



Published in final edited form as:

Cell Rep. 2023 April 25; 42(4): 112373. doi:10.1016/j.celrep.2023.112373.

Elevated enhancer-oncogene contacts and higher oncogene expression levels by recurrent *CTCF* inactivating mutations in acute T cell leukemia

Willem K. Smits^{1,10}, Carlo Vermeulen^{2,3,4,10}, Rico Hagelaar^{1,2,10}, Shunsuke Kimura⁵, Eric M. Vroegindewij¹, Jessica G.C.A.M. Buijs-Gladdines¹, Ellen van de Geer¹, Marjon J.A.M. Verstegen^{2,3}, Erik Splinter⁶, Simon V. van Reijmersdal¹, Arjan Buijs⁷, Niels Galjart⁸, Winfried van Eyndhoven⁹, Max van Min⁶, Roland Kuiper^{1,7}, Patrick Kemmeren¹, Charles G. Mullighan⁵, Wouter de Laat^{2,3}, Jules P.P. Meijerink^{1,11,*}

¹Princess Máxima Center for Pediatric Oncology, Utrecht, the Netherlands

²Oncode Institute, Utrecht, the Netherlands

³Hubrecht Institute-KNAW, Utrecht, the Netherlands

⁴Center for Molecular Medicine, University Medical Center Utrecht, Utrecht, the Netherlands

⁵Laboratory of Pathology, St. Jude's Children's Research Hospital, Memphis TN, USA

⁶Cergentis BV, Utrecht, the Netherlands

⁷Department of Genetics, University Medical Center Utrecht, Utrecht, the Netherlands

⁸Department of Cell Biology, Erasmus University Medical Center Rotterdam, Rotterdam, the Netherlands

⁹Diagnostics and Genomics Group, Agilent Technologies, Amstelveen, the Netherlands

¹⁰These authors contributed equally

¹¹Lead contact

SUMMARY

Monoallelic inactivation of CCCTC-binding factor (CTCF) in human cancer drives altered methylated genomic states, altered CTCF occupancy at promoter and enhancer regions, and deregulated global gene expression. In patients with T cell acute lymphoblastic leukemia (T-ALL), we find that acquired monoallelic CTCF-inactivating events drive subtle and local genomic effects

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

*Correspondence: jules.meijerink@acerta-pharma.com.

AUTHOR CONTRIBUTIONS

W.K.S., C.V., R.H., S.K., E.M.V., J.G.C.A.M.B.-G., E.v.d.G., M.J.A.M.V., E.S., S.V.v.R., A.B., N.G., W.v.E., and M.v.M. designed experiments, performed research, and analyzed data. R.H., P.K., C.G.M., and W.d.L. supervised experiments and wrote the manuscript. J.P.P.M. designed and supervised the study, designed experiments, and wrote manuscript.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.celrep.2023.112373>.

DECLARATION OF INTERESTS

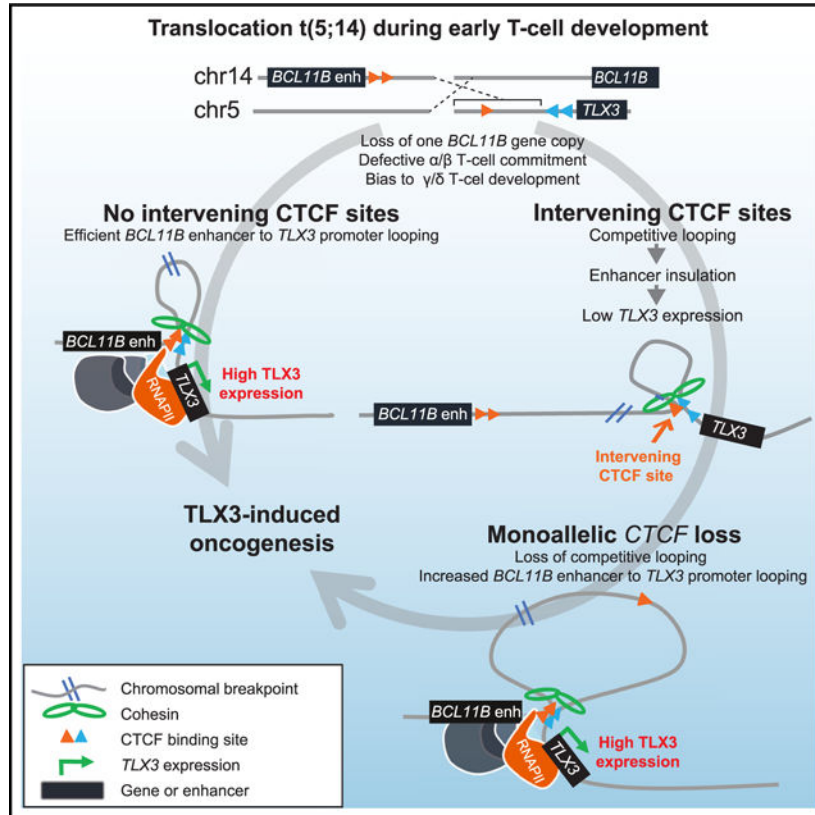
The authors declare no competing interests.

in nearly half of t(5; 14) (q35; q32.2) rearranged patients, especially when CTCF-binding sites are preserved in between the *BCL11B* enhancer and the *TLX3* oncogene. These solitary intervening sites insulate *TLX3* from the enhancer by inducing competitive looping to multiple binding sites near the *TLX3* promoter. Reduced CTCF levels or deletion of the intervening CTCF site abrogates enhancer insulation by weakening competitive looping while favoring *TLX3* promoter to *BCL11B* enhancer looping, which elevates oncogene expression levels and leukemia burden.

In brief

Smits et al. report that *TLX3* chromosomal translocations in T cell acute lymphoblastic leukemia patients recurrently include intervening CTCF-binding sites in the breakpoint area. Pressure to acquire inactivating aberrations in *CTCF* abrogate consequential enhancer insulation that promotes *TLX3* promoter to *BCL11B* enhancer looping boosts higher oncogene expression levels and leukemia burden.

Graphical Abstract



INTRODUCTION

The CCCTC-binding factor (CTCF) is a highly conserved and constitutively expressed transcription factor containing 11 zinc fingers (ZFs) that binds into the major groove of DNA. Over 40,000 target sites in the human genome have been identified across tissues that depend on the variable use of ZFs.¹⁻⁵ Unlike other transcription factors, CTCF binds

more distantly from transcriptional start sites (TSSs).⁶ Various other roles in addition to functioning as a transcriptional repressor or activator have been assigned to CTCF including regulation of antigen receptor or T cell receptor recombination events and serving as a regulator for genomic imprinting. CTCF promotes the formation of DNA loops,^{4,7–20} as first shown at the β -globin and the *Igf2/H19* locus,^{21–25} and is recognized as a most dominant chromatin looping factor in mammalian cells. It often binds to and forms loops between the boundaries of self-interacting chromatin domains, also called topologically associating domains (TADs), structural entities of up to a megabase in size that are appreciable from chromosome conformation capture studies.^{26–29} Chromatin topology is considered a key factor in gene regulation, with CTCF-mediated loops facilitating promoter-enhancer contacts and preventing unwanted contacts.^{30,31} CTCF-mediated looping depends on the cohesin complex,^{32–38} and complete removal of CTCF or cohesin results in loss of TADs and chromatin looping.^{31,39–41} The ring-shaped cohesin complex is composed of structural maintenance of chromosomes 1A and 3 (SMC1A and SMC3, respectively) and RAD21 proteins that are stabilized by STAG2. Cohesin complexes are loaded onto the chromatin fiber and are believed to actively extrude chromatin loops until stalled at CTCF-bound sites.^{42–45} Therefore, cohesin is found to be enriched at CTCF-bound sites in chromatin immunoprecipitation (ChIP) sequencing experiments.^{32,46–50} CTCF-binding sites are directional, and loops are predominantly stabilized among convergent, inward-oriented binding sites.^{43,51} Stable loops are released by the cohesin disassembly factor WAPL.⁵²

Genetic mutations in cohesin components or the cohesin chromatin loader protein nipped-B-like (NIPBL) protein are recurrently found in patients with Cornelia de Lange syndrome. Somatic mutations in *STAG2* are recurrently found in patients with bladder cancer, glioblastoma, melanoma, or acute myeloid leukemia (AML)^{49,53} and, in patients with Ewing's sarcoma, affects the EWS/FLI1 oncogenic program that enhances migration and invasion properties.^{54,55} In the hematopoietic system, inactivation of cohesin components results in the expansion of hematopoietic stem progenitor cells (HSPCs), increasing genomic accessibility near ERG, RUNX1, and GATA2 binding sites and elevating replating efficiencies.^{56,57} Heterozygous deletions of SMC3 increase cellular transformation by FLT3-ITD, which drives myeloproliferative diseases in mice.⁵⁸

CTCF is located on chromosomal band 16q22.1, and its inactivation has also been associated with cancer. Heterozygous loss of *Ctcf* predisposes mice to develop spontaneous or chemical-induced invasive tumors with high proliferation potential.⁵⁹ Monoallelic deletions of CTCF or inactivating missense mutations in specific ZFs have been identified in many human cancers including sporadic breast cancer, prostate cancer, Wilms tumors, and acute lymphoblastic leukemia (ALL).^{60–64} Heterozygous deletions or point mutations have been identified in 57% and 2.7% of patients with breast cancer and in 24% and 21% of patients with uterine endometrial cancers, respectively.^{59,65} In these tumors, CTCF depletion has been associated with large global changes in DNA methylation of divergent CpG islands and poor survival,^{59,66} although the precise role of CTCF aberrations in pathogenesis remains poorly understood.⁶² In this study, we investigated the functional significance and molecular-cytogenetic associations of *CTCF* aberrations in patients with T cell ALL.

RESULTS

***CTCF* aberrations are abundant in patients with *TLX3*-rearranged T-ALL**

DNA copy-number analysis on diagnostic biopsies of 181 pediatric patients with T cell ALL (T-ALL) using array-comparative genomic hybridization and/or multiplex ligation-dependent probe amplification (MLPA) revealed recurrent monoallelic deletions affecting the chromosomal band 16q in sixteen patients (9%; Table S1). The minimal deleted area (MDA) affected exons 3 to 6 of the *CTCF* gene (Figures 1A and 1B). We then screened 146 out of these 181 patients for *CTCF* mutations using single-molecule molecular inversion probe (smMIP) sequencing⁶⁷ (Table S2) and identified an additional 11 patients (8%) harboring mutations with variant allele frequencies of 34% or higher that were considered heterozygous mutations (Table S1), including missense mutations (2 patients), a nonsense mutation (1 patient), and insertion mutations that lead to frameshifts (5 patients). None of the patients had biallelic inactivation of *CTCF*. Furthermore, heterozygous mutations at the -2 or the -12 positions near the splice sites of exons 5 and 8 were identified in three patients (Figure 1B) and resulted in alternative out-of-frame *CTCF* splicing (Figures 1C and 1D). Therefore, *CTCF* aberrations were detected in at least 15% of all patients with T-ALL.

T-ALL is characterized by specific driving oncogene rearrangements that delineate 4 to 5 subtypes that harbor unique expression signatures.^{63,69–71} *CTCF* aberrations were especially frequent in patients with T-ALL with *TLX3*-rearrangements (53%) compared with patients with other driving oncogenic rearrangements (6%, $p = 2.23 \times 10^{-16}$; Figure 1E). In relation to T-ALL subtypes as distinguished by unsupervised cluster analysis of gene expression data ($n = 117$),⁷¹ *CTCF* aberrations were particularly identified in patients belonging to the TLX subtype (9 out of 30), which is highly enriched for *TLX3*-rearranged patients. Only nine out of 87 patients belonging to other subtypes harbor *CTCF* aberrations ($p = 0.01$; Figure S1). Patients with *CTCF* deletions expressed roughly half of *CTCF* levels compared with *CTCF* wild-type patients (Figure 1F; $p < 0.0135$). No significant differences were identified between wild-type and *CTCF* mutant patients. *CTCF* expression levels for 5 patient-derived xenograft (PDX) models derived from 1 *CTCF*-mutated, 1 *CTCF*-deleted, and 3 wild-type patients were conserved between each primary patient sample and its corresponding PDX model (Figure S1).

***CTCF* aberrations in T-ALL do not affect global DNA methylation, gene expression, *CTCF* chromatin binding, or TAD formation patterns**

As monoallelic *CTCF* inactivation in cancer has been associated with strong global changes in gene expression levels and DNA methylation patterns,⁵⁹ we studied whether *CTCF* aberrations would drive similar global effects in patients with *TLX3*-rearranged T-ALL. For this, we compared overall gene expression levels of nine *CTCF*-aberrant patients with 11 *CTCF* wild-type patients. Remarkably, we did not identify a single differentially expressed gene among these patients (Figure S2A). To further identify global effects on DNA methylation, we compared methylation patterns for 4 *CTCF*-deleted versus 3 *CTCF* wild-type patient samples (Table S1) and identified only a single significant and differentially methylated CpG-island probeset out of a total of 853,307 probesets (Figure S2B). So, within the detection limits of our methods, monoallelic loss of *CTCF* in patients

with *TLX3*-rearranged T-ALL therefore does not seem to cause consistent global changes in DNA methylation or gene expression levels.

We then investigated differential CTCF recruitment to chromatin binding sites using ChIP sequencing (ChIP-seq) in these same 6 *TLX3*-rearranged patients with T-ALL with (n = 4) or without (n = 2) *CTCF* aberrations. Two patients with T-ALL who harbored *TLX1* or *NKX2-1* oncogenic rearrangements were included as controls (Table S1). We found only 41 differentially called CTCF-binding peaks out of a total of 31,840 called peaks (Figures 2A and S2; Table S3). Interestingly, 5 out of these 41 binding peaks that were conserved in *CTCF*-deleted patients correspond to the *TCRAD* locus and were lost in *CTCF* wild-type patients with T-ALL due to *TCRD* recombination (Figure 2B). This result is in line with the involvement of CTCF in looping distant *TCRAD* V-gene segments to the recombination center during T cell receptor (TCR) assembly^{35,72-74} and demonstrates that *CTCF*-inactivated patients are compromised in establishing functional rearrangements of the *TCRA* locus. We did not find any further differences in CTCF binding at upstream TSSs (uTSSs), long non-coding RNAs (lncRNAs), or enhancer regions (Figure S2). To further investigate potential global effects of reduced *CTCF* levels in these 6 patients with *TLX3*-rearranged T-ALL (Table S1), we performed HiC to visualize TAD boundaries (Figure 2C). In line with data from *Ctcf* loss-of function mouse models that preserve TAD structures at minimal CTCF levels as low as 15% of normal levels,³¹ we did not find obvious differences in TAD boundaries among *CTCF*-inactivated versus wild-type patients (Figure 2C). The calculated insulation scores for each 50 Kb bin from one CTCF wild-type patient compared with one CTCF-deleted patient using HiCEXplorer⁷⁵ highly correlated with an overall Spearman correlation of 0.92 over all bins (Figure 2D). Therefore, our data suggest that heterozygous loss of CTCF in *TLX3*-rearranged T-ALL does not lead to significant global changes in gene expression, DNA methylation, CTCF chromatin binding, or TAD formation.

Heterozygous loss of CTCF promotes $\gamma\delta$ T cell development

CTCF aberrations are especially found in *TLX3*-rearranged patients, a disease entity that is associated with $\gamma\delta$ T lymphoid lineage development.^{76,77} Most *TLX3*-rearranged patients harbor t(5; 14) (q35; q32) translocations that position the *TLX3* oncogene in close proximity to the *BCL11B* enhancer, which inactivates one functional *BCL11B* allele.⁷⁸ The *BCL11B* transcription factor is a critical regulator for T cell $\alpha\beta$ -lineage commitment, and *Bcl11b* knockout mice demonstrate reduced numbers of $\alpha\beta$ T cells but increased numbers of TCR $\gamma\delta$ + thymocytes.^{79,80} The *BCL11B* enhancer is located approximately 800 Kb downstream of the *BCL11B* gene, and CTCF is essential for *Bcl11b* expression during normal T cell development in mice by facilitating *BCL11B* enhancer to promoter loops.⁸¹ In line with these results, we found that *Ctcf* conditional knockout mice crossed on the *Lck-Cre* transgenic and T lineage-specific background strongly reduced CTCF levels during early T cell development. While the percentage of TCR $\alpha\beta$ + thymocytes seemed unaffected in *Ctcf* heterozygous knockout mice, these dropped to 50% in homozygous knockout mice (Figures 3A and 3C). Despite an overall drop in total thymocyte numbers (Figure 3B), total TCR $\gamma\delta$ + thymocyte numbers in heterozygous and homozygous *Ctcf* knockout mice strongly increased compared with control littermates (Figures 3A and 3D). Furthermore, the

proportion of early thymocyte subsets including double-negative (DN) thymocytes (DN1–3), CD8 intersingle positive thymocytes increased in heterozygous versus homozygous *Ctcf* knockout mice, respectively, which is in line with a partial early T cell development block (Figure S3).

