Blueprint of human thymopoiesis reveals molecular mechanisms of stage-specific TCR enhancer activation

Agata Cieslak, Guillaume Charbonnier, Melania Tesio, Eve-Lyne Mathieu, Mohamed Belhocine, Aurore Touzart, Charlotte Smith, Guillaume Hypolite, Guillaume Andrieu, Joost H.A. Martens, et al.

JEM Journal of Experimental Medicine

ARTICLE

Blueprint of human thymopoiesis reveals molecular mechanisms of stage-specific TCR enhancer activation

Agata Cieslak[1*], Guillaume Charbonnier[2,3*], Melania Tesio[1], Eve-Lyne Mathieu[2,3], Mohamed Belhocine[2,3], Aurore Touzart[1,4], Charlotte Smith[1], Guillaume Hypolite[1], Guillaume P. Andrieu[1], Joost H.A. Martens[5], Eva Janssen-Megens[5], Marta Gut[6,7], Ivo Gut[6,7], Nicolas Boissel[8], Arnaud Petit[9], Denis Puthier[2,3], Elizabeth Macintyre[1], Hendrik G. Stunnenberg[5], Salvatore Spicuglia[2,3**], and Vahid Asnafi[1**]

Cell differentiation is accompanied by epigenetic changes leading to precise lineage definition and cell identity. Here we present a comprehensive resource of epigenomic data of human T cell precursors along with an integrative analysis of other hematopoietic populations. Although T cell commitment is accompanied by large scale epigenetic changes, we observed that the majority of distal regulatory elements are constitutively unmethylated throughout T cell differentiation, irrespective of their activation status. Among these, the *TCRA* gene enhancer (Eα) is in an open and unmethylated chromatin structure well before activation. Integrative analyses revealed that the HOXA5-9 transcription factors repress the Eα enhancer at early stages of T cell differentiation, while their decommission is required for *TCRA* locus activation and enforced αβ T lineage differentiation. Remarkably, the HOXA-mediated repression of Eα is paralleled by the ectopic expression of homeodomain-related oncogenes in T cell acute lymphoblastic leukemia. These results highlight an analogous enhancer repression mechanism at play in normal and cancer conditions, but imposing distinct developmental constraints.

Introduction

T lymphocytes develop from a stepwise process of cell fate choices whereby distinct signaling pathways in the thymus cause hematopoietic precursors to commit to the T cell fate, while mobilizing a T cell gene expression program that prepares the cells for TCR expression, TCR-based repertoire selection, and long, versatile careers as immune effectors (Dik et al., 2005b; Rothenberg, 2019; Spits, 2002). However, major questions remain as the molecular mechanisms involved in this process and the stage-specific regulation of T cell genes are not well defined yet. Despite numerous reports on the dynamics of epigenetic modifications during murine T cell differentiation (Pekowska et al., 2011; Zhang et al., 2012; Hu et al., 2018; Wei et al., 2011), we still have a limited understanding of the epigenetic mechanisms controlling human T cell differentiation. Thus, describing these mechanisms is of crucial importance, given the potential relevance for immune-related diseases (Clave et al., 2018; Kernfeld et al., 2018) as well as for the oncogenic transformation of T cell precursors (Aifantis et al., 2008).

Human T lymphocyte ontogeny in the thymus requires the ordered somatic recombination of V, D, and J gene segments at the *TCR* loci to determine the development into either γδ or αβ T cell lineages (Dik et al., 2005b; Spits, 2002). While the *TCRD*, *TCRG*, and *TCRB* loci rearrange at the early double negative (DN) tCD34 (CD34$^+$/CD3$^-$/CD4$^-$/CD8$^-$) stage (Fig. S1 A), the *TCRA* locus germline expression and rearrangements start at the early cortical (EC) stage and reach high levels of mature TCRA expression by late cortical (LC) and subsequent single positive (SP) stages (Dik et al., 2005b). It has been previously shown that the

[1]Université de Paris (Descartes), Institut Necker-Enfants Malades, Institut National de la Santé et de la Recherche Médicale U1151, and Laboratory of Onco-Hematology, Assistance Publique-Hôpitaux de Paris, Hôpital Necker Enfants-Malades, Paris, France; [2]Aix-Marseille University, Institut National de la Santé et de la Recherche Médicale, Theories and Approaches of Genomic Complexity, UMR1090, Marseille, France; [3]Equipe Labellisée Ligue Contre le Cancer, Marseille, France; [4]Division of Cancer Epigenomics, German Cancer Research Center, Heidelberg, Germany; [5]Department of Molecular Biology, Faculties of Science and Medicine, Radboud Institute for Molecular Life Sciences, Radboud University, Nijmegen, Netherlands; [6]Centro Nacional de Análisis Genómico–Centre for Genomic Regulation, Barcelona Institute of Science and Technology, Barcelona, Spain; [7]Universitat Pompeu Fabra, Barcelona, Spain; [8]Université Paris Diderot, Institut Universitaire d'Hématologie, EA-3518, Assistance Publique-Hôpitaux de Paris, University Hospital Saint-Louis, Paris, France; [9]Department of Pediatric Hematology and Oncology, Assistance Publique-Hôpitaux de Paris, Hôpital Armand Trousseau, Paris, France.

*A. Cieslak and G. Charbonnier contributed equally to this paper; **S. Spicuglia and V. Asnafi contributed equally to this paper; Correspondence to Vahid Asnafi: vahid.asnafi@aphp.fr; Salvatore Spicuglia: salvatore.spicuglia@inserm.fr.

TCRA gene enhancer (Eα enhancer) is essential for the tight regulation of *TCRA* rearrangements and expression during αβ T cell differentiation (Bassing et al., 2003; Sleckman et al., 1997). The Eα enhancer is the only known enhancer associated with the *TCRA* locus and is both necessary and sufficient to provide lineage- and stage-specific *TCRA* rearrangements and expression. The Eα activates transcription originating from the T early α (*TEA*) promoter, located upstream to the Jα gene segments (100 kb upstream to Eα), thus preparing the 5′ Jα chromatin state for the initial recombination events (Abarrategui and Krangel, 2009; Hawwari and Krangel, 2005; Villey et al., 1996). As such, the Eα enhancer represents a paradigm of gene regulation whereby a nucleoprotein complex is assembled during T cell differentiation in order to drive timely and stage-specific rearrangements of the *TCRA* locus (Carico and Krangel, 2015; Giese et al., 1992, 1995; Hernández-Munain et al., 1999; Roberts et al., 1997; Spicuglia et al., 2000). However, the factor, either activators or repressors, governing the stage-specific activation of Eα remains elusive.

To investigate the epigenetic mechanisms guiding the T cell identity, we generated epigenomic and transcriptomic data within the BLUEPRINT Epigenome Project (Stunnenberg and Hirst, 2016), including whole-genome bisulfite sequencing (WGBS) DNA methylation maps, genome-wide maps of six histone modifications, and RNA expression of key thymic subpopulations of human T cell precursors. The integration of other epigenomic dataset from BLUEPRINT and Roadmap, covering the development of the major branches of the hematopoietic system, provided a comprehensive view of the epigenomic dynamics at play at crucial transitions of T cell commitment and development. We observed that most of the distal regulatory elements are constitutively unmethylated throughout T cell differentiation, irrespective of their activation status. This was best exemplified by the Eα enhancer that was found to be in an open and unmethylated chromatin structure well before the activation of the *TCRA* gene.

Our thorough gene expression analysis revealed that the HOXA5-9 transcription factors are down-regulated concomitantly to Eα activation and repress its activity. These findings highlight the key role of HOXA proteins in the epigenetic control of *TCRA* rearrangements and illustrate a general mechanism of "setting the stage" for orchestrated developmental progression as soon as the inhibition is removed.

Results

A reference epigenome of human early T cell differentiation

As part of the BLUEPRINT project, we generated reference epigenomes of sorted human thymocytes encompassing the main differentiation stages (Fig. S1, A and B), including immature DN CD34$^+$ (tCD34: CD34$^+$/CD3$^-$/CD4$^-$/CD8$^-$), EC (TCRαβ$^-$/CD3$^-$/CD4$^+$/CD8$^+$), LC (TCRαβ$^+$/CD3^{+low}/CD4$^+$/CD8$^+$), SP CD4 (SP4: TCRαβ$^+$/CD3$^+$/CD4$^+$/CD8$^-$), and SP CD8 (SP8: TCRαβ$^+$/CD3$^+$/CD4$^-$/CD8$^+$). Reference epigenomes comprise chromatin immunoprecipitation sequencing (ChIP-seq) for four histone marks positively associated with gene expression (H3K36me3, H3K4me1, H3K4me3, and H3K27ac) and two with silent chromatin (H3K27me3 and H3K9me3), WGBS, and RNA sequencing (RNA-seq; details in Table S2). Replicates from different individuals were merged and normalized to provide a consistent set of reference epigenome tracks for each differentiation stage.

To obtain a global view of the epigenomic landscape of human hematopoiesis, we integrated the ChIP-seq data of each thymic subpopulation into 11 chromatin states derived from the six histone modifications (Fig. S1 C) and compared them with chromatin states for other hematopoietic populations generated by the BLUEPRINT and Roadmap consortia. Multiple correspondence analysis (MCA) based on reduced chromatin states (Carrillo-de-Santa-Pau et al., 2017) allowed us to create a multidimensional space, where the different hematopoietic samples were placed depending on their chromatin state vectors across the genome. Clustering based on the first two components grouped samples from the same cell type while the main hematopoietic populations were clearly separated from each other, in agreement with previous results (Carrillo-de-Santa-Pau et al., 2017; Fig. 1 A). The first component revealed two main differentiation branches, representing the myeloid and lymphoid cell lineages, while the hematopoietic stem cells (HSCs) localized in a central position. The second component reflected the differentiation stage and the environmental context of each cell type with respect to the HSCs. Interestingly, the thymocyte subpopulations clustered tightly to the peripheral T cells, except for the most immature T cell precursors (tCD34), which clustered close to HSCs and complied with an incomplete commitment to the T cell lineage (Dik et al., 2005b; Fig. 1 A).

To assess whether the epigenomic states were consistent with the known biological functions of each cell type, we analyzed the enrichment in biological processes of genes associated with active enhancer states (Fig. S1 C). The enhancer regions found in each population were significantly enriched in pathways associated with the corresponding lineage (Fig. 1 B). For instance, thymic and peripheral T cells were associated with TCR signaling, B cells with B cell antigen receptor signaling, and myeloid lineages with phagocytosis and innate immunity.

Lineage-specific genes displayed the expected chromatin dynamics across the hematopoietic populations (Fig. 1 C). The stem cell marker *CD34* was associated with open chromatin states in HSC and tCD34 early T cell precursors; the CD3 cluster of T cell receptors was open in T cell samples; the B cell master transcription factor *PAX5* was associated with open chromatin in B cells; the myeloid marker *CD33* was associated with open chromatin in HSCs, monocytes, and neutrophils but only transcribed in the myeloid lineage. Thus, the epigenomic states, including those from the newly generated thymocyte populations, were able to capture the main biological differences between cell types and were consistent with their known underlying biological processes.

We next investigated more deeply the epigenetic dynamics during human T cell differentiation. Chromatin states clearly separated the main stages of thymic T cell differentiation from HSC and peripheral T cells (Fig. 1 D), suggesting a specific chromatin signature of thymic T cell precursors. While the first dimension (separating T cells from HSC and tCD34 precursors)

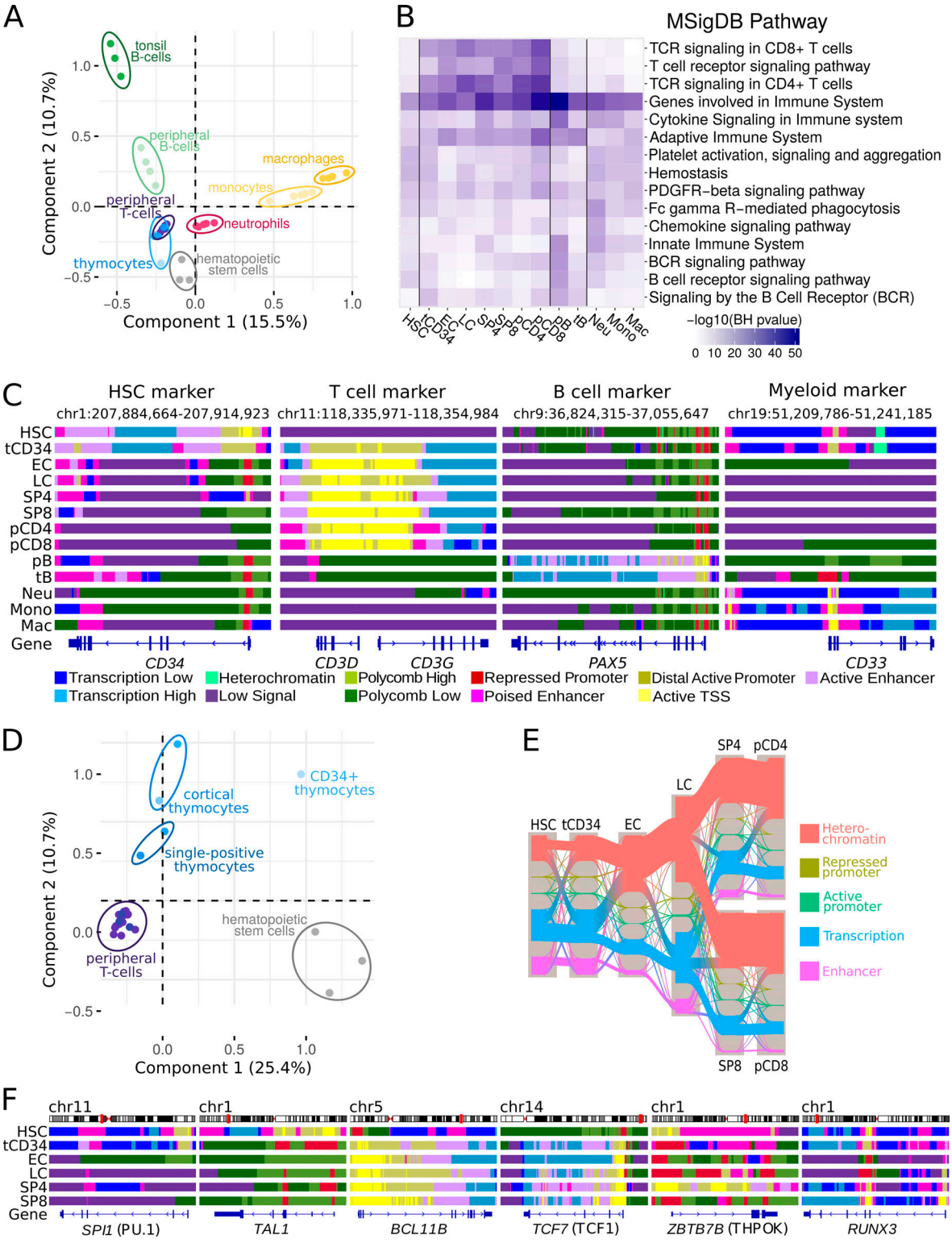


Figure 1. **Reference epigenomes of human hematopoietic and T cell precursors. (A)** MCA of human hematopoietic samples based on chromatin states. Ovals highlight the major subtypes of hematopoietic lineages. **(B)** Top enriched MSigDB pathways associated with active enhancers found in each hematopoietic population of A. Vertical lines delineate the main hematopoietic lineages. **(C)** Chromatin state profiles at representative genes of the different hematopoietic lineages. The color code is indicated at the right of the panel. **(D)** MCA of HSC and human T cell populations based on chromatin states. **(E)** Sankey diagram depicting the main epigenomic transitions between the HSC and the human T cell populations. The width of the arrows is proportional to the number of regions associated with the indicated chromatin state. The color code represents the merged chromatin states as indicated at the right of the panel. **(F)** Chromatin state profiles at representative genes specifically expressed at different stages of T cell differentiation. The color code is as in C. BH, Benjamini–Hochberg. Mac, macrophages; Neu, neutrophils; Mono, monocytes.

was enriched with cell homeostasis and activation functions, the second dimension (separating the thymocyte populations from peripheral T cells and HSC) was specifically enriched with V(D)J recombination (Fig. S1 D). To get a global view of the main epigenetic changes occurring during T cell differentiation, we represented the chromatin state transitions among HSCs, thymic subpopulations, and peripheral T cells using a Sankey diagram (Fig. 1 E). We observed a progressive increase of heterochromatin regions along with a decrease of active enhancers and transcribed regions. This suggests a gradual loss of plasticity from HSCs to mature T cells. Finally, master transcription factors that are either repressed (*SPI1*/PU.1, *TAL1*) or activated (*BCL11B*, *TCF1*/TCF7) during T cell commitment or required for specific CD4$^+$ (*ZBTB7B*/THPOK) or CD8$^+$ (*RUNX3*) T cell differentiation displayed consistent epigenomic changes (Fig. 1 F).

Overall our analysis of the hematopoietic reference epigenomes recapitulated the hematopoietic chromatin signatures and accurately captured the main biological identities of thymic T cell precursors. It demonstrates that our epigenomic data provide an excellent resource to study the regulatory networks underlying early T cell differentiation in humans.

DNA demethylation is a hallmark of distal regulatory elements irrespective of their activation status in T cells

We next analyzed the DNA methylation status of the five thymic subpopulations and the different hematopoietic populations using WGBS data generated by both the BLUEPRINT and Roadmap consortia. Similar to the chromatin state classification, a dimensional reduction based on DNA methylation grouped the different hematopoietic samples according to their cell lineage, while the thymocyte populations were separated from peripheral T cells (Fig. S2 A).

Gene distal DNA hypomethylation has been shown to be a hallmark of enhancer regions, allowing an accurate identification of recognized cis-regulatory sequences (Stadler et al., 2011). To assess the dynamics of enhancer activities during early T cell differentiation, we retrieved all distal hypomethylated regions found in at least one thymic subpopulation and clustered these regions based on the changes of DNA methylation across the thymic subpopulations (Fig. 2 A and Fig. S2 B). Clusters 1–4 corresponded to constitutively hypomethylated regions and represented the majority of distal hypomethylated regions. Cluster 5 corresponded to regions preferentially hypomethylated in tCD34 precursors. Regions of clusters 1–5 were poorly enriched in H3K27ac (Fig. 2 B), and the associated genes were expressed at low levels in thymocytes (Fig. 2 C), displayed low thymus specificity (Fig. 2 D), and were either related to hematopoietic lineages other than T cells or not specifically enriched in relevant biological processes (Fig. 2 E). In contrast to clusters 1–5, clusters 6–8 corresponded to regions that were hypermethylated in tCD34 thymocytes and underwent progressive demethylation through different T cell differentiation stages. Consistent with this, clusters 6–8 were highly enriched with H3K27ac mark (Fig. 2 B), and the associated genes were expressed at a high level in thymocytes (Fig. 2 C), displayed higher thymus specificity (Fig. 2 D), and were related to T cell–specific functions (Fig. 2 E). The results suggest that clusters 6–8 are likely enriched in T cell–related genes that are expressed at all stages, but for which the relative expression level is fine-tuned during differentiation.

