Supplemental Data

Supplemental Methods

Lymphoid panel library preparation and sequencing

An amount of 41-250 ng DNA of each sample was used for library preparation with KAPA HyperPlus (Roche) with the following modifications: fragmentation with 12.5 min incubation, xGen Duplex Seq adapters (3-4 nt unique molecular identifiers (UMI), 0.55 uM, Integrated DNA Technologies) were used for the ligation and xGen Indexing primers (2 mM, with unique dual indices, Integrated DNA Technologies) was used for PCR amplification (5-13 cycles depending on input amount of DNA). Target enrichment was performed in a multiplex fashion with a library amount of 187.5 ng (8-plex). The libraries were hybridized to the capture probes in the GMS Lymphoid panel (Twist Bioscience) with addition of Twist Universal Blockers and Blocking solution for 16 hours. The post-capture PCR was performed with xGen Library Amp Primer (0.5 mM, Integrated DNA Technologies) for 10 cycles. Quality control was performed with Quant-iT dsDNA HS assay (ThermoFisher) and TapeStation HS D1000 assay (Agilent). Sequencing was carried out on the NovaSeq 6000 (Illumina) using paired-end 150 nt readout, aiming at 60 M read pairs per sample. Demultiplexing was performed using Illumina bcl2fastq Conversion Software v2.20.

Bioinformatics analysis

HaloPlex panel: Agilent SureCallTrimmer (v4.0.1) was used to remove adaptor sequences, mask HaloPlex enzyme footprints, and trim low-quality bases. The sequencing reads were aligned to the GRCh37 human reference genome using bwa-mem (v0.7.16) and post-processed using Samtools (v1.8). The Agilent LocatIt tool (v4.0.1) was applied for processing of molecular barcode information. in order to reduce quantitative bias and improve base-calling accuracy. Pisces (v5.2.10.49) was used for detection of single nucleotide variants (SNVs) and small insertions/deletions (indels). with detection thresholds of 40x coverage, minimum 10 reads supporting the variant, and a variant allele frequency (VAF) of 0.5% to also allow for estimation of the background noise-rate. Variants were annotated with population variation databases and Cosmic (v85) using VEP[23] (v91) and SnpEFF (v4.3). Variants included in downstream analysis met the following conditions: (i) have a VAF ≥1%; (ii) be located within an exonic or splicing region; (iii) be non-synonymous; (iv) not have a calculated impact in the LOW-category in VEP and (v) not be listed in the gnomAD database unless recurrently reported in Cosmic v85. Recurrent technical artefacts were filtered by removing variants not previously observed in the Cosmic database and detected in >15 samples. Pysamstats (v1.1.2) was used for detailed investigation of mutations at codon 481 of the *BTK* gene and at selected

PLCG2 hotspots ("per base" analysis*)*. In this analysis, a mutation was called if there was ≥8 (*BTK*) or ≥10 (*PLCG2*) supporting reads for a recurrent base substitution.

Lymphoid panel: BALSAMIC was applied to analyze each of the FASTQ files. In summary, FASTQ files were quality controlled using FastQC v0.11.5 [2]. Adapter sequences and low-quality bases were trimmed using fastp v0.20.0 [3]. Trimmed reads were mapped to the reference genome hg19 using BWA MEM v0.7.15 [4]. The resulting SAM files were converted to BAM files and sorted using samtools v1.6 [5,6]. Duplicated reads were marked using Picard tools MarkDuplicate v2.17.0 [7] and quality controlled using CollectHsMetrics, CollectInsertSizeMetrics, and CollectAligntmentSummaryMetrics functionalities. Results of the quality controlled steps were summarized by MultiQC v1.7 [8]. For each sample, small somatic mutations were called using VarDict v2019.06.04 [9], and copy-number aberrations were called using CNVkit v0.9.4a0 [11]. CNVkit's copy-number ratio (cnr) files were used separately to heatmap all CNVs using "--by-bin –desaturate" options. For filtering CNV calls, all cnr files were first segmented using "segment –method clonal" to produce copy-number segment (cns) files. Thereafter, statistics for each segment was calculated using "segmetrics –ci –pi –sem" which in turn calculated confidence interval, prediction interval, and standard error mean. This generated a new cns file for each case containing copy-number estimates. The new cns files were then subjected to "call –filter ci" to produce a final filtered list of copy-number (cn) regions. Deletions of 8p, 11q, 13q, and 17p were identified by selecting all values with log2 fold change of -0.15 and a region size larger than 5Mb for del(8p), and for trisomy 12, all values with cn=3 and covering both chromosome arms were selected. All SNVs and indels variants were annotated using Ensembl VEP v100 [12] and vcfanno v0.3.2 [13]. For variant filtering the following criteria was applied: read depth (DP) >50, alternative allele depth (AD) >10, allele frequency (AF) >0.1, and GNOMAD AF popmax [14] <0.001. For exact parameters used for each of the softwares, please refer to https://github.com/Clinical-Genomics/BALSAMIC