CTCF inactivation coincides with chromosomal breakpoint locations and increased *TLX3* levels

Given the importance of CTCF as a *BCL11B* regulator, we questioned whether *CTCF* inactivation would negatively impact *BCL11B* and *TLX3* oncogene expression levels that are driven from the *BCL11B* enhancer on the wild-type or translocated alleles in *TLX3*-rearranged patients, respectively. We determined the exact molecular DNA breakpoints for 23 patients with *TLX3-BCL11B*-translocated T-ALL and for 2 *TLX3*-rearranged T-ALL cell lines (HPB-ALL and DND41) using targeted locus amplification (TLA).⁸² For this, TLA was performed on diagnostic patient samples using *TLX3* and *BCL11B* promoter sequences as viewpoints (Table S4). These data were further complemented by whole-genome sequencing breakpoint data from 8 additional patients with *TLX3*-rearranged T-ALL.⁶³ For 27 out of 31 patients and both cell lines, *TLX3* was translocated to the *BCL11B* locus at 14q32 with breakpoints located in or just distal of the *RANBP17* locus that is centromeric to *TLX3* at 5q35 and distal of the *BCL11B* enhancer at 14q32. Most of these breakpoints preserved the DNase hypersensitivity sites HS3 and HS4 and “major peak” in the enhancer region that were identified as pivotal enhancer elements by others before^{83,84} (Figures 4A and 4B). Remarkably, breakpoints in patients P3, P18, and PSJ5 preserved HS3 but not HS4 and major peak, indicating that HS4 and major peak are not essential for enhancer activity (Figure 4C). Three other patients had breakpoints telomeric to *TLX3*, resulting in the insertion of the complete *BCL11B* enhancer downstream of *TLX3* in patient P14 (Figure 4C). One patient (P29) had evidence for a complex translocation to regulatory regions of the CAPSL locus at 5p13.2 (Table S4).

Using qRT-PCR, we found an overall strong correlation between *TLX3* and *BCL11B* expression levels (Figure S4; $p = 0.0002$), suggesting that the *BCL11B* enhancer drives *TLX3* or *BCL11B* by a similar mechanism from the translocated or the wild-type allele, respectively. We noticed that multiple CTCF-binding sites are present in the *BCL11B* enhancer region that are complementary (convergent) to those found in the transcriptional regulatory regions of *BCL11B* and *TLX3* (Figure 4A). Seventeen out of 31 patients and both cell lines contained *CTCF* deletions or mutations, whereas 14 patients were *CTCF* wild type. Calculating exact genomic distances between the *BCL11B* enhancer and the *TLX3* TSS revealed that in *CTCF*-aberrant patients, these genomic distances were generally larger than in *CTCF* wild-type patients (Figure 4D; $p = 0.064$). The more distally located breakpoints from *TLX3*, found in *CTCF*-inactivated patients, kept CTCF-binding motifs in the neighboring *RANBP17* locus in *cis* and therefore in between the *BCL11B* enhancer and the *TLX3* oncogene (Figure 4A). We hypothesized that these intervening CTCF-binding sites insulate the *TLX3* promoter from the *BCL11B* enhancer, possibly through the formation of alternative or competitive DNA loops, and that *CTCF* inactivation is required to alleviate insulation and enable oncogenic *TLX3* expression levels. Remarkably, we indeed found that *CTCF*-inactivated patients expressed lower *BCL11B* levels than *CTCF* wild-type

patients but expressed higher *TLX3* levels that point to efficient *TLX3* promoter interactions with the distal *BCL11B* enhancer in *CTCF*-inactivated patients (Figure S4). Calculating the *TLX3* to *BCL11B* expression ratios revealed significantly higher ratios for *CTCF*-aberrant patients than for wild-type patients (Figures 4E [$p = 0.0073$] and S4). This enhanced oncogene activity may explain the higher white blood cells counts at disease diagnosis as noted for *CTCF*-inactivated patients (Figure 4F; $p = 0.017$).

***BCL11B* enhancer to *TLX3* loops by chromatin conformation capture sequencing**

We then studied promoter interactions of *TLX3* and *BCL11B* with the *CTCF* sites within the distal *BCL11B* enhancer on the der(14) and the normal chromosome 14, respectively. For this, we applied chromatin conformation capture sequencing (4C-seq) analysis to viable diagnosis leukemia cells from 9 pediatric patients with (*TLX3*-rearranged) T-ALL, of whom 7 had *CTCF* aberrations, using viewpoints (VPs) located closely to the *TLX3* or *BCL11B* promoters.⁸⁵ To facilitate data interpretation, we performed *CTCF* and *SMC3* ChIP-seq on these patient samples and plotted their binding profiles alongside the 4C-seq chromatin contact profiles (Table S1; Figure 5). For the wild-type *BCL11B* allele, evidence for DNA loops between the *BCL11B* promoter and its 900 kb downstream distal enhancer was found in all these patients with T-ALL. We also found an additional longdistance loop centromeric of the *BCL11B* enhancer at base-pair position 98,444,890. All 4C-predicted DNA loop interactions coincide with *CTCF*- and cohesin-bound sites. Remarkably, no *BCL11B* promoter loops were identified to sequences in the 900 kb intervening region in between *BCL11B* and the distal enhancer that lacked *CTCF*- or cohesin-bound sites.

When analyzing the chromatin contacts of the *TLX3* gene promoter, we indeed observed in all patient cells that *TLX3* on the translocated allele specifically contacted the *BCL11B* enhancer across the breakpoint. This specific contact was readily appreciable even if the gene and enhancer were far apart, as seen in *CTCF*-aberrant patients (Figure 5). ChIP-seq on *CTCF*-aberrant patient cells with a distal *BCL11B* enhancer confirmed binding of *CTCF* and cohesin to the remaining intervening *CTCF* motifs, despite the lower *CTCF* levels in these patients. 4C-seq further showed that in these patients, the *TLX3* promoter forms secondary loops with these intervening *CTCF* sites. We therefore hypothesized that these intervening *CTCF* sites compete with the distal *BCL11B* enhancer for looping to the *TLX3* gene.

Restoring *CTCF* levels in HPB-ALL enhance competitive loop formation

To test this hypothesis, we developed a functional model system based on HPB-ALL cells. As displayed in Figure 4A, this $t(5; 14)(q35; q32.2)^+$ cell line has preserved a local intervening *CTCF*-binding site in the translocation breakpoint area in between *TLX3* and the distal *BCL11B* enhancer and has inactivated one *CTCF* allele due to a deletion. To increase *CTCF* levels i contacts with the distal na controlled manner in these HPB-ALL cells, they were lentivirally transduced with a doxycycline-inducible *CTCF*-blue fluorescence protein (*mTagBFP*) expression construct (*iCTCF*). To exclude potential lethal effects by altering *TLX3* levels as demonstrated before for DND41 cells⁸⁶ following *CTCF* induction in HPB-ALL cells, we introduced a constitutive Venus-tagged *TLX3* rescue construct via a secondary round of lentiviral transduction (denoted as HPB-ALL-*iCTCF/TLX3* cells; Figure 6A). In addition to bulk cells, G4 and E3 single-cell clones were produced that

express moderate and high CTCF (BFP) levels upon doxycycline treatment, respectively (Figure 6B). As visualized in Figure 6C, induction (+doxycycline [dox]) of CTCF in HPB-ALL-iCTCF/TLX3 bulk cells reduced the contacts between the *TLX3* promoter and the *BCL11B* enhancer but increased contacts with the intervening CTCF site and other proximal sequences just upstream of *TLX3*. Induction of moderate CTCF levels (clone G4) had a similar effect, while induction of higher CTCF levels (clone E3) even further promoted contacts of *TLX3* with its proximal sequences, at the expense of its contacts with the distal *BCL11B* enhancer. To further investigate the role of the intervening CTCF-binding site in these competitive *TLX3* contacts, we removed this site in E3 cells by CRISPR-Cas9-mediated genome editing (Figure 7A). This eliminated the competitive *TLX3* contacts with their proximal sequences, restored efficient contacts with the distal *BCL11B* enhancer (Figure 7B), and led to higher endogenous *TLX3* expression levels (Figure 7C).

DISCUSSION

Impaired maintenance of TAD structures can form the basis of cellular transformation and cancer and has been shown to activate oncogenes from enhancers that are normally located in separate TADs.^{87–89} Recurrent alterations in CTCF binding due to aberrant hypermethylation states in various cancers result in a general loss of CTCF binding near gene promoters while increasing binding near enhancers, often in concert with oncogenic transcription factors that drive expression of their downstream target genes as shown for NOTCH1 in T-ALL.⁹⁰

In addition to boundary loss due to hypermethylation, recurrent inactivation of CTCF by genetic alterations has been observed in up to 50% of patients with breast cancer, endometrium cancer, Wilms tumors, or colon cancer.^{59,61,62} Monoallelic *Ctcf* loss in knockout mice creates a tumor-prone phenotype. In patients, it drives a hypermethylated phenotype and global changes in gene expression levels that enhance survival of cancer cells. For endometrium cancer, it results in altered cellular polarity and poor outcome.^{59,66,91}

We here demonstrate that CTCF aberrations are identified in nearly 15% of pediatric patients with T-ALL, in line with previous studies.⁶³ *CTCF* aberrations are predominantly associated with T-ALL driven by the *TLX3* oncogene. This disease entity is associated with $\gamma\delta$ -lineage development.^{77,92} To date, no consistent oncogenic mechanism for loss-of-function *CTCF* aberrations in T-ALL has been described. In contrast to solid tumors and within the detection limits of the methods used, we here demonstrate that monoallelic *CTCF* aberrations in T-ALL do not globally change DNA methylation, gene expression levels, or TAD organization. Global CTCF genome binding was almost identical among patients with *TLX3*-rearranged T-ALL with or without *CTCF* aberrations, except for the *TRA* locus, which remains unrearranged in *CTCF*-aberrant patients. CTCF and cohesin have been shown essential for productive TRA-recombination events that facilitate locus contraction by looping distant V-gene segments into the vicinity of other gene segments near the VDJ-recombination center.^{35,72,93} These results are in line with observations in *Ctcf* conditional knockout mice in the T cell lineage that result in reduced numbers of $\gamma\delta$ T cells.⁷ Upon further inspection of these *Ctcf* knockout mice, we here reveal that loss of $\alpha\beta$ T cells is accompanied by increased numbers of $\gamma\delta$ T cells.

We identified an unexpected mechanism where monoallelic *CTCF* loss contributes to increased oncogene expression levels by reducing the insulation strength of intervening CTCF sites in the translocation breakpoint region of *TLX3-BCL11B*-rearranged patients. In nearly 50% of *TLX3*-rearranged patients, breakpoints in the *RANBP17* locus, which flanks the *TLX3* oncogene, result in the preservation of intervening CTCF-binding sites in between the *BCL11B* enhancer and the *TLX3* oncogene. *CTCF* deletions, frameshifts, or splice site mutations are almost exclusively found in those patients that preserve these intervening CTCF-binding sites in the breakpoint area. Only a few patients for whom we failed to identify *CTCF* aberrations had retained intervening CTCF motifs in between *TLX3* and the distal *BCL11B* enhancer. As we did not identify inactivating mutations in the intervening CTCF motifs in these patients that impair CTCF binding, one may speculate that these patients may have alternative but functionally equivalent mutations in cohesin components such as STAG 2.

During the t(5;14) chromosomal rearrangement, the *TLX3* oncogene is aberrantly placed in the vicinity of the *BCL11B* enhancer that activates *TLX3* expression. This expression may be insufficient to drive full oncogene potential as the result of preservation of intervening CTCF binding sites in the breakpoint area that form competitive loops with the *TLX3* promoter. Indeed, restoration of higher CTCF levels in the T-ALL line HPB-ALL, which has inactivated one functional copy of *CTCF*, results in increased competitive loop formation between the intervening CTCF-binding site and *TLX3*, which insulates *TLX3* from the *TLX3* enhancer. Elimination of the intervening CTCF-binding site impairs competitive loop formation even at higher CTCF levels, resulting in enhanced *TLX3* expression levels. Therefore, our work is in line with a model where preservation of CTCF-binding sites in the genomic breakpoint area of patients with *TLX3*-rearranged T-ALL yields only moderate *TLX3* activation in preleukemia cells. We hypothesize that this provides a selective pressure on preleukemia cells to reduce CTCF expression levels by acquiring deletions or mutations that raise productive loop formations between *TLX3* and the *BCL11B* enhancer and thus drive higher oncogene expression levels. As demonstrated by 4C-seq, formation of productive loops between *TLX3* and the *BCL11B* enhancer involves multiple convergent CTCF sites. As shown by others, clustered CTCF sites form chromatin loops more robustly than isolated sites.⁹⁴ Higher *TLX3* expression levels in *CTCF*-aberrant patients will boost oncogenic activity, and these patients present with higher numbers of leukemia cells in peripheral blood and bone marrow biopsies at disease diagnosis. As no intervening CTCF-binding sites are found in the normal *BCL11B* locus, this explains the lower *BCL11B* levels in *CTCF*-aberrant patients.

Therefore, the frequent association of *CTCF* aberrations that we find, especially with patients with t(5; 14) (q35; q32.2)-rearranged T-ALL who maintain *TLX3*-proximal CTCF sites, reflects a necessity to neutralize these sites in order to topologically enable the distal *TLX3* enhancer to interact with the *TLX3* oncogene and to boost its expression.

Limitations of the study

In this study, we have identified and characterized the interplay between genetic *CTCF* aberrations and the regulation of the *TLX3* oncogene that is activated by the t(5; 14)

translocation in T-ALL. While focusing on *CTCF* aberrations, we did not investigate whether functional-equivalent mutations may exist in other components of the chromatin organization machinery such as cohesin or WAPL.

Given the relative large areas of chromosomal breakpoint regions in patients with t(5; 14)-translocated T-ALL and the rarity of patients with T-ALL with this translocation (~20%–25% of pediatric patients with T-ALL), we were limited in the number of available primary patient samples for genetic analysis. The unavailability of fresh patient samples also limited the inclusion of additional patient samples for DNA methylation analysis.

Quantitative measurements of CTCF protein levels in living cells did not yield consistent expression levels over different patient samples that related to the copy-number status of the *CTCF* gene. We therefore quantified *CTCF* mRNA levels in relation to the copy number or mutational status of the *CTCF* alleles. For the HPB-ALL model system as developed, restoration of CTCF levels using an inducible expression construct may exceed normal physiological CTCF levels.

STAR+METHODS

RESOURCE AVAILABILITY

Lead contact—Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Dr. Jules P.P. Meijerink (jules.meijerink@acerta-pharma.com).

Materials availability—Plasmids generated in this study are available upon request via the lead contact.

Data and code availability—Data generated by HiC, 4C, BeadChip, array CGH, ChIP-seq, TLA and WGS data have been deposited at GEO and are publicly available as of the date of publication. Accession numbers are listed in the key resources table.

This paper does not report original code.

Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Patient samples—Primary leukemia samples from 189 pediatric T-ALL patients were used (for patient details see Table S1). Eighty-five patients enrolled in the Dutch Childhood Oncology Group (DCOG) ALL-7 (n = 4), ALL-8 (n = 26), ALL-9 (n = 42) and ALL-10 (n = 13) study protocols. Ninety-six patients enrolled in the German Co-operative Study Group for Childhood Acute Lymphoblastic Leukemia (COALL)-97 (n = 30) and COALL-03 (n = 66). Breakpoint data from 8 selected pediatric T-ALL patient samples from the St Jude Research hospital were obtained. Functional analysis was performed on enriched, viably frozen leukemia cell fractions isolated from diagnostic bone marrow or peripheral blood biopsies. Leukemia blasts were enriched to 90–100% purities. The patient's parents

or legal guardians provided informed consent to the use of leftover diagnostic material for research purposes and is approved by the institutional review boards of the Erasmus Medical Center (Rotterdam) and the Princess Máxima Center for Pediatric Oncology (Utrecht) in accordance with the Declaration of Helsinki.

Patient derived xenograft (PDX) models—Leukemic cells ($1-10 \times 10^6$ cells) from pediatric T-ALL patients were intravenously transplanted into 8–12 weeks old female NOD *scid* gamma (NSG) mice^{96,97} that were purchased from Jackson's Laboratory (Bar Harbor, ME). Six weeks after transplantation and onwards, mice were bled on a weekly basis to monitor the percentage of human chimerism in the peripheral blood. At 80% human chimerism or at overt illness, mice were sacrificed and hematopoietic cells from the spleen and bone marrow were isolated and viably frozen or used for second rounds of transplantation.

Cell lines—The t(5; 14)-containing cell line HPB-ALL was obtained via cell repository DSMZ. The cell line was passaged twice per week at 0.35×10^6 cells per mL in 90% RPMI 1640 medium containing 1x Glutamax, and 10% heat-inactivated FBS and grown in a humidified incubator at 37C and 5% CO₂. Integrity of the cell line was checked regularly via short tandem repeat (STR) profiling.

Mice—The *Ctcf*^{fl/fl} mice⁷ were provided by Rudi Hendriks (Erasmus Medical Center Rotterdam). The *Lck-cre* mice⁹⁵ were purchased from Taconic (Rensselaer, NY). Further details are listed in the key resources table. The mice were bred and/or maintained at the animal care facilities of the Erasmus Medical Center Rotterdam and the Hubrecht Institute. Experimental procedures were approved by the Ethical Committees of Animal Welfare of the Erasmus University and Hubrecht Institute. Female and male mice were analyzed at 8–9 weeks of age, unless specified differently in figure legends.