As we were surprised by the elevated proportion of constitutively hypomethylated regions (77% of total distal hypomethylated regions), we decided to further explore the epigenomic dynamics of these potential regulatory regions. Clustering of the constitutively hypomethylated distal DNA regions (corresponding to clusters 1–4 from Fig. 2 A), based on H3K27ac levels (Fig. S2, C and D), revealed five clusters with dynamic H3K27 acetylation (Fig. 2, F and G), representing 17% of the total set of distal hypomethylated regions. Changes in H3K27ac levels roughly followed the expression dynamics of neighboring genes (Fig. 2 H). Consistent H3K27ac signatures at the promoter and constitutively hypomethylated distal regions were observed for early (*RAG1-RAG2*) and late (*CD44*) expressed loci. The *Il7R* locus was associated with distal enhancers activated either at the early (downstream) or the late (upstream) stages; in both cases, the regulated enhancers were constitutively demethylated (Fig. 2 I).

These observations strongly indicate that DNA hypomethylation marks distal regulatory elements independently of their activation status, both preceding to and persisting after enhancer activation.

Chromatin opening precedes enhancer activation

The above analyses suggested that the activation onset of distal regulatory elements might be preceded by DNA demethylation. This raised the question of whether premarked enhancers (DNA hypomethylated) are already in an open chromatin configuration (nucleosome free). To address this issue, we performed assay for transposase-accessible chromatin (ATAC) experiments in tCD34 and EC stages, as these stages reflected the major epigenetic transitions (Fig. 1 and Fig. 2). We identified 2,084 and 2,905 distal ATAC peaks in tCD34 and EC, respectively (Fig. 3, A and B). About one-third of the distal ATAC peaks (1,108) were shared between the two thymic populations. Distal ATAC peaks present in both tCD34 and EC stages and specific to EC were in proximity to genes associated with T cell differentiation and function, while tCD34-specific peaks were associated with more general functions (Fig. 3 C). Consistent with this, analysis of conserved transcription factor binding sites in tCD34-associated peaks showed a preferential association of hematopoietic factors such as TAL1 and AML1 (RUNX1), while transcription factors involved in T cell differentiation and activation such as E47 (TCF3), STAT, and ETS were preferentially found in EC-associated peaks (Fig. 3 D). As predicted, ATAC peak regions were constitutively demethylated in tCD34 and EC stages, independently of whether the regions were specifically open in tCD34 or EC thymocytes (Fig. 3 E). This suggests that DNA demethylation may both precede enhancer opening and hold after enhancer closing. Thus, DNA demethylation seems to represent a primary marking of distal regulatory regions, independently of the actual enhancer activity or chromatin accessibility.

Next, we asked whether constitutively open and demethylated regions could be associated with changes in enhancer

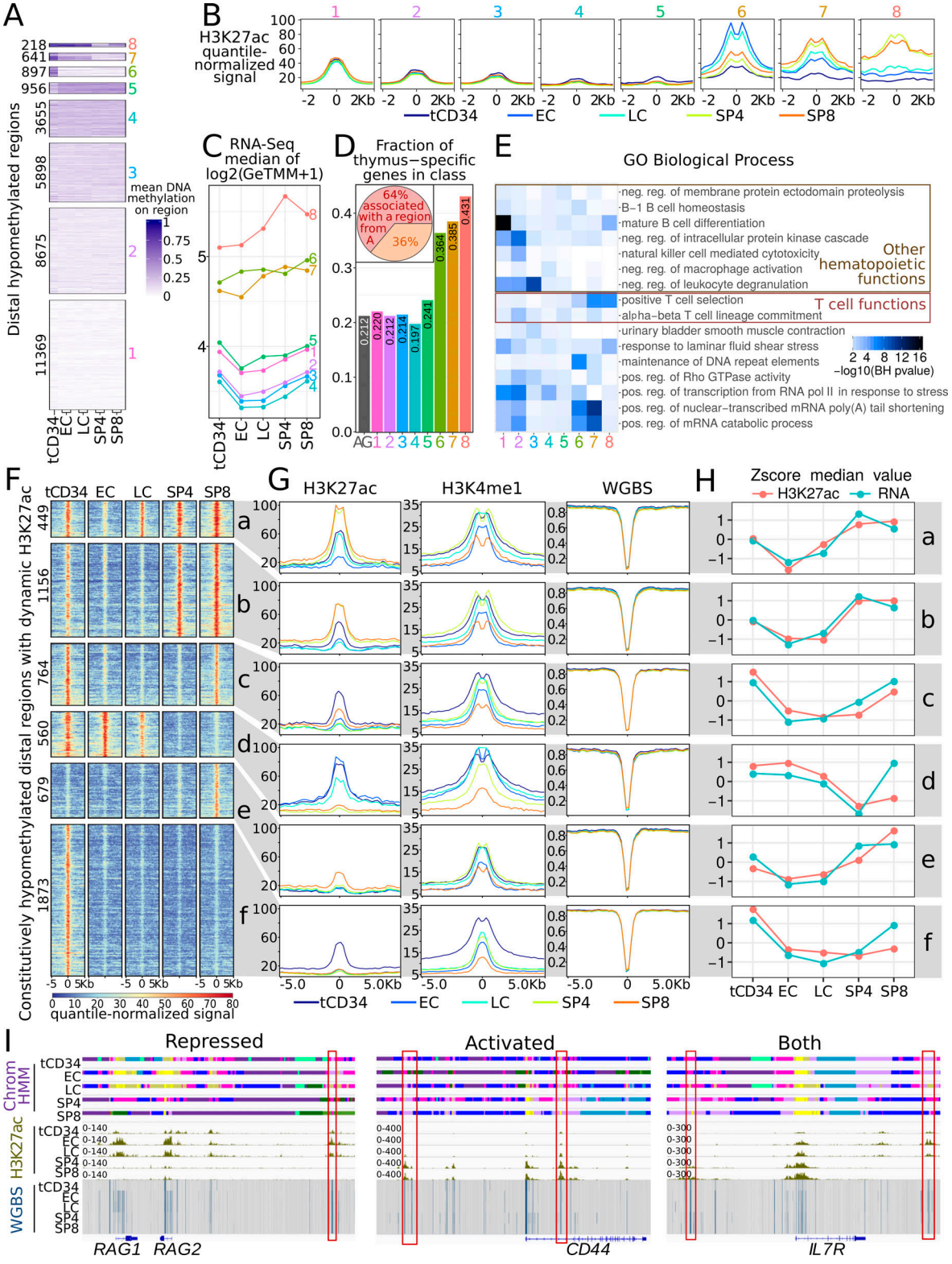


Figure 2. **DNA demethylation is a hallmark of distal regulatory elements irrespective of their activation status in T cells. (A)** Heatmap displaying DNA methylation scores of distal hypomethylated regions in the indicated thymocyte populations, based on WGBS. Regions are clustered into eight groups using k-means. The scale of DNA methylation is indicated at the bottom of the panel. **(B)** H3K27ac profiles in the indicated thymocyte populations, centered on the hypomethylated clusters defined in A. **(C)** Median of the RNA-seq signal of genes associated with the hypomethylated clusters defined in A, in the indicated thymocyte populations. **(D)** Fraction of thymus-specific genes associated with each hypomethylated cluster defined in A and in all genes (AG). The Human Protein Atlas version 19.2 (Uhlén et al., 2015) was used to assess tissue specificity. Inset shows the percentage of thymus-specific genes that is associated with a distal hypomethylated region. **(E)** Top GO biological processes enriched at genes associated with the hypomethylated clusters defined in A. Relevant hematopoietic and T cell related terms are highlighted. Labels are abbreviated as follows: neg, negative; pos, positive; reg, regulation; pol, polymerase.

(F) K-means clustering of DNA hypomethylated regions with dynamics H3K27ac levels among thymocytes subpopulation. The heatmap of H3K27ac levels in the indicated thymocytes subpopulation is shown. **(G)** H3K27ac, H3K4me1, and WGBS profiles in the indicated thymocyte populations centered on the clusters defined in E. Same color code as in B. **(H)** Z-scores of the median H3K27ac signal of the regions defined in E and RNA-seq signal of associated genes in the indicated thymocyte populations of genes associated with clusters defined in E. **(I)** Chromatin states, H3K27ac, and WGBS profiles of representative loci displaying constant hypomethylation, but dynamic H3K27ac enrichments in the indicated thymocyte populations (red rectangles). The regulation of the associated gene, specifically expressed at different stages of T cell differentiation, is indicated at the top of each panel. The normalized signal range is indicated by brackets. For WGBS track, red and blue colors indicate methylated and unmethylated CpGs, respectively. The color code for the ChromHMM track is as in Fig. 1 C.

activity. For this purpose, shared ATAC peaks were ordered relative to the H3K27ac ratio between tCD34 and EC thymocytes (Fig. 3 F). Interestingly, ATAC regions that are most highly H3K27ac marked in tCD34 (tCD34 high) or EC (EC high) cells (Fig. 3 G) were linked with genes that significantly decrease or increase expression, respectively (Fig. 3 H; paired Wilcoxon test). Concordant proportions of differentially regulated genes were also observed (Fig. 3 I). This suggests that constitutively open and demethylated distal regions can be associated with locus activation or repression.

To identify potentially interesting loci, we searched for genes showing strong differential expression between tCD34 and EC/LC thymocytes that were also associated with distal open regions displaying a gain of H3K27ac (Fig. 3 J and Fig. S3). Strikingly, the *TRAC* transcript, associated to the *TCRA* locus, was the most induced gene having an H3K27ac enhancer gain in a region corresponding to the known Eα enhancer (Hawwari and Krangel, 2005; Ho et al., 1989; Winoto and Baltimore, 1989; Fig. 4 A).

The Eα enhancer is in an open but epigenetically silent configuration in immature thymocytes

To better understand the tight regulation of the Eα enhancer, we depicted the epigenetic dynamic at the *TCRA* locus during the human thymopoiesis (Fig. 4 A). In tCD34 thymocytes, the *TCRA* locus was in a relatively closed configuration, with low levels of H3K4me1, H3K4me3, and H3K27ac and relatively high levels of H3K27me3. In contrast, EC and LC cortical thymocytes displayed overall high levels of H3K4me1, H3K4me3, H3K27ac, and H3K36me3, whereas H3K27me3 was strongly reduced. Of note, active histone marks were shifted from the 5′ to the 3′ side of the Jα-Cα locus between the EC and LC stages, in agreement with progressive Vα-to-Jα rearrangements (Hawwari and Krangel, 2005; Huang et al., 2002; Krangel et al., 2004). The adjacent *TCRD* locus displayed a general loss of active marks during T cell differentiation, as expected (Fig. 4 A). In contrast to the *TCRA* locus, the Eα enhancer was found to be in an open chromatin configuration since the very immature stages of thymic maturation (ATAC), when it was fully demethylated (WGBS) and enriched in H3K4me1 (a mark of poised enhancers; Fig. 4, A and B). By the EC and LC stages on, the Eα region had gained substantial levels of H3K27ac and, in particular, H3K4me3, previously shown to be associated with highly active enhancers (Pekowska et al., 2011).

Our findings are consistent with murine data showing that Eα is occupied by transcription factors from the early $CD4^-/CD8^-$ DN stage of thymocyte development, thus preceding the transcriptional activation of *TCRA* locus, which takes place after the β-selection (Hernández-Munain et al., 1999; Spicuglia et al., 2000). We have previously suggested that the Eα enhancer activation occurs via a conformational change of a preassembled nucleo-protein complex (Cauchy et al., 2016; Spicuglia et al., 2000). Analysis of published ChIP-seq experiments for ETS1, RUNX1, E47/TCF3, GATA3, and Ikaros lymphoid transcription factors (Zacarías-Cabeza et al., 2015; Koch et al., 2011; Lepoivre et al., 2013; Wei et al., 2011; Vanhille et al., 2015; Oravecz et al., 2015) confirmed that Eα is already occupied by these activating transcription factors in murine DN thymocytes (Fig. S4 A). Importantly, transcription of short noncoding enhancer RNAs, a hallmark of active enhancers, was only observed in murine late $CD4^+/CD8^+$ double positive (DP) thymocytes (Fig. S4 B). Furthermore, as shown by ChIP–quantitative PCR (ChIP-qPCR) assays (Fig. 4, C and D), Eα is already bound by its activating ETS1 and RUNX1 transcription factors in human immature tCD34 thymocytes. This was also confirmed for RUNX1 by ChIP-seq experiment (Fig. 4 E).

Overall, these data suggest that Eα is in an open chromatin state and is already bound by its activating transcription factors from the very immature stages of human and murine thymic maturation. As Eα remains inactive until the cortical stages, this suggests that a repressive mechanism prevents its activity during the early stages of thymic maturation in mice and humans. While the functions of Eα have been intensively studied (Carey, 1998; Hawwari and Krangel, 2005; Ho et al., 1989, 1990; McMurry and Krangel, 2000), the molecular mechanisms mediating the stage-specific activation of Eα have remained elusive. As such, Eα control of the *TCRA* locus represents an unsolved paradigm of tissue-specific gene regulation by distal regulatory elements.

The *HOXA* locus is progressively repressed during early T cell differentiation

Given that Eα is already bound by key transcription factors required for its activation, we hypothesized that the loss of a transcriptional repressor might explain the stage-specific activation of Eα. To explore this possibility, we analyzed RNA-seq of the main human thymic subpopulations to identify transcripts that were down-regulated at the TCR β-selection. Unsupervised gene expression analysis identified 19 clusters, each corresponding to transcripts with different gene expression kinetics during thymic maturation (Fig. 5 A). Cluster C13, containing 893 different transcripts, presented the expected profile of gene expression modulation, with high expression at the early stages and gradual reduction during maturation up to the extinction of expression after the β-selection stage (immature single positive [ISP] > EC > LC; Fig. 5 A). Interestingly, this cluster

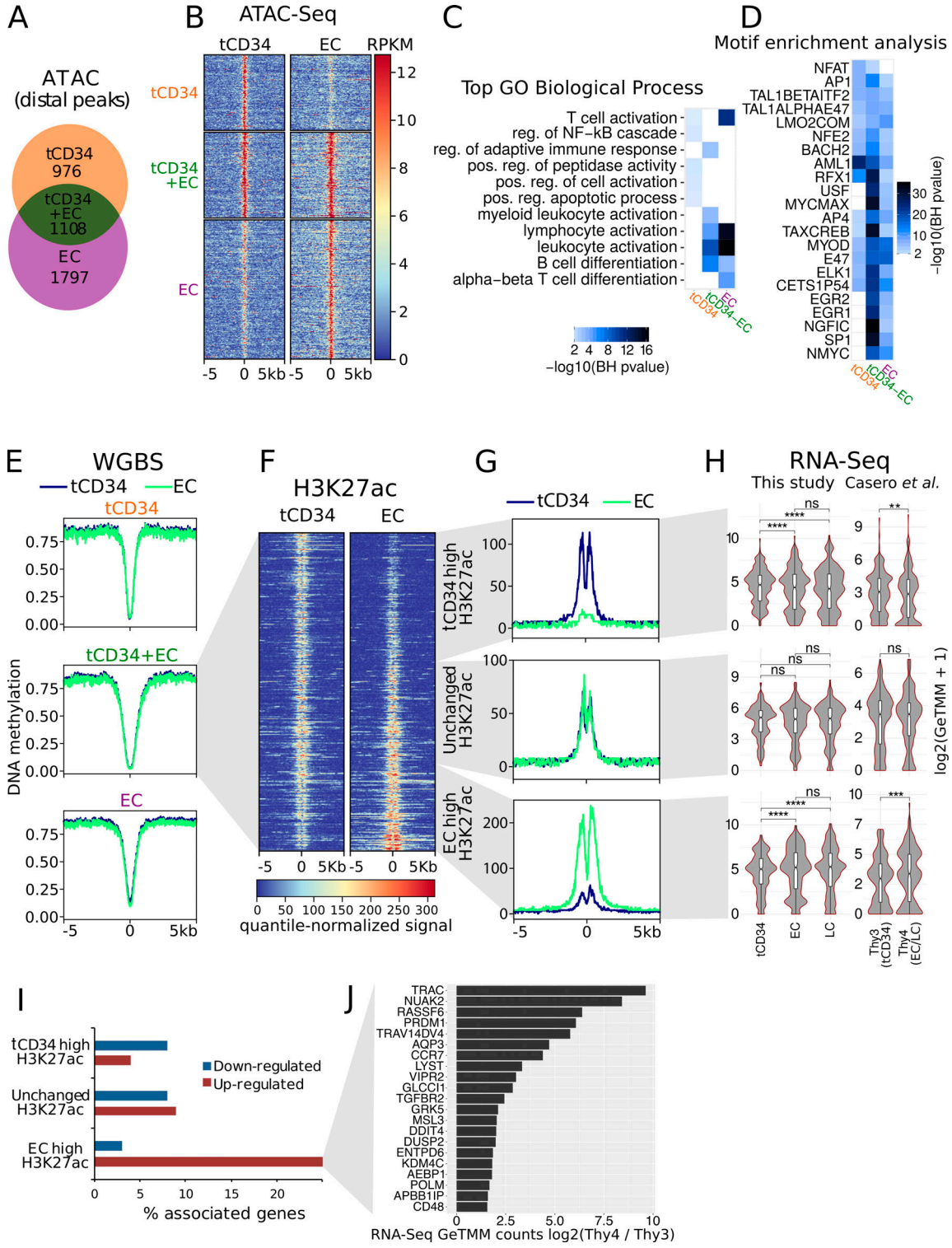


Figure 3. **Chromatin opening precedes enhancer activation. (A)** Overlap between ATAC peaks identified in tCD34 and EC thymocytes. **(B)** Clustered heatmap of ATAC signal in tCD34 and EC thymocytes (left panels). Scores associated with the color code is on the right. **(C)** Top GO biological processes enriched in genes associated with distal ATAC peaks specific to either tCD34 or EC thymocytes, or common to both. **(D)** Conserved TFBS enriched at distal ATAC peaks specific to either tCD34 or EC thymocytes, or common to both. **(E)** Average profiles of WGBS signals corresponding to the ATAC peaks defined in A. **(F)** Heatmap of H3K27ac signals based on common ATAC peaks ordered by the H3K27ac ratio between tCD34 and EC. **(G)** Average H3K27ac profiles centered on ATAC peaks with relative high H3K27ac signal in either tCD34 or EC thymocytes or remained unchanged, as defined in E. **(H)** Violin plots of the RNA-seq signal of genes associated with distal ATAC peaks (as defined in E) in the indicated thymocyte populations. Right panels represent RNA-seq data from Casero et al. (2015). Statistical significance was assessed by a Wilcoxon rank-sum test. **, P ≤ 0.01; ***, P ≤ 0.001; ****, P ≤ 1e-04. **(I)** Percentage of significantly regulated genes associated with distal ATAC peaks defined in E. **(J)** Top significantly induced genes associated with ATAC peaks and H3K27ac gain. The expression fold change between Thy3 (tCD34) and Thy4 (EC/LC) thymocytes is indicated.

