TAL1 activation in T-cell acute lymphoblastic leukemia: a novel oncogenic 3' neo-enhancer

Charlotte Smith,[1]* Ashish Goyal,[2]* Dieter Weichenhan,[2] Eric Allemand,[3] Anand Mayakonda,[2] Umut Toprak,[4,5] Anna Riedel,[2] Estelle Balducci,[1] Manisha Manojkumar,[2] Anastasija Pejkovska,[2] Oliver Mücke,[2] Etienne Sollier,[2] Ali Bakr,[2] Kersten Breuer,[2] Pavlo Lutsik,[2] Olivier Hermine,[3,6] Salvatore Spicuglia,[7] Vahid Asnafi,[1] Christoph Plass[2,8] and Aurore Touzart[1,2]

[1]Université de Paris Cité, Institut Necker Enfants-Malades (INEM), Institut National de la Santé et de la Recherche Médicale (Inserm) U1151, and Laboratory of Onco-Hematology, Assistance Publique-Hôpitaux de Paris, Hôpital Necker Enfants-Malades, Paris, France; [2]Division of Cancer Epigenomics, German Cancer Research Center (DKFZ), Heidelberg, Germany; [3]Université de Paris Cité, Institut Imagine, Inserm U1163, Paris, France; [4]Hopp Children's Cancer Center Heidelberg (KiTZ), Heidelberg, Germany; [5]Division of Neuroblastoma Genomics, German Cancer Research Center (DKFZ), Heidelberg, Germany; [6]Department of Hematology, Hôpital Necker Enfants Malades, AP-HP, Faculté de Médecine Paris Descartes, Paris, France; [7]Aix-Marseille University, Inserm, Theories and Approaches of Genomic Complexity (TAGC), Equipe Labellisée Ligue, UMR1090, Marseille, France and [8]German Cancer Research Consortium (DKTK), Heidelberg, Germany

CS and AG contributed equally as first authors.

Correspondence: A. Touzart
aurore.touzart@aphp.fr

C. Plass
c.plass@dkfz.de

Received: June 16, 2022.
Accepted: December 28, 2022.
Early view: January 12, 2023.

https://doi.org/10.3324/haematol.2022.281583

Abstract

T-cell acute lymphocytic leukemia protein 1 (TAL1) is one of the most frequently deregulated oncogenes in T-cell acute lymphoblastic leukemia (T-ALL). Its deregulation can occur through diverse *cis*-alterations, including SIL-TAL1 microdeletions, translocations with T-cell Receptor loci, and more recently described upstream intergenic non-coding mutations. These mutations consist of recurrent focal microinsertions that create an oncogenic neo-enhancer accompanied by activating epigenetic marks. This observation laid the groundwork for an innovative paradigm concerning the activation of proto-oncogenes via genomic alterations of non-coding intergenic regions. However, for the majority of T-ALL expressing TAL1 (TAL1$^+$), the deregulation mechanism remains 'unresolved'. We took advantage of H3K27ac and H3K4me3 chromatin immunoprecipitation sequencing data of eight cases of T-ALL, including five TAL1$^+$ cases. We identified a putative novel oncogenic neo-enhancer downstream of TAL1 in an unresolved monoallelic TAL1$^+$ case. A rare but recurrent somatic heterozygous microinsertion within this region creates a *de novo* binding site for MYB transcription factor. Here we demonstrate that this mutation leads to increased enhancer activity, gain of active epigenetic marks, and *TAL1* activation via recruitment of MYB. These results highlight the diversity of non-coding mutations that can drive oncogene activation.

Introduction

T-cell acute lymphoblastic leukemia (T-ALL) is a rare, aggressive hematologic malignancy that accounts for 15% of pediatric and 25% of adult ALL.[1,2] The use of intensified chemotherapy regimes has led to recent advances in overall survival.[3] However, 30-35% of patients still relapse and face a dismal prognosis with an overall survival of less than 20% at 5 years.[4,5] T-ALL is caused by the clonal expansion of immature T-cell precursors that are blocked in their thymic differentiation. Several genetic alterations engendering the ectopic expression of key oncogenic transcription factors have been reported, including the HOX gene family members (HOXA,[6,7] TLX1,[8] TLX3,[9,10] NKX1-1 and NXK1-2,[11] basic helix-loop-helix family members (TAL1/2,[12-16] LYL1,[17] BHLHB1,[18] the LMO family members (LMO1, LMO2[19,20]), and MYB.[21] Of these transcription factors, the *TAL1* oncogene is one of the most frequently deregulated in T-ALL. Indeed, increased *TAL1* transcripts are found in up to 60% of patients.[1] Under physiological conditions, TAL1 is an essential transcription factor (TF) for the development of the vascular system and for primary and definitive hematopoiesis. During definitive hematopoiesis, TAL1 is required for erythroid differentiation, yet it is epigenetically repressed during human thymopoiesis.[14,22,23] Oncogenic events leading to the ectopic expression of TAL1 in the T-cell lineage are considered strong drivers of T-cell leukemogenesis. Such deregulation mechanisms occur in *cis*

and involve the 'hijacking' of enhancer elements leading to oncogene activation. The SIL-TAL1 microdeletion is the predominant deregulation mechanism found in 25% of T-ALL. The interstitial deletion places the *TAL1* gene under the regulatory control of the adjacent *STIL* gene promoter.[13,24] In a minority of cases (<5%), the *TAL1* gene is translocated into either the T-cell receptor β (TRB) locus or, more often, the T-cell receptor δ (TRD) locus and under the control of strong T-cell receptor (TCR) *cis*-regulatory elements.[24-26]

We and others have identified a third deregulation mechanism consisting of microinsertions 7kb upstream of the *TAL1* gene in a non-coding region. These microinsertions create a super-enhancer identifiable by its enrichment in transcriptionally activating epigenetic marks[14,15] and *de novo* MYB TF binding sites.[15] The recruitment of MYB to the mutated sequence and subsequent super-enhancer formation leads to ectopic *TAL1* expression.

Lastly, a rare, recurrent intronic point mutation responsible for monoallelic *TAL1* expression was recently discovered in pediatric T-ALL. The mutation creates a *de novo* YY1 TF binding site, is associated with active epigenetic marks, and demonstrates enhancer activity.[27] Nonetheless, these diverse deregulation mechanisms do not account for all TAL1 positive (TAL1$^+$) patients, suggesting 'unresolved' TAL1 deregulation mechanisms.[28-30] We, therefore, hypothesized that the identification of aberrantly formed intergenic histone marks in the vicinity of the TAL1 locus could uncover new TAL1 abnormalities in T-ALL. We discovered a novel recurrent microinsertion downstream of the *TAL1* gene leading to the formation of an MYB-dependent neo-enhancer responsible for *TAL1* activation.

Methods

Antibody-guided chromatin tagmentation

Genome-wide targeting of histone modifications and MYB were performed by antibody-guided chromatin tagmentation (ACT-seq), according to Carter *et al.*[31] pA-Tn5ase protein was isolated from *E. coli* (C3013, New England Biolabs; Ipswich, MA, USA) and transformed with plasmid pET15bpATnp (#121137, Addgene; Watertown, MA, USA). The pA-Tn5 transposome (pA-Tn5ome) was generated by mixing pA-Tn5ase (final concentration either 1.9 μM or 3.3 μM, depending on the pA-Tn5ase preparation) and Tn5ME-A+B load adaptor mix (final concentration 3.3 μM) in complex formation buffer (CB31). The pA-Tn5ome-antibody complexes were formed by mixing 1 μL pA-Tn5ome with 0.8 μL CB and 0.8 μL antibody solution. We used antibodies against histone H3K27ac (#4729, Abcam; Berlin, Germany), H3K4me1 (#8895, Abcam; Berlin, Germany), H3K4me3 (#39915 active motif), IgG (#PP64B, Millipore; Burlington, MA, USA), H2B (#M30930, Hölzel Dianostika; Köln, Germany), and MYB (#45150, Abcam); 100,000 cells were used for the pA-Tn5ome-ab complex binding and tagmentation. Tagmented DNA was purified using MinElute kit (#28004, Qiagen; Venlo, The Netherlands) and eluted with 20 μL elution buffer (EB). Sequencing libraries were generated under real-time conditions with a LightCycler 480 (Roche Professional Diagnostics; Indianapolis, IN, USA) in 50 μL reaction mixes consisting of 20 μL tagmented DNA eluate, 25 μL NEBNext High Fidelity 2X Mix (#M0541, New England Biolabs; Ipswich, MA, USA), 0.5 μL 100x SYBRGreen, 2.5 μL primer Tn5McP1n and 2.5 μL barcode primer. Reaction conditions were 72°C, 5 minutes (min) (gap repair); 98°C, 30 seconds (sec) (initial melting); followed by cycles of: 98°C (10 sec), 63°C (10 sec), 72°C (10 sec). Cycling was stopped when fluorescence units (FU) had increased by 5. Libraries were purified with HighPrep magnetic beads (#220001, Biozym; Hessisch Oldendorf, Germany) with a bead:DNA ratio of 1.4:1 and eluted with 12 μL EB. Quantity and fragment size of the libraries were determined with a Qubit dsDNA HS assay kit (#Q32854, Invitrogen, ThermoFisher; Waltham, MA, USA) and a TapeStation 4150 with D1000 High Sensitivity Assay (#5067- 5585, Agilent; Santa-Clara, CA, USA), respectively. Six to eight differently barcoded libraries were multiplexed and sequenced on a single lane of a NextSeq™ 550 system (paired-end, 75 bp, Illumina; San Diego, CA, USA) with mid-output at the Genome and Proteome Core Facility of DKFZ. ACT-seq data were analyzed as previously described.[32]

