3D Genome Analysis Identifies Enhancer Hijacking Mechanism for High-Risk Factors in Human T-Lineage Acute Lymphoblastic Leukemia

Lu Yang[1,2,3]^, Fengling Chen[4]^, Haichuan Zhu[1,2,3]^§, Yang Chen[4], Bingjie Dong[1,2,3], Minglei Shi[4], Weitao Wang[1,2,3], Qian Jiang[6], Leping Zhang[7], Xiaojun Huang[2,6]*, Michael Q. Zhang[4,5,8]* and Hong Wu[1,2,3,6]*

[1]The MOE Key Laboratory of Cell Proliferation and Differentiation, School of Life Sciences, [2]Peking-Tsinghua Center for Life Sciences and [3]Beijing Advanced Innovation Center for Genomics, Peking University, Beijing 100871, China; [4]MOE Key Laboratory of Bioinformatics, Center for Synthetic and Systems Biology, Bioinformatics Division, BNRist, Department of Automation, [5]School of Medicine, Tsinghua University, Beijing 100084, China; [6]Peking University Institute of Hematology, National Clinical Research Center for Hematologic Disease, [7]Department of Pediatric Hematology/Oncology, Peking University People's Hospital, Beijing 100044, China; [8]Department of Biological Sciences, Center for Systems Biology, The University of Texas, Dallas 800 West Campbell Road, RL11, Richardson, TX 75080-3021, USA

^ These authors contributed equally to this work.
§Current address: Institute of Biology and Medicine, College of Life and Health Sciences, Wuhan University of Science and Technology, Hubei 430081, China.

*Corresponding authors:

hongwu@pku.edu.cn; xjhrm@medmail.com.cn; michael.zhang@utdallus.edu

Lead Contact: hongwu@pku.edu.cn

Recent studies have demonstrated that 3D genome alterations play important roles in tumorigenesis[1-3], including the development of hematological malignancies[4-7]. However, how such alterations may provide key insights into T-lineage acute lymphoblastic leukemia (T-ALL) patients is largely unknown. Here, we report integrated analyses of 3D genome alterations and differentially expressed genes (DEGs) in 18 newly diagnosed T-ALL patients and 4 healthy T cell controls. We found that 3D genome organization at the compartment, topologically associated domains (TAD) and loop levels as well as the gene expression profiles could hierarchically classify different subtypes of T-ALL according to the T cell differentiation trajectory. Alterations in the 3D genome were associated with nearly 45% of the upregulated genes in T-ALL. We also identified 34 previously unrecognized translocations in the noncoding regions of the genome and 44 new loops formed between translocated chromosomes, including translocation-mediated enhancer hijacking of the *HOXA* cluster. Our analysis demonstrated that T-ALLs with *HOXA* cluster overexpression were heterogeneous clinical entities, and ectopic expressions of the *HOXA11*-A*13* genes, but not other genes in the *HOXA* cluster, were associated with immature phenotypes and poor outcomes. Our findings highlight the potentially important roles of 3D genome alterations in the etiology and prognosis of T-ALL.

Keywords: 3D genome alterations in T-ALL, chromosomal translocation, enhancer hijacking, ectopic *HOXA11-A13* expression, poor prognosis of T-ALL.

T-ALL is an aggressive hematological malignancy caused by genetic and epigenetic alterations that affect normal T cell development[8]. Recent whole-exome and RNA sequencing analyses of large T-ALL cohorts are focused on the coding region of the genome and have identified novel driver mutations, dysregulated transcription factors and pathways in T-ALLs[9-12]. To determine whether alterations in the 3D genome organization are associated with malignant transformation of T-ALL, we conducted BL-Hi-C[13] analysis using purified primary leukemic blasts from 18 newly diagnosed T-ALL patients, including 8 early T-cell precursor ALL (ETP ALL) and 10 non-ETP ALL, two clinical subtypes of T-ALL, as well as normal T cells from 4 healthy volunteers (Supplementary Fig. 1a). The maximum resolutions of the chromatin contact maps for ETP, non-ETP ALL and normal samples were approximately 3.5, 3.5 and 10 kb, respectively (Supplementary Table 1).

Principal component analysis (PCA) at the levels of the compartment, TAD and loop structures demonstrated that the T-ALL samples could be separated from the control samples by PC1, while ETP and non-ETP ALL could be separated by PC2 at all three architectural levels (Fig. 1a, upper panels) and be further delineated by hierarchical clustering analysis (Fig. 1a, lower panels). Detailed comparisons of the 3D chromosomal organizations of the T-ALL patients and the healthy controls revealed 1.59% of A-to-B and 1.38% of B-to-A T-ALL-associated switches at the compartment level (Supplementary Fig. 1b), 11% （377/3421） T-ALL-specific TAD boundaries (Supplementary Fig. 1c), and 6% (2330/38464) enhanced and 11% (4073/38464) reduced loops (Supplementary Fig. 1d), implying that the global 3D genome alterations could link to the etiology of T-ALL development.

RNA-seq analysis was also performed on each sample to investigate the impact of 3D genome alterations on gene expression in the T-ALLs. PCA and hierarchical clustering showed that the transcriptome changes were highly consistent with the corresponding 3D genome structural changes (comparing Fig. 1a and 1b) and hence could serve as a biological readout for the functional impact of 3D genome alterations in T-ALL development.

Approximately 29% (996/3392) of the DEGs were associated with 3D genome alterations (Supplementary Table 2). Among them, the genes associated with the B-to-A compartment switches, increased domain scores (D-score) and the corresponding enhanced loops were mostly upregulated (Fig. 1c, red bars and Fig. 1d), and were functionally enriched in pathways such as hematopoietic cell lineage, transcriptional misregulation in cancer and cell cycle (Fig. 1e). In contrast, the genes associated with the A-to-B compartment switches, decreased D-scores and reduced loops were mostly downregulated (Fig. 1c, blue bars and Fig. 1d), and these genes were enriched in pathways such as cytokine-cytokine receptor interaction and T cell receptor signaling (Fig. 1e).

Among the upregulated DEGs, *CDK6* is a potential target for T-ALL treatment[14]. The *CDK6* locus exhibited a strong intra-TAD interaction, and its expression was upregulated in all T-ALL samples (Fig. 1f; Supplementary Fig. 1e). Similarly, several upregulated oncogenic driver genes and T-ALL-associated transcription factors, such as *MYB, MYCN, BCL11A, SOX4* and *WT1,* also had increased D-scores

(Supplementary Fig. 1f).

While the majority of the DEGs were associated with structural changes at 1 or 2 levels, transcription factor *SOX4*, *WT1* and *TFDP2* had structural alterations at all 3 levels (Fig. 1g, Supplementary Fig. 1g and Supplementary Table 2). Importantly, when RcisTarget[15] was used to predict the key transcription factors that might control the upregulated genes in the T-ALLs, *SOX4*, *WT1* and *TFDP2* were among the top-ranking candidates that potentially regulated approximately 28% of the upregulated genes (Supplementary Table 3). After eliminating the overlapping upregulated DEGs, we estimated that 3D genome alterations could associate with approximately 45% of the upregulated genes in T-ALL, of which 56% could be directly related to structural alterations, while the rest could be due to structure alteration-mediated dysregulated transcription factors.

ETP ALL is a subclass of T-ALL stalled at the early T progenitor stage, while non-ETP ALLs are blocked at the later T cell differentiation stages[16]. Our 3D genome landscape analyses could separate the ETP ALL samples from the non-ETP ALL samples (Fig. 1a), suggesting that the chromosomal organizations of T-ALL may represent different “frozen stages” of T cell development[17]. To test this hypothesis, we projected the T-ALL samples onto the T cell developmental trajectory defined by RNA-seq analysis[18]. PCA revealed that most of the ETP ALL samples were arrested at the immature stage, corresponding to the LMPP to Thy1 stages, while the non-ETP samples were arrested at the Thy2 to Thy4 stages (Fig. 2a). Since the ETP and non-ETP ALLs can be better separated at the loop level (Fig. 1a), we further analyzed the differences in loop structures between ETP and non-ETP ALL samples and identified 1820 enhanced and 831 reduced loops in ETP (Fig. 2b). When plotting gene expression changes between ETP and non-ETP ALL against the combined p-value of the loop strength and D-score changes, we found a strong positive correlation (Pearson's correlation coefficient 0.685; Fig. 2c). Approximately 20% and 16% of the upregulated genes in ETP and non-ETP ALL, respectively, harbored chromatin structure changes, including key transcription factors or oncogenes, such as *CEBPA*, *MYCN* and *LYL1* for ETP and *LEF1*, *TCF12* and *PAX9* for non-ETP ALL (Fig. 2c and Supplementary Table 4).

Gene ontology analysis further revealed that genes associated with the ETP ALL enhanced loops were enriched in immune response-activating signal transduction, myeloid cell differentiation and regulation of B cell activation, consistent with the definition of ETP ALL (Fig. 2d). Genes associated with the non-ETP ALL enhanced loops were enriched in terms such as positive regulation of RNA metabolism, transcription, and TCR V(D)J recombination (Fig. 2d). The lack of TCR rearrangement in most of the ETP ALL samples (Supplementary Fig. 2a) and different rearrangements in individual non-ETP samples (Supplementary Fig. 2b) could be easily observed from the Hi-C maps (Fig. 2e). Sample 093 was a unique case as it fell between ETP and non-ETP ALL (Fig. 1a and 1b; Fig. 2a) and had significant TCR rearrangement (Supplementary Fig. 2a). We also observed a lack of *RAG1* and *PTCRA* expression in most of the ETP ALL samples, which are essential for TCR V(D)J rearrangements (Fig. 2f).

The 3D genome alteration analysis also provided a potential explanation for ETP and non-ETP ALL-specific transcription factor expressions. For example, we detected subtype-specific loops and expression patterns in the *MEF2C* locus in ETP and the *PAX9* locus in the non-ETP ALL samples, respectively, which were associated with H3K27ac and CTCF marks in the ETP cell line KE37 and non-ETP cell line Jurkat (Fig. 2g and 2h).

