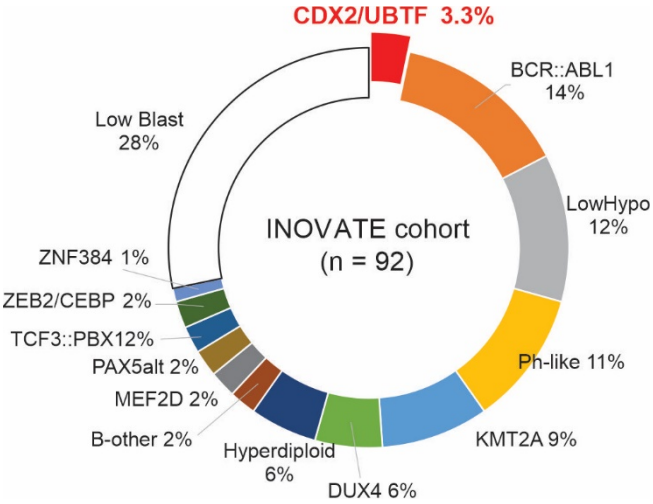
Enhancer retargeting of *CDX2* and *UBTF::ATXN7L3* define a subtype of high-risk B-progenitor acute lymphoblastic leukemia

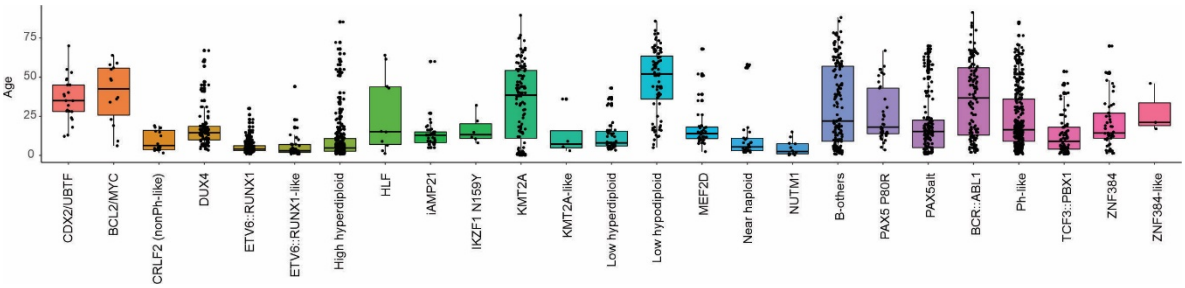
Shunsuke Kimura, Lindsey Montefiori, Ilaria Iacobucci, Yaqi Zhao, Qingsong Gao, Elisabeth M Paietta, Claudia Haferlach, A. Douglas Laird, Paul E. Mead, Zhahoui Gu, Wendy Stock, Mark Litzow, Jacob M. Rowe, Selina M. Luger, Stephen P. Hunger, Georgina L. Ryland, Breon Schmidt, Paul G. Ekert, Alicia Oshlack, Sean M. Grimmond, Jacqueline Rehn, James Breen, David Yeung, Deborah L. White, Ibrahim Aldoss, Elias J. Jabbour, Ching-Hon Pui, Manja Meggendorfer, Wencke Walter, Wolfgang Kern, Torsten Haferlach, Samuel Brady, Jinghui Zhang, Kathryn G. Roberts, Piers Blombery, Charles G. Mullighan

