Science Advances
AAAS

advances.sciencemag.org/cgi/content/full/6/30/eaba3064/DC1

Supplementary Materials for

Diverse noncoding mutations contribute to deregulation of cis-regulatory landscape in pediatric cancers

Bing He, Peng Gao, Yang-Yang Ding, Chia-Hui Chen, Gregory Chen, Changya Chen, Hannah Kim, Sarah K. Tasian, Stephen P. Hunger, Kai Tan*

*Corresponding author. Email: tank1@email.chop.edu

Published 24 July 2020, *Sci. Adv.* **6**, eaba3064 (2020)
DOI: 10.1126/sciadv.aba3064

The PDF file includes:

Other Supplementary Material for this manuscript includes the following:

(available at advances.sciencemag.org/cgi/content/full/6/30/eaba3064/DC1)

Supplementary Methods

Disease-relevant cell types used to construct the enhancer and enhancer-promoter catalogs

For B-cell acute lymphoblastic leukemia (B-ALL), we used data of the following cell types: GM12878, CD19+ B cell, naïve B cell, total B cell. For acute myeloid leukemia (AML), we used data of the following cell types: K562, NB4, monocyte, monocyte progenitor, megakaryocyte, eosinophil, erythroblast, neutrophil, macrophage. For neuroblastoma (NBL), we used data of the following cell types: SK-N-SH, SH-SY5Y, neural crest cell, IMR-5/75. For Wilms tumor (WT), we used data of the following tissue/cell types: adult kidney, HEK293, IMR-5/75, RT407, G401, 786-O. For osteosarcoma (OS), we used data of the following cell types: Osteoblast, U2OS, SaOS2. The sources of the data sets are provided in Table S1.

Prediction of active enhancers using histone mark ChIP-Seq data

Enhancers were predicted using the Chromatin Signature Inference using Artificial Neural Network (CSI-ANN) algorithm (*13*). The inputs to the algorithm are normalized ChIP-Seq signals of three histone marks associated with enhancers (H3K4Me1, H3K4Me3, H3K27Ac). We used public histone mark ChIP-Seq data of human cell/tissue types relevant to the five cancer types in this study (Table S1). Raw ChIP-Seq data were first mapped to human GRCh37 genome using Bowtie2 (v2.1.0). CSI-ANN combines signals of all histone marks and uses an artificial neural network-based classifier to make predictions of active enhancers. A training set for the classifier was prepared using ENCODE data of GM12878, K562, and hESC cells. Specifically, a set of promoter-distal p300 binding sites

(2.5 kbp away from RefSeq TSS) in all three cell types was selected. The top 500 distal p300 sites that overlap with H3K4Me1 and H3K27Ac peaks, but not with H3K4Me3 peaks (promoter mark), were selected as the positive training set. One thousand randomly selected genomic regions and 500 active promoter regions were used as the negative training set. Enhancers were predicted using a false discovery rate (FDR) cutoff of 0.05. Predicted enhancers that overlap by >500 bp were merged by selecting the enhancer with the highest CSI-ANN score.

Prediction of SNVs that disrupt transcription factor binding

We used the Find Individual Motif Occurrences (FIMO) software to scan the 100 bp sequences flanking SNVs. A collection of transcription factor DNA binding motifs from the Cis-BP database (v1.02) was used for motif scan (*48*). Motif score differences between the sequences harboring the alternative alleles of the SNV were computed. To determine if one of the alleles of a SNV causes significance change in the TF motif score, a p-value for the TF motif score difference was calculated using a null distribution computed with all SNPs identified by the 1000 Genome Project. P-values were adjusted for multiple testing using the Benjamini-Hochberg method.

Prediction of enhancer-promoter interactions

Target promoters of enhancers were predicted using the Integrated Method for Predicting Enhancer Targets (IM-PET) algorithm (*14*). It predicts enhancer-promoter interactions by integrating four features derived from transcriptome, epigenome, and multiple genome sequence data, including: 1) enhancer-promoter activity correlation, 2) transcription factor-

promoter co-expression, 3) enhancer-promoter co-evolution, and 4) enhancer-promoter distance. Public histone mark ChIP-Seq and RNA-Seq data (Table S1) were used to compute values of features 1, 2 and 4. Values of feature 3 were computed based on sequence conservation across 15 mammalian species (human, chimp, gorilla, orangutan, gibbon, rhesus, baboon, marmoset, tarsier, mouse lemur, tree shrew, mouse, rat, rabbit, and guinea pig). We used an FDR cutoff of 0.01 for making predictions.

Parameter optimization of the weighted elastic net model

The translocation t(14,X) is known to hijack a super enhancer on chromosome 14 to near *CRLF2* gene and result in the overexpression of *CRLF2*. To optimize parameters of the weighted elastic net (WEN) model, we constructed the mutation count matrix of *CRLF2* disrupting its cis-regulatory elements. We used 10-fold cross-validation to tune the parameters of the WEN model. In each iteration, we used the data of 147 patients for training the model. The trained model was then used to predict *CRLF2* expression using data from the remaining 16 patients. Since our model predicts expression value with cis-regulatory mutations (e.g. enhancer mutations or mutations disrupt enhancer-promoter interactions) (predictors), we define a mean squared error as the difference between expected expression value based on the fitted model and observed expression value of *CRLF2*. The model fitting of the weighted elastic net is done using the Least Angle Regression (LARS) algorithm implemented in the “AdapEnetClass” R package. The WEN model has two adjustable parameters: penalty parameter λ_2 and forward selection step of LARS algorithm. The penalty parameter λ_2 controls the strength of the penalty term in the ridge regression and LASSO regression. The forward step of the LARS algorithm controls

the number of predictors considered by the model. We tested a combination of six different λ_2 values (0, 0.1, 0.5, 1, 2, 3) and fifteen forward steps (1-15). We also performed the tenfold cross validation using permutated CRLF2 expression values. The model generated using real data out-performed the ones generated using permutated data, suggesting the robustness of the trained model (Fig. S1M). We also tested the sensitivity of prediction result on parameter setting. We made predictions using three λ_2 values, 0.1, 0.2, and 0.3. Generally with the increase of λ_2, the number of predicted genes decreased modestly (4% decrease with λ_2 = 0.2 and 15% decrease with λ_2 = 0.3). And the genes predicted with λ_2 = 0.2 or 0.3 were roughly a subset of genes predicted with λ_2 = 0.1 (Fig. S1M). Based on these analyses, we chose to use λ_2 = 0.1 and forward step = 15 to make the final predictions in this study.

The statistical significance of each regression coefficient was computed using the "AdapEnetClass" R package. The p-values were adjusted for multiple-testing with the Benjamini-Hochberg method. As a result, our method selects mutations that are significantly associated with the gene expression change in the patients with an adjusted p-value < 0.05.

Evaluation of false prediction rate using an independent cohort

To validate the recurrence of predicted causal noncoding mutations, we followed the method described in (*49*) to generate the quantile-quantile (QQ) plot. We downloaded simple somatic mutations and somatic structural variants in 2,715 donors from the ICGC data portal. For each genes affected by noncoding mutations, the number of mutated patients, k, was used as the test statistic. In our analysis with TARGET data, 2% of SNVs

in enhancer/promoter regions affect gene expression, and 24% of structural variant break points affect gene expression. To compute the background mutation rate (BMR) for each gene in each ICGC donor, we counted the number of structural variants in 200 kbp region and the number of SNVs in enhancers/promoters and computed the BMR under the assumption that the causal noncoding mutations rate is the same as the TARGET data. Then we computed the probability of having observed k or more mutations in n patients in the gene of interest using a Poisson binomial model:

$$P(K \geq k) = \sum_{l=k}^{n} \sum_{A \in F_l} \prod_{i \in A} p_i \prod_{j \in A^c} (1 - p_j)$$

where F_l is the set of all subsets of k integers that can be selected from {1, 2, …, n}, p_i or p_j is the probability that patient *i* or patient *j* has the mutation; A is a set of k integers that can be selected from {1, 2,…, n} and A^c is the complement of A. We used an approximation for the Poisson binomial distribution implemented in the Python package, poibin. The QQ plot shows the observed P values versus P-values based on random expectation.

Replication timing analysis

Genome-wide replication timing data for GM12878, K562, SK-N-SH, HEK293, and U2OS cells were downloaded from the Replication Domain database (*50*). These data were generated using the repli-seq method (*51*). Briefly, cells were sorted into early (S) and late (G1) phase fractions on the basis of DNA content using flow cytometry. BrdU-labeled DNA from each fraction was immunoprecipitated, amplified, and sequenced. The replication timing was measured as the $\log_2$ ratio of early over late fraction reads in

5 kb bins. The replication timing for each gene was calculated as the average $\log_2$ ratio of early over late fraction reads across the whole gene body.

Regulon disruption of transcription factors in cancer patients

Regulon of a transcription factor is defined as the set of target genes of the TF. To identify the target gene, the enhancer and promoter sequences of the gene is scanned for binding motif hit of the TF using Find Individual Motif Occurrences (FIMO) and an FDR cutoff of 0.01 (*52*). The regulon of a given TF is defined as disrupted if at least one of the following three scenarios is observed: 1) coding region of the TF gene is mutated; 2) at least one TF binding site in the enhancer/promoter of a target gene is mutated; 3) at least one enhancer-promoter interaction involving the TF and a target gene is disrupted by a SV. We ranked the transcription factors according to the number of patients with regulon disruption by combined coding and noncoding mutations.

Mutual exclusivity and co-occurrence of mutations

We tested mutual exclusivity and co-occurrence of mutations for genes mutated in at least 5 patients. For a given gene pair, we performed Fisher's exact test using the Comet R package (*53*). The Benjamini-Hochberg method was used to correct for multiple testing.

Clustering of cancer subtypes

We first generated a joint mutational profile for each patient covering both coding and noncoding mutations. These include: non-silent point mutations in the coding region, copy number alterations in the coding region, gene fusion, SNVs or small indels in enhancers/promoters of the gene, copy number alterations affecting enhancers and

enhancer rearrangement. We performed hierarchical clustering to identify patient groups based on their joint mutational profiles. We used the elbow method to determine the optimal number of clusters. To do so, we cut the hierarchical clustering dendrogram to generate different numbers of clusters ranging from two to fifteen. For each number of clusters, the total within-cluster sum of square (WSS) was calculated. The curve of WSS versus the number of clusters was plotted. The location of a bend in the plot was used to determine the optimal number of clusters**.**

RT-qPCR

Total RNA was isolated using the RNeasy micro kit (Qiagen) including on-column DNase digestion to remove genomic DNA. cDNA was synthesized from total RNA using the high-capacity cDNA reverse transcription kit (Applied Biosystems) according to the manufacturer's instructions. qPCR reactions were performed on a Bio-Rad CFX Connect real-time PCR system with iQ^{TM} SYBR$^{®}$ Green Supermix (Bio-Rad) according to manufacturers' instructions. Relative gene expression was calculated using the $2^{-\Delta\Delta Ct}$ method using *Tbp* as the reference gene. Each sample was assayed in duplicate, and at least three independent samples were analyzed for each experimental condition.

Luciferase reporter assay

Candidate enhancer (~3 kb) was cloned into the luciferase reporter vector pGL3 (Promega) using in-fusion HD cloning kit (Clontech). A super core promoter 1 (SCP1) (*54*) was used as the basal promoter. A negative control region of similar length with no enhancer-associated histone modification signals was also cloned into the same vector as

a negative control. Reporter constructs were co-transfected with the internal control construct pRL-TK (Promega) into Ba/F3 cells by electroporation. 48h post transfection, cells with three independent transfections were harvested and measured for firefly and Renilla luciferase activities using the dual-luciferase reporter assay system (Promega). Firefly luciferase activity of individual transfections was normalized against Renilla luciferase activity.

sgRNA design

The sgRNAs targeting *CHD4* were designed by Feng Zhang's laboratory (*55*). The sequences for non-targeting control sgRNAs were based on a previous publication (*56*). All sequences are listed in Table S4.

