Supplemental Information

Supplemental Materials and Methods

Treatment, sample preparation and next generation sequencing

The backbone of the treatment regimen is identical for all patients treated according to the protocols NHL-BFM 95, LBL Registry, trial EURO-LB02 and NHL-BFM Registry 2012. All protocols are based on the most successful clinical trial for pediatric patients with lymphoblastic lymphoma, the trial NHL-BFM 90[1]. There were only minor changes in the subsequent protocols. The main difference between NHL-BFM 95 and subsequent protocols LBL Registry, EURO-LB 02 and NHL-BFM Registry 2012 are: i) the dose of. E. coli asparaginase in induction which was 5,000 U/m$^{2.}$ compared with 10,000 U/m^2 in subsequent protocols and ii) the recommendation of two additional doses of intrathecal methotrexate in induction for all patients independent of their CNS status at diagnosis. The vast majority of the treatment protocols was identical. The trial EURO-LB 02 addressed two randomized questions[2]. Both questions could not be evaluated on confirmatory statistics due to the early stop of patient recruitment. Descriptive statistics did not show a difference in outcome according to randomization arm. The randomization results for randomized questions are detailed in the supplemental Table 1.

Genomic DNA from primary (n=122) and relapse samples (n=9) was isolated from fresh frozen cells (pleural effusions (n=79), tumor biopsies (n=42), bone marrow (n=4) and pericardial effusions and malignant effusions (n=5)). Samples were selected based on the availability of sufficient tumor material, blast count/pathological cells (80-90%).

Out of 131 samples three samples TP30, TP31, TP55 failed for targeted sequencing because of the poor quality of DNA and TP100 was excluded as it did not have enough blast count. 127 evaluable samples were analyzed which included 118 primary samples that included 21 (~18%) primary samples that suffered relapse, 9 relapse samples (supplementary Figure 1A). 68% of the samples overlap with the published T-LBL cohort[3] and 32% of the samples were newly added from patients registered between 2012-2018 in the NHL-BFM center and have not been analyzed for molecular characterization so far. Genomic DNA was isolated (using QIAamp DNA Mini Kit or QIAamp Blood DNA Mini Kit as per the manufacturer's instructions. Genomic DNA was sequenced as paired-end run (2x 150 bp) on an Illumina NextSeq 500 (Illumina, San Diego, USA). After enzymatic fragmentation of the DNA and pre-capture amplification of the previously selected regions (RefSeq exons of selected genes + 25 bp of the intronic regions) were enriched by pre-pooling hybrid-capture technique. Following hybridization, unique barcodes for sample identification were added by a post-capture amplification. For these steps, the Agilent SureSelectQXT protocol was used (Agilent Technologies, Santa Clara, USA). Quality controls during the protocol and for the final and pooled libraries were performed on a TapeStation (Agilent). Raw data quality was checked using the Illumina Sequencing Viewer (Illumina) and an in-house pipeline.

Targeted sequencing

A targeted panel for T-LBL was designed based on the data obtained from WES of 16 patient samples. The selected genes from WES data for the panel displayed non synonymous somatic mutations and were predicted to be pathogenic or damaging in the functional prediction software tools like PROVEAN (http://provean.jcvi.org/index.php).

Also some genes that were relevant for T-ALL pathogenesis based on literature were also included in the panel.[4,5] Three different versions of the panel were used which had slight variation either as addition or exclusion of very few genes. Version T-LBLv1 was used for analysis of samples from batch 1, T-LBLv2 was used for analysis of samples from batch2 and T-LBLv3 for batch 3, 4 and 5.

Sanger sequencing

PCR products were amplified with forward and reverse primers for respective genes and were treated using ExoSap IT (USB Corporation) and sequenced with ABI Prism BigDye terminator v3.1 (Applied Biosystems). Sequencing reactions were run on an ABI-3730 automated sequencer (Applied Biosystems). All sequences were analyzed using CodonCode Aligner software to identify mutations.

For *BCL11b*, three hotspots spanning exon2[6] and exon4[7] were amplified and sequenced to identify mutations. Detailed information on the primer sequences and mutations identified is provided the supplement Table 6 and supplemental Table 7.

Analysis of WES and targeted sequencing samples

Sequencing data was aligned to the reference genome GRCh37.67 using BWA mem (Li, 2013). For WES Duplicate reads were removed using Picardtools (http://broadinstitute.github.io/picard/).

For WES somatic SNV calling SomaticSniper [8] was used, using a high-confidence filter (mapping quality >40, somatic score >40). For somatic indel calling Strelka[9] was used with the default quality filter settings. Variant calling was performed with the two tools based on their performance in the review articles by Roberts et al.[10] and Xu et al.[11]

To increase sensitivity and specificity of the results, two post-processing steps were applied on the raw variant calling output. 1) Variants with evidence in any germline sample were removed. All variants with minimum VAF of 0.02, minimum number of alternate reads of 2, minimum overall coverage of 10 and minimum base quality of 15 (SNVs only) were considered. This filter allows to remove germline calls that initially have been mislabeled as somatic due to low coverage in the matching germline samples. 2) Variants present in only one, either the primary tumor sample or the relapse sample, were re-evaluated with increased sensitivity (minimum VAF of 0.02, minimum number of alternate reads of 2, minimum overall coverage of 25 and minimum base quality of 15 (SNVs only)). This re-evaluation enabled to detect variants that were present with very low VAF at one of the two time points that were considered.

Subsequently, all calls were annotated with SnpEff[12] using annotations based on RefSeq (hg19). All calls annotated as 3'- or 5'UTR variant, downstream or upstream, intergenic-, intron- or non-coding exon variant, or splice region variant (additionally annotated as intron) were excluded. For the remaining calls, exemplary validation experiments were performed. All samples – germline, tumor and relapse – were re-analyzed by targeted sequencing, analyzing 76-80 genes that were frequently mutated in the WES data and/or were known to be associated with T-LBL. Additionally, few selected set of calls were analyzed with Sanger sequencing.

Altogether, 111 calls were validated as somatic mutations by targeted sequencing and/or Sanger sequencing. Sixty-six calls were identified as being false positive calls (artifacts or germline calls). Based on these validation results, two quality filters were applied on the preliminary results.

1) Analysis of base quality distribution revealed that artifacts were characterized by a significantly lower Phred-score compared to somatic mutations. Density estimation revealed an optimum threshold of 24.94. All calls with average base quality below this threshold were filtered as artifacts. As base quality can only be determined for SNVs, all indels were inspected manually.

2) Germline calls were usually characterized by being present in common polymorphism databases, while somatic mutations were often present in common mutation databases. Analysis was performed for the presence of every call in the polymorphism databases ESP6500 (http://evs.gs.washington.edu/EVS/), 1000 Genomes (The 1000 Genomes Project Consortium, 2015), ExAC[13], dbSNP (build 138 excluding sites after 129)[14] and ClinVar (common and no known medical impact, 03.02.2016).[15] Additionally, dbSNP (build 138), ClinVar (common and clinical, 03.02.2016) and Cosmic were considered.[16] Also the type of the mutation, the VAF and the coverage in the tumor- and germline sample were taken into consideration. All calls present in at least one of the polymorphism databases were manually evaluated by two independent experts. Targeted sequencing data was analyzed using the appreci8 pipeline using default parameters.[17] In addition to automatic classification of the variant calls – as either pathogenic mutations, polymorphisms or artifacts - by the pipeline, we manually investigated all variant calls.

Genotyping Analysis

Genotypes were generated using Illumina's Infinium HumanOmni2-5Exome BeadArray following the manufacturer's automatized workflow. Arrays were scanned using the iScan BeadArray scanner and analyzed by GenomeStudio 2011.1 software both produced by Illumina, San Diego, US. The quality control checked the process controls in GenomeStudio's Control Dashboard and standard performance parameters as call rate,

B allele frequency plots, and excess of heterozygosity. Genotypes were exported in Illumina forward allele designation, positions based on GRCh37.2 genome. Copy number analysis was performed applying Illumina's CNV-Partition v3.1.6 algorithm.

Methylation Arrays

Methylation arrays were performed on 22 T-LBL samples (n=16 primary+6 relapse) representing 16 cases along with corresponding germline samples with Infinium MethylationEPIC BeadChip Kit[18] produced by Illumina, San Diego, US. DNA samples were processed following the manufacturers guidelines for the automatized workflow. Bisulfite conversion was performed using Zymo Research, Freiburg, DE, EZ-96 conversion kits. Quality control was performed using Illumina's GenomeStudio 2011.1, checking process control probes and for probes failing detection p-value thresholds.

Methylation Data Filtering, Imputation and Normalization

Probes were filtered on satisfying any one of the following criteria: probes with a detection p-value above 0.01 in at least 20% of samples; Probes with a beadcount < 3 in at least 5% of samples; non-CpG probes (CpA, CpT, and CpC); Probes with single nucleotide polymorphisms (SNPs) as identified in[18]; Probes that align to multiple locations as identified in[19]; Probes on the X or Y chromosome. Most of these criteria are default settings of ChAMP package for probe filtering[20]. All “not available” (NA) values after filtering were imputed by KNN (k-nearest neighbors, k = 3) algorithm. BMIQ (Beta MIxture Quantile dilation) function[21] was applied for normalization. Probe filtering, imputation and normalization were done using ChAMP package.