Chromosomal translocation is one of the major driving forces for tumorigenesis[19], especially for leukemia[8]. By adapting hic_breakfinder[20], we identified 46 translocations in 14/18 T-ALL samples (Supplementary Table 5), of which 34 were novel events and 26 were interchromosomal events (Fig. 3a, red lines). Among 78 unique breakpoints identified, 47% were located in noncoding regions, and 66% were located in the stable A compartment (Supplementary Fig. 3a). These newly identified translocations not only influenced the expression of the nearest genes (Fig. 3a) but also resulted in the formation of 44 new loops across the translocated chromosomes, which we named translocation-mediated loops (Supplementary Fig. 3b; Supplementary Table 6). Interestingly, the ends of these translocation-mediated loops tend to anchor at the pre-existing loop anchors and CTCF binding sites (Fig. 3b). Importantly, nearly 78% of the translocation-mediated loops with CTCF motifs were linked to pairs of convergently orientated CTCF motifs (Fig. 3c), indicating that these loops may be mediated by loop extrusion mechanism[21-23].

We then investigated the potential mechanisms underlying translocation-mediated T-ALL classification and gene activation. Clinically, non-ETP

ALL can be further classified into the HOXA, TLX and TAL subtypes according to their gene expression profiles[24]. Notably, there was a complete match between loop-based hierarchical clustering and T-ALL subtypes, which were signified by chromosomal translocation-mediated dysregulation of T-ALL-associated transcription factors (Fig. 3d).

Translocations can activate T-ALL-associated transcription factors via either “cis” or “trans” mechanisms. The “cis” mechanism involved translocation-mediated enhancer hijacking in the ETP, TLX and TAL subtypes (Fig. 3d and Supplementary Fig. 3b), of which the ectopically expressed genes, such as *TLX3,* hijacked the enhancers from the translocated *BCL11B* gene via translocation-mediated loops (Fig. 3e and Supplementary Fig. 3c-d). The “cis” mechanism included *BCL11B-TLX3*, *TRB-TAL2*, and 3 novel *HOXA* translocations identified in this study (Supplementary Fig. 3b and Supplementary Table 6; see Fig. 5). Interestingly, most of the hijacked enhancers are from genes that are normally expressed during T cell development, such as *BCL11B* and *TRB*, which lead to ectopic expression of T-ALL-associated transcription factors in the T lineage and block normal differentiation (Supplementary Fig. 3d). The “trans” mechanism involves translocation-mediated gene fusions, such as the *PSIP1-NUP98*, *SET-NUP214,* and *MLL* (*KMT2A-MLLT1*, *PICALM-MLLT10*, and *DDX3X-MLLT10)* gene fusion events, which could epigenetically activate *HOXA* cluster gene expressions[25-28]. These results suggest that translocation-mediated enhancer hijacking or fusion events may drive ectopic transcription factor activation, leading to specific pathogenic gene expression profiles of the T-ALL subgroups.

The dysregulated *HOXA* cluster, which contains 11 genes, is a common feature of T-ALL[24]. However, whether *HOXA*-positive T-ALLs represent a homogeneous clinical entity has not been systematically studied. We conducted unsupervised hierarchical clustering based on *HOXA* gene expressions, which separated 15 T-ALL samples without *HOXA* translocation into *HOXA*-negative (HOXA$^-$) and *HOXA*-positive expression (HOXA$^+$) groups. Translocation-mediated fusion events could be detected in 5/7 HOXA$^+$ samples. The HOXA$^+$ T-ALLs could be further separated into 2 subgroups: the 3'HOXA$^+$ or 5'HOXA$^+$ subgroups, with respect to the location of the *HOXA* genes within the *HOXA* cluster (Fig. 4a). The expression patterns of the three *HOXA* translocation cases (HOXA-T, breakpoints shown in Fig. 4b) were closer to those in the 5'HOXA$^+$ subgroup, characterized by ectopic *HOXA11-A13* expressions (Fig. 4a). Interestingly, the 5'HOXA$^+$ and HOXA-T cases in our Hi-C study are associated with double negative (DN) and ETP phenotypes (Fig. 4a), suggesting that T-ALLs with ectopic *HOXA* cluster gene expression are heterogeneous entities.

The *HOXA* genes are transcriptionally repressed in normal T cells but can be transactivated in T-ALLs by fusion proteins that recruit histone methyltransferase DOT1L to the *HOXA* locus[25,29]. Although this mechanism uncovered how the *HOXA* cluster is activated, it cannot explain the diverse *HOXA* expression patterns associated with different fusion proteins. By integrating Hi-C maps with *HOXA* gene expression patterns, we found that the differential *HOXA* gene expressions were associated with different 3D genome organizations in samples without *HOXA* translocation. Hi-C maps and CTCF motif orientations showed that the 11 *HOXA* genes were partitioned between 2 TADs (Fig. 4b and Supplementary Fig. 4a): the CTCF binding site C11/13 was used as the 3' boundary of the 5' TAD in all samples, while the 5' boundaries of the 3' TADs varied among different samples: C7/9 was used by most of the HOXA$^-$ (6/8) and all 5'HOXA$^+$ samples (2/2), while C10/11 was used by most of the 3'HOXA$^+$ samples (4/5) (red and blue arrows/lines, respectively; Fig. 4b and Supplementary Fig. 4a).

We further identified 6 enhancer regions in each TAD, labeled E1 to E12, which could interact with the *HOXA* cluster (Fig. 4c). Using HOXA$^-$ cases as common denominators (Fig. 4c, top panel), we calculated the overall differential interaction intensities. Although there was no significant difference between the healthy controls and the HOXA$^-$ cases in the 12 interaction regions, we found significantly enhanced interactions between E2-E6 and genes in the 3'HOXA subgroup, as well as between E8, 9, 11, 12 and genes in the 5'HOXA subgroups, either as a group average (Fig. 4c) or individually (Supplementary Fig. 4b-c). ChIP-seq analysis of the HOXA$^+$ Loucy cell line indicated that these increased interactions may be correlated with gains in the H3K27ac histone mark (Fig. 4c). The KMT2A-MLLT1, PICALM-MLLT10, DDX3X-MLLT10 and SET-NUP214 fusions were associated with interaction dynamics and HOXA1-A10 expression in the 3'HOXA subgroup, while the PSIP1-NUP98 fusion event was associated with the 5'HOXA subgroup (Fig. 4d). These results suggest that different translocation-mediated fusion events may epigenetically and preferentially alter the 3D genome interactome within the *HOXA* cluster and control a specific set of *HOXA* gene expression.

To explain the gene expression patterns seen in the three HOXA-T samples, we mapped the breakpoints and found that all the breakpoints were located within the 5' TAD, upstream of the *HOXA13* gene. The translocation partner breakpoints lie in the gene bodies of the *BCL11B* and *CDK6* genes on chromosome 14 and chromosome 7, respectively, as well as upstream of the *ERG* gene on chromosome 21 (Fig. 5 and Supplementary Fig. 5). By examining the TAD structures associated with the translocations, using HOXA$^-$ samples as controls, we found that these translocations mediated new loop formations between the 5’ of the *HOXA* cluster and active enhancers associated with *BCL11B* and *ERG* genes, leading to ectopic expression of *HOXA13* in the case of 076 and *HOXA9-A13* in the case of 077 (blue circles for new loops and green bar graphs for gene expressions; Fig. 5a and b; Supplementary Fig. 3b). For the case of 108, the inter-TAD inversion led to the adoption of active CDK6 enhancers and new loop formation (Fig. 5c), causing ectopic expression of *HOXA11-A13* (Fig. 5c). These results suggest that translocations and inversion associated with the 5’ TAD of the *HOXA* cluster lead to dysregulated 5’ *HOXA* gene expressions through new loop formation-mediated enhancer hijacking. Again, the existence of the CTCF binding sites C11/13 in the case of 076, C7/9 in the case of 077, and C10/11 in the case of 108 (Fig. 5d) could insulate genes located in the 3' TAD from the influence of the hijacked enhancers, leading to 5'HOXA-specific expression patterns (Fig. 4b).

We next investigated whether ectopic *HOXA* cluster expression was associated with poor prognosis. For this, we analyzed a cohort of T-ALL patients with outcome information (see our companion paper). We found that *HOXA11* or *HOXA13* positivity, alone or in combination, but not the expression of other *HOXA* genes, such as the previously used biomarker *HOXA9,* was associated with poor overall and event-free survivals in young adult and pediatric T-ALLs. In contrast, T-ALLs with *TAL1/2* positivity, which was mutually exclusive from that with *HOXA* positivity, were associated with better overall survival, while TLX positivity had no association with the outcome (Fig. 6a and Supplementary Fig. 6a-b). Compared to the *HOXA1-A10*$^+$, *TAL*$^+$ and *TLX*$^+$ samples, the *HOXA11-A13*$^+$ cases were characterized by DN and ETP phenotypes (Fig. 6b and Supplementary Fig. 6c), as well as a higher rate of JAK-STAT pathway gene mutations[9,30] (Fig. 6c). In multivariate analysis, *HOXA13*$^+$ status could serve as an independent predictor for the overall survival of pediatric and young adult T-ALLs (Fig. 6d).

Taken together, our results imply that 3D genome alterations, in addition to previously identified oncogenic driver mutations and dysregulated pathways, may play critical roles in rewiring T-ALL transcriptome and regulating T-ALL development. The newly identified translocations and enhancer hijacking mechanism may provide a potential explanation for how T-ALL-associated transcription factors are ectopically activated and block differentiation at certain T cell developmental stages. Finally, we identified the associations of ectopic *HOXA11-A13* expression and poor survival of T-ALL, corresponding to the subtype currently incurable by standard treatments, and suggested that anti-JAK-STAT inhibitors may benefit this group of patients.

Author contributions

LY, HZ and HW conceived the project; YC designed the Hi-C and RNA-seq experiments; LY, HZ, MS and WW performed the Hi-C and RNA-seq experiments; FC designed the bioinformatic pipelines and performed the Hi-C and RNA-seq integrated analyses, while BD conducted the survival analysis. QJ, LZ and XH contributed the clinical samples and data. LY, FC and BD generated the figures and tables. LY, FC and HW wrote the manuscript with help from all authors. XH was in charge of the clinical study; MZ and YC oversaw the bioinformatics analyses, and HW supervised the entire project.

Acknowledgement

We thank Drs. Meng Lv, Yingjun Chang, and Yan Chang for sample collection; Dr. Cheng Li of Peking University for critically reviewing the manuscript. We also thank Yan Liu, Fei Wang and Xuefang Zhang from the National Center for Protein Sciences Beijing at Peking and Tsinghua Universities for assistance with FACS. This project was supported by the Peking-Tsinghua Center for Life Sciences, Beijing Advanced Innovation Center for Genomics at Peking University for HW and the National Natural Science Foundation of China (81602254 for LY, 31871343 for YC, 31671384 and 81890994 for YC and MZ). WW was supported by the Postdoctoral Fellowship of Peking-Tsinghua Center for Life Sciences.

