Alternative Enhancer Usage and Targeted Polycomb Marking Hallmark Promoter Choice during T Cell Differentiation

Muhammad Ahmad Maqbool, Léo Pioger, Amal Zine El Aabidine, Nezih Karasu, Anne Marie Molitor, Lan T M Dao, Guillaume Charbonnier, Francois van Laethem, Romain Fenouil, Frederic Koch, et al.

▶ **To cite this version:**

Muhammad Ahmad Maqbool, Léo Pioger, Amal Zine El Aabidine, Nezih Karasu, Anne Marie Molitor, et al.. Alternative Enhancer Usage and Targeted Polycomb Marking Hallmark Promoter Choice during T Cell Differentiation. Cell Reports, 2020, 32, 10.1016/j.celrep.2020.108048 . hal-03015018

HAL Id: hal-03015018
https://amu.hal.science/hal-03015018

Submitted on 19 Nov 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

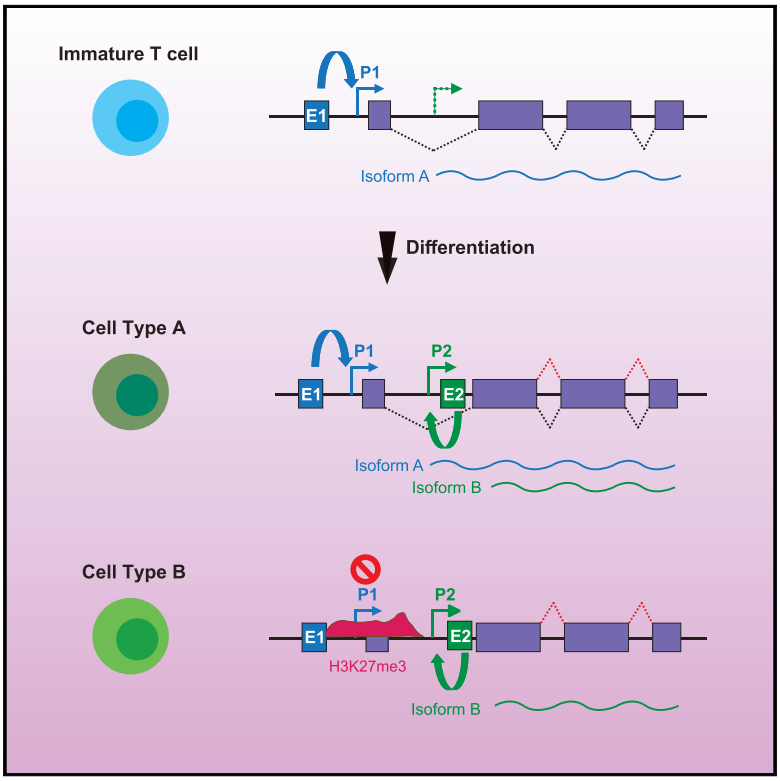
Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

Cell Reports

Resource

Alternative Enhancer Usage and Targeted Polycomb Marking Hallmark Promoter Choice during T Cell Differentiation

Graphical Abstract



Highlights

- Active T cell p-enhancers are highly dynamic during differentiation
- Enhancer diversity might function to select specific isoform expression
- Loss of H3K27me3 combined with enhancer gain hallmark T cell identity
- Promoter choice is regulated by the PRC2 polycomb complex during differentiation

Authors

Muhammad Ahmad Maqbool, Léo Pioger, Amal Zine El Aabidine, ..., Thomas Sexton, Salvatore Spicuglia, Jean-Christophe Andrau

Correspondence

muhammad.maqbool@manchester.ac.uk (M.A.M.), jean-christophe.andrau@igmm.cnrs.fr (J.-C.A.)

In Brief

Development and activation of T lymphocytes are coordinated by lineage-specific transcriptional programs. Here, Maqbool et al. performed a wide epigenomic and transcriptional analysis of mouse T cell differentiation. These data provide new insights into the role of multiple enhancers and PRC2 in controlling alternative promoter choice during differentiation.

Maqbool et al., 2020, Cell Reports *32*, 108048
August 18, 2020 © 2020 The Authors.
https://doi.org/10.1016/j.celrep.2020.108048

CellPress

Cell Reports

CellPress
OPEN ACCESS

Resource

Alternative Enhancer Usage and Targeted Polycomb Marking Hallmark Promoter Choice during T Cell Differentiation

Muhammad Ahmad Maqbool,[1,12,*] Léo Pioger,[1] Amal Zine El Aabidine,[1] Nezih Karasu,[2,3,4,5] Anne Marie Molitor,[2,3,4,5] Lan T.M. Dao,[6] Guillaume Charbonnier,[6] Francois van Laethem,[1] Romain Fenouil,[1] Frederic Koch,[1] Georges Lacaud,[7] Ivo Gut,[8,9] Marta Gut,[8,9] Sebastian Amigorena,[10] Olivier Joffre,[11] Thomas Sexton,[2,3,4,5] Salvatore Spicuglia,[6] and Jean-Christophe Andrau[1,13,*]

[1]Institut de Génétique Moléculaire de Montpellier, Université de Montpellier, CNRS, 1919 Route de Mende, Montpellier 34293, France
[2]Institute of Genetics and Molecular and Cellular Biology (IGBMC), 1 rue Laurent Fries, 67404 Illkirch, France
[3]CNRS UMR7104, 1 rue Laurent Fries, 67404 Illkirch, France
[4]INSERM U1258, 1 rue Laurent Fries, 67404 Illkirch, France
[5]University of Strasbourg, 1 rue Laurent Fries, 67404 Illkirch, France
[6]Aix-Marseille University, UMR-S 1090, TAGC, Marseille 13009, France
[7]CRUK Stem Cell Biology Group, Cancer Research UK Manchester Institute, The University of Manchester, Aderley Park, Macclesfield SK104TG, UK
[8]CNAG-CRG, Centre for Genomic Regulation (CRG), Barcelona Institute of Science and Technology (BIST), Baldiri i Reixac 4, 08028 Barcelona, Spain
[9]Universitat Pompeu Fabra (UPF), Barcelona, Spain
[10]Institut Curie, Université Paris Sciences et Lettres, INSERM U932, 75005 Paris, France
[11]Centre de Physiopathologie de Toulouse Purpan, INSERM UMR1043 CHU Purpan – BP 3028, 31024 Toulouse Cedex 3, France
[12]Present address: CRUK Stem Cell Biology Group, Cancer Research UK Manchester Institute, The University of Manchester, Aderley Park, Macclesfield SK104TG, UK
[13]Lead Contact
*Correspondence: muhammad.maqbool@manchester.ac.uk (M.A.M.), jean-christophe.andrau@igmm.cnrs.fr (J.-C.A.)
https://doi.org/10.1016/j.celrep.2020.108048

SUMMARY

During thymic development and upon peripheral activation, T cells undergo extensive phenotypic and functional changes coordinated by lineage-specific developmental programs. To characterize the regulatory landscape controlling T cell identity, we perform a wide epigenomic and transcriptional analysis of mouse thymocytes and naive CD4 differentiated T helper cells. Our investigations reveal a dynamic putative enhancer landscape, and we could validate many of the enhancers using the high-throughput CapStarr sequencing (CapStarr-seq) approach. We find that genes using multiple promoters display increased enhancer usage, suggesting that apparent "enhancer redundancy" might relate to isoform selection. Furthermore, we can show that two *Runx3* promoters display long-range interactions with specific enhancers. Finally, our analyses suggest a novel function for the PRC2 complex in the control of alternative promoter usage. Altogether, our study has allowed for the mapping of an exhaustive set of active enhancers and provides new insights into their function and that of PRC2 in controlling promoter choice during T cell differentiation.

INTRODUCTION

T cell development in the thymus and differentiation to functionally distinct lineages in peripheral lymphoid tissues require a complex interplay of transcription factors, chromatin modifiers, and genomic regulatory elements to ensure proper cell fate commitment (Rothenberg, 2014). Although many studies have described transcription factors (TFs) important for differentiation and putative regulatory elements in the CD4 lineage (for example, see Cauchy et al., 2016; Fu et al., 2012; and Hollenhorst et al., 2007) or described thymic partial epigenetic landscape (Zhang et al., 2012), none has defined the robust set of enhancers that critically define $CD4^+$ T cell identity through their life cycle.

During these last years, the combinatorial analysis of genome-wide transcriptomic and epigenomic datasets identified H3K27ac deposition by p300 and other histone acetyltransferase (HAT) and H3K4me1 in the presence of low H3K4me3 as the hallmarks of active enhancers in the genome of virtually all cell types in metazoans (De Santa et al., 2010; Koch et al., 2011; Pekowska et al., 2011; Rada-Iglesias et al., 2011). However, and although other genome-wide methods, such as assay for transposase-accessible chromatin combined with high-throughput sequencing (ATAC-seq), allowed to map accessible