METHOD DETAILS

Copy number analysis using array-comparative genomic hybridization—Array-comparative genomic hybridization was successfully performed for diagnostic patient samples of 94 pediatric T-ALL patients using the 2×400k array-CGH microarrays (Agilent Technologies, Santa Clara, CA; Design_ID 014698, 013282, and 021850) as generated using the manufacturer's protocol in a dye-swap experimental design to minimize false positive results as described before.⁹⁸ TIF images obtained by Agilent Scanner (model B and C) were analyzed with Cytogenomics v5.0.0.38 software to detect and visualize copy number variations at specific loci (or associated with the *CTCF* locus), employing settings in the Default Analysis Method – CGH v2 (ADM-2 algorithm with a threshold of 6.0, minimum of 3 consecutive statistically aberrant probes, and a minimum absolute amp/ del average of log₂ratio 0.25). Data is available at GEO repository (key resources table).

Single molecule molecular inversion probe (smMIP)—For detection of *CTCF* mutations, we applied single molecule Molecular Inversion Probe (smMIP) technology according to the original protocol with minor adaptations.⁶⁷ MIPgen software v1.2.1 was used to design Molecular Inversion Probes (MIPs) for all coding exons of *CTCF*

(NM_006565).⁹⁹ Regions of 100bp were targeted for capture. Each MIP contained a 20–24 nucleotide ligation arm, an 8nt molecular tag, a 30nt linker sequence and a 16–20nt elongation arm, for a total of 78 nucleotides per ssDNA molecule. Every position in the region of interest (ROI) was covered by at least two MIPs. Molecules were produced on a 25nm scale and provided in a TE solution (100µM) (Integrated DNA Technologies, Leuven, Belgium). Genomic DNA (100ng) from individual T-ALL patients was hybridized with phosphorylated MIPs and subsequently elongated and ligated; 33000 molecules of the ROI are expected in this reaction. MIPs were added in a 800-fold molar ratio to the genomic template, together with polymerase and ligase. Single-strand DNA circles were formed by a 60°C overnight incubation followed by nuclease treatment to remove non-ligated DNA molecules. Individual samples were barcoded and amplified in a 19-cycle PCR amplification step. Pooled libraries were cleaned up using AMPure XP beads (Beckman Coulter) to enrich for 277bp fragments. Sequencing was performed on the Illumina NextSeq500 Desktop Sequencer using 2×150bp paired-end sequencing. Custom index and sequence primers were added during this procedure. Mutations were called by the SeqNext plugin of SeqPilot software using standard settings (JSI, Ettenheim, Germany), using a minimum of 40 consensus reads that are covered by at least two independent MIPs. All primer and MIP sequences can be found in Table S2.

Gene expression profiling analysis—Affymetrix U133 Plus2 microarray data for the 117 patients as previously published⁷¹ was normalized using Robust Multichip Average (RMA), using Affy package.¹⁰⁰ Data is available at GEO repository (key resources table). From this cohort we selected *TLX3*-rearranged T-ALL patients with *CTCF* aberrations (n = 9) and *CTCF* wild type (WT) (n = 13) patients. *CTCF*-aberrations were compared to WT using Limma.¹⁰¹ $p < 0.05$, FDR < 0.1 and a Log fold-change of >1 was used as cut-off. Results are visualized in a volcano plot using the “ggplot2” package.

DNA methylation arrays—DNA methylation is measured on 853,307 CpG sites for seven samples using Illumina Methylation EPIC BeadChip Infinium microarray, resulting in raw intensity data (idat) files. Quality control on idat-files was performed using an in-house protocol. One sample had several independent and dependent probe failures that were removed from further analysis. Three of seven T-ALL patient samples were *CTCF* wild type, having either *BCL11B-CTCF* translocations (n = 2) or the *CAPSL-TLX3* translocation (n = 1). The remaining samples from 4 *BCL11B-TLX3* translocated T-ALL patients harbored *CTCF* deletions (n = 3) or a *CTCF* mutation (n = 1). Differential methylation analysis was performed comparing *CTCF* wild type patients to patients harboring *CTCF* aberrations based on the betavalues using dmpFinder from R-package minfi (version 1.28.3). The log₂ values of mean probe intensities are calculated for both groups. The output of the differential methylation analysis was flagged for all probes using significance cut-off values of $p < 0.05$, $q < 0.1$ and a fold change >1. Data are available at GEO repository (key resources table).

Chromatin Immunoprecipitation (ChIP) and peak calling—ChIP-seq was performed according to the SimpleChIP® Enzymatic Chromatin IP kit #9003 (Cell Signaling Technologies, Danvers, MA) procedure. Briefly, viably frozen diagnostic patient samples or patient-derived xenograft (PDX) cells were thawed and cross-linked in 1% formaldehyde

(Merck, St. Louis, MO) for 10'. The chromatin was sheared using the Bioruptor Pico (Diagenode, Liege, Belgium; 7 cycles of 30' on, 30' off). After sonication, samples were divided and $\sim 4 \times 10^6$ cells were used for each ChIP experiment. Antibodies were added to concentrations as recommended by the manufacturer and incubated overnight at 4°C under continuous rotation. Anti-CTCF, anti-SMC3 and normal rabbit IgG antibodies are listed in the key resources table. DNA was pelleted with ChIPgrade Protein G Magnetic Beads and washed according to the manufacturer's protocol. After reverse cross-linking, DNA was purified with spin columns from the SimpleChIP kit or the Qiaquick PCR Purification Kit (Qiagen, Hilden, Germany). DNA concentrations were measured using the Qubit HS DNA sensitivity kit (ThermoFischer, Waltham, MA). Libraries were prepared using the NEXTflex Rapid DNA Sequencing Kit (PerkinElmer). Samples were PCR amplified, checked for size and the absence of adaptor dimers on 2% agarose gel. Barcoded libraries were sequenced for 75 bp at a single end using the Illumina NextSeq500 sequencer.

BAM files from sequenced samples were merged using samtools if the total coverage was insufficient in a single sequence run.¹⁰² BAM files were used for peak calling. Peaks were identified by comparing results for ChIP samples versus input DNA control samples using MACS v2.1.1.20160309.¹⁰³ A q-value of 0.01 was applied while other settings were kept at default. We then classified CTCF peaks into:

- i. TAD-associated CTCF peaks located in <3 kb distance from TAD borders.
- ii. Upstream regions (<5kb) from transcriptional start sites (uTSS) of genes as present in the UCSC genome browser (<https://genome.ucsc.edu>).^{104,105}
- iii. Upstream regions (<5kb) from transcriptional start sites of long non-coding RNA genes (uLnc) as present in DeepBase^{106,107} and Lncpedia.¹⁰⁸
- iv. CTCF-peaks within enhancer (Enh) regions as present in GeneHancer.¹⁰⁹
- v. Not classified.

BEDTools was used to create bed files for each category and Deeptools^{110,111} was used to create centered heatmaps. RPKM values from each CTCF peak were visualized with color package "paired" using a scaling between 0 and 50. Following peak calling, we then looked for differentially-bound CTCF peaks comparing data from *CTCF* wild type patients to patients with evidence for heterozygous inactivation of *CTCF* due to deletions or mutations using the edgeR package.¹¹² The resulting p values were corrected for false discovery rate (FDR) using a Benjamini-Hochberg correction for multiple testing.¹¹³ Sites with an FDR <0.05 were considered for further analysis. This differential binding analysis and plots were performed using DiffBind, an R-package created by Rory Stark and Gordon Brown (2011). The package BEDTools¹¹⁴ was used to identify nearest upstream and downstream genes. Data are available at GEO repository (key resources table).

Chromatin conformation capture on chip (4C)—4C template was prepared following as described in the updated 4C-seq protocol⁸⁵ using DpnII (New England Biolabs (NEB), Ipswich, MA) as the primary restriction enzyme and Csp6I as the secondary restriction enzyme (NEB). With several small adaptations using 10 million viable cells as input: Ethanol precipitations of the template were replaced by on-bead Isopropanol precipitation

using NucleoMag P-beads (Macherey-Nagel, Düren, Germany). Inverse PCR was performed in two rounds, initially 16 cycles of PCR were performed with viewpoint specific primers. The PCR product was purified, and size selected using 0.6x AMPure XP beads (Beckman Coulter). Subsequently a second round of PCR (20 cycles) was performed with universal indexed primers. Sequencing was performed on the Illumina Nextseq platform using single-end sequencing (75bps). Data was mapped and analyzed using our in-house pipeline, as described before¹¹⁵. All 4C plots indicate the mean coverage rolling windows spanning 21 restriction fragments. For each sample, 2×10^7 cells were thawed and split into two replicates prior to crosslinking. For primary patient samples, single 4C experiments were performed. For all 4C experiments performed on cell lines, a minimum of 2 replicate experiments were performed. To compare relative interaction frequencies as indicated in Figures 6 and 7 between conditions, we summed the total 4C signals within a given region, i.e., enhancer region (chr14:98,602,411–98,675,204), competitive peak (chr5:170,716,369–170,756,369) or viewpoint (VP; chr5:170,653,532–170,710,663). The resulting frequencies were then divided by each other, and the resulting relative frequencies were compared between the different conditions. Data are available at GEO repository (key resources table).

Hi-C sequencing—The Hi-C protocol was adapted from the *in situ* Hi-C protocol as published.³⁰ Hi-C libraries were sequenced using Illumina Nextseq Paired-End 75bp sequencing. FastQ files were mapped to the human genome (GRCh37) using bwa-mem¹¹⁶ and filtered with removal of duplicates using HiCCUP v0.5.10.¹¹⁷ Chromosomal interaction matrices were generated using Juicer¹¹⁸ at 10 kb resolution and normalized by Knight and Ruiz's matrix balancing algorithm. TAD insulation scores for each 50Kb bin were calculated using HiCEplorer.⁷⁵ Data are available at GEO repository (key resources table).

Breakpoint analysis by targeted locus amplification or WGS—For the preparation of patient samples and cell lines we made use of the Targeted Locus Amplification,⁸² provided as a service by Cergentis BV, Utrecht, the Netherlands. We used the manufacturer's protocol to prepare the samples.¹¹⁹ Briefly, 5–10 million cells were cross-linked by adding 37% Formaldehyde to a final concentration of 1%. Cells were lysed and DNA was digested with NlaIII (New England Biolabs), followed by ligation with T4 DNA ligase. Following a crosslink removal step, the DNA was purified and digested using NspI (New England Biolabs) and ligated. The DNA was purified and a TLA PCR was performed with primers that were specifically designed for this study. The PCR product was purified by AMPure XP beads (Beckman Coulter, Brea, CA) and prepared for Next Generation Sequencing. Sequence data by Illumina MiSeq has been deposited at the Sequence Read Archive (SRA) (key resources table). The breakpoints of 3 T-ALL patients (#10929, #9319, #9452) have been determined before using Complete Genomics WGS sequencing platform¹²⁰ and are available from the European Nucleotide Archive database (<http://www.ebi.ac.uk/ena>) (key resources table). Breakpoint for the St Jude T-ALL patients have been determined by Illumina HiSeq 2000 WGS before,¹²¹ and are accessible through the European Genome-phenomenon Archive (key resources table).

Cloning of plasmids and virus production—*SFFV* promoter, IRES and *mCherry* reporter were stripped from lentiviral pLEGO-iC2 via *ApaI*/*PciI* sites to include a Gateway

compatible attR1-*ccdB-Cm^r*-attR2 cassette. Gateway compatible *attL1/attR5*-flanked pSFFV and *attL5/attL2*-flanked *TLX3*-T2A-Venus dsDNA fragments were synthesized, and then recombined into the Gateway-compatible pLEGO-iC2 backbone (Figure 6A) to generate a *TLX3* expression lentiviral transfer vector. Similarly, dsDNA synthesis and recombination were performed for *attL1/attR5*-flanked doxycycline-inducible promoter, *attL5/attL4*-flanked *mTagBFP-Thosea asigna* virus 2A peptide (T2A)-*CTCF*, *attR4/attR3*-flanked-WPRE-pSFFV, and *attL3/attL2*-flanked *TETon-T2A-NGFR* reporter, to generate an inducible *CTCF* lentiviral transfer vector (Figure 6A). For lentivirus production, HEK293T cells were transfected with transfer vector DNA and helper plasmids pMD2.G (VSV-G), pMDLg/pRRE, and pRSV-REV (Addgene), using 1 μL X-tremeGENE HP DNA Transfection Reagent (Roche) per 1 μg DNA. Transfection was performed in 90% DMEM containing 1× Glutamax, 1% penicillin/streptomycin, and 0.25 mg/mL Fungizone, and 10% heat-inactivated fetal calf serum (FCS) and cultured for 20 h at 37°C and 5% CO₂. Lentivirus particles were collected in Opti-MEM1 (Thermo Fisher Scientific) without serum for up to 48h. Culture medium containing lentiviral particles was filtered through a 0.45 μm filter and concentrated 22-fold using a VIVASPIN 20 concentration column (Sartorius).

Virus transduction and CTCF induction—For transduction, one volume of *TLX3* expression lentivirus was mixed with one volume of HPB-ALL cells (1×10^6 cells/mL) in RPMI 1640 at an end concentration of 1% FCS and incubated for 16 h in a humidified incubator at 37°C and 5% CO₂. Two volumes of 20% FCS-RPMI were added to limit further transduction. Medium was refreshed twice a week, until cell growth was observed under a microscope. Cells were then further cultured as described earlier. Three weeks after transduction, cells were purified for Venus positivity on a cell sorter (Sony Biotechnology, San Jose, CA, USA). Next, this HPB-ALL *TLX3*-Venus line was similarly transduced with a lentivirus containing doxycycline inducible *mTagBFP-T2A-CTCF* and a constitutive truncated *NGFR*. Three weeks after transduction, cells were stained with PE-CD271 (ME20.4–1.H4) (Miltenyi, Bergisch Gladbach, Germany) according to manufacturer's instructions, and purified for PE via cell sorting. Single cell clones were then grown out through a limiting dilution, and clones were chosen for further research based on robust *mTagBFP* expression on a flow cytometer. For these, and subsequent, experiments four days of exposure to doxycycline at a concentration of 200 ng/μL was used.

Genome editing—An *mTagBFP-CTCF* expressing subclone of HPB-ALL was chosen to delete an intervening *CTCF* binding site on the *TLX3* locus via CRISPR-Cas9 mediated technology. Four gRNAs were designed targeting ~150 bp up and downstream of the predicted *CTCF* site at Hg37:chr5:170,736,369, but also in the coding region of *HPRT1* (Table S3); disrupting *HPRT1* allows selection of NHEJ-repaired cells, thus also successfully transfected cells, via addition of 6-thioguanine (6-TG) to the culture medium. 3 μg Cas9 recombinant protein and a cocktail of 5 gRNA's (4.4 pmol each) were mixed in a 2 μL reaction volume and incubated at room temperature for 20 min to generate RNP complexes. 0.1 million cells in 8 μL were added to the RNP complex, and electroporated with a Neon Electroporator with program 1400V, 10 ms, 1 pulse. Cells were allowed to recover for seven days in previously described 1 mL culture medium. Following this, 0.25 μg/mL 6-TG was added during culture for ten days to select for cells without *HPRT1*

expression. Single cells were then grown out through a limiting dilution. DNA was extracted from 1 million cells of each outgrowing clone with EchoLUTION Cell Culture DNA Kit. Genotype, predicted as a 330 bp deletion, was determined on 20 ng DNA, using AmpliTaq Gold polymerase at recommended reaction conditions, modified to 2.5 mM MgCl₂, 58°C annealing for 32 cycles. Primers are listed in the key resources table. PCR products were assessed on a 1.5% agarose gel.

DNA and RNA isolation and cDNA synthesis—DNA and RNA was isolated with TRIzol reagent (Thermo Fisher) according to the guidelines of the manufacturer with minor modifications as described before.⁹⁸ Briefly, an additional phenol-chloroform-isomylalcohol (25:24:1, pH = 4.0) extraction step was performed for RNA purification to remove residual DNA and the RNA was precipitated using isopropanol along with 1µg of (20 µg/mL) glycogen (Roche, Almere, the Netherlands). After precipitation, RNA pellets were dissolved in 20 µL RNase-free TE-buffer (10mM Tris-HCl, 1mM EDTA, pH = 8.0). RNA concentrations were measured using a spectrophotometer. For reverse transcription of RNA into cDNA, 1mg of RNA was denatured for 5' at 80C, and reverse transcribed using a mix of random hexamers (2.5 µM, Life Technologies) and OligodT primers (20nM, Life Technologies). The RT-reaction was performed in a total volume of 25 µL containing 0.2 mM dNTPs, 200U Moloney murine leukemia virus reverse transcriptase (Promega, Madison, WI, USA) and 25U RNAsin (Promega). Conditions for the RT-reaction were 37C for 30' and 42C for 15' followed by an enzyme inactivation step at 94C for 5'. The cDNA was diluted to a final concentration of 1–8 ng/mL and stored at 80C.

Real time quantitative PCR—A DyNAmo HS SYBR Green PCR kit (Thermo Fisher, Waltham, MA) and CFX384 Touch[®] Real-Time PCR detection system (Biorad, Hercules, CA) was used for QRT-PCR in the presence of 3.75 pmol primers and a final concentration of 4 mM MgCl₂ in a total volume of 12.5 µL. Primers used in this study are listed in Table S2.