DMP Analysis and Hierarchical Clustering

Limma powers differential expression analysis and Benjamini-Hochberg (BH) p-value adjustment method from ChAMP's champ. DMP() function was used for detection of differentially methylated position (DMP). Details of Limma package and Limma analysis has been described by Ritchie ME et al.[22] Beta values of detected DMPs or most variant probes based on variance were used to plot heatmap using pheatmap package, sorted by hierarchical clustering. For hierarchical clustering, euclidean distance and complete linkage were used.

Multidimensional Scaling

Non-metric multidimensional scaling was used to discriminate the differential methylation of germline and tumor sample groups. Euclidean distances calculated from beta values of detected DMPs were used to perform Kruskal's non-metric multidimensional scaling with the isoMDS() function in R.

MLPA

LOH6q analysis

LOH6q analysis was performed for the newly added T-LBL samples with SALSA MLPA P200 reference 1 in combination with custom made probes and SALSA MLPA EK1 reagent kit-FAM (MRC Holland) as described in Rohde M, et. al.[23] Briefly, ~50 ng of genomic DNA from the tumor samples, genomic DNA from healthy individuals as reference samples and samples from LOH6q positive samples identified from previous analysis was used as positive controls and MLPA reactions were performed as per the

manufacturer's instruction. The data was evaluated using Coffalyser.Net MLPA analysis software (MRC-Holland).

PTEN deletions

Genomic DNA (~50 ng) isolated from fresh frozen tumor (pleural effusions, bone marrow, lymph nodes) or from peripheral blood from healthy individuals were used as controls (DNeasy Blood and Tissue kit: Qiagen, Hilden, Germany). The commercial panel, SALSA MLPA probemix P383T-A2 T-ALL (MRC-Holland; Amsterdam, The Netherlands) was used to screen *PTEN* deletions. The MLPA assay was performed as per the manufacturer's instructions. The probe mix includes 4 probes for *PTEN* gene spanning exon 1(1probe), exon 7(2 probes) and exon 9(1probe). The data was evaluated using Coffalyser.Net MLPA analysis software (MRC-Holland). The results were normalized and compared with reference samples which included positive controls (genomic DNA from the cell line (CEM, T-ALL cell line) or patient samples which were identified to have deletions in *PTEN*)) and DNA from healthy individuals. The samples which were positive for deletions were analyzed one additional time to reconfirm the results.

ABD analysis

Single tube multiplex TRG PCR was performed as described by Derrieux C et. al.[24] The multiplex PCR assay includes primers for all conserved regions involved in rearrangements in T-cell lymphoid malignancies. PCRs were performed with ~100 ng of genomic DNA isolated from either pleural effusions, tumor biopsy, bone marrow and peripheral blood samples, 1U Ampli*Taq* gold (Life Technology 4317742), MgCl2 and dNTPs were added at final concentrations of 5mM and 0.2mM, respectively. Denaturation

of DNA was performed at 94°C for 8 minutes, followed by 35 cycles comprising of following steps; denaturation for 45 seconds, primer annealing for 1 minute at 57°C, and extension for 1 minute 30 seconds at 72°C. Depending on the sensitivity of the instrument, PCR products were diluted to receive optimal peak heights. PCR products were mixed with Hi-Di formamide and LIZ size standard, denatured at 95°C for 3 minutes and were size separated on a high resolution analysis on an ABI PRISM 3130 GeneScan (Perkin-Elmer Applied Biosystems). The .fsa files are imported and analyzed with the help of software GeneMapper™ (Thermo Fisher Scientific) or “Microsatellite analysis” tool on Thermo Fisher Connect, a cloud-based analysis platform (Thermo Fisher Scientific). The Microsatellite Analysis tool provides all the functions required visualizing the peak pattern for ABD analysis.

The PCR products obtained are in the range of 159-207 bp. The ABD positive samples display a single peak and ABD negative samples display two peaks. Occasionally polyclonal peaks are also detected in ABD positive samples, in these cases the tumor content is checked and confirmed to be of significant percentage.

Supplemental Tables

Supplemental Table 1. List of patient samples

Sample. ID	Protocol	sex	Sample material	EURO-LB 02 Randomization R1 / R2
TP1	EURO-LB 02	male	PL	Dexamethasone / 24 months
TP2	EURO-LB 02	male	PL	Prednisone
TP3	LBL Registry	female	PL	
TP4	EURO-LB 02	male	TU	Prednisone
TP5	EURO-LB 02	male	PL	
TP6	LBL Registry	male	PL	
TP7	EURO-LB 02	male	PL	Dexamethasone / 24 months
TP8	LBL Registry	male	PL	
TP9	EURO-LB 02	female	PL	Prednisone
TP10	LBL Registry	female	PL	
TP11	EURO-LB 02	male	PL	Prednisone / 24 months
TP12	LBL Registry	male	EX	
TP13	LBL Registry	male	TU	
TP14	LBL Registry	male	PL	
TP15	EURO-LB 02	male	PL	
TP16	NHL-BFM Registry 2012	male	TU	
TP17	NHL-BFM Registry 2012	male	PL	
TP18	NHL-BFM Registry 2012	male	PL	
TP19	NHL-BFM Registry 2012	male	TU	
TP20	NHL-BFM Registry 2012	male	TU	
TP21	NHL-BFM Registry 2012	male	TU	
TP22	NHL-BFM Registry 2012	male	TU	
TP23	NHL-BFM Registry 2012	female	PL	
TP24	NHL-BFM Registry 2012	male	PL	
TP25	NHL-BFM Registry 2012	male	PL	
TP26	NHL-BFM Registry 2012	male	PL	
TP27	NHL-BFM Registry 2012	male	TU	
TP28	NHL-BFM Registry 2012	male	TU	
TP29	NHL-BFM Registry 2012	female	PL	
TP30	EURO-LB 02	male	TU	
TP31	EURO-LB 02	male	TU	
TP32	NHL-BFM 95	female	PL	
TP33	EURO-LB 02	female	PL	Prednisone
TP34	NHL-BFM 95	male	PL	

Supplemental Table 1 cont.				
Sample. ID	Study	sex	Sample material	EURO-LB 02 Randomization R1 / R2
TP35	NHL-BFM 95	male	PL	
TP36	NHL-BFM 95	female	PL	
TP37	EURO-LB 02	male	TU	Dexamethasone
TP38	EURO-LB 02	female	PL	
TP39	EURO-LB 02	male	PL	Dexamethasone / 24 months
TP40	EURO-LB 02	female	PL	Prednisone / 18 months
TP41	EURO-LB 02	male	TU	Dexamethasone / 24 months
TP42	EURO-LB 02	male	PL	Prednisone
TP43	LBL Registry	female	PL	
TP44	NHL-BFM 95	female	PL	
TP45	NHL-BFM 95	male	PL	
TP46	NHL-BFM 95	female	PL	
TP47	EURO-LB 02	male	TU	Prednisone / 24 months
TP48	NHL-BFM 95	female	TU	
TP49	NHL-BFM 95	male	PL	
TP50	NHL-BFM 95	male	PL	
TP51	NHL-BFM 95	female	PL	
TP52	NHL-BFM 95	male	TU	
TP53	EURO-LB 02	male	PL	
TP54	LBL Registry	male	PL	
TP55	LBL Registry	female	PL	
TP56	LBL Registry	male	PL	
TP57	LBL Registry	female	PL	
TP58	LBL Registry	female	TU	
TP59	EURO-LB 02	female	TU	Dexamethasone / 24 months
TP60	EURO-LB 02	female	TU	Dexamethasone / 24 months
TP61	EURO-LB 02	male	TU	Prednisone / 18 months
TP62	NHL-BFM 95	female	PL	
TP63	LBL Registry	male	TU	
TP64	NHL-BFM 95	male	TU	
TP65	EURO-LB 02	male	TU	Dexamethasone / 18 months
TP66	NHL-BFM 95	female	PL	
TP67	LBL Registry	male	PL	
TP68	LBL Registry	male	PL	
TP69	LBL Registry	male	PL	
TP70	LBL Registry	female	TU	
TP71	LBL Registry	male	PL	

213

Supplemental Table 1 cont.				
Sample. ID	Study	sex	Sample material	EURO-LB 02 Randomization R1 / R2
TP72	LBL Registry	female	PL	
TP73	LBL Registry	male	PL	
TP74	EURO-LB 02	male	PL	Prednisone / 18 months
TP75	LBL Registry	male	PL	
TP76	LBL Registry	female	PL	
TP77	EURO-LB 02	male	TU	Dexamethasone / 24 months
TP78	EURO-LB 02	male	TU	Dexamethasone
TP79	EURO-LB 02	male	TU	Prednisone
TP80	EURO-LB 02	male	TU	Dexamethasone / 18 months
TP81	EURO-LB 02	male	PL	Dexamethasone / 18 months
TP82	EURO-LB 02	male	PL	Prednisone / 24 months
TP83	EURO-LB 02	male	PL	
TP84	EURO-LB 02	female	PK	Prednisone / 24 months
TP85	LBL Registry	male	TU	
TP86	LBL Registry	male	PL	
TP87	LBL Registry	female	PL	
TP88	LBL Registry	male	PL	
TP89	LBL Registry	male	PK	
TP90	LBL Registry	male	PL	
TP91	LBL Registry	male	PL	
TP92	LBL Registry	female	PL	
TP93	LBL Registry	female	PL	
TP94	LBL Registry	female	TU	
TP95	LBL Registry	male	PL	
TP96	EURO-LB 02	female	TU	
TP97	LBL Registry	male	PL	
TP98	NHL-BFM Registry 2012	male	PL	
TP99	NHL-BFM Registry 2012	male	PK	
TP100	NHL-BFM Registry 2012	male	GL	
TP101	NHL-BFM Registry 2012	male	PL	
TP102	NHL-BFM Registry 2012	male	PL	
TP103	NHL-BFM Registry 2012	male	PL	
TP104	NHL-BFM Registry 2012	male	PL	
TP105	NHL-BFM Registry 2012	male	PL	
TP106	NHL-BFM Registry 2012	male	TU	
TP107	NHL-BFM Registry 2012	male	TU	