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

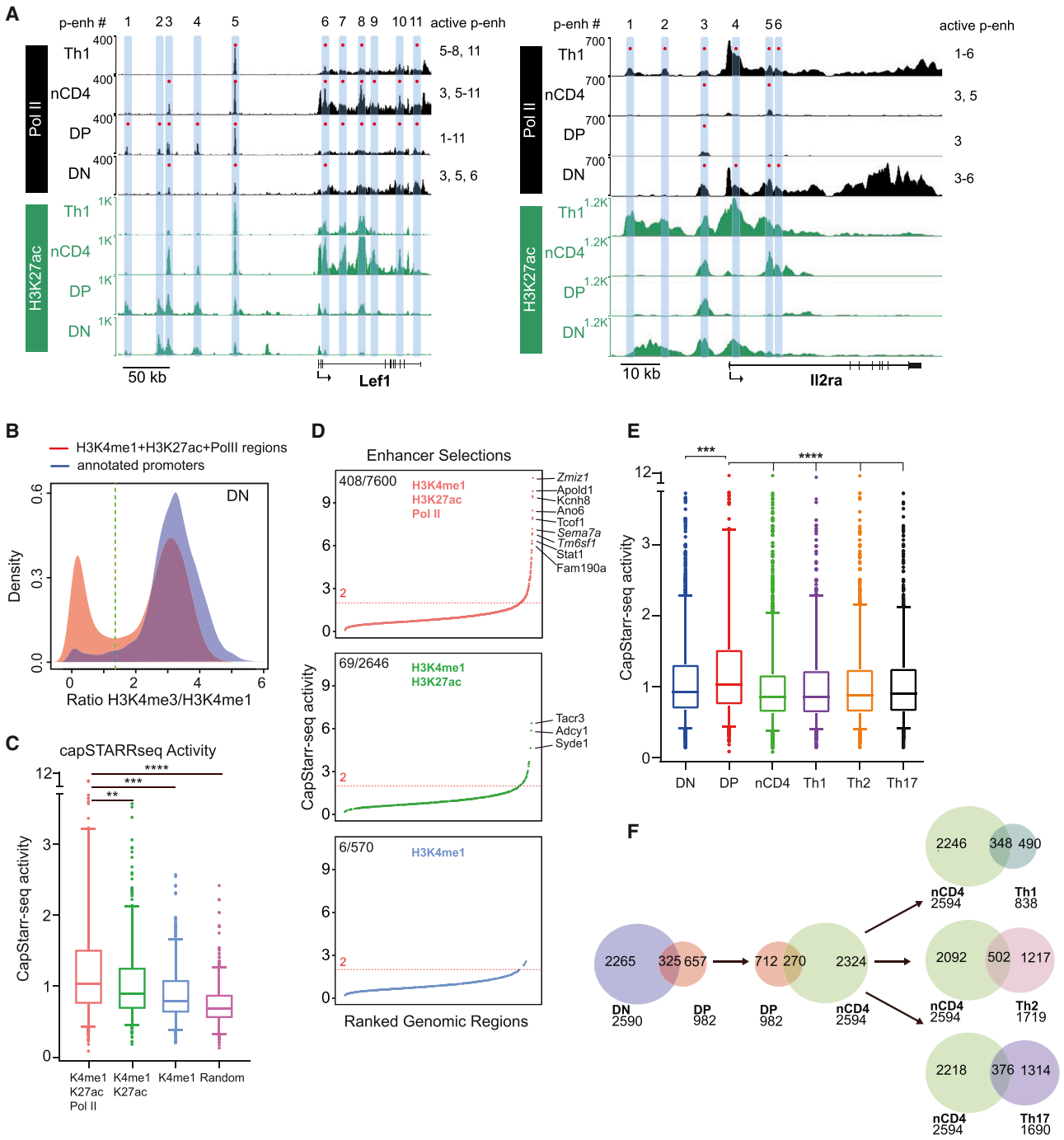


Figure 1. High Enhancer Dynamics in Developing T Cells

(A) ChIP-seq profiles of H3K27ac and Pol II at *Lef1* (left panel) and *Il2ra* (right panel) loci in DN, DP, nCD4, and Th1 cells. Y axes on the left show scaled-normalized ChIP-seq signals as defined in STAR Methods. Light blue vertical bars show putative enhancer regions that are numbered from left to right. Red dots indicate when these enhancers are likely active, and the numbers of active enhancers are shown on the right.

(B) Distribution of H3K4me3/H3K4me1 ratio in DN cells (other stages are shown in Figure S1C). Red curve shows the ratio distribution for all genomic regions enriched with H3K4me1, H3K27ac, and Pol II signal, and the blue curve shows the distribution for annotated promoters only.

(C) CapStarr-seq activity of four classes of *cis*-regulatory elements expressed as fold change over input in P5424 cells. **p = 3.6 × 10^{-7}; ***p = 2 × 10^{-21}; ****p = 2 × 10^{-48} (2-sided Wilcoxon test).

(D) Different classes of putative enhancers were ranked according to their CapStarr-seq activity (fold change over input). Horizontal dashed line indicates the inflection point (FC > 2) above which a region is considered as highly active. Names of putative target genes of selected top-ranked enhancer regions are

(legend continued on next page)

regulatory elements, including in human T cells (Buenrostro et al., 2013; Satpathy et al., 2018), they are not sufficient to accurately predict enhancer activity.

Enhancers and promoters are essential regulatory elements. Although differing in their functions and locations, they share many common characteristics, among which are specific active epigenetic marking and transcription in the initiation or pause mode by RNA polymerase (Pol) II (Core et al., 2014; Koch et al., 2011; Mikhaylichenko et al., 2018). In many species also, including mammals, this transcription is often bidirectional, generally producing small transcripts. Promoters, however, essentially differ from enhancers by their ability to direct Pol II to elongate transcription at protein coding or long non-coding RNA (lncRNA) genes.

Active enhancers are classically defined as regulatory elements able to activate transcription at distance from the promoter in an orientation-independent manner (Andersson and Sandelin, 2020). As mentioned above, putative active enhancers can now be isolated genome wide using both epigenetic and transcriptional hallmarks. One recurrent observation in the field is that genes often harbor several enhancers or putative enhancers in their close proximity or within the same topological associating domain (TAD) (Kim and Shendure, 2019). It has also been observed that, depending on the cell type or developmental stage of an organism, one or another enhancer or enhancer combination can be used by a given gene to ensure a proper spatiotemporal expression pattern or to allow robustness of expression (Dickel et al., 2018; Frankel et al., 2010; Osterwalder et al., 2018). The archetypal example of this model is the *shavenbaby* gene (*svb*) in *Drosophila* (McGregor et al., 2007), which interestingly expresses several isoforms (Salles et al., 2002). These studies, however, do not rule out other possible roles for the need of multiple enhancers. We have shown, for example, in mouse thymocytes that the *Nfatc1* gene makes use of several enhancers, one of which controls the distal promoter, allowing expression of *Nfatc1*α-specific isoform, important for the DN ($CD4^-/CD8^-$) to DP ($CD4^+/CD8^+$) cells transition in T cells (Klein-Hessling et al., 2016).

Here, we generated and analyzed an extensive epigenomic landscape in mouse T cells during development in the thymus and during T helper cells differentiation. Our data indicate that Pol II genomic occupation is an excellent proxy for determining active enhancers as assessed by CapStarr-seq reporter assay (Vanhille et al., 2015). We also show that genes with multiple promoter usage during differentiation harbor more enhancers. In addition, specific enhancer-promoter contacts can be detected on the *Runx3* model locus prior and after activation of alternative promoters. Overall, these observations suggest a novel role for enhancers in isoform selection. Further dynamic mapping of the repressive epigenome indicates that putative enhancer (p-enhancer) gain combined to loss of H3K27me3 is highly cell specific. Finally, we also provide evidence that genes with alternative promoter usage display PRC2 polycomb marking at the inactive promoter or in the first intron of the gene, suggesting an undescribed role for this repressive complex in the control of promoter usage.

RESULTS

Active Enhancer Repertoire during T Cell Development and Differentiation

To characterize the repertoire of active regulatory elements in T cells, we performed genome-wide mapping of active epigenetic marks and Pol II at six distinct stages during mouse T cell differentiation. This analysis included developing primary DN (i.e., $CD4^-CD8^-$) and DP (i.e., $CD4^+CD8^+$) thymocytes, naive $CD4^+$ (nCD4) T cells, and *ex vivo* differentiated Th1, Th2, and Th17 cells. Using these data, we were able to score for active p-enhancers (or p-enh) based on the co-occurrence of active epigenetic marks H3K4me1 and H3K27ac with Pol II, because transcription is a strong hallmark of active and tissue-specific enhancers (Core et al., 2014; De Santa et al., 2010; Kim et al., 2010; Koch et al., 2011). Visual inspection of genomic profiles indicated that putative enhancers were highly dynamic and often stage or lineage specific, as exemplified at the *Lef1*, *Il2ra*, *Ifng*, *Il4*, or *Il17a* loci (Figures 1A and S1A). The *Lef1* locus, which encodes a transcription factor that critically controls early stages of T cell development and represses Th2 cytokines expression, shows at least 11 regulatory elements, and both intensity and apparent combinatorial usage of these elements vary depending on the cell stage, as shown by varying levels of Pol II and H3K27ac (and H3K4me1; not shown). Although p-enh no. 5 appears active at all stages, other elements are present at specific stages in different combinations. A similar situation is observed for the *Il2ra* locus, which encodes the interleukin-2 (IL-2) receptor alpha chain, in which the observed p-enhancer fits well with the previously determined enhancers at this locus in both mouse and human cells (Li et al., 2017; Wong et al., 2017). At *Lef1* and *Il2ra* loci, we did not observe a systematic link between enhancer usage and/or intensity of Pol II/H3K27ac peaks and gene expression level. At other model T cell loci, such as *Ifng* for Th1, *Il4* for Th2, and *Il17a* for Th17, there was, however, a clearer correlation between the presence of p-enhancer/s and *Ifng*/*Il4*/*Il17* expression. Thus, observed p-enhancers are highly dynamic, and their activation, as determined by their epigenomic and transcriptional profile, does not necessarily influence gene expression level.

Next, we isolated more systematically p-enhancers using genomic enrichment profiles. Our p-enhancer isolation procedure (Figure S1B) discriminates potential non-annotated promoters from enhancers using an exclusion threshold of the H3K4me3/me1 signal ratio (Figures 1B and S1C). It should be noted that this strategy implies that some active enhancers will not be selected, such as the ones with high H3K4me3 levels

indicated on the right. Numbers of regions above threshold as well as total number of regions assayed for each class are indicated. Random controls are shown in Figure S1E.

(E) CapStarr-seq activities of all active putative enhancers (H3K4me1/H3K27ac/Pol II) identified at indicated stages of T cell differentiation as in (C). ***$p = 1.6 \times 10^{-8}$ for DN and DP comparison and ****$p < 1.1 \times 10^{-11}$ for DP to all other stages comparison (2-sided Wilcoxon test).

(F) Venn diagram shows the dynamic changes in genome-wide repertoire of active enhancers during T cell differentiation.

(Pekowska et al., 2011), because we do not have a clear mean to discriminate them systematically from promoters of lncRNAs. When plotting the H3K4me3/me1 ratio, we obtain a characteristic curve with two bulks, the most leftward representing putative enhancers and the most rightward promoters. To estimate the accuracy of our enhancer isolation strategy, including the weight of Pol II occupancy in enhancer activity, and compare various p-enhancer selections, we performed a CapStarr-seq enhancer assay, a novel high-throughput method to quantitatively assess enhancer activity (Vanhille et al., 2015). For this, we selected putative regulatory regions at each T cell differentiation stage (DN, DP, nCD4, Th1, Th2, and Th17) based on enrichment of H3K4me1/K27ac/Pol II, H3K4me1/K27ac, or H3K4me1 alone for their activity in the assay (Figure S1D) and assessed their enhancer activity in P5424 mouse cell line that mimics best the DN/DP thymic differentiation stages (Doty et al., 1999; Saadi et al., 2019). As shown in Figures 1C–1E and S1E, the selection making use of the 3 marks shows the highest average enhancer activity in DP cells and contains the most active regions. Although roughly 5% of highly active enhancers (FC > 2) are observed for the first selection, only 2% and 1% are for the two others. Using a lower threshold for isolating weaker enhancers (FC > 1.5) yields 13%, 8%, and 8% for the 3 selections. The order of magnitude of active enhancers observed is comparable to what was described in previous studies (Vanhille et al., 2015; Wang et al., 2018) and reviewed in Santiago-Algarra et al. (2017). Thus, our global enhancer assay allowed validation of our p-enhancer selection approach and also confirmed previous description of the importance of Pol II transcription as a hallmark of active and tissue-specific enhancer activity *in vivo*.