SUPPLEMENTARY FIGURES



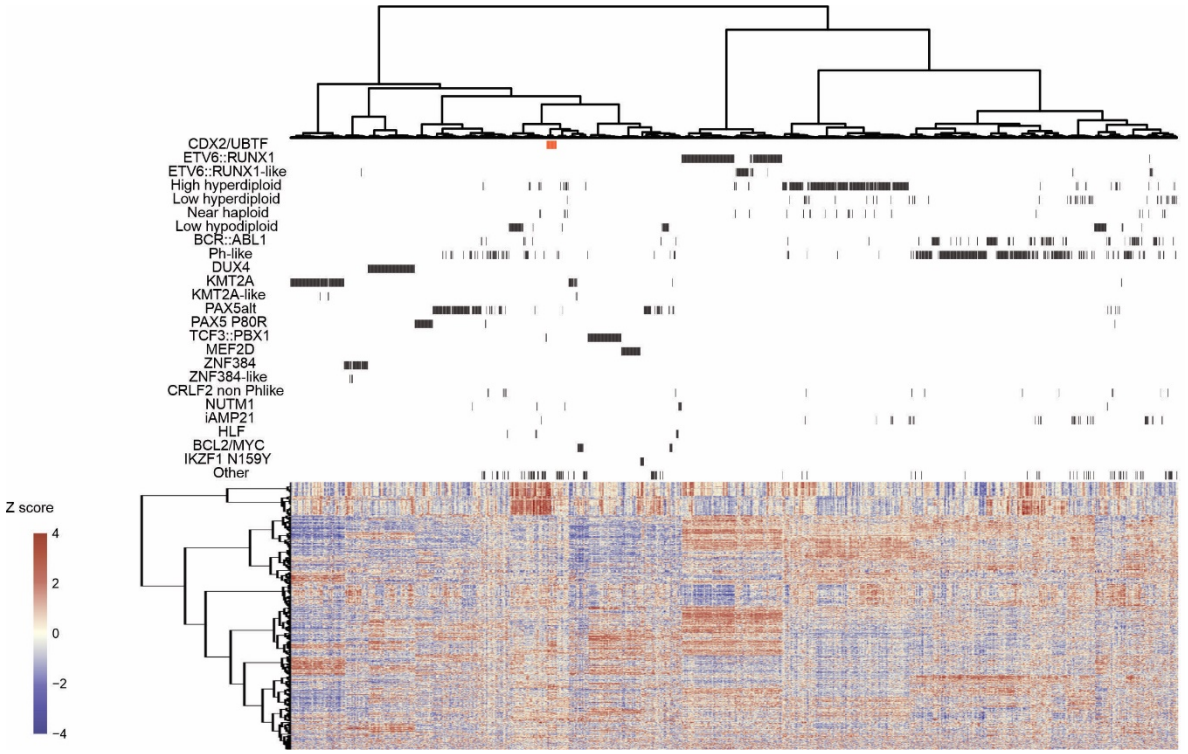
Supplementary Figure 1

Distribution of B-ALL subtypes included in the INOVATE cohort (n = 92) is shown. Only the cases with available RNA-seq are shown. CDX2/UBTF subtype is enriched in relapse/refractory cohort (3.3%).



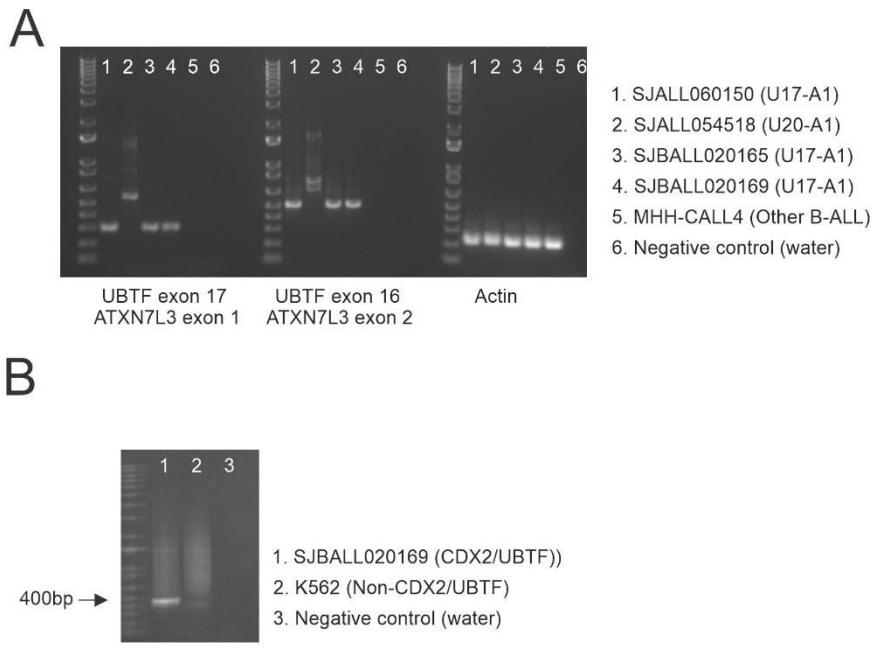
Supplementary Figure 2

Age distribution of each B-ALL subtype from the following institutions is shown; ECOG-ACRIN (n = 731), COG (n = 862), St. Jude (n = 501), and MLL (n = 213).



