

RESEARCH ARTICLE

# Comprehensive Analysis of Disease-Related Genes in Chronic Lymphocytic Leukemia by Multiplex PCR-Based Next Generation Sequencing

Claudia Vollbrecht<sup>1,2,3</sup>, Fabian Dominik Mairinger<sup>4</sup>, Ulrike Koitzsch<sup>1,2,3</sup>, Martin Peifer<sup>3,5</sup>, Katharina Koenig<sup>1,2,3</sup>, Lukas Carl Heukamp<sup>1,2,3,6</sup>, Giuliano Crispantu<sup>2,6,7</sup>, Laura Wilden<sup>2,6</sup>, Karl-Anton Kreuzer<sup>2,6</sup>, Michael Hallek<sup>2,6</sup>, Margarete Odenthal<sup>1,2,3</sup>, Carmen Diana Herling<sup>2,6</sup>\*, Reinhard Buettner<sup>1,2,3</sup>\*

**1** Institute of Pathology, University Hospital Cologne, Cologne, Germany, **2** Center for Integrated Oncology (CIO) Cologne-Bonn, University Hospital Cologne, Cologne, Germany, **3** Center of Molecular Medicine Cologne, University of Cologne, Cologne, Germany, **4** Institute of Pathology, University Hospital Essen, University of Duisburg-Essen, Essen, Germany, **5** Department of Translational Genomics, Cologne Center of Genomics, University of Cologne, Cologne, Germany, **6** Department I of Internal Medicine, University Hospital Cologne, Cologne, Germany, **7** Excellence Cluster for Cellular Stress Response and Aging-Associated Diseases (CECAD), University of Cologne, Cologne, Germany

\* These authors contributed equally to this work.

‡a Current address: Institute for Hematopathology Hamburg, Hamburg, Germany

‡b Current address: New Oncology, Cologne, Germany

\* [reinhard.buettner@uk-koeln.de](mailto:reinhard.buettner@uk-koeln.de) (RB); [carmen.herling@uk-koeln.de](mailto:carmen.herling@uk-koeln.de) (CDH)



CrossMark  
click for updates

## OPEN ACCESS

**Citation:** Vollbrecht C, Mairinger FD, Koitzsch U, Peifer M, Koenig K, Heukamp LC, et al. (2015) Comprehensive Analysis of Disease-Related Genes in Chronic Lymphocytic Leukemia by Multiplex PCR-Based Next Generation Sequencing. PLoS ONE 10(6): e0129544. doi:10.1371/journal.pone.0129544

**Academic Editor:** Christos Hatzis, Yale University, UNITED STATES

**Received:** December 16, 2014

**Accepted:** May 11, 2015

**Published:** June 8, 2015

**Copyright:** © 2015 Vollbrecht et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the paper and its Supporting Information files.

**Funding:** This study was supported by the Deutsche Forschungsgemeinschaft (KFO 286, Speaker MH) to RB, KAK, CDH, and by the German Cancer AID (Onkologisches Spitzenzentrum Köln-Bonn).

**Competing Interests:** The authors have declared that no competing interest exist.

## Abstract

### Background

High resolution molecular studies have demonstrated that the clonal acquisition of gene mutations is an important mechanism that may promote rapid disease progression and drug resistance in chronic lymphocytic leukemia (CLL). Therefore, the early and sensitive detection of such mutations is an important prerequisite for future predictive CLL diagnostics in the clinical setting.

### Material & Methods

Here, we describe a novel, target-specific next generation sequencing (NGS) approach, which combines multiplex PCR-based target enrichment and library generation with ultra-deep high-throughput parallel sequencing using a MiSeq platform. We designed a CLL specific target panel, covering hotspots or complete coding regions of 15 genes known to be recurrently mutated and/or related to B-cell receptor signaling.

### Results

High-throughput sequencing was performed using as little as 40 ng of peripheral blood B-cell DNA from 136 CLL patients and a dilution series of two *ATM*- or *TP53*-mutated cell lines, the latter of which demonstrated a limit of mutation detection below 5%. Using a

stringent functional assessment algorithm, 102 mutations in 8 genes were identified in CLL patients, including hotspot regions of *TP53*, *SF3B1*, *NOTCH1*, *ATM*, *XPO1*, *MYD88*, *DDX3X* and the B-cell receptor signaling regulator *PTPN6*. The presence of mutations was significantly associated with an advanced disease status and molecular markers of an inferior prognosis, such as an unmutated *IGHV* mutation status or positivity for *ZAP70* by flow cytometry.

## Conclusion

In summary, targeted sequencing using an amplicon based library technology allows a resource-efficient and sensitive mutation analysis for diagnostic or exploratory purposes and facilitates molecular subtyping of patient sets with adverse prognosis.

## Introduction

Chronic lymphocytic leukemia (CLL) is an incurable and common type of adult leukemia with significant variability in clinical prognosis that is hard to predict [1, 2]. The current biological understanding is that variable courses of the disease are predominantly caused by molecular inter- and inpatient heterogeneity of leukemic cells and the possibility of clonal disease evolution over time [2–4].

Recent whole-exome and genome sequencing studies have deciphered the mutational landscape in CLL and discovered a variety of somatic mutations and small indels in *NOTCH1*, *SF3B1*, and other candidate genes, which encode for putative and previously unknown drivers of CLL tumorigenesis [5–10]. Some of these mutations seem to be associated with prognosis, however, except for mutations and other genomic aberrations in the *TP53* gene, the clinical consequences to be taken in case a patient presents with one of these mutations, are not clarified [6, 7, 10–14].

Future risk assessment in CLL is now confronted with the need of prospective clinical trials, which systematically integrate mutation and traditional biomarker assessment to determine the parameters with a retained prognostic or predictive value, relevant to clinical practice. This has become of particular importance as new drugs, e.g. inhibitor to protein kinases PI3K and BTK, are entering clinical practice and conveying new mechanisms of treatment resistance compared to standard chemoimmunotherapy [15].

The aim of our study presented here was to develop a targeted genomic sequencing assay, being able to meet such diagnostic and clinical research needs in CLL. Targeted sequencing versus whole-genome or exome-wide massive parallel sequencing (i.e. next generation sequencing, NGS) offers the opportunity to assess genomic changes in areas of specific interest at a coverage as high as deemed appropriate for diagnostic reporting.

In comparison to traditional Sanger sequencing currently used for routine assessment of the *TP53* or *IGHV* genes, NGS allows multiplexing of samples and gene targets in one experimental setup. In addition, the possibility of automation for high-throughput sample processing further minimizes clinical laboratory efforts and final costs per gene and sample [16]. So far, only few studies have implemented targeted NGS technologies for mutation screening in CLL [17–20].

We here describe a multiplex PCR-based sequencing panel suitable for a high-throughput benchtop sequencer as represented by the Illumina MiSeq platform. In addition to genes confirmed to be mutated in previous CLL sequencing studies, such as *ATM*, *CD79B*, *DDX3X*,

*FBXW7*, *MYD88*, *NOTCH1*, *SF3B1*, *TP53*, *XPO* [6, 7, 10, 21–23], we chose target genes directly or indirectly involved in the B-cell receptor (BCR) signaling pathway (*BTK*, *MAPK1*, *PIK3CA*, *PIK3CD*, *PTEN*, *PTPN6*). Using a modified chemistry setup for target enrichment and library preparation in a test cohort of 136 CLL patients and two mutated cell lines, we were able to obtain a high sequencing coverage and a low limit of mutation detection. Previously known and new mutations were detected in coding or hotspot regions of the genes *ATM*, *DDX3X*, *MYD88*, *NOTCH1*, *SF3B1*, *TP53*, *XPO1* and *PTPN6* (SHP-1), and associations between mutations and adverse prognostic markers were investigated.

Overall, our targeted NGS approach resembles a sensitive and resource efficient method for simultaneous mutation analysis of multiple gene regions on a high-throughput sequencing platform and is highly suitable to future diagnostic and clinical research purposes in CLL.