Flow cytometry—Antibodies used were diluted according to the manufacturer's instructions in staining buffer (PBS supplemented with 1% FBS), and included all conjugated anti-mouse antibodies in the key resources table. Cells were stained for 20–30 min on ice, in the dark. After staining the cells were washed and taken up in staining buffer supplemented with DAPI (Biolegend). Flow cytometry analysis was performed on an LSRII (BD Bioscience) and analyzed with FlowJo software (BD Bioscience).^{96,97}

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical tests have been described in text of figure legends with the number of replicates specified. Differences in *CTCF* levels among patients have been determined using the Mann-Whitney U test using SPSS Statistics 26 software. Significant differences based on flow cytometry analyses as used in Figure 3 and for relative *TLX3* expression levels in Figure 7 have been determined using an unpaired t test. Differences among patients in genomic distances between *TLX3* and the *BCL11B* enhancer in t(5; 14) breakpoint (Figure 4D), relative *TLX3* expression to *BCL11B* expression levels (Figure 4E) and white blood cell counts (Figure 4F) was done using the Kruskal-Wallis test using SPSS software. The

p value corrections for false discovery rate (FDR) for various genomic analyses was done using a Benjamini-Hochberg correction for multiple testing¹¹³ in the statistical environment R and have been described in the corresponding sections in STAR Methods.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

W.K.S., R.H., and E.M.V. were sponsored by the Dutch Cancer Society (KWF4691) and the “Kinderen Kankervrij” foundation (grants KiKa 2015–244 to E.M.V., KiKa 2013–219 to R.H., and KiKa 2017–295 to W.K.S.). E.v.d.G. was sponsored by KiKa core funding. W.K.S. and J.G.C.A.M.B.-G. are also sponsored by the foundation “Stichting Kinder Oncologisch Centrum Rotterdam.” C.V. and M.J.A.M.V. were sponsored by the Foundation Leducq (grant 14CVD01). C.V. and M.J.A.M.V. were also sponsored by the Netherlands Scientific Organization NWO: NWO VICI-grant 724.012.003, and C.V. is currently supported by a NWO Vidi Fellowship (639.072.715). C.G.M. is supported by the American Lebanese Syrian Associated Charities of St. Jude Children’s Research Hospital and a National Cancer Institute Outstanding Investigator Award R35 CA197695.

REFERENCES

- Holwerda SJB, and de Laat W. (2013). CTCF: the protein, the bindingpartners, the binding sites and their chromatin loops. *Philos. Trans. R. Soc. Lond. B Biol. Sci* 368, 20120369. 10.1098/rstb.2012.0369.
- Ong CT, and Corces VG (2014). CTCF: an architectural proteinbridging genome topology and function. *Nat. Rev. Genet* 15, 234–246. 10.1038/nrg3663. [PubMed: 24614316]
- Phillips JE, and Corces VG (2009). CTCF: master weaver of thegenome. *Cell* 137, 1194–1211. 10.1016/j.cell.2009.06.001. [PubMed: 19563753]
- Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, Wang Z, Wei G, Chepelev I, and Zhao K. (2007). High-resolution profiling of histone methylations in the human genome. *Cell* 129, 823–837. 10.1016/j.cell.2007.05.009. [PubMed: 17512414]
- Jothi R, Cuddapah S, Barski A, Cui K, and Zhao K. (2008). Genomewide identification of in vivo protein-DNA binding sites from ChIP-Seq data. *Nucleic Acids Res.* 36, 5221–5231. 10.1093/nar/gkn488. [PubMed: 18684996]
- Kim TH, Abdullaev ZK, Smith AD, Ching KA, Loukinov DI, Green RD, Zhang MQ, Lobanenko VV, and Ren B. (2007). Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. *Cell* 128, 1231–1245. 10.1016/j.cell.2006.12.048. [PubMed: 17382889]
- Heath H, Ribeiro de Almeida C, Sleutels F, Dingjan G, van de Nobelen S, Jonkers I, Ling KW, Gribnau J, Renkawitz R, Grosveld F, et al. (2008). CTCF regulates cell cycle progression of alphabeta T cells in the thymus. *EMBO J.* 27, 2839–2850. 10.1038/emboj.2008.214. [PubMed: 18923423]
- Vostrov AA, and Quitschke WW (1997). The zinc finger protein CTCF binds to the APBbeta domain of the amyloid beta-protein precursor promoter. Evidence for a role in transcriptional activation. *J. Biol. Chem* 272, 33353–33359. [PubMed: 9407128]
- Filippova GN, Fagerlie S, Klenova EM, Myers C, Dehner Y, Goodwin G, Neiman PE, Collins SJ, and Lobanenko VV (1996). An exceptionally conserved transcriptional repressor, CTCF, employs different combinations of zinc fingers to bind diverged promoter sequences of avian and mammalian c-myc oncogenes. *Mol. Cell Biol* 16, 2802–2813. [PubMed: 8649389]
- Klenova EM, Nicolas RH, Paterson HF, Carne AF, Heath CM, Goodwin GH, Neiman PE, and Lobanenko VV (1993). CTCF, a conserved nuclear factor required for optimal transcriptional activity of the chicken c-myc gene, is an 11-Zn-finger protein differentially expressed in multiple forms. *Mol. Cell Biol* 13, 7612–7624. [PubMed: 8246978]

11. Köhne AC, Baniahmad A, and Renkawitz R. (1993). NeP1. A ubiquitous transcription factor synergizes with v-ERBA in transcriptional silencing. *J. Mol. Biol* 232, 747–755. 10.1006/jmbi.1993.1428. [PubMed: 8102652]
12. Bell AC, West AG, and Felsenfeld G. (1999). The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. *Cell* 98, 387–396. [PubMed: 10458613]
13. Recillas-Targa F, Bell AC, and Felsenfeld G. (1999). Positional enhancer-blocking activity of the chicken beta-globin insulator in transiently transfected cells. *Proc. Natl. Acad. Sci. USA* 96, 14354–14359. [PubMed: 10588709]
14. Bell AC, and Felsenfeld G. (2000). Methylation of a CTCF-dependent boundary controls imprinted expression of the *Igf2* gene. *Nature* 405, 482–485. 10.1038/35013100. [PubMed: 10839546]
15. Hark AT, Schoenherr CJ, Katz DJ, Ingram RS, Levorse JM, and Tilghman SM (2000). CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/*Igf2* locus. *Nature* 405, 486–489. 10.1038/35013106. [PubMed: 10839547]
16. Kanduri C, Pant V, Loukinov D, Pugacheva E, Qi CF, Wolffe A, Ohlsson R, and Lobanekov VV (2000). Functional association of CTCF with the insulator upstream of the H19 gene is parent of origin-specific and methylation-sensitive. *Curr. Biol* 10, 853–856. [PubMed: 10899010]
17. Saitoh N, Bell AC, Recillas-Targa F, West AG, Simpson M, Pikaart M, and Felsenfeld G. (2000). Structural and functional conservation at the boundaries of the chicken beta-globin domain. *EMBO J.* 19, 2315–2322. 10.1093/emboj/19.10.2315. [PubMed: 10811622]
18. Szabó P, Tang SH, Rentsendorj A, Pfeifer GP, and Mann JR (2000). Maternal-specific footprints at putative CTCF sites in the H19 imprinting control region give evidence for insulator function. *Curr. Biol* 10, 607–610. [PubMed: 10837224]
19. Lobanekov VV, Nicolas RH, Adler VV, Paterson H, Klenova EM, Polotskaja AV, and Goodwin GH (1990). A novel sequence-specific DNA binding protein which interacts with three regularly spaced direct repeats of the CCCTC-motif in the 5' -flanking sequence of the chicken c-myc gene. *Oncogene* 5, 1743–1753. [PubMed: 2284094]
20. Guo C, Yoon HS, Franklin A, Jain S, Ebert A, Cheng HL, Hansen E, Despo O, Bossen C, Vettermann C, et al. (2011). CTCF-binding elements mediate control of V(D)J recombination. *Nature* 477, 424–430. 10.1038/nature10495. [PubMed: 21909113]
21. Splinter E, Heath H, Kooren J, Palstra RJ, Klous P, Grosveld F, Galjart N, and de Laat W. (2006). CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus. *Genes Dev.* 20, 2349–2354. 10.1101/gad.399506. [PubMed: 16951251]
22. Hou C, Zhao H, Tanimoto K, and Dean A. (2008). CTCF-dependent enhancer-blocking by alternative chromatin loop formation. *Proc. Natl. Acad. Sci. USA* 105, 20398–20403. 10.1073/pnas.0808506106. [PubMed: 19074263]
23. Kurukuti S, Tiwari VK, Tavosidana G, Pugacheva E, Murrell A, Zhao Z, Lobanekov V, Reik W, and Ohlsson R. (2006). CTCF binding at the H19 imprinting control region mediates maternally inherited higher-order chromatin conformation to restrict enhancer access to *Igf2*. *Proc. Natl. Acad. Sci. USA* 103, 10684–10689. 10.1073/pnas.0600326103. [PubMed: 16815976]
24. Yoon YS, Jeong S, Rong Q, Park KY, Chung JH, and Pfeifer K (2007). Analysis of the H19 ICR insulator. *Mol. Cell Biol* 27, 3499–3510. 10.1128/MCB.02170-06. [PubMed: 17339341]
25. Tolhuis B, Palstra RJ, Splinter E, Grosveld F, and de Laat W (2002). Looping and interaction between hypersensitive sites in the active beta-globin locus. *Mol. Cell* 10, 1453–1465. 10.1016/S1097-2765(02)00781-5. [PubMed: 12504019]
26. Sexton T, and Cavalli G. (2013). The 3D genome shapes up for pluripotency. *Cell Stem Cell* 13, 3–4. 10.1016/j.stem.2013.06.013. [PubMed: 23827701]
27. Nora EP, Lajoie BR, Schulz EG, Giorgetti L, Okamoto I, Servant N, Piolot T, van Berkum NL, Meisig J, Sedat J, et al. (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* 485, 381–385. 10.1038/nature11049. [PubMed: 22495304]
28. Jin F, Li Y, Dixon JR, Selvaraj S, Ye Z, Lee AY, Yen CA, Schmitt AD, Espinoza CA, and Ren B. (2013). A high-resolution map of the three-dimensional chromatin interactome in human cells. *Nature* 503, 290–294. 10.1038/nature12644. [PubMed: 24141950]

29. Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS, and Ren B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 485, 376–380. 10.1038/nature11082. [PubMed: 22495300]
30. Rao SSP, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I, Omer AD, Lander ES, and Aiden EL (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* 159, 1665–1680. 10.1016/j.cell.2014.11.021. [PubMed: 25497547]
31. Nora EP, Goloborodko A, Valton AL, Gibcus JH, Uebersohn A, Abdennur N, Dekker J, Mirny LA, and Bruneau BG (2017). Targeted degradation of CTCF decouples local insulation of chromosome domains from genomic compartmentalization. *Cell* 169, 930–944.e22. 10.1016/j.cell.2017.05.004. [PubMed: 28525758]
32. Wendt KS, Yoshida K, Itoh T, Bando M, Koch B, Schirghuber E, Tsutsumi S, Nagae G, Ishihara K, Mishiro T, et al. (2008). Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature* 451, 796–801. 10.1038/nature06634. [PubMed: 18235444]
33. Feeney AJ, and Verma-Gaur J. (2012). CTCF-cohesin complex: architect of chromatin structure regulates V(D)J rearrangement. *Cell Res.* 22, 280–282. 10.1038/cr.2011.188. [PubMed: 22105486]
34. Baranello L, Kouzine F, and Levens D. (2014). CTCF and cohesin cooperate to organize the 3D structure of the mammalian genome. *Proc. Natl. Acad. Sci. USA* 111, 889–890. 10.1073/pnas.1321957111. [PubMed: 24398527]
35. Seitan VC, Hao B, Tachibana-Konwalski K, Lavagnolli T, Mira-Bontenbal H, Brown KE, Teng G, Carroll T, Terry A, Horan K, et al. (2011). A role for cohesin in T-cell-receptor rearrangement and thymocyte differentiation. *Nature* 476, 467–471. 10.1038/nature10312. [PubMed: 21832993]
36. Seitan VC, Krangel MS, and Merkenschlager M. (2012). Cohesin, CTCF and lymphocyte antigen receptor locus rearrangement. *Trends Immunol.* 33, 153–159. 10.1016/j.it.2012.02.004. [PubMed: 22440186]
37. Nativio R, Wendt KS, Ito Y, Huddleston JE, Uribe-Lewis S, Woodfine K, Krueger C, Reik W, Peters JM, and Murrell A. (2009). Cohesin is required for higher-order chromatin conformation at the imprinted IGF2-H19 locus. *PLoS Genet.* 5, e1000739. 10.1371/journal.pgen.1000739.
38. Chien R, Zeng W, Kawauchi S, Bender MA, Santos R, Gregson HC, Schmiesing JA, Newkirk DA, Kong X, Ball AR Jr., et al. (2011). Cohesin mediates chromatin interactions that regulate mammalian beta-globin expression. *J. Biol. Chem* 286, 17870–17878. 10.1074/jbc.M110.207365. [PubMed: 21454523]
39. Rao SSP, Huang SC, Glenn St Hilaire B, Engreitz JM, Perez EM, Kieffer-Kwon KR, Sanborn AL, Johnstone SE, Bascom GD, Bochkov ID, et al. (2017). Cohesin loss eliminates all loop domains. *Cell* 171, 305–320.e24. 10.1016/j.cell.2017.09.026. [PubMed: 28985562]
40. Li Y, Haarhuis JHI, Sedeño Cacciatore Á, Oldenkamp R, Ruiten MS, Willems L, Teunissen H, Muir KW, de Wit E, Rowland BD, and Panne D. (2020). The structural basis for cohesin-CTCF anchored loops. *Nature* 578, 472–476. 10.1038/s41586019-1910-z. [PubMed: 31905366]
41. Schwarzer W, Abdennur N, Goloborodko A, Pekowska A, Fudenberg G, Loe-Mie Y, Fonseca NA, Huber W, Haering CH, Mirny L, and Spitz F. (2017). Two independent modes of chromatin organization revealed by cohesin removal. *Nature* 551, 51–56. 10.1038/nature24281. [PubMed: 29094699]
42. Fudenberg G, Imakaev M, Lu C, Goloborodko A, Abdennur N, and Mirny LA (2016). Formation of chromosomal domains by loop extrusion. *Cell Rep.* 15, 2038–2049. 10.1016/j.celrep.2016.04.085. [PubMed: 27210764]
43. Sanborn AL, Rao SSP, Huang SC, Durand NC, Huntley MH, Jewett AI, Bochkov ID, Chinnappan D, Cutkosky A, Li J, et al. (2015). Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proc. Natl. Acad. Sci. USA* 112, E6456–E6465. 10.1073/pnas.1518552112. [PubMed: 26499245]
44. Davidson IF, Bauer B, Goetz D, Tang W, Wutz G, and Peters JM (2019). DNA loop extrusion by human cohesin. *Science* 366, 1338–1345. 10.1126/science.aaz3418. [PubMed: 31753851]
45. Kim Y, Shi Z, Zhang H, Finkelstein IJ, and Yu H. (2019). Human cohesin compacts DNA by loop extrusion. *Science* 366, 1345–1349. 10.1126/science.aaz4475. [PubMed: 31780627]