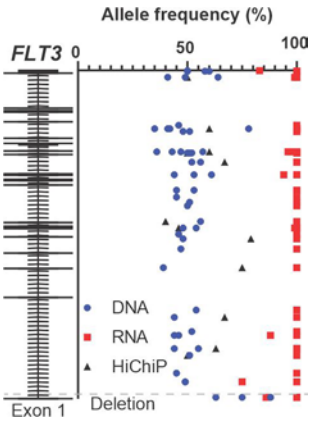
Supplementary Figure 3

Heatmap of gene expression profile from RNA-seq of 2,004 cases including 22 cases with CDX2/UBTF analyzed by hierarchical clustering with Pearson correlation and Ward's clustering method based on the 1,000 most variable genes (evaluated by median absolute deviation) is shown. B-ALL subtypes are annotated at the top of the heatmap. CDX2/UBTF subtype cases were closely clustered (highlighted in red).



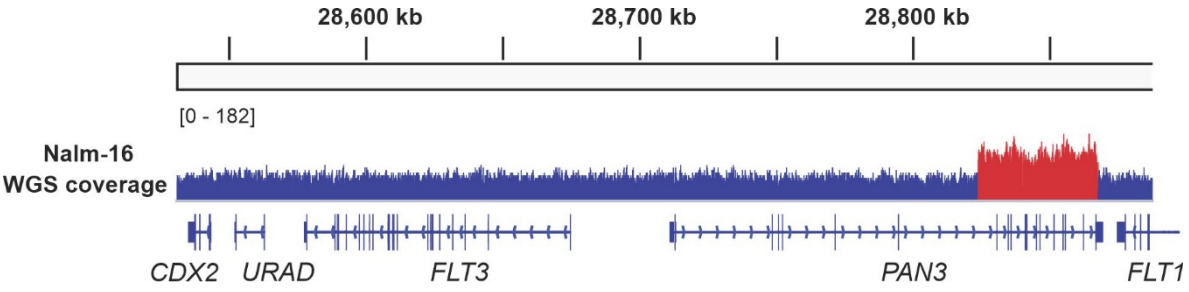
Supplementary Figure 4

(A) Validation of *UBTF::ATXN7L3* fusions by reverse transcription PCR. Two types of *UBTF::ATXN7L3* fusions (exon 17 and exon 20) are shown. Only CDX2/UBTF cases have *UBTF::ATXN7L3* fusions. (B) Validation of deletion at *UBTF* region by genomic PCR. Primers were designed outside of deletion (see Supplementary Methods). Only the CDX2/UBTF case showed the *UBTF* deletion.