## Materials and Methods

### Clinical Samples

The study was approved by the ethical commission of the medical faculty of the University of Cologne (reference no. 13–091) and an informed written consent was obtained from all patients. Between 2012 and 2013, 136 blood samples from CLL patients were collected at the University of Cologne, Germany. All cases demonstrated typical features of CLL as defined by the International Workshop on CLL [24]. Clinical and routine laboratory parameters were retrieved from medical records. CLL-related chromosomal abnormalities were assessed by inter-phase fluorescence-in-situ hybridization (FISH) 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 and the somatic mutation status of *IGHV* genes was determined as described previously [25].

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 B-cell fractions by standard column based purification (DNeasy, Qiagen, Hilden, Germany). DNA quality and quantity was assessed by gel electrophoresis.

### Library Construction and Deep Sequencing

In order to selectively amplify either hotspot or complete coding regions of the following genes *ATM*, *BTK*, *CD79B*, *DDX3X*, *FBXW7*, *MAPK1*, *MYD88*, *NOTCH1*, *PIK3CA*, *PIK3CD*, *PTEN*, *PTPN6*, *SF3B1*, *TP53* and *XPO1*, two panels containing 338 primer pairs in four separate pools were designed using the Ion AmpliSeq algorithm of Life Technologies (Table 1 and S1 Table). Amplifiable DNA was quantified by qPCR (S2 Table). Subsequently, 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). Detailed methods are available on request (<http://www.lungcancergroup.de>). Briefly, a total of 40 ng genomic B-cell DNA was amplified in four separate multiplex PCR reactions per sample. All purification and size selection steps were performed with magnetic beads (Agencourt AMPure XP, Beckman Coulter, Brea, CA, USA) and a Biomek FX<sup>P</sup> workstation (Beckman Coulter). Samples were diluted 10-fold before adenylation and adapter ligation. Finally, library quality was analyzed by microfluidic electrophoresis using the 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA) and amplicons were quantified by qPCR (S3 Table). For sequencing, samples were pooled in an equimolar ratio. 15 pM library pools

**Table 1. 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 pathway	14–16	NM_000061	5
<i>CD79B</i>	B-cell receptor signaling pathway	4–5	NM_021602	2
<i>DDX3X</i>	RNA splicing and processing	7–9, 11, 14	NM_001356	6
<i>FBXW7</i>	Protein ubiquitination	6–9	NM_033632	7
<i>MAPK1</i>	MAP kinase signaling pathway	7	NM_002745	1
<i>MYD88</i>	Toll-like receptor signaling pathway	Complete (5)	NM_002468	9
<i>NOTCH1</i>	Notch signaling pathway	Complete (34)	NM_017617	71
<i>PIK3CA</i>	B-cell receptor signaling pathway	9–11, 20–21	NM_006218	10
<i>PIK3CD</i>	B-cell receptor signaling pathway	21–24	NM_005026	7
<i>PTEN</i>	AKT-mTOR signaling pathway	5–6, 9	NM_000314	7
<i>PTPN6</i> (SHP-1)	B-cell receptor signaling pathway	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
<b>Total number of amplicons</b>				<b>338</b>

doi:10.1371/journal.pone.0129544.t001

including 1% PhiX control library were prepared for sequencing according to the MiSeq System User Guide (Illumina, San Diego, CA, US). Subsequently, sequencing was carried out on a MiSeq instrument (Illumina) using the v2 chemistry as recommended by the manufacturer.

### Estimation of Lowest Detection Rate Using Cell Line DNA Dilutions

The mantle cell lymphoma cell line, Mino (kindly provided by M. Herling, Cologne, Germany), carrying a known homozygous *TP53* mutation (c.440T>G; p.V147G; NM\_000546) [26], and the AT45RM B-cell line (kindly provided by L. Chessa, Rome, Italy) containing an heterozygous *ATM* mutation (c.7792C>T; p.R2598\*; NM\_000051) [27] were used to evaluate the limit of detection (LoD) of our NGS approach. 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, obtained from the American Type Culture Collection ATCC) harboring no known gene mutations.

### Sequencing Data Analysis

Fastq files generated by the MiSeq Reporter Software (Illumina) were analyzed with an in-house developed bioinformatics pipeline, based on the general cancer genome analysis algorithm, which was further optimized for the diagnostic workflow [28]. Briefly, adaptor sequences were first removed from raw sequencing reads. The resulting data was then aligned against NCBI build 37 (hg19) using the Burrows-Wheeler Aligner (BWA, version 0.6.1-r104) [29] with its default settings. In order to capture longer insertion and deletions we realigned unmapped reads with the BLAST-like alignment tool (BLAT) [30, 31]. For variant calling we first determined the background error rate of the sequencer using known single nucleotide polymorphisms (SNPs): Bases diverting other than the possible two variants were counted and set into the relation to the total coverage at the location of the SNP. Finally, variants were called by testing if a mutation was not compatible with the afore mentioned error rate. For this purpose, we set the significance threshold to 0.01, which leads to a slight overcalling of the sequencing data. Spurious calls were subsequently filtered out by the following strategy: Detected

variants were annotated by using the databases dbSNP (<http://www.ncbi.nlm.nih.gov/SNP/>) and the exome variant server (<http://evs.gs.washington.edu/EVS/>). Furthermore, obtained variants were analyzed for their functional impact on the protein by the MutationAssessor (<http://mutationassessor.org>; release 2) [32] and by implementation of the ANNOVAR algorithm [33], which combines the bioinformatic tools SIFT [34], PolyPhen2 [35] and the Mutation Taster [36]. Variants with an allelic frequency below 5%, synonymous and variants without functional impact were removed (Fig 1). Additional, visual analysis of called variants was performed by means of the Integrative Genomic Viewer (IGV, Broad Institute, Cambridge, MA, USA). Potential false positive variants, particularly in repetitive or highly homologous regions of the genome, variants in high background noise, as well as single strand variants, were either eliminated when they were clearly recognizable as artifacts, or were further re-assessed by Sanger sequencing.

## 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) (S4 and S5 Tables). Variants that could not be confirmed were excluded from further analysis.

## Statistical Analysis

Statistical analysis for associations with clinical and/or prognostic covariates was performed for genes with mutations in multiple samples (more than 10) with predicted impact on protein function. Consequentially, the five genes *TP53*, *SF3B1*, *ATM*, *NOTCH1* and *XPO1* were tested for associations with clinical and prognostic parameters (genomic aberration, age, gender, Binet-stage, white blood count (WBC), platelets, ZAP70 and CD38 positivity and *IGHV* mutation status) as available in our dataset. Associations between mutated patient subsets and covariates were assessed applying standard statistical tests (Fisher's exact, Pearson's chi-square, Wilcoxon Mann-Whitney rank sum test). Correlations between linear vectors were tested via Spearman's rho coefficients. Statistical calculations were computed in R version 3.1.0 (R Foundation for Statistical Computing, Vienna, Austria). All reported P-values were considered significant at  $P \leq 0.05$ .