We next evaluated the dynamics of active regulatory elements during differentiation (Figure 1F) by comparing the total number of p-enhancers at each stage and stage transition. Based on our selections, we isolated 8,171 p-enhancers in total, for an average of 1,361 active p-enhancers per stage. We note variation in the absolute number of p-enhancers isolated at each stage that might either simply relate to a difference of signal to noise of individual chromatin immunoprecipitation sequencing (ChIP-seq) experiments or to an actual difference in the requirement of these distal regulatory regions. Strikingly, however, our analysis reveals little overlap of enhancers at each transition as compared to that observed for transcriptome variations of the putative enhancer-dependent genes (Figure S1F). For example, although only one-third of the p-enhancers found in DP cells are already active in DN, 97% of the genes expressed in DP are already expressed in DN. We conclude that, as described by others in various models (He et al., 2016; Wamstad et al., 2012; Wang et al., 2008), T cell p-enhancer activity is highly dynamic and that enhancer usage is changing rapidly during development and differentiation.

To further characterize T cell enhancers, we performed gene ontology and transcription factor binding site (TFBS) analyses. As described earlier (Koch et al., 2011), an enhancer selection based on Pol II recruitment (or enhancer transcription) allows for high level of tissue specificity as determined by gene ontology for all 6 T cell types considered (Figure 2A). Classes, such as immune system, lymphocyte activation, and differentiation, were found enriched for both DN and DP thymic cells although immune response was found only in activated T helper cells. We then performed motif search at enhancer-associated, nucleosome-depleted regions (see STAR Methods) that were specifically active in one stage of differentiation (Figure 2B). This analysis indicated that basic-helix-loop-helix (bHLH) motifs are specific to thymic p-enhancers although Ets-related TFBS are found over-enriched in nCD4 and bZIP and STAT in T helper cell populations. More conventional TFBS search using the genome as a background highlighted once again the importance of the Ets and Runx-Ets in all populations (Cauchy et al., 2016; Hollenhorst et al., 2007; Rothenberg, 2014) as well as that of AP1-related sites in activated Th cells (Figure S2A). In contrast, control T cell promoters showed no or relatively weak enrichment for any motifs (Figure S2B). Finally, to infer whether p-enhancer-dependent genes isolated at a given stage harbor more stage-specific expression, we compared their transcript level and distribution for all stages to that to the overall gene expression in RNA-seq data. Our analysis indicates that this is the case for DN, DP nCD4, and Th1 p-enhancers (Figures 2C and S2C) but less pronounced for other stages (data not shown). Finally, we also analyzed average epigenetic profiling at enhancers or their surrounding genes and found that relative H3K4me3 and H3K27ac levels are slightly more pronounced at enhancers as compared to promoters in DP and CD4 naive T cells, respectively (Figures 2D and S2D). We note that H3K79me2, described as being a hallmark of first intron and some enhancers (Godfrey et al., 2019; Huff et al., 2010), is weakly or not enriched in our p-enhancer selection as compared to promoters, possibly because we disfavored regions with lncRNA features that display enhancer activity. Finally, we also investigated the features of the highly active enhancers as determined in the CapStarr-seq assay in DP cells and found that these regions clearly over-enrich in T cell functions in gene ontology (GO) analyses and their epigenetic and chromatin profiles feature high levels of H3K27ac, Pol II, and increased chromatin openness (Figures S2E and S2F).

In summary, our investigations allowed us to isolate several thousands of transcribed p-enhancers, many of which display enhancer activity in our genome-wide assay. These p-enhancers are highly dynamic and stage specific, but their presence or local abundance do not necessarily correlate with expression levels of putative target genes, raising the question of the role of enhancer redundancy during development or differentiation.

Enhancer Dynamics Moderately Correlate with Long-Distance Interaction Variations

Next, we asked whether enhancer dynamics correlate with changes in long-distance interactions. To address this question, we chose four genomic locations where important p-enhancer dynamics were observed and performed chromatin conformation capture on chip (4C) experiments, using promoters as docking sites, to score for long-distance interactions in DN, DP, and nCD4 T cells. As described above, important p-enhancer changes are observed at the *Lef1* locus during these transitions. In DN cells, *Lef1* promoter establishes several contacts upstream of transcription start site (TSS) and one downstream located close to the 3′ end of the gene. These contacts match

CellPress
OPEN ACCESS

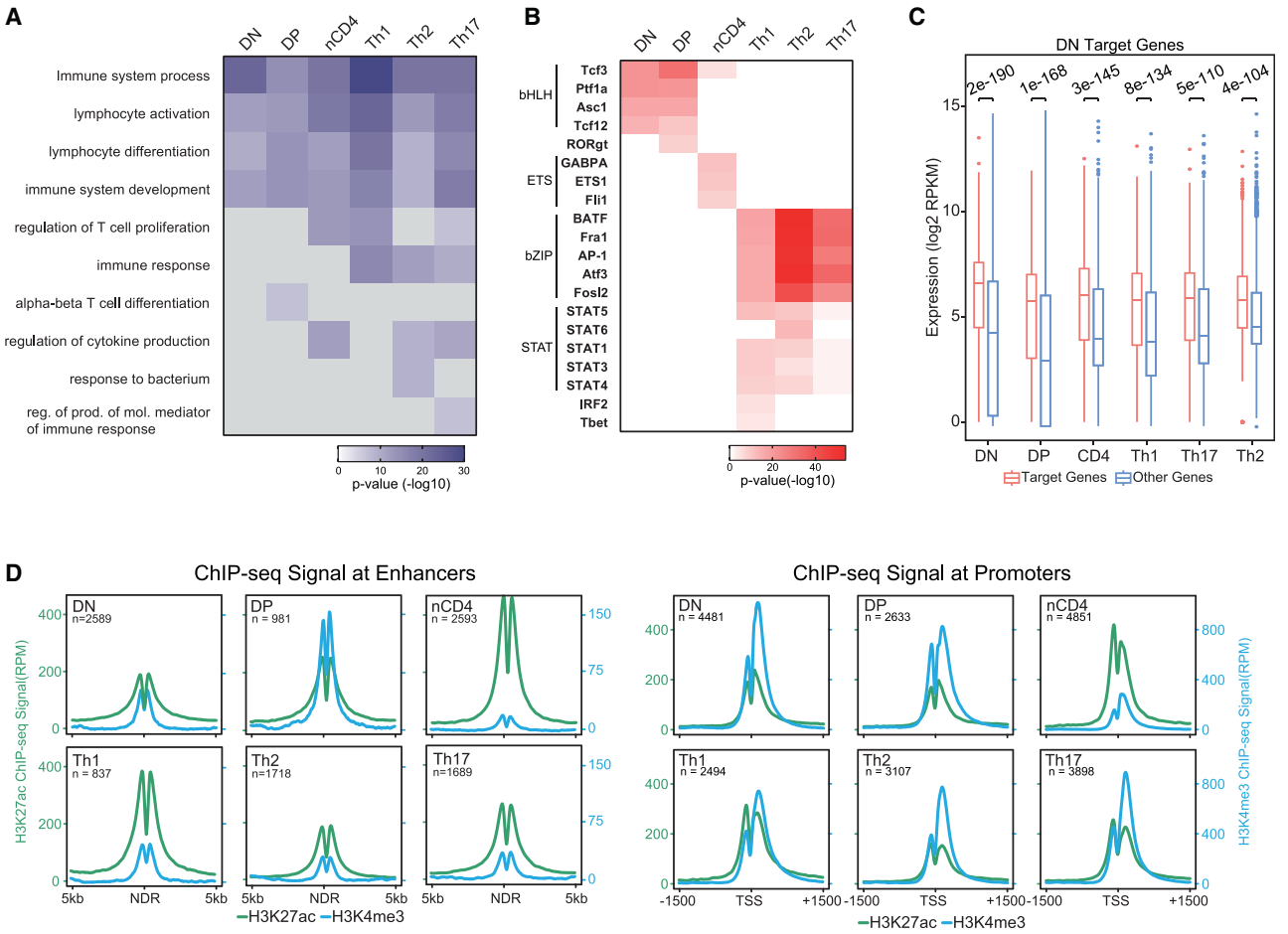


Figure 2. Stage-Specific Gene Expression and Transcription Factor Usage at T Cell Enhancers

(A) Heatmap of gene ontology p values for the putative target genes of active enhancers identified at indicated stage of T cell differentiation as determined by GREAT (McLean et al., 2010). Only top 10 are shown.

(B) Heatmap showing the results of transcription factor binding site (TFBS) analysis of active enhancers unique to indicated stage of T cell differentiation (defined as enhancers that are not found active in any of the five other stages).

(C) Expression levels of putative target genes of enhancers active in DN cells compared to all other genes. Cell stages are ranked according to p value (Mann-Whitney-Wilcoxon test) of difference between expression levels of putative target genes and all non-target genes. Lowest p value indicates highest stage specificity of expression levels of putative target genes. Similar results for DP, nCD4, and Th1 target genes are shown in Figure S2C.

(D) Average enrichment profiles of H3K27ac and H3K4me3 at active enhancer regions (left panels) and active promoter regions (right panels) for indicated cell type. Profiles are docked in the center of enhancer NDRs as described in the STAR Methods or at annotated transcription start sites (TSSs) for promoters; see Figure S2D for other Pol II/epigenetic profiles at active promoters and enhancers.

well with the p-enh no. 1 to no. 5 and nos. 9 to 10 and are essentially conserved in DP and nCD4 cells (Figure 3A). However, and interestingly, both a substantial increase of 4C interaction and H3K27ac signal is observed in an area located in the first intron of Lef1 and matching nos. 6–9 p-enhancers. This region also fits the criteria of definition of transcription initiation platforms (Koch et al., 2011) or enhancer clusters (Lovén et al., 2013; also called super enhancers).

We analyzed 3 other T cell loci, *S1pr1*, *mir181*, and *Nfatc1*, in which important variations of enhancer usage and/or Pol II/H3K27ac signals are observed (Figures 3B–3D). In these cases, more modest changes in contacts between promoters and enhancers were observed during differentiation. At the *S1pr1* locus that becomes activated in nCD4 cells, one of the p-enhancers (no. 3) located upstream of the gene increases its interaction with the promoter. All other p-enhancers in which H3K27 becomes acetylated seem to have pre-existing contacts established prior to activation. A similar situation is observed for the *mir181a* and *Nfatc1* genes, specifically active at the DP stage, in which promoter contacts are present prior to activation and remain afterward.

