Pediatric T-cell lymphoblastic leukemia evolves into relapse by clonal selection, acquisition of mutations and promoter hypomethylation

Joachim B. Kunz,[1-3*] Tobias Rausch,[2,4,13*] Obul R. Bandapalli,[1-3] Juliane Eilers,[1,2] Paulina Pechanska,[1,2] Stephanie Schuessele,[1,2] Yassen Assenov,[5] Adrian M. Stütz,[2,4] Renate Kirschner-Schwabe,[6] Jana Hof,[6,7] Cornelia Eckert,[6] Arend von Stackelberg,[6] Martin Schrappe,[8] Martin Stanulla,[9] Rolf Koehler,[10] Smadar Avigad,[11] Sarah Elitzur,[11] Rupert Handgretinger,[12] Vladimir Benes,[13] Joachim Weischenfeldt,[4] Jan O. Korbel,[2,4] Martina U. Muckenthaler,[1,2**] and Andreas E. Kulozik[1-3**]**

[1]Department of Pediatric Oncology, Hematology and Immunology, Children's Hospital, University of Heidelberg, Germany; [2]Molecular Medicine Partnership Unit, EMBL-University of Heidelberg, Germany; [3]German Cancer Consortium (DKTK), Heidelberg, Germany; [4]European Molecular Biology Laboratory (EMBL), Genome Biology Unit, Heidelberg, Germany; [5]Division of Epigenomics and Cancer Risk Factors, The German Cancer Research Center (DKFZ), Heidelberg, Germany; [6]Department of Pediatric Oncology/Hematology, Charité - Universitätsmedizin Berlin, Germany; [7]German Cancer Consortium (DKTK) and German Cancer Research Center (DKFZ), Heidelberg, Germany; [8]Pediatrics, University Hospital Schleswig-Holstein, Campus Kiel, Germany; [9]Department of Pediatric Hematology/Oncology, Medical School Hannover, Germany; [10]Department of Human Genetics, University of Heidelberg, Germany; [11]Molecular Oncology, Felsenstein Medical Research Center and Pediatric Hematology Oncology, Schneider Children's Medical Center of Israel, Petah Tikva, Israel; [12]Children's Hospital, University Hospital Tübingen, Germany; and [13]European Molecular Biology Laboratory (EMBL), Genomics Core Facility, Heidelberg, Germany

**JBK and TR contributed equally to this work.*

***JOK, MUM and AEK contrubuted equally to this work.*

doi:10.3324/haematol.2015.129692
Manuscript received on April 24, 2015. Manuscript accepted on August 11, 2015.
Correspondence: Martina.Muckenthaler@med.uni-heidelberg.de, Andreas.Kulozik@med.uni-heidelberg.de or jan.korbel@embl-heidelberg.de

Supplemental Methods

Isolation of DNA

Bone marrow or, in one patient with a high proportion of blasts in the periphery, blood samples were enriched for mononuclear cells by Ficoll density gradient centrifugation. DNA was purified from mononuclear cells using the Gentra Puregene Cell Kit (Qiagen, Hilden, Germany).

Exome capture and Illumina sequencing

The Agilent SureSelect Target Enrichment Kit (Agilent, Santa Clara, California; vendor's protocol version 2.0.1) was used to capture all human exons for sequencing of corresponding patient samples obtained at first diagnosis, remission, and relapse. The SureSelect All Exon Kit targets regions of 50 Mb in total size, approximately 1.7% of the human genome. Briefly, 3 µg of genomic DNA was sheared with a Covaris S2 instrument to a mean size of 150 bp. Five hundred nanograms of library was hybridized for 24 hours at 65°C with the SureSelect baits. The captured fragments of the sample were sequenced as 100 bp paired reads using an Illumina HiSeq2000 instrument (based on v3 sequencing chemistry; Illumina, San Diego, California).

Analysis of the whole exome sequencing data

Sequenced DNA reads were preprocessed using Trimmomatic[1] to clip and filter adapter contaminations and to trim reads for low quality bases. The remaining reads > 36bp were mapped to build hg19 of the human reference genome with Stampy[2] v1.0.17 using default parameters, and processed with an adapted version of the EMBL in-house somatic variant analysis pipeline[3]. The number of mapped reads was >98% for all samples, indicating an efficient filtering and preprocessing of DNA reads. The fraction of targets covered >20 fold was >85% for all samples. The mean exome coverage ranged between 68 fold and 304 fold, depending on the efficiency of the capture-based enrichment. Somatic single-nucleotide variants (SNVs) were initially identified using SAMtools mpileup[4] and the Genome Analysis Toolkit (GATK)[5]. To facilitate a fast classification and identification of candidate driver mutations, all identified coding SNVs were comprehensively annotated using the ANNOVAR framework[6]. The annotation features included variants documented in dbSNP v135 (http://www.ncbi.nlm.nih.gov/projects/SNP/), the April 2012 SNV release of the 1000 Genomes Project (http://1000genomes.org), and the NHLBI-ESP project (https://esp.gs.washington.edu). All non-synonymous coding changes were annotated with the corresponding amino acid change, including a computational prediction of the functional impact on the structure and function of the respective protein using SIFT and PolyPhen-2[7, 8]. All SNVs were further annotated with GERP conservation scores[9] and the presence or absence of segmental duplications. To identify possible somatic driver mutations, candidate SNVs were filtered for non-synonymous, stopgain or stoploss SNVs, requiring a SAMtools mpileup SNV quality greater or equal to 50, and requiring absence of segmental duplications. We further removed all inferred somatic SNV sites corresponding to SNPs identified in the 1000 Genomes Project[10] with an allele frequency >1% as well as SNPs present in our EMBL in-house sequencing database of germline genomes, which include sites that are commonly identified as false positive SNVs. Leukemia-specific mutations were identified by filtering against the corresponding remission sample. For somatic insertion and deletion detection, the initial Stampy alignments were realigned using GATK's IndelRealigner and a list of curated known InDels [11]. InDel calling was done with two separate tools, Platypus v0.2.3 [12]and the GATK HaplotypeCaller v2.5 [5]. InDels identified by both methods were annotated using ANNOVAR [6] and filtered for exonic insertions and deletions. Using short reads, the insertion and deletion discovery is limited to small variants of usually less than 10bp although greater InDels can be detected if there is a unique prefix and suffix alignment for a split-read. In the present study the longest detected InDel was 15 bp. Leukemia-specific InDels were identified by removing InDels predicted to be present in the corresponding remission sample. Leukemia-specific SNVs and InDels were validated by Sanger sequencing following PCR amplification.

Targeted deep sequencing

The HaloPlex Target Enrichment Kit (Agilent, Santa Clara, California; vendor's protocol version D.5, May 2013) was used according to the manufacturer's instructions. Starting material was 225 ng of genomic DNA. The captured fragments of the sample were sequenced as 100 bp paired reads using an Illumina HiSeq instrument (Illumina, San Diego, California). As we had started target capture experiments before the results of WES were fully available, two slightly different target capture designs were used, see Supplemental Table 4: Single nucleotide variants (SNVs) and InDels that had been detected during WES (see Supplemental Table 1) were included. Each target region was defined to include 50 bases upstream and downstream of a variant. The total number of target regions was 495, and together they spanned 47.2 kb. The design obtained from Agilent had a total size of 45.4 kb and covered 96.2% of the region of interest (see Supplemental Table 4).