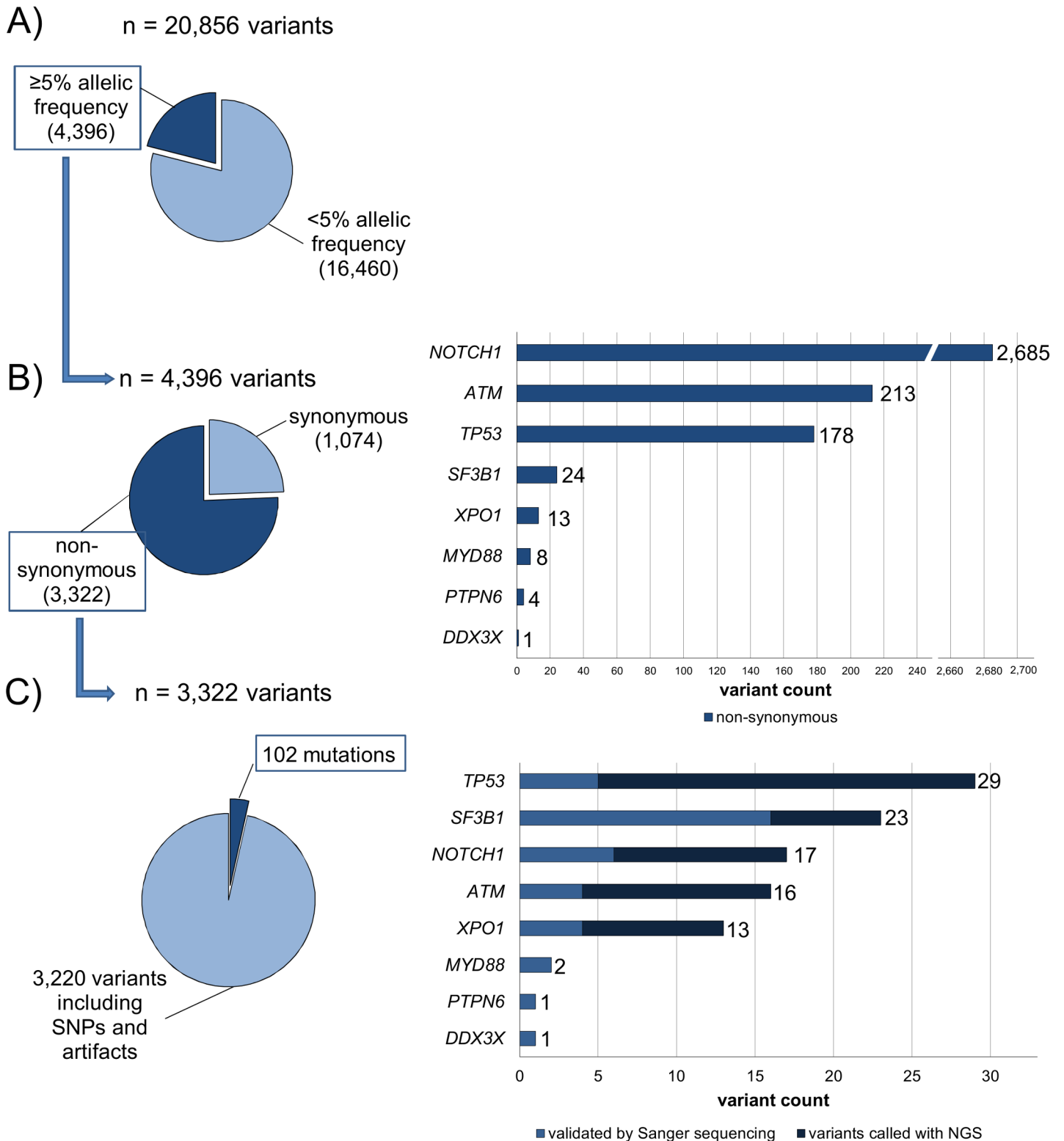
## Results

### Patients' Characteristics

We performed target-specific sequencing on purified B-cell DNA obtained from 96 men and 40 women with confirmed CLL disease, treated and followed at the University of Cologne, Germany. The majority of patients presented with previously untreated (94/70%) and/or early stage disease (Binet stage A, 73 patients/58%), at the time the sample was obtained. A subset of 41 (31%) of all patients had received a median of 2 (1–11) CLL specific treatments prior to inclusion into our study. The median time from diagnosis to sample was 41 months (0–209 months). Among patients from whom FISH analysis was available (81/60%), there was a substantial subset with deletions in chromosome 17p (13 cases/16.0%), most probably due to referral to our institution as a tertiary care center. Other prognostic markers, such as the *IGHV* mutation status, ZAP70-, CD38-surface expression, and serum thymidine kinase were distributed according to expected rates (Table 2 and S6 Table).

### High Levels of Target Coverage and Low Limit of Mutation Detection

All samples successfully completed targeted sequencing in a total of five runs, each producing an average output of  $15.37 \times 10^6$  reads and 4.7 gigabases (S7 Table). 15 genes covered by 338



**Fig 1. Algorithm of variant analysis.** A) Variants with an allelic frequency below 5% were discarded, resulting in 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/or in homopolymeric regions, and single strand variants were visually identified in the Integrative Genomic Viewer (IGV, Broad Institute) and removed. In doubtful cases, Sanger sequencing was performed to prove or disprove an alteration. Furthermore variants without functional impact on the protein determined by at least two of four applied program algorithms as described in material and method were removed. This resulted in 102 final mutations in 60 CLL specimens.

doi:10.1371/journal.pone.0129544.g001

**Table 2. Patient characteristics.**

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

doi:10.1371/journal.pone.0129544.t002

amplicons demonstrated a mean coverage per exon in a range of 0 to 7,156 reads. Only for five exons (3%; *ATM* exon 20, *NOTCH1* exon 27, *SF3B1* exon 5 and 11, *TP53* exon 11) the mean read count was less than 100, but 83% of targeted exons were covered by more than 500 reads (S1 Fig).

Two cell lines (Mino, AT45RM) with known mutations in *TP53* (exon 5) or *ATM* (exon 53) were selected as positive controls to estimate the lowest detection rate of our targeted NGS method. Analyzing fractional dilutions of mutated cell line DNA (5% to 100%), the allelic frequency of the *TP53* and *ATM* mutations detected by NGS followed a linear relationship with increasing amounts of tumor DNA (S2 Fig,  $P \leq 0.003$ ,  $\rho = 1.000$ ). We unambiguously identified the homozygous *TP53* mutation p.V147G in a background of 95% wild type DNA and the heterozygous *ATM* mutation p.R2598\* in up to 90% wild type DNA background (obtained allelic

frequency: 2% and 8%, respectively). Therefore, our NGS method obtained an adequate low LoD to uncover small subsets of mutated CLL cells due to clonal heterogeneity.

Sequencing data analysis resulted in a total of 4,396 variants after raw data alignment and first background removal (Fig 1A and 1B). Exclusion of sequencing errors, synonymous variants, variants without functional impact, and SNPs, led to 102 mutations predicted to affect protein function by at least two of four applied program algorithms as described in material and methods [32, 33]. These 102 mutations including 83 missense mutations, 12 deletions, 6 nonsense mutations, and 1 insertion were detected in eight genes and 60 out of 136 CLL samples (Fig 1C and S8 Table). In the remaining 76 patients (56%) no variants could be identified.

Fastq files are available at European Nucleotide Archive (ENA; <http://www.ebi.ac.uk/ena/data/view/PRJEB9036>).

## Multiplex PCR-based NGS Detects Variants in CLL-Related Genes

The highest frequency of mutations was obtained for *TP53* and *SF3B1* followed by *NOTCH1*, *ATM* and *XPO1*, whereas *MYD88*, *PTPN6* and *DDX3X* showed only two or one variant, respectively (Fig 1C). No mutations were found in *BTK*, *CD79B*, *FBXW7*, *MAPK1*, *PIK3CA*, *PIK3CD* and *PTEN*.

A total of 16 *ATM* mutations appeared in 15 of the 136 CLL patients (11%) and were evenly distributed over the entire gene (Fig 2). Interestingly, most of the samples harboring an *ATM* mutation showed at least one additional mutation in another gene (Fig 3). Three *ATM* mutated patients had also a deletion of *ATM* in the second allele, as assessed by FISH.

Nearly 98% (2,719) of detected *NOTCH1* variants turned out to be SNPs or sequencing errors. The remaining 17 mutations occurred in 17 patients (13%). Seven patients (5%) exhibited the previously reported p.P2514fs mutation, located in the PEST domain encoded by exon 34 [6]. Six variants were located in the NOTCH extracellular part (NEC, 6/17 variants, 35%). Interestingly, one patient exhibited a missense mutation located in exon 26 (p.F1606L) affecting the heterodimerization domain (HD) of the NOTCH1 protein.

*TP53* was the second most frequently mutated gene with 20 of 136 patients (15%) harboring a total of 29 mutations. Most of them (21/29; 72%) occurred in exon 6 to 8 and were identified to disrupt the TP53 DNA binding function. Ten *TP53* mutated patients had also a deletion of *TP53* in the second allele, verified by FISH.

In agreement to previous data from Wang *et al.* *SF3B1* showed a mutation frequency of 15% (21/136 patients) clustering in exon 14 to 16 [10]. The most frequent mutation was determined as p.K700E in exon 15 (4/23, 17%). Typical and functionally relevant exon 15 mutations in *XPO1* occurred in 13 patients (10%, p.E571I/K/Q) [6, 18].

