



OPEN

Epigenetic alteration contributes to the transcriptional reprogramming in T-cell prolymphocytic leukemia

Shulan Tian¹✉, Henan Zhang², Pan Zhang^{3,7}, Michael Kalmbach³, Jeong-Heon Lee⁴, Tamas Ordog⁵, Paul J. Hampel⁶, Timothy G. Call⁶, Thomas E. Witzig⁶, Neil E. Kay⁶, Eric W. Klee¹, Susan L. Slager¹, Huihuang Yan¹ & Wei Ding⁶✉

T cell prolymphocytic leukemia (T-PLL) is a rare disease with aggressive clinical course. Cytogenetic analysis, whole-exome and whole-genome sequencing have identified primary structural alterations in T-PLL, including inversion, translocation and copy number variation. Recurrent somatic mutations were also identified in genes encoding chromatin regulators and those in the JAK-STAT signaling pathway. Epigenetic alterations are the hallmark of many cancers. However, genome-wide epigenomic profiles have not been reported in T-PLL, limiting the mechanistic study of its carcinogenesis. We hypothesize epigenetic mechanisms also play a key role in T-PLL pathogenesis. To systematically test this hypothesis, we generated genome-wide maps of regulatory regions using H3K4me3 and H3K27ac ChIP-seq, as well as RNA-seq data in both T-PLL patients and healthy individuals. We found that genes down-regulated in T-PLL are mainly associated with defense response, immune system or adaptive immune response, while up-regulated genes are enriched in developmental process, as well as WNT signaling pathway with crucial roles in cell fate decision. In particular, our analysis revealed a global alteration of regulatory landscape in T-PLL, with differential peaks highly enriched for binding motifs of immune related transcription factors, supporting the epigenetic regulation of oncogenes and genes involved in DNA damage response and T-cell activation. Together, our work reveals a causal role of epigenetic dysregulation in T-PLL.

T cell prolymphocytic leukemia (T-PLL) is a rare disease with a male predominance, representing ~2% of the mature lymphocytic leukemias in adults^{1,2}. T-PLL is currently considered incurable. As an aggressive disease, T-PLL patients show a rapid clinical course with a median survival of < 1 year³. An overall response of > 90% can be observed in immunotherapy with the monoclonal anti-CD52 antibody alemtuzumab, but relapse is inevitable due to resistance^{1,2}. Relapsed T-PLL typically progresses rapidly with final demise. Thus, the prognosis for patients diagnosed with T-PLL remains far from satisfactory⁴. Recently, combination therapy with epigenetic agents such as pan-histone deacetylase inhibitors (HDACi) vorinostat/romidepsin was shown to overcome alemtuzumab resistance to some extent¹, suggesting the involvement of epigenetic mechanism.

Cytogenetic analysis^{2,3}, array comparative genomic hybridization^{2,4}, whole-exome sequencing (WES) and whole-genome sequencing (WGS)^{4,5} have revealed primary structural alterations in T-PLL, including inv(14)(q11q32) and t(14;14)(q11;q32) that juxtapose the oncogene *TCL1A* at 14q32 to the T cell receptor alpha and delta (TRA/TRD) locus at 14q11, and less frequently, t(X;14)(q28;q11) that juxtaposes the oncogene *MTCPI* (Xq28) to the TRA/TRD locus too. In addition, structural variations are commonly seen on chromosome 8, such

¹Division of Biomedical Statistics and Informatics, Department of Health Sciences Research, Mayo Clinic, Rochester, MN, USA. ²Department of Urology, Mayo Clinic, Rochester, MN, USA. ³Division of Information Management and Analytics, Department of Information Technology, Mayo Clinic, Rochester, MN, USA. ⁴Division of Experimental Pathology and Laboratory Medicine, Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN, USA. ⁵Department of Physiology and Biomedical Engineering, Mayo Clinic, Rochester, MN, USA. ⁶Division of Hematology, Mayo Clinic, Rochester, MN, USA. ⁷Present address: Illinois Informatics Institute, University of Illinois at Urbana-Champaign, Urbana, IL, USA. ✉email: tian.shulan@mayo.edu; Ding.Wei@mayo.edu

ID	Date of sample	Treatment*	Time*	Phenotype	Key FISH results
P1	1/28/16	CampathX2, romidepsin	< 1 year	CD3+, CD4+, CD45RA+, CD45RO-, CCR7+, CD69+	88.5% of nuclei had a rearrangement involving <i>TCL1A</i>
P2	9/29/13	Untreated, observed from 2009 to 2013	4 years	CD3, CD8+, CD45RO-, CD45RA+, CCR7+	35.5% nuclei with inv(14) or t(14;14) involving <i>TCL1A</i> , also with 11q- in 60% nuclei
P3	4/13/16	Campath, methotrexate, pentostatin, pralatrexate. CHOP, Benda	< 2 months	CD3+, CD4+, CD45RA+, CCR7+, CD62L+	74% nuclei has <i>TCL1A</i> rearrangement
P4	2012	Untreated, found in 2008, observed to 2012	4 years	CD3+, CD4+, CD45RA+, CCR7+, CD62L+	70% nuclei had rearrangement involving <i>TCL1A</i> , either inv(14) or t(14;14)
P5	7/14/14	Untreated, but in half a year required therapy for skin rash and cytopenia	0.5 year	CD3+, CD8+, CD45RO+, CCR7+, CD62L+	14q32.1 (<i>TCL1A</i> sep), abnormal in 90.0% nuclei
P6	9/11/15	Untreated, required therapy in 2 months	2 months	CD3+, CD4+, CD45RO+, CCR7+, CD62L+	82% of nuclei had a rearrangement involving <i>TCL1A</i> , 93.0% of nuclei with additional copies of 5'TCRB region, and 68.0% of nuclei with an additional copy of <i>MYC</i> (8q24)
P7	1/23/12	Untreated, with 2 months started alemtuzumab and pentostatin, followed by HyperCVAD, Fludarabine and cyclophosphamide	2 months	CD3+, CD8+, CD45RA+, CCR7+, pCD62L+, pCD95+	98% nuclei rearrangement involved <i>TCL1A</i> , 83% nuclei had three signals for <i>MYC</i> (at 8q24.1), 17p- and del (11q)
P8	7/10/10	Untreated, within 5 months, started alemtuzumab and pentostatin	5 months	CD3+, CD4+, CD45RO+, CCR7+, pCD62L+, pCD95+	<i>TCL1</i> abnormal (83.5%)
P9	11/1/04	Untreated, within 2 months, started alemtuzumab, pentostatin and cyclophosphamide	2 months	CD3+, CD4+, CD45RA+, CCR7+, pCD62L+, pCD95+	46,X,-X,t(6;6)(q27;q15),inv(14)(q13q24),+mar[15]/

Table 1. Clinical information about the nine T-PLL cases. *Treatment history prior to sample collection. Patient P1 received a three-month romidepsin treatment, which was discontinued one month prior to sample collection. +Time from diagnosis to therapy.

as trisomy of 8q²⁻⁴, copy number gains on 8p³, i(8)(q10)²⁻⁴ and der(8)t(8;8)³. Copy number variations (CNVs) were also observed on several other chromosomes²⁻⁴.

WGS^{4,5}, WES^{4,5} and sequencing of PCR amplicons² have uncovered recurrent somatic events, including deletions and missense mutations in the tumor suppressor gene *TP53*², and deletion, missense and frameshift mutations in another tumor suppressor, *ATM*, which is a key regulator of DNA damage response^{2,4}. *ATM* is a serine/threonine kinase responsible for the phosphorylation of I κ B α , which functions as an inhibitor of the NF- κ B transcription factor (TF)⁶. A recent genomic and transcriptome analysis has revealed a central role of *ATM* aberrations in T-PLL pathogenesis⁵, likely through its role in DNA damage repair^{4,5}. In addition, largely mutually exclusive gain-of-function mutations were identified in *IL2RG*, *JAK1*, *JAK3* and *STAT5B* in the JAK-STAT signaling pathway, which were found to enhance cytokine-independent cell proliferation⁴.

Finally, mutations in several epigenetic genes encoding chromatin regulators were also reported, including missense mutations in the histone methyltransferase *EZH2*, which deposits H3K27 methylation⁴, in *TET2*, which hydroxylates 5-methylcytosine to 5-hydroxymethylcytosine³, and in *CREBBP*, which encodes the histone acetyltransferase CREB-binding protein⁴. Despite the identification of recurrent somatic variants, CNVs and genomic rearrangements, the complex mechanism of T-PLL pathogenesis is only partially understood. So far, histone modifications were analyzed only in promoter regions of two genes for repressive marks H3K9me3 and H3K27me3, and for pan-H4 acetylation using ChIP-qPCR¹. Thus, a lack of genome-wide epigenetic data has limited the investigation of the involvement of epigenetic mechanisms in this deadly disease.

In this study, using ChIP-seq data for H3K4me3 and H3K27ac that define promoters and active enhancers, respectively, we found global alterations of regulatory regions in T-PLL, with over 10% of the peaks showing differential occupancy between T-PLL and normal. Through total RNA and mRNA sequencing we identified 1672 and 2364 genes differentially expressed between T-PLL and normal samples, including *TCL1A*, *ATM* and several T-cell receptor regulators that are known to play key roles in T-PLL pathogenesis. Notably, for both histone marks, differential peaks are strongly linked to differential expression of the nearest genes, suggesting a role for epigenetic mechanism in transcriptional reprogramming in T-PLL. We characterized several well-known genes dysregulated in T-PLL in detail and demonstrated that the deregulation of those genes is associated with changes of local chromatin state at their promoters, enhancers, or both. These results have elucidated major epigenetic events and their possible causal roles in T-PLL pathogenesis.

