SUPPLEMENTARY INFORMATION (SI)

KRAS and *CREBBP* mutations: a relapse-linked malicious liaison in childhood high hyperdiploid acute lymphoblastic leukemia

Running title: Concurrent *KRAS* and *CREBBP* mutations in leukemia

Kamilla Malinowska-Ozdowy,[1*] Christian Frech,[1*] Andreas Schönegger,[2*] Cornelia Eckert,[3] Giovanni Cazzaniga,[4] Martin Stanulla,[5] Udo zur Stadt,[6] Astrid Mecklenbräuker,[1] Michael Schuster,[2] Doris Kneidinger,[1] Arend von Stackelberg,[3] Franco Locatelli,[7] Martin Schrappe,[8] Martin A. Horstmann,[6] Andishe Attarbaschi,[9] Christoph Bock,[2†] Georg Mann,[9†] Oskar A. Haas,[1,9†] and Renate Panzer-Grümayer[1†§]

[1]Children's Cancer Research Institute (CCRI), Vienna, Austria; [2]Research Center for Molecular Medicine (CeMM), Vienna, Austria; [3]Department of Pediatric Oncology/Hematology, Charité Universitätsmedizin Berlin, Campus Virchow Klinikum, Berlin, Germany; [4]Centro Ricerca Tettamanti, University of Milano-Bicocca, Ospedale San Gerardo, Monza, Italy; [5]Department of Pediatric Hematology and Oncology, Hannover Medical School, Hannover, Germany; [6]Clinic of Pediatric Hematology and Oncology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany; [7]Department of Pediatric Hematology-Oncology, IRCCS Bambino Gesù Hospital, Rome, Italy; [8]Department of Pediatrics, Medical University of Schleswig Holstein, Kiel, Germany; [9]St. Anna Kinderspital, Medical University Vienna, Vienna, Austria

*These authors contributed equally to this work.

†These authors share senior authorship.

[§] Corresponding author:
Renate Panzer-Grümayer, Children's Cancer Research Institute, Leukemia Biology Group, Zimmermannplatz 10, A-1090 Vienna, Austria, Email: renate.panzer@ccri.at, Phone: +43-1-40470/4030, Fax: +43-1-40470/64030

This study was supported in part by research funding from The Jubiläumsfonds of the Austrian National Bank, OENB 15480 to G.M., the Austrian Science Fund, FWF I-1226-B19 to R.P-G., the FP7-ERA-NET grant TRANSCALL to M.Stanulla, and by a charitable donation of the Kapsch group (http://www.kapsch.net/kapschgroup) to R.P-G.

The authors declare no conflict of interest.

1. **Supplementary Methods**
2. **Supplementary Results**
3. **Supplementary Tables 1-10**
4. **Supplementary Figures 1-8**
5. **Supplementary Information References**

1. Supplementary Methods

Sample preparation and high-throughput sequencing

WES: Briefly, one µg of genomic double stranded DNA of matched samples from 19 cases (UPN 715, A, 545, Y, 842, C, 399, B, 430, 460, 818, 1021247, 786, X, 592, 314, 446, 792, and D) was sheared with Covaris S210 to an average size of 270bp, end-repaired, A-tailed, ligated to barcoded adaptors, amplified by PCR, and quantified using Illumina TruSeq DNA Sample Prep and TruSeq Exome Enrichment kits (Illumina) according to the manufacturer´s recommendations. Six samples were then pooled into one library and hybridized to Capture Target Oligos, and PCR amplified. 100bp paired-end sequencing was performed on three lanes on a Illumina HiSeq 2000 (Illumina) with ≈80x coverage.

Targeted sequencing: In brief, 50 ng of genomic double stranded DNA was enzymatically sheared to an average size of 200bp. Further processing was performed using Illumina Nextera Rapid Capture Custom Kit (Illumina) and 100bp paired-end sequencing was performed on 24 samples per lane on a Illumina HiSeq 2000 (Illumina) to reach a coverage of 100-1 000x.

Read alignment and variant calling in relapsing cases

Read alignment and variant calling were performed following best practices established at the Broad Institute. Specifically, data demultiplexing and initial quality control were performed using Picard (http://picard.sourceforge.net). After demultiplexing, high-quality (pass-filter, PF) reads were aligned to human reference genome hg19/NCBI37 using BWA version 0.5.9.[1] Prior to somatic mutation calling, aligned BAM files were processed using the GATK 2.4-9 software[2, 3] to mark likely PCR duplicates, to realign reads around indels and to recalibrate base quality scores based on common variation. Somatic point mutations were called using MuTect v1.1.4[4] with COSMIC v54 as reference database. Somatic indel calling was performed using IndelGenotyper2 (http://www.broadinstitute.org/cancer/cga/indelocator). For

targeted sequencing, mutation calling was restricted to targeted regions (= coding exons plus splice sites) to reduce the number of spurious detections in non-targeted regions due to insufficient sequencing depth. Alignments in recurrently mutated genes were also inspected manually, which led to the identification of 13 additional somatic mutations falsely rejected by MuTect (six in *CREBBP*, three in *NRAS*, two in *KRAS*, one in *PTPN11*, and one in *TP53*).

Variant annotation

Variant annotation was performed using SnpEff version 3.3h.[5] PolyPhen2,[6] SIFT,[7] SiPhy[8, 9] and G1K variant annotations were obtained by SnpSift[10] using dbNSFP v2.1[11] as annotation source.

We considered *non-silent* variants as variants with predicted SnpEff-effects STOP_GAINED, STOP_LOST, SPLICE_SITE_DONOR, SPLICE_SITE_ACCEPTOR, FRAME_SHIFT, CODON_CHANGE_PLUS_CODON_INSERTION, CODON_DELETION, NON_SYNONYMOUS_CODING, CODON_INSERTION, CODON_CHANGE_PLUS_CODON_DELETION, NON_SYNONYMOUS_START, and START_LOST. *Deleterious* variants included variants with predicted SnpEff effects FRAME_SHIFT, SPLICE_SITE_ACCEPTOR, SPLICE_SITE_DONOR, and STOP_GAINED as well as variants with SnpEff effect NON_SYNONYMOUS_CODING if this non-synonymous variant was predicted to be deleterious by at least two of three methods, including PolyPhen2, SIFT, and SiPhy. PolyPhen2 directly outputs deleteriousness, whereas for SIFT and SiPhy we considered variants deleterious if they had a score <0.05 and ≥12, respectively. In addition, we considered mutations impacting extremely well conserved nucleotides (SiPhy score >20) as deleterious, regardless of PolyPhen2 and SIFT predictions.

Post-calling variant filtering

Following variant calling and annotation, somatic point mutations and indels were further filtered to reduce false-positives. Removed variants included (a) variants overlapping with repetitive regions, segmental duplications, or blacklisted regions (as annotated in UCSC/hg19 tables “rmsk”, “simpleRepeat”, “genomicSuperDups”, and “wgEncodeDacMapabilityConsensusExcludable”); (b) variants corresponding to common SNPs without known medical impact (http://www.ncbi.nlm.nih.gov/variation/docs/human_variation_vcf/#common_no_known; version Sept 30, 2013); (c) variants detected in phase 1 of the 1000 Genomes Project;[12] (d) variants present in remission samples of at least two exome sequencing cases, if in remission these variants had (1) an allelic frequency >5% to not erroneously remove true somatic variants that are detected in remission due to minimal residual disease and (2) were supported by at least 3 reads to not erroneously exclude true somatic variants because of

sequencing errors in remission; and (e) exome sequencing variants with allelic frequencies <10%. This last filter was only applied to exome sequencing variants and not to targeted sequencing variants.