Droplet digital PCR (ddPCR)

The amount of gDNA sample used per reaction was between 12 and 90 ng, depending on material availability. Both assays (*BTK* hotspot C481S: c.1442G>C and c.1441T>A) were run in duplicates. Three positive droplets had to be present for a sample to be considered positive.

References

1. Foroughi-Asl H, Jeggari A, Maqbool K, et al. BALSAMIC: Bioinformatic Analysis Pipeline for Somatic MutatIons in Cancer (v8.2.3). Zenodo. https://doi.org/10.5281/zenodo.5734364

2. Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data. Accessed June 22, 2020. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

3. Chen S, Zhou Y, Chen Y, et al. Fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 2018;34:i884-i890.

4. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 2009;25:1754-1760.

5. Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics 2009;25:2078-2079.

6. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics 2011;27:2987-2993.

7. Picard Tools - By Broad Institute. Accessed June 22, 2020. https://broadinstitute.github.io/picard/

8. Ewels P, Magnusson M, Lundin S, et al. MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics 2016;32:3047-3048.

9. Lai Z, Markovets A, Ahdesmaki M, et al. VarDict: A novel and versatile variant caller for next-generation sequencing in cancer research. Nucleic Acids Research. 2016;44:1–11.

10. Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications | Bioinformatics | Oxford Academic. Accessed June 22, 2020. https://academic.oup.com/bioinformatics/article/32/8/1220/1743909

11. Talevich E, Shain AH, Botton T, et al. CNVkit: Genome-Wide Copy Number Detection and Visualization from Targeted DNA Sequencing. PLOS Computational Biology 2016;12:e1004873.

12. McLaren W, Gil L, Hunt SE, et al. The Ensembl Variant Effect Predictor. Genome Biology 2016;17:122.

13. Pedersen BS, Layer RM, Quinlan AR. Vcfanno: fast, flexible annotation of genetic variants. Genome Biology 2016;17:118.

14. Karczewski KJ, Francioli LC, Tiao G, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature 2020;581:434-443.

Supplemental Results

The average coverage over the entire HaloPlex panel target region after molecular barcodes processing was 1,662x (range 376-5,524x) across 151 samples collected from 98 patients. The recurrently mutated *BTK* positions chrX:100611164 and chrX:100611165 (hg19) reached a mean sequencing-depth of 3,140x (range 619-15,017x). The average sequencing coverage for all exons of *BTK* and *PLCG2* were 1,986x and 2,596x, respectively.

Supplemental Table 1. List of the 13 genes included in the custom targeted resequencing HaloPlex HS panel

Gene	RefSeq	N° of coding exons
ATM	NM_000051	62
BIRC3	NM_001165	8
MYD88	NM_002468	5
NOTCH1	NM_017617	34*
SF3B1	NM_012433	25*
TP53	NM_000546	10
EGR2	NM_000399	2
POT1	NM_015450	15
NFKBIE	NM_004556	6
XPO1	NM_003400	24*
FBXW7	NM_033632	11
BTK	NM_000061	19
PLCG2	NM_002661	32

