SUPPLEMENTAL DATA

SUPPLEMENTAL METHODS

Patient materials and cell line

We collected lymphoma/leukemia samples of the collaborating centers including four previously published cases[1-3]. The study includes a total of 12 patients who were diagnosed as precursor B-cell lymphoblastic lymphoma/leukemia (range 3-75 years at diagnosis) with IG-*MYC* translocation (preBLL) without noticed synchronous or asynchronous mature B-cell lymphoma. Of note is that from Case10, two biopsies from initial diagnosis were available and included in this analysis (10.1 and 10.2). Three patients (cases 1, 4, and 11) suffered from a relapse after initial diagnosis. Hence, three preBLL pairs entered the study of which material from initial diagnosis as well as relapse was available. All samples from initial diagnosis were confirmed by fluorescence *in situ* hybridization (FISH) and/or conventional cytogenetics analysis to carry an IG-*MYC* translocation or respective cytogenetic translocation. Furthermore, using immunohistochemistry, at least the initial diagnosis samples were shown to express Terminal Deoxynucleotidyl Transferase (TdT). The immunophenotypic characterization by immunohistochemistry and/or flow cytometry (FACS) of the cases was performed by the individual centers submitting the cases. Table 1 shows an overview of the clinical, cytogenetic and immunophenotypic characteristics of the preBLL samples herein studied. Supplemental Table 1 gives an overview of materials available and techniques performed in each case.

In addition, we studied the cell line 380, which was obtained from the Leibniz Institute DSMZ-German Collection of Microorganisms and Cell Culture. This cell line was established from the peripheral blood of a 15-year-old boy with acute lymphoblastic leukemia at relapse[4], showing an immature phenotype and carrying an IGH-*MYC* (t(8;14)) as well as an IGH-*BCL2* (t(14;18)) chromosomal translocations. Details on the cell line and both translocations have been published previously[5-7]. The cell line was tested negative for mycoplasma contamination and its authenticity was confirmed by STR analysis using StemElite ID System (Promega).

This study was approved by local review boards and performed in line with the regulations of the Institutional Review Board of the Medical Faculty of the University Kiel (D447/10 and D443/14).

DNA extraction

Extraction of DNA from formalin-fixed paraffin-embedded (FFPE) tissue was performed using either the truXTRAC FFPE DNA micro Tube Kit (Covaris) or the Qiagen FFPE DNA extraction kit (Qiagen) according to manufacturers' protocols. DNA extraction from cryo-preserved material or from cells frozen in DMSO was performed using the PUREGENE DNA Isolation Kit (Qiagen) according to manufacturers´ protocol. To test the quality of the DNA, we performed the High Sensitivity DNA Assay on the 2100 Bioanalyzer (Agilent). Heavily degraded DNA with a main peak below 200bp size, was not considered for exome sequencing. The quantity of the DNA was measured using Qubit fluorometer together with the Quant-iT dsDNA BR Assay Kit (Life Technologies).

Conventional cytogenetics and molecular cytogenetics

If not obtained from previous publications or from the original data from the centers, conventional cytogenetics and FISH analyses were performed as previously described[8]. Karyotypes were described according to ISCN guidelines (2013). FISH was done on 11 of the 12 cases using frozen tissue, FFPE tissues, and/or bone marrow smears as previously described[9,10]. Moreover, for the cell line 380, FISH was done on fixed cells from cell culture. The probes LSI BCL6, LSI MYC, LSI IGH/MYC, CEP8 Tri-color, LSI IGH and LSI BCL2 were applied. All probes were obtained from Abbott Molecular Diagnostics. In addition, for the detection of IGK-*MYC* and IGL-*MYC* translocations non-commercial probes were used[11]. At least 100 nuclei were examined for each probe whenever possible. Digital image acquisition, processing, and evaluation were performed using ISIS digital image analysis version 5.0 (MetaSystems).

Whole exome sequencing

Exome sequencing of six preBLL samples (5 patients) was performed using the Agilent SureSelect v6 exome target enrichment kit. A total of 200 ng of DNA were applied and the standard protocol SureSelectXT Automated Target Enrichment for Illumina paired-end multiplexed sequencing as well as the Agilent Bravo automated liquid handling platform has been used. Since the DNA from the cases was extracted from FFPE tissue, we treated the DNA with NEBNext FFPE DNA Repair Mix (New England Biolabs) according to the manufacturer's instructions to increase the utility for library preparation. Pools of three libraries were generated after validation (2200 TapeStation; Agilent Technologies) and quantification (Qubit System; Invitrogen). These pools were quantified using the KAPA Library Quantification kit (Peqlab) and the 7900HT Sequence Detection System (Applied Biosystems). The library was subsequently sequenced on an Illumina HiSeq 4000 sequencing instrument (Illumina) using a paired-end 2 × 75 bp protocol and an allocation of one pool with three exomes per lane.

Whole genome sequencing

Whole-genome sequencing (WGS) of the cell line 380 was performed within the framework of the ICGC MMML-Seq project. Experimental procedures for whole-genome sequencing of the ICGC MML-Seq have been published previously[12]. WGS data has been deposited in the European Genome-Phenome Archive (EGA) under accession number EGAS00001002968.

WES and WGS data processing

Alignment

Sequences were mapped to hs37d5 NCBI build 37.1 (ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/technical/reference/phase2_reference_assembly_sequence/hs37d5.fa.gz) with bwa-mem (version 0.7.8, minimum base quality threshold set to zero [-T 0], remaining settings left to default)[13]. Coordinates were sorted with bamsort (compression option set to fast [14]) and duplicate read pairs were marked with bammarkduplicates (compression option set to best, both are part of the biobambam package (version 0.0148).

Single nucleotide variant/Indel calling

Single nucleotide variants (SNVs) and indels were called from tumor samples with the internal DKFZ pipeline based on samtools/bcftools 0.1.19 and custom filters (optimized for somatic variant calling by deactivating the pval-threshod in bcftools) and Platypus 0.8.1, respectively as described by Jabs *et* al[15]. Genes were annotated with Annovar (Feb 2016). A 'confidence score' was calculated for each mutation. The maximum allowed allele frequency in ExAC[16] was lowered to 0.0001% (non-TCGA variants). Furthermore, variants with possibly base quality bias or a mapping quality bias (corresponding PV4 p-value < 0.01) were excluded from analysis. More cohort-based filtering criteria were applied downstream of the individual variant calling. In order to remove recurrent artifacts and misclassified germline events, somatic indels, that were identified as germline in at least two patients of the ICGC MMML-Seq WGS cohort were excluded.

Finally, variants in artifact prone regions were removed. These artifact prone regions were identified in an in-house lymphoma exome cohort. Variants in regions with between two and four variants with maximum intermutation distance of 100 bp that were recurrently found in three or more patients were considered artifacts, unless these regions overlapped with recurrent Kataegis regions[17] defined by an in-house whole genome sequencing lymphoma cohort. Of note is, that even thorough filtering was performed, some germline variants most likely remained in the final data set.

IGHV mutation analysis

We subjected 14 samples to IGHV sequence analysis which was performed as described by Kröber *et* al[18]. Of note is that we took a cutoff of 2% mutations to define an IGHV sequence as mutated or unmutated. Cases without evidence for ongoing mutations were not cloned and classified as not harboring ongoing somatic hypermutation. Nevertheless, potential subclone sequences with intraclonal diversity which might constitute 10% of all tumor cells might not be detected by conventional Sanger sequencing. Hence, we cannot conclude for sure that none of the cases harbors a minor subclone with ongoing SHM. Nevertheless, the PCR products for two cases (Case11-ID and Case11-R) showing more than 2% mutations were cloned and at least 5 clones sequenced. Based on those cases we have no evidence for ongoing SHM. Thus, we conclude that the cases do not harbor mutations due to ongoing SHM.

Detection of IG-*MYC* breakpoint junctions

The IG-*MYC* and IG-*BCL2* translocations were called using the SOPHIA algorithm (manuscript in preparation) by a workflow previously described in Sahm et al[19] using the whole-genome sequencing data of the cell line 380. For details refer to López *et* al in preparation.

The IG-*MYC* translocation using the exome sequencing data were identified after uploading the bam files to the IGV viewer (IGV 2.4.9.) and visual screening of the *MYC* and IG loci for reads, which have a mate read on the translocated allele. The mate read sequences were exported and the breakpoint junction retrieved.

Mutation and breakpoint validation by Sanger sequencing

The IG-*MYC* breakpoints identified by WES were validated using Sanger sequencing with primers flanking the breakpoints (Supplemental Table 11). To validate the mutations in *NRAS* and *KRAS* identified by WES and WGS as well as to identify mutations in samples without sequencing data, primers were designed to cover the mutations which are located in the described mutational hotspots affecting codons 12, 13 and 146 of the respective proteins. PCR primers and PCR conditions for Sanger sequencing of *NRAS* and *KRAS* are listed in Supplemental Table 12. PCR products were subjected to Sanger sequencing using the Big Dye Terminator v1.1 Cycle Sequencing Kit (Life Technologies). Sequence analysis was performed using an ABI PRISM 3130 Genetic Analyzer.

Pyrosequencing of *KRAS* hot spot mutations at codon 146

In two cases (Case7-ID and Case8-ID) we were not able to screen for occurrence of a mutation within codon 146 in KRAS by Sanger sequencing. Hence, we used a pyrosequencing assay for detection of the mutations. PCR primers and PCR conditions are listed in Supplemental Table 13. After PCR which was performed using the Pyromark PCR Kit (Qiagen), the product size was verified by gel electrophoresis. Single-strands were prepared using the Vacuum Prep Tool (Qiagen), followed by a denaturation step at 85 °C for 2 min and final sequencing primer hybridization. Pyrosequencing was performed using the PyroMark Q96 ID instrument (Qiagen). Sequencing results were analyzed using the PyroMark Q96 ID software (version 2.5.10, Qiagen).

OncoScan™ CNV FFPE Assay

Copy number analysis was performed on a total of 11 patients (13 samples) using the OncoScan™ CNV FFPE Assay Kit according to standard protocols (Affymetrix). Regions of gain, loss and copy number neutral loss of heterozygosity (CNN-LOH) were evaluated and visually inspected using OncoScan Console 1.3 (Affymetrix) and Nexus Express for OncoScan 3 (Biodiscovery). Human reference genome was GRCh37/hg19. Chromosomal imbalances encompassing at least 20 informative probes and having a minimum size of 100 kb as well as CNN-LOH larger than 5 Mb were considered informative.

DNA methylation analysis

About 0.5-1.0 µg genomic DNA of two preBLL (Supplemental Table 9) was applied for bisulfite conversion using the EZ DNA Methylation kit (ZymoResearch) according to manufacturers' instructions. Genome-wide DNA methylation analysis was performed using the Infinium MethylationEPIC BeadChip (EPIC) (Illumina) according to manufacturers' instruction. Arrays were scanned using the Illumina iScan. The raw DNA methylation data from the EPIC array were processed using the GenomeStudio software (v2011.1; methylation module 1.9.0; Illumina) applying the default settings. For data normalization, the intrinsic controls present on the array were used. Only hybridizations harboring a loci call rate higher than 95 % were further considered for analysis. Loci having a detection p-value >0.01 were excluded from further analysis. DNA methylation data of the preBLL cases is available in a MIAMI compliant format from Gene Expression Omnibus (GSE114210).