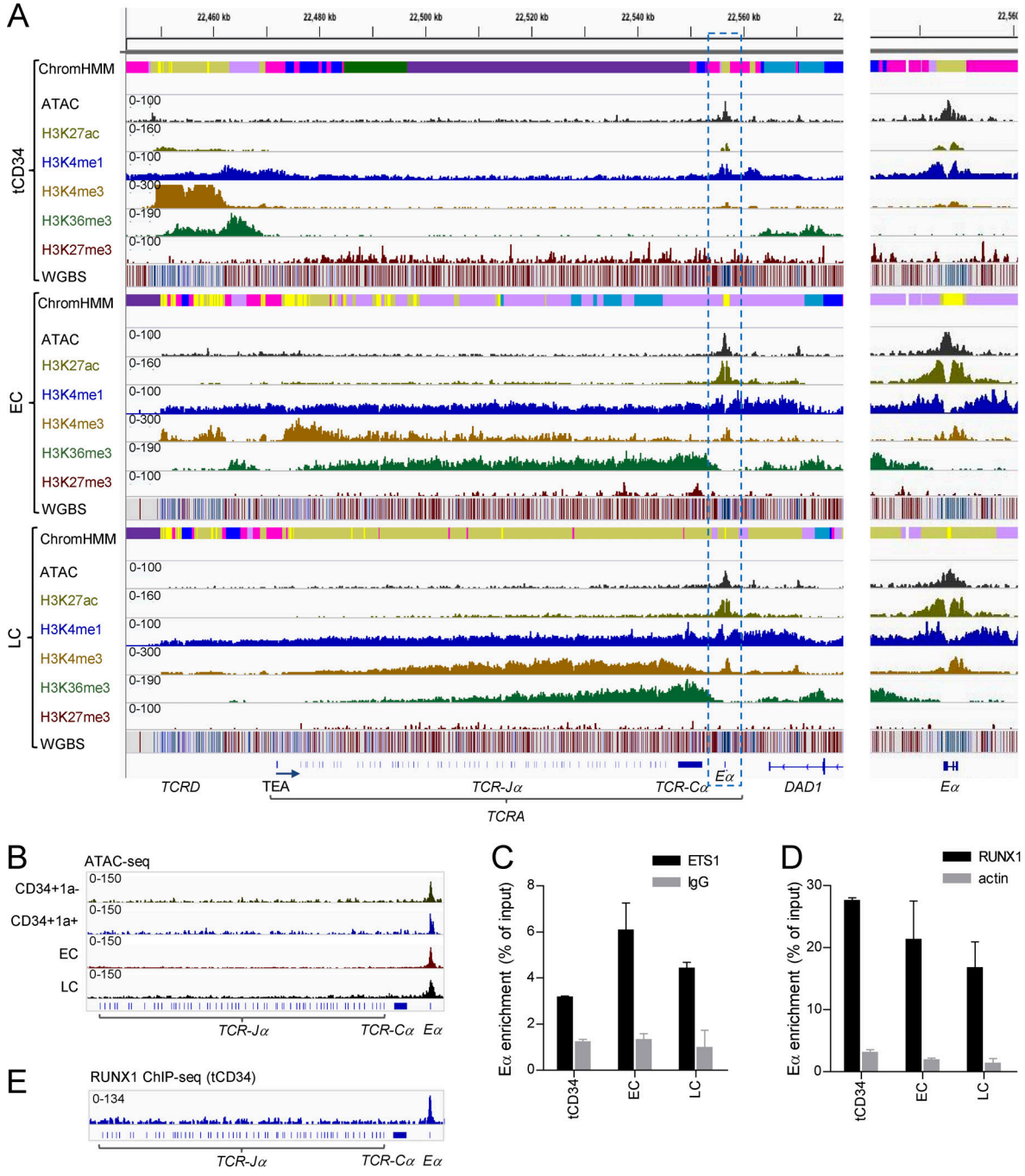


Figure 4. **Epigenetic and transcription factor analysis of the *TCRA* locus in the human thymic subpopulations. (A)** ChromHMM, ATAC-seq, and ChIP-seq profiles for histone modifications (H3K27ac, H3K4me1, H3K4me3, H3K36me3, and H3K27me3) and BS-seq at the *TCRA/D* locus in the human thymic subpopulations tCD34, EC, and LC. A zoom on Eα is presented on the right. The legend is as in Fig. 2 H. **(B)** ATAC profiles of the *TCRA* locus (*TRAJ* gene segments and *TCRA* enhancer) in human thymic subpopulations $CD34^+1a^-$, $CD34^+1a^+$, EC, and LC. Representative of two independent experiments. **(C)** ChIP-qPCR analysis of ETS1 binding to Eα in human thymic subpopulations tCD34, EC, and LC. ChIP enrichment level is presented as percent of input and corresponds to the ratios between the Eα and IgG signal. Data are presented as duplicates of experiments with error bars to represent SEM. **(D)** ChIP-qPCR analysis of RUNX1 binding to Eα in human thymic subpopulations tCD34, EC, and LC. ChIP enrichment level is presented as percent of input and corresponds to the ratios between the Eα and actin signal. IgG isotype control was performed to assess absence of nonspecific Eα ChIP enrichment. Data are presented as duplicates of experiments with error bars to represent SEM. **(E)** RUNX1 ChIP-seq profile in thymic $CD34^+$ population (tCD34).

contained several members of the HOX-like (HOXL) subclass of homeodomain (Hox) transcription factors (Fig. 5 B and Fig. S5, A and B). Of note, ectopic expression (i.e., oncogenic deregulation by chromosomal translocations) of the HOX family of transcription factors (i.e., TLX1 and TLX3 from the NKL subclass) was reported to exert repressive activity on Eα in T cell acute lymphoblastic leukemia (T-ALL), an aggressive cancer developed from thymic precursors (Dadi et al., 2012). We then

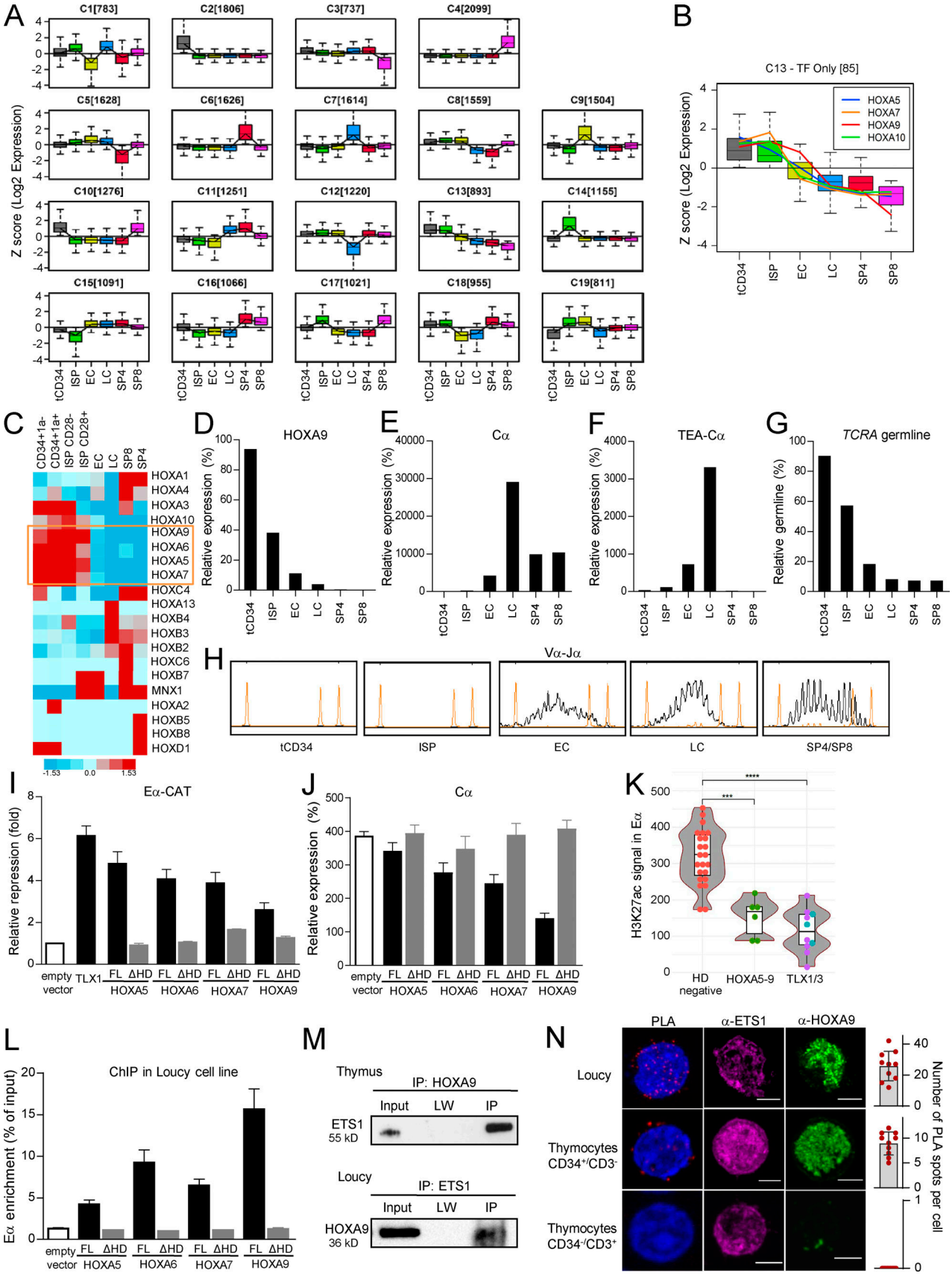


Figure 5. **HOXA5-9 proteins are progressively repressed during human thymopoiesis, bind to Eα, and repress its activity. (A)** An unsupervised RNA-seq gene expression classification identified 19 clusters with different profiles of gene expression modulations. **(B)** Cluster 13 identifies 85 transcription factors in

which expression progressively decreases after β-selection (ISP). The expression of the *HOXA5*, *HOXA7*, *HOXA9*, and *HOXA10* genes is highlighted. **(C)** TLDA analysis of HOXL expression profiles in thymic subpopulations. The expression level is classified from blue to red. Only genes that show expression at any of the thymic subpopulation are presented. Results represent means duplicate reactions. **(D)** RQ-PCR analysis of *HOXA9* gene expression in human thymic subpopulations. Results are relative to the *ABL1* housekeeping gene. Representative of two independent experiments. **(E)** RQ-PCR analysis of *Cα* gene expression in human thymic subpopulations. Results are relative to the *ABL1* housekeeping gene expression. Representative of two independent experiments. **(F)** RQ-PCR analysis of *TEA-Cα* gene expression in human thymic subpopulations. Results are relative to the *ABL1* housekeeping gene. Representative of two independent experiments. **(G)** RQ-PCR analysis of *TCRA* germline from normal human thymocyte subpopulations. Normalization was performed with the albumin housekeeping gene. Representative of two independent experiments. **(H)** Fluorescent PCR Genescan analysis of Vα-Jα rearrangements detected in sorted thymic populations. Representative of two independent experiments. **(I)** Eα-CAT fold repression in HeLa cells following transfection with encoding vectors for ETS1, RUNX1, LEF1, and individual HOXA5-9 (FL and HD-deleted forms). TLX1 is used as a positive control of repression. Eα-CAT signals were normalized to those of control cells transfected with an empty GFP vector. Data are presented as means of at least triplicate measurements with error bars to represent SEM. **(J)** RQ-PCR analysis of *Cα* gene expression in T-ALL cell line transfected with HOXA5-9 (FL or ΔHD forms)–expressing vectors. Results are relative to the *ABL1* housekeeping gene expression. Representative of two independent experiments. **(K)** Violin plots of quantile-normalized H3K27ac signal in Eα in T-ALL patients. Immature T-ALLs (IM0, IMD, and IMB) were excluded. Red dots: T-ALLs with no deregulation of homeodomain genes (n = 23); green dots: T-ALLs overexpressing *HOXA-5*, *-6*, *-7*, and *-9* genes (n = 6); blue dots: *TLX1* positive T-ALLs (n = 3); violet dots: *TLX3* positive T-ALLs (n = 7). Statistical significance was assessed by a Wilcoxon rank-sum test. ***, P ≤ 0.001; ****, P ≤ 1e-04. **(L)** Anti-Flag ChIP-qPCR analysis of HOXA5-9 binding to Eα in the Loucy cell line expressing Flagged HOXA5-9 (FL or ΔHD forms). ChIP enrichment level is presented as a percentage of input and corresponds to the ratios between the Eα and actin signals. ChIP with an empty NTAP vector was performed as a control. Results represent means and SEM of triplicate reactions. **(M)** Upper panel: CoIP using HOXA9 antibody followed by an immunoblotting with anti-ETS1 antibody in human thymocytes. The input corresponds to the 1% of cell extract used in the CoIP experiment. LW, last wash. A representative result of two independent experiments is shown. Lower panel: IP using anti-ETS1 antibody followed by an immunoblotting with anti-HOXA9 antibody in Loucy cell line. A representative result of two independent experiments is shown. The input corresponds to the 1% of cell extract used in the CoIP. **(N)** Duolink PLA and confocal microscopy analysis of cells labeled with anti-ETS1 (Alexa 488, purple) and anti-HOXA9 (Alexa 555, green) antibodies in Loucy cell line and tCD34$^+$/CD3$^-$ and tCD34$^-$/CD3$^+$ thymocytes. Scale bars, 10 µm. On the right, the quantification of the number of PLA dots per cell is presented with the mean values and SD.

focused on HOXL genes pattern expression during human thymopoiesis. A TaqMan Low-Density Array (TLDA) was designed to evaluate the expression of all HOXL members in an independent set of human thymic subpopulations (Table S1). Unsupervised TLDA clustering showed that *HOXA5*, *HOXA6*, *HOXA7*, and *HOXA9* genes best fitted the relevant expression profiles (Fig. 5 C and Fig. S5, C–E). In particular, HOXA9 appeared to be a good candidate, as its down-regulation perfectly coincides with the Cα up-regulation, the presence of TEA-Cα germline transcripts, and the beginning of *TCRA* rearrangements (Fig. 5, D–H).

HOXA5-9 proteins repress Eα activity via their homeodomain

To test the possibility that HOXA5-9 proteins interfere with Eα transcriptional activity, we used a gene reporter assay in which the expression of chloramphenicol acetyltransferase (CAT) is under the control of Eα (Eα-CAT) (Giese et al., 1995). The ectopic expression of HOXA-5, -6, -7, and -9 repressed Eα-regulated CAT expression by three- to fivefold (Fig. 5 I). Next, we evaluated the level of Eα-CAT repression of the truncated forms of HOXA proteins, lacking their homeodomains (ΔHD). As shown in Fig. 5 I, all truncated HOXA forms (ΔHD) exerted significantly reduced repressive activity compared with their respective full-length (FL) proteins. The reduced activity of HOXA5-9 ΔHD proteins was not due to their lack of nuclear localization because they were mainly localized in the nucleus as their FL counterpart (Fig. S5 F). Using the same expression vectors, we tested whether HOXA5-9 proteins can repress *TCRA* activity in the HOXA-negative and TCRαβ-positive T-ALL cell line Jurkat. As shown in Fig. 5 J, HOXA9 overexpression resulted in the strongest repression of the *TCRA* gene expression. Additionally, we found that H3K27ac signal, a hallmark of active enhancers, is significantly lower in Eα in HOX-overexpressing T-ALLs, compared with T-ALLs without deregulation of homeodomain genes (Fig. 5 K). We concluded that HOXA5-9 proteins repress Eα transcriptional activity in a homeodomain-dependent manner.

ChIP assays using Loucy (HOXA overexpressing) or HeLa (HOXA-negative) cell lines nucleotransfected with Flag-tagged expressing vectors for FL *HOXA5*, *-6*, *-7*, or *-9*, showed significantly enriched Eα DNA (Fig. 5 L and Fig. S5 G). Importantly, HOXA proteins without their homeodomain (ΔHD) displayed reduced binding to Eα. As Eα lacks the AT-rich DNA motif TAATNA characteristic of the HOX homeodomain binding, we hypothesized that HOXA proteins exert their Eα repressive activity by interacting with ETS1, as we showed for TLX proteins, members of the NKL subclass of homeodomain proteins (Dadi et al., 2012). Co-immunoprecipitation (IP) assays with anti-HOXA9 antibody in human thymocytes recovered both ETS1 and HOXA9 in the precipitated material (Fig. 5 M, upper panel). CoIP assays in the Loucy cell line, which coexpresses HOXA9 and ETS1, also showed that an anti-ETS1 antibody coimmunoprecipitated HOXA9 (Fig. 5 M, lower panel). Furthermore, proximity ligation assay (PLA) confirmed the HOXA9-ETS1 interaction in Loucy and in immature CD34$^+$/CD3$^-$ thymocytes (Fig. 5 N). However, more mature CD34$^-$/CD3$^+$ thymocytes, as expected, showed no HOXA9-ETS1 interaction (Fig. 5 N).

Taken together, these data suggest that the HOXA9 transcription factor could be recruited to Eα and repress the Eα activity in a homeodomain-dependent manner.

HOXA9 overexpression affects T cell development in humans and mice

To functionally evaluate the role of HOXA homeoproteins in *TCRA* rearrangement and T cell development, we transduced human CD34$^+$ umbilical cord blood cells (UCB CD34$^+$) with either a HOXA9 overexpressing (HOXA9-GFP) or a control (GFP) vector and co-cultured them in vitro on OP9-DL1 stroma under conditions for T cell differentiation. After 7 wk of co-culture, we

found more TCRγδ-expressing cells when HOXA9 was overexpressed (22.0 ± 11.6%) compared with control cells (10.8 ± 4.6%; Fig. 6, A–C). The difference was significantly pronounced after 8 wk of co-culture (HOXA9-GFP 26.8 ± 15.3% vs. control GFP 6.8 ± 3.1%; Fig. 6, A–C). TCRγδ cells in HOXA9-overexpressing conditions were mostly $CD4^+/CD8^+$ DP compared with the TCRγδ cells in controls (Fig. 6 D). We then analyzed the *TCRA* rearrangements and observed fewer *TCRA* rearrangements under HOXA9-overexpressing conditions, with a clear difference by the fifth week of culture (Fig. 6 E). Additionally, using the CRISPR-Cas9 system, we performed knock-out of HOXA9 and deletion of HOXA5-9 genes in $CD34^+$ UCB cells, which were differentiated on OP9-DL1 stroma cells. *TCRA* gene rearrangements were analyzed at day 28 of co-culture, demonstrating a substantial increase in HOXA9 knock-out and HOXA5-9 deletion conditions as compared with controls (Fig. 6 F).

To confirm the role of the HOXA9 protein in *TCRA* rearrangements and T cell development, we performed in vitro and in vivo experiments in mice. Murine DN thymocytes were transduced with either a control vector (GFP) or a HOXA9 overexpressing vector (HOXA9-GFP) and differentiated in vitro on OP9-DL1 stromal cells (Fig. 7 A). In control conditions and following 9 d of co-cultures, 44.4 ± 6.8% of the immature thymocytes developed into TCRαβ-expressing cells. When HOXA9 was overexpressed, however, few thymocytes (14 ± 7%) differentiated into TCRαβ cells, developing instead into $CD4^+/CD8^+$ DP TCRγδ cells (Fig. 7, B–E). To confirm these in vivo data, we transplanted HOXA9-overexpressing DN thymocytes and their control counterparts into sub-lethally irradiated $Rag2^{-/-}\gamma c^{-/-}$ mice (Fig. 7 F). At 3 wk after transplant, control cells became virtually exclusively TCRαβ-positive, whereas HOXA9-overexpressing thymocytes repressed TCRαβ expression and expressed TCRγδ in a significant proportion of T cells (Fig. 7, G–I).

Taken together, these results indicated that the enforced expression of HOXA9 blocks *TCRA* rearrangements, thus affecting αβ T cell development.

