



UPPSALA
UNIVERSITET

*Digital Comprehensive Summaries of Uppsala Dissertations
from the Faculty of Medicine 1259*

Genomic and transcriptomic sequencing in chronic lymphocytic leukemia

DIEGO CORTESE



ACTA
UNIVERSITATIS
UPSALIENSIS
UPPSALA
2016

ISSN 1651-6206
ISBN 978-91-554-9702-6
urn:nbn:se:uu:diva-303703

Dissertation presented at Uppsala University to be publicly examined in Rudbecksalen, Dag Hammarskjölds väg 20, 75237 Uppsala University, Uppsala, Friday, 11 November 2016 at 09:15 for the degree of Doctor of Philosophy (Faculty of Medicine). The examination will be conducted in English. Faculty examiner: Professor Carlos López Otín (University of Oviedo, Spain).

Abstract

Cortese, D. 2016. Genomic and transcriptomic sequencing in chronic lymphocytic leukemia. *Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine* 1259. 63 pp. Uppsala: Acta Universitatis Upsaliensis. ISBN 978-91-554-9702-6.

Identification of recurrent mutations through next-generation sequencing (NGS) has given us a deeper understanding of the molecular mechanisms involved in chronic lymphocytic leukemia (CLL) development and progression and provided novel means for risk assessment in this clinically heterogeneous disease. In **paper I**, we screened a population-based cohort of CLL patients (n=364) for *TP53*, *NOTCH1*, *SF3B1*, *BIRC3* and *MYD88* mutations using Sanger sequencing, and confirmed the negative prognostic impact of *TP53*, *SF3B1* or *NOTCH1* aberrations, though at lower frequencies compared to previous studies. In **paper II**, we assessed the feasibility of targeted NGS using a gene panel including 9 CLL-related genes in a large patient cohort (n=188). We could validate 93% (144/155) of mutations with Sanger sequencing; the remaining were at the detection limit of the latter technique, and technical replication showed a high concordance (77/82 mutations, 94%). In **paper III**, we performed a longitudinal study of CLL patients (n=41) relapsing after fludarabine, cyclophosphamide and rituximab (FCR) therapy using whole-exome sequencing. In addition to known poor-prognostic mutations (*NOTCH1*, *TP53*, *ATM*, *SF3B1*, *BIRC3*, and *NFKBIE*), we detected mutations in a ribosomal gene, *RPS15*, in almost 20% of cases (8/41). In extended patient series, *RPS15*-mutant cases had a poor survival similar to patients with *NOTCH1*, *SF3B1*, or 11q aberrations. *In vitro* studies revealed that *RPS15*^{mut} cases displayed reduced p53 stabilization compared to cases wildtype for *RPS15*. In **paper IV**, we performed RNA-sequencing in CLL patients (n=50) assigned to 3 clinically and biologically distinct subsets carrying stereotyped B-cell receptors (i.e. subsets #1, #2 and #4) and revealed unique gene expression profiles for each subset. Analysis of *SF3B1*-mutated versus wildtype subset #2 patients revealed a large number of splice variants (n=187) in genes involved in chromatin remodeling and ribosome biogenesis. Taken together, this thesis confirms the prognostic impact of recurrent mutations and provides data supporting implementation of targeted NGS in clinical routine practice. Moreover, we provide evidence for the involvement of novel players, such as *RPS15*, in disease progression and present transcriptome data highlighting the potential of global approaches for the identification of molecular mechanisms contributing to CLL development within prognostically relevant subgroups.

Keywords: chronic, lymphocytic, leukemia, CLL, genomics, transcriptomics, DNA, RNA, mutations, NGS, whole-exome, sequencing, prognostic, markers, TP53, SF3B1, RPS15, relapse, stereotyped, subsets.

Diego Cortese, Department of Immunology, Genetics and Pathology, Rudbecklaboratoriet, Uppsala University, SE-751 85 Uppsala, Sweden.

© Diego Cortese 2016

ISSN 1651-6206

ISBN 978-91-554-9702-6

urn:nbn:se:uu:diva-303703 (<http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-303703>)

Nothing in biology makes sense except in the light of evolution”
Theodosius Dobzhansky

List of Papers

This thesis is based on the following papers, which are referred to in the text by their Roman numerals. Reprints were made with permission from the respective publishers.

- I Cortese D**, Sutton L-A, Cahill N, Smedby K E, Geisler C, Gunnarsson R, Juliusson G, Mansouri L and Rosenquist R. On the way towards a ‘CLL prognostic index’: focus on *TP53*, *BIRC3*, *SF3B1*, *NOTCH1* and *MYD88* in a population-based cohort. *Leukemia* 2014; 28(3):710-713.
- II Sutton L-A***, Ljungström V*, Mansouri L, Young E, **Cortese D**, Navrkalova V, Malcikova J, Muggen AF, Trbusek M, Panagiotidis P, Davi F, Belessi C, Langerak AW, Ghia P, Pospisilova S, Stamatopoulos K, Rosenquist R. Targeted next-generation sequencing in chronic lymphocytic leukemia: a high-throughput yet tailored approach will facilitate implementation in a clinical setting. *Haematologica* 2015; 100(3):370-376.
- III Ljungström V***, **Cortese D***, Young E, Pandzic T, Mansouri L, Plevova K, Ntoufa S, Baliakas P, Clifford R, Sutton L-A, Blakemore SJ, Stavroyianni N, Agathangelidis A, Rossi D, Höglund M, Kotaskova J, Juliusson G, Belessi C, Chiorazzi N, Panagiotidis P, Langerak AW, Smedby KE, Oscier D, Gaidano G, Schuh A, Davi F, Pott C, Strefford JC, Trentin L, Pospisilova S, Ghia P, Stamatopoulos K, Sjöblom T, Rosenquist R. Whole-exome sequencing in relapsing chronic lymphocytic leukemia: clinical impact of recurrent *RPS15* mutations. *Blood* 2016; 127(8):1007-16.
- IV Cortese D**, Ljungström V, Plevova K, Rossi D, Stalika E, Agathangelidis A, Scarfò L, Boudjoghra M, Muggen AF, Langerak AW, Pospisilova S, Davi F, Ghia P, Stamatopoulos K, Rosenquist R**, Sutton L-A**. Differential expression of coding/non-coding transcripts and splice variants in stereotyped subsets of chronic lymphocytic leukemia. *Manuscript*.

* Equal first authors, ** Equal senior authors

Related publications

1. Bhoi S, Baliakas P, **Cortese D**, Mattsson M, Engvall M, Smedby KE, Juliusson G, Sutton L-A, Mansouri L. UGT2B17 expression: A novel prognostic marker within IGHV-mutated chronic lymphocytic leukemia? *Haematologica* 2016; 101(2):63-65.
2. Baliakas P, Hadzidimitriou A, Sutton L-A, Rossi D, Minga E, Villamor N, Larrayoz M, Kminkova J, Agathangelidis A, Davis Z, Tausch E, Stalika E, Kantorova B, Mansouri L, Scarfò L, **Cortese D**, Navrkalova V, Rose-Zerilli MJJ, Smedby KE, Juliusson G, Anagnostopoulos A, Makris AM, Navarro A, Delgado J, Oscier D, Belessi C, Stilgenbauer S, Ghia P, Pospisilova S, Gaidano G, Campo E, Strefford JC, Stamatopoulos K, Rosenquist R. Recurrent mutations refine prognosis in chronic lymphocytic leukemia. *Leukemia* 2015; 29(2):329-336.
3. Mansouri L, Sutton L-A, Ljungström V, Bondza S, Arngården L, Bhoi S, Larsson J, **Cortese D**, Kalushkova A, Plevova K, Young E, Gunnarsson R, Falk-Sörqvist E, Lönn P, Muggen AF, Yan X-J, Sander B, Enblad G, Smedby KE, Juliusson G, Belessi C, Rung J, Chiorazzi N, Strefford JC, Langerak AW, Pospisilova S, Davi F, Hellström M, Jernberg-Wiklund H, Ghia P, Söderberg O, Stamatopoulos K, Nilsson M, Rosenquist R. *J Exp Med.* 2015; 212(6):833-843.
4. Strefford JC, Sutton L-A, Baliakas P, Agathangelidis A, Malčíková J, Plevova K, Scarfò L, Davis Z, Stalika E, **Cortese D**, Cahill N, Pedersen LB, di Celle PF, Tzenou T, Geisler C, Panagiotidis P, Langerak AW, Chiorazzi N, Pospisilova S, Oscier D, Davi F, Belessi C, Mansouri L, Ghia P, Stamatopoulos K, Rosenquist R. *Leukemia* 2013; 27(11):2196-2199.
5. Rosenquist R, **Cortese D**, Bhoi S, Mansouri L, Gunnarsson R. Prognostic markers and their clinical applicability in chronic lymphocytic leukemia: where do we stand? *Leuk Lymphoma* 2013; 54(11):2351-2364.

Contents

Introduction	11
Chronic lymphocytic leukemia	12
Clinical and biological heterogeneity	12
The CLL microenvironment	13
Cytogenetic aberrations	14
Immunoglobulin mutational status	15
Prognostic models	16
Novel treatment options	16
Next-generation sequencing in CLL	19
DNA-based sequencing	19
The CLL genome	20
Notch homolog 1, translocation-associated (NOTCH1)	22
Baculoviral IAP Repeat Containing 3 (BIRC3)	22
Myeloid differentiation primary response gene 88 (MYD88)	23
Splicing factor 3B subunit 1 (SF3B1)	23
Tracking clonal evolution in CLL with NGS	24
Next generation RNA-sequencing	26
The CLL transcriptome	27
The deregulation of the spliceosome	28
Immunogenetics	29
The B cell receptor: structure and signaling	29
Somatic recombination	30
B-cell development	31
B-cell receptor diversity	33
Stereotyped subset classification	33
Clinicobiological features of major stereotyped subsets	34
Present investigations	35
Thesis aims	35
Patients and methods	36
Patients	36

Methods	37
PCR amplification and Sanger sequencing	37
Targeted next-generation sequencing	37
Whole-exome sequencing	37
In vitro functional characterization of RPS15	38
Next generation RNA-sequencing	39
Statistical analysis	40
Results and discussion	41
Paper I	41
Paper II	42
Paper III	45
Paper IV	46
Concluding Remarks	49
Acknowledgments	51
References	53

Abbreviations

+12	Trisomy 12
AML	Acute myeloid leukemia
AS	Alternative splicing
<i>ATM</i>	Ataxia telangiectasia mutated
BcR	B-cell receptor
BM	Bone marrow
BMSC	Bone marrow stromal cell
<i>BTK</i>	Bruton tyrosine kinase
cDNA	Complementary DNA
CDR	Complementarity determining region
CLL	Chronic lymphocytic leukemia
CML	Chronic myeloid leukemia
COSMIC	Catalogue of somatic mutations in cancer
CR	Complete remission
CSR	Class switch recombination
DAG	Diacylglycerol
del(11q)	Deletion of long arm of chromosome 11
del(13q)	Deletion of long arm of chromosome 13
del(17p)	Deletion of short arm of chromosome 17
DEU	Differential exon usage
DLBCL	Diffuse large B-cell lymphoma
EBV	Epstein-Barr virus
ER	Endoplasmic reticulum
FC	Fludarabine-cyclophosphamide
FCR	Fludarabine-cyclophosphamide-rituximab
FDR	False discovery rate
FISH	Fluorescence in-situ hybridization
GO	Gene Ontology
IG	Immunoglobulin
IGHV	Immunoglobulin heavy variable
IP3	Inositol triphosphate
ITAM	Immunoreceptor tyrosine-based activation motif
lincRNA	Long intergenic noncoding RNA
LN	Lymph node

lncRNA	Long noncoding RNA
M-CLL	CLL with mutated IGHV genes
MBL	Monoclonal B-cell lymphocytosis
MDS	Myelodysplastic syndrome
miRNA	Micro RNA
mRNA	Messenger RNA
MZ	Marginal zone
NGS	Next-generation sequencing
NHEJ	Non-homologous end joining
NLC	Nurse-like cell
ORR	Overall response rate
OS	Overall survival
PCR	Polymerase chain reaction
PEST	Proline, glutamic acid, serine and threonine
PFS	Progression-free survival
<i>PI3K</i>	Phosphatidylinositol-4,5-bisphosphate 3-kinase
PIP2	Phosphatidylinositol 4,5-bisphosphate
PR	Partial remission
pre-B cell	Precursor B cell
Pre-BcR	Pre B-cell receptor
pro-B cell	Progenitor B cell
RCA	Recurrent chromosomal aberrations
RNA-seq	RNA-sequencing
rRNA	Ribosomal RNA
RSS	Recombination signal sequencing
SHM	Somatic hypermutation
SLL	Small lymphocytic lymphoma
snoRNA	Small nucleolar RNA
SNV	Single nucleotide variant
SR	Somatic recombination
T-ALL	T-cell acute lymphoblastic leukemia
TdT	Terminal deoxynucleotidyl transferase
T _h	T helper
TLR	Toll-like receptor
<i>TP53abn</i>	<i>TP53</i> aberrations
TTT	Time-to-treatment
U-CLL	CLL with unmutated IGHV genes
WES	Whole-exome sequencing
WGS	Whole-genome sequencing

Introduction

B-cell maturation and differentiation are fundamental steps for the development of an efficient immune system in higher organisms such as humans. The accumulation of monoclonal B cells in the blood and their infiltration in both primary and secondary lymphoid organs, as observed in chronic lymphocytic leukemia (CLL), represents a life threatening dysregulation of the immune system. A close interaction with the microenvironment and the accumulation of genetic aberrations both contribute to the development and progression of this disease. Until recent years, the understanding of the molecular basis of CLL was limited, however with the advent of massive parallel sequencing numerous candidate genes involved in the pathogenesis of CLL were revealed. This thesis provides the reader with a comprehensive review of the most recent findings on genetics and immunogenetics in CLL. In addition, prognostic markers, clonal evolution and transcriptome analysis are core sections of the thesis that will be extensively discussed.

Chronic lymphocytic leukemia

Clinical and biological heterogeneity

CLL is the most common leukemia among the elderly in western populations with an incidence of 4.2/100,000/year and a male predominance.¹ The median age at diagnosis is 71 years although about 10% of CLL patients are younger than 55 years.¹ CLL is characterized by the accumulation of small, mature B lymphocytes in blood, bone marrow and secondary lymphoid organs.² CLL cells typically express the B-cell surface receptors CD19, CD20, CD23 and co-express the CD5 antigen.² In addition to surface antigen detection by flow cytometry, the diagnosis of CLL requires a lymphocytosis with a B-cell count equal to or higher than 5.0×10^9 cells/L.^{1,2} A pre-leukemic condition, defined as monoclonal B-cell lymphocytosis (MBL) can exist and, although at low rate (1-2%), may evolve into CLL.³

The clinical outcome of CLL patients can be very diverse ranging from an indolent disease, with no treatment required, to an aggressive disease with reduced survival and refractoriness to treatment.⁴ The majority of CLL patients (85%) are diagnosed at an early disease stage in the absence of symptoms; most patients undergo the 'watch and wait' strategy and will therefore not receive treatment until signs of 'active disease' emerge, while the remaining 15% of patients have a symptomatic disease requiring immediate treatment.⁵⁻⁷ Several therapies are currently available, including chemotherapy and monoclonal antibodies, and more recently small drug inhibitors. Fludarabine-cyclophosphamide-rituximab (FCR) is today the standard treatment for medically fit patients with a response rate in the range of 90%.⁸ Bendamustine-rituximab (BR) is an alternative first-line treatment in patients for whom FCR therapy is not feasible with a response rate similar to FCR.⁹

A large number of biomarkers have been proposed during the last decades in an attempt to refine the prognostication of CLL;¹⁰ however only a few, mainly genetic markers are used today in clinical routine diagnostics to predict disease progression and outcome.¹¹⁻¹⁵ The most plausible explanation for the observed clinical heterogeneity in CLL may reside in the underlying biological heterogeneity of the disease involving both cell-intrinsic, i.e. genetic events, and cell-extrinsic, microenvironmental stimuli.¹⁶ In the next

paragraph, the major mechanisms involved in the crosstalk of CLL cells with the microenvironment are discussed.

The CLL microenvironment

Once considered a disease of resting B cells, CLL has clearly emerged as a proliferative disorder.¹⁷ While circulating CLL cells are predominantly non-dividing, resting cells, about 1% of CLL cells proliferate on a daily basis.¹⁷ Numerous lines of evidence support the microenvironment dependency of CLL cells. Indeed, the establishment of cell lines is particularly difficult in the absence of *Epstein-Barr* virus (EBV) infection and without external stimuli CLL cells undergo apoptosis when cultured *in vitro*.^{18,19} The interaction between stromal cells, T cells and a plethora of chemokines comprise the basis of the complex cross talk between CLL cells and the microenvironment which occurs in proliferative centers located in the bone marrow (BM) and secondary lymphoid organs.²⁰ In the BM, stromal cells (BMSC) are in direct contact with CLL cells and protect them from apoptosis.¹⁹ The anti-apoptotic BCL2 signal is indeed activated by the interaction of VCAM-1 expressed on the surface of BMSC with VLA-4 expressed on the CLL cell surface.¹⁸ At the same time, BMSC release chemokines to sustain the CLL clones. Gene expression studies on CLL cells, derived from co-culture with BMSC, showed a marked upregulation of the pro-survival molecule TCL1.²¹

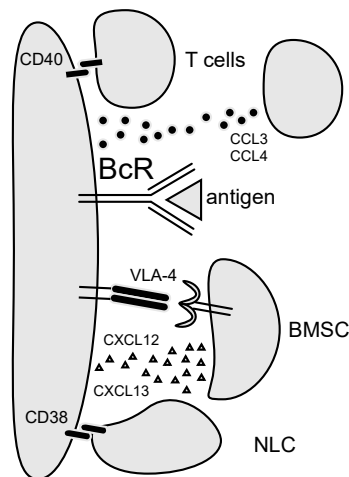


Figure 1. The CLL microenvironment. CLL cells interaction with antigens, T cells, BMSC and NLC is showed. The B cell receptor (BcR) recognizes its cognate antigen on the cell surface; T cells communicated with CLL cells through CD40; bone marrow stromal cells (BMSC) signal to CLL cells via VLA-4 and chemokines (CXCL12, CXCL13); nurse like cells (NLC) communicate with CLL cells through the CD38 receptor.