Overall, our experiments at 4 individual loci with high p-enhancer dynamic activity indicate that, although specific promoter-enhancer contacts can be concomitant with gene/enhancer activation, this situation is not systematic because specific interactions are also observed prior and/or after gene activation, as previously reported by others (Dixon et al., 2015; Ghavi-Helm et al., 2014).

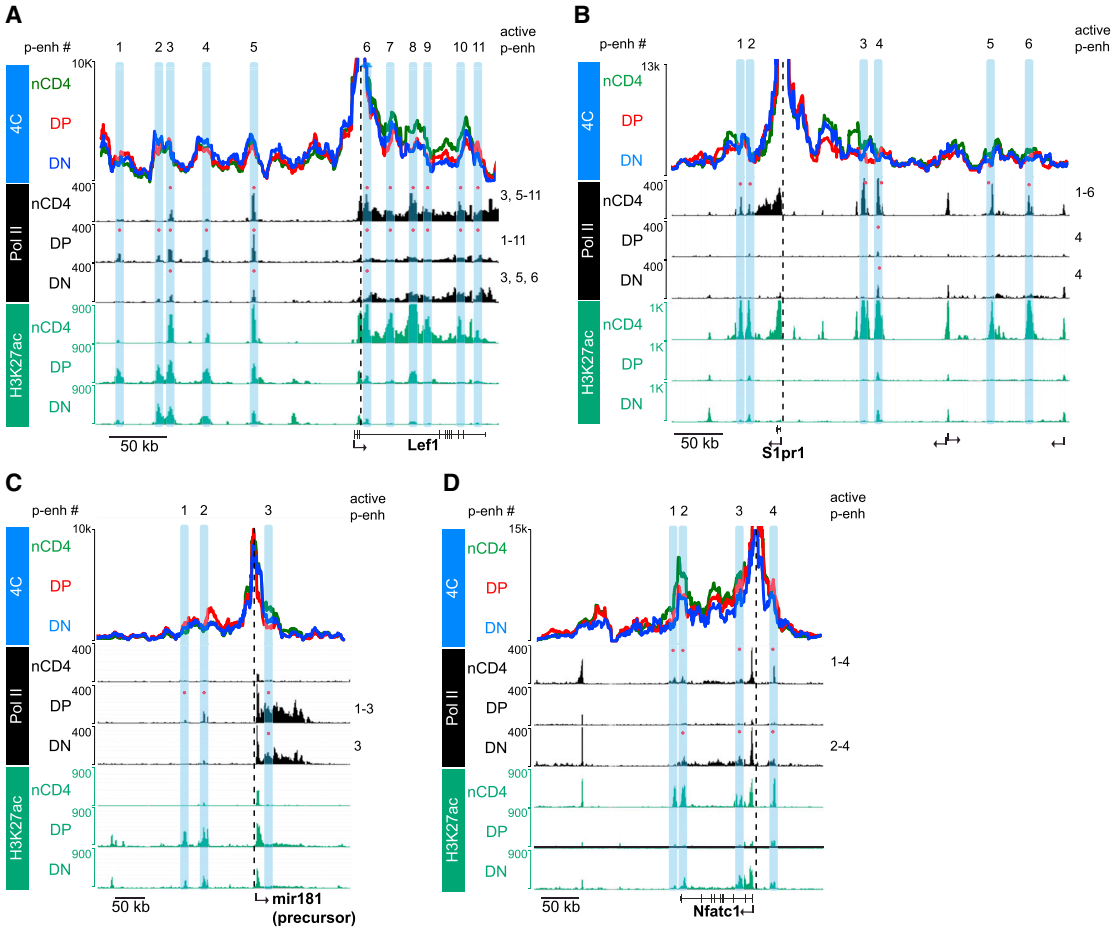


Figure 3. Moderate Dynamic of Enhancer-Promoter Interactions during T Cell Differentiation

(A–D) Chromatin conformation capture (4C-seq), H3K27ac, and Pol II ChIP-seq dynamics during T cell differentiation at *Lef1*, *S1pr1*, *mir181*, and *Nfatc1* loci. Vertical dashed red line indicates the 4C docking point, and light blue bars indicate putative enhancer regions. Red dots indicate most active enhancers based on epigenomic profiles. Y axes represent arbitrary units of 4C signals as defined in the STAR Methods. Only areas corresponding to p-enh nos. 6–9 of the *Lef1* ($p = 0.0041$ between CD4 and DN3 and $p = 0.02$ between CD4 and DN3) and no. 3 of the S1pr1 ($p = 0.047$ between CD4 and DN3 and $p = 0.056$ between CD4 and DN3) were found statistically different using a two-tailed t test.

Genes with Multiple Promoters Display Higher p-Enhancer Dynamics

We found that many genes are transcribed through multiple promoters during T cell differentiation, as determined by H3K4me3 and/or RNA-seq data, giving rise to various isoforms expression as previously described in other cell lines or tissues. To further investigate and analyze the regulation of this process, we first isolated two distinct sets of expressed genes: (1) those making use of at least two distinct promoters during differentiation or (2) those that use one unique promoter (see STAR Methods and Figure S3A). Prior further analysis, we controlled that the 2 sets displayed similar distribution of expression, which was the case (Figure S3B). Strikingly, the first set is highly enriched for genes involved in gene expression control and/or signaling (Figure 4A) although the sets of control genes with one unique promoter are not enriched for any specific gene ontologies (Figures 4A and S3C). The fact that transcription control and signaling ontologies are only enriched when multiple promoters are used is particularly interesting because this potentially introduces a new way of tuning the genetic information during differentiation through regulation of gene isoform expression. We then wondered whether genes with differential promoter usage displayed evidence of alternative enhancer usage. To address this question, we estimated the number of active p-enhancers at proximity (within 50 Kb) of the two gene sets and found that significantly more p-enhancers are present at proximity of genes with multiple promoters (Figure 4B). Similar results were obtained when varying the distance threshold between promoters and p-enhancers or when constraining our search within the same TADs (Figure S3D). We conclude that genes with alternative promoter usage display more active p-enhancers in their proximity and thus possibly more p-enhancer dynamicity.

The *Runx3* locus represents an example of p-enhancer dynamic associated with alternative promoter usage (Figure 4C). Runx3 is an important transcription factor expressed in T cells under the control of two promoters. The Runx3 long isoform is

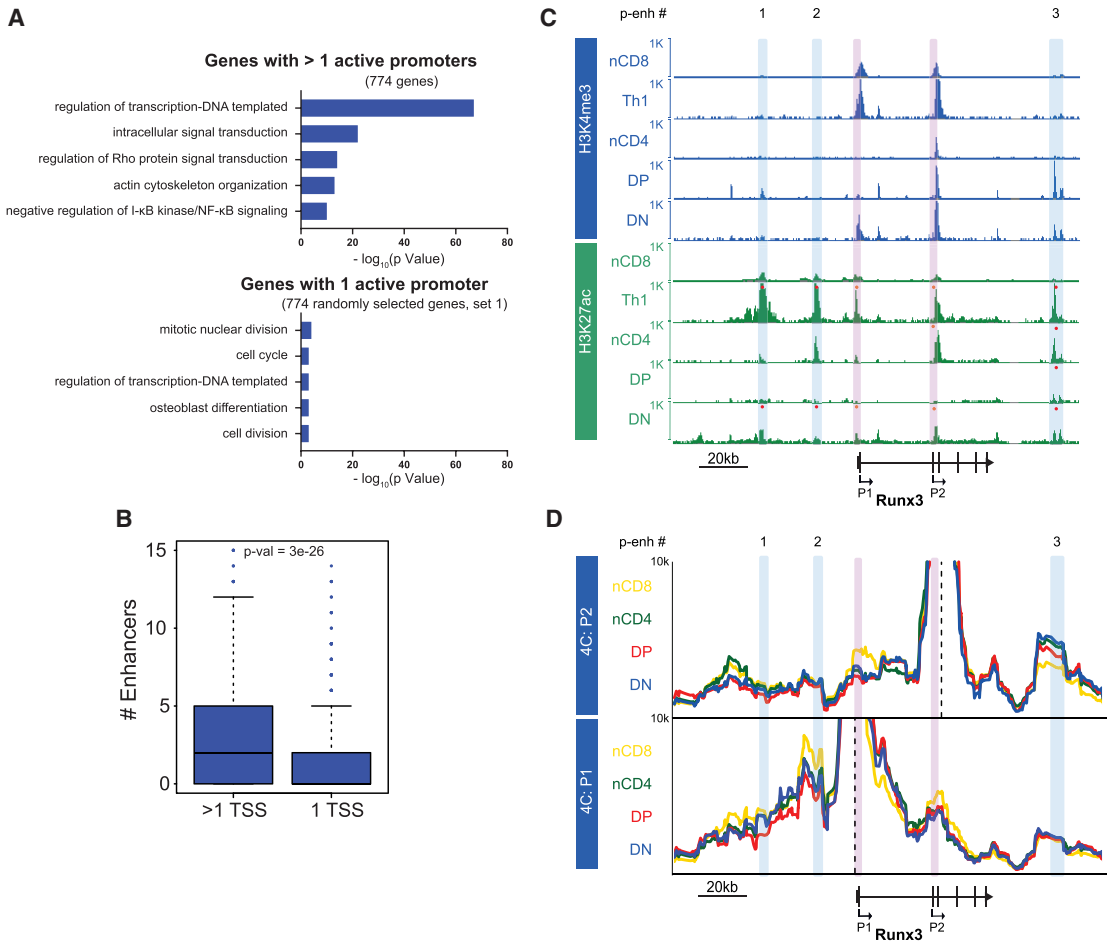


Figure 4. Alternative Promoter Usage Relates to Transition in Enhancer Usage

(A) p values of top 5 significant gene ontology (GO) terms for biological process found with the list of 774 genes with more than one active promoter during T cell differentiation (top panel). Similar results for a set of randomly selected 774 control genes with only one active promoter are shown in lower panel. Figure S3B shows that all gene sets have similar distribution of expression at the DP stage. Figure S3C shows the results of two other independent iterations of control genes.

(B) Comparison of average number of active enhancers in proximity (50 Kb and with no TAD constraint) of the genes with more than one active promoter and control genes that have only one promoter active during T cell differentiation. Figure S3D shows similar analyses for genes in various distance intervals with or without TAD constraints.

(C) H3K27ac and H3K4me3 ChIP-seq profile at *Runx3* locus shows differential activity of alternative promoter (vertical light red bars) correlates with alternative activity of enhancers (vertical light blue bars). See the Figure S3E for H3K4me1 and Pol II ChIP-seq profile at *Runx3* locus. nCD8 cells data originate from He et al. (2016).