Competition growth assay

NALM-6 and REH cells stably expressing Cas9 (NALM6-Cas9 and REH-Cas9) were separately transduced with the lentiviral vectors carrying *CHD4*-sgRNA-GFP or Safe-sgRNA-mCherry. Three days post transduction, transduced cells were pooled together. Depletion of fluorescence signal was measured by flow cytometry at indicated time points.

Supplementary Figures

Figure S1. Identification of putative causal noncoding mutations affecting enhancer/promoter functions.

(A) Pipeline for calling single nucleotide variants (SNVs) and small insertions and deletions (indels) (size < 60bp). We used GATK Haplotypecaller (v3.8) and Freebayes (v1.0.2) to call SNVs and small indels. The p-value is calculated using Fisher's exact test to test whether the mutant read count in the tumor is significantly higher than normal. All p-values were adjusted for multiple testing. **(B)** Quality assessment of identified SNVs and small indels using a set of 735 high-confidence SNVs. The high-confidence set was generated by the TARGET project for B-ALL, AML, NBL, and WT and were validated using multiple experimental protocols including WES, RNA-Seq and PCR. **(C)** Mutation rates of all SNVs, SNVs in coding region, promoter region and enhancer region. **(D)** Mutation rates of SNVs across five cancer types. **(E)** Genomic distribution of identified SNVs and small indels. Other intergenic, intergenic regions other than promoters and enhancers. **(F)** Pipeline for calling structural variants (SVs). We used Delly (v0.7.2) and Lumpy (v0.2.13) to call SVs. **(G)** Number of identified SVs in each cancer type; **(H)** Quality assessment of identified SVs using 12 known SVs in 212 leukemia patients. **(I)** Corroborating evidence for the predicted enhancers in five cancer types. Percentage of overlap of our predicted enhancers with one or more lines of corroborating evidence is shown. Numbers in the brackets denote the number of enhancers identified in the cancer-relevant cell types. Data sources of the public ATAC-Seq data are summarized in Table S1. **(J)** Overlap of enhancer-promoter interactions predicted by IM-PET and published high-resolution chromatin interactions. **(K)** Validation of noncoding mutation recurrence in a pan-cancer cohort from ICGC. The quantile-quantile (QQ) plot shows the observed empirical P values of mutation recurrence (n = 2,706 samples) compared to P values based on random expectation for all noncoding mutations in the TARGET cohort. **(L)** Number of predicted causal mutations for each mutation type normalized by the total number of mutations. **(M)** Parameter optimization of the weighted elastic net model. We tested all combinations of 6 different λ_2 values (0, 0.1, 0.5, 1, 2, 3) and fifteen forward selection steps (1-15) of the LARS algorithm. Left, Mean squared errors of 10-fold cross validation using different parameter settings. Right, Number of genes with predicted causal noncoding mutations using different λ_2 values (0.1, 0.2, 0.3).

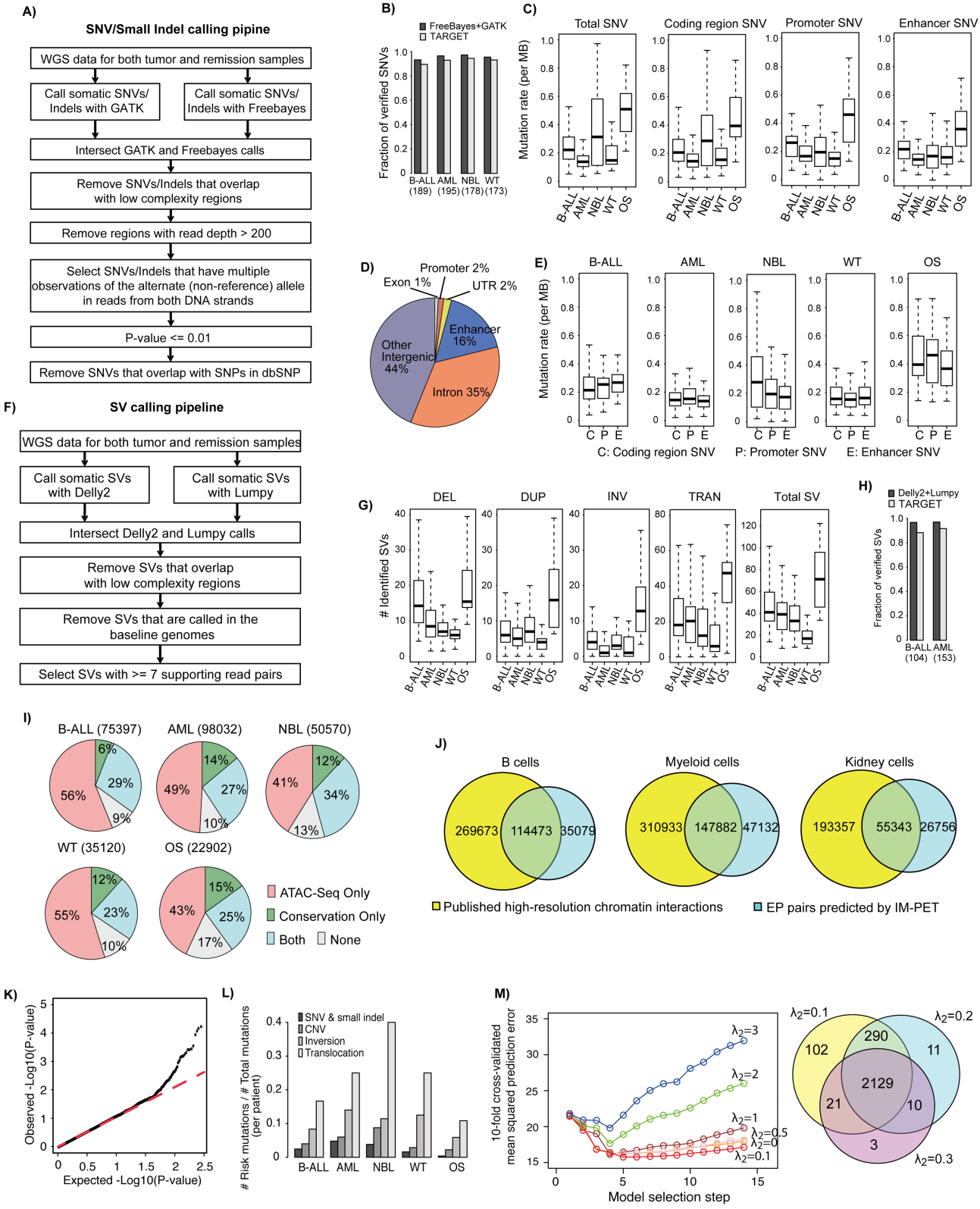

A)
SNV/Small Indel calling pipine
WGS data for both tumor and remission samples
Call somatic SNVs/ Indels with GATK
Call somatic SNVs/ Indels with Freebayes
Intersect GATK and Freebayes calls
Remove SNVs/Indels that overlap with low complexity regions
Remove regions with read depth > 200
Select SNVs/Indels that have multiple observations of the alternate (non-reference) allele in reads from both DNA strands
P-value <= 0.01
Remove SNVs that overlap with SNPs in dbSNP
B)
FreeBayes+GATK
TARGET
Fraction of verified SNVs
B-ALL (189) AML (195) NBL (178) WT (173)
C)
Total SNV
Coding region SNV
Promoter SNV
Enhancer SNV
Mutation rate (per MB)
B-ALL AML NBL WT OS
D)
Promoter 2%
Exon 1%
UTR 2%
Enhancer 16%
Other Intergenic 44%
Intron 35%
E)
B-ALL
AML
NBL
WT
OS
Mutation rate (per MB)
C P E
C: Coding region SNV
P: Promoter SNV
E: Enhancer SNV
F)
SV calling pipeline
WGS data for both tumor and remission samples
Call somatic SVs with Delly2
Call somatic SVs with Lumpy
Intersect Delly2 and Lumpy calls
Remove SVs that overlap with low complexity regions
Remove SVs that are called in the baseline genomes
Select SVs with >= 7 supporting read pairs
G)
DEL
DUP
INV
TRAN
Total SV
Identified SVs
H)
Delly2+Lumpy
TARGET
Fraction of verified SVs
B-ALL (104) AML (153)
I)
B-ALL (75397)
AML (98032)
NBL (50570)
WT (35120)
OS (22902)
ATAC-Seq Only
Conservation Only
Both
None
J)
B cells
Myeloid cells
Kidney cells
269673 114473 35079
310933 147882 47132
193357 55343 26756
Published high-resolution chromatin interactions
EP pairs predicted by IM-PET
K)
Observed -Log10(P-value)
Expected -Log10(P-value)
L)
SNV & small indel
CNV
Inversion
Translocation
Risk mutations / # Total mutations (per patient)
B-ALL AML NBL WT OS
M)
10-fold cross-validated mean squared prediction error
Model selection step
λ2=3
λ2=2
λ2=1
λ2=0.5
λ2=0
λ2=0.1
λ2=0.1
λ2=0.2
λ2=0.3
102
290
11
2129
21
10
3

Figure S2. Example predicted enhancer hijacking events in pediatric cancers and experimental validations.

(A) Left, Genome browser view of enhancer hijacking to *CRLF2* via the t(14;X)(q32;p22) translocation. Shown tracks are histone modification ChIP-Seq data in $CD19^{+}$ B cell and identified SV break points (BPs). The hijacked enhancers predicted to regulate *CRLF2* are highlighted in brown. Right, Expression level of *CRLF2* in patients with and without the translocations. **(B)** Top, Genome browser view of enhancer hijacking to *TERT. TERT* has multiple translocation partners in neuroblastoma patients, including (t(10;5)(p22;p15), t(5;5)(q34;p15), and t(5;5)(q12;p15). Shown tracks are histone modification ChIP-Seq data in normal neural crest cells and identified SV break points (BPs). The hijacked enhancers predicted to regulate *TERT* are highlighted in brown. Bottom, Expression level of *TERT* in patients with and without the translocations. **(C)** We obtained WGS and RNA-Seq data from a recent published pediatric MPAL cohort (Alexander *et al. Nature*. 2018). MPAL, mixed phenotype acute leukemia. Among 94 MPAL patients, 15 had translocations near *ZNF384* and *CHD4*. The expression level of *CHD4* is significantly higher in those patients. **(D)** Time to relapse of patients with and without the *ZNF384*/*CHD4* rearrangement. **(E)** Enrichment of CHD4 targets among down-regulated genes in patients with *CHD4* translocation. CHD4 targets are defined as genes whose enhancer or promoter is bound by CHD4 according to the ChIP-Seq data in GM12878 cells. Down-regulated genes in patients with CHD4 translocation were identified using edgeR (q-val < 0.05). The proportion of all genes bound by CHD4 was calculated as the negative control. Hypergeometric p-value is shown. **(F)** Genome browser view of *PAX5, IRF4, EBF1, and TCF3*. Shown tracks are histone modifications and CHD4 ChIP-Seq signals in GM12878 cells. **(G)** Expression levels of *PAX5, IRF4*, *EBF1*, and *TCF3* in patients with and without CHD4 enhancer hijacking. **(H)** Schematic for the generation of t(6;15)(qF2:qE1) translocation in Ba/f3 cells. The translocation does not create fusion gene involving *CHD4* or *ZNF384*. Instead, it hijacks the enhancer of *EP300* to the vicinity of the *CHD4* promoter. **(I)** Relative mRNA levels of *TCF3*, *PAX5*, and *EBF1* in Ba/F3 cells with and without the introduced translocation. P values were calculated using one-sided Student's t-test (n=2).