214

Supplemental Table 1 cont.				
Sample. ID	Study	sex	Sample material	EURO-LB 02 Randomization R1 / R2
TP108	NHL-BFM Registry 2012	male	TU	
TP109	NHL-BFM Registry 2012	female	PL	
TP110	NHL-BFM Registry 2012	male	TU	
TP111	NHL-BFM Registry 2012	male	TU	
TP112	NHL-BFM Registry 2012	male	TU	
TP113	NHL-BFM Registry 2012	male	PL	
TP114	NHL-BFM Registry 2012	male	TU	
TP115	NHL-BFM Registry 2012	male	PL	
TP116	NHL-BFM Registry 2012	male	TU	
TP117	NHL-BFM Registry 2012	male	PL	
TP118	NHL-BFM 95	male	PL	
TP119	NHL-BFM Registry 2012	male	PL	
TP120	EURO-LB 02	male	TU	
TP121	NHL-BFM Registry 2012	male	PL	
TP122	NHL-BFM Registry 2012	male	PL	
TR11	EURO-LB 02	male	PL	Prednisone / 18 months
TR12	LBL Registry	male	EX	
TR13	LBL Registry	male	TU	
TR14	LBL Registry	male	PL	
TR15	LBL Registry	male	TU	
TR16	NHL-BFM Registry 2012	male	BM	
TR17	NHL-BFM Registry 2012	male	BM	
TR18	NHL-BFM Registry 2012	male	BM	
TR19	NHL-BFM Registry 2012	male	BM	

Supplemental Table 1. List of patient identification (ID), study protocol and sample material of the analyzed cases. PL: pleural effusion (n=79), TU: tumor biopsy (n=42), PK: pericardial effusion (n=3), Ex: malignant effusion (n=2), BM: bone marrow (n=4), GL: germ line. The number of cases from different studies in the analyzed cohort are as follows: NHL-BFM 95, n=16; EURO-LB 02, n=35; LBL Registry, n=39; NHL-BFM Registry 2012, n=41.

Supplemental Table 2. List of genes from the targeted panel

	T-LBLv1	T-LBLv2	T-LBLv3
	batch 1	batch2	batch3,4,5
1	ABCC6	ABCC6	ABCC6
2	AGO2	AGO2	AGO2
3	ANXA2	ANXA2	ANXA2
4	ASS1	ASS1	ASS1
5	BRAF	BRAF	BRAF
6	CAM2G	CAM2G	CAM2G
7	CCNC	CCNC	CCNC
8	CDC27	CDC27	CDC27
9	CDKN1B	CDKN1B	CDKN1B
10	CCND3	CCND3	CCND3
11	CHFR	CHFR	CHFR
12	CISH	CISH	CISH
13	CPEB3	CPEB3	CPEB3
14	CREBBP	CREBBP	CREBBP
15	CSMD1	CSMD1	CSMD1
16	CTCF	CTCF	CTCF
17	DDX3X	DDX3X	DDX3X
18	DNM2	DNM2	DNM2
19	DNMT3A	DNMT3A	DNMT3A
20	EHBP1	EHBP1	EHBP1
21	FBXW7	FBXW7	FBXW7
22	FZD9	FZD9	FZD9
23	GNAI2	GNAI2	GNAI2
24	HCK	HCK	HCK
25	HLA-DRB5	HLA-DRB5	HLA-DRB5
26	HNRNPK	HNRNPK	HNRNPK
27	JAK2	JAK2	JAK2
28	JAK3	JAK3	JAK3
29	KDM3A	KDM3A	KDM3A
30	KDM6B	KDM6B	KDM6B
31	KMT2C	KMT2C	KMT2C
32	KMT2D	KMT2D	KMT2D
33	KRAS	KRAS	KRAS
34	MAPK3	MAPK3	MAPK3
35	MCM10	MCM10	MCM10
36	MCM9	MCM9	MCM9
37	MDC1	MDC1	MDC1
38	MED12	MED12	MED12

Supp.Table 2 contd.			
	T-LBLv1	T-LBLv2	T-LBLv3
	batch 1	batch2	batch3,4,5
39	MFHAS1	MFHAS1	MFHAS1
40	MIB2	MIB2	MIB2
41	MLH1	MLH1	MLH1
42	MTRR	MTRR	MTRR
43	MYCBP2	MYCBP2	MYCBP2
44	NEK1	NEFM	NEFM
45	NOS3	NEK1	NEK1
46	NOTCH1	NOS3	NOS3
47	NOTCH3	NOTCH1	NOTCH1
48	NRAS	NOTCH3	NOTCH3
49	NT5C1A	NRAS	NRAS
50	NT5C2	NT5C1A	NT5C1A
51	PCDH7	NT5C2	NT5C2
52	PHF6	PCDH7	PCDH7
53	PIAS3	PHF6	PHF6
54	PIK3CA	PIAS3	PIAS3
55	PTEN	PIK3CA	PIK3CA
56	PTPRT	PTEN	PTEN
57	RIN1	PTPRT	PTPRT
58	ROR2	RIN1	RIN1
59	RPL22	ROR2	ROR2
60	RUNX1	RPL22	RPL22
61	SETD1B	RUNX1	RUNX1
62	SON	SETD1B	SETD1B
63	STAT5B	SON	SON
64	STK39	STAT5B	STAT5B
65	TAL1	STK39	STK39
66	TCF3	TAL1	TAL1
67	TP53	TCF3	TCF3
68	TRAF7	TFAP4	TFAP4
69	TTN*	TP53	TP53
70	UCK2	TRAF7	TRAF7
71	USH2A	TTN	UCK2
72	WT1	UCK2	USH2A
73	XIST*	USH2A	WT1
74	TRAF6	WT1	ZBTB24
75	MUC4	ZBTB24	XIST*
76	LAMA2	XIST*	USP24*

226

Supp. Table 2 contd.			
	T-LBLv1	T-LBLv2	T-LBLv3
	batch 1	batch2	batch3,4,5
77	ZNF208		USP7
78	ZNF91		IL7R
79			KDM6A
80			ZNF91
Samples	TG1-TG16 TP1-TP16	TG17-TG29 TP17-TP29 TR11-TR16	TP30-TP122 TR116, TR123 and TR124

Supplemental Table 2. List of genes from the different versions of the targeted panel (v1, v2 and v3) used for the targeted sequencing of the extended cohort. (* indicates that for some of the large genes only selected regions were included to design the probes).

Supplemental Table 3. Mutation rates of samples analyzed by whole exome sequencing

Sample	Mutations	Target size	Mutation rate/Mbp	VAF<0.15	%	VAF>=0.15	%
TP1	21	65343296	0.32	4	19.05	17	80.95
TP2	27	65343296	0.41	0	0.00	27	100.00
TP3	20	65343296	0.31	0	0.00	20	100.00
TP4	12	65343296	0.18	1	8.33	11	91.67
TP5	17	65343296	0.26	1	5.88	16	94.12
TP6	28	65343296	0.43	1	3.57	27	96.43
TP7	3	65343296	0.05	0	0.00	3	100.00
TP8	37	65343296	0.57	0	0.00	37	100.00
TP9	23	65343296	0.35	2	8.70	21	91.30
TP10	42	65343296	0.64	0	0.00	42	100.00
TP11	30	65343296	0.46	0	0,00	30	100.00
TR11	56	65343296	0.86	1	1.79	55	98.21
TP12	15	65343296	0.23	1	6.67	14	93.33
TR12	14	65343296	0.21	2	14.29	12	85.71
TP13	16	65343296	0.24	0	0.00	16	100.00
TR13	59	65343296	0.90	0	0.00	59	100.00
TP14	17	65343296	0.26	1	5.88	16	94.12
TR14	518	65343296	7.93	3	0.58	515	99.42
TP15	24	65343296	0.37	1	4.17	23	95.83
TR15	189	65343296	2.89	18	9.52	171	90.48

Supplemental Table 3. Mutation rates of samples from the cases of “limited cohort” screened by whole exome sequencing.

Supplemental Table 4. Top-10 significant genes as reported by MutSigCV

gene	expr	reptime	hic	N_Non silent	N_ silent	N_non coding	n_non silent	n_ silent	n_non coding	nnei	x	X	p	q
NOTCH1	1489379	212	43	364840	103160	0	16	1	0	50	1	1275080	1,53E-07	2,88E-03
FBXW7	117295	501	51	122080	32720	0	6	1	0	50	2	1348780	4,66E-04	1
CDKN1B	791330	136	44	28480	7820	0	2	0	0	50	0	1141080	2,87E-03	1
ZSCAN22	1621097	406	-25	69200	19840	0	2	0	0	50	0	1116840	3,06E-03	1
NRAS	600650	472	11	28140	7260	0	3	0	0	50	2	1181460	3,10E-03	1
STAT5B	2101162	190	31	115400	30760	0	5	0	0	50	0	1214200	8,28E-03	1
PIK3R6	423397	636	11	108020	32380	0	2	0	0	50	0	1162080	8,33E-03	1
DCAF7	934619	247	-1	49440	13800	0	2	0	0	50	0	994720	9,47E-03	1
CNFN	792230	218	42	16500	4560	0	1	0	0	50	1	1377000	1,51E-02	1
HMGN2	1003175	167	31	14140	3440	0	1	0	0	50	2	1095960	1,64E-02	1

Supplemental Table 4. Top-10 significant genes as reported by MutSigCV.