Alpha-cas phasing methods and sgRNA sequences

To phase the engineered heterozygous 3' mutation with the SNP (Hg19: chr1: 47 684 223) in J-3'NE#1 cells, gRNA targeting the region downstream of the SNP (TAL1 SNP DS crRNA) and upstream of the mutation (TAL1 3'NE DS crRNA) were designed. Corresponding Cas9 RNP were prepared as described above and electroporated into J-3'NE cells. Forty-eight hours (h) post electroporation gDNA was isolated and amplified by polymerase chain reaction (PCR). Similarly, Jurkat 5'super-enhancer was phased using Cas9 RNP corresponding to TAL1 SNP DS crRNA and TAL1 5'SE US crRNA followed by PCR. The LMO1 activating point mutation in Jurkat cells was phased to an SNP in LMO1 exon 1 (Hg19: chr11: 8 285 124) using LMO1 SNP DS crRNA and LMO1 MUT US crRNA followed by PCR. The LMO2 activating microinsertion in MOLT4 cells was phased to an SNP in LMO2 exon 6 (Hg19: chr11: 33 881 016) by electroporating Cas9 RNP corresponding to LMO2 SNP DS crRNA and LMO2 MUT US crRNA (100,000 MOLT4 cells were electroporated using the following settings: 1350V, 10ms, 3 pulses), followed by PCR. In all cases the PCR product was cloned into a pCR4 TOPO TA vector using a TOPO TA Cloning Kit (Cat. # 450030, Thermo Scientific; Waltham, MA, USA) and

transformed into electrocompetent Stbl3 cells. Ampicillin-resistant colonies were further inoculated for plasmid miniprep. Plasmids were Sanger sequenced using the M13 reverse primer.

See Online Supplementary Table S1 for primer sequences and *Online Supplementary Table S2* for gRNA sequences. For ACT-seq oligonucleotide sequences see *Online Supplementary Table S3*.

Studies were conducted with informed consent from all patients and in accordance with the principles of the Declaration of Helsinki, and approved by local and multicenter research ethical committees.

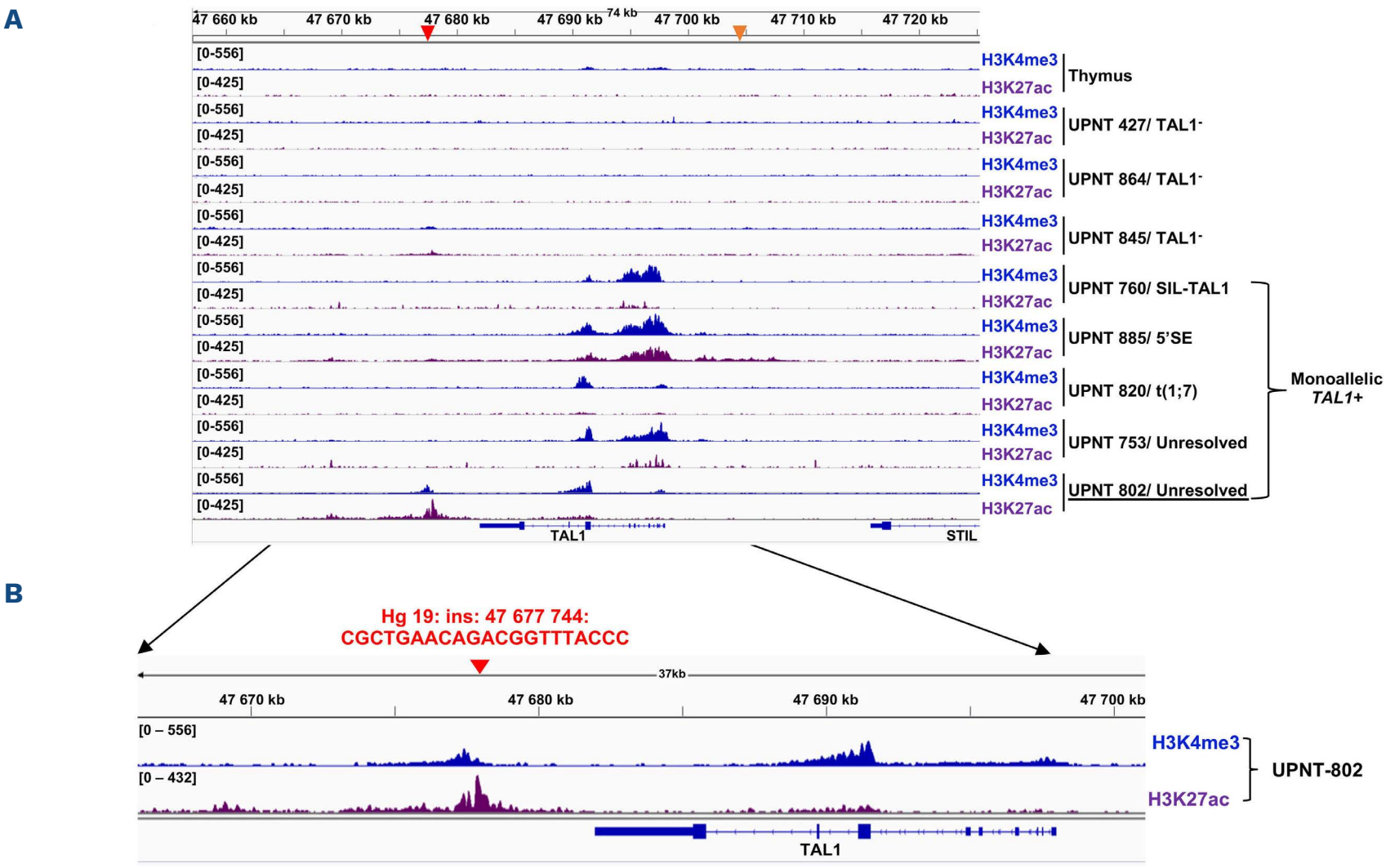
Results

Identification of a putative novel oncogenic neo-enhancer associated with an intergenic 3'*TAL1* microinsertion

To investigate TAL1 deregulation mechanisms, we took advantage of the Blueprint Consortium T-ALL chromatin immunoprecipitation sequencing (ChIP-seq) series.[33,34] We focused on H3K4me3 and H3K27ac enrichment over the *TAL1* locus in eight primary T-ALL samples and the normal thymus (Figure 1A). *TAL1* expression was analyzed using available RNA-seq data (n=6) and/or real-time quantitative (RQ) PCR (n=8). Five patient samples displayed high TAL1 expression and were considered TAL1 positive (TAL1$^+$) and three were considered TAL1 negative (TAL1$^-$) (*Online Supplementary Figure S1A and B*). Allelic expression analysis by RNA-seq and Sanger sequencing revealed that all five TAL1$^+$ samples had monoallelic TAL1 expression, indicating *cis*-deregulation mechanisms (Figure 1C-E, *Online Supplementary Figure S1C*). We screened TAL1$^+$ patients for the known recurrent mechanisms of TAL1 deregulation. Of these five TAL1$^+$ patients, one (UPNT-760) had a SIL-TAL1 microdeletion, one (UPNT-885) had the 5' super-enhancer mutation (5'SE), and one (UPNT-820) had a translocation involving *TAL1* and the *TRB* locus. Two patients (UPNT-753 and UPNT-802) had unidentified TAL1 deregulation mechanisms and were considered unresolved. ChIP-seq data for H3K27ac and H3K4me3 marks showed enrichment over the TAL1 gene body in all five monoallelically expressing TAL1$^+$ patients (Figure 1A). As expected, H3K27ac enrichment encompassing the 5'super-enhancer was observed in the UPNT-885 sample. In addition, we noticed a unique dual monoallelic enrichment over an intergenic region downstream of the *TAL1* gene in one unresolved monoallelic TAL1$^+$ patient (UPNT-802) that was absent in all other TAL1$^+$ patients, suggesting the presence of a novel regulatory element (Figure 1B, *Online Supplementary Figure S2*). We performed whole genome sequencing (WGS) on both tumoral and non-tumoral DNA (bone marrow remission sample with undetectable minimal residual disease [MRD]) from UPNT-802 and identified a heterozygous 21bp microinsertion (variant allele frequency [VAF] 36%) about 4 kb downstream of the *TAL1* gene (Hg 19 chr1: 47 677 744) within the genomic region enriched in H3K27ac and H3K4me3 marks (Figure 1B). This variant was somatic and absent in the non-tumoral sample (Figure 1F). We also confirmed the presence of the mutation by Sanger sequencing. Notably, this mutation was stable in leukemic cells expanded in a patient-derived xenograft (PDX) (Figure 1G). As a final verification, we designed mutation specific primers to amplify the microinsertion in the diagnostic sample (*Online Supplementary Figure S1D*). Collectively, these results suggest the identification of a novel oncogenic neo-enhancer.