Methods

T-PLL patients and healthy subjects. The Mayo Clinic patients who had a clinical diagnosis of T-PLL and donated their blood samples to Mayo Clinic chronic lymphocytic leukemia (Mayo Clinic IRB 1827-00) and lymphoma (Mayo Clinic IRB 118) tissue banks were used in this study. After screening all T-PLL samples in the above tissue banks, only six were stored with more than 50 million fresh peripheral blood mononuclear cells (PBMCs) that were used for ChIP-seq in this study (Table 1). Four of the six samples were used for total RNA sequencing; two of the six samples plus another three samples were used for mRNA sequencing. All but one of the nine T-PLL cases had the characteristic rearrangement involving *TCL1A*. The Mayo Clinic blood bank provided peripheral blood samples from three age-matched healthy individuals for ChIP-seq and total RNA sequencing and another four for mRNA sequencing (Mayo Clinic IRB 07-05543). Samples were collected with written consent and approval from the institutional review board at Mayo Clinic. We used PBMCs from the

T-PLL samples that are over 90% leukemic T cells. For the normal controls, we selected CD3 + enriched PBMCs. All experiments were performed in accordance with relevant guidelines and regulations.

RNA preparation and sequencing. For three normal (N1-N3) and four (P1, P3, P5 and P6) of the nine T-PLL patients, total RNA was extracted with Qiagen miRNeasy Mini Kit on the QIAcube according to manufacturer instructions. Total RNA libraries were prepared using Illumina TruSeq Stranded Total RNA Sample Prep Kit. For the other four normal (N4-N7) and five T-PLL patients (P2, P4, P7-P9), mRNA libraries were prepared using Illumina TruSeq RNA Sample Prep Kit v2. Libraries were sequenced from both ends to 101 bases on a HiSeq2000 platform through the Mayo Clinic Medical Genome Facility.

Chromatin immunoprecipitation and sequencing. H3K4me3 is largely located in the promoters and H3K27ac in the active enhancers. These two histone marks are widely used to map gene regulatory regions. We thus performed ChIP using anti-H3K27ac antibody (Abcam, Cat. # ab4729) and in-house generated anti-H3K4me3 antibody, as previously described⁷. We validated the H3K4me3 antibody by dot blotting with peptide, ChIP-qPCR with known positive and negative targets, and ChIP-seq⁷. ChIP-seq libraries were prepared from about 5 ng ChIP and input DNA using the ThruPLEX DNA-seq Kit V2 (Rubicon Genomics, Ann Arbor, MI). ChIP enrichment was validated using real-time PCR. Libraries were sequenced from both ends to 51 or 101 bases on a HiSeq4000 platform through the Mayo Clinic Medical Genome Facility.

RNA-seq data analysis. Raw reads were aligned to the hg19 reference genome using the spliced transcripts alignment to a reference program (STAR, v2.5.2b, <https://github.com/alexdobin/STAR>)⁸. Read counts per gene were estimated for the Ensembl gene annotation, using the featureCounts tool in the Subread package (v1.5.1, <https://subread.sourceforge.net>)⁹. Gene expression was quantified as reads per kilobase of exon per million mapped reads (RPKM). Protein-coding genes differentially expressed between T-PLL and normal were identified using the edgeR package (v3.16.5, <https://bioconductor.org/packages/edgeR>)¹⁰ at the cutoff of 5% false discovery rate (FDR) and threefold change. Briefly, genes without reads in all libraries were excluded; for the remaining genes, raw read counts were normalized using the trimmed mean of M-values (TMM) method within edgeR. We used conditional maximum likelihood method within edgeR to estimate common negative binomial dispersion across genes, which is proved to be effective for small sample size.

To functionally categorize the differentially expressed genes, pathway and gene set enrichment analyses were done separately for the genes up- and down-regulated in T-PLL. For pathway analysis, we used the ReactomePA package (v1.18.1, <https://bioconductor.org/packages/release/bioc/html/ReactomePA.html>)¹¹ that takes the Reactome database¹². For gene set enrichment analysis, we used the GOSTats package (v2.40.0, <https://bioconductor.org/packages/release/bioc/html/GOSTats.html>)¹³, which calculates p value for the overrepresentation of gene sets in Gene Ontology (GO) Biological Process and Molecular Function.

To highlight the gene expression changes between T-PLL and normal and the level of heterogeneity within T-PLL, we performed unsupervised clustering of expression levels for the top 10,000 protein-coding genes, using the pheatmap package (v1.0.8, <https://cran.r-project.org/web/packages/pheatmap/>) in R. Gene expression RPKM was log₂ transformed and quantile normalization using limma package (v 3.30.13, <https://bioconductor.org/packages/release/bioc/html/limma.html>). The 10,000 genes were selected from autosomes that had RPKM values ≥ 0.5 in at least one of the samples and showed the largest variation across samples. Total RNA and mRNA sequencing data were separated when used in the above differential analysis and unsupervised clustering.

ChIP-seq data analysis. ChIP-seq data were analyzed with the HiChIP pipeline with minor modifications¹⁴. In brief, reads were mapped to the hg19 reference genome using Burrows-Wheeler Aligner (BWA, v0.7.10, <https://bio-bwa.sourceforge.net>)¹⁵. After excluding read pairs with both ends being mapped uniquely or mapped to multiple locations, the remaining reads were re-mapped with Genomic Short-read Nucleotide Alignment Program (GSNAP, gmap_2015_0723, <http://research-pub.gene.com/gmap/archive.html>)¹⁶ to improve mapping rate. BWA and GSNAP alignments were combined and uniquely mapped pairs of reads with mapping quality ≥ 20 were extracted. Duplicates were removed using Picard (v1.67, <https://broadinstitute.github.io/picard/>). To visualize signal in the Integrative Genomics Viewer (IGV, v2.3.93, <http://software.broadinstitute.org/software/igv/>)¹⁷, we used BEDTools¹⁸ and in-house scripts to generate per-million read count over 200-bp sliding windows with 20-bp step size. H3K4me3 and H3K27ac peaks were identified using model-based analysis of ChIP-seq (MACS, v2.0.10, <https://pypi.org/project/MACS2/>)¹⁹, at the cutoff of ≤ 1% FDR and ≥ twofold change over the input. Peaks overlapping the blacklisted regions were excluded from further analysis. The lists of blacklisted regions were downloaded from <http://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeMapability/wgEncodeDacMapabilityConsensusExcludable.bed.gz> and http://mitra.stanford.edu/kundaje/akundaje/release/blacklists/hg19-human/Duke_Hg19SignalRepeatArtifactRegions.bed.gz.

To perform unsupervised clustering for H3K4me3 and H3K27ac peaks, we first merged peaks if they were overlapped by at least 1 bp. For each merged peak in each sample, we estimated the number of mapped pairs in the IP and input using BEDTools. Read counts in IP were input-subtracted, normalized to 10 M mapped reads, log₂ transformed and quantile normalized. We retained merged peaks that were from autosomal chromosomes and present in at least two samples. For H3K27ac, we focused on merged peaks that are located in the enhancers and removed those that overlapped the TSS ± 2.5 kb regions from protein-coding genes, as described^{20,21}. The top 20,000 merged peaks with the largest variation across samples were retained for unsupervised clustering. H3K4me3 peaks were processed similarly, except that we focused on those in the promoters of protein-coding genes, defined as the TSS ± 2 kb regions²². To view the H3K4me3 and H3K27ac changes for genes up- or

down-regulated in T-PLL, we generated aggregate per-million signal density profiles over TSS \pm 2 kb using ngsplot tool (v2.63, <https://github.com/shenlab-sinai/ngsplot>)²³.

We used the DiffBind R package (v2.14.0, <https://bioconductor.org/packages/release/bioc/html/DiffBind.html>)²⁴ to identify the changes of H3K27ac and H3K4me3 between the six T-PLL and three normal subjects, and between the four T-PLL samples and three normal controls that also have total RNA sequencing data. Only merged peaks that were present in \geq 2 samples were included. Data were TMM normalized and differential peaks were identified using the edgeR method within DiffBind, at the cutoff of \leq 5% FDR and \geq twofold change. Differential peaks were assigned to the nearest genes using the closestBed command in BEDTools.

Copy number alteration (CNA) analysis. CNA analysis was done according to²⁵ with minor modifications. Instead of using 1-Mb windows with 500-kb overlap and an absolute z-score cutoff of $> = 2$, we split each autosome into 100-kb windows with a step size of 10 kb, and accordingly used a more stringent cutoff of absolute z-score $> = 3$. For each of the normal and T-PLL cases, alignments from H3K4me3, H3K27ac and input libraries were combined and those in the peaks and blacklisted regions (see above) were excluded. BEDTools was used to estimate per window read pairs in each sample. Windows with no (6.37% of the windows) or less than 100 read pairs (0.87% of the windows) were excluded. The coverage for the remaining windows was calculated as: Coverage = (Window read counts scaled to 10 M)/(Effective window size in kb), where effective window size does not include the base pairs overlapping peaks and blacklisted regions. The coverage in each of the six T-PLL was normalized to the average of the median coverage from the three normal, and log₂ transformed. Z score was calculated using the above median-normalized coverage in each PLL sample. A copy number gain or loss was inferred if the z score $> = 3$ or $< = -3$, respectively. The differential H3K4me3 and H3K27ac peaks were assigned the CNA status of the nearest 100-kb window with the shortest distance between peak center and window middle point.

TF motif enrichment analysis. For differential H3K4me3 and H3K27ac peaks, we identified the nearest open chromatin regions by intersecting with ATAC-seq peaks from blood^{26,27} (see below). Sequences from ATAC-seq peak center \pm 50 bp were extracted and enrichment of known TF motifs was identified using the Homer package (v4.10, <http://homer.ucsd.edu/homer/>).

Public epigenomics data. We downloaded H3K27ac ChIP-seq reads alignments and MACS peaks from the Roadmap reference epigenomes²⁸, which are available at <http://egg2.wustl.edu/roadmap/data/byFileType>. Signal tracks were then generated from alignments using the HiChIP pipeline¹⁴. Raw sequence data for H3K27ac ChIP-seq in four normal B-cell samples were downloaded from SRA, under the accession GSM998996 and GSM998997²⁹, GSM2386722³⁰, and GSM1027287^{31,32}. H3K27ac signal tracks (tdf or bw files) were downloaded from GEO for seven mantle cell lymphoma (GSM1703947, GSM1703949, GSM1703951, GSM1703953, GSM2760324, GSM2760326, and GSM2760328)^{33,34}, three small lymphocytic lymphoma (GSM1703955, GSM1703957, and GSM1703959)³³, and seven high-grade B-cell lymphoma patient samples (GSM1703933, GSM1703935, GSM1703937, GSM1703939, GSM1703941, GSM1703943, and GSM1703945)³³, as well as 10 adult T-cell leukemia/lymphoma (ATL) patients (GSE85695)³⁵. ATAC-seq data were downloaded for 14 human primary blood cell types (GSE74912, 50 samples)²⁶ and 32 immune cell populations (GSE118189, 97 samples)²⁷. Peaks were identified through the HiChIP pipeline, and those with FDR $< = 1e-05$ were retained. Catalog of super-enhancers from 86 cell and tissue types was obtained from³⁶. Finally, to identify enhancer for MYC, we used the promoter capture Hi-C (CHi-C) data from 17 blood cell types³⁷; regions of significant interactions with CHiCAGO score \geq 5 are available at <https://osf.io/u8tzip/>.