Details of the column names: N_nonsilent, number of covered sequenced bases containing non-silent mutations; N_silent, number of covered sequenced bases containing silent mutations; N_noncoding, number of covered sequenced bases containing noncoding mutations; n_nonsilent, number of non-silent mutations; n_silent, number of silent mutations; n_noncoding, number of noncoding mutations; nnei, number of neighboring genes; x, number of mutated bases in neighboring genes; X total number of bases related to neighboring genes; p, p-value; q, q-value, i.e. adjusted p-value. Note. The table is provided as excel files separately

Supplement Table 5. Mutational status of KMT2D, PTEN, NOTCH1 and FBXW7

Sample	KMT2D	PTEN	NOTCH1	FBXW7
TP1	0	0	1	1
TP2	0	0	0	0
TP3	0	0	0	1
TP4	0	0	1	1
TP5	0	0	0	0
TP6	0	0	1	0
TP7	0	0	1	0
TP8	0	0	1	1
TP9	0	0	1	0
TP10	0	0	1	1
TP11	0	0	1	0
TP12	0	0	1	0
TP13	0	0	1	0
TP14	0	0	0	0
TP15	0	0	0	0
TP16	0	1	1	0
TP17	0	0	1	0
TP18	0	0	1	0
TP19	0	0	0	0
TP20	0	0	1	1
TP21	0	0	1	0
TP22	0	0	1	1
TP23	0	0	1	0
TP24	0	0	0	1
TP25	0	0	0	1
TP26	0	0	1	1
TP27	0	0	1	0
TP28	1	0	1	0
TP29	0	0	0	0
TP30	2	2	2	2
TP31	2	2	2	2
TP32	0	0	0	1
TP33	0	0	1	1
TP34	0	1	0	0
TP35	0	0	0	0
TP36	0	0	0	0
TP37	0	0	1	0
TP38	1	0	1	1
TP39	0	0	1	0
TP40	0	0	0	0

Supp. Table 5 cont				
Sample	**KMT2D**	**PTEN**	**NOTCH1**	**FBXW7**
TP41	0	1	1	0
TP42	0	0	0	0
TP43	0	1	1	0
TP44	0	0	0	0
TP45	0	0	0	1
TP46	0	0	0	1
TP47	0	1	1	0
TP48	0	0	1	0
TP49	0	0	1	1
TP50	0	1	1	0
TP51	0	0	1	1
TP52	0	0	0	1
TP53	0	1	0	0
TP54	0	0	0	1
TP55	2	2	2	2
TP56	0	0	0	0
TP57	0	0	0	0
TP58	0	0	1	0
TP59	0	0	0	0
TP60	0	1	0	0
TP61	0	0	0	0
TP62	0	0	1	0
TP63	0	1	1	0
TP64	0	0	1	1
TP65	0	0	1	1
TP66	0	0	1	0
TP67	0	0	0	0
TP68	0	0	0	0
TP69	0	0	1	1
TP70	1	0	1	0
TP71	0	0	1	1
TP72	0	0	1	0
TP73	0	0	1	0
TP74	0	0	1	0
TP75	0	0	0	1
TP76	1	0	0	0
TP77	0	0	0	0
TP78	0	0	1	0
TP79	0	0	0	0
TP80	0	0	1	1

250

Supp. Table 5 cont				
Sample	**KMT2D**	**PTEN**	**NOTCH1**	**FBXW7**
TP81	1	0	1	0
TP82	0	1	1	0
TP83	1	0	0	0
TP84	0	0	0	0
TP85	0	0	1	0
TP86	0	1	1	0
TP87	0	1	0	0
TP88	0	0	1	0
TP89	0	1	1	0
TP90	0	0	0	0
TP91	0	0	1	0
TP92	0	0	1	0
TP93	1	0	0	0
TP94	0	1	0	0
TP95	0	0	1	0
TP96	0	0	1	0
TP97	0	1	0	0
TP98	0	1	0	0
TP99	0	0	1	0
TP100	2	2	2	2
TP101	1	0	1	0
TP102	0	0	0	0
TP103	0	0	0	1
TP104	0	0	1	0
TP105	0	0	1	1
TP106	0	0	0	0
TP107	0	0	1	0
TP108	0	0	0	0
TP109	0	0	0	0
TP110	0	0	0	0
TP111	0	0	0	1
TP112	0	0	1	1
TP113	1	1	0	0
TP114	0	0	0	0
TP115	0	0	1	0
TP116	1	0	0	0
TP117	0	0	1	0
TP118	0	0	1	0
TP119	0	0	0	0
TP120	0	0	1	0
TP121	1	0	1	1
TP122	0	0	1	0

251

Supp. Table 5 cont				
Sample	**KMT2D**	**PTEN**	**NOTCH1**	**FBXW7**
TR11	0	0	1	0
TR12	0	0	0	0
TR13	0	0	0	0
TR14	1	0	0	0
TR15	0	0	1	0
TR16	0	1	1	0
TR116	0	0	0	0
TR123	0	1	0	0
TR124	0	0	0	0

Supplemental Table 5. Supplemental Table 5. Mutational status of *KMT2D, PTEN, NOTCH1* (only exons 26, 27 and 34) and *FBXW7* (only exons 9, 10 and 12) in the extended cohort. 0-not mutated, 1- mutated, 2-failed sequencing,

Purple represents $KMT2D^{mut}$ (n=4) in $NOTCH1^{wt}$ and/or $FBXW7^{wt}$ (N/F^{wt}); pink represents $PTEN^{mut}$ in N/F^{wt} (n=7); green represents $KMT2D^{mut}$ + $PTEN^{mut}$ in N/F^{wt} (n=1); yellow represents samples failed in targeted sequencing; orange represents relapse cases are not included in the assessment of prognostic relevance

Supplemental Table 6. Primers used in the mutational analysis of *BCL11B* and ABD analysis

No.	Name	Primer sequence	Reference
1	BCL11B_ex2_F	TGACCGCCTAAGCCCATCTCTAT	Provided by the Department of Immunology, Erasmus
2	BCL11B_ex2_R	GCCCCGGCTGGTGGCCCAGAGG	Provided by the Department of Immunology, Erasmus
3	BCL11B_ex4B_F	CAGTGCCTTCGACCGAGT	
4	BCL11B_ex4B_R	CGGCTCTCGTTCTCCAGTAG	
5	BCL11B_ex4C_F	GCATCAAGGTGGAGAAGGAC	K. De Keersmaecker Nat Med.2010
6	BCL11B_ex4C_R	CTCGGGTTTCCATAGGACTTC	K. De Keersmaecker Nat. Med.2010
7	TRGV1f	GGTTGTGTTGGAATCAGGAGTCA	Euroclonality Biomed-2 protocols
8	TRGV9	CGGCACTGTCAGAAAGGAATC	Euroclonality Biomed-2 protocols
9	TRGV10	AGCATGGGTAAGACAAGCAA	Euroclonality Biomed-2 protocols
10	TRGV11	TTGCTCAGGTGGGAAGACTA	Euroclonality Biomed-2 protocols
11	TRGJ1/J2_FAM	GTGTTGTTCCACTGCCAAAGAG	Euroclonality Biomed-2 protocols
12	TRGJP1/JP2_FAM	AGTTACTATGAGCYTAGTCCCTT	Euroclonality Biomed-2 protocols

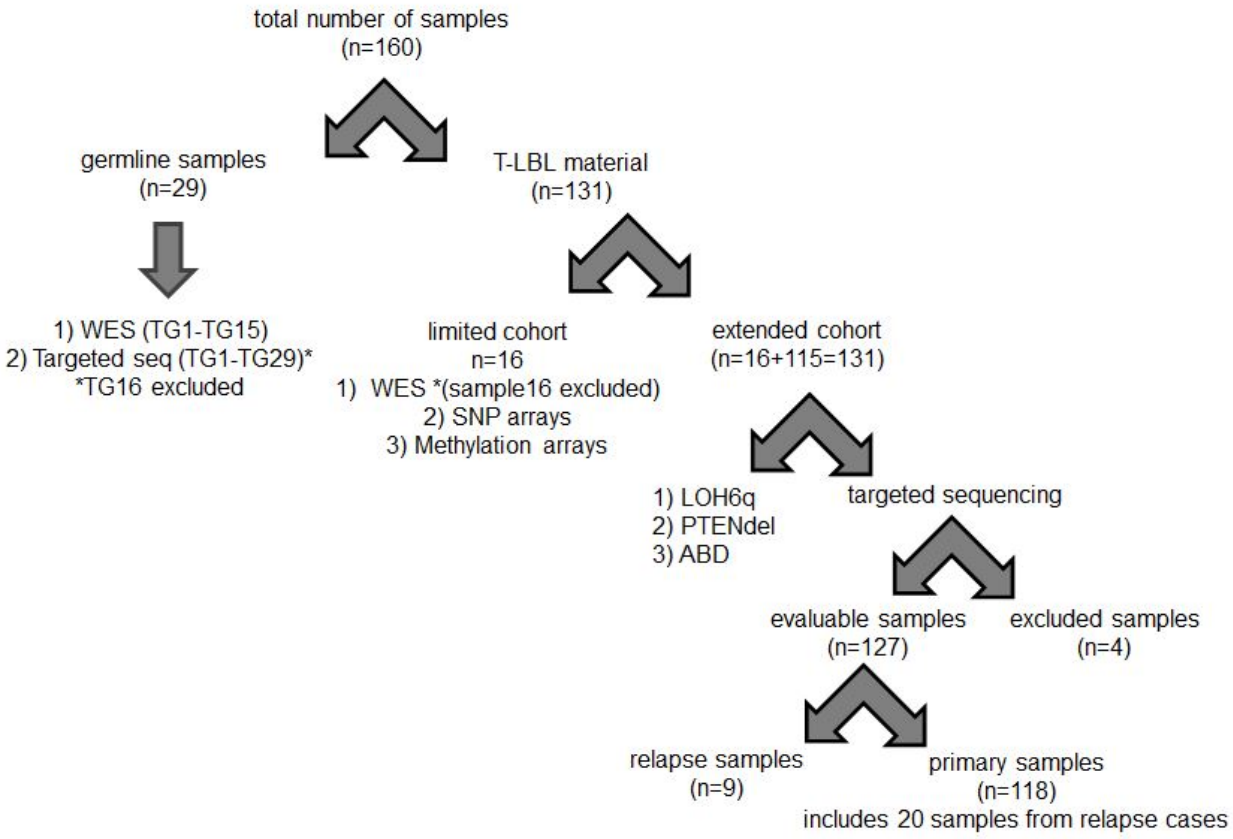
Supplemental Table 6. Primers 1-6 were used for *BCL11B* mutational analysis and primers 7-12 were used for ABD analysis