Conflicts of Interest

The authors declare no competing financial interests.

Materials and Methods

Patients and samples. Eighteen T-ALL patient samples were collected under the protocol approved by the ethics committee of the Institute of Hematology at Peking University. The patient characteristics are described in Supplementary Table 7. ETP status was defined as previously published[16]. Leukemia blast cells were prepared by density-gradient centrifugation of bone marrow samples, and $CD19^-CD14^-CD235^-CD45^+CD7^+$ cells were further purified by FACS analysis using anti-human antibodies for RNA-seq and Hi-C library preparations. Peripheral blood samples were obtained from four healthy donors under the approval of the ethics committee of Peking University. T cells were purified using the EasySep^{TM} Direct Human T Cell Isolation Kit (StemCell Technology #19661).

RNA-seq library preparation, data processing, and differential gene expression analysis. RNA-seq libraries were prepared with TruSeq RNA Library Prep Kit v2 (Illumina). Paired-end RNA-seq reads of the 18 patients and 4 healthy controls were generated with an average depth of 15 million read pairs. Reads were aligned to the hg19 genome with TopHat (v2.1.0) using default settings[31]. Duplicates were removed, and aligned reads were calculated for each protein-coding gene using HTSeq[32], followed by FPKM transformation by normalizing gene exon length and sequencing depth. Raw RNA-seq data for Loucy and Jurkat were downloaded from the GEO database and analyzed as described above.

DESeq2[33] was applied to identify the differentially expressed genes with FDR <0.01 and fold change >2. Genes with fewer than 5 reads in 20% of the samples or with mean reads fewer than 2 were excluded. Signal tracks were generated by using BEDTools[34] genomeCoveragebed to produce bedGraph files scaled to 1 million reads per data set. Then, the UCSC Genome Browser utility[35] bedGraphToBigWig was used with default parameters to generate bigwig files.

ChIP–seq, ATAC-seq data processing and motif analysis. ChIP–seq reads were mapped to the hg19 genome with Bowtie2[36] (v2.3.5) using default parameters, while

ATAC-seq reads were mapped with Bowtie2 using parameter -X 2000 --no-mixed --no-discordant --no-unal. Aligned reads were filtered for a minimum MAPQ of 20, and duplicates were removed using SAMtools[37]. Signal tracks and peaks were generated by using the –SPMR option in MACS2[38]. Then, the UCSC Genome Browser utility bedGraphToBigWig was used with default parameters to transform the bedgraph files to bigwig files. FIMO[39] was used to detect the 20 bp CTCF motif from the Homer motif database in Loucy and Jurkat CTCF peaks with default parameters.

Hi-C and Hi-C data processing. Hi-C was performed on one million cells/sample, according to the BL-Hi-C protocol[13]. Raw BL-Hi-C reads were processed by the in-house HiCpipe framework, which integrated several Hi-C analysis methods for to generate multiple features of the Hi-C data. In particular, ChIA-PET2[40] was used to trim the bridge linkers and HiC-Pro[41] to align reads, filter artifact fragments, and remove duplicates; Juicer[42] was applied to the resulting uniquely mapped contacts to generate individual or merged Hi-C files that could be deposited as contact matrices with multiple resolutions. Knight-Ruiz[43] (KR)-normalized matrices were used in the compartment and TAD analyses.

Compartment and TAD analysis. The compartment was calculated with the eigenvector command of Juicer under 100 kb resolution KR normalized Hi-C matrices. For every 100 kb bin, A or B compartments were defined by the over 70% sample majority rule.

TAD boundaries were calculated by the Insulation score method[44] (with parameters: -is 1000000 -ids 200000 -im mean -nt 0.1) on pooled 40 kb Hi-C matrices of the healthy T cell controls, ETP and non-ETP samples. The resulting TAD boundaries were merged and assigned with relative insulation scores of all samples calculated from HiCDB[45]. Differential TAD boundaries were defined with a t-test FDR<0.01 and a difference between cases and controls higher than 50% quantile of the overall difference.

A TAD was defined when its boundaries were detected in at least two conditions among normal T cells, ETPs and non-ETPs. The domain score[46] was calculated in each sample by dividing the intra-TAD interactions with all interactions connected to the corresponding TAD. Differential domain scores were calculated with a t-test FDR<0.01 and fold change higher than 70% quantile of the overall fold change.

Loop detection and differential loop calling. Loops were called by HiCCUPS[42] at 5 kb and 10 kb resolutions with default parameters (except -d 15000,20000) for pooled Hi-C matrices of the healthy T cell controls, ETP and non-ETP samples, respectively. The differential loop detection method was adapted from Douglas *et al*[47]. Loops were split into two distance ranges (> or < 150 kb) to minimize potential bias (Rubin *et al*.[48]). Differential loops were called within each range (FDR < 0.1) and then combined.

Loop aggregation and functional analysis. Aggregate peak analysis (APA) plots were generated to assess the quality of loop detection and explore the characteristics of different loop classes by the Juicer APA command[42] under 5 kb resolution. Its output matrix was normalized by the loop number that contributed to the matrix generation. For analysis of the function of dynamic loops between non-ETP and ETP, the loop anchors were analyzed by GREAT[49] (v3.0.0) using the nearest gene within 100 kb to generate the enriched biological process. In other sections, genes related to loops were determined if their promoter (5 kb around TSS) overlapped the loop anchors, and DAVID[50] 6.8 was used for KEGG pathway enrichment analysis.

Visualization and V4C plot generation. Tracks of Hi-C maps and ChIP-seq were generated by pyGenomeTracks[51]. Hi-C maps of each condition were normalized by its cis interaction pairs. A visual 4C (V4C) plot for specific loci was generated as the interaction s related to the corresponding viewpoint under 10 kb resolution.

Translocation and translocation-mediated loop detection with Hi-C. hic_breakfinder[20] was adapted to detect translocations in 18 T-ALL patient samples. After we filtered the “translocations” also detected in normal controls, the remaining translocations were manually assessed, and the precise breakpoints were determined. As the average depth of patient Hi-C samples is 486 million read pairs and the read length is 150 bp, Hi-C raw data were treated as single ends to refine the breakpoint locations to single base-pair resolution. Any single ends that could be mapped to two different chromatins without the BL-Hi-C bridge linker in between were chimeric reads. The chimeric reads detected from the BL-Hi-C data overlapped with the aforementioned translocations at 5~20 kb resolution. For each translocation, the exact locations (single base-pair resolution) supported by more than 3 chimeric reads were identified as the actual breakpoints and are reported in Supplementary Table 5.

As translocation-mediated loops were hard to identify by loop detection tools designed for intrachromosomal loop detection and easy to capture by visualization, their locations were manually recorded on interchromosome Hi-C maps with the help of Juicebox[52], which is an interactive visualization software.

Translocation, translocation-mediated loop annotation and visualization. The nearest genes to translocation breakpoints were determined by BEDTools. Known translocations were collected from references 2 and 4 and ChimerPub[53]. A translocation was considered novel if any of the breakpoints was not near any known breakpoint within a 100 kb distance. For translocation-mediated loops, the genes with a promoter or gene body overlapping the loop anchors were annotated as the associated genes. A gene near a breakpoint was considered upregulated if its FPKM was >1- and 2-fold higher than the control samples without nearby breakpoints. Hi-C heatmaps and Visual 4C plots of the reassembled chromatin were generated by MATLAB code.

Statistics. Specific statistical analyses are described in each section. In general, the Wilcoxon rank sum test was employed in R for comparisons of distributions. Survival analysis was performed by a Cox regression model using overall and event-free survival as outcomes. Overall survival was defined as the time from diagnosis to death from any cause. Event-free survival was defined as the time from diagnosis to treatment failure, relapse, or death from any cause. The proportional hazard assumption was tested. Variables tested in the multivariable Cox regression model were sex, age (pediatric vs. adult), white blood cell counts, hemoglobin levels, platelet counts, hepatosplenomegaly, percentage of blasts in the bone marrow and MRD status. Eighty-six patient samples with RNA-seq data were used for DN- and ETP-enrichment analyses, of which 38 samples under the age of 40 with outcome and 63 samples with whole exon sequencing (WES) data were used for survival and mutation analyses, respectively.

Code availability. Custom scripts described in the Online Methods will be made available upon request.

Data availability. All sequencing data are available through the Gene Expression Omnibus (GEO) via accession GSE. Accession codes of the published data used in this study are as follows: CTCF ChIP-seq of CD4+ T cell and Jurkat cell line, GSE12889; CTCF ChIP-seq of Loucy cell line, GSE123214; ATAC-seq of CD4+ T cell, GSE87254; ATAC-seq of Jurkat cell line, GSE115438; H3K27ac ChIP-seq of CD4+ T cell, GSE122826; H3K27ac ChIP-seq of Jurkat cell line, GSE68978; H3K27ac ChIP-seq of Loucy cell line, GSE74311; RNA-Seq of Loucy cell line, GSE100694; RNA-seq of T cell development, GSE69239. The raw sequence data reported in this paper have been deposited in the Genome Sequence Archive[54] in BIG Data Center[55], Beijing Institute of Genomics (BIG), Chinese Academy of Sciences, under accession number HRA000113 that is publicly accessible at **https://bigd.big.ac.cn/gsa**.

References

1 Orlando, G. *et al.* Promoter capture Hi-C-based identification of recurrent noncoding mutations in colorectal cancer. *Nat Genet* **50**, 1375-1380, doi:10.1038/s41588-018-0211-z (2018).

2 Flavahan, W. A. *et al.* Insulator dysfunction and oncogene activation in IDH mutant gliomas. *Nature* **529**, 110-114, doi:10.1038/nature16490 (2016).

3 Yu, M. & Ren, B. The Three-Dimensional Organization of Mammalian Genomes. *Annu Rev Cell Dev Biol* **33**, 265-289, doi:10.1146/annurev-cellbio-100616-060531 (2017).

4 Groschel, S. *et al.* A single oncogenic enhancer rearrangement causes concomitant EVI1 and GATA2 deregulation in leukemia. *Cell* **157**, 369-381, doi:10.1016/j.cell.2014.02.019 (2014).

5 Hnisz, D. *et al.* Activation of proto-oncogenes by disruption of chromosome neighborhoods. *Science* **351**, 1454-1458, doi:10.1126/science.aad9024 (2016).