Lymph nodes (LN) have emerged as particularly crucial sites for CLL cell activation and proliferation. In a study by Herishanu *et al*, gene expression profiling of patient-matched LN and peripheral blood (PB)/BM samples, revealed increased B-cell receptor (BcR) and NFκB activation in LN.²² In fact, tumor proliferation (determined by the c-MYC and E2F expression levels) was highest in the LN compartment compared to PB and BM.²² Furthermore, the presence of nurse-like cells (NLC) is of importance for CLL cell proliferation; gene expression profiling of CLL cells co-cultured with NLC and that of CLL cells derived from patients exhibited striking similarities, with BcR and NFκB pathways being the most activated.²³ In fact, a recent study from Gautam *et al*. proposed the reprogramming of NLC as an effective method to abolish CLL cell survival,²⁴ since interferon-γ shifted NLC towards an effector-like state with enhanced rituximab-mediated phagocytosis of CLL cells.²⁴ In addition to microenvironmental interactions, genetic aberrations are detected in a large proportion of CLL patients.

Cytogenetic aberrations

Fluorescence in situ hybridization (FISH) detection of cytogenetic aberrations is an important tool that assists clinicians in treatment decision-making in CLL.²⁵⁻²⁸ Overall, more than 80% of CLL cases carry recurrent chromosomal aberrations, namely deletions of chromosomes 13q [del(13q)] and 11q [del(11q)], trisomy of chromosome 12 (+12) and deletion of chromosome 17p [del(17p)].²⁵⁻²⁸ The most frequent chromosomal lesion observed is del(13q), which occurs in more than 50% of CLL patients, while del(11q), +12 and del(17p) are less frequent (10-20%, 15-20% and 5-10%, respectively).^{25,29,30}

Despite the high frequency, del(13q), as the sole abnormality, is associated with a more indolent disease course and a better clinical outcome, even when compared to patients without any recurrent chromosomal aberration.^{13,25,31} Deletion of a critical region at 13q14.3 containing two micro-RNAs (miRNA), miR-15a/16-1, known targets of BCL2, has been proposed as the pathogenic mechanism in del(13q) cases.³² Nevertheless, patients with larger del(13q) have an increased risk of progression, shorter time-to-treatment (TTT) and shorter overall survival (OS), implying that other genes within the deleted region could be responsible for disease progression.³³⁻³⁵

CLL cases carrying trisomy +12 showed a better response to treatment with FCR compared to FC alone³⁶, albeit having an intermediate prognosis and heterogeneous clinical outcome. The association of *NOTCH1* mutations with +12 has been shown to confer poorer prognosis than +12 alone, potentially refining the intermediate-risk prognosis of patients carrying +12.¹²

The prognosis for CLL patients with del(11q) has historically been poor, however FCR treatment appears to be of benefit for patients carrying this genetic aberration.⁸ del(11q) encompasses a region containing numerous genes, several of which could be of importance for tumor progression. The *ATM* gene resides in the minimal deleted region of del(11q) and mutations in this gene have been reported in 20-40% of del(11q) suggesting a biallelic inactivation mechanism.³⁷⁻³⁹ However, the association of *ATM* aberrations with tumor progression and survival is still under investigation.

del(17p), which encompasses the *TP53* gene, has been associated with short TTT, poor response to therapy and the worst survival of all CLL patients.^{25,40} While approximately 80% of patients with del(17p) carry a *TP53* mutation on the other allele, patients harboring *TP53* mutations without co-existing del(17p) have an equally poor prognosis as patients with del(17p).^{41,42} The frequency of *TP53*-aberrations (*TP53abn*), i.e. del(17p) and/or *TP53* mutations, steadily increases in more advanced disease stages, and are observed in a high proportion of treatment-refractory patients.^{43,44} While CLL patients carrying *TP53abn* do not benefit from the addition of rituximab to fludarabine and cyclophosphamide (FCR), they appear to respond to two novel inhibitors of the BcR pathway, ibrutinib (BTK) and idelalisib (PI3K).^{45,46} FISH-detection of del(17p) and *TP53* mutation screening are today mandatory before the start of therapy or at relapse, since these aberrations are the only genetic markers with an immediate impact on treatment decisions in CLL.⁴⁷

Immunoglobulin mutational status

In addition to the aforementioned genetic markers, immunoglobulin gene (IG) analysis of CLL patients has dramatically changed the view of the disease. Indeed, IG analysis revealed that CLL consists of two clinically distinct subsets harboring either mutated or unmutated IG heavy variable (IGHV) genes (M-CLL or U-CLL, respectively).⁴⁸⁻⁵⁰ Moreover, the findings of somatically hypermutated IGHV genes changed the prevailing view of CLL cells as antigen-unexperienced, naive B cells.^{4,51-54} An important finding was also the association of the IGHV mutational status with prognosis whereby M-CLL have a significantly better outcome than U-CLL.⁴⁸ In numerous studies, IGHV mutational status has been shown as one of the strongest prognostic markers in CLL.^{55,56} Recently, it was shown for the first time that FCR could potentially represent a cure for M-CLL patients. At the follow-up of 12.8 years, almost 80% of M-CLL who achieved minimal residual disease (MRD) negativity in BM following FCR treatment exhibited long-term disease-free survival.^{57,58}

Prognostic models

As mentioned earlier, due to the inherent disease heterogeneity, it remains a major challenge to accurately predict risk at diagnosis that can also guide treatment decisions. Two staging methods based on physical examination and standard laboratory tests are currently in use world-wide: the Rai and the Binet systems.^{6,7} In the Rai staging system, patients who have lymphocytosis with CLL cells in the blood and/or bone marrow are considered stage 0 (low-risk). Involvement of secondary lymphoid organs (LN, spleen) and liver defines stages I-II (intermediate risk) patients, while the presence of disease-related anemia or thrombocytopenia defines stages III-IV (high risk) patients.¹ In the Binet staging system, organomegaly or enlarged LN in certain areas, such as head and neck, axillae and groins, are taken into consideration as well as anemia and thrombocytopenia. Involvement of up to two areas defines stage A, while three or more areas involved in presence of anemia defines stage B. Stage C is characterized by anemia and thrombocytopenia irrespective of the number of areas with LN involvement.¹

With the introduction of FISH-detection of cytogenetic aberrations, a hierarchical model proposed by Döhner *et al.* is commonly applied for prognostication of CLL patients.²⁵ In this model, del(17p) carries the highest risk/worst outcome, followed by del(11q), trisomy 12, and no aberrations, while del(13q) is associated with the lowest risk and best prognosis.

More recently, the CLL International Prognostic Index (CLL-IPI) working group proposed a novel model combining genetic, biochemical and clinical parameters.⁵⁹ Over 3470 untreated CLL patients were included in the analysis and five independent prognostic factors were identified: *TP53* status, defined as the presence of *TP53* mutations and/or del(17p); IGHV mutational status, with unmutated IGHV associated with a poor prognosis; high β_2 microglobulin concentration in the serum (>3.5 mg/L), age over 65 years and clinical stage (Binet A and Rai 0 vs. Binet B-C and Rai I-IV). The prognostic index was derived from the grading of the aforementioned factors using a score ranging from 0 to 10 and four risk groups were defined: very high, high, intermediate and low. The presence of *TP53abn* is *conditio sine qua non* to assign the patient to the very high-risk group while, in the absence of other risk factors, it alone confers high risk.

Novel treatment options

In the era of chemoimmunotherapy, patients with refractory CLL typically experienced an unfavorable disease course with inferior progression-free survival (PFS) compared to those responding to treatment.¹⁴ Over the past 5

years, novel therapeutic agents targeting the bruton tyrosine kinase (BTK), phosphoinositide-3 kinase (PI3K) and B-cell lymphoma 2 (BCL2) have been studied and approved.^{45,60-64} Introduction of the kinase inhibitors ibrutinib and idelalisib and the BCL2 inhibitor venetoclax has since remarkably improved therapy for relapsed patients.^{45,60-64} Compared to FCR, ibrutinib and idelalisib are less toxic and could hence represent valid future alternatives to chemoimmunotherapy.⁶⁵ However, due to the lack of prospective randomized trials and the current limited experience with these drugs, FCR remains the standard regimen for medically fit patients with untreated CLL and negative for *TP53abn*.⁸

Ibrutinib: Upon BcR activation, the cytoplasmic tyrosine kinase BTK is activated by LYN and SYK to induce proliferation and differentiation.⁶⁶ BTK is highly expressed in B cells but not in T cells and, aside from its role in BcR signaling, is involved in cell adhesion and migration.⁶⁷ Ibrutinib is an irreversible inhibitor of BTK, capable of preventing the crosstalk between the lymphocytes and the microenvironment, thus impeding CLL cell adhesion and homing.⁶⁷ Ibrutinib has been shown to induce apoptosis in the presence of pro-survival factors such as TNF- α , IL-6, IL-4 and CD40L and the drug can inhibit CLL cell survival by blocking the homing of CLL cells in the LN.⁶⁰ Indeed, a characteristic clinical feature of ibrutinib is the redistribution of CLL cells from lymph nodes to the PB. Ibrutinib is currently approved by the US Food and Drug Administration (FDA) for high-risk patients carrying del(17p) and patients with refractory or relapsed CLL.⁶⁸ In a recent report, the response to treatment improved with the duration and induced death of almost 3% of CLL cells per day.⁶¹ Despite the presence of circulating tumor cells, complications arising due to infections decreased during treatment, possibly due to the immune-modulating activity of ibrutinib.⁶¹ However, about 20% of patients discontinued the treatment for causes that need to be further clarified.⁶¹ In these patients, circulating CLL cells might home to lymphoid tissues and the disease may become more aggressive.⁶¹ Despite remarkable responses in patients with refractory and relapsed CLL, 5% of patients treated with ibrutinib still experience disease progression.⁶¹ Although the mechanisms of resistance remain largely unknown, mutations in the binding site of ibrutinib (p.C481S) have been observed.⁶⁹ These mutation were not present prior to the administration of ibrutinib, nor during the response to therapy, thus suggesting a late clonal event, although a subclonal, early event cannot be excluded.⁶⁹

Idelalisib: PI3Ks regulate numerous cellular functions including survival and proliferation.⁷⁰ Several isoforms exist and, while PI3K α and PI3K β are ubiquitous, the expression of the PI3K γ isoform is restricted to the hematopoietic compartment, thus representing an attractive therapeutic target in CLL.^{70,71} Idelalisib is a reversible PI3K inhibitor which is highly selective

for B cells and able to break the effect of the protective CLL microenvironment.⁷² This drug is FDA approved for the treatment of relapsed and refractory CLL patients in combination with rituximab. Similar to ibrutinib, patients receiving idelalisib experience a redistribution of CLL cells to the PB during the initial weeks of treatment.⁷³ In a recent study from Woyach *et al.*, *MYC* and *PI3K* amplifications were suggested to be responsible for the resistance to idelalisib observed in 23% of patients who did not reach an objective response.⁴⁶

Venetoclax: *BCL2* is a regulator of apoptosis with anti-apoptotic properties.⁷⁴ Initially described in follicular lymphoma, *BCL2* has been associated with a number of cancers including CLL.^{75,76} Venetoclax, a selective inhibitor of *BCL2* showed an overall response rate (ORR) comparable to that achieved with ibrutinib or idelalisib and a complete remission (CR) rate of 23% with a median follow-up of 15 months.⁶⁴ Unlike BTK and PI3K inhibitors, venetoclax induces minimal residual disease-negative CRs,⁶³ and it was very recently approved for the treatment of del(17p) relapsed-refractory CLL.

Next-generation sequencing in CLL

DNA-based sequencing

The advent of massively parallel sequencing, or next-generation sequencing (NGS), has revolutionized our view of cancer genetics. On the one hand, NGS offers the possibility to sequence billions of DNA bases allowing us to uncover entire genomes, unravel the mutational landscape of hematological and solid cancers, and identify mutations at an unprecedented rate.^{11,77-79} On the other hand, by using a targeted sequencing approach, it is possible to increase the depth of sequencing to obtain more comprehensive results or to investigate the presence of small tumor clones (subclones) in detail.

Whole-genome sequencing (WGS) allows for the detection of virtually every single base in a genome, including both coding and non-coding regions. A major advantage of WGS is the possibility to identify novel genomic events such as mutations, deletions or insertions without any *a priori* knowledge of candidate genes. However, given the massive number of bases to be sequenced, WGS does not provide high sequence depth, typical output is in the order of 30X, which means that only clonal events can be confidently identified. Another drawback resides in the fact that WGS requires advanced high-throughput sequencing platforms and bioinformatics with obvious economical and technical implications. Despite recent advancements, our understanding of non-coding mutations remains limited⁸⁰ and for this reason exons still represent the most attractive target for investigation in cancer genetics.

Whole-exome sequencing (WES) focuses exclusively on exons thus reducing the volume of data generated. Indeed, the major advantage of WES approach is the possibility to investigate all coding regions within a genome at a significantly higher coverage than WGS. In addition to the detection of clonal events, aberrations that occur at an allelic frequency as low as 10% can be confidently identified with a typical WES coverage being in the range of 100X. However, even at this depth, the detection of microclones, thus well below the 10% allelic frequency, is not plausible using WES, but instead a deep sequencing targeted approach is required.

Using targeted NGS, selected genomic regions are enriched and sequenced, with the possibility to multiplex many samples in a single run. Therefore, a targeted NGS approach requires an additional step during which probes targeting the selected regions are designed. These probes are used in an initial enrichment step during the library preparation. Depending on the instrument to be used for the sequencing run, adapter sequences are added to the amplicons prior to sequencing. The choice of the sequencing platform is based on the desired sequence depth. For small gene panels, the laborious aspect of library preparation can be balanced by the possibility to use lower performance sequencing platforms. Once sequencing has been performed, bioinformatics analysis is necessary to condense the results into manageable formats; this step still represent a bottleneck for all types of NGS due to the lack of robust methods and the heavy computational load required. Targeted NGS offers numerous advantages over Sanger sequencing. For example, it is possible to sequence the entire length of multiple genes, samples can be multiplexed (currently up to 384 in one run) and a confident detection limit for calling mutations can go well below the 10% allelic frequency threshold since the reliability is strongly related to the sequence depth. Additionally, genes comprising a large number of coding exons, which makes the mutational analysis unfeasible by Sanger sequencing in a clinical diagnostic setting, can also be investigated.

In the next paragraph, seminal NGS studies in CLL, in particular WGS and WES, are discussed.

The CLL genome

The typical CLL genome harbors about 1000 somatic mutations which corresponds to almost 1 mutation per megabase, a significantly lower mutational frequency compared to many solid tumors.⁷⁷ In CLL, WGS studies have not only identified key genomic players, but also mapped the pathways used by the tumor cells to escape immune surveillance and gain a proliferative advantage.^{77,81,82} Examples include the DNA damage response and cell cycle control, which are among the cellular functions whose molecular pathways carry a heavy mutational load.^{77,79,83} In a seminal study by Puente *et al.*, recurrent mutations were found in *NOTCH1*, *XPO1*, and *MYD88*, among other genes.⁷⁷ More recently, a study applying WGS and/or WES to 452 CLL patients and 54 cases with MBL, identified additional putative driver aberrations in both coding and non-coding region of the CLL genome.⁸⁰ More specifically, the authors showed that mutations in the 3' region of *NOTCH1* could be functionally relevant for aberrant splicing and increased activity of the gene.⁸⁰

In addition to WGS, WES has also been widely applied in CLL studies. An important result using this latter method in CLL was the finding of recurrent mutations in *SF3B1*, a major component of the alternative splicing machinery.^{11,78} Deregulation of the spliceosome has since been observed in several other tumors and there is a growing interest in targeting its core machinery as a novel therapeutic approach.⁸⁴ More recently, a WES study by Landau *et al.*⁸³ identified 20 subclonal candidate CLL drivers (see Table 1) including *TP53*, *NOTCH1*, *SF3B1*, *ATM*, *MYD88*, *CHD2* and *POT1* previously associated with CLL;^{77,85,86} These 20 candidate genes belong to seven distinct and well-characterized signaling pathways including DNA repair, Notch signaling, Wnt signaling, inflammation, AS, BcR signaling and chromatin modification.⁸⁷ More recently, 24 additional subclonal candidate drivers were identified through WES of over 500 CLL cases collected in a prospective clinical trial.⁸⁷ A recent work from Puente *et al.* extended the list of CLL driver mutations (e.g. *ARID1A*, *ZMYM3*) and identified mutations in the 3' region of *NOTCH1*, supporting a role in CLL for non-coding DNA.⁸⁰

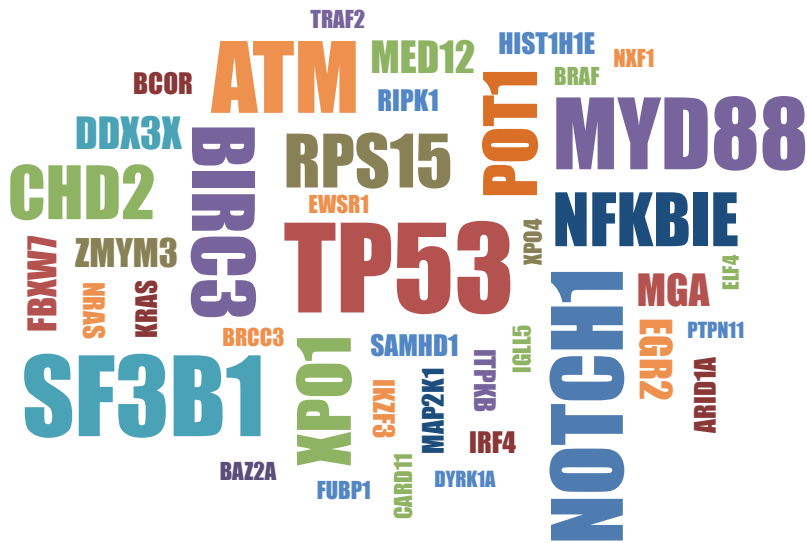


Figure 2. Candidate CLL drivers. Among the more frequently mutated are *TP53*, *NOTCH1*, *BIRC3*, *MYD88* and *SF3B1*.

