
This is the **accepted version** of the journal article:

Belver, Laura [et al.]. «GATA3-controlled nucleosome eviction drives MYC enhancer activity in T-cell development and leukemia». *Cancer Discovery*, Vol. 9 Núm. 12 (december 2019), p. 1774-1791. DOI 10.1158/2159-8290.CD-19-0471, PMID 31519704

This version is available at <https://ddd.uab.cat/record/322399>

under the terms of the  ^{IN}COPYRIGHT license.

Gata3-controlled nucleosome eviction drives *Myc* enhancer activity in T-cell development and leukemia

Authors: Laura Belver¹, Alexander Y. Yang^{1,2}, Robert Alberio^{1,2}, Daniel Herranz^{3,4}, Francesco G. Brundu⁵, S. Aidan Quinn¹, Pablo Pérez-Durán¹, Silvia Álvarez¹, Francesca Gianni¹, Marissa Rashkovan¹, Devya Gurung¹, Pedro P. Rocha⁶, Ramya Raviram^{7,8}, Clara Reglero¹, Jose R. Cortés¹, Anisha J. Cooke¹, Agnieszka A. Wendorff¹, Valentina Cordó⁹, Jules P. Meijerink⁹, Raúl Rabadan^{4,10} and Adolfo A. Ferrando^{1,4,11,12*}

Affiliations:

¹ Institute for Cancer Genetics, Columbia University, New York, NY, 10032, USA.

² These authors contributed equally.

³ Rutgers Cancer Institute of New Jersey, Rutgers University, New Brunswick, 08901, New Jersey, USA.

⁴ Department of Pharmacology, Robert Wood Johnson Medical School, Rutgers University, Piscataway, 08854, New Jersey, USA.

⁵ Department of Systems Biology, Columbia University, New York, NY, 10032, USA.

⁶ Division of Developmental Biology, Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD, 20892, USA.

⁷ Ludwig Institute for Cancer Research, La Jolla, CA, 92093, USA.

⁸ Department of Chemistry and Biochemistry University of California San Diego, La Jolla, CA, 92093, USA.

⁹ Department of Pediatric Oncology/Hematology, Princess Maxima Center for Pediatric Oncology, 3584 CS, Utrecht, the Netherlands.

¹⁰ Department of Biomedical Informatics, Columbia University, New York, NY, 10032, USA.

¹¹ Department of Pediatrics, Columbia University Medical Center, New York, NY, 10032, USA.

¹² Department of Pathology, Columbia University Medical Center, New York, NY, 10032, USA.

* Corresponding author

Running title: MYC enhancer nucleosome eviction in NOTCH1-induced leukemia

Keywords: NOTCH1, T-cell acute lymphoblastic leukemia, T-ALL, enhancer, Myc, Gata3

Corresponding author:

Adolfo Ferrando MD PhD

Institute for Cancer Genetics, Columbia University Medical Center

1130 St Nicholas Ave. ICRC-402A. New York, NY, 10032

Phone: 212-851-4611; FAX: 212-851-5256

E-mail: af2196@columbia.edu

Conflict of interest disclosure:

Adolfo Ferrando: Consulting for Ayala Pharmaceuticals and SpringWorks Therapeutics. Previous research support: Pfizer, Bristol Myers Squibb, Merck, Eli Lilly. Patent and reagent licensing royalties: Novartis, EMD Millipore and Applied Biological Materials.

Word count: 7888

Number of figures and tables: Seven main figures, four supplementary figures and four supplementary tables.

Abstract

Long-range enhancers govern the temporal and spatial control of gene expression, however the mechanisms that regulate enhancer activity during normal and malignant development remain poorly understood. Here we demonstrate a role for aberrant chromatin accessibility in the regulation *Myc* expression in T-cell lymphoblastic leukemia (T-ALL). Central to this process, N-Me, a long-range T-cell specific *Myc* enhancer shows dynamic changes in chromatin accessibility during T-cell specification and maturation and aberrant high degree of chromatin accessibility in mouse and human T-ALL cells. Mechanistically, we demonstrate that Gata3-driven nucleosome eviction dynamically modulates N-Me enhancer activity and is strictly required for NOTCH1-induced T-ALL initiation and maintenance. These results directly implicate aberrant regulation of chromatin accessibility at oncogenic enhancers as mechanism of leukemia transformation.

Statement of significance

MYC is a major effector of NOTCH1 oncogenic programs in T-ALL. Here we show a major role for Gata3-mediated enhancer nucleosome eviction as a driver of *Myc* expression and leukemia transformation. These results support a role of aberrant chromatin accessibility and consequent oncogenic *MYC* enhancer activation in NOTCH1-induced T-ALL.

Introduction

Enhancers are long-range, orientation-independent, cis-acting DNA regulatory elements that control gene expression through physical interaction with proximal regulatory elements located at gene promoters (1-3). Temporal and spatial transcriptional regulation of key developmental factors is frequently coordinated by clusters of distal enhancers organized in regulatory domains (4,5). Active enhancers competent for transcription factor binding and transcriptional regulation show low nucleosome occupancy (6,7), and enhancers that work simultaneously often display coordinated patterns of DNA accessibility, while those that work in mutually exclusive modes show divergent chromatin accessibility profiles (8).

Constitutive activation of NOTCH1 signaling plays a prominent driver role in over 60% of T-cell acute lymphoblastic leukemias (T-ALL) harboring activating mutations in the *NOTCH1* gene (9). Oncogenic NOTCH1 drives T-cell transformation activating a broad transcriptional program that promotes leukemia cell growth and proliferation. Most prominently, NOTCH1 directly activates *MYC* expression and NOTCH1 and *MYC* share multiple common direct target genes driving leukemia cell growth in T-ALL (10). Consistently, N-Me, a NOTCH1-controlled T-cell specific *MYC* long-range enhancer is strictly required for NOTCH1-induced T-ALL (11). Notably, although activating mutations in NOTCH1 are also found in adenoid cystic carcinoma (12,13), chronic lymphocytic leukemia (14) and mantle cell lymphomas (15), N-Me seems to be selectively active only during early T-cell development and in T-ALL (11). This observation supports that as yet unrecognized T-cell specific signaling, transcriptional or epigenetic factors epistatic with NOTCH1 signaling are dominantly required for N-Me enhancer activity and may contribute to leukemia transformation.

Results

Dynamic changes in chromatin accessibility during thymocyte development

T-cell precursors follow an orchestrated developmental program that begins with double negative (DN) 1 cells, the earliest cell entrants in the thymus, and progresses to uncommitted DN2a progenitors, which become T-cell committed as they mature into DN2b cells (16). These early precursors subsequently progress through highly proliferative DN3, DN4 and intermediate single positive (ISP) thymocyte stages, which then exit the cell cycle as they mature into double positive (DP) and ultimately mature single positive CD4 (CD4SP) and CD8 (CD8SP) T cells (16). Analysis of chromatin accessibility by Assay of Transposase-Accessible Chromatin using sequencing (ATAC-seq) in sorted mouse thymocyte precursors identified 69,302 highly accessible regions. Most of these correspond to gene bodies (33,294; 51.8%) and intergenic regions (26,947; 38.8%), and only a fraction reside in gene promoters (9,061; 13%). Interestingly, however, an increased representation of intergenic regions (3,194; 46%; $P = 2^{-28}$) and decreased frequency of promoters (144; 2%; $P = 4.8^{-148}$) is observed in ATAC-seq regions that display variable accessibility through T-cell development stages, suggesting that dynamic control of accessibility at distal regulatory elements may influence thymocyte development. Hierarchical clustering analysis revealed distinct groups of differentially accessible regions that closely clustered thymocyte DN1 and DN2a populations, separate from DN2b and DN3 cells, and DN4, ISP and DP thymocytes distinct from CD4SP and CD8SP populations (**Fig. 1A**). Consensus clustering further highlighted developmental transitions between DN1, DN2a and DN2b cells; placed DN3 closer to the DN4, ISP and DP thymocyte cluster; and distinguished CD4SP and CD8SP cells (**Fig. 1B**). In these analyses, the transition from DN1-DN2a to DN2b, which marks T-cell specification, is associated with marked loss of chromatin accessibility consistent with a restriction of transcriptional potential from uncommitted populations to T-cell progenitors (**Fig. 1A**). Moreover, among the four major differential chromatin accessibility developmental modules, the cluster characterized by high levels of chromatin accessibility in DN1-

DN2a cells accounted for 4,763 (68%) of all differentially accessible segments (**Fig. 1A**). A second cluster composed of 684 (9.8%) segments show orchestrated opening during T-cell specification in DN2b and DN3 cells (**Fig. 1A**). This is followed by the opening of 439 intervals (6.3%) characteristically accessible in DN4-ISP-DP populations and, subsequently, 1,044 intervals (15%) selectively open in mature CD4SP and CD8SP cells (**Fig. 1A**). These results demonstrate a highly dynamic chromatin remodeling landscape during thymocyte development, particularly at non-promoter regulatory regions with discrete clusters of differentially accessible regions controlled by distinct regulatory circuitries. Consistently, transcription factor binding site analyses identified significantly enriched regulatory sites in each of these clusters with prominent representation of PU-box, GATA, Runt-related (RUNX), homeodomain (HOX), helix-loop-helix, ETS, Forkhead-box (FOX) and Krüppel-like (KRAB) transcription factor binding motifs (**Fig. 1C** and **Supplementary Table S1**).

N-Me is a regulatory hub for MYC expression in T-ALL

MYC, a master regulator of cell growth and proliferation in development and cancer, is transcriptionally controlled by a complex array of long-range regulatory elements with tissue and cell-type specific enhancer activities (17). *Myc* expression in developing T-cells is controlled by the NOTCH1-*Myc*-enhancer (N-Me), a discrete long-range enhancer located 1.4 Mb downstream of *Myc* (11,18). Given the importance of *Myc* expression in lymphocyte biology, we examined the regulatory logic and mechanisms responsible for dynamic N-Me regulation during thymocyte development (11,18,19).

Circularized Chromosome Conformation Capture (4C) analyses of NOTCH1-driven human and mouse T-ALL lymphoblasts, where the N-Me enhancer is active (11), confirmed the long range-interaction between the *MYC* proximal promoter and N-Me, but also revealed unanticipated chromatin loops connecting both the *MYC* promoter and the N-Me enhancer with distal elements located centromeric and telomeric from the *MYC* transcription start site, implying a more complex mechanism of transcriptional regulation (**Fig. 2A** and **Supplementary Fig. S1**). Chromatin immunoprecipitation and sequencing (ChIPseq) data showed binding of CTCF and MED1, two factors involved in chromatin-

chromatin interactions, to the N-Me site (**Fig. 2B**). We also observed high densities of chromatin marks characteristic of an active enhancer configuration in the vicinity of N-Me, such as histone H3-K4 monomethylation (H3K4me1) and H3-K27 acetylation (H3K27ac) (11). In addition, analysis of chromatin-associated factors revealed N-Me occupancy by BRD4, a reader of H3K27ac, and KDM6A, which acts to erase histone H3-K27 trimethylation (H3K27me3) (**Fig. 2B**). Moreover, ChIPseq analysis of T-ALL leukemia lymphoblasts showed that, in addition to the expected occupancy of N-Me by NOTCH1 and the RBPJ NOTCH1 DNA-binding partner (11,18), this enhancer is also bound by numerous other transcription factors involved in hematopoietic and lymphoid development, including ERG, ETS1, GATA3, RUNX1, TCF3 and TCF12 (20,21), and by *bona fide* transcription factor oncoproteins with prominent roles in T-ALL pathogenesis, such as HOXA9, MYB, MYC, LMO1, LMO2, TAL1 and TLX1 (22) (**Fig. 2B**). In addition, mass spectrometry analysis of N-Me pull-down preparations identified 25 high-confidence N-Me-associated proteins in nuclear extracts from several T-ALL cell lines (HPB-ALL, ALL-SIL and JURKAT), and 17 additional factors associated with N-Me in at least two of these lines (**Fig. 2C**). Of note, N-Me DNA pulled down developmentally important and ChIP-validated N-Me associated transcription factors (RUNX1, GATA3, TCF3, TCF12 and MYC), as well as numerous additional DNA-binding proteins with major roles in thymocyte development (BCL11B, LEF1, RUNX3, CUX1, CBFA2T3 and IKZF1) (20,21).