6 Petrovic, J. *et al.* Oncogenic Notch Promotes Long-Range Regulatory Interactions within Hyperconnected 3D Cliques. *Mol Cell* **73**, 1174-1190 e1112, doi:10.1016/j.molcel.2019.01.006 (2019).

7 Kloetgen, A., Thandapani, P., Tsirigos, A. & Aifantis, I. 3D Chromosomal Landscapes in Hematopoiesis and Immunity. *Trends Immunol* **40**, 809-824, doi:10.1016/j.it.2019.07.003 (2019).

8 Belver, L. & Ferrando, A. The genetics and mechanisms of T cell acute lymphoblastic leukaemia. *Nat Rev Cancer* **16**, 494-507, doi:10.1038/nrc.2016.63 (2016).

9 Liu, Y. *et al.* The genomic landscape of pediatric and young adult T-lineage acute

lymphoblastic leukemia. *Nat Genet* **49**, 1211-1218, doi:10.1038/ng.3909 (2017).

10 Seki, M. *et al.* Recurrent SPI1 (PU.1) fusions in high-risk pediatric T cell acute lymphoblastic leukemia. *Nat Genet* **49**, 1274-1281, doi:10.1038/ng.3900 (2017).

11 Chen, B. *et al.* Identification of fusion genes and characterization of transcriptome features in T-cell acute lymphoblastic leukemia. *Proc Natl Acad Sci U S A* **115**, 373-378, doi:10.1073/pnas.1717125115 (2018).

12 Zhang, J. *et al.* The genetic basis of early T-cell precursor acute lymphoblastic leukaemia. *Nature* **481**, 157-163, doi:10.1038/nature10725 (2012).

13 Liang, Z. *et al.* BL-Hi-C is an efficient and sensitive approach for capturing structural and regulatory chromatin interactions. *Nat Commun* **8**, 1622, doi:10.1038/s41467-017-01754-3 (2017).

14 Sawai, C. M. *et al.* Therapeutic targeting of the cyclin D3:CDK4/6 complex in T cell leukemia. *Cancer Cell* **22**, 452-465, doi:10.1016/j.ccr.2012.09.016 (2012).

15 Aibar, S. *et al.* SCENIC: single-cell regulatory network inference and clustering. *Nat Methods* **14**, 1083-1086, doi:10.1038/nmeth.4463 (2017).

16 Coustan-Smith, E. *et al.* Early T-cell precursor leukaemia: a subtype of very high-risk acute lymphoblastic leukaemia. *Lancet Oncol* **10**, 147-156, doi:10.1016/S1470-2045(08)70314-0 (2009).

17 Hu, G. *et al.* Transformation of Accessible Chromatin and 3D Nucleome Underlies Lineage Commitment of Early T Cells. *Immunity* **48**, 227-242 e228, doi:10.1016/j.immuni.2018.01.013 (2018).

18 Casero, D. *et al.* Long non-coding RNA profiling of human lymphoid progenitor cells

reveals transcriptional divergence of B cell and T cell lineages. *Nat Immunol* **16**, 1282-1291, doi:10.1038/ni.3299 (2015).

19 Lee, J. K., Choi, Y. L., Kwon, M. & Park, P. J. Mechanisms and Consequences of Cancer Genome Instability: Lessons from Genome Sequencing Studies. *Annu Rev Pathol-Mech* **11**, 283-312, doi:10.1146/annurev-pathol-012615-044446 (2016).

20 Dixon, J. R. *et al.* Integrative detection and analysis of structural variation in cancer genomes. *Nat Genet* **50**, 1388-1398, doi:10.1038/s41588-018-0195-8 (2018).

21 Fudenberg, G. *et al.* Formation of Chromosomal Domains by Loop Extrusion. *Cell Rep* **15**, 2038-2049, doi:10.1016/j.celrep.2016.04.085 (2016).

22 Nuebler, J., Fudenberg, G., Imakaev, M., Abdennur, N. & Mirny, L. A. Chromatin organization by an interplay of loop extrusion and compartmental segregation. *Proc Natl Acad Sci U S A* **115**, E6697-E6706, doi:10.1073/pnas.1717730115 (2018).

23 Davidson, I. F. *et al.* DNA loop extrusion by human cohesin. *Science* **366**, 1338-1345, doi:10.1126/science.aaz3418 (2019).

24 Soulier, J. *et al.* HOXA genes are included in genetic and biologic networks defining human acute T-cell leukemia (T-ALL). *Blood* **106**, 274-286, doi:10.1182/blood-2004-10-3900 (2005).

25 Bernt, K. M. *et al.* MLL-rearranged leukemia is dependent on aberrant H3K79 methylation by DOT1L. *Cancer Cell* **20**, 66-78, doi:10.1016/j.ccr.2011.06.010 (2011).

26 Jo, S. Y., Granowicz, E. M., Maillard, I., Thomas, D. & Hess, J. L. Requirement for Dot1l in murine postnatal hematopoiesis and leukemogenesis by MLL translocation. *Blood* **117**, 4759-4768, doi:10.1182/blood-2010-12-327668 (2011).

27 Nguyen, A. T., Taranova, O., He, J. & Zhang, Y. DOT1L, the H3K79 methyltransferase, is required for MLL-AF9-mediated leukemogenesis. *Blood* **117**, 6912-6922, doi:10.1182/blood-2011-02-334359 (2011).

28 Okada, Y. *et al.* hDOT1L links histone methylation to leukemogenesis. *Cell* **121**, 167-178, doi:10.1016/j.cell.2005.02.020 (2005).

29 Barry, E. R., Corry, G. N. & Rasmussen, T. P. Targeting DOT1L action and interactions in leukemia: the role of DOT1L in transformation and development. *Expert Opin Ther Targets* **14**, 405-418, doi:10.1517/14728221003623241 (2010).

30 de Bock, C. E. *et al.* HOXA9 Cooperates with Activated JAK/STAT Signaling to Drive Leukemia Development. *Cancer Discov* **8**, 616-631, doi:10.1158/2159-8290.CD-17-0583 (2018).

31 Trapnell, C. *et al.* Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc* **7**, 562-578, doi:10.1038/nprot.2012.016 (2012).

32 Anders, S., Pyl, P. T. & Huber, W. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166-169, doi:10.1093/bioinformatics/btu638 (2015).

33 Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550, doi:10.1186/s13059-014-0550-8 (2014).

34 Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842, doi:10.1093/bioinformatics/btq033 (2010).

35 Kuhn, R. M., Haussler, D. & Kent, W. J. The UCSC genome browser and associated tools. *Brief Bioinform* **14**, 144-161, doi:10.1093/bib/bbs038 (2013).

36 Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-359, doi:10.1038/nmeth.1923 (2012).

37 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079, doi:10.1093/bioinformatics/btp352 (2009).

38 Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137, doi:10.1186/gb-2008-9-9-r137 (2008).

39 Grant, C. E., Bailey, T. L. & Noble, W. S. FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27**, 1017-1018, doi:10.1093/bioinformatics/btr064 (2011).

40 Li, G., Chen, Y., Snyder, M. P. & Zhang, M. Q. ChIA-PET2: a versatile and flexible pipeline for ChIA-PET data analysis. *Nucleic Acids Res* **45**, e4, doi:10.1093/nar/gkw809 (2017).

41 Servant, N. *et al.* HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol* **16**, 259, doi:10.1186/s13059-015-0831-x (2015).

42 Rao, S. S. *et al.* A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665-1680, doi:10.1016/j.cell.2014.11.021 (2014).

43 Knight, P. A. & Ruiz, D. A fast algorithm for matrix balancing. *Ima J Numer Anal* **33**, 1029-1047, doi:10.1093/imanum/drs019 (2013).

44 Giorgetti, L. *et al.* Structural organization of the inactive X chromosome in the mouse. *Nature* **535**, 575-579, doi:10.1038/nature18589 (2016).

45 Chen, F., Li, G., Zhang, M. Q. & Chen, Y. HiCDB: a sensitive and robust method for detecting contact domain boundaries. *Nucleic Acids Res* **46**, 11239-11250, doi:10.1093/nar/gky789 (2018).

46 Stadhouders, R. *et al.* Transcription factors orchestrate dynamic interplay between genome topology and gene regulation during cell reprogramming. *Nat Genet* **50**, 238-249, doi:10.1038/s41588-017-0030-7 (2018).

47 Phanstiel, D. H. *et al.* Static and Dynamic DNA Loops form AP-1-Bound Activation Hubs during Macrophage Development. *Mol Cell* **67**, 1037-1048 e1036, doi:10.1016/j.molcel.2017.08.006 (2017).

48 Rubin, A. J. *et al.* Lineage-specific dynamic and pre-established enhancer-promoter contacts cooperate in terminal differentiation. *Nat Genet* **49**, 1522-1528, doi:10.1038/ng.3935 (2017).

49 McLean, C. Y. *et al.* GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol* **28**, 495-501, doi:10.1038/nbt.1630 (2010).

50 Huang da, W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**, 44-57, doi:10.1038/nprot.2008.211 (2009).

51 Ramirez, F. *et al.* High-resolution TADs reveal DNA sequences underlying genome organization in flies. *Nat Commun* **9**, 189, doi:10.1038/s41467-017-02525-w (2018).

52 Durand, N. C. *et al.* Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom. *Cell Syst* **3**, 99-101, doi:10.1016/j.cels.2015.07.012 (2016).

53 Kim, P. *et al.* ChimerDB 2.0--a knowledgebase for fusion genes updated. *Nucleic*

Acids Res **38**, D81-85, doi:10.1093/nar/gkp982 (2010).

54 Wang, Y. *et al.* GSA: Genome Sequence Archive<sup/>. *Genomics Proteomics Bioinformatics* **15**, 14-18, doi:10.1016/j.gpb.2017.01.001 (2017).

55 National Genomics Data Center, M. & Partners. Database Resources of the National Genomics Data Center in 2020. *Nucleic Acids Res* **48**, D24-D33, doi:10.1093/nar/gkz913 (2020).