Identified indels were subjected to additional filtering, because manual inspection revealed many questionable calls in the raw output of IndelGenotyper2. Called indels were rejected if (a) the total read depth in tumor was <10x (= low coverage filter); (b) the fraction of reads supporting the consensus indel among all reads supporting any indel was below 70% (= low consensus filter); (c) the indel was not supported by reads from both strands (= strand bias filter); (d) indel supporting reads had a higher fraction of mismatching bases than reads supporting the reference allele (= misalignment filter); or (e) the average mapping quality of consensus indel-supporting reads was below 40 (= mapping quality filter).

Post-calling filtered variants were removed from all downstream analyses.

Variant calling in non-relapsing cases

For non-relapsing cases, only diagnostic samples without matched remission were sequenced, which necessitated a different variant calling strategy. Reads were first aligned with BWA and processed with GATK as described previously. SAMtools (version 0.1.19)[13] was then used to pileup reads ('*samtools mpileup -q 40*'), followed by VarScan v2.3.6[14] to call non-reference variants ('*VarScan mpileup2cns --variants --strand-filter 1 --min-coverage 10 --min-avg-qual 15 --P-value 1 --min-var-freq 0.1 --min-reads2 4 --output-vcf 1*'). Subsequent variant annotation and post-calling filtering were performed as described previously for MuTect-called variants, which removed all known common germline polymorphisms. Only variants impacting *CREBBP* and RTK/Ras pathway genes (*KRAS, NRAS, PTPN11,* and *FLT3*) were kept for further analysis (Supplemental Table S8).

Pathway analysis

Pathway analysis was performed with Genome MuSiC version 0.4[15] and its integrated PathScan module.[16] Input gene sets were compiled from various sources, including 199 gene sets from KEGG,[17] 10,295 gene sets from MSigDB version 4.0,[18] 224 gene sets from the National Cancer Institute (NCI) Pathway Interaction Database[19], 240 gene sets from WikiPathways,[20] 314 gene sets from BioCarta (http://www.biocarta.com/),[21] 119 gene sets from BBID (http://bbid.grc.nia.nih.gov/), 10,440 Gene Ontology annotations,[22] 7,309 InterPro domain assignments,[23] 4,670 OMIM disease associations (http://www.ncbi.nlm.nih.gov/omim), 3,554 PIR superfamilies,[24] 604 SMART motifs,[25, 26] and 70 COG ontology assignments.[27] KEGG, BioCarta, BBID, GO, InterPro, OMIM, PIR, SMART, and COG gene sets were obtained through the DAVID Bioinformatics Resources, version

6.7.[28] To enter the analysis, gene sets were required to have at least two and a maximum of 400 genes.

PathScan input mutations were provided in MAF file format containing only predicted deleterious variants with allelic frequency ≥20%. Enrichment analysis was performed independently for diagnosis and relapse samples. In addition, for each leukemia occasion, two separate analyses were conducted: one where all mutated genes were considered, and one where the highly recurrently mutated genes *CREBBP* and RTK/Ras pathway genes (*KRAS*, *NRAS* and *PTPN11*) were excluded. Mutations in the following (mostly large cytoskeletal and olfactory) genes were excluded from all pathway analyses: *MESP2*, *TTN*, *TBP*, *CSMD3*, *DNAH5*, *RYR1*, *RYR2*, *RYR3*, *DNAH1*, *DNAH8*, *DNAH9*, *MUC2*, *MUC16*, *MUC12*, *MUC5B*, *OR5H2*, *OR5H2*, *OR11H4*, *OR6F1*, *OR52R1*, *OR2T12*, *OR6V1*, *OR51I2*, *OR5I1*, *OR9Q1*, *OR9Q1*, *PDZD7*. These genes have elevated local background mutation rates that cause them to be found as recurrently mutated in many large-scale mutational screens, although they are most likely not functionally relevant in cancer development or progression.[29]

Enriched gene sets, identified at each leukemia occasion (*i.e.* at diagnosis and relapse) and by each analysis (*i.e.* including and excluding RTK/RAS pathway/*CREBBP* genes), are provided in Supplemental Table S3. The 144 gene sets, significantly enriched at relapse and excluding RTK/RAS pathway/*CREBBP* genes, were clustered to produce Supplemental Figure S4, using the following procedure. First, a binary gene-pathway matrix was constructed, with rows representing mutated genes and columns representing significantly enriched pathways (12 gene sets from MSigDB belonging to categories ‘c1_positional’ and ‘c3_motif’ were excluded in this step). Presence and absence of a gene in a pathway was encoded with 1 and 0, respectively. The binary matrix was then clustered independently in both dimensions with R functions ‘dist’ (distance measure ‘binary’) and ‘hclust’ (agglomeration method ‘average’), resulting in a matrix in which pathways sharing the same genes and genes sharing the same pathways are grouped together. The clustered matrix was color-coded (1=black, 0=white) and plotted alongside hierarchical trees for visual interpretation.

Functional groups identified in Supplemental Figure S4 were then used as scaffold to manually construct an extended gene-case mutation matrix containing both recurrently and non-recurrently mutated genes (Supplemental Figure S5). Recurrently mutated genes were put at the top and non-recurrently genes were put into their corresponding functional groups as identified by pathway analysis. The matrix was manually augmented by genes carrying missense mutations not predicted to be deleterious, which were assigned to functional groups based on manual review of GeneCards annotations (http://www.genecards.org).[30]

Identification of hot-spot mutations in RTK/Ras pathway genes

Mutational hotspots in *KRAS*, *NRAS*, and *PTPN11* genes were identified by first searching the COSMIC database for genomic positions impacted most frequently by somatic mutations. This search identified seven clear mutational hotspots for both *KRAS* and *NRAS* and 17 for *PTPN11.* Using a custom R script, BAM files from targeted sequencing were then inspected at these 31 hotspot positions to find reads supporting non-reference alleles. A mutation was called if (a) it was supported by at least two reads from both strands and (b) the mean base quality of the non-reference allele was at least 22 on both strands. Because *FLT3* mutations were much less clustered than *KRAS*, *NRAS*, and *PTPN11* mutations, we augmented *KRAS, NRAS,* and *PTPN11* hotspot mutations with *FLT3* mutations previously identified by MuTect. All RTK/Ras pathway mutations identified by this procedure are provided in Supplemental Table S7. In a control experiment, we searched 82 remission samples for non-reference Ras pathway alleles using the exact same procedure, but did not call any mutation. This suggests that the false-positive rate of the mutation hotspot caller is very low, at least below one error per 2 500 investigated hotspots.

2. Supplementary Results

Transition versus transversion ratio

The average transition/transversion (Ti/Tv) ratio of mutations at diagnosis and relapse was similar (1.6 and 1.4; P=0.14) and thus not indicative of a genotoxic drug effect or mutator phenotype.