Unlike other hematological tumors such as chronic myeloid leukemia (CML) or Burkitt’s lymphoma where chromosomal translocations are almost universally found and can drive the tumor transformation, deep-sequencing studies have confirmed the absence of a single driver aberration leading to CLL. Rather, it appears that the accumulation of genetic lesions confers a selective advantage to the clone, and it is a combination of these aberrations together with the microenvironment that may represent the general mechanism by which CLL clones evolve.^{83,88,89} By investigating selected pathways,

additional pathogenic mechanisms have recently been identified in CLL.^{90,91} For instance, the finding that *NFKBIE* truncating mutations leads to constitutive NF- κ B activation and a worse patient outcome, underscores the critical role of NF- κ B in the pathobiology of CLL and opens up the possibility of future targeted therapy against components of NF- κ B, at least in certain subsets of patients.⁹²

With more than 1000 CLL exomes sequenced thus far, the number of driver mutations identified has steadily increased. Among the more frequently identified are recurrent mutations in *NOTCH1*, *BIRC3*, *MYD88* and *SF3B1* that have been more extensively investigated. The most important findings related to these gene mutations are discussed below.

Notch homolog 1, translocation-associated (NOTCH1)

Four NOTCH receptors are found in mammals (NOTCH1-4). Following the engagement of the ligand, the cleaved NOTCH intracellular domain translocates to the nucleus to convert the DNA binding protein CSL into an activator of transcription. Genes involved in cell cycle control, such as *MYC*, *CCND1* (cyclin D1) and *CDKN1A* (p21) are then transcribed. Notch activation is terminated through ubiquitination by FBXW7 and subsequent degradation.⁹³ *NOTCH1* has been implicated in apoptosis, cell differentiation and proliferation. A 2 base-pair frameshift deletion (7544_7545delCT) in the C-terminal PEST domain accounts for over 90% of all identified mutations, and was recently shown to confer a stabilizing effect on NOTCH1 signaling.⁷⁹ *NOTCH1* mutations have been reported in 5-15% of CLL patients and are associated with shorter OS and PFS.^{12-14,31,79,94} The association of *NOTCH1* with +12 has been shown to confer poorer prognosis than +12 alone, refining the intermediate-risk prognosis of patients carrying +12.¹² Nevertheless, the independent prognostic value of *NOTCH1* is still under debate.⁹⁵

Baculoviral IAP Repeat Containing 3 (BIRC3)

BIRC3 acts downstream of the tumor necrosis factor (TNF) and mediates cell proliferation, caspase activity, apoptosis and inflammatory signaling.⁹⁶ BIRC3 has E3 ubiquitin-ligase activity and regulates both the canonical and non-canonical NF- κ B pathways.⁹⁶ Although a low frequency of *BIRC3* mutations have been identified at diagnosis (<5%), up to 20% of patients refractory to fludarabine treatment were shown to carry *BIRC3* mutations and/or deletions, suggesting an association of *BIRC3* disruption with chemorefrac-

toriness in patients with wild-type *TP53*.¹⁵ In agreement with this result, *BIRC3* disrupted cases have a poor prognosis similar to patients harboring *TP53abn*, and the two genetic lesions appear to be mutually exclusive.⁹⁵ *BIRC3* is located at 11q22, and similar to mutations within the *ATM* gene, the frequency of mutations within *BIRC3* is higher in patients harboring del(11q).⁹⁵ However, screening of del(11q) patients within the UK CLL4 trial cohort showed that *BIRC3* deletion exclusively co-occurred with *ATM* deletion and that *BIRC3* aberrations had limited impact on OS and PFS.⁹⁷

Myeloid differentiation primary response gene 88 (MYD88)

MYD88 is a cytosolic adapter for interleukin-1 (IL-1) and Toll-like receptor (TLR) signaling. Mutations in the *MYD88* gene are relatively infrequent in CLL (2-5%)^{13,77,95} and are almost exclusively associated with M-CLL patients and thus with a favorable prognosis. However, the prognostic value of *MYD88* mutations is not fully elucidated. A recent study by Martínez-Trillos *et al.* analyzed mutations in the *TLR/MYD88* pathway in a series of 587 CLL patients.⁹⁸ The authors reported that 3.2% of patients (n=19) carried *MYD88* mutations, while aberrations in *IRAK1* (n=2), *TLR2* (n=2), *TLR5* (n=1) and *TLR6* (n=1) were relatively rare. Patients carrying *TLR/MYD88* mutations had a young median age at diagnosis (<50 years) and displayed a favorable outcome.⁹⁸ In a publication by Baliakas *et al.* which investigated 1039 CLL cases, *MYD88* mutations were also found exclusively in M-CLL.⁹⁹ However, when limiting survival analysis to only M-CLL cases, no significant difference was found when comparing the age at diagnosis or survival in *MYD88* mutated versus *MYD88* wild type cases.⁹⁹

Splicing factor 3B subunit 1 (SF3B1)

SF3B1 is a key component involved in pre-mRNA splicing that specifically functions in the recognition of the branch point site by the U2 snRNP complex.¹⁰⁰ Several independent studies have associated *SF3B1* mutations with disease progression, treatment refractoriness and poor prognosis in CLL.^{13,83,94,101-103} Among CLL cases, p.K700E is the predominant mutation; in addition, *SF3B1* mutations have been identified at a remarkably high frequency (45%) in CLL subset #2 patients.^{102,104} The identification of a hot-spot region in the HEAT domain, with the highest frequency reported for p.K700E, G742D and p.K666E suggests a role for *SF3B1* in conferring a selective advantage to the mutated clone.^{11,105} Despite a large number of studies pointing to the involvement of *SF3B1* in the pathogenesis of CLL,¹⁰⁶ the functional implications of these findings remains elusive. A number of

recent studies provided insights into alternative roles of *SF3B1*.^{107–109} For instance, the observed association of *SF3B1* with nucleosomes could be of importance for the correct recognition of the splice-site and could be driven by the chromatin structure.¹¹⁰

Integration of cytogenetic and molecular findings

Since accurate definition of risk groups at diagnosis is a major issue in CLL, an integrated hierarchical model including both chromosomal aberrations and gene mutations could improve the prognostic ability compared to FISH alone³¹. Rossi *et al.* recently proposed a prognostic algorithm including chromosomal abnormalities and gene mutations in which four CLL risk groups were hierarchically classified.³¹ *TP53* and/or *BIRC3* aberrations identified the high-risk group while the intermediate-risk group was characterized by *NOTCH1* and/or *SF3B1* mutations and/or del(11q). Cases harboring +12 were classified among the low-risk group together with patients without recurrent chromosomal aberrations (RCA). A very low-risk group, whose OS did not appear to differ from the general age-matched population, included patients carrying del(13q) as the sole aberration.³¹ Another hierarchical model, considering a comparable in size cohort of patients, has been proposed by Jeromin *et al.* following the analysis of novel genetic, cytogenetic and immunophenotypic markers.¹³ In their model, *SF3B1* mutations, *NOTCH1* mutations and *TP53* disruption identified a group with short OS and TTT.¹³ While *SF3B1* mutations, IGHV mutational status and del(11q) were the only independent genetic markers for TTT, *SF3B1* mutations, IGHV mutational status and *TP53abn* were able to independently predict OS.¹³ In a study from our group based on the screening of 3490 patients for novel genomic markers, we could demonstrate that *NOTCH1* mutations, *SF3B1* mutations and *TP53abn* correlated with shorter TTT.⁹⁵ Interestingly, *SF3B1* mutations and *TP53abn*, but not *NOTCH1*, retained significance independently of the IGHV mutational status.⁹⁵ Differences among the selected populations could account for discrepancies between the proposed hierarchical models. Nevertheless, all of them support the clinical relevance of novel recurrent mutations in CLL highlighting the need for harmonization of screening methods and large-scale screening to reach a consensus model integrating cytogenetic and molecular findings.

Tracking clonal evolution in CLL with NGS

Targeted therapy improves the chances of interfering specifically with a single pathway, thus achieving higher efficacy and safety. Therefore, it is of crucial importance to understand the mechanisms behind the acquisition of resistance to therapy.

Genetic lesions in minor clones within the tumor population, i.e. subclonal events, and their dynamics have only recently been investigated using NGS methodologies.^{30,111–114} In one study of three patients, longitudinal WGS analysis of relapsed CLL cases detailed the series of events following treatment and documented a complex pattern of clonal evolution.⁸¹ Different somatic mutation profiles identified subclones that expanded or declined over time, pointing to heterogeneous clonal evolution patterns at the individual level.⁸¹ In seminal work from Landau *et al*, by applying WES combined with copy-number analysis to 149 CLL cases, the concept of clonal driver mutations, i.e. lesions present in the entire tumor clone (e.g. +12, del(13q)) vs. subclonal events, i.e. detected in a fraction of the clone (e.g. *TP53*) was introduced.⁸³ Moreover, the presence of subclonal driver mutations could independently predict a rapid disease progression, and evolution of subclones with *SF3B1* and *TP53* mutations was associated with treatment relapse.⁸³

In a recent study on FCR relapsing cases from the German CLL8 trial cohort, tracing of clonal evolution of 59 patients pre- and post-treatment revealed mixed patterns, suggesting that CLL evolution after therapy is complex.⁸⁷ In 30% of cases the relapsing clone was detectable before treatment initiation, thus pointing to a possible anticipation of the clonal evolution.⁸⁷ More specifically, the burden of *TP53* and *IKZF3* mutated clones (measured as cancer cell fractions) markedly increased at relapse indicating a fitness advantage over the rest of the cellular populations⁸⁷, while instead, clones harboring *SF3B1* and *ATM* aberrations did not show any clear pattern, possibly due to a lower ‘fitness’ as compared to *TP53abn*.⁸⁷

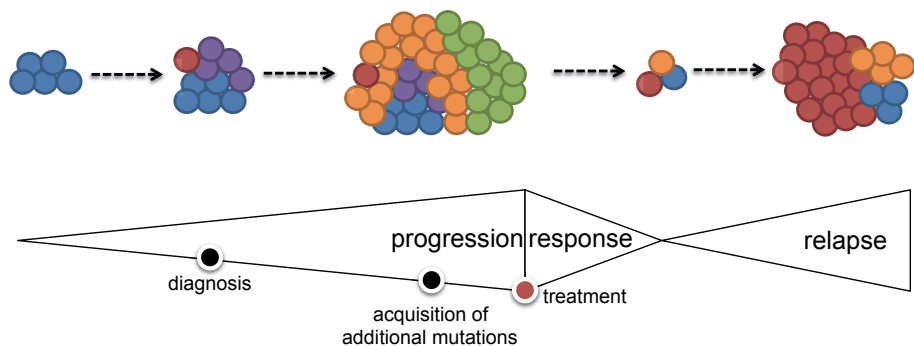


Figure 3. Model of clonal evolution. In the progression phase, mutations are acquired that drive the disease evolution. While therapy may reduce the tumor burden significantly, microclones already present at diagnosis (red circle) may be resistant to treatment and pave the way for a relapse.

Pre-treatment identification of subclones capable of expanding under selective pressure to become clonal at relapse has further elucidated the im-

portance of detecting small clones at diagnosis. A recent work by Rossi *et al.* revealed that *TP53* subclonal mutations already present at diagnosis can impact on the clinical outcome in a similar fashion to clonal *TP53* mutations.¹¹⁵ Malcikova *et al.* tracked therapy-driven clonal evolution in CLL and they confirmed that *TP53*-mutated small subclones can impact on OS and can be detected at diagnosis.¹¹⁶ However, not all subclonal events detectable at diagnosis appear to impact on the clinical outcome. A recent study by Nadeu *et al.* suggested that only *TP53* lesions but not *SF3B1*, *NOTCH1* and *BIRC3* subclonal events were associated with OS.¹¹⁷ However, cases carrying *NOTCH1* subclonal mutations exhibited shorter TTT.

Next generation RNA-sequencing

Next-generation RNA-sequencing (RNA-seq) is a high throughput technology used to detect and quantify the RNA molecules transcribed at a given moment in a population of cells.¹¹⁸ The RNA-seq process requires the fragmentation of total RNA followed by the synthesis of cDNA with the possibility to deplete certain RNA populations prior to sequencing including ribosomal RNA (rRNA) species.¹¹⁹ For example, mRNA can be purified by poly(T) oligos conjugated with beads that will bind the poly(A) tail of coding RNA. Additionally, noncoding populations such as miRNA can be isolated by size selection with magnetic beads, purified and used for sequencing. Since about 90% of total RNA is represented by rRNA, the depletion of this RNA population highly increases the sequencing efficiency and the usefulness of the transcriptomic data obtained.¹²⁰

RNA-seq allows the identification of the whole transcriptome without any *a priori* knowledge of the transcripts of interest thereby providing a considerable advantage over hybridization-based microarrays. When considering only expressed protein coding transcripts, RNA-seq represents a valid method for the detection of small nucleotide variants (SNV), splicing events and fusion transcripts in addition to gene expression.¹¹⁸ Furthermore, RNA-seq detects noncoding transcripts including miRNA, long noncoding RNA (lncRNA), small nucleolar RNA (snoRNA) and pseudogenes. Once the sequencing run is completed, the transcriptome is assembled to the reference genome by using specialized software such as the universal ultrafast RNA-seq aligner STAR.¹²¹ It is then possible to know which genes were transcribed (qualitative information) and at which level (quantitative information), based on the number of mapped reads. There are currently several specialized algorithms which are designed to count mapped reads (e.g. TopHat).¹²² This gene-based approach does not include any information on alternative transcripts and it is useful for gene expression analysis only. Further analysis (e.g. with the R package DESeq2) is performed to produce differential analysis of count data

and calculate normalized gene expression results.^{123,124} The exon-based count, instead, contains information on every exon and can thus be used to study alternative transcripts. Downstream analysis of exon-based counts (e.g. with the R package DEXSeq) can be helpful to infer exon usage and identify alternative splicing events.¹²⁵

In the next paragraph, microarray-based gene expression and RNA-seq studies in CLL are discussed.

The CLL transcriptome

Seminal microarray-based studies revealed distinct gene expression signatures for U-CLL and M-CLL, although the number of differentially expressed genes were relatively low.^{51,126–129} From these studies, it was demonstrated that both the U-CLL and M-CLL expression profile was more similar to memory B cells rather than germinal center B cells or naïve B cells.¹³⁰ In a recent study by Seifert *et al.*,¹³¹ a novel post-germinal center B cell subset (CD5⁺ CD27⁺) was postulated as the cell of origin for M-CLL,¹³¹ while U-CLL instead was suggested to derive from an unmutated CD5⁺ B-cell subset.

Using RNA-seq, a comprehensive transcriptome characterization¹³² of a cohort of 98 CLL patients revealed that genes related to metabolic pathways showed higher expression when compared to B cells obtained from healthy donors, while genes involved in the ribosome, proteasome and spliceosome were down-regulated. Moreover, RNA-seq quantification uncovered two separate transcriptional groups, defined as C1 and C2, which retained clinical significance independent of the IGHV mutational status.¹³² Cases belonging to the C1 group exhibited a more favorable outcome, while patients in the C2 group experienced a more aggressive disease course.¹³² Intriguingly, the analysis of splicing variants revealed a large number of deregulated transcripts. In particular, splicing variants in *SF3B1*-mutated cases displayed the usage of alternative 3' cryptic splicing sites, as previously described in other studies and malignancies.^{133,134} The annotation of non-coding transcripts unraveled a notable deregulation of lncRNA and pseudogenes in CLL compared to normal B cells.¹³² Currently, our knowledge on lncRNA is still limited and therefore, future functional studies are required in order to dissect the molecular basis behind noncoding RNA and cell regulation in general and CLL in particular. That said, a recent study of over 200 CLL cases identified three snoRNA that could independently predict survival.¹³⁵

The potential of RNA-seq in elucidating the pathogenesis of stereotyped subsets of CLL (for a detailed description of BcR stereotypy please refer to the Immunogenetics chapter) has been recently put forward in a pilot study

that uncovered profound differences in gene expression and the existence of novel splice variants among different subsets.¹³⁶ In this thesis, a comprehensive analysis of the transcriptome of stereotyped subsets is presented in paper IV. An overview of the deregulation of the alternative splicing machinery is given in the next paragraph.