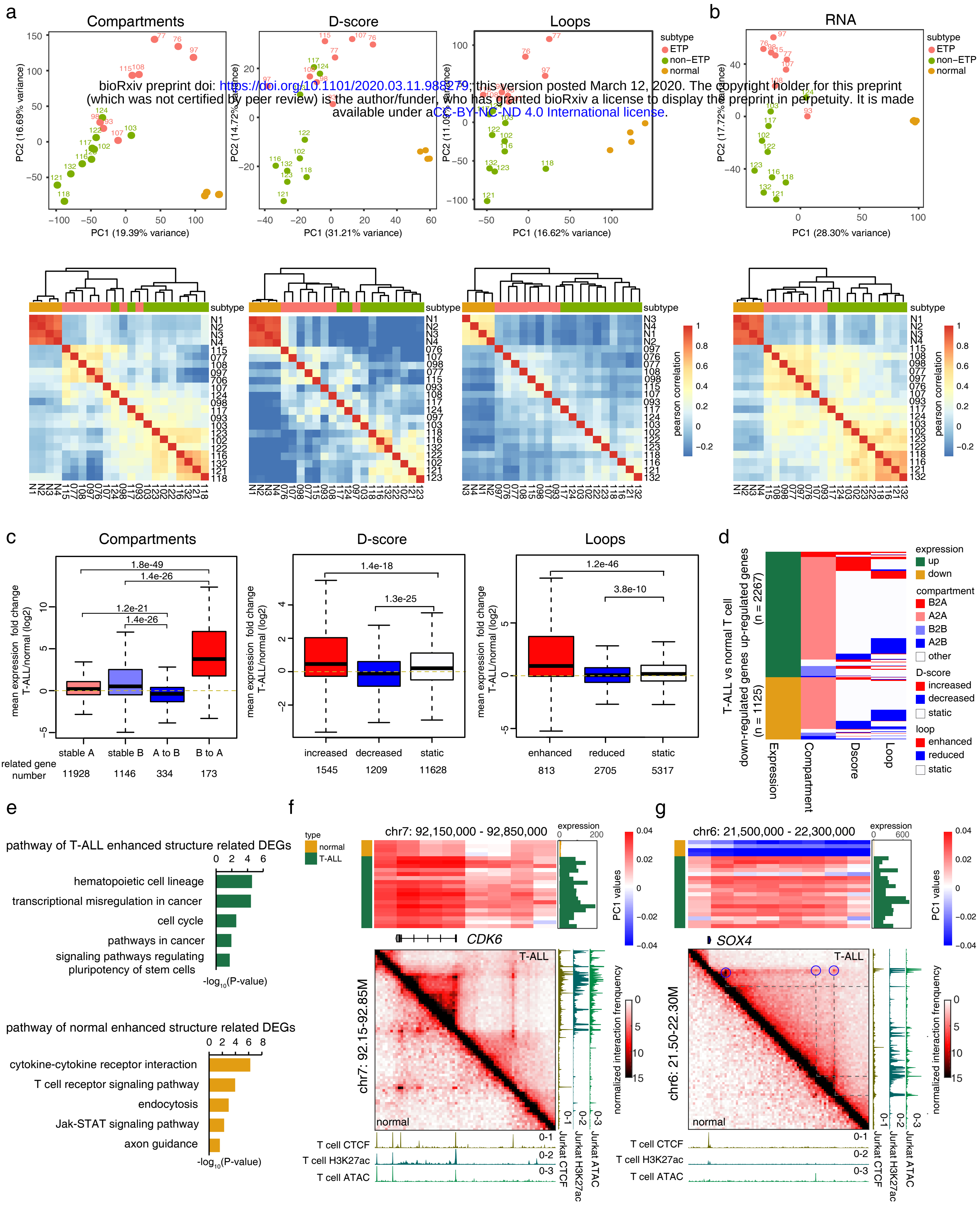

a
Compartments
PC1 (19.39% variance)
PC2 (16.69% variance)
D-score
PC1 (31.21% variance)
PC2 (14.72% variance)
Loops
PC1 (16.62% variance)
PC2 (11.09% variance)
subtype
ETP
non−ETP
normal
b
RNA
PC1 (28.30% variance)
PC2 (17.72% variance)
pearson correlation
c
Compartments
mean expression fold change T-ALL/normal (log2)
1.8e-49
1.4e-26
1.2e-21
1.4e-26
stable A
stable B
A to B
B to A
related gene number
11928
1146
334
173
D-score
1.4e-18
1.3e-25
increased
decreased
static
1545
1209
11628
Loops
1.2e-46
3.8e-10
enhanced
reduced
static
813
2705
5317
d
T-ALL vs normal T cell
up-regulated genes (n = 2267)
down-regulated genes (n = 1125)
Expression
Compartment
Dscore
Loop
expression
up
down
compartment
B2A
A2A
B2B
A2B
other
D-score
increased
decreased
static
loop
enhanced
reduced
static
e
pathway of T-ALL enhanced structure related DEGs
hematopoietic cell lineage
transcriptional misregulation in cancer
cell cycle
pathways in cancer
signaling pathways regulating pluripotency of stem cells
-log10(P-value)
pathway of normal enhanced structure related DEGs
cytokine-cytokine receptor interaction
T cell receptor signaling pathway
endocytosis
Jak-STAT signaling pathway
axon guidance
-log10(P-value)
f
type
normal
T-ALL
chr7: 92,150,000 - 92,850,000
expression
CDK6
chr7: 92.15-92.85M
T-ALL
normal
PC1 values
normalized interaction frenquency
T cell CTCF
T cell H3K27ac
T cell ATAC
Jurkat CTCF
Jurkat H3K27ac
Jurkat ATAC
g
chr6: 21,500,000 - 22,300,000
expression
SOX4
chr6: 21.50-22.30M
T-ALL
normal
PC1 values
normalized interaction frenquency
T cell CTCF
T cell H3K27ac
T cell ATAC
Jurkat CTCF
Jurkat H3K27ac
Jurkat ATAC

Fig.1 Global 3D genome architectures in T-ALLs. (a) PCA (upper) and unsupervised hierarchical clustering analysis (lower) of compartment, domain score (D-score) and loop in normal T cells and T-ALLs. (b) PCA (upper) and unsupervised hierarchical clustering analysis (lower) of gene expression profiles in normal T cells and T-ALLs. (c) Association of DEGs and genomic alterations at the levels of compartment, TAD and loop. Red, upregulation; blue, downregulated. The p values were calculated using Wilcoxon rank sum test. (d) A summary of DEGs and their corresponding chromatin structure changes. (e) KEGG analysis for enriched pathways in T-ALLs and normal T cells based on DEGs that are associated with chromatin structural changes. (f) and (g) Top, heatmaps show the compartment scores for each sample across the genomic regions of the *CDK6* and *SOX4* loci, respectively. Bar plots on the right show the gene expression level of *CDK6* and *SOX4* for each sample. Bottom, Hi-C contact maps of the *CDK6* and *SOX4* loci in normal T cell and T-ALL; blue circles: enhanced loops in T-ALL. ATAC-seq tracks and ChIP-seq tracks of CTCF and H3K27ac of normal T cells and T-ALL Jurket cells are also included.