Analysis of the targeted deep sequencing data

Raw DNA sequencing reads were preprocessed using cutadapt[13] to remove the HaloPlex-specific adapter sequences. Trimmomatic[1] was used to filter and trim low quality bases and to crop the first 5bp that can be reference-biased due to the HaloPlex restriction enzyme footprint. All trimmed reads shorter than 36bp were discarded. The remaining pairs were aligned to build hg19 of the human reference genome with bwa mem v0.7.7 [14]. A custom python script was used to annotate all somatic SNVs identified previously by WES with the minor allele frequency in the deep coverage targeted sequencing data.

In order to identify subclonal SNVs that were selected for during treatment, leukemia specific SNVs were filtered for an allele frequency of <0.05 in primary disease and remission and of >0.05 in relapse. For the remaining SNVs, the allele frequency in a sample of interest was compared to the background frequency that was calculated as average allele frequencies in the samples from all other patients that did not, by chance, carry the same mutation. Samples with an allele frequency that exceeded the background frequency by more than 3 standard deviations were considered to carry that specific allele at a subclonal level. No further correction for multiple testing was applied.

Because in most patients we were restricted to few mutations in order to define a subclone, we defined all mutations to belong to the major clone that had an allele frequency of at least one third of the most abundant heterozygous mutation in this patient.

DNA Methylation Analysis Using 450k BeadChip Arrays

Genomic DNA (200 ng) was bisulfite converted using the EZ DNA Methylation Gold Kit (Zymo Research). The Infinium methylation assay (Illumina) was carried out as previously described [15]. Data from the 450k Human Methylation Array were normalized by the Beta Mixture Quantile (BMIQ) method[16] using the RnBeads analysis software package[17]. The methylation level of a CpG locus is expressed as beta value (β) which represents the proportion of methylated alleles of all alleles. A gene promoter was defined as the region spanning 1.5 kilobases (Kb) upstream and 0.5 Kb downstream of the respective transcription start site.

Integrated analysis

In order to evaluate functions that were altered at relapse, an integrated pathway analysis was performed using Ingenuity Pathway Analysis (IPA, Version 21249400). This tool provides information about diseases, molecular function and biological process categories, as well as biological pathways related to genes. In addition, IPA maps each gene within a global molecular network developed from information contained in the Ingenuity Pathways Knowledge Base. Gene networks are generated algorithmically based on their connectivity in terms of expression, activation, transcription, and/or inhibition. A network in IPA is defined as a graphical representation of the molecular relationships between genes, represented with nodes, and the biological relationship between them represented by connecting lines. All connections are supported by published data stored in the Ingenuity Pathways Knowledge Base. IPA ranks all genes based on their connectivity, using a generalization of the concept of node degree, which measures the number of single genes to which a gene is connected.
For analysis of mutations, the list of genetic changes (detected either by MLPA or by WES) that were present in the major clone at both primary leukemia and relapse was compared to the list of genes that were mutated in a relapse-specific manner. If one gene in the same patient carried different mutations in primary disease and in relapse, this was considered as a single change that was not relapse specific.

For analysis of genes somatically mutated or deleted in leukemia, genetic alterations detected by WES and MLPA were combined. The list of genes that was altered at both time points (primary disease and relapse) was compared to the list of genes altered only in relapse. As a reference set the Ingenuity Knowledge Base (genes only) was selected.

After methylation analysis by Illumina 450k array, promoters that were represented by at least 3 probes on the 450k array and that had a gene symbol assigned were analyzed (in total promoters of 8297 genes). Promoters that were hypomethylated (β in relapse < β in primary – 0.2) in at least 3 different patients at relapse were included in Ingenuity Pathway Analysis, as a reference set the total 8297 genes entering the analysis were selected.

References

1. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30(15):2114-2120.

2. Lunter G, Goodson M. Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads. Genome Res. 2011;21(6):936-939.

3. Rausch T, Jones DT, Zapatka M, et al. Genome sequencing of pediatric medulloblastoma links catastrophic DNA rearrangements with TP53 mutations. Cell. 2012;148(1-2):59-71.

4. Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25(16):2078-2079.

5. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010;20(9):1297-1303.

6. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010;38(16):e164.

7. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc. 2009;4(7):1073-1081.

8. Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. Nat Methods. 2010;7(4):248-249.

9. Cooper GM, Stone EA, Asimenos G, et al. Distribution and intensity of constraint in mammalian genomic sequence. Genome Res. 2005;15(7):901-913.

10. Consortium GP. An integrated map of genetic variation from 1,092 human genomes. Nature. 2012;491(7422):56-65.

11. Mills RE, Pittard WS, Mullaney JM, et al. Natural genetic variation caused by small insertions and deletions in the human genome. Genome Res. 2011;21(6):830-839.

12. Rimmer A, Phan H, Mathieson I, et al. Integrating mapping-, assembly- and haplotype-based approaches for calling variants in clinical sequencing applications. Nat Genet. 2014;46(8):912-918.

13. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. . EMBnetjournal. 2011;17(1):10-12.

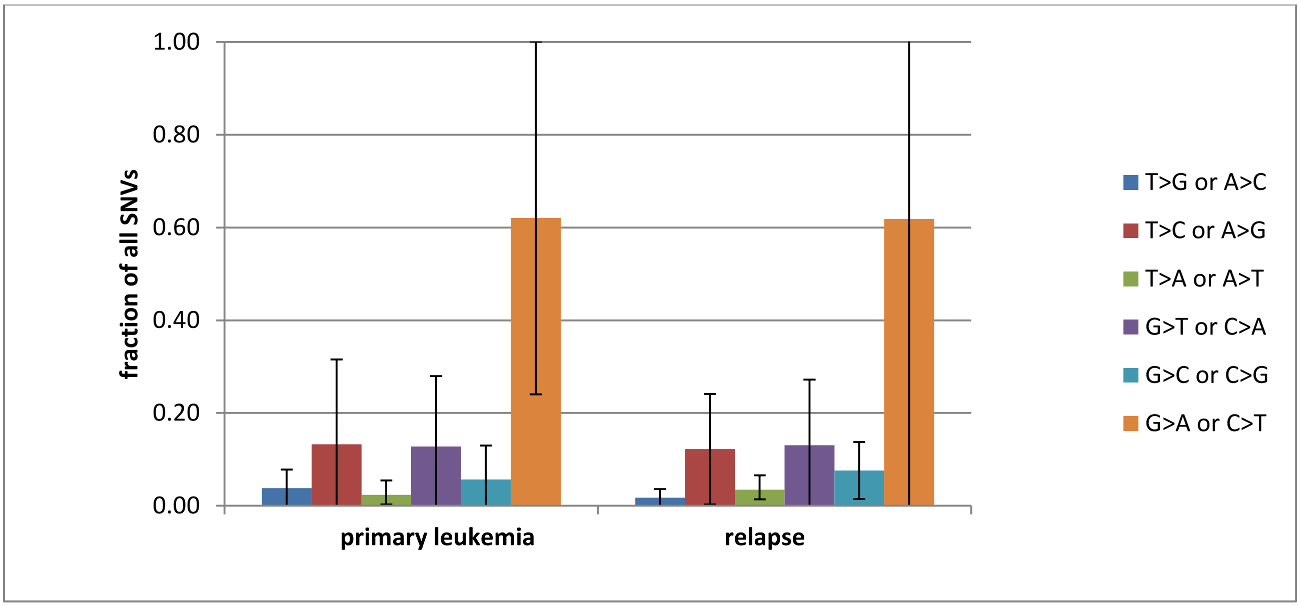
14. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25(14):1754-1760.

15. Bibikova M, Le J, Barnes B, et al. Genome-wide DNA methylation profiling using Infinium(R) assay. Epigenomics. 2009;1(1):177-200.

