

Aus dem Zentrum für Pathologie der Universität zu Köln
Institut für Allgemeine Pathologie und Pathologische Anatomie
Direktor: Universitätsprofessor Dr. med. R. Büttner

**Targeting Disease Related Genes in Chronic Lymphocytic
Leukemia by a Multiplex PCR Approach Followed by Massive
Parallel Sequencing**

Inaugural-Dissertation zur Erlangung der Würde eines doctor rerum medicinalium
der Hohen Medizinischen Fakultät
der Universität zu Köln

vorgelegt von
Claudia Vollbrecht
aus Berlin

promoviert am 10. Juni 2015

COPY CLARA GbR, Berlin

Gedruckt mit Genehmigung der Medizinischen Fakultät der Universität zu Köln, 2015

Dekan: Universitätsprofessor Dr. med. Dr. h. c. Th. Krieg

1. Berichterstatter: Universitätsprofessor Dr. med. R. Büttner
2. Berichterstatter: Universitätsprofessor Dr. med. C. Reinhardt

Erklärung

Ich erkläre hiermit, dass ich die vorliegende Dissertationsschrift ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe; die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sind als solche kenntlich gemacht.

Bei der Auswahl und Auswertung des Materials, sowie bei der Herstellung des Manuskripts habe ich Unterstützungsleistungen von folgenden Personen erhalten:

Privatdozentin Dr. Margarete Odenthal (Institut für Pathologie der Universitätsklinik Köln)

Dr. Carmen-Diana Herling (geb. Schweighofer) (Klinik für Innere Medizin I der Universitätsklinik Köln)

Weitere Personen waren an der geistigen Herstellung der vorliegenden Arbeit nicht beteiligt. Insbesondere habe ich nicht die Hilfe einer Promotionsberaterin/eines Promotionsberaters in Anspruch genommen. Dritte haben von mir weder unmittelbar noch mittelbar geldwerte Leistungen für Arbeiten erhalten, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen.

Die Dissertationsschrift wurde von mir bisher weder im Inland noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt.

Köln, den 08.12.2014

Die dieser Arbeit zugrunde liegenden Patientendaten, sowie die DNA-Aufreinigung wurden im Labor der Klinik für Inneren Medizin I, Universitätsklinik Köln unter der Leitung von Frau Dr. Carmen-Diana Herling (Klinik für Innere Medizin I) zusammengetragen und durchgeführt.

Der zur primären Datenauswertung verwendete Algorithmus wurde von Dr. Martin Peifer (Department of Translational Genomics, Cologne Center of Genomics, Universität zu Köln) speziell auf die diagnostische Anwendung angepasst. Die Auswertung der Daten wurde von mir selbst durchgeführt.

Die statistische Auswertung der Ergebnisse erfolgte mit Unterstützung von Fabian D. Mairinger (Institut für Pathologie, Universitätsklinikum Essen, Universität Duisburg-Essen).

Sofern nicht anders Text erwähnt, wurden keine weiteren Unterstützungsleistungen in Anspruch genommen.

Acknowledgments

My deepest gratitude goes to **Professor Dr. Reinhard Büttner**, for giving me the opportunity to facilitate my work at the Institute of Pathology and for supporting me on my way.

I especially want to thank **Dr. Margarete Odenthal** for reviewing and providing insightful comments on my thesis. Her extraordinary patience and accessibility provided the best conditions for my work. She taught me to always look on the bright side.

Thanks to **Dr. Carmen-Diana Herling** for supporting me by providing the clinical samples for my work and for her expert advice and support.

I also want to express my special thanks to **Ulrike Koitzsch** and **Carina Heydt** and to all my **lab colleagues** for their warmly friendship and for always having a helping hand in technical questions.

Special thanks to **my parents**, who supported me a long way, believed in me and made this work possible.

Hearty and most sincere thanks to **Fabian**, for his unselfish support, his tireless patience and the long and critical discussions, which encouraged me every time anew!

Table of Contents

Glossary	I
1. Introduction	1
1.1 Aspects of Chronic Lymphocytic Leukemia	1
1.2 How to Diagnose Chronic Lymphocytic Leukemia	1
1.2.1 Morphological Analysis.....	2
1.2.2 Clinical Markers.....	3
1.2.2.1 The Staging Systems	3
1.2.2.2 Serum Markers	4
1.2.3 Prognostic Factors	5
1.2.3.1 Genetic Aberrations	5
1.2.3.2 Protein-Based Markers	7
1.3 Therapy Strategies for Chronic Lymphocytic Leukemia	8
1.4 New Strategies for Biomarker Discovery and Application on Personalized CLL Treatment	10
2. Aim of the Study.....	15
3. Material and Methods.....	16
3.1 Reagents and Kits	16
3.1.1 Quantification of Input DNA.....	16
3.1.2 Library Construction	16
3.1.3 Quality Assessment and Quantification of Libraries	17
3.1.4 High-Throughput Sequencing	18
3.2 Clinical Samples.....	18
3.3 B-Cell Isolation and DNA Extraction	20
3.4 DNA Quantification	20
3.5 NGS of CLL DNA Samples Using Multiplex PCR Target Enrichment	21
3.5.1 NGS Analysis Using a Leukemia Gene Panel.....	21

3.5.1.1	NGS Library Construction by Qiagen GeneRead Technology	22
3.5.1.2	Massive Parallel Sequencing of an Amplicon Library of a Commercially Available Primer Set	22
3.5.1.3	Company Provided Online Data Analysis	23
3.5.2	Targeted Deep Sequencing of CLL Related Genes	23
3.5.2.1	Design of a CLL Specific Panel.....	23
3.5.2.2	Construction of Amplicon Libraries	24
3.5.2.3	Quality Assessment and Quantification of the Constructed Libraries Followed by Sequencing	27
3.5.3	Improvement of Adapter Ligation Efficiency	30
3.5.4	Automation of Sample Preparation Process.....	30
3.6	Determination of Sequencing Sensitivity	30
3.7	Sequencing Data Analysis.....	31
3.8	Variant Confirmation.....	32
3.9	Statistical Analysis.....	33
4.	Results	34
4.1	MiSeq Sequencing Using a Multiplex PCR Approach for Leukemia Specific Gene Enrichment	34
4.2	Protocol Development for Target Specific Sequencing of Native B-Cell DNA.....	35
4.2.1	Determination of the Amount of Input B-Cell DNA	35
4.2.2	Optimization of the Adapter Ligation Step	36
4.2.3	Final Quantification of Constructed Libraries.....	37
4.2.4	Automated Sample Processing	38
4.3	Target Specific Sequencing of Native B-Cell DNA	39
4.4	High Sensitivity in <i>TP53</i> and <i>ATM</i> Mutant Identification	43
4.5	Variant Validation by Sanger Sequencing	44
4.6	Cluster of Genes Affected by High Frequency Variants	44
4.7	Associations with Clinical Parameters	48

5. Discussion	53
6. Summary.....	62
7. Zusammenfassung.....	63
8. References.....	64
9. Preliminary Publications	78
Research Articles	78
Technical Reports	78
Reviews.....	78
10. Appendix.....	79
10.1 Target Regions and Primer Sequences of CLL Panel 1 and 2.....	79
10.2 Primer for Sanger Sequencing	86
10.3 Sequencing Data.....	88
10.4 Statistical Analysis	92
11. Lebenslauf	93

Glossary

Gene names are written italic

Abbreviation	Explanation
A	Alemtuzumab
AD	Activation Domain
ALL	Acute Lymphocytic Leukemia
alloSCT	allogeneic Stem Cell Transplantation
AML	Acute Myeloid Leukemia
ANK	Ankryin Repeat Domain
ATM	<i>Ataxia Telangiectasia Mutated</i>
B	Bendamustine
BCR	B-Cell Receptor
BLAT	Blast (basic local alignment search tool)-Like Alignment Tool
BSC	Best Supportive Care
BTK	<i>Bruton agammaglobulinemia Tyrosine Kinase</i>
BTK	Tyrosine-protein Kinase BTK
CD	Cluster of Differentiation
CD79B	<i>CD79b molecule, immunoglobulin-associated beta</i>
CD79B	B-cell antigen receptor complex-associated protein beta chain
CIB	Chlorambucil
CLL	Chronic Lymphocytic Leukemia
CML	Chronic Myeloid Leukemia
COSMIC	Catalogue Of Somatic Mutations in Cancer
CR	Complete Remission
dbSNP	Single Nucleotide Polymorphism database

Gene names are written italic

Abbreviation	Explanation
<i>DDX3X</i>	<i>DEAD (Asp-Glu-Ala-Asp) box polypeptide 3, X-linked</i>
DDX3X	ATP-dependent RNA helicase DDX3X
Del11q	Deletion affecting the long arm of chromosome 11
Del13q	Deletion affecting the long arm of chromosome 13
Del17p	Deletion affecting the short arm of chromosome 17
DF	Dilution Factor
DLBCL	Diffuse Large B-Cell Lymphoma
DNA	Deoxyribonucleic Acid
dNTP	2'-deoxyribonucleoside triphosphate
EGFR-like	Epidermal Growth Factor Receptor like
EMBOSS	European and Molecular Biology Open Software Suite
emPCR	Emulsion PCR (polymerase chain reaction)
FA	Fluorescence Absorbance
FAT	FRAP-ATM-TRAPP
FATC	FAT-C-term domain
FBXW7	F-Box/WD repeat-containing protein 7
<i>FBXW7</i>	<i>F-Box/WD repeat domain containing 7</i>
FCR	Fludarabin, Cyclophosphamid, Rituximab
FISH	Fluorescence In Situ Hybridization
GATK	Genome Analysis Tool Kit
gDNA	Genomic DNA
HD	Heterodimerization Domain
HD-C	Heterodimerization Domain C-terminus
HD-N	Heterodimerization Domain N-terminus
HEAT repeats	Huntington, Elongation factor 3, protein phosphatase 2A, target of rapamycin 1 repeats
HEK cells	Human Embryonic Kidney cells

Gene names are written italic

Abbreviation	Explanation
Hg19	Human genome version 19
Ig	Immunoglobulin
IGHV	Immunoglobulin Heavy Chain Variable region
IGV	Integrative Genomic Viewer
IκB α	Nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor, alpha
Indels	Insertions/deletions
ITIM	Immunoreceptor Tyrosin-based Inhibition Motif
KD	Protein Kinase Domain
LNR	Lin-12 NOTCH Repeats
<i>MAPK1</i>	<i>Mitogen-Activated Protein Kinase 1</i>
MAPK1	Mitogen-Activated Protein Kinase 1
M-CLL	IGHV mutated CLL patient
mRNA	messenger RNA
MOSFET	Metal-Oxide-Semiconductor Field-Effect Transistor
MYD88	myeloid differentiation primary response protein MyD88
<i>MYD88</i>	<i>myeloid differentiation primary response gene (88)</i>
NaOH	Sodium hydroxide
NCI-WG	National Cancer Institute-Working Group
ND	Negative regulation Domain
NES	Nuclear Export Sequence
NF κ B	Nuclear factor kappa light chain enhancer of activated B-cells
NGS	Next Generation Sequencing
NICD	Notch Intracellular Domain
<i>NOTCH1</i>	<i>Notch homolog 1, translocation-associated (Drosophila)</i>
NOTCH1	Neurogenic locus notch homolog protein 1

OS	Overall Survival
PCR	Polymerase Chain Reaction
PD	Progressive Disease
PD	Proline-rich Domain
PEST domain	proline-glutamic acid-serine-threonine motif
PI3K	Phosphatidylinositol 4,5-bisphosphat 3-kinase
PIK3CA	Phosphatidylinositol 4,5-bisphosphat 3-kinase catalytic subunit alpha isoform
<i>PIK3CA</i>	<i>Phosphatidylinositide-3-kinase, catalytic, alpha polypeptide</i>
PIK3CD	Phosphatidylinositol 4,5-bisphosphat 3-kinase catalytic subunit delta isoform
<i>PIK3CD</i>	<i>Phosphatidylinositide-3-kinase, catalytic, delta polypeptide</i>
PolyPhen	Polymorphism Phenotyping
PR	Partial Remission
PRD	PIKK-Regulatory Domain
PTEN	Phosphatidyl 3,4,5-triphosphate 3-phosphatase and dual-specific protein phosphatase PTEN
<i>PTEN</i>	<i>Phosphatase and Tensin homolog</i>
PTPN6	Tyrosine-protein phosphatase non-receptor type 6
qPCR	quantitative real-time polymerase chain reaction
R	Rituximab
SD	Stable Disease
SF3B1	SF3B1 protein
<i>SF3B1</i>	<i>Splicing factor 3b, subunit 1, 155kDa</i>
SHP1	Tyrosine-protein phosphatase non-receptor type 6
SIFT	Sorting tolerant from intolerant
snRNP	small nuclear ribonucleoprotein

STAT3	Signal Transducer and Activator of Transcription 3
TD	Tetramerization Domain
TE buffer	Tris EDTA (EthyleneDiamineTetraacetic Acid) buffer
TIR domain	Toll/IL-1R homology domain
TK	Serum Thymidine Kinase count
TLR pathway	Toll-Like Receptor pathway
<i>TP53</i>	<i>Tumor Protein p53</i>
TP53	cellular Tumor antigen p53
TTT	Time-To-Treatment
U-CLL	IGHV unmutated CLL
WBC	White Blood Count
ww	watchful waiting
XPO1	Exportin-1
<i>XPO1</i>	<i>Exportin 1 (CRM1 homolog, yeast)</i>

1. Introduction

1.1 Aspects of Chronic Lymphocytic Leukemia

Leukemic diseases originate from the bone marrow. Depending on their clinical course, which can be chronic or acute, and the involvement of blood cells, they are divided into different subgroups. This classification includes acute lymphocytic leukemia (ALL), acute myeloid leukemia (AML), chronic myeloid leukemia (CML) and chronic lymphocytic leukemia (CLL). Each of them is characterized by substantial differences with respect to epidemiology, biology and prognosis of the disease.

In the western world CLL is the most common leukemia in adults, representing about 24-30% of all leukemia types, mainly affecting elderly individuals with a higher incidence in males [13, 17, 22]. In 2010, about 11,500 new cases of leukemia were diagnosed in Germany and more than one third belonged to CLL [69]. The incidence rate is about two to six new cases per 100,000 individuals per year and increases with age, reaching 12.8 new cases per year at 65 years of age, which is the mean age at diagnosis [22].

Until now, CLL is an incurable disease characterized by the accumulation of CD5 positive B-lymphocytes in the peripheral blood, bone marrow and secondary lymphoid organs (lymph nodes and spleen) and an increase in the lymphocyte count in the peripheral blood [17, 22, 94]. The CLL is an extensively heterogeneous disease showing a highly variable clinical course [13]. Whereas half of the patients can live for 20 years or more with no need for treatment, others show rapidly progression, leading to substantial morbidity and mortality within a few months [13, 14, 94].

1.2 How to Diagnose Chronic Lymphocytic Leukemia

CLL is diagnosed either incidentally with an abnormally high white blood count (WBC) in asymptomatic patients, or due to symptoms that result from cytopenias (anemia, thrombocytopenia), adenopathy or constitutional symptoms as outlined by the 2008 International Workshop on CLL [26, 40]. Patients can then present with fatigue and increasing intolerance to physical exercise caused by progressive anemia, due to the detrimental effects of the expanding malignant cell population in the bone marrow and in the peripheral blood [22, 27]. In other cases, patients

1. Introduction

present with lymphoma-like symptoms as enlargement of superficial lymph nodes, liver and spleen due to the infiltration of secondary lymphoid organs. Patients can also show systemic symptoms such as weight loss, fever without infections, night sweat and fatigue [22, 27]. To ensure a clear distinction of CLL from other lymphoproliferative diseases, the evaluation of the blood count, blood smear, and the immune phenotype of the circulating lymphoid cells according to the National Cancer Institute-Working Group (NCI-WG) criteria is obligatory [26, 27].

1.2.1 Morphological Analysis

The diagnosis of CLL requires the presence of at least 5×10^9 B-lymphocytes per liter (5,000/ μ L) in the peripheral blood [26, 27]. The clonality of the circulating B-lymphocytes is confirmed by flow cytometry, whereas the morphological analysis is performed on the peripheral blood smears from the patients by panoptic staining (Figure 1) [26, 27]. The accumulation of small, homogeneous, mature B-lymphocytes with a narrow border of cytoplasm and a dense nucleus, lacking discernible nucleoli and having partially aggregated chromatin is characteristic for the disease [22, 26, 27]. These characteristic B-lymphocytes (B-cells) might be mixed with larger atypical cells, cleaved cells, or pro-lymphocytes, which may comprise 55% of the blood lymphocytes [26, 27]. The B-cells are also characterized by a fragility of the cell membrane that leads to frequent rupture of the leukemic cells. In addition, when preparing the blood film, they create the so-called “Gumprecht nuclear shadows” [22, 26, 27].

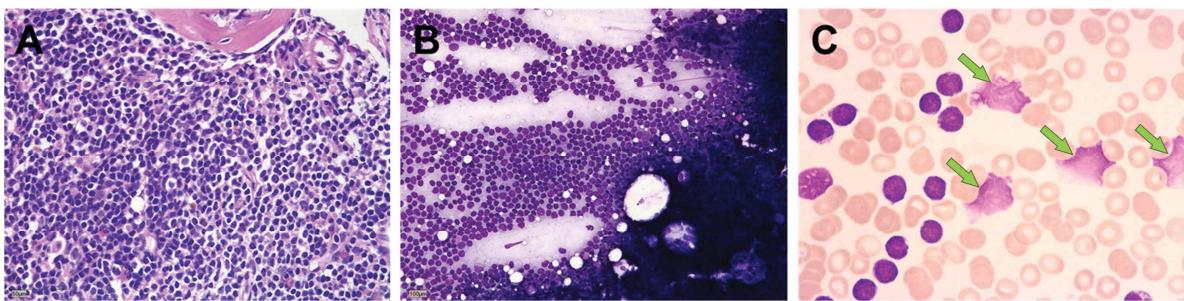


Figure 1: Typical B-CLL morphology. **A:** 400 fold magnifications of an iliac crest biopsy stained by hematoxylin and eosin, showing a predominant population of small lymphocytes. Myeloid hematopoiesis is massively decreased. **B:** 400 fold magnification of a bone marrow smear stained by May-Grünwald-Giemsa (MGG) method, showing a high degree infiltration of monomorphic small lymphocytes replacing the physiological hematopoiesis. **C:** 1,000 fold magnification of a peripheral blood smear stained by MGG method showing small lymphocytes with round or oval nuclei and erythrocytes. Very striking is the absence of platelets and granulocytes. Typical “Gumprecht nuclear shadows” are labeled by green arrows. The light microscopy images are kindly provided by Dr. M. Engels, University Hospital Cologne, Institute of Pathology (2014).

Importantly, an immunophenotyping analysis is required to verify a diagnosis of CLL (1.2.3.2) [22, 26]. Marrow aspirates and biopsies are generally not required for the diagnosis [26, 27]. Even if the type of marrow infiltration (diffuse or non-diffuse) reflects the tumor burden and provides some prognostic information, the prognostic value of bone marrow biopsies may be superseded by new prognostic markers [26]. However, it can help to evaluate factors that might contribute to cytopenias that may or may not be directly related to leukemia-cell infiltration of the marrow [26]. Because such factors could influence the susceptibility to drug-induced cytopenias, a marrow biopsy is recommended before initiating a therapy [26].

1.2.2 Clinical Markers

1.2.2.1 The Staging Systems

Two independent clinical staging systems are currently used to define the disease progression, the prognosis and especially the clinical management [6, 22, 26, 64, 71].

The modified Rai classification categorizes patients by their monoclonal lymphocytosis and is more frequently used in the United States. The Binet staging is the most common staging system in Europe and is based on the number of involved tissue lymphatic systems. Important criteria are the enlargement of lymph

1. Introduction

nodes (greater than 1 cm in diameter), organomegaly, anemia or thrombocytopenia (Table 1) [26, 27].

Table 1: Classification of chronic lymphocytic leukemia [6, 22, 26, 64]

Modified Rai Staging System

Low-risk patients	Only lymphocytosis ($>5,000$ monoclonal B-lymphocytes) in the peripheral blood and bone marrow (stage 0)
Intermediate-risk patients	Lymphocytosis and lymphadenopathy (stage I) and/or hepatosplenomegaly (stage II)
High-risk patients	With lymphocytosis together with anemia ($Hb <11$ g/dL) (stage III) and/or thrombocytopenia (platelets $<100,000/\mu L$) (stage IV)

Binet Staging System

Stage A	$Hb \geq 100$ g/L, platelets $\geq 100 \times 10^9/L$ and up to two sites involved
Stage B	$Hb \geq 100$ g/L, platelets $\geq 100 \times 10^9/L$ and organomegaly greater than defined for stage A (more areas of nodal or organ enlargement)
Stage C	$Hb <100$ g/L and/or platelets $<100 \times 10^9/L$ regardless of organomegaly

Hb hemoglobin

Both staging systems are easy to apply because they are based on physical examination and standardized blood tests. Furthermore, they may help to divide patients into risk groups and facilitate a good prognostication of the patient's response, in particular because patients with Binet stage A or Rai stage 0 have a long overall survival (OS). In comparison, patients with Binet stage B and Rai stage I/II show an intermediate median survival of five to seven years, whereas patients in the high-risk groups (Binet stage C, Rai stage III/IV), have a shorter median survival of less than three years [71].

Nevertheless, these clinical staging systems have limitations because patients may progress to a higher stage. This makes it problematic to assess the disease course of patients in the low-risk group and to provide any information about patient's treatment response [71].

1.2.2.2 Serum Markers

Beside the clinical staging systems serum markers are used as clinical markers in CLL to facilitate a useful prognosis. A high level of the serum marker thymidine kinase (TK) is found in CLL patients with a more aggressive disease, and is

associated with unmutated immunoglobulin heavy chain variable (*IGHV*) genes and high-risk genomic aberrations [71]. Despite the fact that serum markers have a role in CLL prognostication, the advent of novel molecular biomarkers has shifted the focus to factors providing more disease-specific information regarding survival and treatment responses [71].

1.2.3 Prognostic Factors

1.2.3.1 Genetic Aberrations

For many years, it has been well established that CLL cells are characterized by several recurrent genomic aberrations [13, 22, 24, 71]. In contrast to other lymphoid tumors, translocations are rarely observed in CLL. The most typical aberrations lead to loss or gain of genetic material [17, 22]. Deletions on the long arm of chromosome 13 (13q14) are considered as an initiating event in the tumorigenesis of nearly 50% of CLL cases, followed by the accumulation of additional alterations during tumor progression, such as deletions of 11q (in 10-20% of the cases), where the *ataxia telangiectasia mutated* (*ATM*) tumor suppressor gene is localized, 17p deletion (del17p) or trisomy 12 (in 10-20% of the cases) [13, 17, 22, 71]. Importantly, patients with deletions in chromosomal bands 17p13 (in <10% of the cases), which affect the tumor suppressor gene *TP53*, show the most aggressive phenotype and this aberration is the only marker currently recommended to direct treatment decisions [13, 17, 22, 26, 60, 71]. More than 50% of the CLL cases show only a single abnormality, 20% carry two and about 10% more than two aberrations [22]. Multivariate analysis showed that genomic aberrations are independent prognostic factors. For example, 13q deletions are associated with favorable prognostic effects, whereas the remaining abnormalities are associated with a progressively poorer prognosis (Figure 2) [13, 22, 71]. Conventional cytogenetic analysis of genomic aberrations by monitoring chromosome banding turned out as not useful due to the low proliferation rate of the CLL cells [13, 22, 99]. Therefore, fluorescence in situ hybridization (FISH) has become the gold standard method in clinical diagnostics for the detection of known recurrent genomic aberrations. It enables the detection and direct visualization of chromosomal abnormalities not only in dividing cells but also in interphase nuclei, in 80% of the CLL patients [13, 22, 26, 71, 99].

1. Introduction

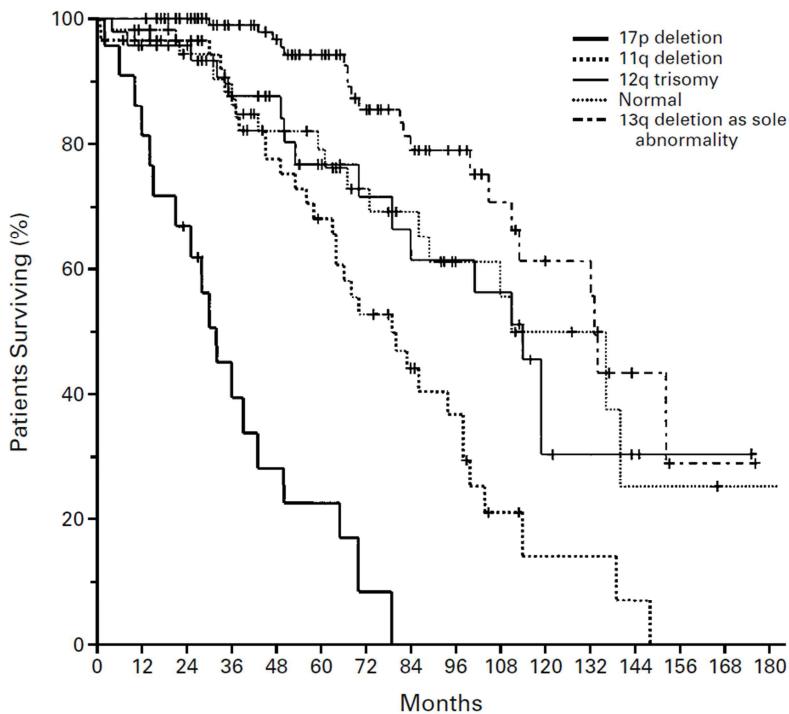


Figure 2: Probability of survival from the date of diagnosis among patients in five genetic categories. The median survival times for the groups with 17p deletion, 11q deletion, 12q trisomy, normal karyotype, and 13q deletion as the sole abnormality were 32, 79, 114, 111, and 133 months, respectively. The image was taken from Döhner *et al* [13].

As mature B-lymphocytes, CLL cells express a B-cell receptor (BCR) on the cell surface, which is encoded by the immunoglobulin (Ig) genes [22]. The molecular analysis of these genes and in particular of the *IGHV* genes showed in more than 50% of the CLL cases the presence of somatic hypermutations, which were associated to prognosis [11, 22, 71]. The distinction between mutated (M-CLL) and unmutated (U-CLL) *IGHV* in CLL cells is based on the degree of identity between the CLL *IGHV* sequence and the closest germline sequence [95]. Even if the cutoff for this distinction varies between 95% and 98%, depending on the performing institute, the presence or absence of these mutations can be used to distinguish between two patient subsets [23, 95]. U-CLL patients show a more aggressive disease, high-risk cytogenetics and significantly decreased median survival in contrast to M-CLL patients, which are associated with a more favorable clinical course often with no need for treatment [71, 95]. This important subclassification is currently one of the most commonly used prognostic markers and also widely accepted as one of the most stable and reliable indicators of clinical outcome [23, 42, 71].

1.2.3.2 Protein-Based Markers

The CLL phenotype is characterized by leukemia B-cells co-expressing the mature B-cell markers CD19, CD20, CD23 and high levels of CD5 [22, 26, 40]. CLL B-lymphocytes typically express low levels of surface Ig, CD20 and CD79b in the peripheral blood, bone marrow, lymph nodes and spleen compared with those found in normal B-cells [22, 26, 40]. All leukemia cells are monoclonal and are restricted to expression of either kappa or lambda Ig light chains [22, 26].

Extensive efforts to find surrogate markers that can substitute the intricate *IGHV* gene analysis have identified some markers on the protein level, with an independent prognostic value. One of them is the transmembrane glycoprotein CD38, which is expressed by cells of hematopoietic origin, showing a high expression of activated B- and T-cells, as well as natural killer and dendritic cells [12, 31, 46, 71]. Nevertheless, the assessment of a prognostic statement can be difficult due to technical aspects like the choice of an optimal cutoff for the number of CD38 positive cells determined by flow cytometry. Furthermore, CD38 expression of the leukemia clone may change during the course of the disease [71, 95].

Another important protein-based marker is the zeta-chain associated protein kinase 70 kDa (ZAP70), a tyrosine kinase normally expressed by natural killer cells and T-lymphocytes [9, 71]. Rosenwald *et al* were the first that identified an overexpression of ZAP70 in *IGHV* unmutated CLL patients [71, 72]. Several studies validated this finding and reported ZAP70 expression as a surrogate marker predicted for the distinction of *IGHV* mutation status and the identification of patients with different clinical course [10, 71, 95]. A high ZAP70 expression correlates with an unfavorable disease course in terms of progression and OS (Table 2) [10, 71, 95].

1. Introduction

Table 2: Prognostic factors in chronic lymphocytic leukemia [8, 71]

Prognostic Markers	Status	Prognostic Indication	Evaluation Method
Clinical Markers			
Rai staging (0-IV)	Low stage High stage	Low risk High risk	Physical examination and blood tests
Binet staging (A-C)	Low stage High stage	Low risk High risk	Physical examination and blood tests
TK serum level	High level	High risk	Serum level
Genetic Markers			
<i>IGHV</i> gene mutation status	Unmutated (≥ 98% accordance to germline sequence)	High risk	Sequencing
Del11q	i.e. deletion of <i>ATM</i>	Intermediate to high risk	FISH or microarray
Del13q	i.e. deregulation of miR15a/16-1	Low risk	FISH or microarray
Del17p	i.e. deletion of <i>TP53</i>	High risk	FISH or microarray
Trisomy 12	Amplification	Intermediate risk	FISH or microarray
Protein-Based Markers			
CD38 level	High expression (>20% positive cells)	High risk	Flow cytometry
ZAP70 level	High expression (>20% positive cells)	High risk	Flow cytometry or RNA expression

1.3 Therapy Strategies for Chronic Lymphocytic Leukemia

For asymptomatic patients and for those in early stages (Binet A or B, Rai 0-II), watchful waiting is still the treatment of choice, since these patients show no survival benefit from an early initiation of treatment [29, 40, 82]. At present, there is still no consensus if there is a need for early treatment of patients with high-risk features like del17p [29]. Otherwise, patients with advanced stages (Binet C, Rai III-IV) or symptoms need aggressive treatment strategies. They benefit from the advances of chemotherapy regimens combining purine analogs and alkylating agents with monoclonal antibodies such as rituximab (R) against CD20 [29, 40].

However, these effective chemoimmunotherapies like FCR (fludarabin, cyclophosphamid, rituximab) cannot be applied with the same success to all CLL patients [29]. For example, the FCR therapy comes up with a wide range of side effects and can be very toxic especially for elderly patients and for patients with comorbidities, respectively, and is therefore reserved for physically fit patients (Figure 3) [22, 29, 67].

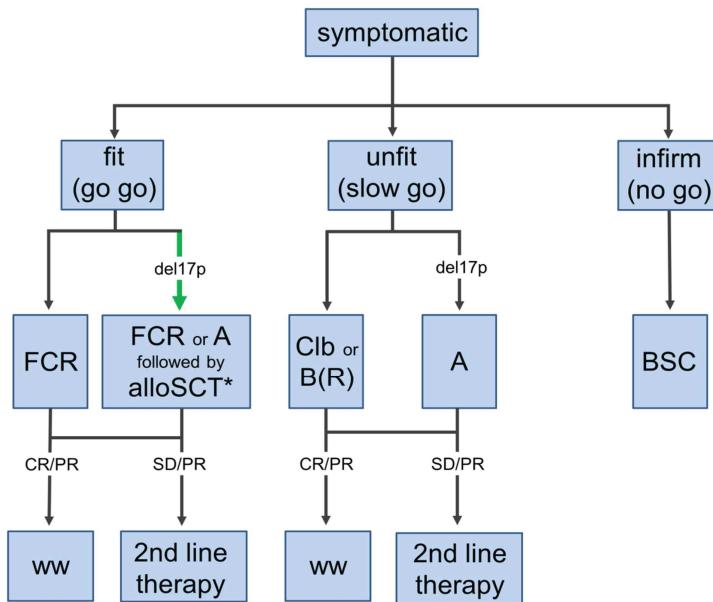


Figure 3: The scheme for the first-line treatment of symptomatic CLL patients shows the differentiation of patients according to their physical fitness. FCR fludarabin, cyclophosphamide and rituximab; A alemtuzumab; alloSCT allogeneic stem cell transplantation; Clb chlorambucil; B bendamustine; R rituximab; BSC best supportive care; CR complete remission; PR partial remission; SD stable disease; PD progressive disease; ww watchful waiting; green arrow indicates the curative therapy approach

Similarly, patients with specific molecular features like deletions of the short arm of chromosome 17 (del17p), leading to a disruption of *TP53*, do not have an optimal benefit from FCR chemoimmunotherapy [29]. The tumor suppressor TP53 is a tetrameric nuclear phosphoprotein that codes for a key regulator of the DNA-damage response and usually its activation leads to cell-cycle arrest, DNA repair, apoptosis or senescence [75]. Therefore, it plays a crucial role in mediating pro-apoptotic and anti-proliferative mechanisms of DNA-damaging chemotherapeutic agents [75]. Chemotherapy with alkylating agents like chlorambucil (Clb), cyclophosphamid (C) and bendamustin (B) introducing single and double strand

1. Introduction

breaks, as well as purine analogs like fludarabin (F) are used for standard first-line treatment in CLL and mediate cell death and apoptosis through DNA damage and subsequent TP53 dependent apoptosis [3, 22, 27]. Consequently, patients with a loss of TP53 have an extremely high probability of failing standard treatments and are known to be associated with treatment resistance and shortened survival in CLL [3]. Therefore, it is recommended that these patients undergo alternative regimens to overcome chemofractoriness or even they should be considered for an early allogeneic stem cell transplantation [26, 29, 60, 75]. These patients can also benefit from a monotherapy with a humanized IgG1 antibody against CD52, namely alemtuzumab (A) [22, 29, 53, 82, 85]. Due to the high density of the cell surface antigen CD52 on T- and B-cells in CLL, this antibody shows very good response rates. Furthermore, normal hematopoietic stem cells (erythrocytes and thrombocytes) do not show CD52 expression and are therefore not affected, reducing the toxic side effects [22, 29, 53, 82, 85].

However, approximately 25-50% of the patients' relapses within two years of first- or second-line therapy, respectively [2, 28, 68]. In general, first-line therapy may be repeated if the response achieved more than one year (more than two years if FCR or a similar potent regimen was applied) [29]. Treatment options are more difficult for therapy-refractory patients. The principle in those cases is that therapy should be switched [29]. Despite the improvement of therapeutic options and increased survival rates, CLL remains largely incurable and the high rates of relapsed patients underline the urgent need for the detection of new predictive markers.