Supplemental Table 7. Detailed description of identified mutations in BCL11B

Sample ID	exon	DNA (NM_138576.4)	Codon change	Mutation type
TP35	2	c.122_c.123insCCGTTTCGTCTTATAGATGGCAAGACGAGGGTCTGGAGATAGAGGAGCC het	P41Pfs5*	frameshift
TP11	4	c.2561A>G het	Y854C	missense
TP18	4	c.1156_1165delinsTAAACTTT het	N386*	nonsense
TP20	4	c.1284_1285insAAAATTTACGGG	S428_C429insKIYG	indel
TP32	4	c.1089_1088insCCC	F363_S364insP	indel
TP34	4	c.1391G>A het	C464Y	missense
TP35	4	c.1483G>A het	A495T	missense
TP41	4	c.1378T>C het	C460R	missense
TP45	4	c.2655_2676delinsCCGTTGG het	V886Pfs106*	frameshift
TP51	4	c.1283_1284insACCTAGCC het	S428Sfs48*	frameshift
TP120	4	c.1349C>T het	T450M	missense
TR116	4	c.1483G>A het	A495T	missense

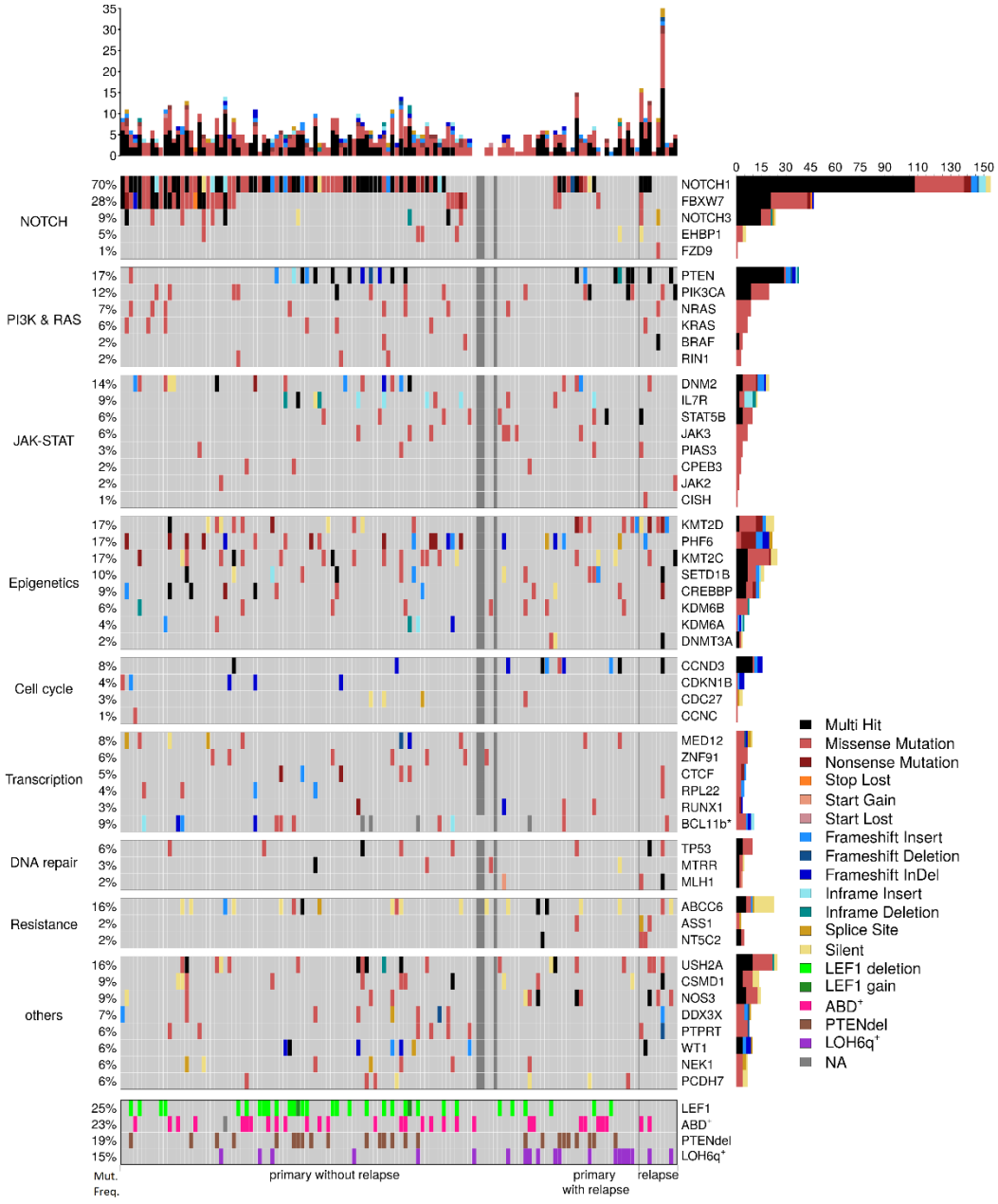
Supplemental Figures

Supplemental Figure 1. Overview of samples used in the current project



Supplemental Figure 1. Schematic overview of samples and mutational spectrum of T-LBL. Overview of samples used in the current project. In total 131 samples including 118 primary samples and 9 relapse samples (including 7 matched relapse samples (TR11-TR17) and 2 single relapse samples (TR18 and TR19)).

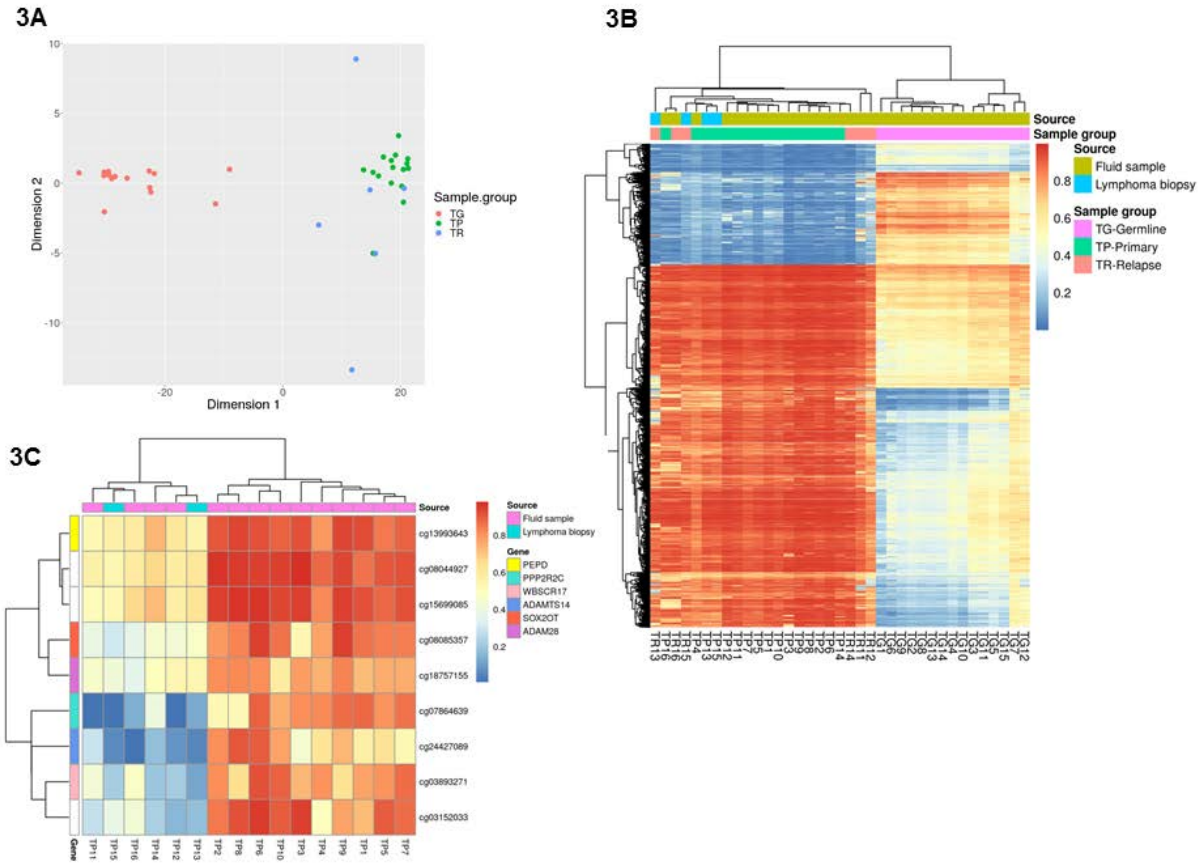
Supplemental Figure 2. Mutational spectrum of T-LBL (VAF>1%)