N-Me GATA site motifs control thymocyte development

The landscape of chromatin interactions and DNA binding factors associated with the N-Me enhancer suggests that N-Me acts as a regulatory hub that receives multiple developmental and oncogenic cues to control *MYC* expression in T-ALL. Despite this apparent complexity, we hypothesized that N-Me activity would be governed by highly conserved cis-acting elements. Indeed, multispecies conservation analysis of N-Me sequences across vertebrates revealed a tight clustering of enhancers from placental mammals and a more distant relationship of these with the N-Me sequences of marsupials and monotremes, which clustered closer to those of reptilian and avian species (**Fig. 2D**). Moreover,

phylogenetic footprint analyses of vertebrate N-Me enhancer sequences (**Fig. 2E** and **Supplementary Table S2**) revealed multiple highly conserved regulatory elements present in mammals, birds and reptiles. Given the strict requirement of both *Myc* expression and N-Me activity for early in *Myc*-driven highly proliferative DN3, DN4 and ISP thymocytes (11), we tested the impact of targeted mutations at these highly conserved motifs on mouse T-cell development (**Fig. 2E**, **Supplementary Fig. S2A** and **Supplementary Table S3**). In these analyses, mutant mice harboring homozygous disruption of two independent highly-conserved N-Me HOX-PBX motifs, a RUNX binding site, a LEF/TCF binding site and a PAX binding site showed no clear alterations in the distributions of thymocyte populations (**Supplementary Fig. S2B** and **S2C**). Similarly, mice carrying a genetic disruption of a highly conserved RBPJ binding site showed no clear phenotypical alterations in thymus (**Supplementary Fig. S2B** and **S2C**), despite effective abrogation of NOTCH1 binding to DNA induced by this mutation (**Supplementary Fig. S2D**). These results support a redundant role for multiple transcription factor binding sites in the control of N-Me activity and *Myc* expression in the thymus. In contrast, mutation of an ultraconserved GATA transcription factor binding motif (hereafter named GATA site 1, GS1) (**Supplementary Fig. S3A**) revealed an accumulation of DN2, DN3, DN4 and ISP populations with preserved DP and SP cells ($P < 0.05$) (**Fig. 3A-E** and **Supplementary Fig. S2B** and **S2C**), while homozygous disruption of a second highly conserved GATA-binding motif (hereafter named GATA site 2, GS2), across all mammalian species, showed a more modest phenotype with accumulation of DN3 and ISP populations with otherwise largely preserved thymocyte development (**Fig. 3A-E** and **Supplementary Fig. S2B** and **S2C**). In line with these findings, N-Me-MYC promoter luciferase reporter assays showed a marked reduction in transcriptional activity in a GS1-mutant N-Me construct compared with the wild-type N-Me control, and a more moderate decrease in enhancer function in the reporter containing a mutation in the less conserved N-Me GS2 site (**Fig. 4A**). Based on these results, we generated mice homozygous for combined mutations in these two regulatory elements. Notably, N-Me GS1- and GS2-deficient mice (hereafter named GS1+2mut) showed a marked early T-cell developmental defect with small thymi and a dramatic reduction in thymocyte numbers (**Fig. 3A** and

3B) accompanied by accumulation of DN3, DN4 and ISP cells, and decreased numbers of DP and SP CD4 and CD8 thymocytes (**Fig. 3C-E**). Lack of apoptosis in this model supports a developmental block phenotype (**Supplementary Fig. S3B**). Moreover, analysis of spleen and lymph nodes from GS1+2mut mice showed a consistent reduction in the numbers of mature CD4⁺ and CD8⁺ T-cell populations (**Supplementary Fig. S3C and S3D**). These results support a cooperative and partially redundant role of the GS1 and GS2 N-Me GATA sites in the control of N-Me activity and *Myc* expression during thymocyte development.

To test this hypothesis, we performed RNAseq analysis of wild type and GS1+2 mutant DN3 thymocytes. These analyses revealed an effective down-regulation of *Myc* messenger RNA and *Myc*-controlled gene expression programs in GS1+2 mutant DN3 cells compared with wild type controls (**Fig. 4B and 4C**). Moreover, single cell RNAseq analyses showed decreased numbers of *Myc*-expressing cells in the thymus of homozygous N-Me GS1+2mut animals (**Fig. 4D and 4E**). Similarly, immunohistochemical and flow cytometry analysis revealed markedly reduced numbers of *Myc*-positive cells in the thymi of N-Me GS1+2 homozygous mutant mice, and reduced *Myc* levels in the DN4 and ISP populations compared with wild type controls (**Fig. 4F and 4G**).

N-Me GATA sites are required for NOTCH1 induced leukemia

Next we evaluated the impact of N-Me GS1+2 mutations in T-ALL transformation. Towards this goal, we infected hematopoietic progenitors from N-Me wild type (N-Me^{+/+}), N-Me GS1+2 heterozygous (N-Me^{+/GS1+2}) and N-Me GS1+2 homozygous (N-Me^{GS1+2/GS1+2}) animals with retroviruses driving the expression of an oncogenic constitutively active form of NOTCH1 (Δ E-NOTCH1), which specifically induces development of MYC-driven T-ALL in mice, and transplanted them into sublethally irradiated recipients (11,23). Consistent with previous reports (11), mice transplanted with Δ E-NOTCH1-infected N-Me wild type cells showed a transient wave of CD4 CD8 double-positive cells in peripheral blood at 21 days post-transplant (**Fig. 5A**) and developed overt T-ALL with a median latency of 8 weeks (**Fig. 5B and 5C**). In contrast, mice transplanted with Δ E-NOTCH1-expressing N-Me GS1+2mut

heterozygous progenitors showed a blunted DP wave and impaired tumor development ($P < 0.01$) (**Fig. 5A-C**), while animals transplanted with ΔE -NOTCH1 N-Me GS1+2 homozygous mutant cells remained leukemia free 100 days post-transplant ($P < 0.001$) (**Fig. 5A-C**). To explore the requirement for Gata3 N-Me binding in tumor progression and maintenance, we analyzed the capacity of the N-Me GS1+2 mutant allele to support cell growth and proliferation in established NOTCH1-induced leukemia. To this end, we generated ΔE -NOTCH1-induced T-ALL tumors from mice expressing tamoxifen-inducible Cre recombinase ($Rosa26^{TMCre}$; $Rosa26^{+/CreERT2}$) that are also compound heterozygous for the conditional N-Me knockout ($N-Me^{flox}$) (11) and the N-Me GS1+2 mutant alleles ($Rosa26^{+/CreERT2}$ $N-Me^{flox/GS1+2mut}$). Following leukemia development, we injected T-ALL cells into secondary recipients and then treated these animals with vehicle only ($N-Me^{flox/GS1+2mut}$ group) or tamoxifen, to induce Cre-mediated deletion of the $N-Me^{flox}$ allele ($N-Me^{-/GS1+2mut}$ group). Vehicle-treated mice bearing compound heterozygous N-Me GS1+2 leukemia cells ($N-Me^{flox/GS1+2mut}$) developed overt leukemia and died of disease with a median survival of 33 days. In contrast, tamoxifen treated animals harboring isogenic hemizygous N-Me GS1+2 mutant cells ($N-Me^{-/GS1+2mut}$) showed a markedly impaired tumor progression with a median survival of 53 days ($P < 0.0001$) (**Fig. 5D**). Consistently, analysis of $Rosa26^{+/CreERT2}$ $N-Me^{flox/GS1+2mut}$ leukemia lymphoblasts treated with vehicle only ($N-Me^{flox/GS1+2mut}$) or tamoxifen ($N-Me^{-/GS1+2mut}$) showed reduced *Myc* expression and markedly impaired growth, proliferation and survival in N-Me GS1+2 hemizygous leukemia cells (tamoxifen treated, $N-Me^{-/GS1+2mut}$), compared with isogenic N-Me GS1+2 heterozygous (vehicle treated, $N-Me^{flox/GS1+2mut}$) controls (**Fig. 5E-H**). In all, these results demonstrate a strict requirement for N-Me GATA site-mediated enhancer activity in the pathogenesis of T-ALL.

Gata3 mediates nucleosome eviction at the N-Me enhancer

GATA3, a GATA binding transcription factor upregulated during early stages of thymocyte maturation and critically implicated in T-cell development (20,21,24-26), prominently interacts with the N-Me enhancer (**Fig. 2B** and **2C**). To evaluate the potential role of Gata3 as driver of N-Me activity and effector factor mediating N-Me GATA site function, we analyzed the representation of wild type and

GATA site mutant sequences in Gata3 ChIP DNA isolated from heterozygous GS1+2mut DN3 thymocytes (**Fig. 6A**). In these experiments Gata3 ChIP DNA preparations contained only wild type N-Me sequences in support of complete disruption of Gata3 binding to GS1+2mut N-Me (**Fig. 6A**). In contrast, analysis of Gata3 ChIPs from heterozygous GS1mut and GS2mut DN3 cells showed only a partial reduction in Gata3 binding to N-Me single GATA site mutant DNA (**Fig. 6A**), a result in agreement with the observed partial redundancy of the GS1 and GS2 GATA motifs in *Myc* regulation and thymocyte development.

Active enhancers characteristically correspond to regions of low nucleosome occupancy (7,27,28) flanked by areas of high density of nucleosomes containing H3K4me1 and H3K27ac (29,30). To evaluate the potential impact of N-Me GATA site mutations and impaired Gata3 occupancy in the epigenetic landscape surrounding the N-Me enhancer, we performed ChIP for the H3K4me1 and H3K27ac chromatin marks in DN3 thymocytes from heterozygous GS1+2 mutant mice and evaluated the recovery of the N-Me mutant over wild type enhancer by DNA sequencing as before. These analyses revealed no apparent difference in the recovery of active enhancer-associated chromatin marks in N-Me GS1+2 mutant chromosomes over wild type (**Fig. 6B**), suggesting that Gata3 N-Me binding is not required for the establishment and maintenance of active enhancer histone marks. In addition, and to further explore the mechanisms responsible for the loss of effective N-Me-mediated transcriptional activity in N-Me Gata3-binding deficient thymocytes, we evaluated the effect of GATA site mutations in the establishment and maintenance of N-Me-*Myc* promoter chromatin loops by interphase fluorescent in situ hybridization (FISH) using DNA probes mapping to the *Myc* promoter region and the N-Me enhancer. Analysis of N-Me wild type cells showed probe co-localization indicative of effective interaction between N-me and the *Myc* promoter in DN3 thymocytes, but not in B cells, where the N-Me enhancer is not active (**Fig. 6C**). In addition, and most notably, we observed a similar pattern of probe co-localization in DN3 cells from homozygous N-Me GS1+2 mutant mice, ruling out a role for Gata3 N-Me binding in N-Me enhancer-*Myc* promoter long-range chromatin looping (**Fig. 6C**). To further explore a potential role of changes in chromatin looping in the loss of N-Me activity in GS1+2

mutant cells, and to rule out a potential confounding effect of secondary effects derived from low *Myc* expression in GS1+2 mutant thymocytes, we crossed N-Me GS1+2 mutant mice with tamoxifen inducible *Rosa26-Cre-ERT2* and *Rosa26-LSL-Myc* knock-in animals, generating in this way *Rosa26-Cre-ERT2 Rosa26-LSL-Myc N-Me^{GS1+2/GS1+2}* mice, which are defective in Gata3 N-Me binding, but have the capacity to express *Myc* ectopically from the *Rosa26* locus after tamoxifen treatment. Hematopoietic progenitors from this model were infected with oncogenic Δ E-NOTCH1 expressing retroviruses and treated with tamoxifen to generate T-ALL N-Me GS1+2 mutant *Rosa26-Myc* tumors. Chromatin configuration by Chromatin Conformation Capture (3C) analysis in N-Me GS1+2 *Rosa26-Myc* T-ALL lymphoblasts and N-Me wild type T-ALL controls confirmed effective interaction between the N-Me enhancer and *Myc* promoter sequences in GS1+2 mutant cells (**Fig. 6D**).

Given the broad and dynamic changes in enhancer accessibility observed during thymocyte development we evaluated if regulation of chromatin accessibility at the N-Me enhancer could function in the control *Myc* expression. Analysis of ATAC-seq data obtained at different stages of thymocyte development revealed a closed N-Me enhancer configuration in early DN1 and DN2a progenitors; acquisition of an open chromatin conformation in DN2b cells; high levels of chromatin accessibility in *Myc*-expressing DN3, DN4 and ISP populations; and a closed chromatin configuration in DP thymocytes and mature single positive CD4 and CD8 T cells, which express low levels of *Myc* (31) (**Fig. 7A and 7B**). These results place N-Me as part of the cluster of regulatory sites gaining accessibility as cells commit to the T-cell lineage, which is characteristically enrichment in GATA site motifs (239/684; 35%) suggesting that N-Me accessibility in the thymus could be controlled by a GATA transcription factor. To test this hypothesis, we performed ATAC-seq in sorted wild type and homozygous N-Me GS1+2 mutant DN3 cells. These analyses revealed a marked decrease in chromatin accessibility in N-Me GS1+2 mutant thymocytes compared with wild type controls (**Fig. 7C**). Moreover, nucleosome position analysis demonstrated a nucleosome exclusion area at the N-Me enhancer flanked by two regions of prominent nucleosome occupancy in wild type DN3 thymocytes (**Fig. 7D**). In contrast, and consistent with decreased chromatin accessibility, N-Me GS1+2 mutant DN3 cells showed prominent

nucleosome invasion and a consequent marked reduction in the nucleosome-free region surrounding the GS1 and GS2 GATA sites (**Fig. 7D**). Notably, ATAC-seq analyses in N-Me GS1+2 mutant *Rosa26-Myc* cells show similar loss of chromatin accessibility and nucleosome invasion of the N-Me enhancer (**Supplementary Fig. S4A and S4B**) supporting that loss of enhancer accessibility at the N-Me site is linked to loss of Gata3 binding and not an indirect effect resulting of decreased Myc expression. Moreover, Gata3 protein enhanced DNase I digestion of N-Me histone H1-compacted nucleosome arrays (32), further supporting a role as pioneer factor promoting N-Me enhancer accessibility (**Supplementary Fig. S4C and S4D**).