The deregulation of the spliceosome

Among all possible mechanisms responsible for malignant transformation of a cell, alternative splicing (AS) was not previously among the most-likely suspects. However, numerous pieces of evidence, including the selective usage of certain transcripts and the abundance of aberrantly spliced isoforms in the transcriptome of neoplastic cells, pointed to the involvement of the spliceosome in cancer.¹³⁷ In humans, over 90% of genes contain multiple exons and AS is considered to be a critical mechanism; however the question still remains how the spliceosome could be deregulated without causing catastrophic events for the cell. WES studies revealed the presence of mutations in numerous key players of the spliceosome in MDS, acute myeloid leukemia (AML) and CLL but also in breast cancer and uveal melanoma. Although *SF3B1*, to date, has been reported to be the most dysregulated member of the spliceosome, numerous other splicing factors including *SF3A1*, *SRSF2*, *PRP40B*, *U2AF35*, *ZRSR2*, *SF1* and *U2AF65* have been identified at various frequencies in different malignancies, thus indicating that dysregulation of the spliceosome could be a new hallmark of cancer.^{138–140}

Immunogenetics

Though genome and transcriptome studies have provided great insights into CLL biology and identified previously unrecognized players with important impact on disease progression and outcome, genetics is only one aspect of the disease. Indeed, B cells have a master regulator, the BcR. The entire development process of a B lymphocyte revolves around building a functional BcR and mature B cells are dependent on the activation of the BcR to exert their immunological function. Immunogenetics studies, focused on IG gene sequences, have provided revolutionary insights into the pathobiology of CLL. In the next paragraphs, the BcR and B cell development are described. Subsequently, the revelations made from immunogenetic studies in CLL are detailed.

The B cell receptor: structure and signaling

The BcR is a transmembrane protein located on the outer surface of B cells. It is composed of two main parts: (i) the IG molecule (IgM, IgD, IgA, IgG or IgE) which is involved in antigen recognition; and (ii) the CD79 heterodimer, which is composed of two distinct chains called Ig- α and Ig- β . This CD79 heterodimer is associated with the IG molecule and upon antigen recognition, propagates the signal to the cell through the immunoreceptor tyrosine-based activation motifs (ITAMs) located within its intracellular domains.¹⁴¹ BcR signaling is initiated by the encounter with the cognate antigen.¹⁴² The duration and amplitude of BcR signaling are determined by the availability of adaptor molecules (e.g. GAB1, GRB2) and are influenced by the activity of downstream phosphatases, i.e. SHP-1 and kinases, i.e. LYN and SYC.¹⁴² BcR activation induces the recruitment of transcription factors in the nucleus to promote cell proliferation through the NF- κ B pathway.¹⁴³ Briefly, the CD79 molecule signals to phospholipase C (PLC- γ) producing the cleavage of phosphatidylinositol 4,5-bisphosphate (PIP2) into inositol 1,4,5-trisphosphate (IP3) and diacylglycerol (DAG). IP3 amplifies the signal by inducing the release of calcium from the endoplasmic reticulum (ER) to, in turn, activate the NF- κ B signaling complex. The I κ B kinase (IKK) is activated and phosphorylates the NF- κ B inhibitor I κ B that, once tagged for degradation, allows NF- κ B to shuttle to the nucleus.¹⁴³

Two heavy and two light chains form the IG molecule and interaction between the chains is via disulphide bridges leading to the formation of heterodimers.¹⁴⁴ Both the IG heavy and light chains contain a variable (V) domain which is responsible for antigen recognition and a constant (C) domain which is important for anchoring and signaling.¹⁴⁴ The IG heavy (IGH) locus is located on chromosome 14 and contains 129 IGH variable (IGHV) genes upstream of 27 IGH diversity (IGHD) genes and 9 IGH joining (IGHJ) genes.¹⁴⁴ Nine IGH constant (IGHC) genes are also located at this locus. IGHV genes have been grouped based on sequence homology into seven subgroups (IGHV1-7); the IGHV3 group is the largest and contains 21 genes. IGHV4 contains 10 genes while 9 genes form the IGHV1 subgroup.¹⁴⁴ However, not all gene segments can produce a functional IGH rearrangement and of the 129 IGHV genes only 46 are functional. The two IG light chain loci are located on different chromosomes: the IG kappa (IGK) locus is found on chromosome 2, while the IG lambda (IGL) locus is on chromosome 22.¹⁴⁴ Between 34-37 functional IGKV genes are found at the IGK locus and 30-33 functional IGLV genes at the IGL locus.¹⁴⁴ In addition, 5 IGKJ genes, 4-5 IGLJ genes and only 1 IGLC gene have been identified.¹⁴⁴ IG light chain genes lack any diversity gene segments.

The variable domains of both the IG heavy and light chains contain distinctive regions: (i) the framework region (FR); and (ii) the complementarity determining region (CDR).¹⁴⁴ Each chain is comprised of four FRs whose function is to maintain the structural integrity of the molecule and three CDRs that form the actual binding pocket for the antigen.¹⁴⁴ While FRs have a relatively conserved sequence, the CDRs display large variability. In particular, the heavy chain CDR3 (VH CDR3) is believed to be the most diverse among all the CDRs and the main determinant of antigen specificity.¹⁴⁴

Somatic recombination

IG genes go through a rearrangement process, termed somatic recombination (SR), during the maturation step of a B cell in order to produce a functional IG heavy and light chain.¹⁴⁵ SR generates a vast repertoire of IG molecules (specifically IgM/IgD) capable of recognizing antigens derived from bacteria, parasites, worms and viruses.¹⁴⁵ This complex IG gene rearrangement process involves IGHV-IGHD-IGHJ genes and occurs at both the IGH and IGK/IGL loci.¹⁴⁵ SR is a tightly controlled event in which only pro-B cells that carry a productive IG rearrangement (about 50%) are capable of further differentiating.¹⁴⁵ SR follows a “cut and rejoin” model with the RAG1 and RAG2 enzymes being responsible for the cleavage and non-homologous end joining (NHEJ) factors involved in the rejoining of the cleaved fragments.¹⁴⁶ RAG1 and RAG2 introduce double strand breaks following their binding to

specific recombination signal sequencing (RSS) flanking the IGHV, IGHD and IGHJ regions.¹⁴⁶ The excised DNA is removed and the ends of the fragments are joined together.¹⁴⁶ During the SR process, a further source of variability is introduced by the terminal transferase (TdT) enzyme which is capable of inserting nontemplated nucleotides at the end of the DNA segments.¹⁴⁶ In addition to their role in cleaving the DNA, RAG1 and RAG2 can also increase IG variability through the introduction of palindromic nucleotides which inversely repeat the DNA sequence thus favoring the formation of helix loops.¹⁴⁶ SR begins with the formation of the IGHD-IGHJ (DJ) complex that involves the joining of an IGHD gene with an IGHJ gene.¹⁴⁵ An IGHV gene then joins the DJ complex to form the IGHV-IGHD-IGHJ (VDJ) complex.¹⁴⁵ A functional rearrangement at the IGH locus is completed when the IGHC gene joins the VDJ gene complex.¹⁴⁵ During the recombination of the IGK and IGL loci, the IGK locus rearranges first the IGKV and the IGKJ genes to produce a functional kappa light chain. In case of failure on both the kappa alleles, rearrangement at the IGL locus will begin.¹⁴⁵

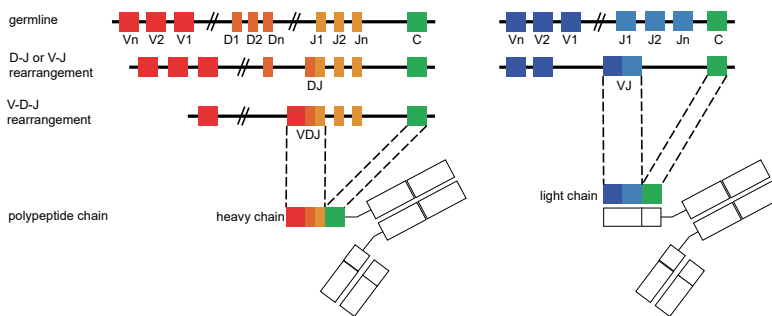


Figure 4. Somatic recombination of the IG genes. At the IGH locus, V-D-J rearrangement produces the functional heavy chain, while V-J rearrangement at the IGL(K) locus produces the functional light chain.

B-cell development

Differentiated plasma and memory B cells have the scope to protect the host from pathogens and the whole defense process relies on the ability of the B cells to recognize their cognate antigens.¹⁴⁷ During their long developmental path, B cells are initially primed in the BM to reach maturity and express IgM and IgD on their surface with unique antigen specificity.¹⁴⁸ In the maturation step, pro-B cells differentiate into pre-B cells upon SR of the IGH locus.¹⁴⁸ At this stage, the BcR is not yet formed and the μ chain is mounted on a surrogate light chain.¹⁴⁸ Subsequently, the crosslinking by stromal-cell ligands causes allelic exclusion and induces the rearrangement of the light chain genes.¹⁴⁸ The successful IG light chain rearrangement replaces the surrogate chain to form a fully functional IG molecule and the cells, now

termed immature B cells, leave the BM to circulate through lymph and blood before reaching the secondary lymphoid organs.¹⁴⁸ Full maturation requires co-expression of IgM and IgD on the surface membrane and at this stage a large proportion of cells become anergic or die due to the lack of IgD. A switch in the RNA processing of the heavy chain allows the transcription of two mRNAs to produce both μ and γ chains to be co-expressed on the surface of mature naïve B cells.¹⁴⁸

In secondary sites, following the encounter with their specific antigens, B cells undergo clonal expansion and differentiate into antibody-secreting plasma and memory B cells.¹⁴⁸ However, of the 5×10^6 naïve B cells that exit the BM, approximately 90% undergo apoptosis within a few weeks due to not encountering their cognate antigen.¹⁴⁸ Germinal centers are the anatomical sites where mature B cells (follicular B cells) proliferate and differentiate.¹⁴⁹ A first crucial event is represented by the isotype switch; following the acquisition of the antigen from dendritic follicular cells, mature B cells, now called centroblasts, present the antigen to T helper (T_H) cells in the dark area at the border between the B-cell follicles and the T cell area.¹⁴⁹ In turn, the centroblasts receive the signal to undergo isotype switch. Class switch recombination (CSR) allows the expression of IgA, IgE and IgG genes instead of IgM and IgD.¹⁴⁹ The isotype switch does not affect the affinity of the IG molecule for binding to the antigen since the IGHV region of the antibody remains unaltered. Upon the completion of several rounds of expansion, centroblasts are subjected to somatic hypermutation (SHM) in order to generate pools of different clones broadening the ability of the immune system to recognize various antigens.¹⁴⁹ In fact, during the SHM process, the coding regions of the IG molecule involved in the antigen recognition acquire mutations that change their affinity for the ligand.¹⁴⁵ Once SHM is completed, centroblasts migrate to the light zone of the germinal center (GC) to become centrocytes capable of expressing antibodies on their surfaces.¹⁴⁹ In a process known as GC selection, because of an activated death program, centrocytes need survival signals to avoid apoptosis. Such selection probably relies on the affinity of the antigen with the antibody. Indeed, only those centrocytes that underwent efficient SHM can be rescued and receive a differentiation signal to enable them to exit the GC and become plasma or memory B cells.¹⁴⁹

In addition to the follicular B cells that proliferate in the GC and require contact with T_H cells, marginal zone (MZ) B cells mature at the boundary between the white and the red pulp of the spleen to provide a barrier against T cell independent antigens such as polymeric proteins, mostly of bacterial origin.¹⁵⁰ Sequencing of the IGHV gene of MZ B cells reported the presence of both unmutated and mutated IGHV, thus suggesting the co-existence of both naïve and memory B cells in the same site.¹⁵¹

B-cell receptor diversity

SR, SHM and additional changes in the sequences of IG heavy and light chain genes contribute to the creation of an enormous variability among the IG molecules; the possibility to find an identical BcR on two independent clones is very unlikely, if not negligible. Indeed, the permutation of V, D and J genes at the IGH locus produces over 7×10^3 combinations that coupled with the probability of two clones to carry the same IGK or IGL rearrangement, results in a probability in the range of 10^{-6} .¹⁴⁵ Furthermore, the SHM, that catalyzes nucleotide changes at a frequency 10^5 higher than the average mutational frequency throughout the genome, greatly contributes to the enhanced variability, thus lowering the probability of finding two healthy individuals carrying the same BcR to $1:10^{12}$.¹⁴⁵

Stereotyped subset classification

Immunogenetics analysis in CLL boomed over the years revealing the strongest evidence for antigen selection in the disease pathogenesis: unrelated groups of CLL patients can carry similar or quasi-identical IG sequences in their BcR.^{152,153} This phenomenon is now known as BcR stereotypy.¹⁵⁴ The identification of restricted repertoire of variable regions in the IG led to the current definition of BcR stereotypy. To begin with, SR appears to be a random process in healthy individuals but the repertoire of IG genes utilized by CLL patient is biased towards a restricted number of genes:^{155,156} IGHV1-69, IGHV3-7, IGHV3-21 and IGHV4-34.¹⁵² The SHM process also shows remarkable differences; while IGHV1-69 displays no or few mutations, IGHV3-7 and IGHV3-21 exhibit higher mutational load.¹⁵⁷ This skewed prevalence of mutation indicates that SHM is not uniform among different IGHV genes, but instead it is influenced by the IG usage. Moreover, despite the low probability of finding two B-cells with the same BcR, subgroups of CLL patients were observed that carried highly similar VH CDR3 with shared amino acid motifs. The immense effort to characterize those similarities resulted in a conclusive study involving over 7500 CLL patients and 16 laboratories showing that approximately 30% of all CLL patients can be assigned to a stereotyped subset and that 19 major subsets exist.¹⁵³ Both M-CLL and U-CLL cases can exhibit stereotyped BcR, however stereotyped subsets are more frequently found among U-CLL.¹⁵³ For instance, subset #2 (IGHV3-21/IGLV3-21) accounts for 5-6% of CLL, thus representing the largest subset, and comprises both M-CLL and U-CLL.¹⁵⁸ Among the subsets including only U-CLL, subset #1 (Clan I genes/IGKV1(D)-39) is the largest (2%) while subset #4 (IGHV4-34/IGKV2-30) is the most frequent among M-CLL cases (1%).¹⁵³ Subset classification is challenging and IG sequences have to meet three criteria in order to be assigned to the same

stereotyped subset: (a) use the same IGHV gene or a member of the same phylogenetic clan, (b) the respective VH CDR3 sequence must have at least 70% similarity and (c) 50% amino acid identity.¹⁵³

Clinicobiological features of major stereotyped subsets

Mounting evidence suggests an involvement of BcR features beyond antigen recognition. Indeed, stereotyped subsets appear to share IG sequences but also genomic aberrations, DNA methylation and gene expression profiles.^{104,128,159,160} Stereotyped subsets may hence represent distinct clinicobiological entities of CLL. For instance, the acquisition of genomic aberrations does not appear to be stochastic, instead enrichment of certain aberrations appears to associate with certain subsets. When focusing on gene mutations, *SF3B1* mutations are highly enriched within subset #2 (45%) and subset #3 (46%), while they are rare in all other subsets.^{102,104} Subset #2 distribution of *SF3B1* does not appear to be SHM-dependent since both M-CLL and U-CLL subset #2 cases carry the mutations at relatively similar frequencies.¹⁰⁴ Of note, in subset #2, about 57% of *SF3B1* mutations are represented by the substitution p.K700E and in about 20% codon p.G742D is affected, with the latter being virtually absent in subset #3.¹⁰⁴ Mutations within exon 34 of the *NOTCH1* gene are more frequent in subsets #1, #58, #99 and #6 while subsets #3, #5 and #7 display a relatively low frequency of *NOTCH1* mutations.¹⁰⁴ Of note, all the above-mentioned subsets concern U-CLL and utilize the same IGHV1-69 gene. Intriguingly, subset #8, which carries a high risk for developing Richter's transformation, is also highly enriched for *NOTCH1* mutations.¹⁰⁴ *TP53* mutations appear to be enriched in subsets #1 and #99 while subsets #2, #4 and #8 display a low frequency of these mutations.¹⁰⁴

The association of genomic features to distinct stereotyped subsets could imply that the manner in which stereotyped BcRs interact with the microenvironment is associated with acquisition of certain genetic lesions, which ultimately may determine clonal evolution and the disease outcome.