Frequency and description of mutated genes from targeted sequencing

Recurrent targets of sequence mutations were RTK/Ras pathway genes (63%), *CREBBP* (24%), *TP53* (5%), *ATM* (4%), *TRAPP* (3%), *MLL2* (3%), *WHSC1* (3%), *ZNF516* (3%), *JAK2* (2%), *PI3KCB* (2%), *TLX3* (2%), *FBXL7* (2%), *GDPD2* (2%), *MAGI1* (2%), *NDC80* (2%), and *USP9X* (2%).

The chromatin modifier group comprises genes mediating acetylation (*CREBBP*, *ATM*, and *TRRAP*) and methylation (*WHSC1*, *MLL2*). While all of them, except *TRRAP*, are known players in the pathogenesis of various leukemias or have previously been associated with childhood ALL,[31-36] *TRRAP* mutations have only recently been described in colon cancer and lymphoma.[37] All three cases carried the same novel *TRRAP* mutation (E3107K) affecting the tetratricopeptide repeat (TPR) domain that is responsible for ligand binding. Considering the similar function of TRRAP and CREBBP, they both acetylate histone and non-histone

proteins, and their mutually exclusive presence of *TRRAP* mutations thus might substitute for *CREBBP* mutations in these cases at relapse.

The signaling cohort comprised not only mutations in genes of the RTK/Ras pathway, but also in *JAK2*, *PI3KCB*, *USP9X*, and *MAGI1*. JAK2 is a non-receptor tyrosine kinase involved in regulation of cell growth, development, and differentiation. One case harbored a gain of function mutation (R683) at relapse that affects the pseudokinase domain and was previously identified in 10% of pediatric cases with high risk ALL and in about 20% of Down Syndrome-associated ALL.[38-40] The second case had a conserved mutation (D1128G), annotated in COSMIC, that targets the tyrosine kinase domain. The other mutated signaling genes were found in solid and hematopoietic malignancies and were, in part, associated with resistance to treatment.[41-47]

The third category, termed "others", contains genes that are associated with oncogenesis such as *TP53* and *TLX3,*[48] and potential novel players *FBXL7,*[49] *ZNF516*, *GDPD2*, and *NDC80*. *TP53* mutations are particularly frequent in childhood low-hypodiploid ALL[50] and, less often though, in relapsed ALL and seem to confer poor prognosis.[51, 52] We detected two *TP53* mutations already at diagnosis and three additional mutations at relapse. Four of these five *TP53* mutations target the DNA binding domain including one reported before in hypodiploid ALL.[52]

Altered pathways at diagnosis and relapse

Pathway analyses revealed enrichment of deleterious mutations in genes implicated in transcription, cell signaling, cytoskeleton, cell cycle, development or differentiation (P<1e-6) (Supplemental Table S3 and Figure S4). Together with pathways responsible for drug resistance and hypoxia regulation, they were significantly more common at relapse. When we performed the analyses without *CREBBP* and RTK/Ras, the central hubs in many of these pathways,[53, 54] the most prominent of the remaining ones at diagnosis was MSigDB gene set prostaglandin E2 DN (P=2.4e-8), while at relapse all other pathways remained statistically significant. These findings may suggest that other relapse-specific mutations in these pathways either substitute for mutated RTK/Ras and/or *CREBBP* genes, or when they co-occur, even synergize in their function and progression towards relapse (Supplemental Figure S5).

3. Supplementary Tables

Supplementary Table S1. Clinical characteristics of relapsing and non-relapsing HD ALL cases at diagnosis

	Diagnosis		
	Relapsing cases	**Non-relapsing cases***	***P***
Total	**n=66 (100%)**	**n=51 (100%)**	
Gender (%)			0.36
Male	37 (56.1)	24 (47.1)	
Female	29 (43.9)	27 (52.9)	
Age, y (%)			**0.04**
1-10	55 (83.3)	49 (96)	
≥10	11 (16.7)	2 (4)	
WBC count (%)			0.76
< 50 000/µL	60 (91)	45 (88.2)	
≥ 50 000/µL	6 (9)	6 (11.8)	
Mod. no. chrom. (%)			**0.001**
51-53	8 (14.5)	3 (6.1)	
54-57	22 (40)	37 (75.5)	
58-66	25 (45.5)	9 (18.4)	
NA	11	2	
Pred response (%)			0.1
Poor	1 (1.9)	5 (9.8)	
Good	53 (98.1)	46 (90.2)	
Na	11		
MRD risk (%)			0.17
Standard	10 (16.4)	11 (33.3)	
Intermediate /SER	33/16 (54.1/26.2)	19/2 (57.6 /6.1)	
High	2 (3.3)	1 (3)	
NA	5	18	
Time to relapse (%)			
Very early	5 (7.7)		
Early	18 (27.7)		
Late	42 (64.6)		
NA	1		

Mod. no. chrom., modal number of chromosomes; NA, not available; na, not applicable; SER, slow early responder; *, median follow up 103.6 months (range, 67.2-251.6 months). Associations between categorical variables were examined using Fisher's exact test.

Thus, relapsing HD ALL cases tended to be older at initial diagnosis and had a higher modal number of chromosomes than non-relapsing cases. The latter finding contrasts data from a recent EORTC study suggesting that children with a higher modal number of chromosomes have a superior outcome.[55]

Supplementary Table S2. Somatic variants identified by whole exome sequencing in 19 relapsing HD ALL cases.

Supplementary Table S3. Enriched pathways identified by pathway analysis. The Excel file contains two sheets, one showing the results considering all recurrently mutated genes and the second one the results after excluding RTK/RAS pathway and *CREBBP* genes.

Supplementary Table S4. Genes and their genomic regions selected for targeted sequencing. The panel comprises genes identified by WES as recurrently mutated in 19 HD ALL relapsing cases, (non-recurrently) mutated genes also annotated in the COSMIC database, and genes manually curated from the literature.

Supplementary Table S5. Targeted sequencing variants identified in the entire relapsing cohort. List of somatic mutations identified by MuTect in the 60 targeted genes of the relapsing cohort.

Supplementary Table S6. Mutation clusters from case UPN 715. List contains mutated genes in various clusters.

Supplementary Table S7. RTK/Ras pathway mutations in relapsing cases. This list combines mutations found by the RTK/Ras pathway hotspot mutation caller and FLT3 mutations identified by MuTect.

Supplementary Table S8. RTK/Ras pathway and *CREBBP* variants identified by targeted sequencing in the non-relapsing cohort.

Supplementary Table S9. MRD risk and outcome of HD ALL cases after first relapse

Relapsed Cases (n=84)*	
MRD risk	n (%)
Standard	33 (53.2)
High	29 (46.8)
NA	22
Outcome (%)	
2nd CCR	56 (70)
Death after treatment	5 (6.25)
TRM	7 (8.75)
Subsequent relapse	12 (15)
NA	4

*, median follow up 113.9 months (range, 38.5-192.5).
NA, not available; CCR, continuous complete remission;
TRM, treatment related mortality.