Mechanistically, abrogation of Gata3 binding in N-me GS1+2 mutant cells resulted in loss of N-Me binding by the Smarca4 SWI/SNF core factor, a result consistent with a role of Gata3 in promoting enhancer chromatin opening via the recruitment of chromatin remodeling complexes implicated in nucleosome repositioning (**Fig. 7E**). In this context, we predicted that nucleosome invasion could result in broad abrogation of transcription factor binding to N-Me. Chromatin immunoprecipitation analyses in GS1+2 mutant *Rosa26-Myc* T-ALL cells revealed that Gata site mutations impaired transcription factor binding for Runx1, Ets1, Notch1, Tcf1, Tcf3 and Tcf12 in support of a defect in multi-transcription factor combinatorial regulation at this enhancer (**Fig. 7F**). In all, these results demonstrate a driving role for Gata3 binding in promoting N-Me enhancer nucleosome eviction, and a distinct and strict requirement for enhancer accessibility in the control of N-Me enhancer activity and *Myc* expression in immature T-cells.

Aberrant N-Me enhancer accessibility in T-ALL

Distinct clinico-biological groups of T-ALL leukemia show parallels in their immunophenotypes and gene expression signatures with those of immature thymocytes at different stages of development supporting broadly common developmental and transcriptional regulatory circuitries between tumor cells and their normal cell counterparts (33). However, analysis of mouse enhancer accessibility at the N-Me enhancer in normal DP thymocytes and in DP T-ALL cells showed a closed enhancer

configuration in normal DP thymocytes and an open conformation in DP T-ALL lymphoblasts (**Fig. 7G**). ATAC-seq analysis also revealed markedly divergent profiles in human normal and leukemic DP cells, with normal DP thymocytes showing a closed chromatin N-Me profile and human DP T-ALL lymphoblasts displaying high levels of chromatin enhancer accessibility (**Fig. 7H**). These results support a potential oncogenic role for deregulated N-Me enhancer accessibility as driver of *MYC* expression in T-ALL.

Discussion

Enhancers represent distal transcription regulatory elements with high density of transcription factor binding sites and the capacity to interact with proximal regulatory elements in the vicinity of transcription start sites via long range chromatin loops (34,35). The activity of enhancers is dependent on multiple transcription factors with additive and cooperative roles in transcription regulation, which together establish a combinatorial logic responsible for tuning transcriptional control in response to developmental and signaling cues (36). Transcription factor binding motif redundancy and overlapping control by multiple transcription factors are considered key enhancer features responsible for robust and fine-tuned transcriptional regulation (37). In addition, enhancer activity depends on chromatin looping and long range chromatin interactions and is closely associated with the presence of a distinct pattern of chromatin marks with low levels of H3-K27 trimethylation, and high density of H3-K4 monomethylation and H3-K27 acetylation (38). Moreover, active enhancers are highly accessible chromatin regions with highly mobile H2A.Z-containing nucleosomes (39) and low nucleosome occupancy (6). Pioneer transcription factors capable of binding to their target sequences on nucleosomal DNA and eliciting changes in local chromatin structure are particularly important to license enhancer activity during development (40).

By performing global analysis of ATAC-seq data in early T-cell progenitors we observed dynamic changes in enhancer accessibility consistent with a prominent role of pioneer factors during thymocyte development. Moreover, the presence of distinct enrichment of transcription factor binding motifs in

enhancers with coordinated opening at specific stages of thymocyte development supports the presence of a combinatorial logic in enhancer regulation. Among these highly dynamically regulated regulatory elements, N-Me a long-range *Myc* enhancer, is prominently dependent on NOTCH1 signaling (11). Consistent with its role as a long-range *Myc* enhancer, N-Me directly interacts with *Myc* promoter sequences (11). However, extended mapping of chromatin interactions by 4C shows a more extensive interaction landscape involving multiple intergenic sites conserved between human and mouse of potential functional importance. Moreover, while NOTCH1 is active in multiple tissues during development, N-Me is selectively licensed and active only in early T-cell progenitors, suggesting a role for additional, T-cell specific, transcriptional regulators.

N-Me is highly conserved with multiple transcription factor binding sites commonly present among reptiles, avian species and mammals in support for a combinatorial multi-transcription factor mode of enhancer regulation. However, analysis of mice harboring deleterious mutations in multiple highly conserved transcription factor binding motifs showed no apparent defects in thymocyte development, a result consistent with transcription factor binding site redundancy and overlapping transcription factor activities in enhancer regulation. In contrast, while mice harboring mutations in one of two N-Me Gata3 binding sites showed only mild developmental defects, mutation of these two GATA motifs in cis resulted in abrogation of N-Me enhancer activity and marked defects in thymocyte development. These results indicate cooperation between these two GATA sites and a prominent role for Gata3 in N-Me regulation. It is worth noting that N-Me Gata3-binding motifs appeared asynchronously during evolution as the GS1 site can be found in N-Me sequences in turtle and alligator species, dating back 250 million years; while the GS2 site is present only in mammals and seems to have emerged early in monotreme development 170 million years ago. It has not escaped our attention that some of the developmental phenotypes in GS1+2 mutant mice are slightly less prominent than those observed in the thymus of mice harboring a full N-Me enhancer knock-out (11), a result that indicates at least partial functional overlap with other N-Me-binding transcriptional regulators. Consistent with this hypothesis, reverse ChIP and ChIP-seq data, indicates prominent transcription factor occupancy at the N-Me enhancer and

points to this regulatory element as a major node controlling *Myc* expression with input from multiple developmentally important transcription factors including NOTCH1, and from T-ALL transcription factor oncogenes activated by chromosomal translocations (TAL1, LMO1, LMO2, MYB and TLX1).

Importantly, functional characterization of mechanisms involved in loss of enhancer activity in N-Me GATA site mutant thymocytes and T-ALL cells showed loss of chromatin accessibility and enhancer nucleosome invasion, but preserved active enhancer chromatin marks and chromatin looping between N-Me and *Myc* promoter sequences. Loss of Gata3 binding in N-Me is linked to defective occupancy by Smarca4, a core subunit of SWI/SNF chromatin remodeling complexes mediating an open chromatin configuration. Moreover, nucleosome invasion secondary to Gata3 binding defects results in broad abrogation of enhancer binding by multiple other transcription factors. The model that emerges from these observations is that Gata3 binding facilitates the recruitment of SWI/SNF chromatin remodeling complexes to N-Me and promotes nucleosome eviction in early T-cell progenitor cells, which in turn enables multi-transcription factor control of *Myc* expression. Consistently, Gata3 binding and N-Me enhancer accessibility are epistatic over NOTCH1 signaling in T-cell transformation. Finally, the presence of an aberrant open chromatin configuration at the N-Me enhancer in DP T-ALL cells supports a role for deregulated *MYC* enhancer accessibility in the pathogenesis of T-ALL.

Materials and methods

Patient samples

DNA from leukemic ALL blasts were provided by the Princess Máxima Center for Pediatric Oncology tumor bank. Normal thymus samples were obtained from Morgan Stanley Children's New York Presbyterian Hospital. Written informed consent was obtained at study entry and samples were collected under the supervision of local Institutional Review Boards for participating institutions and analyzed under the supervision of the Columbia University Medical Center Institutional Review Board (Protocol Numbers: IRB-AAAB3250 and IRB-AAAC1660).

Cell Culture

We performed cell culture of cell lines in standard conditions in a humidified atmosphere at 37°C under 5% CO₂. We obtained ALL-SIL (RRID: CVCL_1805), HPB-ALL (RRID: CVCL_1820) and JURKAT (RRID: CVCL_0065) T-ALL cell lines from the Deutsche Sammlung von Mikroorganismen und Zellkulturen (DSMZ) cell line repository. We purchased HEK293T (RRID: CVCL_006) cells from American Type Culture Collection (ATCC).

We cultured primary mouse lineage negative bone marrow cells in Opti-MEM media (Life Technologies 51985091) supplemented with 10% FBS, 100 U mL⁻¹ penicillin G, 100 µg mL⁻¹ streptomycin, 55 µM β-mercaptoethanol, 10 ng mL⁻¹ IL3 (PeproTech 213-13), 10 ng mL⁻¹ IL6 (PeproTech 216-16), 25 ng mL⁻¹ IL7 (PeproTech 217-17), 50 ng mL⁻¹ SCF (PeproTech 250-03), and 50 ng mL⁻¹ Flt3L (PeproTech 250-31L). Primary mouse leukemia lymphoblasts in Opti-MEM media supplemented with 10% FBS, 100 U mL⁻¹ penicillin G, 100 µg mL⁻¹ streptomycin, 55 µM β-mercaptoethanol and 10 ng mL⁻¹ murine IL7 (PeproTech 217-17).

Generation of N-Me transcription factor binding site mutant mice

We generated N-Me mutant mice at the Herbert Irving Comprehensive Cancer Center Transgenic Shared Resource (41) by injecting fertilized eggs from B6CBAF1 (GATA site 1, GATA site 2 and RBPJ site) or B6 (RUNX site, LEF/TCF site, PAX site, HOX site 1, and HOX site 2) females with Cas9 mRNA (TriLink Biotechnologies L-6125) and the corresponding sgRNA in each case (Synthego, Menlo Park, CA) and transferring them into the oviducts of Swiss Webster foster females.

We performed mouse genotyping by N-Me PCR amplification (GATA and RBPJ sites: forward primer 5'-GTGAAAAATTACAAGGATGGG-3' and reverse primer 5'-CATCAGAGTAGAGTACAGTGC-3'; HOX, RUNX, LEF/TCF and PAX sites: forward primer 5'-GACCTTTGCTGCACTTGCATC-3' and reverse primer 5'-TGACACAATCACCAGGTTTCAG-3') and Sanger DNA sequencing at Genewiz.

Animal procedures

All animals were maintained in specific pathogen-free facilities at the Irving Cancer Research Center at Columbia University Medical Campus. The Columbia University Institutional Animal Care and Use Committee (IACUC) approved all animal procedures. Animal experiments were conducted in compliance with all relevant ethical regulations.

To generate NOTCH1-induced T-ALL in mice, we infected lineage negative enriched cells from bone marrow of N-Me wild type, GS1+2 heterozygous mutant and GS1+2 homozygous mutant donors with retroviral particles expressing oncogenic NOTCH1 (ΔE -NOTCH1) (23) and the green fluorescent protein (GFP) as previously described (11,42), and transplanted them in sublethally irradiated (500 cGy) NRG mice (RRID: IMSR_JAX:007799).

We crossed N-Me GS1+2 mutant mice (N-Me^{+/GS1+2mut}), N-Me conditional knockout (N-Me^{+/flox}; Rr38^{+/tm1.1Aafo}) (11) and Rosa26^{TM-Cre} (Cre-ERT2) mice (43) to generate NOTCH1-induced leukemia from resulting Rosa26^{+/CreERT2} N-Me^{flox/GS1+2mut} mice as before (11,42). We transplanted lymphoblasts from spleens of diseased animals into secondary hosts and treated mice with 3 mg of tamoxifen or with corn oil vehicle by intraperitoneal injection two days after transplant and every five days thereafter.

To generate GS1+2 mutant Myc-rescued tumors, we crossed N-Me GS1+2 homozygous mutant mice, Rosa26^{TM-Cre} mice and Rosa26^{StopFLMYC} (LSL-Myc) mice (44). We infected lineage negative enriched cells from bone marrow of Rosa26^{CreERT2/LSL-Myc} N-Me^{GS1+2mut/GS1+2mut} donors with oncogenic NOTCH1 and transplanted them in NRG mice as described before. We treated mice with 3 mg of tamoxifen by intraperitoneal injection one week after transplant.

Circularized Chromosome Conformation Capture (4C)

We performed 4C analysis in Jurkat cells and mouse primary T-ALL lymphoblast as described before (45), using the restriction enzymes *HindIII* and *DpnII*. We constructed sequencing libraries from 4C DNA including barcoded Illumina adapters to the 5' end of each PCR primer (human *MYC* promoter viewpoint: forward primer – 5'-AGACGTGGGGGCTAAAGCTT-3' and reverse primer 5'-TGGGTATTTGGTTTGGCCTAT-3'; human N-Me viewpoint – forward primer 5'-

CCAAAGTACCCTACAAGCTT-3' and reverse primer 5'-GCTGACAGTTGTTAGCAGGG-3'; mouse *Myc* promoter viewpoint: forward primer 5'-TAAAGGATGACCGGAAGCTT-3' and reverse primer 5'-GGGAGTAGATGAACCCATCC-3'; mouse N-Me viewpoint – forward primer 5'-CTAATTTATTTTCTAAGCTT-3' and reverse primer 5'-CATGAAATTCCATTGCTTCAG-3'). We sequenced pooled libraries using a HiSeq 2500 sequencer (Illumina).

We analyzed 4C sequencing results using 4Cseqpipe (46). Samples containing less than 0.6 million mapped reads (following removal of undigested and self-ligated fragments) were discarded. Additionally, samples with fewer than 0.2 cis/trans ratio of mapped reads and with fewer than 40% read coverage for all HindIII sites in the 1 Mb surrounding the bait region were discarded.