Results

Transcriptional dysregulation in T-PLL. We performed total RNA sequencing in PBMCs from four T-PLL patients (P1, P3, P5, and P6) and three healthy individuals (N1-N3), and mRNA sequencing for another five T-PLL (P2, P4, and P7-P9) and four normal (N4-N7). Of the nine T-PLL patients, P1 and P3 had already received treatment when samples were collected, while the other seven patients were treatment naive (Table 1). In total RNA sequencing, a total of 12,375 protein-coding genes had RPKM \geq 0.5 in at least one of the samples. The top 10,000 most variable genes were used in unsupervised clustering, which revealed distinctive differences between T-PLL and normal samples (Fig. 1a). We observed noticeable heterogeneity of gene expression within T-PLL. In mRNA sequencing, a total of 13,728 protein-coding genes had RPKM \geq 0.5 in at least one of the samples. Unsupervised clustering using the top 10,000 most variable genes revealed a similar pattern of expression changes between T-PLL and normal and heterogeneity across T-PLL patients (Supplemental Fig. 1a).

Using total RNA sequencing, we identified 807 and 865 genes that were up- and down-regulated, respectively, in T-PLL compared to normal (Fig. 1b). In addition, mRNA sequencing revealed 1466 up- and 898 down-regulated genes in T-PLL (Supplemental Fig. 1b). Of those, 280 up- and 279 down-regulated genes were shared between the two RNA-seq datasets. mRNA sequencing identified 1.8 \times as many up-regulated protein-coding genes in T-PLL compared to total RNA sequencing. This discrepancy is likely due to difference in sequencing depth, platform-specific bias of the two RNA-seq protocols, and the heterogeneity among T-PLL cases. To gain a better understanding of the possible causes, we first estimated the sequencing depth by counting the number of reads mapped to all 57,773 genes and to the subset of 20,327 protein-coding genes. While normal samples had largely comparable number of reads mapped to protein-coding genes in both datasets, T-PLL samples indeed had lower coverage of protein-coding genes in total RNA sequencing (17.87–37.13 M) than in mRNA sequencing (41.52–52.74 M) (Supplemental Fig. 2a). Next, we assessed expression level for the protein-coding genes up-regulated in T-PLL between the two datasets, using gene maximum RPKM values across both normal and T-PLL cases. Up-regulated genes identified in both datasets showed similar levels of expression (Mann–Whitney

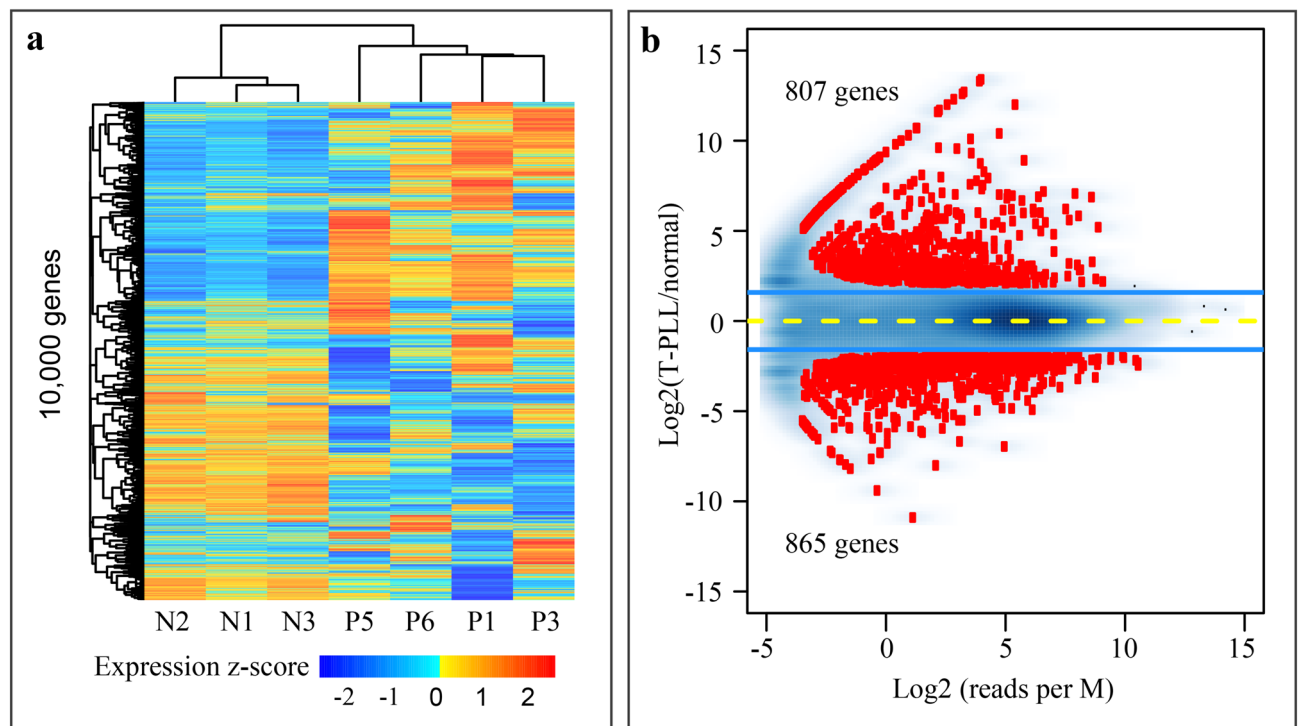


Figure 1. Gene expression changes in T-PLL. **(a)** Unsupervised clustering of expression level from the top 10,000 genes. Gene expression was quantified by total RNA sequencing as RPKM values and transformed to \log_2 scale with an offset of 0.1, following by quantile normalization. These genes were selected from autosomes that had RPKM ≥ 0.5 in at least one of the samples and showed the largest variation across samples. Unlike the three normal, the T-PLL cases showed remarkable heterogeneity in expression profile. N1–N3, normal individuals; P1, P3, P5 and P6, T-PLL patients. **(b)** Protein-coding genes differentially expressed between T-PLL and normal. Raw reads were aligned to the hg19 reference genome using STAR and the number of reads in exons per gene was estimated with featureCounts. The 1672 (807 + 865) differentially expressed genes were identified using edgeR at the cutoff of 5% FDR and threefold change. Y-axis, fold-change at the \log_2 scale; x-axis, sum of the normalized read count per M from the normal and T-PLL group.

U test $p = 0.09$, left panel in Supplemental Fig. 2b). Not surprisingly, for the genes identified as up-regulated in T-PLL by mRNA sequencing alone, they had much higher expression levels in mRNA sequencing than in total RNA sequencing ($p < 2.2e-16$, middle panel in Supplemental Fig. 2b), and vice versa ($p < 2.2e-16$, right panel in Supplemental Fig. 2b). It is known that genes with low counts are less likely to be significant in differential expression test due to a lack of statistical power^{38,39}.

For both datasets, Gene Ontology analysis indicated a remarkable enrichment of the down-regulated genes in immune response (Table 2 and Supplemental Table 1). Similarly, pathway enrichment analysis using the Reactome database revealed enrichment in immune and inflammatory responses, including the “chemokine receptors”, “immune system”, and “immunoregulatory interactions between a lymphoid and a non-lymphoid cell” (Table 2 and Supplemental Table 1). On the other hand, Gene Ontology analysis revealed that the up-regulated genes were enriched in “WNT-protein binding” and “WNT-activated receptor activity”, as well as in many other categories of developmental related biological processes (Supplemental Tables 2 and 3). This result indicates a similar role for WNT signaling network in T-PLL, as documented in B-cell chronic lymphocytic leukemia^{40–43} and T-lineage acute lymphoblastic leukemia⁴⁴.

Global alteration of regulatory regions in T-PLL. To understand the regulatory landscape and the possible roles of epigenetic mechanisms in T-PLL pathogenesis, we generated ChIP-seq data for H3K4me3 and H3K27ac, which are mainly located in the promoters and active enhancers, respectively. To reveal the extent of global similarities in H3K4me3 and H3K27ac distribution across normal and T-PLL, we estimated pairwise Pearson correlation using normalized read counts from merged peaks present in two or more samples (Supplemental Fig. 3a,b). As expected, the three normal samples were highly similar in the occupancy of both H3K4me3 ($R = 0.96–0.97$) and H3K27ac ($R = 0.90–0.92$). On the other hand, the six T-PLL were more divergent ($R = 0.82–0.92$ for H3K4me3 and $R = 0.65–0.86$ for H3K27ac), in particular for H3K27ac, suggesting an obvious heterogeneity of epigenomic landscape. Of the six T-PLL, P1 received romidepsin treatment one month before sample collection. Romidepsin is a histone deacetylase (HDAC) class I-selective agent for HDAC1, HDAC2, HDAC3, and HDAC8. Romidepsin treatment is through to increase histone acetylation, thus altering gene expression, in particular activating tumor suppressor genes^{25,45}. P1 is a non-responder to romidepsin, and was thus included in the analysis together with the other T-PLL. To support this, we assessed global enrichment of H3K27ac over the