16. Teschendorff AE, Marabita F, Lechner M, et al. A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. Bioinformatics. 2013;29(2):189-196.

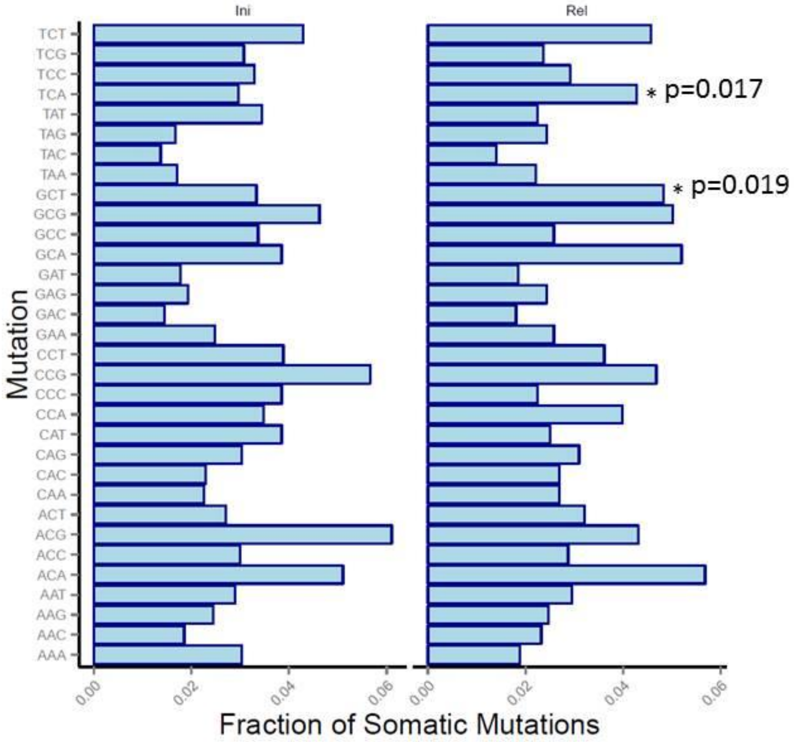
17. Assenov Y, Muller F, Lutsik P, Walter J, Lengauer T, Bock C. Comprehensive analysis of DNA methylation data with RnBeads. Nat Methods. 2014;11(11):1138-1140.

Supplemental Figure 1 A



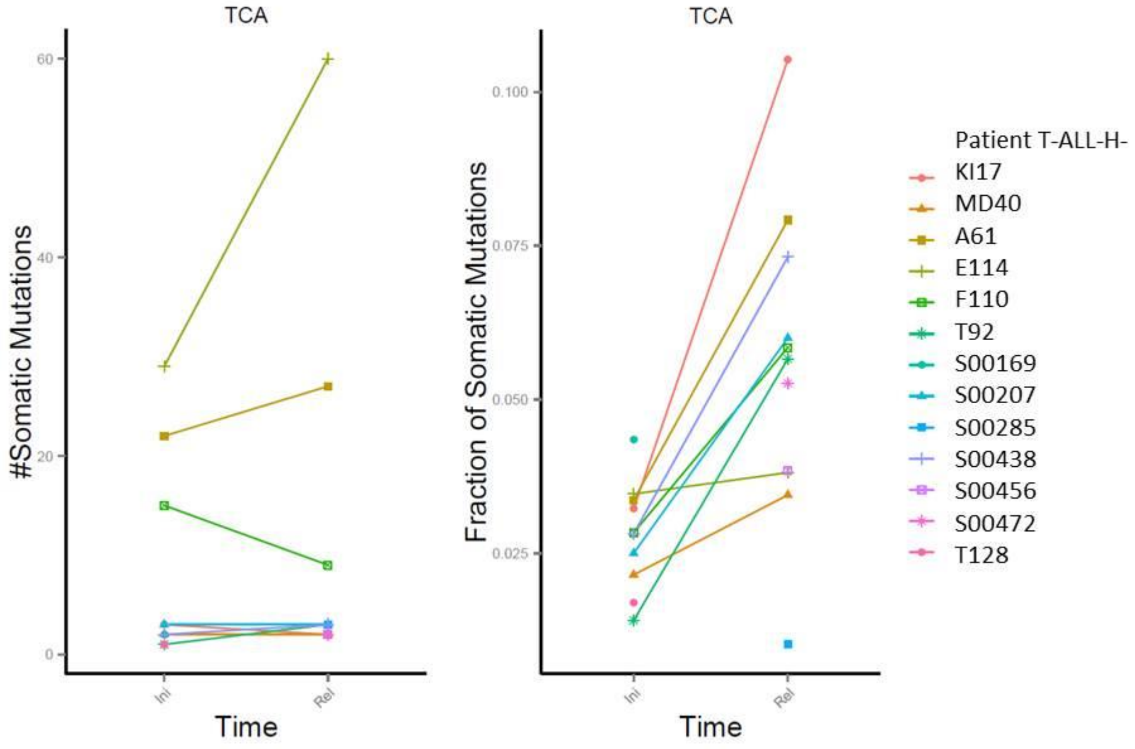
Distribution of mutation types does not differ between primary and relapsed leukemia: The fractions of each type of somatic SNV (both synonymous and non-synonymous) for all 13 patients are given in average. Error bars denote standard deviations.

Supplemental Figure 1 B



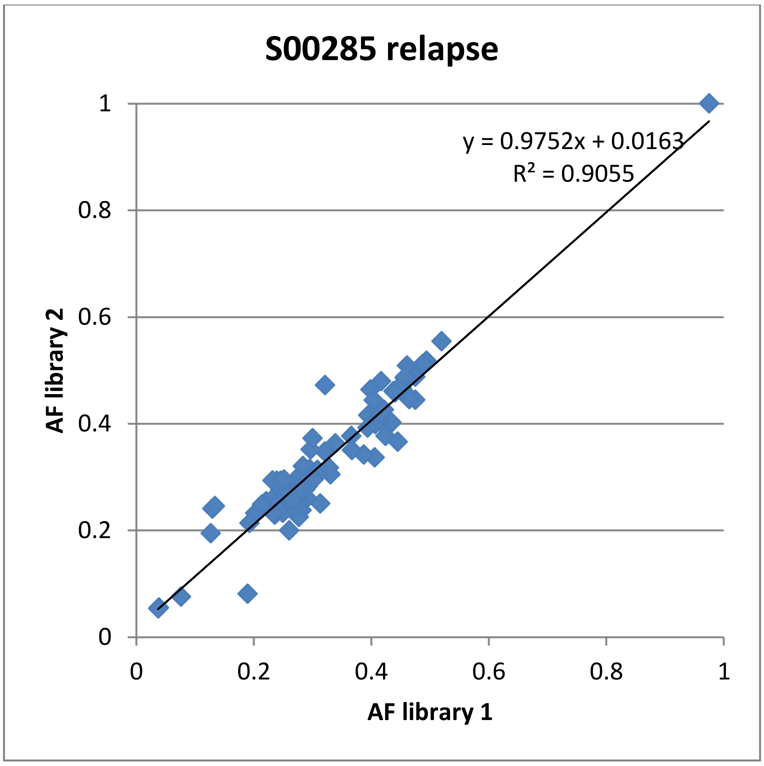
Distribution of mutation types in primary and relapsed leukemia: All genomic mutations (exonic, intronic, intergenic) were called and the fraction of mutations occurring at the central position of each possible trinucleotide sequence was determined. Left panel, mutations in primary leukemia; right panel, mutations specific for relapse. p-values (t-test) below 0.02 are reported.