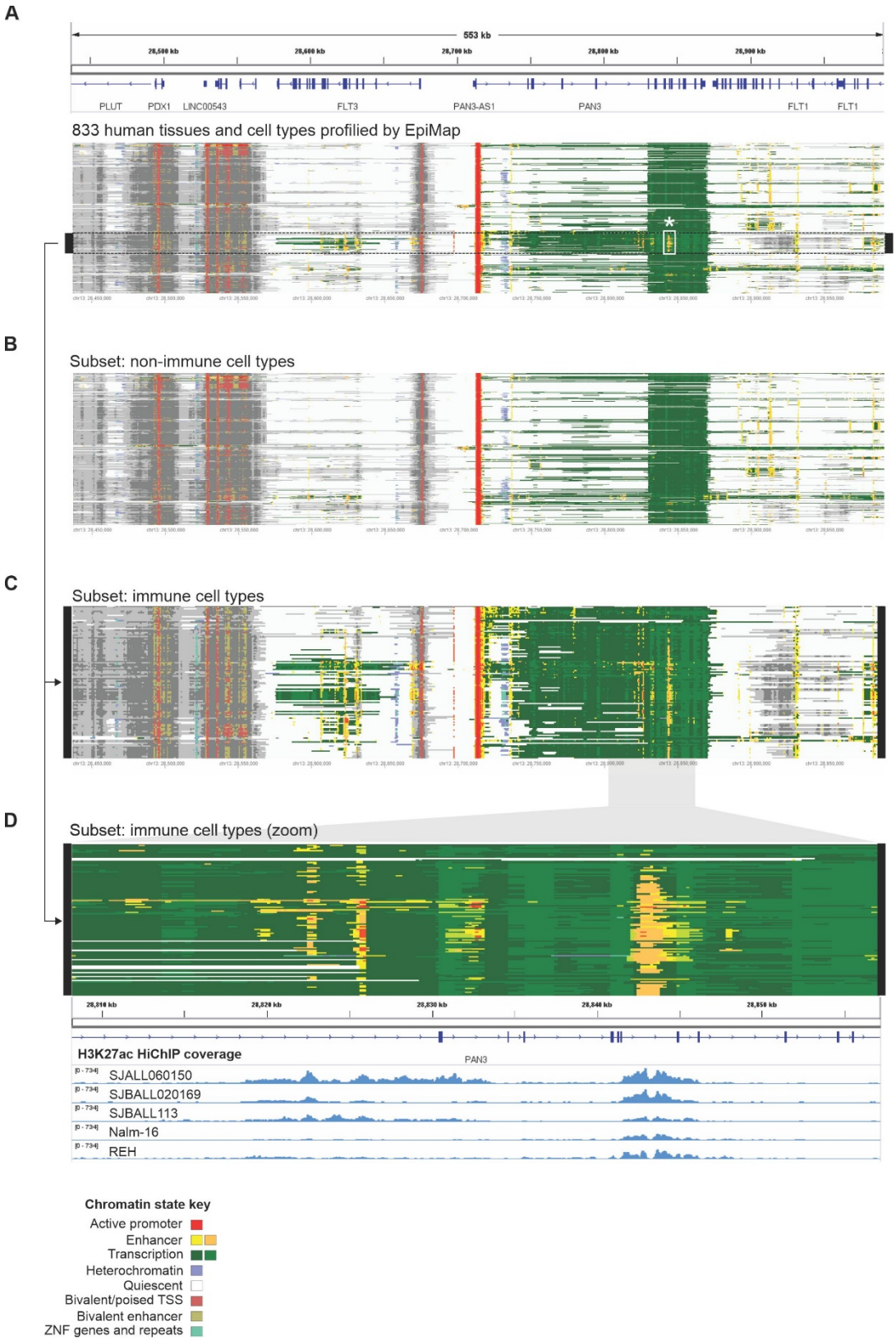
Supplementary Figure 5

Allele-specific expression of *FLT3* is confirmed by comparison of allele frequencies at bases of heterozygous single nucleotide polymorphism among RNA-seq (red), whole genome sequencing (blue), and HiChIP (black).