Name	No. gene	Q value	Source
Immune System	214	7.34E-26	Reactome
Immunoregulatory interactions between a Lymphoid and a non-Lymphoid cell	43	1.31E-21	Reactome
Neutrophil degranulation	70	8.87E-14	Reactome
Innate Immune System	129	1.96E-13	Reactome
Adaptive Immune System	87	2.47E-09	Reactome
Cytokine Signaling in Immune system	80	7.29E-09	Reactome
Costimulation by the CD28 family	17	3.58E-06	Reactome
PD-1 signaling	10	4.66E-06	Reactome
Chemokine receptors bind chemokines	13	2.19E-05	Reactome
Interferon gamma signaling	18	2.69E-05	Reactome
Immune response (GO:0006955)	185	4.48E-38	GO
Immune system process (GO:0002376)	242	4.48E-38	GO
Cell activation (GO:0001775)	115	2.29E-24	GO
Defense response (GO:0006952)	162	5.41E-27	GO
Regulation of immune system process (GO:0002682)	158	3.03E-31	GO
Leukocyte activation (GO:0045321)	97	1.89E-21	GO
Regulation of immune response (GO:0050776)	110	4.00E-23	GO
Immune effector process (GO:0002252)	76	8.04E-13	GO
Leukocyte mediated immunity (GO:0002443)	41	4.75E-10	GO
Myeloid leukocyte activation (GO:0002274)	30	2.84E-10	GO

Table 2. Enriched pathways and GO terms for genes down-regulated in T-PLL. Gene expression was quantified by total RNA sequencing for three normal (N1-N3) and four T-PLL (P1, P3, P5, and P6).

respective input (Supplemental Fig. 4a), as well as the signal levels within consensus peaks shared either by all six T-PLL or by at least two (Supplemental Fig. 4b). P1 did not show clearly excess enrichment relative to the other T-PLL. Also, we examined the expression of tumor suppressor genes⁴⁶, revealing that P1 had the lowest median expression among the four T-PLL with total RNA sequencing data (Supplemental Fig. 4c).

Specifically, for H3K27ac we focused on distal peaks, defined as those that are ≥ 2.5 kb away from gene TSS; while for H3K4me3, we focused on proximal peaks in the promoter regions, defined as $TSS \pm 2.0$ kb. Using the top 20,000 most variable (out of 52,248) distal H3K27ac peaks that are present in two or more samples, unsupervised clustering identified a preferential loss of active enhancers in T-PLL (Fig. 2a). Similar results were obtained with the top 10,000 and 30,000 distal H3K27ac peaks (Supplemental Fig. 5a,b). There are 16,629 proximal H3K4me3 peaks within $TSS \pm 2.0$ kb. A similar pattern was identified for the 10,000 most variable H3K4me3 peaks (Fig. 2b). Therefore, there is a genome-wide change of regulatory landscape in T-PLL. As expected, we observed obvious heterogeneity in the occupancy of both H3K4me3 and H3K27ac, showing two sub-groups with P2, P4 and P5 belonging to one and the other three forming the second one (Fig. 2a,b and Supplemental Fig. 5a,b).

We next used the DiffBind package to identify the changes of H3K4me3 and H3K27ac between the four T-PLL (P1, P3 and P5-P6) and three normal that also had total RNA sequencing data. The analysis identified differential occupancy for 15.1% (9362) of the 62,016 H3K27ac peaks and 10.5% (3935) of the 37,306 H3K4me3 peaks (Table 3). As revealed by unsupervised clustering (Fig. 2a,b), there is a strong tendency of losing both marks in T-PLL compared to normal. To understand whether differential peaks are associated with the change of gene expression, we assigned them to the nearest genes and checked the expression changes of those genes. Indeed, for both marks, peaks that showed increased occupancy in T-PLL were preferentially associated with up-regulated in T-PLL, compared to non-differential peaks, and vice versa (enrichment: 3.8–19.2, chi-square test $p < 2.2e-16$) (Table 3). To correlate the changes of gene expression with those of proximal occupancy of H3K4me3 and H3K27ac, for the 865 down- and 807 up-regulated genes identified by total RNA sequencing, we also plotted the average signal profiles of H3K4me3 and H3K27ac over the $TSS \pm 2$ kb (Fig. 2c,d). As expected, overall, the down-regulated genes were associated with decreased H3K4me3 and H3K27ac, and vice versa; the correlated changes were more obvious for the up-regulated genes (Fig. 2d).

To understand which TFs might be involved in the perturbation of T-PLL epigenome, we first used diffbind to identify differential H3K4me3 and H3K27ac peaks on chr1-22 between the six T-PLL and three normal, followed by k-means clustering of the differential peaks. The analysis identified 2445 (1728 + 717) of the 2456 H3K27ac peaks with increased signal in T-PLL and all 5587 peaks with decreased signal as discriminatory peaks (Fig. 3a). To a less extent, 80% (647 + 124 + 129 = 900) of the 1128 H3K4me3 peaks with increased signal in T-PLL and 90% (2658) of the 2971 peaks with decreased signal were retained (Fig. 3b). The other differential peaks showed less difference between T-PLL and normal. We then used the Homer package to identify known TF motifs that are enriched in these discriminatory peaks (Fig. 3c,d). The differential H3K27ac peaks (Fig. 3c), as well as H3K4me3 peaks gained in T-PLL (upper panel in Fig. 3d) are enriched for binding motifs of E-twenty-six (ETS) TFs, such as ERG, ETS1, and ELF4. On the other hand, differential H3K4me3 peaks lost in T-PLL (lower panel in Fig. 3d) are particularly enriched for Interferon regulatory factor (IRF) TF motifs. Finally, for both marks, the differential

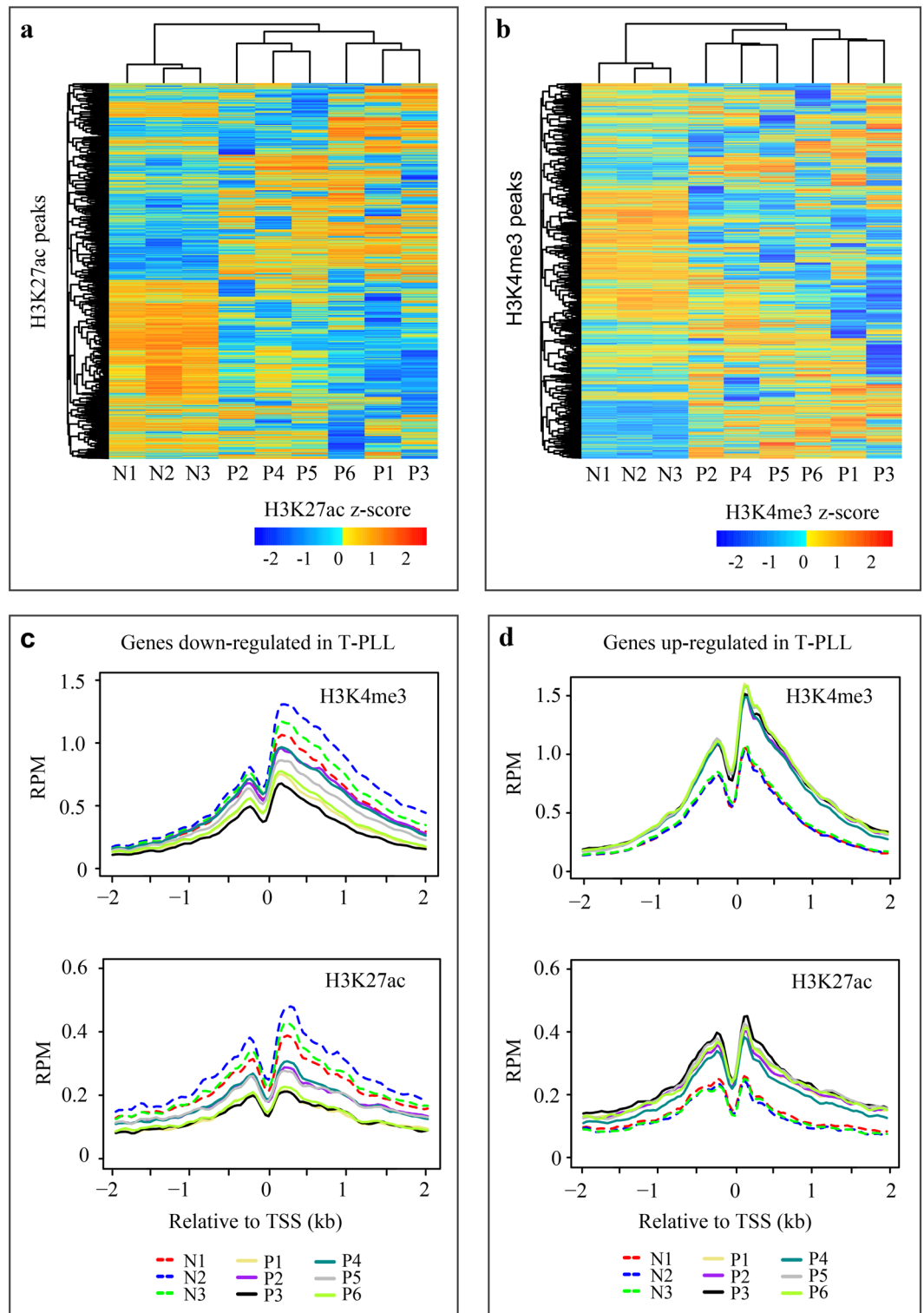


Figure 2. Global alteration of regulatory regions in T-PLL. **(a)** Unsupervised clustering of 20,000 H3K27ac peaks. Peaks from all nine samples were first merged if they were overlapped by at least 1 bp. For each merged peak, the input-subtracted read counts were normalized to 10 M mapped reads, log₂ transformed and quantile normalized. The top 20,000 merged peaks were selected from autosomes that were each present in at least two samples, not in the TSS \pm 2.5 kb regions, and had the largest between-sample variation. **(b)** Unsupervised clustering of 10,000 H3K4me3 peaks. The H3K4me3 peaks were processed similarly as above, except that only those in the TSS \pm 2 kb regions of protein-coding genes were selected. **(c)** Average read density profile over TSS \pm 2 kb for genes down-regulated in T-PLL. H3K4me3 and H3K27ac signals in 40-bp bins were estimated using the ngs.plot software. RPM, reads per M. **(d)** Average read density profile over TSS \pm 2 kb for genes up-regulated in T-PLL.