Supplemental Figure 1 C



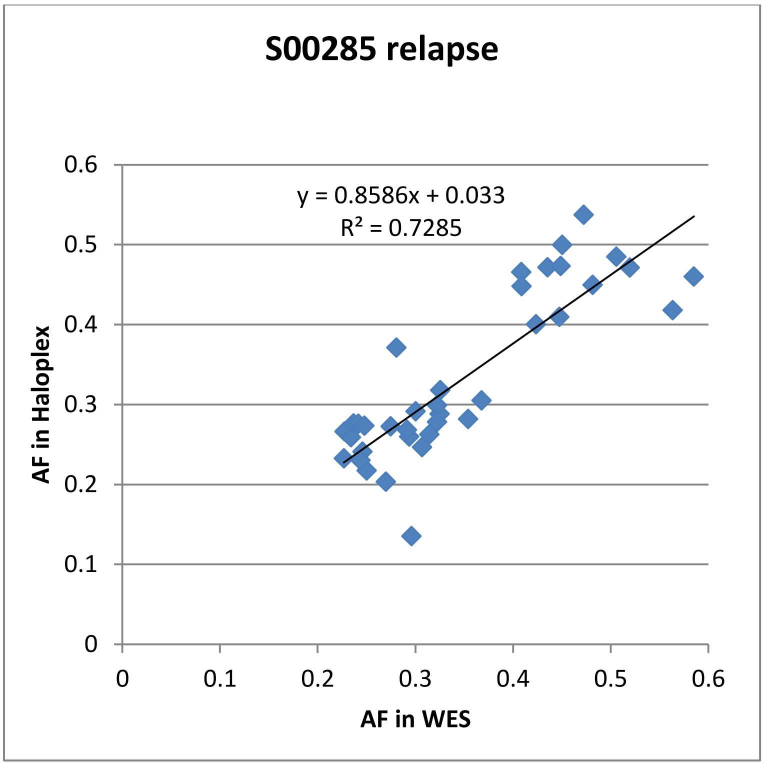
Frequency of somatic mutations at the central base of the trinucleotide sequence $^{5'}$TpCpA$^{3'}$/$^{5'}$TpGpA$^{3'}$ in primary leukemia and in relapse. The number (left panel) and fraction (right panel) of somatic mutations occurring at the central position of the trinucleotide sequence $^{5'}$TpCpA$^{3'}$/$^{5'}$TpGpA$^{3'}$ were determined. The numbers for initial diagnosis were compared to the numbers of relapse-specific mutations. Because of different capture efficiencies, the number of genomic mutations that can be analyzed varies much more between patients than exonic mutations.

Supplemental Figure 2 A



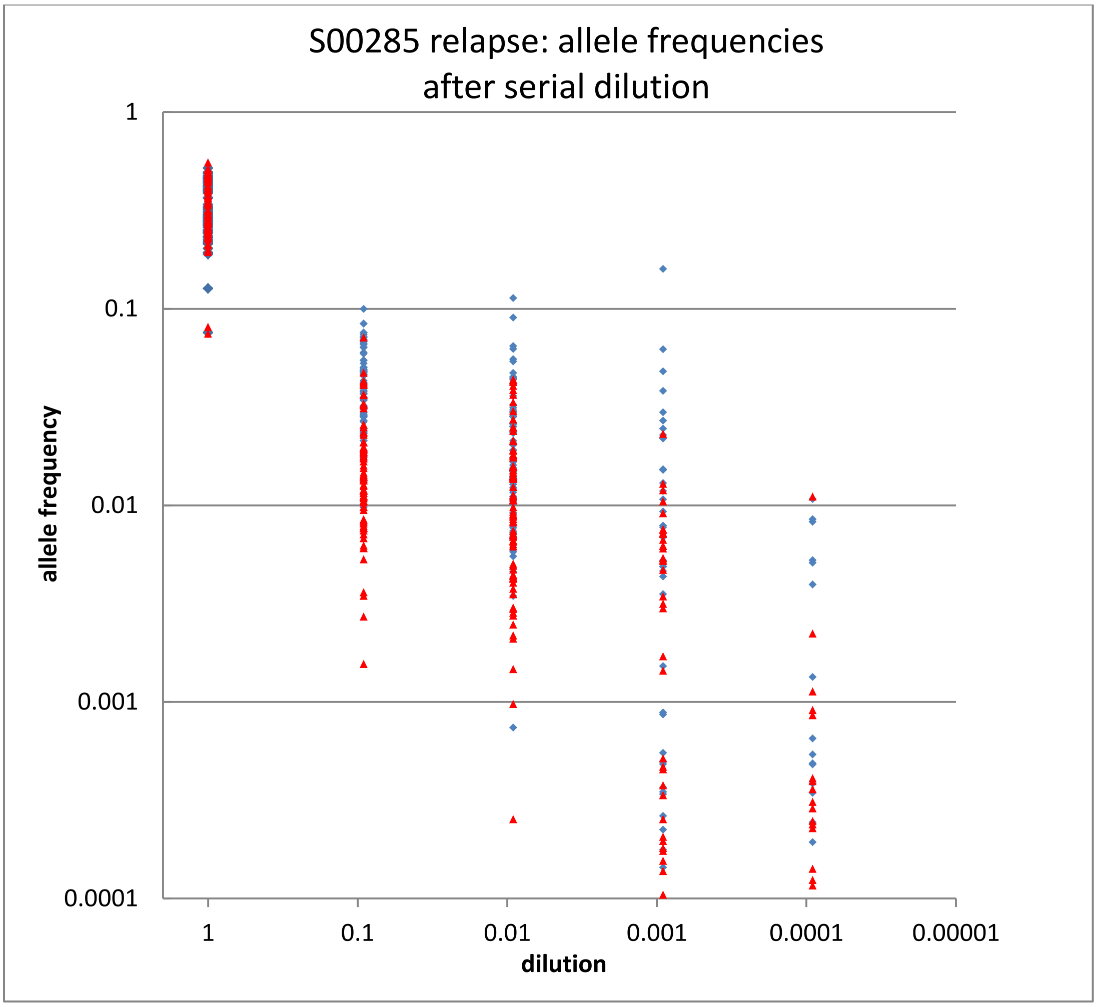
Deep sequencing after target capture with HaloPlex is reproducible: The same DNA sample (patient S-00285, relapse) was subjected to target capture by HaloPlex two times independently. For all SNVs that reached an allele frequency (AF) of >0.05 in one of the libraries, AFs in library 2 were plotted over AF in library 1.

Supplemental Figure 2 B



HaloPlex reproduces allele frequencies found in well covered loci in WES: For all loci that were covered more than 150x in WES, allele frequencies found by HaloPlex were plotted over allele frequencies found by WES.