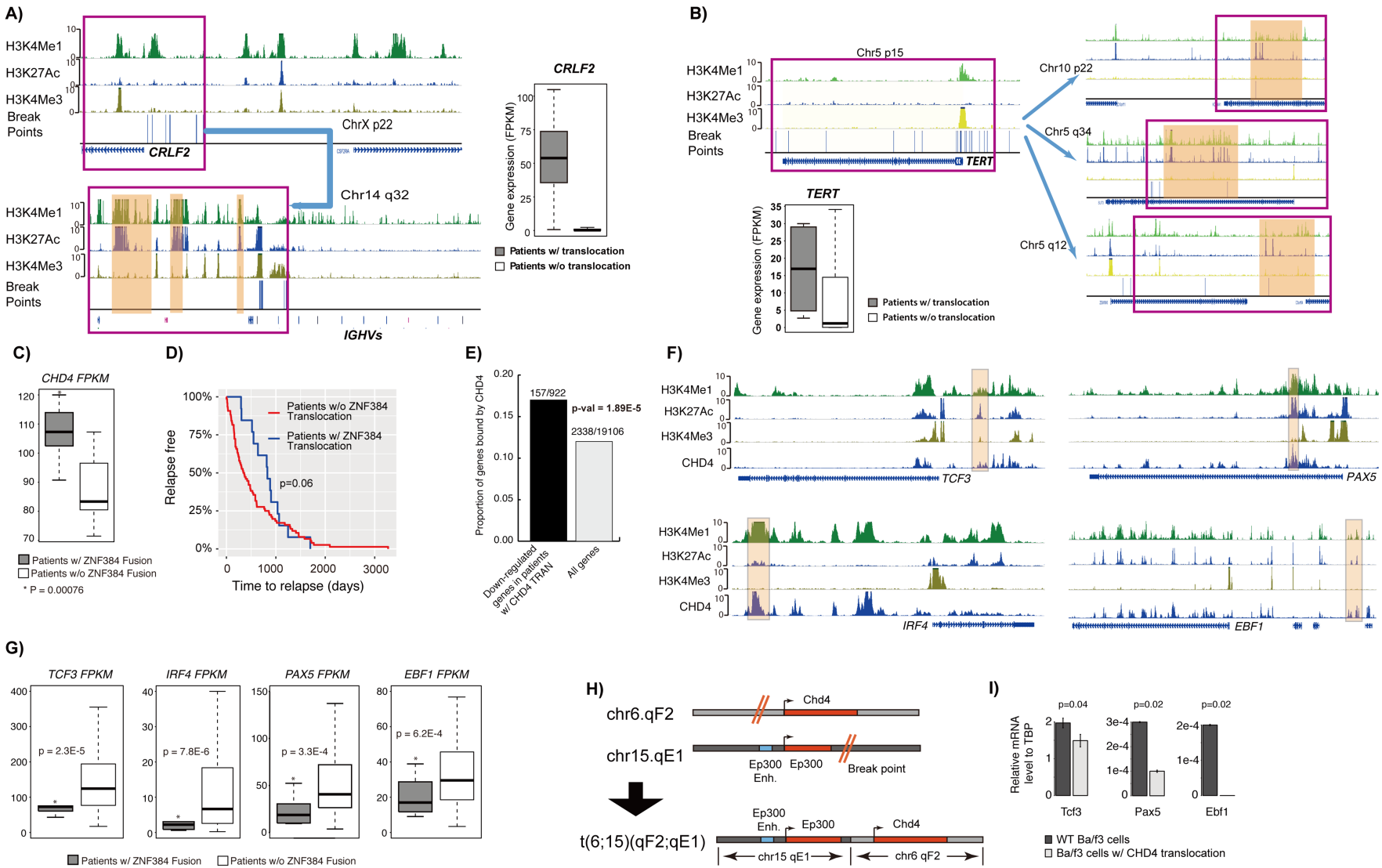

A)
H3K4Me1
H3K27Ac
H3K4Me3
Break Points
ChrX p22
CRLF2
Chr14 q32
IGHVs
CRLF2
Gene expression (FPKM)
Patients w/ translocation
Patients w/o translocation
B)
Chr5 p15
TERT
Chr10 p22
Chr5 q34
Chr5 q12
TERT
Gene expression (FPKM)
Patients w/ translocation
Patients w/o translocation
C)
CHD4 FPKM
Patients w/ ZNF384 Fusion
Patients w/o ZNF384 Fusion
* P = 0.00076
D)
Relapse free
Patients w/o ZNF384 Translocation
Patients w/ ZNF384 Translocation
p=0.06
Time to relapse (days)
E)
Proportion of genes bound by CHD4
157/922
p-val = 1.89E-5
2338/19106
Down-regulated genes in patients w/ CHD4 TRAN
All genes
F)
H3K4Me1
H3K27Ac
H3K4Me3
CHD4
TCF3
PAX5
IRF4
EBF1
G)
TCF3 FPKM
IRF4 FPKM
PAX5 FPKM
EBF1 FPKM
p = 2.3E-5
p = 7.8E-6
p = 3.3E-4
p = 6.2E-4
Patients w/ ZNF384 Fusion
Patients w/o ZNF384 Fusion
H)
chr6.qF2
chr15.qE1
Chd4
Ep300 Enh.
Ep300
Break point
t(6;15)(qF2;qE1)
chr15 qE1
chr6 qF2
I)
Relative mRNA level to TBP
p=0.04
p=0.02
p=0.02
Tcf3
Pax5
Ebf1
WT Ba/f3 cells
Ba/f3 cells w/ CHD4 translocation

Figure S3. Additional examples of enhancer alterations in pediatric cancers.

(A) Genome browser view of *ATG3* enhancer deletion. The enhancers are highlighted in brown. Tracks shown are average H3K4Me1, H3K4Me3, and H3K27Ac signals in human CD19+ B cells. P-value of one-sided t-test is shown (n=153). **(B)** Genome browser view of *GATA2* gene and its regulating enhancers. The enhancers are highlighted in brown. Tracks shown are average H3K4Me1, H3K4Me3, and H3K27Ac signals in 13 myeloid cell types, and frequency of the identified SNVs in AML patients. *IDH2* expression values in two groups of patients are shown in the right panel. P-value of one-sided t-test is shown (n=153). **(C)** Genome browser view of *GFI1B* gene and its regulating enhancers. P-value of one-sided t-test is shown (n=153).

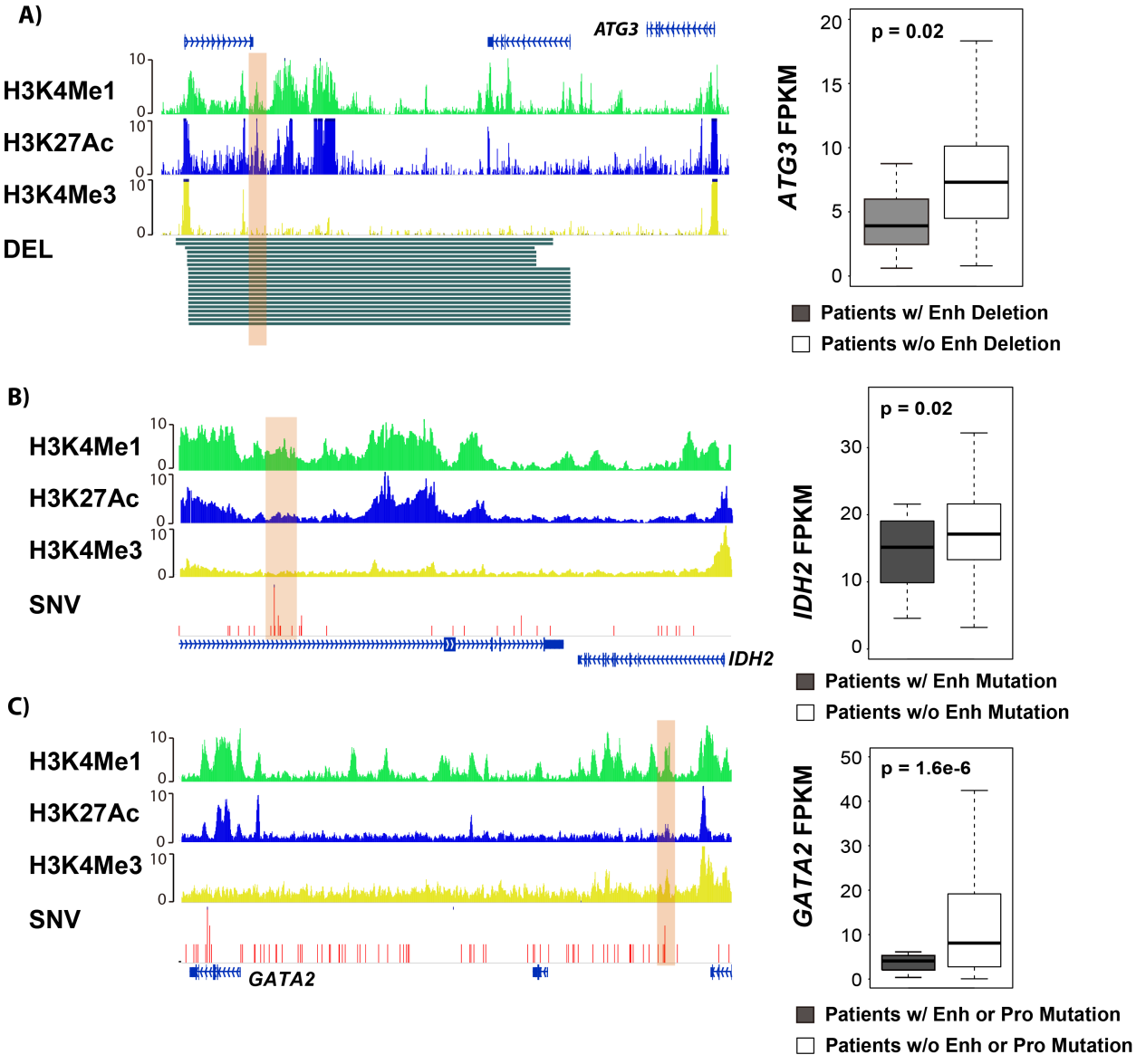
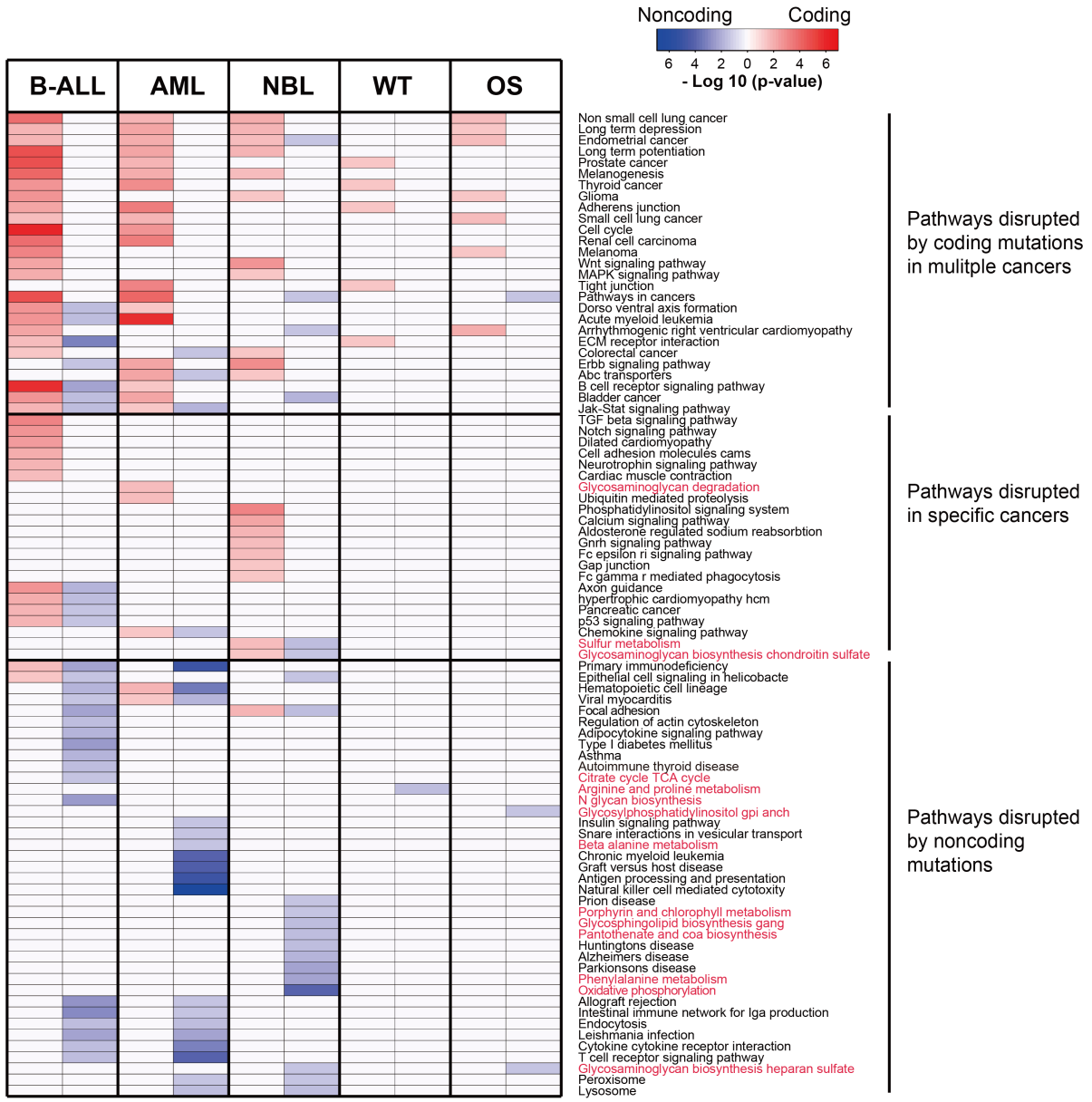