(D) 4C-seq profiles viewed from two alternate promoters (as indicated by vertical red lines) of *Runx3* gene. See Figure S3F for 4C-seq profiles viewed from the enhancer sites of *Runx3* gene.

under the control of the distal promoter (P1) and expressed in Th1 and CD8$^+$ T cells (Egawa and Littman, 2008; Egawa et al., 2007). Its short isoform, under the control of the proximal promoter (P2), is expressed in other T cell subsets, including thymocytes (Egawa et al., 2007; Taniuchi et al., 2002). However, our own RNA-seq and H3K4me3 data indicate that the long isoform is also expressed in DN thymic cells. Based on the criteria described above, we found 3 main active p-enhancers present in at least one differentiation stage (Figures 4C and S3E). On one hand, when both p-enh no. 1 and no. 2, located upstream of the TSS, were marked active in DN and Th1 cells, both promoters were used for gene transcription. On the other hand, p-enh no. 3 was active at all stages, including when no. 1 and/or no. 2 were not active. This suggests that P1 usage depends on p-enh no. 1 and/or no. 2 although P2 depends on no. 3. To tackle this question and directly connect promoters and enhancers of this locus, we performed 4C experiments with P1, P2, and p-enh no. 1, no. 2, and no. 3 as docking points (Figures 4D and S3F) in DN, DP, nCD4, and CD8 T cells. Our results shed light on major interactions of P2 with P1 and p-enh no. 3 although P1 preferably interacts with the no. 1/2 area. When analyzing the p-enhancer long-range interactions, we observed consistently major contacts between no. 1 and no. 2 with P1, although no. 3 interacted with P2. Another striking observation is the apparent lack of strong dynamic of these various interactions, suggesting that promoters-enhancers (p-e) interactions are primed prior and

remain after p-enhancer activation, as described above for the *mir181a* and *Nfatc1* loci (Figure 3).

All in all, we found that alternative promoter usage is linked to genes important for signaling and transcription in T cells and that the corresponding genes display more enhancer usage on average, suggesting a novel role of enhancers in transcription regulation. A detailed investigation of the *Runx3* locus further showed that alternative promoters can connect to specific p-enhancer sets. In this case, the long-range interactions associated to p-e pairs are not dynamic during differentiation, suggesting pre-established and stable p-e contacts during differentiation.

Regulation of Promoter Choice through PRC2-Specific Repression of Alternative Promoters

To further characterize the epigenome of developing T cells and more specifically the interplay between activation and repression of regulatory elements, we performed profiling of the H3K27me3 repressive chromatin mark, deposited by the PRC2 polycomb complex, at all six stages of differentiation. We first analyzed active enhancers being repressed by PRC2 and gaining H3K27me3 during Th activation or being de-repressed by losing H3K27me3 with a gain of H3K27ac (Figure 5A) that are exemplified by the *Fasl* and *Bhlhe40* loci, respectively (Figure 5B). We found 174–303 regions with enhancer loss and 48–114 with enhancer gain from nCD4 to either Th1, Th2, and Th17. Although a substantial overlap was found for repressed enhancers between the 3 lineages, little or no overlap was found for activated enhancers, suggesting that these latter loci harbored more specificity in each lineage. When analyzing the TFBS associated with these regions, Runx and GFi1b motifs were predominant at repressed loci although Zbtb12 and Zbtb3 were the major sites found at activated areas (Figure 5C). We note that Runx and Gfi1b can function as transcriptional repressors in hematopoietic lineage and that Zbtb12 is also described as a transcriptional repressor, suggesting that gain or loss of repressor sites could play a role in polycomb-mediated dynamic regulation. Finally, we investigated the usage specificity of known motifs and found that enhancer activation TFBS enriched in PRC2-repressed regions was specific as compared to enhancer repressed by PRC2 (Figure 5D), consistent with the specific enhancers observed in each of the 3 lineages Th1, Th2, or Th17 (Figure 5A). Overall, these data thus indicate that Th enhancers activated from PRC2-repressed regions are lineage specific although PRC2-repressed areas arising following Th differentiation are generally shared between lineages.

We next examined whether the PRC2 complex could play a role in regulating alternative promoter selection. For this analysis, we again used our set of genes with one or multiple promoter usage (Figure S3A) and examined their overlap with H3K27me3 signal. Although, at the global level, 25% of unique promoter genes harbored H3K27me3 signal in at least one differentiation stage, this fraction climbed to over 60% for genes with multiple promoters (Figure 6A). To further narrow this result and avoid potential bias, we selected, at each differentiation stage, expressed genes (RPKM > 0.5). Although the number of genes in each group associated with H3K27me3 decreased, PRC2-occupied promoters remained clearly higher at each stage in the multi-TSS group (Figure S4A). These results suggest that PRC2 could play a role in the expression of specific isoforms through repression of the other existing promoters of the genes. We observed such repression for genes such as *Runx3* or *Ptpre* (Figure 6B), for which H3K27me3 repressive signal encompassed either the repressed promoter(s) and/or the first intron(s). To further confirm this observation at the genome-wide scale, we performed metagene profiling of our two sets of genes with one/multiple promoters by docking the analysis at the most distal annotated promoter. This showed that genes with multiple promoters clearly harbor more H3K27me3 signal on average around their transcription start sites although lower signal is observed for Pol II, H3K27ac, or K4me3 (Figures 6C, 6D, S4B, and S4C). Over gene bodies, multiple-promoter genes harbored decreased H3K27me3, comparable to that of the unique promoter group, suggesting specific promoter repression although Pol II or H3K79me2 signal remained unchanged (Figures 6E, S4D, and S4E). Finally, global analysis further indicated that multiple-promoter genes harbor significantly longer first intron (Figure 6F).

In sum, our analyses have revealed that p-enhancers' dynamic activity can be regulated by gain or loss of polycomb PRC2 repression, making use of repressor-binding-sites-associated regulatory regions. We also shed light on a new potential role for the PRC2 complex in regulating promoter choice during differentiation through repression of alternative promoter-driven specific isoforms.

DISCUSSION

In this article, we describe a global epigenomic landscape of murine T cell differentiation in the thymus and in *in vitro* differentiated T helper cell populations. Besides providing a large and robust resource for scientists to further study the dynamic regulome of T cells, our observations emphasize novel findings on the role of regulatory elements hallmarked by enhancer signature.

First, we used Pol II recruitment as a critical hallmark of active enhancers, an approach that we validated by genome-wide enhancer assay. Our rationale was supported by previous descriptions of enhancer transcription as a strong determinant of enhancer activity (Core et al., 2014; De Santa et al., 2010; Kim et al., 2010; Koch et al., 2011). Although other studies have described dynamics of open regions or used combination of one or more epigenetic marks to describe T cell p-enhancers, our large dataset offers a more accurate picture of dynamic activity of p-enhancers during differentiation. In our study, we used stringent selection criteria disfavoring regions with relatively high H3K4me3 levels, thus displaying features of noncoding RNA promoters. This helped us, at least in part, to identify the p-enhancers that show highly dynamic and cell-type-specific activity and to describe some characteristics of regulatory elements associated to enhancer redundancy.

Enhancer redundancy has been a long-standing question in the field and previously proposed to support spatiotemporal gene expression during development (Osterwalder et al., 2018) but also to fine-tune transcription rate and promoter bursting (Fukaya et al., 2016; Larsson et al., 2019). It has also been described that developmental genes, such as the

CellPress
OPEN ACCESS

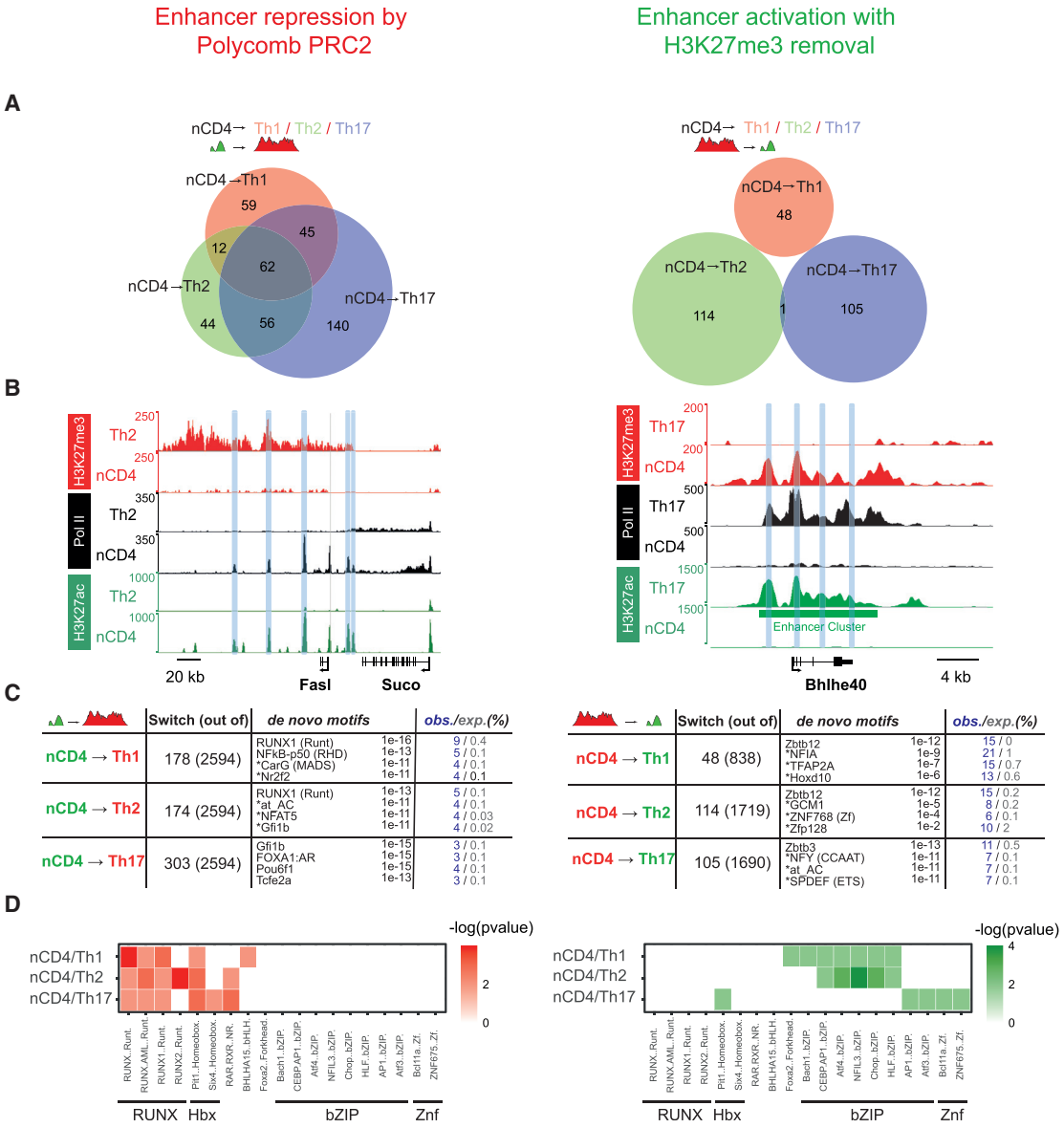


Figure 5. Switch of Enhancer Activity and PRC2 Repressed Regions during Th Differentiation

(A) Overlap of repressed (left panel) and activated (right panel) switch regions during Th differentiation associated with gain or loss of PRC2 activity, respectively.