Supplemental Figure 2 C



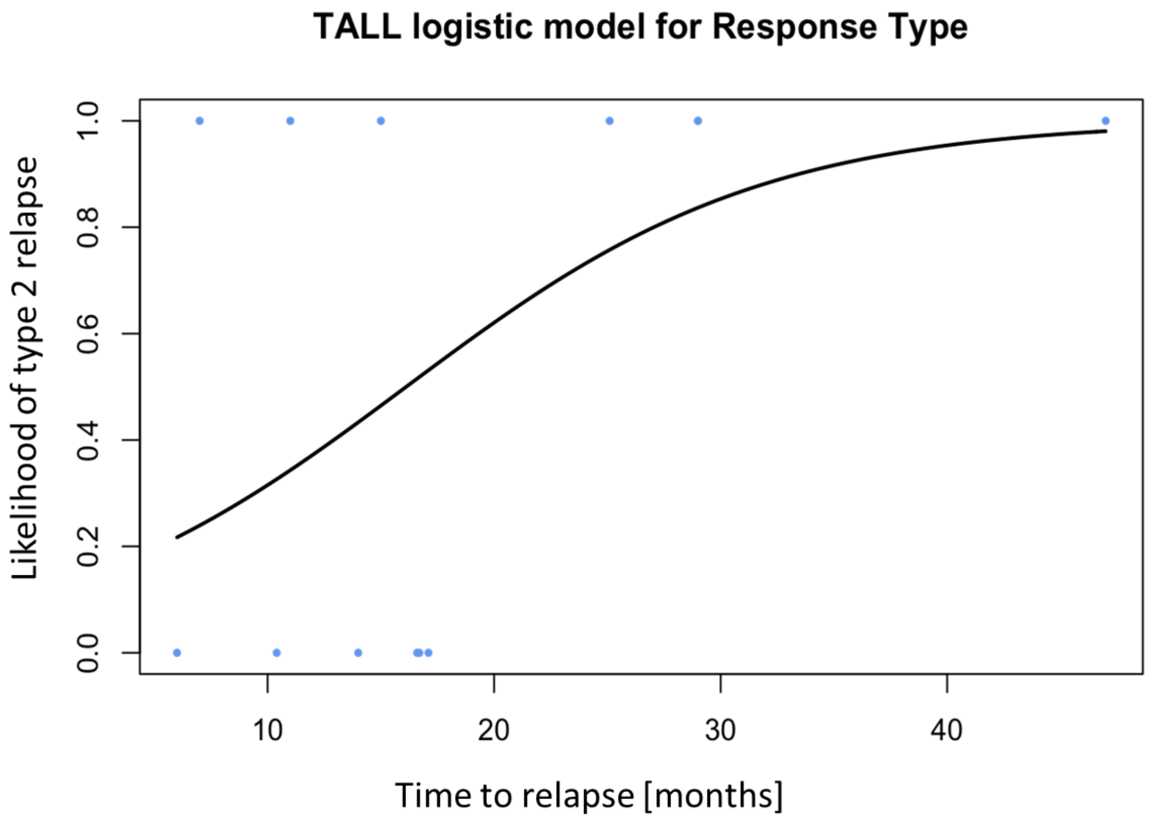
HaloPlex results are consistent after serial dilution: DNA samples from relapse were diluted by mixing 11 samples from different patients in equal amounts. This mix was further serially diluted by mixing with DNA from a healthy control in a ratio of 1 in 10. Background allele frequencies (AF) detected in the healthy control DNA were subtracted. Here, AFs for all SNVs that were detected in the relapse sample from patient S00285 (AF > 0.05) were plotted over the dilution factor. The average Pearson correlation coefficient after logarithmic transformation was 0.91 (standard deviation 0.13), indicating that the dilution factor was truly reflected in the allele frequencies. The experiment was done in duplicate (blue squares/ red triangles). Please note that all allele frequencies that were determined are reported, including those that did not exceed the threshold of specific detection (see Supplemental Figure 2 D).