Phylogenetic tree based on DNA methylation level

The aim of the DNA methylation analysis was to identify the cell of origin of the preBLL samples based on the DNA methylation pattern. For classification of the cell of origin we used the approach recently described by Oakes *et* al[20], who showed that specific loci become differentially methylated during B-cell development. To this end, we used the Infinium HumanMethylationEPIC BeadChip array data from two preBLL and eight BL samples (Supplemental Table 9) and extracted those loci which were defined by Oakes *et* al.[20] to monitor B-cell development, ending up with 1404 CpG loci. The DNA methylation pattern of these 1404 CpG loci was compared to published HumanMethylation450 BeadChip array data sets of a total of 80 benign B-cell subpopulations[20-22] as well as to 26 pediatric acute lymphoblastic leukemias (ALL) carrying *KMT2A* rearrangements[23]. In addition, we downloaded 183 data sets from representative pediatric ALL samples, corresponding to 10 ALL subsets, from published B-ALL cases (GSE49031)[24] (Supplemental Table 10).

For reasons of clarity and comprehensibility, we calculated the mean methylation for each of the 1,404 loci for each ALL subgroup. The diversification of the benign as well as tumor samples was displayed in a phylogenetic tree using the ape package for R[25].

SUPPLEMENTAL RESULTS AND DISCUSSION

IG-MYC- and IGH-BCL2-translocation analysis

We performed WES in six samples (five patients) with suitable DNA quality and WGS of the cell line 380[4]. The breakpoint junctions of the IG-*MYC* translocations could be identified in 4/5 cases with exome (except for Case6) and genome (cell line 380) sequencing data, and were verified by PCR-based Sanger sequencing in three cases (Figure 1A). The breakpoints in the *MYC* locus were located in intron 1 in case 1 (identical in ID and R samples), and upstream of *MYC* in Case2 (~108 kb) and Case3 (~1.7 kb) as well as in the cell line 380 (~142 kb) as reported previously[7] (Supplemental Table 2). Moreover, the breakpoint in Case4 (identical in ID and R samples) was located downstream of *MYC* as typical for the IG-*MYC* variant translocations. The breakpoints within the IGH locus were located within the VDJ region, and showed evidence that a mislead VDJ recombination caused the IGH-*MYC* rearrangement (see Supplemental Figure 1 for details). In addition, in Case4 harboring an IGK-*MYC* translocation, the breakpoint was located in JK4 region mediated also by mistake during V-J-recombination. Given that DJ and VDJ recombination occurs in the bone marrow during an early stage of B cell development, in which MYC is expressed[26] which is considered to be a prerequisite for a translocation to occur, these findings in preBLL contrasts the situation in BL. The IG-*MYC* translocations in BL arise due to aberrant somatic hypermutation (SHM) or class switch recombination during the germinal center reaction (unpublished data, [27]). In line, we also failed in the six preBLL samples subjected to WES to observe *MYC* mutations which are introduced in around 70% of BL due to somatic hypermutation activity (unpublished data, [28]). Nevertheless, for cases 7 and 8 MYC expression could be shown by immunohistochemistry[1].

The IGH-*BCL2* translocation in the cell line 380, which was also identified by WGS, showed two breakpoints 15 kb and 30 kb downstream of *BCL2* and affecting JH-DH genes (Supplemental Table 3 and Supplemental Figure 2 and 3) as described previously[6]. Thus, this translocation also derived from a mistaken VDJ recombination in precursor B-cells of the bone marrow[27]. The simultaneous presence of *MYC* translocations and *BCL2* rearrangements as present in some preBLL is also observed in a subset of mature aggressive B-cell lymphomas which constitute a provisional entity frequently abbreviated as double hit (DH) lymphomas[29,30]. However, the latter entity is restricted to germinal center B cell derived (GCB) lymphomas[30] in which the *MYC*-translocation arises in the germinal center again due to aberrant SHM or CSR (Hübschmann *et* al. in preparation). Hence, VDJ associated DH preBLL should be differentiated from SHM/CSR associated DH GCB lymphoma.

SUPPLEMENTAL TABLES

Supplemental Table 1

Overview of preBLL cases and available data.

Case	Type	Material	Diagnosis	Karyotype and/or FISH	Sequencing data	OncoScan	EPIC	IGHV
Cell line 380	Cell line	Cell culture	Relapse	available	WGS	available	n.a.	failure
Case1-ID	tissue	fresh frozen	ID	available	WES	*	n.a.	not possible
Case1-R	tissue	fresh frozen	Relapse	available	WES	available	n.a.	available
Case2-ID	tissue	FFPE	ID	available	WES	available	n.a.	failure
Case3-ID	tissue	DMSO	ID	available	WES	available	n.a.	available
Case4-ID	tissue	DMSO	ID	available	WES	available	available	failure
Case4-R	tissue	fresh frozen	Relapse	n.a.	n.a.	n.a.	n.a.	failure
Case5-ID	tissue	FFPE	ID	available	n.a.	available	n.a.	available
Case6-ID	tissue	fresh frozen	ID	available	WES	n.a.	available	failure
Case7-ID	tissue	FFPE	ID	available	n.a.	available	n.a.	not possible
Case8-ID	tissue	blood smear	ID	available	n.a.	available	n.a.	available
Case9-ID	tissue	blood smear	ID	available	n.a.	available	n.a.	failure
Case10.1-ID	tissue	FFPE	ID	available	n.a.	available	n.a.	failure
Case10.2-ID	tissue	FFPE	ID	available	n.a.	available	n.a.	available
Case11-ID	tissue	fresh frozen	ID	available	n.a.	available	n.a.	available
Case11-R	tissue	fresh frozen	Relapse	available	n.a.	available	n.a.	available
Case12-ID	tissue	n.a.	ID	available	n.a.	n.a.	n.a.	n.a.

* of this case only aCGH data as published by Roug et al.[3] are available

Diagnosis: ID: initial diagnosis, Material for DNA extraction: Refers to the material type from which the DNA was extracted, FFPE: formalin fixed paraffin embedded, DMSO: viable cells frozen in media dosed with dimethylsulfoxide; WGS: whole-genome sequencing, WES: whole-exome sequencing, OncoScan: SNP array, EPIC: HumanMethylationEPIC BeadChip Array, IGHV: Sequence analysis of immunoglobulin heavy chain locus; n.a.: not available, failure: data not available due to technical failure, not possible: no DNA left for analysis

Supplemental Table 2

IG-*MYC* breakpoints which could be retrieved from WES and WGS in four preBLL cases as well as the cell line model which were verified by PCR and Sanger sequencing in four samples (Cases1-ID and -R, Case3-ID and Case4-ID). In addition, we used the breakpoint sequence detected in Case4-ID to check if the same breakpoints were detectable by Sanger sequencing in the corresponding relapse (Case4-R). Genomic coordinates are given in bp according to hg19. The breakpoints of the cell line 380 have been previously described[7].

Case	Detected by WES/WGS	Chr8 genomic position in bp	*MYC* breakpoint position[§]	Chr. partner	Chr. partner genomic position in bp	Validated by Sanger Sequencing	IG breakpoint position	IG-remodeling process involved in translocation event
Case1-ID	WES	128750090	Intron 1	14	106815788	Yes	IGHV3-33	VDJ recombination, 12 N nucleotides
Case1-R	WES	128750090	Intron 1	14	106815788	Yes	IGHV3-33	VDJ recombination, 12 N nucleotides
Case2-ID	WES	128639954	Upstream	14	106363826	No DNA left	IGHD2-15	VDJ recombination, 6 N nucleotides
Case3-ID	WES	128746552	Upstream	14	106330415	Yes	IGHJH4	VDJ recombination, 6 N nucleotides
Case4-ID	WES	129033832	Downstream	2	89160441	Yes	IGHJ4-J3	VDJ recombination, 5 N nucleotides
Case4-R	n.a.	129033832	Downstream	2	89160441	Yes	IGHJ4-J3	VDJ recombination, 5 N nucleotides
Cell line 380	WGS	128606227	Upstream	14	106370361	No	IGHD3-10	VDJ recombination, 7 N nucleotides
	WGS	128606239	Upstream	14	106329449	No	IGHJH6	VDJ recombination, 22 N nucleotides

[§] based on MYC transcript NM_002467.4.

Supplemental Table 3

IGH-*BCL2* breakpoints detected by WGS in cell line 380. Genomic coordinates are given in bp based on hg19. The breakpoints of the IGH-*BCL2* translocation were previously described[5,6].

Case	Chr18 genomic position	*BCL2* § breakpoint position	Chr14 genomic position	IGH breakpoint position	IG-remodeling process involved in translocation event
Cell line 380	60774828	downstream	106330066	IGHJH5	VDJ recombination, 7 N-nucleotides
	60760638	downstream	106375769	IGHD1-7	

§ based on BCL2 transcript NM_000633.2

Supplemental Table 4

Results of the IGHV gene analysis. In none of the analyzed cases Sanger sequencing of the clonal VDJ rearrangement of the IGH locus indicated ongoing somatic hypermutation.

Case	IGHV gene	IGHD gene	IGHJ gene	IGHV mutation (% of mutated bases)	Original V-gene rearrangement	Consequence of IGHV mutation
Case1-R	IGHV1-8	IGHD2-2	IGHJ6	unmutated (0%)	productive	-
Case3-ID	IGHV3-22	IGHD6-13	IGHJ4	unmutated (0%)	unproductive (out-of-frame)	-
Case5-ID	IGHV3-15	IGHD-2	IGHJ5	unmutated* (0.5%)	productive	unproductive (1bp insertion causing frameshift)
Case8-ID	IGHV3-23	IGHD3-22	IGHJ3	unmutated (0%)	productive	-
Case10.2-ID	IGHV3-64	Not conclusive	IGHJ6	unmutated (0%)	unproductive (out-of-frame)	-
Case11-ID	IGHV3-53	IGHD1-26	IGHJ4	mutated (7.84%)	productive	Unproductive (stopgain in FR3)
Case11-R	IGHV3-53	IGHD1-26	IGHJ4	mutated (7.84%)	productive	Unproductive (stopgain in FR3)

* Less than 2% mutations in IGHV sequence are considered as not mutated

“Original V-gene rearrangement“: indicates the productivity of the original V(D)J rearrangement. As the rearrangements of cases 3 and 10 lack point mutations, it is highly likely that the non-productivity stemmed from the rearrangement process, which generated an out-of-frame rearrangement. For cases 5 and 11, the original IGH rearrangements were in-frame and without stop codons in CDRIII, and hence productive. In these cases, mutations within the IGHV genes caused a frameshift (Case5) and a stopgain (Case11).

Supplemental Table 5

Overview of the chromosomal imbalances and mutations affecting the IG-loci as well as expression of surface immunoglobulin.

Case	Diagnosis	IG-*MYC* status	IGH-*BCL2* status	IGH Locus translocation	IGHV rearrangement	Expression sIg
Case1-ID	ID	IGH-*MYC*	negative	Monoallelic trans.	n.a.	negative
Case1-R	Relapse	IGH-*MYC*	negative	Monoallelic trans.	productive	negative
Case2-ID	ID	IGH-*MYC*	negative	Monoallelic trans.	failure	n.a.
Case3-ID	ID	IGH-*MYC*	negative	Monoallelic trans.	unproductive	n.a.
Case4-ID	ID	IGK-*MYC*	negative	Monoallelic trans.	failure	negative
Case4-R	Relapse	n.a.	n.a.	n.a.	failure	negative
Case5-ID	ID	IGH-*MYC*	negative	Monoallelic trans.	unproductive	negative
Case6-ID	ID	IGL-*MYC*	negative	Monoallelic trans.	failure	n.a.
Case7-ID	ID	IGH-*MYC*	negative	Monoallelic trans.	n.a.	positive
Case8-ID	ID	IGH-*MYC*	negative	Monoallelic trans.	productive	n.a.
Case9-ID	ID	IGH-*MYC*	negative	Monoallelic trans.	failure	n.a.
Case10.1-ID	ID	IGH-*MYC*	positive	Biallelic trans.	failure	n.a.
Case10.2-ID	ID	IGH-*MYC*	positive	Biallelic trans.	unproductive	n.a.
Case11-ID	ID	IGL-*MYC*	positive	Monoallelic trans.	unproductive	n.a.
Case11-R	Relapse	IGL-*MYC*	positive	Monoallelic trans.	unproductive	n.a.
Case12	ID	IGH-*MYC*	positive	Biallelic trans.	n.a.	n.a.
Cell line 380	Relapse	IGH-*MYC*	positive	Biallelic trans.	failure	negative

n.a.: not available, IGH-*BCL2* status: negative: no IGH-*BCL2* translocation, positive: IGH-*BCL2* translocation detected, failure: data not available due to technical failure; Expression of surface immunoglobulin (sIg): positive: sIg expression detected, negative: no sIg expression detected; trans.: translocation.