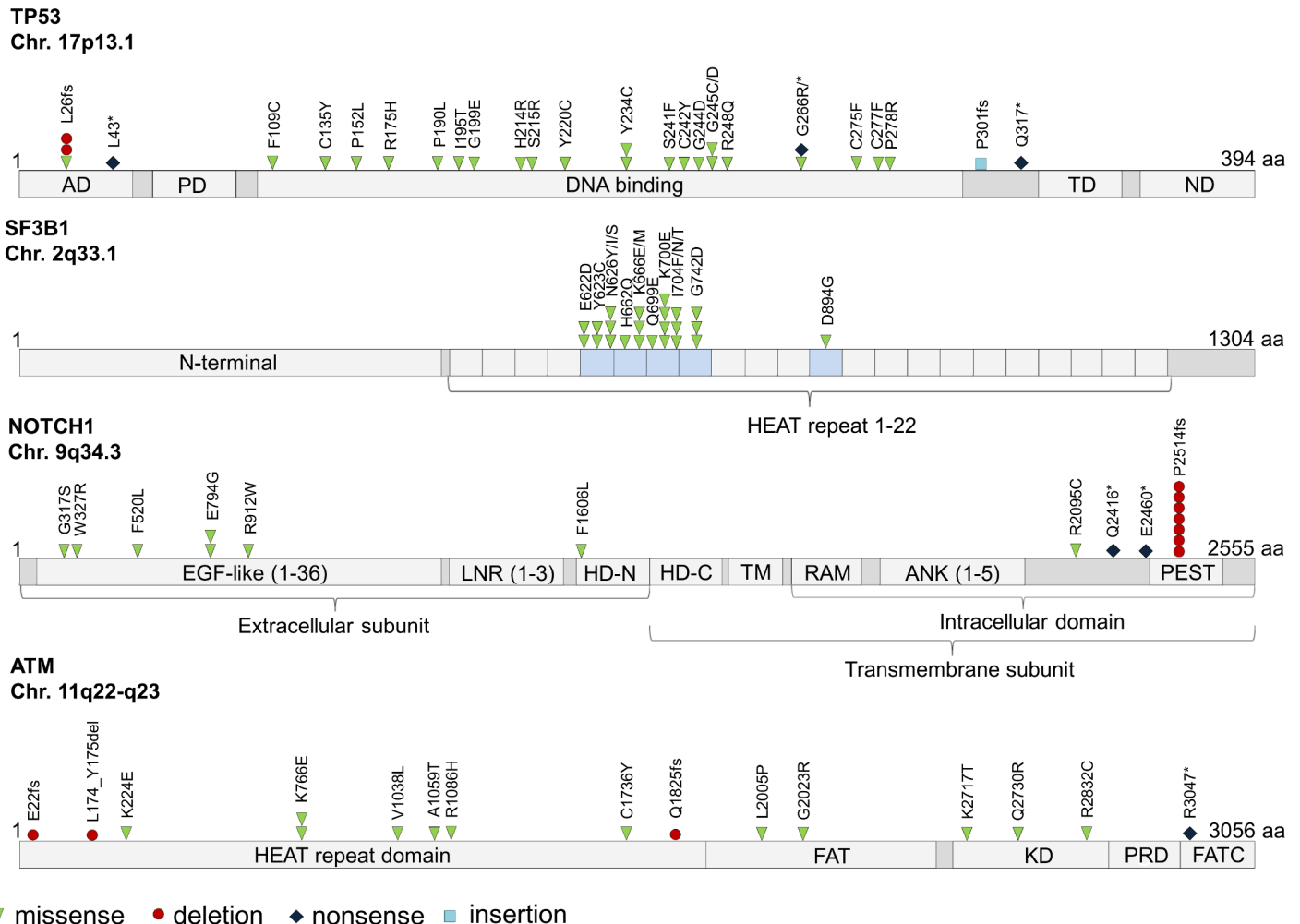
Furthermore, we found two *MYD88* mutations (p.V217F in exon 3 and p.L265P in exon 5) in two of the 136 patients (2%).

Only one patient (1%) exhibited a mutation in exon 9 of *DDX3X* (p.T275P) and one patient in exon 11 of *PTPN6* (p.V451M). The latter one was located in the highly conserved catalytic protein-tyrosine phosphatase domain of the growth factor regulator SHP-1 (*PTPN6*) and occurred with an allelic frequency of 51%.

## Variants Detected by NGS Associate with Clinical and Prognostic Parameters

Genes with mutations in at least ten samples obtained by our analysis, i.e. *TP53*, *SF3B1*, *NOTCH1*, *ATM* and *XPO1*, were tested for associations with clinical and prognostic parameters, as available in our dataset. The majority of patients without any DNA alteration detected by our sequencing panel presented significantly more frequently with early stage (Binet A/B,





**Fig 2. Alteration type, number of occurrence and location of detected mutations in *TP53*, *SF3B1*, *NOTCH1* and *ATM* are shown.** *TP53*: 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*: The majority of *SF3B1* alterations were clustered in the region encoding the highly conserved HEAT (huntingtin, elongation factor 3, protein phosphatase 2A, target of rapamycin 1) repeats 5–8. Only one alteration occurred in the N-terminal (amino acids 1–450), domain, which is an important docking or binding domain for numerous splicing factor partners like U2AF1/2, and cyclin E. *NOTCH1*: (EGF)-like epidermal growth factor 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); *ATM*: 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); aa amino acid

doi:10.1371/journal.pone.0129544.g002

47/53, 89%;  $P = 0.03$ ) or previously untreated CLL (42/53, 79%;  $P = 0.001$ ) at the time of sampling. Presence of mutations in *NOTCH1*, *SF3B1*, *TP53* and *XPO1* was associated with at least one unfavorable prognostic marker such as an unmutated *IGHV* gene status or positivity for ZAP70 or CD38 (Table 3). *SF3B1* mutated patients were significantly more frequent of male gender (20/96 males vs. 1/40 females,  $P = 0.008$ ) and *IGHV* unmutated (15/21, 72% vs. 5/21, 24%,  $P = 0.03$ ) than *SF3B1* wild type cases (S3 Fig). In the untreated patient cohort, the presence of *SF3B1* mutations significantly correlated with positivity for CD38 assessed by flow cytometry ( $P = 0.03$ ) (S4 Fig).

Similarly, an unmutated *IGHV* status occurred more frequently in *TP53* mutants (15/20, 75% vs. 4/20, 20%,  $P = 0.04$ ) and in untreated patients with a *NOTCH1* PEST domain mutation (4/4, 100% vs. 0/4, 0%,  $P = 0.04$ ), compared to their wild type counterparts.

	<i>TP53</i>	<i>SF3B1</i>	<i>NOTCH1</i>	<i>ATM</i>	<i>XPO1</i>	<i>PTPN6</i>	<i>DDX3X</i>	<i>MYD88</i>
P111	1	1	1					
P091	1	1			1			
P086	1	1						
P101	1	1						
P110	1	1						
P094	1		1		1			
P119	1			1			1	
P058	1			1				
P085	1			1				
P105	1			1				
P115	1			1				
P032	1							
P082	1							
P117	1							
P040	2							
P095	2							
P112	2							
P132	3	1						
P064	3							
P067	3							
P071		1	1					
P079		1		1				
P122		1		1				
P057		1			1			
P108		1						
P011		1						
P015		1						
P021		1						
P044		1						
P050		1						
P056		1						
P084		1						
P088		1						
P035		2						
P100		2						
P074			1	1				
P093			1	1				
P134			1	1				
P114			1	2				
P001			1		1			
P028			1		1			
P072			1		1			
P080			1		1			
P016			1			1		
P039			1					
P046			1					
P061			1					
P062			1					
P097			1					
P048				1	1			
P031				1				
P042				1				
P076				1				
P052					1			
P037					1			
P068					1			
P096					1			
P130					1			
P078								1
P103	60							1

**Fig 3. Genetic profile of 60 CLL samples carrying gene mutations determined by NGS.** Each row represents the variants of one patient, each column summarizes the mutations occurring in one specific gene. Per each gene the number of mutations is given per patient. Dark blue samples indicate patients with aberration on chromosome 11 (del11q) for *ATM* mutated cases or on chromosome 17 (del17p) for *TP53* mutated cases, determined by FISH.

doi:10.1371/journal.pone.0129544.g003

Patients with mutations in *XPO1* exhibited significantly increased WBC, possibly reflecting the proliferative capacity of CLL cells, compared to patients without *XPO1* mutations (mean: 134 vs. 65 x10<sup>9</sup>/L, *P*<0.001). In treated patients, positivity for ZAP70 was significantly overrepresented in patients with a mutated *XPO1* gene (*P* = 0.02).