Supplemental Figure 2. Mutational spectrum of T-LBL: Overview of recurrently mutated genes (VAF cutoff of>1%) and alterations in T-LBL. The samples are sorted in three sections; Primary samples from cases that did not suffer relapse, primary samples from cases that suffered relapse and third section includes relapse cases as indicated at the bottom of the panel. The frequency of mutations (Mut. Freq.) and name of pathways

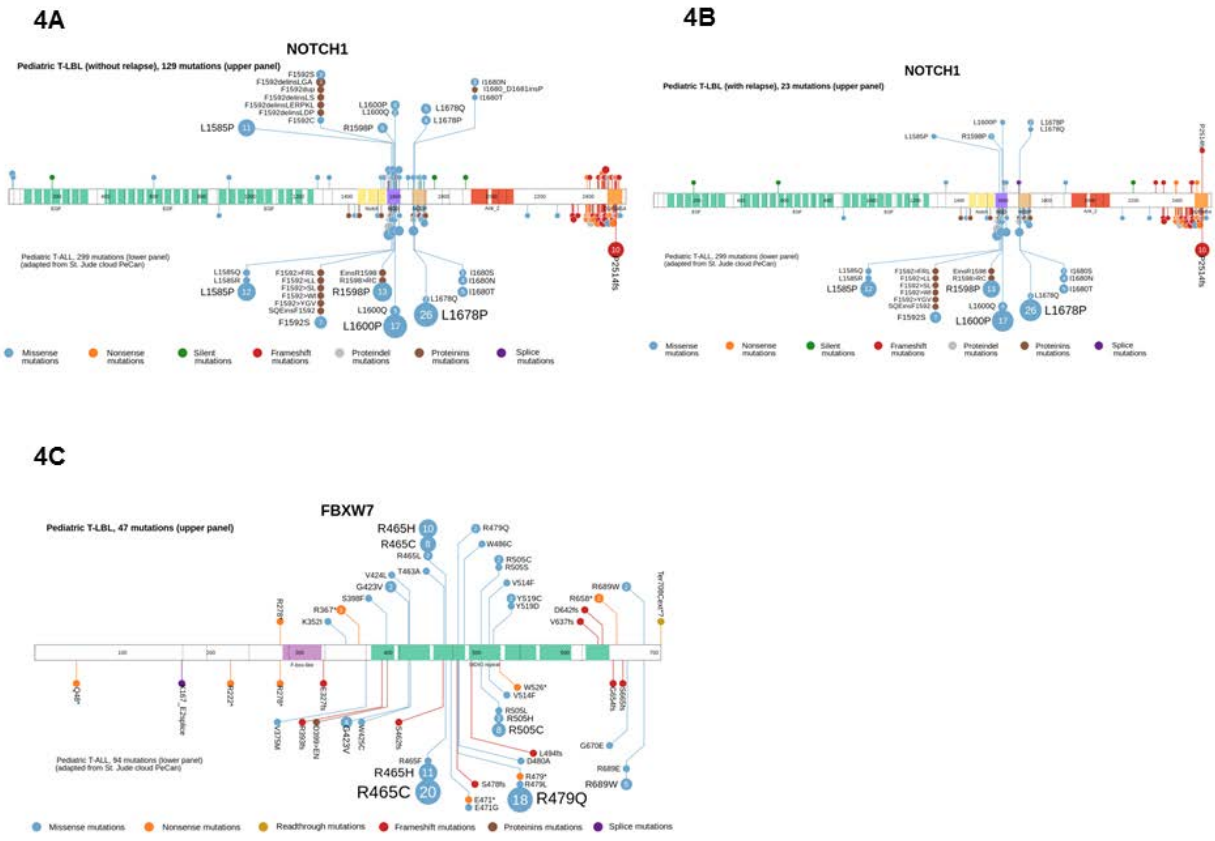
is indicated on the left side of the panel. Name of the genes and the type of mutations are indicated on the right side. The number of mutations identified by targeted sequencing is displayed at the top of the plot as bar plots. TRG rearrangements (ABD), *PTEN* deletions (*PTEN*del) and LOH6q alterations are displayed in a sub-panel below. Colored boxes on the right side indicate the type of mutation. (*) indicates that *BCL11b* was not part of targeted panel and analyzed by Sanger sequencing. Samples that failed in targeted sequencing are represented in dark grey boxes.

Supplemental Figure 3. Overview of differential methylation between different samples from the "limited cohort"



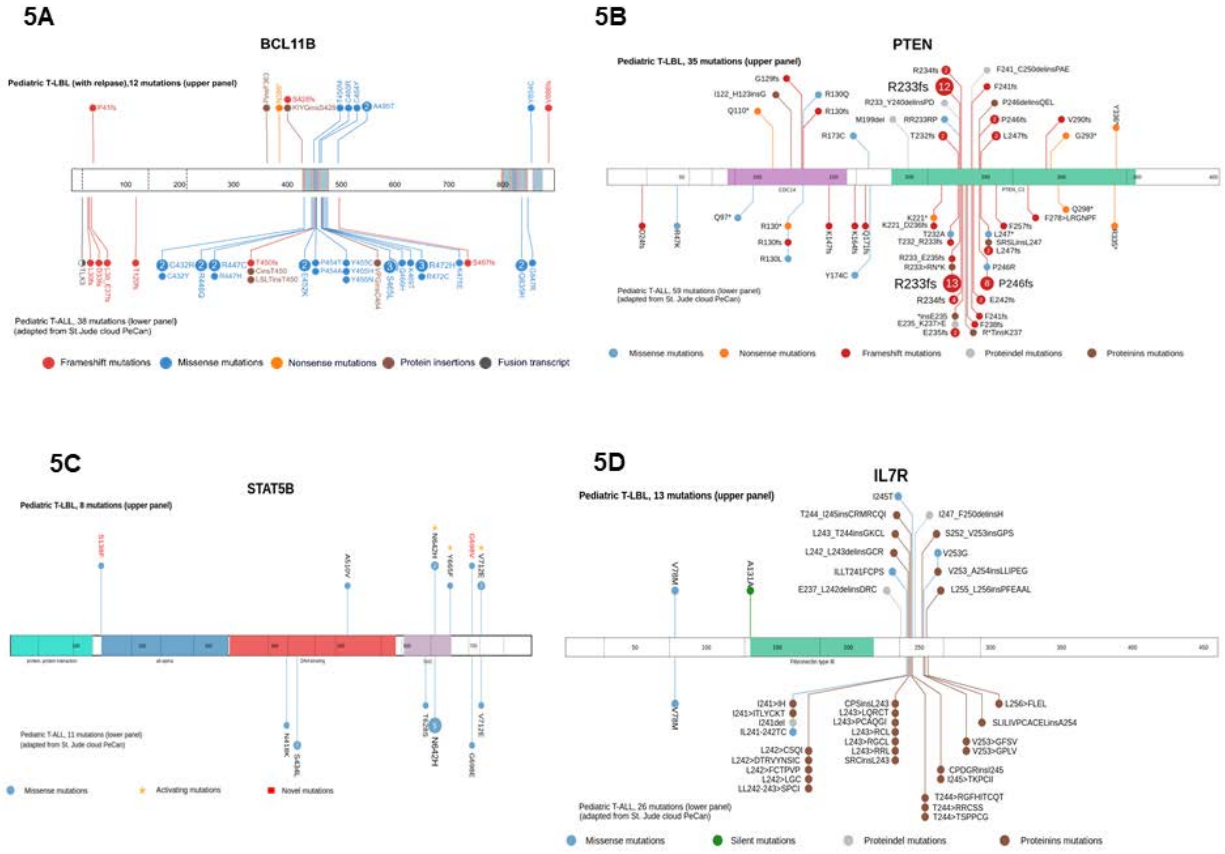
Supplemental Figure 3. (A) Non-metric multidimensional scaling (NMDS) analysis representing the variation in methylation epigenotypes between samples **(B)** Supervised hierarchical clustering conducted using 8045 probes that were significantly differentially methylated between germline (TG) and tumor samples (TP- primary and TR-relapse). *p* value=$1e^{-12}$. **(C)** Differential methylated genes between primary samples from relapse(-) and relapse(+) cases. *p* value=0.05**.** Array used: Infinium MethylationEPIC BeadChip '850K'

Supplemental Figure 4. Schematic display of localization and frequencies of somatic mutations identified in *NOTCH1* and *FBXW7*



Supplemental Figure 4. Schematic display of localization and frequencies of somatic mutations identified for, **(A)** *NOTCH1* in relapse^{-} **(B)** *NOTCH1* in relapse^{+} samples **(C)** *FBXW7* in total samples Note: The data for mutations **(4 A-C)** from pediatric T-ALL was imported from St. Jude Pediatric Cancer Genomic data portal and are displayed at lower part of the protein/Gene and mutations identified in current T-LBL project are displayed on the upper part of the protein structure.