Figure S4. Coding and noncoding mutations affect different pathways.

(A) Heatmap showing p-values of enriched pathways among genes with coding mutations (red) and noncoding mutations (blue). Pathways highlighted in red are metabolic pathways. **(B)** Replication timing of metabolic genes. Genome-wide replication timing data for GM12878, K562, SK-N-SH, HEK293, and U2OS cells were downloaded from the Replication Domain database. The replication timing was measured as the $\log_2$ ratio of the array signals in early (S) phase over the array signals in late (G1) phase in 5 kb bins. The list of metabolic genes (2071) is curated from the metabolic pathways of KEGG, Reactome, and NCI-Nature pathway databases. P-value of one-sided t-test is shown (n=21,841).

A)


Noncoding
Coding
6 4 2 0 2 4 6
- Log 10 (p-value)
B-ALL
AML
NBL
WT
OS
Non small cell lung cancer
Long term depression
Endometrial cancer
Long term potentiation
Prostate cancer
Melanogenesis
Thyroid cancer
Glioma
Adherens junction
Small cell lung cancer
Cell cycle
Renal cell carcinoma
Melanoma
Wnt signaling pathway
MAPK signaling pathway
Tight junction
Pathways in cancers
Dorso ventral axis formation
Acute myeloid leukemia
Arrhythmogenic right ventricular cardiomyopathy
ECM receptor interaction
Colorectal cancer
Erbb signaling pathway
Abc transporters
B cell receptor signaling pathway
Bladder cancer
Jak-Stat signaling pathway
TGF beta signaling pathway
Notch signaling pathway
Dilated cardiomyopathy
Cell adhesion molecules cams
Neurotrophin signaling pathway
Cardiac muscle contraction
Glycosaminoglycan degradation
Ubiquitin mediated proteolysis
Phosphatidylinositol signaling system
Calcium signaling pathway
Aldosterone regulated sodium reabsorbtion
Gnrh signaling pathway
Fc epsilon ri signaling pathway
Gap junction
Fc gamma r mediated phagocytosis
Axon guidance
hypertrophic cardiomyopathy hcm
Pancreatic cancer
p53 signaling pathway
Chemokine signaling pathway
Sulfur metabolism
Glycosaminoglycan biosynthesis chondroitin sulfate
Primary immunodeficiency
Epithelial cell signaling in helicobacte
Hematopoietic cell lineage
Viral myocarditis
Focal adhesion
Regulation of actin cytoskeleton
Adipocytokine signaling pathway
Type I diabetes mellitus
Asthma
Autoimmune thyroid disease
Citrate cycle TCA cycle
Arginine and proline metabolism
N glycan biosynthesis
Glycosylphosphatidylinositol gpi anch
Insulin signaling pathway
Snare interactions in vesicular transport
Beta alanine metabolism
Chronic myeloid leukemia
Graft versus host disease
Antigen processing and presentation
Natural killer cell mediated cytotoxity
Prion disease
Porphyrin and chlorophyll metabolism
Glycosphingolipid biosynthesis gang
Pantothenate and coa biosynthesis
Huntingtons disease
Alzheimers disease
Parkionsons disease
Phenylalanine metabolism
Oxidative phosphorylation
Allograft rejection
Intestinal immune network for Iga production
Endocytosis
Leishmania infection
Cytokine cytokine receptor interaction
T cell receptor signaling pathway
Glycosaminoglycan biosynthesis heparan sulfate
Peroxisome
Lysosome
Pathways disrupted by coding mutations in mulitple cancers
Pathways disrupted in specific cancers
Pathways disrupted by noncoding mutations

B)

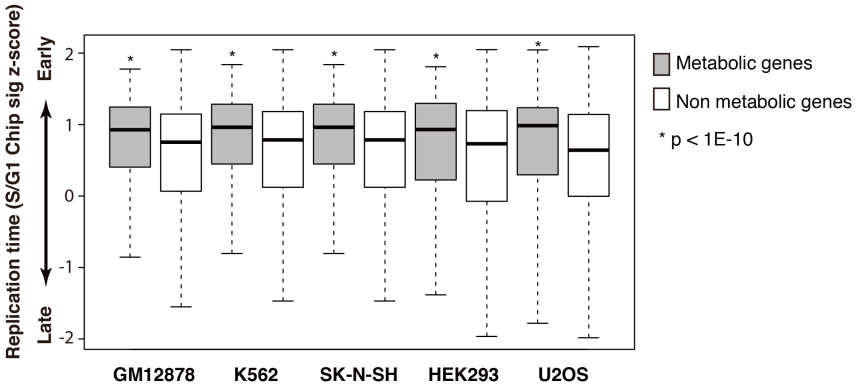

Replication time (S/G1 Chip sig z-score)
Early
Late
2
1
0
-1
-2
GM12878
K562
SK-N-SH
HEK293
U2OS
Metabolic genes
Non metabolic genes
* p < 1E-10

Figure S5. Putative causal noncoding mutations defines novel B-ALL subgroups.

(A) Genome browser view of enhancer hijacking of *SUPT7L*. Shown tracks are histone modification ChIP-Seq data in $CD19^{+}$ B cells and SV break points (BP). The hijacked enhancer predicted to regulate *SUPT7L* is highlighted in brown. **(B)** Expression level of *SUPT7L* in patients with and without the inversions. **(C)** Contingency matrix indicates patient mutation status of *TCF3* and *SUPT7L*. **(D)** Time to relapse of patients with *TCF3-PBX1* and Inv(2) mutations. P-value of one-sided log-rank test is shown (n=159). **(E)** Genome browser view of enhancer rearrangement of *MIR663B.* Shown tracks are histone modifications in CD19+ B cells, and identified SV break points. The enhancer regulating *MIR663B* is highlighted in brown. **(F)** Expression levels of *ANKRD30BL, MIR663B, CCL17, CD40, PIK3CD* in patients with and without the translocation. P-values of one-sided t-test are shown (n=163).

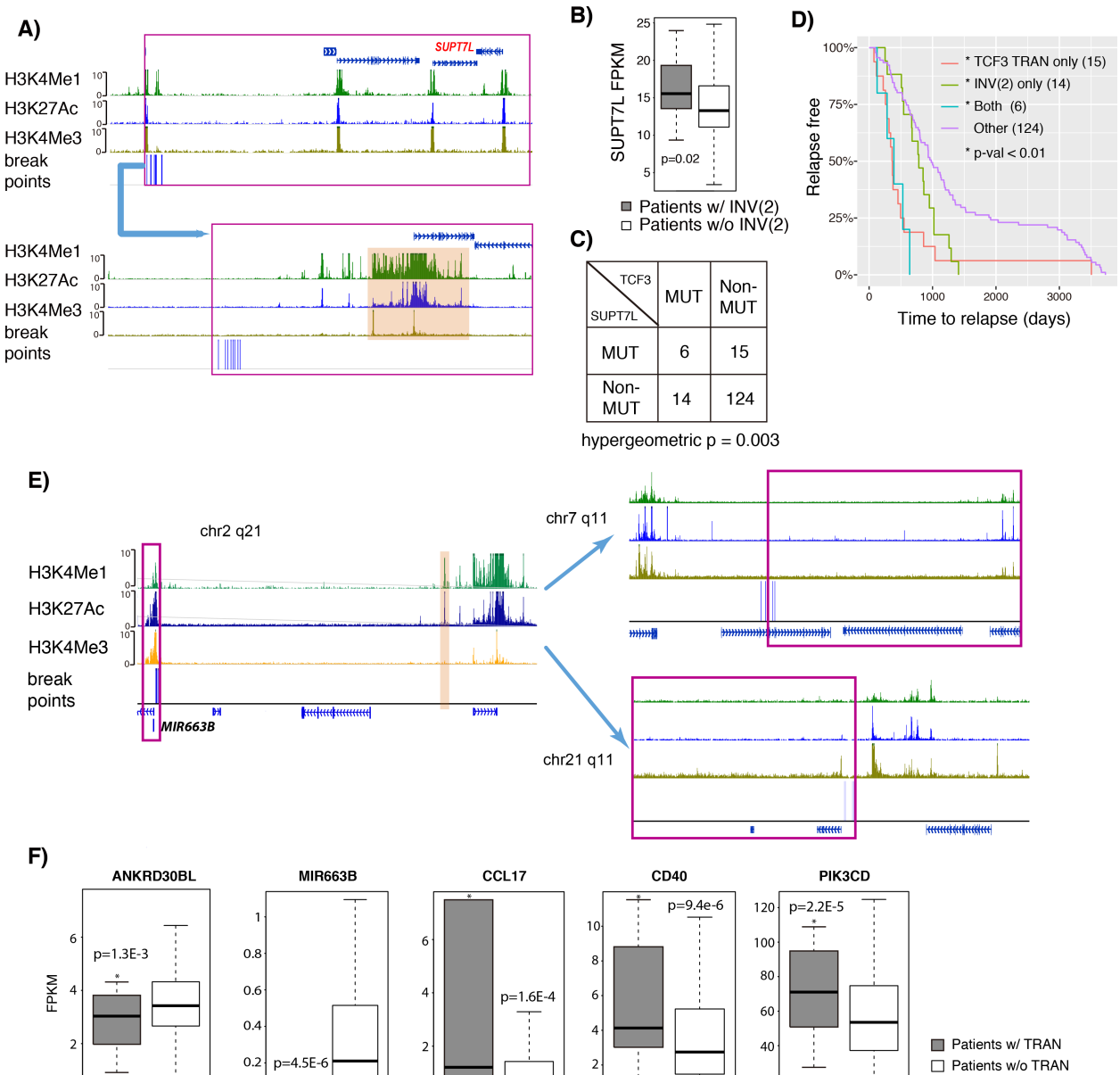
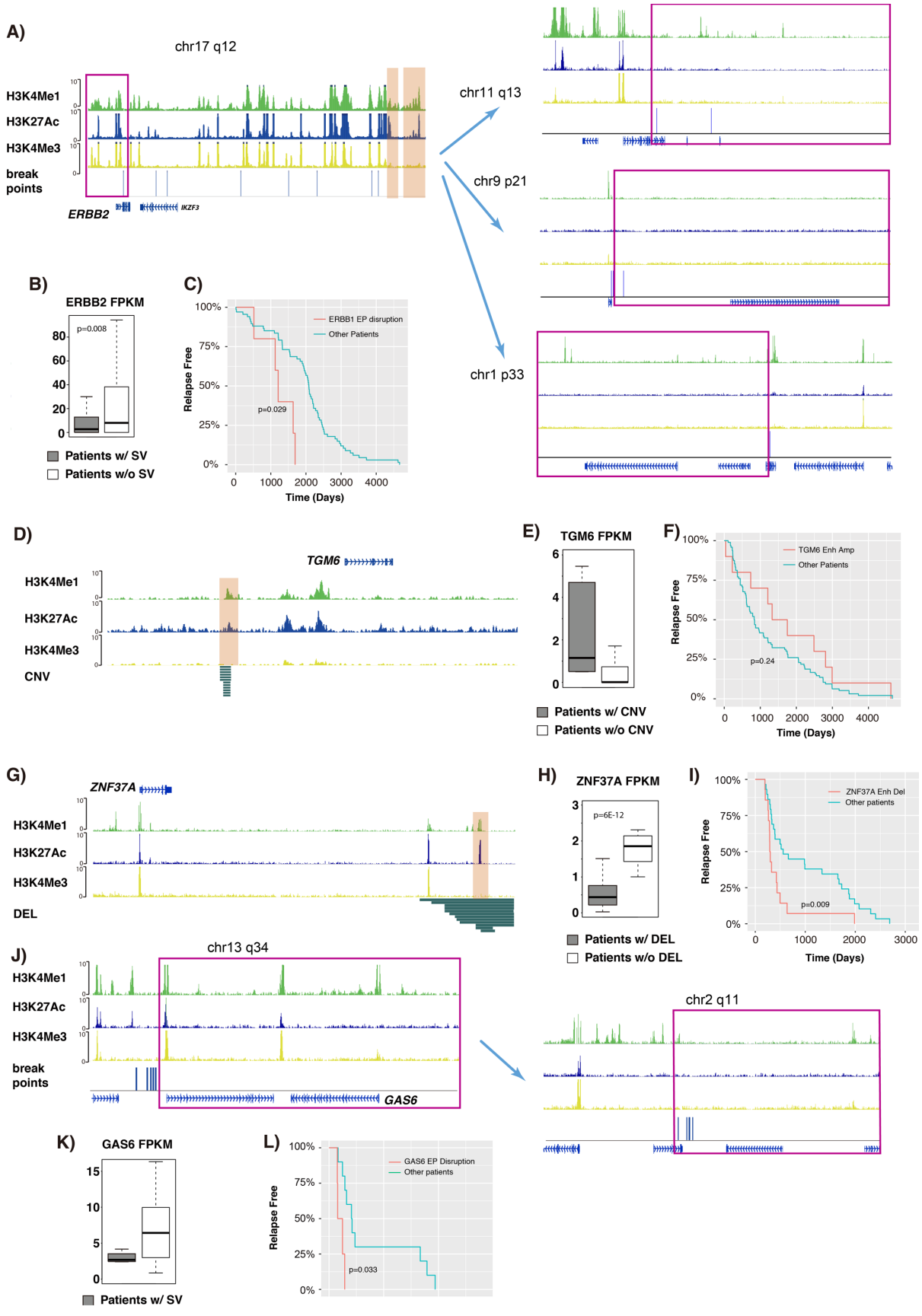