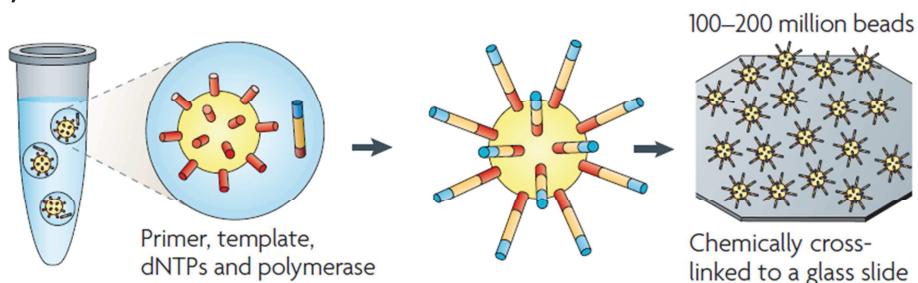
1.4 New Strategies for Biomarker Discovery and Application on Personalized CLL Treatment

Although the clinical markers are very useful to make a prognosis concerning the survival of the CLL patients, they cannot be used to predict the individual risk of disease progression, or treatment response. Therefore, additional prognostic and therapeutically relevant markers are urgently needed.

The advent of a new level of sequencing methods, namely next generation, or second generation sequencing (NGS) has enabled sequencing with a high capacity and thus allowing the analysis of complete genomes or exomes of a patient in a single run. Various approaches have been developed by different

companies each consisting of their own specific sequencing chemistry and sample preparation protocol. In general, after shearing of the genomic DNA, the template fragments get attached or immobilized to a solid surface for a following clonal amplification [52]. The most common methods are emulsion PCR on primer-coated beads (emPCR) [15] and solid-phase amplification on a primer carrying glass slide carrying (Figure 4) [18, 52].

A) Emulsion PCR



B) Solid-phase amplification

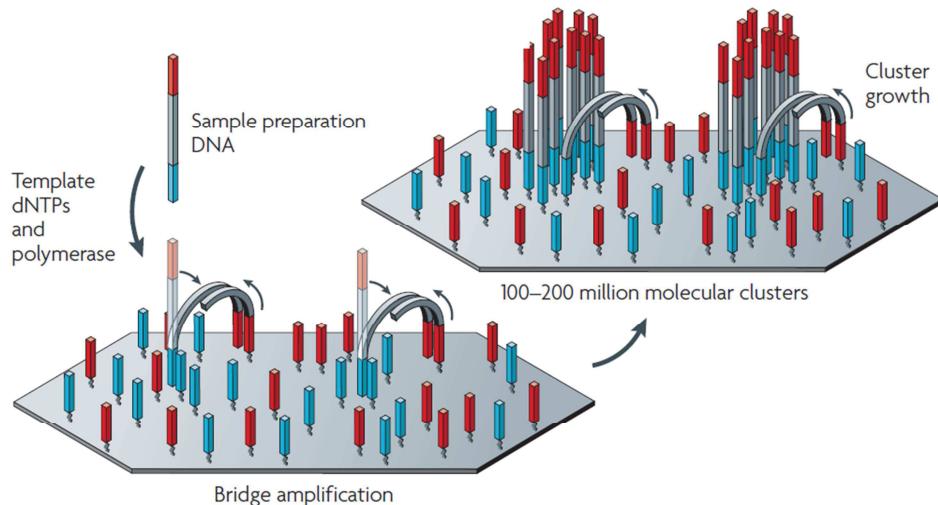


Figure 4: Template immobilization strategies: A) In emulsion PCR (emPCR), a reaction mixture consisting of an oil-aqueous emulsion is created to form bead-DNA complexes within single droplets. Following, PCR amplification is performed in these droplets to create beads containing several thousand copies of the same template sequence. Afterwards, beads are chemically attached to a glass slide (IonTorrent sequencing, Life Technologies) or deposited into PicoTiterPlate (454 sequencing, Roche) wells. B) The solid-phase amplification is based on two basic steps: initial priming and extending of the single-stranded, single molecule-template, and a bridge amplification of the immobilized template with adjacent primers to form clusters. dNTP 2'-deoxyribonucleoside triphosphate; the image is taken from Metzker *et al* [52]

1. Introduction

Commonly, the templates carry adapter specific for the sequencing platform. Samples prepared by emPCR are sequenced either by pyrosequencing (454, Roche) or ion semiconductor sequencing (IonTorrent, Life Technologies) [52]. Parallel sequencing of DNA clones by pyrosequencing is based on the release of the pyrophosphate after the incorporation of an appropriate nucleotide into the target sequence [43, 51]. The pyrophosphate gets converted and a final luciferase reaction leads to light emission that is detected by a sensitive camera [43, 51].

The ion semiconductor approach is focused on ultrasensitive measurements of pH changes. In order to detect the proton that is released during the nucleotide incorporation, sequencing of the clonal DNA templates occurs in picolitre cartridges of a MOSFET (metal-oxide-semiconductor field-effect transistor) flow cell [77].

In contrast samples for sequencing by synthesis (HiSeq, Illumina) were immobilized on a glass slide (flow cell) for generation of clonal DNA cluster by bridge amplification. The sequencing is performed by DNA strand synthesis using fluorochrome-labeled nucleotides containing a reversible terminator [5, 52]. Each sequencing cycles starts with the incorporation of all four nucleotides, each labeled with a different dye, imaging, cleaving the terminator and regenerating the 3'OH group for the next cycle [5, 52]. The sequencing can be performed in either only one (single end) or both directions of the template strand (paired-end).

The application of these whole genome/exome (WGS/ WES) NGS approaches has helped to discover novel genetic alterations and recently identified *NOTCH1*, *SF3B1* and *MYD88* as the most frequently mutated genes in CLL [17, 32, 33, 61, 63, 74, 94]. Numerous somatic mutations, among them putative driver mutations and/or indels (insertions/deletions), were identified in the examined genes [17, 32, 33, 40, 61, 63, 74, 94]. *SF3B1* and *NOTCH1* have been reported to be mutated in approximately 2-10% of CLL cases with a higher frequency in patients with progressive or high-risk disease and showed a significant correlation to shorter time-to-treatment (TTT) and OS (Figure 5) [4, 17, 32, 57, 61, 63, 74, 94].

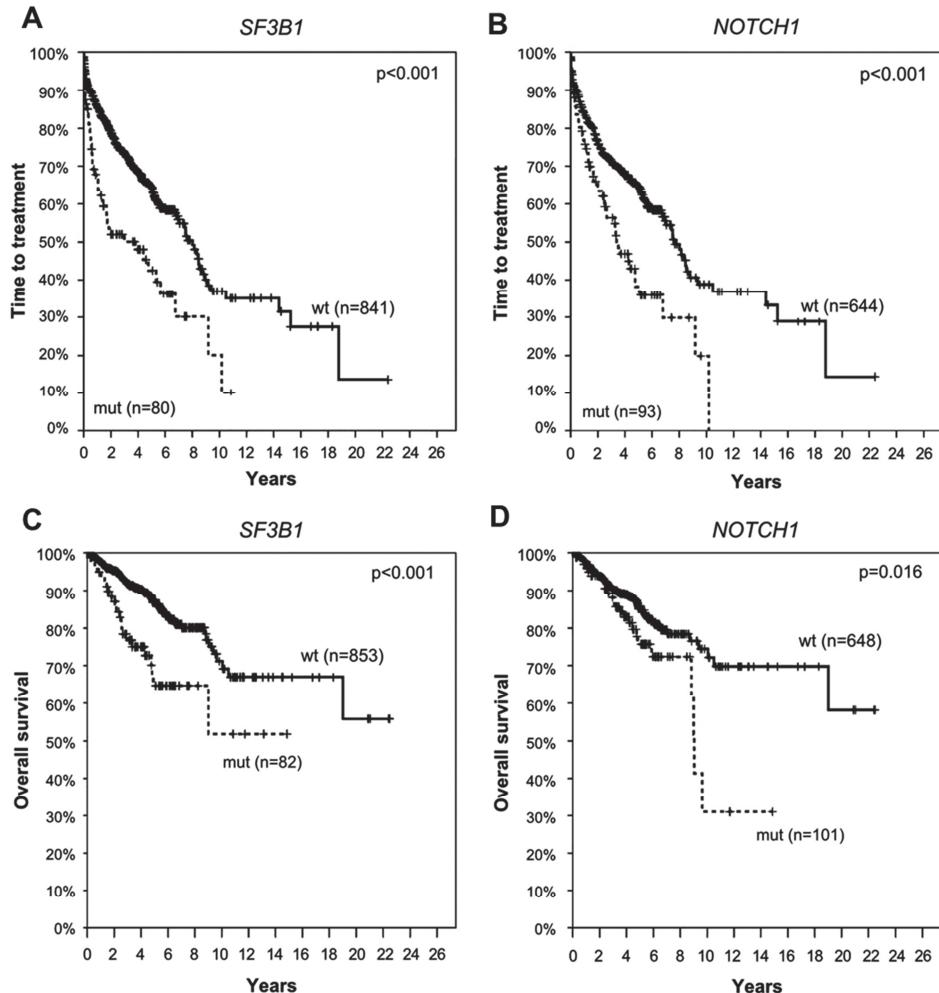


Figure 5: Kaplan-Meier plots of time-to-treatment (A, B) and overall survival (C, D) from diagnosis for CLL patients analyzed for *SF3B1* (A, C) and *NOTCH1* (B, D) shows a significantly poorer outcome in mutated cases. Mutated cases are represented by dotted lines. Number of patients with available follow-data is given in brackets. The image is taken from Jeromin et al [32].

Various recurrent mutations in genes like *MYD88*, *FBXW7* and *XPO1* have been reported in rare cases with frequencies below 10%. The clinical impact of these mutations remains unclear, as the data in CLL patients are currently limited [32, 61]. The corresponding proteins are part of diverse cellular processes: *FBXW7* plays a major role in *NOTCH1* degradation, *MYD88* has functions in immune response and *XPO1* in nucleocytoplasmic transport [30, 56].

Knowing that these genes are related to CLL, some of them may constitute new therapeutic targets and can become important on the way towards a personalized medicine. For example, different *NOTCH1* inhibitors have been under clinical investigation and *in vivo* studies using gamma-secretase inhibitors to induce

1. Introduction

apoptosis of CLL cells showed promising results [2, 54, 56, 70]. Moreover, therapies targeting the BCR signaling pathway, with inhibitors as ibrutinib (targeting tyrosine-protein kinase BTK) and idelalisib (targeting PI3K), are also promising [8, 20, 40, 78, 96]. Furthermore, MYD88 aberrations can be targeted through direct inhibition of the corresponding pathway, through proteasomal inhibition or even through inhibition of BTK [40, 55]. Other gene aberrations will probably be used in the future to predict the highly variable course of CLL.

Currently, genes with prognostic or predictive impact in cancer therapy are generally analyzed by conventional methods like polymerase chain reaction (PCR) or Sanger sequencing, which is considered as the gold standard for routine diagnostic [16, 83]. These methods are simple to implement in laboratory processes, are well established and accuracy and sensitivity (20%) are well known [16, 89, 90]. However, the number of identified new potential therapeutically relevant markers in CLL has enlarged the spectrum of genes to be sequenced and have aroused the need for more efficient, higher-capacity methods. NGS methods improved the scale and efficiency of DNA sequencing and thereby enabled the possibility to keep up with the rapid pace of research [50, 52]. Massive parallel sequencing technologies (NGS) now provide the opportunity of systematically discovering the genetic alterations that underlie the disease and to identify new therapeutic targets and predictive biomarkers [94]. Nevertheless, since WGS/WES approaches produce only a low coverage, targeted sequencing approaches are needed to cover all promising CLL related gene regions with a high coverage and a high sensitivity. Especially targeted sequencing on current benchtop platforms facilitate deep sequencing of a comprehensive number of targets and samples, which makes this an advantageous, reasonable tool for the clinical setting.

All attempted efforts towards a better understanding of CLL and identification of potential therapeutically relevant prognostic and predictive markers have clarified the need for a standardized workflow, which allows a comprehensive analysis of currently important disease relevant targets in clinical practice.

Accordingly, the establishment of a comprehensive panel of recurrent genetic abnormalities in CLL for application in clinical practice is urgently needed.

2. Aim of the Study

The ability to predict a more aggressive disease course has improved with the establishment of tests for prognostic markers including chromosomal aberrations and *IGHV* gene mutation status. Nevertheless, these markers have limitations and still the prediction of the disease course is not highly reliable.

The present study focused on the application of next generation sequencing technology as a diagnostic tool for comprehensive mutation analysis of CLL. First a panel of putative diagnostic relevant genes should be compiled and primer sets for multiplex enrichment should be designed. Next, it was aimed to establish the library preparation by multiplex PCR amplicon enrichment and link the approach to parallel sequencing using an Illumina MiSeq platform. Finally, the method should be applied to clinical samples and mutation status of diagnostic relevant genes should be identified.

3. Material and Methods

3.1 Reagents and Kits

All reagents and kits used for DNA quantification, library preparation and quality assessment of constructed libraries, as well as reagents for NGS are listed the chapters 3.1.1 to 3.1.4 (Table 3 to 6).

3.1.1 Quantification of Input DNA

Quantification of input DNA was performed by fluorometry measurement and quantitative real-time PCR (qPCR) using kits and primer sets as listed in Table 3.

Table 3: Reagents and kits for quantification of DNA templates used for target enrichment by multiplex PCR

Reagent	Reagent/Kit	Company
Fluorescence Absorbance		
QuantiFluor	QuantiFluor™ dsDNA System	Promega
qPCR		
Mastermix	GoTaq® qPCR Master Mix	Promega
Primer	<i>HFE</i> primer forward 5'ATG GAT GCC AAG GAG TTC GAA CC	Eurofins Genomics
	<i>HFE</i> primer reverse 5'GCC ATA ATT ACC TCC TCA GGC AC	Eurofins Genomics

3.1.2 Library Construction

Disease specific targets were enriched by multiplex PCR, ligated to barcode adapters specific for the sequencing platform and amplified by a final adapter-specific PCR.

The protocol contained diverse purification and size selection steps to cull amplicons with the expected fragment lengths.

All used reagents are summarized in Table 4.

Table 4: Reagents and kits used for next generation sequencing library construction

Step	Reagent/Kit	Company
Multiplex PCR for Target Enrichment	2x Ion AmpliSeq™ Primer Pool Panel 1 IAD27641 (pool1: 39 primer pairs, pool 2: 34 primer pairs; 400nM each)	Life Technologies
	2x Ion AmpliSeq™ Primer Pool Panel 2 IAD39491 (pool1: 134 primer pairs, pool 2: 131 primer pairs; 400nM each)	Life Technologies
	5x Ion AmpliSeq™ HiFi Master Mix (Ion AmpliSeq™ Library Kit 2.0)	Life Technologies
3' Adenylation	NEXTflex™ Adenylation Mix	Bioo Scientific
Primer Digestion	FuPa reagent (Ion AmpliSeq™ Library Kit 2.0)	Life Technologies
Purification and Size Selection	Agencourt® AMPure® XP	Beckman Coulter
Adapter Ligation	NEXTflex™ DNA barcodes – 48	Bioo Scientific
	Switch Solution (Ion AmpliSeq™ Library Kit 2.0)	Life Technologies
	DNA Ligase (Ion AmpliSeq™ Library Kit 2.0)	Life Technologies
Adapter-Specific PCR	Platinum® PCR Super Mix (Ion AmpliSeq™ Library Kit 2.0)	Life Technologies
	NEXTflex™ Primer Mix (NEXTflex™ DNA Barcode Kit)	Bioo Scientific

3.1.3 Quality Assessment and Quantification of Libraries

Prior to sequencing, quality of finalized libraries was checked by microfluidic electrophoresis and concentrations were measured by two different methods. The reagents for quality assessment and quantification of adapter-linked amplicons are listed in Table 5.

3. Material and Methods

Table 5: Reagents and kits used for quality assessment and quantification of the generated libraries

Step	Reagent/Kit	Company
Quality Assessment		
Microfluidic Electrophoresis	High Sensitivity DNA Kit	Agilent Technologies
Quantification		
qPCR of Adapter-Carrying Amplicons	Illumina Adapter Primer Forward 5'AAT GAT ACG GCG ACC ACC GAG ATC TAC AC	Eurofins Genomics
	Illumina Adapter Primer Reverse 5'CAA GCA GAA GAC GGC ATA CGA GAT	Eurofins Genomics
	GoTaq® qPCR Master Mix	Promega
	PhiX Control v3	Illumina
Fluorescence Absorbance	QuantiFluor™ dsDNA System	Promega

3.1.4 High-Throughput Sequencing

High-throughput sequencing of the CLL samples was performed on a MiSeq (Illumina, San Diego, CA, USA) instrument using reagents shown in Table 6.

Table 6: Reagents and kits used for next generation sequencing on the MiSeq platform

Step	Reagent/Kit	Company
Next Generation Sequencing	MiSeq Reagent Kit v2 (300 cycles)	Illumina
	PhiX Control v3	Illumina

3.2 Clinical Samples

136 blood samples from CLL patients were collected at the University Hospital Cologne, Germany, between 2012 and 2013. The study was approved by the institutional ethical review board (reference no. 13-091) and an informed consent was obtained from all patients. All cases demonstrated typical features of CLL as defined by the International Workshop on CLL [25]. Clinical and routine laboratory parameters were retrieved from medical records. CLL-related chromosomal abnormalities were assessed by interphase FISH within the routine diagnostic process using commercially available probes, detecting trisomy 12 and deletions

on chromosomes 6q21 (*SEC63*), 11q22.3 (*ATM*), 13q34 (*D13S319*) and 17p13.1 (*TP53*) (Abbott, Abbott Park, IL, USA). In addition, CLL immunophenotypes including CD38 and ZAP70 surface expression, as well as the somatic mutation status of *IGHV* genes were determined as described previously [81].

Patient characteristics are summarized in Table 7.

Table 7: Clinical and laboratory characteristics of 136 CLL patients

	Absolute
Age at sampling (years)	
Median (range)	63 (29-86)
Gender	
Male	96
Female	40
Binet stage (n=127/136)	
A	73
B	30
C	24
Treatment status (n=135/136)	
Untreated	94
Treated	41
White blood count (WBC) [10⁹/L]	
Median (range)	48.8 (10.8-483.8)
IGHV somatic mutation status (n=124/136)	
Mutated	59
Unmutated	65
Serum thymidine kinase (TK) (n=78/136)	
>10U/L	50
Median (range)	16.1 (3.5-330.0)
ZAP70 expression (n=86/136)	
Positive	30
Negative	56
CD38 expression (n=84/136)	
<30%	56
≥30%	28
FISH positivity (hierarchical model, n=81/136)	
Del13q as sole abnormality	35
Trisomy 12	7
Del11q	9
Del17p	13
Normal	20
Median time from diagnosis to sampling (range)	61.5 (0-296.1)

3. Material and Methods

3.3 B-Cell Isolation and DNA Extraction

B-cells were enriched by negative selection using RosetteSep®-based cell removal (Stemcell Technologies, Vancouver, BC, Canada) followed by Pancoll human density centrifugation (Pan Biotech, Aidenbach, Germany).

Genomic DNA was extracted from the B-cell fractions by standard column based purification (DNeasy, Qiagen, Hilden, Germany). DNA quality was assessed by gel electrophoresis.

3.4 DNA Quantification

Two different quantification methods were used to determine the amount of input DNA. B-cell DNA was quantified by fluorescence absorbance using the QuantiFluor™ dsDNA System (Promega, Fitchburg, WI, USA) and the QuantiFluor™-ST fluorometer (Promega) according to the manufacturer's instructions. Briefly, 1 µl of DNA sample was diluted 200-fold in 1x TE buffer and measured on the fluorometer. Prior to measurements, a two-point calibration curve was established using the supplied standard lambda DNA, at 0 ng/µl and 200 ng/µl.

Alternatively, the extracted B-cell DNA was measured quantitatively by qPCR using *human hemochromatosis gene (HFE)* as amplifying target (234 bp). DNA samples were quantified from genomic DNA standard curves that were prepared from native DNA extracted from human embryonic kidney cells (HEK-293, obtained from the American Type Culture Collection ATCC) without known gene modification (Table 8). Nine 2-fold dilutions of HEK-293 cell DNA in a range of 0.195 to 50 ng/µl were used. All samples were measured in duplicates.

Table 8: Set up of *HFE* qPCR and amplification conditions

A) <i>HFE</i> qPCR Setup	
Component	Volume [µl]
Nuclease-free water	7.4
<i>HFE</i> primer forward (10 µM) ^{*1}	0.8
<i>HFE</i> primer reverse (10 µM) ^{*2}	0.8
GoTaq® qPCR Master Mix (Promega)	10.0
Total	19.0
gDNA (10 ng)	1.0
Total	20.0

^{*1}5'ATG GAT GCC AAG GAG TTC GAA CC^{*2} 5'GCC ATA ATT ACC TCC TCA GGC AC**B) *HFE* qPCR Amplification Conditions**

Stage	Temperature	Time
Hold	94°C	3 Minutes
	94°C	30 Seconds
55 Cycles	60°C	30 Seconds
	72°C	30 Seconds

Plate read and melting curve

3.5 NGS of CLL DNA Samples Using Multiplex PCR Target Enrichment

The feasibility of the MiSeq NGS platform was determined by analyzing B-cell DNA from CLL patients and exons of 20 leukemia relevant genes that were enriched by a company provided multiplex PCR approach. In a further step a customized CLL specific panel was designed and a protocol for library construction was established to analyze 136 CLL samples on the evaluated MiSeq system (3.5.2). Additionally, the library preparation protocol was optimized by improving the adapter ligation efficiency and implementing automated purification and size selection steps (3.5.3).

3.5.1 NGS Analysis Using a Leukemia Gene Panel

To check the feasibility of the NGS platform with amplicons generated by a multiplex PCR approach, twelve B-cell DNA samples were analyzed. The DNA was prepared to perform NGS with a target specific exon panel of Qiagen that

3. Material and Methods

covers all coding regions of the following 20 leukemia related genes: *ABL1*, *ASXL1*, *CDKN2A*, *CEBPA*, *FLT3*, *GATA1*, *GATA2*, *IDH1*, *IDH2*, *JAK2*, *KIT*, *KRAS*, *MPL*, *NPM1*, *NRAS*, *PTPN11*, *RUNX1*, *TET2*, *TP53* and *WT1*. In total 1,272 gene regions were covered by the multiplex PCR using the GeneRead DNAseq Leukemia Gene Panel (Qiagen). First, target enrichment was performed by multiplex PCR, followed by adapter ligation, quantification and sequencing on a MiSeq platform and data analysis. Sequencing data was analyzed by an online based analysis pipeline provided by the manufacturer.

3.5.1.1 NGS Library Construction by Qiagen GeneRead Technology

NGS sample preparation was performed according to the manufacturer's instructions (GeneRead DNAseq Gene Panel Handbook 11/2012, Appendix B, Qiagen; NEBNext DNA Library Prep Master Mix Set for Illumina, New England Biolabs, Ipswich, MA, USA).

For an equal input of amplifiable template DNA, a qPCR using human *HFE* gene as amplifying target was performed (Table 8). A total of 80 ng purified genomic B-cell DNA was amplified in four separate PCR reactions per sample and purified with Agencourt® AMPure® XP magnetic beads (Beckman Coulter, Brea, CA, USA) afterwards. Subsequently, PCR amplicons of each sample were pooled. After barcode adapter ligation (NEBNext Multiplex Oligos for Illumina, Index Primer 1-12; New England BioLabs), another purification and size selection of the fragments was performed culling amplicons of around 300 bp. A final amplification of the adapter carrying amplicons was done by PCR and fragments were re-purified with Agencourt® AMPure® XP magnetic beads.

Finally, library concentration was determined by qPCR according to the manufacturer's protocol (GeneRead DNAseq Library Quant Array, Protocol 3, Qiagen). Before sequencing, samples were diluted to a final concentration of 2 nM and pooled (sample library pool).

3.5.1.2 Massive Parallel Sequencing of an Amplicon Library of a Commercially Available Primer Set

In order to denaturize the DNA, 0.2 N sodium hydroxide (NaOH) was added to the sample pool followed by incubation for five minutes. A PhiX v3 spike-in control (Illumina) was denaturized in the same manner. Both, the sample pool and the PhiX control were then diluted to a final concentration of 8 pM. Next, 600 µl of the

finalized library including 1% PhiX control was applied to the MiSeq cartridge (Illumina) according to the manufacturer's instructions. Paired-end sequencing was performed using a MiSeq instrument (Illumina) and the v2 chemistry as recommended by Illumina.

3.5.1.3 Company Provided Online Data Analysis

Run parameters were evaluated with the Sequence Analysis Viewer (Illumina). For further analysis of the sequencing data, Fastq files generated by the MiSeq Reporter software (Illumina) were uploaded into a cloud space for automatic variant analysis (Qiagen). Somatic variant analysis and paired-end read mode was chosen. The primary, so called preliminary alignment of the raw data, including the full read data set, was done using Bowtie 2. This first alignment was followed by trimming of primer sequences and quality filtering, which excluded reads with an untrimmed length of less than 45 bp. In the final alignment the trimmed reads were mapped against the human reference genome 19 (hg19). Applied parameters were identical to those used in the preliminary alignment.

The results of the final read alignment were used further for variant calling, which was performed with the Genome Analysis Tool Kit (GATK) Lite version 2.1-8 (GATK Unified Genotyper program, Broad Institute Cambridge, USA). Variant filtering was done automatically in two steps: first, variants that failed some of the thresholds for variant calling were marked, and second, variants with less than 4% as well as indels with less than 20% variant allele frequency were removed.

3.5.2 Targeted Deep Sequencing of CLL Related Genes

Additionally, the development of a CLL-specific targeted genomic sequencing assay, which was able to meet diagnostic and clinical research needs, was focused. In this approach, a multiplex PCR-based sequencing panel, suitable for a high-throughput sequencing platform (MiSeq, Illumina) is described. This panel covers multiple informative gene regions, known or suspected to be mutated in CLL.

3.5.2.1 Design of a CLL Specific Panel

To select hotspot regions of genes for the target specific CLL panel, first I screened COSMIC (catalogue of somatic mutations in cancer) [19], BioGPS [97], UniProt [91] and EMBOSS (The European and Molecular Biology Open Software Suite, 2000) [66] databases, previously published data on CLL and genes involved

3. Material and Methods

in BCR signaling and related pathways. Based on the previous inquiry, a first panel contained 146 primer pairs in two primer pools covering 73 hotspot regions of the following 15 genes: *ATM*, *BTK*, *CD79B*, *DDX3X*, *FBXW7*, *MAPK1*, *MYD88*, *NOTCH1*, *PIK3CA*, *PIK3CD*, *PTEN*, *PTPN6*, *SF3B1*, *TP53* and *XPO1*. Due to many requests from clinicians, the panel was extended with targets covering the whole coding regions of the five genes *ATM*, *MYD88*, *NOTCH1*, *SF3B1* and *TP53* resulting in a second panel with 265 primers in two more primer pools. Both panels were designed using the Ion AmpliSeq™ algorithm of Life Technologies (Carlsbad, CA, USA). In total, the panels contained primer for 338 target regions and were delivered in four primer pools (Table 9, Table S1).

Table 9: Overview of the genes covered by the CLL panels

Gene	Biological Process	Exons	Transcript ID	n Amplicons
<i>ATM</i>	DNA damage/ cell cycle control	Complete (62)	NM_000051	117
<i>BTK</i>	B-cell receptor signaling	14-16	NM_000061	5
<i>CD79B</i>	B-cell receptor signaling	4-5	NM_021602	2
<i>DDX3X</i>	RNA splicing and processing	7-9, 11, 14	NM_001356	6
<i>FBXW7</i>	Ubiquitination of proteins	6-9	NM_033632	7
<i>MAPK1</i>	MAP kinase signaling	7	NM_002745	1
<i>MYD88</i>	Toll-like receptor signaling	Complete (5)	NM_002468	9
<i>NOTCH1</i>	Notch signaling	Complete (34)	NM_017617	71
<i>PIK3CA</i>	B-cell receptor signaling	9-11, 20-21	NM_006218	10
<i>PIK3CD</i>	B-cell receptor signaling	21-24	NM_005026	7
<i>PTEN</i>	AKT-mTOR signaling	5-6, 9	NM_000314	7
<i>PTPN6</i>	B-cell receptor signaling	11-12	NM_080548	2
<i>SF3B1</i>	RNA splicing and processing	Complete (25)	NM_012433	52
<i>TP53</i>	DNA damage/ cell cycle control	Complete (9)	NM_000546	16
<i>XPO1</i>	RNA splicing and processing	12-13, 15	NM_003400	7
Total number of amplicons				338

3.5.2.2 Construction of Amplicon Libraries

Target enrichment and library preparation followed the instructions of the “Ion AmpliSeq™ Library Kit 2.0” (Life Technologies) and the “NEXTflex™ DNA Sequencing Kit, Manual V11.12” (Bioo Scientific, Austin, TX, USA). Briefly, a total of 40 ng genomic B-cell DNA was amplified in four separate multiplex PCR

reactions per sample (Table 10).

Table 10: Multiplex PCR reaction mix and amplification conditions

A) Multiplex PCR Setup

Component	20µl Approach	10µl Approach
	Volume [µl]	Volume [µl]
5x Ion AmpliSeq™ HiFi Master Mix (Ion AmpliSeq™ Kit 2.0, Life Technologies)	4	2
2x Ion AmpliSeq™ primer pool* (400nM each) (Life Technologies)	10	5
gDNA, ~10 ng	Y	Y
Nuclease-free water	6-Y	3-Y
Total	20	10

*IAD27641 pool1: 39 primer pairs, pool 2: 34 primer pairs;

*IAD39491 pool1: 134 primer pairs, pool 2: 131 primer pairs

B) PCR Amplification Conditions

Stage	Temperature	Time
Hold	99°C	2 Minutes
30 Cycles	99°C	15 Seconds
	60°C	4 Minutes
Hold	10°C	∞

Then, two of the four PCR reactions per sample were pooled and subjected to enzymatic digestion of the primer sequences according to the manufacturer's protocol (Life Technologies) (Table 11).

Table 11 Protocol for enzymatic digestion of the amplicon primer sequences

Temperature	Time
50°C	10 Minutes
55°C	10 Minutes
65°C	20 Minutes
10°C	For up to 1 hour

Afterwards, each reaction per sample was pooled and amplification products were purified from half of the reaction volume. All purification and size selection steps

3. Material and Methods

were performed in a 96 well format using Agencourt® AMPure® XP magnetic beads and a Biomek® FXp workstation (Beckman Coulter). Each purification and size selection step contained two washing steps with 200 µl 80% ethanol. Purification step I was done with 1.6-fold reaction volume of magnetic beads to remove fragments <100 bp. DNA was eluted with 20 µl nuclease-free water. Subsequently, 10-fold diluted samples were adenylated (Table 12) and ligated to NEXTflex™ DNA barcodes - 48 (Bioo Scientific) (Table 13).

Table 12: Adenylation reaction mix and conditions

Component	Volume [µl]
Purified amplicons	17.0
NEXTflex™ Adenylation Mix (Bioo Scientific)	4.0
Total	21.0
Incubation for 30 minutes at 37°C	

Table 13: Adapter ligation reaction mix

A) Components for Adapter Ligation	Volume [µl]
3' Adenylated DNA	20.0
NEXTflex™ DNA Barcode Adapter 1-48 (Bioo Scientific) (12.5 µM)*	2.0
Switch Solution (Ion AmpliSeq™ Kit 2.0, Life Technologies)	4.0
DNA Ligase (Ion AmpliSeq™ Kit 2.0, Life Technologies)	2.0
Nuclease-free water	2.0
Total	30.0

* 5'AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC T

* 5'GAT CGG AAG AGC ACA CGT CTG AAC TCC AGT CAC XXX XXX ATC TCG TAT GCC GTC TTC TGC TTG

B) Conditions

Temperature	Time
22°C	30 Minutes
72°C	10 Minutes
10°C	Hold

Further, purification step II was done with 1.8-fold reaction volume of magnetic beads to remove fragments <100 bp. DNA was eluted with 40 µl nuclease-free water. To enrich amplicons with library sizes between 200 and 400 bp size selection was performed with 0.8-fold and 0.6-fold reaction volume of magnetic

beads. The DNA was eluted with 23 µl nuclease-free water. The adapter carrying amplicons were enriched by a final PCR step (Table 14) and purified afterwards.

Table 14: PCR parameters for the final PCR amplification of adapter ligated DNA libraries; A) PCR components; B) PCR conditions

A) PCR Setup for Final Adapter-Specific PCR	
Component	Volume [µl]
Platinum® PCR SuperMix High Fidelity (Ion AmpliSeq™ Kit 2.0, Life Technologies)	25.0
NEXTflex™ Primer Mix (Bioo Scientific) (12.5 µM)*	2.0
DNA library	23.0
Total	50.0

B) PCR Amplification Parameters		
Stage	Temperature	Time
Hold	98°C	2 Minutes
10 Cycles	98°C	15 Seconds
	60°C	60 Seconds
Hold	10°C	∞

The final purification step III was done with 1-fold reaction volume of magnetic beads and elution of DNA with 20 µl nuclease-free water.

3.5.2.3 Quality Assessment and Quantification of the Constructed Libraries Followed by Sequencing

The quality of the enriched libraries was analyzed by microfluidic electrophoresis using a High Sensitivity DNA Assay on the 2100 Bioanalyzer (Agilent Technologies) according to the manufacturer's instructions. Briefly, the samples were diluted 1:4 and 1µl was run on a primed chip along with DNA markers for size determination.

Amplicon library quantity was determined by qPCR using primers covering the Illumina adapter sequences. Five 5-fold dilutions of PhiX Control V3 (Illumina) in a range from 0.064 up to 40 pM served as reference standard (Figure 6).