Supplemental Table 6

Overview of identified unique and shared nonsynonymous exonic variants in initial diagnosis and relapse of Case1.

	Gene	Chromosomal position (hg19)	Transcript	consequence on protein level	exonic classification
Unique in ID	*GTF2H4*	chr6:30878447A>C	ENST00000376316.2	p.K127T	missense
	ARMCX3	chrX:100880781 C>A	ENST00000341189.4	p.A271D	missense
	MIA2,RP11-407N17.3	chr14:39703370 A>G	ENST00000557148.1	p.K18E	missense
	GJA3	chr13:20716935 C>A	ENST00000241125.3	p.A165S	missense
	FAM160A2	chr11:6238333 G>T	ENST00000524416.1	p.S828Y	missense
	POMZP3	chr7:76240898 C>T	ENST00000310842.4	p.V150M	missense
	DCHS1	chr11:6643744 T>C	ENST00000299441.3	p.S3055G	missense
Unique in Relapse	*EPHA8*	chr1:22915675 G>A	ENST00000374644.4	p.V431I	missense
	AC096670.3	chr2:111858638 A>T	.		ncRNA
	ANK2	chr4:113970896 A>T	ENST00000504454.1	p.E4D	missense
	HFE	chr6:26092918 C>G	ENST00000357618.5	p.P208A	missense
	KDELR2	chr7:6505719 T>C	ENST00000258739.4	p.Y196C	missense
	TRPA1	chr8:72948599 G>A	ENST00000262209.4	p.P827S	missense
	GLDC	chr9:6645306 C>A	ENST00000321612.6	p.R65L	missense
	SLC17A6	chr11:22399126 A>T	ENST00000263160.3	p.D530V	missense
	MAML2	chr11:95825839 G>T	ENST00000524717.1	p.H452Q	missense
	FLT1	chr13:29001930 G>T	ENST00000282397.4	p.S412X	stopgain
	SOX21	chr13:95364020 G>A	ENST00000376945.2	p.P95L	missense
	MYEF2	chr15:48443733 G>A	ENST00000267836.6	p.R415X	stopgain
	CACNA1H	chr16:1252054 G>A	ENST00000565831.1	p.S535N	missense
	CACNA1H	chr16:1252360 C>T	ENST00000565831.1	p.A637V	missense
	CLCN7	chr16:1515338 G>A	ENST00000564568.1	p.S13L	missense
	SRRM2-AS1	chr16:2790164 A>T	.		ncRNA

	RNMTL1	chr17:685856 G>A	ENST00000574509.1	p.A80T	missense
	TMEM200C	chr18:5891066 C>A	ENST00000581347.2	p.A333S	missense
	CD3EAP	chr19:45911908 G>A	ENST00000309424.3	p.G228R	missense
	SEZ6L	chr22:26695016 A>C	ENST00000403121.1	p.H183P	missense
Shared	*NRAS*	chr1:115258744 C>T	ENST00000369535.4	p.G13D	missense
	CLCN6	chr1:11897176 G>A	ENST00000376487.3	p.E679K	missense
	RCOR3	chr1:211451541 G>A	ENST00000528926.1	p.R232H	missense
	RP11-439E19.3	chr1:246939671 C>T	.		ncRNA
	AK2	chr1:33487289 C>T	ENST00000467905.1	p.V79I	missense
	PTGER3	chr1:71513175 C>T	ENST00000354608.5	p.R29H	missense
	JMJD1C	chr10:64975061 T>C	ENST00000402544.1	p.K107R	missense
	NPAT	chr11:108032143 C>T	ENST00000278612.8	p.G1224S	missense
	KMT2A	chr11:118344402 C>G	ENST00000531904.2	p.S876C	missense
	C11orf96	chr11:43965017 G>T	ENST00000528572.1	p.E304D	missense
	CKAP5	chr11:46781927 T>C	.		ncRNA
	PATL1	chr11:59419942 T>A	ENST00000300146.9	p.K474M,	missense
	SCT	chr11:626510 C>T	ENST00000176195.3	p.W96X	stopgain
	NUMA1	chr11:71724411 G>A	ENST00000393695.3	p.R1380C	missense
	SLC5A8	chr12:101560444 T>C	ENST00000536262.2	p.I452V	missense
	ZNF140	chr12:133682813 C>G	ENST00000355557.2	p.T317S	missense
	MUC19	chr12:40928816 C>A	UNKNOWN		
	CDK17	chr12:96728586 A>C	ENST00000551816.1	p.L10R	missense
	KIF26A	chr14:104605456 C>T	ENST00000423312.2	p.P40L	missense
	OR11H7	chr14:20697802 T>C	UNKNOWN		unknown
	AKAP6	chr14:33292435 A>G	ENST00000280979.4	p.I1806V	missense
	FAM177A1	chr14:35548202 T>C	ENST00000280987.4	p.I159T	missense
	DIS3L	chr15:66625534 A>T	ENST00000319194.5	p.I934F	missense
	SNUPN	chr15:75890892 TG>G	ENST00000567134.1	p.T297fs	frameshift deletion
	DNASE1	chr16:3707214 T>A	ENST00000246949.5	p.N192K	missense
	RAB11FIP3	chr16:569844 A>C	ENST00000457159.1	p.S650R	missense
	JUP	chr17:39928039 T>C	ENST00000591690.1	p.D23G	missense
	ABCA10	chr17:67181617 C>T	.		ncRNA
	CHD3	chr17:7798247 G>A	ENST00000330494.7	p.V428I	missense
	RNF213	chr17:78345677 T>G	ENST00000582970.1	p.F4143L	missense
	DOCK6	chr19:11354498 T>G	ENST00000294618.7	p.K362Q	missense
	ATP5D	chr19:1242604 C>G	ENST00000215375.2	p.Y97X	stopgain
	SLC1A6	chr19:15067297 G>A	ENST00000221742.3	p.T387M	missense
	ANO8	chr19:17434564 G>A	ENST00000159087.4	p.P1154L	missense
	ZNF431	chr19:21352648 G>A	ENST00000599296.1	p.S121N	missense
	ZNF181	chr19:35231566 T>C	ENST00000492450.1	p.Y94H	missense
	ZNF568	chr19:37428129 G>T	ENST00000333987.7	p.G115W	missense
	PTPRS	chr19:5225777 C>T	ENST00000372412.4	p.A820T	missense
	TSEN34	chr19:54695442 C>G	ENST00000455798.1	p.S76C	missense
	AC007249.3	chr2:10594009 C>T	.		ncRNA
	GLI2	chr2:121729595 G>A	ENST00000314490.11	p.E52K	missense
	AGAP1	chr2:236403421 T>TA	ENST00000304032.8	p.Y31_E32delinsX	stopgain
	CDH26	chr20:58562553 C>T	ENST00000244047.5	p.R295C	missense
	COL20A1	chr20:61950934 G>A	ENST00000326996.6	p.G968R	missense
	DSCR9	chr21:38592901 G>A	.		ncRNA
	AP001059.5	chr21:45671184 G>A	.		ncRNA
	MIEF1	chr22:39907911 A>G	ENST00000433117.2	p.K68E	missense
	NXPE3	chr3:101520335 T>C	ENST00000477909.1	p.F117S	missense
	ACPP	chr3:132047130 G>A	ENST00000351273.7	p.R47Q	missense
	NUP210	chr3:13373810 T>TA	ENST00000254508.5	p.I1306fs	frameshift insertion

TM4SF18	chr3:149042763 CCAGAAAAAGCAAT>C	ENST00000470080.1	p.I101fs	frameshift deletion
CACNA2D2	chr3:50417184 G>C	:ENST00000479441.1	p.S322C	missense
STIM2	chr4:27021029 A>T	ENST00000477474.2	p.I193	missense
SLC2A9	chr4:9836525 G>T	ENST00000309065.3	p.L438I	missense
FBXL17	chr5:107717280 G>C	ENST00000542267.1	p.P38R	missense
PCDH1	chr5:141233830 T>G	ENST00000287008.3	p.Q1164P	missense
GUSBP1	chr5:21585688 G>A	.		ncRNA
PDZD2	chr5:32074593 C>G	ENST00000282493.3	p.H1127Q	missense
ANKRD31	chr5:74491508 T>A	ENST00000506364.2	p.E322V	missense
MTX3	chr5:79284414 C>A	ENST00000512528.1	p.W125C	missense
MYB	chr6:135523805 C>T	ENST00000316528.8	p.P651L	missense
AC073342.1	chr7:142919291 G>A	ENST00000595842.1	p.R24C	missense
TBRG4	chr7:45145243 T>A	ENST00000258770.3	p.K178X	stopgain
SLCO5A1	chr8:70744085 C>G	ENST00000526750.1	p.G275A	missense
C9orf84	chr9:114464394 A>C	ENST00000374287.3	p.L963X	stopgain
PAPPA	chr9:118982300 C>T	ENST00000328252.3	p.P668L	missense
TLR4	chr9:120475271 C>T	ENST00000355622.6	p.R289X	stopgain
SDCCAG3	chr9:139299100 C>T	ENST00000371725.3	p.E278K	missense
ADAMTSL1	chr9:18533267 C>T	ENST00000327883.7	p.R72X	stopgain
GLDC	chr9:6592166 T>C	ENST00000321612.6	p.I487V	missense

Supplemental Table 7

Overview of *NRAS* and *KRAS* mutations detected in preBLL.

PID	Gene	Chromosomal location (hg19)	Consequence on protein level	Annovar Function	Detected by	VAF
Case1-ID	*NRAS*	Chr1: 115258744C>T	p.G13D	exonic	WES/Sanger	0.316
Case1-R		Chr1: 115258744C>T	p.G13D	exonic	WES/Sanger	0.175
Case3-ID		Chr1: 115258745C>G	p.G13R	exonic	WES/Sanger	0.253
Case4-ID		Chr1: 115258744C>T	p.G13D	exonic	WES/Sanger	0.049
Case4-R		Chr1: 115258747C>A	p.G12D	exonic	Sanger	-
Case7-ID		Chr1: 115258744C>T	p.G13D	exonic	Sanger	-
Case8-ID		Chr1: 115258745C>G	p.G13R	exonic	Sanger	-
Cell line 380	*KRAS*	Chr12: 25398284C>T	p.G12D	exonic	WGS/Sanger	0.488
Case2-ID		Chr12: 25378562C>T	p.A146T	exonic	WES/Sanger	0.385
Case4-ID		Chr12:25398281C>T	p.G13D	exonic	WES/Sanger	0.286

Gene: *NRAS*: ENST00000369535.4, *KRAS*: ENST00000311936.3; detected by: WES: whole-exome sequencing, WGS: whole-genome sequencing, Sanger: Sanger sequencing, VAF: variant allele frequency as determined by WES or WGS, this value was no corrected for purity of the sample.