Only *TP53* mutations were found to be enriched in treated versus untreated patients (12/41, 29% vs. 8/95, 8%, *P*<0.001), indicating a possible selection of this genetic alteration due to prior treatments. Further, patients with *TP53* mutations exhibited significantly more frequent deletions in the second allele on chromosome 17p, resulting in a complete disruption of the TP53 protein function (*P*<0.001) (Table 3).

### Discussion

CLL is a socioeconomically relevant disease of older adults with a currently rapidly changing field of new drugs entering clinical practice and an evolving discovery of genomic mutations with major clinical relevance [3, 15]. Future diagnostics and research in CLL and cancer in general will require the implementation of mutational screening assays, which are resource efficient, sensitive, and rapidly adaptable to clinical and scientific needs.

Here, we present a targeted sequencing assay, which combines library and sequencing chemistry beyond the boundaries of manufacturers. For this approach we optimized target amplification, sequencing output and data analysis for routine application. We performed a multiplex PCR-based library amplification combining an Ion AmpliSeq primer design (Life Technologies) with a modified library preparation chemistry that allows sequencing on an Illumina instrument. Our assay targeted the complete coding regions of five most frequently mutated “CLL-genes” (*ATM*, *MYD88*, *NOTCH1*, *SF3B1* and *TP53*) [6, 7, 10] and additionally ten genes with a more exploratory driven interest, e.g. the kinase domains of the drug targets *BTK* and *PIK3CD* [15]. Our method is performable within three days from sample DNA extraction to data analysis and offers suitable flexibility by the replacement or addition of target regions during primer design.

**Table 3. Statistical correlations between gene mutation status and clinical and biological parameters.**

	Mutated Gene	Clinical/Prognostic Parameter	P-Value
Overall Cohort	<i>SF3B1</i>	Male sex	0.00759
	<i>SF3B1</i>	Unmutated <i>IGHV</i>	0.03029
	<i>SF3B1</i>	Decreased platelet count	0.02467
	<i>TP53</i>	Binet stage	0.00724
	<i>TP53</i>	Chromosome 17p deletion	0.00000
	<i>TP53</i>	Unmutated <i>IGHV</i>	0.03842
	<i>TP53</i>	Treatment status	0.00071
	<i>XPO1</i>	Unmutated <i>IGHV</i>	0.00015
	<i>XPO1</i>	Increased White blood count (WBC)	0.00070
Untreated Cohort	<i>NOTCH1</i> (p.P2514fs)	Unmutated <i>IGHV</i>	0.03646
	<i>SF3B1</i>	CD38 positivity	0.03096
Treated Cohort	<i>XPO1</i>	ZAP70 positivity	0.02362

doi:10.1371/journal.pone.0129544.t003

While targeted NGS offers the advantage to assess multiplexed samples and genes in one experimental setup—thus being relatively cost-efficient compared to Sanger sequencing—, one disadvantage is that the ability to detect mutations in a distinct gene depends on the achieved coverage/depth of reads in this specific region. Coverage and sequencing depth can vary substantially, depending on the gene region itself (e.g. GC rich, homopolymers, etc.), enzyme chemistry and sequencing platform. With our technology we were able to cover 83% of targeted exons with a minimum coverage of 500 reads. Only for five exons the mean number of reads was below 100, a threshold under which we would consider calling of mutations not possible for diagnostic purposes and apply either repeated NGS or Sanger sequencing.

Most studies on targeted CLL sequencing published so far have implemented NGS methods without giving details on the performance of the technology. Sutton *et al.* are the first to report details on assay quality and analytical requirements of targeted NGS results using the HaloPlex probe technology from Agilent in CLL [20]. This technology offers the advantage of target specific probe hybridization without PCR amplification. They investigated a set of 188 patients with poor prognostic features for gene alterations in *ATM*, *BIRC3*, *KLHL6*, *MYD88*, *NOTCH1*, *POT1*, *SF3B1*, *TP53*, and *XPO1*. For final analysis, they only included patients, for which they obtained at least 100 reads for 80% of the targeted bases (96% of their samples). Thus, their assay achieved reasonable quality results in terms of coverage and uniformity of read depth, comparable to ours. As discussed by the authors, cutoffs currently chosen for quality parameters to evaluate targeted NGS data are more or less arbitrary. Further studies are needed to standardize and harmonize such parameters for comparability of different datasets and clinical implementation.

Another difficulty in studies implementing high-throughput NGS in a clinical setting is the necessity to distinguish tumor acquired mutations from germline or non-tumor-specific variants. Non-tumor tissue biopsies for DNA comparison are difficult to obtain in clinical routine and additional sequencing is cost intensive. Most targeted sequencing studies in CLL published so far filtered their data according to variant information available from public databases such as dbSNP (<http://www.ncbi.nlm.nih.gov/SNP/>) or COSMIC (<http://cancer.sanger.ac.uk/cosmic>). Similarly, we applied a systematic annotation of variants detected by our assay for potential SNPs listed in dbSNP and the Exome Variant server (EVS). However, not all genetic variants described in these databases are listed as SNPs of healthy individuals in the general population, for example the activating mutation p.L858R in exon 21 of the EGF receptor gene (rs121434568). It is also conceivable, that low frequency SNPs with functional relevance on a protein level might be relevant for the CLL pathophysiology. Therefore—instead of eliminating all variants found in dbSNP or EVS—we prioritized our data for variants with functional impact on a protein level as assessed by two of four implemented program tools (MutationTaster, MutationAssessor, SIFT, Polyphen) [32, 34–36]. Only SNPs reported as benign were eliminated from the analysis. Some of the mutations detected in our study occur with approximately 100% or 50% allele frequency but show additionally a deletion on the second allele. Therefore, these mutations cannot necessarily be determined as SNPs. Hence, it is not excluded that our final list of mutations still contains variants which are not CLL-specific but so far unproven germline SNPs. For example, we describe a p.V451M mutation in the catalytic phosphatase domain of SHP-1 (*PTPN6*) which was detected at an allelic frequency of 51% and predicted to impair protein function [5, 6, 18, 37, 38]. This variant has a dbSNP entry (rs62621988) but no information about the allele origin or the clinical significance. It was described at a rare frequency of 0.0005 in the 1000 genomes project (<http://www.1000genomes.org>). Although allelic frequency and location in a highly conserved region point to a potential SNP, the functional relevance of this alteration is still interesting to report. SHP-1 (*PTPN6*) is a known repressor of

BCR signaling. Therefore inactivating mutations could constitutively activate BCR signaling in CLL cells and therefore influence disease development and outcome [39, 40].

In general, the mutation rates obtained by our NGS assay for genes known to be mutated in CLL (*TP53*, *SF3B1*, *ATM*, *NOTCH1*, *XPO1*, *MYD88*, *DDX3X*) are comparable to other studies [6–8, 10, 13, 17–19]. Most of the mutations detected are located at typical hotspot locations, such as the p.K700E and p.G742D mutations in *SF3B1* (predominantly found in male patients), the p.L256P mutation in *MYD88*, the p.T275P mutation in exon 9 of *DDX3X*, or the p.E571K mutation in *XPO1* [6, 7, 10, 18, 41]. One CLL case exhibited a *MYD88* p.V217F mutation, an alteration previously described in diffuse large B-cell lymphoma (DLBCL) by Ngo *et al.* [41].

Targeted NGS studies in CLL published to date frequently omitted sequencing of *ATM*, due to the lack of hotspots regions in this relatively large gene and size limitations of their assay [13, 17–19]. In our hands, *ATM* sequencing within a larger gene panel was feasible and mutations detected at a rate of 11%, comparable to the 12% rate reported by Austen and colleagues [22]. Interestingly, we found *ATM* variants occurring more frequently in combination with other variants, in particular with *NOTCH1* or *TP53* (Fig 3), an aspect also confirmed by the study of Sutton *et al.*, described above. For *NOTCH1* our data analysis obtained a high incidence of non-functional variants, which might be attributable to technical issues during target enrichment and/or sequencing, e.g. by polymerase reading errors in GC- or homopolymeric regions. Beside variants in the EGF-like and PEST domains, we detected a gain-of-function mutation (p.F1606L) in the HD domain of the *NOTCH1* extracellular subunit. Only the p.P2514 frameshift deletion in the PEST domain revealed a significant correlation with an unmutated *IGHV* status (Table 3), indicating that these mutations are preferentially enriched in CLL patients with adverse prognosis.