The intergenic 3'-*TAL1* microinsertion is responsible for TAL1 deregulation in cell-line models

To study the function of the new putative neo-enhancer, we introduced the heterozygous 3´ microinsertion in Jurkat (TAL1$^+$) and Peer (TAL1$^-$) T-ALL cell lines using CRISPR-Cas9 technology. In the Jurkat cell line, the 3' microinsertion was introduced and the original 5'super-enhancer subsequently deleted (J-3'NE #1 and J-3'NE #2 derivative cell lines). Likewise, we introduced the 3' mutation into the Peer cell line (P-3'NE #1, P-3'NE #2, P-3'NE #3 derivative cell lines), and also engineered a Jurkat cell line that was deleted for the original 5' super-enhancer (J-del derivative cell line) (Figure 2A). Genotyping by Sanger sequencing confirmed successful genomic editing in the derivative cell lines with the insertion of the heterozygous microinsertion (*Online Supplementary Figure S3A and B*). As expected, a strong decrease in TAL1 expression was observed in J-del cells, which was rescued upon introduction of the 3'microinsertion in J-3'NE #1 and J-3'NE #2 derivative cell lines (Figure 2B and G). Whereas cell proliferation was affected in J-del cells, J-3'NE #1 cells displayed unaffected proliferation (similar to normal unedited Jurkat) (Figure 2C). The introduction of this mutation also led to a strong TAL1 activation in the Peer derivative cell lines (P-3'NE #1, P-3'NE #2, P-3'NE #3) (Figure 2D and G). Importantly, whereas wild-type (WT) Peer showed low and biallelic TAL1 expression, the three Peer derivative cell lines displayed monoallelic *TAL1* expression in line with *cis*-activation (Figure 2E). The mutated sequence was cloned in a PGL4.23 plasmid (containing the luciferase gene and a minimal promoter) and demonstrated significantly increased enhancer activity (P=0.004) compared to the WT sequence in the luciferase reporter assay (Figure 2F), confirming the enhancer function of this variant. Taken together, these results suggest that the novel 3´ microinsertion leads to the creation of a new regulatory element able to activate



Continued on following page.

Figure 1. Identification of a novel oncogenic regulatory element. Blueprint Consortium chromatin immunoprecipitation sequencing (ChIP-seq) tracks for H3K4me3 (in blue) and H3K27ac (in purple) over the *TAL1* locus in the normal thymus and eight primary T-cell acute lymphoblastic leukemia (T-ALL) samples. H3K4me3 and H3K27ac enrichment corresponded with *TAL1* expression. Type of *TAL1* deregulation mechanism is shown. Red arrow indicates a unique 3' H3K4me3 and H3K27ac enrichment in the T-ALL sample UPNT-802 underlined in black; orange arrow indicates the 5'super-enhancer mutation. Tracks were aligned to the Hg19 genome reference, normalized to normal thymic cell populations where TAL1 is epigenetically silenced and to input samples. Blue arrows on the *TAL1* gene indicate the direction of transcription. (B) ChIP-seq tracks for H3K4me3 (in blue) and H3K27ac (in purple) over the *TAL1* locus in the unresolved monoallelically-expressing *TAL1* patient UPNT-802. Tracks show a unique dual enrichment downstream of the *TAL1* gene. Red arrow highlights the somatic 21 bp microinsertion found by whole genome sequencing (WGS). (C) RNA-sequencing analysis confirmed high *TAL1* expression in UPNT-802 compared to $CD34^+$ thymocytes. (D) RNA sequencing reads centered on the informative SNP (rs 1010812) in the TAL1- 3' UTR confirmed that UPNT-802 had monoallelic *TAL1* expression. (E) Monoallelic expression was further validated by Sanger sequencing of this 3'UTR heterozygous SNP in genomic DNA (left) and cDNA (right). (F) WGS sequencing reads of tumoral and non-tumoral samples from patient UPNT-802. Red arrow highlights the genomic location of the heterozygous 21 bp microinsertion (Hg19: chr1: 47,677,744); the variant allele frequency (VAF) of the microinsertion was 36%. (G) Sanger sequencing chromatograms from tumoral (Diag), non-tumoral (remission sample), and patient-derived xenograft (PDX) DNA samples confirming the somatic micoinsertion.

TAL1 expression and sustain cell proliferation.

Aberrant *TAL1* expression from the mutated allele

In order to associate the 3' microinsertion with the observed aberrant monoallelic *TAL1* expression in J-3'NE #1 cells (Figure 3A), and to prove that the expressed allele carried the mutation, we phased the mutation and the heterozygous 3´-UTR SNP used to study the allelic expression in this cell line. Since the mutation and the SNP are approximately 6.5 kb apart, this region was difficult to clone. To circumvent this problem, we used CRISPR-Cas9 to delete a large DNA fragment between the SNP and the mutation to reduce the distance between the two variants. The region was then PCR amplified. The small PCR product was either cloned into a plasmid and Sanger sequenced, or barcoded and directly sequenced using Mi-seq (Illumina; San Diego, CA, USA). We called this allele-phasing method "Alpha-Cas" (see Figure 3B). After deletion of the intervening DNA and cloning into the plasmid, 9 clones were sequenced. Sequences of two clones contained the unexpressed SNP G and the WT sequence; conversely, six clones contained the expressed SNP A and the mutated sequence, demonstrating the phasing of the mutation on the expressed allele. In addition, we found one clone with the unexpressed SNP G and the microinsertion, which likely resulted from a low rate of allelic exchange during the CRISPR-Cas9 cutting and recombination process (technical artefact) (Figure 3C). Mi-seq produced 56,054 aligned sequenced reads with 81% of reads having the expected allele phasing (47.9% of reads contained the expressed SNP A and microinsertion, 33.1% of reads had the unexpressed SNP G and WT sequence), and a minority of reads (19%) demonstrating allelic exchange (Figure 3D). We and others have previously described the *TAL1* monoallelic expression associated with the 5' super-enhancer mutation;[14,28] however, allele phasing was not performed at the time of its discovery. To validate the Alpha-Cas method, we used it to phase the 5' mutation and the expressed allele in the Jurkat cell line. We deleted the 20 kb intervening region between the SNP and 5' mutation before phasing. Alpha-Cas found the expected SNP and mutation phasing by both cloning and Mi-seq methods: 86.8% of sequencing reads had the expected phased alleles (51.4% of reads with the expressed SNP A and the microinsertion, 35.4% of reads with unexpressed SNP G and the WT sequence). Similarly, we observed a low rate of allelic exchange (13.1%) (*Online Supplementary Figure S4A-D*). This method was also suitable to phase intergenic mutations associated with other T-ALL oncogenes. The *LMO1* gene is also activated and monoallelically expressed by the creation of a neo-enhancer bound by MYB transcription factor in the Jurkat cell line.[35] The mutation and the informative SNP are 4.3 kb apart (*Online Supplementary Figure S5A and B*). We compared the direct cloning of the entire 4.3 kb region (*Online Supplementary Figure S5C*) to the Alpha-Cas method for allele-phasing (*Online Supplementary Figure S5D and E*). Direct cloning and Alpha-Cas produced comparable results with the expected phasing observed in 91% of the Mi-seq sequencing reads, thus validating this method. In the Molt-4 cell line, *LMO2* is activated via a distal microinsertion, creating once again a *de novo* MYB binding motif which leads to its ectopic monoallelic expression[36,37] (*Online Supplementary Figure S6A and B*). This mutation is situated at a long distance from the informative SNP (75.7 kb) making direct cloning impossible, and the Alpha-Cas method proved, therefore, very useful. Alpha-Cas revealed the expected phasing in 93.8% of reads (*Online Supplementary Figure S6C*). Despite a low rate of allelic exchange occurring during Cas9 recombination, Alpha-Cas is a simple, efficient, and reliable method for allele-phasing, especially in the case of very distant variants.

3' neo-enhancer: epigenetic profiling

Using the recently described ACT-seq method as an alternative to conventional ChIP-sequencing (see Methods), we observed the creation of the active 3' neo-enhancer by the increased H3K27ac and H3K4me1 enrichment at the microinsertion in J-3'NE #1, J-3'NE #2 and P-3'NE #1 de-