46. Rubio ED, Reiss DJ, Welch PL, Distèche CM, Filippova GN, Baliga NS, Aebersold R, Ranish JA, and Krumm A. (2008). CTCF physically links cohesin to chromatin. *Proc. Natl. Acad. Sci. USA* 105, 8309–8314. 10.1073/pnas.0801273105. [PubMed: 18550811]
47. Parelho V, Hadjur S, Spivakov M, Leleu M, Sauer S, Gregson HC, Jarmuz A, Canzonetta C, Webster Z, Nesterova T, et al. (2008). Cohesins functionally associate with CTCF on mammalian chromosome arms. *Cell* 132, 422–433. 10.1016/j.cell.2008.01.011. [PubMed: 18237772]
48. Stedman W, Kang H, Lin S, Kissil JL, Bartolomei MS, and Lieberman PM (2008). Cohesins localize with CTCF at the KSHV latency control region and at cellular c-myc and H19/Igf2 insulators. *EMBO J.* 27, 654–666. 10.1038/emboj.2008.1. [PubMed: 18219272]
49. Losada A. (2014). Cohesin in cancer: chromosome segregation and beyond. *Nat. Rev. Cancer* 14, 389–393. 10.1038/nrc3743. [PubMed: 24854081]
50. Remeseiro S, and Losada A. (2013). Cohesin, a chromatin engagement ring. *Curr. Opin. Cell Biol.* 25, 63–71. 10.1016/j.ceb.2012.10.013. [PubMed: 23219370]
51. de Wit E, Vos ESM, Holwerda SJB, Valdes-Quezada C, Verstegen MJAM, Teunissen H, Splinter E, Wijchers PJ, Krijger PHL, and de Laat W. (2015). CTCF binding polarity determines chromatin looping. *Mol. Cell* 60, 676–684. 10.1016/j.molcel.2015.09.023. [PubMed: 26527277]
52. Haarhuis JHI, van der Weide RH, Blomen VA, Yáñez-Cuna JO, Amendola M, van Ruiten MS, Krijger PHL, Teunissen H, Medema RH, van Steensel B, et al. (2017). The cohesin release factor WAPL restricts chromatin loop extension. *Cell* 169, 693–707.e14. 10.1016/j.cell.2017.04.013. [PubMed: 28475897]
53. Williams MS, and Somervaille TCP (2015). Leukemogenic activity of cohesin rings true. *Cell Stem Cell* 17, 642–644. 10.1016/j.stem.2015.11.008. [PubMed: 26637939]
54. Surdez D, Zaidi S, Grossetete S, Laud-Duval K, Ferre AS, Mous, L, Vourc’h T, Tirode F, Pierron G, Raynal V, et al. (2021). STAG2 mutations alter CTCF-anchored loop extrusion, reduce cis-regulatory interactions and EWSR1-FLI1 activity in Ewing sarcoma. *Cancer Cell* 39, 810–826.e9. 10.1016/j.ccell.2021.04.001. [PubMed: 33930311]
55. Adane B, Alexe G, Seong BKA, Lu D, Hwang EE, Hnisz D, Lareau CA, Ross L, Lin S, Dela Cruz FS, et al. (2021). STAG2 loss rewires oncogenic and developmental programs to promote metastasis in Ewing sarcoma. *Cancer Cell* 39, 827–844.e10. 10.1016/j.ccell.2021.05.007. [PubMed: 34129824]
56. Mazumdar C, Shen Y, Xavy S, Zhao F, Reinisch A, Li R, Corces MR, Flynn RA, Buenrostro JD, Chan SM, et al. (2015). Leukemia-associated cohesin mutants dominantly enforce stem cell programs and impair human hematopoietic progenitor differentiation. *Cell Stem Cell* 17, 675–688. 10.1016/j.stem.2015.09.017. [PubMed: 26607380]
57. Mullenders J, Aranda-Orgilles B, Lhoumaud P, Keller M, Pae J, Wang K, Kayembe C, Rocha PP, Raviram R, Gong Y, et al. (2015). Cohesin loss alters adult hematopoietic stem cell homeostasis, leading to myeloproliferative neoplasms. *J. Exp. Med* 212, 1833–1850. 10.1084/jem.20151323. [PubMed: 26438359]
58. Viny AD, Ott CJ, Spitzer B, Rivas M, Meydan C, Papalexi E, Yelin D, Shank K, Reyes J, Chiu A, et al. (2015). Dose-dependent role of the cohesin complex in normal and malignant hematopoiesis. *J. Exp. Med* 212, 1819–1832. 10.1084/jem.20151317. [PubMed: 26438361]
59. Kemp CJ, Moore JM, Moser R, Bernard B, Teater M, Smith LE, Rabaia NA, Gurley KE, Guinney J, Busch SE, et al. (2014). CTCF haploinsufficiency destabilizes DNA methylation and predisposes to cancer. *Cell Rep.* 7, 1020–1029. 10.1016/j.celrep.2014.04.004. [PubMed: 24794443]
60. Filippova GN, Lindblom A, Meincke LJ, Klenova EM, Neiman PE, Collins SJ, Doggett NA, and Lobanov VV (1998). A widely expressed transcription factor with multiple DNA sequence specificity, CTCF, is localized at chromosome segment 16q22.1 within one of the smallest regions of overlap for common deletions in breast and prostate cancers. *Genes Chromosomes Cancer* 22, 26–36. [PubMed: 9591631]
61. Filippova GN, Qi CF, Ulmer JE, Moore JM, Ward MD, Hu YJ, Loukinov DI, Pugacheva EM, Klenova EM, Grundy PE, et al. (2002). Tumor-associated zinc finger mutations in the CTCF transcription factor selectively alter its DNA-binding specificity. *Cancer Res.* 62, 48–52. [PubMed: 11782357]

62. Debaugny RE, and Skok JA (2020). CTCF and CTCFL in cancer. *Curr. Opin. Genet. Dev* 61, 44–52. 10.1016/j.gde.2020.02.021. [PubMed: 32334335]
63. Liu Y, Easton J, Shao Y, Maciaszek J, Wang Z, Wilkinson MR, McCastlain K, Edmonson M, Pounds SB, Shi L, et al. (2017). The genomic landscape of pediatric and young adult T-lineage acute lymphoblastic leukemia. *Nat. Genet* 49, 1211–1218. 10.1038/ng.3909. [PubMed: 28671688]
64. Ding LW, Sun QY, Tan KT, Chien W, Mayakonda A, Yeoh AEJ, Kawamata N, Nagata Y, Xiao JF, Loh XY, et al. (2017). Mutational landscape of pediatric acute lymphoblastic leukemia. *Cancer Res.* 77, 390–400. 10.1158/0008-5472.CAN-16-1303. [PubMed: 27872090]
65. Cancer Genome Atlas Research Network; Kandath C, Schultz N, Cherniack AD, Akbani R, Liu Y, Shen H, Robertson AG, Pashtan I, Shen R, et al. (2013). Integrated genomic characterization of endometrial carcinoma. *Nature* 497, 67–73. 10.1038/nature12113. [PubMed: 23636398]
66. Damaschke NA, Gawdzik J, Avilla M, Yang B, Svaren J, Roopra A, Luo JH, Yu YP, Keles S, and Jarrard DF (2020). CTCF loss mediates unique DNA hypermethylation landscapes in human cancers. *Clin. Epigenetics* 12, 80. 10.1186/s13148-020-00869-7. [PubMed: 32503656]
67. Hiatt JB, Pritchard CC, Salipante SJ, O’Roak BJ, and Shendure J. (2013). Single molecule molecular inversion probes for targeted, high accuracy detection of low-frequency variation. *Genome Res.* 23, 843–854. 10.1101/gr.147686.112. [PubMed: 23382536]
68. Zhou X, Edmonson MN, Wilkinson MR, Patel A, Wu G, Liu Y, Li Y, Zhang Z, Rusch MC, Parker M, et al. (2016). Exploring genomic alteration in pediatric cancer using ProteinPaint. *Nat. Genet* 48, 4–6. 10.1038/ng.3466. [PubMed: 26711108]
69. Soulier J, Clappier E, Cayuela JM, Regnault A, García-Peydró M, Dombret H, Baruchel A, Toribio ML, and Sigaux F. (2005). HOXA genes are included in genetic and biologic networks defining human acute T-cell leukemia (T-ALL). *Blood* 106, 274–286. [PubMed: 15774621]
70. Ferrando AA, Neuberg DS, Staunton J, Loh ML, Huard C, Raimondi SC, Behm FG, Pui CH, Downing JR, Gilliland DG, et al. (2002). Gene expression signatures define novel oncogenic pathways in T cell acute lymphoblastic leukemia. *Cancer Cell* 1, 75–87. [PubMed: 12086890]
71. Homminga I, Pieters R, Langerak AW, de Rooi JJ, Stubbs A, Verstegen M, Vuerhard M, Buijs-Gladdines J, Kooi C, Klous P, et al. (2011). Integrated transcript and genome analyses reveal NKX2-1 and MEF2C as potential oncogenes in T cell acute lymphoblastic leukemia. *Cancer Cell* 19, 484–497. 10.1016/j.ccr.2011.02.008. [PubMed: 21481790]
72. Chaumeil J, and Skok JA (2012). The role of CTCF in regulating V(D)J recombination. *Curr. Opin. Immunol* 24, 153–159. 10.1016/j.coi.2012.01.003. [PubMed: 22424610]
73. Shih HY, Verma-Gaur J, Torkamani A, Feeney AJ, Galjart N, and Krangel MS (2012). Tcr α gene recombination is supported by a Tcr α enhancer- and CTCF-dependent chromatin hub. *Proc. Natl. Acad. Sci. USA* 109, E3493–E3502. 10.1073/pnas.1214131109. [PubMed: 23169622]
74. Shih HY, and Krangel MS (2013). Chromatin architecture, CCCTC-binding factor, and V(D)J recombination: managing long distance relationships at antigen receptor loci. *J. Immunol* 190, 4915–4921. 10.4049/jimmunol.1300218. [PubMed: 23645930]
75. Wolff J, Rabbani L, Gilsbach R, Richard G, Manke T, Backofen R, and Gruning BA (2020). Galaxy HiCExplorer 3: a web server for reproducible Hi-C, capture Hi-C and single-cell Hi-C data analysis, quality control and visualization. *Nucleic Acids Res.* 48, W177–W184. 10.1093/nar/gkaa220. [PubMed: 32301980]
76. van Grotel M, Meijerink JPP, Beverloo HB, Langerak AW, Buys-Gladdines JGCAM, Schneider P, Poulsen TS, den Boer ML, Horstmann M, Kamps WA, et al. (2006). The outcome of molecular-cytogenetic subgroups in pediatric T-cell acute lymphoblastic leukemia: a retrospective study of patients treated according to DCOG or COALL protocols. *Haematologica* 91, 1212–1221. [PubMed: 16956820]
77. Asnafi V, Beldjord K, Libura M, Villarese P, Millien C, Ballerini P, Kuhlein E, Lafage-Pochitaloff M, Delabesse E, Bernard O, and Macintyre E. (2004). Age-related phenotypic and oncogenic differences in T-cell acute lymphoblastic leukemias may reflect thymic atrophy. *Blood* 104, 4173–4180. 10.1182/blood-2003-11-3944. [PubMed: 15054041]
78. Bernard OA, Busson-LeConiat M, Ballerini P, Mauchauffé M, Della Valle V, Monni R, Nguyen Khac F, Mercher T, Penard-Lacronique V, Pasturaud P, et al. (2001). A new recurrent and specific

- cryptic translocation, t(5;14)(q35;q32), is associated with expression of the Hox11L2 gene in T acute lymphoblastic leukemia. *Leukemia* 15, 1495–1504. [PubMed: 11587205]
79. Wakabayashi Y, Watanabe H, Inoue J, Takeda N, Sakata J, Mishima Y, Hitomi J, Yamamoto T, Utsuyama M, Niwa O, et al. (2003). Bcl11b is required for differentiation and survival of alphabeta T lymphocytes. *Nat. Immunol* 4, 533–539. 10.1038/ni927. [PubMed: 12717433]
 80. Li L, Leid M, and Rothenberg EV (2010). An early T cell lineage commitment checkpoint dependent on the transcription factor Bcl11b. *Science* 329, 89–93. 10.1126/science.1188989. [PubMed: 20595614]
 81. Isoda T, Moore AJ, He Z, Chandra V, Aida M, Denholtz M, Pietvan Hamburg J, Fisch KM, Chang AN, Fahl SP, et al. (2017). Non-coding transcription instructs chromatin folding and compartmentalization to dictate enhancer-promoter communication and T cell fate. *Cell* 171, 103–119.e18. 10.1016/j.cell.2017.09.001. [PubMed: 28938112]
 82. de Vree PJP, de Wit E, Yilmaz M, van de Heijning M, Klous P, Verstegen MJAM, Wan Y, Teunissen H, Krijger PHL, Geeven G, et al. (2014). Targeted sequencing by proximity ligation for comprehensive variant detection and local haplotyping. *Nat. Biotechnol* 32, 1019–1025. 10.1038/nbt.2959. [PubMed: 25129690]
 83. Nagel S, Scherr M, Kel A, Hornischer K, Crawford GE, Kaufmann M, Meyer C, Drexler HG, and MacLeod RAF (2007). Activation of TLX3 and NKX2–5 in t(5;14)(q35;q32) T-cell acute lymphoblastic leukemia by remote 3'-BCL11B enhancers and coregulation by PU.1 and HMGA1. *Cancer Res.* 67, 1461–1471. 10.1158/0008-5472.CAN-06-2615. [PubMed: 17308084]
 84. Li L, Zhang JA, Dose M, Kueh HY, Mosadeghi R, Gounari F, and Rothenberg EV (2013). A far downstream enhancer for murine Bcl11b controls its T-cell specific expression. *Blood* 122, 902–911. 10.1182/blood-2012-08-447839. [PubMed: 23741008]
 85. Krijger PHL, Geeven G, Bianchi V, Hilvering CRE, and de Laat W (2020). 4C-seq from beginning to end: a detailed protocol for sample preparation and data analysis. *Methods* 170, 17–32. 10.1016/j.ymeth.2019.07.014. [PubMed: 31351925]
 86. Dadi S, Le Noir S, Payet-Bornet D, Lhermitte L, Zacarias-Cabeza J, Bergeron J, Villarè se, Vachez, Dik, Millien C, et al. (2012). TLX homeodomain oncogenes mediate T cell maturation arrest in T-ALL via interaction with ETS1 and suppression of TCRalpha gene expression. *Cancer Cell* 21, 563–576. 10.1016/j.ccr.2012.02.013. [PubMed: 22516263]
 87. Flavahan WA, Drier Y, Liao BB, Gillespie SM, Venteicher AS, Stemmer-Rachamimov AO, Suvà ML, and Bernstein BE (2016). Insulator dysfunction and oncogene activation in IDH mutant gliomas. *Nature* 529, 110–114. 10.1038/nature16490. [PubMed: 26700815]
 88. Kloetgen A, Thandapani P, Ntziachristos P, Ghebrechristos Y, Nomikou S, Lazaris C, Chen X, Hu H, Bakogianni S, Wang J, et al. (2020). Three-dimensional chromatin landscapes in T cell acute lymphoblastic leukemia. *Nat. Genet* 52, 388–400. 10.1038/s41588-020-0602-9. [PubMed: 32203470]
 89. Flavahan WA, Drier Y, Johnstone SE, Hemming ML, Tarjan DR, Hegazi E, Shareef SJ, Javed NM, Raut CP, Eschle BK, et al. (2019). Altered chromosomal topology drives oncogenic programs in SDH-deficient GISTs. *Nature* 575, 229–233. 10.1038/s41586-019-1668-3. [PubMed: 31666694]
 90. Fang C, Wang Z, Han C, Safgren SL, Helmin KA, Adelman ER, Serafin V, Basso G, Eagen KP, Gaspar-Maia A, et al. (2020). Cancer-specific CTCF binding facilitates oncogenic transcriptional dysregulation. *Genome Biol.* 21, 247. 10.1186/s13059-02002152-7. [PubMed: 32933554]
 91. Marshall AD, Bailey CG, Champ K, Vellozzi M, O'Young P, Metierre C, Feng Y, Thoeng A, Richards AM, Schmitz U, et al. (2017). CTCF genetic alterations in endometrial carcinoma are protumorigenic. *Oncogene* 36, 4100–4110. 10.1038/onc.2017.25. [PubMed: 28319062]
 92. van Grotel M, Meijerink JPP, van Wering ER, Langerak AW, Beverloo HB, Buijs-Gladdines JGCAM, Burger NB, Passier M, van Lieshout EM, Kamps WA, et al. (2008). Prognostic significance of molecular-cytogenetic abnormalities in pediatric T-ALL is not explained by immunophenotypic differences. *Leukemia* 22, 124–131. 10.1038/sj.leu.2404957. [PubMed: 17928886]
 93. Zhao H, Li Z, Zhu Y, Bian S, Zhang Y, Qin L, Naik AK, He J, Zhang Z, Krangel MS, and Hao B. (2020). A role of the CTCF binding site at enhancer Ealpha in the dynamic chromatin organization of the Tcrα-Tcrδ locus. *Nucleic Acids Res.* 48, 9621–9636. 10.1093/nar/gkaa711. [PubMed: 32853367]

94. Kentepozidou E, Aitken SJ, Feig C, Stefflova K, Ibarra-Soria X, Odom DT, Roller M, and Flicek P. (2020). Clustered CTCF binding is an evolutionary mechanism to maintain topologically associating domains. *Genome Biol.* 21, 5. 10.1186/s13059-019-1894-x. [PubMed: 31910870]
95. Lee PP, Fitzpatrick DR, Beard C, Jessup HK, Lehar S, Makar KW, Pérez-Melgosa M, Sweetser MT, Schlissel MS, Nguyen S, et al. (2001). A critical role for Dnmt1 and DNA methylation in T cell development, function, and survival. *Immunity* 15, 763–774. [PubMed: 11728338]
96. Ishikawa F, Yasukawa M, Lyons B, Yoshida S, Miyamoto T, Yoshimoto G, Watanabe T, Akashi K, Shultz LD, and Harada M. (2005). Development of functional human blood and immune systems in NOD/SCID/IL2 receptor γ chain(null) mice. *Blood* 106, 1565–1573. 10.1182/blood-2005-02-0516. [PubMed: 15920010]
97. Shultz LD, Lyons BL, Burzenski LM, Gott B, Chen X, Chaleff S, Kotb M, Gillies SD, King M, Mangada J, et al. (2005). Human lymphoid and myeloid cell development in NOD/LtSz-scid IL2R γ null mice engrafted with mobilized human hemopoietic stem cells. *J. Immunol.* 174, 6477–6489. [PubMed: 15879151]
98. Van Vlierberghe P, van Grotel M, Beverloo HB, Lee C, Helgason T, Buijs-Gladdines J, Passier M, van Wering ER, Veerman AJP, Kamps WA, et al. (2006). The cryptic chromosomal deletion del(11)(p12p13) as a new activation mechanism of LMO2 in pediatric T-cell acute lymphoblastic leukemia. *Blood* 108, 3520–3529. 10.1182/blood-2006-04-019927. [PubMed: 16873670]
99. Boyle EA, O’Roak BJ, Martin BK, Kumar A, and Shendure J (2014). MIPgen: optimized modeling and design of molecular inversion probes for targeted resequencing. *Bioinformatics* 30, 2670–2672. 10.1093/bioinformatics/btu353. [PubMed: 24867941]
100. Irizarry RA, Bolstad BM, Collin F, Cope LM, Hobbs B, and Speed TP (2003). Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res.* 31, e15. 10.1093/nar/gng015. [PubMed: 12582260]
101. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, and Smyth GK (2015). Limma powers differential expression analyses for RNAsequencing and microarray studies. *Nucleic Acids Res.* 43, e47. 10.1093/nar/gkv007. [PubMed: 25605792]
102. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, and Durbin R; 1000 Genome Project Data Processing Subgroup (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079. 10.1093/bioinformatics/btp352. [PubMed: 19505943]
103. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, and Liu XS (2008). Model-based analysis of ChIP-seq (MACS). *Genome Biol.* 9, R137. 10.1186/gb-2008-9-9-r137. [PubMed: 18798982]
104. Karolchik D, Hinrichs AS, Furey TS, Roskin KM, Sugnet CW, Haussler D, and Kent WJ (2004). The UCSC Table Browser data retrieval tool. *Nucleic Acids Res.* 32, D493–D496. 10.1093/nar/gkh103. [PubMed: 14681465]
105. Casper J, Zweig AS, Villarreal C, Tyner C, Speir ML, Rosenbloom KR, Raney BJ, Lee CM, Lee BT, Karolchik D, et al. (2018). The UCSC Genome Browser database: 2018 update. *Nucleic Acids Res.* 46, D762–D769. 10.1093/nar/gkx1020. [PubMed: 29106570]
106. Yang JH, Shao P, Zhou H, Chen YQ, and Qu LH (2010). deepBase: a database for deeply annotating and mining deep sequencing data. *Nucleic Acids Res.* 38, D123–D130. 10.1093/nar/gkp943. [PubMed: 19966272]
107. Zheng LL, Li JH, Wu J, Sun WJ, Liu S, Wang ZL, Zhou H, Yang JH, and Qu LH (2016). deepBase v2.0: identification, expression, evolution and function of small RNAs, LncRNAs and circular RNAs from deep-sequencing data. *Nucleic Acids Res.* 44, D196–D202. 10.1093/nar/gkv1273. [PubMed: 26590255]
108. Volders PJ, Anckaert J, Verheggen K, Nuytens J, Martens L, Mestdagh P, and Vandesompele J. (2019). LNCipedia 5: towards a reference set of human long non-coding RNAs. *Nucleic Acids Res.* 47, D135–D139. 10.1093/nar/gky1031. [PubMed: 30371849]
109. Fishilevich S, Nudel R, Rappaport N, Hadar R, Plaschkes I, Iny Stein T, Rosen N, Kohn A, Twik M, Safran M, et al. (2017). GeneHancer: genome-wide integration of enhancers and target genes in GeneCards. *Database* 2017, bax028. 10.1093/database/bax028.