Present investigations

Thesis aims

The main aim of this thesis was to perform genomic and transcriptomic sequencing to investigate the prognostic impact of recurrent mutations in CLL and to unravel novel mechanisms behind CLL evolution and progression. More specifically, the aims were as follow:

- I To describe the frequencies of recently reported, recurrently mutated genes and to evaluate their prognostic impact in a population-based cohort of CLL. We focused on *TP53*, *BIRC3*, *SF3B1*, *NOTCH1* and *MYD88* and sequenced 364 patients using Sanger sequencing.
- II To explore the feasibility of targeted NGS as a novel strategy to assess the mutation status of genes with prognostic potential in CLL. We applied targeted NGS to 188 CLL cases with poor prognosis.
- III To uncover novel molecular mechanisms behind clinical aggressiveness and therapy resistance. We performed longitudinal WES in 41 cases, uniformly treated with FCR, that relapsed after a median of 2 years.
- IV To characterize the transcriptome of stereotyped subsets of CLL and to identify potentially aberrant transcriptomic events. We applied RNA-sequencing to 50 CLL cases assigned to subsets #1, #2 or #4 to investigate gene expression and alternative splicing.

Patients and methods

Patients

All CLL cases were diagnosed according to the iwCLL guidelines¹ and displayed a typical CLL phenotype.^{161,162} Written informed consent was collected according to the Declaration of Helsinki and the local review committees of participating institutions granted ethical approval. PB was the main source of material. In **paper I**, 364 CLL patients from the Scandinavian population-based case-control study called SCALE (Scandinavian Lymphoma Etiology) were included. In **paper II**, 188 CLL patients with poor-prognostic features were included from seven collaborating institutions (73% IGHV unmutated and 27% IGHV3-21 subset #2 patients). In **paper III**, three different patient cohorts were investigated. In the discovery cohort, 41 CLL patients from 7 collaborating institutions were screened using WES. All patients had received FCR treatment; 32 patients obtained a CR and 9 patients had a partial remission (PR), and relapsed at a median of 2.17 years (range, 1-11 years). Response and relapse criteria were adopted according to the iwCLL guidelines.¹⁶³ PB samples were collected before the start of FCR treatment and after treatment relapse; constitutional DNA from sorted T cells or buccal swabs was available for 28 patients. In the first validation cohort, targeted sequencing of *RPS15* exon 4 was performed on 790 CLL samples as well as in 30 cases with Richter's transformation. In the second validation cohort, samples from 329 untreated patients enrolled in the UK Leukaemia Research Fund Chronic Lymphocytic Leukaemia Trial 4 (CLL4) clinical trial were investigated. In **paper IV**, 50 CLL patients from 7 collaborating centers were selected based on their assignment to subset #1 (n=13), subset #2 (n=27) or subset #4 (n=10).

Methods

PCR amplification and Sanger sequencing

PCR amplification and direct Sanger sequencing was performed using the Big Dye terminator cycle sequencing reaction kit (Applied Biosystems) and the automated DNA sequencer ABI3730XL for the following hotspot exons: *MYD88* (exon 5), *TP53* (exons 4-8), *BIRC3* (exons 6-9), *SF3B1* (exons 14-16), *NOTCH1* (p.P2514 on exon 34) in **paper I** and *RPS15* (exon 4) in **paper III**. In **paper III**, direct Sanger sequencing utilizing the pCMV6-Entry specific primers confirmed the presence of mutations in the RSP15^{P131S} and RPS15^{G132A} expression vector.

Targeted next-generation sequencing

In **paper II**, the Haloplex probe design for the 9 genes submitted to the SureDesign service included 2,309 amplicons covering 41,962 bases; all coding exons within selected genes were targeted and comprised 202 regions, with 99.1% of target bases covered by at least one probe. The target regions were captured using the Haloplex Target Enrichment kit according to the manufacturer's instructions. Paired-end sequencing (100 bp reads) was performed across two lanes on the HiSeq2000 instrument (Illumina, San Diego, CA, USA) and the mean read depth within the regions of interest was ~1500 reads/base. Illumina sequencing adapters were removed using Cutadapt (v.0.9.5) and trimmed reads were aligned to the hg19 human reference genome (February 2009 assembly) using MosaikAligner version 2.1.33 (<http://arxiv.org/abs/1309.1149>). Tumor variants were identified using VarScan 2 in mpileup2cns mode with detection thresholds of 20x coverage, variant allele frequency (VAF) 10%, a minimum of 10 reads supporting the variant, and an average base quality of 15 at the variant position. Variants had to meet the following conditions to be included in downstream analysis: (i) be located within an exonic or splicing region; (ii) be non-synonymous; (iii) not be listed in the European 1,000 genomes variant database; (iv) not be listed in dbSNP137 unless also listed in the Catalogue of Somatic Mutations in Cancer (COSMIC) 65 database; and (v) have a VAF >10%.

Whole-exome sequencing

Libraries for WES were constructed using the TruSeq Exome Enrichment Kit (Illumina) and sequenced on an Illumina HiSeq 2000. The raw sequencing reads were processed using the bcbio-nextgen framework. Reads were

aligned to the Hg19 reference genome using BWA-mem, version 0.7.10.21 realigned using GATK, version 3.2.22 and polymerase chain reaction duplicates were marked using Sambamba, version 0.4.7. SNV and indels were detected using VarScan2, version 2.3.6.23. Samples with matched germline DNA were analyzed using the somatic mode and the remaining samples using the mpileup2cns mode, both with a 10% VAF cutoff. Samples with matched constitutional DNA were analyzed in all targeted regions as specified by the manufacturer. From this screening, genes mutated in ≥ 2 samples at relapse were screened in samples without matched normal DNA. Somatic copy-number aberrations were derived from WES data using EXCAVATOR and the output was segmented using the R package DNACopy.

In vitro functional characterization of RPS15

All methods described under this section refer to **paper III** in which we described a novel role for RPS15 in p53 stabilization *in vitro*.

Cell culture Because of the inherent difficulties in transfecting existing CLL cell lines, a well-characterized TP53wt colorectal cancer cell line HCT116 (ATCC) was selected.

Transient transfections Myc-DDK tagged RPS15 (RC210640), MDM2 (RG219518), and MDMX constructs (RG209620) were purchased from Origene. To generate RPS15 expression vectors with p.P131S or p.G132A mutations, site-directed mutagenesis was performed using the QuikChange II kit (Agilent Technologies). Transfection of HCT116 colorectal cancer cells was performed using lipofectamine 2000 (Thermo Fisher Scientific Inc.) according to the manufacturer's instructions.

Western blot Cells were lysed in NP40 buffer supplemented with phosphatase/protease inhibitors (Roche); an equal amount of total protein was run on sodium dodecyl sulfate-polyacrylamide gel electrophoresis gel and transferred to nitrocellulose membranes. Following incubation with primary antibodies, the membranes were washed in phosphate-buffered saline and incubated with secondary horseradish peroxidase-coupled goat anti-mouse and horse anti-rabbit antibodies. Immunoreactive proteins were visualized using ECL western blotting detection reagent (GE Healthcare) on the ImageQuant LAS 4000 imaging system (GE Healthcare).

Co-immunoprecipitation HCT116 cells were co-transfected, as described previously, using an equal amount of plasmids. For immunoprecipitation, cells were harvested 24 hours post transfection.

Cycloheximide-chase assay The stability of p53 in the presence of RPS15 was detected by blocking protein synthesis using cycloheximide. Briefly, after transfection, HCT116 cells were treated with cycloheximide and harvested at specified time points.

Ubiquitination assay For p53 ubiquitination assay, after transfection with indicated plasmids, cells were treated with MG132 (Selleckchem, S2619).

Next generation RNA-sequencing

Total RNA was used to prepare the sequencing libraries using the TruSeq stranded total RNA library preparation kit with ribosomal depletion using Ribo-Zero (Illumina Inc., San Francisco, CA). Libraries were multiplexed and sequenced across two flow cells, 5 samples per lane, on an Illumina HiSeq 2500 system using v4 sequencing chemistry and 125 cycle paired-end sequencing. Sequencing resulted in 21.1-176.4x10⁶ aligned reads per sample (median: 68.3 x 10⁶). FASTQ files were filtered prior to mapping, and quality trimming and adapter sequence removal was performed using Cutadapt (v.0.9.5). Quality control was performed using Fastqc. Processed paired-end reads were mapped to the human reference genome (hg19) using STAR version 2.5.

Gene expression analysis Gene counts or abundance estimates were calculated using the feature count tool version 0.6.1. These counts were subsequently used for differential gene expression analysis utilizing the DESeq2 Bioconductor package (version 1.10.1). DESeq2 relies on the negative binomial distribution in order to assess the significance of any detected changes in expression levels. After normalizing the observed counts to enable comparison across samples, variance-mean dependence was estimated for the input count data and differences among two or more subgroups of patients were tested. Adjusted p-values (Benjamini and Hochberg correction) were used to rank genes based on a significant difference in expression ($p < 0.05$) between subgroups. Spearman correlation coefficients were calculated by comparing transcript relative abundances across samples in a pairwise fashion and used to cluster samples using the Genesis software.¹⁶⁴

Differential splicing Changes in splicing can be assessed through the identification of either differential exon usage (DEU) or differential transcript usage events. The Bioconductor package DEXSeq v1.7.0, was applied to infer DEU between various groups and a false discovery rate (FDR) threshold of 0.01 was set to assess significance (Benjamini and Hochberg correction). DEXSeq, which relies on the same algorithmic principles as the aforementioned DESeq2, enables the identification of significant differences

in the proportion of reads that overlap each exon, relative to the total number of reads that overlap the corresponding gene.

Gene Ontology (GO) Enrichment analysis and functional annotation clustering were performed using the Database for Annotation, Visualization and Integrated Discovery (DAVID) version 6.7. Annotation clustering was carried out using the “highest stringency” settings and only groups with an enrichment score above 1.3 (equivalent to a non-log scale value of 0.05) were considered. An adjusted p-value of 0.05 was used for the identification of significant GO terms (Benjamini and Hochberg correction). Non-coding RNAs and IG genes were excluded from this analysis.

Statistical analysis

The main endpoints of the statistical analysis were TTT and OS. TTT was defined as the time between diagnosis and date of treatment, while OS was measured from the date of diagnosis to the date of death or last follow-up. Survival curves were constructed using the non-parametric Kaplan-Meier method. Cox regression analysis was utilized to determine the prognostic impact of each factor; following univariate regression, multiple regression was performed to test the association of single factors with OS taking into account the effect on the outcome of the remaining factors. Tested factors were ranked accordingly to their hazard ratio. Modeling of the hypothesis was performed under normal distribution conditions and all tests were two-sided. Significance was defined as a p-value < 0.05. Statistical analysis was performed with the Statistica Software 10.0 (StatSoftInc, Tulsa, OK).

Results and discussion

Paper I

Reflecting the diverse nature of the investigated CLL cohorts, the novel recurrently mutated genes *TP53*, *BIRC3*, *SF3B1*, *NOTCH1* and *MYD88* have been reported with variable frequencies.^{10–13,29,37} In order to get a better estimate of the true mutational frequencies at diagnosis, we screened a large population-based cohort of CLL patients (n=364) for mutations in these genes using Sanger sequencing. Since several of the novel recurrent genes had been demonstrated to have a prognostic impact, we also aimed to test their independent prognostic value in our sample cohort. Finally, given the risk model previously suggested by Rossi *et al.*,³¹ we tested their algorithm to verify its validity in our population material.

We screened 364 CLL patients for the mutations within the hotspot regions of 5 recurrently mutated genes; *TP53* exons 4 to 8, *BIRC3* exons 6 to 9, *SF3B1* exons 14 to 16, *NOTCH1* exon 34 and *MYD88* exon 5 by PCR amplification and Sanger sequencing. Of the 364 cases, 19 (5.2%) carried a *TP53* mutation and/or deletion, 17 patients (4.7%) had a hotspot *NOTCH1* P2524fs mutation, 13 patients (3.6%) harbored mutations within *SF3B1*, 6 patients (2.3%) had a p.L265P hotspot mutation in *MYD88* while only 1 case (0.4%) carried a *BIRC3* mutation. We next clustered cases where all 5 genes had been investigated (n=262) according to their cytogenetic and mutational profiles and observed that *NOTCH1*, *SF3B1* and *MYD88* mutations were mutually exclusive. As expected, *TP53*-mutated cases often carried del(17p) and less frequently showed +12 and del(11q). A small number of cases (n=3) showed co-existing *TP53* and *NOTCH1* or *SF3B1* mutations.

OS analysis for patients carrying *TP53abn*, *NOTCH1*, *SF3B1* and *MYD88* mutations, +12, del(11q) or no RCA/del(13q) revealed a worse outcome for *SF3B1* or *NOTCH1* mutated patients compared to patients with del(11q) or +12. Notably, no difference was observed between *SF3B1/NOTCH1* mutated and *TP53abn* cases. Following Cox regression analysis on OS we assessed the prognostic independent value of *TP53abn* and *SF3B1/NOTCH1* mutations. In addition to confirming that Binet stage, IGHV mutational status and

age were independent factors for OS prediction, the presence of *TP53abn* and/or *SF3B1/NOTCH1* mutations was independent factors for a dismal outcome. Although our data partly supported the prognostic risk stratification model proposed by Rossi *et al.* differences were also observed since they questioned whether *SF3B1/NOTCH1*-mutated cases should always be considered as intermediate-risk patients. Of note, M-CLL cases carrying *TP53abn* showed a better OS compared to U-CLL (142 vs. 49 months).

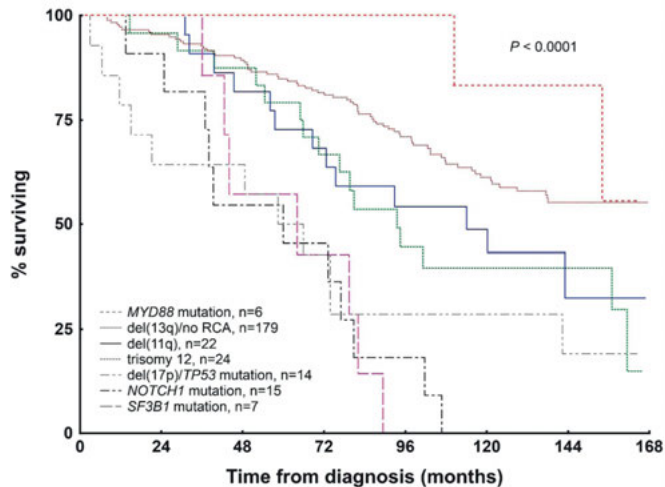


Figure 5. Prognostic impact of available data for all genes analyzed. Two cases harboring both a *TP53* disruption and *NOTCH1* deletion were included in the ‘*del(17p)/TP53* mutation’ group; cases with both *del(11q)* and *SF3B1* mutation were included in the ‘*SF3B1* mutation’ group. *BIRC3* was excluded as only one patient carried a mutation.

A limitation of our study was the number of patients analyzed since prognostic models can only be accurately tested with very large cohorts. That said, despite the smaller cohort size, our data was highly concordant with the recent study by Jeromin *et al.* in which 1160 cases were analyzed.¹³ Screening for selected exons may also negatively affect our ability to identify mutations outside those regions even though the number of mutations found outside the mutational hotspot regions by WGS or WES is extremely small. Furthermore, by applying Sanger sequencing, instead of a target deep sequencing method, important subclonal events may remain undetected. Nevertheless, this study was designed when Sanger sequencing was still the method of choice for mutational screening in clinical routine.

Paper II

Targeted NGS represents an attractive methodology for the identification of genomic aberrations not only in research but also in the clinical setting. Before this method can be transferred to routine practice the quality of the re-

sults in terms of representation and coverage need to be assessed, as well as the reliability of the method in terms of sensitivity, specificity and reproducibility. Our study was specifically designed to address these issues and to determine whether targeted NGS gene panels could be successfully adopted within clinical routine. We selected 188 CLL cases with poor-prognostic features to be screened for all coding exons in 9 genes (*SF3B1*, *NOTCH1*, *TP53*, *XPO1*, *POT1*, *ATM*, *BIRC3*, *MYD88* and *KLHL6*), which have previously been shown to have prognostic relevance and/or to be recurrently mutated in CLL,^{13,77,78} using HaloPlex probes for targeted enrichment.

After sequencing, the panel design rendered targeted gene coverage of 99.1% and an average depth of 1500 reads per base. To ensure high quality of the variants identified, we only included in the analysis those samples in which at least 80% of bases reached 100% coverage in the targeted region. Screening of the 9 genes revealed that 63% of patients carried at least one mutation; mutations in *SF3B1*, *NOTCH1*, *TP53*, *XPO1*, and *POT1* accounted for almost half of the total mutational hits (80/177). In concordance with previous observations,^{14,77,78} the great majority of *SF3B1* mutations were localized within exons 14-16. Of note, *ATM* mutations were dispersed among almost all coding regions, further underscoring the need of targeted NGS for this particular gene. Furthermore, we observed that several patients carried more than one mutation (114 cases and 177 mutations); in particular, *ATM* mutations co-occurred with *SF3B1* in eight cases and with *NOTCH1* in six cases.