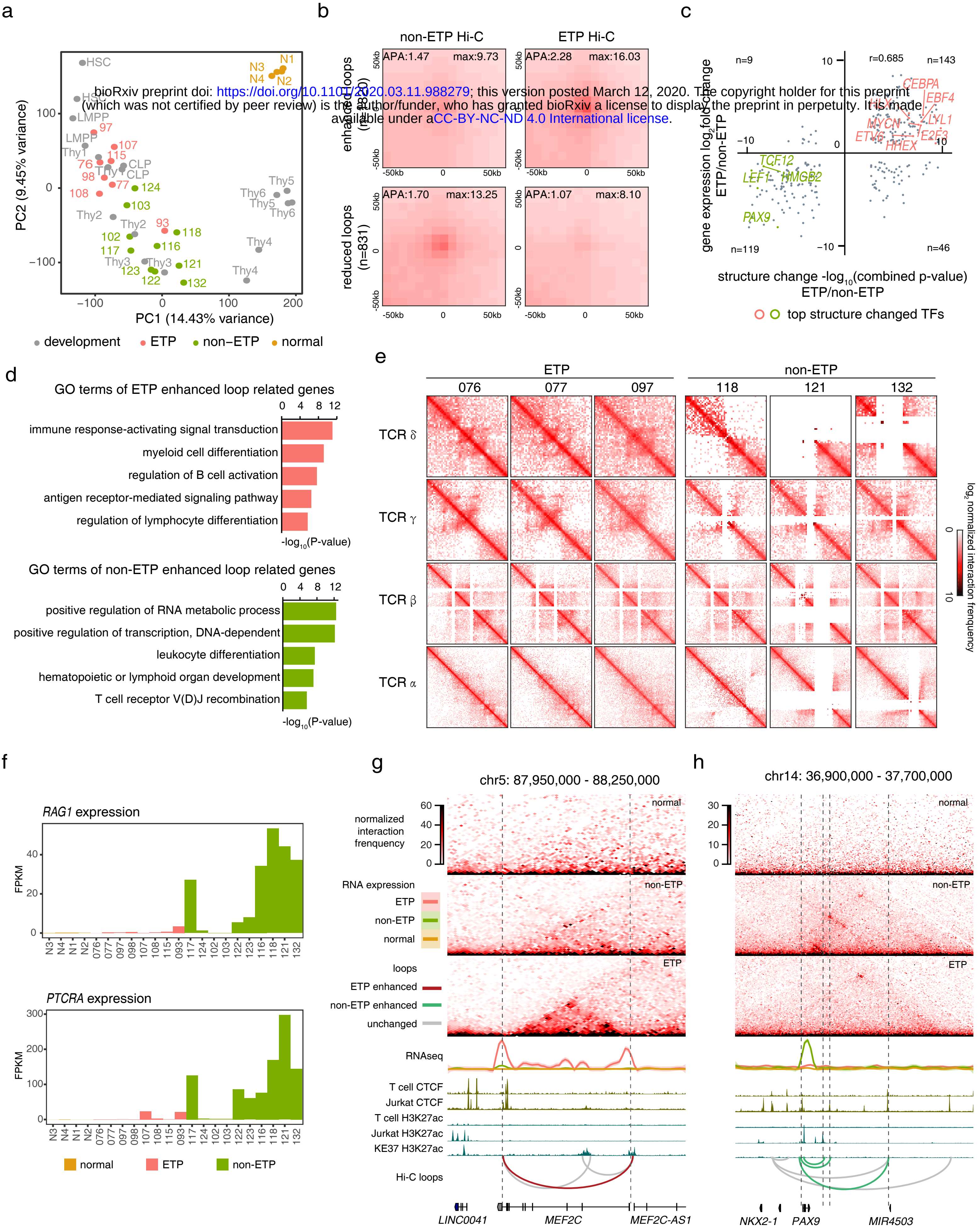

a
PC1 (14.43% variance)
PC2 (9.45% variance)
HSC
LMPP
CLP
Thy1
Thy2
Thy3
Thy4
Thy5
Thy6
N1
N2
N3
N4
development
ETP
non−ETP
normal
b
non-ETP Hi-C
ETP Hi-C
enhanced loops (n=1880)
reduced loops (n=831)
APA:1.47 max:9.73
APA:2.28 max:16.03
APA:1.70 max:13.25
APA:1.07 max:8.10
-50kb
0
50kb
c
gene expression log2 fold change ETP/non-ETP
structure change -log10(combined p-value) ETP/non-ETP
r=0.685
n=9
n=143
n=119
n=46
CEBPA
EBF4
HLX
LYL1
MYCN
E2F3
ETV6
HHEX
TCF12
LEF1
HMGB2
PAX9
top structure changed TFs
d
GO terms of ETP enhanced loop related genes
immune response-activating signal transduction
myeloid cell differentiation
regulation of B cell activation
antigen receptor-mediated signaling pathway
regulation of lymphocyte differentiation
-log10(P-value)
GO terms of non-ETP enhanced loop related genes
positive regulation of RNA metabolic process
positive regulation of transcription, DNA-dependent
leukocyte differentiation
hematopoietic or lymphoid organ development
T cell receptor V(D)J recombination
-log10(P-value)
e
ETP
non-ETP
076
077
097
118
121
132
TCR δ
TCR γ
TCR β
TCR α
log2 normalized interaction frequency
f
RAG1 expression
PTCRA expression
FPKM
normal
ETP
non-ETP
g
chr5: 87,950,000 - 88,250,000
normalized interaction frenquency
RNA expression
ETP
non-ETP
normal
loops
ETP enhanced
non-ETP enhanced
unchanged
RNAseq
T cell CTCF
Jurkat CTCF
T cell H3K27ac
Jurkat H3K27ac
KE37 H3K27ac
Hi-C loops
LINC0041
MEF2C
MEF2C-AS1
h
chr14: 36,900,000 - 37,700,000
NKX2-1
PAX9
MIR4503

Fig. 2 ETP and non-ETP ALLs have different loop structures. (a) PCA shows the association between ETP and non-ETP ALL gene expression profiles and T cell developmental trajectory. (b) APA plots for loops that are enhanced (top) or reduced (bottom) in ETP compared with non-ETP ALL. (c) Scatterplot shows the correlation between gene expression changes and structural changes. The structural change was defined by the combined p-value of D-score and loop strength change. Top ETP and non-ETP ALL-associated transcription factors with structural changes are highlighted in red and green, respectively. (d) GO terms for genes adjacent to enhanced loop anchors in ETP (top) and non-ETP (bottom). (e) Hi-C contact maps for the TCR genomic regions in ETP and non-ETP ALLs. (f) *RAG1* and *PTCRA* expression levels. (g) and (h) Hi-C contact maps for TADs enclosing the genomic loci of ETP expressed *MEF2C* (g) and non-ETP expressed *PAX9* (h). Anchors of differential loops are labeled with dashed lines.

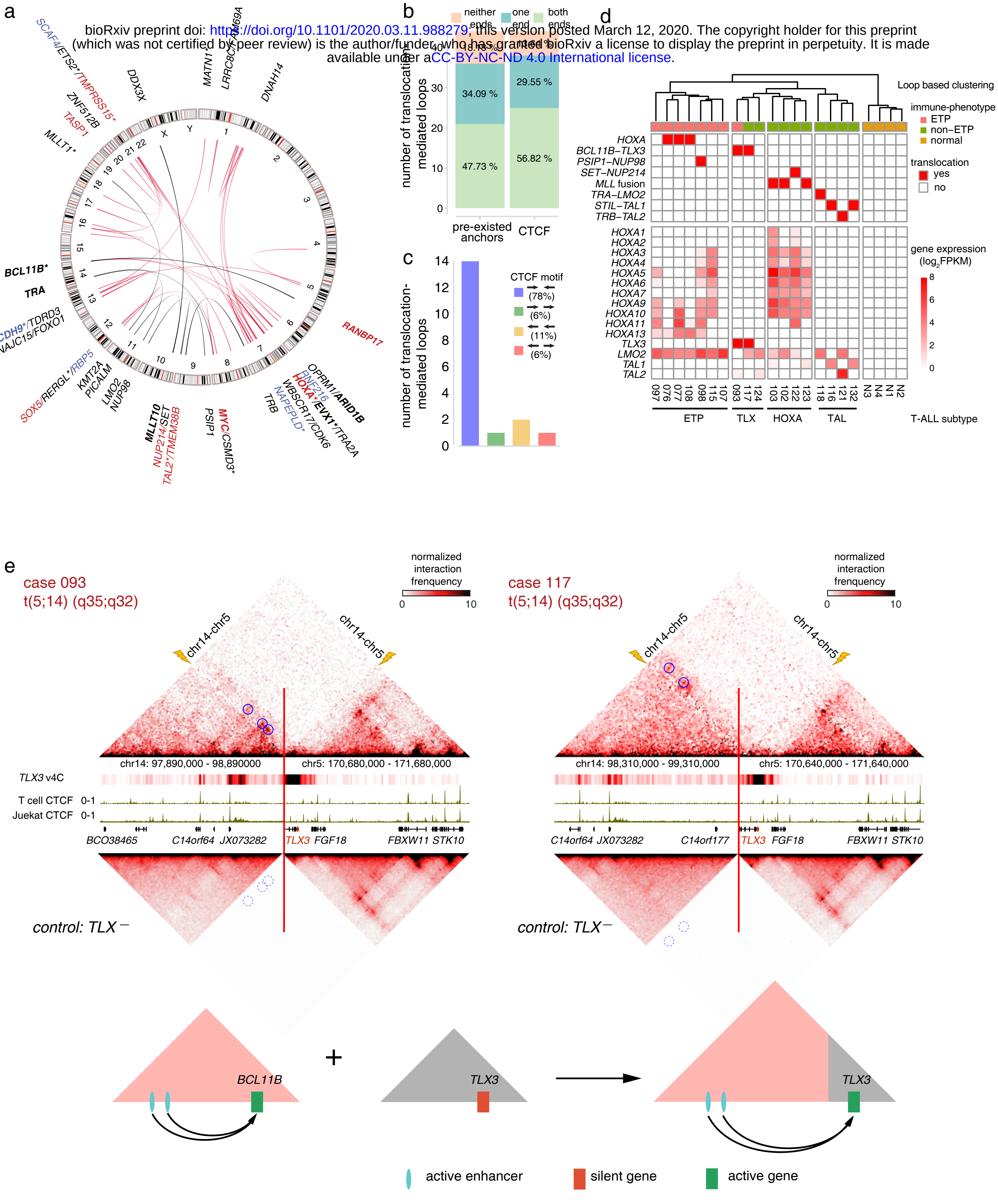

a
SCAF4/ETS2*/TMPRSS15*
ZNF512B
TASP1
MLLT1*
DDX3X
MATN1*
LRRC8C/FAM69A
DNAH14
BCL11B*
TRA
PCDH9*/TDRD3
DNAJC15/FOXO1
SOX5/REFGL*/RBP5
KMT2A
PICALM
LMO2
NUP98
MLLT10
NUP214/SET
TAL2*/TMEM38B
PSIP1
MYC/CSMD3*
TRB
NAPEPLD*
WBSCR17/CDK6
HOXA*/EVX1*/TRA2A
RNF216
OPRM1/ARID1B
RANBP17
b
neither ends
one end
both ends
number of translocation-mediated loops
18.18 %
34.09 %
47.73 %
13.64 %
29.55 %
56.82 %
pre-existed anchors
CTCF
c
number of translocation-mediated loops
CTCF motif
(78%)
(6%)
(11%)
(6%)
d
Loop based clustering
immune-phenotype
ETP
non−ETP
normal
translocation
yes
no
HOXA
BCL11B−TLX3
PSIP1−NUP98
SET−NUP214
MLL fusion
TRA−LMO2
STIL−TAL1
TRB−TAL2
HOXA1
HOXA2
HOXA3
HOXA4
HOXA5
HOXA6
HOXA7
HOXA9
HOXA10
HOXA11
HOXA13
TLX3
LMO2
TAL1
TAL2
gene expression ($\log_2$FPKM)
097 076 077 108 098 115 107 093 117 124 103 102 122 123 118 116 121 132 N3 N4 N1 N2
ETP
TLX
HOXA
TAL
T-ALL subtype
e
case 093
t(5;14) (q35;q32)
normalized interaction frenquency
0
10
chr14-chr5
chr14-chr5
chr14: 97,890,000 - 98,890000
chr5: 170,680,000 - 171,680,000
TLX3 v4C
T cell CTCF 0-1
Juekat CTCF 0-1
BCO38465
C14orf64
JX073282
TLX3
FGF18
FBXW11
STK10
control: TLX −
case 117
t(5;14) (q35;q32)
chr14: 98,310,000 - 99,310,000
chr5: 170,640,000 - 171,640,000
C14orf64
JX073282
C14orf177
TLX3
FGF18
FBXW11
STK10
control: TLX −
BCL11B
+
TLX3
TLX3
active enhancer
silent gene
active gene

Fig. 3 Chromosomal rearrangements in T-ALLs. (a) The genomic landscape of translocations discovered by Hi-C. The novel translocation partners are connected by red lines and known translocations are shown by black lines. Breakpoints nearest genes with increased or decreased expressions are highlighted by red and blue, respectively. Recurrent breakpoints nearest genes are labeled in bold. The genes closest to the breakpoint in noncoding regions are marked by star symbols. (b) Majority of the translocation-mediated loop anchors are pre-existing loop anchors and contains CTCF binding sites. (c) Majority of translocation-mediated loops have convergent CTCF motif orientation. (d) The loop-based clustering overlaps with leukemogenic transcription factor- or translocations-based clustering in T-ALLs. *STIL-TAL1* fusions were detected by RT-qPCR, the other translocations were detected by Hi-C. (e) Upper panels: Hi-C heatmaps for cases with *BCL11B-TLX3* translocation (top), visual 4C plots generated from Hi-C contact maps using *TLX3* promoter as the viewpoint (middle) and averaged Hi-C heatmaps for cases without *BCL11B-TLX3* translocation as controls (bottom). Breakpoints are marked by the red lines and yellow lightning bolts. Translocation-mediated loops are highlighted by blue circles, corresponding loop locations at controls are highlighted by dotted line circles. Lower panels: a schematic illustration for the consequence of the *BCL11B-TLX3* translocation.