Homeodomain protein deregulation leads to a TCRγδ bias in human T-ALL

In an effort to evaluate the implication of the homeoprotein-mediated inhibition of Eα activity, we took advantage of the T-ALL model. T-ALL is a rare disease resulting from the leukemic transformation of thymic precursors arrested at specific stages of differentiation (Asnafi et al., 2003; Ferrando et al., 2002). Approximately 30% of T-ALLs have a "mature" phenotype as they express the sCD3/TCR. Intriguingly, among these $sTCR^+$ T-ALLs, an unexpectedly high fraction of cases (≅50%) express a TCRγδ. The molecular mechanism of this bias was unidentified (Asnafi et al., 2004; Trinquand et al., 2016). We analyzed a series of 150 TCR-expressing T-ALLs (45 TCRαβ T-ALLs vs. 95 TCRγδ T-ALLs) and evaluated HOXA9 expression. The TCRγδ T-ALLs showed higher (39/95; 41%) HOXA9 overexpression compared with TCRαβ T-ALLs (Fig. 8 A). A further 32/95 (34%) TCRγδ T-ALLs overexpress homeodomain genes from the NKL subclass (i.e., *TLX1* or *TLX3*). In total, 71/95 (75%) TCRγδ T-ALLs demonstrated overexpression of homeodomain genes (HD-overexpressing), compared with 5/45 TCRαβ T-ALLs ($P < 0.0001$). We hypothesized that HD overexpression could impose TCRγδ expression in αβ-lineage oriented precursors by preventing TCRαβ expression. In support of this, the expression of the pre-T cell receptor α (pTα) was significantly higher in HD-overexpressing TCRγδ T-ALLs, compared with TCRγδ T-ALLs without HD deregulation ($P = 0.04$; Fig. 8 B). The opposite was observed for *SOX13*, a transcription factor involved in normal γδ T cell differentiation (Melichar et al., 2007), which is expressed at lower levels in the HD-overexpressing TCRγδ T-ALLs ($P = 0.003$; Fig. 8 C). Next, we analyzed the *TCRB* rearrangements in TCRγδ-expressing T-ALLs and demonstrated a high incidence of complete VDJ *TCRB* rearrangement in the HD-overexpressing group (45/71 vs. 3/24; $P < 0.0001$; Table 1). These data suggest that the molecular basis for the skewed TCRγδ expression in T-ALL could be related to homeodomain proteins (HOXA, TLX1/3) deregulation, preventing Eα activation and *TCRA* rearrangements, with the compensatory expression of a surface TCRγδ.

Discussion

We took advantage of our comprehensive epigenomic resource encompassing the main thymic populations of human T cell precursors to shed light on the epigenomic dynamics of a well-defined developmental program. First, we found that the epigenetic landscape positions thymocyte populations between HSC and mature T cells within the lymphoid lineage. T cell differentiation was accompanied by major remodeling of histone modifications involving a progressive closing of the chromatin landscape, clearly reflecting the Waddington model of cellular differentiation (Waddington, 1957). Second, our analysis revealed an unexpected persistence of DNA demethylation at distal regulatory regions. Interestingly, distal hypomethylated regions were associated with ∼64% of tissue-specific genes (Fig. 2 D, inset). A high proportion of distal regulatory regions were found to be constitutively hypomethylated irrespectively of their activation status throughout T lymphoid development, thus suggesting that DNA hypomethylation might work as a major epigenetic hallmark of enhancer preassembly and memory. Finally, we disentangled the regulatory mechanisms leading to stage-specific activation of the Eα *TCRA* enhancer, a paradigm of enhancer organization and function (Carey, 1998). Eα is found in a hypomethylated and open chromatin configuration at the very early stages of T cell differentiation, well before the activation of the *TCRA* locus, where the HOXA proteins function as developmentally regulated repressors of Eα.

During T cell thymopoiesis, Eα regulates the chromatin structure of *TCRA* gene segments by its ability to recruit a unique combination of transcription factors and induce modification of histone marks (McMurry and Krangel, 2000). We here demonstrated that Eα is already bound by its activating transcription factors, but stays nonfunctional at the earliest stages of the thymic maturation. Indeed, at these stages, Eα is found in an open chromatin state, fully demethylated and associated with histone marks characteristics of poised enhancer defined by enrichment of H3K4me1 and reduced levels of H3K27ac. Additionally, Eα is surrounded by the presence of H3K27me3, indicating a repressive mechanism acting on its

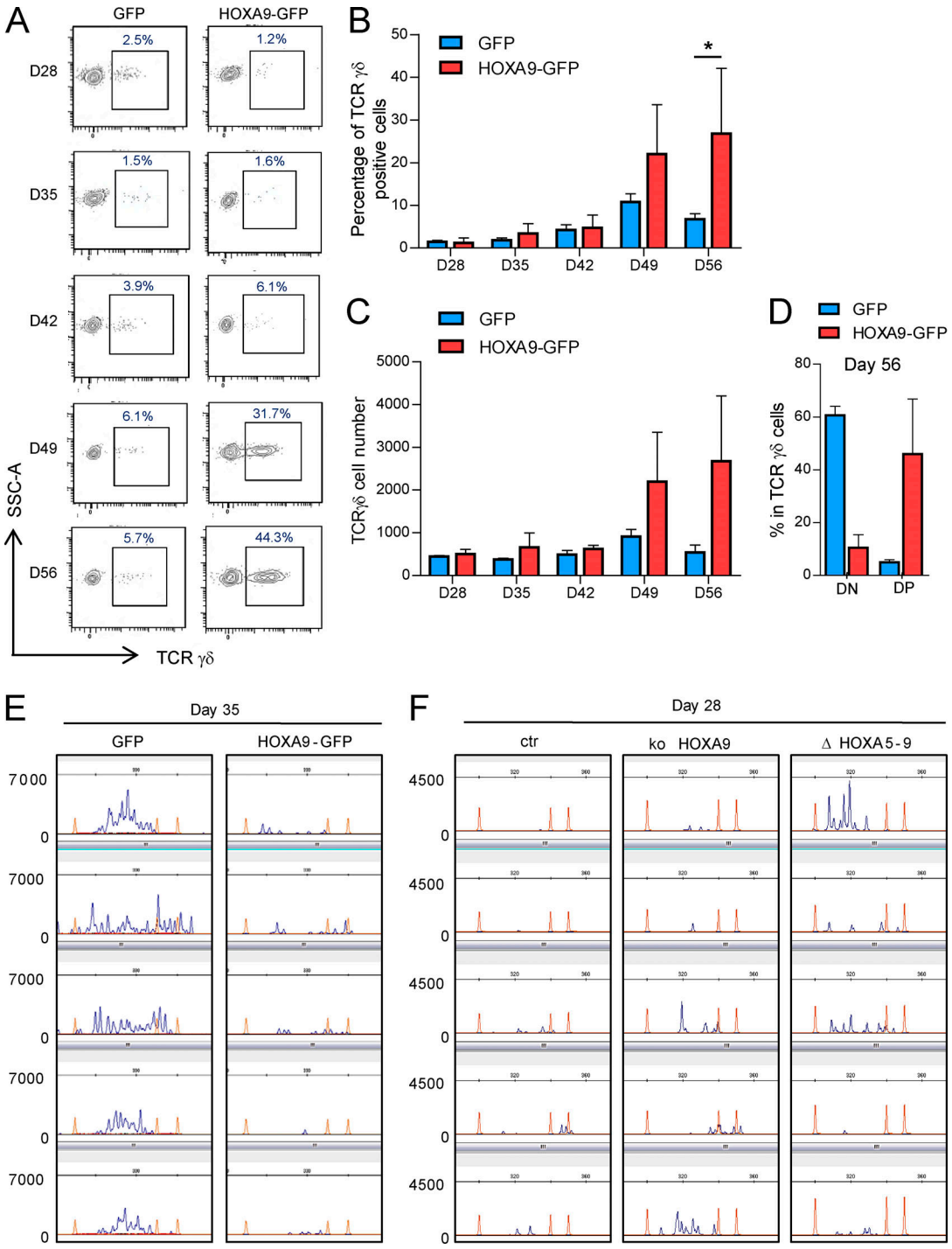


Figure 6. **HOXA9 overexpression biases T cell development in humans. (A)** Representative FACS plots representing TCRγδ expressing cells during T cell differentiation of human CD34⁺ UCB cells transduced with a HOXA9-GFP expressing vector or a control GFP vector. Presented cells are gated on the CD45⁺CD7⁺CD1a⁺ population. **(B)** Quantification of TCRγδ positive cells (as in A). Data are presented as means of three independent experiments with error bars as SEM. **(C)** Cell number of TCRγδ-positive cells (as in A and B). Data are presented as means of three independent experiments with error bars as SEM. **(D)** CD4 and CD8 expression of TCRγδ-positive cells at day 56 of human CD34⁺ UCB differentiation. Data are presented as means of two experiments with error bars as SEM. **(E)** Fluorescent PCR Genescan analysis of *TCRA* rearrangements detected in differentiated CD34⁺ UCB cells in HOXA9-overexpressing experiments. Representative of two different experiments. **(F)** Fluorescent PCR Genescan analysis of *TCRA* rearrangements detected in differentiated CD34⁺ UCB cells in CRISPR-Cas9 mediated HOXA9 knock-out and HOXA5-9 deletion experiments. Representative of two different experiments. P values were calculated by Student's test. *, $P < 0.05$. SSC-A, side scatter area. ko, knockout; ctr, control.

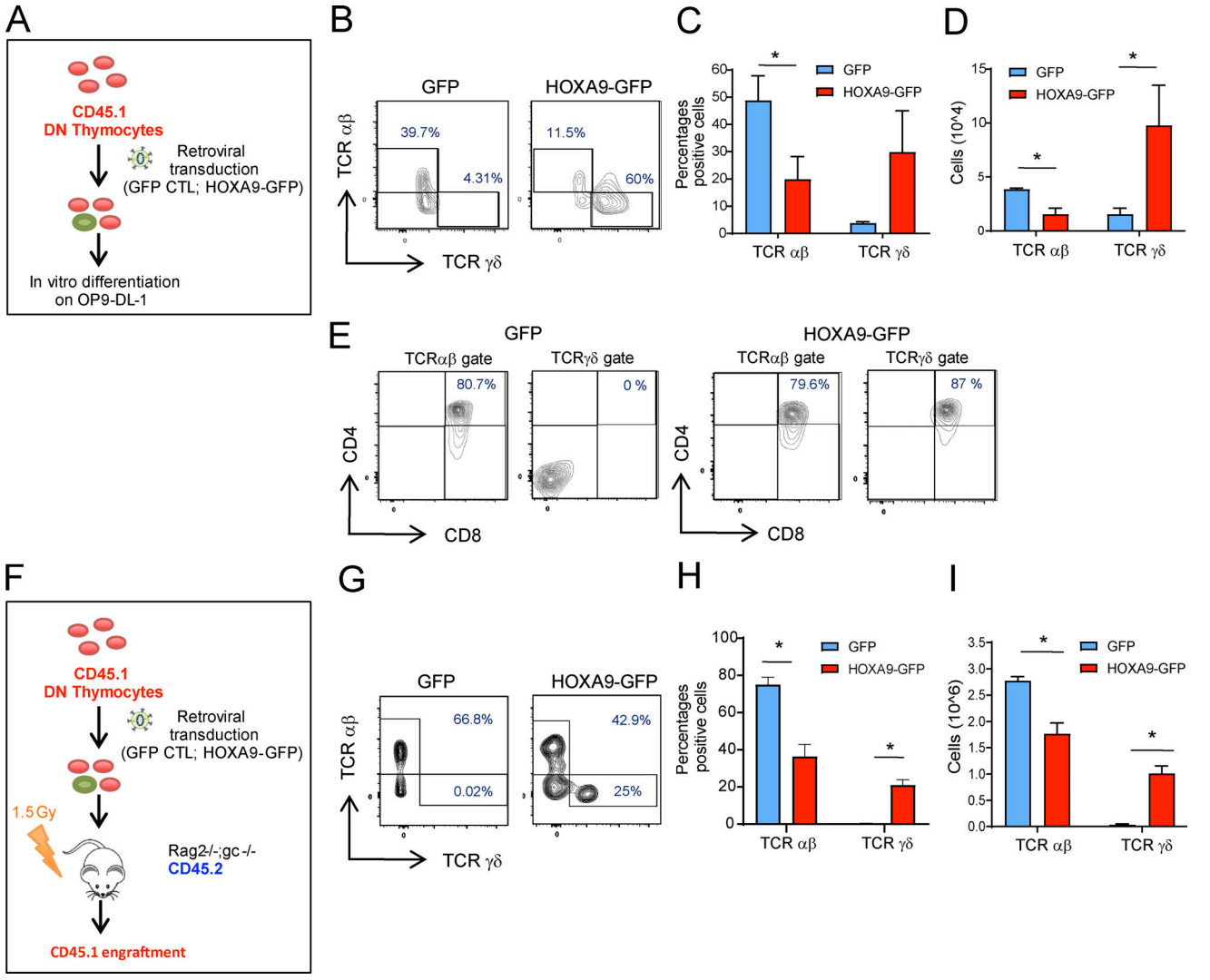


Figure 7. **HOXA9 overexpression biases T cell development in mice. (A)** Schematic representation of the in vitro experimental design in mice. **(B)** Representative FACS plots and quantification of TCRαβ- and TCRγδ-expressing cells in DN thymocytes transduced with a HOXA9-GFP overexpressing vector or a GFP control construct cultivated in vitro for 9 d on OP9-DL1 stromal cells. **(C)** Quantification of TCRαβ- and TCRγδ-expressing cells in mice in vitro assay (as in B). Quantification represents data of three independent experiments. **(D)** Cell number in TCRαβ- and TCRγδ-expressing cells in mice in an in vitro assay (A). Quantification represents data of three independent experiments with SEM. **(E)** Representative FACS plots of CD4 and CD8 expression in TCRαβ- and TCRγδ-expressing cells in an in vitro mice experiment (A). **(F)** Schematic representation of the in vivo experimental design in mice. **(G)** Representative FACS plots of TCRαβ- and TCRγδ-expressing cells in donor-derived thymocytes isolated from Rag2$^{-/-}$γc$^{-/-}$ mice at 3 wk after transplantation. **(H)** Quantification of TCRαβ- and TCRγδ-expressing cells in donor-derived thymocytes (as in G). Quantification represents data of three independent experiments with error bars as SEM. **(I)** Cell number of TCRαβ- and TCRγδ-expressing cells in an in vivo mice experiment (E). Quantification represents data of three independent experiments with SEM. P values were calculated by Student's test. *, P < 0.05.

activity, as suggested for developmentally regulated enhancers (Cruz-Molina et al., 2017; Rada-Iglesias et al., 2011; Zentner et al., 2011). Our results are also reminiscent of previous findings showing that cell type–restricted enhancers are premarked by DNA hypomethylation and binding of embryonic stem cell transcription factors, although they do not exhibit traditional enhancer epigenetic marks in embryonic stem cells (Kim et al., 2018; Xu et al., 2007).

HOXA proteins belong to a family of genes that shares a characteristic homeodomain protein fold, consisting of a 60–amino acid helix-turn-helix structure responsible for DNA binding and interactions with other proteins. HOXA functions are well established in the antero-posterior axis definition of body segment identity specification during embryogenesis (Carroll, 1995; Goodman, 2002; Lewis, 1978). Importantly, they also play a key role in controlling cell identity and differentiation of HSCs and progenitors (Lawrence et al., 1996; Magli et al., 1997). *HOX* genes are highly expressed in HSCs and progenitors, and their expression is silenced as cells become fully mature. Progressive down-modulation of *HOXA* transcripts was also reported during thymic cell maturation (Taghon et al., 2003). We now report that HOXA proteins repress the Eα activity in a homeodomain-dependent manner. All HOX homeodomains bind highly similar AT-rich DNA motifs (Berger et al.,

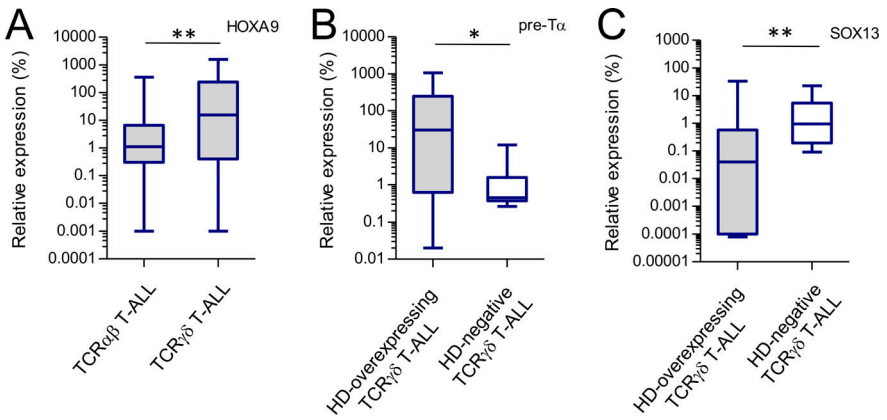


Figure 8. **HOXA9 overexpression leads to a TCRγδ bias in human T-ALL. (A)** RQ-PCR analysis of the HOXA9 gene expression in the TCRαβ (n = 45) and TCRγδ (n = 96) T-ALLs. Results are relative to the *ABL1* housekeeping gene. **, P = 0.003 (two-tailed Mann–Whitney test). **(B)** RQ-PCR analysis of the *preTα* gene expression in the HD-overexpressing (n = 30) and HD-negative (n = 8) TCRγδ T-ALL. Results are relative to the *ABL1* housekeeping gene expression. *, P = 0.02 (two-tailed Mann–Whitney test). **(C)** RQ-PCR analysis of the *SOX13* gene expression in the HD-overexpressing (n = 34) and HD-negative (n = 12) TCRγδ T-ALL. Results are relative to the *ABL1* housekeeping gene expression. **, P = 0.003 (two-tailed Mann–Whitney test).

2008; Mann et al., 2009; Noyes et al., 2008). The TAATNA recognition sequence seems to be critically important for DNA binding, as >98% of genome-wide HOXA9 binding sites in transformed myeloblasts contain a HOX motif (Huang et al., 2012). Additional specificity of HOXA proteins is achieved through a combination of motif affinity, interaction with co-factors, and context-specific chromatin accessibility (Choo et al., 2011; Slattery et al., 2011). The TAATNA HOX recognition sequence is, however, absent in the core Eα, suggesting that HOXA recruitment to Eα is mediated by interactions with components of the Eα enhanceosome. The suppressive function of HOXA proteins on Eα activity could be mediated by ETS1, as has been previously shown for TLX1/TLX3 homeoproteins (Dadi et al., 2012), although we do not exclude alternative protein–protein interactions and mediation of a suppressive effect by other transcription factors, such as RUNX1 (Hollenhorst et al., 2009).

Among HOX genes, *HOXA9* is the most preferentially expressed in CD34$^+$ HSCs and early hematopoietic progenitors and is subsequently down-regulated during differentiation (Taghon et al., 2003). Moreover, *HoxA9* is known to impact mouse thymocyte differentiation, and its knock-down induces the most dramatic hematopoietic phenotype of all *HOX* genes (Izon et al., 1998; Lawrence et al., 1997). Hoxa9$^{-/-}$ mice show an increase in the percentage of DP TCRαβ^{hi} cells and the total number of TCRαβ positive cells and accelerated progression of DP cells to mature CD4$^+$ and CD8$^+$ SP when compared with wild-type cells (Izon et al., 1998). This suggests that lifting the repression exerted by HOXA9 on the *TCRA* enhancer accelerates the maturation of thymocytes. Inversely, our HOXA9 overexpression experiments suggest a shift in lineage choice toward TCRγδ-expressing cells in both in vitro and in vivo experiments. This hints at an important role for physiological repression of Eα activity by HOXA proteins on thymocyte development. These data also pave the way for further investigation of the HOXA proteins role in the αβ versus γδ T cell lineage choice.