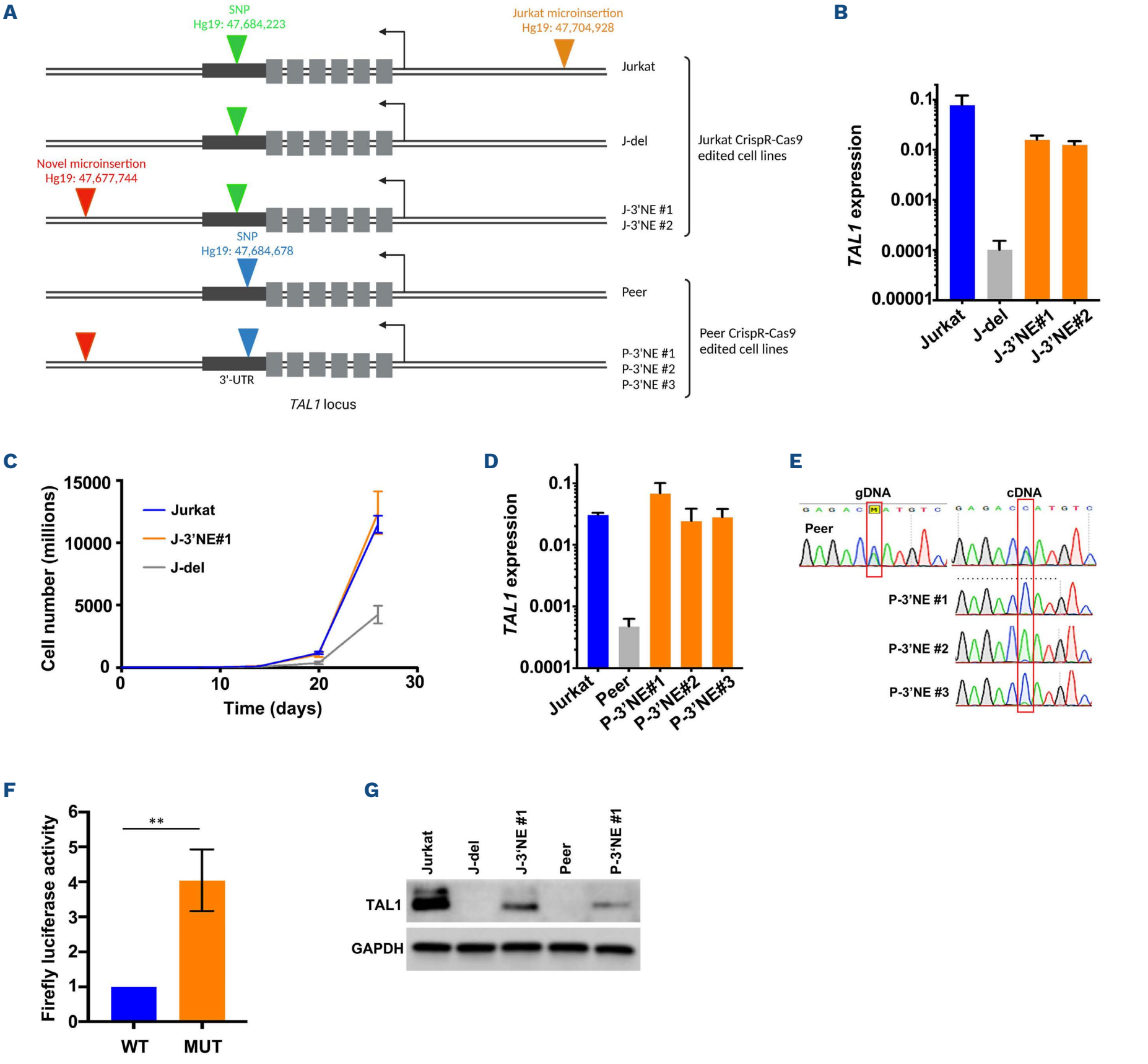


Figure 2. CrispR-cas9 engineered cell lines mimic the 3´ neo-enhancer and proximal mutation. (A) Graphic representation of the different CrispR-cas9 derivative cell lines engineered. J-del derives from Jurkat. In this clone, the 5´super-enhancer (orange arrow) was deleted. J-3'NE #1 and J-3'NE #2 also derive from Jurkat. In these clones, the 21 bp microinsertion (red arrow) was introduced 3´ to the TAL1 locus and the 5´super-enhancer was deleted. P-3'NE #1, P-3'NE #2 and P-3'NE #3 derive from the TAL1^{-} cell line Peer. In these clones, the 21 bp microinsertion was introduced 3´ to TAL1. Green and blue arrows represent the respective heterozygous SNP in the TAL1 3' UTR. (B) *TAL1* expression determined by real-time quantitative-polymerase chain reaction (RQ-PCR) in Jurkat, J-del, J-3'NE #1 and J-3'NE #2 derivative cell lines. Expression was normalized to the housekeeping gene GAPDH (n= 3, standard deviation [SD] is shown). (C) Cell proliferation analysis of Jurkat, J-3'NE #1 and J-del cell lines. Cell number was determined using the automated cell counter Countess (Invitrogen; ThermoFisher, Waltham, MA, USA). Three independent experiments were performed; mean cell number and SD (error bar) are shown. (D) *TAL1* expression normalized to *GAPDH* in Jurkat, Peer and Peer derivative cell lines P-3'NE #1, P-3'NE #2 and P-3'NE #3 (n= 3, SD is shown). (E) *TAL1* allelic expression in Peer derivative cell lines. Sanger sequencing chromatograms of the informative SNP in the 3' UTR of *TAL1* from genomic DNA (gDNA, left) or complementary DNA (cDNA, right) made from DNAse treated RNA in Peer, P-3'NE #1, P-3'NE #2, and P-3'NE #3. Red boxes highlight the SNP. (F) Luciferase reporter assay. A 556 bp fragment around the 3´neo-enhancer containing either the wild-type sequence (WT) or the mutant allele (MUT) was cloned upstream of luciferase and a minimal promoter in PGL4.23 plasmid. Constructs were electroporated into Jurkat cells. Firefly luciferase activity was measured after 24 hours, normalized to renilla luciferase activity to control for cell number and transfection efficiency, and expressed as a ratio relative to the activity of the WT sequence construct. Error bars are ±SD from three independent experiments. ***P*=0.004. (G) TAL1 Western blot in Jurkat, J-3'NE#1, Peer and P-3'NE#1 cell lines.

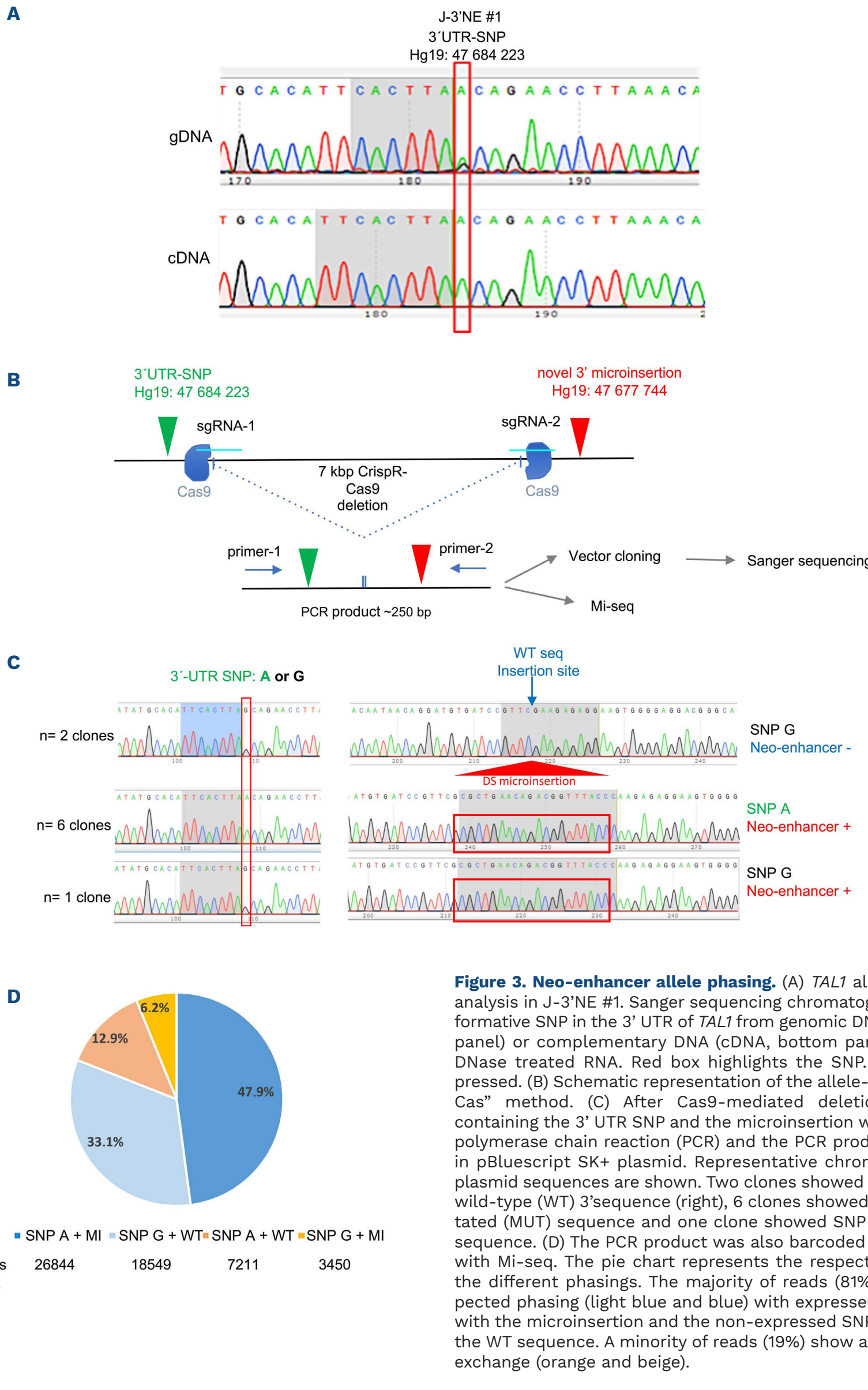


Figure 3. Neo-enhancer allele phasing. (A) *TAL1* allelic expression analysis in J-3'NE #1. Sanger sequencing chromatograms of the informative SNP in the 3' UTR of *TAL1* from genomic DNA (gDNA, upper panel) or complementary DNA (cDNA, bottom panel) made from DNase treated RNA. Red box highlights the SNP. Allele A is expressed. (B) Schematic representation of the allele-phasing "Alpha-Cas" method. (C) After Cas9-mediated deletion, the region containing the 3' UTR SNP and the microinsertion was amplified by polymerase chain reaction (PCR) and the PCR product was cloned in pBluescript SK+ plasmid. Representative chromatograms of 9 plasmid sequences are shown. Two clones showed SNP G (left) and wild-type (WT) 3'sequence (right), 6 clones showed SNP A and mutated (MUT) sequence and one clone showed SNP G and mutated sequence. (D) The PCR product was also barcoded and sequenced with Mi-seq. The pie chart represents the respective fractions of the different phasings. The majority of reads (81%) show the expected phasing (light blue and blue) with expressed SNP A phased with the microinsertion and the non-expressed SNP G phased with the WT sequence. A minority of reads (19%) show artefactual allele exchange (orange and beige).