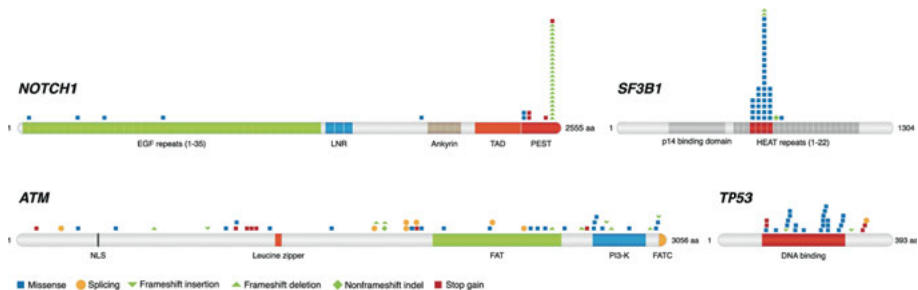


Figure 6. Localization and frequencies of mutations within *ATM*, *NOTCH1*, *SF3B1* and *TP53*. This schematic illustrates how mutations within *NOTCH1*, *SF3B1* and *TP53* predominantly localize to hotspots; recurrent mutations within *SF3B1* are targeted to its highly conserved C-terminal domain, with all but one mutation in this study being detected between the fifth to eighth HEAT domains (exons 14-16). Similarly, a recurrent two base-pair deletion in exon 34 (PEST domain) accounts for the majority of *NOTCH1* mutations observed in CLL. For *TP53*, 24/27 (89%) mutations occurred within the DNA-binding motifs (exons 4-8); remaining mutations were located in exons 9 and 10. In contrast, mutations in the *ATM* gene are unique and distributed throughout the gene's 62 coding exons. To date, no mutational hotspots have been identified in *ATM*.

Sanger sequencing was then used to assess the reliability and capability of the targeted NGS panel to detect mutations and of the 155 mutations addi-

tionally analyzed by Sanger we could confirm 93%, hence demonstrating a high concordance. Notably, the discordant mutations showed low VAFs (range 11-27%), bordering on the limitations of mutation detection by Sanger sequencing and thus possibly explaining this discrepancy. Sixty-three CLL cases from the initial screening that harbored at least one mutation were selected to be included in a validation target NGS panel including 6 genes to assess the reproducibility of our method. We demonstrated high concordance between the targeted NGS panels (94%) and no additional mutations were found in the validation screen, providing us with zero false-positives. However, the latter screen failed to detect five variants from the initial screening, resulting in a 3% false negative calling. Three of the low-frequent variants that could not be confirmed by Sanger sequencing were validated by the validation NGS panel, thus highlighting an advantage of NGS over Sanger sequencing in terms of sensitivity.

Summarizing our results, we show that targeted re-sequencing is a feasible approach for mutations screening within the clinical setting, however stringent quality criteria must be determined and adhered to. The possibility to investigate the full coding region of genes, such as *ATM*, is very attractive. However, the finding of mutations outside hotspot codons or exons warrants caution, in particular if they have not been verified as somatic or have not been previously reported. In our study, we validated the initial screen by repeating the assay and obtained promising results considering the low rate of false positives and false negatives. However, to fully assess the reproducibility of our approach, several targeted NGS methods need to be compared to cross-validate the findings in order to definitively rule out the need of validation by Sanger sequencing, an activity that is ongoing within the European Research Initiative in CLL (ERIC).

The recent identification of minor *TP53* mutated subclones with negative impact on OS in CLL patients¹¹⁵, raises the question as to whether not only clonal but also subclonal events should be investigated and reported at diagnosis. With adequate sequencing depth coverage, targeted re-sequencing can help with the identification of such subclonal populations. In our study, we successfully called, and subsequently validated as true variants, minor subclones with a variant allele frequency ranging from 1-9.9%. The future usage of unique molecular identifiers (UMI) or molecular barcodes to track single DNA molecules may be a powerful tool to accurately detect mutations with low VAFs.

Paper III

Even though CLL patients treated with FCR generally exhibit high ORR to treatment a proportion of patients will eventually relapse.¹⁶⁶ In the German CLL Study Group CLL8 trial, the likelihood of relapse after FCR treatment was 5.6% at 12 months and 14.3% at 24 months.³⁶ While studies aimed at identifying mechanisms leading to FC/FCR resistance typically included cases at a single time point, limiting the identification of clonal evolution events, we set out to whole-exome sequence longitudinal DNA samples from 41 CLL patients who relapsed after FCR treatment.

All patients had received FCR as first-line therapy and demonstrated a good response initially; 32 patients achieved a CR, while nine patients had a partial remission (PR) at <1 year, and relapsed within a median of 26 months (12-132 months). We confirmed somatic mutations for 28 patients with matched germline DNA. On average, 15.2 non-synonymous mutations were found per case in the pre-treatment samples and 17.6 at relapse, implying that the mutational load is not significantly increased following chemoimmunotherapy *per se*.^{11,83} However, analyzing the transversion pattern (substitutions of two ring purines into single ring pyrimidines or *vice versa*) we observed a shift from cytosine to guanine into adenine to thymine for samples at relapse. Furthermore, we report enrichment of certain mutations (e.g. *TP53*, *EGR2*) at relapse; in fact, 26 out of 41 screened cases (63.4%) harbored mutations in *NOTCH1*, *TP53*, *ATM*, *SF3B1*, *MGA*, *BIRC3*, *NFKBIE* or *EGR2* prior to treatment, while 33 patients (80.5%) carried mutations in these genes at relapse.

Unexpectedly, we detected mutations (mostly missense single nucleotide variants) in the ribosomal protein S15 (*RPS15*) in almost 20% of cases (8/41). *RPS15* encodes for a small component of the ribosomal 40S subunit and mutations within this gene were not previously associated with CLL pathobiology. When *RPS15* mutations were present at both time points, we detected them at a high allelic frequency suggesting *RPS15* mutations as early initiating events in CLL. Copy-number WES data did not indicate the presence of small genomic lesions encompassing the *RPS15* gene. All *RPS15* mutated cases were U-CLL and six cases carried *TP53abn* or del(11q). Screening of exon 4 in *RPS15* in an extended CLL series comprising 604 patients with adverse prognostic features revealed mutations in 36 cases (6%). Similar to the discovery cohort, we observed an enrichment of *TP53abn*, *SF3B1* mutations and +12 in *RPS15*-mutated cases. *RPS15*^{mut}/*TP53*^{wt} cases had OS similar to *NOTCH1*^{mut}/*SF3B1*^{mut}/del(11q) and a lower 10-year survival compared to *RPS15*^{wt}/*TP53*^{wt}, pointing to a poor prognosis for *RPS15*-mutated CLL cases.

Since MDM2, a direct regulator of p53, has previously been shown to bind

RPS15, we hypothesized that *RPS15* mutations could impair the ability of *RPS15* to regulate p53 activation via the MDM2/MDM4 axis. We therefore investigated the functional implications of two *RPS15* mutations, *RPS15*^{P131S} and *RPS15*^{G132A}, on the regulation of endogenous p53 *in vitro*. *RPS15*^{mut} displayed a major reduction of p53 stabilization compared to *RPS15*^{wt}.

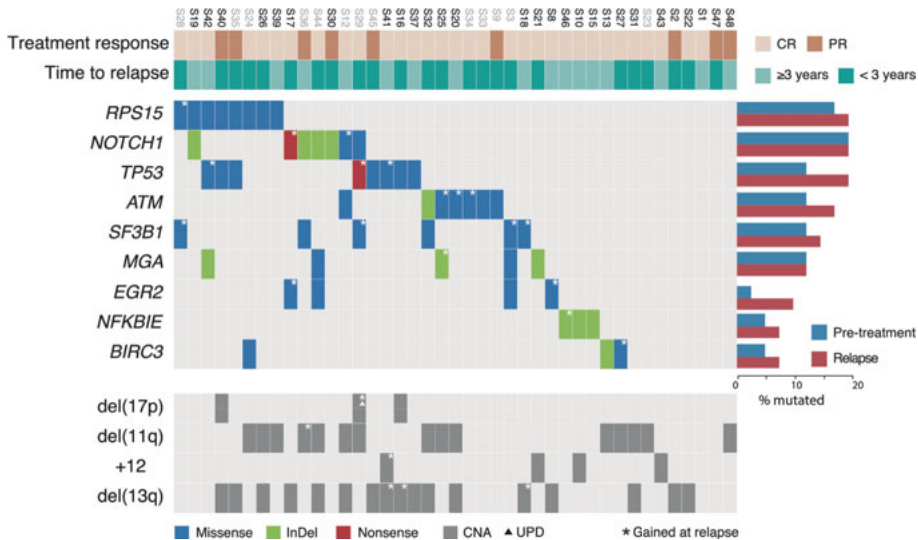


Figure 7. Recurrently mutated genes. Columns represent patients (n=41) and rows genes or genetic lesions. Color-coding indicates the type of mutation or genomic alteration. Case names in gray were analyzed without matched normal DNA. The majority of cases with *TP53* aberrations harbored a mutation without coexisting *del(17p)*; this is explained by the fact that the *TP53* mutation status was not known in most cases before the start of the FCR regime in contrast to fluorescence in situ hybridization detection of *del(17p)*, which had been performed in all cases. All *TP53* mutations were deemed damaging and have been reported previously. CNA, copy-number aberration; PR, partial relapse; UPD, uniparental disomy.

In summary, we performed WES on patients relapsing after FCR therapy and identify mutations in a ribosomal protein as a potential new mechanism involved in CLL evolution. Our findings on *RPS15* are strengthened by the results from Landau *et al.* which proposed *RSP15* as a candidate driver mutation in a large series of FCR relapsed patients.⁸⁷ However, the cellular implications of *RPS15* mutations are still largely unknown and investigating the impact on the ribosomal machinery in presence of *RPS15* mutations may give novel insight into CLL progression and relapse.

Paper IV

The characterization of the full transcriptomic landscape of CLL and the potential impact of transcript diversity on CLL pathogenesis is currently

limited, since only a few studies have thus far been performed using RNA-sequencing.^{132,136} We thus sought to investigate the transcriptome of subgroups of patients carrying stereotyped BcRs and exhibiting distinct clinical and biological characteristics i.e. the poor-prognostic subsets #1 and #2 and the more favorable-prognostic subset #4 (in total 50 patients) using RNA-seq.

We first sought to characterize the subset-specific gene expression signatures and our analysis revealed major differences between the individual subsets. Not surprisingly, every subset showed marked upregulation of distinct IGHV/IGKV/IGLV genes in agreement with the subset-specific IG rearrangements; which also served to confirm our analysis. Comparison of subset #1 and #4 cases demonstrated that more than 1460 gene transcripts were differentially expressed. Two thirds (60.5%) of these transcripts displayed upregulation in subset #1 and were associated with cell proliferation and activation, mirroring the aggressive clinical course observed in patients assigned to this subset.¹⁰⁴ On the other hand, subset #4 cases exhibited a lower level of transcription indicating a lower proliferative ability of the CLL cells in these patients. This finding fits well the more indolent disease course observed in patients assigned to subset #4.¹⁰⁴ The comparison of subset #2 and #4 identified over 640 genes differentially expressed; cell-cycle control and regulation of transcription were among the most enriched pathways. Finally, the comparison between the two poor-prognostic subsets revealed the lowest number of differentially expressed transcripts (n=172).

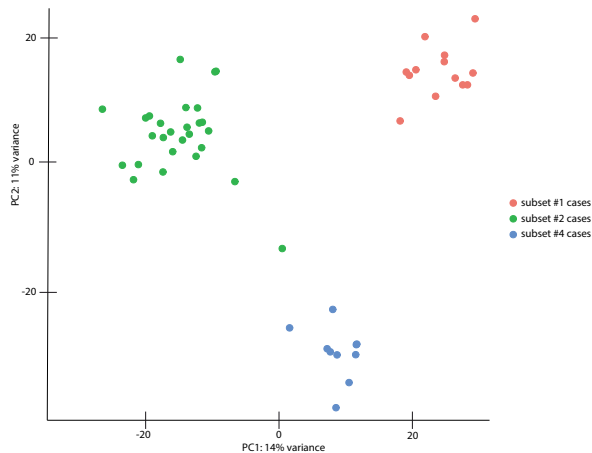


Figure 8. PCA based on gene count of CLL cases assigned to subset #1, #2 and #4.

Moving our focus to non-coding elements, we identified lincRNA as one of the most deregulated classes of non-coding genes amongst the subsets ana-

lyzed. In particular, 98 lincRNA distinguished subset #1 from subset #4. Interestingly, the most differentially expressed lincRNAs, ZNF667-AS1 is one such example, were completely absent in one subset while being highly overexpressed in another. These extreme differences in expression indicate a role for lincRNAs in CLL pathogenesis, potentially through a dysregulated enhancer/silencer mechanism.

A growing body of evidence points towards the involvement of AS in the pathogenesis of CLL. Our group recently reported that approximately 45% of subset #2 cases harbor *SF3B1* mutations^{102,104} but the biological significance of such mutations remains unclear. We therefore thought to compare the transcriptome of subset #2 cases carrying *SF3B1* mutations with subset #2 cases wildtype for *SF3B1*, however, no significant difference in gene expression were observed. That said, analysis of splicing events between these two subgroups revealed 187 differentially expressed transcripts, many of which are involved in chromatin remodeling, NF- κ B pathway and ribosome biogenesis. Intriguingly, alternative transcripts were seen in *SF3B1* mutated cases in *RPL22L1*, *RPL24*, *RPL31* and *RPL32*, all involved in non-sense mediated decay, a surveillance mechanism linked to splicing to reduce errors in translation. Of note, results from a study on alternative splicing events associated to *SF3B1* mutations in breast cancer, largely overlap with our findings. In particular, Maguire *et al.* identified *RPL24* and *RPL31* as gene targets of deregulation¹⁶⁷ and we could confirm the inclusion/exclusion of the same exons, supporting the idea that the splicing deregulation is highly selective.

Concluding Remarks

As described in this thesis, by applying different types of NGS technologies, we could show the feasibility of targeted NGS for clinical diagnostics and identify novel putative drivers involved in progression and relapse in CLL. Furthermore, NGS not only enabled us to investigate the genomic landscape but also to explore the transcriptome of CLL.

Considering the expanding availability of NGS technologies, the rate at which we can identify novel mutations/mechanisms is rapidly increasing; however for many of these gene mutations we do not know their functional role in disease ontogeny and evolution. For several of these mutations, we could confirm their proposed prognostic relevance and showed that targeted NGS could be used to screen for those mutations. In the end, risk-stratification of early-stage CLL patients will probably benefit from integration of recurrent mutations with chromosomal aberrations and the IGHV mutational status, although this has to be further investigated in larger patient cohorts and in prospective trials.

In recent years, several studies have reported that subclonal *TP53* mutations, at time of diagnosis, can represent a risk for later disease progression and relapse. Hence, the detection of low frequency mutations at an early stage could become essential for risk-based design of therapy. Sanger sequencing is limited by a detection threshold of 10-20%, far from the allelic frequencies of subclonal events. Targeted NGS panels incorporating UMI with high sequence depth for areas of recurrently mutated genes of particular interest could here represent a good solution, especially considering the fast pace at which sequencing technologies are evolving. Moreover, an NGS approach that would allow simultaneous detection of SNVs, indels as well as chromosomal aberrations would be highly warranted.

The molecular mechanisms behind CLL progression have not been fully elucidated. In our longitudinal WES study, we identified mutations in the ribosomal *RSP15* gene that appeared to associate with a worse prognosis. While we found a link between *RPS15* mutations and impaired p53 stability, further analysis is now needed to define the functional consequences of these

mutations including studies of the impact on translation. In this regard, it is noteworthy to mention that additional ribosomal proteins have been found mutated in other malignancies, such as in T-cell acute lymphoblastic leukemia (T-ALL).

By transcriptome sequencing, we revealed distinct gene expression signatures associated with each stereotyped CLL subset and identified extreme differences in lincRNA expression between subsets, a finding that merits further investigation. Furthermore, analysis of *SF3B1*-mutated subset #2 cases revealed over 180 deregulated transcripts of genes involved in chromatin remodeling, NF- κ B pathway and ribosomal biogenesis, further supporting a role for AS in CLL pathobiology. We hence conclude that a global approach such as RNA-seq has the potential to advance our knowledge of the molecular mechanisms responsible for CLL development. Indeed, the integration of genomic data with information from gene expression and AS can move the focus from the specific mutation to the dysregulated pathway providing an attractive model for the identification of prognostic markers and therapeutic targets.

Acknowledgments

Everything I did in the lab over the last four years began with a list of cases to be investigated. Every row in that list is a patient, is a person with wishes and dreams that unluckily was diagnosed with leukemia. I never met in person those **patients** but I feel the moral need to thank them for their trust in Medical Research in the attempt to give hope to future patients.

I want to express my most sincere gratitude to my main supervisor, Professor **Richard Rosenquist**. He accepted me in his lab and gave me the chance to grow in the leukemia field; when I first heard about DNA sequencing I thought it was a dream. Thank you Richard for this huge opportunity; I learnt a lot from you.

I want to thank my supervisors **Larry Mansouri** and **Lesley-Ann Sutton** for their precious guidance and for all the time spent working on my project to make them work. Larry, thank you for being always concerned about me; moving abroad it was nice to feel safe just by coming to your office. Lesley, thank you for making the projects running and to keep up when things went wrong. We know very well that sequencing hasn't always been nice to us but you constantly had a positive thinking and found the energy to start again.

My colleagues and friends **Panos, Viktor, Emma, Sujata, Mattias, Karin** and **Aron**. You are among the nicest and most helpful people I ever met. Wherever we will end up, I will bring with me memories of good moments I spent with any of you. Thank you! **Tanja**, you taught me all the cell work. You really have a lot of patience, even when I wanted to kick myself out of the lab, you made me feel like everything would be fine. Thank you!