A)
SUPT7L
H3K4Me1
H3K27Ac
H3K4Me3
break points
B)
SUPT7L FPKM
p=0.02
Patients w/ INV(2)
Patients w/o INV(2)
C)
TCF3
SUPT7L
MUT
Non-MUT
MUT 6 15
Non-MUT 14 124
hypergeometric p = 0.003
D)
Relapse free
Time to relapse (days)
* TCF3 TRAN only (15)
* INV(2) only (14)
* Both (6)
Other (124)
* p-val < 0.01
E)
chr2 q21
MIR663B
chr7 q11
chr21 q11
F)
ANKRD30BL
MIR663B
CCL17
CD40
PIK3CD
FPKM
p=1.3E-3
p=4.5E-6
p=1.6E-4
p=9.4e-6
p=2.2E-5
Patients w/ TRAN
Patients w/o TRAN

Figure S6. Putative causal noncoding mutations defines novel patient subgroups for AML, NBL, WT.

(A) Genome browser view of enhancer rearrangement of *ERBB2*. Shown tracks are histone modification ChIP-Seq data in neural crest cell (*Prescott et al. Cell. 2015*) and identified SV break points. The enhancers predicted to regulate *ERBB2* are highlighted in brown. **(B)** Expression level of *ERBB2* in patients with and without the EP disruption. P-value of one-sided t-test is shown (n=100). **(C)** Time to relapse of patients with and without the *ERBB2* EP disruption. P-value of one-sided log-rank test is shown (n=100). **(D)** Genome browser view of enhancer copy number change of *TGM6*. Shown tracks are histone modification ChIP-Seq data in neural crest cells and identified CNVs. The enhancer predicted to regulate *TGM6* is highlighted in brown. **(E)** Expression level of *TGM6* in patients with and without enhancer duplication. P-value of one-sided t-test is shown (n=100). **(F)** Time to relapse of patients with and without the *TGM6* enhancer duplication. P-value of one-sided log-rank test is shown (n=100). **(G)** Genome browser view of enhancer deletion of *ZNF37A* in AML patients. Shown tracks are histone modification ChIP-Seq data in human myeloid cells, and identified deletions. The enhancer predicted to regulate *ZNF37A* is highlighted in brown. **(H)** Expression level of *ZNF37A* in patients with and without the enhancer deletion. P-value of one-sided t-test is shown (n=153). **(I)** Time to relapse of patients with and without the *ZNF37A* enhancer deletion. P-value of one-sided log-rank test is shown (n=153). **(J)** Genome browser view of enhancer rearrangement of *GAS6*. Shown tracks are histone modification ChIP-Seq data in HEK293 cells, and identified SV break points. **(K)** Expression level of *GAS6* in patients with and without the EP disruption. P-value of one-sided t-test is shown (n=53). **(L)** Time to relapse of patients with and without the *GAS6* EP disruption. P-value of one-sided log-rank test is shown (n=53).


A)
chr17 q12
H3K4Me1
H3K27Ac
H3K4Me3
break points
ERBB2
IKZF3
chr11 q13
chr9 p21
chr1 p33
B)
ERBB2 FPKM
p=0.008
Patients w/ SV
Patients w/o SV
C)
ERBB1 EP disruption
Other Patients
Relapse Free
p=0.029
Time (Days)
D)
TGM6
H3K4Me1
H3K27Ac
H3K4Me3
CNV
E)
TGM6 FPKM
Patients w/ CNV
Patients w/o CNV
F)
TGM6 Enh Amp
Other Patients
Relapse Free
p=0.24
Time (Days)
G)
ZNF37A
H3K4Me1
H3K27Ac
H3K4Me3
DEL
H)
ZNF37A FPKM
p=6E-12
Patients w/ DEL
Patients w/o DEL
I)
ZNF37A Enh Del
Other patients
Relapse Free
p=0.009
Time (Days)
J)
chr13 q34
H3K4Me1
H3K27Ac
H3K4Me3
break points
GAS6
chr2 q11
K)
GAS6 FPKM
Patients w/ SV
L)
GAS6 EP Disruption
Other patients
Relapse Free
p=0.033

Supplementary Tables

Table S1. Published data and software used in this study.

Data	Source	Accession # / Link
TARGET B-ALL WGS and RNA-Seq	TARGET	SRP011998, SRP011999
TARGET AML WGS and RNA-Seq	TARGET	SRP012000
TARGET NBL WGS and RNA-Seq	TARGET	SRP012002
TARGET WT WGS and RNA-Seq	TARGET	SRP012006
TARGET OS WGS and RNA-Seq	TARGET	SRP012003
Histone modification ChIP-Seq and RNA-Seq data for GM12878, CD19+ B cell, K562, NB4, adult kidney, SK-N-SH, HEK293, osteoblast cells	ENCODE	https://www.encodeproject.org/matrix/?type=Experiment&status=released
Histone modification ChIP-Seq and RNA-Seq data for monocyte, monocyte progenitor, megakaryocyte, eosinophil, erythroblast, neutrophil, macrophage	BLUEPRINT	https://epigenomesportal.ca/ihec/
Histone modification ChIP-Seq and RNA-Seq data for IMR-5/75 and SH-SY5Y cells	Henrich et al. Cancer Res. 2016	GSE80197, GSE80397
Histone modification ChIP-Seq and RNA-Seq data for neural crest cell	Prescott et al. Cell. 2015	GSE70751
Histone modification ChIP-Seq and RNA-Seq data for RT407 and G401 cells	Wang et al. Nat. Genet. 2017	GSE71506
Histone modification ChIP-Seq and RNA-Seq data for 786-O cell	Platt et al. EMBO Rep. 2016	GSE67237, GSE78113
Histone modification ChIP-Seq and RNA-Seq data for U2OS cell	Walz et al. Nature. 2014	GSE44672
ATAC-Seq data for GM12878	Buenrostro et al. Nat Methods, 2013	GSE47753
ATAC-Seq data for CD19+ B cell, Erythroblast and NK cells	Corces et al., Nat Genet, 2014	GSE74912
ATAC-Seq data for K562	Schmidl et al., Nat Methods, 2015	GSE70482
ATAC-Seq data for neural crest cell	Prescott et al. Cell. 2015	GSE70751
ATAC-Seq data for HEK293 cell	Karabacak et al., Genome Biol, 2019	GSE108513
ATAC-Seq data for SaOS2 cell	Morris et al., Nat Genet, 2019	GSE120755
Hi-C data for GM12878 and K562 cells	Rao et al., Cell, 2014	GSE63525
Hi-C data for monocyte and macrophage	Phanstiel et al., Mol Cell, 2017	http://www.aidenlab.org/juicebox/
Hi-C data for HEK293	Zuin et al., PNAS, 2014	GSE44267
Hi-C data for SK-N-SH	Guo et al., Cell, 2015	GSE71072
Hi-C data for G401	Lajoie et al. Methods, 2015	GSE105235
Capture Hi-C for naïve B cell, total B cell, megakaryocyte, eosinophil, erythroblast, neutrophil	Javierre et al., Cell, 2016	https://osf.io/u8tzp/
ChIA-PET for GM12878 and K562 cells	Heidari et al. Genome Res., 2014	GSE59395

ChIA-PET for NB4 cell	Li et al., Cell, 2012	GSE33664
Software	**Source**	**Link**
Samtools	Li et al., 2009	http://samtools.sourceforge.net/
GATK (v3.8)	McKenna et al., 2010	https://software.broadinstitute.org/gatk/
Freebayes (v1.0.2)	Garrison and Marth, 2012	https://github.com/ekg/freebayes
Delly2 (v0.7.2)	Rausch et al., 2012	https://github.com/dellytools/delly
Lumpy (v0.2.13)	Layer et al., 2014	https://github.com/arq5x/lumpy-sv
STAR (v2.7)	Dobin et al., 2013	https://github.com/alexdobin/STAR
Cufflinks (v2.2.1)	Mortazavi et al., 2008	http://cole-trapnell-lab.github.io/cufflinks
CSI-ANN	Firpi et al., 2010	https://github.com/tanlabcode/CSI-ANN
IM-PET	He et al., 2014	https://github.com/tanlabcode/IM-PET
Bowtie2 (v2.1.0)	Langmead et al., 2012	http://bowtie-bio.sourceforge.net/bowtie2/index.shtml
FIMO	Grant et al., 2011	http://meme-suite.org/doc/fimo.html
Cis-BP database	Weirauch et al., 2014	http://cisbp.ccbr.utoronto.ca/
Comet R package	Leiserson et al., 2015	https://bioconductor.org/packages/release/bioc/html/coMET.html

Table S2. Enhancer-promoter pairs that are disrupted by noncoding mutations.
See excel file.