rivative cell lines (Figure 4A and B). The corresponding sequencing reads contained uniquely the mutated sequence confirming the monoallelic epigenetic activation of the mutated allele (Figure 4C). As control, we also observed an enrichment at the native 5' super-enhancer in the unedited Jurkat cell line, which was lost in the J-3'NE #1, J-3'NE #2 cell lines (Figure 4A). Using primers spanning the 3' microinsertion site, we performed H3K27ac ChIP-qPCR and observed an increased enrichment in J-3'NE #1 cells compared to Jurkat cells (Figure 4D). We also carried out an allele specific H3K27ac ChIP-qPCR in J-3'NE #1 cells and observed an increased enrichment of this histone mark at the mutated allele compared to the WT allele (Figure 4E).

Microinsertions downstream of the *TAL1* gene are recurrent in T-cell acute lymphoblastic leukemia and neo-enhancer activity depends on MYB transcription factor binding.

The 3´neo-enhancer is a rare but recurrent oncogenic event, as we identified another patient (UPNT-613) with a 6 bp microinsertion at the same genomic location from an independent series of 189 T-ALL samples screened for this mutation by Sanger sequencing. This mutation was also detected in its corresponding PDX model (*Online Supplementary Figure S7A*). Moreover, this patient displayed a high unresolved *TAL1* expression (Figure 5A) and, like UPNT-802, this high *TAL1* expression was stable in the PDX model (Figure 5B). Using Oxford Nanopore sequencing technology (Oxford Nanopore Technologies; Oxford, UK), we performed a targeted analysis of TAL1 expression in the two 3'NE samples (UPNT-802 and UPNT-613 PDX cells) and in two 5'SE samples (Jurkat cells and UPNT-525 PDX cells) (Figure 5C, *Online Supplementary Figure S7B-D*). Our results revealed differences in the transcription of the 5'region of the *TAL1* gene between the 3'NE and 5'SE samples. Long amplicons were detected (S1/R2 primer pair) uniquely in 5'SE samples whereas 3'NE samples initiated transcription downstream of exon 1 without affecting the *TAL1* open reading frame (ORF) (S3/R2 primer pair). We also detected transcription activated downstream of the *TAL1* gene near the 3'NE. However, transcription initiated here was not specific to the 3'NE regulatory element but specific to TAL1^{+} samples, as it was detectable in both 3'NE and 5'SE samples and absent in the TAL1^{-} sample (DND-41) (S9b/R5 primer pair). Similarly to the previously described 5' super-enhancer mutation, both the 3´ microinsertions in UPNT-802 and UPNT-613 patient samples were predicted (JASPAR) to create a *de novo* binding site for MYB transcription factor and other known members of the TAL1 complex[38] (Figure 5D, *Online Supplementary Table S4*) suggesting the creation of the 3' neo-enhancer is dependent on MYB binding. Using primers spanning either the 5' super-enhancer site in Jurkat cells (5'SE) or the 3' microinsertion site (3'NE), we performed MYB ChIP qPCR and observed an increased MYB enrichment at the 3'NE in J-3'NE #1 and P-3'NE #1 cells. As controls, we verified MYB enrichment at the 5'SE in Jurkat WT cells and its absence in Peer cells lacking MYB-driven super-enhancers (Figure 6A). MYB binding at the 3'NE was also detected in UPNT-802 PDX cells (*Online Supplementary Figure S8A*). Furthermore, to prove that the observed MYB binding was specific to the 3'mutated sequence, we performed an allele specific MYB ChIP-qPCR in J-3'NE #1, P-3'NE #1 cells and UPNT-802 cells and observed an increased MYB enrichment at the mutated allele compared to the WT allele (Figure 6B, *Online Supplementary Figure S8B*). In order to definitively link the MYB-dependent epigenetic activation of *TAL1*, we repressed MYB expression in Jurkat cells using an inducible dCas9-KRAB-MECP2 system with two sgRNA targeting the *MYB* promoter (*Online Supplementary Figure S8C*) and measured enhancer activity using a luciferase reporter assay. *MYB* silencing significantly reduced the enhancer activity associated with the mutated sequence (Figure 6C). Using the same inducible MYB repression system in J'3-NE#1 cells (Figure 6D) resulted in significantly reduced transcriptional and protein TAL1 expression (Figure 6E), demonstrating the essential role of MYB TF for TAL1 expression in these cells. *LMO1*, whose expression is also driven by a MYB-dependent neo-enhancer in Jurkat cells, served as a positive control (*Online Supplementary Figure S8D*).

Discussion

TAL1 is a major oncogene in both adult and pediatric T-ALL. Intriguingly, the mechanisms leading to TAL1 overexpression are extremely diverse. Genetic rearrangements of the *TAL1* gene such as SIL-TAL1 microdeletions and *TAL1* translocations with TCR loci (*TRD* or *TRB*) have long been described.[12,13,24-26] More recently, we and others[14,28] uncovered novel mutations in a non-coding region leading to oncogene activation via the creation of an oncogenic neo-enhancer upstream of *TAL1*. The discovery of these mutations represented a major conceptual advancement, highlighting the underestimated contribution of intergenic mutations in genomic alterations driving cancers. In this circumstance, somatic heterozygous microinsertions of variable size create *de novo* binding sites for MYB TF, leading to the recruitment of a multi-protein complex, epigenetic activation and *TAL1* overexpression in leukemic cells. Since this discovery, similar 5' non-coding mutations affecting two other major T-ALL oncogenes, *LMO1* and *LMO2*, that are frequently co-deregulated with *TAL1* were found.[35,37]

Intergenic non-coding mutations leading to oncogene deregulation are not limited to T-ALL. A pivotal example is the frequent *TERT* promoter mutations found in many aggressive solid cancers such as melanomas, glioblastomas,