3. Material and Methods

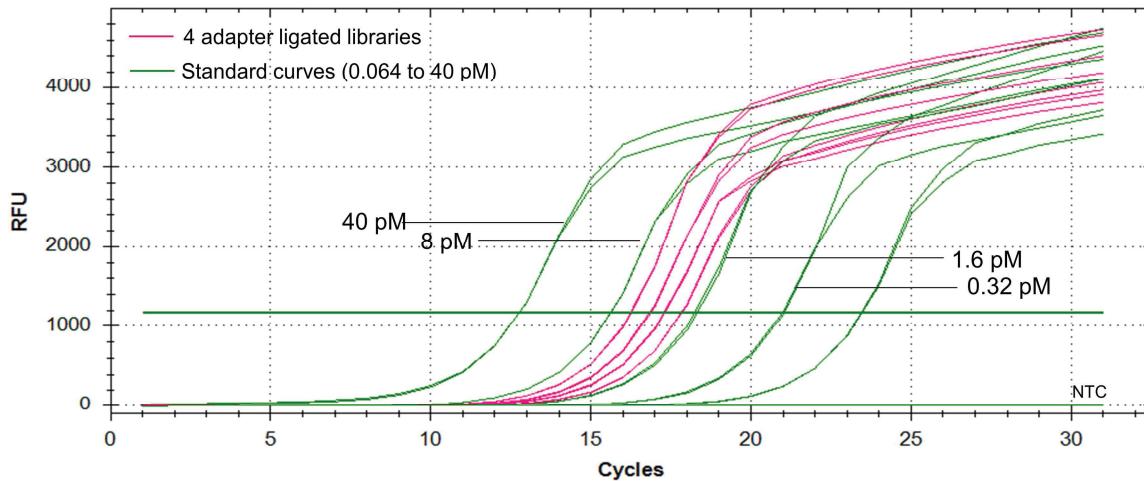


Figure 6: Amplification curves of the established standard series (green) and four different adapter ligated libraries (pink)

The library samples were diluted 1:4,000 and measured in duplicates (Table 15).

Table 15: PCR parameters for quantification of the constructed libraries by qPCR; A) PCR components; B) PCR conditions

A) qPCR Setup

Component	Volume [μl]
Nuclease-free water	7.4
Forward primer (10 μM) ^{*1}	0.8
Reverse primer (10 μM) ^{*2}	0.8
GoTaq® qPCR Master Mix (Promega)	10.0
Total	19.0
DNA library (diluted 1:4,000)	1.0
Total	20.0

^{*1} 5'AAT GAT ACG GCG ACC ACC GAG ATC TAC AC

^{*2} 5'CAA GCA GAA GAC GGC ATA CGA GAT

B) qPCR Amplification Parameters

Stage	Temperature	Time
Hold	94°C	3 Minutes
	94°C	30 Seconds
30 Cycles	60°C	30 Seconds
	72°C	30 Seconds

For sequencing, samples were pooled in an equimolar ratio according to results of the qPCR. 12-15 pM library pools including 1%-2.5% PhiX Control V3 library were prepared for sequencing according to MiSeq System User Guide (Illumina). Finally, paired-end sequencing was carried out on a MiSeq instrument (Illumina) using the v2 chemistry and 300 cycle cartridges as recommended by the manufacturer.

An overview of the developed NGS workflow is shown in Figure 7.

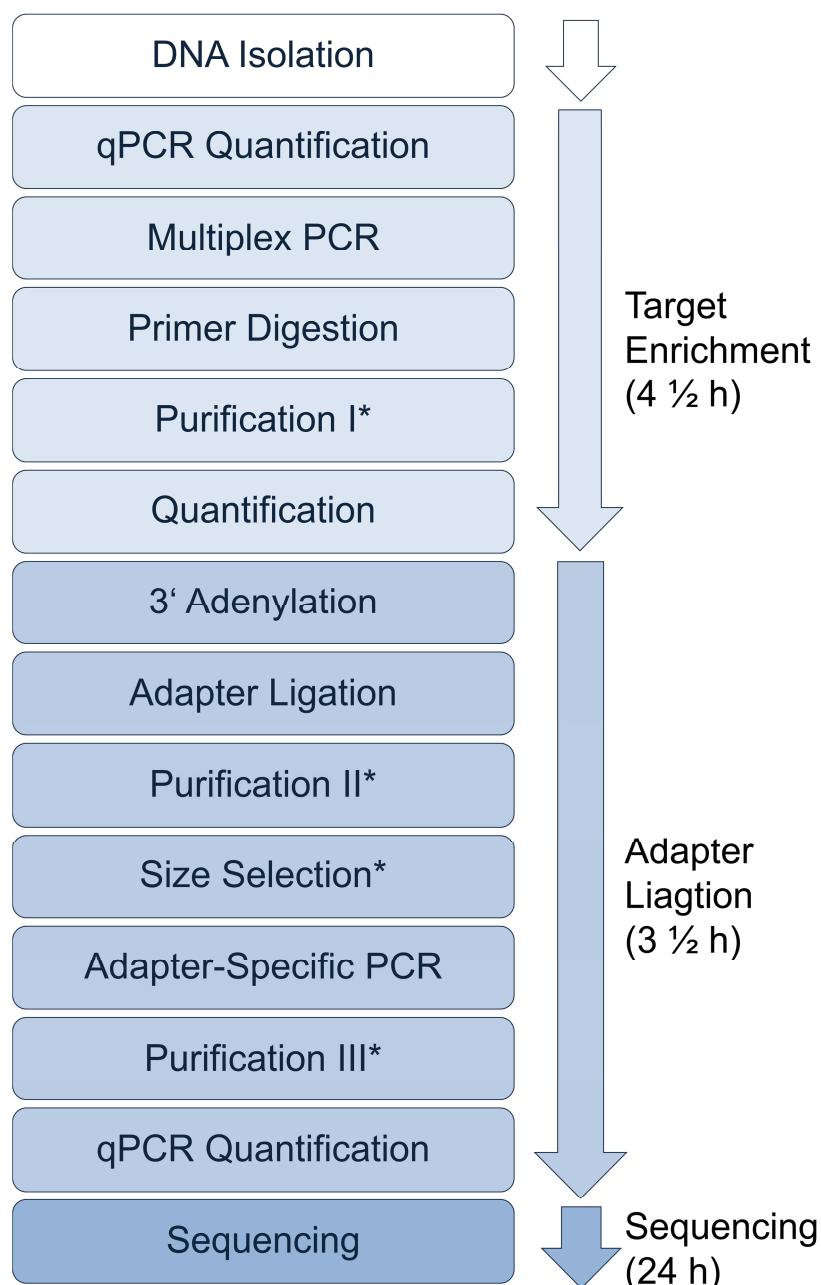


Figure 7: The flow chart shows an overview of the developed workflow for targeted next generation sequencing of CLL samples; *automated process.

3. Material and Methods

3.5.3 Improvement of Adapter Ligation Efficiency

To prove the efficiency of the adapter ligation, generated libraries were quantified by fluorescence absorbance (QuantiFluor™; Promega) according to the manufacturer's instructions and by qPCR with primers covering the Illumina adapter sequences (Table 15). The results were compared and the difference between both methods was used to assess adapter ligation efficiency. Furthermore, library sizes before and after adapter ligation were analyzed by microfluidic gelelectrophoresis (2100 Bioanalyzer, Agilent Technologies).

3.5.4 Automation of Sample Preparation Process

To check the feasibility of robotics for NGS process automation, CLL samples were subjected to automated purification and size selection with a Biomek® FXp workstation (Beckman Coulter). Manual sample preparation served as a reference control. Quality assessment and quantification of all steps was performed either by microfluidic electrophoresis using the 2100 Bioanalyzer (Agilent Technologies) or by qPCR analysis as described above. Finally, libraries were constructed and sequenced with the MiSeq instrument as described above (Illumina).

3.6 Determination of Sequencing Sensitivity

The mantle cell lymphoma cell line, Mino (kindly provided by M. Herling, Cologne, Germany), carrying a known *TP53* mutation (c.440T>G; p.V147G, exon 5) [39] and the AT45RM cell line (kindly provided by L. Chessa, Rome, Italy) containing an *ATM* mutation (c.7792C>T; p.R2598, exon 53) [45] were used to evaluate the sensitivity of NGS mutation detection. Cells were cultured according to standard protocols. DNA was extracted and sequenced as described above. 200 to 9,000 genomic copies of each cell line DNA were diluted in wild type DNA from human embryonic kidney cells (HEK-293, American Type Culture Collection ATCC) harboring no known gene modification (Table 16).

Table 16: Dilution of different mutated cell line DNA; *calculation of cells based on 7 pg DNA per cell

Cell Line	Dilution 1	Dilution 2	Dilution 3	Dilution 4	Dilution 5	Dilution 6
Mino	5% 1.50ng 214 cells*	10% 3.00ng 429 cells*	15% 4.50ng 643 cells*	20% 6.00ng 857 cells*	50% 15.00ng 2,143 cells*	100% 30.00ng 4,286 cells*
HEK	95% 28.50ng 4,071 cells*	90% 27.00ng 3,857 cells*	85% 25.50ng 3,643 cells*	80% 24.00ng 3,429 cells*	50% 15.00ng 2,143 cells*	-
AT45RM	5% 0.75ng 107 cells*	10% 1.50ng 214 cells*	15% 2.25ng 321 cells*	20% 3.00ng 429 cells*	50% 7.50ng 1,071 cells*	100% 15.00ng 2,143 cells*
HEK	95% 14.25ng 2,036 cells*	90% 13.50ng 1,929 cells*	85% 12.75ng 1,821 cells*	80% 12.00ng 1,714 cells*	50% 7.50ng 1,071 cells*	-

3.7 Sequencing Data Analysis

Fastq files generated by the MiSeq Reporter Software (Illumina) were analyzed with an in-house developed bioinformatics pipeline based on our general cancer genome analysis tool, which was further optimized for the diagnostic workflow [58]. Before aligning raw reads to the reference genome, adaptor sequences were trimmed. The resulting data was then aligned to the coordinates of the target amplicons according to hg19 reference genome (NCBI build 37) using the Burrows-Wheeler alignment tool [44]. In order to increase the sensitivity for longer deletions or insertions, we realigned all unmapped reads to the reference by using the BLAT (BLAST (basic local alignment tool)-like alignment tool) [36]. Sequencing errors were estimated from known germline polymorphisms and each sequenced genomic position was analyzed for variants that exceed this background rate. Detected variants were finally annotated and filtered to remove germline variants by using the databases dbSNP (single nucleotide polymorphism database; <http://www.ncbi.nlm.nih.gov/SNP/>) and the exome variant server (<http://evs.gs.washington.edu/EVS/>).

Obtained variants with an allelic frequency below 5% and synonymous variants were removed. Additional, visual analysis of called variants was performed by means of the Integrative Genomic Viewer (IGV, Broad Institute, Cambridge, MA, USA). False positive variants, particularly in repetitive or highly homologous regions of the genome, variants in high background noise, as well as single strand

3. Material and Methods

variants, were either eliminated when they were clearly recognizable as artifacts, or were further re-assessed by Sanger sequencing.

Furthermore, variants were analyzed for their functional impact on the protein by MutationAssessor (<http://mutationassessor.org>, release 2) [65] and implementation of the ANNOVAR algorithm [93], which combines the bioinformatics tools SIFT (sorting tolerant from intolerant) [38], PolyPhen2 (polymorphism phenotyping 2) [1] and MutationTaster [80].

3.8 Variant Confirmation

A subset of variants, including variants with less than 100 reads, was confirmed by conventional Sanger sequencing using the BigDye® Terminator v3.1 Cycle Sequencing Kit (Life Technologies) (Table 17, Table S2). Variants that could not be confirmed were excluded from further analysis.

Table 17: PCR reaction mix and conditions for Sanger sequencing; A1 and 2) components and conditions for amplification of target regions by PCR I; B1 and 2) components and conditions for Sanger sequencing of target regions by PCR II

A1) PCR Setup		B1) Sanger PCR Setup			
Component	Volume [μ l]	Component	Volume [μ l]		
Nuclease-free water	11.5-Y	Nuclease-free water	13.5		
Primer forward (10 μ M)	0.5	Primer forward or reverse (10 μ M)	0.5		
Primer reverse (10 μ M)	0.5	5x Big Dye sequencing buffer (Life Technologies)	4.5		
Multiplex PCR Master Mix (Qiagen)	12.5	Big Dye® Terminator v3.1 Cycle (Life Technologies)	0.5		
Total	24.0	Total	24.0		
gDNA (10 ng)	Y	DNA amplicons	1.0		
Total	25.0	Total	20.0		
A2) PCR amplification parameters			B2) Sanger PCR amplification parameters		
Stage	Temperature	Time	Stage		
Hold	94°C	15 min	Hold	95°C	1 min
40 Cycles	94°C	30 sec	34 Cycles	96°C	30 sec
	X°C*	90 sec		X°C*	60 sec
	72°C	60 sec		60°C	4 min
	Hold	72°C	Hold	60°C	10 min

* Annealing temperature of specific primer pair

3.9 Statistical Analysis

Association between mutation subsets of six of the eight mutated genes harboring at least thirteen variants (*ATM*, *DDX3X*, *NOTCH1*, *SF3B1*, *TP53*, *XPO1*) and following clinical and/or prognostic covariates were assessed applying standard statistical tests: gender (female/male), age >65 years (yes/no), CD38 surface expression and positivity, ZAP70 surface expression and positivity, platelet count, WBC, TK count, Binet staging A/B vs C, unmutated *IGHV* status, treatment status, FISH hierarchy. Furthermore, we tested chromosomal aberration status of del17p and *TP53* mutation status, del11q and *ATM* mutation status as well as trisomy 12 and *NOTCH1* mutation status.

Additionally to the overall cohort of patient samples, a subgroup analysis of treated and untreated patients was done.

Statistical calculations were computed in R version 3.1.0 (R Foundation for Statistical Computing, Vienna, Austria). Associations to dichotomous variables (e.g. gender) were analyzed using Exact Wilcoxon Mann-Whitney Rank Sum test. Double dichotomous contingency tables were analyzed using Fisher's exact test. To test dependence to ranked parameters with more than two groups (e.g. Binet staging) the Pearson's Chi-squared test was used.

All reported P-values were considered significant at $P \leq 0.05$.

4. Results

In the present study the multiplex PCR target enrichment linked to NGS was successfully tested as a novel tool for clinical diagnosis of CLL. In a first approach a commercial gene panel for analysis of leukemia relevant marker genes was applied to a limited number of CLL DNA samples (4.1). Since the genes covered leukemia associated genes in general, but not genes and hotspots that are particularly important for CLL prognosis and therapeutical evaluation, in a second approach (4.2) primer sets for CLL specific gene enrichment were designed and used for the multiplex PCR linked to NGS. The technology was optimized concerning library construction (4.2), then it was applied to more than 100 B-cell DNA samples from patients with leukemia (4.3-4.6) and proven for association to clinical parameters (4.7).

4.1 MiSeq Sequencing Using a Multiplex PCR Approach for Leukemia Specific Gene Enrichment

The sequencing of the twelve CLL samples prepared with a commercially available enrichment panel produced 12.97×10^6 reads and an output of 3.7 gigabases in one NGS run. Equimolar sample pooling according to the results of the final library quantification by qPCR resulted in a heterogeneous distribution of reads per sample (Figure 8). Remarkably, sample 1, 2 and 5 demonstrated less reads than the other samples.

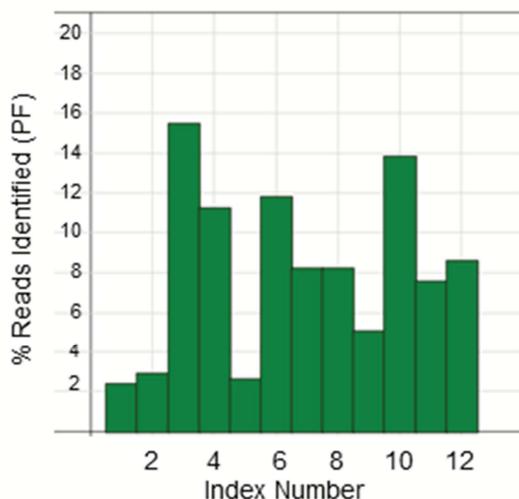


Figure 8: Distribution of reads per sample in percentage of the sequencing run including twelve CLL samples each with 1,272 amplicons covering 20 leukemia relevant genes

Read quality parameters of the sequencing run showed Q30 scores greater than 86%. A score of Q30 corresponds to a probability of base-call error of 1 in 1,000 or 99.9% accuracy, respectively. The bridge amplification of the target regions resulted in a cluster density of 665 K/mm² with 93% passing the quality filter parameters of the instrument.

The mean coverage per gene was 11,028 reads. However, some genes demonstrated remarkable deviations from the mean read count. Thus *CEBPA*, *RUNX1* and *GATA2* were clearly underrepresented in comparison to the other 17 sequenced genes (Figure 9). For these three genes a mean coverage of 2,120, 1,653 and 2,744 reads per gene and run was generated.

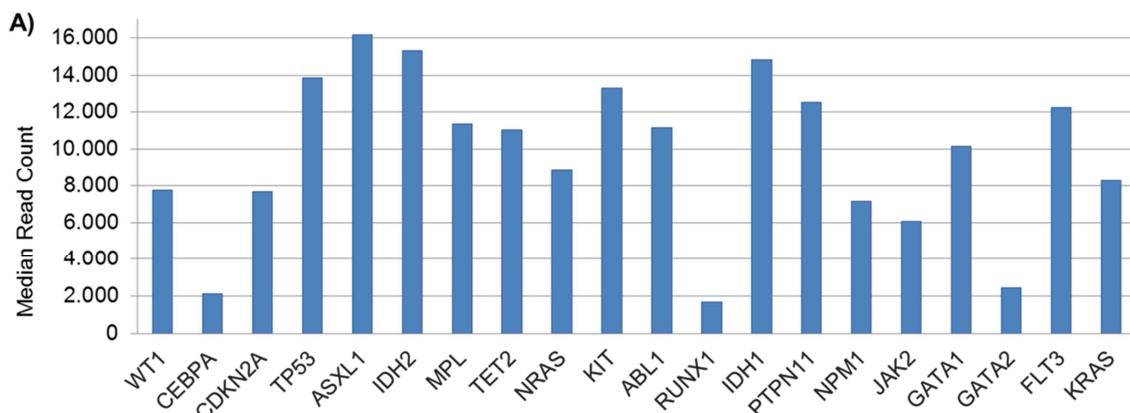


Figure 9: Median coverage per gene and run are shown; 20 leukemia related genes were sequenced with 12 CLL samples. *CEBPA*, *RUNX1* and *GATA2* showed remarkably lower reads compared to the other 17 tested genes.

The variant analysis revealed 83% of the detected variants as single nucleotide variant. By far, most variants were found in the genes *TET2* (16%), *FLT3* (13%), *ASXL* (10%), *TP53* (10%) and *WT1* (8%).

4.2 Protocol Development for Target Specific Sequencing of Native B-Cell DNA

4.2.1 Determination of the Amount of Input B-Cell DNA

All B-cell DNA samples were evaluated quantitatively by fluorescence absorbance and by real-time qPCR method using *HFE* gene as the amplifying target. The

4. Results

values of DNA concentration ranged from 0.34 to 29.06 ng/ μ l by fluorescence measurement and from 0.26 to 18.37 ng/ μ l by qPCR (Table 18).

Table 18: Comparison of DNA concentrations of 136 CLL samples measured by two different quantification methods

Method	Mean	Median	Minimum	Maximum
Fluorescence absorbance [ng/ μ]	6.04	3.04	0.34	29.06
qPCR [ng/ μ l]	5.13	2.52	0.26	18.37

Even if the fluorescence absorbance method showed 1.8-fold higher DNA concentrations in mean, a correlation with the qPCR approach ($P<0.0001$, $\rho=0.92$, Figure 10) was observed.

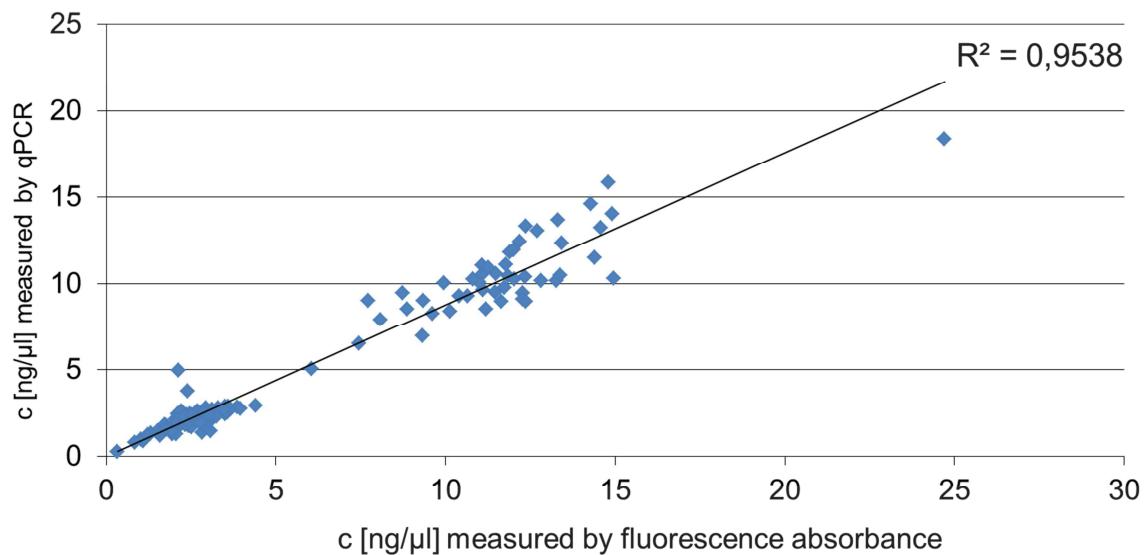


Figure 10: Correlation of the B-cell DNA concentrations, measured by two different quantification methods are shown; statistical outliers (second degree) were excluded ($n=2/136$).

4.2.2 Optimization of the Adapter Ligation Step

Quality assessment of the generated amplicons by microfluidic electrophoresis showed that target amplicon libraries had a size in the range of 100 to 200 bp and indicated high library DNA concentrations of approximately 300 ng/ μ l (Figure 11).

After adapter ligation, the molecular size of the libraries was shifted to sizes around 300 bp. However, a following adenlyation and adapter ligation of the libraries without prior dilution of the samples was inefficient leading only to 30% adapter ligated amplicons. A slight improvement of the adapter ligation was

achieved with 5-fold diluted libraries. 100% efficiency of the adenylation and adapter ligation reaction was reached with 10-fold diluted amplicons (Figure 11).

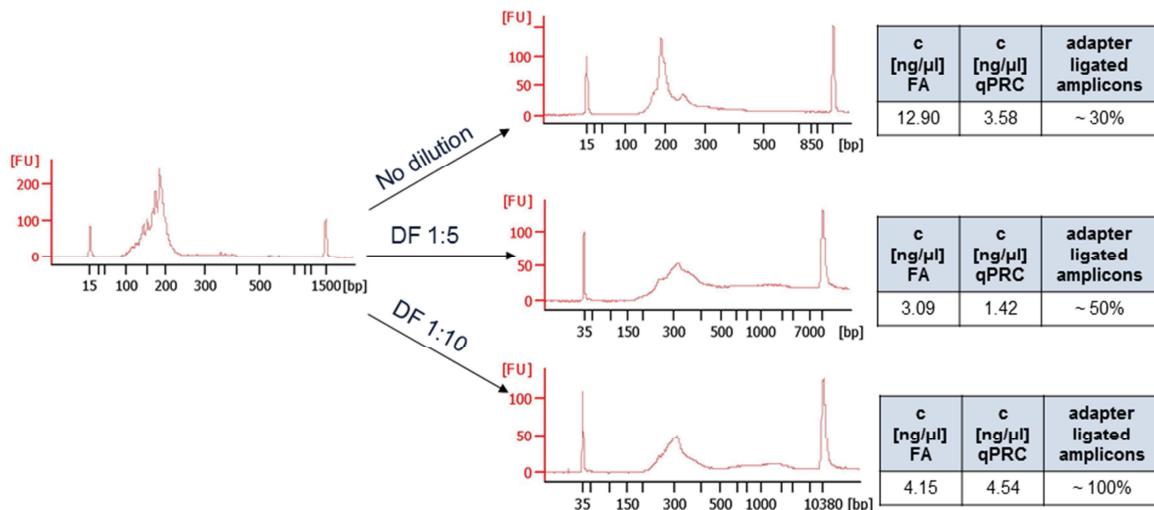


Figure 11: Amplicon quality assessment by microfluidic electrophoresis and quantification by fluorescence absorbance (FA) and qPCR demonstrated the relation between amplicon dilution (DF) and adapter ligation efficiency. Left electropherogram: amplicons after multiplex PCR amplification, right electropherograms: amplicons after adapter ligation; blue boxes: DNA concentrations measured by FA and qPCR and adapter ligation efficiency. Less input in adapter ligation after dilution of multiplex PCR libraries showed clearly improved ligation efficiency.

4.2.3 Final Quantification of Constructed Libraries

The quantification of the finalized target libraries showed differences in the results depending on the method that was used. When libraries were not diluted prior adapter ligation, ligation efficiency was less than 100% (4.2.2), fluorescence absorbance quantification of the final library resulted in higher amplicon concentrations compared to qPCR results.

Equimolar sample pooling according to the results of fluorescence absorbance measurement for sequencing run 1 led to a mean read count per sample of 2.78% with a high range from 0.74% to 7.13% (Figure 12). Samples that were pooled according to the library quantification by qPCR showed much lower deviation of reads between the samples. The three runs demonstrated mean read counts of 3.44% (range 2.43%-4.50%), 3.5% (range 3.66%-4.93%) and 3.44% (range 2.20%-4.93%), respectively.

4. Results

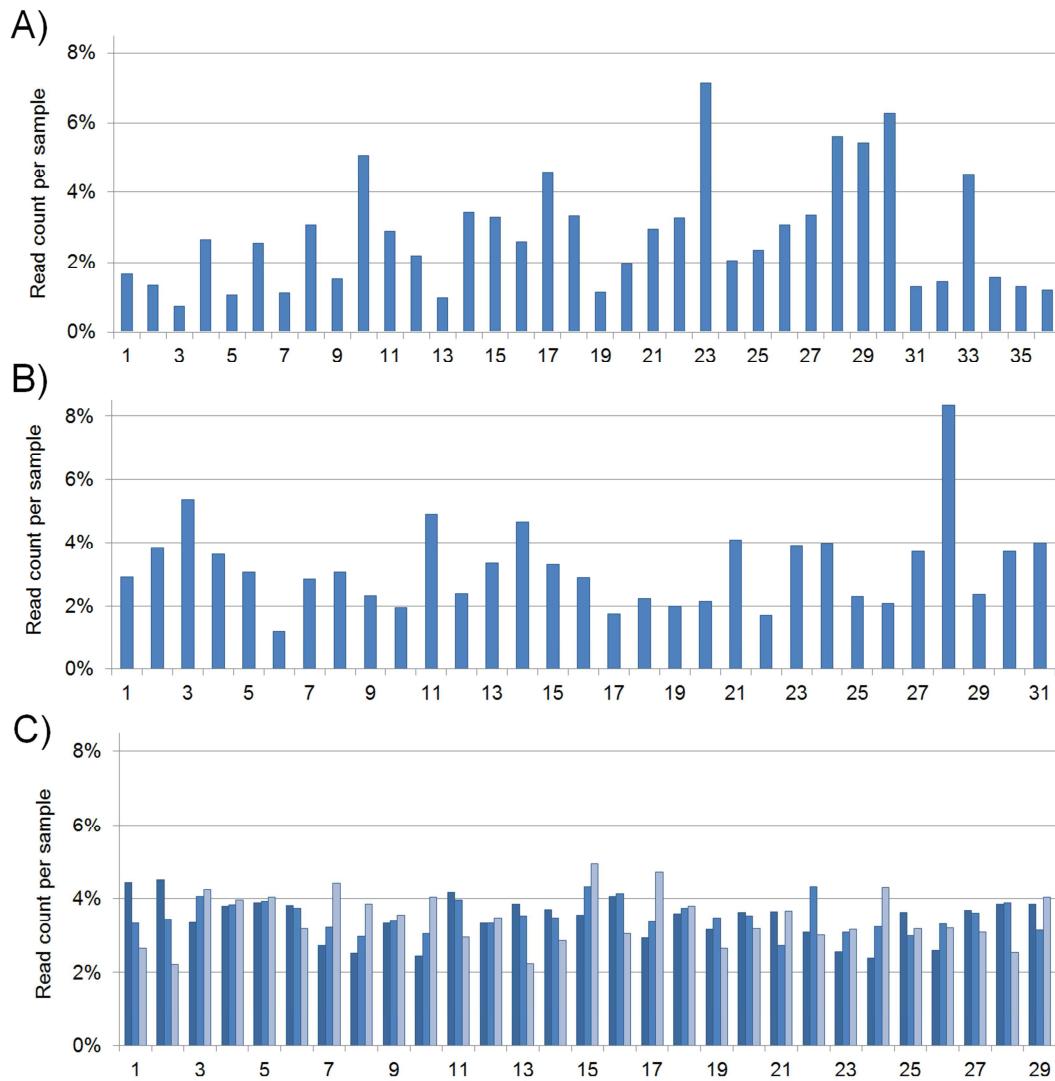


Figure 12: Improvement of read count per sample after protocol changes; A) run 1 containing 36 CLL samples and 2.04×10^6 reads with 2.78% mean read count per sample (range 0.74% to 7.13%) and B) run 2 containing 31 CLL samples and 6.63×10^6 reads with 3.23% mean read count per sample (range 1.21% to 8.33%) were performed according to the standard protocol. C) After quantification control by qPCR and adjustment of DNA concentration prior adapter ligation and sample loading, the deviation of gene coverages were diminished shown by run 3-5 containing 29 samples each and 13.86×10^6 , 10.48×10^6 and 14.66×10^6 reads with mean read count per sample of 3.44% (range 2.43% to 4.50%), 3.51% (range 3.66%-4.93%) and 3.44% (range 2.20%-4.93%), respectively.

4.2.4 Automated Sample Processing

The feasibility of automated purification and size selection for the NGS workflow was checked by comparing automatically and manually prepared CLL samples.

No significant difference of quality or quantity was determined between the automated and the manually prepared libraries. Each sample either prepared by manual handling or by automated process showed good sequencing results as exemplarily demonstrated in Figure 13.

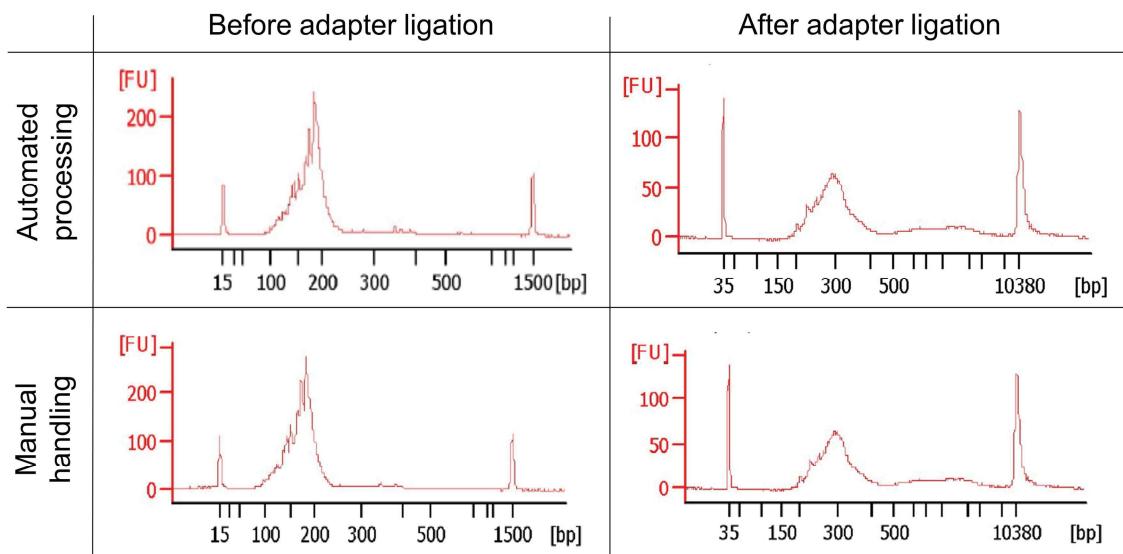


Figure 13: Quality assessment by microfluidic electrophoresis of amplicon libraries either manually or automatically processed showed no quality differences.

4.3 Target Specific Sequencing of Native B-Cell DNA

After successful optimization of the sample preparation procedure and assessment of a standardized workflow, target specific sequencing was performed on B-cell DNA from 136 CLL patients. Although the selected specimens, which were applied to target specific NGS evaluation, originate from a cohort that was rather heterogeneous in terms of age, gender, stage and prognostic factors. They included representative samples of a wide range of CLL patients with heterogeneous disease progression.

In particular, the cohort contained a broad proportion of samples obtained from untreated patients with an early CLL stage (Binet stage A), for which improved diagnostic and prognostic parameters are highly important. The clinical data of all patients included in our study are summarized in Table 7.

All samples could be successfully applied to five NGS runs, producing an average of 15.37×10^6 reads and an average output of 4.7 gigabases (Table 19). The best

4. Results

sequencing results were generated with library pools of 14 to 15 pM, including 0.8 mM NaOH and 1% PhiX control library.

Table 19: Run parameter of the five sequencing runs

Parameter	Run 1	Run 2	Run 3	Run 4	Run 5	
Entity	CLL	CLL	CLL	CLL	CLL	
Samples	36	31	29	30	29	
Amplicons [n]	338	338	338	338	338	
PhiX Control	2.5%	2.5%	1%	1%	1%	
NaOH	1.2 mM	1.2 mM	0.8 mM	0.8 mM	0.8 mM	
Library Pool	12 pM	12 pM	15 pM	15 pM	14 pM	Average
Cluster Density [K/mm²]	192	590	1,277	1,382	1,380	964
Cluster Passed Filter	90.07%	93.77%	89.27%	69.78%	88.67%	86.31%
Reads	3.75M	11.85M	24.38M	23.76M	26.37M	18.02M
Reads Passed Filter	3.38M	11.11M	21.76M	17.23M	23.38M	15.37M
Total Yield	1.00Gb	3.4Gb	6.6Gb	5.3Gb	7.1Gb	4.7Gb
≥ Q30 Score	97.70%	97.70%	95.30%	90.60%	95.00%	95.26%

Q30 Score 0.1% chance of wrong base call; M million; Gb gigabases

Data analysis of the fifteen genes covered by 338 amplicons demonstrated a mean coverage per exon in a range of 0 to 7,156 reads. Only for five of 167 exons (3%; *ATM* exon 20, *NOTCH1* exon 27, *SF3B1* exon 5 and 11 and *TP53* exon 11) the mean read count was less than 100, 83% of the targets were covered by more than 500 reads (Figure 14).