My gratitude goes also to Professor **Giorgio Inghirami** that back in 2010 accepted to his lab in Turin a master student with a burning interest in leukemia; he showed me the magic of lymphocytes and introduced me to the world of medical research.

I want to take the chance to thank **Christina Magnusson** and **Helene Norlin** from the administration for their valuable support. You also made this thesis possible, thank you!

I am grateful to the **Funding Agencies** that made the research behind this thesis possible. In particular the **Nordic Cancer Union**; the **Swedish Cancer Society**; the **Swedish Medical Research Council**; **Lion's Cancer fund, Uppsala**, and the **Selander's Foundation**, Uppsala, Sweden; and "AEGLE, An analytics framework for integrated and personalized healthcare services in Europe," by the European Commission. The **SNP&SEQ Technology Platform, SciLifeLab** at Uppsala University, performed sequencing. The computations were performed on resources provided by **Swedish National Infrastructure for Computing** through Uppsala Multidisciplinary Center for Advanced Computational Science. Fluorescence-activated cell sorting was performed at the **BioVis imaging facility**, SciLifeLab, Uppsala.

My source of ancient greek words **Argyris, Grammatiki** and **Maria** ... you are excellent teachers, but most of all you are splendid friends!

My dearest friends **Andrea, Checco, Elisa, Erika, Meo** and **Sergio**. We are far apart but we are very close, thank you for your true friendship.

My "little Italy" **Alberto, Arianna, Belinda, Chiara, Giuseppe, Juan, Luca, Paola, Richi, Toni, Valentina, Valeria**. I am glad we met; with you and some good wine Sweden is a warmer place. **Matteo**, thank you also for being my PhD mate and for coffee in Navet.

My Argentinian hermana **Belen, Daniela** and **Pelle, Xing, Jens** and **Astrid**, you are precious friends.

Francesca "socc"; I feel so fortunate we met again in Sweden, thank you for all the laugh and for our very long scientific discussions.

My wife **Francesca**; you have always been with me. Thank you for believing more than me that I could make this PhD. My little **Niccolò**, thank you for keeping me awake at nights; you already know that it is when ideas rise.

My roots: my dad **Enrico**, my mom **Claudia**, my brother **Mauro** and **Antonella**, my uncle **Valter**; you supported and encouraged me all the way since childhood to chase my dreams. Thank you. Grazie alle mie radici: papá, mamma, Mauro e Anto, zio Valter; mi incoraggiate da sempre a inseguire i miei sogni. Questo successo é anche vostro.

References

1. Hallek, M. *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* **111**, 5446–56 (2008).
2. Eichhorst, B., Dreyling, M., Robak, T., Montserrat, E. & Hallek, M. Chronic lymphocytic leukemia: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann. Oncol.* **22 Suppl 6**, vi50-4 (2011).
3. Strati, P. *et al.* Monoclonal B-cell lymphocytosis and early-stage chronic lymphocytic leukemia: diagnosis, natural history, and risk stratification. *Blood* **126**, 454–62 (2015).
4. Chiorazzi, N., Rai, K. R. & Ferrarini, M. Chronic lymphocytic leukemia. *N. Engl. J. Med.* **352**, 804–15 (2005).
5. Baliakas, P., Mattsson, M., Stamatopoulos, K. & Rosenquist, R. Prognostic indices in chronic lymphocytic leukaemia: where do we stand how do we proceed? *J. Intern. Med.* **279**, 347–357 (2016).
6. Rai, K. R. *et al.* Clinical staging of chronic lymphocytic leukemia. *Blood* **46**, 219–34 (1975).
7. Binet, J. L. *et al.* A new prognostic classification of chronic lymphocytic leukemia derived from a multivariate survival analysis. *Cancer* **48**, 198–206 (1981).
8. Fischer, K. *et al.* Long-term remissions after FCR chemoimmunotherapy in previously untreated patients with CLL: updated results of the CLL8 trial. *Blood* **127**, 208–15 (2016).
9. Fischer, K. *et al.* Bendamustine in combination with rituximab for previously untreated patients with chronic lymphocytic leukemia: a multicenter phase II trial of the German Chronic Lymphocytic Leukemia Study Group. *J. Clin. Oncol.* **30**, 3209–16 (2012).
10. Rosenquist, R., Cortese, D., Bhoi, S., Mansouri, L. & Gunnarsson, R. Prognostic markers and their clinical applicability in chronic lymphocytic leukemia: where do we stand? *Leuk. Lymphoma* **54**, 2351–64 (2013).
11. Wang, L. *et al.* SF3B1 and other novel cancer genes in chronic lymphocytic leukemia. *N. Engl. J. Med.* **365**, 2497–506 (2011).
12. Del Giudice, I. *et al.* NOTCH1 mutations in +12 chronic lymphocytic leukemia (CLL) confer an unfavorable prognosis, induce a distinctive

- transcriptional profiling and refine the intermediate prognosis of +12 CLL. *Haematologica* **97**, 437–41 (2012).
13. Jeromin, S. *et al.* SF3B1 mutations correlated to cytogenetics and mutations in NOTCH1, FBXW7, MYD88, XPO1 and TP53 in 1160 untreated CLL patients. *Leukemia* **28**, 108–17 (2014).
 14. Oscier, D. G. *et al.* The clinical significance of NOTCH1 and SF3B1 mutations in the UK LRF CLL4 trial. *Blood* **121**, 468–75 (2013).
 15. Rossi, D. *et al.* Disruption of BIRC3 associates with fludarabine chemorefractoriness in TP53 wild-type chronic lymphocytic leukemia. *Blood* **119**, 2854–62 (2012).
 16. Sutton, L.-A. & Rosenquist, R. The complex interplay between cell-intrinsic and cell-extrinsic factors driving the evolution of chronic lymphocytic leukemia. *Semin. Cancer Biol.* **34**, 22–35 (2015).
 17. Chiorazzi, N. Cell proliferation and death: forgotten features of chronic lymphocytic leukemia B cells. *Best Pract. Res. Clin. Haematol.* **20**, 399–413 (2007).
 18. Burger, J. A. & Gribben, J. G. The microenvironment in chronic lymphocytic leukemia (CLL) and other B cell malignancies: Insight into disease biology and new targeted therapies. *Semin. Cancer Biol.* **24**, 71–81 (2014).
 19. Burger, J. A. Nurture versus nature: the microenvironment in chronic lymphocytic leukemia. *Hematology Am. Soc. Hematol. Educ. Program* **2011**, 96–103 (2011).
 20. Fabbri, G. & Dalla-Favera, R. The molecular pathogenesis of chronic lymphocytic leukaemia. *Nat. Rev. Cancer* **16**, 145–162 (2016).
 21. Sivina, M. *et al.* Marrow Stromal Cells Induce TCL1 Expression in Chronic Lymphocytic Leukemia B Cells. *Blood* **114**, 2347–2347 (2009).
 22. Herishanu, Y. *et al.* The lymph node microenvironment promotes B-cell receptor signaling, NF- κ B activation, and tumor proliferation in chronic lymphocytic leukemia. *Blood* **117**, (2011).
 23. Burger, J. A. *et al.* Blood-derived nurse-like cells protect chronic lymphocytic leukemia B cells from spontaneous apoptosis through stromal cell-derived factor-1. *Blood* **96**, 2655–63 (2000).
 24. Gautam, S. *et al.* Reprogramming Nurse-like Cells with Interferon γ to Interrupt Chronic Lymphocytic Leukemia Cell Survival. *J. Biol. Chem.* **291**, 14356–14362 (2016).
 25. Döhner, H. *et al.* Genomic aberrations and survival in chronic lymphocytic leukemia. *N. Engl. J. Med.* **343**, 1910–6 (2000).
 26. Gahrton, G., Robèrt, K. H., Friberg, K., Zech, L. & Bird, A. G. Extra chromosome 12 in chronic lymphocytic leukaemia. *Lancet (London, England)* **1**, 146–7 (1980).
 27. Juliusson, G. *et al.* Prognostic subgroups in B-cell chronic lymphocytic leukemia defined by specific chromosomal abnormalities. *N. Engl. J. Med.* **323**, 720–4 (1990).

28. Zech, L. & Mellstedt, H. Chromosome 13--a new marker for B-cell chronic lymphocytic leukemia. *Hereditas* **108**, 77–84 (1988).
29. Matutes, E. *et al.* Trisomy 12 defines a group of CLL with atypical morphology: correlation between cytogenetic, clinical and laboratory features in 544 patients. *Br. J. Haematol.* **92**, 382–8 (1996).
30. Gunnarsson, R. *et al.* Array-based genomic screening at diagnosis and during follow-up in chronic lymphocytic leukemia. *Haematologica* **96**, 1161–9 (2011).
31. Rossi, D. *et al.* Integrated mutational and cytogenetic analysis identifies new prognostic subgroups in chronic lymphocytic leukemia. *Blood* **121**, 1403–12 (2013).
32. Cimmino, A. *et al.* miR-15 and miR-16 induce apoptosis by targeting BCL2. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 13944–9 (2005).
33. Parker, H. *et al.* 13q deletion anatomy and disease progression in patients with chronic lymphocytic leukemia. *Leukemia* **25**, 489–97 (2011).
34. Ouillette, P. *et al.* The prognostic significance of various 13q14 deletions in chronic lymphocytic leukemia. *Clin. Cancer Res.* **17**, 6778–90 (2011).
35. Dal Bo, M. *et al.* 13q14 deletion size and number of deleted cells both influence prognosis in chronic lymphocytic leukemia. *Genes. Chromosomes Cancer* **50**, 633–43 (2011).
36. Hallek, M. *et al.* Addition of rituximab to fludarabine and cyclophosphamide in patients with chronic lymphocytic leukaemia: a randomised, open-label, phase 3 trial. *Lancet* **376**, 1164–74 (2010).
37. Schaffner, C., Stilgenbauer, S., Rappold, G. A., Döhner, H. & Lichter, P. Somatic ATM mutations indicate a pathogenic role of ATM in B-cell chronic lymphocytic leukemia. *Blood* **94**, 748–53 (1999).
38. Stankovic, T. *et al.* Inactivation of ataxia telangiectasia mutated gene in B-cell chronic lymphocytic leukaemia. *Lancet (London, England)* **353**, 26–9 (1999).
39. Austen, B. *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.* **25**, 5448–57 (2007).
40. Zenz, T. *et al.* TP53 mutation and survival in chronic lymphocytic leukemia. *J. Clin. Oncol.* **28**, 4473–9 (2010).
41. Zenz, T. *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* **112**, (2008).
42. Malcikova, J. *et al.* Monoallelic and biallelic inactivation of TP53 gene in chronic lymphocytic leukemia: selection, impact on survival, and response to DNA damage. *Blood* **114**, (2009).
43. Schnaiter, A. *et al.* NOTCH1, SF3B1, and TP53 mutations in

- fludarabine-refractory CLL patients treated with alemtuzumab: results from the CLL2H trial of the GCLLSG. *Blood* **122**, 1266–70 (2013).
44. Zenz, T. *et al.* Detailed analysis of p53 pathway defects in fludarabine-refractory chronic lymphocytic leukemia (CLL): dissecting the contribution of 17p deletion, TP53 mutation, p53-p21 dysfunction, and miR34a in a prospective clinical trial. *Blood* **114**, 2589–97 (2009).
 45. Burger, J. A. *et al.* Safety and activity of ibrutinib plus rituximab for patients with high-risk chronic lymphocytic leukaemia: a single-arm, phase 2 study. *Lancet. Oncol.* **15**, 1090–9 (2014).
 46. Woyach, J. A. & Johnson, A. J. Targeted therapies in CLL: mechanisms of resistance and strategies for management. *Blood* **126**, 471–477 (2015).
 47. Rosenquist, R. *et al.* Clinical impact of recurrently mutated genes on lymphoma diagnostics: state-of-the-art and beyond. *Haematologica* **101**, 1002–9 (2016).
 48. Hamblin, T. J., Davis, Z., Gardiner, A., Oscier, D. G. & Stevenson, F. K. Unmutated Ig V(H) genes are associated with a more aggressive form of chronic lymphocytic leukemia. *Blood* **94**, 1848–54 (1999).
 49. Damle, R. N. *et al.* Ig V Gene Mutation Status and CD38 Expression As Novel Prognostic Indicators in Chronic Lymphocytic Leukemia. *Blood* **94**, (1999).
 50. Maloum, K. *et al.* Expression of unmutated V H genes is a detrimental prognostic factor in chronic lymphocytic leukemia. *Blood* **96**, (2000).
 51. Klein, U. *et al.* Gene expression profiling of B cell chronic lymphocytic leukemia reveals a homogeneous phenotype related to memory B cells. *J. Exp. Med.* **194**, 1625–38 (2001).
 52. Rosenwald, A. *et al.* Relation of gene expression phenotype to immunoglobulin mutation genotype in B cell chronic lymphocytic leukemia. *J. Exp. Med.* **194**, 1639–47 (2001).
 53. Stevenson, F. K. & Caligaris-Cappio, F. Chronic lymphocytic leukemia: revelations from the B-cell receptor. *Blood* **103**, 4389–95 (2004).
 54. Damle, R. N. *et al.* B-cell chronic lymphocytic leukemia cells express a surface membrane phenotype of activated, antigen-experienced B lymphocytes. *Blood* **99**, (2002).
 55. Oscier, D. G. *et al.* Multivariate analysis of prognostic factors in CLL: clinical stage, IGVH gene mutational status, and loss or mutation of the p53 gene are independent prognostic factors. *Blood* **100**, 1177–84 (2002).
 56. Bulian, P. *et al.* IGHV gene mutational status and 17p deletion are independent molecular predictors in a comprehensive clinical-biological prognostic model for overall survival prediction in chronic

- lymphocytic leukemia. *J. Transl. Med.* **10**, 18 (2012).
57. Thompson, P. A. *et al.* Fludarabine, cyclophosphamide, and rituximab treatment achieves long-term disease-free survival in IGHV-mutated chronic lymphocytic leukemia. *Blood* **127**, (2016).
 58. Awan, F. T. Cure for CLL? *Blood* **127**, (2016).
 59. Rai, K. *et al.* An international prognostic index for patients with chronic lymphocytic leukaemia (CLL-IPI): a meta-analysis of individual patient data. *Lancet Oncol.* **17**, 779–790 (2016).
 60. Maffei, R. *et al.* Targeting neoplastic B cells and harnessing microenvironment: the “double face” of ibrutinib and idelalisib. *J. Hematol. Oncol.* **8**, 60 (2015).
 61. Forconi, F. *et al.* Three years of ibrutinib in CLL. *Blood* **125**, 2455–6 (2015).
 62. Barr, P. M. *et al.* Phase 2 study of idelalisib and entospletinib: pneumonitis limits combination therapy in relapsed refractory CLL and NHL. *Blood* **127**, 2411–5 (2016).
 63. Cervantes-Gomez, F. *et al.* Pharmacological and protein profiling suggest ABT-199 as optimal partner with ibrutinib in chronic lymphocytic leukemia. *Clin. Cancer Res.* (2015). doi:10.1158/1078-0432.CCR-14-2809
 64. Roberts, A. W. *et al.* Targeting BCL2 with Venetoclax in Relapsed Chronic Lymphocytic Leukemia. *N. Engl. J. Med.* **374**, 311–322 (2016).
 65. Falchi, L., Baron, J. M., Orlikowski, C. A. & Ferrajoli, A. BCR Signaling Inhibitors: an Overview of Toxicities Associated with Ibrutinib and Idelalisib in Patients with Chronic Lymphocytic Leukemia. *Mediterr. J. Hematol. Infect. Dis.* **8**, e2016011 (2016).
 66. Mohamed, A. J. *et al.* Bruton’s tyrosine kinase (Btk): function, regulation, and transformation with special emphasis on the PH domain. *Immunol. Rev.* **228**, 58–73 (2009).
 67. Hendriks, R. W., Yuvaraj, S. & Kil, L. P. Targeting Bruton’s tyrosine kinase in B cell malignancies. *Nat. Rev. Cancer* **14**, 219–232 (2014).
 68. Ibrutinib Improves Survival in Patients with Previously Treated Chronic Lymphocytic Leukemia - National Cancer Institute. Available at: <http://www.cancer.gov/types/leukemia/research/ibrutinib-ctl>. (Accessed: 29th July 2015)
 69. Furman, R. R. *et al.* Ibrutinib Resistance in Chronic Lymphocytic Leukemia. *N. Engl. J. Med.* **370**, 2352–2354 (2014).
 70. Davids, M. S. & Brown, J. R. Phosphoinositide 3’-kinase inhibition in chronic lymphocytic leukemia. *Hematol. Oncol. Clin. North Am.* **27**, 329–39 (2013).
 71. Ortiz-Maldonado, V., García-Morillo, M. & Delgado, J. The biology behind PI3K inhibition in chronic lymphocytic leukaemia. *Ther. Adv. Hematol.* **6**, 25–36 (2015).