(B) H3K27ac, H3K27me3, and Pol II ChIP-seq profiles at *Fasl* and *Bhlhe40* loci show the repression (left panel) and activation (right panel) of enhancers (vertical light blue bars).

(C) Tables showing the first hits of *de novo* TFBS analyses of repressed (left panel) or activated enhancers (right panel) during Th differentiation, with as background non-repressed enhancers, of the same stage without H3K27me3 (* possible false positive).

(D) Heatmaps showing the results of known TFBS analyses of repressed (left) or activated enhancers (right).

classical Hox or shadow enhancers, tend to have more enhancers and larger gene loci (reviewed in Bolt and Duboule, 2020 and Spitz and Furlong, 2012). Here, we propose a working model in which enhancers also condition alternative promoter usage and thus potentially favor new regulation modes through differential isoform expression. Strikingly, functions of genes associated with expression of more than one promoter-dependent isoform during differentiation are linked to transcription regulation or signal transduction. This suggests that enhancers introduce an additional level of regulation of genes coding for proteins with strong effector potential by controlling the activity of alternative promoters, thus influencing the differential expression of gene isoforms. In the same line, a previous study proposed that enhancer clusters could control microRNA miRNA maturation (Suzuki et al., 2017). Based on this new possible function for enhancers, it is tempting to speculate that enhancers could also regulate isoform expression through alternative splicing or 3′ end selection, an idea that was

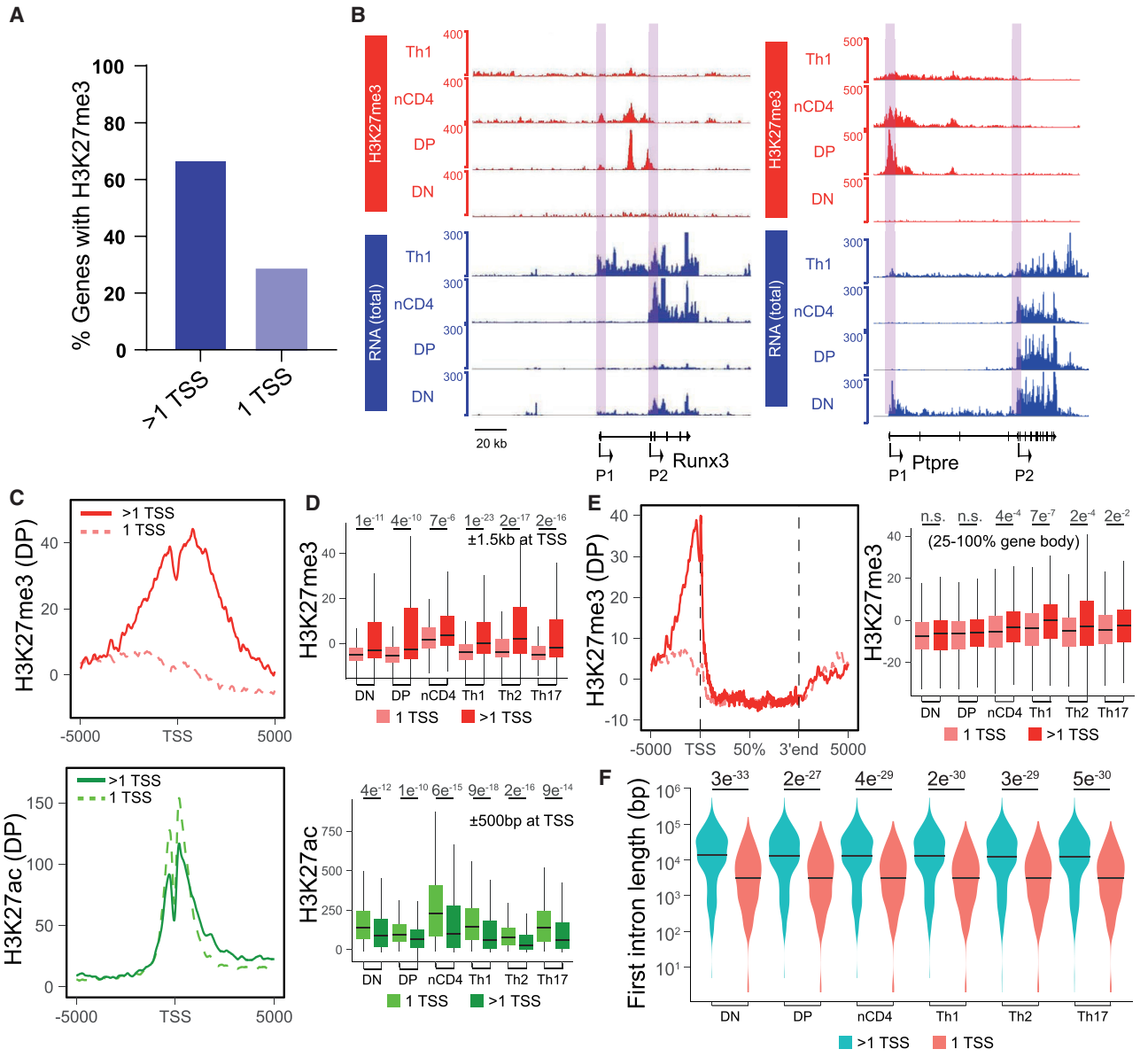


Figure 6. Polycomb Repression of Unused Promoter during Activation of Alternative Promoter

(A) Fraction of genes with significant H3K27me3 signal over the gene body in at least one stage of differentiation in gene sets with one or multiple promoter usage. Results for expressed genes only at each stage are presented in Figure S4A.

(B) H3K27me3 ChIP-seq and RNA-seq signal at the *Runx3* and *Ptpre* loci suggest PRC2-mediated repression of P1 distal promoter and/or first exon.

(C) ChIP-seq metagene profile analyses of H3K27me3 and H3K27ac around TSS of gene sets with RPKM > 0.5 associated to one or multiple promoters at the DP stage. See Figure S4B for Pol II and H3K4me3 ChIP-seq metagene profiles.

(D) ChIP-seq of H3K27me3 and H3K27ac mean signal at all stages, around TSS of gene sets with one or multiple active promoters, and filtered by RPKM > 0.5 at a given stage. See Figure S4C for Pol II and H3K4me3 signals at TSS at all stages.

(E) H3K27me3 ChIP-seq metagene profile analysis on the gene body at DP stage (left) and mean signal at all stages on part of the gene body (right), with gene sets filtered at each stage by RPKM > 0.5. See Figure S4D for Pol II and H3K79me2 and Figure S4E for Pol II and H3K79me2 signals at all stages.

(F) Distribution of first intron length for gene sets with one or multiple promoter usage at each differentiation stage. The difference of distribution for each group is significant using a Mann-Whitney-Wilcoxon test (p values are indicated on top).

previously suggested based on DNase I profiles (Mercer et al., 2013) and that will further deserve future attention.

Second, our observations on five model loci indicate that promoter-enhancer long-distance interactions can be pre-existing before or remain after activation (Runx3, S1pr1, and Nfatc1) but also induced during activation (Lef1 and mir181). These several modes of action observed in T cells support the existing view on how enhancers might be at play in various models or cell

types (Dixon et al., 2015; Ghavi-Helm et al., 2014; Zhang et al., 2013). In the case of *Runx3*, our model for alternative promoter usage in CD4 and CD8 subtypes, long-distance interactions between p-enh no. 1 and 2 with P1 (distal promoter) and no. 3 with P2 also support pre-exiting contacts in the case of alternative promoter choice.

Third, analysis of dynamic PRC2 activity in T helper cells indicates that, although PRC2-mediated enhancer repression (decommissioning) is common to the 3 helper cell types, enhancer activation or re-activation from repressed regions is more specific to each lineage. A simple explanation could be the set of transcription factors and/or repressors used in each case. This difference could also involve more drastic mechanistic difference in each case, resulting in a more specific path for T helper cell fate determination in the case of H3K27me3 removal and/or transcription factors recruited. This could take place through a different nature of 3D hubs involved when PRC2 activity is induced or erased in the genome of T cells.

Finally, another important observation of this study is that PRC2 could be involved in promoter selection by repressing one specific TSS location but leaving another free for activation, thereby influencing the expression of a specific gene isoform. We have also observed that, at many promoters showing stage-specific H3K27me3 marking, H3K4me3 was also present at a substantial or moderate level, thus defining an apparent bivalent locus (data not shown). In early days of genome-wide epigenetic characterization, bivalency of active and inactive marks were described in embryonic stem cells (ESCs) as a mechanism to allow flexibility in adopting one or another developmental fate (Azuara et al., 2006; Bernstein et al., 2006; Mikkelsen et al., 2007). Our finding offers a novel possible function for this apparent bivalency. That H3K27me3 hallmarks at least a fraction of repressed alternate promoters does not demonstrate a direct causal effect of PRC2 on this repression that might also involve other complexes or mechanisms. When analyzing transcriptome data for Ezh1/Ezh2-depleted ESCs (Lavarone et al., 2019) or inhibiting PRC2 enzymatic activity in human T cells, we did not observe a specific de-repression of the promoters that are generally H3K27me3 enriched in wild type (WT) or untreated cells (data not shown). It is then possible that, even though PRC2 would be involved in promoter repression, it is not sufficient for the complete blocking of repressed promoter and that PRC1 could be also involved. Finally, our data do not exclude the possibility that PRC2 may act downstream of transcriptional repressors to maintain already repressed promoters in an off state as previously suggested (Riising et al., 2014).