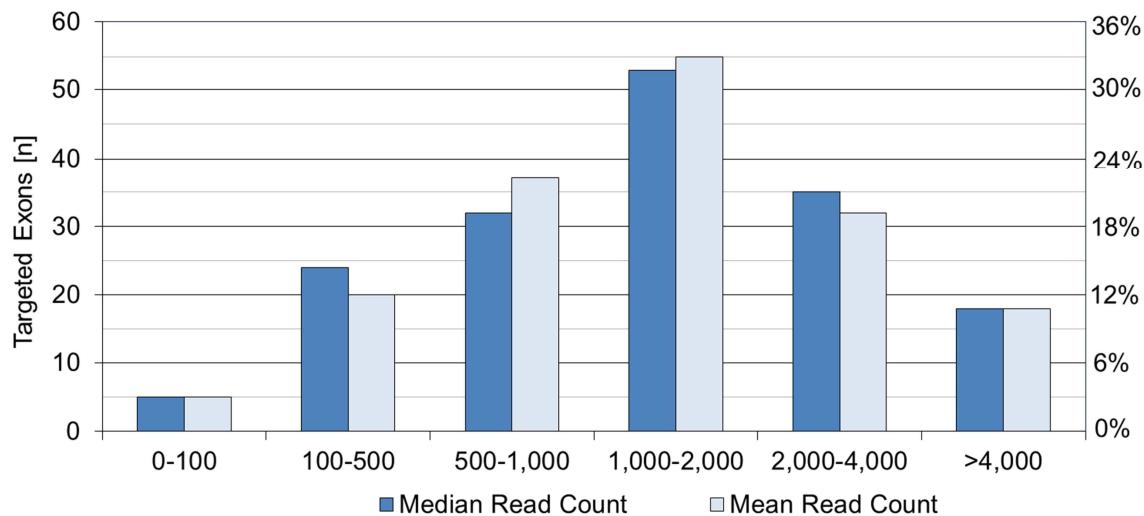


Figure 14: Median and mean read count per exon for 15 genes covered by 167 exons are shown; results reflect data from five NGS runs. Number of targeted exons is presented as absolute number [n] and as percentage of total number [%].

CLL patient samples produced a total of 4,396 variants after raw data alignment and first (automated) background removal. Exclusion of sequencing artifacts, synonymous variants, and polymorphisms, led to 167 significant variants (133 missense, 12 deletions, 8 indels, 8 splice site, 5 nonsense, 1 insertion) in eight genes in 86 of the 136 CLL samples (Figure 15). Ninety-two (55%) of the detected variants were predicted to functionally affect the corresponding protein by at least two of four program algorithms (Table S3). In the remaining 50 patients (36.8%) no variants could be identified.

4. Results

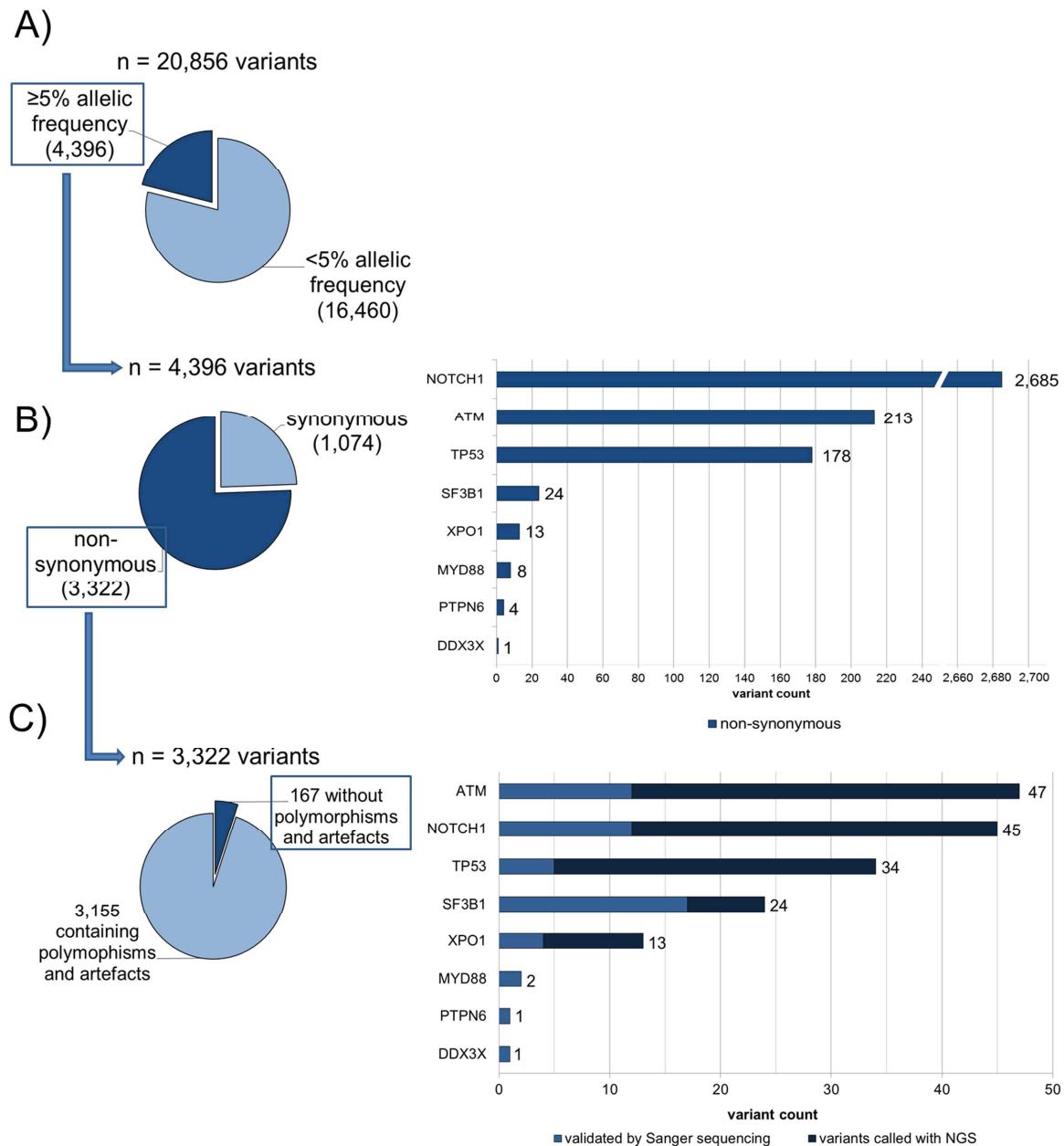


Figure 15: Algorithm of variant analysis A) Variants with an allelic frequency below 5% were discarded, resulting in a total of 4,396 variants. B) Only the 3,322 non-synonymous variants were used for further analysis. The variant count per gene is represented in the bar chart. C) Variants located in areas of high background noise and homopolymeric regions, probable phasing errors and single strand variants were visually identified in the Integrative Genomic Viewer (IGV, Broad Institute) and were removed. In doubtful cases, Sanger sequencing was performed to prove or disprove an alteration. This resulted in 167 final variants in 86 CLL sample.

4.4 High Sensitivity in *TP53* and *ATM* Mutant Identification

Since *TP53* and *ATM* mutations are of particular diagnostic interest, we tested the detection sensitivity of the target specific NGS approach. Therefore, the two cell lines Mino and AT45RM with known mutations in *TP53* exon 5 or *ATM* exon 53 had been selected as positive controls to estimate the sensitivity of our targeted NGS methods.

Analyzing the mutation rate in the fractional dilutions from 5% to 100%, the allelic frequency of the *TP53* and *ATM* cell line mutations detected by NGS followed a linear relationship with increasing amounts of tumor DNA (P -value 0.003, rho 1.000) (Figure 16). The *ATM* mutation p.R2598* was detected with a sensitivity of 10%. Dilutions of 5% and 10% *TP53*-mutated DNA (Mino cell line) were detected even with 2% and 5% allelic frequency, respectively.

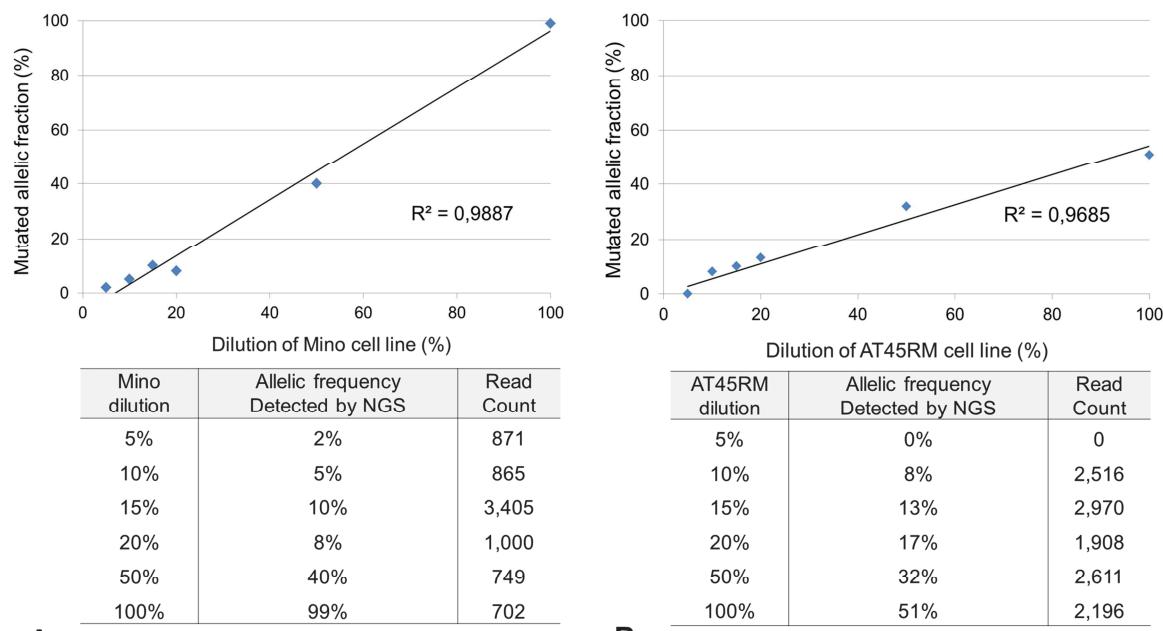


Figure 16: Linear relationship of mutation rate and allele frequency detected by NGS; sequencing of two dilution series of cell line DNA with known A) *TP53* c.440T>G; p.V147G mutation (Mino cell line) and B) *ATM* c.7792C>T; p.R2598* (AT45RM cell line) demonstrated a linear relationship between dilution and the mutation allele frequency obtained by NGS. Further, the data point to the sensitivity achieved by our NGS panel by detecting at least 214 *ATM* mutated AT45RM cells in a background of 2,036 wild type HEK-293 cells and 214 *TP53* mutated Mino cells in a background of 4,071 wild type HEK-293 cells.

4. Results

4.5 Variant Validation by Sanger Sequencing

One hundred-seventy-three variants in eight genes were detected. Eight of them were located in splice sites, namely five in *ATM* and three in *TP53*.

Fifty-eight variants (34%) were selected for further validation by Sanger sequencing (Figure 15). Fifty-four variants of the 58 were confirmed (93%), whereas four *NOTCH1* variants (p.S854F, p.C885R and p.V856A) in exon 16, which appeared with a low coverage of ≤ 190 reads, were not confirmed. Therefore, these variants were excluded for further analysis.

4.6 Cluster of Genes Affected by High Frequency Variants

The highest frequency of variants was detected in *ATM* and *NOTCH1* followed by *TP53*, *SF3B1* and *XPO1*, whereas *MYD88*, *PTPN6* and *DDX3X* showed only two or one variant, respectively (Figure 15C). No variants were found in *BTK*, *CD79B*, *FBXW7*, *MAPK1*, *PIK3CA*, *PIK3CD* and *PTEN*.

A total of 42 *ATM* variants appeared in 40 of the 136 CLL patients (29.4%), and were evenly distributed over the entire gene (Figure 17).

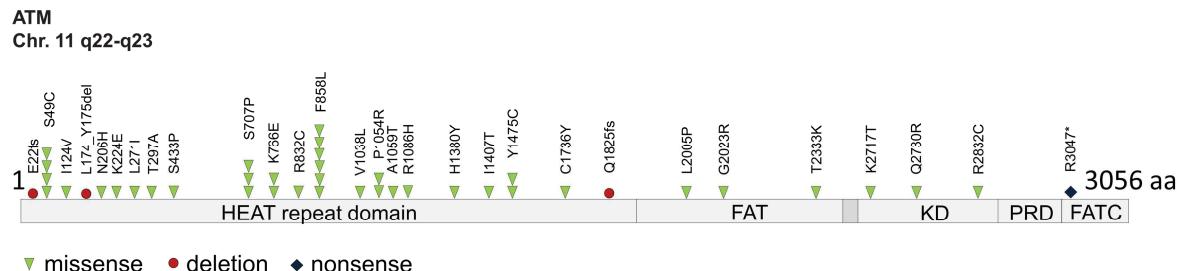


Figure 17: Alteration type, number of occurrence and location of detected variants in the *ATM* gene are shown. FAT FRAP-ATM-TRRAP (amino acid 1960-2566); KD protein kinase domain (amino acid 2712-2962); PRD PIKK-regulatory domain (amino acid 2961-3025); FATC FAT-c-term domain (amino acid 3024-3056)

Five patients revealed two different *ATM* variants within one sample. The most frequent alteration was c.2572T>C, p.F858L in exon 19 (6/40, 15.0%) (Table S3). Interestingly, most of the samples harboring an *ATM* variant showed at least one additional variant in another gene (Figure 18).

		ATM	NOTCH1	TP53	SF3B1	XPO1	PTPN6	DDX3X	MYD88
P079	1	1	1		1				
P025		1	1						
P074		1	1						
P093		1	1						
P123		1	1						
P134		1	1						
P053		1	1						
P031		1	1						
P028		1	2			1			
P044		1		1	1				
P101		1		1	1				
P119		1		1			1		
P115		1		1					
P058		1		1					
P085		1		1					
P105		1		1					
P117		1		1					
P095		1		2					
P132		1		3	1				
P108		1			1				
P088		1				1			
P122		1				1			
P052		1					1		
P004		1							
P012		1							
P017		1							
P065		1							
P076		1							
P083		1							
P124		1							
P009		1							
P042		1							
P114		2	1						
P125		2	1						
P084		2			1				
P048		2				1			
P133	37	2							
P032			1	1					
P071			1		1				
P080			1			1			
P016			1				1		
P023			1						
P027			1						
P030			1						
P034			1						
P038			1						
P039			1						
P046			1						
P059			1						
P061			1						
P062			1						
P066			1						
P090			1						
P097			1						
P109			1						
P120			1						
P121			1						
P094		2	1			1			
P001		2				1			
P072		2				1			
P130		2				1			
P126		2							
P111		3	1	1					
P091			1	1		1			
P110			1	1					
P086			1	1					
P045			1						
P082			1						
P040			2						
P112			2						
P064			3						
P067			3						
P057					1	1			
P021					1				
P011					1				
P015					1				
P050					1				
P056					1				
P081					1				
P035					2				
P100					2				
P037						1			
P068						1			
P096						1			
P078							1		
P103	86								

Figure 18: Genetic profile of 86 CLL samples carrying gene variants determined by NGS.

Each row summarizes the variants of one patient, each column summarizes variants occurring in one specific gene. Per each gene the number of variants is given per patient. Dark blue samples indicate patients with aberrations on chromosome 11 (del11q) for ATM mutated cases or on chromosome 17 (del17p) for TP53 mutated cases.

4. Results

Five *ATM* mutated patients had also a heterozygous deletion of *ATM* detected by FISH analysis.

Nearly 98% (2,691) of detected *NOTCH1* variants were polymorphisms or sequencing errors. From the remaining 45 variants occurring in 42 patients (30.8%) 15 were functional (Table S3). They covered the complete coding region but were mostly located in the notch extracellular part (NEC, 30/45, 67%) (Figure 19).

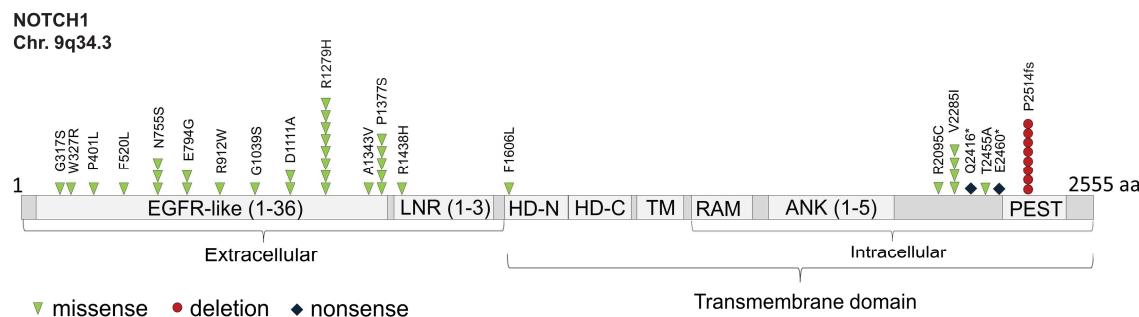


Figure 19: Alteration type, number of occurrence and location of detected variants in the NOTCH1 gene are shown. The NOTCH1 extracellular domain contains epidermal growth factor (EGFR)-like repeats (amino acid 20-1426); LNR Lin-12 NOTCH repeats (amino acid 1449-1571); HD-N/C heterodimerization domain (N-terminus, C-terminus); RAM RAM domain; ANK ankyrin repeat domain (amino acid 1927-2089); PEST Pro-Glu-Ser-Thr motif for degradation (amino acid 2507-2526)

One CLL patient had three, and six patients exhibited two different variants. Interestingly, one patient showed a missense mutation located in exon 26 (c.4816 T>C, p.F1606L) affecting the heterodimerization domain (HD) of the Notch1 protein. Seven (7/136, 5.1%) patients exhibited the known c.7541_7542delCT, p.P2514fs mutation located in the PEST domain encoded by exon 34 [61].

In 24/136 (17.7%) patients we found a total of 34 variants in *TP53*. Most of them (23/24; 68%) occurred in exon 6 to 8, which encode the DNA binding domain of the tumor suppressor and were identified to disrupt the *TP53* DNA binding function (Figure 20).

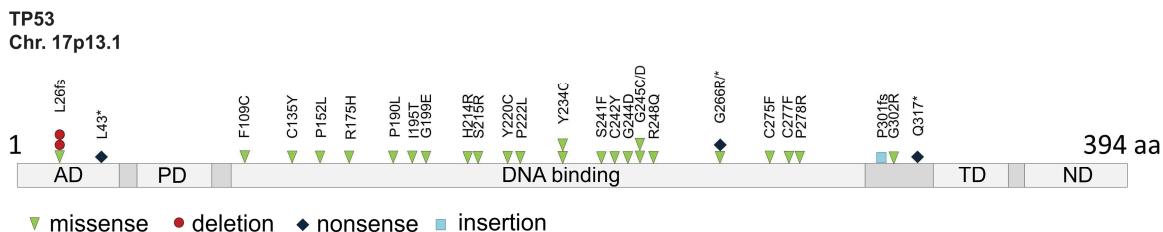


Figure 20: Alteration type, number of occurrence and location of detected variants in the TP53 gene are shown. AD activation domain (amino acid 1-50); PD proline-rich domain (amino acid 63-97); TD tetramerization domain (amino acid 323-356); ND negative regulation domain (amino acid 363-393)

SF3B1 showed a variant frequency of 16% (22/136 patients) clustering in exon 14 to 16 coding for HEAT repeats, which are responsible for the assembling of the splicing machinery (Figure 21) [94].

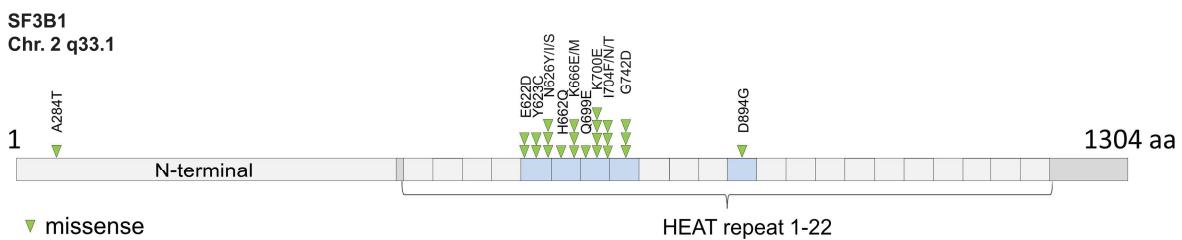


Figure 21: Alteration type, number of occurrence and location of detected variants in the SF3B1 gene are shown. The majority of SF3B1 alterations are clustered in the region encoding the highly conserved HEAT (huntingtin, elongation factor 3, protein phosphatase 2A, target of rapamycin 1) repeats 5-8. Multiple specific and overlapping docking or binding domains for numerous splicing factor partners occur within the N-terminal domain (amino acids 1-450), including sites important for interactions with U2AF1/2, and cyclin E.

The most frequent mutation was determined as c.2098A>G, p.K700E in exon 15 (4/24, 16.7%). All except one variant were predicted to cause a dysfunctional protein (Table S3).

Highly frequent and functionally relevant exon 15 mutations in XPO1 occurred in 13 patients (9.6%, p.E571I/K/Q) [32, 61].

Furthermore, we found two MYD88 mutations (c.649G>T, p.V217F, exon 3 and c.613T>C, p.L265P, exon 5) in two of the 136 patients (1.5%), both were determined as functionally damaging.

Only one patient (0.7%) exhibited a mutation in exon 9 of DDX3X (c.823A>C, p.T275P) and one patient in exon 11 of PTPN6 (c.1351G>A, p.V453M). The latter

4. Results

one was located in the highly conserved catalytic protein-tyrosine phosphatase domain of SHP-1 (*PTPN6*) and is predicted to impair the protein function.

Non-synonymous variants with an allelic frequency $\geq 5\%$ are summarized in Table S3.

4.7 Associations with Clinical Parameters

The overall 92 variants with predicted functional impact were tested for associations with clinical and prognostic parameters as available in our dataset. The majority of patients without any predicted damaging DNA alteration detected by our sequencing panel presented significantly more frequently with early stage (Binet stages A/B, 45/50, 90%; $p=0.03$) or previously untreated CLL (40/50, 80%; $p=0.001$) at time of sample of sample collection.

The presence of mutations in *NOTCH1*, *SF3B1*, *TP53* and *XPO1* was associated with at least one unfavorable prognostic marker such as unmutated *IGHV* gene status or positivity for ZAP70 or CD38 expression (Table 20). *SF3B1* mutated patients showed significantly more often an unmutated *IGHV* status (16/22, 73% vs. 5/22, 23%, $P=0.03$) and were of male gender (95% vs. 5%, $P=0.008$) (Figure 22).

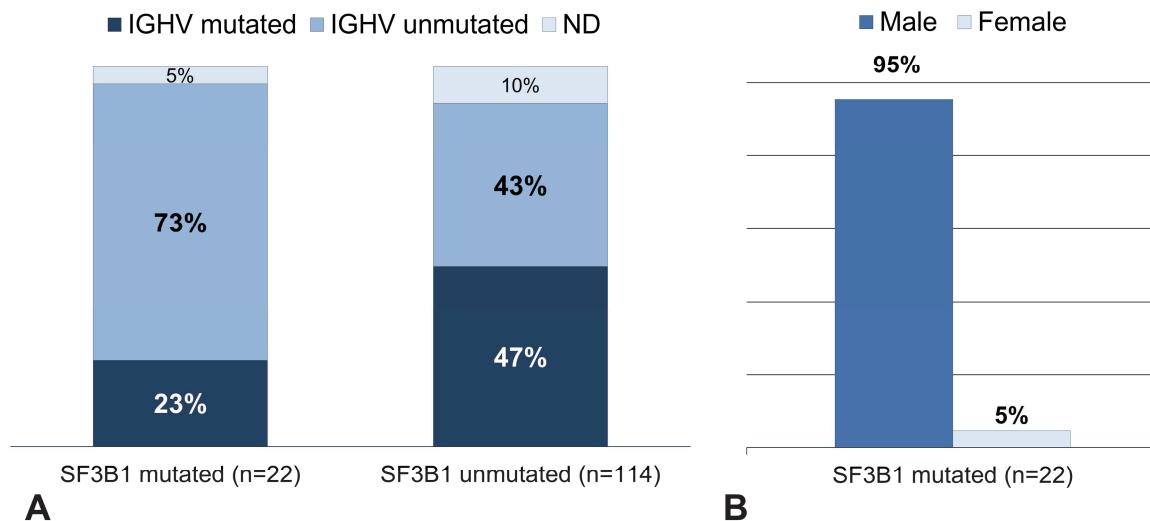


Figure 22: Associations of the identified functional *SF3B1* mutations with *IGHV* mutational status and gender are shown. A) CLL patients with *SF3B1* mutations were mostly negative for mutated *IGHV*, in contrast to patients with no *SF3B1* mutation who show a normal mutated *IGHV* status ($P=0.03$). B) *SF3B1* mutated B-cells occur significantly more frequent in male than in female ($P=0.008$).

Similarly, an unmutated *IGHV* status occurred more frequently in *TP53* mutants (16/22, 73% vs. 5/22, 23%, P=0.04), and in untreated patients with *NOTCH1* PEST domain mutation (6/7, 86% vs. 1/7, 14%, P=0.04), compared to their respective wild type counterparts (Figure 23).

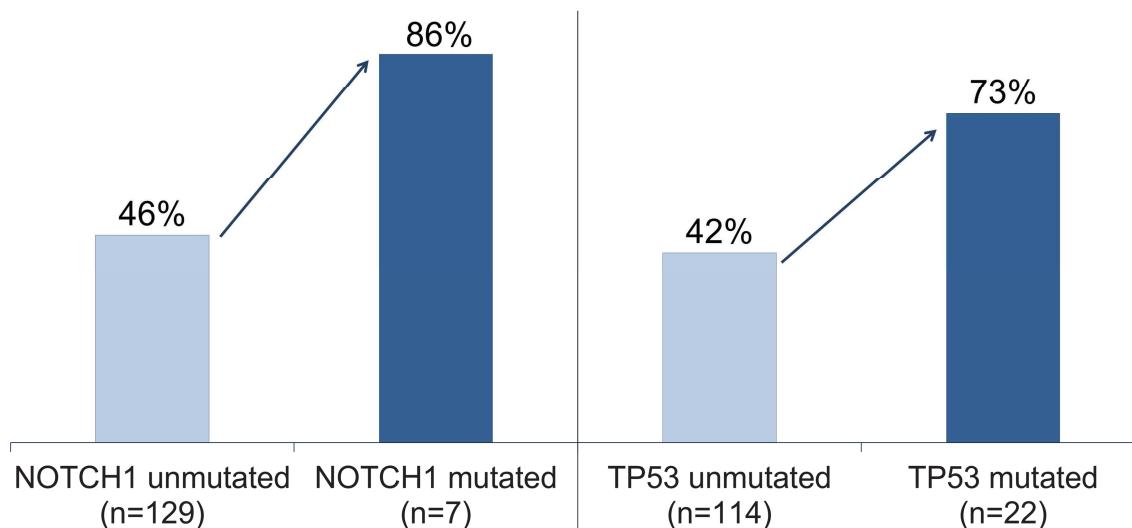


Figure 23: Distribution of unmutated *IGHV* status in patients with or without *NOTCH1* (P=0.04) and *TP53* mutations (P=0.04).

In the untreated patient cohort, the presence of *SF3B1* mutations significantly correlated with positivity for CD38 expression (P=0.03).

In treated patients, ZAP70 positivity was found more often in patients with mutated *XPO1* B-cells (P=0.02). Patients with mutations in *XPO1* also exhibited significantly increased WBC, possibly reflecting the proliferative capacity of tumor B-cells, compared to patients without mutations in the exportin-1 gene (mean: 134 vs. $65 \times 10^9/L$, P<0.001) (Figure 24).

4. Results

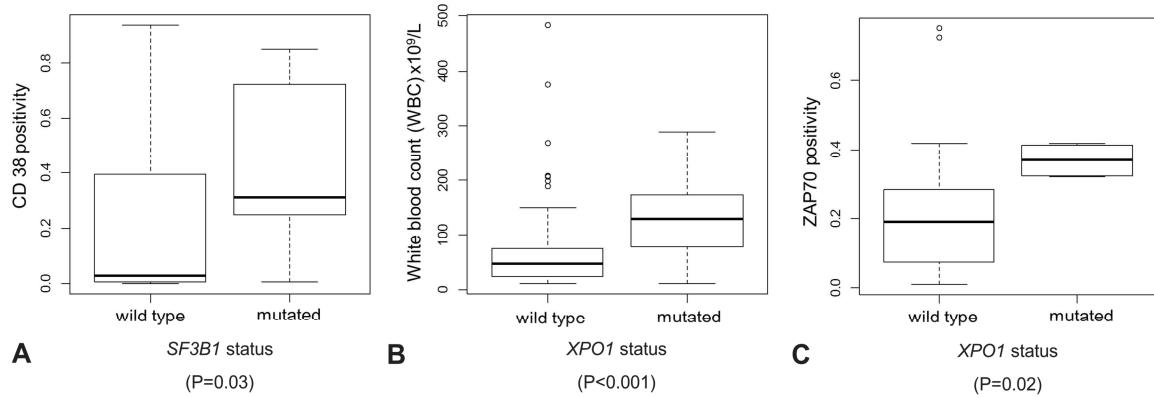


Figure 24: The presence of functional relevant mutations in *SF3B1* (A), and *XPO1* (B and C) was associated with the unfavorable prognostic markers like positivity for ZAP70 or CD38.

Only *TP53* mutations were found to be enriched in treated versus untreated patients (12/41, 30% vs. 10/95, 11%, $P<0.001$), indicating a possible selection of these mutations due to prior therapies. Furthermore, in patients with *TP53* mutations in addition deletions in chromosome 17 were found very often. These deletions then affect the second *TP53* allele and result in a complete disruption of the TP53 protein function ($P<0.001$) (Table 20).

In 127 patients Binet stages were available. *TP53* mutations were found to be particularly more frequent in patients with intermediate and advanced stage with a high need for treatment (Binet stage B/C). Thus, 64% of the patients (14/22) with a Binet stage B/C CLL have got *TP53* mutated B-cells, but only 23% of patients (5/22) with an early CLL stage (Binet stage A) carry TP53 mutations (Figure 25).

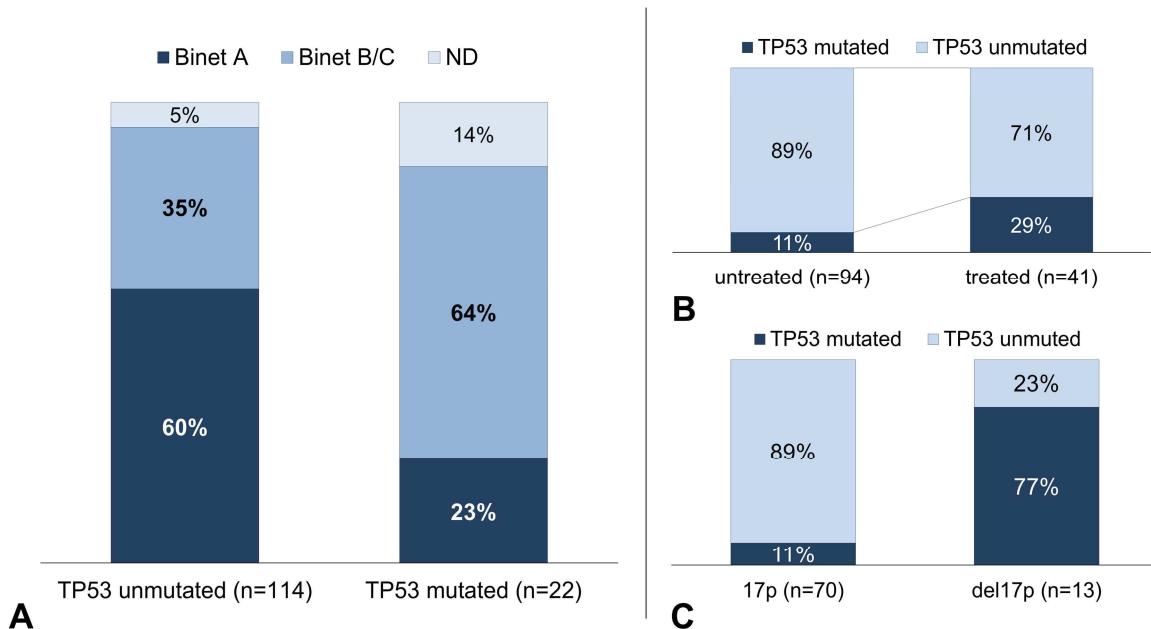


Figure 25: Associations of detected functional *TP53* mutations with different clinical and prognostic parameters are shown. A) *TP53* mutations were found particularly more frequent in intermediate and advanced stage (Binet B/C) compared with patients in an early stage (stage A) ($P=0.008$). B) and C) The mutations were also frequently more detected in treated patients ($P<0.001$) and in patients with genomic aberrations on chromosome 17 (del17p) ($P<0.001$)

No correlations between mutations and biological or clinical parameters were found for *ATM* and *DDX3X*.

4. Results

Table 20: Statistical correlations* between gene mutation status and clinical and biological parameters are summarized; correlations were calculated only for variants that were predicted to have a functional impact on the protein.

	Gene	Clinical/Prognostic Parameter	P-Value
Overall Cohort	<i>SF3B1</i>	Male sex	0.008
		Unmutated <i>IGHV</i>	0.030
		Decreased platelet count	0.025
	<i>TP53</i>	Binet stage	0.008
		Chromosome 17p deletion	<0.001
		Unmutated <i>IGHV</i>	0.038
		Treatment status	<0.001
	<i>NOTCH1</i> (<i>p.P2514fs</i>)	Trisomy 12	0.022
	<i>XPO1</i>	Unmutated <i>IGHV</i>	<0.001
		Increased white blood count (WBC)	<0.001
Untreated Cohort	<i>NOTCH1</i> (<i>p.P2514fs</i>)	Unmutated <i>IGHV</i>	0.037
	<i>SF3B1</i>	CD38 positivity	0.031
Treated Cohort	<i>XPO1</i>	ZAP70 positivity	0.024

* Statistical tests chosen for correlation analysis are described in material and methods 3.9 and are listed in Table S4

5. Discussion

CLL is the most frequent leukemia in adults of the Western world and shows a very heterogeneous clinical course. Therefore, the detection of recurrent mutations is promising for predictive diagnostics with respect to a future personalized clinical management of CLL patients. In the present study, we tested the feasibility of analyzing B-cell DNA from CLL patients for mutations of leukemia relevant genes by deep sequencing on a MiSeq NGS platform.