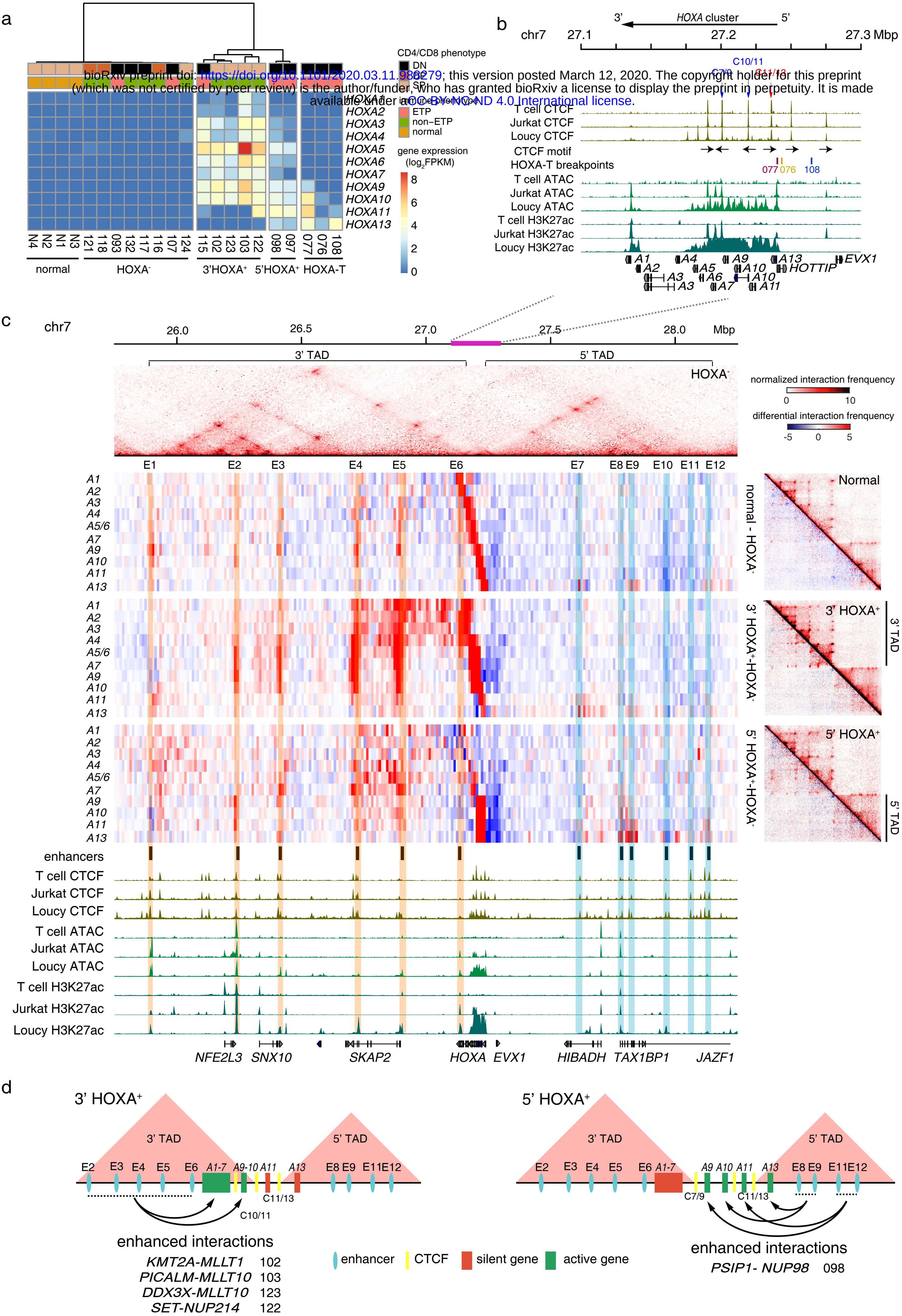

a
CD4/CD8 phenotype
DN
SP
immunophenotype
ETP
non−ETP
normal
gene expression
(log2FPKM)
8
6
4
2
0
HOXA1
HOXA2
HOXA3
HOXA4
HOXA5
HOXA6
HOXA7
HOXA9
HOXA10
HOXA11
HOXA13
N4 N2 N1 N3 121 118 093 132 117 116 107 124 115 102 123 103 122 098 097 077 076 108
normal
HOXA-
3'HOXA+
5'HOXA+
HOXA-T
b
chr7
27.1
27.2
27.3 Mbp
3'
HOXA cluster
5'
C7/9
C10/11
C11/13
T cell CTCF
Jurkat CTCF
Loucy CTCF
CTCF motif
HOXA-T breakpoints
077
076
108
T cell ATAC
Jurkat ATAC
Loucy ATAC
T cell H3K27ac
Jurkat H3K27ac
Loucy H3K27ac
A1 A2 A3 A3 A4 A5 A6 A7 A9 A10 A10 A11 A13 HOTTIP EVX1
c
chr7
26.0
26.5
27.0
27.5
28.0
Mbp
3' TAD
5' TAD
HOXA-
normalized interaction frenquency
0
10
differential interaction frenquency
-5
0
5
E1 E2 E3 E4 E5 E6 E7 E8 E9 E10 E11 E12
A1 A2 A3 A4 A5/6 A7 A9 A10 A11 A13
normal - HOXA-
3' HOXA+-HOXA-
5' HOXA+-HOXA-
Normal
3' HOXA+
3' TAD
5' HOXA+
5' TAD
enhancers
T cell CTCF
Jurkat CTCF
Loucy CTCF
T cell ATAC
Jurkat ATAC
Loucy ATAC
T cell H3K27ac
Jurkat H3K27ac
Loucy H3K27ac
NFE2L3 SNX10 SKAP2 HOXA EVX1 HIBADH TAX1BP1 JAZF1
d
3' HOXA+
3' TAD
5' TAD
E2 E3 E4 E5 E6 A1-7 A9-10 A11 A13 E8 E9 E11 E12
C11/13
C10/11
enhanced interactions
KMT2A-MLLT1 102
PICALM-MLLT10 103
DDX3X-MLLT10 123
SET-NUP214 122
enhancer
CTCF
silent gene
active gene
5' HOXA+
3' TAD
5' TAD
E2 E3 E4 E5 E6 A1-7 A9 A10 A11 A13 E8 E9 E11 E12
C7/9
C11/13
enhanced interactions
PSIP1- NUP98 098

Fig. 4 Chromatin interaction profile and expression patterns of the HOXA cluster in T-ALLs. (a) T-ALL subtypes with different *HOXA* gene expression patterns. HOXA-translocation (HOXA-T) negative cases are grouped by unsupervised hierarchical clustering based on the *HOXA* gene expressions. (b) ChIP-seq tracks of CTCF and H3K27ac and ATAC-seq tracks in T cell, HOXA$^-$ Jurkat and HOXA$^+$ Loucy T-ALL cells in the genomic region corresponding to the pink bar of Fig. 4c. CTCF motif orientations, 5'TAD and 3'TAD boundaries and HOXA-T breakpoints are also included. (c) Top: a Hi-C heatmap shows the average interaction intensity of HOXA$^-$ cases in chr7: 25,750,000-28,250,000 (hg19), which includes *HOXA* gene cluster and its 3' and 5' TADs. Middle left: from top to bottom, Hi-C heatmaps show differential interaction intensities between normal T cell vs. HOXA$^-$, 3' HOXA$^+$ vs. HOXA$^-$ and 5' HOXA$^+$ vs. HOXA$^-$ cases using visual 4C plot. Each line represents interactome of a *HOXA* gene shown on the left. *HOXA5* and *HOXA6* share one line as they located in the same bin. The main enhancers are highlighted with orange in the 3' TAD and blue in the 5' TAD. Middle right: Top right of each heatmap represents the averaged interaction intensities of normal T cell, 3' HOXA$^+$ cases and 5' HOXA$^+$ cases, respectively; bottom left of each heatmap represents the differential interaction intensities between each group and HOXA$^-$ cases. Bottom: ChIP-seq tracks for CTCF and H3K27ac and ATAC-seq tracks in T cell, HOXA$^-$ Jurkat and HOXA$^+$ Loucy T-ALL cells. (d) Schematic illustrations of the associations among 3D genomic interactions, *HOXA* cluster expressions and associated fusion events in 3' HOXA cases (left) and 5' HOXA cases (right).