Table S3. Known cancer-relevant functions of the top affected TFs in each cancer type.

TF	Cancer	Reported function (reference PMID)
LMO2	B-ALL	"HOX-mediated LMO2 expression in embryonic mesoderm is recapitulated in acute leukaemias" (23708655) "The Lmo2 oncogene initiates leukemia in mice by inducing thymocyte self-renewal" (20093438)
GLIS2	B-ALL	GLIS2 fusion in leukemia patients
RUNX1	B-ALL	ETV6-RUNX1 frequently occur in pediatric B-ALL
PAX5	B-ALL	"Pax5 loss imposes a reversible differentiation block in B-progenitor acute lymphoblastic leukemia" (24939936)
TCF3	B-ALL	TCF3-PBX1 frequently occur in pediatric B-ALL
MGA	AML	Frequently mutated in the AML patients with partial tandem duplication of MLL (MLL-PTD) (27389053)
KLF1	AML	KLF1 enhancer is found methylated in AML patients. (28749240) "EKLF/KLF1-regulated cell cycle exit is essential for erythroblast enucleation" (27480112)

GATA1	AML	“GATA1 mutations in acute leukemia in children with Down syndrome” (16631446)
RARG	AML	CPSF6-RARG, NUP98-RARG fusion in AML patients. (29568099)
CREB1	AML	“CREB Increases Chemotherapy Resistance through Regulation of the DNA Damage Repair Pathway in AML Cells” (15837624)
STAT2	AML	JAK-STAT in AML
RUNX1	AML	RUNX1-RUNX1T1 in AML
ELK1	NBL	“The presence of Elk-1 in the cytoplasm of neurons or neuroblastomas was confirmed.” (21441990)
KLF4	NBL	“KLF4 inhibits tumor growth in various cancers such as neuroblastoma and lung cancers by inducing the expression of CDK inhibitors and inhibiting cyclin D1 and FOXM1 expression.” (23045286)
WT1	NBL	“The WT1 may govern cell differentiation and suppress cell proliferation in NBL” (21292082)
CREB1	NBL	“CREB was specifically cleaved by caspases in neuroblastoma extracts, and in cells induced to undergo apoptosis by staurosporine.” (11119719)
ETV1	NBL	“The small molecule inhibitor YK-4-279 disrupts mitotic progression of neuroblastoma cells, overcomes drug resistance and synergizes with inhibitors of mitosis” (28602975)
PLAG1	WT	“PLAG1, the Main Translocation Target in Pleomorphic Adenoma of the Salivary Glands, Is a Positive Regulator of IGF-II” (10646861)
YY1	WT	“Survivin selective inhibitor YM155 induce apoptosis in SK-NEP-1 Wilms tumor cells” (23267699)
E2F4	WT	“Loss of heterozygosity on chromosome 16q increases relapse risk in Wilms’ tumor” (29029528)
PAX3	OS	“Pax3 induced a mesenchymal to epithelial transition (MET) in human SaOS-2 osteosarcomas” (15688035)
GLI2	OS	“GLI2 is a novel therapeutic target for metastasis of osteosarcoma” (25082385)

Table S4. PCR and sgRNA target sequences used in this study.

Name	Sequence (5' to 3')	Purpose
mrCHD4_F1	ATGGACGCACTTCTGAACAACA	
mrCHD4_R1	GCTTTGGAGTCTCTGCTTCG	
mrTBP_F2	GTAAACTTGACCTAAAGACCATTGC	
mrTBP_R2	ACGCAGTTGTCCGTGGCT	
mrNOP2_F1	CGAAAACAGAAGGGTGCGGAGAC	
mrNOP2_R1	ACGACTCGACAGCCTCTTGGA	
mrNOP2_F2	CGTCCCTAAGCCAAATAAGTCTCCT	RT-qPCR
mrNOP2_R2	TCTTCCTCCTCGTCCCCATCACT	
mrZFP384_F1	GCCTTCTATCCCCACAGTCTCAG	
mrZFP384_R1	CCACAGCCCTTCTCTGGCAACA	
mrZFP384_F2	CACACAGTGAAACACGCCAAGGT	
mrZFP384_R2	ATCAGGAGGGTTGTGTTTGCG	
mrIRF4_F1	GGGCAAGCAGGACTACAATCGT	

mrIRF4_R1	ATCCCTTCTCGGAACTTGCCTTTA	
mrIRF4_F2	GACCAGTCACACCCAGAAATCCCATA	
mrIRF4_R2	GGGACTCAGGTGGGGCACAAGCATA	
mrPAX5_F1	AGATGTAGTCCGCCAAAGGATAGTG	
mrPAX5_R1	CGGCTTGATGCTTCCTGTCTC	
mrPAX5_F2	GCCGACACCAACAAACGCAAG	
mrPAX5_R2	GCCATTCGGCACTGGAGACT	
mrEP300_F1	TCCAGAAGGAACTAGAAGAGAAACG	
mrEP300_R1	CCATGTTCGACCCAGTATTCATAGGA	
mrFlt3_F1	ACAGAGACCCAGGCAGGAGAATAC	
mrFlt3_R1	GCGTCCTGGTTTTCCATCTTCCTA	
mrFlt3_F2	TCAAAGCACCCCAGCCAGTCA	
mrFlt3_R2	TGACTGAGAAGCAGAACTTTTCGTAC	
mrTcf3_F1	ACTTCAGTGACTCCCACAGCAG	
mrTcf3_R1	CTCCCAAAGGTGGCATAGGCATTC	
mrTcf3_F2	CAGATACTCAGCCGAAGAAGGTCC	
mrTcf3_R2	ATCCCTGCTGTAGCTGTCACCT	
mrEbf1_F1	GCTGTGGCAACCGAAATGAGACT	
mrEbf1_R1	CACGTGGGTTTCCTGCATTCTTTAG	
mrEbf1_F2	TGTCCACAATAACTCCAAGCACGG	
mrEbf1_R2	GCTGATGGCTTTGATACAGGGAGT	
mpEP300E_F2	CCGAGCTCTTACGCGTGCTGCTTTTTCAGAAGACCT	
mpEP300E_R2	CCGGGCTAGCACGCGTGTAGAGTCTCCAAGATGGTAGT	
mpNCE_F1	CCGAGCTCTTACGCGTTGTTTGTAGTTTGAGTTCCAC	
mpNCE_R1	CCGGGCTAGCACGCGTTTGAACCAAATGCCTACCT	
hpNCE_F1	CCGAGCTCTTACGCGTCTTTTATTCTGTGGGTTGTCTC	PCR
hpNCE_R1	CCGGGCTAGCACGCGTCTACCACTCCCAAAGCATAA	
mpT418_F1	CCCTAAACCTTTGTTGTCAGA	
mpT486_R1	TGACCCAAACTTAGCACATC	
mpT418_R2	TAACCATAATGTGTTTCATCCTCC	
CHD4_1	GAAGGGGATGGCGTCGGGCC	
CHD4_2	TTCCGGCGCGCCGAGTCCTT	
CHD4_3	AGGTGGTGGTGCAACCTCAG	
non-targeting control_1	GCCCCGCCGCCCTCCCCTCC	sgRNA
non-targeting control_2	CCAGTTGCTCTGGGGGAACA	
non-targeting control	TCAGCAAAGGACGAAACAAA	
Translocation_1	CATGCAAACGCACACCATAC *AGG*	
Translocation_2	CAGTATTCTAAGTTACGGGC *AGG*	

REFERENCES AND NOTES

1. X. Ma, Y. Liu, Y. Liu, L. B. Alexandrov, M. N. Edmonson, C. Gawad, X. Zhou, Y. Li, M. C. Rusch, J. Easton, R. Huether, V. Gonzalez-Pena, M. R. Wilkinson, L. C. Hermida, S. Davis, E. Sioson, S. Pounds, X. Cao, R. E. Ries, Z. Wang, X. Chen, L. Dong, S. J. Diskin, M. A. Smith, J. M. Guidry Auvil, P. S. Meltzer, C. C. Lau, E. J. Perlman, J. M. Maris, S. Meshinchi, S. P. Hunger, D. S. Gerhard, J. Zhang, Pan-cancer genome and transcriptome analyses of 1,699 paediatric leukaemias and solid tumours. *Nature* **555**, 371–376 (2018).

2. S. N. Gröbner, B. C. Worst, J. Weischenfeldt, I. Buchhalter, K. Kleinheinz, V. A. Rudneva, P. D. Johann, G. P. Balasubramanian, M. Segura-Wang, S. Brabetz, S. Bender, B. Hutter, D. Sturm, E. Pfaff, D. Hübschmann, G. Zipprich, M. Heinold, J. Eils, C. Lawerenz, S. Erkek, S. Lambo, S. Waszak, C. Blattmann, A. Borkhardt, M. Kuhlen, A. Eggert, S. Fulda, M. Gessler, J. Wegert, R. Kappler, D. Baumhoer, S. Burdach, R. Kirschner-Schwabe, U. Kontny, A. E. Kulozik, D. Lohmann, S. Hettmer, C. Eckert, S. Bielack, M. Nathrath, C. Niemeyer, G. H. Richter, J. Schulte, R. Siebert, F. Westermann, J. J. Molenaar, G. Vassal, H. Witt; ICGC Ped Brain-Seq Project; ICGC MMML-Seq Project, B. Burkhardt, C. P. Kratz, O. Witt, C. M. van Tilburg, C. M. Kramm, G. Fleischhack, U. Dirksen, S. Rutkowski, M. Frühwald, Katja von Hoff, S. Wolf, T. Klingebiel, E. Koscielniak, P. Landgraf, J. Koster, A. C. Resnick, J. Zhang, Y. Liu, X. Zhou, A. J. Waanders, D. A. Zwijnenburg, P. Raman, B. Brors, U. D. Weber, P. A. Northcott, K. W. Pajtler, M. Kool, R. M. Piro, J. O. Korbel, M. Schlesner, R. Eils, D. T. W. Jones, P. Lichter, L. Chavez, M. Zapatka, S. M. Pfister, The landscape of genomic alterations across childhood cancers. *Nature* **555**, 321–327 (2018).

3. F. K. Lorbeer, D. Hockemeyer, TERT promoter mutations and telomeres during tumorigenesis. *Curr. Opin. Genet. Dev.* **60**, 56–62 (2020).

4. D. A. Oldridge, A. C. Wood, N. Weichert-Leahey, I. Crimmins, R. Sussman, C. Winter, L. D. McDaniel, M. Diamond, L. S. Hart, S. Zhu, A. D. Durbin, B. J. Abraham, L. Anders, L. Tian, S. Zhang, J. S. Wei, J. Khan, K. Bramlett, N. Rahman, M. Capasso, A. Iolascon, D. S. Gerhard, J. M. Guidry Auvil, R. A. Young, H. Hakonarson, S. J. Diskin, A. Thomas Look, J. M. Maris, Genetic predisposition to neuroblastoma mediated by a LMO1 super-enhancer polymorphism. *Nature* **528**, 418–421 (2015).