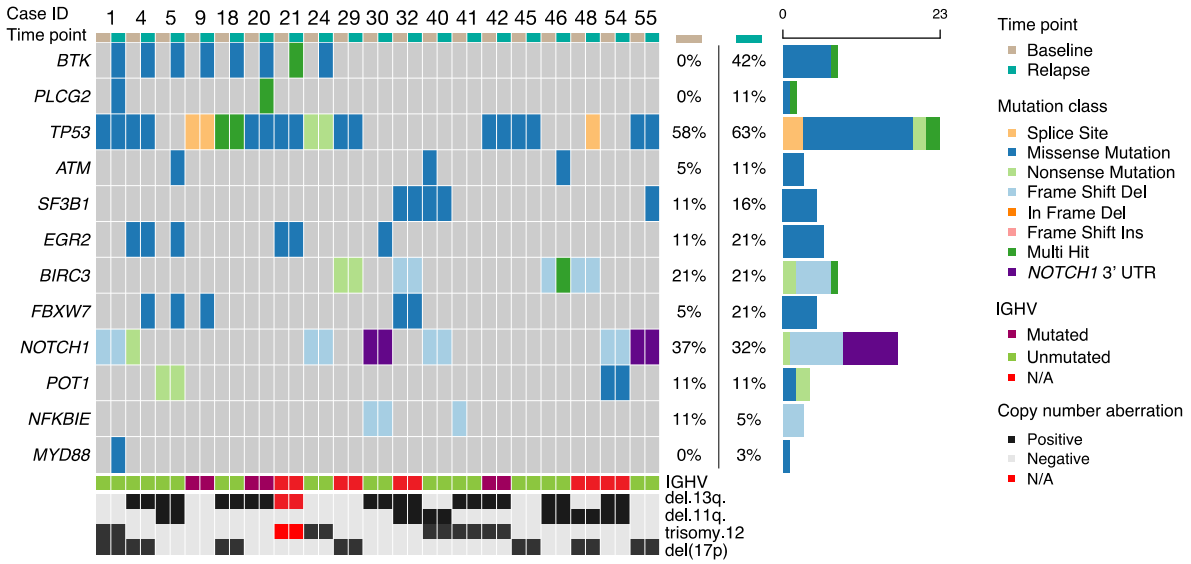
*Only the mutational hotspot regions of the following genes were sequenced: *NOTCH1* (exon 34+3'UTR), *SF3B1* (exons 14-16 and 18) and *XPO1* (exon 15). All coding exons in the remaining genes were sequenced.

Supplemental Table 2. Somatic variants detected by HaloPlex panel in the relapsed and responsive cohorts (excel file)

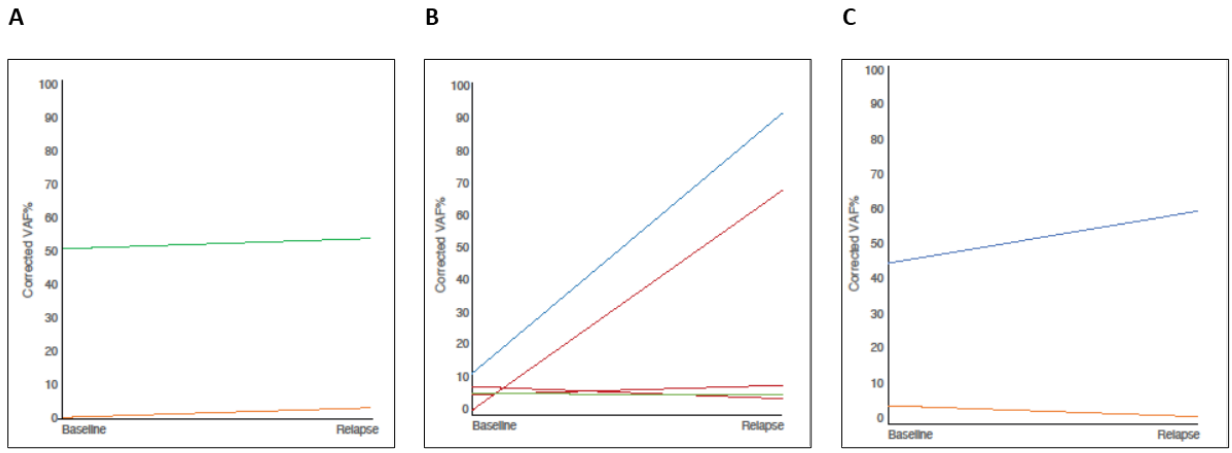
Supplemental Table 3. Somatic variants detected by the capture lymphoid panel in the relapsed and responsive cohorts (excel file)

Supplemental Table 4. Del(8p) detected by the capture lymphoid panel in the relapsed and responsive cohorts (excel file)