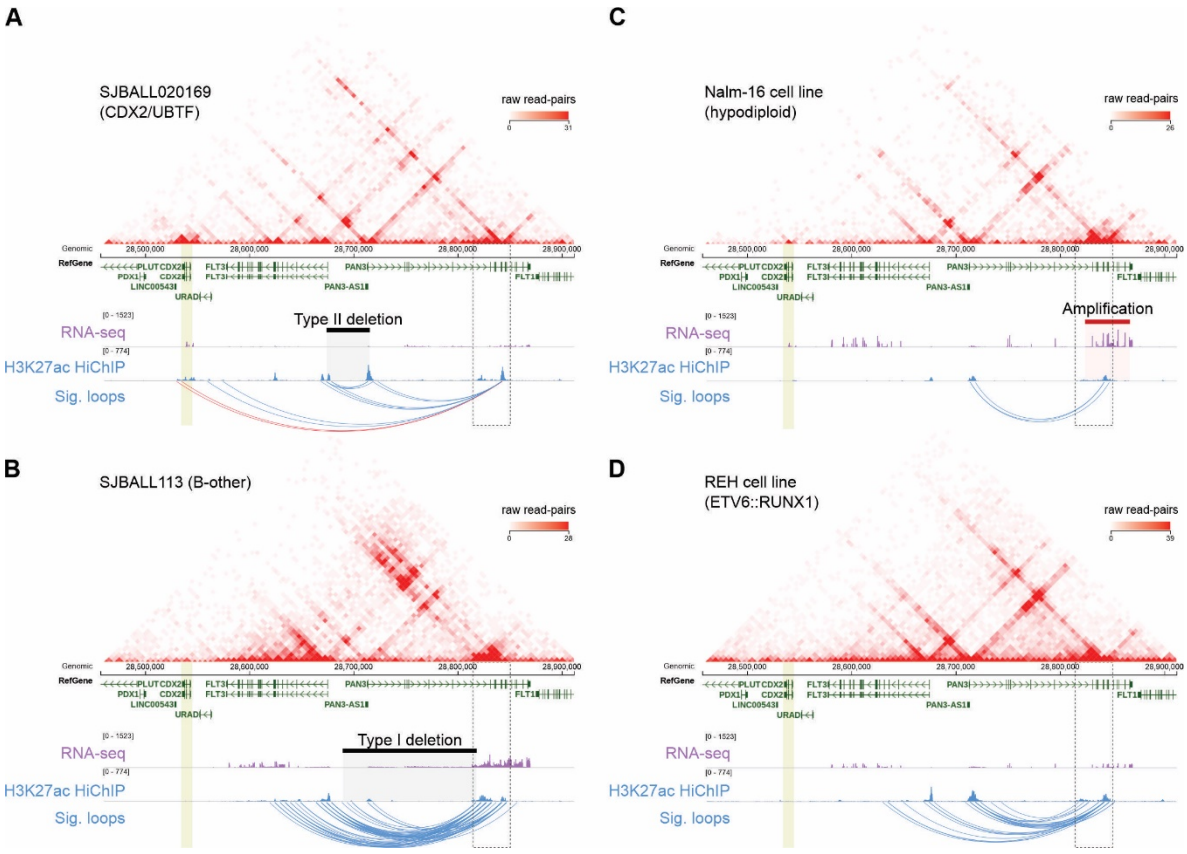
Supplement Figure 6

The coverage track of whole genome sequencing for Nalm-16 (hypodiploid) cell line. Amplification of the enhancer in the *PAN3* gene is marked as red.


A
553 kb
28,500 kb
28,600 kb
28,700 kb
28,800 kb
28,900 kb
PLUT
PDX1
LINC00543
FLT3
PAN3-AS1
PAN3
FLT1
FLT1
833 human tissues and cell types profilied by EpiMap
B
Subset: non-immune cell types
C
Subset: immune cell types
D
Subset: immune cell types (zoom)
28,810 kb
28,820 kb
28,830 kb
28,840 kb
28,850 kb
H3K27ac HiChIP coverage
PAN3
SJALL060150
SJBALL020169
SJBALL113
Nalm-16
REH
Chromatin state key
Active promoter
Enhancer
Transcription
Heterochromatin
Quiescent
Bivalent/poised TSS
Bivalent enhancer
ZNF genes and repeats

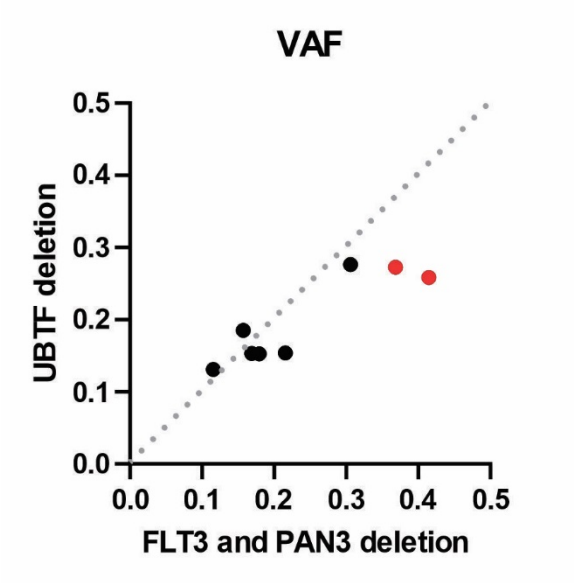
Supplementary Figure 7

Chromatin state of the enhancers in the *PAN3* gene across human tissues and cell types. Snapshots from the Epilogos browser[1] (https://epilogos.altius.org/) showing chromatin state of the *CDX2* and *PAN3* genomic region. Each row is a different cell type and each nucleotide is colored by the chromatin state as determined by chromHMM.[2] (A) All 833 samples with the subset of immune-related cell types indicated with the black outline. The enhancers in the *PAN3* gene are outlined with the white box and asterisk. (B) Same as in (A) but only showing non-immune cell types. (C) same as in (A) but only showing immune cell types. (D) Zoomed in view of the enhancers in the *PAN3* gene of immune cell types. Shown below are H3K27ac HiChIP coverage tracks from the 5 samples profiled by HiChIP.