The blockage of Eα activity and subsequent lack of *TCRA* rearrangements during thymic differentiation can have severe implications. The cortical thymic maturation arrest observed in T-ALLs overexpressing TLX1 or TLX3 proteins is caused by the recruitment of TLX1/TLX3 to Eα by interacting with ETS1, leading to reduced enhancer activity and consequent inhibition of *TCRA* VJ recombination, maturation arrest of DP thymocytes, and leukemic transformation (Dadi et al., 2012). This maturation block can be overcome by TLX1/3 abrogation or by downstream TCRαβ expression, which leads to *TCRA* rearrangement and apoptosis. The deregulation of HOX genes has also been reported in acute leukemias (Alharbi et al., 2013), especially those with mixed-lineage leukemia/KMT2A translocations. Mixed-lineage leukemia fusion proteins constitutively up-regulate *HOXA9* expression (Ferrando et al., 2003). *HOX* genes are also overexpressed in T-ALLs with *PICALM-MLLT10* or *SET-NUP214* translocations or rearrangement of *HOXA* into a TCR locus (HOXA-TCR; Dik et al., 2005a; Soulier et al., 2005; Speleman et al., 2005; Van Vlierberghe et al., 2008). These T-ALLs are predominantly immature or express a TCRγδ, suggesting a role for *HOXA* gene overexpression and Eα inhibition in the development of these leukemias. Taken together, we propose that the endogenous HOXA5-9 transcription factors restrain the activity of a fully assembled Eα enhanceosome and prevent premature *TCRA* recombination in early thymocyte precursors, thereby providing the tight epigenetic control of Eα activation.

Overall, these epigenomic resources provide a detailed molecular framework to guide future studies on early T cell differentiation in human and their potential implications in leukemogenesis as well as thymic origins of immune-related disorders (Clave et al., 2018; Kernfeld et al., 2018).

Table 1. ***TCRB* rearrangement analysis in the TCRγδ T-ALLs according to their deregulation of homeodomain genes**

		TCRB rearrangements		
TCRγδ T-ALL patients (n = 95)	**No. (%)**	**VDJ**	**DJ**	**GL**
HD-overexpressing	71 (75)	45	23	3
HOXA	39 (41)	16	20	3
TLX1/3	32 (34)	29	3	0
HD-negative	24 (25)	3	10	11

Abbreviations: DJ, incomplete *TCRB* rearrangements; GL, germline (*TCRB* not rearranged); HD, homeodomain gene (i.e., HOXA, TLX1, TLX3).

Materials and methods

Cell lines

HeLa cells were maintained in DMEM, high glucose (Thermo Fisher Scientific, 41965062) supplemented with 10% FBS (Thermo Fisher Scientific, 10270106) and 1% penicillin/streptomycin.

Loucy (ACC-394) and Jurkat cell lines (ACC-282) were cultured in RPMI-1640 medium (Thermo Fisher Scientific, 21875091) supplemented with 20% and 10% FBS, respectively, 1% penicillin/streptomycin, and 1× sodium pyruvate.

Platinum-E (the retroviral packaging cell line) was maintained in DMEM supplemented with 10% FBS (Thermo Fisher Scientific, 10270106), 1% penicillin/streptomycin, 10 µg/ml blasticidin, and 1 µg/ml puromycin.

OP-DL1 stromal cell lines were maintained in MEM α and GlutaMAX supplement medium (Thermo Fisher Scientific, 32561094) supplemented with 20% FBS Hyclone SH30070.03HI (Thermo Fisher Scientific, 10772634) and 1% penicillin/streptomycin.

Primary cells preparation

Human CD34$^+$ UCB cells were isolated from cord blood using lymphocyte separation medium (Eurobio, CMSMSL01-01). Next, immature CD34$^+$ UCB cells were positively selected with direct CD34$^+$ progenitor cell isolation kit (Miltenyi Biotec, 130–046-702) using the LS separation columns (Miltenyi Biotec, 130–042-401). The CD34$^+$-positive cell population was further purified by sorting with FACS ARIA III to obtain 98–99% purity of CD34$^+$ UCB cells.

Human thymic subpopulations were obtained from children undergoing heart surgery. Informed consent was obtained from the parents. Thymi were dissociated until the single-cell suspension. Specific subpopulations were purified by sorting using FACS ARIA III (BD Biosciences). For the immature subpopulation, thymocytes were prepurified by depletion of CD3- and CD8-positive fractions by magnetic-activated cell sorting using CD3 MicroBeads (Miltenyi Biotec, 130–050-101) and CD8 MicroBeads (Miltenyi Biotec, 130–045-201) respectively. Antibodies used for cell labeling of the CD3$^-$/CD8$^-$ thymic fraction were as follows: CD1a FITC, clone NA1/34 (Dako, F7141), CD34 APC (BD Biosciences, 345804), CD8 PC7 (Beckman, 737661), CD3 Alexa 700 (BD Biosciences, 557943), CD4 V450, clone RPA-T4 (BD Biosciences, 560345), and CD45 V500, clone HI30 (BD Biosciences, 560777). Final purity after sorting was over 95%.

Mouse thymic cells from 6- to 8-wk-old C57Bl/6Rj mice were extracted and mechanistically disrupted on a 75-µm nylon cell strainer. Cells were washed twice with cold PBS supplemented with 2% of FBS. DN thymocytes were isolated by depletion of CD4- and CD8-positive fractions using magnetic beads (Miltenyi Biotec, 130–117-043 and 130–117-044, respectively). For maximum purity, the sorting of DN cells was performed using a FACS ARIA III.

Rodents

C57Bl/6 (CD45.2) mice were purchased from Janvier Labs. Rag2$^{-/-}$γc$^{-/-}$ (CD45.1) mice used in the in vivo experiments were purchased from The Jackson Laboratory. All experiments were performed according to procedures approved by the Committee of Paris Descartes University.

T-ALL patients

150 adult and pediatric TCRαβ- and TCRγδ-expressing T-ALLs were selected from the Necker-Enfants Malades Bio-bank collection (http://www.biobanques.eu/fr/nous-connaitre/membres/item/prb-necker-paris) based on their DNA/cDNA availability. Studies were conducted in accordance with the Declaration of Helsinki and approved by local and multicenter research ethical committees (Group for Research in Adult Acute Lymphoblastic Leukemia—GRAALL NCT00222027 and NCT00327678). Informed consent was obtained from all patients. Leukemic blasts were isolated from peripheral blood of T-ALL patients using Ficoll gradient separation.

Reference epigenomes of human thymopoiesis

Five sorted human thymic subpopulations (tCD34, EC, LC, SP4, and SP8) were used for ChIP-seq (H3K4me3, H4K4me1, H3K27ac, H3K27me3, H3K36me3, and H3K9me3) and WGBS following the BLUEPRINT protocol (http://dcc.blueprint-epigenome.eu/#/md/methods). Briefly, for ChIP-seq of histone marks, sorted thymic samples were incubated with 1% formaldehyde (Merck, F8775) in PBS for 10 min at room temperature. Next, 1/10 volumes of quenching solution of 1.25 M glycine (Merck, GE17-1323-01) were added to the medium and were shaken for 5 min at room temperature. Cells were collected by centrifugation and washed once with PBS-BSA 0.5%. For WGBS sequencing, DNA was extracted from sorted thymic subpopulations using the Nucleon BACC2 extraction kit (Merck, GERPN8502) according to the supplier's instructions.

Processed data were retrieved from the BLUEPRINT project. To complement the reference epigenomes, additional ChIP samples were sequenced in house in single-end 75 nt mode using the NextSeq 500/550 (Illumina) according to the manufacturer's instructions, and processed following the BLUEPRINT protocol. Aligned ChIP-Seq reads from multiple donors were merged for each thymic subpopulation (details in Table S2 and Table S6).

RNA-seq of human thymopoiesis

Six sorted human thymic subpopulations (tCD34, ISP, EC, LC, SP4, and SP8) were used for poly(A)-enriched RNA-Seq. Fragments were sequenced in stranded paired-end mode (2 × 50 bp) using the SOLiD HQ5500XL platform (Life Technologies; Table S2 and Table S6). SOLiD RNA-seq data were processed using LifeScope (Life Technologies) and TopHat using default parameters. Thy3 and Thy4 RNA-seq samples from Casero et al. (2015) were retrieved and aligned using STAR and Ensembl GRCh38 release 93 annotation track. Quantification in genes for both experiments was done with Subread FeaturesCounts. GeTMM count normalization (Table S3; Smid et al., 2018), and differential expression analysis were done using EdgeR (Table S4).

ATAC-seq of human thymopoiesis

Four sorted human thymic subpopulations (CD34$^+$1a$^-$, CD34$^+$1a$^+$, EC, and LC) from two donors were used for ATAC-seq. 5 × 10^4

thymocytes were washed with cold PBS and lysed in 50 µl of cold lysis buffer (10 mM Tris-HCl, pH 7.4; 10 mM NaCl; 3 mM $MgCl_2$; and 0.1% Igepal CA-630). Transposition reaction was performed in 50 µl of 1× Tagment DNA reaction buffer supplemented with 2.5 µl of TDE1 Tn5 Transposase from Nextera DNA Library Prep Kit (Illumina, FC-121-1030) at 37°C for 30 min. Directly after transposition, purification using MinElute PCR Purification kit (Qiagen, 28004) was performed, followed by sequencing. Samples were sequenced in house in single-end 75 nt mode using the NextSeq 500/550 (Illumina) according to the manufacturer's instructions. Reads were trimmed with sickle and aligned with bowtie2 using default settings. Aligned reads from $tCD34^+1a^-$ and $tCD34^+1a^+$ populations from the two donors were merged to generate the tCD34 ATAC sample (Table S2 and Table S6).

Chromatin segmentation and space generation

Merged ChIP-seq for each thymic subpopulations were segmented with ChromHMM (Ernst and Kellis, 2017) using the 11-state chromatin model from (Carrillo-de-Santa-Pau et al., 2017) to complement their published chromatin states from other hematopoietic cells. Regions with variable chromatin states across hematopoietic cell types were extracted using Chromdet and the five-state collapse model from Carrillo-de-Santa-Pau et al. (2017). MCA was done using FactoMineR R package, and cell-type delimitation ellipses were manually added afterward. The Sankey plot was generated using the riverplot R package. Genome browser views were produced with IGV, then compacted with Inkscape.

Functional enrichment analysis

Functional enrichment analyses were produced using a custom pipeline to automate multi-sample queries to the Genomic Regions Enrichment of Annotations Tool (GREAT) web service. In summary, features were converted to hg19 assembly using crossmap and queried with rGREAT R package against Gene Ontology (GO) Biological Process and Molecular Signatures Database (MSigDB) Pathways. Genomic association rules with genes were done using default GREAT "basal plus extension." Significant terms were filtered as those having binomial fold enrichment >2 and both binomial and hypergeometric tests Benjamini–Hochberg–adjusted P values <0.05. Additionally, MSigDB Pathway terms were further filtered for heatmap display to keep only the best five terms for each sample according to the binomial P value. For GO biological process, GOSemSim R package was used to compute Wang similarity distance between all terms. Terms with a similarity distance higher than 0.1 were grouped, unless specified otherwise. Only the best term according to binomial fold enrichment for each group and sample was kept for heatmap display. Color scale on heatmaps displays binomial Benjamini–Hochberg–adjusted P values.

Definition of distal regulatory elements and association to genes

Hypomethylated CpG regions were defined from WGBS following the BLUEPRINT protocol (http://dcc.blueprint-epigenome.eu/#/md/bs_seq_grch38). Distal regulatory regions were defined as hypomethylated regions in at least one thymic subpopulation falling between 2 and 2,000 kb from the closest gene, based on the subset of Ensembl release 93 annotation track containing only protein-coding as well as TCR and immunoglobulin loci. Regions overlapping repeats from repeatMasker, mitochondrial, and sexual chromosomes or exhibiting standard deviation >0.05 between donor samples from the same thymic subpopulation were filtered out. Constitutive hypomethylated distal regulatory elements were defined as those with CpG mean methylation signal <0.25 for all thymic subpopulation samples. Distal cis-regulatory regions from WGBS and ATAC peaks were associated with both their nearest upstream and downstream genes as long as they were within 2,000 kb range.

Clustering

Clustering was done by k-means using Lloyd algorithm and Pearson distance on mean CpG signal in hypomethylated regions (Fig. 2 A), on qualitative H3K27ac peak presence (Fig. 2 F), and on qualitative ATAC peak presence (Fig. 3).

Normalization and peak-calling

ChIP-seq coverage tracks were computed using DANPOS and quantile-normalized against SP8 samples, as an arbitrary reference. ATAC-seq coverage tracks were computed using deepTools bamCoverage with RPKM normalization. Peaks for both ChIP-seq and ATAC-seq were called using Macs2. Profiles and heatmaps were generated using deepTools plotProfile and plotHeatmap, respectively.

Tissue specificity

Consensus RNA-seq signals from 61 tissues were retrieved from the Human Protein Atlas version 19.2 (http://www.proteinatlas.org/download/proteinatlas.tsv.zip; Uhlén et al., 2015). Preferential expression measure (Huminiecki et al., 2003) was used to score differential expression of genes in thymus versus all other tissues with the exception of mature T cells, total peripheral blood mononuclear cells, and spleen. Genes with a preferential expression measure >0.1 were classified as thymus specific.

Motif enrichment

Conserved transcription factor binding sites were retrieved from the University of California, Santa Cruz genome browser. Binding motifs for the same transcription factor were merged. Clustered distal ATAC peaks were queried for enrichment against the merged set of motifs using the OLOGRAM tool (Ferré et al., 2019) from the Pygtftk package (Lopez et al., 2019).

Gene expression analysis

RNA was extracted using the RNeasy Mini Kit (Qiagen, 74104) and converted into cDNA using SuperScript III Reverse transcription (Thermo Fisher Scientific, 18080093). Real-time PCR reactions were performed using Taqman Universal PCR Master Mix (Thermo Fisher Scientific, 4304437) or SYBR Green PCR Master Mix (Thermo Fisher Scientific, 4309155). Real-time qPCR (RQ-PCR) was performed on an Applied Biosystems 7900HT Fast Real-Time PCR system. Primer sequences are in Table S5.

Multiplex *TCRA* RT-PCR analysis

TCRA cDNA was amplified from Cα (6-fluorescein amidite fluorophore–labeled) and various Vα in five multiplex RT-PCRs (Villarese et al., 2018). Briefly, 20 ng of cDNA were amplified per PCR tube with 1U HotStarTaq DNA Polymerase (Qiagen, 203203), 2 mM of $MgCl_2$, 20% Q solution, 10 mM deoxynucleoside triphosphate final concentration, and 10 pmol of each primer. The Taq polymerase was activated for 15 min at 95°C. Then the DNA was amplified with 37 cycles at 94°C for 30 s, at 63°C for 45 s, and at 72°C for 1 min 30 s. The final elongation was 72°C for 10 min. Primer sequences are in Table S5. The analysis of rearrangements by multiplex fluorescent PCR was performed by separation of single-stand (denaturated) PCR products in a capillary sequencing polymer and detected via automated laser scanning (3130 Genetic Analyzer, Applied Biosystems).

CAT reporter assay

HeLa cell lines were transfected using Lipofectamine 2000 Transfection Reagent (Thermo Fisher Scientific, 11668019) in the Opti-MEM I Reduced Serum Medium (Thermo Fisher Scientific, 51985026) according to the manufacturer's instructions. Cells at 70% confluency were cotransfected with the Eα-CAT reporter construct together with expression vectors for ETS1-HA-His, LEF1-HA, and RUNX1-T7. To test the CAT repression, cells were additionally transfected with expressing vectors for FL forms of HOXA5-9 or homeodomain deleted forms: HOXA5-ΔHD (aa 1–187), HOXA6-ΔHD (aa 1–147), HOXA7-ΔHD (aa 1–126), and HOXA9-ΔHD (aa 1–193). All HOXA proteins were cloned into the pEGFP-C1 vector (Clontech, 6084–1), where their N terminus were fused with GFP. After 24 h of incubation, expression of CAT activity was determined using the CAT ELISA kit (Merck, 11363727001) according to the manufacturer's recommendations.

TCRA enhancer repression in Jurkat cell line

Jurkat cells were transected with FL and homeodomain depleted forms (ΔHD) of HOXA5-9 proteins (same as in CAT-ELISA experience). Electroporation transfection was performed using the Neon Transfection System according to the manufacturer's instructions. At 48 h after transfection, the expression of the constant region of the *TCRA* gene (Cα transcript) was evaluated by quantitative RT-PCRs.

Quantification of H3K27ac level at Eα enhancer in T-ALL patients

H3K27ac ChIP-seq for mature T-ALL samples was quantile-normalized against an SP8 sample using DANPOS. Then H3K27ac coverage around Eα (chr14:22554500-22558500 in hg38 assembly) was quantified using DeepTools multiBigwigSummary. T-ALL were divided into three groups according to homeodomain gene expression and classified as homeodomain-negative, HOXA5-9–positive, or TLX1/TLX3-positive.

ChIPs

tCD34$^+$, EC, and LC thymocytes were cross-linked for 10 min with 1% formaldehyde at 20°C and sonicated using a Branson 450 Sonifier Cell Disruptor to obtain DNA fragments of ∼500 bp. ChIP was conducted with anti-RUNX1 (Abcam, ab23980), anti-ETS1 (C20X; Santa Cruz, sc-350) or anti-IgG antibody (Abcam, ab37415) using Dynabeads Protein G for Immunoprecipitation (Thermo Fisher Scientific, 10003D). Extracted ChIP DNA was purified by QIAquick PCR Purification Kit (Qiagen, 28104). ChIPed DNA was quantified in RQ-PCR assay using primers presented in Table S5.

ChIP analysis of HOXA binding was performed in two cell line models: Loucy and HeLa. Loucy cells were electroporated using Neon Transfection System (1 pulse of 1,700 V, 20 ms) with expressing vectors for HOXA5-9 (FL and HD-deleted forms) tagged with streptavidin-binding peptide (SBP)–Flag epitope or empty vector NTAP as control (vectors are listed in Table S6). After 48 h incubation, cells were cross-linked with 1% formaldehyde for 10 min. ChIP was performed as described above using anti-Flag antibody (Merck, F1804). Samples were analyzed by RQ-PCR using Ealpha-F/Ealpha-R for enhancer α binding and actin-F/actin-R for control. Primers are listed in Table S5. HeLa cells were transiently transfected using Lipofectamine 2000 Transfection Reagent (Thermo Fisher Scientific, 11668019) in the Opti-MEM I Reduced Serum Medium (Thermo Fisher Scientific, 51985026) with expressing vectors for ETS1-HA-His, LEF1-HA, RUNX1-T7, and Eα-CAT reporter constructs together with individual expressing vectors for HOXA5-9 (FL or HD-deleted forms) tagged with SBP-Flag or the empty vector SBP-Flag (NTAP) as control. After 24 h of incubation, cells were cross-linked with 1% formaldehyde for 10 min. ChIPs were performed with anti-Flag antibody (Merck, F1804). Samples were analyzed by RQ-PCR with the following primers: vect-Ealpha-F/vect-Ealpha-R for enhancer binding and vect-ctr-F/vect-ctr-R for control. Primers are listed in Table S5.

IP

2×10^8 of Loucy cells or total thymocytes were lysed for 30 min with 8 ml of 1× radioimmunoprecipitation assay buffer (Cell Signaling, 9806) complemented with Complete EDTA-free Protease Inhibitor Cocktail (Merck, 11873580001). The lysates were incubated overnight at 4°C with 10 µg of anti-ETS1 (C20X; Santa Cruz, sc-350) or anti-HOXA9 antibody (13C11, noncommercial production), followed by 2 h of incubation at 4°C with 20 µl of Protein G Agarose, Fast Flow (Merck, 16–266). After four washes with washing buffer (100 mM NaCl and 15 mM Tris-HCl, pH 7.8), bound proteins were eluted and detected by Western blot with anti-HOXA9 (13C11) or anti-ETS1 (C20X) antibodies.