5. M. R. Mansour, B. J. Abraham, L. Anders, A. Berezovskaya, A. Gutierrez, A. D. Durbin, J. Etchin, L. Lawton, S. E. Sallan, L. B. Silverman, M. L. Loh, S. P. Hunger, T. Sanda, R. A. Young, A. T. Look, Oncogene regulation. An oncogenic super-enhancer formed through somatic mutation of a noncoding intergenic element. *Science* **346**, 1373–1377 (2014).

6. International Cancer Genome Consortium, International network of cancer genome projects. *Nature* **464**, 993–998 (2010).

7. P. A. Northcott, C. Lee, T. Zichner, A. M. Stütz, S. Erkek, D. Kawauchi, D. J. H. Shih, V. Hovestadt, M. Zapatka, D. Sturm, D. T. W. Jones, M. Kool, M. Remke, F. M. G. Cavalli, S. Zuyderduyn, G. D. Bader, S. VandenBerg, L. A. Esparza, M. Ryzhova, W. Wang, A. Wittmann, S. Stark, L. Sieber, H. Seker-Cin, L. Linke, F. Kratochwil, N. Jäger, I. Buchhalter, C. D. Imbusch, G. Zipprich, B. Raeder, S. Schmidt, N. Diessl, S. Wolf, S. Wiemann, B. Brors, C. Lawerenz, J. Eils, H.-J. Warnatz, T. Risch, M.-L. Yaspo, U. D. Weber, C. C. Bartholomae, C. von Kalle, E. Turányi, P. Hauser, E. Sanden, A. Darabi, P. Siesjö, J. Sterba, K. Zitterbart, D. Sumerauer, P. van Sluis, R. Versteeg, R. Volckmann, J. Koster, M. U. Schuhmann, M. Ebinger, H. L. Grimes, G. W. Robinson, A. Gajjar, M. Mynarek, K. von Hoff, S. Rutkowski, T. Pietsch, W. Scheurlen, J. Felsberg, G. Reifenberger, A. E. Kulozik, A. von Deimling, O. Witt, R. Eils, R. J. Gilbertson, A. Korshunov, M. D. Taylor, P. Lichter, J. O. Korbel, R. J. Wechsler-Reya, S. M. Pfister, Enhancer hijacking activates GFI1 family oncogenes in medulloblastoma. *Nature* **511**, 428–434 (2014).

8. R. J. H. Ryan, Y. Drier, H. Whitton, M. J. Cotton, J. Kaur, R. Issner, S. Gillespie, C. B. Epstein, V. Nardi, A. R. Sohani, E. P. Hochberg, B. E. Bernstein, Detection of enhancer-associated rearrangements reveals mechanisms of oncogene dysregulation in B-cell lymphoma. *Cancer Discov.* **5**, 1058–1071 (2015).

9. S. K. Tasian, M. L. Loh, Understanding the biology of CRLF2-overexpressing acute lymphoblastic leukemia. *Crit. Rev. Oncog.* **16**, 13–24 (2011).

10. J. M. Zook, B. Chapman, J. Wang, D. Mittelman, O. Hofmann, W. Hide, M. Salit, Integrating human sequence data sets provides a resource of benchmark SNP and indel genotype calls. *Nat. Biotechnol.* **32**, 246–251 (2014).

11. J. A. Perry, A. Kiezun, P. Tonzi, E. M. Van Allen, S. L. Carter, S. C. Baca, G. S. Cowley, A. S. Bhatt, E. Rheinbay, C. S. Pedamallu, E. Helman, A. Taylor-Weiner, A. McKenna, D. S. DeLuca, M. S. Lawrence, L. Ambrogio, C. Sougnez, A. Sivachenko, L. D. Walensky, N. Wagle, J. Mora, C. de Torres, C. Lavarino, S. Dos Santos Aguiar, J. A. Yunes, S. R. Brandalise, G. E. Mercado-Celis, J. Melendez-Zajgla, R. Cárdenas-Cardós, L. Velasco-Hidalgo, C. W. M. Roberts, L. A. Garraway, C. Rodriguez-Galindo, S. B. Gabriel, E. S. Lander, T. R. Golub, S. H. Orkin, G. Getz, K. A. Janeway, Complementary genomic approaches highlight the PI3K/mTOR pathway as a common vulnerability in osteosarcoma. *Proc. Natl. Acad. Sci. U.S.A.* **111**, E5564–E5573 (2014).

12. K. G. Roberts, Y. Li, D. Payne-Turner, R. C. Harvey, Y.-L. Yang, D. Pei, K. McCastlain, L. Ding, C. Lu, G. Song, J. Ma, J. Becksfort, M. Rusch, S.-C. Chen, J. Easton, J. Cheng, K. Boggs, N. Santiago-Morales, I. Iacobucci, R. S. Fulton, J. Wen, M. Valentine, C. Cheng, S. W. Paugh, M. Devidas, I.-M. Chen, S. Reshmi, A. Smith, E. Hedlund, P. Gupta, P. Nagahawatte, G. Wu, X. Chen, D. Yergeau, B. Vadodaria, H. Mulder, N. J. Winick, E. C. Larsen, W. L. Carroll, N. A. Heerema, A. J. Carroll, G. Grayson, S. K. Tasian, A. S. Moore, F. Keller, M. Frei-Jones, J. A. Whitlock, E. A. Raetz, D. L. White, T. P. Hughes, J. M. Guidry Auvil, M. A. Smith, G. Marcucci, C. D. Bloomfield, K. Mrózek, J. Kohlschmidt, W. Stock, S. M. Kornblau, M. Konopleva, E. Paietta, C.-H. Pui, S. Jeha, M. V. Relling, W. E. Evans, D. S. Gerhard, J. M. Gastier-Foster, E. Mardis, R. K. Wilson, M. L. Loh, J. R. Downing, S. P. Hunger, C. L. Willman, J. Zhang, C. G. Mullighan, Targetable kinase-activating lesions in Ph-like acute lymphoblastic leukemia. *N. Engl. J. Med.* **371**, 1005–1015 (2014).

13. H. A. Firpi, D. Ucar, K. Tan, Discover regulatory DNA elements using chromatin signatures and artificial neural network. *Bioinformatics* **26**, 1579–1586 (2010).

14. B. He, C. Chen, L. Teng, K. Tan, Global view of enhancer-promoter interactome in human cells. *Proc. Natl. Acad. Sci. U.S.A.* **111**, E2191–E2199 (2014).

15. A. Khan, X. Zhang, dbSUPER: A database of super-enhancers in mouse and human genome. *Nucleic Acids Res.* **44**, D164–D171 (2016).

16. Y. Drier, M. J. Cotton, K. E. Williamson, S. M. Gillespie, R. J. H. Ryan, M. J. Kluk, C. D. Carey, S. J. Rodig, L. M. Sholl, A. H. Afrogheh, W. C. Faquin, L. Queimado, J. Qi, M. J. Wick, A. K. El-Naggar, J. E. Bradner, C. A. Moskaluk, J. C. Aster, B. Knoechel, B. E. Bernstein, An oncogenic

MYB feedback loop drives alternate cell fates in adenoid cystic carcinoma. *Nat. Genet.* **48**, 265–272 (2016).

17. M. Peifer, F. Hertwig, F. Roels, D. Dreidax, M. Gartlgruber, R. Menon, A. Krämer, J. L. Roncaioli, F. Sand, J. M. Heuckmann, F. Ikram, R. Schmidt, S. Ackermann, A. Engesser, Y. Kahlert, W. Vogel, J. Altmüller, P. Nürnberg, J. Thierry-Mieg, D. Thierry-Mieg, A. Mariappan, S. Heynck, E. Mariotti, K.-O. Henrich, C. Gloeckner, G. Bosco, I. Leuschner, M. R. Schweiger, L. Savelyeva, S. C. Watkins, C. Shao, E. Bell, T. Höfer, V. Achter, U. Lang, J. Theissen, R. Volland, M. Saadati, A. Eggert, B. de Wilde, F. Berthold, Z. Peng, C. Zhao, L. Shi, M. Ortmann, R. Büttner, S. Perner, B. Hero, A. Schramm, J. H. Schulte, C. Herrmann, R. J. O'Sullivan, F. Westermann, R. K. Thomas, M. Fischer, Telomerase activation by genomic rearrangements in high-risk neuroblastoma. *Nature* **526**, 700–704 (2015).

18. M. Qian, H. Zhang, S. K.-Y. Kham, S. Liu, C. Jiang, X. Zhao, Y. Lu, C. Goodings, T.-N. Lin, R. Zhang, T. Moriyama, Z. Yin, Z. Li, T. C. Quah, H. Ariffin, A. M. Tan, S. Shen, D. Bhojwani, S. Hu, S. Chen, H. Zheng, C.-H. Pui, A. E.-J. Yeoh, J. J. Yang, Whole-transcriptome sequencing identifies a distinct subtype of acute lymphoblastic leukemia with predominant genomic abnormalities of EP300 and CREBBP. *Genome Res.* **27**, 185–195 (2017).

19. T. B. Alexander, Z. Gu, I. Iacobucci, K. Dickerson, J. K. Choi, B. Xu, D. Payne-Turner, H. Yoshihara, M. L. Loh, J. Horan, B. Buldini, G. Basso, S. Elitzur, V. de Haas, C. M. Zwaan, A. Yeoh, D. Reinhardt, D. Tomizawa, N. Kiyokawa, T. Lammens, B. de Moerloose, D. Catchpoole, H. Hori, A. Moorman, A. S. Moore, O. Hrusak, S. Meshinchi, E. Orgel, M. Devidas, M. Borowitz, B. Wood, N. A. Heerema, A. Carrol, Y.-L. Yang, M. A. Smith, T. M. Davidsen, L. C. Hermida, P. Gesuwan, M. A. Marra, Y. Ma, A. J. Mungall, R. A. Moore, S. J. M. Jones, M. Valentine, L. J. Janke, J. E. Rubnitz, C.-H. Pui, L. Ding, Y. Liu, J. Zhang, K. E. Nichols, J. R. Downing, X. Cao, L. Shi, S. Pounds, S. Newman, D. Pei, J. M. Guidry Auvil, D. S. Gerhard, S. P. Hunger, H. Inaba, C. G. Mullighan, The genetic basis and cell of origin of mixed phenotype acute leukaemia. *Nature* **562**, 373–379 (2018).

20. C. Dege, J. Hagman, Mi-2/NuRD chromatin remodeling complexes regulate B and T-lymphocyte development and function. *Immunol. Rev.* **261**, 126–140 (2014).

21. J. Sperlazza, M. Rahmani, J. Beckta, M. Aust, E. Hawkins, S. Z. Wang, S. Zu Zhu, S. Podder, C. Dumur, K. Archer, S. Grant, G. D. Ginder, Depletion of the chromatin remodeler CHD4 sensitizes AML blasts to genotoxic agents and reduces tumor formation. *Blood* **126**, 1462–1472 (2015).

22. T. J. Pugh, O. Morozova, E. F. Attiyeh, S. Asgharzadeh, J. S. Wei, D. Auclair, S. L. Carter, K. Cibulskis, M. Hanna, A. Kiezun, J. Kim, M. S. Lawrence, L. Lichenstein, A. McKenna, C. S. Pedamallu, A. H. Ramos, E. Shefler, A. Sivachenko, C. Sougnez, C. Stewart, A. Ally, I. Birol, R. Chiu, R. D. Corbett, M. Hirst, S. D. Jackman, B. Kamoh, A. H. Khodabakshi, M. Krzywinski, A. Lo, R. A. Moore, K. L. Mungall, J. Qian, A. Tam, N. Thiessen, Y. Zhao, K. A. Cole, M. Diamond, S. J. Diskin, Y. P. Mosse, A. C. Wood, L. Ji, R. Sposto, T. Badgett, W. B. London, Y. Moyer, J. M. Gastier-Foster, M. A. Smith, J. M. G. Auvil, D. S. Gerhard, M. D. Hogarty, S. J. M. Jones, E. S. Lander, S. B. Gabriel, G. Getz, R. C. Seeger, J. Khan, M. A. Marra, M. Meyerson, J. M. Maris, The genetic landscape of high-risk neuroblastoma. *Nat. Genet.* **45**, 279–284 (2013).