Supplemental Figure 2 D

					AF in dilution series A					AF in dilution series B					mean AF of	SD of AF in un-	fraction of
#chr	start	ref	alt	gene	undiluted	1 : 11	1 : 110	1 : 1100	1 : 11000	undiluted	1 : 11	1 : 110	1 : 1100	1 : 11000	unrelated samples	related samples	false positives
chrX	86919844	G	A	KLHL4	0.269582	0.063909	0.044547	0.000261	0	0.251214	0.042256	0.042579	0.007358	0	0.000162709	0.000170285	0/21
chrX	122551596	C	T	GRIA3	0.220854	0.041684	0.000883	0	0	0.254545	0.071552	0.0406	0	0	0.000122223	0.000230318	1/21
chrX	153588214	C	T	FLNA	0.291433	0.050383	0.009143	0.159218	0	0.30131	0.020986	0.018908	0	0.000907	0.000173329	0.000376292	0/21
chr13	49772563	C	T	FNDC3A	0.202374	0.01386	0.008022	0.004473	0	0.232547	0.007997	0.004374	0.005432	0.000211	0.00015651	0.000126724	0/21
chr13	86369194	G	T	SLITRK6	0.192903	0.021534	0	0	0	0.213851	0.017472	0	0.00011	0	8.86793E-05	9.00806E-05	0/21
chr12	4737889	C	T	AKAP3	0.457111	0.059998	0.064624	0.000266	0.005455	0.486265	0.036778	0.015831	0.000573	0.000339	0.000428122	0.000424697	0/21
chr12	7015028	C	T	LRRC23	0.270792	0.037064	0.014851	0	0	0.254402	0.006421	0.006694	0.000879	0.002588	0.000130261	0.000181779	1/21
chr12	51723604	C	G	CELA1	0.239223	0.059896	0.028238	0.014346	0.012706	0.293027	0.026054	0.022984	0.011057	0.015587	0.01478301	0.008482216	0/21
chr12	51771098	A	G	GALNT6	0.276632	0.032363	0.009321	0.000148	0	0.259959	0.011322	0.009991	0	0	7.60195E-05	9.09248E-05	0/21
chr12	110815336	C	T	ANAPC7	0.292539	0.048955	0.022375	0.00019	0	0.283643	0.012999	0.018874	0.002547	0.000159	0.000257379	0.000474091	1/21
chr11	46905462	G	A	LRP4	0.278896	0.054979	0.012618	0.000575	0.000406	0.303756	0.018248	0.003407	0.000169	0.000307	0.000230183	0.000215616	1/21
chr11	57564152	G	A	CTNND1	0.275614	0.034793	0	0	0	0.24198	0.014655	0.043745	0.000511	0.000463	0.000309741	0.000303273	0/21
chr11	119059247	G	A	PDZD3	0.245102	0.035646	0.019312	0	0	0.276776	0.006433	0.006996	0.000566	0	0.000226266	0.000267226	0/21
chr11	134184273	G	A	GLB1L3	0.434602	0.048672	0.090754	0	0	0.402103	0.012713	0.022123	0.012746	0.000359	0.000211645	0.000423057	1/21
chr10	12143158	C	A	DHTKD1	0.330411	0.047225	0.014145	0.000133	0	0.304759	0.013976	0.014239	0.000273	0.000385	0.000354833	0.000202597	0/21
chr10	61967816	G	T	ANK3	0.398827	0.071718	0.026412	0.000208	0	0.464112	0.041511	0.018471	0.000638	0.00065	0.000348382	0.000271028	0/21
chr10	99160258	A	G	RRP12	0.424161	0.041072	0.007168	0.024162	0	0.376628	0.012896	0.015358	0.007584	0.000221	0.00054558	0.000458107	1/21
chr10	104852955	C	T	NT5C2	0.494052	0.08448	0.062868	0.012436	0	0.517861	0.047926	0.044233	0.002313	0.000493	0.000554055	0.0003585	1/18
chr10	120354498	G	A	PRLHR	0.245751	0.029439	0.008124	0.000248	0	0.246968	0.014827	0.004857	0.005155	0.000151	0.000588991	0.000556759	0/21
chr17	36718035	G	A	SRCIN1	0.290931	0.035745	0.017001	0	0	0.26032	0.013154	0.014545	0.000238	0	0.000284286	0.000353859	0/21
chr17	62493004	C	A	POLG2	0.281615	0.022523	0.004829	0	0	0.237958	0.010784	0.009267	0.000182	0.000521	0.000335075	0.000261129	0/21
chr17	73124825	G	A	ARMC7	0.395059	0.03247	0.016918	0.00027	0	0.415675	0.007675	0.010179	0.000127	0.000155	0.000942564	0.00149865	1/21
chr17	76157321	G	A	C17orf99	0.459743	0.050181	0.030849	0.005889	0	0.482422	0.01939	0.009207	0.000772	0.000452	0.000235015	0.000423168	1/21
chr17	76991157	C	T	CANT1	0.231156	0.031579	0	0	0	0.258065	0	0	0	0	0	0	0/21
chr16	3421856	A	G	MTRNR2L4	0.47476	0.065632	0.012702	9.96E-05	0.00099	0.488123	0.021511	0.004633	0.000438	0.00016	0.000457214	0.00083723	1/21
chr16	8998333	G	A	USP7	0.276439	0.035237	0.0129	0.000305	0	0.224825	0.012044	0.006662	0	0.000187	0.000200168	0.000180914	0/21
chr16	31309192	G	A	ITGAM	0.295277	0.028658	0.021307	0.000657	0	0.314832	0.014206	0.00061	0	0	0.000227281	0.000278356	0/21
chr16	57507576	A	G	DOK4	0.405504	0.028654	0.016216	0.000793	0	0.336461	0.021219	0.007427	0.005468	0.000604	0.000297769	0.000348456	0/21
chr16	57760020	C	T	CCDC135	0.290613	0.027207	0.024106	0	0	0.281777	0.010083	0.007696	0	0.000151	0.000201618	0.000266505	0/21
chr16	69377359	C	T	TMED6	0.277654	0.02908	0.014078	0.000749	0.009346	0.254344	0.007985	0.003884	0.001316	0.000244	0.000447613	0.000409561	0/21
chr15	43814245	A	G	MAP1A	0.320517	0.049497	0.012814	0.038751	0	0.347219	0.019241	0.004943	0.000142	0.000295	0.00029626	0.000207993	0/21
chr15	69011148	G	A	CORO2B	0.463599	0.074854	0.008314	0	0	0.451294	0.014894	0.014737	0.00038	0.000292	0.000148041	0.00032496	1/21
chr15	91292615	G	T	BLM	0.445114	0.046118	0.041543	0.021855	0	0.366006	0.019425	0.011007	0	0	7.90597E-05	0.000149964	1/21
chr15	91422743	G	A	FURIN	0.263673	0.026597	0.015001	0.005333	0.000339	0.281424	0.007799	0.00938	0.000251	7.98E-05	8.79644E-05	0.000111644	0/21
chr14	24655963	C	T	IPO4	0.308041	0.023885	0.030785	0.005173	0.000648	0.313209	0.011443	0.001774	0.000327	0.000169	0.000457287	0.000403356	0/21
chr14	24656631	G	T	IPO4	0.271802	0.041831	0.020299	0	0.000532	0.285581	0.008454	0.002152	0.00012	0.000907	0.000127797	0.000119073	0/21
chr14	81610309	G	A	TSHR	0.297017	0.017954	0.113798	0	0	0.351879	0.041734	0	0.023628	0.000776	0.000102102	0.000280862	1/21
chr19	2717401	G	A	DIRAS1	0.265199	0.036981	0.007368	0	0.01087	0.270885	0.011312	0.004419	0.000375	0.000369	0.000313763	0.000518492	1/21
chr19	11346351	C	T	DOCK6	0.47121	0.069177	0.054022	0.000947	0.000277	0.495546	0.023516	0.014751	0.007651	0.000312	0.000196831	0.000182223	0/21
chr19	18420607	C	T	LSM4	0.24148	0.031701	0.011394	0.015494	0	0.251336	0.008851	0.003341	7.46E-05	0	0.000297927	0.000264471	0/21
chr19	33617614	G	C	GPATCH1	0.387204	0.099839	0	0	0	0.341789	0.010765	0.004476	0	0	0.000204961	0.00024555	0/21
chr19	38573448	C	A	SIPA1L3	0.251446	0.03091	0.028316	0	0	0.269174	0.012188	0.01403	0.000135	0.000161	0.00027619	0.000463971	1/21
chr19	55489171	G	A	NLRP2	0.452325	0.049153	0.045405	0.024928	0	0.466325	0.025349	0.03391	0.000117	0.000216	0.000151	0.000151576	0/21
chr19	58265360	G	A	ZNF776	0.299905	0.039989	0.01414	0	0.001081	0.372432	0.019042	0.014646	0.010916	0	0.000322786	0.000263039	0/21
chr18	21426278	G	T	LAMA3	0.414466	0.05325	0.017913	0.000259	0.000256	0.404487	0.012314	0.007106	0.003639	0.00017	0.000345618	0.000228589	0/21
chr20	60899229	C	T	LAMA5	0.276755	0.02955	0.005079	0.001037	0	0.281448	0.008332	0.005815	0.009999	0.000717	0.000656135	0.000331664	0/21
chr7	5267763	G	T	WIPI2	0.282795	0.029223	0.007385	0.00092	0.00112	0.304014	0.010408	0.005624	0.000573	0.000568	0.000703316	0.000287632	0/21
chr7	5540303	G	T	FBXL18	0.28806	0.038779	0.005056	0.004923	0	0.277028	0.011155	0.007375	0.000133	0	9.17958E-05	0.000102649	0/21
chr7	100470908	G	A	TRIP6	0.519869	0.075804	0.014291	0.003917	0	0.554357	0.019505	0.009273	0.000243	0	0.000404589	0.000475475	0/21
chr7	141385384	C	T	KIAA1147	0.407218	0.069441	0.03003	0.015772	0.004511	0.406366	0.016041	0.012971	0.000223	0	0.000176304	0.000193633	0/21
chr6	16306766	C	T	ATXN1	0.43804	0.0636	0.023529	0.011008	0	0.460715	0.023138	0.015997	0.000107	0.000342	0.00013223	9.80952E-05	0/21
chr6	30576618	G	T	PPP1R10	0.282588	0.024277	0.011209	0.027986	0	0.269517	0.02057	0.016796	0.001111	0.001415	0.000951126	0.000583126	1/21
chr6	31611683	C	T	BAG6	0.366099	0.038526	0.012223	0.00092	0	0.376957	0.010932	0.003318	0.000907	0	0.000321805	0.000217424	0/21
chr6	42579942	G	T	UBR2	0.235967	0.050997	0.008889	0	0	0.229021	0.024461	0.011848	0.014374	0.000219	0.002073394	0.008100359	1/21
chr5	893136	T	C	TRIP13	0.327586	0.028885	0	0	0	0.317631	0.010224	0	0.00105	0.001854	0.000401181	0.000688189	0/21
chr5	72469299	G	A	TMEM174	0.487402	0.071798	0.055998	0.001476	0	0.511759	0.043573	0.019759	0.000222	0.000189	0.000403887	0.000638165	1/21
chr5	78301114	G	T	DMGDH	0.44075	0.066651	0.030762	0.000288	0	0.458752	0.033358	0.030992	0	0.000126	0.000452831	0.000601325	1/21
chr5	167887746	G	A	WWC1	0.45463	0.069707	0.019393	0.048276	0	0.476349	0.018276	0.021552	0.000356	0.000199	0.000185434	0.000180718	0/21
chr5	179992901	C	T	CNOT6	0.283929	0.038928	0.006839	0.000312	0	0.320301	0.015535	0.005603	0	0.000161	0.000458788	0.000554067	1/21
chr4	7045536	T	A	TADA2B	0.214144	0.059547	0.02406	0	0	0.248397	0.00416	0.00153	0	0	0.000160349	0.000362266	0/21
chr4	15629579	C	T	FBXL5	0.249151	0.031523	0.01459	0.001113	0.000198	0.232603	0.0204	0.018939	0.00081	0.000602	0.000738772	0.000372819	1/21
chr4	91230147	G	A	CCSER1	0.421733	0.055001	0.025415	0.000848	0	0.425968	0.031495	0.039	0	0	0.000150058	0.000236767	0/21
chr4	153249385	G	A	FBXW7	0.46078	0.07031	0.029049	0.013635	0.001139	0.508878	0.043027	0.037082	0.000129	0.000488	0.001342073	0.002761087	1/21
chr3	49166532	C	T	LAMB2	0.189514	0.023612	0.014157	0	0.001396	0.08096	0.003525	0.002241	0.000153	0.000159	0.000263277	0.000482107	1/21
chr3	196388585	G	A	LRRC33	0.267561	0.02453	0.010657	0.0001	0.00836	0.241651	0.008228	0.017609	0.013021	0.000103	0.000118899	0.000106854	0/21
chr2	27353440	C	T	ABHD1	0.232072	0.023188	0.015312	0.000326	0.000486	0.293207	0.009552	0.008407	0.006776	0	0.000231285	0.000369097	1/21
chr2	55566692	C	T	CCDC88A	0.458211	0.04903	0.047987	0.007782	0.000597	0.488164	0.020274	0.025878	0.008163	0.000419	0.000748554	0.000659671	0/21
chr2	213872610	C	T	IKZF2	0.367043	0.034942	0.017967	0.008415	0	0.350183	0.017281	0.014444	0.000667	0.000624	0.000587023	0.000289871	0/21
chr1	17413177	G	A	PADI2	0.404208	0.067474	0.015726	0	0	0.444131	0.020704	0.016669	0.00695	0.000201	0.000366848	0.000288095	1/21
chr1	22928152	G	A	EPHA8	0.406849	0.021574	0.044548	0.029973	0	0.399035	0.013877	0.009241	0	0.000398	0.000301448	0.000239284	0/21
chr1	23233293	C	T	EPHB2	0.417457	0.084088	0.010966	0	0	0.418131	0.023978	0.006497	0.003766	0.000377	0.000685121	0.000874487	1/21
chr1	44360122	G	A	ST3GAL3	0.393741	0.050469	0.014319	0.009821	0.000507	0.392531	0.017907	0.016341	0.000245	0.011619	0.000678337	0.000536802	0/21
chr1	55474129	C	A	BSND	0.12677	0.013517	0.003662	0.000209	0	0.194409	0.002929	0.002684	0.000414	0.000148	0.000125797	0.000135525	0/21
chr1	78478866	C	T	DNAJB4	0.245364	0.04745	0.007126	0	0	0.29313	0.011338	0.00042	0.000636	0.00011	0.000217274	0.000252615	1/21
chr1	150444785	C	T	RPRD2	0.302209	0.043588	0.015365	0.002044	0.005629	0.296038	0.011532	0.011722	0.000376	0.000504	0.000294485	0.000152011	0/21
chr1	155257083	C	T	HCN3	0.243505	0.032109	0.025553	0.005566	0.000569	0.238726	0.007139	0.00725	0.000358	0.000453	0.000278785	0.000201729	1/21
chr1	166039591	G	A	FAM78B	0.474944	0.048009	0.037681	0.000823	0.001138	0.444902	0.027509	0.015853	0.007923	0.001397	0.001211658	0.000681834	0/21
chr1	181750598	T	C	CACNA1E	0.279745	0.035702	0.007148	0	0.000235	0.271989	0.016263	0.007326	0.005733	0.000137	0.000210032	0.000167364	0/21
chr1	183515184	G	C	SMG7	0.416591	0.073192	0.032237	0	0	0.479496	0.037397	0.024856	0.000165	0.000434	0.000177448	0.000403321	1/21
chr1	186092230	G	A	HMCN1	0.265703	0.024847	0.024254	0.00022	0	0.257022	0.021318	0.011103	0.000589	0.000352	0.000579582	0.00049831	0/21
chr1	200827060	G	T	CAMSAP2	0.313262	0.047915	0.007309	0.062305	0	0.25	0.0127	0.013777	0	0	8.39305E-05	0.000193346	1/21
chr1	203743327	G	A	LAX1	0.278013	0.042625	0.03292	0	0	0.277971	0.014609	0.011004	0.000216	0.000141	0.000193852	0.000177632	0/21
chr9	4622462	T	A	SPATA6L	0.46481	0.068235	0.026307	0	0	0.445812	0.03326	0.025245	0.003377	0.000482	0.000175527	0.000150367	0/21
chr9	130580470	C	T	ENG	0.251308	0.029742	0.005813	0	0	0.295139	0.007943	0.004073	0.000505	0.000253	0.000187259	0.000298543	1/21
chr9	139397768	A	G	NOTCH1	0.076385	0.010509	0.011003	0.001267	0.000596	0.075386	0.002327	0.004306	0.000465	0.001079	0.000867658	0.000530601	1/20
chr9	139399365	A	G	NOTCH1	0.260086	0.07585	0.044001	0.000439	0	0.199699	0.032661	0.027765	0.000295	0.000161	0.000450329	0.00105099	1/19
chr8	91657509	C	T	TMEM64	0.321278	0.024666	0.015522	0.000375	0	0.472202	0.018813	0.011399	0	0.00052	0.000236374	0.000255184	0/21
chr8	103850976	C	T	AZIN1	0.338676	0.037709	0.026695	0.000477	0.00037	0.363113	0.019522	0.004505	9.52E-05	0.000246	0.000203394	0.000188769	1/21
chr8	145112992	T	C	OPLAH	0.234773	0.014386	0.012275	0.008925	0	0.231206	0.008873	0.016251	0.008696	0	0.000325521	0.00040878	1/21