Supplemental Figure 5. Schematic display of localization and frequencies of somatic mutations identified in candidate genes



Supplemental Figure 5. Schematic display of localization and frequencies of somatic mutations identified for **(A)** *BCL11B* **(B)** *PTEN* **(C)** *STAT5B* **(D)** *IL7R* in total samples Note: For figures **(B-D)** the mutations displayed are with VAF cutoff >1%. The data for mutations **(5A-D)** from pediatric T-ALL was imported from St. Jude Pediatric Cancer Genomic data portal and are displayed at lower part of the protein/Gene and mutations identified in current T-LBL project are displayed on the upper part of the protein structure.

Supplemental Figure 6 Gene relation plot

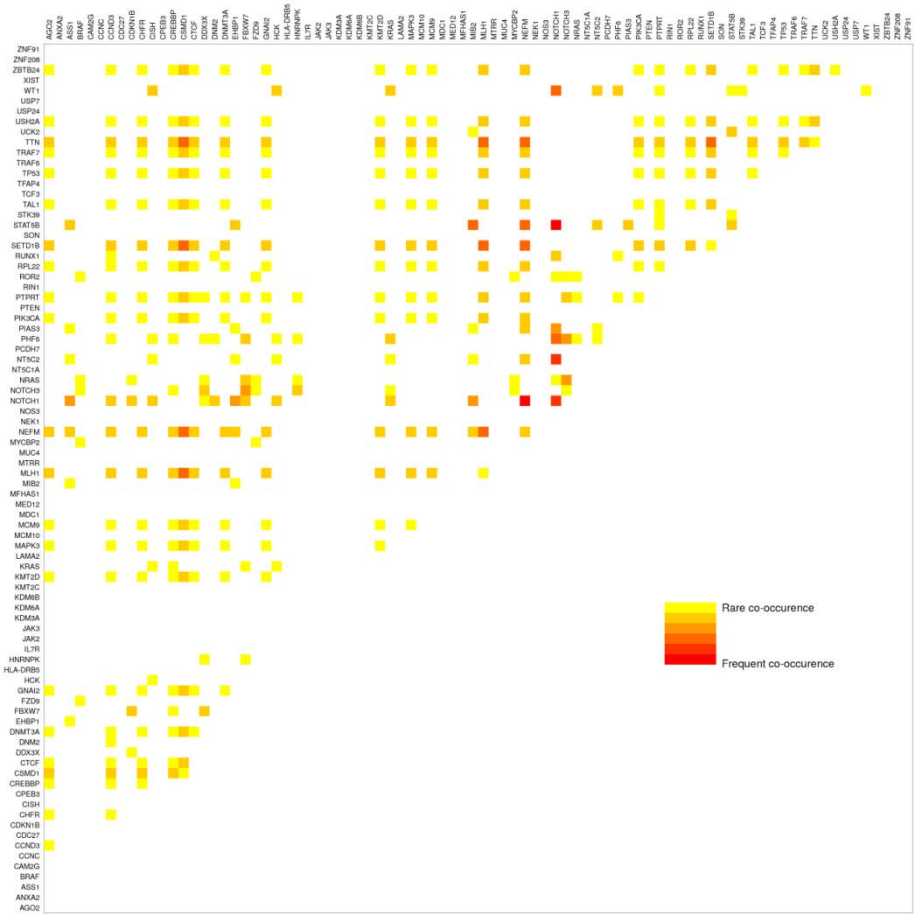
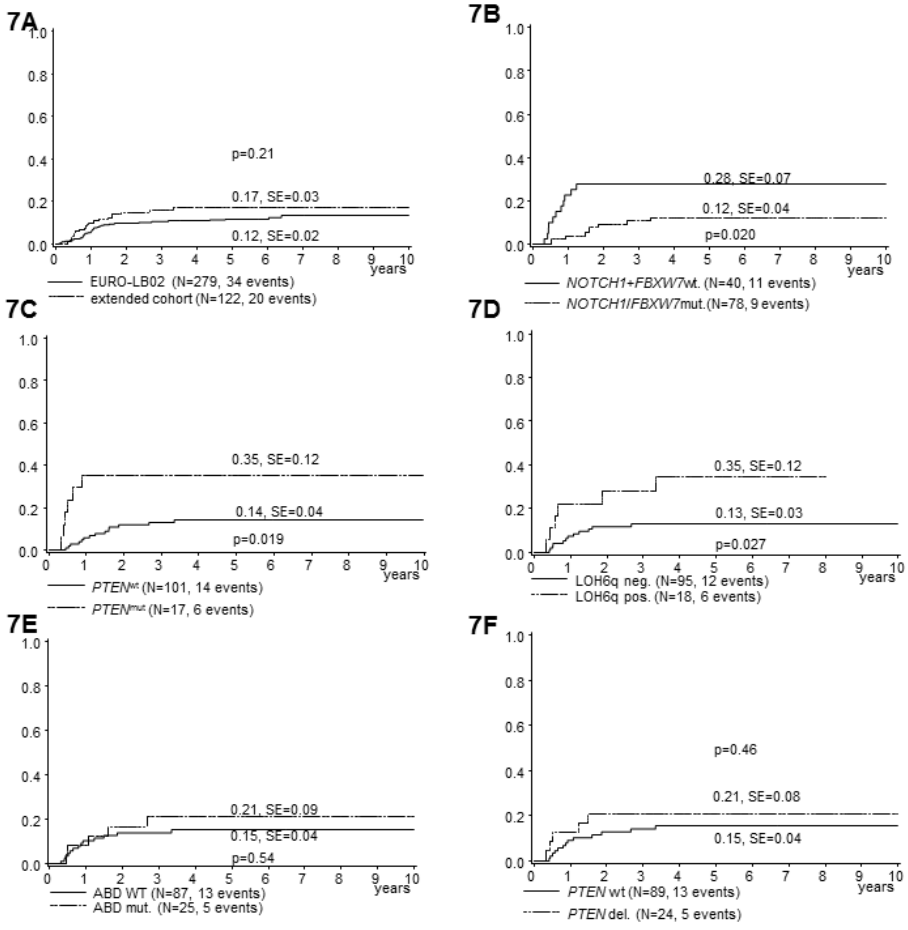


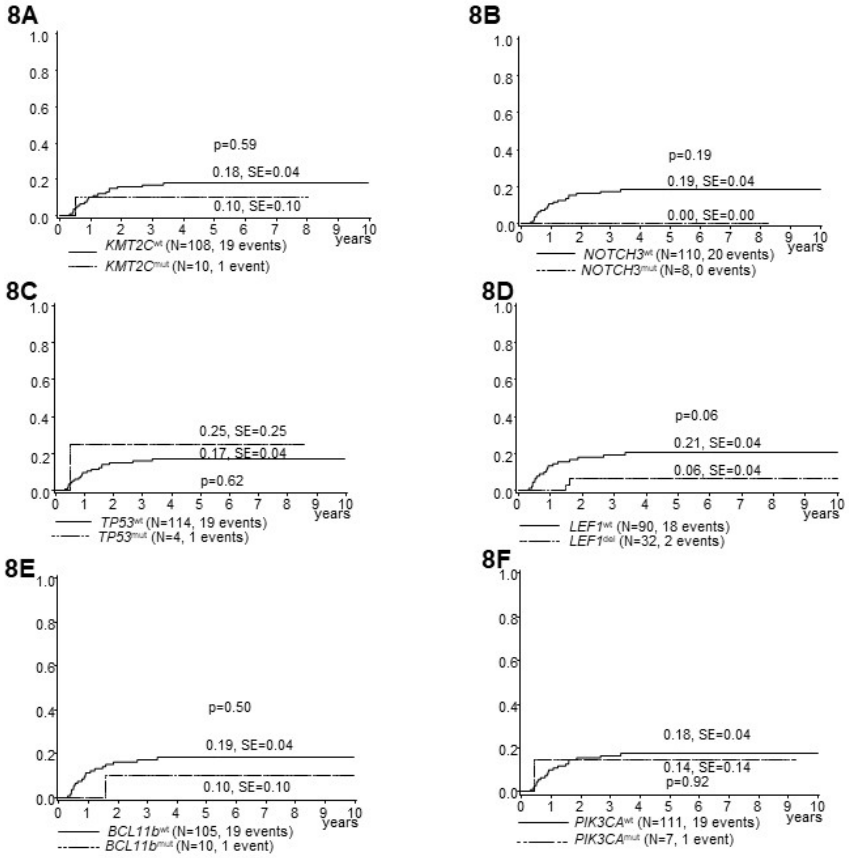
Figure 6. Gene relation plot depicting correlation of mutated genes from the targeted sequencing.