First, target enrichment was carried out by a commercially available multiplex PCR approach addressing leukemia involved genes, in general. All samples were successfully tested for 20 leukemia relevant genes covered by 1,272 target regions. The runs produced a good quality and a high output of data. Since the commercial multiplex PCR setup missed the genes which might be involved in CLL, but was designed for universal use of leukemia analysis, we designed a disease specific targeted genomic sequencing assay that was able to meet diagnostic and clinical research needs in order to get a detailed insight in CLL relevant gene alterations. Therefore, we established a customized protocol for library construction optimized for native B-cell DNA and multiplex PCR enriched targets, enabling a high throughput of samples in a time-effective manner.

Achieving high quality sequencing data with high specificity requires an exact knowledge about DNA quality and amount of DNA template molecules that are used for the target enrichment by multiplex PCR. Comparison of different DNA quantification methods, in particular fluorescence-based DNA quantification and “absolute real-time PCR” (qPCR) measurements, demonstrated the superiority of qPCR. Real-time PCR approaches were already described as the gold standard in the field of forensic as it is an accurate, reproducible, and time-effective technique to identify not only the total amount of DNA but the amplifiable molecules in a sample. Though it is still not standard for NGS sample processing, the determination of the exact number of DNA molecules used for the multiplex PCR enrichment of target genes is necessary and highly recommended in particular when DNA of low quality has to be used. Sah *et al* described extensively that an input of low accessible DNA copy number can produce an high number of false positive results in the sequencing data [79]. Therefore, we chose the *HFE* gene for determination of amplifiable DNA templates by qPCR. As expected, we found the

5. Discussion

B-cell DNA concentrations measured by fluorescence absorbance were higher than the DNA amount measured by qPCR. These differences can be explained by the fact that DNA damages won't lead to a change in concentration, but will reduce the amount of amplifiable DNA. Additionally, it has been shown that DNA concentrations measured by fluorometry can be impaired by proteins causing high background fluorescence [87]. Our quantitative results were concordant with the data reported by other authors who used fluorescence absorbance and qPCR methods for evaluation of DNA concentrations. This again underlines the importance to use the right method before starting with the amplicon generation, especially when using formalin-fixed paraffin embedded tissues [79, 84, 87].

Another critical step during the sample processing is the adapter ligation. The multiplex PCR approach leads to a massive amplification of target regions. We proved a correlation between adapter ligation efficiency and DNA input. In contrast to the first target enrichment PCR, which should be done with an adequate input of at least around 5,000 amplifiable DNA copies, adapter ligation needs limited amounts of DNA input because care has to be addressed that the reaction is not saturated. Tests of different amounts of DNA input in adapter ligation showed that 100% adapter ligation efficiency was achieved when samples were diluted by a factor of 10 after multiplex PCR, corresponding to approximately 30-50 ng DNA input. This step in the workflow is posed to be a bottleneck and our results indicate the need to evaluate the right ratio of input DNA and adapters.

Furthermore, the amount of DNA loaded on the sequencer is of great importance and exact quantification of adapter carrying amplicons is essential. If too much DNA is loaded, the clusters generated during bridge amplification will overlap and thereby affect the quality of the sequencing data [7, 62]. Currently, the used methods for library quantification like UV-spectrophotometry, or fluorometry are not exact enough as they also detect DNA fragments without adapters. If target amplicons lacking adapters are present in the DNA library pool, this can result in a lower cluster density than expected when the standard concentration is loaded [7]. Quantification of DNA libraries by qPCR accurately determines only fragments with adapters ligated to both ends of the target region, because only these fragments can be amplified. Thus, overestimation of the DNA concentration from fragments carrying no or only one adapter can be avoided.

We used a dilution series of PhiX library control with a well-defined phage genome containing approximately 45% GC (guanine, cytosine) and 55% AT (adenine, tyrosine) and 375 bp target length as a quantification standard. This approach ensures an accurate quantification of libraries with similar properties. It also generates reproducible and comparable results of independent libraries. A concern, however, can be complex differences between the standard DNA and the library sample. Libraries containing elements that are difficult to amplify, such as targets rich in guanine and cytosine bases, can lead to a negative bias [7]. This impaired quantification can result in an underestimation of the sample library concentration because targets with GC rich content may not be amplified efficiently.

Nevertheless, the present results showed a more homogenous distribution of reads per sample when the libraries were pooled according to the results of qPCR quantification.

Notably, the NGS approach, established in the present study, allows the simultaneous testing of 48 samples for 338 disease specific targets. However, the protocol for library construction consists of many time-consuming, error-prone pipetting and incubation steps. Furthermore, dealing with many samples containing thousands of amplified targets and different patient specific barcode adapters, the risk of contamination is high. Therefore, we improved the library preparation protocol by implementation of automated purification and size selection. A comparison of manually and automatically processed samples showed no differences in quality and quantity. We found the automated process to be very time-effective and helpful for sample preparation scale-up as it allows a simultaneous processing of up to 96 samples. Importantly, automated sample processing by robotic system defines the pipetting, ensures against the risk of mixing samples up and it allows an improved process control and comprehensive sample tracking.

In this study, I further used the here established, novel target-specific NGS approach for high-throughput mutation analysis of CLL DNA samples from 136 patients focusing on an informative gene set for CLL.

Although this assay targets a limited number of genes, for the first time I provide ultra-deep sequence data addressing nucleotide variants in a gene panel adapted to BCR signaling in CLL, including relevant drug targets and all coding exons of

5. Discussion

the considerably large *ATM* and *NOTCH1* genes. I was able to confirm various somatic mutations in previously described CLL associated driver genes such as *ATM*, *NOTCH1*, *TP53*, *SF3B1*, *XPO1*, *MYD88*, *DDX3X*, and discovered a novel mutation in *PTPN6*. A total of 36% patients revealed single variants, whereas 27% showed at least two variants in different genes. Although non-tumor DNA was not available for comparison, our stringent sequencing data analysis and filtering enabled us to confirm known somatic hotspot mutations in *SF3B1*, *XPO1* and *MYD88* with nearly the same frequency as described in previous studies [61, 63, 92, 94]. Since our diagnostic panel included complete coding regions of *ATM* and *NOTCH1*, we detected more variants than previously described in these genes. Our assay was performable in three days from the time of sample entry to data output and it achieved a significant higher sensitivity of mutation detection than Sanger sequencing, as I have demonstrated by the mutation analysis of DNA dilution series established by reference DNA derived from cell lines with defined mutation status.

Twenty-four functionally relevant *SF3B1* variants in 22 of 136 patients (16%) were identified. The splicing factor 3b, subunit 1 (*SF3B1*) is a component of the SF3B complex, which is associated with the U2 small nuclear ribonucleoprotein (snRNP) at the catalytic center of the spliceosome [24, 92, 94]. As part of the splicing machinery, *SF3B1* catalyzes the removal of introns from precursor messenger RNA (mRNA) to produce mature mRNA [35, 75]. Recently, *SF3B1* has been discovered as one of the most frequent mutated genes in CLL in 5% to 18% of patients, depending on the composition of the various cohorts [92].

In particular the p.K700E and p.G742D hotspot mutations are of high interest. They are localized within the highly conserved C-terminal PP2A-repeat regions and are supposed to lead to protein dysfunction expressed by an altered protein interaction and incorrect splicing of the target mRNAs [24, 32, 92]. In this context, it has been shown that CLL patients with *SF3B1* mutations and defective splicing activity have a high ratio of unspliced *BRD2*, *RIOK3* and *FOXP1* mRNA amongst other [24, 63, 94]. Additionally, recent data suggests that *SF3B1* mutated patients frequently experience a more aggressive course of disease, shorter TTT, shorter time to disease progression and lower 10-year OS rates [24, 32, 63, 94]. The data of my study show that *SF3B1* mutations occur almost exclusively in male patients,

as 93% of the investigated female patients were *SF3B1* wild type, confirming the findings of Jeromin *et al* [32].

In addition, I also identified the prominent p.L256P mutation in *MYD88*, which has recently been described in CLL patients and is part of the toll-like receptor (TLR) pathway, by the novel target-specific NGS approach. The *MYD88* gene encodes a cytosolic adapter protein that mediates the contact with the TLR after signaling activation [75]. More precisely, upon ligand binding and TLR stimulation, *MYD88* is recruited to the activated receptor complexes that trigger and activate the NF κ B pathway, which then induces inflammation, immune regulation, survival and proliferation [55, 75]. The detected alteration changing a leucine to proline is located in the well-conserved Toll/IL-1R homology (TIR) domain of the protein and presents a gain-of-function driver mutation that leads to selective advantage during tumor evolution [32, 55, 61, 63, 94]. Additionally, we identified in one CLL sample the p.V217F variant in the highly conserved B-B loop of the TIR domain, that was previously described in diffuse large B-cell lymphoma (DLBCL) by Ngo *et al* [55]. Both mutations are not linked to alterations in other tested genes and hence, they seem to represent epistatically linked genetic events in CLL progression. Interestingly, *MYD88* mutated CLL cells were shown to have a higher level of activated signal transducer and activator of transcription 3 (STAT3), I κ B α and NF κ B p65 subunit together with an altered DNA binding activity of NF κ B [24]. Together with the identification of *MYD88* mutations in other lymphomas, this data suggests that *MYD88* is a novel proto-oncogene and - from a therapeutic point of view -, the signaling complex coordinated by *MYD88* mutations represents an enticing target [24, 55].

Two genes in our study showed only one variant in a single patient. One of them was the p.T275P mutation in exon 9 of the RNA helicase DDX3X. This mutation was already described in CLL patients by Wang *et al*. It is located in the well-conserved core helicase domain of the gene [21, 94]. DDX3X is a member of the DEAD-box protein family that constitutes a subfamily of RNA helicases, which play an essential role in almost all aspects of RNA metabolism from transcription, splicing and decay, to translation.[21, 94] Remarkably, DDX3X interacts directly with XPO1, which we also identified to be mutated in 13 CLL samples. Exportin-1 (XPO1) is a member of the importin- β superfamily of transport receptors, and the functionally relevant p.E571X mutations could lead to an impaired recognition of

5. Discussion

nuclear export sequences (NES) [30, 37, 41]. Interestingly, variants were detected in all genes involved in the splicing process (*SF3B1*, *DDX3X*, *XPO1*), which underlines the importance of RNA processing in CLL.

Importantly, for the first time we describe a novel mutation in a pathogenic driver pathway for CLL. The p.V451M mutation in the catalytic phosphatase domain of SHP1 (*PTPN6*) was predicted to impair the protein function, leading to decreased activity and thereby to an increase of tyrosine-phosphorylated proteins in the CLL cells. Since SHP1 is a known repressor of BCR signaling, inactivating mutations are predicted to constitutively activate BCR signaling in CLL cells. SHP1 is primarily expressed in hematopoietic cells and is known to be a negative regulator for hematopoietic cell function [88, 98]. In lymphocytes, SHP1 binds to the immunoreceptor tyrosine-based inhibition motif (ITIM) of the inhibitor receptors, such as CD22, CD72 and KIR, through its SH2 domains. There, it dephosphorylates the downstream proteins and subsequently either terminates the activated BCR pathway or activates other pathways such as apoptosis [98]. Furthermore, dysfunctions of SHP1 were found to be a potential trigger for hematopoietic malignancies in mice and predicted to induce leukemia [98, 100]. Thus, our results suggest that disruption of SHP1 in CLL may be oncogenic by causing constitutive activation of BCR signaling.

NOTCH1 encodes a class one transmembrane protein functioning as a ligand-activated transcription factor and was one of the most altered genes in our study [17, 48, 75]. The protein undergoes several proteolytic cleavages that allow its intracellular part (NICD, notch intracellular domain) to translocate into the nucleus, thus leading to transcriptional activation of multiple target genes, including *MYC* and *CDKN1A* [17, 48, 75]. 98% of the *NOTCH1* variants were already sorted out during the first steps of the filtering algorithm because they were identified as synonymous variants or sequencing errors. Furthermore, only 33% of the remaining *NOTCH1* variants were predicted to affect the protein function. This high incidence of non-functional variants indicates that most of the variants might be a matter of technical problems of target enrichment and sequencing. In particular, reading errors of the Taq-polymerase might be a problem of the here used multiplex PCR approach. Furthermore, the high GC content and many homopolymeric regions of the gene – 17 of the 34 exons contain more than 65% GC – impair the sequencing results. Thus, due to the high

sequencing errors, it is difficult to identify putative driver mutations in this gene. Beside many variants in the EGF-like domains, we could detect a p.F1606L variant in the HD domain of the NOTCH1 extracellular part, which was predicted to have an impact on the protein function. Malecki *et al* reported that mutations in the HD domain can behave as true gain-of-function mutations, leading to a ligand independent stimulation of the NOTCH1 receptor [48]. I also found an eminent two base pair frameshift deletion p.P2514fs in the NICD that causes a truncated protein through the disruption of the C-terminal PEST (proline-glutamic acid-serine-threonine motif for degradation) domain by a premature stop codon. This mutation results in an accumulation of a more stable and activated protein [17, 24, 32, 61]. Thus, the nature of *NOTCH1* mutations identified by the here presented, novel NGS approach, results in a constitutive activation of NOTCH1 signaling. Since constitutive activation of NOTCH1 was shown to increase cell survival and apoptosis resistance in CLL cells [24, 32], the hereby detected *NOTCH1* alterations might be crucial for CLL progression. Interestingly, these *NOTCH1* alterations have already been described in CLL patients and were associated with unmutated *IGHV* genes and poor prognosis [94]. CLL patients with *NOTCH1* mutations were characterized by significant shorter TTT and OS, compared with their wild type counterpart, independent of other prognostic factors [17, 49, 57, 61, 76]. Although other studies showed significant correlations between *NOTCH1* mutation status and clinical parameter, we could not confirm these findings. However, a statistical analysis, focusing only on the p.P2514fs deletion in the PEST domain, revealed a significant correlation with unmutated *IGHV* (Table 20).

ATM is known to be frequently mutated in CLL. Although its precise role in the pathogenesis of CLL has not been fully understood, mutations seem to be associated with drug resistance [3, 59, 86]. In this study I show a so far unrecognized, exceptionally high mutation frequency of 31% (Figure 15C). Variants are spread over the entire *ATM* gene and 43% of these variants are associated with protein dysfunction (Table S3). Most of the 57% variants with no functional relevance had allele frequencies around 50% or 100%. This points to polymorphisms originated from the germline or by spontaneous alterations during tumor development.

One was located in the 33-amino acid FATC domain that is required for the kinase activity and plays a crucial role in interactions with the histone acetyltransferase

5. Discussion

KAT5 (*TIP60*) [34]. Jiang *et al* found KAT5 acetylating ATM after exposure to ionizing radiation, leading to the activation of ATM kinase activity [34]. They also found that a deletion of the FATC domain blocked the DNA damage-induced activation of ATM kinase activity and prevented ATM from regulating cellular radiosensitivity [34]. Interestingly, we found a p.R3047* point mutation in the ATM FATC domain resulting in an inhibition of KAT5-ATM interaction and potentially leading to a decreased radiation therapy response. Since this serine/threonine-protein kinase coordinates an integrated cellular response to DNA damage by double-stranded DNA breaks, it is also implicated in TP53 activation [3, 59]. Interestingly, we found *ATM* variants occurring more frequently in combination with other variants, particular either with *NOTCH1* or *TP53* (Figure 18) that seem to be mutually exclusive. We detected twelve patients harboring mutations in *TP53* and *ATM*. These finding agree with the results of Austen *et al* [3]. ATM acts upstream of TP53 in the apoptotic pathway, but also activates several TP53-independent cellular pathways, and it is conceivable that a mutation in both genes could increase the survival advantage of a malignant clone [3]. Murine studies demonstrated that ATM and TP53 can cooperate with respect to their function as tumor suppressors [3].

In contrast to *ATM* mutations that did not correlate with chromosomal deletions of the *ATM* gene locus, we found a highly significant correlation between *TP53* mutation status and deletions in its chromosomal region (17p13) ($P<0.001$) (Table 20). This clearly points to a two-hit mechanism, indicating that patients with a heterozygous loss of *TP53* seem to acquire an additional inactivating mutation of the second allele, resulting in a homozygous *TP53* inactivation leading to an impaired apoptosis and DNA repair mechanism and higher proliferation of the CLL cells. As expected, the statistical analysis revealed also a highly significant correlation between mutation status of *TP53* and treatment status of the patients ($P<0.001$). The untreated cohort showed 12% *TP53* mutations, whereas the treated cohort showed nearly three times more mutations. This data emphasize the rational of comprehensive biomarker analysis, especially because *TP53* inactivation by mutations is a known mechanism of therapy resistance, and an increase of mutation frequency in the refractory disease has been recently described [3, 47, 73]. Nearly 77% of the detected variants were located in the DNA binding domain, which is known to be a hotspot region for inactivating mutations

and therefore, were predicted to impair DNA repair mechanisms, cell cycle control and apoptosis [47, 59, 94].

In conclusion, we further demonstrated that targeted NGS is a viable high sensitive method, which has the potential to replace less comprehensive methods in the clinical setting, including Sanger sequencing for the detection of therapeutic and prognostic relevant DNA mutations. Importantly, the hereby presented data supports previously published findings that describe the potential prognostic impact of significant mutations in CLL. Furthermore, large genes like *ATM*, which has been described as frequently mutated in CLL negatively as affecting the progression-free survival and OS, are now amenable to routine diagnostics. Unlike currently used methods in the clinical molecular diagnostic, multiplex NGS panels offer great scalability and flexibility in the panel design and require less time. Panels can be customized and easily updated by adding new target regions or genes, which offers the opportunity to integrate new prognostic and diagnostic relevant markers very quickly. In comparison to whole exome or whole genome approaches, targeted NGS approaches as here presented benefit from the high gene coverage, needed for the diagnostic evaluation of the impact of the respective mutant in progress. Finally, we here show that NGS sample preparation can be automated and therefore, require minimal technician time for the setup, allowing a higher sample throughput compared to standard methods. Therefore, this approach has the potential to replace more time-intensive methods currently used to detect gene mutations in the clinical setting.

6. Summary

Chronic lymphocytic leukemia (CLL) is the most frequent leukemia in adults of the western world and shows a very heterogeneous clinical course. Whereas half of the patients can live for 20 years or more with no need for treatment, others show rapid progression, leading to substantial morbidity and mortality within a few months. Beside the well-known clinical markers, whole exome/genome studies identified mutations in CLL associated genes, which are potential new prognostic and predictive markers. Therefore, the detection of recurrent mutations is promising with respect to a future personalized clinical management of CLL patients. Aim of the present study was to establish a target-specific next generation sequencing approach. For this purpose, target enrichment focused on informative gene sets by library generation technologies was combined with ultra-deep and high-throughput parallel sequencing. First, a commercially available multiplex PCR approach covering leukemia relevant gene loci in general was used to set up target enrichment by PCR technology. Using dilutions of mutated cell lines, our method was able to reach sensitivities of approximately 5%. Next, in order to analyze CLL specific genes, for the first time a comprehensive CLL specific mutation hotspot and gene panel was designed. A total of 15 genes, known to be frequently mutated in CLL and/or involved in B-cell receptor signaling were selected for CLL-specific target enrichment and library preparation was optimized in terms of adapter ligation, process automation and sequencing quality. Using this novel ultra-deep sequencing approach, B-cell DNA samples from 136 CLL patients were sequenced on an Illumina MiSeq platform. By stringent filtering and a functional assessment algorithm, I identified 167 variants in 8 genes including hotspot regions of *ATM*, *MYD88*, *NOTCH1*, *SF3B1*, *TP53*, and *XPO1* as well as a new mutation in the phosphatase-domain of *PTPN6*, a negative regulator of BCR signaling. Furthermore, *TP53* mutations occurred more often in patients after therapy and most notably, in tumors with chromosomal p17 deletion, leading to complete *TP53* silencing.

In summary, targeted ultra-deep sequencing, allowing comprehensive, but highly cost-effective, sensitive and reliable detection of low frequency alleles in CLL, identified mutations with a marked clinical impact.

7. Zusammenfassung

Chronisch lymphatische Leukämie ist die häufigste Leukämieerkrankung unter Erwachsenen der westlichen Welt und zeigt einen sehr heterogenen klinischen Verlauf. Neben den in der klinischen Diagnostik fest etablierten prognostischen Markern, haben jüngste „whole exome/genome“ Analysen neue Mutationen in CLL-assoziierten Genen gefunden. Daher ist es für die prädiktive Diagnostik von besonderem Interesse diese neuen, teils Treibermutationen zu erfassen, um somit einen weiteren Schritt Richtung personalisierter Medizin zu ermöglichen.

Die hier vorliegende Arbeit beschreibt einen neuen Ansatz zur zielgerichteten „next generation“ Sequenzierung, der verschiedene Methoden zur Anreicherung von Zielregionen und zum Herstellen von Probenbibliotheken mit Hochdurchsatz- und Tiefensequenzierung verbindet. Dadurch wird es möglich eine Gruppe von Patienten zeitgleich auf ein umfassendes Spektrum an krankheitsrelevanten Genen und Mutations-Hotspots zu untersuchen.

Mit der Sequenzierung von Verdünnungsreihen mutierter Zelllinien konnte für die verwendete Methode eine Sensitivität von etwa 5% nachgewiesen werden.

Zum ersten Mal ist hier ein umfassendes CLL-spezifisches Panel zusammengestellt worden mit denen Mutationen, in den 15 am häufigsten in CLL mutierten Genen (u.a. aus dem B-Zell Rezeptor Signalweg), abdeckt werden. Die CLL-spezifischen Zielregionen wurden mittels Multiplex-PCR angereichert und 136 B-Zell DNA Proben von CLL Patienten anschließend mit der Illumina MiSeq Plattform sequenziert. Durch die Anwendung dieser Tiefensequenzierung, die stringente Auswertung und die Anwendung von Algorithmen zur Funktionsanalyse konnten 167 Varianten in 8 Genen, darunter Hotspots in *ATM*, *MYD88*, *NOTCH1*, *SF3B1*, *TP53*, und *XPO1*, sowie eine neue Mutation in der Phosphatase-Domäne von *PTPN6*, einem Negativregulator des B-Zell Rezeptor Signalwegs, gefunden werden. Des Weiteren konnte ein häufigeres Auftreten von *TP53* Mutationen in bereits therapierten Patienten, so wie in Tumoren mit einer Deletion des kurzen Arms von Chromosom 17, die einen kompletten Verlust von *TP53* zur Folge haben, beobachtet werden. Zusammenfassend zeigt die vorliegende Arbeit, dass die zielgerichtete Tiefensequenzierung eine umfassende, aber sehr kosten-effektive, sensitive und verlässliche Methode zum Nachweis von niedrig-frequenten Mutationen mit klinischer Relevanz darstellt.

8. References

- [1] Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR (2010). A method and server for predicting damaging missense mutations. *Nat Methods.* 7(4): 248-9.
- [2] Alsolami R, Knight SJ, Schuh A (2013). Clinical application of targeted and genome-wide technologies: can we predict treatment responses in chronic lymphocytic leukemia? *Per Med.* 10(4): 361-76.
- [3] Austen B, Powell JE, Alvi A, Edwards I, Hooper L, Starczynski J, Taylor AM, Fegan C, Moss P, Stankovic T (2005). Mutations in the ATM gene lead to impaired overall and treatment-free survival that is independent of IGVH mutation status in patients with B-CLL. *Blood.* 106(9): 3175-82.
- [4] Baliakas P, Hadzidimitriou A, Sutton LA, Rossi D, Minga E, Villamor N, Larrayoz M, Kminkova J, Agathangelidis A, Davis Z, Tausch E, Stalika E, Kantorova B, Mansouri L, Scarfo L, Cortese D, Navrkalova V, Rose-Zerilli MJ, 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 (2014). Recurrent mutations refine prognosis in chronic lymphocytic leukemia. *Leukemia:* 1-8.
- [5] Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, Hall KP, Evers DJ, Barnes CL, Bignell HR, Boutell JM, Bryant J, Carter RJ, Keira Cheetham R, Cox AJ, Ellis DJ, Flatbush MR, Gormley NA, Humphray SJ, Irving LJ, Karbelashvili MS, Kirk SM, Li H, Liu X, Maisinger KS, Murray LJ, Obradovic B, Ost T, Parkinson ML, Pratt MR, Rasolonjatovo IM, Reed MT, Rigatti R, Rodighiero C, Ross MT, Sabot A, Sankar SV, Scally A, Schroth GP, Smith ME, Smith VP, Spiridou A, Torrance PE, Tzonev SS, Vermaas EH, Walter K, Wu X, Zhang L, Alam MD, Anastasi C, Aniebo IC, Bailey DM, Bancarz IR, Banerjee S, Barbour SG, Baybayan PA, Benoit VA, Benson KF, Bevis C, Black PJ, Boodhun A, Brennan JS, Bridgham JA, Brown RC, Brown AA, Buermann DH, Bundu AA, Burrows JC, Carter NP, Castillo N, Chiara ECM, Chang S, Neil Cooley R, Crake NR, Dada OO, Diakoumako KD, Dominguez-Fernandez B, Earnshaw DJ, Egbujor UC, Elmore DW, Etchin SS, Ewan MR, Fedurco M, Fraser LJ, Fuentes Fajardo KV, Scott Furey W, George

- D, Gietzen KJ, Goddard CP, Golda GS, Granieri PA, Green DE, Gustafson DL, Hansen NF, Harnish K, Haudenschild CD, Heyer NI, Hims MM, Ho JT, Horgan AM, Hoschler K, Hurwitz S, Ivanov DV, Johnson MQ, James T, Huw Jones TA, Kang GD, Kerelska TH, Kersey AD, Khrebtukova I, Kindwall AP, Kingsbury Z, Kokko-Gonzales PI, Kumar A, Laurent MA, Lawley CT, Lee SE, Lee X, Liao AK, Loch JA, Lok M, Luo S, Mammen RM, Martin JW, McCauley PG, McNitt P, Mehta P, Moon KW, Mullens JW, Newington T, Ning Z, Ling Ng B, Novo SM, O'Neill MJ, Osborne MA, Osnowski A, Ostadan O, Paraschos LL, Pickering L, Pike AC, Chris Pinkard D, Pliskin DP, Podhasky J, Quijano VJ, Raczy C, Rae VH, Rawlings SR, Chiva Rodriguez A, Roe PM, Rogers J, Rogert Bacigalupo MC, Romanov N, Romieu A, Roth RK, Rourke NJ, Ruediger ST, Rusman E, Sanches-Kuiper RM, Schenker MR, Seoane JM, Shaw RJ, Shiver MK, Short SW, Sizto NL, Sluis JP, Smith MA, Ernest Sohma Sohma J, Spence EJ, Stevens K, Sutton N, Szajkowski L, Tregidgo CL, Turcatti G, Vandevondele S, Verhovsky Y, Virk SM, Wakelin S, Walcott GC, Wang J, Worsley GJ, Yan J, Yau L, Zuerlein M, Mullikin JC, Hurles ME, McCooke NJ, West JS, Oaks FL, Lundberg PL, Kleinerman D, Durbin R, Smith AJ (2008). Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*. 456(7218): 53-9.
- [6] Binet JL, Auquier A, Dighiero G, Chastang C, Piguet H, Goasguen J, Vaugier G, Potron G, Colona P, Oberling F, Thomas M, Tchernia G, Jacquillat C, Boivin P, Lesty C, Duault MT, Monconduit M, Belabbes S, Gremy F (1981). A new prognostic classification of chronic lymphocytic leukemia derived from a multivariate survival analysis. *Cancer*. 48(1): 198-206.
- [7] Buehler B, Hogrefe HH, Scott G, Ravi H, Pabon-Pena C, O'Brien S, Formosa R, Happe S (2010). Rapid quantification of DNA libraries for next-generation sequencing. *Methods*. 50(4): S15-8.
- [8] Burger JA (2012). Inhibiting B-cell receptor signaling pathways in chronic lymphocytic leukemia. *Curr Hematol Malig Rep*. 7(1): 26-33.
- [9] Chan AC, Iwashima M, Turck CW, Weiss A (1992). ZAP-70: a 70 kd protein-tyrosine kinase that associates with the TCR zeta chain. *Cell*. 71(4): 649-62.
- [10] Crespo M, Bosch F, Villamor N, Bellosillo B, Colomer D, Rozman M, Marce S, Lopez-Guillermo A, Campo E, Montserrat E (2003). ZAP-70 expression as a

8. References

- surrogate for immunoglobulin-variable-region mutations in chronic lymphocytic leukemia. *N Engl J Med.* 348(18): 1764-75.
- [11] Damle RN, Wasil T, Fais F, Ghiotto F, Valetto A, Allen SL, Buchbinder A, Budman D, Dittmar K, Kolitz J, Lichtman SM, Schulman P, Vinciguerra VP, Rai KR, Ferrarini M, Chiorazzi N (1999). Ig V gene mutation status and CD38 expression as novel prognostic indicators in chronic lymphocytic leukemia. *Blood.* 94(6): 1840-7.
- [12] Del Poeta G, Maurillo L, Venditti A, Buccisano F, Epiceno AM, Capelli G, Tamburini A, Suppo G, Battaglia A, Del Principe MI, Del Moro B, Masi M, Amadori S (2001). Clinical significance of CD38 expression in chronic lymphocytic leukemia. *Blood.* 98(9): 2633-9.
- [13] Döhner H, Stilgenbauer S, Benner A, Leupolt E, Kröber A, Bullinger L, Döhner K, Bentz M, Lichter P (2000). Genomic Aberrations and Survival in Chronic Lymphocytic Leukemia. *N Engl J Med.* 343(26): 1910-6.
- [14] Domenech E, Gomez-Lopez G, Glez-Pena D, Lopez M, Herreros B, Menezes J, Gomez-Lozano N, Carro A, Grana O, Pisano DG, Dominguez O, Garcia-Marco JA, Piris MA, Sanchez-Beato M (2012). New mutations in chronic lymphocytic leukemia identified by target enrichment and deep sequencing. *PLoS One.* 7(6): e38158.
- [15] Dressman D, Yan H, Traverso G, Kinzler KW, Vogelstein B (2003). Transforming single DNA molecules into fluorescent magnetic particles for detection and enumeration of genetic variations. *Proc Natl Acad Sci U S A.* 100(15): 8817-22.
- [16] Duncavage EJ, Abel HJ, Szankasi P, Kelley TW, Pfeifer JD (2012). Targeted next generation sequencing of clinically significant gene mutations and translocations in leukemia. *Mod Pathol.* 25(6): 795-804.
- [17] Fabbri G, Rasi S, Rossi D, Trifonov V, Khiabanian H, Ma J, Grunn A, Fangazio M, Capello D, Monti S, Cresta S, Gargiulo E, Forconi F, Guarini A, Arcaini L, Paulli M, Laurenti L, Larocca LM, Marasca R, Gattei V, Oscier D, Bertoni F, Mullighan CG, Foa R, Pasqualucci L, Rabidan R, Dalla-Favera R, Gaidano G (2011). Analysis of the chronic lymphocytic leukemia coding genome: role of NOTCH1 mutational activation. *J Exp Med.* 208(7): 1389-401.

- [18] Fedurco M, Romieu A, Williams S, Lawrence I, Turcatti G (2006). BTA, a novel reagent for DNA attachment on glass and efficient generation of solid-phase amplified DNA colonies. *Nucleic Acids Res.* 34(3): e22.
- [19] Forbes SA, Bhamra G, Bamford S, Dawson E, Kok C, Clements J, Menzies A, Teague JW, Futreal PA, Stratton MR (2008). The Catalogue of Somatic Mutations in Cancer (COSMIC). *Curr Protoc Hum Genet.* Chapter 10: Unit 10 1.
- [20] Furman RR, Sharman JP, Coutre SE, Cheson BD, Pagel JM, Hillmen P, Barrientos JC, Zelenetz AD, Kipps TJ, Flinn I, Ghia P, Eradat H, Ervin T, Lamanna N, Coiffier B, Pettitt AR, Ma S, Stilgenbauer S, Cramer P, Aiello M, Johnson DM, Miller LL, Li D, Jahn TM, Dansey RD, Hallek M, O'Brien SM (2014). Idelalisib and rituximab in relapsed chronic lymphocytic leukemia. *N Engl J Med.* 370(11): 997-1007.
- [21] Garbelli A, Beermann S, Di Cicco G, Dietrich U, Maga G (2011). A motif unique to the human DEAD-box protein DDX3 is important for nucleic acid binding, ATP hydrolysis, RNA/DNA unwinding and HIV-1 replication. *PLoS One.* 6(5): e19810.
- [22] Ghia P, Ferreri AM, Caligaris-Cappio F (2007). Chronic lymphocytic leukemia. *Crit Rev Oncol Hematol.* 64(3): 234-46.
- [23] Ghia P, Stamatopoulos K, Belessi C, Moreno C, Stilgenbauer S, Stevenson F, Davi F, Rosenquist R (2007). ERIC recommendations on IGHV gene mutational status analysis in chronic lymphocytic leukemia. *Leukemia.* 21(1): 1-3.
- [24] Gunnarsson R, Mansouri L, Rosenquist R (2013). Exploring the genetic landscape in chronic lymphocytic leukemia using high-resolution technologies. *Leuk Lymphoma.* 54(8): 1583-90.
- [25] Hallek M, Cheson BD, Catovsky D, Caligaris-Cappio F, Dighiero G, Dohner H, Hillmen P, Keating MJ,Montserrat E, Rai KR, Kipps TJ (1996). 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.
- [26] Hallek M, Cheson BD, Catovsky D, Caligaris-Cappio F, Dighiero G, Dohner H, Hillmen P, Keating MJ, Montserrat E, Rai KR, Kipps TJ (2008). Guidelines for

8. References

- 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(12): 5446-56.
- [27] Hallek M, Emmerich B (2009). Chronische Lymphatische Leukämie. 4 ed. UNIMED SCIENCE: UNI-MED Verlag AG.
- [28] Hallek M, Fischer K, Fingerle-Rowson G, Fink AM, Busch R, Mayer J, Hensel M, Hopfinger G, Hess G, von Grunhagen U, Bergmann M, Catalano J, Zinzani PL, Caligaris-Cappio F, Seymour JF, Berrebi A, Jager U, Cazin B, Trneny M, Westermann A, Wendtner CM, Eichhorst BF, Staib P, Buhler A, Winkler D, Zenz T, Bottcher S, Ritgen M, Mendila M, Kneba M, Dohner H, Stilgenbauer S (2010). Addition of rituximab to fludarabine and cyclophosphamide in patients with chronic lymphocytic leukaemia: a randomised, open-label, phase 3 trial. *Lancet*. 376(9747): 1164-74.
- [29] Hallek M, Pflug N (2011). State of the art treatment of chronic lymphocytic leukaemia. *Blood Rev*. 25(1): 1-9.
- [30] Hutten S, Kehlenbach RH (2007). CRM1-mediated nuclear export: to the pore and beyond. *Trends Cell Biol*. 17(4): 193-201.
- [31] Ibrahim S, Keating M, Do KA, O'Brien S, Huh YO, Jilani I, Lerner S, Kantarjian HM, Albitar M (2001). CD38 expression as an important prognostic factor in B-cell chronic lymphocytic leukemia. *Blood*. 98(1): 181-6.
- [32] Jeromin S, Weissmann S, Haferlach C, Dicker F, Bayer K, Grossmann V, Alpermann T, Roller A, Kohlmann A, Haferlach T, Kern W, Schnittger S (2014). SF3B1 mutations correlated to cytogenetics and mutations in NOTCH1, FBXW7, MYD88, XPO1 and TP53 in 1160 untreated CLL patients. *Leukemia*. 28(1): 108-17.
- [33] Jethwa A, Hullein J, Stolz T, Blume C, Sellner L, Jauch A, Sill M, Kater AP, te Raa GD, Geisler C, van Oers M, Dietrich S, Dreger P, Ho AD, Paruzynski A, Schmidt M, von Kalle C, Glimm H, Zenz T (2013). Targeted resequencing for analysis of clonal composition of recurrent gene mutations in chronic lymphocytic leukaemia. *Br J Haematol*. 163(4): 496-500.
- [34] Jiang X, Sun Y, Chen S, Roy K, Price BD (2006). The FATC domains of PIKK proteins are functionally equivalent and participate in the Tip60-dependent activation of DNA-PKcs and ATM. *J Biol Chem*. 281(23): 15741-6.