Supplemental Table 8

Copy number alterations in preBLL cases as identified by OncoScan analysis.

Case	CNA	Chromosomal region in bp (hg19)	Cytoband	Length in Mb
Case1-R	CN Gain	chr1:144,009,053-249,212,878	1q21.1 - q44	105.2
	LOH	chr6:204,909-33,641,721	6p25.3 - p21.31	33.4
	CN Gain	chr6:1-171,115,067	6p25.3 - q27	171.2
	CN Gain	chr7:1-159,138,663	7p22.3 - q36.3	159.2
	CN Gain	chr8:1-146,364,022	8p23.3 - q24.3	146.4
	CN Gain	chr14:20,219,083-106,325,013	14q11.2 - q32.33	86.1
	CN Loss	chr15:59,296,491-78,984,214	15q22.1 - q25.1	19.7
	CNN-LOH	chr17:400,959-19,277,144	17p13.3 - p11.2	18.9
	CN Gain	chr20:1-63,025,520	20p13 - q13.33	63.0
	High Level Gain	chr21:47,542,220-48,097,610	21q22.3	0.6
	CN Gain	chr22:30,047,732-30,331,290	22q12.2	0.3
	CN Gain	chrX:2,699,968-154,929,412	Xp22.33 - q28	152.2
Case2-ID	CN Gain	chr1:146,501,028-147,136,284	1q21.1 - q21.2	0.6
	CN Gain	chr1:149,742,045-249,212,878	1q21.2 - q44	99.5
	CN Loss	chr16:67,574,408-69,003,205	16q22.1	1.4
	CN Loss	chr19:247,232-1,213,911	19p13.3	1.0
Case3-ID	CN Gain	chr1:144,009,053-249,212,878	1q21.1 - q44	105.2
	High Level Gain	chr4:1,864,994-1,965,201	4p16.3	0.1
	CN Loss	chr6:134,471,451-134,537,590	6q23.2	0.1
	Homozygous Loss	chr7:27,157,756-27,288,041	7p15.2	0.1
	CN Loss	chr8:128,737,699-128,761,102	8q24.21	0.02
	CN Loss	chr9:20,739,231-21,422,316	9p21.3	0.7
	Homozygous Loss	chr9:21,438,684-23,979,730	9p21.3	2.5
	CN Loss	chr9:24,004,490-25,119,796	9p21.3	1.1
	CN Loss	chr11:6,227,083-6,621,516	11p15.4	0.4
	CN Gain	chr14:100,709,309-101,021,558	14q32.2	0.3
	CN Gain	chr15:40,784,285-41,038,214	15q15.1	0.3
Case4-ID	High Level Gain	chr8:142,604,299-143,302,737	8q24.3	0.7
	CN Gain	chr11:64,156,585-64,499,482	11q13.1	0.3
	CN Gain	chr21:1-48,129,895	21p13 - q44	48.1
Case5-ID	CN Gain	chr1:144,936,742-249,212,878	1q21.1 - q21.3	104.3
	CNN-LOH	chr14:56,823,320-62,569,592	14q22.3 - q23.2	5.7
Case7-ID	CN Loss	chr3:63,411-129,391,127	3p26.3 - q22.1	129.3
	CN Loss	chr5:54,178,160-54,579,914	5q11.2	0.4
	CN Loss	chr5:81,673,168-154,353,012	5q14.2 - q33.2	72.7
	CNN-LOH	chr17:400,959-12,936,513	17p13.3 - p12	12.5
	CN Gain	chr20:1-63,025,520	20p13 - q13.33	63.0
	CN Gain	chr22:16,054,713-51,213,826	22q11.1 - q13.33	35.2
Case8-ID	CN Gain	chr1:144,009,053-189,648,793	1q21.1 - q31.1	45.6
	CN Gain	chr1:189,670,204-249,212,878	1q31.1 - q44	59.5
	CN Gain	chr8:115,654,443-127,636,944	8q23.3 - q24.21	12.0
	CN Loss	chr8:127,658,147-128,094,460	8q24.21	0.4
	CNN-LOH	chr14:58,259,785-68,953,817	14q23.1 - q24.1	10.7
Case9-ID	CN Gain	chr19:8,518,415-10,275,660	19p13.2	1.8
Case10.1-ID	CN Gain	chr1:144,009,053-147,826,658	1q21.1 - q21.2	3.8
	CN Loss	chr1:147,830,492-149,815,536	1q21.2	2.0
	CN Gain	chr1:149,849,171-249,212,878	1q21.2 - q44	99.4
	CNN-LOH	chr2:23,370,317-28,580,150	2p24.1 - p23.2	5.2
	CN Loss	chr3:70,550,462-71,099,541	3p13	0.6
	CNN-LOH	chr10:7,199,054-16,979,661	10p14 - p13	9.8
	CN Gain	chr17:55,523,434-55,694,141	17q22	0.2
	CN Gain	chr21:9,648,315-48,097,610	21p11.2 - q22.3	38.5
Case10.2-ID	CNN-LOH	chr2:23,485,507-28,572,491	2p24.1 - p23.2	5.1
	CN Loss	chr3:70,550,462-71,070,749	3p13	0.5
	CNN-LOH	chr10:7,199,054-16,979,661	10p14 - p13	9.8
	CN Gain	chr21:9,648,315-48,097,610	21p11.2 - q22.3	38.5
Case11-ID	CN Gain	chr1:145,268,038-249,212,878	1q21.1 - q44	103.9
	CNN-LOH	chr6:204,909-51,838,372	6p25.3 - p12.2	51.6
	CN Gain	chr7:1-159,138,663	7p22.3 - q36.3	159.1

	Homozygous Loss	chr9:21,968,549-22,047,117	9p21.3	0.08
Case11-R	CN Gain	chr1:145,268,038-249,212,878	1q21.1 - q44	103.9
	CNN-LOH	chr6:204,909-51,838,372	6p25.3 - p12.2	51.6
	CN Gain	chr7:1-159,138,663	7p22.3 - q36.3	159.1
	Homozygous Loss	chr9:21,968,549-22,047,117	9p21.3	0.08
Cell line 380	CN Loss	chr5:158,194,413-158,522,577	5q33.3	0.3
	CNN-LOH	chr9:204,738-21,033,557	9p24.3 - p21.3	20.8
	Homozygous Loss	chr9:21,042,299-22,300,493	9p21.3	1.3
	CNN-LOH	chr9:22,318,163-37,539,319	9p21.3 - p13.2	15.2
	CN Gain	chr10:75,905,198-76,256,536	10q22.2	0.4
	CN Loss	chr13:48,984,722-49,143,427	13q14.2	0.2

CN: copy number, CNN-LOH: copy number neutral loss of heterozygosity

Supplemental Table 9

Overview of cases with HumanMethylationEPIC BeadChip array data included in the DNA methylation analysis.

Case	**Cohort**	**Diagnose**	**Case published in**
Case4-ID	preBL	preBL	
Case6-ID	preBL	preBL	
R3A	MMML-MYC-SYS	BL	Wagener et al., in preparation
R6A	MMML-MYC-SYS	BL	Wagener et al., in preparation
N1B	MMML-MYC-SYS	BL	Wagener et al., in preparation
4182393	ICGC MMML-Seq	BL	[31]
4133511	ICGC MMML-Seq	BL	[31]
4177856	ICGC MMML-Seq	BL	[31]
4194218	ICGC MMML-Seq	BL	[31]
4194891	ICGC MMML-Seq	BL	[31]

Supplemental Table 10

Overview of cases with HumanMethylation450 BeadChip Array data included in the DNA methylation analysis.

Type	Subtyp	Subtyp abbreviation	Number of cases	Case and data published in
benign	Hematopoietic precursor cell	HPC	6	[22]
	Precursor 1 B-cell	preB1C	6	[22]
	Precursor 2 B-cell	preB2C	6	[22]
	Immature B-cell	iBC	4	[22]
	Naïve B-cells	naiBC	7	[20,21]
	Tonsillar naïve B-cells	t-naiBC	5	[21]
	Germinal center founder cells	GCF	3	[20]
	Early non-class switched memory B-cells	Early nonCS MBC	4	[20]
	Germinal center B-cells	gcBC	9	[21]
	Centroblast	CB	3	[20]
	Centrocyte	CC	3	[20]
	Splenic marginal zone B-cell	sMGZ	3	[20]
	Non-class switched memory B-cell	nonCS MBC	3	[20]
	Class-switched memory B-cell	CS MBC	2	[20]
	Memory B-cells	MBC	5	[21]
	Tonsillar plasma cell	t-PC	8	[21]
	Bone marrow plasma cell	bm-PC	3	[21]
tumor	ALL with KMT2A/MLL rearrangement	KMT2A+ALL	26	[23]
	High hyperdiploid ALL	HeH	57	[24]
	ALL with t(12;21)ETV6/RUNX1	t(12;21)	40	[24]
	Undefined ALL	Undefined	40	[24]
	Non-recurrent ALL	Non-recurrent	22	[24]
	ALL with 11q23/MLL alteration	11q23/MLL	10	[24]
	ALL with t(1;19)TCF3/PBX1	t(1;19)	5	[24]
	ALL with dic(9;20)	dic(9;20)	2	[24]
	ALL with t(9;22)BCR/ABL1	t(9;22)	4	[24]
	ALL with iAMP21	iAMP21	2	[24]
	ALL with <45 chromosomes	<chr45	1	[24]

Supplemental Table 11

Overview of PCR primers and PCR conditions for Sanger sequencing of IG-*MYC* breakpoints in three preBLL cases.

Case	Gene	Genomic region (hg19)	Forward primer (5'-3') Reverse primer (5'-3')	Annealing Temp.[1]	PCR[2] (bp)
Case1	*MYC*	chr8:128,750,092-128,750,111	TCCTTCAGGTGGCGCAAAAC	55°C	132
	IGH	chr14:106,573,263-106,573,282	AAATGAACAGCCTGAGAGCC		
Case3	*MYC*	chr8:128,746,566-128,746,585	CAGCTCAGCGTTCAAGTGTT	60°C	155
	IGH	chr14:106,330,398-106,330,417	AGCAGGAGAGAGGTTGTGAG		
Case4	*MYC*	chr8:129033895-129033914	TTTTAGCACACTCACCCTGC	55°C	193
	IGK	chr2:89160526-89160547	GGGGAAATGGGGAAGAAATAGC		

[1] Annealing temperature in °C
[2] Length of PCR product in base pairs (bp)

Supplemental Table 12

PCR primers and PCR conditions for Sanger sequencing of hotspot mutations at protein codons 12 and 13 in NRAS and codons 12, 13 as well as 146 in KRAS.