Enhancer pull down assays (Reverse ChIP)

We performed reverse ChIP assays as previously described (47). Briefly, we generated N-Me DNA bait sequences by PCR from human genomic DNA using a N-Me biotinylated forward primer (5'-CCCTAATTTCTATCCCCACTGTC-3') and an unmodified N-Me reverse primer (5'-ATTTTTTTCCTGTAAATATGCTGTAC-3'). Then, we conjugated DNA baits to streptavidin beads and incubated them with nuclear protein extracts from ALL-SIL, HPB-ALL, or JURKAT cells. We used non-conjugated beads as negative control. N-Me pulled down proteins were analyzed by mass spectrometry at the Proteomics Laboratory at the New York University School of Medicine. The MS/MS spectra were searched against the Uniprot human reference proteome database using Sequest within Proteome Discoverer. A 1% False discovery rate (FDR) cut off was applied on the peptide level using a standard target-decoy database strategy. All proteins identified with less than two unique peptides were excluded from analysis. Thus, we recovered 362 proteins from HPB-ALL extracts, 321 proteins from ALL-SIL extracts, and 219 from JURKAT extracts. We normalized data to background signal (KRT1) and filtered against signal-to-noise ratio (protein signal / KRT1 signal > 0.30) and the Contaminant Repository for Affinity Purification (CRAPome) (48) contaminant list ($\leq 10\%$ of the CRAPome). We identified 79 N-Me

associated proteins in HPB-ALL extracts, 50 in ALL-SIL extracts, and 38 in JURKAT extracts. Proteins identified in all three cell line extracts were considered high confidence N-Me associated proteins.

For site-specific reverse ChIP of RUNX and RBPJ binding sites, we annealed biotinylated complementary primers encompassing the binding motifs and their mutated forms (wild type RUNX site 5'- TGAGATGATCAGTTTTACCACAGTTCACTACACTC-3', mutant RUNX site 5'- TGAGATGATCAGTTTTAgactAGTTCACTACACTC-3', wild type RBPJ site 5'- CAGAGATGGGGTTCCTCAGGGTGTTCAGGG-3', mutant RBPJ site 5'- CAGAGATGGGGTTCgCgtGGGTGTTCAGGG-3'). We analyzed RUNX1 and activated NOTCH1 binding to wild type and mutated sequences by western blot using RUNX1 (Santa Cruz 365644, RRID: AB_10843207) and activated NOTCH1 (Cell Signaling Technologies 2421, RRID:AB_2314204) antibodies respectively.

N-Me evolutionary conservation analysis

We analyzed evolutionary conservation of N-Me sequences (**Supplementary Table S2**) using the DiAlign TF tool (RRID: SCR_008036) (49). We aligned N-Me sequences from 28 vertebrate species and identified putative transcription factor binding sites in highly conserved regions. We calculated evolutionary conservation scores (phyloP score) at individual alignment sites using Ion Reporter Software (50). We generated a phylogenetic tree using the iTOL tool (51).

Single guide RNA design for N-Me transcription factor binding site targeting

We designed single guide RNAs (sgRNAs) overlapping the N-Me transcription factor binding sites of interest and evaluated for potential off-targets using E-CRISPR (Deutsches Krebsforschungszentrum) (52). SgRNA sequences can be found in **Supplementary Figure S2**.

Flow cytometry analyses and cell sorting

All flow cytometry data were collected on a FACSCanto II flow cytometer (BD Biosciences) using FACSDiva software (BD Biosciences, RRID: SCR_001456) and analyzed with FlowJo software (Tree Star, RRID: SCR_008520).

To analyze thymic populations, we stained single cell suspensions of thymocytes with a lineage marker biotinylated antibody cocktail against CD11b, Gr1, NK1.1, Ter119, CD19 and B220, and then with fluorochrome-conjugated streptavidin and antibodies against CD3e, CD4, CD8a, CD25 and CD44 (**Supplementary Table S4**). Lineage negative cells were represented in a CD4 versus CD8a plot and CD4/CD8 double positive (DP, CD4⁺ CD8a⁺) and CD4 single positive (CD4SP, CD4⁺ CD8a⁻) populations were gated. Then CD4/CD8 double negative (DN) cells were plotted in a CD44 versus CD25 plot and CD8 positive cells in a CD3 histogram, to characterize double negative 1 (DN1, CD4⁻ CD8a⁻ CD44⁺ CD25⁻), 2 (DN2, CD4⁻ CD8a⁻ CD44⁺ CD25⁺), 3 (DN3, CD4⁻ CD8a⁻ CD44⁻ CD25⁺), and 4 (DN4, CD4⁻ CD8a⁻ CD44⁻ CD25⁻) populations, and intermediate single positive (ISP, CD4⁻ CD8a⁺ CD3e⁻) and CD8 single positive (CD8SP, CD4⁻ CD8a⁺ CD3e⁺) populations respectively.

To analyze mature T-cell populations in peripheral lymphoid tissues, we stained single cell suspensions of spleen and lymph nodes with antibodies against CD4 and CD8a (**Supplementary Table S4**).

For Myc intracellular staining, we fixed and permeabilized membrane marker-labeled cells using Fixation/Permeabilization Solution Kit (BD Biosciences 554714) and stained them with a fluorochrome-conjugated anti-Myc antibody (**Supplementary Table S4**).

In tumor generation experiments, we stained bone marrow infected cell preparations with anti-Sca1 antibody (eBioscience 17-5981, RRID: AB_469488) and evaluated infection efficiency by assessment of the percentage of GFP⁺ Sca1⁺ cells. To analyze the emergence of the post-transplant CD4⁺ CD8⁺ double positive wave, we bled mice 21 days after transplant, lysed erythrocytes, and stained white blood cells with antibodies against CD4 (BD Pharmingen 553051, RRID: AB_398528) and CD8a (eBioscience 25-0081, RRID: AB_469584).

For apoptosis and cell cycle analysis, we harvested primary mouse tumor cells from the spleen of leukemic mice and cultured them with vehicle only (ethanol) or (Z)-4-hydroxytamoxifen for 3 days. To analyze apoptosis, we stained 10^5 cells with APC-conjugated Annexin-V (BD Biosciences 550475) in Annexin-V Binding Buffer (BD Biosciences 556454) for 15 minutes at room temperature. Then, we incubated cells with $5 \mu\text{g mL}^{-1}$ of DAPI (Invitrogen D3571) and determined the percentages of live (Annexin-V⁻ DAPI⁻), apoptotic (Annexin-V⁺ DAPI⁻), and dead cell (DAPI⁺) by flow cytometry. To analyze cell cycle, we stained 10^6 cells with $5 \mu\text{g mL}^{-1}$ of Hoechst 33342 (Sigma-Aldrich B2261) in complete Opti-MEM media for one hour at 37°C. We then washed the cells and incubated them with 50 ng mL^{-1} of 7AAD (BD Biosciences 559925) as a viability dye, and analyzed cell cycle progression in the GFP⁺ 7AAD⁻ population by flow cytometry.

For isolation of DN3 cells, we enriched CD4⁻ CD8⁻ double negative cells from thymi from 6-8 week old mice by labeling with a lineage marker biotinylated antibody cocktail against CD11b, Gr1, NK1.1, Ter119, CD4, CD8a, CD19 and B220 (**Supplementary Table S4**) followed by magnetic depletion of antibody-labeled cells using streptavidin microbeads (Miltenyi Biotec 130-048-101). We stained non-depleted cells with fluorochrome-conjugated streptavidin and antibodies against CD4, CD8, CD25 and CD44 (**Supplementary Table S4**), and sorted DN3 cells. For isolation of B cells, we stained splenocytes with a fluorochrome-conjugated anti-B220 antibody (Miltenyi 130-102-187, RRID: AB_2660443) and sorted B-cells (B220⁺). All sortings were performed on a SH800 cell sorter (SONY Biotechnologies).

Luciferase reporter assays

We performed luciferase reporter assays using a pBV-Luc – MYC promoter luciferase construct(53) together with a plasmid driving the expression of the Renilla luciferase gene (pCMV-Renilla) used as an internal control in Jurkat cells as described before (11).

Chromatin immunoprecipitation (ChIP)

To perform H3K27 acetylation, H3K4 monomethylation and Gata3 ChIP in GS1+2 heterozygous thymocytes, we crosslinked DN3 cells isolated from 6 week old mice with 1% formaldehyde in PBS for 10 minutes at room temperature. We quenched the reaction by adding glycine up to 0.125 M and incubated it for 5 minutes at room temperature. We performed cell lysis, chromatin shearing, chromatin immunoprecipitation (ChIP) and purification of precipitated chromatin using the Auto iDeal ChIP-seq Kit for Histones x100 (Diagenode C01010171), the Auto IPure Kit v2 x100 (Diagenode C03010010) and the Diagenode Automated Platform SX-8G IP-Star Compact, following the manufacturer's protocol and using antibodies recognizing H3K27 acetylation (Abcam ab4729, RRID: AB_2118291), H3K4 monomethylation (Diagenode C15410194, RRID: AB_2637078), and Gata3 (Cell Signaling Technologies 5852S, RRID: AB_10835690). For differential allele binding analysis, we amplified N-Me DNA by PCR from immunoprecipitated chromatin (forward primer 5'-GTGAAAAATTACAAGGATGGG-3' and reverse primer 5'-CATCAGAGTAGAGTACAGTGC-3') and performed Sanger sequencing of the PCR products at Genewiz.

To perform Smarca4 ChIP from GS1+2 heterozygous thymocytes, we crosslinked DN3 cells with 1mg/mL disuccinimidyl glutarate in PBS for 30 minutes, followed by a second fixation with 1% formaldehyde in PBS for 10 minutes and 0.125 M glycine quenching for 5 minutes. We lysate pelleted nuclei in lysis buffer containing 10mM Tris-HCl (pH 7.5), 0.1% SDS, 1mM EDTA, 0.1% sodium deoxycholate, 1% Triton X-100, 150mM NaCl and protease inhibitor cocktail, and sonicated them on a Bioruptor (Diagenode). Anti-Smarca4 antibody (Abcam ab110641, RRID: AB_10861578) was adsorbed to Dynabeads Protein A/G (Invitrogen), added to the diluted chromatin complex, incubated overnight at 4°C, washed and eluted for one hour at 65°C in ChIP elution buffer containing 100mM NaCO₃ and 1% SDS. We treated eluted samples with RNase A (Invitrogen) and proteinase K (Ambion), and cleaned up chromatin samples using MicroChIP DiaPure columns (Diagenode). We performed differential allele binding analysis as described before.

To analyze the binding of transcription factors to N-Me, we used wild type and GS1+2 Myc-rescued tumor cells and performed ChIP as described before for Smarca4 using antibodies against cleaved Notch1 (Santa Cruz 6014-R, RRID: AB_650335), Ets1 (Santa Cruz 350, RRID: AB_2100688), Runx1 (Abcam ab23980, RRID: AB_2184205), Tcf1 (Santa Cruz 271453, RRID: AB_10649799), Tcf3/E2A (Santa Cruz 349X), and Tcf12/HEB (Cell Signaling 11825, RRID: AB_2797736). We analyzed N-Me enrichment over the input chromatin by quantitative real-time PCR (qRT-PCR) with a QuantStudio 3 Real-Time PCR System (Applied Biosystems) using FastStart Universal SYBR Green (Roche) (forward primer 5'-AACCTGAACCTGGTGATTG-3' and reverse primer 5'-GCCAAGAACTCCTCTGTGCT-3').

Histology

We fixed thymi in 3.7% buffered formalin and embedded them in paraffin using standard procedures at the Herbert Irving Cancer Center Molecular Pathology Core. We stained 5 µm tissue sections with hematoxylin and eosin and performed Myc and Caspase 3 staining following standard procedures (Histowiz). We prepared cell smears from heparinized blood preparations and stained them with May Grünwald (Sigma-Aldrich MG500-500ml) and Giemsa (Sigma-Aldrich GS500-500ml) following standard protocols.

RNA-seq

We extracted RNA from DN3 thymocytes isolated by FACS using RNeasy Micro Kit (Qiagen 74004). We performed RNA library preparations and next generation sequencing using the SMART-Seq v4 Ultra Low Input Kit for Sequencing (Clontech 634888) for full-length cDNA synthesis and amplification at Genewiz. Illumina Nextera XT Library Preparation Kit (Illumina FC-131-1024) was used for library preparation. Sequencing was performed on an Illumina HiSeq platform (Illumina).

We aligned RNAseq raw reads to the mouse genome (mm10) using STAR (54) with 2-pass mapping. We computed gene-level raw counts in R using featureCounts (55) for all genes in the Gencode vM16

reference annotation. We performed differential gene expression analysis in R using the Bioconductor package DESeq2 (56).

Single-cell RNAseq

We loaded thymocyte suspensions (5×10^3 cells per sample) isolated from 6 week old mice on a 10× Chromium instrument (10× Genomics) and prepared single-cell RNAseq libraries using the Chromium Single Cell 3' v2 Reagent Kit (10× Genomics CG00052) and sequenced them on an Illumina HiSeq instrument (Illumina) at the JP Sulzberger Columbia Genome Center.

We demultiplexed BCL files with 10x Cell Ranger's mkfastq command and performed analysis and alignment with Cell Ranger's count command with Cell Ranger's reference mm10 version 1.2.0 using Cell Ranger 2.0.0. We loaded single-cell data into count matrices (pandas v0.23.0) and removed ribosomal and mitochondrial genes. We normalized each cell for library size using CPM normalization and log-transformed. We then performed principal component analysis (PCA) on the resulting matrices (scanpy v1.4.3). To identify significant principal components, we exploited Random Matrix Theory methodologies through the randomly algorithm (57). We used UMAP (58) to visualize the distribution of cells in the projection of the significant principal components. We then extracted markers for the different populations from data produced by the IMMGEN project (GSE15907). Briefly, we selected as genes that maximized the difference between the mean expression in the population of interest and the rest of the populations. Each gene score was then divided by the standard deviation of the gene in the samples outside the population of interest to penalize genes with high variability. The top genes were defined as markers for each population. These genes were subsequently used to score cells in the scRNA-Seq samples. Cells were assigned to a specific population if scoring in the top percentile (for up-regulated genes) or bottom percentile (for down-regulated genes) for such particular population. The percentiles for each population were defined using known population percentages determined by flow cytometry analysis in each genotype as ground truth. Cells with multiple or no population call were not assigned to any subset and were excluded from the final analysis.