110. Ramı ez F, Dundar F, Diehl S, Gr€uning BA, and Manke T. (2014). deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res.* 42, W187–W191. 10.1093/nar/gku365. [PubMed: 24799436]
111. Ramı ez F, Ryan DP, Gruning B, Bhardwaj V, Kilpert F, Richter €AS, Heyne S, Dundar F, and Manke T. (2016). deepTools2: a next€generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* 44, W160–W165. 10.1093/nar/gkw257. [PubMed: 27079975]
112. Robinson MD, McCarthy DJ, and Smyth GK (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140. 10.1093/bioinformatics/btp616. [PubMed: 19910308]
113. Hochberg Y, and Benjamini Y. (1990). More powerful procedures for multiple significance testing. *Stat. Med* 9, 811–818. [PubMed: 2218183]
114. Quinlan AR, and Hall IM (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842. 10.1093/bioinformatics/btq033. [PubMed: 20110278]
115. van de Werken HJG, de Vree PJP, Splinter E, Holwerda SJB, Klous P, de Wit E, and de Laat W. (2012). 4C technology: protocols and data analysis. *Methods Enzymol.* 513, 89–112. 10.1016/B978-0-12-391938-0.00004-5. [PubMed: 22929766]
116. Li H, and Durbin R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760. 10.1093/bioinformatics/btp324. [PubMed: 19451168]
117. Wingett S, Ewels P, Furlan-Magaril M, Nagano T, Schoenfelder S, Fraser P, and Andrews S. (2015). HiCUP: pipeline for mapping and processing Hi-C data. *F1000Res.* 4, 1310. 10.12688/f1000research.7334.1. [PubMed: 26835000]
118. Durand NC, Shamim MS, Machol I, Rao SSP, Huntley MH, Lander ES, and Aiden EL (2016). Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst.* 3, 95–98. 10.1016/j.cels.2016.07.002. [PubMed: 27467249]
119. Hottentot QP, van Min M, Splinter E, and White SJ (2017). Targeted locus amplification and next-generation sequencing. *Methods Mol. Biol* 1492, 185–196. 10.1007/978-1-4939-6442-0_13.
120. Li Y, Buijs-Gladdines JGCAM, Canté -Barrett K, Stubbs AP, Vroegindewij EM, Smits WK, van Marion R, Dinjens WNM, Horstmann M, Kuiper RP, et al. (2016). IL-7 receptor mutations and steroid resistance in pediatric T cell acute lymphoblastic leukemia: a genome sequencing study. *PLoS Med.* 13, e1002200. 10.1371/journal.pmed.1002200.
121. Brady SW, Roberts KG, Gu Z, Shi L, Pounds S, Pei D, Cheng C, Dai Y, Devidas M, Qu C, et al. (2022). The genomic landscape of pediatric acute lymphoblastic leukemia. *Nat. Genet* 54, 1376–1389. 10.1038/s41588-022-01159-z. [PubMed: 36050548]

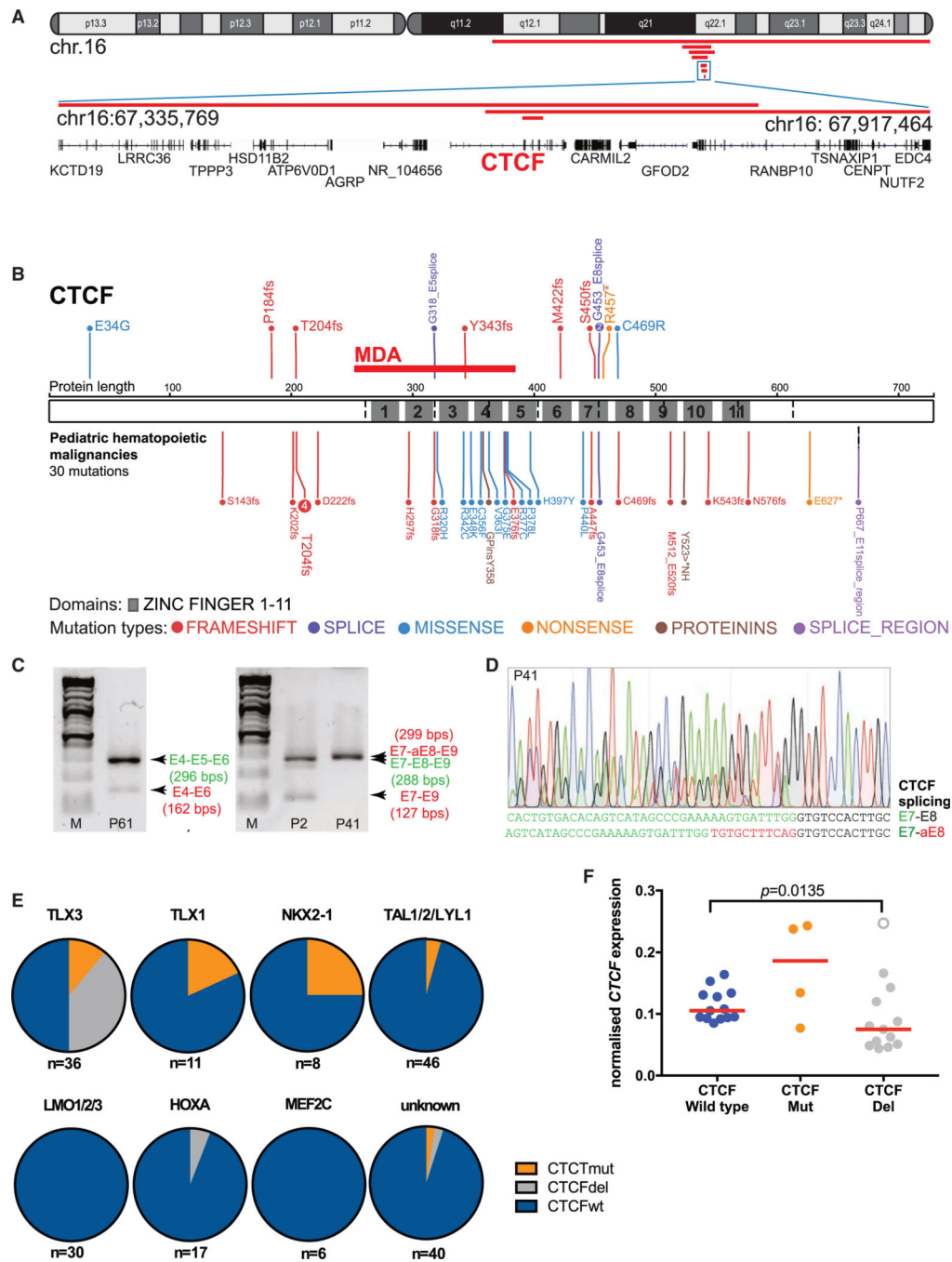


Figure 1. *CTCF* aberrations are abundant in pediatric patients with *TLX3*-rearranged T-ALL
 (A) Array comparative genomic hybridization (array-CGH) results of 7 pediatric patients with T-ALL with loss of heterozygosity of chromosomal arm 16q (GRCh37). Heterozygous deletions are displayed (red bars). The minimal deleted area (MDA; see inset) affects the *CTCF* locus including all 11 zinc fingers.
 (B) Positions of heterozygous point mutations in or near splice acceptor sites (in dark blue) and missense (in light blue), nonsense (in orange), and insertion mutations (in red) in the coding regions of *CTCF* transcript ENST00000264010. Exon boundaries are indicated by

vertical dashed lines, and the 11 zinc finger domains are displayed by numbered gray blocks, whereas the position of the MDA (see A) is shown by a red bar. Bottom: positions of 30 mutations as identified in children with hematopoietic malignancies from the St. Jude Research Hospital ProteinPaint database.⁶⁸

(C) Reverse transcription PCR results from 3 patients with T-ALL with alternative exon 4 to 6 (P61) or exon 7 to 9 (P2) splicing due to (–2) splice acceptor site mutations upstream of exons 5 or 8, respectively. P41 demonstrates alternative exon 7 to 8 splicing due to a C to G transversion mutation at the –12 position upstream of exon 8 that creates a new splice acceptor site. PCR product sizes in base pairs (bps) have been indicated. The 1 kb plus ladder (New England Biolabs) is displayed, with band sizes starting from 100, 200, 300, and higher bps.

(D) The alternative exon 7 (green sequence) to exon 8 (black sequence) splicing (E7-aE8) is displayed for P41. Bases from the intron that are misspliced are shown in red.

(E) Relative distribution of *CTCF* deletions (percentage, marked in gray), mutations (marked in orange), or wild type (marked in blue) as displayed for patients with T-ALL harboring specific oncogene rearrangements.

(F) Relative *CTCF* expression levels normalized to housekeeping gene *GAPDH* as determined by qRT-PCR in diagnostic leukemia cells from pediatric patients with *TLX3*-rearranged T-ALL who are *CTCF* wild type (n = 13, blue circles) or who harbor mutations (n = 4, orange circles) or deletions (n = 13, gray circles). Each circle represents two technical replicates. Median expression values are displayed by red bars. qRT-PCR results for the patient who has the smallest *CTCF* deletion affecting exons 3 to 6 is indicated by an open gray circle; this deletion falls outside the exons used for qRT-PCR analysis. *CTCF*-deleted patients express significantly lower *CTCF* levels than *CTCF* wild-type patients (Mann-Whitney U test), with exclusion of data from the patient with the smallest *CTCF* deletion.

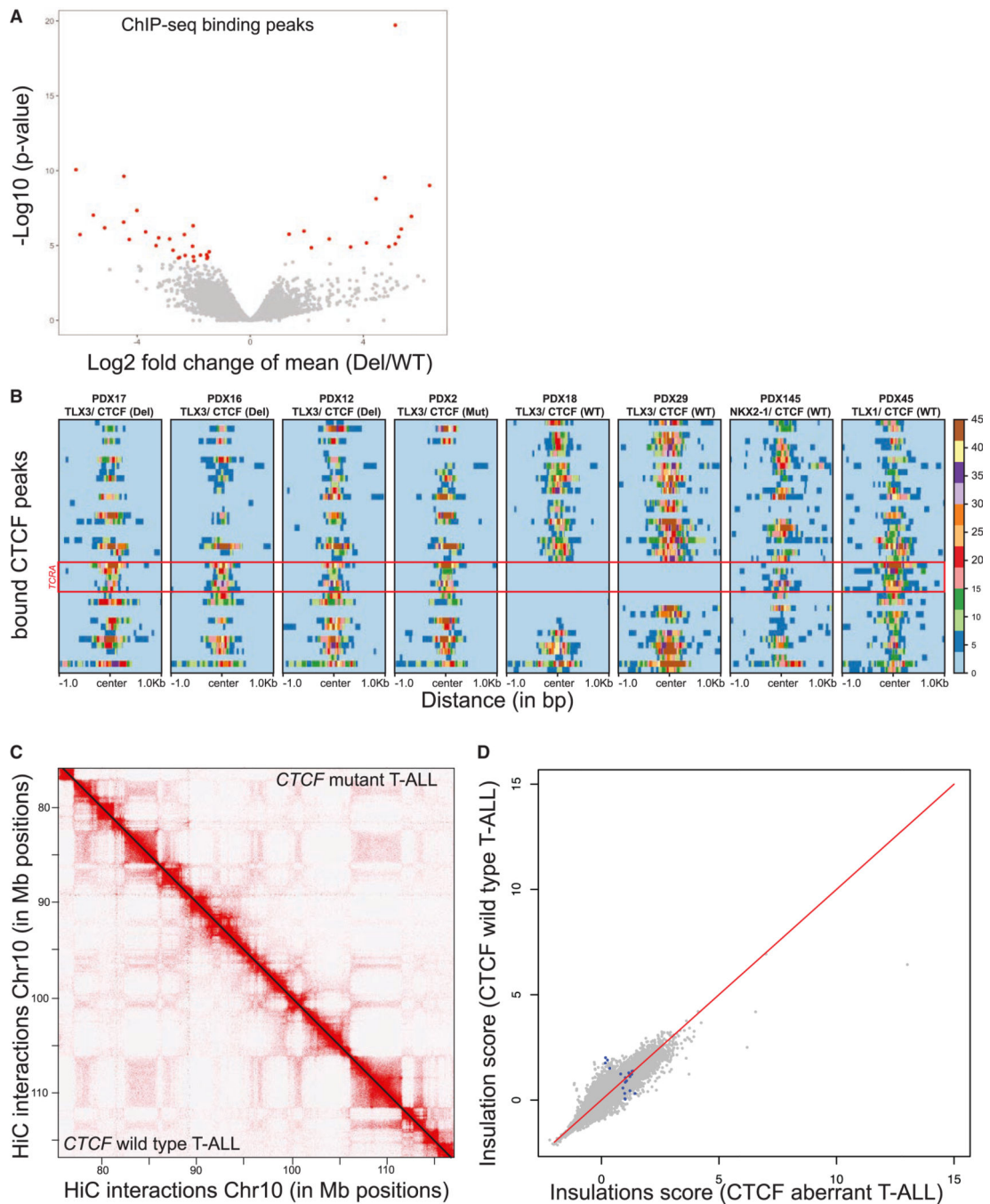


Figure 2. CTCF aberrations in T-ALL do not affect global DNA methylation, gene expression, CTCF chromatin binding, or TAD formation patterns

(A) Volcano plot of differentially CTCF-bound DNA sites for patients as described in (B). Significant binding peaks (significance levels < 0.05 , false discovery rate [FDR] < 0.1 , and fold change > 1) are indicated in red.

(B) Visualization of the 41 differentially bound CTCF peaks for patients as described in (C) and two control patients with T-ALL with *TLX1* or *NKX2-1* oncogene rearrangements. *TCRA* locus binding peaks are marked by the red square. Different peak intensity levels are indicated by colors as shown.

(C) The collapsed HiC interaction profile revealing TADs within a representative chromosomal region (chr10:75–117 Mb region [GRCh37]) for the 2 samples from patients with *CTCF* wild-type versus 4 samples from patients with *CTCF*-inactivated T-ALL (as also used for the data in B of Figure S2).

(D) TAD insulation scores for each 50 Kb bin for a patient with *CTCF* wild-type versus a patient with *CTCF*-aberrant, *TLX3*-rearranged T-ALL as calculated using HiCExplorer. Blue dots refer to the bins from the chr14:98,000,000–99,000,000 region (GRCh37/hg19) that encompass the *BCL11B* enhancer region. Patient characteristics as used for this study have been summarized in Table S1.