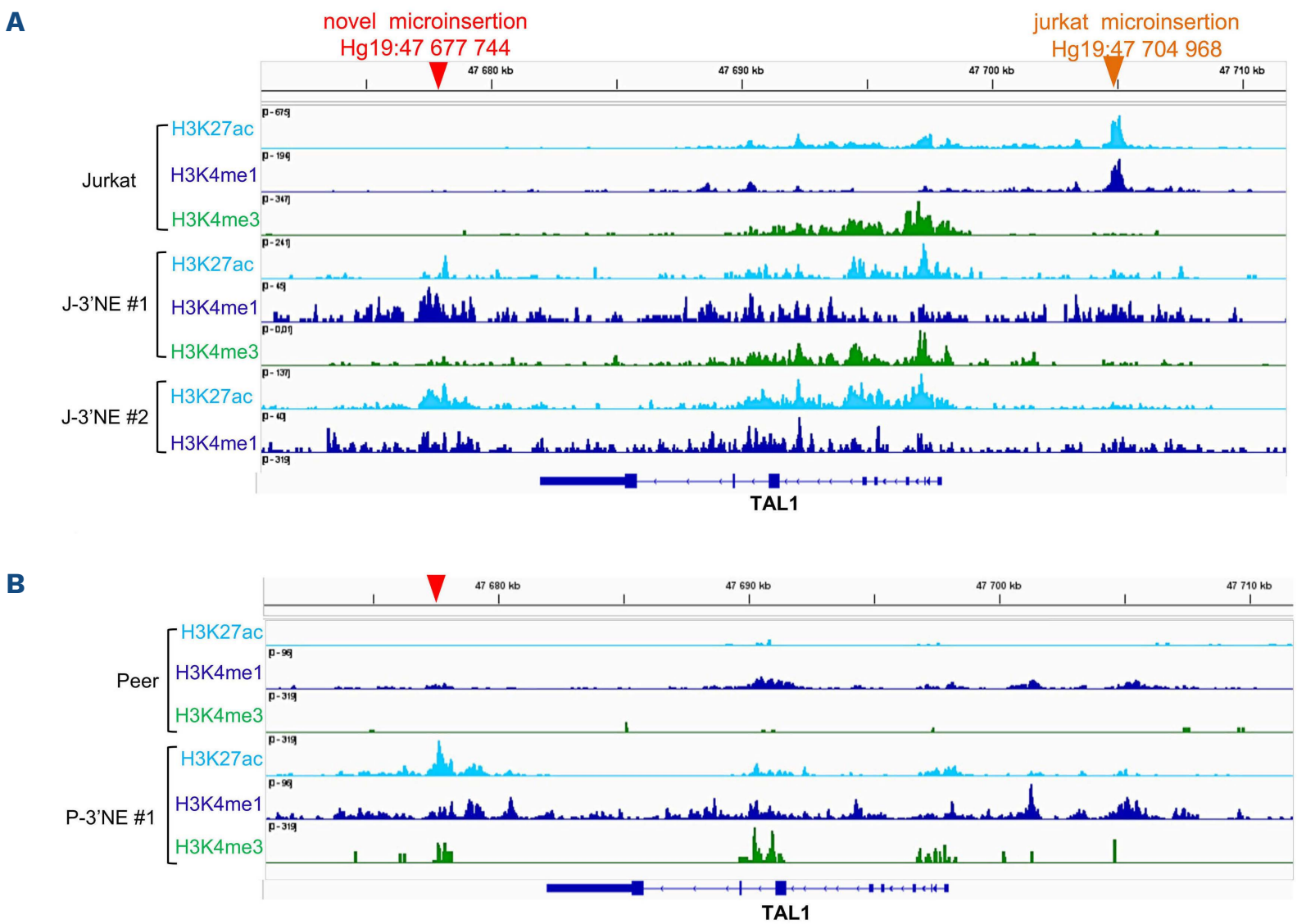


Figure 4. Epigenetic profiling in Jurkat and Peer derivative cell lines. (A) ACT-sequencing (seq) tracks for H3K27ac, H3K4me1 and H3K4me3 histone marks in Jurkat, J-3'NE #1, and J-3'NE #2 derivative cell lines centered over the *TAL1* locus. Orange arrow highlights the 5'super-enhancer present in wild-type (WT) Jurkat; red arrow highlights the 3'microinsertion present in J-3'NE #1 and J-3'NE #2. (B) ACT-seq tracks for H3K27ac, H3K4me1, and H3K4me3 histone marks over the TAL1 *locus* in Peer and P-3'NE #1 derivative cell lines. Red arrow highlights the 3'microinsertion present in P-3'NE #1 cell line. (C) H3K27ac and H3K4me1 ACT-seq reads in J-3'NE #2 and P-3'NE #1 over the 3'microinsertion site. All reads contain the mutation. (D) Enrichment of H3K27ac at the 3'microinsertion site in Jurkat and J-3'NE #1 cells as assayed by chromatin immunoprecipitation sequencing (ChIP-seq)-quantitative polymerase chain reaction (qPCR) using primers spanning the 3'microinsertion. Mean of three independent experiments with the Standard Deviation (SD) error bar is shown. (E) Enrichment of H3K27ac of the WT or mutant (MUT) allele at the 3'microinsertion site in J-3'NE #1 cells as assayed by ChIP-qPCR using WT specific or MUT specific primers. The mean of three independent experiments and the SD error bar are shown.

medulloblastomas, and hepatocellular carcinomas.[39] These non-coding mutations occur in the promoter region of *TERT* and create *de novo* binding sites for ETS TF, increase chromatin accessibility of the mutant alleles, and cause an epigenetic switch. Finally, a rare, recurrent non-coding mutation in the promoter region of TAL1 was discovered in two pediatric T-ALL samples. The authors showed that this mutation creates a *de novo* binding site for YY1 TF and leads to *TAL1* activation.[27]

In the present work, by combining histone ChIP-seq and

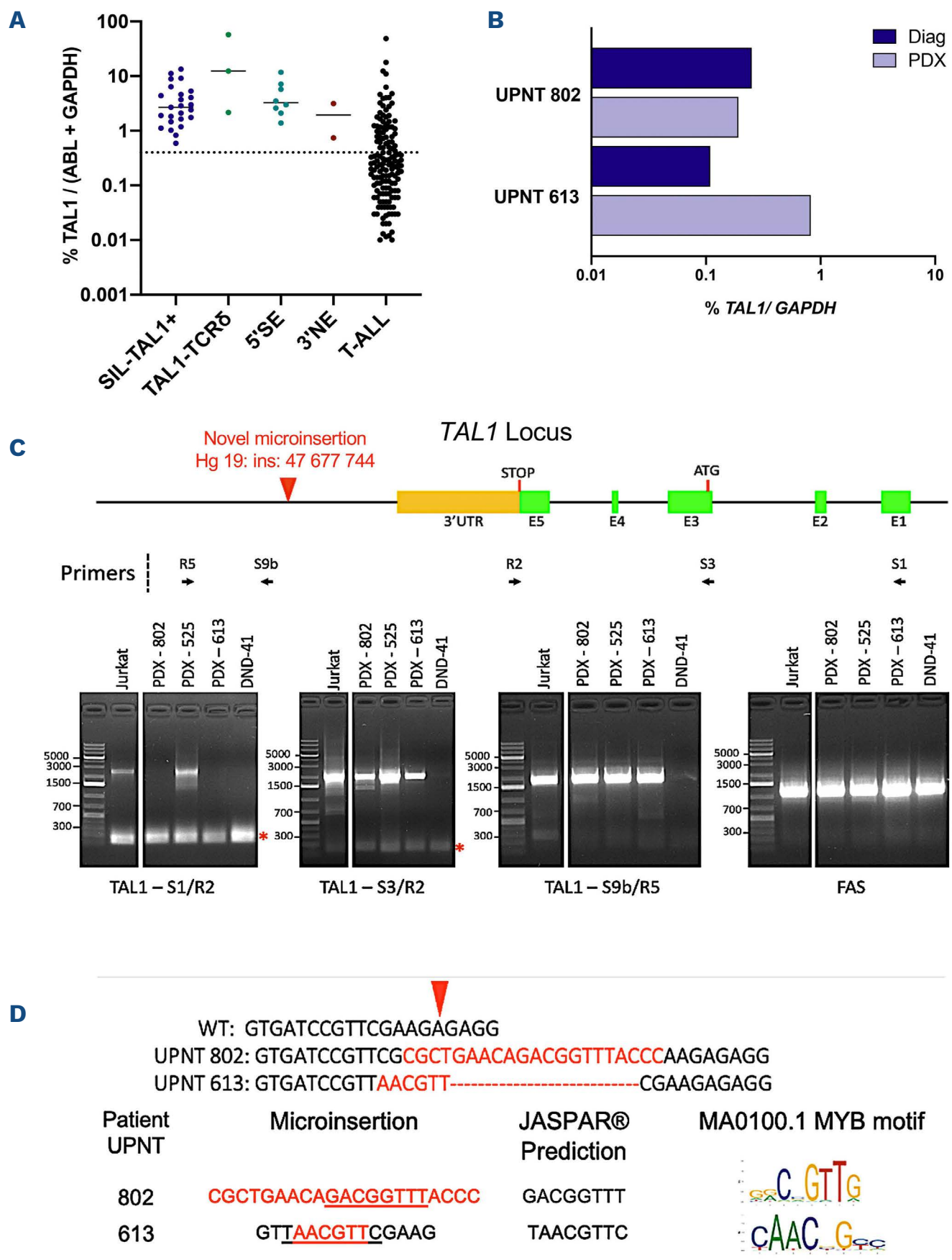
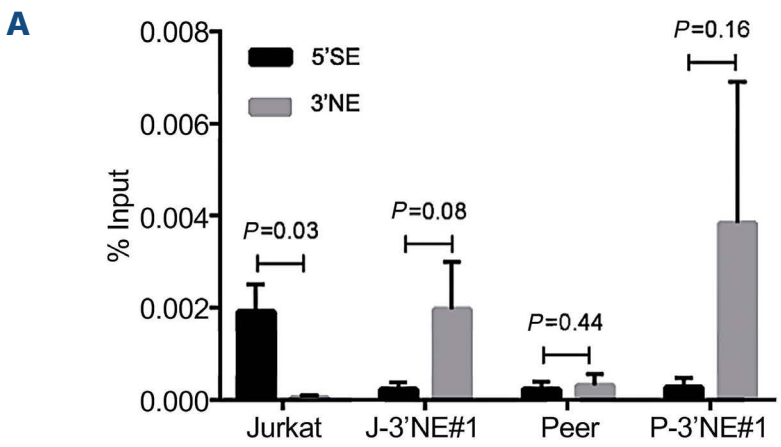
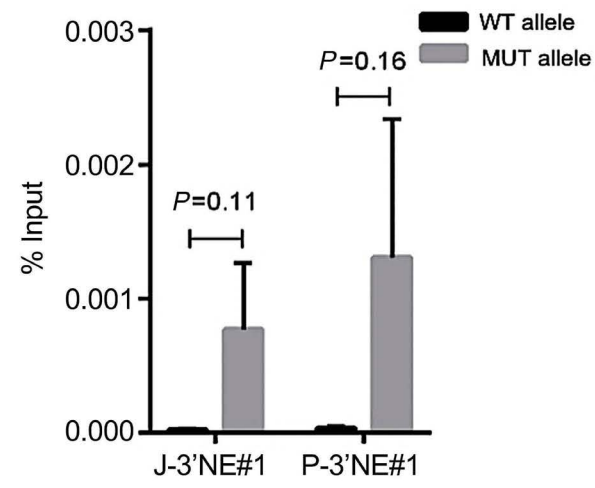
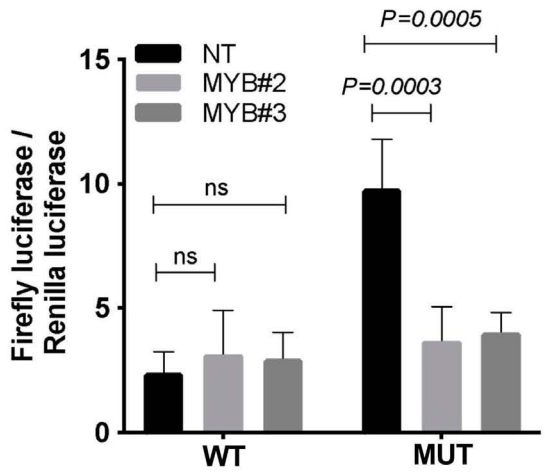