72. Herman, S. E. M. *et al.* Phosphatidylinositol 3-kinase- δ inhibitor CAL-101 shows promising preclinical activity in chronic lymphocytic leukemia by antagonizing intrinsic and extrinsic cellular survival signals. *Blood* **116**, 2078–88 (2010).
73. Caligaris-Cappio, F. *et al.* Directing CLL-cell traffic. *Blood* **126**, 1267–8 (2015).
74. Czabotar, P. E., Lessene, G., Strasser, A. & Adams, J. M. Control of apoptosis by the BCL-2 protein family: implications for physiology and therapy. *Nat. Rev. Mol. Cell Biol.* **15**, 49–63 (2013).
75. Yip, K. W. & Reed, J. C. Bcl-2 family proteins and cancer. *Oncogene* **27**, 6398–6406 (2008).
76. Schimmer, A. D., Munk-Pedersen, I., Minden, M. D. & Reed, J. C. Bcl-2 and apoptosis in chronic lymphocytic leukemia. *Curr. Treat. Options Oncol.* **4**, 211–8 (2003).
77. Puente, X. S. *et al.* Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia. *Nature* **475**, 101–5 (2011).
78. Quesada, V. *et al.* Exome sequencing identifies recurrent mutations of the splicing factor SF3B1 gene in chronic lymphocytic leukemia. *Nat. Genet.* **44**, 47–52 (2012).
79. Balatti, V. *et al.* NOTCH1 mutations in CLL associated with trisomy 12. *Blood* **119**, 329–31 (2012).
80. Puente, X. S. *et al.* Non-coding recurrent mutations in chronic lymphocytic leukaemia. *Nature* **526**, 519–524 (2015).
81. Schuh, A. *et al.* Monitoring chronic lymphocytic leukemia progression by whole genome sequencing reveals heterogeneous clonal evolution patterns. *Blood* **120**, 4191–6 (2012).
82. Rose-Zerilli, M. J. J. *et al.* Longitudinal copy number, whole exome and targeted deep sequencing of ‘good risk’ IGHV-mutated CLL patients with progressive disease. *Leukemia* **30**, 1301–10 (2016).
83. Landau, D. A. *et al.* Evolution and impact of subclonal mutations in chronic lymphocytic leukemia. *Cell* **152**, 714–26 (2013).
84. Quidville, V. *et al.* Targeting the deregulated spliceosome core machinery in cancer cells triggers mTOR blockade and autophagy. *Cancer Res.* **73**, 2247–58 (2013).
85. Fabbri, G. *et al.* Analysis of the chronic lymphocytic leukemia coding genome: role of NOTCH1 mutational activation. *J. Exp. Med.* **208**, 1389–401 (2011).
86. Rodríguez, D. *et al.* Mutations in CHD2 cause defective association with active chromatin in chronic lymphocytic leukemia. *Blood* **126**, 195–202 (2015).
87. Landau, D. A. *et al.* Mutations driving CLL and their evolution in progression and relapse. *Nature* **526**, 525–30 (2015).
88. Puente, X. S. & López-Otín, C. The evolutionary biography of chronic lymphocytic leukemia. *Nat. Genet.* **45**, 229–31 (2013).

89. Sutton, L.-A. & Rosenquist, R. Deciphering the molecular landscape in chronic lymphocytic leukemia: time frame of disease evolution. *Haematologica* **100**, (2014).
90. Wan, Y. & Wu, C. J. SF3B1 mutations in chronic lymphocytic leukemia. *Blood* **121**, 4627–34 (2013).
91. Mansouri, L. *et al.* Functional loss of $\text{I}\kappa\text{B}\epsilon$ leads to NF- κB deregulation in aggressive chronic lymphocytic leukemia. *J. Exp. Med.* **212**, 833–43 (2015).
92. Tuveson, D. & Rai, K. R. Augmenting NF- κB in poor-risk CLL: A general paradigm for other cancers? *J. Exp. Med.* **212**, 830–1 (2015).
93. Dotto, G. P. Crosstalk of Notch with p53 and p63 in cancer growth control. *Nat. Rev. Cancer* **9**, 587–95 (2009).
94. Mansouri, L. *et al.* Short telomere length is associated with NOTCH1/SF3B1/TP53 aberrations and poor outcome in newly diagnosed chronic lymphocytic leukemia patients. *Am. J. Hematol.* **88**, 647–51 (2013).
95. Baliakas, P. *et al.* Recurrent mutations refine prognosis in chronic lymphocytic leukemia. *Leukemia* (2014). doi:10.1038/leu.2014.196
96. Bertrand, M. J. M. *et al.* cIAP1/2 are direct E3 ligases conjugating diverse types of ubiquitin chains to receptor interacting proteins kinases 1 to 4 (RIP1-4). *PLoS One* **6**, e22356 (2011).
97. Rose-Zerilli, M. J. J. *et al.* ATM mutation rather than BIRC3 deletion and/or mutation predicts reduced survival in 11q-deleted chronic lymphocytic leukemia: data from the UK LRF CLL4 trial. *Haematologica* **99**, 736–42 (2014).
98. Martínez-Trillos, A. *et al.* Mutations in TLR/MYD88 pathway identify a subset of young chronic lymphocytic leukemia patients with favorable outcome. *Blood* **123**, 3790–6 (2014).
99. Baliakas, P. *et al.* Prognostic relevance of MYD88 mutations in CLL: the jury is still out. *Blood* **126**, 1043–4 (2015).
100. Will, C. L. & Lührmann, R. Spliceosome structure and function. *Cold Spring Harb. Perspect. Biol.* **3**, (2011).
101. Rossi, D. *et al.* Mutations of the SF3B1 splicing factor in chronic lymphocytic leukemia: association with progression and fludarabine-refractoriness. *Blood* **118**, 6904–8 (2011).
102. Strefford, J. C. *et al.* Distinct patterns of novel gene mutations in poor-prognostic stereotyped subsets of chronic lymphocytic leukemia: the case of SF3B1 and subset #2. *Leukemia* **27**, 2196–9 (2013).
103. Cortese, D. *et al.* On the way towards a ‘CLL prognostic index’: focus on TP53, BIRC3, SF3B1, NOTCH1 and MYD88 in a population-based cohort. *Leukemia* **28**, 710–3 (2014).
104. Sutton, L. A. *et al.* Different spectra of recurrent gene mutations in subsets of chronic lymphocytic leukemia harboring stereotyped B-cell receptors. *Haematologica* haematol.2016.141812- (2016).

- doi:10.3324/haematol.2016.141812
105. Wan, Y. & Wu, C. J. SF3B1 mutations in chronic lymphocytic leukemia. *Blood* **121**, 4627–34 (2013).
 106. Gentien, D. *et al.* A common alternative splicing signature is associated with SF3B1 mutations in malignancies from different cell lineages. *Leukemia* **28**, 1355–7 (2014).
 107. Te Raa, G. D. *et al.* The impact of SF3B1 mutations in CLL on the DNA-damage response. *Leukemia* (2014). doi:10.1038/leu.2014.318
 108. Sundaramoorthy, S., Vázquez-Novelle, M. D., Lekomtsev, S., Howell, M. & Petronczki, M. Functional genomics identifies a requirement of pre-mRNA splicing factors for sister chromatid cohesion. *EMBO J.* **33**, 2623–42 (2014).
 109. Watrin, E., Demidova, M., Watrin, T., Hu, Z. & Prigent, C. Sororin pre-mRNA splicing is required for proper sister chromatid cohesion in human cells. *EMBO Rep.* **15**, 948–55 (2014).
 110. Kfir, N. *et al.* SF3B1 association with chromatin determines splicing outcomes. *Cell Rep.* **11**, 618–29 (2015).
 111. Shanafelt, T. D. *et al.* Karyotype evolution on fluorescent in situ hybridization analysis is associated with short survival in patients with chronic lymphocytic leukemia and is related to CD49d expression. *J. Clin. Oncol.* **26**, e5-6 (2008).
 112. Stilgenbauer, S. *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* **92**, 1242–5 (2007).
 113. Grubor, V. *et al.* Novel genomic alterations and clonal evolution in chronic lymphocytic leukemia revealed by representational oligonucleotide microarray analysis (ROMA). *Blood* **113**, 1294–303 (2009).
 114. Ouillette, P. *et al.* Clonal evolution, genomic drivers, and effects of therapy in chronic lymphocytic leukemia. *Clin. Cancer Res.* **19**, 2893–904 (2013).
 115. Rossi, D. *et al.* Clinical impact of small TP53 mutated subclones in chronic lymphocytic leukemia. *Blood* **123**, 2139–47 (2014).
 116. Malcikova, J. *et al.* Detailed analysis of therapy-driven clonal evolution of TP53 mutations in chronic lymphocytic leukemia. *Leukemia* **29**, 877–85 (2015).
 117. Rasi, S. *et al.* Clinical impact of small subclones harboring NOTCH1, SF3B1 or BIRC3 mutations in chronic lymphocytic leukemia. *Haematologica* **101**, (2016).
 118. Wang, Z., Gerstein, M. & Snyder, M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* **10**, 57–63 (2009).
 119. Zhao, W. *et al.* Comparison of RNA-Seq by poly (A) capture, ribosomal RNA depletion, and DNA microarray for expression profiling. *BMC Genomics* **15**, 419 (2014).

120. O'Neil, D., Glowatz, H. & Schlumpberger, M. Ribosomal RNA depletion for efficient use of RNA-seq capacity. *Curr. Protoc. Mol. Biol.* **Chapter 4**, Unit 4.19 (2013).
121. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
122. Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105–11 (2009).
123. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
124. Robinson, M. D. & Oshlack, A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* **11**, R25 (2010).
125. Anders, S., Reyes, A. & Huber, W. Detecting differential usage of exons from RNA-seq data. *Genome Res.* **22**, 2008–17 (2012).
126. Haslinger, C. *et al.* Microarray gene expression profiling of B-cell chronic lymphocytic leukemia subgroups defined by genomic aberrations and VH mutation status. *J. Clin. Oncol.* **22**, 3937–49 (2004).
127. Vasconcelos, Y. *et al.* Gene expression profiling of chronic lymphocytic leukemia can discriminate cases with stable disease and mutated Ig genes from those with progressive disease and unmutated Ig genes. *Leukemia* **19**, 2002–2005 (2005).
128. Marincevic, M. *et al.* Distinct gene expression profiles in subsets of chronic lymphocytic leukemia expressing stereotyped IGHV4-34 B-cell receptors. *Haematologica* **95**, 2072–9 (2010).
129. Ecker, S. *et al.* Higher gene expression variability in the more aggressive subtype of chronic lymphocytic leukemia. *Genome Med.* **7**, 8 (2015).
130. Klein, U. *et al.* Transcriptional analysis of the B cell germinal center reaction. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 2639–44 (2003).
131. Seifert, M. *et al.* Cellular origin and pathophysiology of chronic lymphocytic leukemia. *J. Exp. Med.* **209**, 2183–98 (2012).
132. Ferreira, P. G. *et al.* Transcriptome characterization by RNA sequencing identifies a major molecular and clinical subdivision in chronic lymphocytic leukemia. *Genome Res.* **24**, 212–26 (2014).
133. Darman, R. B. *et al.* Cancer-Associated SF3B1 Hotspot Mutations Induce Cryptic 3 Splice Site Selection through Use of a Different Branch Point Splice Site Selection through Use of a Different Branch Point. *CellReports* **13**, 1033–1045 (2015).
134. Alsafadi, S. *et al.* Cancer-associated SF3B1 mutations affect alternative splicing by promoting alternative branchpoint usage. *Nat. Commun.* **7**, 10615 (2016).
135. Ronchetti, D. *et al.* Small nucleolar RNAs as new biomarkers in chronic lymphocytic leukemia. *BMC Med. Genomics* **6**, 27 (2013).

136. Mansouri, L. *et al.* Next generation RNA-sequencing in prognostic subsets of chronic lymphocytic leukemia. *Am. J. Hematol.* **87**, 737–40 (2012).
137. Surprising Splicing: The New Most Frequent Class of Genetic Alteration in Myelodysplastic Syndromes. (2016).
138. Lodomery, M. Aberrant alternative splicing is another hallmark of cancer. *Int. J. Cell Biol.* **2013**, 463786 (2013).
139. Oltean, S. & Bates, D. O. Hallmarks of alternative splicing in cancer. *Oncogene* **33**, 5311–5318 (2014).
140. Sveen, A., Kilpinen, S., Ruusulehto, A., Lothe, R. A. & Skotheim, R. I. Aberrant RNA splicing in cancer; expression changes and driver mutations of splicing factor genes. *Oncogene* **35**, 2413–2427 (2016).
141. Rickert, R. C. New insights into pre-BCR and BCR signalling with relevance to B cell malignancies. *Nat. Rev. Immunol.* **13**, 578–591 (2013).
142. Treanor, B. B-cell receptor: from resting state to activate. *Immunology* **136**, 21–7 (2012).
143. Staudt, L. M. Oncogenic activation of NF-kappaB. *Cold Spring Harb. Perspect. Biol.* **2**, a000109 (2010).
144. Schroeder, H. W. & Cavacini, L. Structure and function of immunoglobulins. *J. Allergy Clin. Immunol.* **125**, S41-52 (2010).
145. Li, Z., Woo, C. J., Iglesias-Ussel, M. D., Ronai, D. & Scharff, M. D. The generation of antibody diversity through somatic hypermutation and class switch recombination. *Genes Dev.* **18**, 1–11 (2004).
146. Davis, A. J. & Chen, D. J. DNA double strand break repair via non-homologous end-joining. *Transl. Cancer Res.* **2**, 130–143 (2013).
147. Kurosaki, T., Kometani, K. & Ise, W. Memory B cells. *Nat. Rev. Immunol.* **15**, 149–159 (2015).
148. Pieper, K., Grimbacher, B. & Eibel, H. B-cell biology and development. *J. Allergy Clin. Immunol.* **131**, 959–71 (2013).
149. De Silva, N. S. & Klein, U. Dynamics of B cells in germinal centres. *Nat. Rev. Immunol.* **15**, 137–148 (2015).
150. Cerutti, A., Cols, M. & Puga, I. Marginal zone B cells: virtues of innate-like antibody-producing lymphocytes. *Nat. Rev. Immunol.* **13**, 118–132 (2013).
151. Stamatopoulos, K. *et al.* Immunoglobulin heavy- and light-chain repertoire in splenic marginal zone lymphoma. *Mol. Med.* **10**, 89–95 (2004).
152. Murray, F. *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* **111**, 1524–33 (2008).
153. Agathangelidis, A. *et al.* Stereotyped B-cell receptors in one-third of chronic lymphocytic leukemia: a molecular classification with implications for targeted therapies. *Blood* **119**, 4467–75 (2012).

154. Messmer, B. T. *et al.* Multiple distinct sets of stereotyped antigen receptors indicate a role for antigen in promoting chronic lymphocytic leukemia. *J. Exp. Med.* **200**, 519–25 (2004).
155. Schroeder, H. W. & Dighiero, G. The pathogenesis of chronic lymphocytic leukemia: analysis of the antibody repertoire. *Immunol. Today* **15**, 288–94 (1994).
156. Efremov, D. G. *et al.* Restricted immunoglobulin VH region repertoire in chronic lymphocytic leukemia patients with autoimmune hemolytic anemia. *Blood* **87**, 3869–76 (1996).
157. Fais, F. *et al.* Chronic lymphocytic leukemia B cells express restricted sets of mutated and unmutated antigen receptors. *J. Clin. Invest.* **102**, 1515–25 (1998).
158. Tobin, G. *et al.* Somatically mutated Ig V(H)3-21 genes characterize a new subset of chronic lymphocytic leukemia. *Blood* **99**, 2262–4 (2002).
159. Strefford, J. C. *et al.* Distinct patterns of novel gene mutations in poor-prognostic stereotyped subsets of chronic lymphocytic leukemia: the case of SF3B1 and subset #2. *Leukemia* **27**, 2196–9 (2013).
160. Cahill, N. *et al.* 450K-array analysis of chronic lymphocytic leukemia cells reveals global DNA methylation to be relatively stable over time and similar in resting and proliferative compartments. *Leukemia* **27**, 150–8 (2013).
161. Swederlow, S. *et al.* *WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues*. (IARC, 2008).
162. Eichhorst, B. *et al.* Chronic lymphocytic leukaemia: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann. Oncol.* **26 Suppl 5**, v78-84 (2015).
163. Hallek, M. *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* **111**, 5446–56 (2008).
164. Sturn, A., Quackenbush, J. & Trajanoski, Z. Genesis: cluster analysis of microarray data. *Bioinformatics* **18**, 207–8 (2002).
165. Sutton, L.-A. *et al.* Targeted next-generation sequencing in chronic lymphocytic leukemia: a high-throughput yet tailored approach will facilitate implementation in a clinical setting. *Haematologica* **100**, 370–6 (2015).
166. Brown, J. R. The treatment of relapsed refractory chronic lymphocytic leukemia. *Hematology Am. Soc. Hematol. Educ. Program* **2011**, 110–8 (2011).
167. Maguire, S. L. *et al.* SF3B1 mutations constitute a novel therapeutic target in breast cancer. *J. Pathol.* **235**, 571–580 (2015).

Acta Universitatis Upsaliensis

*Digital Comprehensive Summaries of Uppsala Dissertations
from the Faculty of Medicine 1259*

Editor: The Dean of the Faculty of Medicine

A doctoral dissertation from the Faculty of Medicine, Uppsala University, is usually a summary of a number of papers. A few copies of the complete dissertation are kept at major Swedish research libraries, while the summary alone is distributed internationally through the series Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine. (Prior to January, 2005, the series was published under the title "Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine".)

Distribution: publications.uu.se
urn:nbn:se:uu:diva-303703



ACTA
UNIVERSITATIS
UPSALIENSIS
UPPSALA
2016