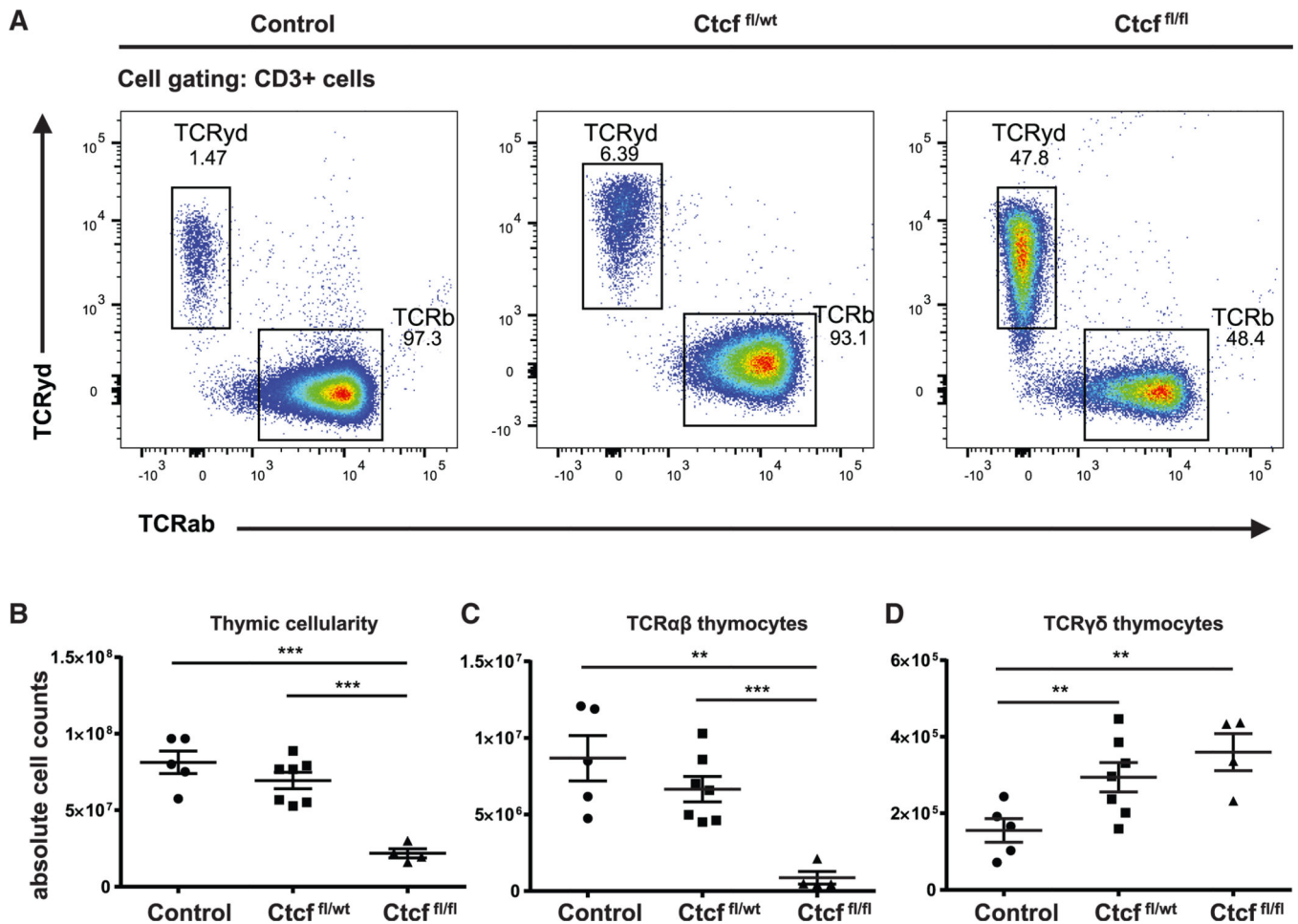


Figure 3. *Ctcf* loss impairs αβ T cell, but not γδ T cell, development

(A) TCRαβ and TCRγδ thymocyte development in *Lck-cre:Ctcf^{fl/fl}* or *Lck-cre:Ctcf^{WT/fl}* mice compared with *Ctcf^{WT/fl}* or *Ctcf^{fl/fl}* control littermates at 9 weeks of age. Relative distribution (percentage) of CD3⁺ TCRαβ⁺ or TCRγδ⁺ thymocytes are displayed for representative examples from *Lck-cre:Ctcf^{fl/fl}* (n = 4), *Lck-cre:Ctcf^{WT/fl}* (n = 7), or *Ctcf^{WT/fl}* or *Ctcf^{fl/fl}* control mice (n = 5) are shown.

(B–D) Total thymic cellularity (B), total numbers of TCRαβ⁺ thymocytes (C), and total numbers of TCRγδ⁺ thymocytes (D) are displayed. Significance levels are determined by using an unpaired t test: *p < 0.05, **p < 0.005, and ***p < 0.0005.

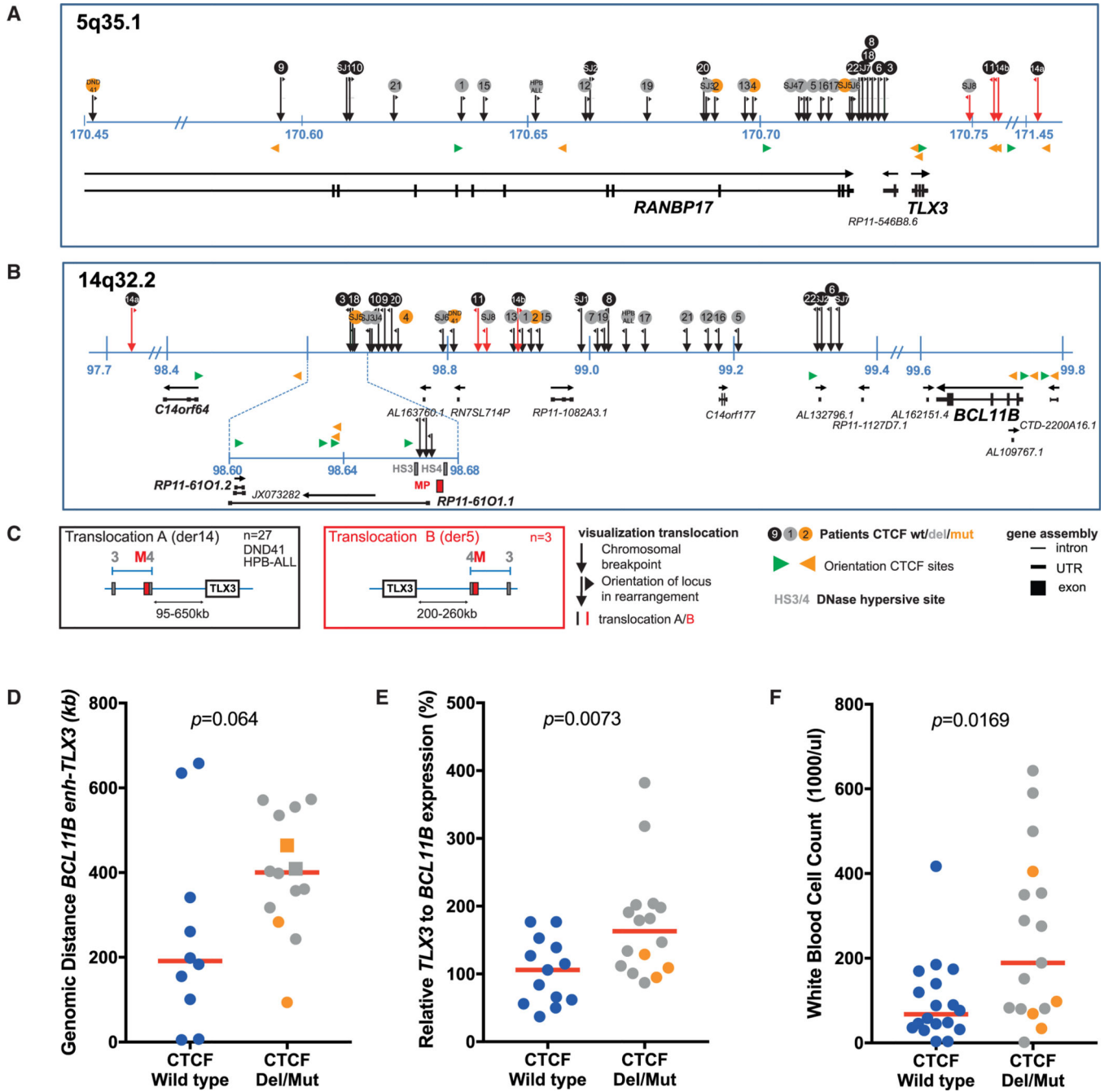


Figure 4. *CTCF* inactivation coincides with chromosomal breakpoints and increased *TLX3* levels TLA breakpoints (arrows) from derivative chromosomes 14 or 5 for 30 pediatric patients with t(5; 14) (q35; q32)-translocated T-ALL and cell lines HPB-ALL and DND41 as displayed on the 5q35.1 (A) and 14q32 (B) chromosomal regions. Positions of genomic breakpoints (indicated by arrows) in patient blasts or the HPBALL or DND41 cell lines in the intervening region in between the *BCL11B* enhancer and the *BCL11B* gene and centromeric to *TLX3* (black arrow, translocation type A) and telomeric to *TLX3* (red arrow, translocation type B) are

shown. Flags connected to arrows point to the direction of sequences that are included in the der14 (n = 27) or the der5 (n = 3, red arrows) chromosomal junctions. Patients with T-ALL or cell lines with heterozygous *CTCF* deletions (gray patient numbers) or mutations (yellow patient numbers) or who are *CTCF* wild type (black patient numbers) are displayed. Green and yellow arrow points mark the relative position and orientation of CTCF-binding sites that are bound by CTCF using chromatin immunoprecipitation sequencing (displayed in Figure 5). DNase hypersensitivity sites (HS3 and HS4) and major peak (MP) have been indicated.

(C) Summary of *TLX3* translocation types as found for the 30 pediatric patients with T-ALL as displayed in (A) and (B).

(D) Genomic distances between the *TLX3* oncogene and the *BCL11B* enhancer for 22 pediatric patients with *CTCF* wild type (blue circles), *CTCF* mutant (yellow circles), or *CTCF*-deleted (gray circles) *TLX3*-rearranged T-ALL (St. Jude patients not included). Genomic distances for HPB-ALL and DND41 are indicated by a gray or a yellow square, respectively.

(E) Relative *TLX3* expression levels compared with *BCL11B* expression levels for patients as described in (D). Each circle represents two technical replicates.

(F) White blood cell counts at disease presentation for patients with T-ALL as described in (D). p values have been determined using Kruskal-Wallis significance testing.

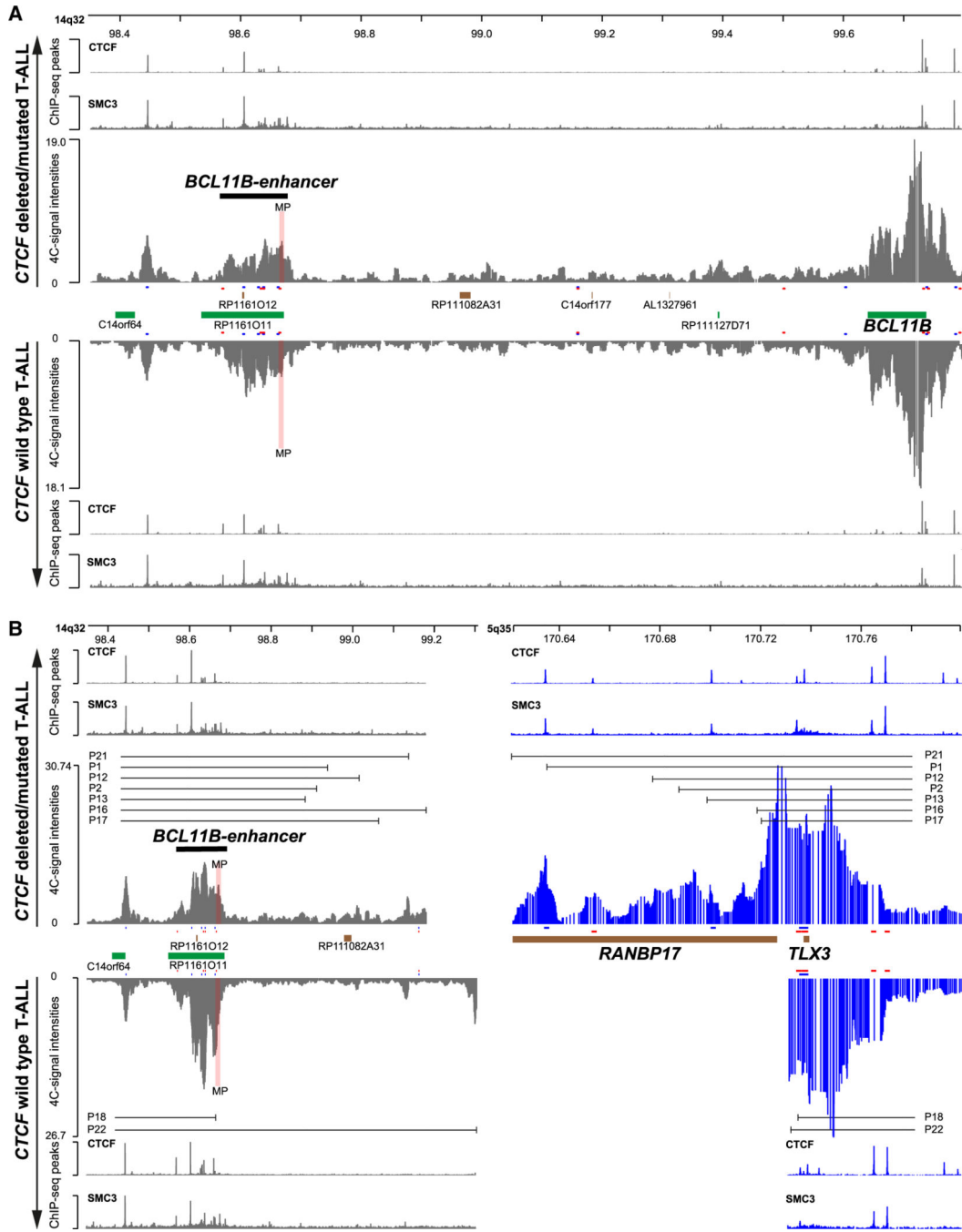


Figure 5. *BCL11B* enhancer to *TLX3* loops by chromatin conformation capture sequencing
 (A) The average chromatin conformation capture (4C) interaction plots of the *BCL11B* promoter to local or distant *BCL11B* enhancer sequences (loop interaction) is displayed for the wild-type *BCL11B* locus from 4C data of 7 pediatric patients with *TLX3*-translocated (t(5; 14) (q35; q32)) T-ALL with heterozygous *CTCF*-inactivating events and mirrored against the average interaction plot for two patients with t(5; 14)-translocated T-ALL that are *CTCF* wild type. The 4C viewpoint was positioned in the *BCL11B* promoter. Averaged and stacked *CTCF* and cohesin (*SMC3*) chromatin binding peaks

by chromatin immunoprecipitation (ChIP)-seq for indicated CTCF-deleted/-mutated and wild-type patients are displayed above or below the corresponding 4C tracks, respectively. The *BCL11B* enhancer region (4C interaction region) has been indicated by a black line above the tracks, and the GRCh37/hg19 positions of MP (red box), lncRNAs, and the *BCL11B* (legend continued on next page) gene have been indicated. lncRNA and *BCL11B* reading frames that are positioned in telomeric (brown) or centromeric (green) orientations have been indicated. CTCF-binding sites in telomeric (red bar) or centromeric (blue bar) orientations are displayed in between 4C tracks.

(B) Mirrored average 4C interaction plots for the der14 chromosome of the same patients with *TLX3*-translocated (t(5; 14) (q35; q32)) T-ALL as used in (A) are displayed. The 4C viewpoint was positioned in the *TLX3* promoter. 4C interaction tracks that cover 5q35 genomic sequences are indicated in blue, whereas interaction tracks that cover 14q32 genomic sequences are displayed in gray. Patient numbers and their relative 5q35 or 14q32 sequences and breakpoint involved in the der14 chromosomal junctions are displayed for each patient. Averaged and stacked CTCF and cohesin (SMC3) binding peaks for parental HPBALL cells and the relative positions of MP (red box), lncRNAs, genes, and CTCF-binding sites are indicated as described in (A).