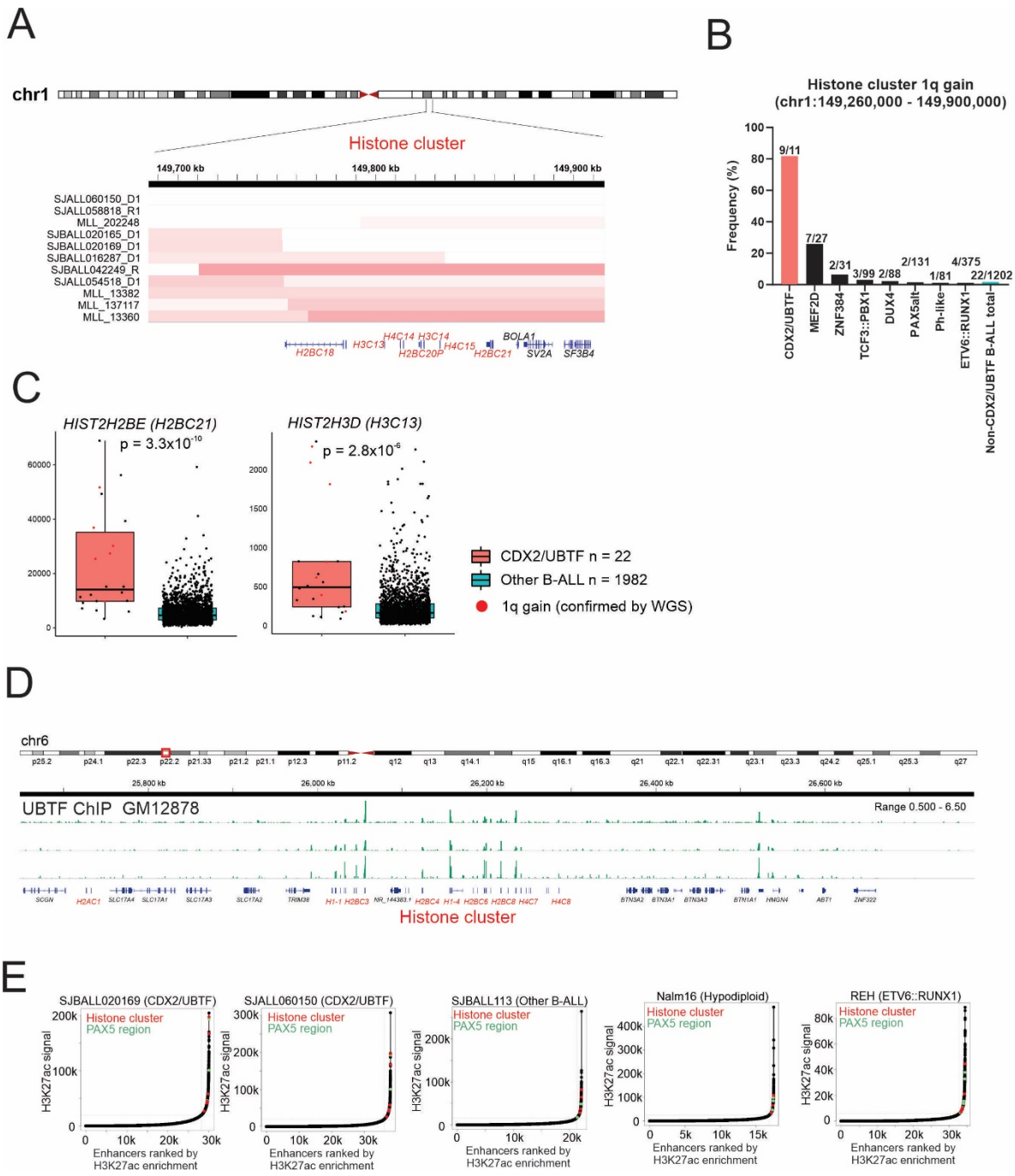
Supplementary Figure 8

3D configuration and enhancer landscape of patient samples and cell lines with *FLT3/PAN3* SVs. Each panel shows the H3K27ac HiChIP heatmap and corresponding H3K27ac and RNA-seq coverage tracks as described in Figure 3D. For each sample, the total number of reads spanning each pair-wise 5kb genomic bin are shown in the heatmap and significant (FDR < 0.01) chromatin loops are shown as blue arcs. Loops linking the *PAN3* genic enhancers to the *CDX2* gene are shown in red. The *CDX2* gene itself is highlighted in gold and the *PAN3* genic enhancers are outlined in the black dotted box. The genomic window corresponds to chr3:28457037-28911626.



Supplementary Figure 9

Variant allele frequencies (VAFs) of two different structural variants (deletion of *UBTF* and *FLT3/PAN3* regions) detected in CDX2/UBTF cases. Two cases (SJBALL020169 and SJBALL042249; annotated in red) had higher VAFs for *FLT3/PAN3* deletion than *UBTF* deletion, suggesting *FLT3/PAN3* deletion occurred before *UBTF* deletion.