One advantage of NGS technologies for mutation analysis is that the achievement of a high sequencing coverage allows the more sensitive determination of small subclones carrying mutations. It has been demonstrated that such subclones can evolve over time and drive CLL progression and transformation [4, 42]. In our assay, the allelic frequency of mutations ranged from 5% to 100%. The smallest clonal fraction was determined for two *NOTCH1* mutations (p.W327R; 4,765 reads and p.F1606L; 166 reads) and one *TP53* mutation (p.Y234C; 2,635 reads). Furthermore, sequencing of mutated cell lines allowed us to estimate the low LoD of our method, which depicted 5% of *TP53* mutated cells diluted in wild type background at a 2% allelic frequency. Thus, our method would be clearly able to pick up small mutated subclones in CLL, presumably beyond the detection limit of Sanger sequencing. This compares to other NGS studies reporting a sensitivity or LoD of targeted NGS at 2–3% [17, 18].

In conclusion, we have developed a targeted NGS panel and high-throughput assay for mutation analysis in CLL, which is resource-efficient and highly sensitive for the detection of low frequency alleles and fast enough to be applicable to clinical decision processes. Applying a systematic functional data assessment, we found various alterations including known hotspot mutations and one interesting *PTPN6* mutation in the BCR, without the need of non-tumor DNA sequencing. Our NGS methodology can be easily translated to molecular diagnostics of other types of cancer and may pave the way for a fast-throughput combination of morphological and molecular diagnostics in hematologic and non-hematologic malignancies.

## Supporting Information

**S1 Fig. Median and mean read count per exon for 167 exons of the 15 genes are shown.** Results reflect data from five NGS runs.  
(TIF)

**S2 Fig. Linear relationship of mutation rate and allele frequency detected by NGS.** Sequencing of two dilution series of cell line DNA with a known A) homozygous *TP53* c.440T>G; p.V147G mutation (Mino cell line) and B) heterozygous *ATM* c.7792C>T; p.R2598\* mutation (AT45RM cell line) demonstrated a linear relationship of the fractional dilution rate and the mutation allele frequency obtained by NGS. Further, the data point to the detection limit achieved by our NGS approach 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.

(TIF)

**S3 Fig. Associations of *SF3B1* and *TP53* mutations with clinical and prognostic parameters are shown.** A) *SF3B1* mutated patients were mainly *IGHV* unmutated, in contrast to *SF3B1* wild type patients that showed a normal mutated *IGHV* status ( $P = 0.03$ ). B) *SF3B1* mutated patients were significantly more of male gender ( $P = 0.008$ ). C) *TP53* mutations were found particularly more frequent in intermediate and advanced stage with a need for treatment (Binet stage B/C) compared with patients in an early stage (Binet stage A) ( $P = 0.008$ ). D and E) *TP53* mutations were also frequently more detected in treated patients ( $P < 0.001$ ) and in patients with genomic aberrations on chromosome 17 (del17p) ( $P < 0.001$ ). ND not determined.

(TIFF)

**S4 Fig. The presence of functional relevant mutations in *SF3B1* and *XPO1* was associated with the unfavorable prognostic marker like positivity for *ZAP70* or *CD38*.** A) *SF3B1* mutated untreated patients showed an increased *CD38* expression ( $p < 0.04$ ). B and C) Patients harboring *XPO1* mutations showed an increased WBC ( $p < 0.001$ ) and treated patients presented a higher *ZAP70* expression compared to their wild type counterparts ( $p < 0.03$ ).

(TIF)

**S1 Table. Target regions of the CLL panel 1 and 2.**

(DOCX)

**S2 Table. Parameters for the *HFE* qPCR.** DNA quantification was done using native DNA from HEK-293 (human embryonic kidney) cells without known gene mutations; all samples were measured in duplicates.

(DOCX)

**S3 Table. PCR parameters for quantification of the constructed libraries by qPCR.** A) PCR components; B) PCR conditions. Amplicon library quantification was performed with 5-fold dilutions of PhiX Control V3 (Illumina, San Diego, CA, USA) in a range from 0.064 up to 40 pM as reference standard. The library samples were diluted 1:4000 and measured in duplicates.

(DOCX)

**S4 Table. Primer used for Sanger sequencing validation.**

(DOCX)

**S5 Table. Parameters for PCR and Sanger sequencing.** A1 and 2) Components and conditions for amplification of target regions by PCR; B1 and 2) Components and conditions for Sanger sequencing reaction.

(DOCX)

**S6 Table. Clinical information of the patients analyzed in this study.**

(XLS)

**S7 Table. Run parameters from the five MiSeq sequencing runs.**

(DOCX)

**S8 Table. Complete list of detected mutations.** Bold mutations were confirmed by Sanger sequencing. fs frame shift; \* stop gained; Freq frequency; Cov coverage; dbSNP single nucleotide variants database.

(DOCX)

## Acknowledgments

We greatly appreciate the help of Thomas Landwehr (CLL Biobank, Department I of Internal Medicine, University of Cologne, Cologne, Germany) for sample collection and the support of Jasmin Bahlo from the German CLL Study Group (GCLLSG) in statistical data interpretation. We thank Christian Reinhardt (Department I of Internal Medicine, Center for Integrated Oncology (CIO) Cologne-Bonn, and Excellence Cluster for Cellular Stress Response and Aging-Associated Diseases (CECAD), University of Cologne, Cologne, Germany) for many helpful discussions.

## Author Contributions

Conceived and designed the experiments: CV CDH MO KK. Performed the experiments: CV LW UK. Analyzed the data: CV FDM GC CDH MO. Contributed reagents/materials/analysis tools: MO CDH KAK MH MP LCH. Wrote the paper: CV MO CDH RB.