ATAC-seq

We performed ATAC-seq analysis from sorted DN3 thymocytes from 6 week old mice, from NOTCH1-induced wild type and Myc-rescued GS1+2 tumors, and from DP leukemia lymphoblasts obtained from NOTCH1-induced leukemia-bearing mice and human lymphoblasts from two patients with DP (CD4⁺ CD8⁺) T-ALL. We generated transposed DNA fragments as described before (59) and amplified them by PCR using NEBNext High-Fidelity 2 PCR Master Mix (NEB M0541) and custom primer indexes to generate ATAC-seq libraries (59). We purified PCR products using Agencourt AMPure XP beads (Beckman Coulter A63880) and sequenced them on an Illumina NextSeq instrument (Illumina).

BCL files were demultiplexed and FASTQ files generated on the BaseSpace platform (Illumina). Reads were trimmed of contaminating adapter sequences using cutadapt and aligned to the GRCh38 (mm10) build mouse genome and the GRCh37 (hg19) build human genome respectively using Bowtie2 (60). Peaks of transposase accessible chromatin were called using MACS2 v2.1.1 (61).

We analyzed chromatin accessibility during T-cell development from mouse T-cell precursor ATAC-seq data from the Immunological Genome Project (GSE100738) as described before. All the peaks from each population were merged together using the merge function of bedtools to generate a consensus peak table for the T-cell development program (69,302 peaks). Genomic analysis and visualization of highly variable ATAC-seq peaks was performed using R packages (R version 3.5.0). Briefly, we studied the 10% of peaks (6930) with higher coefficient of variation. Consensus clustering of the samples was done using the ConsensusClusterPlus package in R. Unsupervised clustering and heatmap representation was done with gplot package. Analysis of motif enrichment (AME) was conducted the Meme suite tools v 5.0.2 with default parameters, using the Joma 2013 database. For downstream analysis, we considered motifs with E-value < 0.05 and a minimum percentage of 30% of true positive (TP).

DNA fluorescent in situ hybridization (FISH)

We designed oligonucleotide fluorescent probes for DNA FISH (Agilent) covering 200 Kb windows encompassing N-Me (mm9 chr15:62977370-63177630) or the *Myc* promoter (mm9 chr15:61721767-61924188). N-Me probe (64% coverage) was conjugated to FITC. The *Myc* promoter probe (70% coverage) was conjugated to Cy3.

We performed DNA FISH analyses of splenic B-cells and DN3 thymocytes from 6 week old mice. We incubated cells at 37°C for 30 minutes on poly-L-lysine coated slides and fixed them with 4% PFA in PBS at room temperature for 10 minutes. Then, cells were permeabilized in 0.5% Triton X-100 in PBS. Probes were hybridized following Agilent's FISH Protocol on FFPE Samples. After hybridization, coverslips were mounted on slides using ProLong Diamond Antifade Mountant with DAPI (Invitrogen P36962). We imaged cells with the CSU-X1 confocal spinning disk system (Yokogawa Life Sciences) on an Eclipse TiE microscope stand (Nikon Instruments, Melville, NY) using a 100x Oil Apo TIRF oil-immersion objective and an Andor Zyla 4.2 sCMOS camera.

We analyzed FISH images using ImageJ software (U. S. National Institutes of Health, RRID: SCR_003070) (62) and measured distances between FISH foci. Briefly, nuclei were segmented using a mask of the DAPI staining. Green and red channels in each nuclei (FISH labeling) were thresholded to create binary images and FISH foci were detected using 3D Objects Counter plugin. Volumes and distances between objects were measured using 3D ROI Manager plugin (63). Measurements were filtered to analyze the shortest centroid to centroid distance between NMe and *Myc* promoter foci.

Chromosome Conformation Capture (3C)

We performed 3C as previously described (64) using *MboI* as restriction enzyme. Bacterial artificial chromosome (BAC) clones were used as control template to cover the genomic region under study. We analyzed 3C libraries by quantitative real-time PCR (qRT-PCR) with a QuantStudio 3 Real-Time PCR System (Applied Biosystems) using FastStart Universal SYBR Green (Roche) and primers for the *Myc* promoter (5'-TGCCTTCCCCGCGAGATGGAGTGGCTGTTT-3'), N-Me (5'-

TGCAAGTGGAGTTGGCCATTGGGTGGCACC-3'), and a N-Me neighboring region (5'-TCACCCCAAGCCCAGTGCCTGTCATATGGGA-3').

Drugs

(Z)-4-hydroxytamoxifen (Santa Cruz SC-3542) was dissolved in ethanol and added to the media at a final concentration of 1 μ M. Tamoxifen (Sigma T5648) was dissolved in corn oil to a final concentration of 30 g mL⁻¹.

Quantitative RT-PCR

We analyzed expression of *Myc* and *Actb* by quantitative PCR (*Myc*: forward primer 5'-AGTGCTGCATGAGGAGACAC-3' and reverse primer 5'-GGTTTGCCTCTTCTCCACAG-3'; *Actb*: forward primer 5'-AGGTGACAGCATTGCTTCTG-3' and reverse primer 5'-GCTGCCTCAACACCTCAAC-3') using FastStart Universal SYBR Green (Roche 4913850001) in a 7500 Real-Time PCR system (Applied Biosystems).

Nucleosome assembly assay

We generated N-Me DNA sequences by PCR from mouse genomic DNA using a N-Me biotinylated forward primer (5'-ACTTCTACTGTATGCAGAATG-3') and an unmodified N-Me reverse primer (5'-GTAATAAAAGACCTCTCTTCC-3'). We assembled extended nucleosome arrays using Chromatin Assembly Kit (Active Motif), following the manufacturer's protocol. We generated compacted nucleosome arrays by adding histone H1 (Active motif) to extended nucleosome arrays and incubating for one hour at 27°C. We incubated extended and compacted nucleosome arrays with GATA3 (Origene) for two hours at room temperature in binding buffer containing 10mM Tris pH 7.5, 1mM β -mercaptoethanol, 40mM KCl, 5mM DTT, 250 μ g/mL BSA, 1% ficoll and 5% glycerol. We digested nucleosome arrays with 40U/mL of DNase I (NEB) for one minute at room temperature. DNA was purified, run in an agarose gel and transferred into a nylon membrane for chemoluminescence detection.

using Streptavidin-AP conjugate (Roche) and CDP-Star (Roche). DNA smear products were quantified by plot profile analysis using Fiji and normalized to the total DNA content.

ChIP-seq analysis

We analyzed N-Me occupancy of chromatin marks, epigenetic factors and transcription factors using the following T-ALL publicly available ChIPseq data sets from GEO: GSM1697882, GSM1314139, GSM2218755, GSM1689152, GSM1581344, GSM1442004, GSM722168, GSM3243670, GSM1442005, GSM2218756, GSM722167, GSM722166, GSM1193664, GSM449525, GSM2474553, GSM837992, GSM1410327, GSM732905, GSM2274676, GSM722165 and GSM1524254.

Statistical analyses

For analysis of mouse thymus development phenotypes, Myc intracellular expression in thymic populations, CD4⁺ CD8⁺ double positive wave in tumor generation experiments, *MYC* promoter luciferase reporter activity assays, and *Myc* expression, cell growth, apoptosis and cell cycle in *in vitro* experiments, we evaluated statistical significance using two-tailed Student's t test assuming normality and equal distribution of variance between the different groups analyzed. Distances between foci in DNA FISH experiments were analyzed using non-parametric Kolmorov-Smirnov test. Survival in mouse experiments was represented with Kaplan-Meier curves, and significance was estimated with the log-rank test. We performed all statistical analyses using Prism GraphPad 6 (RRID: SCR_002798) and considered statistical significance at $P < 0.05$.

Data availability

GEO Series accession numbers: RNA-seq data, GSE117483; single-cell RNAseq data, GSE117412; ATAC-seq data, GSE117573, GSE124175, and GSE124223.

Acknowledgements

We are grateful to T. Ludwig (The Ohio State University Comprehensive Cancer Center) for the Rosa26^{+/Cre-ERT2} mouse. We thank Esperanza Agullo-Pascual, Sofie Demeyer, Victor Lin and Beatrix

Ueberheide for outstanding technical assistance. This work was supported by the National Institute of Health grants R35 CA210065 (AF), U54 CA193313 (RR), R01 CA185486 (RR), U54 CA209997 (RR), and P30 CA013696 (Confocal and Specialized Microscopy Shared Resource and Transgenic Animal Shared Resource, Molecular Pathology Shared Resource, Herbert Irving Comprehensive Cancer Center). R.A. and S.A. are supported by Leukemia and Lymphoma Society postdoctoral fellowships. D.H. is supported by the US National Institutes of Health Grant K99/R00 CA197869 and an Alex Lemonade Stand Foundation Young Investigator grant. F.G. is supported by the American-Italian Cancer Foundation postdoctoral fellowship. M.R. is supported by a Damon-Runyon Sohn Pediatric Cancer fellowship. J.R.C is supported by a Lady Tata Memorial Trust fellowship. V.C. is supported by the Dutch Cancer League (KWF 2016-10355).

References

1. Levine M, Cattoglio C, Tjian R. Looping back to leap forward: transcription enters a new era. *Cell* **2014**;157(1):13-25 doi 10.1016/j.cell.2014.02.009.
2. Ong CT, Corces VG. Enhancer function: new insights into the regulation of tissue-specific gene expression. *Nature reviews Genetics* **2011**;12(4):283-93 doi 10.1038/nrg2957.
3. Shlyueva D, Stampfel G, Stark A. Transcriptional enhancers: from properties to genome-wide predictions. *Nature reviews Genetics* **2014**;15(4):272-86 doi 10.1038/nrg3682.
4. Whyte WA, Orlando DA, Hnisz D, Abraham BJ, Lin CY, Kagey MH, *et al.* Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **2013**;153(2):307-19 doi 10.1016/j.cell.2013.03.035.
5. Hnisz D, Abraham BJ, Lee TI, Lau A, Saint-Andre V, Sigova AA, *et al.* Super-enhancers in the control of cell identity and disease. *Cell* **2013**;155(4):934-47 doi 10.1016/j.cell.2013.09.053.
6. He HH, Meyer CA, Shin H, Bailey ST, Wei G, Wang Q, *et al.* Nucleosome dynamics define transcriptional enhancers. *Nature genetics* **2010**;42(4):343-7 doi 10.1038/ng.545.
7. Mito Y, Henikoff JG, Henikoff S. Histone replacement marks the boundaries of cis-regulatory domains. *Science (New York, NY)* **2007**;315(5817):1408-11 doi 10.1126/science.1134004.
8. Cao Q, Anyansi C, Hu X, Xu L, Xiong L, Tang W, *et al.* Reconstruction of enhancer-target networks in 935 samples of human primary cells, tissues and cell lines. *Nature genetics* **2017**;49(10):1428-36 doi 10.1038/ng.3950.
9. Weng AP, Ferrando AA, Lee W, Morris JPt, Silverman LB, Sanchez-Irizarry C, *et al.* Activating mutations of NOTCH1 in human T cell acute lymphoblastic leukemia. *Science (New York, NY)* **2004**;306(5694):269-71 doi 10.1126/science.1102160.
10. Palomero T, Lim WK, Odom DT, Sulis ML, Real PJ, Margolin A, *et al.* NOTCH1 directly regulates c-MYC and activates a feed-forward-loop transcriptional network promoting leukemic cell growth. *Proceedings of the National Academy of Sciences of the United States of America* **2006**;103(48):18261-6 doi 10.1073/pnas.0606108103.

11. Herranz D, Ambesi-Impiombato A, Palomero T, Schnell SA, Belper L, Wendorff AA, *et al.* A NOTCH1-driven MYC enhancer promotes T cell development, transformation and acute lymphoblastic leukemia. *Nature medicine* **2014**;20(10):1130-7 doi 10.1038/nm.3665.
12. Ho AS, Kannan K, Roy DM, Morris LG, Ganly I, Katabi N, *et al.* The mutational landscape of adenoid cystic carcinoma. *Nature genetics* **2013**;45(7):791-8 doi 10.1038/ng.2643.
13. Stephens PJ, Davies HR, Mitani Y, Van Loo P, Shlien A, Tarpey PS, *et al.* Whole exome sequencing of adenoid cystic carcinoma. *The Journal of clinical investigation* **2013**;123(7):2965-8 doi 10.1172/jci67201.
14. Puente XS, Pinyol M, Quesada V, Conde L, Ordonez GR, Villamor N, *et al.* Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia. *Nature* **2011**;475(7354):101-5 doi 10.1038/nature10113.
15. Kridel R, Meissner B, Rogic S, Boyle M, Telenius A, Woolcock B, *et al.* Whole transcriptome sequencing reveals recurrent NOTCH1 mutations in mantle cell lymphoma. *Blood* **2012**;119(9):1963-71 doi 10.1182/blood-2011-11-391474.
16. Yui MA, Rothenberg EV. Developmental gene networks: a triathlon on the course to T cell identity. *Nature reviews Immunology* **2014**;14(8):529-45 doi 10.1038/nri3702.
17. Shen Y, Yue F, McCleary DF, Ye Z, Edsall L, Kuan S, *et al.* A map of the cis-regulatory sequences in the mouse genome. *Nature* **2012**;488(7409):116-20 doi 10.1038/nature11243.
18. Yashiro-Ohtani Y, Wang H, Zang C, Arnett KL, Bailis W, Ho Y, *et al.* Long-range enhancer activity determines Myc sensitivity to Notch inhibitors in T cell leukemia. *Proceedings of the National Academy of Sciences of the United States of America* **2014**;111(46):E4946-53 doi 10.1073/pnas.1407079111.
19. Bahr C, von Paleske L, Uslu VV, Remeseiro S, Takayama N, Ng SW, *et al.* A Myc enhancer cluster regulates normal and leukaemic haematopoietic stem cell hierarchies. *Nature* **2018**;553(7689):515-20 doi 10.1038/nature25193.