Gene	Hotspot	Genomic region (hg19)	Forward primer (5'-3') Reverse primer (5'-3')	Annealing Temp.[1]	PCR[2] (bp)
NRAS	Codons 12/13	chr1:115,258,602-115,258,849	AGTACTGTAGATGTGGCTCGC GAATATGGGTAAAGATGATCCGAC	60°C	248
KRAS	Codons 12/13	chr12:25,398,147-25,398,426	TTTGTATTAAAAGGTACTGGTGGAG CCTTTATCTGTATCAAAGAATGGTC	55°C	280
KRAS	Codon 146	chr12:25,378,396-25,378,842	GGAAAATTTGGTGTAGTGGAAAC AGAAGCAATGCCCTCTCAAG	60°C	447

[1] Annealing temperature in °C, [2] Length of PCR product in base pairs (bp)

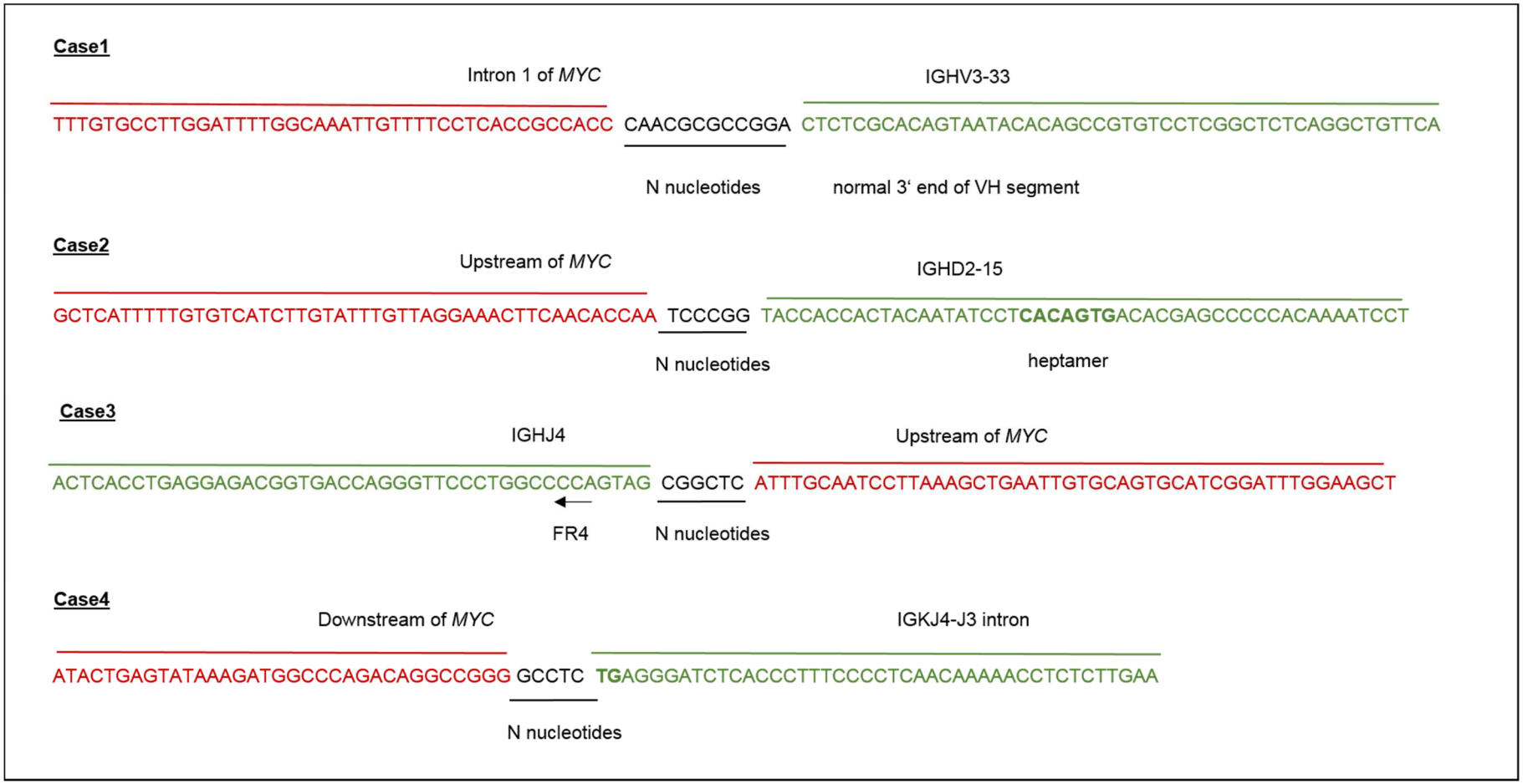
Supplemental Table 13

PCR primers and sequencing primer for pyrosequencing of the hotspot mutation in codon 146 of *KRAS* gene.

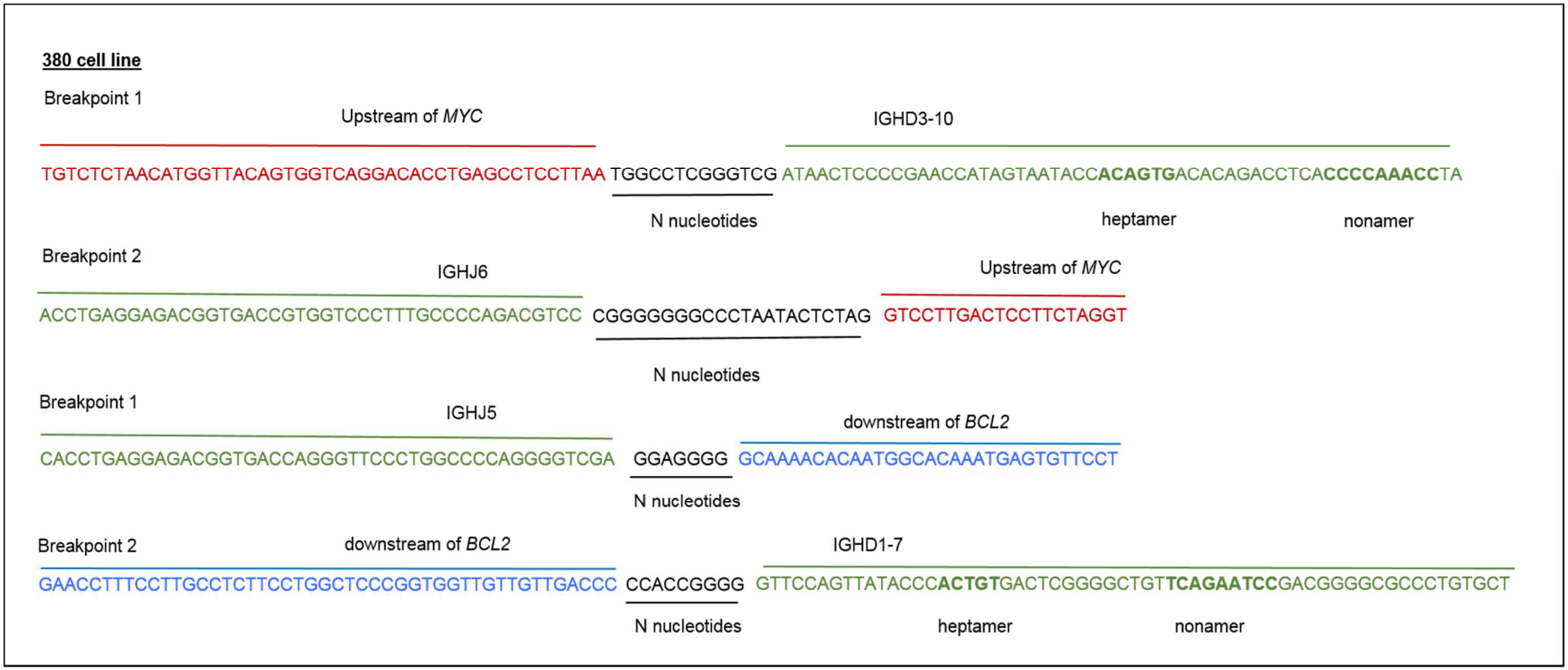
Hotspot	Forward primer (5'-3') Reverse primer (5'-3')	Sequencing primer (5'-3')	Annealing Temp.[1]	PCR[2] (bp)
Codon 146	ACCCAGAGAACAAATTAAAAGAGTT AGTTATGATTTTGCAGAAAACAGA	AATTCCTTTTATTGAAACAT	55°C	215

[1] Annealing temperature in °C, [2] Length of PCR product in base pairs (bp)

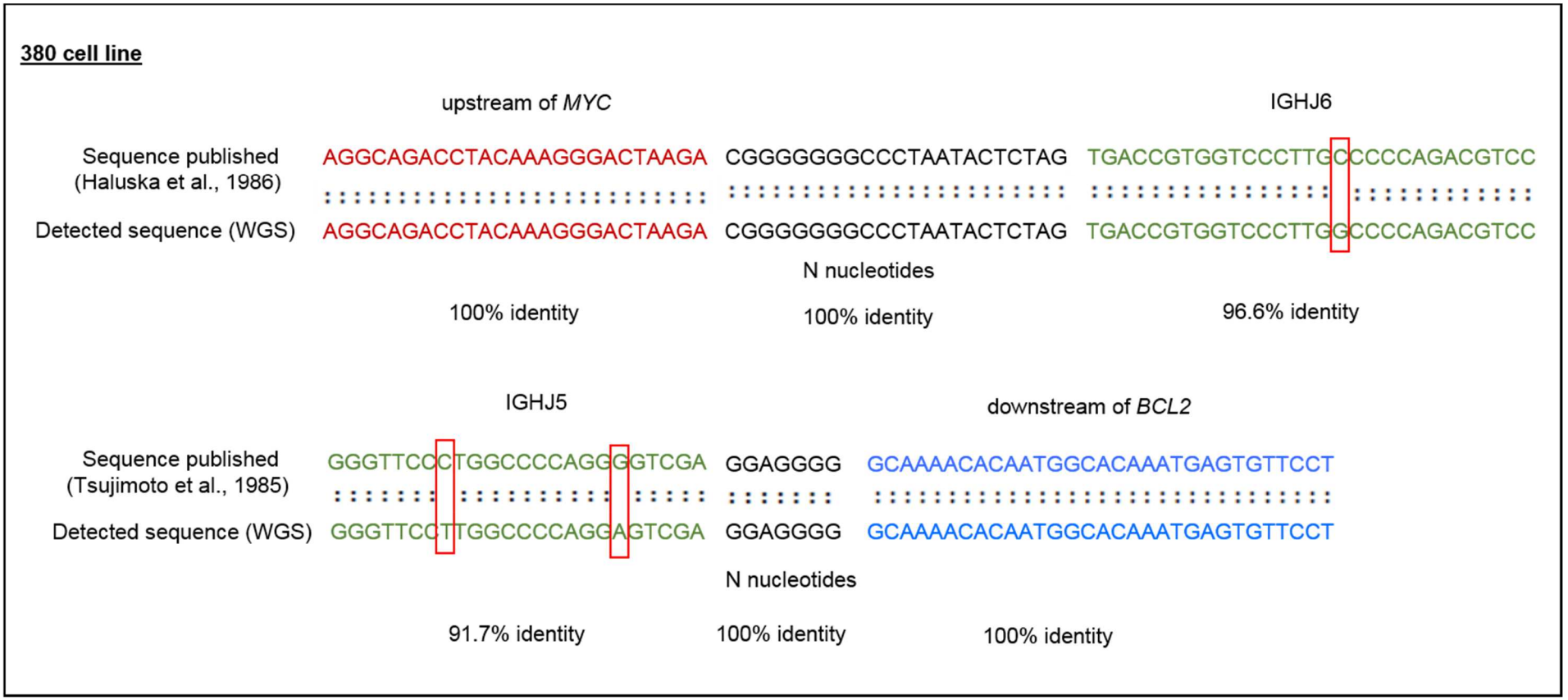
SUPPLEMENTAL FIGURES



Supplemental Figure 1: Detailed breakpoint analyses of IGH-*MYC* translocations in preBLL cases. The sequences in red correspond to chromosome 8, the sequences in green to IG loci, and the sequences in black indicate the presence of N nucleotides. In Case1 the breakpoint affects the 3' end of the IGHV gene. In Case2 the breakpoint maps to the middle of IGHD2-15, the heptamer sequence of the 5' RSS of IGHD2-15 is highlighted in bold green. In Case3 the breakpoint is located at the 5' end of IGHJ4, with loss of some bases from the 5' end, as typical for JH rearrangements. There are N nucleotides at the breakpoint. Case4 exhibits the breakpoint at the 5' end of IGKJ4. The first 2 bases with homology to the IGK locus (TG, labeled in bold) represent the 5' end of the RSS of IGKJ4, followed by the intron between IGKJ3 and IGKJ4. The breaks in all the cases do not fall into full VDJ rearrangements, but seemed rather to have occurred during a DH to JH recombination in a pro B cell or during IGK rearrangement.

