Med. 2008;14:210-218.
184. Deaglio S, Vaisitti T, Aydin S, Ferrero E, Malavasi F. In-tandem insight from basic science combined with clinical research: CD38 as both marker and key component of the pathogenetic network underlying chronic lymphocytic leukemia. Blood. 2006;108:1135-1144.
185. Chang CC, Cleveland RP. Conversion of CD38 and/or myeloid-associated marker expression status during the course of B-CLL: association with a change to an aggressive clinical course. Blood. 2002;100:1106.
186. Hamblin TJ, Orchard JA, Ibbotson RE, et al. CD38 expression and immunoglobulin variable region mutations are independent prognostic

variables in chronic lymphocytic leukemia, but CD38 expression may vary during the course of the disease. Blood. 2002;99:1023-1029.
187. Ibrahim S, Keating M, Do KA, et al. CD38 expression as an important prognostic factor in B-cell chronic lymphocytic leukemia. Blood. 2001;98:181-186.
188. Deaglio S, Vaisitti T, Aydin S, et al. CD38 and ZAP-70 are functionally linked and mark CLL cells with high migratory potential. Blood. 2007;110:4012-4021.
189. Deaglio S, Vaisitti T, Bergui L, et al. CD38 and CD100 lead a network of surface receptors relaying positive signals for B-CLL growth and survival. Blood. 2005;105:3042-3050.
190. Chan TO, Rittenhouse SE, Tsichlis PN. AKT/PKB and other D3 phosphoinositide-regulated kinases: kinase activation by phosphoinositide-dependent phosphorylation. Annu Rev Biochem. 1999;68:965-1014.
191. Wiestner A, Rosenwald A, Barry TS, et al. ZAP-70 expression identifies a chronic lymphocytic leukemia subtype with unmutated immunoglobulin genes, inferior clinical outcome, and distinct gene expression profile. Blood. 2003;101:4944-4951.
192. Krober A, Bloehdorn J, Hafner S, et al. Additional genetic high-risk features such as 11q deletion, 17p deletion, and V3-21 usage characterize discordance of ZAP-70 and VH mutation status in chronic lymphocytic leukemia. J Clin Oncol. 2006;24:969-975.
193. Efremov DG, Gobessi S, Longo PG. Signaling pathways activated by antigen-receptor engagement in chronic lymphocytic leukemia B-cells. Autoimmun Rev. 2007;7:102-108.
194. Durig J, Nuckel H, Cremer M, et al. ZAP-70 expression is a prognostic factor in chronic lymphocytic leukemia. Leukemia. 2003;17:2426-2434.
195. Crespo M, Bosch F, Villamor N, et al. ZAP-70 expression as a surrogate for immunoglobulin-variable-region mutations in chronic lymphocytic leukemia. N Engl J Med. 2003;348:1764-1775.
196. Scielzo C, Camporeale A, Geuna M, et al. ZAP-70 is expressed by normal and malignant human B-cell subsets of different maturational stage. Leukemia. 2006;20:689-695.
197. Buhl AM, Jurlander J, Geisler CH, et al. CLLU1 expression levels predict time to initiation of therapy and overall survival in chronic lymphocytic leukemia. Eur J Haematol. 2006;76:455-464.
198. Buhl AM, Novotny GW, Josefsson P, et al. The CLLU1 expression level is a stable and inherent feature of the chronic lymphocytic leukemia clone. Leukemia. 2009;23:1182-1186.

199. Mamputu JC, Renier G. Differentiation of human monocytes to monocyte-derived macrophages is associated with increased lipoprotein lipase-induced tumor necrosis factor-alpha expression and production: a process involving cell surface proteoglycans and protein kinase C. Arterioscler Thromb Vasc Biol. 1999;19:1405-1411.
200. Nuckel H, Huttmann A, Klein-Hitpass L, et al. Lipoprotein lipase expression is a novel prognostic factor in B-cell chronic lymphocytic leukemia. Leuk Lymphoma. 2006;47:1053-1061.
201. Van Bockstaele F, Pede V, Janssens A, et al. Lipoprotein lipase mRNA expression in whole blood is a prognostic marker in B cell chronic lymphocytic leukemia. Clin Chem. 2007;53:204-212.
202. Nowell PC. The clonal evolution of tumor cell populations. Science. 1976;194:23-28.
203. Carman MD, Schornagel JH, Rivest RS, et al. Resistance to methotrexate due to gene amplification in a patient with acute leukemia. J Clin Oncol. 1984;2:16-20.
204. Curt GA, Carney DN, Cowan KH, et al. Unstable methotrexate resistance in human small-cell carcinoma associated with double minute chromosomes. N Engl J Med. 1983;308:199-202.
205. Gorre ME, Ellwood-Yen K, Chiosis G, Rosen N, Sawyers CL. BCR-ABL point mutants isolated from patients with imatinib mesylate-resistant chronic myeloid leukemia remain sensitive to inhibitors of the BCR-ABL chaperone heat shock protein 90. Blood. 2002;100:3041-3044.
206. Juliusson G, Friberg K, Gahrton G. Consistency of chromosomal aberrations in chronic B-lymphocytic leukemia. A longitudinal cytogenetic study of 41 patients. Cancer. 1988;62:500-506.
207. Juliusson G, Friberg K, Gahrton G. Chromosomal aberrations in progressive and indolent chronic B-lymphocytic leukaemia. A longitudinal study. Acta Oncol. 1988;27:473-477.
208. Nowell PC, Moreau L, Growney P, Besa EC. Karyotypic stability in chronic B-cell leukemia. Cancer Genet Cytogenet. 1988;33:155-160.
209. Oscier D, Fitchett M, Herbert T, Lambert R. Karyotypic evolution in B-cell chronic lymphocytic leukaemia. Genes Chromosomes Cancer. 1991;3:16-20.
210. Berkova A, Zemanova Z, Trneny M, et al. Clonal evolution in chronic lymphocytic leukemia studied by interphase fluorescence in-situ hybridization. Neoplasma. 2009;56:455-458.