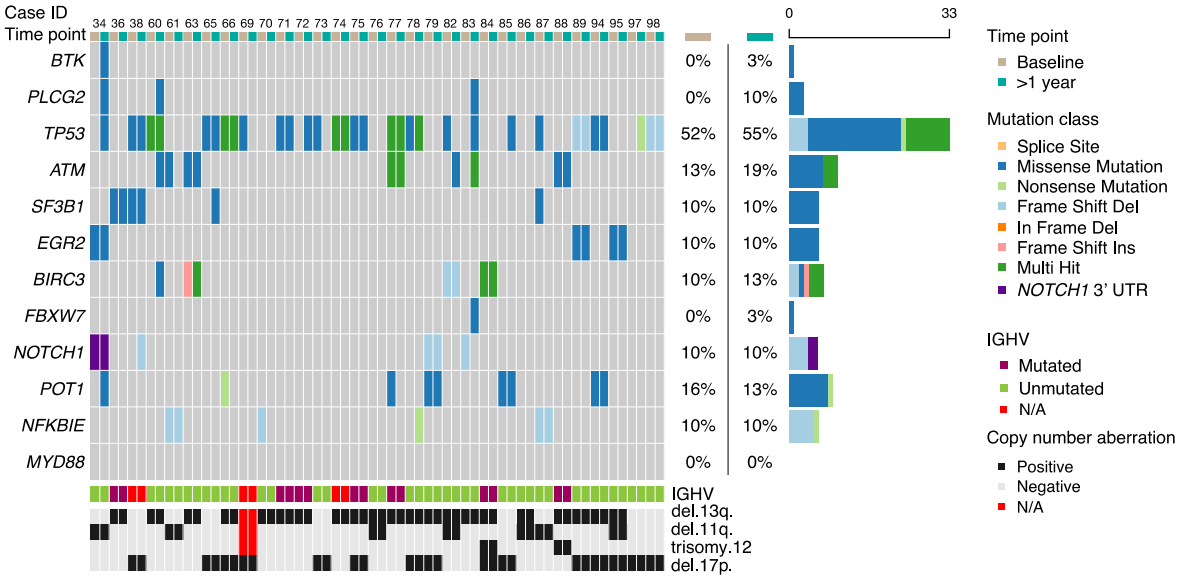
Supplemental Figure 1. Oncoplot displaying the somatic variants detected in the 13 genes analyzed by a HaloPlex HS custom panel in 19 patients with paired samples (at baseline and at relapse) of the relapsed cohort. Illustrated are the distribution of the somatic variants, IGHV mutational status, cytogenetic profile determined by fluorescence in situ hybridization, and the mutation frequency of each gene, at baseline and at relapse, respectively. Patient identifiers are indicated across the top row. For each patient, both the baseline (grey) and the relapse (light blue) time points are represented.



Supplemental Figure 2. *EGR2, BIRC3* and *NFKBIE* clonal dynamics in the relapsed cohort. (A) *EGR2* clonal dynamics in 3 mutated patients; one patient was not included in the chart due to missing $CD19^+$ purity data. (B) *BIRC3* clonal dynamics in 3 mutated patients; one patient carried three different *BIRC3* mutations; one patient was not included in the chart due to missing $CD19^+$ purity data. (C) *NFKBIE* clonal dynamics in 2 mutated patients.
Each patient is represented by a different color. The vertical axis displays the mutation VAF%, corrected according to the % of $CD19^+$ cells in the sample.



Supplemental Figure 3. Oncoplot displaying the somatic variants detected in the 13 genes analyzed by HaloPlex HS custom panel in 31 patients with paired samples (at baseline and at ≥1 year time point) of the responsive cohort. Illustrated are the distribution of the somatic variants, IGHV mutational status, cytogenetic profile determined by fluorescence in situ hybridization, and the mutation frequency of each gene, at baseline and at ≥1 year time point, respectively. Patient identifiers are indicated across the top row. For each patient, both the baseline (grey) and the ≥1 year (light blue) time points are represented.



Supplemental Figure 4. Oncoplot displaying the recurrent somatic variants detected by the capture lymphoid panel in the relapsed and responding cohorts. Illustrated are the distribution of the somatic variants, IGHV mutational status, cytogenetic profile determined by fluorescence in situ hybridization or lymphoid panel (del(8p)), and the mutation frequency of each gene in the relapsed and responding cohorts, respectively. Patient identifiers are indicated across the top row.