- [35] Kaida D, Motoyoshi H, Tashiro E, Nojima T, Hagiwara M, Ishigami K, Watanabe H, Kitahara T, Yoshida T, Nakajima H, Tani T, Horinouchi S, Yoshida M (2007). Spliceostatin A targets SF3b and inhibits both splicing and nuclear retention of pre-mRNA. *Nat Chem Biol.* 3(9): 576-83.
- [36] Kent WJ (2002). BLAT--the BLAST-like alignment tool. *Genome Res.* 12(4): 656-64.
- [37] Kudo N, Matsumori N, Taoka H, Fujiwara D, Schreiner EP, Wolff B, Yoshida M, Horinouchi S (1999). Leptomycin B inactivates CRM1/exportin 1 by covalent modification at a cysteine residue in the central conserved region. *Proc Natl Acad Sci U S A.* 96(16): 9112-7.
- [38] Kumar P, Henikoff S, Ng PC (2009). Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc.* 4(7): 1073-81.
- [39] Lai R, McDonnell TJ, O'Connor SL, Medeiros LJ, Oudat R, Keating M, Morgan MB, Curiel TJ, Ford RJ (2002). Establishment and characterization of a new mantle cell lymphoma cell line, Mino. *Leuk Res.* 26(9): 849-55.
- [40] Landau DA, Wu CJ (2013). Chronic lymphocytic leukemia: molecular heterogeneity revealed by high-throughput genomics. *Genome Med.* 5(5): 47.
- [41] Langer K, Dian C, Rybin V, Muller CW, Petosa C (2011). Insights into the function of the CRM1 cofactor RanBP3 from the structure of its Ran-binding domain. *PLoS One.* 6(2): e17011.
- [42] Langerak AW, Davi F, Ghia P, Hadzidimitriou A, Murray F, Potter KN, Rosenquist R, Stamatopoulos K, Belessi C (2011). Immunoglobulin sequence analysis and prognostication in CLL: guidelines from the ERIC review board for reliable interpretation of problematic cases. *Leukemia.* 25(6): 979-84.
- [43] Leamon JH, Lee WL, Tartaro KR, Lanza JR, Sarkis GJ, deWinter AD, Berka J, Weiner M, Rothberg JM, Lohman KL (2003). A massively parallel PicoTiterPlate based platform for discrete picoliter-scale polymerase chain reactions. *Electrophoresis.* 24(21): 3769-77.
- [44] Li H, Durbin R (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics.* 25(14): 1754-60.
- [45] Magliozi M, Piane M, Torrente I, Sinibaldi L, Rizzo G, Savio C, Lulli P, De Luca A, Dallapiccola B, Chessa L (2006). DHPLC screening of ATM gene in

8. References

- Italian patients affected by ataxia-telangiectasia: fourteen novel ATM mutations. *Dis Markers.* 22(4): 257-64.
- [46] Malavasi F, Deaglio S, Funaro A, Ferrero E, Horenstein AL, Ortolan E, Vaisitti T, Aydin S (2008). Evolution and function of the ADP ribosyl cyclase/CD38 gene family in physiology and pathology. *Physiol Rev.* 88(3): 841-86.
- [47] Malcikova J, Pavlova S, Kozubik KS, Pospisilova S (2014). TP53 mutation analysis in clinical practice: lessons from chronic lymphocytic leukemia. *Hum Mutat.* 35(6): 663-71.
- [48] Malecki MJ, Sanchez-Irizarry C, Mitchell JL, Histen G, Xu ML, Aster JC, Blacklow SC (2006). Leukemia-associated mutations within the NOTCH1 heterodimerization domain fall into at least two distinct mechanistic classes. *Mol Cell Biol.* 26(12): 4642-51.
- [49] Mansouri L, Cahill N, Gunnarsson R, Smedby KE, Tjonnfjord E, Hjalgrim H, Juliusson G, Geisler C, Rosenquist R (2013). NOTCH1 and SF3B1 mutations can be added to the hierarchical prognostic classification in chronic lymphocytic leukemia. *Leukemia.* 27(2): 512-4.
- [50] Mardis ER (2011). A decade's perspective on DNA sequencing technology. *Nature.* 470(7333): 198-203.
- [51] Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgesen S, Ho CH, Irzyk GP, Jando SC, Alenquer ML, Jarvie TP, Jirage KB, Kim JB, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H, Makijani VB, McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, Wang SH, Wang Y, Weiner MP, Yu P, Begley RF, Rothberg JM (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature.* 437(7057): 376-80.
- [52] Metzker ML (2010). Sequencing technologies - the next generation. *Nat Rev Genet.* 11(1): 31-46.
- [53] Mone AP, Cheney C, Banks AL, Tridandapani S, Mehter N, Guster S, Lin T, Eisenbeis CF, Young DC, Byrd JC (2006). Alemtuzumab induces caspase-independent cell death in human chronic lymphocytic leukemia cells through a lipid raft-dependent mechanism. *Leukemia.* 20(2): 272-9.

- [54] Nefedova Y, Gabrilovich D (2008). Mechanisms and clinical prospects of Notch inhibitors in the therapy of hematological malignancies. *Drug Resist Updat.* 11(6): 210-8.
- [55] Ngo VN, Young RM, Schmitz R, Jhavar S, Xiao W, Lim KH, Kohlhammer H, Xu W, Yang Y, Zhao H, Shaffer AL, Romesser P, Wright G, Powell J, Rosenwald A, Muller-Hermelink HK, Ott G, Gascoyne RD, Connors JM, Rimsza LM, Campo E, Jaffe ES, Delabie J, Smeland EB, Fisher RI, Brazil RM, Tubbs RR, Cook JR, Weisenburger DD, Chan WC, Staudt LM (2011). Oncogenically active MYD88 mutations in human lymphoma. *Nature.* 470(7332): 115-9.
- [56] O'Neil J, Grim J, Strack P, Rao S, Tibbitts D, Winter C, Hardwick J, Welcker M, Meijerink JP, Pieters R, Draetta G, Sears R, Clurman BE, Look AT (2007). FBW7 mutations in leukemic cells mediate NOTCH pathway activation and resistance to gamma-secretase inhibitors. *J Exp Med.* 204(8): 1813-24.
- [57] Oscier DG, Rose-Zerilli MJ, Winkelmann N, Gonzalez de Castro D, Gomez B, Forster J, Parker H, Parker A, Gardiner A, Collins A, Else M, Cross NC, Catovsky D, Strefford JC (2013). The clinical significance of NOTCH1 and SF3B1 mutations in the UK LRF CLL4 trial. *Blood.* 121(3): 468-75.
- [58] Peifer M, Fernandez-Cuesta L, Sos ML, George J, Seidel D, Kasper LH, Plenker D, Leenders F, Sun R, Zander T, Menon R, Koker M, Dahmen I, Muller C, Di Cerbo V, Schildhaus HU, Altmuller J, Baessmann I, Becker C, de Wilde B, Vandesompele J, Bohm D, Ansen S, Gabler F, Wilkening I, Heynck S, Heuckmann JM, Lu X, Carter SL, Cibulskis K, Banerji S, Getz G, Park KS, Rauh D, Grutter C, Fischer M, Pasqualucci L, Wright G, Wainer Z, Russell P, Petersen I, Chen Y, Stoelben E, Ludwig C, Schnabel P, Hoffmann H, Muley T, Brockmann M, Engel-Riedel W, Muscarella LA, Fazio VM, Groen H, Timens W, Sietsma H, Thunnissen E, Smit E, Heideman DA, Snijders PJ, Cappuzzo F, Ligorio C, Damiani S, Field J, Solberg S, Brustugun OT, Lund-Iversen M, Sanger J, Clement JH, Soltermann A, Moch H, Weder W, Solomon B, Soria JC, Validire P, Besse B, Brambilla E, Brambilla C, Lantuejoul S, Lorimier P, Schneider PM, Hallek M, Pao W, Meyerson M, Sage J, Shendure J, Schneider R, Buttner R, Wolf J, Nurnberg P, Perner S, Heukamp LC, Brindle PK, Haas S, Thomas RK (2012). Integrative genome analyses identify key somatic driver mutations of small-cell lung cancer. *Nat Genet.* 44(10): 1104-10.

8. References

- [59] Pettitt AR, Sherrington PD, Stewart G, Cawley JC, Taylor AM, Stankovic T (2001). p53 dysfunction in B-cell chronic lymphocytic leukemia: inactivation of ATM as an alternative to TP53 mutation. *Blood*. 98(3): 814-22.
- [60] Pospisilova S, Gonzalez D, Malcikova J, Trbusek M, Rossi D, Kater AP, Cymbalista F, Eichhorst B, Hallek M, Dohner H, Hillmen P, van Oers M, Gribben J, Ghia P, Montserrat E, Stilgenbauer S, Zenz T (2012). ERIC recommendations on TP53 mutation analysis in chronic lymphocytic leukemia. *Leukemia*. 26(7): 1458-61.
- [61] Puente XS, Pinyol M, Quesada V, Conde L, Ordonez GR, Villamor N, Escaramis G, Jares P, Bea S, Gonzalez-Diaz M, Bassaganyas L, Baumann T, Juan M, Lopez-Guerra M, Colomer D, Tubio JM, Lopez C, Navarro A, Tornador C, Aymerich M, Rozman M, Hernandez JM, Puente DA, Freije JM, Velasco G, Gutierrez-Fernandez A, Costa D, Carrio A, Guijarro S, Enjuanes A, Hernandez L, Yague J, Nicolas P, Romeo-Casabona CM, Himmelbauer H, Castillo E, Dohm JC, de Sanjose S, Piris MA, de Alava E, San Miguel J, Royo R, Gelpi JL, Torrents D, Orozco M, Pisano DG, Valencia A, Guigo R, Bayes M, Heath S, Gut M, Klatt P, Marshall J, Raine K, Stebbings LA, Futreal PA, Stratton MR, Campbell PJ, Gut I, Lopez-Guillermo A, Estivill X, Montserrat E, Lopez-Otin C, Campo E (2011). Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia. *Nature*. 475(7354): 101-5.
- [62] Quail MA, Kozarewa I, Smith F, Scally A, Stephens PJ, Durbin R, Swerdlow H, Turner DJ (2008). A large genome center's improvements to the Illumina sequencing system. *Nat Methods*. 5(12): 1005-10.
- [63] Quesada V, Conde L, Villamor N, Ordonez GR, Jares P, Bassaganyas L, Ramsay AJ, Bea S, Pinyol M, Martinez-Trillo A, Lopez-Guerra M, Colomer D, Navarro A, Baumann T, Aymerich M, Rozman M, Delgado J, Gine E, Hernandez JM, Gonzalez-Diaz M, Puente DA, Velasco G, Freije JM, Tubio JM, Royo R, Gelpi JL, Orozco M, Pisano DG, Zamora J, Vazquez M, Valencia A, Himmelbauer H, Bayes M, Heath S, Gut M, Gut I, Estivill X, Lopez-Guillermo A, Puente XS, Campo E, Lopez-Otin C (2012). Exome sequencing identifies recurrent mutations of the splicing factor SF3B1 gene in chronic lymphocytic leukemia. *Nat Genet*. 44(1): 47-52.

- [64] Rai KR, Sawitsky A, Cronkite EP, Chanana AD, Levy RN, Pasternack BS (1975). Clinical staging of chronic lymphocytic leukemia. *Blood*. 46(2): 219-34.
- [65] Reva B, Antipin Y, Sander C (2011). Predicting the functional impact of protein mutations: application to cancer genomics. *Nucleic Acids Res*. 39(17): e118.
- [66] Rice P, Longden I, Bleasby A (2000). EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet*. 16(6): 276-7.
- [67] Riches JC, Ramsay AG, Gribben JG (2011). Chronic lymphocytic leukemia: an update on biology and treatment. *Curr Oncol Rep*. 13(5): 379-85.
- [68] Robak T, Dmoszynska A, Solal-Celigny P, Warzocha K, Loscertales J, Catalano J, Afanasiev BV, Larratt L, Geisler CH, Montillo M, Zyuzgin I, Ganly PS, Dartigues C, Rosta A, Maurer J, Mendila M, Saville MW, Valente N, Wenger MK, Moiseev SI (2010). Rituximab plus fludarabine and cyclophosphamide prolongs progression-free survival compared with fludarabine and cyclophosphamide alone in previously treated chronic lymphocytic leukemia. *J Clin Oncol*. 28(10): 1756-65.
- [69] Robert-Koch-Institut, Krebs in Deutschland 2005/2006. Häufigkeiten und Trends, in Beiträge zur Gesundheitsberichterstattung des Bundes, e.V. RK-IudGdeKiD, Editor 2010: Berlin. p. 120.
- [70] Rosati E, Sabatini R, De Falco F, Del Papa B, Falzetti F, Di Ianni M, Cavalli L, Fettucciaro K, Bartoli A, Scarpanti I, Marconi P (2013). gamma-Secretase inhibitor I induces apoptosis in chronic lymphocytic leukemia cells by proteasome inhibition, endoplasmic reticulum stress increase and notch down-regulation. *Int J Cancer*. 132(8): 1940-53.
- [71] Rosenquist R, Cortese D, Bhoi S, Mansouri L, Gunnarsson R (2013). Prognostic markers and their clinical applicability in chronic lymphocytic leukemia: where do we stand? *Leuk Lymphoma*. 54(11): 2351-64.
- [72] Rosenwald A, Alizadeh AA, Widhopf G, Simon R, Davis RE, Yu X, Yang L, Pickeral OK, Rassenti LZ, Powell J, Botstein D, Byrd JC, Grever MR, Cheson BD, Chiorazzi N, Wilson WH, Kipps TJ, Brown PO, Staudt LM (2001). Relation of gene expression phenotype to immunoglobulin mutation genotype in B cell chronic lymphocytic leukemia. *J Exp Med*. 194(11): 1639-47.
- [73] Rosenwald A, Chuang EY, Davis RE, Wiestner A, Alizadeh AA, Arthur DC, Mitchell JB, Marti GE, Fowler DH, Wilson WH, Staudt LM (2004). Fludarabine

8. References

- treatment of patients with chronic lymphocytic leukemia induces a p53-dependent gene expression response. *Blood*. 104(5): 1428-34.
- [74] Rossi D, Bruscaggin A, Spina V, Rasi S, Khiabanian H, Messina M, Fangazio M, Vaisitti T, Monti S, Chiaretti S, Guarini A, Del Giudice I, Cerri M, Cresta S, Deambrogi C, Gargiulo E, Gattei V, Forconi F, Bertoni F, Deaglio S, Rabadan R, Pasqualucci L, Foa R, Dalla-Favera R, Gaidano G (2011). Mutations of the SF3B1 splicing factor in chronic lymphocytic leukemia: association with progression and fludarabine-refractoriness. *Blood*. 118(26): 6904-8.
- [75] Rossi D, Ciardullo C, Spina V, Gaidano G (2013). Molecular bases of chronic lymphocytic leukemia in light of new treatments. *Immunol Lett*. 155(1-2): 51-5.
- [76] Rossi D, Rasi S, Fabbri G, Spina V, Fangazio M, Forconi F, Marasca R, Laurenti L, Bruscaggin A, Cerri M, Monti S, Cresta S, Fama R, De Paoli L, Bulian P, Gattei V, Guarini A, Deaglio S, Capello D, Rabadan R, Pasqualucci L, Dalla-Favera R, Foa R, Gaidano G (2012). Mutations of NOTCH1 are an independent predictor of survival in chronic lymphocytic leukemia. *Blood*. 119(2): 521-9.
- [77] Rothberg JM, Hinz W, Rearick TM, Schultz J, Mileski W, Davey M, Leamon JH, Johnson K, Milgrew MJ, Edwards M, Hoon J, Simons JF, Marran D, Myers JW, Davidson JF, Branting A, Nobile JR, Puc BP, Light D, Clark TA, Huber M, Branciforte JT, Stoner IB, Cawley SE, Lyons M, Fu Y, Homer N, Sedova M, Miao X, Reed B, Sabina J, Feierstein E, Schorn M, Alanjary M, Dimalanta E, Dressman D, Kasinskas R, Sokolsky T, Fidanza JA, Namsaraev E, McKernan KJ, Williams A, Roth GT, Bustillo J (2011). An integrated semiconductor device enabling non-optical genome sequencing. *Nature*. 475(7356): 348-52.
- [78] Ruderman EM, Pope RM (2011). More than just B-cell inhibition. *Arthritis Res Ther*. 13(4): 125.
- [79] Sah S, Chen L, Houghton J, Kemppainen J, Marko AC, Zeigler R, Latham GJ (2013). Functional DNA quantification guides accurate next-generation sequencing mutation detection in formalin-fixed, paraffin-embedded tumor biopsies. *Genome Med*. 5(8): 77.
- [80] Schwarz JM, Rodelsperger C, Schuelke M, Seelow D (2010). MutationTaster evaluates disease-causing potential of sequence alterations. *Nat Methods*. 7(8): 575-6.

- [81] Schweighofer CD, Huh YO, Luthra R, Sargent RL, Ketterling RP, Knudson RA, Barron LL, Medeiros LJ, Keating MJ, Abruzzo LV (2011). The B cell antigen receptor in atypical chronic lymphocytic leukemia with t(14;19)(q32;q13) demonstrates remarkable stereotypy. *Int J Cancer.* 128(11): 2759-64.
- [82] Schweighofer CD, Wendtner CM (2010). First-line treatment of chronic lymphocytic leukemia: role of alemtuzumab. *Onco Targets Ther.* 3: 53-67.
- [83] Slatko BE, Kieleczawa J, Ju J, Gardner AF, Hendrickson CL, Ausubel FM (2011). "First generation" automated DNA sequencing technology. *Curr Protoc Mol Biol.* Chapter 7: Unit7 2.
- [84] Sozzi G, Conte D, Mariani L, Lo Vullo S, Roz L, Lombardo C, Pierotti MA, Tavecchio L (2001). Analysis of circulating tumor DNA in plasma at diagnosis and during follow-up of lung cancer patients. *Cancer Res.* 61(12): 4675-8.
- [85] Stanglmaier M, Reis S, Hallek M (2004). Rituximab and alemtuzumab induce a nonclassic, caspase-independent apoptotic pathway in B-lymphoid cell lines and in chronic lymphocytic leukemia cells. *Ann Hematol.* 83(10): 634-45.
- [86] Stankovic T, Stewart GS, Fegan C, Biggs P, Last J, Byrd PJ, Keenan RD, Moss PA, Taylor AM (2002). Ataxia telangiectasia mutated-deficient B-cell chronic lymphocytic leukemia occurs in pregerminal center cells and results in defective damage response and unrepaired chromosome damage. *Blood.* 99(1): 300-9.
- [87] Szpechcinski A, Struniawska R, Zaleska J, Chabowski M, Orlowski T, Roszkowski K, Chorostowska-Wynimko J (2008). Evaluation of fluorescence-based methods for total vs. amplifiable DNA quantification in plasma of lung cancer patients. *J Physiol Pharmacol.* 59 Suppl 6: 675-81.
- [88] Tibaldi E, Brunati AM, Zonta F, Frezzato F, Gattazzo C, Zambello R, Gringeri E, Semenzato G, Pagano MA, Trentin L (2011). Lyn-mediated SHP-1 recruitment to CD5 contributes to resistance to apoptosis of B-cell chronic lymphocytic leukemia cells. *Leukemia.* 25(11): 1768-81.
- [89] Tol J, Dijkstra JR, Vink-Borger ME, Nagtegaal ID, Punt CJ, Van Krieken JH, Ligtenberg MJ (2010). High sensitivity of both sequencing and real-time PCR analysis of KRAS mutations in colorectal cancer tissue. *J Cell Mol Med.* 14(8): 2122-31.

8. References

- [90] Tsiantis AC, Norris-Kirby A, Rich RG, Hafez MJ, Gocke CD, Eshleman JR, Murphy KM (2010). Comparison of Sanger sequencing, pyrosequencing, and melting curve analysis for the detection of KRAS mutations: diagnostic and clinical implications. *J Mol Diagn.* 12(4): 425-32.
- [91] UniProtConsortium (2014). Activities at the Universal Protein Resource (UniProt). *Nucleic Acids Res.* 42(Database issue): D191-8.
- [92] Wan Y, Wu CJ (2013). SF3B1 mutations in chronic lymphocytic leukemia. *Blood.* 121(23): 4627-34.
- [93] Wang K, Li M, Hakonarson H (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* 38(16): e164.
- [94] Wang L, Lawrence MS, Wan Y, Stojanov P, Sougnez C, Stevenson K, Werner L, Sivachenko A, DeLuca DS, Zhang L, Zhang W, Vartanov AR, Fernandes SM, Goldstein NR, Folco EG, Cibulskis K, Tesar B, Sievers QL, Shefler E, Gabriel S, Hacohen N, Reed R, Meyerson M, Golub TR, Lander ES, Neuberg D, Brown JR, Getz G, Wu CJ (2011). SF3B1 and other novel cancer genes in chronic lymphocytic leukemia. *N Engl J Med.* 365(26): 2497-506.
- [95] Wiestner A, Rosenwald A, Barry TS, Wright G, Davis RE, Henrickson SE, Zhao H, Ibbotson RE, Orchard JA, Davis Z, Stetler-Stevenson M, Raffeld M, Arthur DC, Marti GE, Wilson WH, Hamblin TJ, Oscier DG, Staudt LM (2003). ZAP-70 expression identifies a chronic lymphocytic leukemia subtype with unmutated immunoglobulin genes, inferior clinical outcome, and distinct gene expression profile. *Blood.* 101(12): 4944-51.
- [96] Woyach JA, Johnson AJ, Byrd JC (2012). The B-cell receptor signaling pathway as a therapeutic target in CLL. *Blood.* 120(6): 1175-84.
- [97] Wu C, Orozco C, Boyer J, Leglise M, Goodale J, Batalov S, Hodge CL, Haase J, Janes J, Huss JW, 3rd, Su AI (2009). BioGPS: an extensible and customizable portal for querying and organizing gene annotation resources. *Genome Biol.* 10(11): R130.
- [98] Wu C, Sun M, Liu L, Zhou GW (2003). The function of the protein tyrosine phosphatase SHP-1 in cancer. *Gene.* 306: 1-12.
- [99] Zenz T, Mertens D, Dohner H, Stilgenbauer S (2011). Importance of genetics in chronic lymphocytic leukemia. *Blood Rev.* 25(3): 131-7.

- [100] Zhang J, Somani AK, Siminovitch KA (2000). Roles of the SHP-1 tyrosine phosphatase in the negative regulation of cell signalling. *Semin Immunol.* 12(4): 361-78.

9. Preliminary Publications

Research Articles

- Claudia Vollbrecht, Fabian Dominik Mairinger, Carmen-Diana Schweighofer, Lukas Carl Heukamp, Sabine Merkelbach-Bruse, Reinhard Buettner, Margarete Odenthal: *Targeting Cancer Related Genes by Multiplex PCR Followed by High Throughput Parallel Sequencing*. International Journal of Genomic Medicine. 03/2014; 2(1).
- Claudia Vollbrecht, Fabian D. Mairinger, Ulrike Koitzsch, Martin Peifer, Katharina Koenig, Lukas C. Heukamp, Giuliano Crispazza, Laura Wilden, Karl-Anon Kreuzer, Michael Hallek, Margarete Odenthal, Carmen-Diana Herling, Reinhard Buettner. Comprehensive analysis of informative genes in chronic lymphocytic leukemia by multiplex PCR and deep sequencing. PLoS One submitted in December 2014.

Technical Reports

- Kerstin Becker, Claudia Vollbrecht, Ulrike Koitzsch, Katharina Koenig, Jana Fassunke, Sebastian Huss, Peter Nuernberg, Lukas C Heukamp, Reinhard Buettner, Margarete Odenthal, Janine Altmueller, Sabine Merkelbach-Bruse: *Deep ion sequencing of amplicon adapter ligated libraries: a novel tool in molecular diagnostics of formalin fixed and paraffin embedded tissues*. Journal of clinical pathology 04/2013. Equally contributed first author

Reviews

- Claudia Vollbrecht, Reinhard Büttner: *[Prädiktive und prognostische Biomarker in der Krebstherapie: Was ist heute technisch möglich?]*. Onkologie heute 01/2014 1:9-12.
- Claudia Vollbrecht, Katharina König, Lukas Heukamp, Reinhard Büttner, Margarete Odenthal: *[Molecular pathology of the lungs: New perspectives by next generation sequencing]*. Der Pathologe 02/2013; 34(1):16-24.

10. Appendix

10.1 Target Regions and Primer Sequences of CLL Panel 1 and 2

Table S1: Target region of the CLL panel 1 and 2; Chr. chromosome

Chr.	Start	Stop	AmplonID_Gene_Exon
Panel 1			
chr1	9784025	9784137	AMPL790398636_PIK3CD_Exon21#
chr1	9784136	9784275	AMPL790401743_PIK3CD_Exon21#
chr1	9784294	9784427	AMPL789461861_PIK3CD_Exon22#
chr1	9784427	9784605	AMPL789494496_PIK3CD_Exon22#
chr1	9784840	9785017	AMPL789532005_PIK3CD_Exon23#
chr1	9786876	9787019	AMPL1023269090_PIK3CD_Exon24#
chr1	9787019	9787160	AMPL965242047_PIK3CD_Exon24#
chr2	61719324	61719485	AMPL604752115_XPO1_Exon15#
chr2	61719483	61719592	AMPL586208437_XPO1_Exon15#
chr2	61719586	61719726	AMPL631468458_XPO1_Exon15#
chr2	61719927	61720095	AMPL867845130_XPO1_Exon13#
chr2	61720123	61720278	AMPL631330509_XPO1_Exon13#
chr2	61720986	61721129	AMPL586209492_XPO1_Exon12#
chr2	61721126	61721268	AMPL586224083_XPO1_Exon12#
chr2	198266439	198266537	AMPL529499302_SF3B1_Exon16#
chr2	198266522	198266691	AMPL529502820_SF3B1_Exon16#
chr2	198266719	198266860	AMPL529499628_SF3B1_Exon15#
chr2	198267149	198267321	AMPL540707664_SF3B1_Exon14#
chr2	198267321	198267421	AMPL529510076_SF3B1_Exon14#
chr2	198267421	198267556	AMPL529523970_SF3B1_Exon14#
chr3	38182190	38182374	AMPL461708423_MYD88_Exon4#
chr3	38182550	38182651	AMPL461320681_MYD88_Exon5#
chr3	38182651	38182774	AMPL461335271_MYD88_Exon5#
chr3	178928117	178928286	AMPL391663014_PIK3CA_Exon9#
chr3	178928286	178928456	AMPL391678526_PIK3CA_Exon9#
chr3	178935835	178936001	AMPL431898893_PIK3CA_Exon10#
chr3	178936022	178936106	AMPL393459515_PIK3CA_Exon10#
chr3	178936092	178936180	AMPL766334235_PIK3CA_Exon10#
chr3	178936892	178937043	AMPL439014560_PIK3CA_Exon11#
chr3	178937043	178937114	AMPL391540460_PIK3CA_Exon11#
chr3	178947889	178948058	AMPL929426275_PIK3CA_Exon20#
chr3	178948056	178948184	AMPL393716297_PIK3CA_Exon20#
chr3	178952028	178952205	AMPL392251341_PIK3CA_Exon21#
chr4	153249276	153249418	AMPL749428714_FBXW7_Exon9#
chr4	153249418	153249550	AMPL411831344_FBXW7_Exon9#
chr4	153250711	153250857	AMPL731873228_FBXW7_Exon8#
chr4	153250857	153250992	AMPL410695952_FBXW7_Exon8#
chr4	153251860	153252032	AMPL687137193_FBXW7_Exon7#
chr4	153253709	153253847	AMPL408308387_FBXW7_Exon6#
chr4	153253847	153254000	AMPL422723675_FBXW7_Exon6#

10. Appendix

Chr.	Start	Stop	AmpliconID_Gene_Exon
chr9	139390107	139390280	AMPL389071515_NOTCH1_Exon34#
chr9	139390278	139390410	AMPL1023279207_NOTCH1_Exon34#
chr10	89692658	89692775	AMPL703872780_PTEN_Exon5#
chr10	89692775	89692952	AMPL703903618_PTEN_Exon5#
chr10	89711806	89711979	AMPL968354751_PTEN_Exon6#
chr10	89711979	89712106	AMPL623452574_PTEN_Exon6#
chr10	89724917	89725058	AMPL703848673_PTEN_Exon9#
chr10	89725058	89725147	AMPL391543592_PTEN_Exon9#
chr10	89725174	89725341	AMPL391543631_PTEN_Exon9#
chr11	108173533	108173698	AMPL637661621_ATM_Exon36#
chr11	108173698	108173792	AMPL851616633_ATM_Exon36#
chr12	7067072	7067247	AMPL789687311_PTPN6_Exon11#
chr12	7069032	7069212	AMPL790061137_PTPN6_Exon12#
chr17	7576995	7577174	AMPL705603720_TP53_Exon8#
chr17	7577356	7577509	AMPL388733576_TP53_Exon7#
chr17	7577508	7577611	AMPL387805586_TP53_Exon7#
chr17	7578150	7578333	AMPL387845397_TP53_Exon6#
chr17	7578249	7578425	AMPL387854314_TP53_Exon5#
chr17	7578425	7578555	AMPL707393388_TP53_Exon5#
chr17	62006055	62006236	AMPL755777037_CD79B_Exon6#
chr17	62006765	62006930	AMPL551731808_CD79B_Exon5#
chr22	22127125	22127297	AMPL527206685_MAPK1_Exon7#
chrX	41202453	41202597	AMPL801574439_DDX3X_Exon7#
chrX	41202947	41203112	AMPL801449024_DDX3X_Exon8#
chrX	41203250	41203371	AMPL801524167_DDX3X_Exon9#
chrX	41204387	41204468	AMPL801406997_DDX3X_Exon11#
chrX	41204468	41204644	AMPL801418667_DDX3X_Exon11#
chrX	41205740	41205913	AMPL801337081_DDX3X_Exon14#
chrX	100609563	100609737	AMPL551294195_BTK_Exon16#
chrX	100610912	100611086	AMPL552372528_BTK_Exon15#
chrX	100611086	100611256	AMPL552429753_BTK_Exon15#
chrX	100611641	100611795	AMPL611084046_BTK_Exon14#
chrX	100611795	100611946	AMPL551583129_BTK_Exon14#

Panel 2

chr2	198256985	198257139	AMPL529513592_SF3B1_Exon25#
chr2	198257139	198257306	AMPL562020001_SF3B1_Exon25#
chr2	198257715	198257825	AMPL3454705572_SF3B1_Exon24#
chr2	198257825	198257984	AMPL3503992962_SF3B1_Exon24#
chr2	198260663	198260809	AMPL576642473_SF3B1_Exon23#
chr2	198260809	198260936	AMPL529499472_SF3B1_Exon23#
chr2	198260936	198261057	AMPL529510434_SF3B1_Exon23#
chr2	198262638	198262809	AMPL529510759_SF3B1_Exon22#
chr2	198262809	198262964	AMPL587295092_SF3B1_Exon22#
chr2	198263136	198263280	AMPL3673302042_SF3B1_Exon21#
chr2	198263278	198263399	AMPL576333997_SF3B1_Exon21#
chr2	198264644	198264775	AMPL604500653_SF3B1_Intron20#
chr2	198264775	198264937	AMPL587280519_SF3B1_Exon20#
chr2	198264832	198265012	AMPL604524655_SF3B1_Exon20#_Exon19#
chr2	198265012	198265184	AMPL2758507014_SF3B1_Exon19#