20. Anderson MK. At the crossroads: diverse roles of early thymocyte transcriptional regulators. *Immunological reviews* **2006**;209:191-211 doi 10.1111/j.0105-2896.2006.00352.x.
21. Rothenberg EV, Ungerback J, Champhekar A. Forging T-Lymphocyte Identity: Intersecting Networks of Transcriptional Control. *Advances in immunology* **2016**;129:109-74 doi 10.1016/bs.ai.2015.09.002.
22. Belver L, Ferrando A. The genetics and mechanisms of T cell acute lymphoblastic leukaemia. *Nature reviews Cancer* **2016**;16(8):494-507 doi 10.1038/nrc.2016.63.
23. Schroeter EH, Kisslinger JA, Kopan R. Notch-1 signalling requires ligand-induced proteolytic release of intracellular domain. *Nature* **1998**;393(6683):382-6 doi 10.1038/30756.
24. Hattori N, Kawamoto H, Fujimoto S, Kuno K, Katsura Y. Involvement of transcription factors TCF-1 and GATA-3 in the initiation of the earliest step of T cell development in the thymus. *The Journal of experimental medicine* **1996**;184(3):1137-47.
25. Ting CN, Olson MC, Barton KP, Leiden JM. Transcription factor GATA-3 is required for development of the T-cell lineage. *Nature* **1996**;384(6608):474-8 doi 10.1038/384474a0.
26. Hendriks RW, Nawijn MC, Engel JD, van Doorninck H, Grosveld F, Karis A. Expression of the transcription factor GATA-3 is required for the development of the earliest T cell progenitors and correlates with stages of cellular proliferation in the thymus. *European journal of immunology* **1999**;29(6):1912-8 doi 10.1002/(sici)1521-4141(199906)29:06<1912::aid-immu1912>3.0.co;2-d.
27. Tallen G, Ratei R, Mann G, Kaspers G, Niggli F, Karachunsky A, *et al.* Long-term outcome in children with relapsed acute lymphoblastic leukemia after time-point and site-of-relapse stratification and intensified short-course multidrug chemotherapy: results of trial ALL-REZ BFM 90. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **2010**;28(14):2339-47 doi 10.1200/jco.2009.25.1983.
28. Song L, Zhang Z, Gräsfeder LL, Boyle AP, Giresi PG, Lee BK, *et al.* Open chromatin defined by DNaseI and FAIRE identifies regulatory elements that shape cell-type identity. *Genome research* **2011**;21(10):1757-67 doi 10.1101/gr.121541.111.

29. Heintzman ND, Stuart RK, Hon G, Fu Y, Ching CW, Hawkins RD, *et al.* Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nature genetics* **2007**;39(3):311-8 doi 10.1038/ng1966.
30. Rada-Iglesias A, Bajpai R, Swigut T, Brugmann SA, Flynn RA, Wysocka J. A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* **2011**;470(7333):279-83 doi 10.1038/nature09692.
31. Dose M, Sleckman BP, Han J, Bredemeyer AL, Bendelac A, Gounari F. Intrathymic proliferation wave essential for Valpha14+ natural killer T cell development depends on c-Myc. *Proceedings of the National Academy of Sciences of the United States of America* **2009**;106(21):8641-6 doi 10.1073/pnas.0812255106.
32. Cirillo LA, Lin FR, Cuesta I, Friedman D, Jarnik M, Zaret KS. Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. *Molecular cell* **2002**;9(2):279-89.
33. Ferrando AA, Neuberg DS, Staunton J, Loh ML, Huard C, Raimondi SC, *et al.* Gene expression signatures define novel oncogenic pathways in T cell acute lymphoblastic leukemia. *Cancer cell* **2002**;1(1):75-87.
34. Arnone MI, Davidson EH. The hardwiring of development: organization and function of genomic regulatory systems. *Development (Cambridge, England)* **1997**;124(10):1851-64.
35. Borok MJ, Tran DA, Ho MC, Drewell RA. Dissecting the regulatory switches of development: lessons from enhancer evolution in *Drosophila*. *Development (Cambridge, England)* **2010**;137(1):5-13 doi 10.1242/dev.036160.
36. Canver MC, Smith EC, Sher F, Pinello L, Sanjana NE, Shalem O, *et al.* BCL11A enhancer dissection by Cas9-mediated in situ saturating mutagenesis. *Nature* **2015**;527(7577):192-7 doi 10.1038/nature15521.
37. Payne JL, Wagner A. Mechanisms of mutational robustness in transcriptional regulation. *Frontiers in genetics* **2015**;6:322 doi 10.3389/fgene.2015.00322.

38. Calo E, Wysocka J. Modification of enhancer chromatin: what, how, and why? *Molecular cell* **2013**;49(5):825-37 doi 10.1016/j.molcel.2013.01.038.
39. Brunelle M, Nordell Markovits A, Rodrigue S, Lupien M, Jacques PE, Gevry N. The histone variant H2A.Z is an important regulator of enhancer activity. *Nucleic acids research* **2015**;43(20):9742-56 doi 10.1093/nar/gkv825.
40. Zaret KS, Mango SE. Pioneer transcription factors, chromatin dynamics, and cell fate control. *Current opinion in genetics & development* **2016**;37:76-81 doi 10.1016/j.gde.2015.12.003.
41. Wang H, Yang H, Shivalila CS, Dawlaty MM, Cheng AW, Zhang F, *et al.* One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell* **2013**;153(4):910-8 doi 10.1016/j.cell.2013.04.025.
42. Herranz D, Ambesi-Impiombato A, Sudderth J, Sanchez-Martin M, Belver L, Tosello V, *et al.* Metabolic reprogramming induces resistance to anti-NOTCH1 therapies in T cell acute lymphoblastic leukemia. **2015**;21(10):1182-9 doi 10.1038/nm.3955.
43. Guo K, McMinn JE, Ludwig T, Yu YH, Yang G, Chen L, *et al.* Disruption of peripheral leptin signaling in mice results in hyperleptinemia without associated metabolic abnormalities. *Endocrinology* **2007**;148(8):3987-97 doi 10.1210/en.2007-0261.
44. Calado DP, Sasaki Y, Godinho SA, Pellerin A, Kochert K, Sleckman BP, *et al.* The cell-cycle regulator c-Myc is essential for the formation and maintenance of germinal centers. *Nature immunology* **2012**;13(11):1092-100 doi 10.1038/ni.2418.
45. Rocha PP, Raviram R, Fu Y, Kim J, Luo VM, Aljoufi A, *et al.* A Damage-Independent Role for 53BP1 that Impacts Break Order and Igh Architecture during Class Switch Recombination. *Cell reports* **2016**;16(1):48-55 doi 10.1016/j.celrep.2016.05.073.
46. van de Werken HJ, de Vree PJ, Splinter E, Holwerda SJ, Klous P, de Wit E, *et al.* 4C technology: protocols and data analysis. *Methods in enzymology* **2012**;513:89-112 doi 10.1016/b978-0-12-391938-0.00004-5.

47. Unnikrishnan A, Guan YF, Huang Y, Beck D, Thoms JA, Peirs S, *et al.* A quantitative proteomics approach identifies ETV6 and IKZF1 as new regulators of an ERG-driven transcriptional network. *Nucleic acids research* **2016**;44(22):10644-61 doi 10.1093/nar/gkw804.
48. Mellacheruvu D, Wright Z, Couzens AL, Lambert JP, St-Denis NA, Li T, *et al.* The CRAPome: a contaminant repository for affinity purification-mass spectrometry data. *Nature methods* **2013**;10(8):730-6 doi 10.1038/nmeth.2557.
49. Cartharius K, Frech K, Grote K, Klocke B, Haltmeier M, Klingenhoff A, *et al.* MatInspector and beyond: promoter analysis based on transcription factor binding sites. *Bioinformatics (Oxford, England)* **2005**;21(13):2933-42 doi 10.1093/bioinformatics/bti473.
50. Pollard KS, Hubisz MJ, Rosenbloom KR, Siepel A. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome research* **2010**;20(1):110-21 doi 10.1101/gr.097857.109.
51. Letunic I, Bork P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic acids research* **2016**;44(W1):W242-5 doi 10.1093/nar/gkw290.
52. Heigwer F, Kerr G, Boutros M. E-CRISP: fast CRISPR target site identification. **2014**;11(2):122-3 doi 10.1038/nmeth.2812.
53. He TC, Sparks AB, Rago C, Hermeking H, Zawel L, da Costa LT, *et al.* Identification of c-MYC as a target of the APC pathway. *Science (New York, NY)* **1998**;281(5382):1509-12.
54. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics (Oxford, England)* **2013**;29(1):15-21 doi 10.1093/bioinformatics/bts635.
55. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics (Oxford, England)* **2014**;30(7):923-30 doi 10.1093/bioinformatics/btt656.
56. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology* **2014**;15(12):550 doi 10.1186/s13059-014-0550-8.

57. Aparicio L, Bordyuh M, Blumberg AJ, Rabadan R. Quasi-universality in single-cell sequencing data. arXiv e-prints2018.
58. McInnes L, Healy J, Melville J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. arXiv e-prints2018.
59. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nature methods* **2013**;10(12):1213-8 doi 10.1038/nmeth.2688.
60. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nature methods* **2012**;9(4):357-9 doi 10.1038/nmeth.1923.
61. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome biology* **2008**;9(9):R137 doi 10.1186/gb-2008-9-9-r137.
62. Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. *Nature methods* **2012**;9(7):671-5.
63. Bolte S, Cordelieres FP. A guided tour into subcellular colocalization analysis in light microscopy. *Journal of microscopy* **2006**;224(Pt 3):213-32 doi 10.1111/j.1365-2818.2006.01706.x.
64. He B, Chen C, Teng L, Tan K. Global view of enhancer-promoter interactome in human cells. *Proceedings of the National Academy of Sciences of the United States of America* **2014**;111(21):E2191-9 doi 10.1073/pnas.1320308111.

Figure legends

Figure 1. Chromatin accessibility dynamics during T-cell development. (A-B) Analysis of active genomic intervals in thymocyte populations. Unsupervised clustering heatmap (A) and consensus clustering (k=6) (B) of the 10% most variable ATAC-seq peaks (n=6930) through the different T-cell precursor populations are shown. (C) Chromatin accessibility profiles (upper panel) and transcription factor binding site enrichment analysis (lower panel) in active genomic intervals associated with the most relevant T-cell developmental stages. Bar graphs represent the percentage of active genomic intervals that contain a significant enrichment in transcription factor binding motifs for the PU-box, GATA, Runt-related (RUNX), homeodomain (HOX), helix-loop-helix, ETS, Forkhead-box (FOX) and Krüppel-like (KRAB) transcription factor families.

Figure 2. Functional and structural characterization of N-Me. (A) Normalized 4C contact profiles in Jurkat cells (upper panel) and NOTCH1-induced mouse T-ALL cells (lower panel). Viewpoint is located in the MYC promoter (top tracks) or in N-Me (bottom tracks). 4C signal is merged across three replicates. The median, 20th and 80th percentiles of sliding 25Kb windows determine the main trend line. Color scale represents read coverage of sliding windows sized from 2 to 50Kb. (B) Analysis epigenetic of marks (yellow), epigenetic factor (gray) and transcription factor (blue) N-Me occupancy by ChIPseq in human T-ALL cells. Dotted lines mark the boundaries of N-Me. Scale bar is represented in the upper left corner. (C) Reverse ChIP identification of potential N-Me-binding factors. A N-Me DNA bait was incubated in the presence of nuclear extracts from Jurkat, ALL-SIL and HPB-ALL cells and recovered peptides were analyzed by mass spectrometry. The diagram represents the proteins recovered in one (purple), two (red) or all three (blue) cell lines analyzed. (D) N-Me evolutionary conservation tree. (E) Predicted ultraconserved transcription factor binding motifs in the N-Me sequence. PhyloP scores are shown above the sites.