Immunofluorescence analyses

Cells were fixed on poly-L-lysine (0.01%) precoated slides with PBS-BSA 1%/3.5% formaldehyde solution for 20 min at room temperature, and then permeabilized in PBS supplemented with 10% FBS and 1% Triton X-100 for 5 min (cell lines) or 10 min (thymocytes). Cells were incubated at 4°C overnight with anti-HOXA9 (13C11; dilution 1/20) or anti-ETS1 (C20X; dilution 1/250) antibodies, and then labeled with secondary antibodies goat anti-mouse IgG Alexa Fluor A555 (Thermo Fisher, A-21422) and goat anti-rabbit IgG Alexa Fluor 488 (Thermo Fisher, A-11008) in a dilution of 1/200. For PLA, cells were fixed, permeabilized, and incubated with primary antibodies as described above. The

following Duolink Probes were used for protein detection: Duolink In Situ PLA Probe Anti-Mouse MINUS (Merck, DUO92004) for HOXA9 and Duolink In Situ PLA Probe Anti-Rabbit PLUS (Merck, DUO92002) for ETS1. The interactions between proteins were detected with Duolink In Situ Detection Reagents Red (Merck, DUO92008) following the manufacturer's instructions. For EGFP fluorescence analysis, HeLa cells were transfected with expressing vectors for FL or HD-deleted forms of HOXA5-9 genes fused to eGFP in their N terminus (pEGFP-C1 vector, same as for CAT-ELISA experiments). After 24 h, cells were fixed as described above. Images were collected on a confocal microscope (Carl Zeiss LSM 700) with Zen 2011 software using 63× objectives at room temperature. Images were processed using ImageJ software. The colocalization coefficients were monitored with the JACoP plugin of ImageJ (Bolte and Cordelières, 2006). Quantification of PLA dots was performed using Icy Image Processing Software (de Chaumont et al., 2012).

Human T cell differentiation

$CD34^+$ UCB cells were stimulated for 16 h before transduction in the CellGenix GMP SCGM Stem Cell Growth Medium (CellGenix, 20802–500) supplemented with cytokines: 20 ng/ml human thrombopoietin, 100 ng/ml human stem cell factor, and 100 ng/ml hFLT3-L (Miltenyi Biotec). Stimulated $CD34^+$ UCB cells were transduced with HOXA9-GFP or control GFP VSV-G lentiviral vectors. 48 h after transduction, GFP-expressing cells were sorted using FACS ARIA III. $CD34^+$ UCB cells were also electroporated with Cas9-gRNA ribonucleoprotein complexes for HOXA9 knock-out and HOXA5-9 deletion. gRNAs were synthesized from Integrated DNA Technologies as Alt-R CRISPR-Cas9 crRNA. The functional gRNA was created after annealing with Alt-R tracrRNA (Integrated DNA Technologies). Editing efficiency evaluated with TIDE algorithm (https://tide.nki.nl/) was around 40%. Next, the cells were cultured on OP9-DL1 stroma cells in homemade MEMα (Thermo Fisher Scientific, 12000063) supplemented with 20% FBS HycloneSH30070.03HI (Thermo Fisher Scientific, 10772634) and cytokines: 5 ng/ml rFLT3-L, 10 ng/ml human stem cell factor, and 2 ng/ml hIL7 (Miltenyi Biotec; Six et al., 2011). Stromal OP9-DL1 cells were changed every week. At different time points of culture, cells were collected and analyzed by FACS for surface expression of TCRγδ and TCRαβ, and *TCRA* rearrangements by multiplex RT-PCR analysis.

Production of retrovirus particles and transduction of DN thymocytes

The retroviral vectors pMSCV-GFP and pMSCV-HOXA9-GFP were purchased from Addgene. For the production of retroviral vector supernatants, Platinum-E packaging cells (3×10^5 cells/well) were cultured in six-well plates and transiently transfected with pMSCV-GFP or pMSCV-HOXA9-GFP vectors (3.3 µg/well) using FuGENE HD Transfection Reagent (Thermo Fisher Scientific, PRE2311) and Opti-MEM Reduced Serum Media. Medium was changed 24 h after transfection. Vector supernatants were collected 48 h after transfection and were freshly used for thymocyte transduction. Sorted DN thymocytes were resuspended in vector supernatants with 10 µg/ml polybrene and spinoculated for 3 h at 3,500 rpm at room temperature. Next, the sorted cells were cultured in vitro on OP9-DL1 or injected into $Rag2^{-/-}\gamma c^{-/-}$ Ly5.1 mice.

Mouse T cell differentiation

DN thymocytes were FACS-sorted from C57/Bl6 mice Ly5.2. During in vitro studies, transduced DN cells were cultured for 9 d on an OP9-DL1 stroma layer in homemade MEMα (Thermo Fisher Scientific, 12000063) supplemented with 20% FBS Hyclone SH30070.03HI (Thermo Fisher Scientific, 10772634) and recombinant murine cytokines: 5 ng/ml rFLT3-L (R&D Systems, 427-FL), and 5 ng/ml rIL7 (407-ML). During in vivo studies, transduced DN cells were retro-orbitally injected into $Rag2^{-/-}\gamma c^{-/-}$ Ly5.1 mice that had been previously irradiated (2 Gy). 3 wk after transplantation, recipient mice were euthanized and the donor-derived cells analyzed.

Accession numbers

The high-throughput sequencing data produced were deposited in the Gene Expression Omnibus (Edgar et al., 2002) under accession no. GSE137718.

Online supplemental material

Fig. S1 shows a schematic representation of the major of human thymopoiesis, gating strategy for cell sorting, and chromatin states derived from six histone modifications. Fig. S2 shows methylation profiles around the eight clustered classes based on hypomethylation dynamics during T cell differentiation and H3K27-acetylation dynamics at constitutively hypomethylated regions. Fig. S2 also shows GREAT gene enrichment analysis. Fig. S3 shows the epigenomic profile of the RASSF6 gene. Fig. S4 shows transcription factor binding to *TCRA* enhancer and short RNA-seq profiles of the TCRA locus in the mouse thymocytes. Fig. S5 shows chromatin state profiles of the human *HOXA* locus, *HOXA5-9* gene expression evaluated by TLDA, and RQ-PCR. Fig. S5 also shows HOXA5-9 nuclear localization and Eα binding. Table S1 presents the design of the TLDA array. Table S2 shows the list of samples from high-throughput sequencing. Table S3 shows normalized counts for RNA-seq samples. Table S4 presents differential expression analysis. Table S5 presents a list of oligonucleotides and Table S6 lists key resources used in the study.

Acknowledgments

We thank the Cell Imaging Platform of Institut des Maladies Génétiques Imagine and the Transcriptomics and Genomics Marseille-Luminy sequencing platform. We thank the Pediatric Cardiac Surgery of Hôpital-Necker-Enfants-Malades for the human thymic samples. We thank the Cell Therapy Unit of Hôpital Saint-Louis for UCB samples. We thank Erika Brunet from the Institut des Maladies Génétiques Imagine for providing Cas9 protein. We thank Christoph Plass from the German Cancer Research Center in Heidelberg for reading and constructive criticism.

A. Cieslak was supported by grants from the Association pour la Recherche sur le Cancer, the Fondation pour la Recherche

Médicale, and the BLUEPRINT project. M. Tesio was supported by grants from Fondation de France and Association pour la Recherche sur le Cancer. G.P. Andrieu was supported by Fondation de France (post-doctoral fellow). Work in the V. Asnafi laboratory was supported by Association pour la Recherche contre le Cancer (Equipelabelisée) and Institut National du Cancer (PLBIO and PRTK). Work in the laboratory of S. Spicuglia was supported by recurrent funding from Institut National de la Santé et de la Recherche Médicale and Aix-Marseille University and by specific grants from the European Union's FP7 Program (agreement 282510-BLUEPRINT), the Foundation for Cancer Research, Association pour la Recherche sur le Cancer (ARC PJA 20151203149), A*MIDEX (ANR-11-IDEX-0001-02), Plan Cancer 2015 (C15076AS), PlBio-INCA, and Equipe Labelisée Ligue contre le Cancer.

Author contributions: A. Cieslak, G. Charbonnier, M. Tesio, S. Spicuglia, and V. Asnafi wrote the manuscript. A. Cieslak, M. Tesio, E.-L. Mathieu, C. Smith, G. Hypolite, A. Touzart, and G.P. Andrieu performed research and data analysis. G. Charbonnier, M. Belhocine, D. Puthier, and S. Spicuglia performed bioinformatics analysis. J.H.A. Martens, E. Janssen-Megens, M. Gut, I. Gut, and H.G. Stunnenberg are members of the BLUEPRINT consortium. N. Boissel and A. Petit provided the T-ALL patient samples. All authors validated the manuscript. V. Asnafi and S. Spicuglia oversaw conceptual development of the project.

Disclosures: The authors declare no competing interests exist.

Submitted: 17 December 2019
Revised: 3 April 2020
Accepted: 15 May 2020

References

Abarrategui, I., and M.S. Krangel. 2009. Germline transcription: a key regulator of accessibility and recombination. *Adv. Exp. Med. Biol.* 650: 93–102. https://doi.org/10.1007/978-1-4419-0296-2_8

Aifantis, I., E. Raetz, and S. Buonamici. 2008. Molecular pathogenesis of T-cell leukaemia and lymphoma. *Nat. Rev. Immunol.* 8:380–390. https://doi.org/10.1038/nri2304

Alharbi, R.A., R. Pettengell, H.S. Pandha, and R. Morgan. 2013. The role of HOX genes in normal hematopoiesis and acute leukemia. *Leukemia*. 27: 1000–1008. https://doi.org/10.1038/leu.2012.356

Asnafi, V., K. Beldjord, E. Boulanger, B. Comba, P. Le Tutour, M.H. Estienne, F. Davi, J. Landman-Parker, P. Quartier, A. Buzyn, et al. 2003. Analysis of TCR, pT alpha, and RAG-1 in T-acute lymphoblastic leukemias improves understanding of early human T-lymphoid lineage commitment. *Blood*. 101:2693–2703. https://doi.org/10.1182/blood-2002-08-2438

Asnafi, V., K. Beldjord, M. Libura, P. Villarese, C. Millien, P. Ballerini, E. Kuhlein, M. Lafage-Pochitaloff, E. Delabesse, O. Bernard, et al. 2004. Age-related phenotypic and oncogenic differences in T-cell acute lymphoblastic leukemias may reflect thymic atrophy. *Blood*. 104:4173–4180. https://doi.org/10.1182/blood-2003-11-3944

Bassing, C.H., R.E. Tillman, B.B. Woodman, D. Canty, R.J. Monroe, B.P. Sleckman, and F.W. Alt. 2003. T cell receptor (TCR) alpha/delta locus enhancer identity and position are critical for the assembly of TCR delta and alpha variable region genes. *Proc. Natl. Acad. Sci. USA*. 100: 2598–2603. https://doi.org/10.1073/pnas.0437943100

Berger, M.F., G. Badis, A.R. Gehrke, S. Talukder, A.A. Philippakis, L. Peña-Castillo, T.M. Alleyne, S. Mnaimneh, O.B. Botvinnik, E.T. Chan, et al. 2008. Variation in homeodomain DNA binding revealed by high-resolution analysis of sequence preferences. *Cell*. 133:1266–1276. https://doi.org/10.1016/j.cell.2008.05.024

Bolte, S., and F.P. Cordelières. 2006. A guided tour into subcellular colocalization analysis in light microscopy. *J. Microsc.* 224:213–232. https://doi.org/10.1111/j.1365-2818.2006.01706.x

Carey, M.. 1998. The enhanceosome and transcriptional synergy. *Cell*. 92:5–8. https://doi.org/10.1016/S0092-8674(00)80893-4

Carico, Z., and M.S. Krangel. 2015. Chromatin Dynamics and the Development of the TCRα and TCRδ Repertoires. *Adv. Immunol.* 128:307–361. https://doi.org/10.1016/bs.ai.2015.07.005

Carrillo-de-Santa-Pau, E., D. Juan, V. Pancaldi, F. Were, I. Martin-Subero, D. Rico, and A. Valencia; BLUEPRINT Consortium. 2017. Automatic identification of informative regions with epigenomic changes associated to hematopoiesis. *Nucleic Acids Res.* 45:9244–9259. https://doi.org/10.1093/nar/gkx618

Carroll, S.B.. 1995. Homeotic genes and the evolution of arthropods and chordates. *Nature*. 376:479–485. https://doi.org/10.1038/376479a0

Casero, D., S. Sandoval, C.S. Seet, J. Scholes, Y. Zhu, V.L. Ha, A. Luong, C. Parekh, and G.M. Crooks. 2015. Long non-coding RNA profiling of human lymphoid progenitor cells reveals transcriptional divergence of B cell and T cell lineages. *Nat. Immunol.* 16:1282–1291. https://doi.org/10.1038/ni.3299

Cauchy, P., M.A. Maqbool, J. Zacarias-Cabeza, L. Vanhille, F. Koch, R. Fenouil, M. Gut, I. Gut, M.A. Santana, A. Griffon, et al. 2016. Dynamic recruitment of Ets1 to both nucleosome-occupied and -depleted enhancer regions mediates a transcriptional program switch during early T-cell differentiation. *Nucleic Acids Res.* 44:3567–3585. https://doi.org/10.1093/nar/gkv1475

Choo, S.W., R. White, and S. Russell. 2011. Genome-wide analysis of the binding of the Hox protein Ultrabithorax and the Hox cofactor Homothorax in Drosophila. *PLoS One*. 6. e14778. https://doi.org/10.1371/journal.pone.0014778

Clave, E., I.L. Araujo, C. Alanio, E. Patin, J. Bergstedt, A. Urrutia, S. Lopez-Lastra, Y. Li, B. Charbit, C.R. MacPherson, et al; Milieu Intérieur Consortium. 2018. Human thymopoiesis is influenced by a common genetic variant within the *TCRA-TCRD* locus. *Sci. Transl. Med.* 10. eaao2966. https://doi.org/10.1126/scitranslmed.aao2966

Cruz-Molina, S., P. Respuela, C. Tebartz, P. Kolovos, M. Nikolic, R. Fueyo, W.F.J. van Ijcken, F. Grosveld, P. Frommolt, H. Bazzi, et al. 2017. PRC2 Facilitates the Regulatory Topology Required for Poised Enhancer Function during Pluripotent Stem Cell Differentiation. *Cell Stem Cell*. 20: 689–705.e9. https://doi.org/10.1016/j.stem.2017.02.004

Dadi, S., S. Le Noir, D. Payet-Bornet, L. Lhermitte, J. Zacarias-Cabeza, J. Bergeron, P. Villarèse, E. Vachez, W.A. Dik, C. Millien, et al. 2012. TLX homeodomain oncogenes mediate T cell maturation arrest in T-ALL via interaction with ETS1 and suppression of TCRα gene expression. *Cancer Cell*. 21:563–576. https://doi.org/10.1016/j.ccr.2012.02.013

de Chaumont, F., S. Dallongeville, N. Chenouard, N. Hervé, S. Pop, T. Provoost, V. Meas-Yedid, P. Pankajakshan, T. Lecomte, Y. Le Montagner, et al. 2012. Icy: an open bioimage informatics platform for extended reproducible research. *Nat. Methods*. 9:690–696. https://doi.org/10.1038/nmeth.2075

Dik, W.A., W. Brahim, C. Braun, V. Asnafi, N. Dastugue, O.A. Bernard, J.J. van Dongen, A.W. Langerak, E.A. Macintyre, and E. Delabesse. 2005a. CALM-AF10+ T-ALL expression profiles are characterized by overexpression of HOXA and BMI1 oncogenes. *Leukemia*. 19:1948–1957. https://doi.org/10.1038/sj.leu.2403891

Dik, W.A., K. Pike-Overzet, F. Weerkamp, D. de Ridder, E.F. de Haas, M.R. Baert, P. van der Spek, E.E. Koster, M.J. Reinders, J.J. van Dongen, et al. 2005b. New insights on human T cell development by quantitative T cell receptor gene rearrangement studies and gene expression profiling. *J. Exp. Med.* 201:1715–1723. https://doi.org/10.1084/jem.20042524

Edgar, R., M. Domrachev, and A.E. Lash. 2002. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* 30:207–210. https://doi.org/10.1093/nar/30.1.207

Ernst, J., and M. Kellis. 2017. Chromatin-state discovery and genome annotation with ChromHMM. *Nat. Protoc.* 12:2478–2492. https://doi.org/10.1038/nprot.2017.124

Ferrando, A.A., D.S. Neuberg, J. Staunton, M.L. Loh, C. Huard, S.C. Raimondi, F.G. Behm, C.H. Pui, J.R. Downing, D.G. Gilliland, et al. 2002. Gene expression signatures define novel oncogenic pathways in T cell acute lymphoblastic leukemia. *Cancer Cell*. 1:75–87. https://doi.org/10.1016/S1535-6108(02)00018-1

Ferrando, A.A., S.A. Armstrong, D.S. Neuberg, S.E. Sallan, L.B. Silverman, S.J. Korsmeyer, and A.T. Look. 2003. Gene expression signatures in MLL-rearranged T-lineage and B-precursor acute leukemias: dominance of

HOX dysregulation. *Blood*. 102:262–268. https://doi.org/10.1182/blood-2002-10-3221

Ferré, Q., G. Charbonnier, N. Sadouni, F. Lopez, Y. Kermezli, S. Spicuglia, C. Capponi, B. Ghattas, and D. Puthier. 2019. OLOGRAM: Determining significance of total overlap length between genomic regions sets. *Bioinformatics*. btz810. https://doi.org/10.1093/bioinformatics/btz810

Giese, K., J. Cox, and R. Grosschedl. 1992. The HMG domain of lymphoid enhancer factor 1 bends DNA and facilitates assembly of functional nucleoprotein structures. *Cell*. 69:185–195. https://doi.org/10.1016/0092-8674(92)90129-Z

Giese, K., C. Kingsley, J.R. Kirshner, and R. Grosschedl. 1995. Assembly and function of a TCR alpha enhancer complex is dependent on LEF-1-induced DNA bending and multiple protein-protein interactions. *Genes Dev*. 9:995–1008. https://doi.org/10.1101/gad.9.8.995

Goodman, F.R.. 2002. Limb malformations and the human HOX genes. *Am. J. Med. Genet*. 112:256–265. https://doi.org/10.1002/ajmg.10776

Hawwari, A., and M.S. Krangel. 2005. Regulation of TCR delta and alpha repertoires by local and long-distance control of variable gene segment chromatin structure. *J. Exp. Med*. 202:467–472. https://doi.org/10.1084/jem.20050680

Hernández-Munain, C., B.P. Sleckman, and M.S. Krangel. 1999. A developmental switch from TCR delta enhancer to TCR alpha enhancer function during thymocyte maturation. *Immunity*. 10:723–733. https://doi.org/10.1016/S1074-7613(00)80071-0

Ho, I.C., L.H. Yang, G. Morle, and J.M. Leiden. 1989. A T-cell-specific transcriptional enhancer element 3′ of C alpha in the human T-cell receptor alpha locus. *Proc. Natl. Acad. Sci. USA*. 86:6714–6718. https://doi.org/10.1073/pnas.86.17.6714