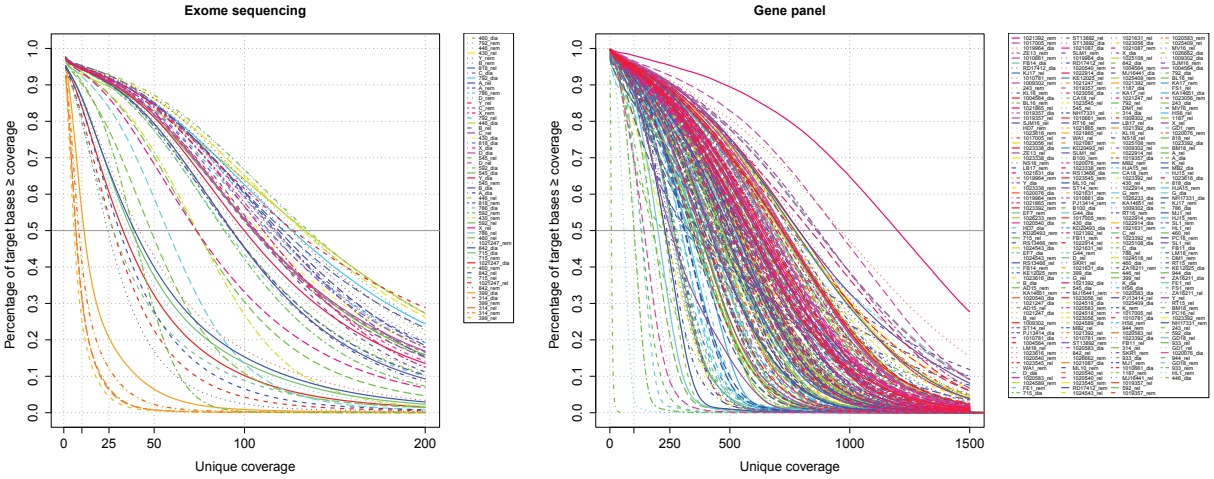
Supplementary Table S10. Relation of relapse-specific *CREBBP* and RTK/Ras pathway mutations with clinical features at first relapse

Mutational status of genes	Characteristics			
	Time of relapse*		MRD risk	
	Very early/ early	Late	Standard	High
No. of cases (%)	22 (27)	61 (73)	33 (53)	29 (47)
CREBBP				
M	9 (41)	15 (24.6)	7 (21)	12 (41)
Wt	13 (59)	46 (75.4)	26 (79)	17 (59)
P	0.18		0.1	
RTK/Ras pw				
m	17 (77)	37 (61)	18 (54.5)	20 (69)
wt	5 (23)	24 (39)	15 (45.5)	9 (31)
P	0.2		0.3	
KRAS				
m	10 (45.5)	9 (15)	3 (9)	10 (34.5)
wt	12 (54.5)	52 (85)	30 (91)	19 (65.5)
P	**0.0065**		**0.026**	
NRAS				
m	5 (23)	14 (23)	5 (15)	8 (27.6)
wt	17 (77)	47 (77)	28 (85)	21 (72.4)
P	1		0,35	
PTPN11				
m	1 (4.5)	12 (20)	8 (24)	2 (7)
wt	21 (95.5)	49 (80)	25 (76)	27 (93)
P	0.17		0.088	
FLT3				
m	1 (4.5)	4 (6.6)	3 (9)	1 (3.4)
wt	21 (95.5)	57 (93.4)	30 (91)	28 (96.6)
P	1		0,62	
***CREBBP*/*KRAS* double mutant**				
pos	7 (31.8)	5 (8.2)	3 (9.1)	5 (17.2)
neg	15 (68.2)	56 (91.8)	30 (90.9)	24 (82.8)
P	**0.012**		0.45	

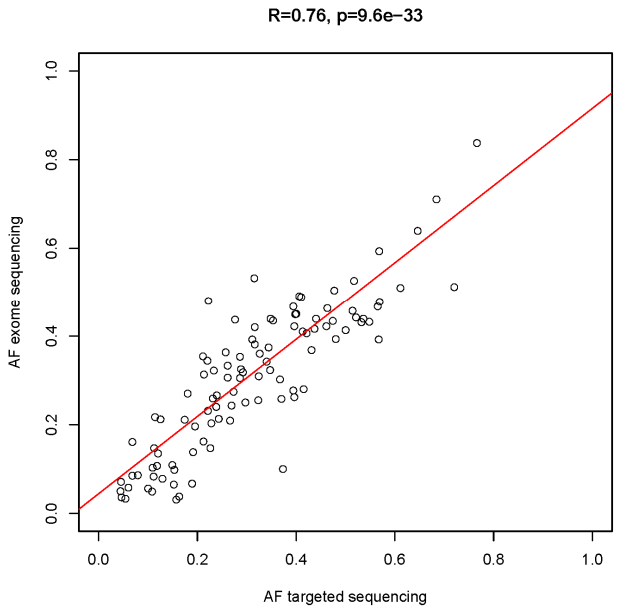
*, based on first remission duration: very early, ≤18 months; early, between 19-30 month; late, >30 months; m, mutated; wt, wild type; pw, pathway; pos, positive; neg, negative.

4. Supplementary Figures

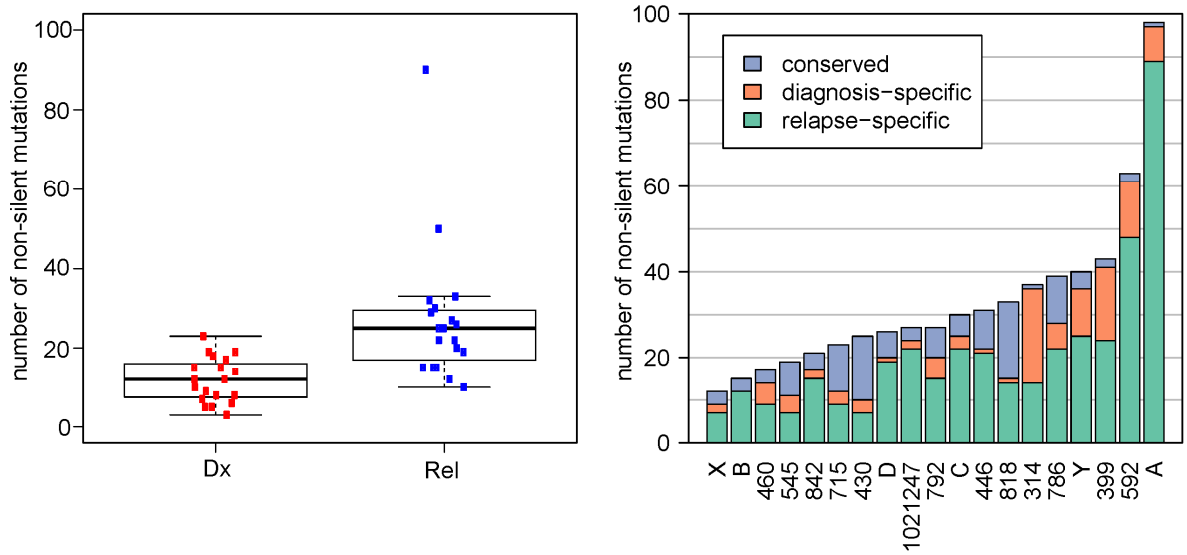
Supplementary Figure S1. Unique coverage achieved by whole exome sequencing (left, n=57) and targeted sequencing (right, n=285). For each sample, the plot shows the percentage of target bases (y-axis) covered by a certain minimum number of reads (x-axis), not counting reads with mapping quality 0 and duplicate reads. Samples in the legend are sorted column-wise by their y-axis value at coverage 50 (whole exome sequencing) and 250 (targeted sequencing).