Overall, our study presents an extensive epigenetic map of T cell differentiation that opens the path to understanding novel mechanisms, such as the role of enhancers in controlling alternative promoter usage as well as on that of the function of PRC2 and apparent bivalency in differentiated cells. Future work, building on these data, will further help deciphering the precise mechanism of these processes.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead Contact
  - Materials Availability
  - Data and Code Availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Thymic $CD4^-/CD8^-$ and $CD4^+/CD8^+$ cell isolation
  - $CD4^+$ and $CD8^+$ T cell purification
  - $CD4^+$ T cell cultures
- METHOD DETAILS
  - I- Experimental Procedures
  - II- Bioinformatic Procedures
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.celrep.2020.108048.

ACKNOWLEDGMENTS

We thank Chales Lecellier for his help and advice on TFBS analyses. This work was supported in J.-C.A.'s lab by grants from Fondation pour la Recherche Médicale (FRM) "Amorçage Jeunes Equipes" AJE20130728183, Agence Nationale de la Recherche (ANR) isplice (ANR-11-BSV8-0013), ANR ChromaTin (ANR-10-BLAN-1326), ITMO INCA "Dys3Dpoly," ESGI 2011 (FP7 funding for high-throughput sequencing), and Ligue Régionale contre le Cancer (Hérault). L.P. was supported by a grant from FRM (4eme année thèse FDT201904007910). The T.S. lab was supported by funds from the European Research Council (ERC) (H2020, Starting Grant 678624 - CHROMTOPOLOGY), the ATIP-Avenir program, and the grant ANR-10-LABX-0030-INRT, under program ANR-10-IDEX-0002-02. A.M.M. is supported by funds from IDEX (University of Strasbourg) and the Institut National Du Cancer.

AUTHOR CONTRIBUTIONS

J.-C.A. and M.A.M. conceived the study and most of the experimental frame. M.A.M. and O.J. prepared the sorted primary and *in vitro* differentiated T cells used in this study. M.A.M. performed all the ChIP-seq and RNA-seq experiments, including QCs and library preparations. F.K. contributed in some ChIP-seq experiments. N.K., A.M.M., and T.S. performed 4C experiments and analyzed results. L.T.M.D., G.C., and S.S. performed CapStarr-seq experiments and analyzed the results. M.A.M., along with L.P. and A.Z.E.A., performed bioinformatics analyses using in-house scripts, some of which were developed by R.F. I.G. and M.G. sequenced the ChIP-seq and RNA-seq libraries. M.A.M., L.P., T.S., and J.-C.A. prepared and finalized the figures. M.A.M. and J.-C.A. wrote the manuscript, which was reviewed by all authors.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: February 17, 2020
Revised: April 13, 2020
Accepted: July 28, 2020
Published: August 18, 2020

REFERENCES

Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq–a Python framework to work with high-throughput sequencing data. Bioinformatics *31*, 166–169.

Andersson, R., and Sandelin, A. (2020). Determinants of enhancer and promoter activities of regulatory elements. Nat. Rev. Genet. *21*, 71–87.

Azuara, V., Perry, P., Sauer, S., Spivakov, M., Jørgensen, H.F., John, R.M., Gouti, M., Casanova, M., Warnes, G., Merkenschlager, M., and Fisher, A.G. (2006). Chromatin signatures of pluripotent cell lines. Nat. Cell Biol. *8*, 532–538.

Ben Zouari, Y., Platania, A., Molitor, A.M., and Sexton, T. (2020). 4See: a flexible browser to explore 4C data. Front. Genet. *10*, 1372.

Bernstein, B.E., Mikkelsen, T.S., Xie, X., Kamal, M., Huebert, D.J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., et al. (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell *125*, 315–326.

Bolt, C.C., and Duboule, D. (2020). The regulatory landscapes of developmental genes. Development *147*, dev171736.

Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat. Methods *10*, 1213–1218.

Cauchy, P., Maqbool, M.A., Zacarias-Cabeza, J., Vanhille, L., Koch, F., Fenouil, R., Gut, M., Gut, I., Santana, M.A., Griffon, A., 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.

Core, L.J., Martins, A.L., Danko, C.G., Waters, C.T., Siepel, A., and Lis, J.T. (2014). Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers. Nat. Genet. *46*, 1311–1320.

Dao, L.T.M., Galindo-Albarrán, A.O., Castro-Mondragon, J.A., Andrieu-Soler, C., Medina-Rivera, A., Souaid, C., Charbonnier, G., Griffon, A., Vanhille, L., Stephen, T., et al. (2017). Genome-wide characterization of mammalian promoters with distal enhancer functions. Nat. Genet. *49*, 1073–1081.

De Santa, F., Barozzi, I., Mietton, F., Ghisletti, S., Polletti, S., Tusi, B.K., Muller, H., Ragoussis, J., Wei, C.L., and Natoli, G. (2010). A large fraction of extragenic RNA pol II transcription sites overlap enhancers. PLoS Biol. *8*, e1000384.

Descostes, N., Heidemann, M., Spinelli, L., Schüller, R., Maqbool, M.A., Fenouil, R., Koch, F., Innocenti, C., Gut, M., Gut, I., et al. (2014). Tyrosine phosphorylation of RNA polymerase II CTD is associated with antisense promoter transcription and active enhancers in mammalian cells. eLife *3*, e02105.

Dickel, D.E., Ypsilanti, A.R., Pla, R., Zhu, Y., Barozzi, I., Mannion, B.J., Khin, Y.S., Fukuda-Yuzawa, Y., Plajzer-Frick, I., Pickle, C.S., et al. (2018). Ultraconserved enhancers are required for normal development. Cell *172*, 491–499.e15.

Dixon, J.R., Jung, I., Selvaraj, S., Shen, Y., Antosiewicz-Bourget, J.E., Lee, A.Y., Ye, Z., Kim, A., Rajagopal, N., Xie, W., et al. (2015). Chromatin architecture reorganization during stem cell differentiation. Nature *518*, 331–336.

Doty, R.T., Xia, D., Nguyen, S.P., Hathaway, T.R., and Willerford, D.M. (1999). Promoter element for transcription of unrearranged T-cell receptor beta-chain gene in pro-T cells. Blood *93*, 3017–3025.

Egawa, T., and Littman, D.R. (2008). ThPOK acts late in specification of the helper T cell lineage and suppresses Runx-mediated commitment to the cytotoxic T cell lineage. Nat. Immunol. *9*, 1131–1139.

Egawa, T., Tillman, R.E., Naoe, Y., Taniuchi, I., and Littman, D.R. (2007). The role of the Runx transcription factors in thymocyte differentiation and in homeostasis of naive T cells. J. Exp. Med. *204*, 1945–1957.

Fenouil, R., Descostes, N., Spinelli, L., Koch, F., Maqbool, M.A., Benoukraf, T., Cauchy, P., Innocenti, C., Ferrier, P., and Andrau, J.C. (2016). Pasha: a versatile R package for piling chromatin HTS data. Bioinformatics *32*, 2528–2530.

Frankel, N., Davis, G.K., Vargas, D., Wang, S., Payre, F., and Stern, D.L. (2010). Phenotypic robustness conferred by apparently redundant transcriptional enhancers. Nature *466*, 490–493.

Fu, W., Ergun, A., Lu, T., Hill, J.A., Haxhinasto, S., Fassett, M.S., Gazit, R., Adoro, S., Glimcher, L., Chan, S., et al. (2012). A multiply redundant genetic switch 'locks in' the transcriptional signature of regulatory T cells. Nat. Immunol. *13*, 972–980.

Fukaya, T., Lim, B., and Levine, M. (2016). Enhancer control of transcriptional bursting. Cell *166*, 358–368.

Ghavi-Helm, Y., Klein, F.A., Pakozdi, T., Ciglar, L., Noordermeer, D., Huber, W., and Furlong, E.E. (2014). Enhancer loops appear stable during development and are associated with paused polymerase. Nature *512*, 96–100.

Godfrey, L., Crump, N.T., Thorne, R., Lau, I.J., Repapi, E., Dimou, D., Smith, A.L., Harman, J.R., Telenius, J.M., Oudelaar, A.M., et al. (2019). DOT1L inhibition reveals a distinct subset of enhancers dependent on H3K79 methylation. Nat. Commun. *10*, 2803.

He, B., Xing, S., Chen, C., Gao, P., Teng, L., Shan, Q., Gullicksrud, J.A., Martin, M.D., Yu, S., Harty, J.T., et al. (2016). $CD8^+$ T cells utilize highly dynamic enhancer repertoires and regulatory circuitry in response to infections. Immunity *45*, 1341–1354.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell *38*, 576–589.

Hollenhorst, P.C., Shah, A.A., Hopkins, C., and Graves, B.J. (2007). Genome-wide analyses reveal properties of redundant and specific promoter occupancy within the ETS gene family. Genes Dev. *21*, 1882–1894.

Huff, J.T., Plocik, A.M., Guthrie, C., and Yamamoto, K.R. (2010). Reciprocal intronic and exonic histone modification regions in humans. Nat. Struct. Mol. Biol. *17*, 1495–1499.

Kim, S., and Shendure, J. (2019). Mechanisms of interplay between transcription factors and the 3D genome. Mol. Cell *76*, 306–319.

Kim, T.K., Hemberg, M., Gray, J.M., Costa, A.M., Bear, D.M., Wu, J., Harmin, D.A., Laptewicz, M., Barbara-Haley, K., Kuersten, S., et al. (2010). Widespread transcription at neuronal activity-regulated enhancers. Nature *465*, 182–187.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. *14*, R36.

Klein-Hessling, S., Rudolf, R., Muhammad, K., Knobeloch, K.P., Maqbool, M.A., Cauchy, P., Andrau, J.C., Avots, A., Talora, C., Ellenrieder, V., et al. (2016). A threshold level of NFATc1 activity facilitates thymocyte differentiation and opposes notch-driven leukaemia development. Nat. Commun. *7*, 11841.

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

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. *10*, R25.

Larsson, A.J.M., Johnsson, P., Hagemann-Jensen, M., Hartmanis, L., Faridani, O.R., Reinius, B., Segerstolpe, Å., Rivera, C.M., Ren, B., and Sandberg, R. (2019). Genomic encoding of transcriptional burst kinetics. Nature *565*, 251–254.

Lavarone, E., Barbieri, C.M., and Pasini, D. (2019). Dissecting the role of H3K27 acetylation and methylation in PRC2 mediated control of cellular identity. Nat. Commun. *10*, 1679.

Li, P., Mitra, S., Spolski, R., Oh, J., Liao, W., Tang, Z., Mo, F., Li, X., West, E.E., Gromer, D., et al. (2017). STAT5-mediated chromatin interactions in superenhancers activate IL-2 highly inducible genes: Functional dissection of the *Il2ra* gene locus. Proc. Natl. Acad. Sci. USA *114*, 12111–12119.