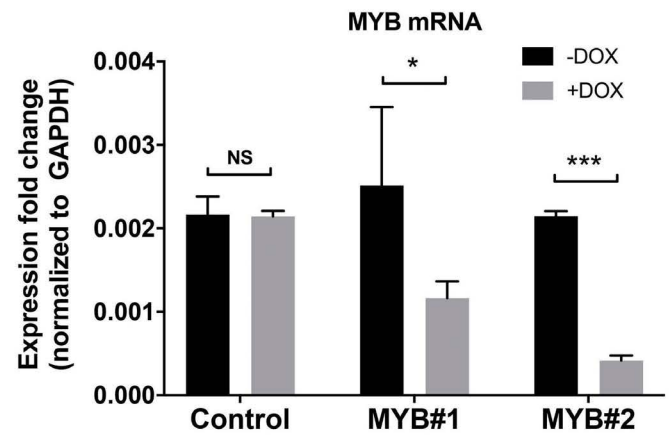
Figure 5. 3'NE *TAL1* transcription analysis. (A) *TAL1* expression was measured by real-time-quantitative polymerase chain reaction RQ-PCR in an independent series of 189 T-cell acute lymphoblastic leukemia (T-ALL). *TAL1* expression was normalized to *ABL* and *GAPDH* expression and presented according to TAL1 deregulation mechanism, when known. 3'microinsertions are annotated 3'NE. (B) *TAL1* expression in diagnostic and patient-derived xenograft (PDX) 3'NE samples. Expression was normalized to *GAPDH*. (C) Schematic representation of the *TAL1* locus showing the positions of the several primers used to amplify different *TAL1* transcripts by non-quantitative reverse transcriptase-PCR. *TAL1* exons are shown in green (top). Agarose gel analysis of *TAL1* and *FAS* amplicons before long read sequencing. Red stars denote PCR primer dimers. FAS was used as a positive PCR control (bottom). (D) 3'mutations are predicted to create MYB binding sites (JASPAR).

gene expression data analysis we identified a rare but recurrent novel oncogenic neo-enhancer responsible for TAL1 activation in T-ALL with unknown deregulation mechanisms. By a mechanism analogous to TAL1 5'SE mutations, we demonstrated that 3'NE mutations create a *de novo* binding site for MYB TF, which drives the deposition of the observed epigenetic marks and aberrant expression of TAL1, probably via its recruitment of CBP/p300 and members of the TAL1 complex (or core regulatory circuit, CRC).[38] Interestingly, these novel microinsertions are the first oncogenic neo-enhancers found downstream of the *TAL1* gene. This characteristic is in line with the enhancers' independent position. Enhancers can be located upstream or downstream of a regulated gene and interact with proximal regulatory elements in the vicinity of transcription start sites, sometimes via long-range chromatin loops.[40,41] These types









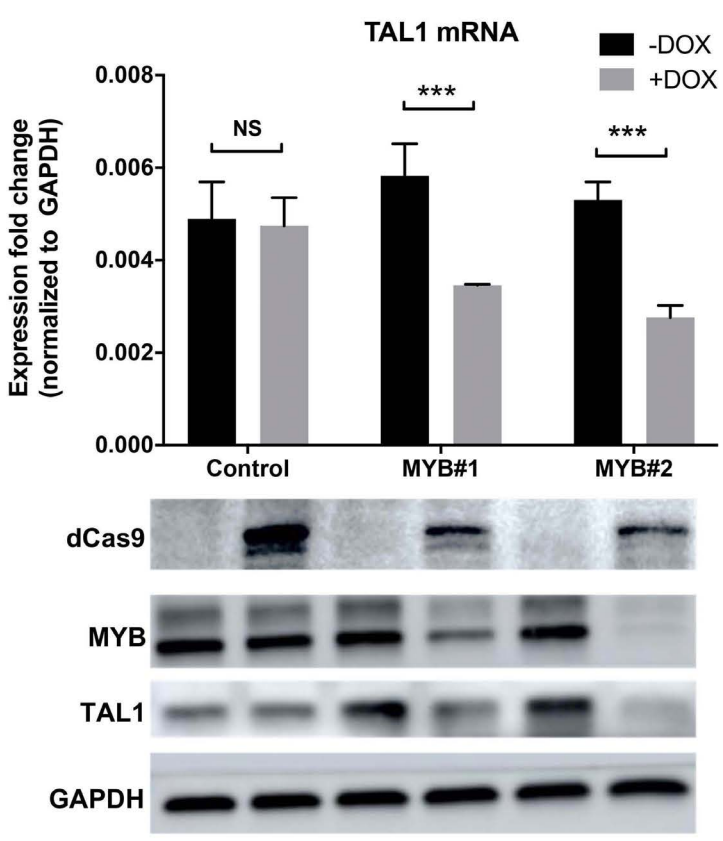


Figure 6. The neo-enhancer activity is MYB dependent. (A) Enrichment of MYB at 5'super-enhancer (5'SE) or 3'microinsertion site (3'NE) in Jurkat, J-3'NE#1, Peer, and P-3'NE#1 cells using primers spanning the 5'SE or 3'NE, respectively. Mean of three independent experiments and the standard deviation (SD error) bar are shown. (B) Enrichment of MYB at wild-type (WT) or mutated (MUT) alleles at the 3'NE in J-3'NE #1 or P-3'NE #1 cells as assayed by chromatin immunoprecipitation (ChIP)-quantitative-polymerase chain reaction (qPCR) using WT-specific or MUT-specific primers. The mean of three independent experiments and the Standard Deviation (SD) error bar are shown. (C) Luciferase reporter assay using either the WT or MUT allele sequence upon MYB knockdown. Firefly luciferase activity was measured after 24 hours, normalized to renilla luciferase activity to control for cell number and transfection efficiency, and expressed as a ratio relative to the activity of the WT sequence construct. Mean and SD error bar from three independent experiments. (D) Using a repressive doxycycline inducible-dCas9-KRAB-MeCP2 system and two different sgRNA targeting the *MYB* locus, MYB was knockdown in the J-3'NE #1 cell line resulting in significantly decreased MYB transcriptional expression (n=4, mean and SD shown). (E) MYB inhibition led to significantly decreased transcriptional *TAL1* expression assessed by RQ-PCR and normalized to GAPDH compared to sgRNA control (n=4, mean and SD shown) (top) and decreased TAL1 protein expression: Western blot showing dCas9, MYB, TAL1 and *GAPDH* protein expression corresponding to the upper expression graph (bottom). ns: not significant.

of long-range interactions may be a consequence of genetic alterations at more distant loci and could be pertinent deregulation mechanisms underlining outstanding unresolved monoallelic *TAL1* expression in T-ALL. A systematic analysis of histone ChIP-seq data in such cases could be an efficient way to discover potential distant oncogenic neo-enhancers and should be investigated in further studies.

In addition to identifying 3'microinsertions, we demonstrated the mutation's exclusive presence on the aberrantly-expressed allele using allele phasing. This is also true for *TAL1* 5'SE mutations and *LMO1* and *LMO2* intergenic neo-enhancer mutations. These results were expected but had not yet been proven. We circumvented the challenges of phasing distant variants in these cases by developing a new technique based on CRISPR-Cas9 genome editing technology, called Alpha-Cas, to bring these variants closer and facilitate allele phasing analysis. This method is easy to implement and could provide a useful tool to phase potentially very distant mutations and oncogenes.

Disclosures

No conflicts of interest to disclose.

Contributions

VA, AT and CP conceived and designed the study. AT, CS, AG, DW, EA, AR, MM, AP and OM performed the experiments. AM, UT, ES and KB analyzed the sequencing data. All authors critically reviewed the manuscript.

Acknowledgments

The authors would like to thank the Genomics and Proteomics Core facilities at DKFZ and the Transcriptomics and Genomics, Marseille-Luminy platform for sequencing the ChIP-seq samples. We would also like to thank Guillaume Charbonnier for his help processing the NGS data and Frederic Tores from the bioinformatics platform of Imagine Institute for the ONT analysis.

Funding

The study was supported in part by the Helmholtz-Foundation. AT was supported by a DKFZ postdoctoral fellowship. The work in the VA lab was supported by ARC-Labellisation and the associations "Force Hémato" and "Laurette Fugain." This study was also supported in the lab by grants from INCA PLBIO18-031, PLBIO 2018-00252, and ITMO Cancer Epig-2015 (to VA and SS). SS was supported by the Ligue Contre le Cancer (labeling 2018) and the RNA-sequencing was funded by LIGUE. This work was also supported by La Fondation pour la Recherche Médicale with grant FDT202012010638 (to CS). The ONT sequencing was also supported by ANR funding KREM-AIF (ANR-21-CE17-0014-02) (to EA).