Ho, I.C., N.K. Bhat, L.R. Gottschalk, T. Lindsten, C.B. Thompson, T.S. Papas, and J.M. Leiden. 1990. Sequence-specific binding of human Ets-1 to the T cell receptor alpha gene enhancer. *Science*. 250:814–818. https://doi.org/10.1126/science.2237431

Hollenhorst, P.C., K.J. Chandler, R.L. Poulsen, W.E. Johnson, N.A. Speck, and B.J. Graves. 2009. DNA specificity determinants associate with distinct transcription factor functions. *PLoS Genet*. 5. e1000778. https://doi.org/10.1371/journal.pgen.1000778

Hu, G., K. Cui, D. Fang, S. Hirose, X. Wang, D. Wangsa, W. Jin, T. Ried, P. Liu, J. Zhu, et al. 2018. Transformation of Accessible Chromatin and 3D Nucleome Underlies Lineage Commitment of Early T Cells. *Immunity*. 48:227–242.e8. https://doi.org/10.1016/j.immuni.2018.01.013

Huang, C.Y., R. Golub, G.E. Wu, and O. Kanagawa. 2002. Superantigen-induced TCR alpha locus secondary rearrangement: role in tolerance induction. *J. Immunol*. 168:3259–3265. https://doi.org/10.4049/jimmunol.168.7.3259

Huang, Y., K. Sitwala, J. Bronstein, D. Sanders, M. Dandekar, C. Collins, G. Robertson, J. MacDonald, T. Cezard, M. Bilenky, et al. 2012. Identification and characterization of Hoxa9 binding sites in hematopoietic cells. *Blood*. 119:388–398. https://doi.org/10.1182/blood-2011-03-341081

Huminiecki, L., A.T. Lloyd, and K.H. Wolfe. 2003. Congruence of tissue expression profiles from Gene Expression Atlas, SAGEmap and TissueInfo databases. *BMC Genomics*. 4:31. https://doi.org/10.1186/1471-2164-4-31

Izon, D.J., S. Rozenfeld, S.T. Fong, L. Kömüves, C. Largman, and H.J. Lawrence. 1998. Loss of function of the homeobox gene Hoxa-9 perturbs early T-cell development and induces apoptosis in primitive thymocytes. *Blood*. 92:383–393. https://doi.org/10.1182/blood.V92.2.383

Kernfeld, E.M., R.M.J. Genga, K. Neherin, M.E. Magaletta, P. Xu, and R. Maehr. 2018. A Single-Cell Transcriptomic Atlas of Thymus Organogenesis Resolves Cell Types and Developmental Maturation. *Immunity*. 48:1258–1270.e6. https://doi.org/10.1016/j.immuni.2018.04.015

Kim, H.S., Y. Tan, W. Ma, D. Merkurjev, E. Destici, Q. Ma, T. Suter, K. Ohgi, M. Friedman, D. Skowronska-Krawczyk, et al. 2018. Pluripotency factors functionally premark cell-type-restricted enhancers in ES cells. *Nature*. 556:510–514. https://doi.org/10.1038/s41586-018-0048-8

Koch, F., R. Fenouil, M. Gut, P. Cauchy, T.K. Albert, J. Zacarias-Cabeza, S. Spicuglia, A.L. de la Chapelle, M. Heidemann, C. Hintermair, et al. 2011. Transcription initiation platforms and GTF recruitment at tissue-specific enhancers and promoters. *Nat. Struct. Mol. Biol*. 18:956–963. https://doi.org/10.1038/nsmb.2085

Krangel, M.S., J. Carabana, I. Abbarategui, R. Schlimgen, and A. Hawwari. 2004. Enforcing order within a complex locus: current perspectives on the control of V(D)J recombination at the murine T-cell receptor alpha/delta locus. *Immunol. Rev*. 200:224–232. https://doi.org/10.1111/j.0105-2896.2004.00155.x

Lawrence, H.J., G. Sauvageau, R.K. Humphries, and C. Largman. 1996. The role of HOX homeobox genes in normal and leukemic hematopoiesis. *Stem Cells*. 14:281–291. https://doi.org/10.1002/stem.140281

Lawrence, H.J., C.D. Helgason, G. Sauvageau, S. Fong, D.J. Izon, R.K. Humphries, and C. Largman. 1997. Mice bearing a targeted interruption of the homeobox gene HOXA9 have defects in myeloid, erythroid, and lymphoid hematopoiesis. *Blood*. 89:1922–1930. https://doi.org/10.1182/blood.V89.6.1922

Lepoivre, C., M. Belhocine, A. Bergon, A. Griffon, M. Yammine, L. Vanhille, J. Zacarias-Cabeza, M.A. Garibal, F. Koch, M.A. Maqbool, et al. 2013. Divergent transcription is associated with promoters of transcriptional regulators. *BMC Genomics*. 14:914. https://doi.org/10.1186/1471-2164-14-914

Lewis, E.B.. 1978. A gene complex controlling segmentation in Drosophila. *Nature*. 276:565–570. https://doi.org/10.1038/276565a0

Lopez, F., G. Charbonnier, Y. Kermezli, M. Belhocine, Q. Ferre, N. Zweig, M. Aribi, A. Gonzalez, S. Spicuglia, and D. Puthier. 2019. Explore, edit and leverage genomic annotations using Python GTF toolkit. *Bioinformatics*. 35:3487–3488. https://doi.org/10.1093/bioinformatics/btz116

Magli, M.C., C. Largman, and H.J. Lawrence. 1997. Effects of HOX homeobox genes in blood cell differentiation. *J. Cell. Physiol*. 173:168–177. https://doi.org/10.1002/(SICI)1097-4652(199711)173:2<168::AID-JCP16>3.0.CO;2-C

Mann, R.S., K.M. Lelli, and R. Joshi. 2009. Hox specificity unique roles for cofactors and collaborators. *Curr. Top. Dev. Biol*. 88:63–101. https://doi.org/10.1016/S0070-2153(09)88003-4

McMurry, M.T., and M.S. Krangel. 2000. A role for histone acetylation in the developmental regulation of VDJ recombination. *Science*. 287:495–498. https://doi.org/10.1126/science.287.5452.495

Melichar, H.J., K. Narayan, S.D. Der, Y. Hiraoka, N. Gardiol, G. Jeannet, W. Held, C.A. Chambers, and J. Kang. 2007. Regulation of gammadelta versus alphabeta T lymphocyte differentiation by the transcription factor SOX13. *Science*. 315:230–233. https://doi.org/10.1126/science.1135344

Noyes, M.B., R.G. Christensen, A. Wakabayashi, G.D. Stormo, M.H. Brodsky, and S.A. Wolfe. 2008. Analysis of homeodomain specificities allows the family-wide prediction of preferred recognition sites. *Cell*. 133:1277–1289. https://doi.org/10.1016/j.cell.2008.05.023

Oravecz, A., A. Apostolov, K. Polak, B. Jost, S. Le Gras, S. Chan, and P. Kastner. 2015. Ikaros mediates gene silencing in T cells through Polycomb repressive complex 2. *Nat. Commun*. 6:8823. https://doi.org/10.1038/ncomms9823

Pekowska, A., T. Benoukraf, J. Zacarias-Cabeza, M. Belhocine, F. Koch, H. Holota, J. Imbert, J.C. Andrau, P. Ferrier, and S. Spicuglia. 2011. H3K4 tri-methylation provides an epigenetic signature of active enhancers. *EMBO J*. 30:4198–4210. https://doi.org/10.1038/emboj.2011.295

Rada-Iglesias, A., R. Bajpai, T. Swigut, S.A. Brugmann, R.A. Flynn, and J. Wysocka. 2011. A unique chromatin signature uncovers early developmental enhancers in humans. *Nature*. 470:279–283. https://doi.org/10.1038/nature09692

Roberts, J.L., P. Lauzurica, and M.S. Krangel. 1997. Developmental regulation of VDJ recombination by the core fragment of the T cell receptor alpha enhancer. *J. Exp. Med*. 185:131–140. https://doi.org/10.1084/jem.185.1.131

Rothenberg, E.V.. 2019. Programming for T-lymphocyte fates: modularity and mechanisms. *Genes Dev*. 33:1117–1135. https://doi.org/10.1101/gad.327163.119

Six, E.M., F. Benjelloun, A. Garrigue, D. Bonhomme, E. Morillon, J. Rouiller, L. Cacavelli, J. Blondeau, K. Beldjord, S. Hacein-Bey-Abina, et al. 2011. Cytokines and culture medium have a major impact on human in vitro T-cell differentiation. *Blood Cells Mol. Dis*. 47:72–78. https://doi.org/10.1016/j.bcmd.2011.04.001

Slattery, M., T. Riley, P. Liu, N. Abe, P. Gomez-Alcala, I. Dror, T. Zhou, R. Rohs, B. Honig, H.J. Bussemaker, et al. 2011. Cofactor binding evokes latent differences in DNA binding specificity between Hox proteins. *Cell*. 147:1270–1282. https://doi.org/10.1016/j.cell.2011.10.053

Sleckman, B.P., C.G. Bardon, R. Ferrini, L. Davidson, and F.W. Alt. 1997. Function of the TCR alpha enhancer in alphabeta and gammadelta T cells. *Immunity*. 7:505–515. https://doi.org/10.1016/S1074-7613(00)80372-6

Smid, M., R.R.J. Coebergh van den Braak, H.J.G. van de Werken, J. van Riet, A. van Galen, V. de Weerd, M. van der Vlugt-Daane, S.I. Bril, Z.S. Lalmahomed, W.P. Kloosterman, et al; MATCH study group. 2018. Gene length corrected trimmed mean of M-values (GeTMM) processing of RNA-seq data performs similarly in intersample analyses while improving intrasample comparisons. *BMC Bioinformatics*. 19:236. https://doi.org/10.1186/s12859-018-2246-7

Soulier, J., E. Clappier, J.M. Cayuela, A. Regnault, M. García-Peydró, H. Dombret, A. Baruchel, M.L. Toribio, and F. Sigaux. 2005. HOXA genes are included in genetic and biologic networks defining human acute

T-cell leukemia (T-ALL). *Blood*. 106:274–286. https://doi.org/10.1182/blood-2004-10-3900

Speleman, F., B. Cauwelier, N. Dastugue, J. Cools, B. Verhasselt, B. Poppe, N. Van Roy, J. Vandesompele, C. Graux, A. Uyttebroeck, et al. 2005. A new recurrent inversion, inv(7)(p15q34), leads to transcriptional activation of HOXA10 and HOXA11 in a subset of T-cell acute lymphoblastic leukemias. *Leukemia*. 19:358–366. https://doi.org/10.1038/sj.leu.2403657

Spicuglia, S., D. Payet, R.K. Tripathi, P. Rameil, C. Verthuy, J. Imbert, P. Ferrier, and W.M. Hempel. 2000. TCRalpha enhancer activation occurs via a conformational change of a pre-assembled nucleo-protein complex. *EMBO J*. 19:2034–2045. https://doi.org/10.1093/emboj/19.9.2034

Spits, H.. 2002. Development of alphabeta T cells in the human thymus. *Nat. Rev. Immunol*. 2:760–772. https://doi.org/10.1038/nri913

Stadler, M.B., R. Murr, L. Burger, R. Ivanek, F. Lienert, A. Schöler, E. van Nimwegen, C. Wirbelauer, E.J. Oakeley, D. Gaidatzis, et al. 2011. DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature*. 480:490–495. https://doi.org/10.1038/nature10716

Stunnenberg, H.G., and M. Hirst; International Human Epigenome Consortium. 2016. The International Human Epigenome Consortium: A Blueprint for Scientific Collaboration and Discovery. *Cell*. 167:1897. https://doi.org/10.1016/j.cell.2016.12.002

Taghon, T., K. Thys, M. De Smedt, F. Weerkamp, F.J. Staal, J. Plum, and G. Leclercq. 2003. Homeobox gene expression profile in human hematopoietic multipotent stem cells and T-cell progenitors: implications for human T-cell development. *Leukemia*. 17:1157–1163. https://doi.org/10.1038/sj.leu.2402947

Trinquand, A., N.R. Dos Santos, C. Tran Quang, F. Rocchetti, B. Zaniboni, M. Belhocine, C. Da Costa de Jesus, L. Lhermitte, M. Tesio, M. Dussiot, et al. 2016. Triggering the TCR Developmental Checkpoint Activates a Therapeutically Targetable Tumor Suppressive Pathway in T-cell Leukemia. *Cancer Discov*. 6:972–985. https://doi.org/10.1158/2159-8290.CD-15-0675

Uhlén, M., L. Fagerberg, B.M. Hallström, C. Lindskog, P. Oksvold, A. Mardinoglu, Å. Sivertsson, C. Kampf, E. Sjöstedt, A. Asplund, et al. 2015. Proteomics. Tissue-based map of the human proteome. *Science*. 347. 1260419. https://doi.org/10.1126/science.1260419

Van Vlierberghe, P., M. van Grotel, J. Tchinda, C. Lee, H.B. Beverloo, P.J. van der Spek, A. Stubbs, J. Cools, K. Nagata, M. Fornerod, et al. 2008. The recurrent SET-NUP214 fusion as a new HOXA activation mechanism in pediatric T-cell acute lymphoblastic leukemia. *Blood*. 111:4668–4680. https://doi.org/10.1182/blood-2007-09-111872

Vanhille, L., A. Griffon, M.A. Maqbool, J. Zacarias-Cabeza, L.T. Dao, N. Fernandez, B. Ballester, J.C. Andrau, and S. Spicuglia. 2015. High-throughput and quantitative assessment of enhancer activity in mammals by CapStarr-seq. *Nat. Commun*. 6:6905. https://doi.org/10.1038/ncomms7905

Villarese, P., C. Lours, A. Trinquand, S. Le Noir, M. Belhocine, L. Lhermitte, A. Cieslak, M. Tesio, A. Petit, M. LeLorch, et al. 2018. TCRα rearrangements identify a subgroup of NKL-deregulated adult T-ALLs associated with favorable outcome. *Leukemia*. 32:61–71. https://doi.org/10.1038/leu.2017.176

Villey, I., D. Caillol, F. Selz, P. Ferrier, and J.P. de Villartay. 1996. Defect in rearrangement of the most 5′ TCR-J alpha following targeted deletion of T early alpha (TEA): implications for TCR alpha locus accessibility. *Immunity*. 5:331–342. https://doi.org/10.1016/S1074-7613(00)80259-9

Waddington, C.H.. 1957. The Strategy of the Genes; a Discussion of Some Aspects of Theoretical Biology. Allen & Unwin, London.

Wei, G., B.J. Abraham, R. Yagi, R. Jothi, K. Cui, S. Sharma, L. Narlikar, D.L. Northrup, Q. Tang, W.E. Paul, et al. 2011. Genome-wide analyses of transcription factor GATA3-mediated gene regulation in distinct T cell types. *Immunity*. 35:299–311. https://doi.org/10.1016/j.immuni.2011.08.007

Winoto, A., and D. Baltimore. 1989. A novel, inducible and T cell-specific enhancer located at the 3′ end of the T cell receptor alpha locus. *EMBO J*. 8:729–733. https://doi.org/10.1002/j.1460-2075.1989.tb03432.x

Xu, J., S.D. Pope, A.R. Jazirehi, J.L. Attema, P. Papathanasiou, J.A. Watts, K.S. Zaret, I.L. Weissman, and S.T. Smale. 2007. Pioneer factor interactions and unmethylated CpG dinucleotides mark silent tissue-specific enhancers in embryonic stem cells. *Proc. Natl. Acad. Sci. USA*. 104:12377–12382. https://doi.org/10.1073/pnas.0704579104

Zacarías-Cabeza, J., M. Belhocine, L. Vanhille, P. Cauchy, F. Koch, A. Pekowska, R. Fenouil, A. Bergon, M. Gut, I. Gut, et al. 2015. Transcription-dependent generation of a specialized chromatin structure at the TCRβ locus. *J. Immunol*. 194:3432–3443. https://doi.org/10.4049/jimmunol.1400789

Zentner, G.E., P.J. Tesar, and P.C. Scacheri. 2011. Epigenetic signatures distinguish multiple classes of enhancers with distinct cellular functions. *Genome Res*. 21:1273–1283. https://doi.org/10.1101/gr.122382.111

Zhang, J., L. Ding, L. Holmfeldt, G. Wu, S.L. Heatley, D. Payne-Turner, J. Easton, X. Chen, J. Wang, M. Rusch, et al. 2012. The genetic basis of early T-cell precursor acute lymphoblastic leukaemia. *Nature*. 481:157–163. https://doi.org/10.1038/nature10725

JEM

Supplemental material

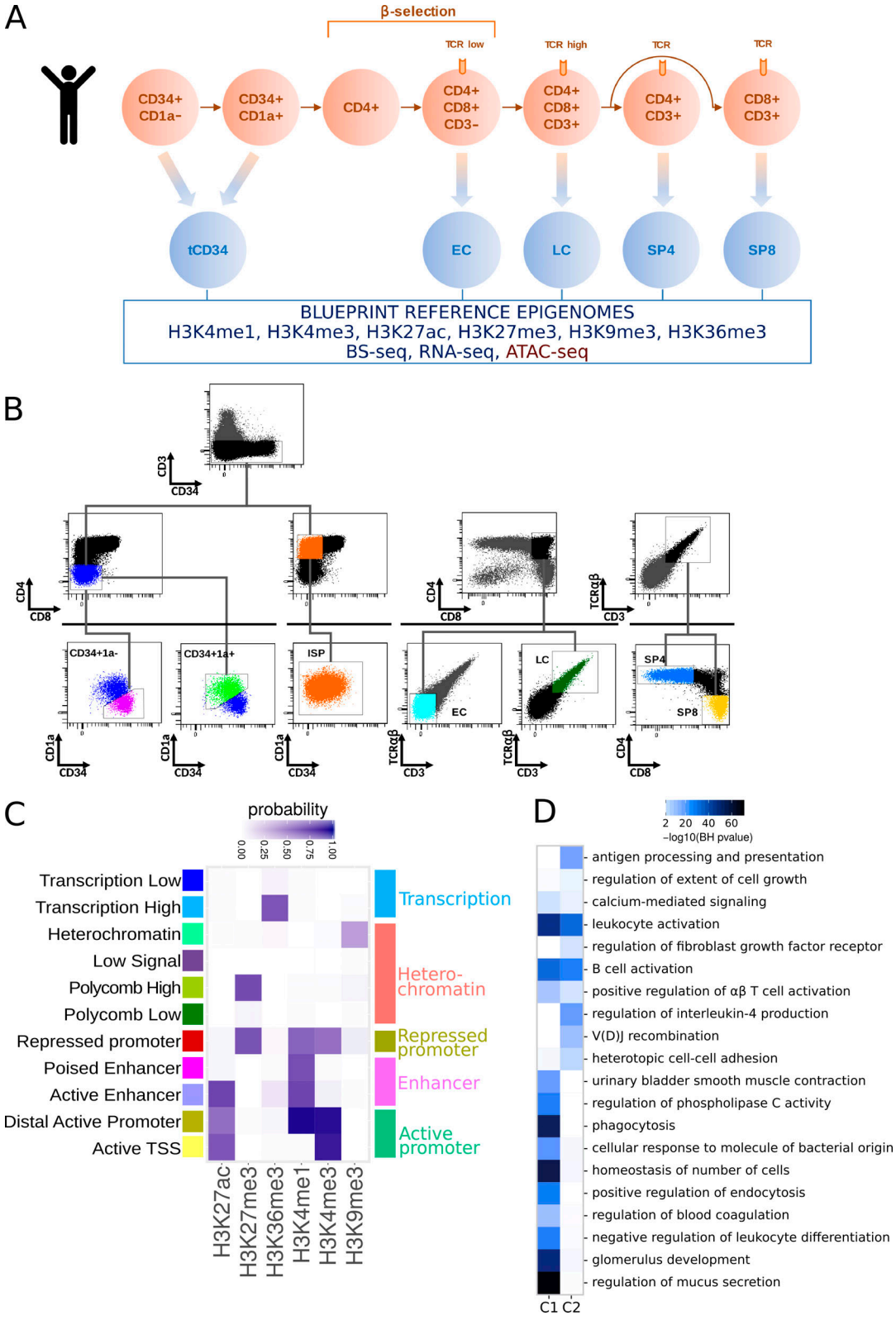


Figure S1. **Major human T cell differentiation stages. (A)** Schematic representation of the major stages of human thymopoiesis (orange). The thymic subpopulations used to prepare the BLUEPRINT reference epigenomes are presented in a blue color. **(B)** Plots showing the gating strategy used to sort the human thymic subpopulations. Purity after sorting was between 95–99%. tCD34, immature DN CD34$^+$ (CD34$^+$/CD3$^-$/CD4$^-$/CD8$^-$); EC (TCRαβ$^-$/CD3$^-$/CD4$^+$/CD8$^+$); LC (TCRαβ$^+$/CD3^{+low}/CD4$^+$/CD8$^+$); SP4, SP CD4$^+$ (TCRαβ$^+$/CD3$^+$/CD4$^+$/CD8$^-$); SP8, SP CD8$^+$ (TCRαβ$^+$/CD3$^+$/CD4$^-$/CD8$^+$). **(C)** State emissions for the used chromatin segmentation model and their biological description. **(D)** GREAT gene enrichment analysis for the genomic regions highly correlating with the first (C1, 21,523 regions with correlation > 0.9) and second (C2, 1,787 regions with correlation > 0.8) dimensions of the MCA from Fig. 1 D. Only the top 10 GO biological process terms passing a 0.5 Wang similarity threshold are shown.