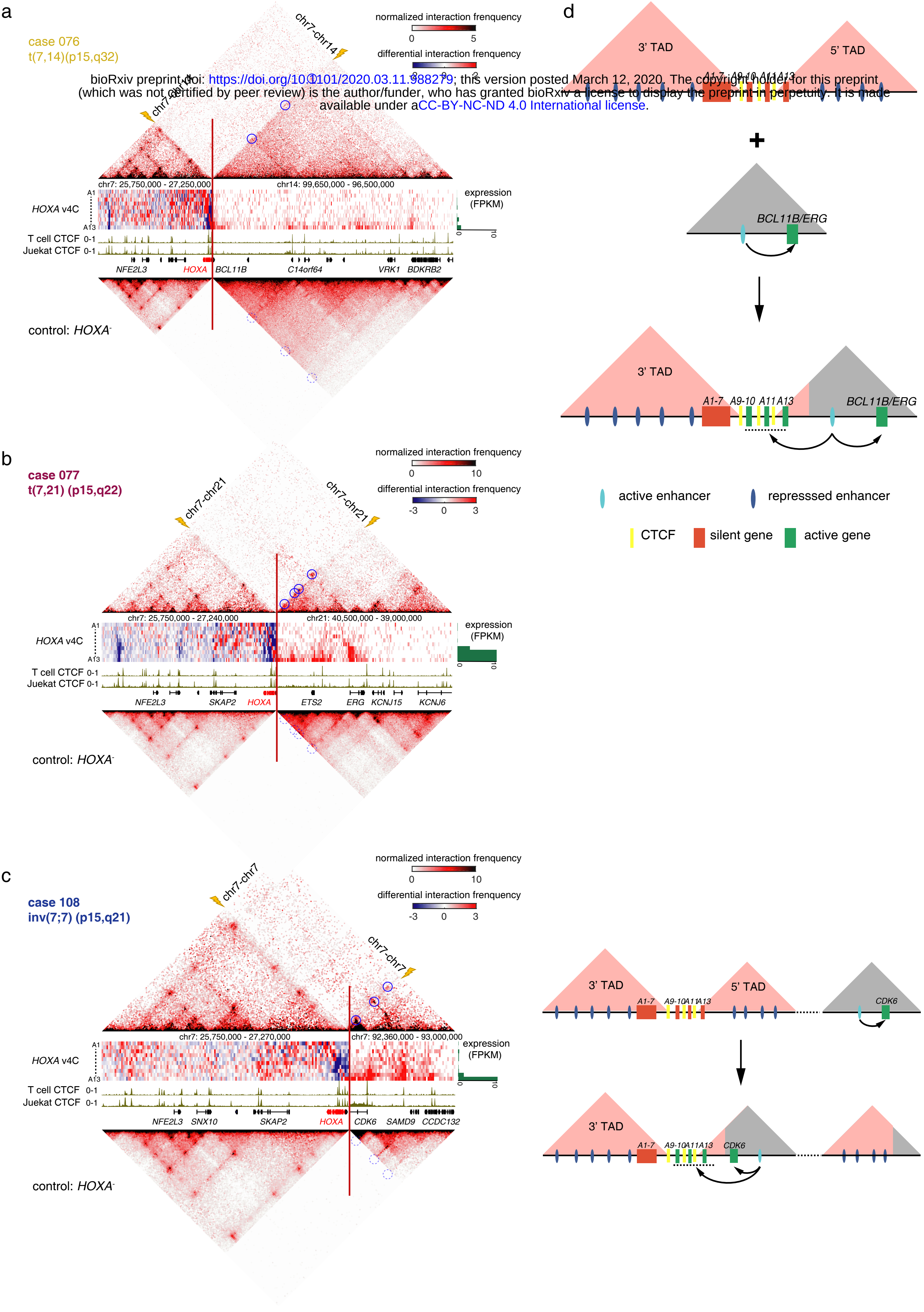
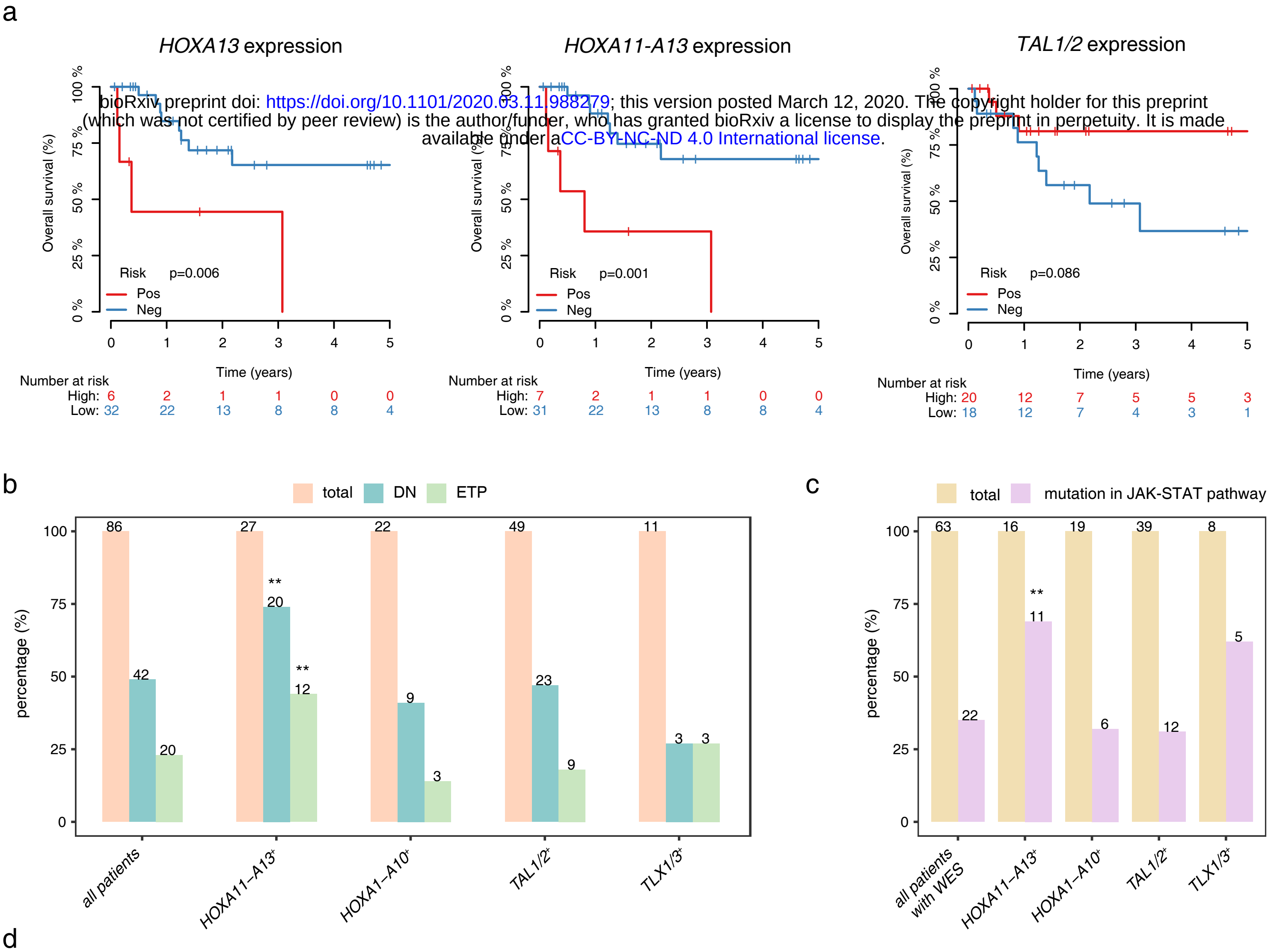

a
case 076
t(7,14)(p15,q32)
chr7-chr14
chr7-chr14
normalized interaction frenquency
0
5
differential interaction frenquency
-2
0
2
chr7: 25,750,000 - 27,250,000
chr14: 99,650,000 - 96,500,000
A1
HOXA v4C
A13
expression
(FPKM)
0
10
T cell CTCF 0-1
Juekat CTCF 0-1
NFE2L3
HOXA
BCL11B
C14orf64
VRK1
BDKRB2
control: HOXA-
b
case 077
t(7,21) (p15,q22)
chr7-chr21
chr7-chr21
normalized interaction frenquency
0
10
differential interaction frenquency
-3
0
3
chr7: 25,750,000 - 27,240,000
chr21: 40,500,000 - 39,000,000
A1
HOXA v4C
A13
expression
(FPKM)
0
10
T cell CTCF 0-1
Juekat CTCF 0-1
NFE2L3
SKAP2
HOXA
ETS2
ERG
KCNJ15
KCNJ6
control: HOXA-
c
case 108
inv(7;7) (p15,q21)
chr7-chr7
chr7-chr7
normalized interaction frenquency
0
10
differential interaction frenquency
-3
0
3
chr7: 25,750,000 - 27,270,000
chr7: 92,360,000 - 93,000,000
A1
HOXA v4C
A13
expression
(FPKM)
0
10
T cell CTCF 0-1
Juekat CTCF 0-1
NFE2L3
SNX10
SKAP2
HOXA
CDK6
SAMD9
CCDC132
control: HOXA-
d
3' TAD
5' TAD
A1-7
A9-10
A11
A13
+
BCL11B/ERG
3' TAD
A1-7
A9-10
A11
A13
BCL11B/ERG
active enhancer
represssed enhancer
CTCF
silent gene
active gene
3' TAD
A1-7
A9-10
A11
A13
5' TAD
CDK6
3' TAD
A1-7
A9-10
A11
A13
CDK6

Fig. 5. Translocation-mediated enhancer hijacking and ectopic *HOXA* gene expressions in T-ALLs. (a-c) Individual Hi-C heatmaps for three HOXA-T cases (top) and averaged Hi-C heatmaps of HOXA$^-$ control cases (bottom). Middle panels are differential visual 4C plots between HOXA-T cases and HOXA$^-$ using *HOXA1-A13* genes (from top to bottom) as the viewpoint. The green bar graphs on the right show the expression level of each *HOXA* gene. Breakpoints are marked by the red lines and yellow lightning bolts. Translocation-mediated loops are highlighted by blue circles, corresponding locations in control samples are marked by dotted circles. (d) Schematic illustrations of the consequences of *HOXA*-related translocations (upper) and inversion (lower).



d

	Univariate			Multivariate (HOXA13)			Multivariate (HOXA11-13)		
	HR	95% CI	P value	HR	95% CI	P value	HR	95% CI	P value
HOXA13 expression (FPKM) ≥ 1 vs. < 1	4.675	1.39-15.69	0.013	38.25	1.37-1071	0.032			
HOXA11-13 expression (FPKM) ≥ 1 vs. < 1	5.872	1.84-18.74	0.003				8.84	0.82-95.24	0.072
Gender Males vs. Female	1.13	0.42-3.01	0.806	0.959	0.05-17.78	0.978	0.89	0.066-12.015	0.93
Age Pediatric vs. Adult	0.21	0.07-0.64	0.006	0.154	0.006-3.684	0.248	0.34	0.014-8.444	0.512
WBC Low vs. High	1.46	0.62-3.43	0.381	73.52	1.59-3390	0.028	31.95	1.5-680.5	0.026
Hemoglobin	1	0.99-1.02	0.571	1.02	0.94-1.11	0.574	1.008	0.94-1.077	0.813
Platelet	1	0.99-1	0.952	0.986	0.968-1.004	0.127	0.987	0.971-1.004	0.139
Hepatosplenomegaly Yes vs. No	1.47	0.34-6.33	0.605	0.24	0.004-15.882	0.504	1.67	0.022-128.414	0.817
Blasts in BM	1.01	0.98-1.03	0.621	1.107	0.906-1.352	0.319	1.08	0.9-1.294	0.407

Fig. 6 Ectopic *HOXA11-A13* expressions are correlated with poor outcomes in pediatric and young adult T-ALLs. (a) Kaplan-Meier overall survival curves for patients with (red) and without (blue) ectopic *HOXA13* expression (left), *HOXA11-A13* expressions (middle) and *TAL1/TAL2* expressions (right) in pediatric and young adult patients. (b) The associations between ectopic *HOXA11-A13* expression and DN and ETP phenotypes. *P*-values are calculated by Fisher exact test (**: p<0.01). (c) The proportion of cases with JAK-STAT pathway mutations in each T-ALL subgroup, *P*-values are calculated by Fisher exact test (**: p<0.01). (d) Univariable and multivariable analysis of overall survival. HR, hazard ratio; CI, confidence interval.