Mark	Region	Differential peaks			Non-differential peaks		Enrichment
		Type	Total	Linked to up-regulated gene	Total	Linked to up-regulated gene	
H3K27ac	> TSS ± 2.5 kb	Up in normal	5462	1001 (265)	35,788	1372 (323)	4.78
H3K27ac	> TSS ± 2.5 kb	Up in T-PLL	2833	583 (167)	35,788	1112 (287)	6.62
H3K27ac	≤ TSS ± 2.5 kb	Up in normal	524	212 (136)	16,866	355 (213)	19.22
H3K27ac	≤ TSS ± 2.5 kb	Up in T-PLL	543	166 (123)	16,866	270 (187)	19.1
H3K4me3	> TSS ± 2.0 kb	Up in normal	1959	272 (175)	17,596	635 (321)	3.83
H3K4me3	> TSS ± 2.0 kb	Up in T-PLL	1086	165 (102)	17,596	626 (300)	4.27
H3K4me3	≤ TSS ± 2.0 kb	Up in normal	742	212 (186)	15,775	471 (403)	9.57
H3K4me3	≤ TSS ± 2.0 kb	Up in T-PLL	148	85 (84)	15,775	509 (394)	17.8

Table 3. Association of differential H3K27ac and H3K4me3 occupancy with differential gene expression. Total RNA-seq differential expression analysis and ChIP-seq DiffBind were performed on four T-PLL (P1, P3, P5 and P6) and three normal (N1-N3). Peaks were split into those that showed differential occupancy between T-PLL and normal and others that showed no differential occupancy based on DiffBind analysis (FDR ≤ 0.05, fold-change ≥ 2). Peaks were assigned to the nearest genes. For differential peaks showing increased occupancy in normal (“Up in normal”), the number in parentheses indicates the number of up-regulated genes in normal. For differential peaks showing increased occupancy in T-PLL (“Up in T-PLL”), the number in parentheses indicates the number of up-regulated genes in T-PLL. Enrichment level is estimated as: (Number of differential peaks linked to differentially-expressed genes/total differential peaks) / (Number of non-differential peaks linked to differentially-expressed genes/total non-differential peaks).

peaks lost in T-PLL are enriched for basic-region leucine zipper (bZIP) TF motifs including BATF, AP-1 and JUNB (Fig. 3c,d, in blue) that are not enriched in peaks gained in T-PLL. In fact, many of them are immune related TFs, including lineage-specific TFs such as SCL, PU.1, and TCF12^{47,48}. For example, BATF, JUNB, and PU.1 interact with IRF4, which plays critical roles in lymphoid development and immune response⁴⁹.

To reveal the extent by which CNA may contribute to epigenomic change, we identified CNAs in individual T-PLL cases based on sequence coverage in 100-kb windows along chr1-22. At the cutoff of absolute z-score > 3, there are on average 0.34% and 0.37% of the windows that showed CN gain and loss, respectively. However, peaks are > sevenfold over-represented in the windows with CN gain compared to those with CN loss. Specifically, for the consensus peaks shared by at least two T-PLL, 0.78–5.20% of the H3K27ac peaks and similar proportions of the H3K4me3 peaks (0.66–4.82%) fall in windows with CN gain, versus ≤ 0.6% in windows with CN loss (Supplemental Table 4). This bias is unlikely due to peak calling, as a similar trend was also observed for additional two consensus peak sets of each mark, i.e., those present in > = 2 T-PLL and > = 2 normal, as well as those present in > = 2 normal but absent from all T-PLL. In these additional peak sets, < 0.5% of peaks overlap windows with CN loss, changing little (< 1%) at a less stringent cutoff of z-score < = - 2. The results indicate that regions with CN loss are less likely to be enriched with H3K4me3 and H3K27ac modifications. Finally, we investigated the overlap of differential peaks with windows carrying CNAs (Supplemental Table 4). Intersecting both marks with windows carrying CN gain revealed overlap with about 3–4% the peaks with increased signal in T-PLL, versus with 0.6% (H3K27ac) and 1.94% (H3K4me3) of the peaks with decreased signal in T-PLL. Thus, even for peaks falling within windows with CN gain, some showed reduced histone modification levels in T-PLL instead. On the other hand, similar to the whole peak set, differential peaks are much less associated with windows with CN loss. It is possible that, with a large (100-kb) window size and a modest coverage, small CNAs overlapping differential peaks may not have been identified, and that this method may be less effective in the detection of CN loss compared to that of CN gain. Nevertheless, CNAs were found to contribute to the changes of H3K4me3 and H3K27ac in a small portion of peaks, including those associated with key genes (such as *MYC* and *AGO2*) described below.

Changes in chromatin states of key dysregulated genes. As revealed by a recent study⁵ and our RNA-seq data described above, T-PLL is characterized by the dysregulation of several oncogenes, such as *TCL1A*, *MYC* and *EZH2*, as well as genes implicated in DNA damage repair (e.g., *ATM*) and T-cell receptor regulation (such as *CTLA4* and SLAM family members). Below we focused on these key genes. Specifically, we aimed to illustrate the epigenetic state in the regulatory regions of these genes and to understand how epigenetic alterations have modulated the transcriptional outcome.

The hallmark of genetic alterations in T-PLL includes t(14;14)(q11;q32) and inv(14)(q11q32) that involve the proto-oncogene *TCL1A*. *TCL1A* activates the serine/threonine kinase AKT that is implicated in cell survival regulation, as well as the NF-κB signaling pathway that plays key roles in the initiation and promotion of cancer^{6,50}. Increased *TCL1A* expression is associated with less favorable clinical outcome in chronic lymphocytic leukemia^{6,51}. *TCL1A* was significantly up-regulated in T-PLL compared to normal ($\log_2FC = 5.60$ and $p = 3.30e-08$ in total RNA-seq; $\log_2FC = 7.87$ and $p = 1.08e-18$ in mRNA-seq) (Fig. 4a, right panel). All six T-PLL cases demonstrated a gain of H3K4me3 and H3K27ac over its promoter and gene-body (Fig. 4a), which, we believe, contributes to *TCL1A* activation in addition to the known rearrangement of *TCL1A* locus. Based on the signal level in the corresponding input, at least two of the T-PLL samples (P1 and P5) showed copy number gains in *TCL1A*, which may have also resulted in increase of *TCL1A* expression, as well as of H3K4me3 and H3K27ac occupancy.

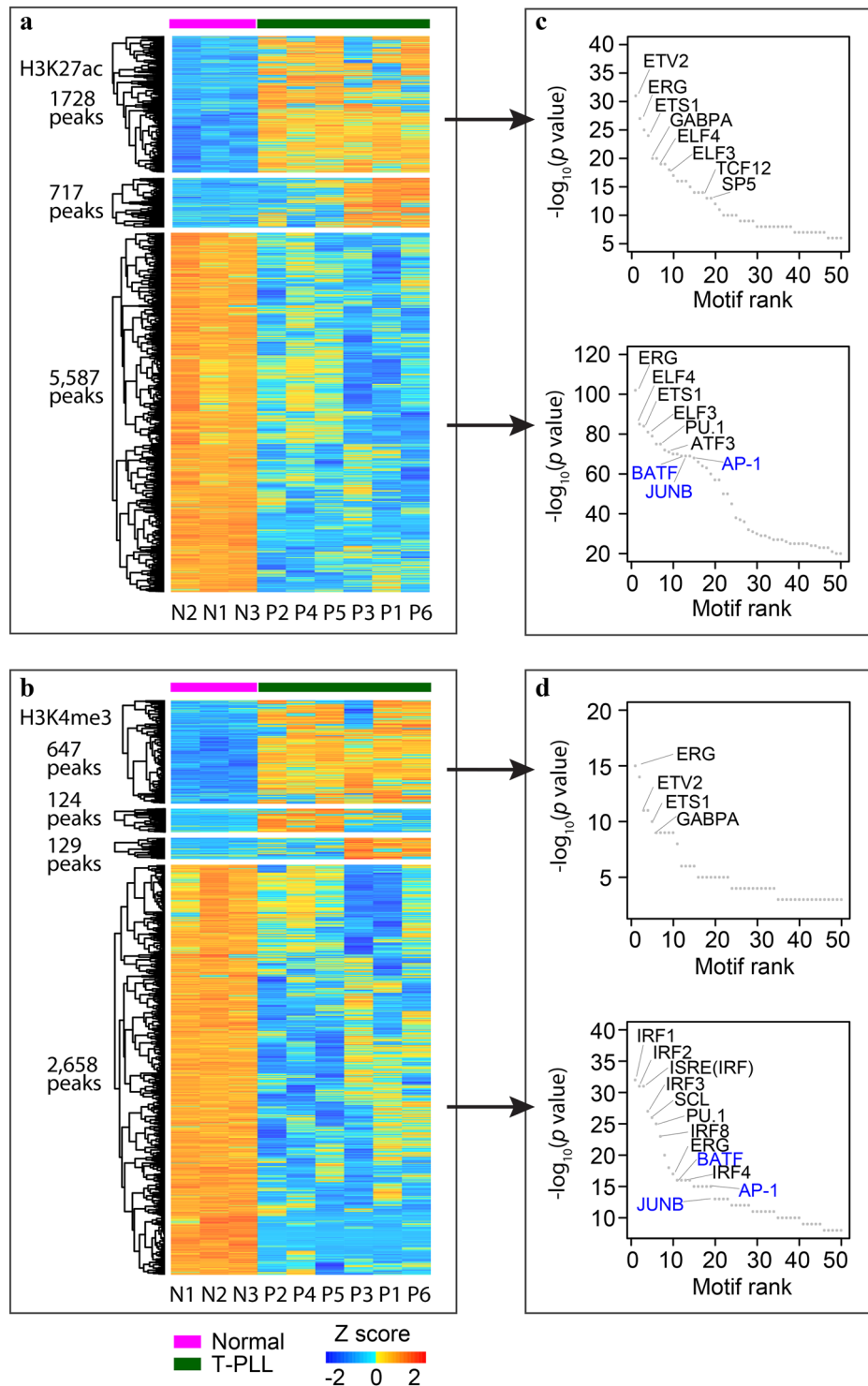


Figure 3. Heatmap of differential peaks and enriched TF binding motifs. (a) H3K27ac peaks with increased (upper panel) or decreased (lower panel) signal in T-PLL. There are 717 peaks (middle panel) that show increased occupancy in three of the T-PLL (P3/P1/P6). (b) H3K4me3 peaks with increased (upper panel) or decreased (lower panel) signal in T-PLL. There are 124 and 129 peaks (two middle panels) that show increased signals in P2/P4/P5 and P3/P1/P6, respectively. Differential peaks were identified using the diffbind package, using merged peaks present in at least two of the nine samples. Discriminatory peaks identified in k-means clustering were shown in the heatmap. K-means clustering and heatmap were performed using the number of reads per kb per 10 M, after input subtraction, log₂ transformation and quantile normalization. (c) TF motifs enriched in the differential H3K27ac peaks gaining (upper panel) or losing occupancy (lower panel) in T-PLL. (d) TF motifs enriched in the differential H3K4me3 peaks gaining (upper panel) or losing occupancy (lower panel) in T-PLL. Top 50 motifs with the lowest p values from Homer analysis (<http://homer.ucsd.edu/homer/>) were displayed. Basic leucine zipper TF family members BATE, JUNB and AP-1 were shown in blue.