Supplemental Figure 7. Five year cumulative incidence of relapse (CIR) for Euro-LB02 vs extended cohort and previously identified candidate genes



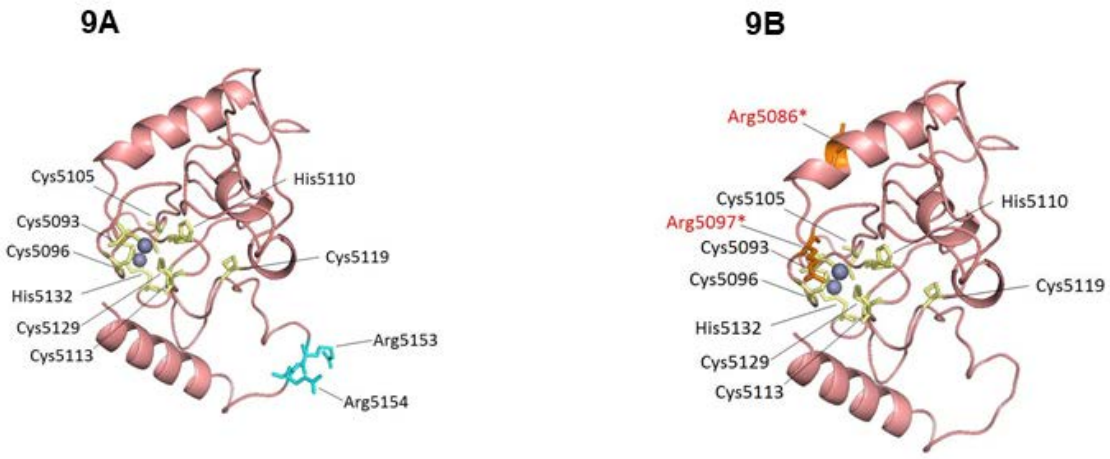
Supplemental Figure 7. Five year cumulative incidence of relapse (CIR) for **(A)** Euro-LB02 vs extended cohort, **(B)** *NOTCH1* and/or *FBXW7* mutational status, **(C)** *PTEN* mutational status, **(D**) LOH6q status, **(E)** absence of biallelic T-cell receptor (TCR) deletion (ABD) status **(F**) *PTEN*del status

Supplemental Figure 8. Five year cumulative incidence of relapse (CIR) for newly identified potential candidate genes



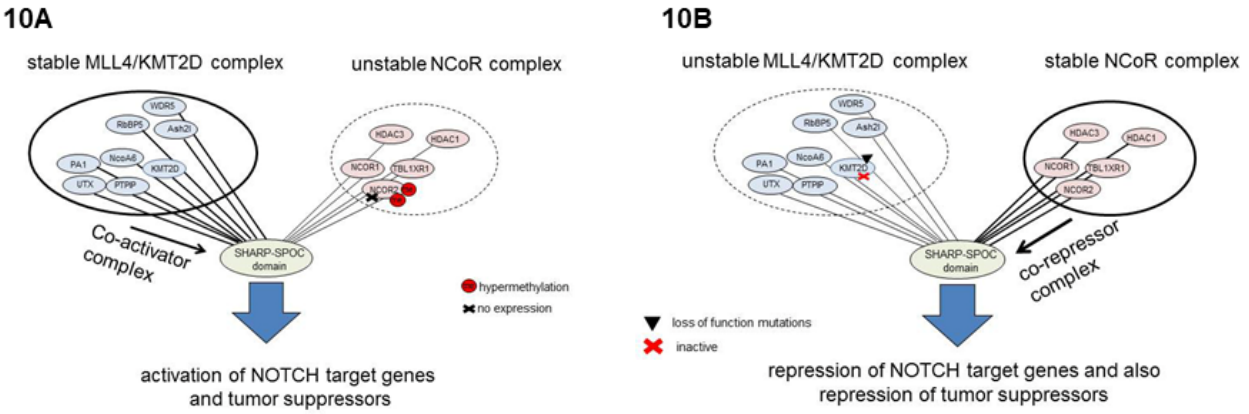
Supplemental Figure 8. Five year cummulative incidence of relapse (CIR) for **(A)** *KMT2C* mutational status, **(B)** *NOTCH3* mutational status, **(C)** *TP53* mutational status, **(D**) *LEF1* mutational status, **(E)** *BCL11b* mutational status **(F)** *PIK3CA* mutational status

Supplemental Figure 9. Modelling of *KMT2D* domain structures and comparison with the mutated domain structures.



Supplemental Figure 9. (A) Domain structure of PHD_7 of *KMT2D*^{wt} was generated using Phyre2, Cys_4–His-Cys_3- His motif is displayed in stick structures (yellow) and two zinc molecules (purple) were modeled into the structure using PyMOL. The two nonsense mutations at the respective positions (red) are indicated by (*). (**B)** Schematic display of Arginine residues (cyan) at respective positions near PHD_7 of *KMT2D*^{wt}. Mutations post PHD_7 at positions R5153P (cyan) and R5154Q (cyan), the latter identified in T-ALL, are represented in stick structures.

Supplemental Figure 10



Supplemental Figure 10. Schematic representation of mechanism of regulation of Notch target genes and altered pathways in pediatric T-LBL

(**A**) Schematic representation of mechanism of regulation of Notch target genes and tumor suppressors via a dynamic interplay between the *KMT2D* co-activator complex and NCoR repressor complex. In absence of loss of function mutations for *KMT2D* or post-translational modification which can inactivate *KMT2D* and in presence of NCoR2 hypermethylation which might result in unstable repressor complex, the *KMT2D* co-activator complex binds the SHARP protein to activate Notch target genes and tumor suppressors. (**B**) In case of loss of function mutations in *KMT2D* or post-translational modification such as phosphorylation which can inactivate *KMT2D* presumably results in binding of stable repressor complex to the SHARP protein to repress notch target genes and tumor suppressors. (Figures 4A, 4B adapted from Oswald F, et. al. Nucleic Acids Research 2016)**.**

References

1. Reiter A, Schrappe M, Ludwig WD, et al. Intensive ALL-type therapy without local radiotherapy provides a 90% event-free survival for children with T-cell lymphoblastic lymphoma: a BFM group report. *Blood*. 2000;95(2):416-421.

2. Landmann E, Burkhardt B, Zimmermann M, et al. Results and conclusions of the European Intergroup EURO-LB02 trial in children and adolescents with lymphoblastic lymphoma. *Haematologica*. 2017;102(12):2086-2096.

3. Bonn BR, Rohde M, Zimmermann M, et al. Incidence and prognostic relevance of genetic variations in T-cell lymphoblastic lymphoma in childhood and adolescence. *Blood*. 2013;121(16):3153-3160.

4. Liu Y, Easton J, Shao Y, et al. The genomic landscape of pediatric and young adult T-lineage acute lymphoblastic leukemia. *Nat Genet*. 2017;49(8):1211-1218.

5. Ma X, Liu Y, Liu Y, et al. Pan-cancer genome and transcriptome analyses of 1,699 paediatric leukaemias and solid tumours. *Nature*. 2018;555(7696):371-376.

6. Kraszewska MD, Dawidowska M, Kosmalska M, et al. BCL11B, FLT3, NOTCH1 and FBXW7 mutation status in T-cell acute lymphoblastic leukemia patients (vol 50, pg 33, 2013). *Blood Cells Molecules and Diseases*. 2013;51(1):66-68.

7. De Keersmaecker K, Real PJ, Della Gatta G, et al. The TLX1 oncogene drives aneuploidy in T cell transformation. *Nature Medicine*. 2010;16(11):1321-+.

8. Larson DE, Harris CC, Chen K, et al. SomaticSniper: identification of somatic point mutations in whole genome sequencing data. *Bioinformatics*. 2012;28(3):311-317.

9. Saunders CT, Wong WS, Swamy S, Becq J, Murray LJ, Cheetham RK. Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. *Bioinformatics*. 2012;28(14):1811-1817.

10. Roberts ND, Kortschak RD, Parker WT, et al. A comparative analysis of algorithms for somatic SNV detection in cancer. *Bioinformatics*. 2013;29(18):2223-2230.

11. Xu H, DiCarlo J, Satya RV, Peng Q, Wang Y. Comparison of somatic mutation calling methods in amplicon and whole exome sequence data. *BMC Genomics*. 2014;15:244.

12. Cingolani P, Platts A, Wang le L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly (Austin)*. 2012;6(2):80-92.

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

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. Landrum MJ, Lee JM, Benson M, et al. ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res*. 2016;44(D1):D862-868.

16. Forbes SA, Beare D, Gunasekaran P, et al. COSMIC: exploring the world's knowledge of somatic mutations in human cancer. *Nucleic Acids Res*. 2015;43(Database issue):D805-811.

17. Sandmann S, Karimi M, de Graaf AO, et al. appreci8: a pipeline for precise variant calling integrating 8 tools. *Bioinformatics*. 2018;34(24):4205-4212.

18. Zhou W, Laird PW, Shen H. Comprehensive characterization, annotation and innovative use of Infinium DNA methylation BeadChip probes. *Nucleic Acids Res*. 2017;45(4):e22.

19. 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.

20. Tian Y, Morris TJ, Webster AP, et al. ChAMP: updated methylation analysis pipeline for Illumina BeadChips. *Bioinformatics*. 2017;33(24):3982-3984.

21. Teschendorff AE, Widschwendter M. Differential variability improves the identification of cancer risk markers in DNA methylation studies profiling precursor cancer lesions. *Bioinformatics*. 2012;28(11):1487-1494.

22. Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015;43(7):e47.

23. Rohde M, Bonn BR, Zimmermann M, et al. Multiplex ligation-dependent probe amplification validates LOH6q analyses and enhances insight into chromosome 6q aberrations in pediatric T-cell lymphoblastic leukemia and lymphoma. *Leuk Lymphoma*. 2015;56(6):1884-1887.

24. Derrieux C, Trinquand A, Bruneau J, et al. A Single-Tube, EuroClonality-Inspired, TRG Clonality Multiplex PCR Aids Management of Patients with Enteropathic Diseases, including from Formaldehyde-Fixed, Paraffin-Embedded Tissues. *J Mol Diagn*. 2019;21(1):111-122.