Supplementary Figure 10

(A) The results of copy number alteration analysis by CNVkit revealed frequent gain of histone cluster on 1q in CDX2/UBTF (9 out of 11cases). (B) The frequencies of

gain of histone cluster (chr1: 149,260,000-149,900,000) in each B-ALL representative subtype were shown. CDX2/UBTF exhibited higher frequency of gain (81.8%; 9/11) compared to other B-ALL cases (1.83%; 22/1202). (C) Expression of *HIST2H2BE* and *HIST2H3D* genes (on histone cluster in chromosome 1q) by normalized read counts comparing CDX2/UBTF and other B-ALL cases are shown. CDX2/UBTF exhibited higher expression of these genes. (D) ChIP-seq coverage tracks (replicates) of UBTF[3] for GM12878 (Epstein-Barr virus transformed normal B cell line) in histone cluster on chromosome 6 are shown. UBTF binds to the promoters of histone genes. (E) Rank ordering of super-enhancers analyzed by ROSE software using H3K27ac HiChIP data of representative patient samples (n = 3, SJALL060150 and SJBALL020169 (CDX2/UBTF) and SJBALL113 (B-other)) and cell lines (n = 2, Nalm-16 (hypodiploid) and Reh (*ETV6::RUNX1*)). Enhancers at histone clusters and *PAX5* regions were enriched in CDX2/UBTF subtype compared to others.

SUPPLEMENTARY METHODS

UBTF PCR

DNA and RNA were extracted and reverse transcribed using random hexamer primers and Superscript III (Invitrogen, #18081-051) as previously described.[4] Genomic UBTF region was amplified from genomic DNA using Phusion High-Fidelity DNA polymerase (NEB, #M0530S) using the following primers: GTTACCAGAAGTTCTCCCAGGAG (chr17:42275927-42275948) and GTCTGGCCTGGATAACAGCAAA (chr17:42286423-42286444). UBTF fusion transcripts were amplified from genomic DNA using Phusion High-Fidelity DNA polymerase (NEB, #M0530S) using the following primers: CCTCCAGCTGCTACAAATTCTTCC (UBTF exon 16), GTTACCAGAAGTTCTCCCAGGAG (UBTF exon 17), TTTGCTGTTATCCAGGCCAGAC (ATXN7L3 exon 1), and CTTCATGCTATCAGGGTCCGTG (ATXN7L3 exon 2).

Transfection and Immunoblotting

The 293T cells lines were purchased from ATCC (Manassas, VA) and were identified by DNA profiling (STR). The 293T cell line was cultured in Dulbecco's modified Eagle's Medium (DMEM) with 10% FBS and L-glutamine, 100 U/mL penicillin-streptomycin. 293T cells were transfected with MSCV-IRES-GFP/Ametrine as an empty vector or containing the wild-type gene of interest by using FuGENE HD

(Promega). Transfected cells were harvested 48 hour post transfection. Transfected cells were lysed in RIPA buffer supplemented with protease and posphatase inhibitors (Thermo Scientific, #1861281). Lysed protein (40 µg) was electrophoresed through 4–12% NuPage Bis-Tris gels (Life Technologies) at 110 V for 100 min. Blots were probed with anti-UBTF (Abcam, #ab244287), anti-ATXN7L3 (Thermo Scientific, #MA3-084), and anti-beta actin (C4) (Santa Cruz Biotechnology, #sc-47778) antibodies as previously described.[5]

REFERENCES

1. Boix CA, James BT, Park YP, et al: Regulatory genomic circuitry of human disease loci by integrative epigenomics. Nature 590:300-307, 2021
2. Ernst J, Kellis M: ChromHMM: automating chromatin-state discovery and characterization. Nat Methods 9:215-6, 2012
3. Moore JE, Pratt HE, Purcaro MJ, et al: A curated benchmark of enhancer-gene interactions for evaluating enhancer-target gene prediction methods. Genome Biol 21:17, 2020
4. Mullighan CG, Goorha S, Radtke I, et al: Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia. Nature 446:758-64, 2007
5. Gu Z, Churchman M, Roberts K, et al: Genomic analyses identify recurrent MEF2D fusions in acute lymphoblastic leukaemia. Nat Commun 7:13331, 2016