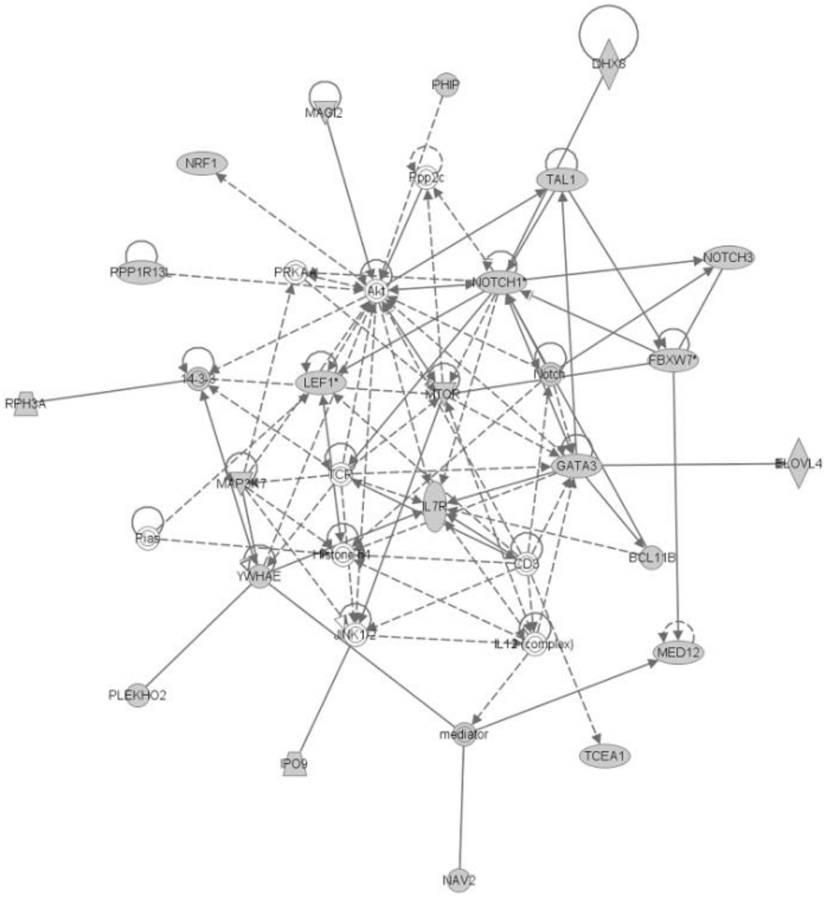
SNVs can be detected by HaloPlex ultradeep sequencing in dilutions of up to 1:11.000. Here, AFs for all SNVs that were detected in the relapse sample from patient S00285 (AF > 0.05; see also Supplemental Fig. 2C) are compared to the average AFs of all other, unrelated samples. The threshold of detection was set to the average allele frequency of unrelated samples plus three standard deviations. AFs exceeding the threshold for detection are marked in orange. The experiment was done in duplicate (dilution series A/B). At least 91% of leukemia-specific SNVs were detected after diluting leukemia DNA in control DNA in a ratio of 1:110, 25% were detected in a dilution of 1:1.100 and 3% in a dilution of 1:11.000. The rate of mutations that were detected as false positives is 1.8 % (35 false positives in 1.863 comparisons, rightmost column).