211. Shanafelt TD, Witzig TE, Fink SR, et al. Prospective evaluation of clonal evolution during long-term follow-up of patients with untreated early-stage chronic lymphocytic leukemia. J Clin Oncol. 2006;24:4634-4641.
212. Stilgenbauer S, Sander S, Bullinger L, et al. Clonal evolution in chronic lymphocytic leukemia: acquisition of high-risk genomic aberrations associated with unmutated VH, resistance to therapy, and short survival. Haematologica. 2007;92:1242-1245.
213. Grubor V, Krasnitz A, Troge JE, et al. Novel genomic alterations and clonal evolution in chronic lymphocytic leukemia revealed by representational oligonucleotide microarray analysis (ROMA). Blood. 2009;113:1294-1303.
214. Smedby KE, Hjalgrim H, Melbye M, et al. Ultraviolet radiation exposure and risk of malignant lymphomas. J Natl Cancer Inst. 2005;97:199-209.
215. Rabbee N, Speed TP. A genotype calling algorithm for affymetrix SNP arrays. Bioinformatics. 2006;22:7-12.
216. Olshen AB, Venkatraman ES, Lucito R, Wigler M. Circular binary segmentation for the analysis of array-based DNA copy number data. Biostatistics. 2004;5:557-572.
217. Goransson H, Edlund K, Rydaker M, et al. Quantification of normal cell fraction and copy number neutral LOH in clinical lung cancer samples using SNP array data. PLoS One. 2009;4:e6057.
218. McCarroll SA, Kuruvilla FG, Korn JM, et al. Integrated detection and population-genetic analysis of SNPs and copy number variation. Nat Genet. 2008;40:1166-1174.
219. Redon R, Ishikawa S, Fitch KR, et al. Global variation in copy number in the human genome. Nature. 2006;444:444-454.
220. Sebat J, Lakshmi B, Troge J, et al. Large-scale copy number polymorphism in the human genome. Science. 2004;305:525-528.
221. Venkatraman ES, Olshen AB. A faster circular binary segmentation algorithm for the analysis of array CGH data. Bioinformatics. 2007;23:657-663.
222. Greshock J, Feng B, Nogueira C, et al. A comparison of DNA copy number profiling platforms. Cancer Res. 2007;67:10173-10180.
223. Hehir-Kwa JY, Egmont-Petersen M, Janssen IM, Smeets D, van Kessel AG, Veltman JA. Genome-wide Copy Number Profiling on High-density Bacterial Artificial Chromosomes, Single-nucleotide Polymorphisms, and Oligonucleotide Microarrays: A Platform Comparison based on Statistical Power Analysis. DNA Res. 2007;14:1-11.

224. Coe BP, Ylstra B, Carvalho B, Meijer GA, Macaulay C, Lam WL. Resolving the resolution of array CGH. Genomics. 2007;89:647-653.
225. Gaasenbeek M, Howarth K, Rowan AJ, et al. Combined array-comparative genomic hybridization and single-nucleotide polymorphism-loss of heterozygosity analysis reveals complex changes and multiple forms of chromosomal instability in colorectal cancers. Cancer Res. 2006;66:3471-3479.
226. Kloth JN, Oosting J, van Wezel T, et al. Combined array-comparative genomic hybridization and single-nucleotide polymorphism-loss of heterozygosity analysis reveals complex genetic alterations in cervical cancer. BMC Genomics. 2007;8:53.
227. Gunn SR, Mohammed MS, Gorre ME, et al. Whole-genome scanning by array comparative genomic hybridization as a clinical tool for risk assessment in chronic lymphocytic leukemia. J Mol Diagn. 2008;10:442-451.
228. Hoeijmakers JH. Genome maintenance mechanisms for preventing cancer. Nature. 2001;411:366-374.
229. Shiloh Y. ATM and related protein kinases: safeguarding genome integrity. Nat Rev Cancer. 2003;3:155-168.
230. Ouillette P, Erba H, Kujawski L, Kaminski M, Shedden K, Malek SN. Integrated genomic profiling of chronic lymphocytic leukemia identifies subtypes of deletion 13q14. Cancer Res. 2008;68:1012-1021.
231. Geisler C, Jurlander J, Bullinger L, et al. Danish CLL2-Study revisited: FISH on a cohort with a 20-yr follow-up confirms the validity of the hierarchical model of genomic aberrations in chronic lymphocytic leukaemia. Eur J Haematol. 2009;83:156-158.