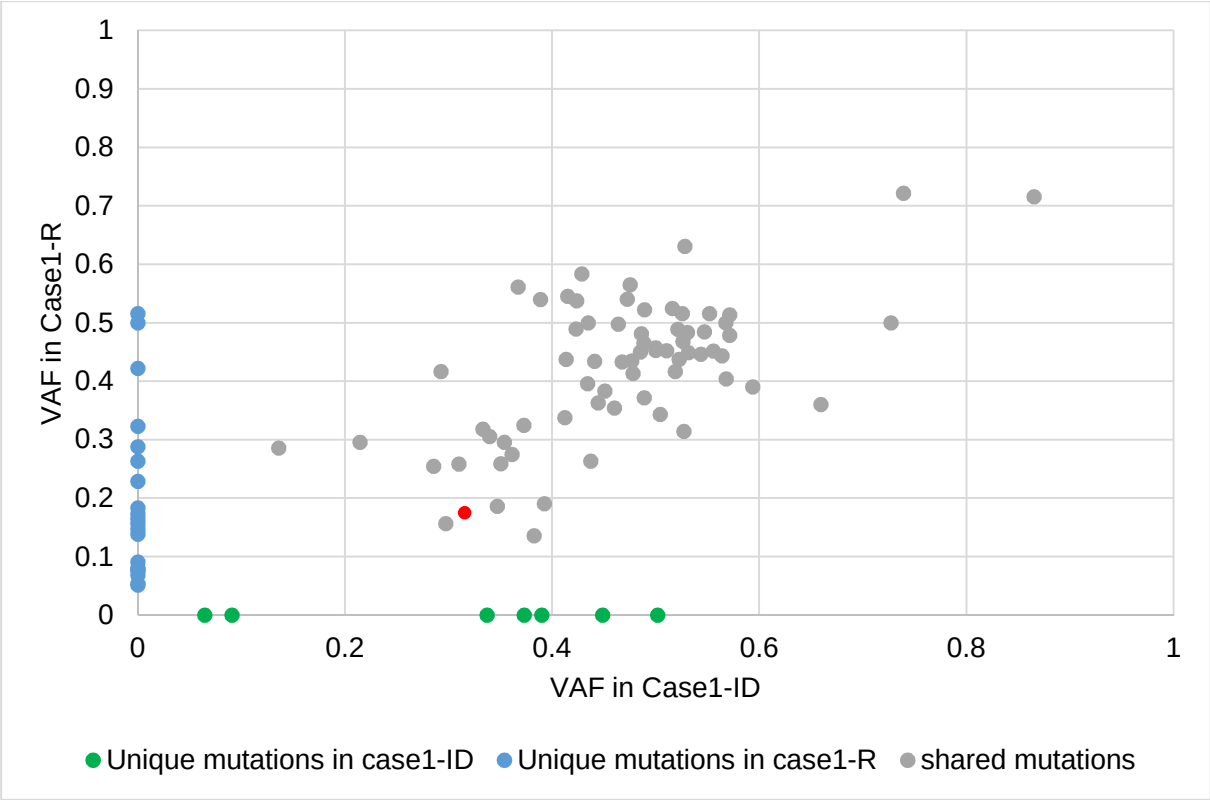


Supplemental Figure 2: Detailed breakpoint analysis of the IGH-*MYC* and IGH-*BCL2* translocations in cell line 380. The sequences for chromosome 8 are labeled in red, the sequences for chromosome 14 (IGH locus) in green, the sequences for chromosome 18 (*BCL2* locus) in blue, and N nucleotides are represented in black. IGHD genes have heptamer and nonamer sequences highlighted in bold letters. In both translocations one break is at the 5' end of JH with typical removal of some bases and addition of some N nucleotides; the reciprocal breaks in both are at the 3' end of an IGHD gene, also here with removal of some bases and addition of N nucleotides. The breakpoints identified in cell line 380 for IGH-*MYC* and IGH-*BCL2* translocations are in line with the previous published studies[6,7] (Supplemental Figure 3) and occurred during IGHD to IGHJ recombination in a pro B cell, as mistakes of VDJ recombination.

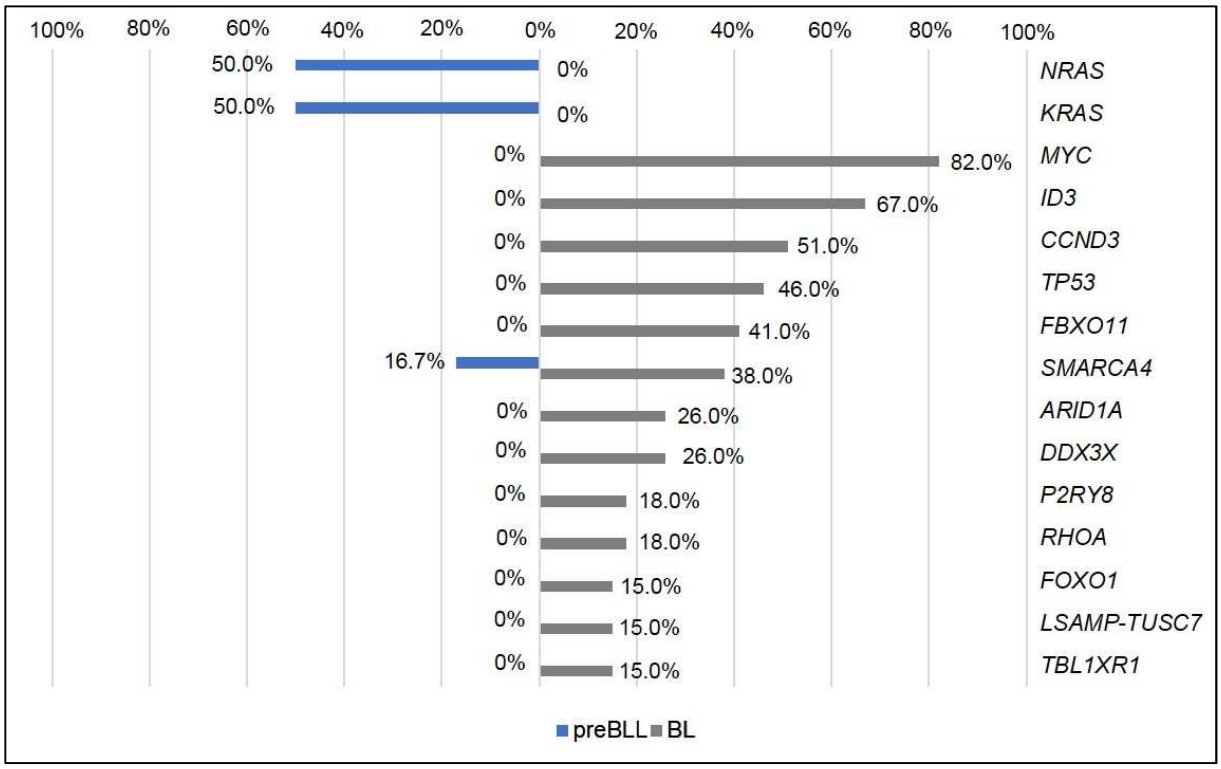


Supplemental Figure 3: Alignment of published and detected breakpoint sequence of the IGH-*MYC* and IGH-*BCL2* translocations in cell line 380[5-7]. The sequence of the IGH-*MYC* translocation shows 100% identity to the previously published breakpoint sequence in chromosome 8 (upstream of *MYC* locus), including the sequence of the observed N nucleotides. The sequence of the IGHJ6 breakpoints observed by WGS shows 96.6% identity to the published sequence, differing only in one nucleotide (highlighted with the red box). The sequence determined by WGS of the IGH-*BCL2* translocation, shows 91.7% identity in the IGHJ5 sequence to the published sequence, differing in two nucleotides (highlighted with the red box) and 100% identity to the sequence of chromosome 18 (downstream of *BCL2* locus) as well as to the N nucleotides.

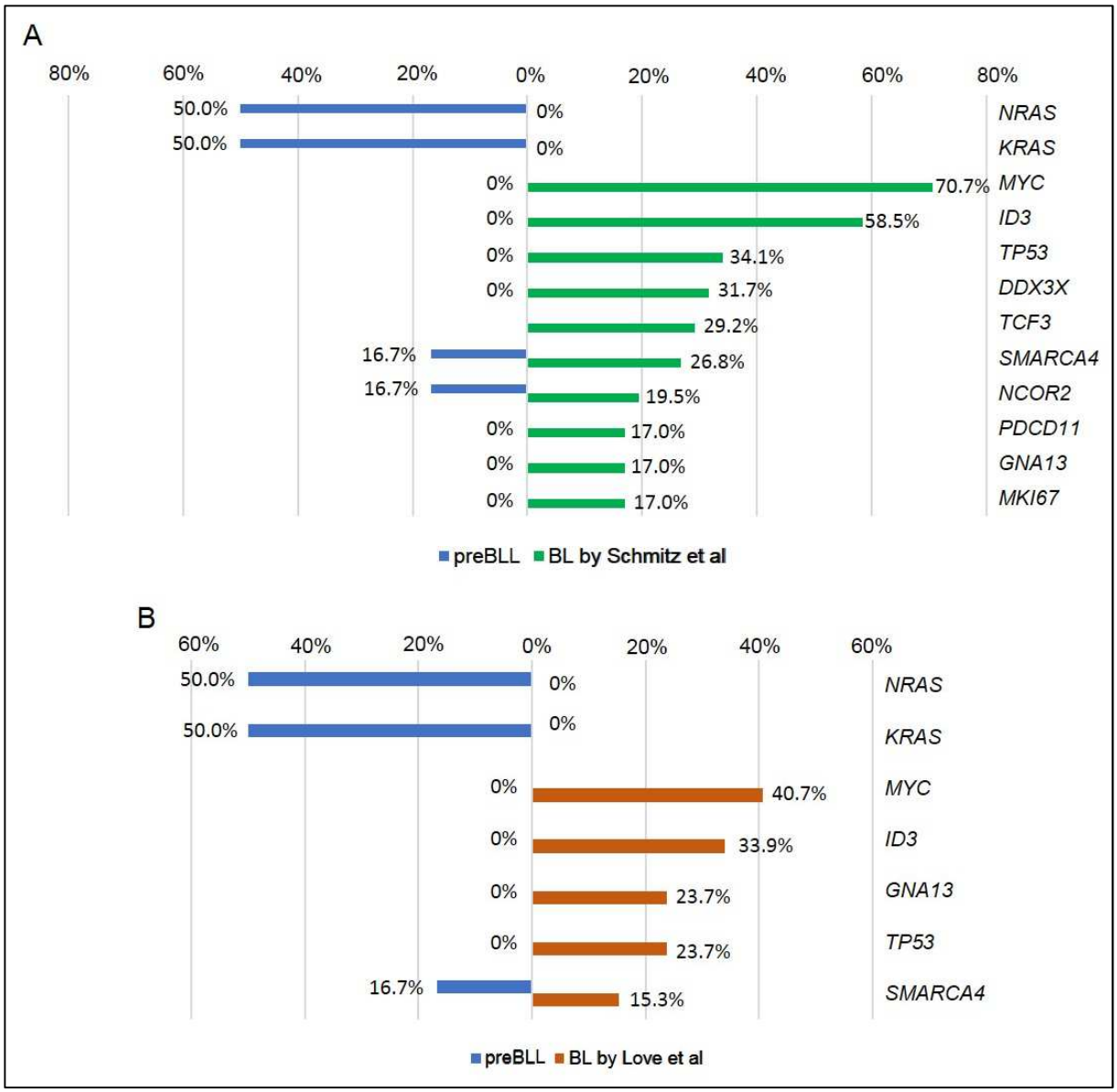
.