Chr.	Start	Stop	AmpliconID_Gene_Exon
chr2	198265322	198265491	AMPL576642062_SF3B1_Exon18#
chr2	198265466	198265565	AMPL529512492_SF3B1_Exon18#
chr2	198265565	198265681	AMPL529515581_SF3B1_Exon18#
chr2	198266092	198266205	AMPL555912744_SF3B1_Exon17#
chr2	198266178	198266278	AMPL529500194_SF3B1_Exon17#
chr2	198266443	198266526	AMPL1413798090_SF3B1_Exon16#
chr2	198266522	198266691	AMPL529502820_SF3B1_Exon16#
chr2	198266575	198266719	AMPL529503894_SF3B1_Exon16#_Exon15#
chr2	198267149	198267321	AMPL540707664_SF3B1_Exon14#
chr2	198267321	198267421	AMPL529510076_SF3B1_Exon14#
chr2	198267421	198267556	AMPL529523970_SF3B1_Exon14#
chr2	198267590	198267708	AMPL723505021_SF3B1_Exon13#
chr2	198267708	198267862	AMPL556015216_SF3B1_Exon13#
chr2	198268218	198268391	AMPL529521989_SF3B1_Exon12#
chr2	198268391	198268493	AMPL529531715_SF3B1_Exon12#
chr2	198269791	198269860	AMPL1407830555_SF3B1_Exon11#
chr2	198269822	198269979	AMPL529515562_SF3B1_Exon11#
chr2	198270069	198270162	AMPL529500017_SF3B1_Exon10#
chr2	198270162	198270303	AMPL576519838_SF3B1_Exon10#
chr2	198272616	198272785	AMPL3628433904_SF3B1_Exon9#
chr2	198272785	198272954	AMPL3504129708_SF3B1_Exon9#
chr2	198273036	198273138	AMPL1652590115_SF3B1_Exon8#
chr2	198273133	198273239	AMPL3673265777_SF3B1_Exon8#
chr2	198273238	198273312	AMPL3673283970_SF3B1_Exon8#
chr2	198274486	198274636	AMPL529516469_SF3B1_Exon7#
chr2	198274634	198274752	AMPL529527985_SF3B1_Exon7#
chr2	198281477	198281607	AMPL529514659_SF3B1_Exon6#
chr2	198281607	198281679	AMPL529525116_SF3B1_Exon6#
chr2	198283221	198283357	AMPL529502653_SF3B1_Exon5#
chr2	198283577	198283740	AMPL529501161_SF3B1_Exon4#
chr2	198285008	198285164	AMPL723964157_SF3B1_Exon4#
chr2	198285164	198285332	AMPL529528213_SF3B1_Exon4#
chr2	198285709	198285814	AMPL529499999_SF3B1_Exon3#
chr2	198285814	198285948	AMPL529503419_SF3B1_Exon3#
chr2	198288382	198288555	AMPL723166508_SF3B1_Exon2#
chr2	198288555	198288728	AMPL529524655_SF3B1_Exon2#
chr2	198299613	198299790	AMPL529521539_SF3B1_Exon1#
chr3	38180094	38180255	AMPL460903538_MYD88_Exon1#
chr3	38180243	38180382	AMPL460906876_MYD88_Exon1#
chr3	38180382	38180555	AMPL460925384_MYD88_Exon1#
chr3	38181339	38181514	AMPL461139916_MYD88_Exon2#
chr3	38181785	38181959	AMPL1186414652_MYD88_Exon3#
chr3	38181959	38182137	AMPL1184868952_MYD88_Exon3#
chr3	38182190	38182374	AMPL461708423_MYD88_Exon4#
chr3	38182550	38182651	AMPL461320681_MYD88_Exon5#
chr3	38182651	38182805	AMPL461335314_MYD88_Exon5#
chr9	139390444	139390610	AMPL393483900_NOTCH1_Exon34#
chr9	139390609	139390742	AMPL393490709_NOTCH1_Exon34#
chr9	139390724	139390880	AMPL1733131768_NOTCH1_Exon34#

10. Appendix

Chr.	Start	Stop	AmpliconID_Gene_Exon
chr9	139390868	139391002	AMPL687198331_NOTCH1_Exon34#
chr9	139390991	139391125	AMPL393548366_NOTCH1_Exon34#
chr9	139391115	139391275	AMPL393560912_NOTCH1_Exon34#
chr9	139391275	139391446	AMPL393565691_NOTCH1_Exon34#
chr9	139391397	139391575	AMPL393570760_NOTCH1_Exon34#
chr9	139391574	139391740	AMPL393579722_NOTCH1_Exon34#
chr9	139391734	139391922	AMPL393587649_NOTCH1_Exon34#
chr9	139391919	139392066	AMPL393593091_NOTCH1_Exon34#
chr9	139393302	139393490	AMPL1133733423_NOTCH1_Exon33#
chr9	139393467	139393591	AMPL395255812_NOTCH1_Exon32#
chr9	139393575	139393746	AMPL393462275_NOTCH1_Exon32#
chr9	139394896	139395014	AMPL393550305_NOTCH1_Exon31#
chr9	139395014	139395193	AMPL393552977_NOTCH1_Exon31#
chr9	139395176	139395311	AMPL393558073_NOTCH1_Exon31#
chr9	139396066	139396226	AMPL395395395_NOTCH1_Exon30#
chr9	139396224	139396370	AMPL393876564_NOTCH1_Exon30#
chr9	139396395	139396568	AMPL393468476_NOTCH1_Exon29#
chr9	139396665	139396782	AMPL394076136_NOTCH1_Exon28#
chr9	139396782	139396963	AMPL389056250_NOTCH1_Exon28#
chr9	139397611	139397796	AMPL1406509564_NOTCH1_Exon27#
chr9	139399124	139399292	AMPL3659581187_NOTCH1_Exon26#
chr9	139399289	139399458	AMPL3659582003_NOTCH1_Exon26#
chr9	139399455	139399547	AMPL389057328_NOTCH1_Exon26#
chr9	139399546	139399674	AMPL576385984_NOTCH1_Exon26#
chr9	139399726	139399884	AMPL393813581_NOTCH1_Exon25#
chr9	139399884	139400065	AMPL3317928598_NOTCH1_Exon25#
chr9	139400065	139400193	AMPL1489840519_NOTCH1_Exon25#
chr9	139400159	139400339	AMPL393827630_NOTCH1_Exon25#
chr9	139400930	139401112	AMPL393523557_NOTCH1_Exon24#
chr9	139401115	139401274	AMPL393762385_NOTCH1_Exon23#
chr9	139401270	139401438	AMPL393768721_NOTCH1_Exon23#
chr9	139401722	139401904	AMPL974297171_NOTCH1_Exon22#
chr9	139402300	139402478	AMPL395582111_NOTCH1_Exon21#
chr9	139402478	139402631	AMPL3659630892_NOTCH1_Exon21#
chr9	139402655	139402843	AMPL393826115_NOTCH1_Exon20#
chr9	139403260	139403426	AMPL393847611_NOTCH1_Exon19#
chr9	139403426	139403608	AMPL393851212_NOTCH1_Exon19#
chr9	139404133	139404315	AMPL393835150_NOTCH1_Exon18#
chr9	139404315	139404423	AMPL393843865_NOTCH1_Exon18#
chr9	139405071	139405234	AMPL393483551_NOTCH1_Exon17#
chr9	139405234	139405392	AMPL3318173741_NOTCH1_Exon17#
chr9	139405558	139405647	AMPL3659461802_NOTCH1_Exon16#
chr9	139405645	139405763	AMPL393716240_NOTCH1_Exon16#
chr9	139407433	139407620	AMPL393674542_NOTCH1_Exon15#
chr9	139407826	139408005	AMPL393475135_NOTCH1_Exon14#
chr9	139408878	139409020	AMPL393872773_NOTCH1_Exon13#
chr9	139409019	139409203	AMPL393887068_NOTCH1_Exon13#
chr9	139409696	139409872	AMPL2686674775_NOTCH1_Exon12#
chr9	139409884	139410024	AMPL394049893_NOTCH1_Exon11#

Chr.	Start	Stop	AmpliconID_Gene_Exon
chr9	139410024	139410192	AMPL394052386_NOTCH1_Exon11#
chr9	139410300	139410474	AMPL1838015592_NOTCH1_Exon10#
chr9	139410471	139410589	AMPL393840759_NOTCH1_Exon10#
chr9	139411745	139411898	AMPL395901075_NOTCH1_Exon9#
chr9	139412137	139412280	AMPL393459313_NOTCH1_Exon8#
chr9	139412275	139412438	AMPL393467389_NOTCH1_Exon8#
chr9	139412495	139412680	AMPL3299434711_NOTCH1_Exon7#
chr9	139412679	139412864	AMPL396036289_NOTCH1_Exon7#
chr9	139412951	139413132	AMPL393577583_NOTCH1_Exon6#
chr9	139413132	139413282	AMPL3386297428_NOTCH1_Exon6#
chr9	139413836	139413982	AMPL1406505740_NOTCH1_Exon5#
chr9	139413981	139414083	AMPL1406505750_NOTCH1_Exon5#
chr9	139417188	139417356	AMPL3299973779_NOTCH1_Exon4#
chr9	139417311	139417501	AMPL393534998_NOTCH1_Exon4#
chr9	139417501	139417645	AMPL393537721_NOTCH1_Exon4#
chr9	139418097	139418203	AMPL393677786_NOTCH1_Exon3#
chr9	139418201	139418337	AMPL393683751_NOTCH1_Exon3#
chr9	139418335	139418441	AMPL393687962_NOTCH1_Exon3#
chr9	139438446	139438603	AMPL393839516_NOTCH1_Exon2#
chr11	108098340	108098500	AMPL392261218_ATM_Exon2#
chr11	108098395	108098552	AMPL391997777_ATM_Exon2#_Exon#3
chr11	108098552	108098686	AMPL623973469_ATM_Exon3#
chr11	108099804	108099959	AMPL391948634_ATM_Exon4#
chr11	108099959	108100072	AMPL391991577_ATM_Exon4#
chr11	108106358	108106525	AMPL391522368_ATM_Exon5#
chr11	108106505	108106603	AMPL391536984_ATM_Exon5#
chr11	108114587	108114757	AMPL447455904_ATM_Exon6#
chr11	108114728	108114818	AMPL3660815253_ATM_Exon6#
chr11	108114808	108114959	AMPL409275345_ATM_Exon6#
chr11	108115454	108115549	AMPL391579143_ATM_Exon7#
chr11	108115549	108115700	AMPL391589757_ATM_Exon7#
chr11	108115658	108115784	AMPL411498022_ATM_Exon7#
chr11	108117627	108117790	AMPL1482570325_ATM_Exon8#
chr11	108117717	108117865	AMPL699213618_ATM_Exon8#
chr11	108119695	108119765	AMPL391558752_ATM_Exon9#
chr11	108119765	108119883	AMPL624222416_ATM_Exon9#
chr11	108121359	108121498	AMPL623995564_ATM_Exon10#
chr11	108121498	108121589	AMPL392330115_ATM_Exon10#
chr11	108121589	108121758	AMPL3660799045_ATM_Exon10#
chr11	108121758	108121911	AMPL3431977515_ATM_Exon10#
chr11	108122552	108122622	AMPL1549898614_ATM_Exon11#
chr11	108122622	108122789	AMPL1482037524_ATM_Exon11#
chr11	108123527	108123694	AMPL392071108_ATM_Exon12#
chr11	108124511	108124639	AMPL391670871_ATM_Exon13#
chr11	108124639	108124805	AMPL687292966_ATM_Exon13#
chr11	108126829	108126991	AMPL391466017_ATM_Exon14#
chr11	108126989	108127160	AMPL437230929_ATM_Exon14#
chr11	108128145	108128253	AMPL1186328350_ATM_Exon15#
chr11	108128247	108128318	AMPL391971013_ATM_Exon15#

10. Appendix

Chr.	Start	Stop	AmpliconID_Gene_Exon
chr11	108129604	108129757	AMPL391704487_ATM_Exon16#
chr11	108129757	108129913	AMPL391707792_ATM_Exon16#
chr11	108137856	108137954	AMPL3660734218_ATM_Exon17#
chr11	108137953	108138110	AMPL392298588_ATM_Exon17#
chr11	108139103	108139224	AMPL3660785676_ATM_Exon18#
chr11	108139224	108139372	AMPL3660795306_ATM_Exon18#
chr11	108141801	108141965	AMPL391926752_ATM_Exon19#
chr11	108141821	108141993	AMPL394748692_ATM_Exon19#_Exon#20#
chr11	108143142	108143309	AMPL405925555_ATM_Exon21#
chr11	108143283	108143417	AMPL1186569991_ATM_Exon21#
chr11	108143454	108143598	AMPL3660734964_ATM_Exon22#
chr11	108150159	108150253	AMPL394889034_ATM_Exon23#
chr11	108150253	108150428	AMPL437719184_ATM_Exon23#
chr11	108151630	108151799	AMPL622791648_ATM_Exon24#
chr11	108151799	108151957	AMPL391714171_ATM_Exon24#
chr11	108153428	108153506	AMPL3660881930_ATM_Exon25#
chr11	108153499	108153674	AMPL394898140_ATM_Exon25#
chr11	108154887	108154996	AMPL392039731_ATM_Exon26#
chr11	108154993	108155128	AMPL1493143374_ATM_Exon26#
chr11	108155125	108155205	AMPL3660806661_ATM_Exon26#
chr11	108158173	108158344	AMPL398028992_ATM_Exon27#
chr11	108158344	108158491	AMPL3660799178_ATM_Exon27#
chr11	108159655	108159818	AMPL394737832_ATM_Exon28#
chr11	108159759	108159838	AMPL410139130_ATM_Exon28#
chr11	108160294	108160451	AMPL1484991368_ATM_Exon29#
chr11	108160381	108160553	AMPL391647032_ATM_Exon29#
chr11	108163272	108163437	AMPL624516207_ATM_Exon30#
chr11	108163434	108163563	AMPL396418276_ATM_Exon30#
chr11	108164008	108164168	AMPL424101530_ATM_Exon31#
chr11	108164140	108164253	AMPL411271719_ATM_Exon31#
chr11	108165630	108165801	AMPL391523578_ATM_Exon32#
chr11	108167987	108168078	AMPL1034497409_ATM_Exon33#
chr11	108170412	108170552	AMPL548404054_ATM_Exon34#
chr11	108170537	108170648	AMPL3660881219_ATM_Exon34#
chr11	108172329	108172409	AMPL392000436_ATM_Exon35#
chr11	108172409	108172542	AMPL886331227_ATM_Exon35#
chr11	108173541	108173697	AMPL3660816598_ATM_Exon36#
chr11	108173697	108173792	AMPL3660816898_ATM_Exon36#
chr11	108175384	108175481	AMPL391465960_ATM_Exon37#
chr11	108175446	108175525	AMPL391468297_ATM_Exon37#
chr11	108175525	108175638	AMPL564204458_ATM_Exon37#
chr11	108178586	108178747	AMPL3660792931_ATM_Exon38#
chr11	108180811	108180961	AMPL401725247_ATM_Exon39#
chr11	108180961	108181074	AMPL391988106_ATM_Exon39#
chr11	108182987	108183139	AMPL1105767645_ATM_Exon40#
chr11	108183139	108183282	AMPL391653694_ATM_Exon40#
chr11	108186543	108186680	AMPL391887107_ATM_Exon41#
chr11	108186710	108186875	AMPL391557630_ATM_Exon42#
chr11	108188069	108188163	AMPL403099818_ATM_Exon43#

Chr.	Start	Stop	AmpliconID_Gene_Exon
chr11	108188144	108188275	AMPL392139446_ATM_Exon43#
chr11	108190618	108190692	AMPL3660792146_ATM_Exon44#
chr11	108190692	108190824	AMPL392158627_ATM_Exon44#
chr11	108191947	108192064	AMPL391469365_ATM_Exon45#
chr11	108192064	108192167	AMPL391477050_ATM_Exon45#
chr11	108196027	108196166	AMPL392048296_ATM_Exon46#
chr11	108196166	108196330	AMPL392159002_ATM_Exon46#
chr11	108196788	108196881	AMPL395528661_ATM_Exon47#
chr11	108196802	108196889	AMPL392242365_ATM_Exon47#
chr11	108198241	108198405	AMPL433902805_ATM_Exon48#
chr11	108198405	108198495	AMPL391740586_ATM_Exon48#
chr11	108199749	108199822	AMPL1706343703_ATM_Exon49#
chr11	108199817	108199906	AMPL391485841_ATM_Exon49#
chr11	108199906	108200019	AMPL391496996_ATM_Exon49#
chr11	108200917	108201010	AMPL637532870_ATM_Exon50#
chr11	108201010	108201153	AMPL3660813389_ATM_Exon50#
chr11	108202073	108202231	AMPL391465932_ATM_Exon51#
chr11	108202231	108202400	AMPL425276545_ATM_Exon51#
chr11	108202507	108202681	AMPL391795564_ATM_Exon52#
chr11	108202681	108202847	AMPL624574569_ATM_Exon52#
chr11	108203399	108203517	AMPL623313100_ATM_Exon53#
chr11	108203517	108203651	AMPL391664576_ATM_Exon53#
chr11	108204595	108204673	AMPL391502501_ATM_Exon54#
chr11	108204689	108204827	AMPL426749282_ATM_Exon54#
chr11	108205569	108205735	AMPL410139362_ATM_Exon55#
chr11	108205732	108205847	AMPL391694137_ATM_Exon55#
chr11	108206500	108206621	AMPL391771149_ATM_Exon56#
chr11	108206621	108206697	AMPL651267058_ATM_Exon56#
chr11	108213815	108213971	AMPL397937009_ATM_Exon57#
chr11	108213971	108214138	AMPL391562624_ATM_Exon57#
chr11	108216525	108216652	AMPL3660877791_ATM_Exon58#
chr11	108217986	108218131	AMPL1035082608_ATM_Exon59#
chr11	108224475	108224641	AMPL3660764478_ATM_Exon60#
chr11	108225506	108225661	AMPL3660740856_ATM_Exon61#
chr11	108235661	108235830	AMPL1312142110_ATM_Exon62#
chr11	108235830	108235957	AMPL1503005649_ATM_Exon62#
chr11	108235983	108236158	AMPL1035106763_ATM_Exon63#
chr11	108236158	108236296	AMPL391548043_ATM_Exon63#
chr17	7572847	7573025	AMPL388592559_TP53_Exon11#
chr17	7573863	7574041	AMPL387894777_TP53_Exon10#
chr17	7576424	7576537	AMPL4296982438_TP53_Exon9#
chr17	7576566	7576697	AMPL4297343428_TP53_Exon9#
chr17	7576789	7576968	AMPL1514447650_TP53_Exon9#
chr17	7576995	7577174	AMPL705603720_TP53_Exon8#
chr17	7577356	7577509	AMPL388733576_TP53_Exon7#
chr17	7577508	7577613	AMPL387805648_TP53_Exon7#
chr17	7578150	7578333	AMPL387845397_TP53_Exon6#
chr17	7578249	7578425	AMPL2449945349_TP53_Exon6#_Exon7#
chr17	7578425	7578560	AMPL707393441_TP53_Exon5#

10. Appendix

Chr.	Start	Stop	AmpliconID_Gene_Exon
chr17	7579234	7579385	AMPL624460718_TP53_Exon4#
chr17	7579385	7579522	AMPL651471237_TP53_Exon4#
chr17	7579522	7579684	AMPL388679445_TP53_Exon4#
chr17	7579609	7579766	AMPL1547505015_TP53_Exon3#
chr17	7579782	7579964	AMPL388009515_TP53_Exon2#

10.2 Primer for Sanger Sequencing

Table S2: Primer used for Sanger sequencing validation; chr. chromosome; bp base pairs; hg human genome

Target	Genomic Position (hg19)	Target Length	Sequence (5'->3')
ATM Exon 3	chr.11:108098525-108098716	192 bp	GCCTGATTGAGATCCTGAAACAATTAA
			GCAAAGATAAATGTTAACACTTACACACAA
ATM Exon 5	chr11:108106323-108106618	295 bp	TAAATAGTTGCCATTCCAAGTGTC
			TGGTGAAGTTTCATTCATGAGG
ATM Exon 6	chr.11:108114558-108114987	271 bp	ATTAGTGCAGTTTAAAATCCTTTCTGT
			TCTCTTAGGAATCCACTAGTTCTGTTA
ATM Exon 17	chr.11:108137832-108138138	307 bp	AAATTTTGACTACAGCATGCTCCTG
			GAGGCCTCTTATACTGCCAAATCAATAT
ATM Exon 22	chr.11:108143432-108143623	192 bp	CAGTTCTTTCCCGTAGGCTGAT
			CATCTGCAGCATTCCAAATACTTCA
ATM Exon 35	chr.11:108172301-108172571	271 bp	TTTCAGTGGAGGTTAACATTCAAGA
			ACAGAACTGTTTAGATATGCTGGGTATT
ATM Exon 48	chr.11:108198217-108198523	307 bp	ATAGTTGTATGGCAAAAGCAGATGA
			CTAAGTAACTATCTAAGGGTTGCTCCA
ATM Exon 56	chr11:108206495-108206788	294 bp	TTGCTATTCTCAGATGACTCTGTG
			GCCTCCCAAAGCATTATGAATATG
DDX3X Exon 9	chr.X:41203225-41203398	174 bp	TGATGAACTTTCAAACAGGGTAGGT
			CTGGAACTCATACTTACCTTCTGGCT
MYD88 Exon 3	chr.3:38182940-38182156	217 bp	GCAGGAGATGATCCGGCAAC
			GCTGGACAGTGCACAGCTA
MYD88 Exon 5	chr.3:38182530-38182829	300 bp	GGGATGGCTGTTGTTAACCCCT
			GTACATGGACAGGCAGACAGATAC
NOTCH1 Exon 14	chr.9:139407809-139408023	215 bp	CCTCCCTCGACCTGCAGT
			GGCCCTCTGCAC TGAGAA
NOTCH1 Exon 16	chr.9:139405541-139405783	243 bp	CCGATTGGGAGATCCCTCT
			CCTGTGTCCCCGAGACAT
NOTCH1 Exon 23	chr.9:139401097-139401457	361 bp	GCTTGGGCCACTGACGAAA
			GGTAAGAGCAGGGCAGTGA
NOTCH1 Exon 25	chr.9:139399866-139400359	494bp	ACCGTCCCTGTCTTCCCTCTC
			CAGAGACTGCGTGCAGTTC
NOTCH1 Exon 34	chr.9:139390592-139391023	432 bp	TTACAGATGCAGCAGCAGAAC
			GGACCAGTCGGAGACGTT
PTPN6 Exon 11	chr.12:7067054-7067272	219 bp	GGGCACTGACCCTATGTCC
			AGCTGTCACTACTACCACAGGATT

Target	Genomic Position (hg19)	Target Length	Sequence (5'->3')
SF3B1 Exon 7	chr.2:198274465-198274659	195 bp	GCAACCCCAGGCTAAAAATATG
			CCCACACACCCATACTCCACTA
SF3B1 Exon 14	chr.2:198267128-198267397	457 bp	ACTCATGACTGTCTTCTTTGTTACA
			ACAGGCTGTGTGTACCTCTA
SF3B1 Exon 15	chr.2:198266692-198266886	195 bp	GGCATAGTTAACACCTGTGTTGGTT
			GTAATTGGTGGATTACCTTCCTCTGT
SF3B1 Exon 16	chr.2:198266409-198266719	311 bp	ACAGAGGAAAGGTAAATCCACCAATTAC
			ATTCTGTTAGAACCATGAAACATATCCAGTT
SF3B1 Exon 18	chr.2:198265302-198265519	218 bp	GGAGCAGCAGATATTGATCATAAACTTG
			TATCGTTGGTAACCCCCCTGA
TP53 Exon 5	chr.17:7,578,228-7,578,579	352 bp	CTCTGTCTCCTTCCTCTTACA
			ATCCAATACTCCACACGCAA
TP53 Exon 7	chr.17:7,577,489-7,577,634	146 bp	CATCTTGGGCCTGTGTTATCTCC
			GGCTCCTGACCTGGAGTCCT
XPO1 Exon 15	chr.2: 61719302-61719749	448 bp	GCAATGCATGAAGAGGACGAAAA
			ACAAGCCATATCCTGGACTCCAT

10. Appendix

10.3 Sequencing Data

Table S3: A complete list of non-synonymous variants ≥5% allelic frequency is shown. Bold variants were confirmed by Sanger sequencing. Freq frequency; Cov coverage; FI functional impact; fs frame shift; * stop gained; ND not determined

Gene Name	Start Position	Transcript	Variant Type	cDNA Change	Protein Change	Allelic Freq.	Cov.	Patient	FI
ATM	chr11:108098417	NM_000051	fsdeletion	c.66delA	p.E22fs	80%	10,198	P114	ND
ATM	chr11:108098576	NM_000051	missense	c.146C>G	p.S49C	93%	783	P108	Yes
ATM	chr11:108098576	NM_000051	missense	c.146C>G	p.S49C	77%	1,507	P132	Yes
ATM	chr11:108098576	NM_000051	missense	c.146C>G	p.S49C	48%	1,011	P052	Yes
ATM	chr11:108098576	NM_000051	missense	c.146C>G	p.S49C	50%	1,152	P065	Yes
ATM	chr11:108106435	NM_000051	missense	c.370A>G	p.I124V	48%	2,036	P084	No
ATM	chr11:108114703	NM_000051	deletion	c.520_525delCTCTAT	p.L174_Y175del	41%	137	P031	ND
ATM	chr11:108114799	NM_000051	missense	c.616A>C	p.N206H	57%	1,523	P025	No
ATM	chr11:108115522	NM_000051	missense	c.670A>G	p.K224E	54%	1,070	P134	Yes
ATM	chr11:108115663	NM_000051	missense	c.811C>A	p.L271I	55%	3,062	P133	No
ATM	chr11:108115741	NM_000051	missense	c.889A>G	p.T297A	6%	120	P044	No
ATM	chr11:108121489	NM_000051	missense	c.1297T>C	p.S433P	5%	174	P124	No
ATM	chr11:108124761	NM_000051	missense	c.2119T>C	p.S707P	48%	2,985	P123	No
ATM	chr11:108124761	NM_000051	missense	c.2119T>C	p.S707P	50%	1,542	P028	No
ATM	chr11:108124761	NM_000051	missense	c.2119T>C	p.S707P	53%	324	P009	No
ATM	chr11:108128253	NM_000051	missense	c.2296A>G	p.K766E	8%	307	P085	Yes
ATM	chr11:108128253	NM_000051	missense	c.2296A>G	p.K766E	55%	369	P093	Yes
ATM	chr11:108137925	NM_000051	missense	c.2494C>T	p.R832C	51%	1,107	P117	No
ATM	chr11:108138003	NM_000051	missense	c.2572T>C	p.F858L	49%	4,339	P101	Yes
ATM	chr11:108138003	NM_000051	missense	c.2572T>C	p.F858L	48%	504	P012	Yes
ATM	chr11:108138003	NM_000051	missense	c.2572T>C	p.F858L	49%	4,146	P125	Yes
ATM	chr11:108138003	NM_000051	missense	c.2572T>C	p.F858L	49%	3,255	P133	Yes
ATM	chr11:108138003	NM_000051	missense	c.2572T>C	p.F858L	52%	459	P017	Yes
ATM	chr11:108138003	NM_000051	missense	c.2572T>C	p.F858L	56%	158	P004	Yes
ATM	chr11:108143293	NM_000051	missense	c.3112G>C	p.V1038L	47%	1,428	P115	Yes
ATM	chr11:108143456	NM_000051	missense	c.3161C>G	p.P1054R	44%	16	P125	Yes
ATM	chr11:108143456	NM_000051	missense	c.3161C>G	p.P1054R	54%	25	P083	Yes
ATM	chr11:108143470	NM_000051	missense	c.3175G>A	p.A1059T	48%	5,088	P119	Yes
ATM	chr11:108143552	NM_000051	missense	c.3257G>A	p.R1086H	50%	2,477	P042	No
ATM	chr11:108159732	NM_000051	missense	c.4138C>T	p.H1380Y	52%	426	P084	No
ATM	chr11:108159814	NM_000051	missense	c.4220T>C	p.I1407T	6%	162	P095	No
ATM	chr11:108160516	NM_000051	missense	c.4424A>G	p.Y1475C	50%	1,764	P053	No
ATM	chr11:108160516	NM_000051	missense	c.4424A>G	p.Y1475C	47%	1,613	P088	No
ATM	chr11:108172404	NM_000051	missense	c.5207G>A	p.C1736Y	66%	130	P122	Yes
ATM	chr11:108173733	NM_000051	fsdeletion	c.5473delC	p.Q1825fs	99%	3,383	P079	ND
ATM	chr11:108186557	NM_000051	missense	c.6014T>C	p.L2005P	6%	3,100	P114	Yes
ATM	chr11:108186610	NM_000051	missense	c.6067G>A	p.G2023R	48%	3,394	P076	Yes
ATM	chr11:108198394	NM_000051	missense	c.6998C>A	p.T2333K	98%	189	P048	No
ATM	chr11:108205835	NM_000051	missense	c.8150A>C	p.K2717T	8%	414	P058	Yes
ATM	chr11:108206609	NM_000051	missense	c.8189A>G	p.Q2730R	45%	5,059	P105	Yes
ATM	chr11:108216545	NM_000051	missense	c.8494C>T	p.R2832C	7%	2,831	P048	Yes
ATM	chr11:108236203	NM_000051	nonsense	c.9139C>T	p.R3047*	28%	1,052	P074	ND
DDX3X	chrX:41203340	NM_001356	missense	c.823A>C	p.T275P	96%	572	P119	Yes
MYD88	chr3:38182025	NM_001172567	missense	c.649G>T	p.V217F	44%	2,427	P103	ND
MYD88	chr3:38182641	NM_002468	missense	c.613T>C	p.L265P	34%	3,209	P078	Yes
NOTCH1	chr9:139413193	NM_017617	missense	c.949G>A	p.G317S	48%	6,738	P074	Yes
NOTCH1	chr9:139413163	NM_017617	missense	c.979T>C	p.W327R	5%	4,765	P061	No

Gene Name	Start Position	Transcript	Variant Type	cDNA Change	Protein Change	Allelic Freq.	Cov.	Patient	FI
NOTCH1	chr9:139412642	NM_017617	missense	c.1202C>T	p.P401L	12%	405	P079	No
NOTCH1	chr9:139410544	NM_017617	missense	c.1558T>C	p.F520L	7%	104	P134	Yes
NOTCH1	chr9:139407932	NM_017617	missense	c.2264_2265delinsGC	p.N755S	50%	1,243	P109	No
NOTCH1	chr9:139407932	NM_017617	missense	c.2264_2265delinsGC	p.N755S	26%	883	P126	No
NOTCH1	chr9:139407932	NM_017617	missense	c.2264_2265delinsGC	p.N755S	50%	1,203	P130	No
NOTCH1	chr9:139407559	NM_017617	missense	c.2381A>G	p.E794G	8%	148	P039	Yes
NOTCH1	chr9:139407559	NM_017617	missense	c.2381A>G	p.E794G	7%	155	P093	Yes
NOTCH1	chr9:139405111	NM_017617	missense	c.2734C>T	p.R912W	40%	45	P097	No
NOTCH1	chr9:139403378	NM_017617	missense	c.3115G>A	p.G1039S	48%	2,096	P126	No
NOTCH1	chr9:139402584	NM_017617	missense	c.3332_3333delinsCT	p.D1111A	28%	1,624	P111	No
NOTCH1	chr9:139402584	NM_017617	missense	c.3332_3333delinsCT	p.D1111A	25%	3,868	P090	No
NOTCH1	chr9:139401234	NM_017617	missense	c.3835C>T	p.R1279C	56%	462	P059	No
NOTCH1	chr9:139401233	NM_017617	missense	c.3836G>A	p.R1279H	49%	851	P111	No
NOTCH1	chr9:139401233	NM_017617	missense	c.3836G>A	p.R1279H	57%	2,021	P121	No
NOTCH1	chr9:139401233	NM_017617	missense	c.3836G>A	p.R1279H	54%	162	P027	No
NOTCH1	chr9:139401233	NM_017617	missense	c.3836G>A	p.R1279H	44%	262	P028	No
NOTCH1	chr9:139401233	NM_017617	missense	c.3836G>A	p.R1279H	48%	257	P038	No
NOTCH1	chr9:139401233	NM_017617	missense	c.3836G>A	p.R1279H	48%	706	P053	No
NOTCH1	chr9:139401233	NM_017617	missense	c.3836G>A	p.R1279H	50%	1,296	P072	No
NOTCH1	chr9:139401233	NM_017617	missense	c.3836G>A	p.R1279H	53%	1,197	P094	No
NOTCH1	chr9:139400320	NM_017617	missense	c.4028C>T	p.A1343V	49%	33	P025	No
NOTCH1	chr9:139400219	NM_017617	missense	c.4129C>T	p.P1377S	51%	2,144	P125	No
NOTCH1	chr9:139400219	NM_017617	missense	c.4129C>T	p.P1377S	54%	772	P030	No
NOTCH1	chr9:139400219	NM_017617	missense	c.4129C>T	p.P1377S	15%	194	P031	No
NOTCH1	chr9:139400219	NM_017617	missense	c.4129C>T	p.P1377S	56%	473	P034	No
NOTCH1	chr9:139400219	NM_017617	missense	c.4129C>T	p.P1377S	50%	4,135	P066	No
NOTCH1	chr9:139400035	NM_017617	missense	c.4313G>A	p.R1438H	48%	147	P032	No
NOTCH1	chr9:139399327	NM_017617	missense	c.4816T>C	p.F1606L	5%	166	P046	Yes
NOTCH1	chr9:139391908	NM_017617	missense	c.6283C>T	p.R2095C	52%	651	P094	No
NOTCH1	chr9:139391338	NM_017617	missense	c.6853G>A	p.V2285I	50%	57	P001	No
NOTCH1	chr9:139391338	NM_017617	missense	c.6853G>A	p.V2285I	51%	818	P120	No
NOTCH1	chr9:139391338	NM_017617	missense	c.6853G>A	p.V2285I	55%	1,092	P123	No
NOTCH1	chr9:139391338	NM_017617	missense	c.6853G>A	p.V2285I	53%	672	P130	No
NOTCH1	chr9:139390945	NM_017617	nonsense	c.7246C>T	p.Q2416*	62%	165	P028	ND
NOTCH1	chr9:139390828	NM_017617	missense	c.7363A>G	p.T2455A	42%	255	P023	No
NOTCH1	chr9:139390813	NM_017617	nonsense	c.7378G>T	p.E2460*	25%	2,511	P072	ND
NOTCH1	chr9:139390649	NM_017617	fsdeletion	c.7541_7542delCT	p.P2514fs	29%	48	P001	Yes
NOTCH1	chr9:139390649	NM_017617	fsdeletion	c.7541_7542delCT	p.P2514fs	13%	954	P111	Yes
NOTCH1	chr9:139390649	NM_017617	fsdeletion	c.7541_7542delCT	p.P2514fs	54%	2,479	P114	Yes
NOTCH1	chr9:139390649	NM_017617	fsdeletion	c.7541_7542delCT	p.P2514fs	46%	36	P016	Yes
NOTCH1	chr9:139390649	NM_017617	fsdeletion	c.7541_7542delCT	p.P2514fs	60%	1,879	P062	Yes
NOTCH1	chr9:139390649	NM_017617	fsdeletion	c.7541_7542delCT	p.P2514fs	12%	1,852	P071	Yes
NOTCH1	chr9:139390649	NM_017617	fsdeletion	c.7541_7542delCT	p.P2514fs	58%	3,499	P080	Yes
PTPN6	chr12:7067226	NM_080549	missense	c.1351G>A	p.V451M	51%	1,191	P016	Yes
SF3B1	chr2:198274548	NM_012433	missense	c.850G>A	p.A284T	49%	4,440	P081	No
SF3B1	chr2:198267491	NM_012433	missense	c.1866G>C	p.E622D	14%	28,824	P132	Yes
SF3B1	chr2:198267491	NM_012433	missense	c.1866G>T	p.E622D	49%	11,601	P056	Yes
SF3B1	chr2:198267489	NM_012433	missense	c.1868A>G	p.Y623C	42%	11,120	P108	Yes
SF3B1	chr2:198267489	NM_012433	missense	c.1868A>G	p.Y623C	10%	13,019	P086	Yes
SF3B1	chr2:198267481	NM_012433	missense	c.1876A>T	p.N626Y	48%	19,047	P088	Yes
SF3B1	chr2:198267480	NM_012433	missense	c.1877A>T	p.N626I	9%	3,760	P035	Yes
SF3B1	chr2:198267480	NM_012433	missense	c.1877A>G	p.N626S	7%	8,908	P079	Yes
SF3B1	chr2:198267371	NM_012433	missense	c.1986C>A	p.H662Q	45%	852	P015	Yes