Figure 3. Phenotypical analysis of N-Me GATA site mutant mice. (A) Morphology in the thymi of 6-week old N-Me wild type (WT), GATA site 1 (GS1), GATA site 2 (GS2), and GATA site 1 and 2

(GS1+2) homozygous mutant mice. **(B)** Cellularity of thymi as in **A**. **(C)** Representative flow cytometry plots of thymocyte populations stained with anti-CD4 and anti-CD8 antibodies in wild type and mutant mouse thymus as in **A**. **(D)** Absolute numbers of thymic populations in wild type and mutant mouse thymus as in **A**. **(E)** Absolute numbers of DN thymic subpopulations in wild type and mutant mouse thymus as in **A**. Individual values for single mice are shown (n=5). Box plots indicate values from the 25th through 75th percentile, median is indicated by the horizontal bar and whiskers extend down to the minimum and up to the maximum value. *P* values correspond to two-tailed Student's t-test.

Figure 4. N-Me GATA site mutations impair *Myc* expression. **(A)** Luciferase reporter assay of the *MYC* promoter alone or coupled to wild type or GATA site mutant N-Me. **(B)** RNAseq gene expression analysis in sorted DN3 thymocytes from 6-week old N-Me wild type (WT) and GATA site 1 and 2 (GS1+2mut) homozygous mutant mice (n=3). Representative Gene Set Enrichment Analysis (GSEA) plot of genes regulated by *MYC* and bar graph representation of normalized enrichment scores for the top *MYC*-related gene signatures from MSigDB. **(C)** Heatmap representation of the top 50 differentially expressed genes between wild type and homozygous GS1+2 mutant DN3 cells. Scale bar shows color-coded differential expression, with red indicating higher levels of expression and blue indicating lower levels of expression. **(D)** Single-cell RNAseq analysis of total thymus (upper panel) and CD4⁺ CD3⁺ thymocytes (lower panel) from 6-week old N-Me wild type and GS1+2 homozygous mutant mice. UMAP embeddings (left) show the cells annotated to each thymic population. Dotplots (right) represent the expression of *Myc*. Size of the dots is proportional to the percentage of cells expressing *Myc* in each population, color of the dots represents *Myc* average expression. **(E)** UMAP embeddings representing single-cell *Myc* expression in total CD4⁺ CD3⁺ thymocytes as in **D**. **(F)** Histology and immunohistochemical analysis of *Myc* expression in thymic tissue from 6-week old N-Me wild type, GS1, GS2, and GS1+2 homozygous mutant mice. Scale bar = 50 μ m. **(G)** Flow cytometry analysis of *Myc* expression levels in thymic populations from wild type and mutant mice as in **F**. Kinetics (left) and representative histograms with individual median fluorescence intensity (MFI) values for single mice (right) are shown (n=3). *P* values in **A** and **G** correspond to two-tailed Student's t-test.

Figure 5. N-Me GATA sites are essential for NOTCH1-induced leukemia development and maintenance. (A) Quantification of the CD4⁺ CD8⁺ preleukemic cells in peripheral blood of mice transplanted with Δ E-NOTCH1-infected wild type (GS1+2 +/+), GS1+2 heterozygous (GS1+2 +/-mut), and GS1+2 homozygous (GS1+2 mut/mut) mutant bone marrow progenitors. (B) Representative blood smear preparations in mice transplanted as in A eight weeks after transplant. (C) Kaplan-Meier survival curves (n=7) in mice transplanted as in A. (D) Survival analysis of mice transplanted with Rosa26TM-Cre NOTCH1-induced leukemias harboring a N-Me conditional and a N-Me GS1+2 mutant allele, and treated with vehicle only (N-Me^{flox/GS1+2mut}) or tamoxifen (N-Me^{-/GS1+2mut}) (n = 8 per group). (E) RT-qPCR analysis of *Myc* expression in vehicle treated (N-Me^{flox/GS1+2mut}) and tamoxifen treated (N-Me^{-/GS1+2mut}) NOTCH1-induced T-ALL N-Me^{flox/GS1+2mut} tumor cells. (F) Growth curve of vehicle treated (N-Me^{flox/GS1+2mut}) and tamoxifen treated (N-Me^{-/GS1+2mut}) NOTCH1-induced T-ALL N-Me^{flox/GS1+2mut} tumor cells. (G) Cell cycle analysis of NOTCH1-induced T-ALL N-Me^{flox/GS1+2mut} tumor cells treated with vehicle (N-Me^{flox/GS1+2mut}) or tamoxifen (N-Me^{-/GS1+2mut}) for 3 days. (H) Analysis of apoptosis and cell death in NOTCH1-induced T-ALL N-Me^{flox/GS1+2mut} tumor cells treated with vehicle (N-Me^{flox/GS1+2mut}) or tamoxifen (N-Me^{-/GS1+2mut}) for 3 days. The *P* value in A, E, F, G and H was calculated using two-tailed Student's *t*-test. The *P* value in C and D was calculated using log-rank test.

Figure 6. N-Me epigenetic marks and promoter-enhancer looping in GS1+2 mutant thymocytes and T-ALL cells are preserved. (A) Gata3 binding to N-Me in sorted DN3 thymocytes from GS1, GS2, and GS1+2 heterozygous mutant mice. N-Me was PCR-amplified from total input chromatin (left) or Gata3-immunoprecipitated chromatin (right) and analyzed by Sanger sequencing. (B) H3K27ac and H3K4me1 occupancy of N-Me in DN3 thymocytes from GS1+2 heterozygous mutant mice. Expected sequences for the wild type and mutant GATA site alleles are indicated below each chromatogram. (C) FISH analysis of N-Me-*Myc* promoter interaction in DN3 thymocytes and naïve B-cells from wild type and GS1+2 homozygous mutant (GS1+2mut) mice. Graph represents distances between foci. Horizontal bars represent the median values. Representative nuclei (single z-slice) are shown with red signal corresponding to the *Myc* promoter probe and green signal to the N-Me probe. Scale bar = 5 μ m.

Statistical significance was calculated using Kolmogorov-Smirnov test. **(D)** 3C quantitative PCR analysis of the relative interaction between DNA sequences flanking a *Mbol* restriction site in the MYC promoter and in the vicinity of the N-Me enhancer. Individual points represent independent library preparations. PCR signal is normalized to bacterial artificial chromosome (BAC) templates and to a N-Me neighboring region. *P* value corresponds to two-tailed Student's t-test.

Figure 7. Chromatin accessibility at the N-Me enhancer in GS1+2 mutant thymocytes. **(A)** ATAC-seq chromatin accessibility analysis of N-Me during T-cell differentiation. Dotted lines mark the boundaries of N-Me. Scale bar is represented in the upper left corner. **(B)** Heatmap representation of *Myc* expression in developing thymocytes. *Myc* RNAseq mRNA levels are color coded with red indicated higher levels and blue lower levels of expression. **(C)** ATAC-seq chromatin accessibility analysis of the N-Me enhancer in sorted DN3 thymocytes from 6-week old N-Me wild type and GS1+2 homozygous mutant mice (n=3). Normalized signal tracks for each genotype and differential chromatin accessibility heatmap are shown. **(D)** Nucleosome occupancy profiles as in **C**. Black bars indicate nucleosome-free regions. **(E)** Smarca4 occupancy of N-Me in DN3 thymocytes from GS1+2 heterozygous mutant mice. Expected sequences for the wild type and mutant GATA site alleles are indicated below the chromatograms. **(F)** Transcription factor occupancy of N-Me in wild type and *Myc*-rescued GS1+2 mutant NOTCH1-induced T-ALL lymphoblasts. PCR signal is normalized to input chromatin and to the average signal in wild type tumors. Error bars represent standard deviation between technical replicates. **(G)** ATAC-seq chromatin accessibility analysis of the N-Me enhancer in mouse DP thymocytes and in DP T-ALL lymphoblast cells. Dotted lines mark the boundaries of N-Me. Scale bar is represented in the upper right corner. **(H)** ATAC-seq chromatin accessibility analysis of the N-Me enhancer in human DP thymocytes and in two independent DP T-ALL samples as in **G**.