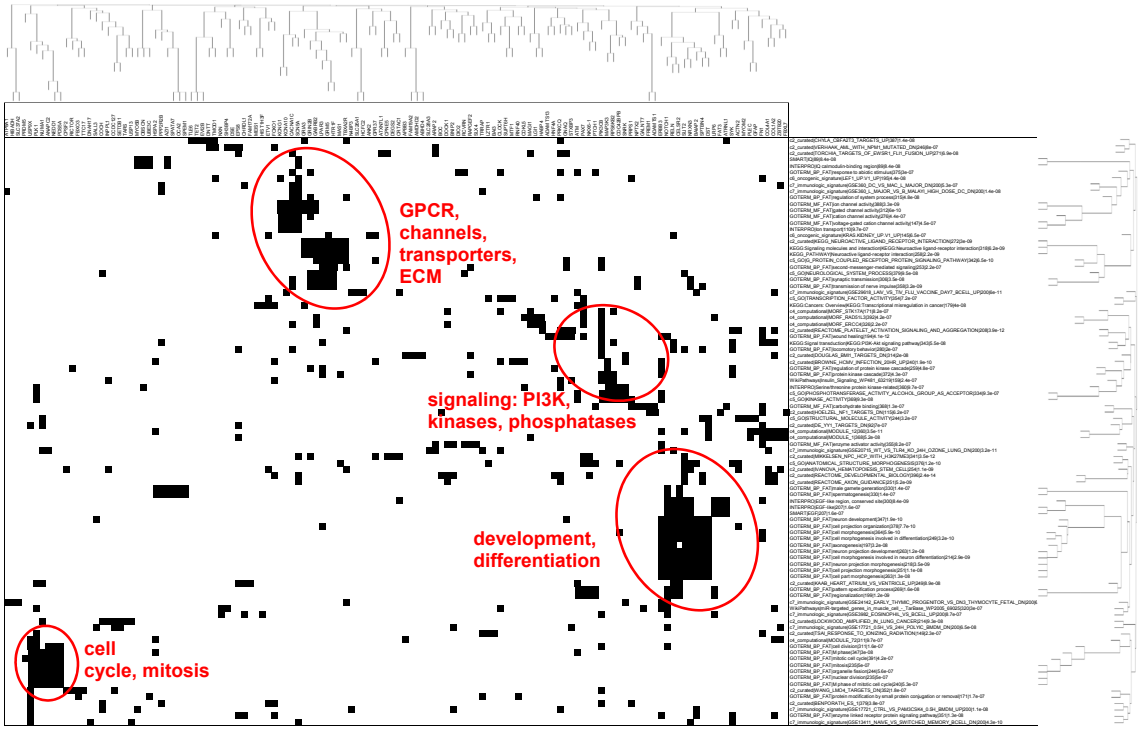
Supplemental Figure S2. Correlation of allelic frequency (AF) estimates from exome sequencing *vs.* targeted sequencing. Allelic frequency estimates obtained by exome and targeted sequencing are highly correlated (R=0.76, *P*=9.6e-33), suggesting that allelic frequencies obtained by both sequencing approaches provide a good estimate of true allelic frequencies. Plotted are AFs of mutations detected by both exome and targeted sequencing with a minimum total coverage of 10 reads.



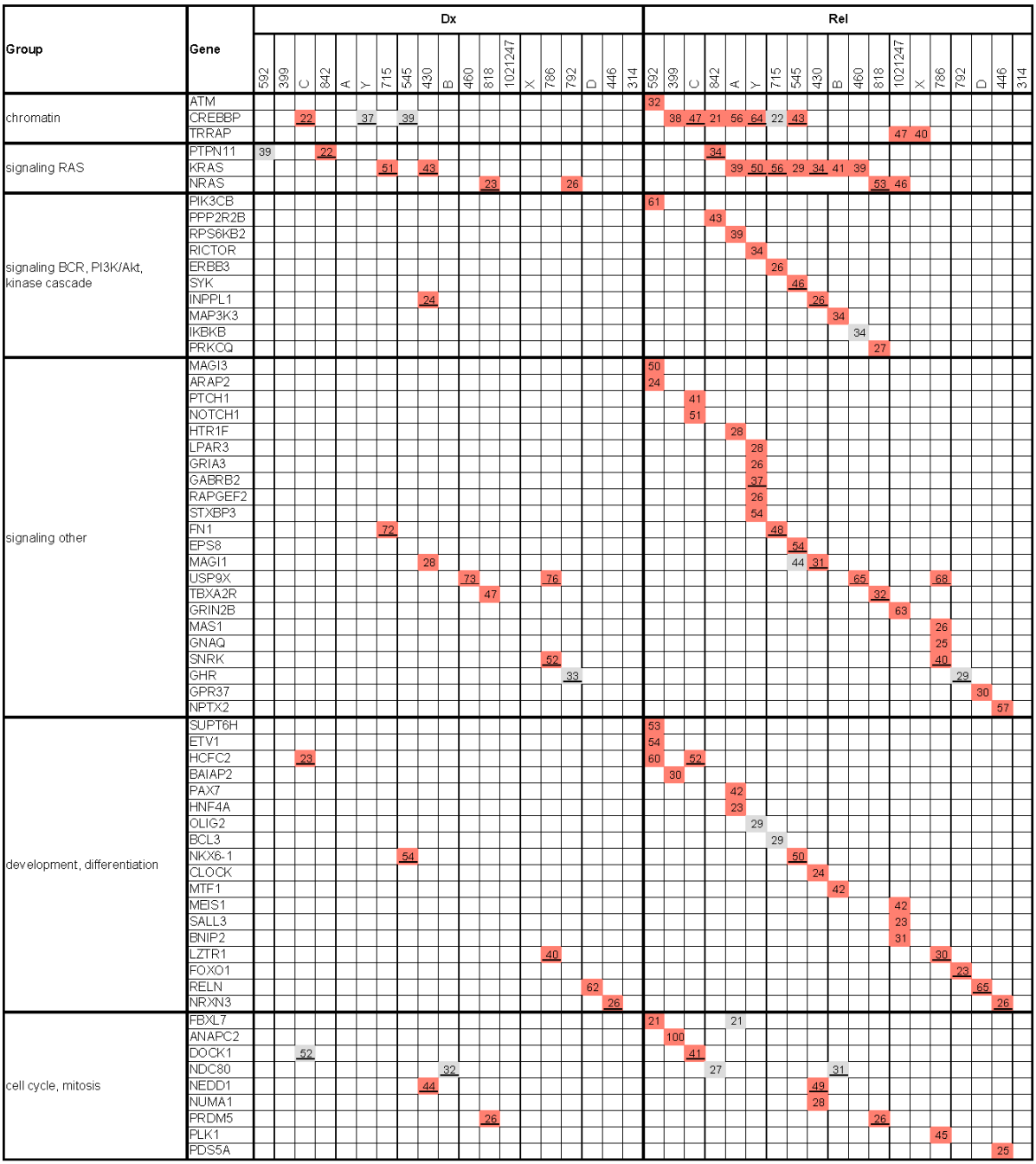
Supplemental Figure S3. Number of non-silent somatic mutations identified in the discovery cohort (n=19) by exome sequencing. Left: Overall number of non-silent somatic mutations identified per case. The median number of mutations is 12 at diagnosis and 25 at relapse (*P*=5.4e-05, Kruskal-Wallis test). Right: Number of non-silent somatic mutations per case stratified by conserved, diagnosis-specific, and relapse-specific mutations.