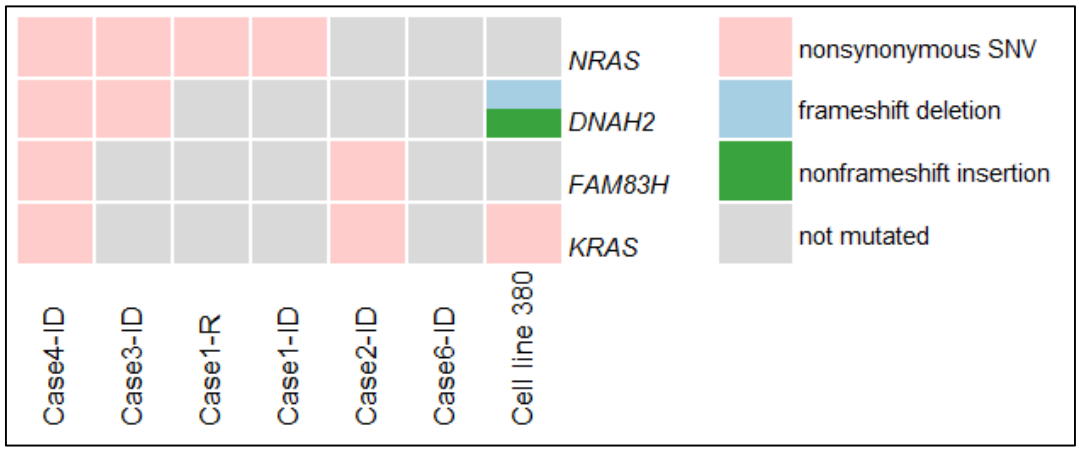
Supplemental Figure 4: Overview on potentially protein-changing variants in Case1. Plotted are the variant allele frequencies (VAF) of the mutations identified in the ID and the relapse. Of note is, that the VAF was not corrected for the tumor cell content. Mutations which are not identified in the ID or relapse (R), i.e. are unique, have a VAF of 0. Mutations unique to ID are colored in green, mutations unique to the relapse are blue and shared mutations in grey. Highlighted in red is the *NRAS* mutation shared by ID and relapse sample.



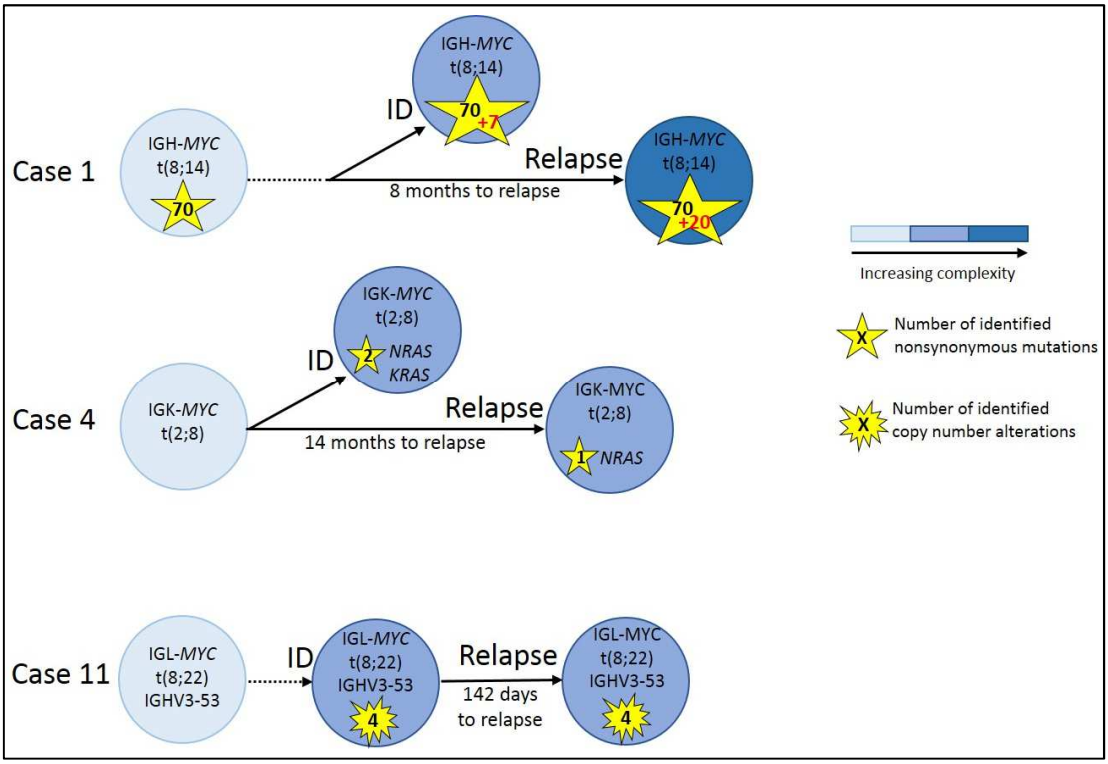
Supplemental Figure 5: Comparison of frequently mutated genes in BL (grey) and preBLL (blue, based on 7 cases (6 patients) with sequencing data) showing no overlap of recurrently affected genes between both lymphoma subgroups. Included were those genes which are mutated in more than 2/5 preBLL patients (*NRAS, KRAS*) and the 13 genes recurrently mutated in 6/39 BL patients (based on unpublished data which are accessible at www.icgc.org).



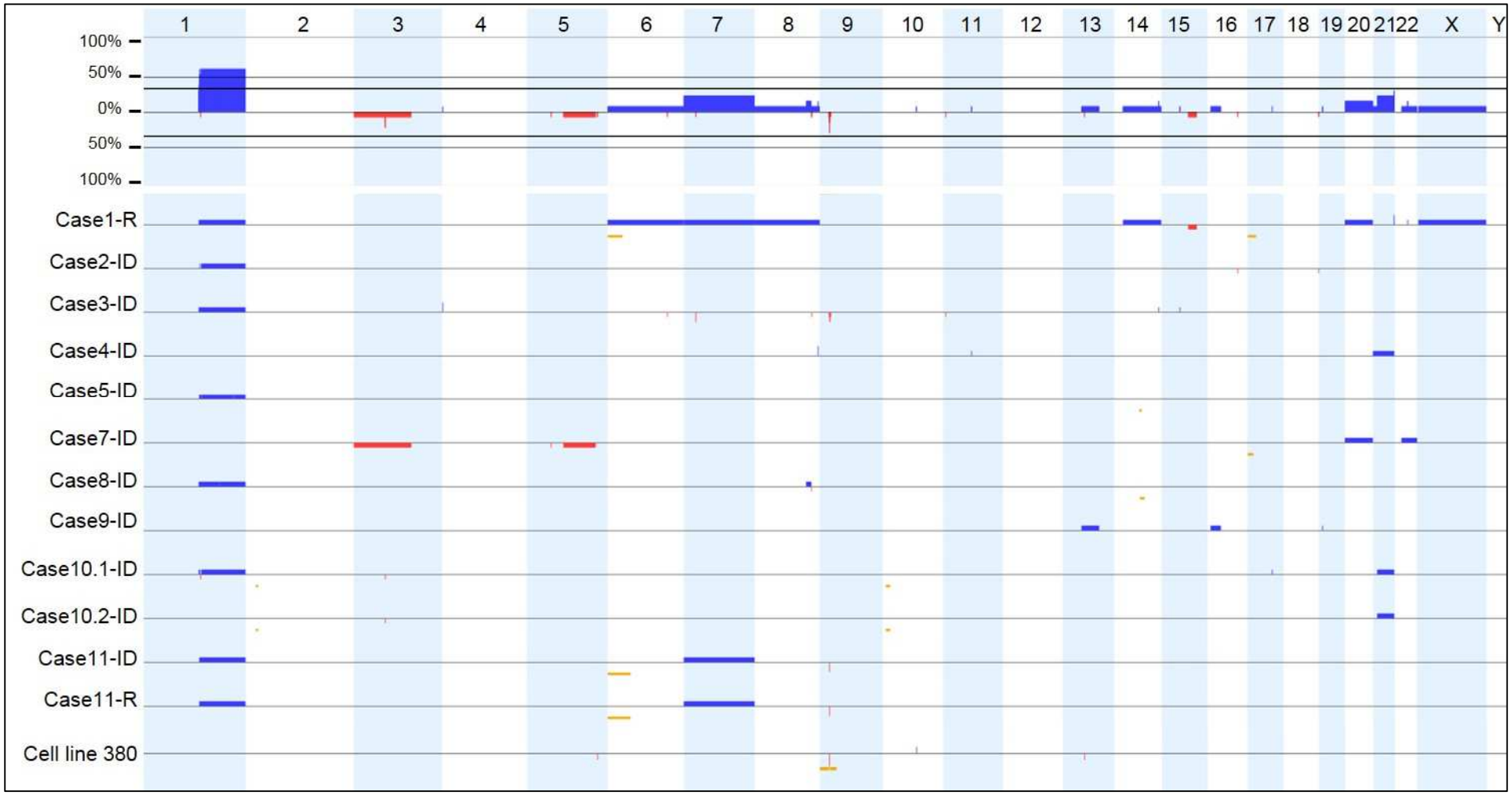
Supplemental Figure 6: Comparison of frequently mutated genes in preBLL (blue, based on 7 cases (6 patients) with sequencing data) and in BL (in green and brown). Included are those genes which were reported by Schmitz et al[28] (A) and Love et al[32] (B) to be recurrently mutated in at least 15% of cases in the respective studies as well as *NRAS* and *KRAS* (mutated in more than 2/5 preBLL patients). Using the published mutational landscapes of BL, highly recurrently mutated genes in BL were not altered in preBLL which is in agreement with the findings of the yet unpublished data (Supplemental Figure 5).