10. Appendix

Gene Name	Start Position	Transcript	Variant Type	cDNA Change	Protein Change	Allelic Freq.	Cov.	Patient	FI
<i>SF3B1</i>	chr2:198267361	NM_012433	missense	c.1996A>G	p.K666E	6%	11,613	P101	Yes
<i>SF3B1</i>	chr2:198267361	NM_012433	missense	c.1996A>G	p.K666E	51%	3,169	P122	Yes
<i>SF3B1</i>	chr2:198267360	NM_012433	missense	c.1997A>T	p.K666M	48%	5,418	P057	Yes
<i>SF3B1</i>	chr2:198266837	NM_012433	missense	c.2095C>G	p.Q699E	51%	3,254	P071	Yes
<i>SF3B1</i>	chr2:198266834	NM_012433	missense	c.2098A>G	p.K700E	11%	360	P011	Yes
<i>SF3B1</i>	chr2:198266834	NM_012433	missense	c.2098A>G	p.K700E	44%	362	P021	Yes
<i>SF3B1</i>	chr2:198266834	NM_012433	missense	c.2098A>G	p.K700E	48%	2,448	P084	Yes
<i>SF3B1</i>	chr2:198266834	NM_012433	missense	c.2098A>G	p.K700E	9%	4,504	P091	Yes
<i>SF3B1</i>	chr2:198266822	NM_012433	missense	c.2110A>T	p.I704F	48%	5,892	P110	Yes
<i>SF3B1</i>	chr2:198266821	NM_012433	missense	c.2111T>A	p.I704N	19%	1,651	P100	Yes
<i>SF3B1</i>	chr2:198266821	NM_012433	missense	c.2111T>C	p.I704T	8%	719	P035	Yes
<i>SF3B1</i>	chr2:198266611	NM_012433	missense	c.2225G>A	p.G742D	14%	13,125	P111	Yes
<i>SF3B1</i>	chr2:198266611	NM_012433	missense	c.2225G>A	p.G742D	13%	2,079	P044	Yes
<i>SF3B1</i>	chr2:198266611	NM_012433	missense	c.2225G>A	p.G742D	40%	3,085	P050	Yes
<i>SF3B1</i>	chr2:198265476	NM_012433	missense	c.2681A>G	p.D894G	54%	631	P100	Yes
<i>TP53</i>	chr17:7579719	NM_000546	fsdeletion	c.77delT	p.L26fs	89%	1,610	P064	ND
<i>TP53</i>	chr17:7579719	NM_000546	missense	c.77T>C	p.L26P	20%	169	P064	Yes
<i>TP53</i>	chr17:7579718	NM_000546	fsdeletion	c.78delT	p.L26fs	9%	1,481	P064	ND
<i>TP53</i>	chr17:7579559	NM_000546	nonsense	c.128T>A	p.L43*	6%	417	P110	ND
<i>TP53</i>	chr17:7579361	NM_000546	missense	c.326T>G	p.F109C	81%	6,515	P101	Yes
<i>TP53</i>	chr17:7578526	NM_000546	missense	c.404G>A	p.C135Y	70%	18,906	P095	Yes
<i>TP53</i>	chr17:7578475	NM_000546	missense	c.455C>T	p.P152L	9%	17,762	P132	Yes
<i>TP53</i>	chr17:7578406	NM_000546	missense	c.524G>A	p.R175H	16%	5,040	P095	Yes
<i>TP53</i>	chr17:7578280	NM_000546	missense	c.569C>T	p.P190L	10%	11,695	P067	Yes
<i>TP53</i>	chr17:7578265	NM_000546	missense	c.584T>C	p.I195T	8%	15,913	P086	Yes
<i>TP53</i>	chr17:7578253	NM_000546	missense	c.596G>A	p.G199E	58%	386	P032	Yes
<i>TP53</i>	chr17:7578208	NM_000546	missense	c.641A>G	p.H214R	11%	3,170	P111	Yes
<i>TP53</i>	chr17:7578204	NM_000546	missense	c.645T>A	p.S215R	97%	2,456	P085	Yes
<i>TP53</i>	chr17:7578190	NM_000546	missense	c.659A>G	p.Y220C	96%	1,683	P091	Yes
<i>TP53</i>	chr17:7578184	NM_000546	missense	c.665C>T	p.P222L	69%	156	P045	No
<i>TP53</i>	chr17:7577580	NM_000546	missense	c.701A>G	p.Y234C	6%	7,440	P058	Yes
<i>TP53</i>	chr17:7577580	NM_000546	missense	c.701A>G	p.Y234C	5%	2,635	P067	Yes
<i>TP53</i>	chr17:7577559	NM_000546	missense	c.722C>T	p.S241F	97%	4,136	P119	Yes
<i>TP53</i>	chr17:7577556	NM_000546	missense	c.725G>A	p.C242Y	99%	3,730	P117	Yes
<i>TP53</i>	chr17:7577550	NM_000546	missense	c.731G>A	p.G244D	7%	3,609	P112	Yes
<i>TP53</i>	chr17:7577548	NM_000546	missense	c.733G>T	p.G245C	7%	21,069	P132	Yes
<i>TP53</i>	chr17:7577547	NM_000546	missense	c.734G>A	p.G245D	10%	6,840	P040	Yes
<i>TP53</i>	chr17:7577538	NM_000546	missense	c.743G>A	p.R248Q	98%	4,972	P082	Yes
<i>TP53</i>	chr17:7577142	NM_000546	missense	c.796G>A	p.G266R	13%	9,074	P115	Yes
<i>TP53</i>	chr17:7577142	NM_000546	nonsense	c.796G>T	p.G266*	20%	7,472	P094	ND
<i>TP53</i>	chr17:7577113	NM_000546	missense	c.824_825delinsTC	p.C275F	6%	5,282	P040	Yes
<i>TP53</i>	chr17:7577108	NM_000546	missense	c.830G>T	p.C277F	13%	18,021	P105	Yes
<i>TP53</i>	chr17:7577105	NM_000546	missense	c.833C>G	p.P278R	7%	9,393	P112	Yes
<i>TP53</i>	chr17:7577036	NM_000546	fsinsertion	c.902_903insC	p.P301fs	13%	9,566	P067	ND
<i>TP53</i>	chr17:7577034	NM_000546	missense	c.904G>A	p.G302R	7%	1,511	P044	No
<i>TP53</i>	chr17:7576897	NM_000546	nonsense	c.949C>T	p.Q317*	13%	3,408	P132	ND
<i>XPO1</i>	chr2:61719471	NM_003400	missense	c.1711_1712delinsAT	p.E571I	24%	3,789	P052	Yes
<i>XPO1</i>	chr2:61719472	NM_003400	missense	c.1711G>A	p.E571K	26%	570	P001	Yes
<i>XPO1</i>	chr2:61719472	NM_003400	missense	c.1711G>A	p.E571K	51%	4,247	P130	Yes
<i>XPO1</i>	chr2:61719472	NM_003400	missense	c.1711G>A	p.E571K	48%	4,581	P028	Yes
<i>XPO1</i>	chr2:61719472	NM_003400	missense	c.1711G>A	p.E571K	44%	1,955	P037	Yes
<i>XPO1</i>	chr2:61719472	NM_003400	missense	c.1711G>A	p.E571K	49%	9,922	P057	Yes
<i>XPO1</i>	chr2:61719472	NM_003400	missense	c.1711G>A	p.E571K	45%	5,696	P068	Yes

Gene Name	Start Position	Transcript	Variant Type	cDNA Change	Protein Change	Allelic Freq.	Cov.	Patient	FI
XPO1	chr2:61719472	NM_003400	missense	c.1711G>A	p.E571K	18%	5,262	P072	Yes
XPO1	chr2:61719472	NM_003400	missense	c.1711G>A	p.E571K	52%	3,721	P080	Yes
XPO1	chr2:61719472	NM_003400	missense	c.1711G>A	p.E571K	50%	4,457	P091	Yes
XPO1	chr2:61719472	NM_003400	missense	c.1711G>A	p.E571K	49%	4,781	P094	Yes
XPO1	chr2:61719472	NM_003400	missense	c.1711G>A	p.E571K	52%	2,197	P096	Yes
XPO1	chr2:61719472	NM_003400	missense	c.1711G>C	p.E571Q	13%	3,121	P048	Yes

10.4 Statistical Analysis

Table S4: Statistical correlations* between gene mutation status and clinical and biological parameters are summarized; correlations were calculated only for variants that were predicted to have a functional impact on the protein.

	Gene	Clinical/Prognostic Parameter	Statistical Test	P-Value	Z	Chi-Square
Overall Cohort	<i>SF3B1</i>	Male sex	Fisher's Exact Test	0.008	-	-
		Unmutated <i>IGHV</i>	Fisher's Exact Test	0.030	-	-
		Decreased platelet count	Exact Wilcoxon Mann-Whitney Rank Sum Test	0.025	2.2344	-
	<i>TP53</i>	Binet stage	Pearson's Chi-squared test	0.008	-	13.0847
		Chromosome 17p deletion	Fisher's Exact Test	<0.001	-	-
		Unmutated <i>IGHV</i>	Fisher's Exact Test	0.038	-	-
		Treatment status	Fisher's Exact Test	<0.001	-	-
	<i>NOTCH</i> (<i>p.P2514fs</i>)	Trisomy 12	Fisher's Exact Test	0.022	-	-
Untreated Cohort	<i>XPO1</i>	Unmutated <i>IGHV</i>	Fisher's Exact Test	<0.001	-	-
		Increased white blood count (WBC)	Exact Wilcoxon Mann-Whitney Rank Sum Test	<0.001	-3.2751	-
	<i>NOTCH1</i> (<i>p.P2514fs</i>)	Unmutated <i>IGHV</i>	Fisher's Exact Test	0.037	-	-
Treated Cohort	<i>SF3B1</i>	CD38 positivity	Exact Wilcoxon Mann-Whitney Rank Sum Test	0.031	-2.1416	-
	<i>XPO1</i>	ZAP70 positivity	Exact Wilcoxon Mann-Whitney Rank Sum Test	0.024	-2.2129	

* Statistical tests chosen for correlation analysis are described in material and methods 3.9

11. Lebenslauf

Personal Information

Name	Claudia Vollbrecht, B.Sc.
Date of birth	3 November 1984
Nationality	German

Education

Since October 2012	Postgraduate studies at the Institute of Pathology, University Hospital Cologne in Cologne, Germany
Oct 20011 – Sep 2012	Scientific research at the Institute of Pathology, University Hospital Cologne in Cologne, Germany
Oct 2007 – Feb 2011	Bachelor's Degree in Biotechnology Studies at the Brandenburg University of Technology Cottbus in Senftenberg, Germany
Oct 2010 – Feb 2011	Bachelor-Thesis at the Institute for Pathology at the Medical University in Graz, Austria: " <i>Deep sequencing of mutations of the Epidermal Growth Factor Receptor (EGFR) in adenocarcinomas of the lung</i> "
Sep 2009 – Feb 2010	Internship at the Institute for Pathology at the Medical University in Innsbruck, Austria: " <i>The relevance of p16 and p27 for recurrent-free survival in low and high grade urothelial bladder cancer</i> "
Oct 2004 – Sep 2007	Studies in Biology at the Johann-Wolfgang Goethe University in Frankfurt/ Main, Germany
Aug 1998 – Jul 2004	Final secondary school examination at the Lise-Meitner Gymnasium in Falkensee, Germany
Aug 1991 – Jul 1997	Primary school Schönwalde, Germany

Further Expertise

May 2014 – May 2014	Automation Biomek Continuum Training, Beckman Coulter, Nyon, Switzerland
Feb 2014 – Feb 2014	Medical Statistics, University of Cologne, Cologne, Germany
Sep 2013 – Sep 2013	Genome Sequencer Software Summer School, Roche Applied Sciences, Mannheim, Germany
Jan 2013 – Aug 2013	Small Business Management, University of Duisburg-Essen, Duisburg, Germany

<i>Mar 2010 – Sep 2010</i>	Project work in the working group of Prof. Dr. rer. nat. J.-H. Küpper at the Brandenburg University of Technology Cottbus in Senftenberg, Germany: “ <i>Investigation of Cytochrome P450 expression in proliferating human liver cell lines</i> ”
<i>May 2009 – Jul 2009</i>	Trainee in the working group of Prof. Dr. rer. nat. Ch. Schröder at the Brandenburg University of Technology Cottbus in Senftenberg, Germany working on: “ <i>Homogenous microbead-based semiquantitative detection of nucleic acids in a multiplex format</i> ”

Awards and Grants

<i>Jul 2014</i>	Award: EACR Meeting Bursary
<i>Jul 2013</i>	Award: 2. Preis Gründungsinitiative Innovation Duisburg (GRIID)
<i>Sep 2009</i>	Scholarship: ERASMUS for an internship within the EU

Skills and Activities

<i>Languages</i>	English, German
<i>Scientific Memberships</i>	European Association for Cancer Research (EACR)
<i>Interests</i>	Since February 2013 voluntary service at the retirement home “Johanniter-Stift Köln-Kalk” Cologne, Germany April 2012 – January 2013: Voluntary service for the “SK-Stiftung CSC – Cologne Science” at the Odysseum, Cologne, Germany

Journal Publications

15 articles (4 first authorships)

- Fabian Dominik Mairinger, Saskia Ting, Robert Werner, Robert Fred Henry Walter, Thomas Hager, Claudia Vollbrecht, Daniel Christoph, Karl Worm, Thomas Mairinger, Sien-Yi Sheu-Grabellus, Dirk Theegarten, Kurt Werner Schmid, Jeremias Wohlschlaeger: *Different micro-RNA expression profiles distinguish subtypes of neuroendocrine tumors of the lung: results of a profiling study*. Modern Pathology. 05/2014.
- Fabian D Mairinger, Robert FH Walter, Saskia Ting, Claudia Vollbrecht, Jens Kollmeier, Sergei Griff, Thomas Hager, Thomas Mairinger, Daniel C Christoph, Dirk Theegarten, Schmid Kurt Werner, Jeremias Wohlschlaeger: *Mdm2*

protein expression is strongly associated with survival in malignant pleural mesothelioma. Future Oncology. 05/2014; 10(6):995-1005.

- Claudia Vollbrecht, Fabian Dominik Mairinger, Carmen-Diana Schweighofer, Lukas Carl Heukamp, Sabine Merkelbach-Bruse, Reinhard Buettner, Margarete Odenthal: *Targeting Cancer Related Genes by Multiplex PCR Followed by High Throughput Parallel Sequencing*. International Journal of Genomic Medicine. 03/2014; 2(1).
- Fabian D Mairinger, Claudia Vollbrecht, Anna Streubel, Andreas Roth, Olfert Landt, Henry F R Walter, Jens Kollmeier, Thomas Mairinger: *The "COLD-PCR Approach" for Early and Cost-Effective Detection of Tyrosine Kinase Inhibitor Resistance Mutations in EGFR-Positive Non-Small Cell Lung Cancer*. Applied immunohistochemistry & molecular morphology: AIMM / official publication of the Society for Applied Immunohistochemistry 02/2014; 22(2):114-8.
- Claudia Vollbrecht, Reinhard Büttner: *[Prädiktive und prognostische Biomarker in der Krebstherapie: Was ist heute technisch möglich?]*. Onkologie heute 01/2014 1:9-12.
- Fabian D Mairinger, Robert F H Walter, Robert Werner, Daniel C Christoph, Saskia Ting, Claudia Vollbrecht, Konstantinos Zarogoulidis, Haidong Huang, Qiang Li, Kurt W Schmid, Jeremias Wohlschlaeger, Paul Zarogoulidis: *Activation of angiogenesis differs strongly between pulmonary carcinoids and neuroendocrine carcinomas and is crucial for carcinoid tumourgenesis*. Journal of Cancer. 01/2014; 5(6):465-71.
- Fabian D Mairinger, Robert FH Walter, Claudia Vollbrecht, Thomas Hager, Karl Worm, Saskia Ting, Jeremias Wohlschläger, Paul Zarogoulidis, Konstantinos Zarogoulidis, Kurt W Schmid: *Isothermal multiple displacement amplification: a methodical approach enhancing molecular routine diagnostics of microcarcinomas and small biopsies*. OncoTargets and Therapy 01/2014; 7:1441-7.
- Fabian Dominik Mairinger, Robert Fred Henry Walter, Dirk Theegarten, Thomas Hager, Claudia Vollbrecht, Daniel Christian Christoph, Karl Worm, Saskia Ting, Robert Werner, Georgios Stamatis, Thomas Mairinger, Hideo Baba, Konstantinos Zarogoulidis, Haidong Huang, Qiang Li, Kosmas Tsakiridis, Paul Zarogoulidis, Schmid Kurt Werner, Jeremias Wohlschlaeger: *Gene Expression Analysis of the 26S Proteasome Subunit PSMB4 Reveals*

Significant Upregulation, Different Expression and Association with Proliferation in Human Pulmonary Neuroendocrine Tumours. Journal of Cancer. 01/2014; 5(8):646-54.

- Robert Fred Henry Walter, Fabian Dominik Mairinger, Jeremias Wohlschlaeger, Karl Worm, Saskia Ting, Claudia Vollbrecht, Kurt Werner Schmid, Thomas Hager: *FFPE tissue as a feasible source for gene expression analysis - A comparison of three reference genes and one tumor marker.* Pathology - Research and Practice 09/2013.
- Fabian Mairinger, Claudia Vollbrecht, Thomas Mairinger, Helmut Popper: *The issue of studies evaluating biomarkers which predict outcome after pemetrexed-based chemotherapy in malignant pleural mesothelioma.* Journal of thoracic oncology: official publication of the International Association for the Study of Lung Cancer 08/2013; 8(8):e80-2.
- Kerstin Becker, Claudia Vollbrecht, Ulrike Koitzsch, Katharina Koenig, Jana Fassunke, Sebastian Huss, Peter Nuernberg, Lukas C Heukamp, Reinhard Buettner, Margarete Odenthal, Janine Altmueller, Sabine Merkelbach-Bruse: *Deep ion sequencing of amplicon adapter ligated libraries: a novel tool in molecular diagnostics of formalin fixed and paraffin embedded tissues.* Journal of clinical pathology 04/2013.
- Fabian Mairinger, Claudia Vollbrecht, Iris Halbwedl, Martina Hatz, Elvira Stacher, Christian Gölly, Franz Quehenberger, Susann Stephan-Falkenau, Jens Kollmeier, Andreas Roth, Thomas Mairinger, Helmut Popper: *Reduced Folate Carrier and Folylpolyglutamate Synthetase, but not Thymidylate Synthase Predict Survival in Pemetrexed-Treated Patients Suffering from Malignant Pleural Mesothelioma.* Journal of thoracic oncology: official publication of the International Association for the Study of Lung Cancer 02/2013.
- Claudia Vollbrecht, Katharina König, Lukas Heukamp, Reinhard Büttner, Margarete Odenthal: *[Molecular pathology of the lungs: New perspectives by next generation sequencing].* Der Pathologe 02/2013; 34(1):16-24.
- K Pütz, M Engels, C Vollbrecht, LC Heukamp, AC Adam, R Büttner: *[Indications and limitations of fresh frozen sections in the pulmonary apparatus].* Der Pathologe 07/2012; 33(5):402-6.

Conference Proceedings

(This list is not exhaustive)

32 conference presentations (11 as first and/or presenting author, 5 oral presentations)

- Fabian Dominik Mairinger, Robert Fred Henry Walter, Robert Werner, Claudia Vollbrecht, Saskia Ting, Dirk Theegarten, Daniel Christian Christoph, Kurt Werner Schmid, Jeremias Wohlschläger: *mRNA biomarker screening in neuroendocrine pulmonary malignancies*. Biennial Congress of the European Association for Cancer Research, Munich (GER); 07/2014.
- Claudia Vollbrecht, Fabian Dominik Mairinger, Carmen Schweighofer, Lukas Heukamp, Sabine Merkelbach-Bruse, Reinhardt Büttner, Margarete Odenthal: *Testing cancer relating gene panels for high throughput parallel sequencing*. Biennial Congress of the European Association for Cancer Research, Munich (GER); 07/2014.
- Fabian Dominik Mairinger, Saskia Ting, Robert Werner, Robert Fred Henry Walter, Claudia Vollbrecht, Daniel Christian Christoph, Thomas Mairinger, Dirk Theegarten, Kurt Werner Schmid: *miRNA-regulated pathways in neuroendocrine tumors of the lung*. Biennial Congress of the European Association for Cancer Research, Munich (GER); 07/2014.
- Claudia Vollbrecht, Katharina König, Lukas Heukamp, Martin Peifer, Fabian Dominik Mairinger, Reinhardt Büttner, Margarete Odenthal, Carmen Schweighofer: *Detection of hot spot and pathway related variants in chronic lymphocytic leukemia by multiplexed amplicon sequencing*. Biennial Congress of the European Association for Cancer Research, Munich (GER); 07/2014.
- Claudia Vollbrecht, Ulrike Koitzsch, Katharina Koenig, Michael Kloth, Reinhard Buettner, Margarete Odenthal: *Process-automation of next generation sequencing for high-throughput mutation analysis in cancer*. 23rd Biennial Congress of the European Association For Cancer Research (EACR), Munich, (GER); 07/2014.
- Robert Fred Henry Walter, Fabian Dominik Mairinger, Saskia Ting, Claudia Vollbrecht, Thomas Hager, Karl Worm, Wilfried Eberhardt, Jens Kollemier, Thomas Mairinger, Daniel Christian Christoph, Kurt Werner Schmid, Lutz Freitag, Jeremias Wohlschläger: *Inhibition of MDM2 via Nutlin-3 - the next*

therapy for pleural mesothelioma with poor prognosis? 98. Jahrestagung der deutschen Gesellschaft für Pathologie e. V., Berlin (GER); 06/2014.

- Robert Fred Henry Walter, Fabian Dominik Mairinger, Robert Werner, Saskia Ting, Claudia Vollbrecht, Dirk Theegarten, Georgis Stamatis, Denise Zwanziger, Wilfried Eberhardt, Martin Schuler, Daniel Christian Christoph, Lutz Freitag, Kurt Werner Schmid, Jeremias Wohlschläger: *Screening 80 neuroendocrine lung cancer patients by using NanoString Technology implies several targets for new tailored therapies.* 98. Jahrestagung der deutschen Gesellschaft für Pathologie e. V., Berlin (GER); 06/2014.
- Fabian Dominik Mairinger, Saskia Ting, Robert Werner, Robert Fred Henry Walter, Thomas Hager, Claudia Vollbrecht, Daniel Christoph, Karl Worm, Thomas Mairinger, Sien-Yi Sheu-Grabellus, Dirk Theegarten, Kurt Werner Schmid, Jeremias Wohlschläger: *Exploring miRNAs in neuroendocrine lung tumors.* 98. Jahrestagung der deutschen Gesellschaft für Pathologie e. V., Berlin (GER); 06/2014.
- Robert Fred Henry Walter, Fabian Dominik Mairinger, Robert Werner, Saskia Ting, Claudia Vollbrecht, Dirk Theegarten, Georgis Stamatis, Denise Zwanziger, Wilfried Eberhardt, Martin Schuler, Daniel Christian Christoph, Lutz Freitag, Kurt Werner Schmid, Jeremias Wohlschläger: *Is the mRNA expression profile of apoptosis and cell cycle related genes a marker for the aggressiveness of neuroendocrine tumors of the lung?* 98. Jahrestagung der deutschen Gesellschaft für Pathologie e. V., Berlin (GER); 06/2014.
- Fabian Dominik Mairinger, Robert Fred Henry Walter, Robert Werner, Saskia Ting, Thomas Hager, Claudia Vollbrecht, Dirk Theegarten, Daniel Christian Christoph, Kurt Werner Schmid, Jeremias Wohlschläger: *Investigating angiogenesis in pulmonary neuroendocrine tumours - results of a gene expression profile.* 98. Jahrestagung der deutschen Gesellschaft für Pathologie e. V., Berlin (GER); 06/2014.
- Fabian Dominik Mairinger, Robert Fred Henry Walter, Robert Werner, Claudia Vollbrecht, Saskia Ting, Thomas Hager, Kurt Werner Schmid, Jeremias Wohlschläger, Daniel Christian Christoph: *Folic acid metabolism and DNA-mismatch repair in pulmonary neuroendocrine tumors: identification of potential chemotherapy-resistance mechanisms.* 98. Jahrestagung der deutschen Gesellschaft für Pathologie e. V., Berlin (GER); 06/2014.

- Claudia Vollbrecht, Carmen Schweighofer, Ulrike Koitzsch, Laura Wilden, Katharina König, Lukas Heukamp, Reinhard Büttner, Margarete Odenthal: *Establishment of multiplexed amplicon sequencing to assess hotspot and pathway related variants in chronic lymphocytic leukemia.* 98. Jahrestagung der Deutschen Gesellschaft für Pathologie e.V., Berlin (GER); 06/2014.
- Fabian Dominik Mairinger, Robert Fred Henry Walter, Thomas Hager, Claudia Vollbrecht, Daniel Christoph, Karl Worm, Kurt Werner Schmid, Jeremias Wohlschläger: *The proteasome subunits PSMA5 and PSMB4 are potent markers for proliferation in pulmonary carcinoids.* European Respiratory Society Annual Congress 2013, Barcelona (ESP); 09/2013.
- Robert Fred Henry Walter, Fabian Dominik Mairinger, Robert Werner, Saskia Ting, Thomas Hager, Claudia Vollbrecht, Daniel Christoph, Kurt Werner Schmid: *Gene expression screening using NanoString nCounter technology in lung tumors.* European Respiratory Society Annual Congress 2013, Barcelona (ESP); 09/2013.
- Fabian Dominik Mairinger, Saskia Ting, Robert Werner, Robert Fred Henry Walter, Thomas Hager, Claudia Vollbrecht, Daniel Christoph, Karl Worm, Kurt Werner Schmid, Jeremias Wohlschläger: *Micro-RNA screening in neuroendocrine lung tumors: Differential expression of miR-29 family members.* European Respiratory Society Annual Congress 2013, Barcelona (ESP); 09/2013.
- Fabian Dominik Mairinger, Saskia Ting, Claudia Vollbrecht, Robert Fred Henry Walter, Jens Kollmeier, Sergej Grief, Thomas Mairinger, Thomas Hager, Daniel Christian Christoph, Kurt Werner Schmid, Jeremias Wohlschläger: *E3 ubiquitin-protein ligase (mouse double minute 2; Mdm2) protein expression is strongly associated with poor overall survival in malignant pleural mesothelioma.* European Respiratory Society Annual Congress 2013, Barcelona (ESP); 09/2013.
- Robert Fred Henry Walter, Robert Werner, Claudia Vollbrecht, Karl Worm, Thomas Hager, Kurt Werner Schmid, Fabian Dominik Mairinger: *Melting peak analysis by SMART-PCR: at the peak of epigenetics.* Jahrestagung der deutschen Gesellschaft für Pathologie e. V., Heidelberg (GER); 05/2013.
- Fabian Dominik Mairinger, Saskia Ting, Claudia Vollbrecht, Robert Fred Henry Walter, Jens Kollmeier, Sergej Grief, Thomas Hager, Thomas Mairinger,

Daniel Christian Christoph, Kurt Werner Schmid, Jeremias Wohlschläger: *P53 E3 ubiquitin protein ligase homolog (mouse) (MDM2)- immunoexpression could improve clinical management of patients suffering by malignant pleural mesothelioma*. Jahrestagung der deutschen Gesellschaft für Pathologie e. V., Heidelberg (GER); 05/2013.

- Claudia Vollbrecht, Katharina König, Ulrike Koitzsch, Sabine Merkelbach-Bruse, Lukas Heukamp, Reinhard Büttner, Margarete Odenthal: *Comparison of new next generation sequencing methods for diagnostic in molecular pathology*. 97. Jahrestagung der Deutschen Gesellschaft für Pathologie, Heidelberg (GER); 05/2013.
- Robert Fred Henry Walter, Fabian Dominik Mairinger, Jeremias Wohlschläger, Karl Worm, Saskia Ting, Claudia Vollbrecht, Kurt Werner Schmid, Thomas Hager: *FFPE-tissue for gene-expression analysis - a comparision of multiple reference genes*. Jahrestagung der deutschen Gesellschaft für Pathologie e. V., Heidelberg (GER); 05/2013.
- Fabian Dominik Mairinger, Robert Fred Henry Walter, Thomas Hager, Claudia Vollbrecht, Robert Werner, Daniel Christoph, Karl Worm, Saskia Ting, Kurt Werner Schmid, Jeremias Wohlschläger: *Deregulated proteasome expression in neuroendocrine lung tumors*. Jahrestagung der deutschen Gesellschaft für Pathologie e. V., Heidelberg (GER); 05/2013.
- Fabian Dominik Mairinger, Claudia Vollbrecht, Anna Streubel, Andreas Roth, Olfert Landt, Wolfram Grüning, Jens Kollmeier, Thomas Mairinger: *Methodical enhancement of molecular-pathological routine diagnosis - the next generation of sensitivity?* Jahrestagung der deutschen Gesellschaft für Pathologie e. V., Berlin (GER); 06/2012.
- Fabian Dominik Mairinger, Claudia Vollbrecht, Karl Worm, Wolfram Grüning, Thomas Mairinger, Kurt Werner Schmid: *Implementation of IMDA in pathological routine diagnosis of samples with insufficient size*. Jahrestagung der deutschen Gesellschaft für Pathologie e. V., Berlin (GER); 05/2012.
- Fabian Dominik Mairinger, Claudia Vollbrecht, Anna Streubel, Andreas Roth, Olfert Landt, Worlfram Grüning, Jens Kollmeier, Thomas Mairinger: *COLD-PCR: a powerful tool in routine-diagnostic for cost-neutral detection of minor clones using real-time PCR or sequencing*. Kongress der Deutschen

Gesellschaft für Pneumologie und Beatmungsmedizin e. V., Nürnberg (GER); 04/2012.

- Claudia Vollbrecht, Fabian Dominik Mairinger, Slave Trajanoski, Iris Halbwedl, Elvira Stacher, Theresa Maierhofer, Gabriele Michelitsch, Christian Güly, Martin Filipits, Jens Kollmeier, Thomas Mairinger, Helmut Popper: *Evaluation von EGFR-Mutationen in Adenokarzinomen der Lunge durch 454 Sequenzierung*. Kongress der Deutschen Gesellschaft für Pneumologie und Beatmungsmedizin e. V., Nürnberg (GER); 04/2012.
- Claudia Vollbrecht, Fabian Dominik Mairinger, Slave Trajanoski, Iris Halbwedl, Elvira Stacher, Theresa Maierhofer, Martin Filipits, Jens Kollmeier, Thomas Mairinger, Helmut Popper: *Very early detection of small clones harbouring EGFR mutations in NSCLC by 2nd generation sequencing*. European Respiratory Society Annual Congress 2011, Amsterdam (NED); 09/2011.
- Fabian Dominik Mairinger, Claudia Vollbrecht, Iris Halbwedl, Martina Hatz, Thomas Mairinger, Helmut Popper: *The influence of members of the folic acid metabolism on response to pemetrexed-based chemotherapy in MPM*. World Conference on Lung Cancer, Amsterdam (NED); 07/2011.
- Claudia Vollbrecht, Fabian Dominik Mairinger, Iris Halbwedl, Theresa Maierhofer, Christian Güly, Thomas Mairinger, Helmut Popper: *Very early detection of small clones harboring EGFR mutations in NSCLC by 2nd generation sequencing*. World Conference on Lung Cancer, Amsterdam (NED); 07/2011.
- Fabian Dominik Mairinger, Claudia Vollbrecht, Iris Halbwedl, Martina Hatz, Christian Güly, Franz Quehenberger, Susann Stefan-Falkenau, Jens Kollmeier, Daniel Misch, Thomas Mairinger, Helmut Popper: *Association between TS expression levels and response to pemetrexed in MPM: Fact or Fiction*. 95. Jahrestagung der Deutschen Gesellschaft für Pathologie e.V., Leipzig (GER); 06/2011.
- Fabian Dominik Mairinger, Claudia Vollbrecht, Georg Schäfer, Petra Massoner, Michael Ladurner-Renner, Thomas Mairinger, Gregor Mikuz, Helmut Klocker: *TMPRSS2-ERG gene fusion in prostatic adenocarcinoma – Analysing methods, prognosis and potential influence to clinical outcome*. 95. Jahrestagung der Deutschen Gesellschaft für Pathologie e.V., Leipzig (GER); 06/2011.

- Claudia Vollbrecht, Fabian Dominik Mairinger, Iris Halbwedl, Christian Gölly, Theresa Maierhofer, Gabriele Michelitsch, Thomas Mairinger, Jens Kollmeier, Martin Filipitz, Helmut Popper: *Preclinical identification of epidermal growth factor receptor mutations in non-small-cell lung cancer by 2nd generation sequencing*. 95. Jahrestagung der Deutschen Gesellschaft für Pathologie e.V., Leipzig (GER); 06/2011.
- Claudia Vollbrecht: [Deep-Sequencing – Die Zukunft der EGFR Analyse]. MOON Molekulare Onkologie, Fuschl (AUT); 01/2011.

Puplications in Progress

- Inga Grünwald, Claudia Vollbrecht, Jeannine Meinrath, Moritz Meyer, Lukas C. Heukamp, Uta Drebber, Alexander Quaas, Dirk Beutner, Karl-Bernd Hüttner, Eva Wardelmann, Wolfgang Hartmann, Reinhard Büttner, Margarete Odenthal, Markus Stenner: Next generation sequencing of parotid gland cancer uncovers genetic heterogeneity and indicates options for an individualized molecular tumor therapy. Clinical Cancer Research. Submitted in November 2014.
- Claudia Vollbrecht, Fabian D. Mairinger, Ulrike Koitzsch, Martin Peifer, Koenig K, Lukas C. Heukamp, Giuliano Crispazza, Laura Wilden, Karl-Anon Kreuzer, Michael Hallek, Margarete Odenthal, Carmen-Diana Herling, Reinhard Buettner. Comprehensive analysis of informative genes in chronic lymphocytic leukemia by multiplex PCR and deep sequencing. PLoS One submitted in December 2014.