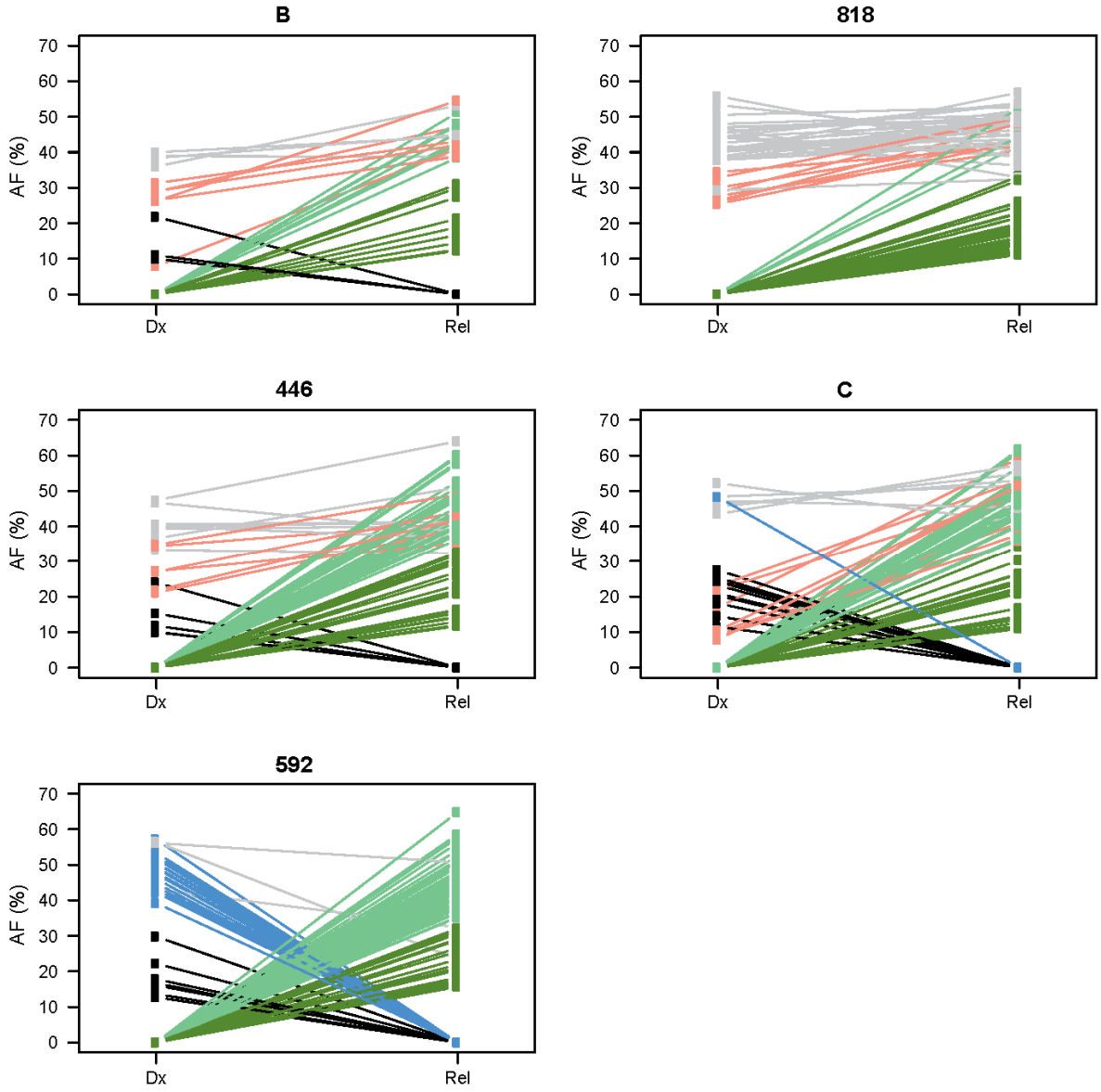
Supplementary Figure S4. Deleterious mutations are enriched in signaling genes, genes involved in development and differentiation, cell cycle genes, and GPCR/chanel/ transporter/ECM genes. The matrix shows significantly (*P*<1e-6) enriched pathways at the right and genes impacted by deleterious mutations at the top. Individual black dots indicate which genes are assigned to which pathways. Genes and pathways were hierarchically clustered as described in the method section, resulting in clusters of black dots ("blobs") representing groups of genes and pathways with similar biological functions. The pathway identifier to the right consists of multiple entries separated by '|', including the source database, the pathway name, the size of the pathway in genes, and the enrichment *P*-value.



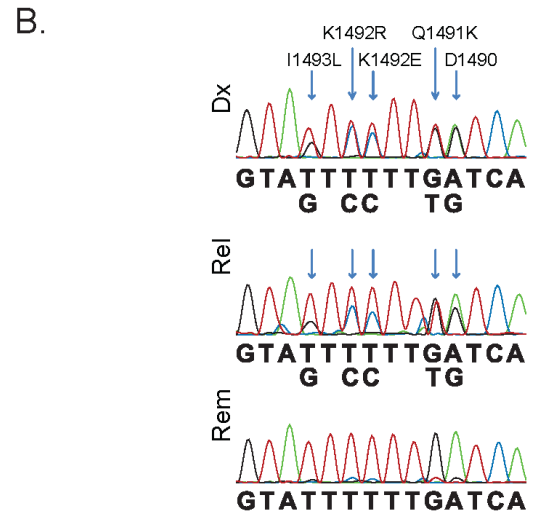
Supplementary Figure S5. Extended mutation matrix with both recurrently and non-recurrently mutated genes identified by WES. The matrix shows selected mutated genes (rows) from all cases in the discovery cohort at diagnosis and relapse (columns). A mutated gene was included if it was recurrently mutated or if it belonged to a significantly enriched functional cluster as identified by pathway analysis. Numbers represent allelic frequencies in percent. Red squares indicate predicted deleterious mutations, grey squares non-silent predicted benign mutations, and underscores indicate mutations conserved between diagnosis and relapse. Genes are ordered by their mutation status at relapse.