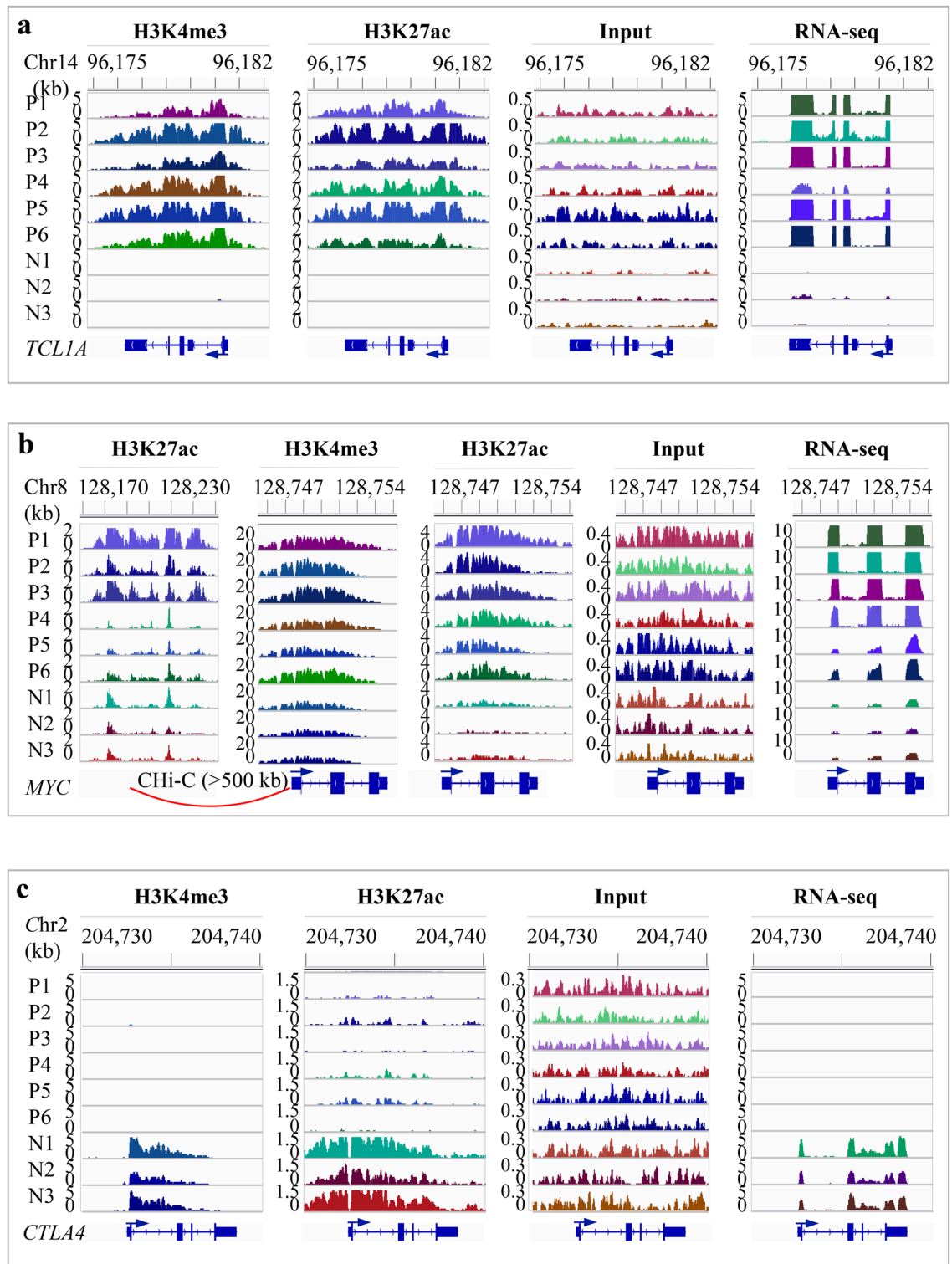


Figure 4. IGV snapshots showing the changes of histone marks and gene expression in T-PLL. **(a)** Increased expression of *TCLA1* and the associated gains of H3K4me3 and H3K27ac. *TCLA1* was actively expressed in T-PLL but not in normal. All T-PLL cases gained both H3K4me3 and H3K27ac peaks (~6-kb) that covered the promoter and gene body. Based on the read density in the inputs, there are copy number gains in two T-PLL (P1 and P5). **(b)** Increased expression of *MYC* and the associated increase of H3K4me3 and H3K27ac. *MYC* expression was up-regulated in T-PLL. There was an overall increase in H3K4me3 and H3K27ac occupancy over an ~6 kb region that covers the promoter and gene body. The signal level in the inputs suggested a copy number gain in P1, P5 and P6. Promoter capture Hi-C data from 17 blood cell types suggested that an ~50-kb enhancer located >500 kb upstream interacted with *MYC* promoter in total B, naïve B, fetal thymus and five types of T cells. This enhancer showed a marked increase of H3K27ac in three of the T-PLL (P1-P3) compared to the normal. **(c)** Markedly reduced expression of *CTLA4* and the associated loss of H3K4me3 and H3K27ac. The signal in the inputs suggested no copy number alteration within *CTLA4* in T-PLL. N1-N3, normal; P1-P6, T-PLL. Y-axis indicates the number of reads per 200-bp (for ChIP-seq and input) or per base pair (for RNA-seq).

To examine whether this ~6-kb H3K27ac peak exists in other tissue and cell types, we first checked the 98 human reference epigenomes that have H3K27ac data²⁸. H3K27ac enrichment was detected only in a primary B-cell sample and in GM12878, which is a B cell-derived lymphoblastoid cell line. Based on other publicly available H3K27ac ChIP-seq data we collected, this broad peak was identified in four of the five normal B-cell samples, in patient biopsies from all seven mantle cell lymphoma, three small lymphocytic lymphoma and four of the seven high-grade B-cell lymphoma subjects (Supplemental Fig. 6), but not in any of the 10 adult T-cell leukemia/lymphoma (ATL) patients³⁵. In addition, based on the catalog of super-enhancers from 86 cell and tissue types³⁶, this enhancer is part of a B-cell specific super-enhancer. These results suggest a possibility that requisition of this cell-type restricted enhancer activates the oncogenic expression of *TCL1A* in T-PLL.

Oncogenic activation of *MYC* in T cell leukemia⁵² and B cell lymphoma³³ involves interaction between its promoter with subtype specific enhancers. T-PLL is also characterized by the activation of *MYC*⁵. We detected up-regulation of *MYC* expression ($\log_2FC = 2.83$ and $p = 5.13e-03$ in total RNA-seq; $\log_2FC = 3.37$ and $p = 9.49e-06$ in mRNA-seq). To support a role for epigenetic mechanism in *MYC* activation, ChIP-seq data showed an overall increase of H3K4me3 and H3K27ac occupancy over an ~6 kb region in all T-PLL cases that covers *MYC* promoter and gene body (Fig. 4b). The signal in the inputs suggested copy number gains in P1, P5 and P6. To identify the enhancer that potentially interacts with the *MYC* promoter, we used the promoter ChIP-C data from 17 blood cell types³⁷. The analysis identified an ~50-kb enhancer located over 500 kb upstream that interacted with the *MYC* promoter in total B, naïve B, fetal thymus and five of the six types of T cells (Fig. 4b). This broad enhancer showed marked increase of H3K27ac in three of the T-PLL cases (P1-P3). Thus, in these three cases, the up-regulation of *MYC* expression appears to be driven by the increased H3K27ac in the enhancer together with increased H3K4me3 and H3K27ac in the promoter. For the other three cases (P4-P6), the increased *MYC* expression is associated with a gain of H3K27ac in the promoter in all three, and a gain of H3K4me3 in P6.

EZH2 is an oncogene encoding a histone methyltransferase that deposits H3K27me3, a histone mark associated with transcriptional repression⁵³. *EZH2* is highly expressed in several B-cell lymphomas⁵³. We found that *EZH2* was up-regulated in four of the T-PLL samples (P2-P4 and P6) (Supplemental Fig. 7). Three of the four (P3, P4, and P6) also showed increased occupancy of both H3K4me3 and H3K27ac in the promoter (Supplemental Fig. 7), with the exception of P2 that had a copy-number loss indicated by the lower signal in the input. Overall, the elevated *EZH2* expression in T-PLL is associated with increased H3K4me3 and H3K27ac in its promoter.

In the argonaute (AGO) family, only *AGO2* has the catalytic activity and plays an essential role in small RNA-guided gene silencing⁵⁴. Often up-regulated in cancer, *AGO2* is involved in chromatin remodeling and alternative splicing⁵. *AGO2* was up-regulated in P1 and P2 (Supplemental Fig. 8). All the six T-PLL samples showed increased H3K4me3 and H3K27ac in its promoter (Supplemental Fig. 8). The signal level in inputs suggested that four (P1, P3, P5 and P6) had copy-number gains, which likely have contributed to the increased H3K4me3 and H3K27ac in these cases. For P2 that showed no copy number alteration, its up-regulation may be directly due to the increased H3K4me3 and H3K27ac.