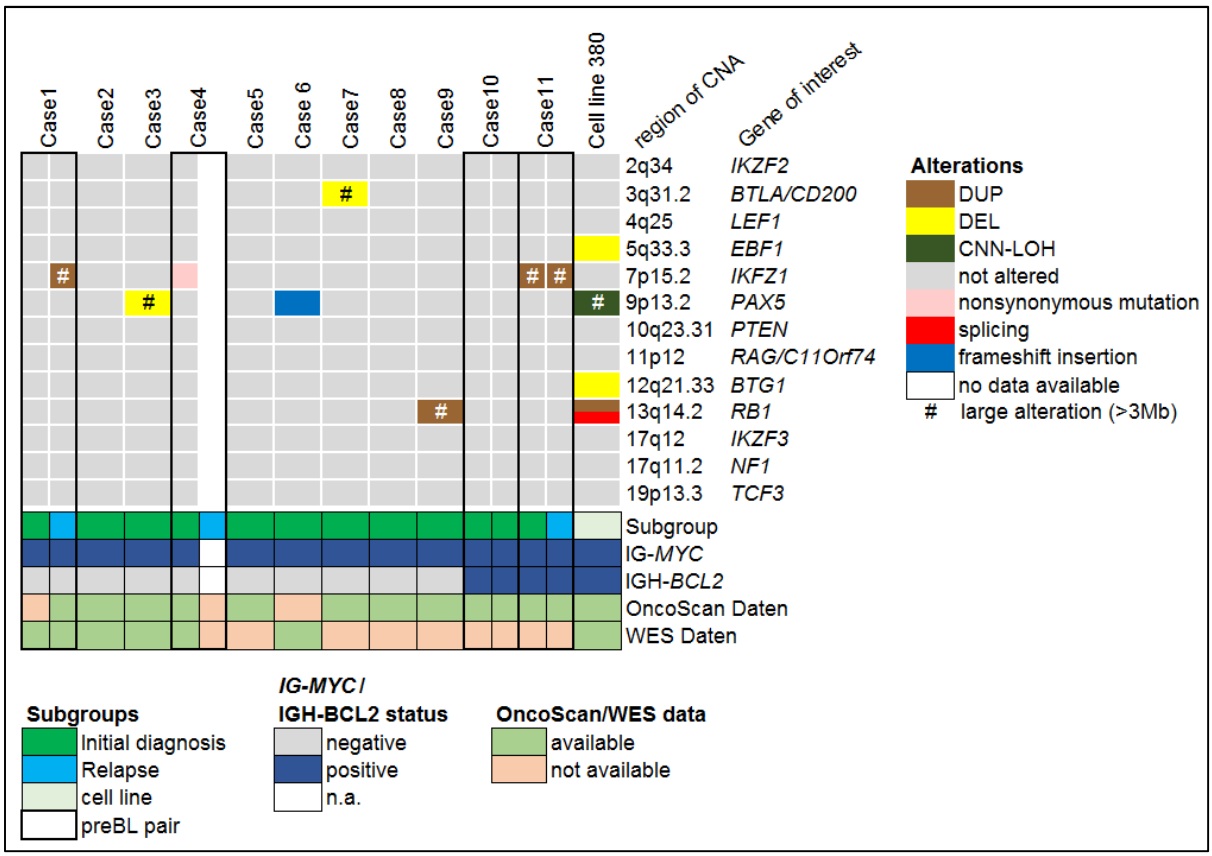
Supplemental Figure 7: Overview on recurrently mutated genes in preBLL cases. Recurrence was defined as a gene being mutated in at least 2/5 primary patients (excluding the cell line). Note that within the oncoprint Case1 is displayed with the samples of the initial diagnose as well as the relapse. *DNAH2* encoding *Dynein Axonemal Heavy Chain 2* and *FAM83H* encoding *Family With Sequence Similarity 83 Member H* are both large genes and mutations affecting those genes might represent passenger mutations.



Supplemental Figure 8: Overview on the clonal evolution of the three preBLL pairs for which samples from initial diagnosis (ID) and relapse were available.
Case1, harboring an IGH-*MYC* translocation developed a relapse eight months after ID. Both, ID and relapse, share the same IGH-*MYC* translocation as well as 70 nonsynonymous mutations as identified by exome sequencing, whereas ID harbors 7 (in red) unique mutations and the relapse 20 (in red). Refer to Supplemental Table 6 for an overview on the mutated genes.
Case4, harboring an IGK-*MYC* translocation developed a relapse 14 months after ID. The *NRAS* mutations differ in ID and relapse; in addition, ID harbors a *KRAS* mutation which was not observed in the relapse. This indicates a divergent clonal evolution of case 4 as well as that the IG-*MYC* translocation occurred prior to the *RAS* mutations.
Case11, harboring an IGL-*MYC* translocation developed a relapse 142 days after ID. The IGHV rearrangement is the same for ID and relapse (refer to Supplemental Table 4) and they share the same copy number alterations.



Supplemental Figure 9: Overview on the copy number alterations in preBLL cases analyzed by OncoScan array. The top panel depicts the recurrent alterations found in the twelve preBLL samples and the cell line 380. In the lower panel, each line corresponds to one case. Blue bars depict copy number gains, red bards copy number losses and yellow bars copy number neutral loss of heterozygosity. The most recurrently altered regions are gains in 1q as well as trisomy 7 and 21.



Supplemental Figure 10: Overview on focal copy number alterations (CNA, 100 kb – 3 Mb) as well as large alterations (>3 Mb, highlighted by #) in genes described to harbor recurrent focal CNA in ALL[33]. In addition, the mutation status of these genes was checked for cases in which WES data were available. Of note is that if a gene is indicated as not altered, this is limited due to the fact that of Case1-ID and Case6 only WES data and for cases 7 to 11 only OncoScan data are available. This analysis shows that the primary preBLL do not harbor focal alterations in genes recurrently altered in ALL. Checking the mutation status of those genes revealed only in 2/5 preBLL cases mutations affecting *IKZF1* and *PAX5* in each one preBLL.

SUPPLEMENTAL REFERENCES

1. Li Y, Gupta G, Molofsky A, et al. B lymphoblastic leukemia/lymphoma with burkitt-like morphology and IGH/MYC rearrangement: Report of 3 cases in adult patients. *Am J Surg Pathol*. 2017.

2. Mufti GJ, Hamblin TJ, Oscier DG, Johnson S. Common ALL with pre-B-cell features showing (8;14) and (14;18) chromosome translocations. *Blood*. 1983;62(5):1142-1146.

3. Roug AS, Wendtland P, Bendix K, Kjeldsen E. Supernumerary isochromosome 1, idic(1)(p12), leading to tetrasomy 1q in burkitt lymphoma. *Cytogenet Genome Res*. 2014;142(1):7-13.

4. Pegoraro L, Palumbo A, Erikson J, et al. A 14;18 and an 8;14 chromosome translocation in a cell line derived from an acute B-cell leukemia. *Proc Natl Acad Sci U S A*. 1984;81(22):7166-7170.

5. Willis TG, Jadayel DM, Coignet LJ, et al. Rapid molecular cloning of rearrangements of the IGHJ locus using long-distance inverse polymerase chain reaction. *Blood*. 1997;90(6):2456-2464.

6. Tsujimoto Y, Gorham J, Cossman J, Jaffe E, Croce CM. The t(14;18) chromosome translocations involved in B-cell neoplasms result from mistakes in VDJ joining. *Science*. 1985;229(4720):1390-1393.

7. Haluska FG, Finver S, Tsujimoto Y, Croce CM. The t(8; 14) chromosomal translocation occurring in B-cell malignancies results from mistakes in V-D-J joining. *Nature*. 1986;324(6093):158-161.

8. Schlegelberger B, Zwingers T, Harder L, et al. Clinicopathogenetic significance of chromosomal abnormalities in patients with blastic peripheral B-cell lymphoma. kiel-wien-lymphoma study group. *Blood*. 1999;94(9):3114-3120.

9. Gesk S, Martin-Subero JI, Harder L, et al. Molecular cytogenetic detection of chromosomal breakpoints in T-cell receptor gene loci. *Leukemia*. 2003;17(4):738-745.

10. Ventura RA, Martin-Subero JI, Jones M, et al. FISH analysis for the detection of lymphoma-associated chromosomal abnormalities in routine paraffin-embedded tissue. *J Mol Diagn*. 2006;8(2):141-151.

11. Martin-Subero JI, Harder L, Gesk S, et al. Interphase FISH assays for the detection of translocations with breakpoints in immunoglobulin light chain loci. *Int J Cancer*. 2002;98(3):470-474.

12. Richter J, Schlesner M, Hoffmann S, et al. Recurrent mutation of the ID3 gene in burkitt lymphoma identified by integrated genome, exome and transcriptome sequencing. *Nat Genet*. 2012;44(12):1316-1320.

13. Li H, Durbin R. Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics*. 2009;25(14):1754-1760.

14. Sherry ST, Ward MH, Kholodov M, et al. dbSNP: The NCBI database of genetic variation. *Nucleic Acids Res*. 2001;29(1):308-311.

15. Jabs J, Zickgraf FM, Park J, et al. Screening drug effects in patient-derived cancer cells links organoid responses to genome alterations. *Mol Syst Biol*. 2017;13(11):955.

16. Lek M, Karczewski KJ, Minikel EV, et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature*. 2016;536(7616):285-291.

17. Nik-Zainal S, Alexandrov LB, Wedge DC, et al. Mutational processes molding the genomes of 21 breast cancers. *Cell*. 2012;149(5):979-993.

18. Krober A, Seiler T, Benner A, et al. V(H) mutation status, CD38 expression level, genomic aberrations, and survival in chronic lymphocytic leukemia. *Blood*. 2002;100(4):1410-1416.

19. Sahm F, Toprak UH, Hubschmann D, et al. Meningiomas induced by low-dose radiation carry structural variants of NF2 and a distinct mutational signature. *Acta Neuropathol*. 2017;134(1):155-158.

20. Oakes CC, Seifert M, Assenov Y, et al. DNA methylation dynamics during B cell maturation underlie a continuum of disease phenotypes in chronic lymphocytic leukemia. *Nat Genet*. 2016;48(3):253-264.

21. Kulis M, Merkel A, Heath S, et al. Whole-genome fingerprint of the DNA methylome during human B cell differentiation. *Nat Genet*. 2015;47(7):746-756.

22. Lee ST, Muench MO, Fomin ME, et al. Epigenetic remodeling in B-cell acute lymphoblastic leukemia occurs in two tracks and employs embryonic stem cell-like signatures. *Nucleic Acids Res*. 2015;43(5):2590-2602.

23. Bergmann AK, Castellano G, Alten J, et al. DNA methylation profiling of pediatric B-cell lymphoblastic leukemia with KMT2A rearrangement identifies hypomethylation at enhancer sites. *Pediatr Blood Cancer*. 2017;64(3):10.1002/pbc.26251. Epub 2016 Oct 27.

24. Nordlund J, Backlin CL, Wahlberg P, et al. Genome-wide signatures of differential DNA methylation in pediatric acute lymphoblastic leukemia. *Genome Biol*. 2013;14(9):r105-2013-14-9-r105.

25. Paradis E, Claude J, Strimmer K. APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics*. 2004;20(2):289-290.

26. Habib T, Park H, Tsang M, et al. Myc stimulates B lymphocyte differentiation and amplifies calcium signaling. *J Cell Biol*. 2007;179(4):717-731.

27. Küppers R, Dalla-Favera R. Mechanisms of chromosomal translocations in B cell lymphomas. *Oncogene*. 2001;20(40):5580-5594.

28. Schmitz R, Young RM, Ceribelli M, et al. Burkitt lymphoma pathogenesis and therapeutic targets from structural and functional genomics. *Nature*. 2012;490(7418):116-120.

29. Aukema SM, Kreuz M, Kohler CW, et al. Biological characterization of adult MYC-translocation-positive mature B-cell lymphomas other than molecular burkitt lymphoma. *Haematologica*. 2014;99(4):726-735.

30. Swerdlow SH, Campo E, Pileri SA, et al. The 2016 revision of the world health organization classification of lymphoid neoplasms. *Blood*. 2016;127(20):2375-2390.

31. Kretzmer H, Bernhart SH, Wang W, et al. DNA methylome analysis in burkitt and follicular lymphomas identifies differentially methylated regions linked to somatic mutation and transcriptional control. *Nat Genet*. 2015;47(11):1316-1325.

32. Love C, Sun Z, Jima D, et al. The genetic landscape of mutations in burkitt lymphoma. *Nat Genet*. 2012;44(12):1321-1325.

33. Mullighan CG, Downing JR. Genome-wide profiling of genetic alterations in acute lymphoblastic leukemia: Recent insights and future directions. *Leukemia*. 2009;23(7):1209-1218.