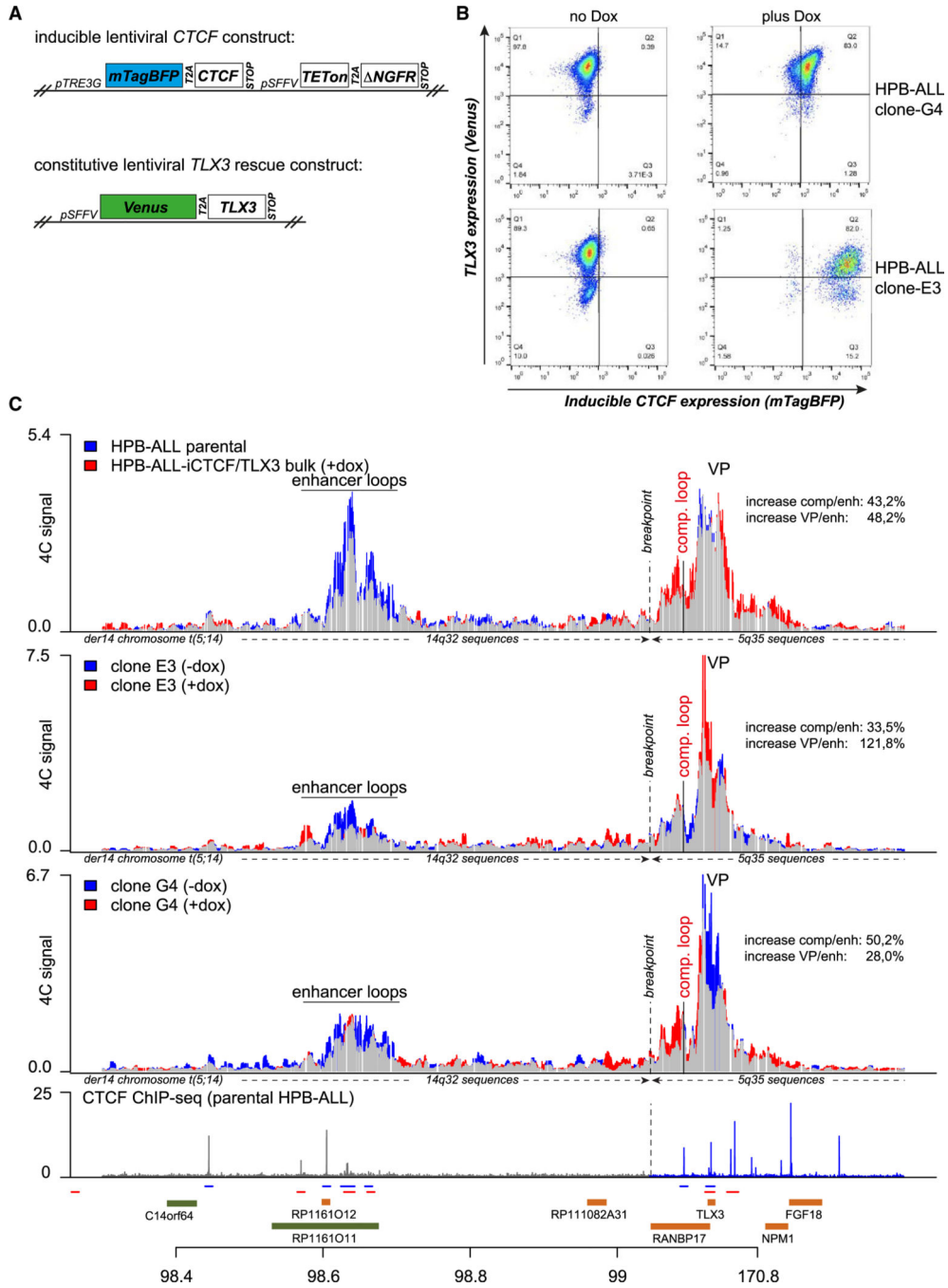


Figure 6. Increased CTCF levels in HPB-ALL enhance competitive loop formation
 (A) Schematic overview of the lentiviral doxycycline-inducible blue fluorescent protein (BFP)-T2A-CTCF construct and the constitutively active Venusgreen-T2ATLX3 rescue construct are displayed that have been used to produce the HPB-ALL derivative HPB-ALL-iCTCF/TLX3 bulk line.
 (B) Flow cytometry analysis of single-cell clones G4 and E3 that have been derived from bulk HPB-ALL-iCTCF/TLX3 cells and that constitutively express *TLX3* (in addition to endogenous *TLX3* oncogene expression) as visualized by Venusgreen intensities but

moderate (G4) or high (E3) CTCF expression levels (as determined by BFP expression) following doxycycline exposure.

(C) Display of representative examples of normalized 4C interaction plots for the der14 chromosomal region of HPB-ALL-iCTCF/TLX3 bulk cells or the G4 or E3 single-cell clones that were cultured in the absence (-dox, blue tracing) or presence (+dox, red tracing) of doxycycline for 2 days. Gray tracing represents equal levels of interactions among -dox and +dox conditions. The 4C viewpoint was positioned in the *TLX3* promoter as indicated. The relative positions of the *BCL11B* enhancer, the chromosomal breakpoint, and the regions that were used to calculate changes in chromosomal looping for the enhancer loop (enh) area (Hg37:chr14:98,602,411–98,675,204), the competitive loop area (comp; Hg37:chr5:170,716,369–170,756,369), and the proximal viewpoint (VP) area (Hg37:chr5:170,653,532–170,710,663) have been indicated. Relative gain in competitive or proximal VP loops related to *BCL11B* enhancer loops has been indicated for the +dox (induced CTCF) versus the dox conditions. Stacked CTCF-binding peaks for parental HPB-ALL cells by CHIP-seq and the relative Hg37 positions of lncRNAs, genes, and CTCF-binding sites have been indicated.

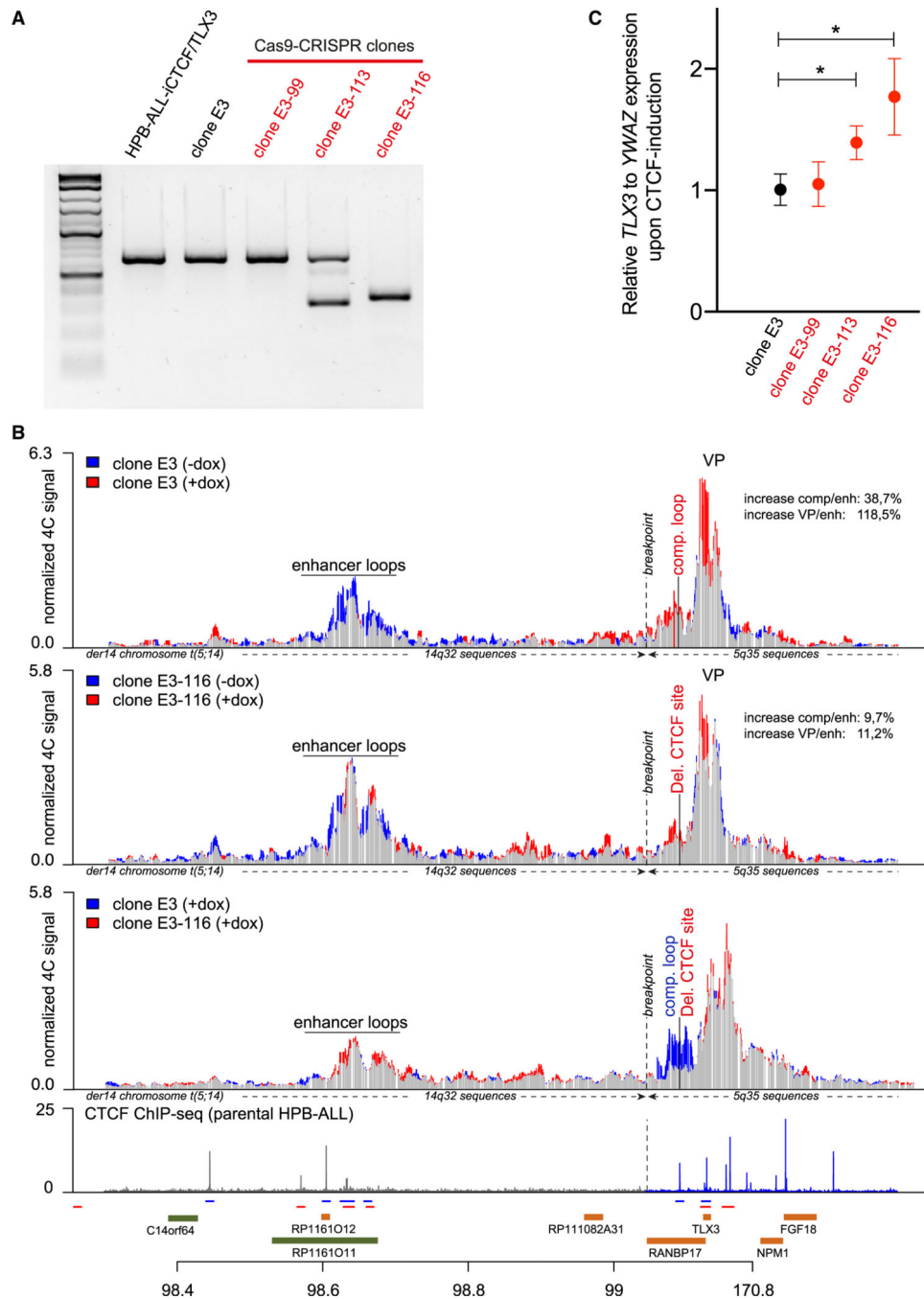


Figure 7. Competitive loops are lost upon removal of intervening CTCF-binding sites and enhance *TLX3* expression levels

(A) Genotype PCR results for CRISPR-Cas9 genome edited HPB-ALL-iCTCF/TLX3 clone E3 derived single-cell clones (#E3-99, #E3-113, and #E3-116) to remove the competitive intervening CTCF-binding site (Hg37:chr5:170,700,483–170,700,501) are displayed. HPB-ALL-iCTCF/TLX3 bulk cells and single-cell clone E3 cells serve as controls.

(B) Display of representative examples of normalized 4C interaction plots for the der14 chromosomal region of HPB-ALL-iCTCF/TLX3 clone E3 cells or derivative #E3-116 cells that were cultured in the absence (-dox) or presence (+dox) of doxycycline for 2 days.

The 4C VP was positioned in the *TLX3* promoter. Tracing colors, the relative positions of the *BCL11B* enhancer, the chromosomal breakpoint, the location of the competitive CTCF-binding site, and the 4C VP as well as the regions to calculate the relative gain in comp or proximal VP loops relative to *BCL11B* enhs are as described in the legend of Figure 6C. The bottom 4C interaction plot compares the 4C interactions from the *TLX3* VP for the E3 and E3–116 clones following induction of CTCF (+dox). Competitive and proximal interactions for clone E3 (blue tracing) are lost in clone E3–116 (red tracing) upon removal of the single intervening CTCF-binding site, which results in enhanced *BCL11B* enhancer interactions. Stacked CTCF chromatin binding peaks for HPB-ALL by ChIP-seq and the relative Hg37 positions of lncRNAs, genes, and CTCF-binding sites have been indicated.

(C)Relative change in endogenous *TLX3* oncogene expression levels as measured by qRT-PCR for doxycycline-induced versus non-induced E3 cells or derivative single-cell clones as described in (A). Significance levels as indicated are calculated by an unpaired t test. Circles and brackets represent the mean with its standard deviation of three technical replicates, respectively.

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
anti-human CTCF; clone D31H2	Cell Signaling Technology	Cat# 3418; RRID:AB_2086791
anti-human SMC3	Abcam	Cat# ab9263; RRID:AB_307122
Normal rabbit IgG	CellSignaling Technology	Cat# 2729; RRID:AB_1031062
anti-mouse CD8a VioGreen; clone 53–6.7	Miltenyi Biotec	Cat# 130–102-805; RRID:AB_2659890
anti-mouse CD4 Alexa Fluor 488; clone GK1.5	BioLegend	Cat# 100425; RRID:AB_493520
anti-mouse TCRy δ PE; clone GL3	BD Biosciences	Cat# 553178; RRID:AB_394689
anti-mouse TCRb PE-Cy5; clone H57–597	BD Biosciences	Cat# 561080; RRID:AB_10584335
anti-mouse CD3e PE-Cy7; clone 145–2C11	BD Biosciences	Cat# 552774; RRID:AB_394460
anti-mouse CD117 APC; clone 2B8	BD Biosciences	Cat# 561074; RRID:AB_10563203
anti-mouse CD25 APC-Cy7; clone PC61	BD Biosciences	Cat# 561038; RRID:AB_2034002
anti-mouse CD45R/B220; clone RA3–6B2	BioLegend	Cat# 103211; RRID:AB_312996
anti-human CD271 (LNGFR) PE; clone ME20.4–1.H4	Miltenyi Biotec	Cat# 130–113-421, RRID:AB_2733795
Chemicals, peptides, and recombinant proteins		
Ampure XP Reagent	Beckman Coulter	Cat# A63880
Formaldehyde solution	Merck	Cat# 252549; CAS: 50–00-0
NlaIII	New England Biolabs	Cat# R0125S
T4 DNA ligase	New England Biolabs	Cat# M0202S
NspI	New England Biolabs	Cat# R0602S
TRIzol Reagent	Thermo Fisher Technology	Cat# 15596026
Acid-Phenol:Chloroform pH4.5	Thermo Fisher Technology	Cat# AM9722
Glycogen	Roche	Cat# 10901393001
M-MLV, RNase H minus, Point mutant	Promega	Cat# M3681
RNasin ribonuclease inhibitor	Promega	Cat# 2115
DAPI	BioLegend	Cat# 422801
DpnII restriction enzyme	New England Biolabs	Cat# R0543S
Csp6I restriction enzyme	Thermo Fisher Scientific	Cat# ER0211
Nucleomag beads	Bioke	Cat# 744100.34
Expand long template PCR system	Roche	Cat# 11681834001
Doxycycline hydrochloride	Merck	Cat# D3447; CAS: 10592–13-9
X-tremeGENE HP DNA Transfection Reagent	Roche	Cat# 6366546001
IVSS VIVASPIN 20 centrifugation concentration columns	Sartorius	Cat# Z614653–48EA
Thioguanine	Merck	Cat# A4882; CAS: 154–42-7
TrueCut Cas9 Protein v2	Thermo Fisher Scientific	Cat# A36498
AmpliTaq Gold	Thermo Fisher Scientific	Cat# N8080245
Critical commercial assays		

REAGENT or RESOURCE	SOURCE	IDENTIFIER
SimpleChIP®Enzymatic Chromatin IP Kit	Cell Signaling Technology	Cat# 9003
QIAquick PCR Purification Kit	QIAGEN	Cat# 28104
Qubit dsDNA HS Assay Kit	Thermo Fisher Scientific	Cat# Q32851
NEXTflex™ Rapid DNA Sequencing Kit	PerkinElmer	Cat# NOVA-5144-03
Targeted Locus Amplification Analysis kit	Cergentis BV	N/A
DyNAmo HS SYBR Green qPCR kit	Thermo Fisher Scientific	Cat# F410XL
Gateway Vector Conversion System	Thermo Fisher Scientific	Cat# 11828029
EchoLUTION Cell Culture DNA Kit	BioEcho	Cat# 010-006-250
Deposited data		
GEO superseries containg complete set for: Elevated enhancer- oncogene contacts and higher oncogene expression levels by recurrent CTCF inactivating mutations in acute T cell leukemia	RRID:SCR_005012	GEO: GSE182317
HiC data	RRID:SCR_005012	GEO: GSE182316
4C data	RRID:SCR_005012	GEO: GSE182315
Bead chip Methylation data	RRID:SCR_005012	GEO: GSE182313
Array CGH data	RRID:SCR_005012	GEO: GSE182312
ChIPseq data	RRID:SCR_005012	GEO: GSE181759
Sequence Read Archive (breakpoint data by TLA)	RRID:SCR_001370	PRJNA945826
European Nucleotide Archive (breakpoint data by Complete Genomics WGS)	RRID:SCR_006515	ERS934791
European Genome-phenome Archive (breakpoint data by Illumina WGS)	RRID:SCR_005012	EGAS00001005250
Experimental models: Cell lines		
HPB-ALL	DSMZ	RRID:CVCL_1820
Experimental models: Organisms/strains		
Lck-Cre mice (B6.Cg-Tg(Lck-cre)1Cwi N9)	Taconic	4197-M, Lee et al. ⁹⁵
CTCF fl/fl mice	Erasmus Medical Center Rotterdam, Rudi Hendriks lab	Heath et al. ⁷
Oligonucleotides		
Genome editing Genotyping Forward Primer ATCCAGCACATCTCTCTCA	IDTDNA	N/A
Genome editing Genotyping Reverse Primer GAACCAGATGGAAATAAAAATATC	IDTDNA	N/A
Recombinant DNA		
Lego-iC2	Addgene	27345
pMD2.G	Addgene	12259
pMDLg/pRRE	Addgene	12251
pRSV-Rev	Addgene	12253

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and algorithms		
MIPgen software v1.2.1	Github (RRID:SCR_002630)	RRID:SCR_003325
Sequence Pilot; SeqNext	JSI medical systems	https://www.jsi-medisys.de/
SAMTOOLS	Github (RRID:SCR_002630)	RRID:SCR_002105
FlowJo	Becton Dickinson	RRID:SCR_008520
4C-seq pipeline	Krijger et al. ⁸⁵	https://github.com/deLaatLab/pipe4C
Agilent CytoGenomics software	Agilent	RRID:SCR_010917
Affy	Bioconductor (RRID:SCR_006442)	RRID:SCR_012835
Limma	Bioconductor (RRID:SCR_006442)	RRID:SCR_010943
Minfi	Bioconductor (RRID:SCR_006442)	RRID:SCR_012830
MACS	Github (RRID:SCR_002630)	RRID:SCR_013291
UCSC genome browser	UCSC	RRID:SCR_005780
Deeptools	Github (RRID:SCR_002630)	RRID:SCR_016366
edgeR	Bioconductor (RRID:SCR_006442)	RRID:SCR_012802
DiffBind	Bioconductor (RRID:SCR_006442)	RRID:SCR_012918
BEDtools	Github (RRID:SCR_002630)	RRID:SCR_006646
Other		
MiSeqDx system	Illumina (RRID:SCR_010233)	Illumina miniseq; RRID:SCR_016380
NextSeq 500	Illumina (RRID:SCR_010233)	Illumina NextSeq 500; RRID:SCR_014983
Bioruptor Pico sonication device	Diagenode	Cat# B01060010
CFX384 touch thermal cycler	Bio-Rad	N/A
LSRII flow cytometer	Becton Dickinson	N/A
SH800S Cell Sorter	Sony	N/A
Neon Transfection System	Thermo Fisher Scientific	MPK5000