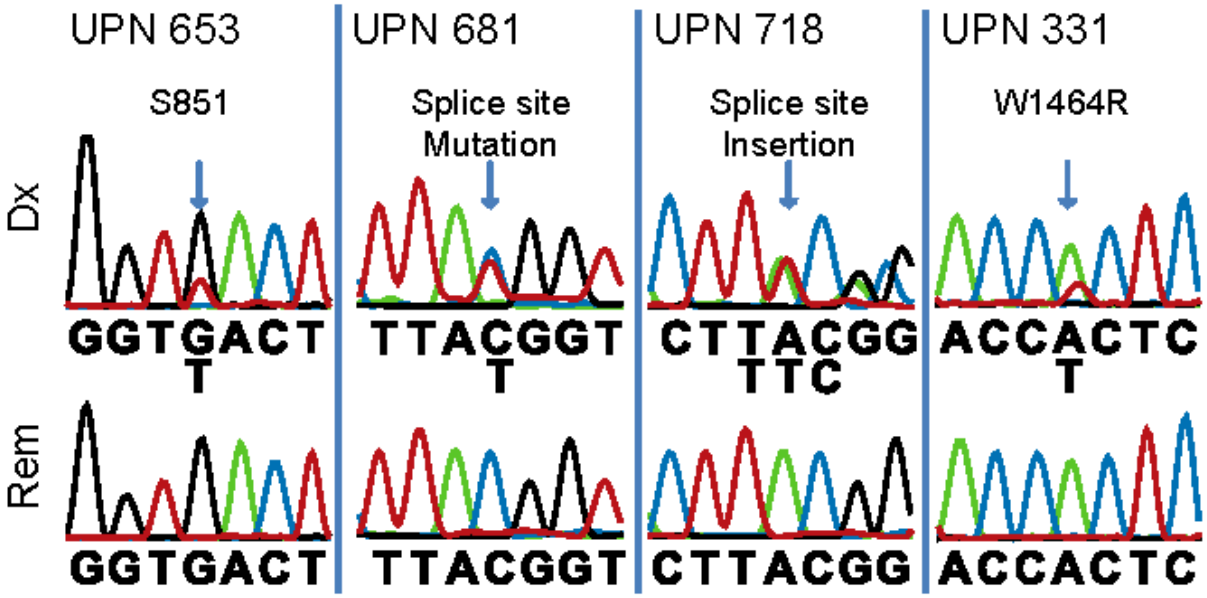
Supplementary Figure S6. Patterns of mutational kinetics in relapse evolution. Three cases (UPN B, 818 and 446) follow the clonal progression pattern and in two cases (UPN C and 592) the relapse is derived from a pre-diagnostic ancestral clone. Plotted are AFs of non-silent mutations with AF ≥10% at either diagnosis or relapse. AFs were adjusted to the copy number of the respective gene obtained from SNP array analysis. Diagnosis and relapse samples contained ≥90% blasts.



Supplementary Figure S7. Multiple somatically acquired *CREBBP* mutations in case UPN KD20493. (A) Selected reads aligning to exon 27 of *CREBBP* in diagnostic (Dx), relapse (Rel), and remission (Rem) samples (top to bottom), indicating five clustered somatic mutations at diagnosis and relapse. For each time point, the upper track shows summarized coverage information and the lower track selected aligned reads from both strands in different colors. The five mutations are present on ~35% of reads and always co-occur within the same read, providing evidence they impact a single allele. Image generated with Golden Helix GenomeBrowse. (B) Sanger sequence traces confirming the five mutations (indicated by an arrow) in diagnostic (top) and relapse (middle) samples compared to matched remission sample (bottom). Four of these mutations have a predicted deleterious outcome.



Supplementary Figure S8. Verification of somatic *CREBBP* mutations in leukemic samples of non-relapsing HD ALL cases. Sanger sequencing was performed to confirm the somatic nature of *CREBBP* mutations with a predicted deleterious outcome. UPN, position of the respective mutation in the diagnostic sample (Dx; top) and their absence in the remission sample (Rem; bottom) are shown.



5. Supplementary Information References

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

2. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C*, et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 2011**; 43:** 491-498.

3. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A*, et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research* 2010**; 20:** 1297-1303.

4. Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C*, et al.* Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol* 2013**; 31:** 213-219.

5. Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang 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* 2012**; 6:** 80-92.

6. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P*, et al.* A method and server for predicting damaging missense mutations. *Nat Methods* 2010**; 7:** 248-249.

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:** 1073-1081.

8. Garber M, Guttman M, Clamp M, Zody MC, Friedman N, Xie X. Identifying novel constrained elements by exploiting biased substitution patterns. *Bioinformatics* 2009**; 25:** i54-62.

9. Lindblad-Toh K, Garber M, Zuk O, Lin MF, Parker BJ, Washietl S*, et al.* A high-resolution map of human evolutionary constraint using 29 mammals. *Nature* 2011**; 478:** 476-482.

10. Cingolani P, Patel VM, Coon M, Nguyen T, Land SJ, Ruden DM*, et al.* Using Drosophila melanogaster as a Model for Genotoxic Chemical Mutational Studies with a New Program, SnpSift. *Front Genet* 2012**; 3:** 35.

11. Liu X, Jian X, Boerwinkle E. dbNSFP v2.0: a database of human non-synonymous SNVs and their functional predictions and annotations. *Human Mutation* 2013**; 34:** E2393-2402.

12. Genomes Project C, Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM*, et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* 2012**; 491:** 56-65.

13. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N*, et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009**; 25:** 2078-2079.

14. Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L*, et al.* VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Research* 2012**; 22:** 568-576.

15. Dees ND, Zhang Q, Kandoth C, Wendl MC, Schierding W, Koboldt DC*, et al.* MuSiC: identifying mutational significance in cancer genomes. *Genome Research* 2012**; 22:** 1589-1598.

16. Wendl MC, Wallis JW, Lin L, Kandoth C, Mardis ER, Wilson RK*, et al.* PathScan: a tool for discerning mutational significance in groups of putative cancer genes. *Bioinformatics* 2011**; 27:** 1595-1602.

17. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Research* 2000**; 28:** 27-30.

18. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA*, et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA* 2005**; 102:** 15545-15550.

19. Schaefer CF, Anthony K, Krupa S, Buchoff J, Day M, Hannay T*, et al.* PID: the Pathway Interaction Database. *Nucleic Acids Research* 2009**; 37:** D674-679.

20. Pico AR, Kelder T, van Iersel MP, Hanspers K, Conklin BR, Evelo C. WikiPathways: pathway editing for the people. *PLoS Biology* 2008**; 6:** e184.

21. Nishimura D. BioCarta. *Biotech Software \& Internet Report: The Computer Software Journal for Scient* 2001**; 2:** 117-120.

22. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM*, et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature Genetics* 2000**; 25:** 25-29.

23. Hunter S, Jones P, Mitchell A, Apweiler R, Attwood TK, Bateman A*, et al.* InterPro in 2011: new developments in the family and domain prediction database. *Nucleic Acids Research* 2012**; 40:** D306-312.

24. Wu CH, Nikolskaya A, Huang H, Yeh LS, Natale DA, Vinayaka CR*, et al.* PIRSF: family classification system at the Protein Information Resource. *Nucleic Acids Research* 2004**; 32:** D112-114.

25. Letunic I, Doerks T, Bork P. SMART 7: recent updates to the protein domain annotation resource. *Nucleic Acids Research* 2012**; 40:** D302-305.

26. Schultz J, Copley RR, Doerks T, Ponting CP, Bork P. SMART: a web-based tool for the study of genetically mobile domains. *Nucleic Acids Research* 2000**; 28:** 231-234.