Supplemental Figure 3

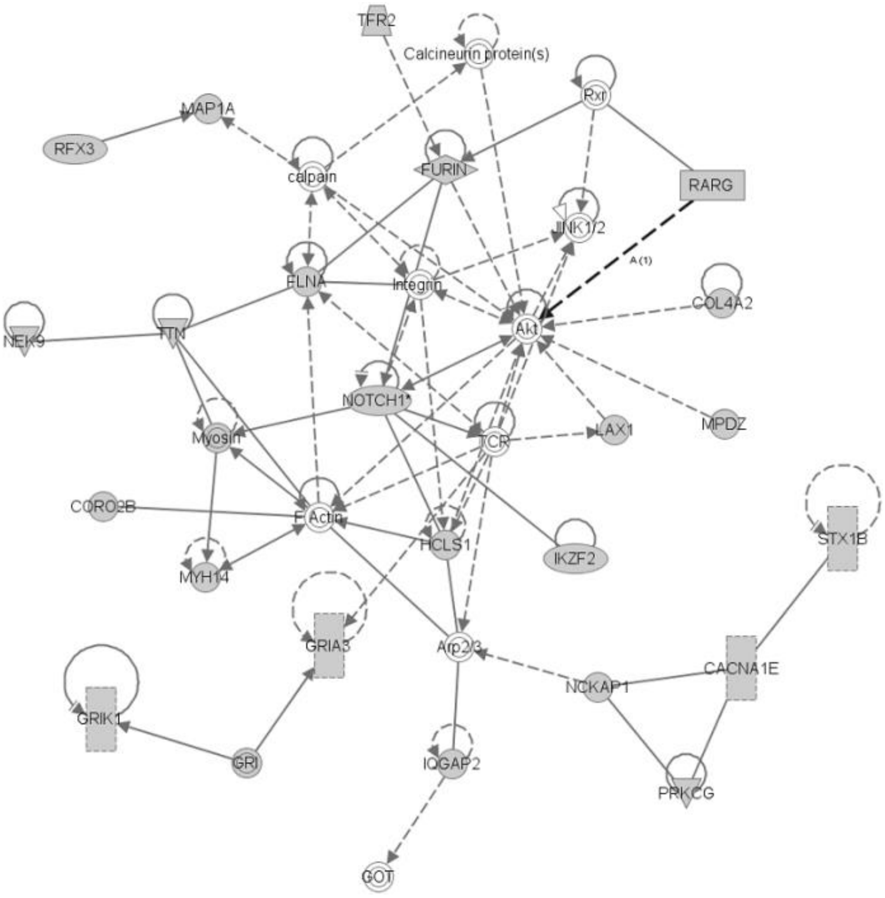


Logit Model for type of relapse dependent on time to relapse: Logistic regression was used to fit the time to relapse in months to type 1 and type 2 relapse. For every increase in month, the odds ratio of type 2 increases by a factor of 1.135 (coefficients as odds-ratios). Blue dots- observed data.

Supplemental Figure 4 A



Supplemental Figure 4 B



Ingenuity Pathway Analysis Key regulatory networks: Genes are represented as nodes, and the biological relationship between nodes is represented as an edge (line). Grey nodes signify genes that are mutated, colorless nodes are “missing links” added by the software. All edges are supported by at least 1 reference from the Ingenuity Knowledge Base.

A Network was constructed from genes that were mutated both in primary leukemia and in relapse.

B Network was constructed from genes that were mutated in relapse only.

All supplemental tables are provided in a single Supplemental Excel file.

Supplemental Table1: WES results

Leukemia-specific mutations detected by WES: Mutations detected in primary disease are in black, mutations detected in relapse in red, mutations detected at both time points additionally in green. For loci poorly covered by WES but well covered in HaloPlex, allele frequencies from HaloPlex sequencing (see Supplemental Table 5) are reported and marked with an asterix*.

Supplemental Table 2: Copy number variants detected by MLPA

Copy numbers detected with the MRC Holland P383-A1 probemix are reported. Because of contamination of some samples with nonleukemic cells, zygosity cannot be safely determined in all samples (see footnotes below table).

Supplemental Table 3: MRD markers of BFM patients

For all patients that have been analysed for MRD markers (rearrangements of T-cell receptor and immunoglobulin genes), the markers that have been identified and the corresponding junctional regions in primary leukemia and in relapse are reported.

Supplemental Table 4: HaloPlex design

Loci that were identified as mutated by WES were included in the HaloPlex design. As WES results became available sequentially, two slightly differing designs were used (design 1 and 2).

Supplemental Table 5: HaloPlex results

For each locus included in the HaloPlex design, coverage, nucleotide distribution and allele frequencies are given. For HaloPlex design 1, a dilution curve of DNAs from relapse of 11 patients was included. This dilution curve was the basis for Supplemental Figure 2 C and D.

Supplemental Table 6: Mutations detected in remission

All mutations that were detected in a remission sample are reported. A mutation was considered to be present if its allele frequency (AF) was higher in the sample of interest than the average of unrelated samples plus three standard deviations (SD).

Supplemental Table 7: Mutations undergoing selection

All mutations that were detected in a primary disease sample with a minor (<0.05) allele frequency (AF) are reported. A mutation was considered to be present if its allele frequency was higher in the sample of interest than the average of unrelated samples plus three standard deviations (SD).

Supplemental Table 8: Gene lists analyzed by Ingenuity Pathway Analysis

The lists of genes subjected to Ingenuity Pathway Analysis, which was the basis for Table 4 and Figure 3: Column A, all genes mutated both in primary leukemia and relapse of the same patient; column B, all genes mutated only in relapse in at least one patient; column C, all genes with promoters that are hypomethylated in at least 3 patients.