On the other hand, we also checked the chromatin states for several T-cell receptor regulators and *ATM*. As a major negative regulator of T-cell response, *CTLA4* was markedly down-regulated in T-PLL ($\log_2FC = -6.96$ and $p = 8.46e-14$ in total RNA-seq; $\log_2FC = -6.27$ and $p = 1.08e-16$ in mRNA-seq), which was associated with an obvious reduction of H3K4me3 and H3K27ac (Fig. 4c). Input signal indicated no CNV in any of the T-PLL cases. The SLAM family members also act as the negative regulators of T-cell response. Of the nine members, seven reside in a 378-kb region on chromosome 1 (Supplemental Fig. 9). Of the seven, *CD84* (also known as *SLAMF5*, $\log_2FC = -3.69$ and $p = 2.22e-05$ in total RNA-seq; $\log_2FC = -4.47$ and $p = 2.85e-10$ in mRNA-seq), *SLAMF7* ($\log_2FC = -3.12$ and $p = 4.44e-04$ in total RNA-seq; $\log_2FC = -3.78$ and $p = 7.42e-08$ in mRNA-seq) and *CD244* (*SLAMF4*, $\log_2FC = -2.17$ and $p = 2.26e-02$ in total RNA-seq; $\log_2FC = -2.34$ and $p = 1.16e-03$ in mRNA-seq) were down-regulated in all T-PLL cases. Accordingly, these three genes showed lower levels of H3K4me3 in their promoters compared to the normal. In addition, *SLAMF6* was down-regulated in P2, P4 and P5, showing lower H3K4me3 in promoter in P4 and P5; *SLAMF1* was down-regulated in P1, P3 and P4, with lower H3K4me3 in promoter in the first two. Clearly, for genes involved in the regulation of T-cell response, alterations of chromatin state in the promoters are linked to their dysregulation.

In B-cell chronic lymphocytic leukemia, *ATM* was found to physically interact with *TCL1A* protein, and this association activates the NF- κ B pathway in the regulation of apoptosis^{6,55}. *ATM* aberration is viewed as a key initiating lesion in T-PLL⁵. *ATM* showed lower H3K4me3 occupancy in the promoter in all T-PLL cases, with down-regulation of expression in four of them (except in P1 and P5) (Supplemental Fig. 10). Based on the input signal within *ATM*, there was a copy-number loss in five T-PLL cases (P2-P6), which may have led to the observed changes in epigenetic state and expression. The result suggested a link between the reduction of H3K4me3 in promoter and *ATM* down-regulation. Collectively, our analysis demonstrated a central role for epigenetic mechanisms in the dysregulation of key oncogenes, and genes with roles in DNA damage repair and T-cell receptor regulation in T-PLL.

Discussion

T-PLL was best characterized by oncogenic expression of *TCL1A* and dysfunction of *ATM*⁵. Our analysis validated the previous genomic findings of dysregulated genes and further suggested key roles for epigenetic mechanisms in transcriptional dysregulation. We identified genome-wide alterations of chromatin states at both promoters and active enhancers in T-PLL, which likely involves immune related TFs that are highly enriched in differential peaks. The results supported the epigenetic regulation of oncogene activation and repression of genes involved in DNA damage and T-cell responses. Importantly, we provided evidence for the gained enhancers in T-PLL that contribute to the overexpression of *TCL1A* and *MYC*. We believe the same mechanism underlies their activation in several other hematological malignancies. For example, a super-enhancer within *TCL1A* was

identified in high-grade B-cell lymphoma, mantle cell lymphoma and small lymphocytic lymphoma^{33,34}. Accordingly, *TCLIA* showed overexpression in these B-cell lymphomas^{56,57}. In B-cell chronic lymphocytic leukemia, *TCLIA* up-regulation was also associated with the gained super-enhancer⁵⁸ and with DNA hypomethylation in its promoter⁵⁵. Additionally, in many other cancers including T cell leukemia, chronic myeloid leukemia and multiple myeloma, *MYC* expression was linked to the super-enhancers acquired in tumor cells³⁶. Thus, this study has enhanced our understanding of epigenetic regulation of gene expression in T-PLL.

Our study is limited by a small sample size because T-PLL is a rare mature T cell leukemia. Nevertheless, we revealed an obvious heterogeneity in gene regulatory landscape within T-PLL, with P2, P4 and P5 in one group and the other three in the second group. We also identified marked differences in gene expression across T-PLL cases. Further study using a larger collection of samples is needed to identify the epigenetic and transcriptional programs associated with disease progression.

Our analysis supported roles for epigenetic mechanisms in gene dysregulation in T-PLL. Genetic alterations were found to contribute to transcriptional abnormality in T-PLL⁵. Based on input signals, two to four of the T-PLL cases had copy-number gains in *AGO2*, *MYC* and *TCLIA* that were accompanied by increased H3K27ac and H3K4me3, while copy-number loss was observed in *ATM* in five T-PLL cases (P2–P6) that was accompanied by decreased H3K4me3. Complementarily, CNA analysis with 100-kb windows identified CN gains covering *AGO2* and *MYC*, and also *TCLIA* at a less stringent cutoff of $z = 2$. For these four genes, the CNVs inferred from input signal spanned both gene-body and promoter. In fact, copy-number gain in *AGO2* and *MYC* and copy-number loss in *ATM* are the major genomic aberrations in T-PLL⁵. Genomic rearrangements in regulatory regions represent an important mechanism for oncogenic activation. For example, in acute myeloid leukemia⁵⁹ and acute lymphoblastic leukemia⁶⁰, the distal super-enhancers, 1.7 and 1.3 Mb away from the *MYC* promoter, respectively, were located in regions of focal amplification. Integrative analysis of genetic variation, epigenetic profiles with gene expression will help decipher the roles of interplay between genetic and epigenetic factors in shaping the transcriptional program in T-PLL.

In summary, the analyses have identified the transcriptional changes for genes enriched in immune and inflammatory responses, as well as for those in WNT signaling network. Using H3K4me3 and H3K27ac ChIP-seq data, we revealed genome-wide alterations of chromatin states at both promoters and active enhancers in T-PLL. Integrative analysis with RNA-seq data suggests the changes of regulatory regions as a major mechanism underlying transcriptional reprogramming. Importantly, epigenetic changes in regulatory regions are evident at key oncogenes and genes involved in DNA damage response and T-cell receptor regulation, which are tightly linked to their transcriptional dysregulation, the core lesions of T-PLL. Collectively, this study enhances our understanding of the roles of epigenetic mechanisms in T-PLL pathogenesis.

Data availability

The processed RNA-seq and ChIP-seq data generated in this study are available in GEO under the accession GSE143374.

Received: 23 September 2020; Accepted: 5 April 2021

Published online: 15 April 2021

References

- Hasanali, Z. S. *et al.* Epigenetic therapy overcomes treatment resistance in T cell prolymphocytic leukemia. *Sci. Transl. Med.* **7**, 293ra102. <https://doi.org/10.1126/scitranslmed.aaa5079> (2015).
- Stengel, A. *et al.* Genetic characterization of T-PLL reveals two major biologic subgroups and JAK3 mutations as prognostic marker. *Genes Chromosomes Cancer* **55**, 82–94. <https://doi.org/10.1002/gcc.22313> (2016).
- Lopez, C. *et al.* Genes encoding members of the JAK-STAT pathway or epigenetic regulators are recurrently mutated in T-cell prolymphocytic leukaemia. *Br. J. Haematol.* **173**, 265–273. <https://doi.org/10.1111/bjh.13952> (2016).
- Kiel, M. J. *et al.* Integrated genomic sequencing reveals mutational landscape of T-cell prolymphocytic leukemia. *Blood* **124**, 1460–1472. <https://doi.org/10.1182/blood-2014-03-559542> (2014).
- Schrader, A. *et al.* Actionable perturbations of damage responses by *TCL1/ATM* and epigenetic lesions form the basis of T-PLL. *Nat. Commun.* **9**, 697. <https://doi.org/10.1038/s41467-017-02688-6> (2018).
- Gaudio, E. *et al.* *Tcl1* interacts with *Atm* and enhances NF-kappaB activation in hematologic malignancies. *Blood* **119**, 180–187. <https://doi.org/10.1182/blood-2011-08-374561> (2012).
- Zhong, J. *et al.* Purification of nanogram-range immunoprecipitated DNA in ChIP-seq application. *BMC Genom.* **18**, 985. <https://doi.org/10.1186/s12864-017-4371-5> (2017).
- Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21. <https://doi.org/10.1093/bioinformatics/bts635> (2013).
- Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930. <https://doi.org/10.1093/bioinformatics/btt656> (2014).
- Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140. <https://doi.org/10.1093/bioinformatics/btp616> (2010).
- Yu, G. & He, Q. Y. ReactomePA: an R/Bioconductor package for reactome pathway analysis and visualization. *Mol. Biosyst.* **12**, 477–479. <https://doi.org/10.1039/c5mb00663e> (2016).
- Croft, D. *et al.* Reactome: a database of reactions, pathways and biological processes. *Nucl. Acids Res.* **39**, D691–697. <https://doi.org/10.1093/nar/gkq1018> (2011).
- Falcon, S. & Gentleman, R. Using GOSTats to test gene lists for GO term association. *Bioinformatics* **23**, 257–258 (2007).
- Yan, H. *et al.* HiChIP: a high-throughput pipeline for integrative analysis of ChIP-Seq data. *BMC Bioinform.* **15**, 280. <https://doi.org/10.1186/1471-2105-15-280> (2014).
- Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760. <https://doi.org/10.1093/bioinformatics/btp324> (2009).
- Wu, T. D. & Nacu, S. Fast and SNP-tolerant detection of complex variants and splicing in short reads. *Bioinformatics* **26**, 873–881. <https://doi.org/10.1093/bioinformatics/btq057> (2010).
- Robinson, J. T. *et al.* Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26. <https://doi.org/10.1038/nbt.1754> (2011).

18. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842. <https://doi.org/10.1093/bioinformatics/btq033> (2010).
19. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137. <https://doi.org/10.1186/gb-2008-9-9-r137> (2008).
20. Farh, K. K. *et al.* Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature* **518**, 337–343. <https://doi.org/10.1038/nature13835> (2015).
21. Ooi, W. F. *et al.* Epigenomic profiling of primary gastric adenocarcinoma reveals super-enhancer heterogeneity. *Nat. Commun.* **7**, 12983. <https://doi.org/10.1038/ncomms12983> (2016).
22. Yao, L., Tak, Y. G., Berman, B. P. & Farnham, P. J. Functional annotation of colon cancer risk SNPs. *Nat. Commun.* **5**, 5114. <https://doi.org/10.1038/ncomms6114> (2014).
23. Shen, L., Shao, N., Liu, X. & Nestler, E. ngs.plot: Quick mining and visualization of next-generation sequencing data by integrating genomic databases. *BMC Genom.* **15**, 284. <https://doi.org/10.1186/1471-2164-15-284> (2014).
24. Ross-Innes, C. S. *et al.* Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature* **481**, 389–393. <https://doi.org/10.1038/nature10730> (2012).
25. Qu, K. *et al.* Chromatin accessibility landscape of cutaneous T cell lymphoma and dynamic response to HDAC inhibitors. *Cancer Cell* **32**, 27–41.e24. <https://doi.org/10.1016/j.ccell.2017.05.008> (2017).
26. Corces, M. R. *et al.* Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. *Nat. Genet.* **48**, 1193–1203. <https://doi.org/10.1038/ng.3646> (2016).
27. Calderon, D. *et al.* Landscape of stimulation-responsive chromatin across diverse human immune cells. *Nat. Genet.* **51**, 1494–1505. <https://doi.org/10.1038/s41588-019-0505-9> (2019).
28. Kundaje, A. *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330. <https://doi.org/10.1038/nature14248> (2015).
29. Andersson, R. *et al.* An atlas of active enhancers across human cell types and tissues. *Nature* **507**, 455–461. <https://doi.org/10.1038/nature12787> (2014).
30. Zhang, J. *et al.* The CREBBP acetyltransferase is a haploinsufficient tumor suppressor in B-cell lymphoma. *Cancer Discov.* **7**, 322–337. <https://doi.org/10.1158/2159-8290.cd-16-1417> (2017).
31. Bernstein, B. E. *et al.* The NIH roadmap epigenomics mapping consortium. *Nat. Biotechnol.* **28**, 1045–1048. <https://doi.org/10.1038/nbt1010-1045> (2010).
32. Bernstein, B. E. *et al.* An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74. <https://doi.org/10.1038/nature11247> (2012).
33. Ryan, R. J. *et al.* Detection of enhancer-associated rearrangements reveals mechanisms of oncogene dysregulation in B-cell lymphoma. *Cancer Discov.* **5**, 1058–1071. <https://doi.org/10.1158/2159-8290.cd-15-0370> (2015).
34. Ryan, R. J. H. *et al.* A B cell regulome links notch to downstream oncogenic pathways in small B cell lymphomas. *Cell Rep.* **21**, 784–797. <https://doi.org/10.1016/j.celrep.2017.09.066> (2017).
35. Wong, R. W. J. *et al.* Enhancer profiling identifies critical cancer genes and characterizes cell identity in adult T-cell leukemia. *Blood* **130**, 2326–2338. <https://doi.org/10.1182/blood-2017-06-792184> (2017).
36. Hnisz, D. *et al.* Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934–947. <https://doi.org/10.1016/j.cell.2013.09.053> (2013).
37. Javierre, B. M. *et al.* Lineage-specific genome architecture links enhancers and non-coding disease variants to target gene promoters. *Cell* **167**, 1369–1384. <https://doi.org/10.1016/j.cell.2016.09.037> (2016).
38. Sonesson, C. & Delorenzi, M. A comparison of methods for differential expression analysis of RNA-seq data. *BMC Bioinform.* **14**, 91. <https://doi.org/10.1186/1471-2105-14-91> (2013).
39. Chen, Y., Lun, A. T. & Smyth, G. K. From reads to genes to pathways: differential expression analysis of RNA-Seq experiments using Rsubread and the edgeR quasi-likelihood pipeline. *F1000Res* **5**, 1438. <https://doi.org/10.12688/f1000research.8987.2> (2016).
40. Lu, D. *et al.* Activation of the Wnt signaling pathway in chronic lymphocytic leukemia. *Proc. Natl. Acad. Sci. U S A* **101**, 3118–3123. <https://doi.org/10.1073/pnas.0308648100> (2004).
41. Janovska, P. & Bryja, V. Wnt signalling pathways in chronic lymphocytic leukaemia and B-cell lymphomas. *Br. J. Pharmacol.* **174**, 4701–4715. <https://doi.org/10.1111/bph.13949> (2017).
42. Rosati, E. *et al.* NOTCH1 aberrations in chronic lymphocytic leukemia. *Front. Oncol.* **8**, 229. <https://doi.org/10.3389/fonc.2018.00229> (2018).
43. Gianfelici, V. Activation of the NOTCH1 pathway in chronic lymphocytic leukemia. *Haematologica* **97**, 328–330. <https://doi.org/10.3324/haematol.2012.061721> (2012).
44. Liu, Y. *et al.* The genomic landscape of pediatric and young adult T-lineage acute lymphoblastic leukemia. *Nat. Genet.* **49**, 1211–1218. <https://doi.org/10.1038/ng.3909> (2017).
45. Suraweera, A., O’Byrne, K. J. & Richard, D. J. Combination therapy with histone deacetylase inhibitors (HDACi) for the treatment of cancer: achieving the full therapeutic potential of HDACi. *Front Oncol* **8**, 92. <https://doi.org/10.3389/fonc.2018.00092> (2018).
46. Zhao, M., Kim, P., Mitra, R., Zhao, J. & Zhao, Z. TSGene 2.0: an updated literature-based knowledgebase for tumor suppressor genes. *Nucl. Acids Res.* **44**, D1023–1031. <https://doi.org/10.1093/nar/gkv1268> (2016).
47. Lichtinger, M. *et al.* RUNX1 reshapes the epigenetic landscape at the onset of haematopoiesis. *EMBO J.* **31**, 4318–4333. <https://doi.org/10.1038/emboj.2012.275> (2012).
48. Koizumi, S. I. *et al.* JunB regulates homeostasis and suppressive functions of effector regulatory T cells. *Nat. Commun.* **9**, 5344. <https://doi.org/10.1038/s41467-018-07735-4> (2018).
49. Li, P. *et al.* BATF-JUN is critical for IRF4-mediated transcription in T cells. *Nature* **490**, 543–546. <https://doi.org/10.1038/nature11530> (2012).
50. Staudt, L. M. Oncogenic activation of NF-kappaB. *Cold Spring Harb. Perspect. Biol.* **2**, a000109. <https://doi.org/10.1101/cshperspect.a000109> (2010).
51. Herling, M. *et al.* High TCL1 levels are a marker of B-cell receptor pathway responsiveness and adverse outcome in chronic lymphocytic leukemia. *Blood* **114**, 4675–4686. <https://doi.org/10.1182/blood-2009-03-208256> (2009).
52. Yashiro-Ohtani, Y. *et al.* Long-range enhancer activity determines Myc sensitivity to Notch inhibitors in T cell leukemia. *Proc. Natl. Acad. Sci. U S A* **111**, E4946–4953. <https://doi.org/10.1073/pnas.1407079111> (2014).
53. Souroullas, G. P. *et al.* An oncogenic Ezh2 mutation induces tumors through global redistribution of histone 3 lysine 27 trimethylation. *Nat. Med.* **22**, 632–640. <https://doi.org/10.1038/nm.4092> (2016).
54. Ye, Z., Jin, H. & Qian, Q. Argonaute 2: a novel rising star in cancer research. *J. Cancer* **6**, 877–882. <https://doi.org/10.7150/jca.11735> (2015).
55. Garding, A. *et al.* TCL1A and ATM are co-expressed in chronic lymphocytic leukemia cells without deletion of 11q. *Haematologica* **98**, 269–273. <https://doi.org/10.3324/haematol.2012.070623> (2013).
56. Weng, J. *et al.* TCL1: a shared tumor-associated antigen for immunotherapy against B-cell lymphomas. *Blood* **120**, 1613–1623. <https://doi.org/10.1182/blood-2011-09-382838> (2012).
57. Paduano, F. *et al.* T-Cell Leukemia/Lymphoma 1 (TCL1): an Oncogene Regulating Multiple Signaling Pathways. *Front Oncol.* **8**, 317. <https://doi.org/10.3389/fonc.2018.00317> (2018).

58. Fabbri, G. *et al.* Common nonmutational NOTCH1 activation in chronic lymphocytic leukemia. *Proc. Natl. Acad. Sci. U S A* **114**, E2911–e2919. <https://doi.org/10.1073/pnas.1702564114> (2017).
59. Shi, J. *et al.* Role of SWI/SNF in acute leukemia maintenance and enhancer-mediated Myc regulation. *Genes Dev.* **27**, 2648–2662. <https://doi.org/10.1101/gad.232710.113> (2013).
60. Herranz, D. *et al.* A NOTCH1-driven MYC enhancer promotes T cell development, transformation and acute lymphoblastic leukemia. *Nat. Med.* **20**, 1130–1137. <https://doi.org/10.1038/nm.3665> (2014).

Acknowledgements

This study was supported by the National Institute of Health Lymphoma SPORE CA 97274 and K23 CA160345, Richard M. Schulze Family Foundation for Awards in Cancer Research, Fraternal Order of Eagles Cancer Research Fund, and the Predolin Foundation.

Author contributions

S.T., H.Z., H.Y., and W.D. designed the study. S.T., P.Z., H.Y., and W.D. analyzed the data. H.Z. collected samples. M.K. provided IT support. J.H.L. prepared ChIP samples. T.O., P.H., T.W., E.W.K., and S.L.S. edited the paper. T.C., T.W., N.K., and S.L.S. provided samples. S.T., H.Y., and W.D. wrote the manuscript. All authors reviewed and approved the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-87890-9>.

Correspondence and requests for materials should be addressed to S.T. or W.D.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021