Data-sharing agreement

Data generated during this study have been submitted to the NCBI Gene Expression Omnibus repository under the accession number GSE200860.

References

1. Ferrando AA, Neuberg DS, Staunton J, et al. Gene expression signatures define novel oncogenic pathways in T cell acute lymphoblastic leukemia. Cancer Cell. 2002;1(1):75-87.
2. Marks DI, Paietta EM, Moorman AV, et al. T-cell acute lymphoblastic leukemia in adults: clinical features, immunophenotype, cytogenetics, and outcome from the large randomized prospective trial (UKALL XII/ECOG 2993). Blood. 2009;114(25):5136-5145.
3. Rivera GK, Raimondi SC, Hancock ML, et al. Improved outcome in childhood acute lymphoblastic leukaemia with reinforced early treatment and rotational combination chemotherapy. Lancet. 1991;337(8733):61-66.
4. Huguet F, Chevret S, Leguay T, et al. Intensified therapy of acute lymphoblastic leukemia in adults: report of the randomized GRAALL-2005 clinical trial. J Clin Oncol. 2018;36(24):25142523.
5. Desjonquères A, Chevallier P, Thomas X, et al. Acute lymphoblastic leukemia relapsing after first-line pediatric-inspired therapy: a retrospective GRAALL study. Blood Cancer J. 2016;6(12):e504.
6. Bond J, Graux C, Lhermitte L, et al. Early response-based therapy stratification improves survival in adult early thymic precursor acute lymphoblastic leukemia: a Group for Research on Adult Acute Lymphoblastic Leukemia study. J Clin Oncol. 2017;35(23):2683-2691.
7. Dik WA, Brahim W, Braun C, et al. CALM-AF10+ T-ALL expression profiles are characterized by overexpression of HOXA and BMI1 oncogenes. Leukemia. 2005;19(11):1948-1957.
8. Kennedy MA, Gonzalez-Sarmiento R, Kees UR, et al. HOX11, a homeobox-containing T-cell oncogene on human chromosome 10q24. Proc Natl Acad Sci U S A. 1991;88(20):8900-8904.
9. Bernard OA, Busson-LeConiat M, Ballerini P, et al. 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. 2001;15(10):1495-1504.
10. Van Vlierberghe P, Homminga I, Zuurbier L, et al. Cooperative genetic defects in TLX3 rearranged pediatric T-ALL. Leukemia. 2008;22(4):762-770.
11. Homminga I, Pieters R, Langerak AW, et al. Integrated transcript and genome analyses reveal NKX2-1 and MEF2C as potential oncogenes in T cell acute lymphoblastic leukemia. Cancer Cell. 2011;19(4):484-497.
12. Begley CG, Aplan PD, Davey MP, et al. Chromosomal translocation in a human leukemic stem-cell line disrupts the T-cell antigen receptor delta-chain diversity region and results in a previously unreported fusion transcript. Proc Natl Acad Sci U S A. 1989;86(6):2031-2035.
13. Brown L, Cheng J-T, Chen Q, et al. Site-specific recombination of the tal-1 gene is a common occurrence in human T cell leukemia. EMBO J. 1990;9(10):3343-3351.
14. Navarro J-M, Touzart A, Pradel LC, et al. Site- and allele-specific polycomb dysregulation in T-cell leukaemia. Nat Commun. 2015;6:6094.
15. Mansour MR, Abraham BJ, Anders L, et al. An oncogenic super-enhancer formed through somatic mutation of a noncoding

intergenic element. Science. 2014;346(6215):1373-1377.
16. Xia Y, Brown L, Yang CY, et al. TAL2, a helix-loop-helix gene activated by the (7;9)(q34;q32) translocation in human T-cell leukemia. Proc Natl Acad Sci U S A. 1991;88(24):11416-11420.
17. Mellentin JD, Smith SD, Cleary ML. lyl-1, a novel gene altered by chromosomal translocation in T cell leukemia, codes for a protein with a helix-loop-helix DNA binding motif. Cell. 1989;58(1):77-83.
18. Wang J, Jani-Sait SN, Escalon EA, et al. The t(14;21)(q11.2;q22) chromosomal translocation associated with T-cell acute lymphoblastic leukemia activates the BHLHB1 gene. Proc Natl Acad Sci U S A. 2000;97(7):3497-3502.
19. Royer-Pokora B, Loos U, Ludwig WD. TTG-2, a new gene encoding a cysteine-rich protein with the LIM motif, is overexpressed in acute T-cell leukaemia with the t(11;14)(p13;q11). Oncogene. 1991;6(10):1887-1893.
20. McGuire EA, Hockett RD, Pollock KM, Bartholdi MF, O'Brien SJ, Korsmeyer SJ. The t(11;14)(p15;q11) in a T-cell acute lymphoblastic leukemia cell line activates multiple transcripts, including Ttg-1, a gene encoding a potential zinc finger protein. Mol Cell Biol. 1989;9(5):2124-2132.
21. Clappier E, Cuccuini W, Kalota A, et al. The C-MYB locus is involved in chromosomal translocation and genomic duplications in human T-cell acute leukemia (T-ALL), the translocation defining a new T-ALL subtype in very young children. Blood. 2007;110(4):1251-1261.
22. Herblot S, Steff A, Hugo P, Aplan PD, Hoang T. SCL and LMO1 alter thymocyte differentiation: inhibition of E2A-HEB function and pre-Tα chain expression. Nat Immunol. 2000;1(2):138-144.
23. Zhang JA, Mortazavi A, Williams BA, Wold BJ, Rothenberg EV. Dynamic transformations of genome-wide epigenetic marking and transcriptional control establish T cell identity. Cell. 2012;149(2):467-482.
24. Bernard O, Guglielmi P, Jonveaux P, et al. Two distinct mechanisms for the SCL gene activation in the t(1;14) translocation of T-cell leukemias. Genes Chromosom Cancer. 1990;1(3):194-208.
25. Cauwelier B, Dastugue N, Cools J, et al. Molecular cytogenetic study of 126 unselected T-ALL cases reveals high incidence of TCRβ locus rearrangements and putative new T-cell oncogenes. Leukemia. 2006;20(7):1238-1244.
26. Le Noir S, Ben Abdelali R, Lelorch M, et al. Extensive molecular mapping of TCRalpha/delta- and TCRbeta-involved chromosomal translocations reveals distinct mechanisms of oncogene activation in T-ALL. Blood. 2012;120(16):3298-3309.
27. Liu Y, Li C, Shen S, et al. Discovery of regulatory noncoding variants in individual cancer genomes by using cis-X. Nat Genet. 2020;52(8):811-818.
28. Mansour MR, Abraham BJ, Anders L, et al. Oncogene regulation. An oncogenic super-enhancer formed through somatic mutation of a noncoding intergenic element. Science. 2014;346(6215):1373-1377.
29. Ferrando AA, Herblot S, Palomero T, et al. Biallelic transcriptional activation of oncogenic transcription factors in T-cell acute lymphoblastic leukemia. Blood. 2004;103(5):1909-1911.
30. Tan TK, Zhang C, Sanda T. Oncogenic transcriptional program driven by TAL1 in T-cell acute lymphoblastic leukemia. Int J Hematol. 2019;109(1):5-17.
31. Carter B, Ku WL, Kang JY, et al. Mapping histone modifications in low cell number and single cells using antibody-guided chromatin tagmentation (ACT-seq). Nat Commun. 2019;10(1):3747.
32. Liu C-S, Toth R, Bakr A, et al. Epigenetic modulation of radiation-induced diacylglycerol kinase alpha expression prevents pro-fibrotic fibroblast response. Cancers. 2021;13(10):2455.
33. Cieslak A, Charbonnier G, Tesio M, et al. Blueprint of human thymopoiesis reveals molecular mechanisms of stage-specific TCR enhancer activation. J Exp Med. 2020;217(9):e20192360.
34. Belhocine M, Simonin M, Abad Flores JD, et al. Dynamic of broad H3K4me3 domains uncover an epigenetic switch between cell identity and cancer-related genes. Genome Res. 2022;32(7):13281342.
35. Hu S, Qian M, Zhang H, et al. Whole-genome noncoding sequence analysis in T-cell acute lymphoblastic leukemia identifies oncogene enhancer mutations. Blood. 2017;129(24):3264-3268.
36. Abraham BJ, Hnisz D, Weintraub AS, et al. Small genomic insertions form enhancers that misregulate oncogenes. Nat Commun. 2017;8(1):14385.
37. Rahman S, Magnussen M, León TE, et al. Activation of the LMO2 oncogene through a somatically acquired neomorphic promoter in T-cell acute lymphoblastic leukemia. Blood. 2017;129(24):3221-3226.
38. Sanda T, Lawton LN, Barrasa MI, et al. Core transcriptional regulatory circuit controlled by the TAL1 complex in human T cell acute lymphoblastic leukemia. Cancer Cell. 2012;22(2):209-221.
39. Rachakonda S, Hoheisel JD, Kumar R. Occurrence, functionality and abundance of the TERT promoter mutations. Int J Cancer. 2021;149(11):1852-1862.
40. Kim S, Shendure J. Mechanisms of interplay between transcription factors and the 3D genome. Mol Cell. 2019;76(2):306-319.
41. Bulger M, Groudine M. Functional and mechanistic diversity of distal transcription enhancers. Cell. 2011;144(3):327-339.