Figure 1

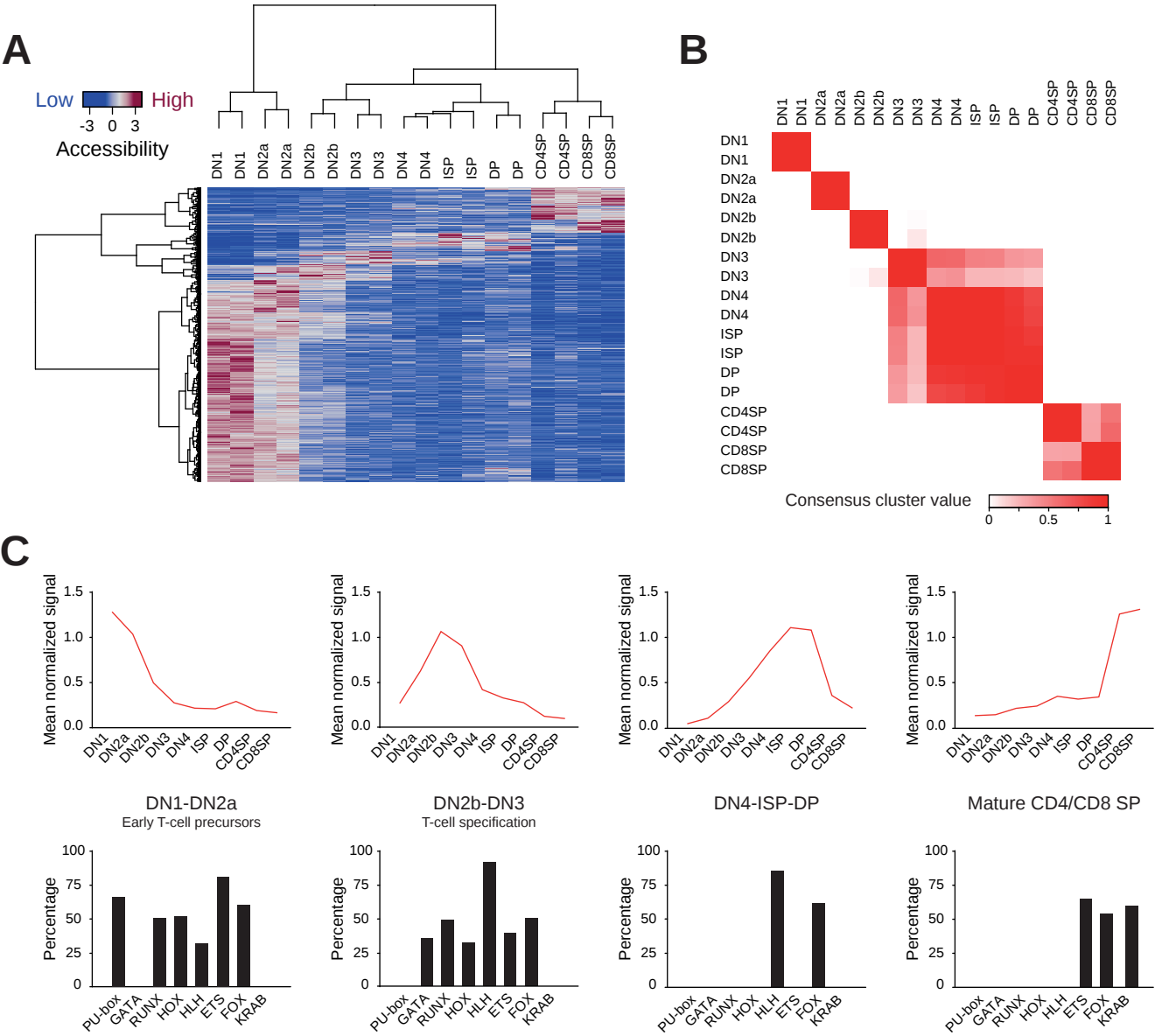


Figure 3

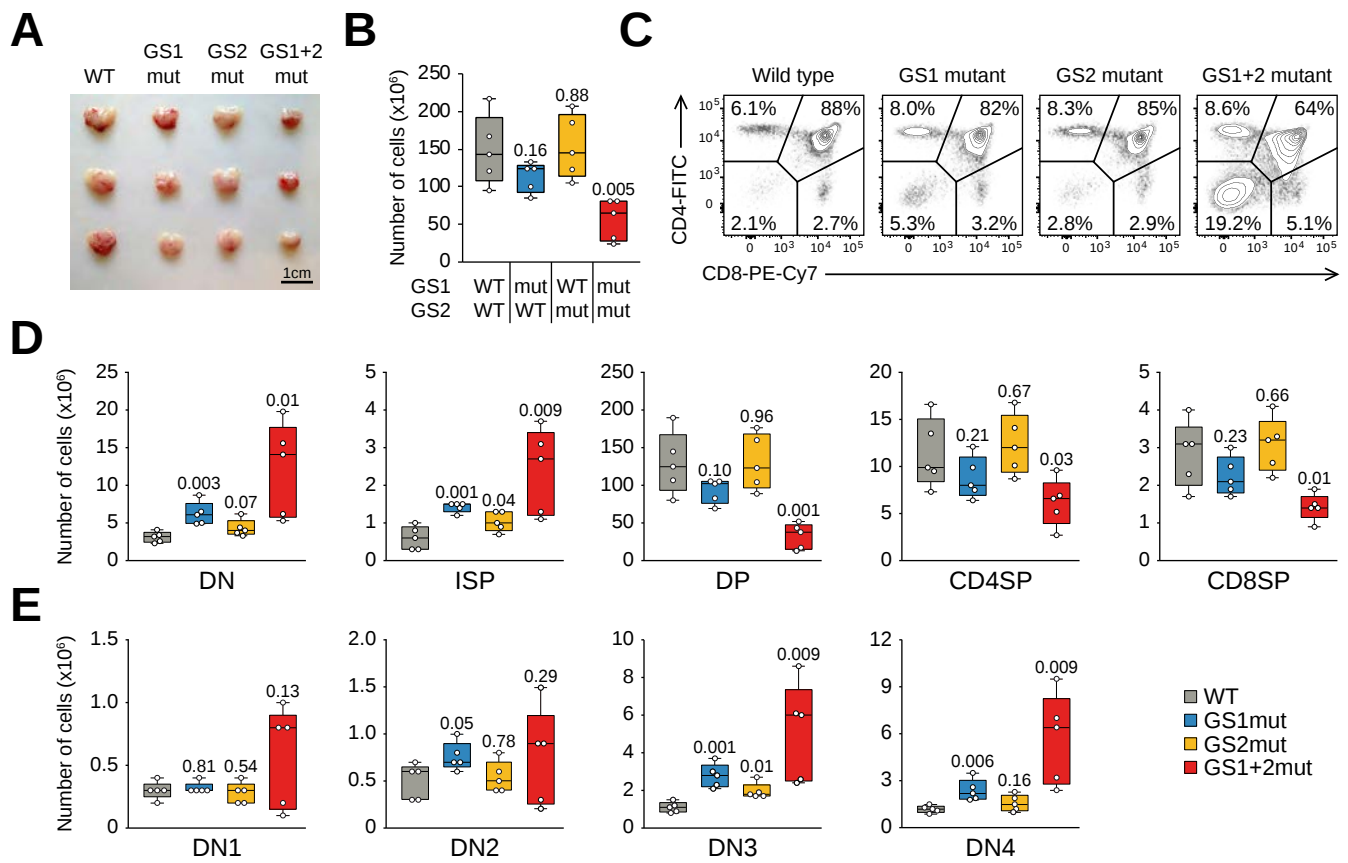


Figure 4

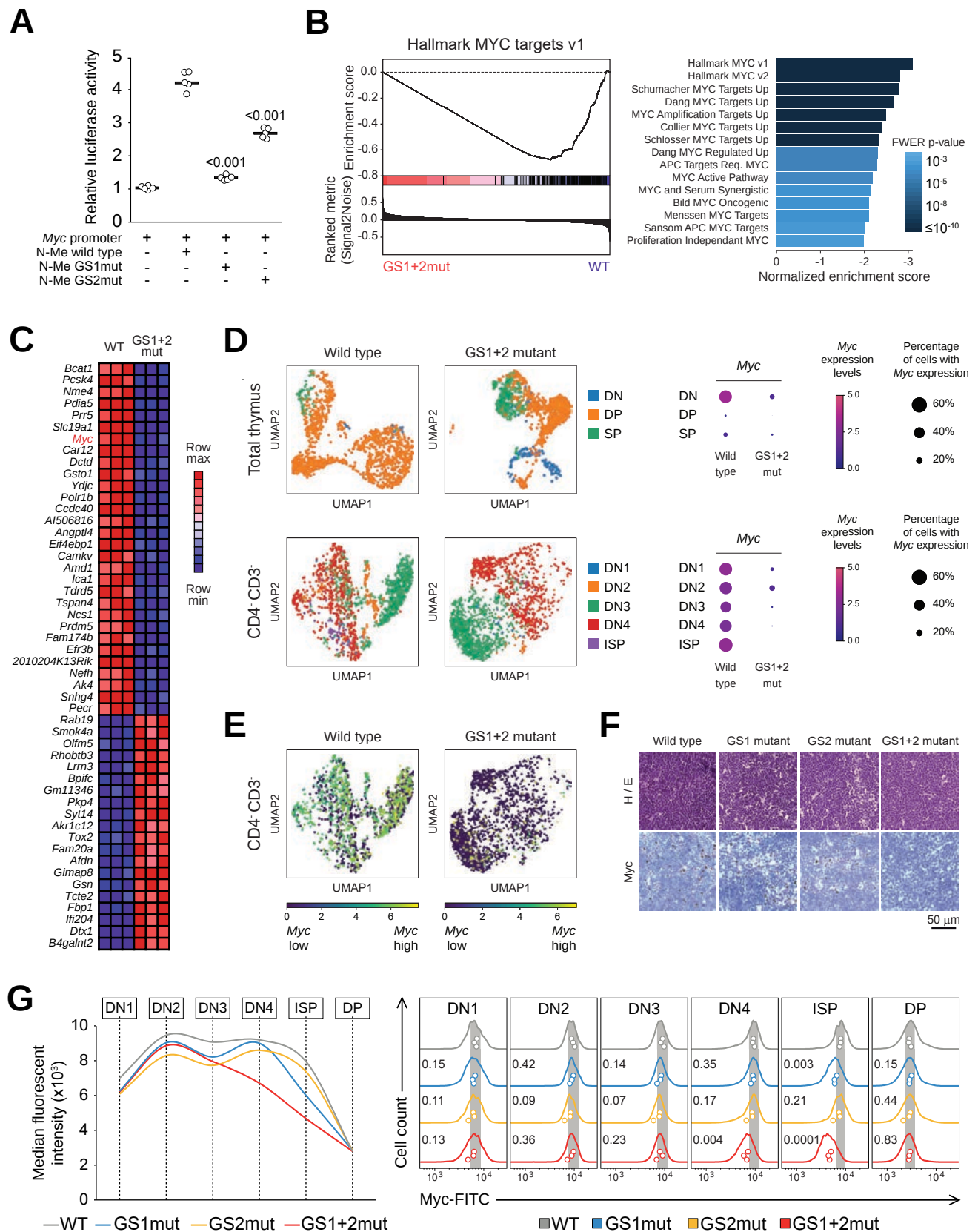


Figure 5

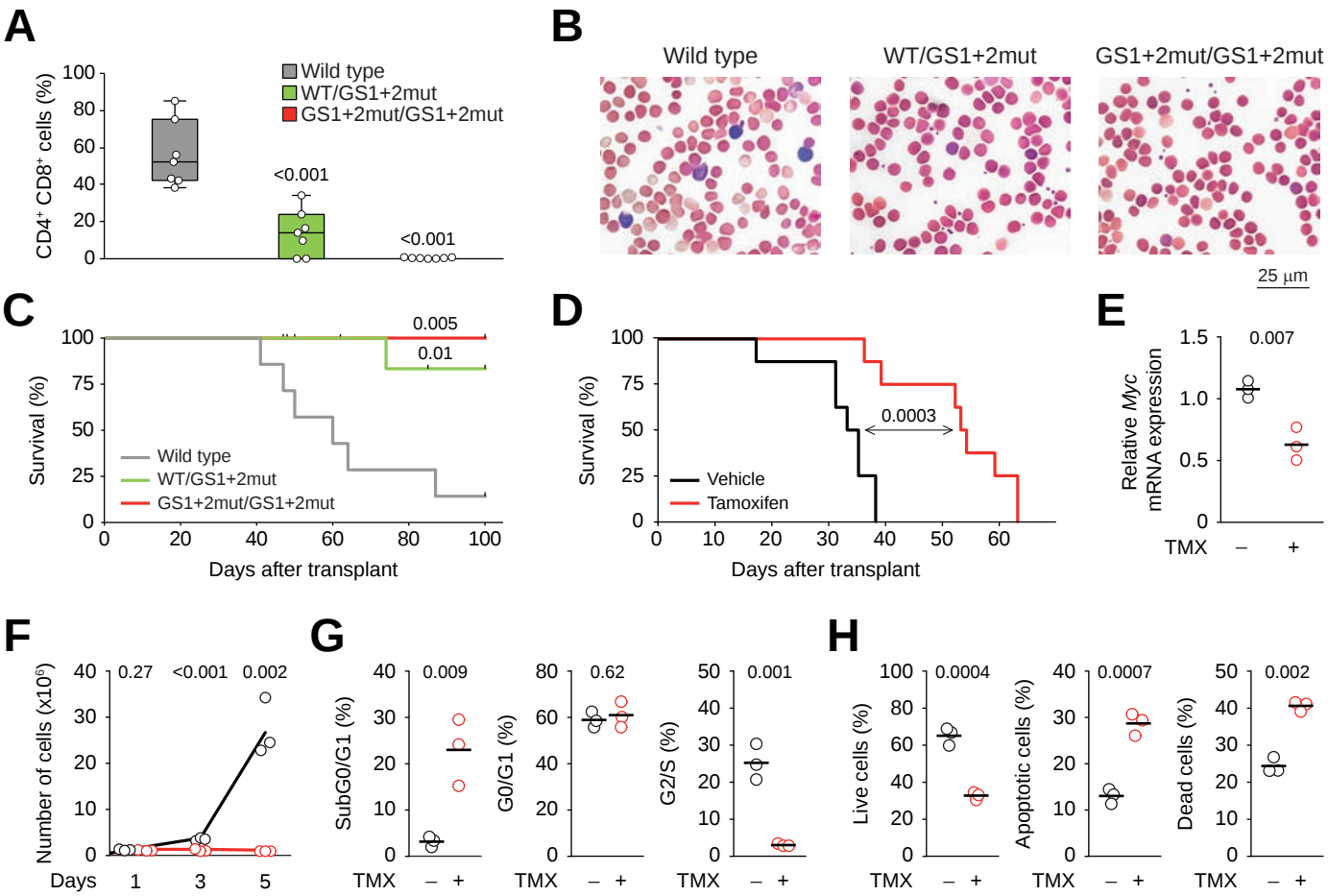


Figure 6

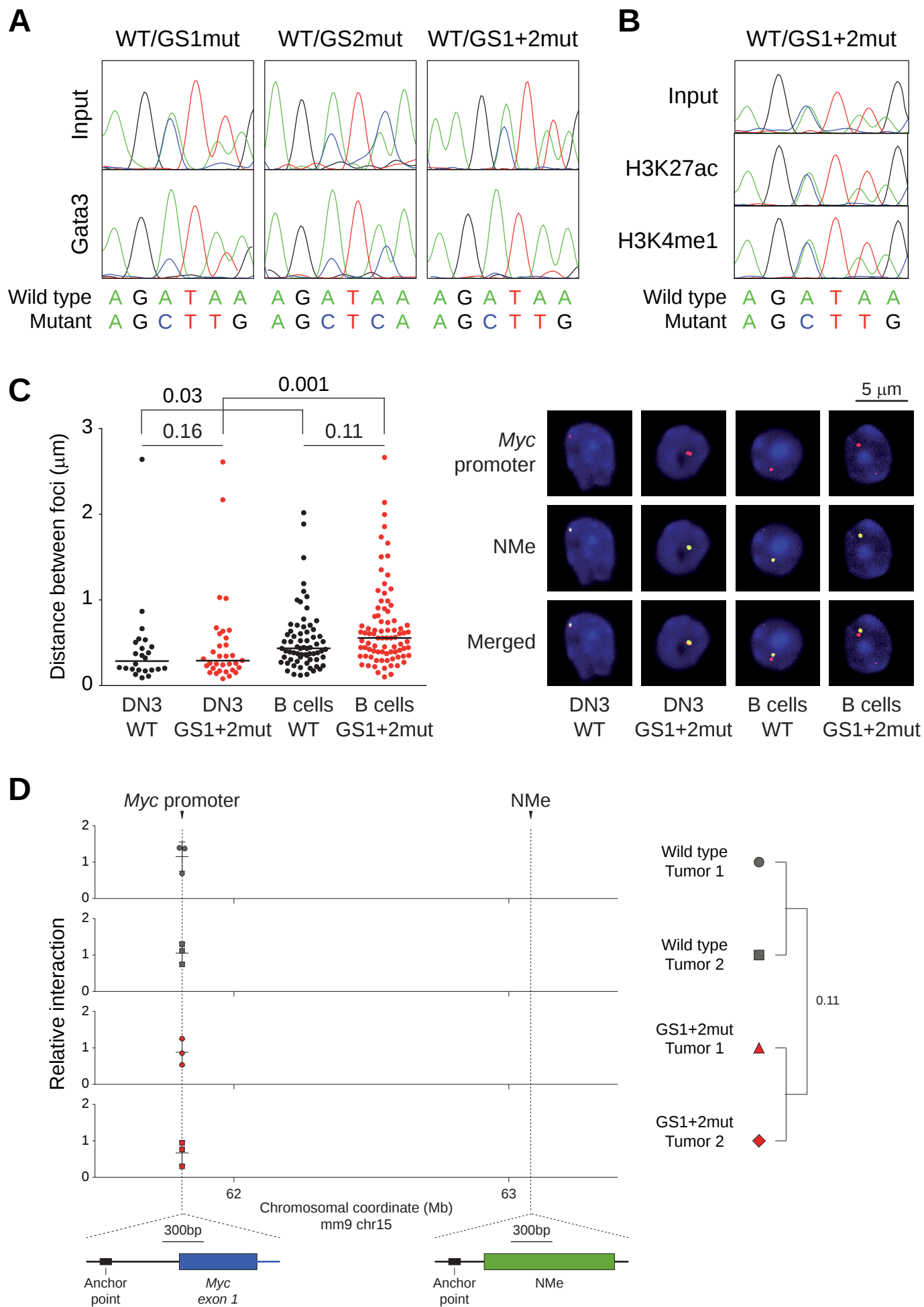


Figure 7