23. L. Zhuang, Y. Ma, Q. Wang, J. Zhang, C. Zhu, L. Zhang, X. Xu, Atg3 overexpression enhances bortezomib-induced cell death in SKM-1 cell. *PLOS ONE* **11**, e0158761 (2016).

24. V. Moignard, I. C. Macaulay, G. Swiers, F. Buettner, J. Schütte, F. J. Calero-Nieto, S. Kinston, A. Joshi, R. Hannah, F. J. Theis, S. E. Jacobsen, M. F. de Bruijn, B. Göttgens, Characterization of transcriptional networks in blood stem and progenitor cells using high-throughput single-cell gene expression analysis. *Nat. Cell Biol.* **15**, 363–372 (2013).

25. L. Vassen, H. Beauchemin, W. Lemsaddek, J. Krongold, M. Trudel, T. Möröy, Growth factor independence 1b (gfi1b) is important for the maturation of erythroid cells and the regulation of embryonic globin expression. *PLOS ONE* **9**, e96636 (2014).

26. A. Thivakaran, L. Botezatu, J. M. Hönes, J. Schütte, L. Vassen, Y. S. Al-Matary, P. Patnana, A. Zeller, M. Heuser, F. Thol, R. Gabdoulline, N. Olberding, D. Frank, M. Suslo, R. Köster, K. Lennartz, A. Görgens, B. Giebel, B. Opalka, U. Dührsen, C. Khandanpour, Gfi1b: A key player in the genesis and maintenance of acute myeloid leukemia and myelodysplastic syndrome. *Haematologica* **103**, 614–625 (2018).

27. D. Hnisz, B. J. Abraham, T. I. Lee, A. Lau, V. Saint-André, A. A. Sigova, H. A. Hoke, R. A. Young, Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934–947 (2013).

28. F. Damm, F. Thol, I. Hollink, M. Zimmermann, K. Reinhardt, M. M. van den Heuvel-Eibrink, C. M. Zwaan, V. de Haas, U. Creutzig, J.-H. Klusmann, J. Krauter, M. Heuser, A. Ganser, D. Reinhardt, C. Thiede, Prevalence and prognostic value of IDH1 and IDH2 mutations in childhood AML: A study of the AML-BFM and DCOG study groups. *Leukemia* **25**, 1704–1710 (2011).

29. J. L. Patel, J. A. Schumacher, K. Frizzell, S. Sorrells, W. Shen, A. Clayton, R. Jattani, T. W. Kelley, Coexisting and cooperating mutations in NPM1-mutated acute myeloid leukemia. *Leuk. Res.* **56**, 7–12 (2017).

30. S. Gröschel, M. A. Sanders, R. Hoogenboezem, E. de Wit, B. A. M. Bouwman, C. Erpelinck, V. H. J. van der Velden, M. Havermans, R. Avellino, K. van Lom, E. J. Rombouts, M. van Duin, K. Döhner, H. B. Beverloo, J. E. Bradner, H. Döhner, B. Löwenberg, P. J. M. Valk, E. M. J. Bindels, W. de Laat, R. Delwel, A single oncogenic enhancer rearrangement causes concomitant EVI1 and GATA2 deregulation in leukemia. *Cell* **157**, 369–381 (2014).

31. J. Sima, D. M. Gilbert, Complex correlations: Replication timing and mutational landscapes during cancer and genome evolution. *Curr. Opin. Genet. Dev.* **25**, 93–100 (2014).

32. S. K. Tasian, S. P. Hunger, Genomic characterization of paediatric acute lymphoblastic leukaemia: An opportunity for precision medicine therapeutics. *Br. J. Haematol.* **176**, 867–882 (2017).

33. M. Pieraccioli, S. Nicolai, C. Pitolli, M. Agostini, A. Antonov, M. Malewicz, R. A. Knight, G. Raschellà, G. Melino, ZNF281 inhibits neuronal differentiation and is a prognostic marker for neuroblastoma. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 7356–7361 (2018).

34. J. Dzieran, A. Rodriguez Garcia, U. K. Westermark, A. B. Henley, E. Eyre Sánchez, C. Träger, H. J. Johansson, J. Lehtiö, M. Arsenian-Henriksson, MYCN-amplified neuroblastoma maintains an aggressive and undifferentiated phenotype by deregulation of estrogen and NGF signaling. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E1229–E1238 (2018).

35. Z. J. Faber, X. Chen, A. L. Gedman, K. Boggs, J. Cheng, J. Ma, I. Radtke, J.-R. Chao, M. P. Walsh, G. Song, A. K. Andersson, J. Dang, L. Dong, Y. Liu, R. Huether, Z. Cai, H. Mulder, G. Wu, M. Edmonson, M. Rusch, C. Qu, Y. Li, B. Vadodaria, J. Wang, E. Hedlund, X. Cao, D. Yergeau, J. Nakitandwe, S. B. Pounds, S. Shurtleff, R. S. Fulton, L. L. Fulton, J. Easton, E. Parganas, C.-H. Pui, J. E. Rubnitz, L. Ding, E. R. Mardis, R. K. Wilson, T. A. Gruber, C. G. Mullighan, R. F. Schlenk, P. Paschka, K. Döhner, H. Döhner, L. Bullinger, J. Zhang, J. M. Klco, J. R. Downing, The genomic landscape of core-binding factor acute myeloid leukemias. *Nat. Genet.* **48**, 1551–1556 (2016).

36. T. Holmlund, M. J. Lindberg, D. Grander, A. E. Wallberg, GCN5 acetylates and regulates the stability of the oncoprotein E2A-PBX1 in acute lymphoblastic leukemia. *Leukemia* **27**, 578–585 (2013).

37. L. De Cecco, M. Capaia, S. Zupo, G. Cutrona, S. Matis, A. Brizzolara, A. M. Orengo, M. Croce, E. Marchesi, M. Ferrarini, S. Canevari, S. Ferrini, Interleukin 21 controls mRNA and microRNA expression in CD40-activated chronic lymphocytic leukemia cells. *PLOS ONE* **10**, e0134706 (2015).

38. E. Izycka-Swieszewska, A. Wozniak, J. Kot, W. Grajkowska, A. Balcerska, D. Perek, B. Dembowska-Baginska, T. Klepacka, E. Drozynska, Prognostic significance of HER2 expression in neuroblastic tumors. *Modern Pathol.* **23**, 1261–1268 (2010).

39. H. Thomas, K. Beck, M. Adamczyk, P. Aeschlimann, M. Langley, R. C. Oita, L. Thiebach, M. Hils, D. Aeschlimann, Transglutaminase 6: A protein associated with central nervous system development and motor function. *Amino Acids* **44**, 161–177 (2013).

40. A. S. Algarni, A. J. Hargreaves, J. M. Dickenson, Activation of transglutaminase 2 by nerve growth factor in differentiating neuroblastoma cells: A role in cell survival and neurite outgrowth. *Eur. J. Pharmacol.* **820**, 113–129 (2018).

41. N. Stransky, E. Cerami, S. Schalm, J. L. Kim, C. Lengauer, The landscape of kinase fusions in cancer. *Nat. Commun.* **5**, 4846 (2014).

42. L. Wang, Y. Sun, Y. Sun, L. Meng, X. Xu, First case of AML with rare chromosome translocations: A case report of twins. *BMC Cancer* **18**, 458 (2018).

43. G. Wu, Z. Ma, W. Hu, D. Wang, B. Gong, C. Fan, S. Jiang, T. Li, J. Gao, Y. Yang, Molecular insights of Gas6/TAM in cancer development and therapy. *Cell Death Dis.* **8**, e2700 (2017).

44. L. Yang, L. J. Luquette, N. Gehlenborg, R. Xi, P. S. Haseley, C.-H. Hsieh, C. Zhang, X. Ren, A. Protopopov, L. Chin, R. Kucherlapati, C. Lee, P. J. Park, Diverse mechanisms of somatic structural variations in human cancer genomes. *Cell* **153**, 919–929 (2013).

45. I. Janoueix-Lerosey, P. Hupé, Z. Maciorowski, P. La Rosa, G. Schleiermacher, G. Pierron, S. Liva, E. Barillot, O. Delattre, Preferential occurrence of chromosome breakpoints within early replicating regions in neuroblastoma. *Cell Cycle* **4**, 1842–1846 (2005).

46. P. Kim, F. Cheng, J. Zhao, Z. Zhao, ccmGDB: A database for cancer cell metabolism genes. *Nucleic Acids Res.* **44**, D959–D968 (2016).

47. R. Torres, M. C. Martin, A. Garcia, J. C. Cigudosa, J. C. Ramirez, S. Rodriguez-Perales, Engineering human tumour-associated chromosomal translocations with the RNA-guided CRISPR-Cas9 system. *Nat. Commun.* **5**, 3964 (2014).

48. M. T. Weirauch, A. Yang, M. Albu, A. G. Cote, A. Montenegro-Montero, P. Drewe, H. S. Najafabadi, S. A. Lambert, I. Mann, K. Cook, H. Zheng, A. Goity, H. van Bakel, J.-C. Lozano, M. Galli, M. G. Lewsey, E. Huang, T. Mukherjee, X. Chen, J. S. Reece-Hoyes, S. Govindarajan, G. Shaulsky, A. J. M. Walhout, F.-Y. Bouget, G. Ratsch, L. F. Larrondo, J. R. Ecker, T. R. Hughes, Determination and inference of eukaryotic transcription factor sequence specificity. *Cell* **158**, 1431–1443 (2014).

49. W. Zhang, A. Bojorquez-Gomez, D. O. Velez, G. Xu, K. S. Sanchez, J. P. Shen, K. Chen, K. Licon, C. Melton, K. M. Olson, M. K. Yu, J. K. Huang, H. Carter, E. K. Farley, M. Snyder, S. I. Fraley, J. F. Kreisberg, T. Ideker, A global transcriptional network connecting noncoding mutations to changes in tumor gene expression. *Nat. Genet.* **50**, 613–620 (2018).

50. N. Weddington, A. Stuy, I. Hiratani, T. Ryba, T. Yokochi, D. M. Gilbert, ReplicationDomain: A visualization tool and comparative database for genome-wide replication timing data. *BMC Bioinformatics* **9**, 530 (2008).

51. R. S. Hansen, S. Thomas, R. Sandstrom, T. K. Canfield, R. E. Thurman, M. Weaver, M. O. Dorschner, S. M. Gartler, J. A. Stamatoyannopoulos, Sequencing newly replicated DNA reveals widespread plasticity in human replication timing. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 139–144 (2010).

52. C. E. Grant, T. L. Bailey, W. S. Noble, FIMO: Scanning for occurrences of a given motif. *Bioinformatics* **27**, 1017–1018 (2011).

53. M. D. M. Leiserson, H.-T. Wu, F. Vandin, B. J. Raphael, CoMEt: A statistical approach to identify combinations of mutually exclusive alterations in cancer. *Genome Biol.* **16**, 160 (2015).

54. T. Juven-Gershon, S. Cheng, J. T. Kadonaga, Rational design of a super core promoter that enhances gene expression. *Nat. Methods* **3**, 917–922 (2006).

55. N. E. Sanjana, O. Shalem, F. Zhang, Improved vectors and genome-wide libraries for CRISPR screening. *Nat. Methods* **11**, 783–784 (2014).

56. J. G. Doench, N. Fusi, M. Sullender, M. Hegde, E. W. Vaimberg, K. F. Donovan, I. Smith, Z. Tothova, C. Wilen, R. Orchard, H. W. Virgin, J. Listgarten, D. E. Root, Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat. Biotechnol.* **34**, 184–191 (2016).