27. Tatusov RL, Galperin MY, Natale DA, Koonin EV. The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Research* 2000**; 28:** 33-36.

28. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols* 2009**; 4:** 44-57.

29. Lawrence MS, Stojanov P, Polak P, Kryukov GV, Cibulskis K, Sivachenko A*, et al.* Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* 2013**; 499:** 214-218.

30. Rebhan M, Chalifa-Caspi V, Prilusky J, Lancet D. GeneCards: a novel functional genomics compendium with automated data mining and query reformulation support. *Bioinformatics* 1998**; 14:** 656-664.

31. Guo C, Chang CC, Wortham M, Chen LH, Kernagis DN, Qin X*, et al.* Global identification of MLL2-targeted loci reveals MLL2's role in diverse signaling pathways. *Proc Natl Acad Sci U S A* 2012**; 109:** 17603-17608.

32. Inthal A, Zeitlhofer P, Zeginigg M, Morak M, Grausenburger R, Fronkova E*, et al.* CREBBP HAT domain mutations prevail in relapse cases of high hyperdiploid childhood acute lymphoblastic leukemia. *Leukemia* 2012**; 26:** 1797-1803.

33. Jaffe JD, Wang Y, Chan HM, Zhang J, Huether R, Kryukov GV*, et al.* Global chromatin profiling reveals NSD2 mutations in pediatric acute lymphoblastic leukemia. *Nat Genet* 2013**; 45:** 1386-1391.

34. Mullighan CG, Zhang J, Kasper LH, Lerach S, Payne-Turner D, Phillips LA*, et al.* CREBBP mutations in relapsed acute lymphoblastic leukaemia. *Nature* 2011**; 471:** 235-239.

35. Oyer JA, Huang X, Zheng Y, Shim J, Ezponda T, Carpenter Z*, et al.* Point mutation E1099K in MMSET/NSD2 enhances its methyltranferase activity and leads to altered global chromatin methylation in lymphoid malignancies. *Leukemia* 2014**; 28:** 198-201.

36. Stankovic T, Skowronska A. The role of ATM mutations and 11q deletions in disease progression in chronic lymphocytic leukemia. *Leuk Lymphoma* 2014**; 55:**1227-1239.

37. Mouradov D, Sloggett C, Jorissen RN, Love CG, Li S, Burgess AW*, et al.* Colorectal cancer cell lines are representative models of the main molecular subtypes of primary cancer. *Cancer Res* 2014**; 74:** 3238-3247.

38. Kearney L, Gonzalez De Castro D, Yeung J, Procter J, Horsley SW, Eguchi-Ishimae M*, et al.* Specific JAK2 mutation (JAK2R683) and multiple gene deletions in Down syndrome acute lymphoblastic leukemia. *Blood* 2009**; 113:** 646-648.

39. Roll JD, Reuther GW. CRLF2 and JAK2 in B-progenitor acute lymphoblastic leukemia: a novel association in oncogenesis. *Cancer Research* 2010**; 70:** 7347-7352.

40. Wu QY, Guo HY, Li F, Li ZY, Zeng LY, Xu KL. Disruption of E627 and R683 interaction is responsible for B-cell acute lymphoblastic leukemia caused by JAK2 R683G(S) mutations. *Leuk Lymphoma* 2013**; 54:** 2693-2700.

41. Dbouk HA, Khalil BD, Wu H, Shymanets A, Nurnberg B, Backer JM. Characterization of a tumor-associated activating mutation of the p110beta PI 3-kinase. *PLoS One* 2013**; 8:** e63833.

42. Dupont S, Mamidi A, Cordenonsi M, Montagner M, Zacchigna L, Adorno M*, et al.* FAM/USP9x, a deubiquitinating enzyme essential for TGFbeta signaling, controls Smad4 monoubiquitination. *Cell* 2009**; 136:** 123-135.

43. Junyent F, de Lemos L, Verdaguer E, Folch J, Ferrer I, Ortuno-Sahagun D*, et al.* Gene expression profile in JNK3 null mice: a novel specific activation of the PI3K/AKT pathway. *J Neurochem* 2011**; 117:** 244-252.

44. Kuang SQ, Tong WG, Yang H, Lin W, Lee MK, Fang ZH*, et al.* Genome-wide identification of aberrantly methylated promoter associated CpG islands in acute lymphocytic leukemia. *Leukemia* 2008**; 22:** 1529-1538.

45. Sun H, Kapuria V, Peterson LF, Fang D, Bornmann WG, Bartholomeusz G*, et al.* Bcr-Abl ubiquitination and Usp9x inhibition block kinase signaling and promote CML cell apoptosis. *Blood* 2011**; 117:** 3151-3162.

46. Trivigno D, Essmann F, Huber SM, Rudner J. Deubiquitinase USP9x confers radioresistance through stabilization of Mcl-1. *Neoplasia* 2012**; 14:** 893-904.

47. Zhang G, Liu T, Wang Z. Downregulation of MAGI1 associates with poor prognosis of hepatocellular carcinoma. *J Invest Surg* 2012**; 25:** 93-99.

48. Dadi S, Le Noir S, Payet-Bornet D, Lhermitte L, Zacarias-Cabeza J, Bergeron J*, et al.* TLX homeodomain oncogenes mediate T cell maturation arrest in T-ALL via interaction with ETS1 and suppression of TCRalpha gene expression. *Cancer Cell* 2012**; 21:** 563-576.

49. Coon TA, Glasser JR, Mallampalli RK, Chen BB. Novel E3 ligase component FBXL7 ubiquitinates and degrades Aurora A, causing mitotic arrest. *Cell Cycle* 2012**; 11:** 721-729.

50. Holmfeldt L, Wei L, Diaz-Flores E, Walsh M, Zhang J, Ding L*, et al.* The genomic landscape of hypodiploid acute lymphoblastic leukemia. *Nature Genetics* 2013**; 45:** 242-252.

51. Hof J, Krentz S, van Schewick C, Korner G, Shalapour S, Rhein P*, et al.* Mutations and deletions of the TP53 gene predict nonresponse to treatment and poor outcome in first relapse of childhood acute lymphoblastic leukemia. *Journal of Clinical Oncology* 2011**; 29:** 3185-3193.

52. Stengel A, Schnittger S, Weissmann S, Kuznia S, Kern W, Kohlmann A*, et al.* TP53 mutations occur in 15.7% of ALL and are associated with MYC-rearrangement, low hypodiploidy and a poor prognosis. *Blood* 2014**; 124:** 251-258.

53. Schubbert S, Shannon K, Bollag G. Hyperactive Ras in developmental disorders and cancer. *Nature Reviews Cancer* 2007**; 7:** 295-308.

54. Bedford DC, Kasper LH, Fukuyama T, Brindle PK. Target gene context influences the transcriptional requirement for the KAT3 family of CBP and p300 histone acetyltransferases. *Epigenetics* 2010**; 5:** 9-15.

55. Dastugue N, Suciu S, Plat G, Speleman F, Cave H, Girard S*, et al.* Hyperdiploidy with 58-66 chromosomes in childhood B-acute lymphoblastic leukemia is highly curable: 58951 CLG-EORTC results. *Blood* 2013**; 121:** 2415-2423.