## References

1. Dighiero G, Hamblin TJ. Chronic lymphocytic leukaemia. *Lancet*. 2008; 371(9617):1017–29. Epub 2008/03/25. doi: [10.1016/S0140-6736\(08\)60456-0](https://doi.org/10.1016/S0140-6736(08)60456-0) PMID: [18358929](https://pubmed.ncbi.nlm.nih.gov/18358929/).
2. Cramer P, Hallek M. Prognostic factors in chronic lymphocytic leukemia-what do we need to know? *Nat Rev Clin Oncol*. 2011; 8(1):38–47. Epub 2010/10/20. doi: [10.1038/nrclinonc.2010.167](https://doi.org/10.1038/nrclinonc.2010.167) PMID: [20956983](https://pubmed.ncbi.nlm.nih.gov/20956983/).
3. Gruber M, Wu CJ. Evolving understanding of the CLL genome. *Semin Hematol*. 2014; 51(3):177–87. Epub 2014/07/23. doi: [10.1053/j.seminhematol.2014.05.004](https://doi.org/10.1053/j.seminhematol.2014.05.004) PMID: [25048782](https://pubmed.ncbi.nlm.nih.gov/25048782/); PubMed Central PMCID: PMC4107366.
4. Landau DA, Carter SL, Stojanov P, McKenna A, Stevenson K, Lawrence MS, et al. Evolution and impact of subclonal mutations in chronic lymphocytic leukemia. *Cell*. 2013; 152(4):714–26. Epub 2013/02/19. doi: [10.1016/j.cell.2013.01.019](https://doi.org/10.1016/j.cell.2013.01.019) PMID: [23415222](https://pubmed.ncbi.nlm.nih.gov/23415222/); PubMed Central PMCID: PMC3575604.
5. Fabbri G, Rasi S, Rossi D, Trifonov V, Khiabani H, Ma J, et al. Analysis of the chronic lymphocytic leukemia coding genome: role of NOTCH1 mutational activation. *The Journal of experimental medicine*. 2011; 208(7):1389–401. Epub 2011/06/15. doi: [10.1084/jem.20110921](https://doi.org/10.1084/jem.20110921) PMID: [21670202](https://pubmed.ncbi.nlm.nih.gov/21670202/); PubMed Central PMCID: PMC3135373.
6. Puente XS, Pinyol M, Quesada V, Conde L, Ordonez GR, Villamor N, et al. Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia. *Nature*. 2011; 475(7354):101–5. Epub 2011/06/07. doi: [10.1038/nature10113](https://doi.org/10.1038/nature10113) PMID: [21642962](https://pubmed.ncbi.nlm.nih.gov/21642962/); PubMed Central PMCID: PMC3322590.
7. Quesada V, Conde L, Villamor N, Ordonez GR, Jares P, Bassaganyas L, et al. Exome sequencing identifies recurrent mutations of the splicing factor SF3B1 gene in chronic lymphocytic leukemia. *Nature genetics*. 2012; 44(1):47–52. Epub 2011/12/14. doi: [10.1038/ng.1032](https://doi.org/10.1038/ng.1032) PMID: [22158541](https://pubmed.ncbi.nlm.nih.gov/22158541/).
8. Wan Y, Wu CJ. SF3B1 mutations in chronic lymphocytic leukemia. *Blood*. 2013; 121(23):4627–34. Epub 2013/04/10. doi: [10.1182/blood-2013-02-427641](https://doi.org/10.1182/blood-2013-02-427641) PMID: [23568491](https://pubmed.ncbi.nlm.nih.gov/23568491/); PubMed Central PMCID: PMC3674664.
9. Villamor N, Lopez-Guillermo A, Lopez-Otin C, Campo E. Next-generation sequencing in chronic lymphocytic leukemia. *Semin Hematol*. 2013; 50(4):286–95. Epub 2013/11/20. doi: [10.1053/j.seminhematol.2013.09.005](https://doi.org/10.1053/j.seminhematol.2013.09.005) PMID: [24246696](https://pubmed.ncbi.nlm.nih.gov/24246696/).
10. Wang L, Lawrence MS, Wan Y, Stojanov P, Sougnez C, Stevenson K, et al. SF3B1 and other novel cancer genes in chronic lymphocytic leukemia. *The New England journal of medicine*. 2011; 365(26):2497–506. Epub 2011/12/14. doi: [10.1056/NEJMoa1109016](https://doi.org/10.1056/NEJMoa1109016) PMID: [22150006](https://pubmed.ncbi.nlm.nih.gov/22150006/); PubMed Central PMCID: PMC3685413.

11. Rossi D, Rasi S, Fabbri G, Spina V, Fangazio M, Forconi F, et al. Mutations of NOTCH1 are an independent predictor of survival in chronic lymphocytic leukemia. *Blood*. 2012; 119(2):521–9. Epub 2011/11/15. doi: [10.1182/blood-2011-09-379966](https://doi.org/10.1182/blood-2011-09-379966) PMID: [22077063](https://pubmed.ncbi.nlm.nih.gov/22077063/); PubMed Central PMCID: PMC3257017.
12. Villamor N, Conde L, Martinez-Trillos A, Cazorla M, Navarro A, Bea S, et al. NOTCH1 mutations identify a genetic subgroup of chronic lymphocytic leukemia patients with high risk of transformation and poor outcome. *Leukemia*. 2013; 27(5):1100–6. Epub 2013/01/09. doi: [10.1038/leu.2012.357](https://doi.org/10.1038/leu.2012.357) PMID: [23295735](https://pubmed.ncbi.nlm.nih.gov/23295735/).
13. Cortese D, Sutton LA, Cahill N, Smedby KE, Geisler C, Gunnarsson R, et al. On the way towards a 'CLL prognostic index': focus on TP53, BIRC3, SF3B1, NOTCH1 and MYD88 in a population-based cohort. *Leukemia*. 2014; 28(3):710–3. Epub 2013/11/13. doi: [10.1038/leu.2013.333](https://doi.org/10.1038/leu.2013.333) PMID: [24217197](https://pubmed.ncbi.nlm.nih.gov/24217197/).
14. Rossi D, Fangazio M, Rasi S, Vaisitti T, Monti S, Cresta S, et al. Disruption of BIRC3 associates with fludarabine chemorefractoriness in TP53 wild-type chronic lymphocytic leukemia. *Blood*. 2012; 119(12):2854–62. Epub 2012/02/07. doi: [10.1182/blood-2011-12-395673](https://doi.org/10.1182/blood-2011-12-395673) PMID: [22308293](https://pubmed.ncbi.nlm.nih.gov/22308293/).
15. Woyach JA, Furman RR, Liu TM, Ozer HG, Zapotka M, Ruppert AS, et al. Resistance mechanisms for the Bruton's tyrosine kinase inhibitor ibrutinib. *The New England journal of medicine*. 2014; 370(24):2286–94. Epub 2014/05/30. doi: [10.1056/NEJMoa1400029](https://doi.org/10.1056/NEJMoa1400029) PMID: [24869598](https://pubmed.ncbi.nlm.nih.gov/24869598/).
16. Chang F, Li MM. Clinical application of amplicon-based next-generation sequencing in cancer. *Cancer Genet*. 2013; 206(12):413–9. Epub 2013/12/18. doi: [10.1016/j.cancergen.2013.10.003](https://doi.org/10.1016/j.cancergen.2013.10.003) PMID: [24332266](https://pubmed.ncbi.nlm.nih.gov/24332266/).
17. Jethwa A, Hullein J, Stolz T, Blume C, Sellner L, Jauch A, et al. Targeted resequencing for analysis of clonal composition of recurrent gene mutations in chronic lymphocytic leukaemia. *British journal of haematology*. 2013; 163(4):496–500. Epub 2013/09/17. doi: [10.1111/bjh.12539](https://doi.org/10.1111/bjh.12539) PMID: [24032483](https://pubmed.ncbi.nlm.nih.gov/24032483/).
18. Jeromin S, Weissmann S, Haferlach C, Dicker F, Bayer K, Grossmann V, et al. SF3B1 mutations correlated to cytogenetics and mutations in NOTCH1, FBXW7, MYD88, XPO1 and TP53 in 1160 untreated CLL patients. *Leukemia*. 2014; 28(1):108–17. Epub 2013/10/12. doi: [10.1038/leu.2013.263](https://doi.org/10.1038/leu.2013.263) PMID: [24113472](https://pubmed.ncbi.nlm.nih.gov/24113472/).
19. Baliakas P, Hadzidimitriou A, Sutton LA, Rossi D, Minga E, Villamor N, et al. Recurrent mutations refine prognosis in chronic lymphocytic leukemia. *Leukemia*. 2014. Epub 2014/06/20. doi: [10.1038/leu.2014.196](https://doi.org/10.1038/leu.2014.196) PMID: [24943832](https://pubmed.ncbi.nlm.nih.gov/24943832/).
20. Sutton LA, Ljungstrom V, Mansouri L, Young E, Cortese D, Navrkalova V, et al. Targeted next-generation sequencing in chronic lymphocytic leukemia: a high-throughput yet tailored approach will facilitate implementation in a clinical setting. *Haematologica*. 2015; 100(3):370–6. Epub 2014/12/07. doi: [10.3324/haematol.2014.109777](https://doi.org/10.3324/haematol.2014.109777) PMID: [25480502](https://pubmed.ncbi.nlm.nih.gov/25480502/); PubMed Central PMCID: PMC4349276.
21. Thompson AA, Talley JA, Do HN, Kagan HL, Kunkel L, Berenson J, et al. Aberrations of the B-cell receptor B29 (CD79b) gene in chronic lymphocytic leukemia. *Blood*. 1997; 90(4):1387–94. Epub 1997/08/15. PMID: [9269755](https://pubmed.ncbi.nlm.nih.gov/9269755/).
22. Austen B, Powell JE, Alvi A, Edwards I, Hooper L, Starczynski J, et al. 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*. 2005; 106(9):3175–82. Epub 2005/07/15. doi: [10.1182/blood-2004-11-4516](https://doi.org/10.1182/blood-2004-11-4516) PMID: [16014569](https://pubmed.ncbi.nlm.nih.gov/16014569/).
23. Zenz T, Eichhorst B, Busch R, Denzel T, Habe S, Winkler D, et al. TP53 mutation and survival in chronic lymphocytic leukemia. *J Clin Oncol*. 2010; 28(29):4473–9. Epub 2010/08/11. doi: [10.1200/JCO.2009.27.8762](https://doi.org/10.1200/JCO.2009.27.8762) PMID: [20697090](https://pubmed.ncbi.nlm.nih.gov/20697090/).
24. Hallek M, Cheson BD, Catovsky D, Caligaris-Cappio F, Dighiero G, Dohner H, et al. Guidelines for the diagnosis and treatment of chronic lymphocytic leukemia: a report from the International Workshop on Chronic Lymphocytic Leukemia up dating the National Cancer Institute-Working Group 1996 guidelines. *Blood*. 1996; 111:5446–56.
25. Schweighofer CD, Huh YO, Luthra R, Sargent RL, Ketterling RP, Knudson RA, et al. The B cell antigen receptor in atypical chronic lymphocytic leukemia with t(14;19)(q32;q13) demonstrates remarkable stereotypy. *International journal of cancer Journal international du cancer*. 2011; 128(11):2759–64. Epub 2010/08/18. doi: [10.1002/ijc.25605](https://doi.org/10.1002/ijc.25605) PMID: [20715110](https://pubmed.ncbi.nlm.nih.gov/20715110/).
26. Lai R, McDonnell TJ, O'Connor SL, Medeiros LJ, Oudat R, Keating M, et al. Establishment and characterization of a new mantle cell lymphoma cell line, Mino. *Leukemia research*. 2002; 26(9):849–55. Epub 2002/07/20. PMID: [12127561](https://pubmed.ncbi.nlm.nih.gov/12127561/).
27. Magliozzi M, Piane M, Torrente I, Sinibaldi L, Rizzo G, Savio C, et al. DHPLC screening of ATM gene in Italian patients affected by ataxia-telangiectasia: fourteen novel ATM mutations. *Disease markers*. 2006; 22(4):257–64. Epub 2006/11/25. PMID: [17124347](https://pubmed.ncbi.nlm.nih.gov/17124347/); PubMed Central PMCID: PMC3862285.
28. Peifer M, Fernandez-Cuesta L, Sos ML, George J, Seidel D, Kasper LH, et al. Integrative genome analyses identify key somatic driver mutations of small-cell lung cancer. *Nature genetics*. 2012; 44(10):1104–10. Epub 2012/09/04. doi: [10.1038/ng.2396](https://doi.org/10.1038/ng.2396) PMID: [22941188](https://pubmed.ncbi.nlm.nih.gov/22941188/).



29. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009; 25(14):1754–60. Epub 2009/05/20. doi: [10.1093/bioinformatics/btp324](https://doi.org/10.1093/bioinformatics/btp324) PMID: [19451168](https://pubmed.ncbi.nlm.nih.gov/19451168/); PubMed Central PMCID: PMC2705234.
30. Kent WJ. BLAT—the BLAST-like alignment tool. *Genome Res*. 2002; 12(4):656–64. Epub 2002/04/05. doi: [10.1101/gr.229202](https://doi.org/10.1101/gr.229202) Article published online before March 2002. PMID: [11932250](https://pubmed.ncbi.nlm.nih.gov/11932250/); PubMed Central PMCID: PMC187518.
31. Bhagwat M, Young L, Robison RR. Using BLAT to find sequence similarity in closely related genomes. *Current protocols in bioinformatics / editorial board, Andreas D Baxevanis [et al]*. 2012;Chapter 10: Unit10 8. Epub 2012/03/06. doi: [10.1002/0471250953.bi1008s37](https://doi.org/10.1002/0471250953.bi1008s37) PMID: [22389010](https://pubmed.ncbi.nlm.nih.gov/22389010/); PubMed Central PMCID: PMC4101998.
32. Reva B, Antipin Y, Sander C. Predicting the functional impact of protein mutations: application to cancer genomics. *Nucleic acids research*. 2011; 39(17):e118. Epub 2011/07/06. doi: [10.1093/nar/gkr407](https://doi.org/10.1093/nar/gkr407) PMID: [21727090](https://pubmed.ncbi.nlm.nih.gov/21727090/); PubMed Central PMCID: PMC3177186.
33. Landau DA, Wu CJ. Chronic lymphocytic leukemia: molecular heterogeneity revealed by high-throughput genomics. *Genome Med*. 2013; 5(5):47. Epub 2013/06/05. doi: [10.1186/gm451](https://doi.org/10.1186/gm451) PMID: [23731665](https://pubmed.ncbi.nlm.nih.gov/23731665/); PubMed Central PMCID: PMC3706960.
34. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc*. 2009; 4(7):1073–81. Epub 2009/06/30. doi: [10.1038/nprot.2009.86](https://doi.org/10.1038/nprot.2009.86) PMID: [19561590](https://pubmed.ncbi.nlm.nih.gov/19561590/).
35. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. *Nat Methods*. 2010; 7(4):248–9. Epub 2010/04/01. doi: [10.1038/nmeth0410-248](https://doi.org/10.1038/nmeth0410-248) PMID: [20354512](https://pubmed.ncbi.nlm.nih.gov/20354512/); PubMed Central PMCID: PMC2855889.
36. Schwarz JM, Rodelsperger C, Schuelke M, Seelow D. MutationTaster evaluates disease-causing potential of sequence alterations. *Nat Methods*. 2010; 7(8):575–6. Epub 2010/08/03. doi: [10.1038/nmeth0810-575](https://doi.org/10.1038/nmeth0810-575) PMID: [20676075](https://pubmed.ncbi.nlm.nih.gov/20676075/).
37. Malecki MJ, Sanchez-Irizarry C, Mitchell JL, Histen G, Xu ML, Aster JC, et al. Leukemia-associated mutations within the NOTCH1 heterodimerization domain fall into at least two distinct mechanistic classes. *Mol Cell Biol*. 2006; 26(12):4642–51. Epub 2006/06/02. doi: [10.1128/MCB.01655-05](https://doi.org/10.1128/MCB.01655-05) PMID: [16738328](https://pubmed.ncbi.nlm.nih.gov/16738328/); PubMed Central PMCID: PMC1489116.
38. Gunnarsson R, Mansouri L, Rosenquist R. Exploring the genetic landscape in chronic lymphocytic leukemia using high-resolution technologies. *Leukemia & lymphoma*. 2013; 54(8):1583–90. Epub 2012/11/22. doi: [10.3109/10428194.2012.751530](https://doi.org/10.3109/10428194.2012.751530) PMID: [23167608](https://pubmed.ncbi.nlm.nih.gov/23167608/).
39. Tibaldi E, Brunati AM, Zonta F, Frezzato F, Gattazzo C, Zambello R, et al. Lyn-mediated SHP-1 recruitment to CD5 contributes to resistance to apoptosis of B-cell chronic lymphocytic leukemia cells. *Leukemia*. 2011; 25(11):1768–81. Epub 2011/06/28. doi: [10.1038/leu.2011.152](https://doi.org/10.1038/leu.2011.152) PMID: [21701493](https://pubmed.ncbi.nlm.nih.gov/21701493/).
40. Wu C, Sun M, Liu L, Zhou GW. The function of the protein tyrosine phosphatase SHP-1 in cancer. *Gene*. 2003; 306:1–12. Epub 2003/03/27. PMID: [12657462](https://pubmed.ncbi.nlm.nih.gov/12657462/).
41. Ngo VN, Young RM, Schmitz R, Jhavar S, Xiao W, Lim KH, et al. Oncogenically active MYD88 mutations in human lymphoma. *Nature*. 2011; 470(7332):115–9. Epub 2010/12/24. doi: [10.1038/nature09671](https://doi.org/10.1038/nature09671) PMID: [21179087](https://pubmed.ncbi.nlm.nih.gov/21179087/).
42. Landau DA, Carter SL, Getz G, Wu CJ. Clonal evolution in hematological malignancies and therapeutic implications. *Leukemia*. 2014; 28(1):34–43. Epub 2013/08/28. doi: [10.1038/leu.2013.248](https://doi.org/10.1038/leu.2013.248) PMID: [23979521](https://pubmed.ncbi.nlm.nih.gov/23979521/); PubMed Central PMCID: PMC3934006.