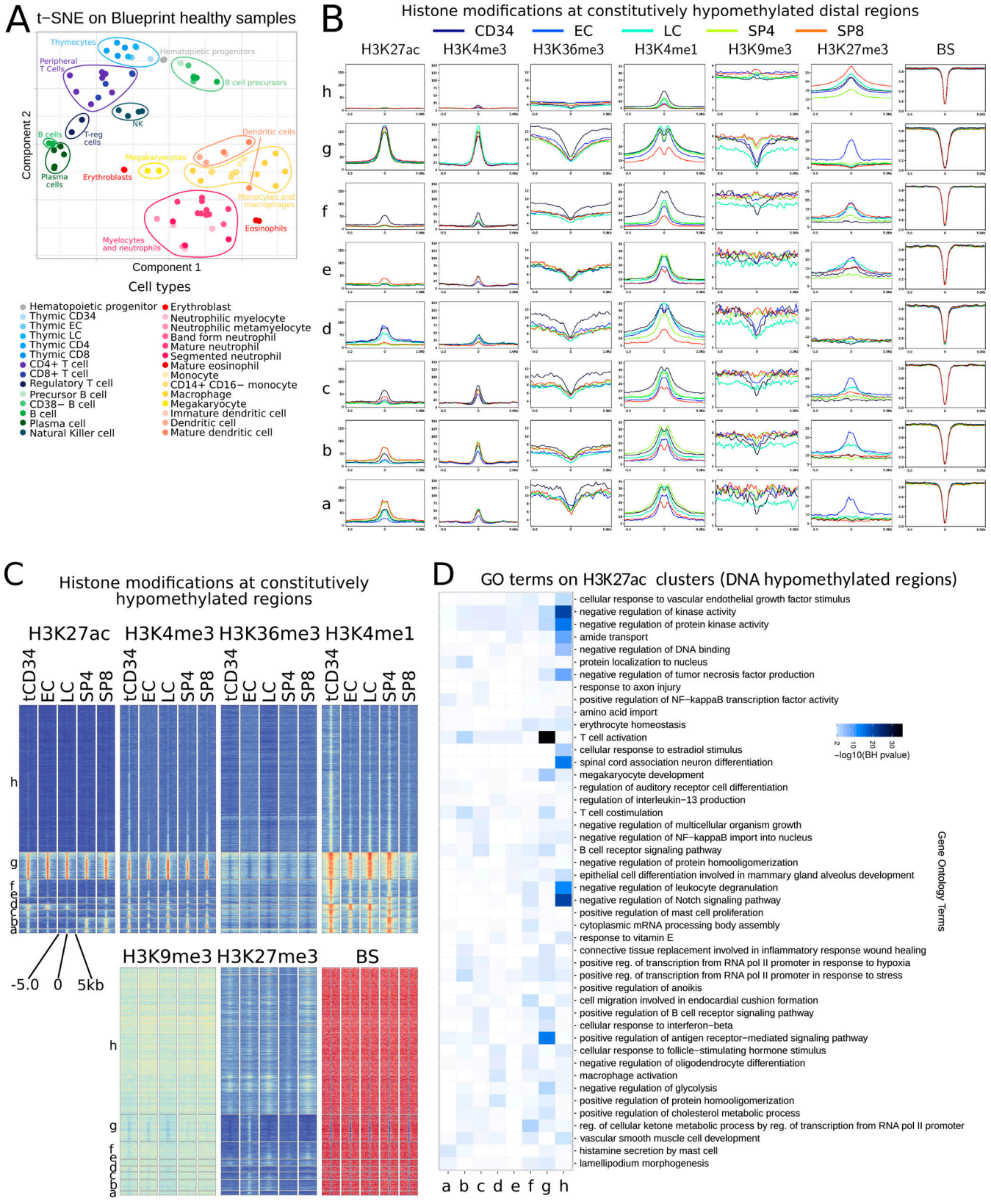


Figure S2. **Integration of DNA methylation and histone modifications. (A)** t-SNE on WGBS from all BLUEPRINT healthy hematopoietic samples. **(B)** Histone modifications and methylation profiles around the eight clustered classes based on hypomethylation dynamics during T cell differentiation (Fig. 2 A). **(C)** Histone modifications and methylation profiles around the eight clustered classes based on the H3K27-acetylation dynamics during T cell differentiation at constitutively hypomethylated regions. **(D)** GREAT gene enrichment analysis for the 8 clustered classes based on H3K27-acetylation dynamics during T cell differentiation at constitutively hypomethylated regions. Only GO biological process terms passing a 0.2 Wang similarity threshold are shown.

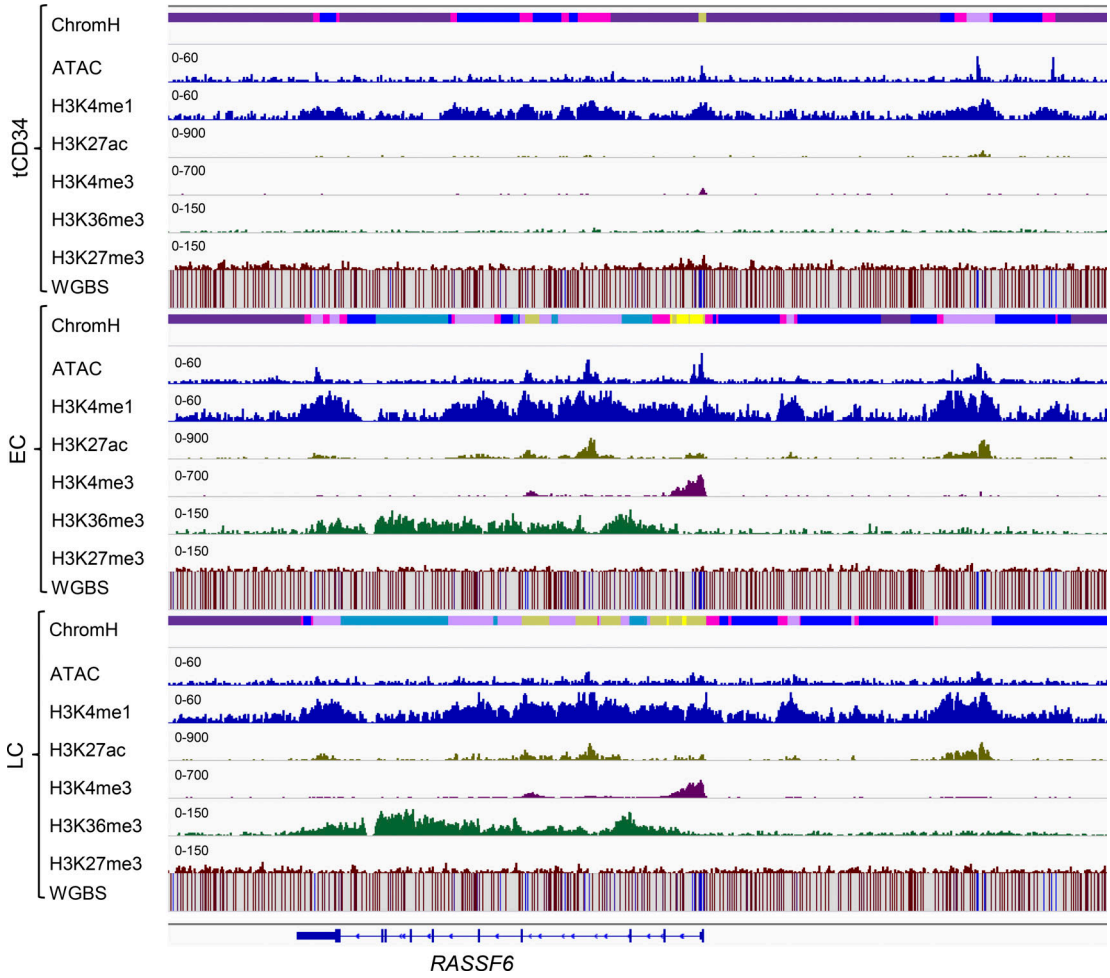


Figure S3. **Epigenomic profile of the induced gene *RASSF6*.** The normalized signal range is indicated by brackets.

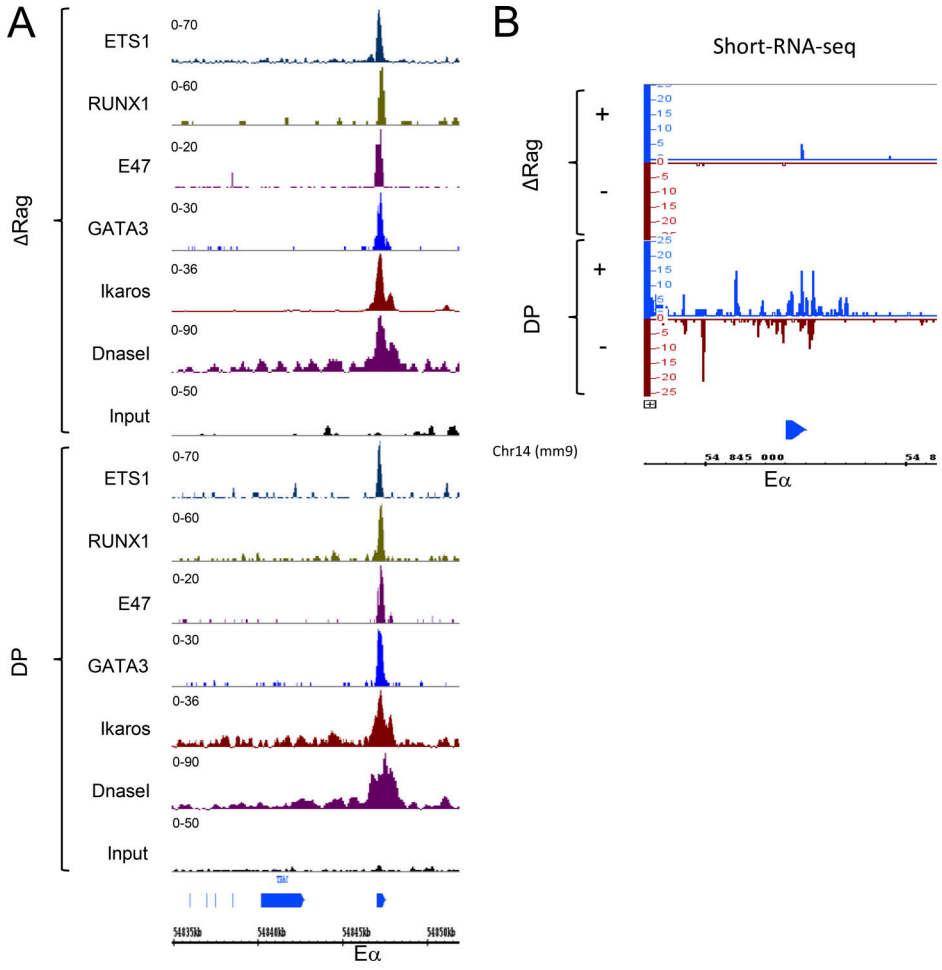
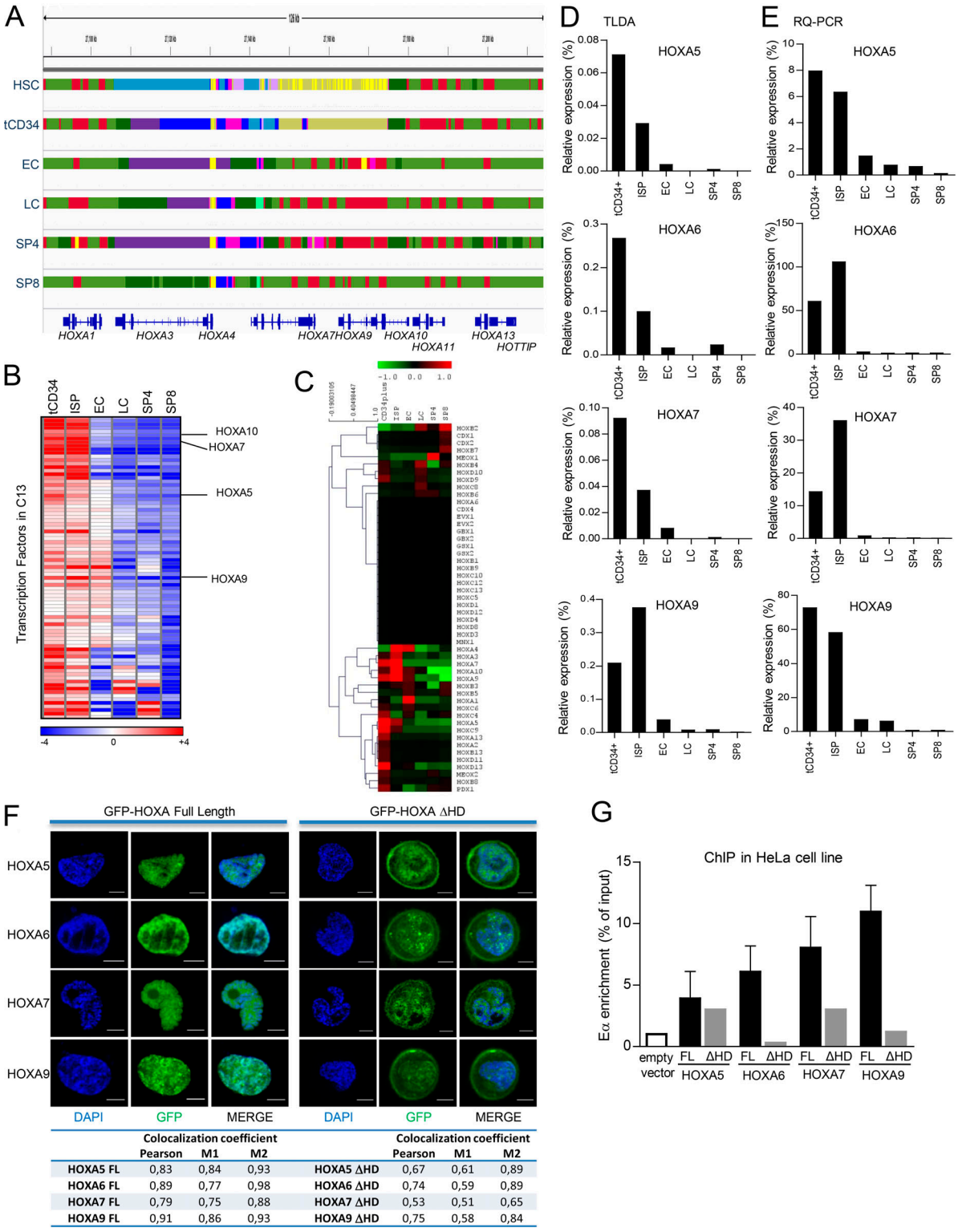


Figure S4. **ChIP-seq and short-RNA-seq profiles of the *TCRA* locus in the mouse thymocytes. (A)** DNaseI-seq and ChIP-seq analysis for ETS1, RUNX1, E47, GATA3, and Ikaros binding to Eα in the ΔRag and DP mouse thymocytes. **(B)** Short-RNA-seq analysis of the enhancer RNA at the Eα enhancer in the ΔRag and DP murine thymocytes. Data sources are detailed in Table S6.



	Colocalization coefficient				Colocalization coefficient		
	Pearson	M1	M2		Pearson	M1	M2
HOXA5 FL	0,83	0,84	0,93	**HOXA5 ΔHD**	0,67	0,61	0,89
HOXA6 FL	0,89	0,77	0,98	**HOXA6 ΔHD**	0,74	0,59	0,89
HOXA7 FL	0,79	0,75	0,88	**HOXA7 ΔHD**	0,53	0,51	0,65
HOXA9 FL	0,91	0,86	0,93	**HOXA9 ΔHD**	0,75	0,58	0,84

Figure S5. **Expression of *HOXA5-9* genes in human thymic subpopulations, their nuclear localization, and binding to Eα. (A)** Chromatin state profiles of the human *HOXA* locus. The color code is as in Fig. 1 C. **(B)** Heatmap representation of the cluster C13. **(C)** Unsupervised RNA-seq gene expression clustering of the HOXL gene family. **(D)** Histogram of the *HOXA5-9* gene expression analyzed by TLDA. Results are normalized to the *GAPDH* housekeeping gene expression. **(E)** RQ-PCR of the *HOXA5-9* gene expression in human thymic subpopulations. Results are relative to the *ABL1* housekeeping gene expression. **(F)** Examples of fluorescence microscopic analysis of the HOXA5-9 expression in HeLa cells transfected with expression vectors for FL or HD-deleted forms (ΔHD) of HOXA5-9. The white scale bar is 10 mm. In the table below the colocalization coefficients, Pearson's and Mander's (M1 and M2) are presented for each condition. Values range from 0 (no colocalization) to 1 (perfect colocalization). **(G)** Anti-Flag ChIP-qPCR analysis of HOXA5-9 binding to Eα in HeLa cells transfected with ETS1, RUNX1, LEF1, TCRA-CAT, and Flagged HOXA5-9 (FL or ΔHD forms) expressing vectors. Enrichment at Eα at the substrate TCRA-CAT vector is shown relative to input and normalized to empty vector control. Results represent means and SEM of triplicate reactions.

Tables S1–S6 are provided online. Table S1 shows the design of HOXL assay in TLDA. Table S2 lists samples and meta-samples from high-throughput sequencing approaches. Table S3 shows normalized counts for RNA-seq samples. Table S4 shows differential expression analysis. Table S5 lists the oligonucleotides and gRNAs used in the study. Table S6 lists key resources.