Lovén, J., Hoke, H.A., Lin, C.Y., Lau, A., Orlando, D.A., Vakoc, C.R., Bradner, J.E., Lee, T.I., and Young, R.A. (2013). Selective inhibition of tumor oncogenes by disruption of super-enhancers. Cell *153*, 320–334.

McGregor, A.P., Orgogozo, V., Delon, I., Zanet, J., Srinivasan, D.G., Payre, F., and Stern, D.L. (2007). Morphological evolution through multiple cis-regulatory mutations at a single gene. Nature *448*, 587–590.

McLean, C.Y., Bristor, D., Hiller, M., Clarke, S.L., Schaar, B.T., Lowe, C.B., Wenger, A.M., and Bejerano, G. (2010). GREAT improves functional interpretation of cis-regulatory regions. Nat. Biotechnol. *28*, 495–501.

Mercer, T.R., Edwards, S.L., Clark, M.B., Neph, S.J., Wang, H., Stergachis, A.B., John, S., Sandstrom, R., Li, G., Sandhu, K.S., et al. (2013). DNase I-hypersensitive exons colocalize with promoters and distal regulatory elements. Nat. Genet. *45*, 852–859.

Mikhaylichenko, O., Bondarenko, V., Harnett, D., Schor, I.E., Males, M., Viales, R.R., and Furlong, E.E.M. (2018). The degree of enhancer or promoter activity is reflected by the levels and directionality of eRNA transcription. Genes Dev. *32*, 42–57.

Mikkelsen, T.S., Ku, M., Jaffe, D.B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T.K., Koche, R.P., et al. (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature *448*, 553–560.

Osterwalder, M., Barozzi, I., Tissières, V., Fukuda-Yuzawa, Y., Mannion, B.J., Afzal, S.Y., Lee, E.A., Zhu, Y., Plajzer-Frick, I., Pickle, C.S., et al. (2018). Enhancer redundancy provides phenotypic robustness in mammalian development. Nature *554*, 239–243.

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

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842.

Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S.A., Flynn, R.A., and Wysocka, J. (2011). A unique chromatin signature uncovers early developmental enhancers in humans. Nature *470*, 279–283.

Riising, E.M., Comet, I., Leblanc, B., Wu, X., Johansen, J.V., and Helin, K. (2014). Gene silencing triggers polycomb repressive complex 2 recruitment to CpG islands genome wide. Mol. Cell *55*, 347–360.

Rothenberg, E.V. (2014). The chromatin landscape and transcription factors in T cell programming. Trends Immunol. *35*, 195–204.

Saadi, W., Kermezli, Y., Dao, L.T.M., Mathieu, E., Santiago-Algarra, D., Manosalva, I., Torres, M., Belhocine, M., Pradel, L., Loriod, B., et al. (2019). A critical regulator of Bcl2 revealed by systematic transcript discovery of lncRNAs associated with T-cell differentiation. Sci. Rep. *9*, 4707.

Salles, C., Mével-Ninio, M., Vincent, A., and Payre, F. (2002). A germline-specific splicing generates an extended ovo protein isoform required for Drosophila oogenesis. Dev. Biol. *246*, 366–376.

Santiago-Algarra, D., Dao, L.T.M., Pradel, L., España, A., and Spicuglia, S. (2017). Recent advances in high-throughput approaches to dissect enhancer function. F1000Res. *6*, 939.

Satpathy, A.T., Saligrama, N., Buenrostro, J.D., Wei, Y., Wu, B., Rubin, A.J., Granja, J.M., Lareau, C.A., Li, R., Qi, Y., et al. (2018). Transcript-indexed ATAC-seq for precision immune profiling. Nat. Med. *24*, 580–590.

Shinkai, Y., Rathbun, G., Lam, K.P., Oltz, E.M., Stewart, V., Mendelsohn, M., Charron, J., Datta, M., Young, F., Stall, A.M., et al. (1992). RAG-2-deficient mice lack mature lymphocytes owing to inability to initiate V(D)J rearrangement. Cell *68*, 855–867.

Spitz, F., and Furlong, E.E. (2012). Transcription factors: from enhancer binding to developmental control. Nat. Rev. Genet. *13*, 613–626.

Suzuki, H.I., Young, R.A., and Sharp, P.A. (2017). Super-enhancer-mediated RNA processing revealed by integrative microRNA network analysis. Cell *168*, 1000–1014.e15.

Taniuchi, I., Osato, M., Egawa, T., Sunshine, M.J., Bae, S.C., Komori, T., Ito, Y., and Littman, D.R. (2002). Differential requirements for Runx proteins in CD4 repression and epigenetic silencing during T lymphocyte development. Cell *111*, 621–633.

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

Wamstad, J.A., Alexander, J.M., Truty, R.M., Shrikumar, A., Li, F., Eilertson, K.E., Ding, H., Wylie, J.N., Pico, A.R., Capra, J.A., et al. (2012). Dynamic and coordinated epigenetic regulation of developmental transitions in the cardiac lineage. Cell *151*, 206–220.

Wang, Z., Zang, C., Rosenfeld, J.A., Schones, D.E., Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Peng, W., Zhang, M.Q., and Zhao, K. (2008). Combinatorial patterns of histone acetylations and methylations in the human genome. Nat. Genet. *40*, 897–903.

Wang, X., He, L., Goggin, S.M., Saadat, A., Wang, L., Sinnott-Armstrong, N., Claussnitzer, M., and Kellis, M. (2018). High-resolution genome-wide functional dissection of transcriptional regulatory regions and nucleotides in human. Nat. Commun. *9*, 5380.

Wong, R.W.J., Ngoc, P.C.T., Leong, W.Z., Yam, A.W.Y., Zhang, T., Asamitsu, K., Iida, S., Okamoto, T., Ueda, R., Gray, N.S., et al. (2017). Enhancer profiling identifies critical cancer genes and characterizes cell identity in adult T-cell leukemia. Blood *130*, 2326–2338.

Zhang, J.A., Mortazavi, A., Williams, B.A., Wold, B.J., and Rothenberg, E.V. (2012). Dynamic transformations of genome-wide epigenetic marking and transcriptional control establish T cell identity. Cell *149*, 467–482.

Zhang, Y., Wong, C.H., Birnbaum, R.Y., Li, G., Favaro, R., Ngan, C.Y., Lim, J., Tai, E., Poh, H.M., Wong, E., et al. (2013). Chromatin connectivity maps reveal dynamic promoter-enhancer long-range associations. Nature *504*, 306–310.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Rabbit polyclonal anti-H3K4me1	Abcam	ab8895
Rabbit polyclonal anti-H3K4me3	Abcam	ab8580
Rabbit polyclonal anti-H3K27ac	Abcam	ab4729
Rabbit polyclonal anti-H3K27me3	Millipore	07-449
Rabbit polyclonal anti-H3K79me2	Abcam	ab3594
Rabbit polyclonal anti-RNA Polymerase II	Santa Cruz Biotech	Sc-899
Purified Rat Anti-Mouse CD4 - Clone GK1.5	BD Biosciences	553727
Hamster Anti-Mouse TCR β Chain- Clone H57-597	BD Biosciences	553174
Rat Anti-Mouse CD44 - Clone 1M7	BD Biosciences	553133
Rat Anti-Mouse CD62L - Clone MEL-14	BD Biosciences	553152
Mouse Anti-Mouse I-A[b]	BD Biosciences	553551
Chemicals, Peptides, and Recombinant Proteins		
RNaseIII	Thermo Fisher	AM2290
Turbo DNA-Free	Thermo Fisher	AM1907
Deposited Data		
ChIP-seq, RNA-seq, CapSTARR-seq and 4C-seq data	This study	GSE144586
ChIP-seq data	Cauchy et al., 2016	GSE56395
ChIP-seq data	Fenouil et al., 2016	GSE38577
ChIP-seq data	Koch et al., 2011	GSE29362
MNase-seq data	Cauchy et al., 2016	GSE56395
Experimental Models: Cell Lines		
P5424	Dr Salvatore Spicuglia, TAGC Aix-Marseille University, France	N/A
Experimental Models: Organisms/Strains		
C57BL/6 wild-type	IGMM	N/A
RAG 2−/− mice	IGMM	N/A
Software and Algorithms		
PASHA	Fenouil et al., 2016	https://cran.r-project.org/web/packages/Pasha/
Bowtie2	Langmead and Salzberg, 2012	http://bowtie-bio.sourceforge.net/bowtie2/index.shtml
TopHat2	Kim et al., 2013	http://ccb.jhu.edu/software
BEDTools	Quinlan and Hall, 2010	https://github.com/arq5x/bedtools2
GREAT	McLean et al., 2010	http://great.stanford.edu/public/html/

RESOURCE AVAILABILITY

Lead Contact

Further information and requests for reagents should be directed to and will be fulfilled by Lead Contact Jean-Christophe Andrau (jean-christophe.andrau@igmm.cnrs.fr).

Materials Availability

This study did not generate new unique reagents.

Data and Code Availability

All high throughput sequencing data used in this study have been deposited at GEO under accession number GSE144586. Wiggle files of ChIP-seq and RNA-seq data can be viewed on a dedicated UCSC genome browser (http://genome-euro.ucsc.edu/cgi-bin/hgHubConnect?redirect=manual&source=genome.ucsc.edu) Track Hub using this link (https://ndownloader.figshare.com/files/22633064?private_link=3df2f43ff40a9ae78875).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Thymic $CD4^-/CD8^-$ and $CD4^+/CD8^+$ cell isolation

Rag2−/− (ΔRag) (Shinkai et al., 1992) male mice were used to isolate thymic $CD4^-/CD8^-$ cells as described previously (Pekowska et al., 2011).

To isolate $CD4^+/CD8^+$ cells, thymuses were harvested from 5–6-week-old C57BL/6 wild-type mice. After homogenization of the tissue, cells to be used for ChIP were cross-linked by adding 1/10th volume of 10X crosslinking solution (100mM NaCl, 1mM EDTA pH 8, 0.5mM EGTA pH 8, 50mM HEPES pH 7.8 and 11% formaldehyde), while those used for RNA-seq were not and stored in Trizol as described. CD8-R-phycoerythrin antibody and anti-phycoerythrin multisort beads were used to sort CD8+ cells on an AutoMACS system (Miltenyi). After release from beads, sorted cells were subsequently sorted again using anti-CD4 MACS beads (Miltenyi). Purity of the sorted samples was checked using FACS analysis.

$CD4^+$ and $CD8^+$ T cell purification

For purification of naive $CD4^+$ T cells in 5–6-week-old C57BL/6 wild-type mice, spleen and lymph nodes (mesenteric, inguinal, axillary, brachial, superficial cervical, lumbar) were collected and single-cell suspensions were obtained following digestion with Liberase TM (Sigma) and DNase I (Sigma). After red blood cells lysis (Red Blood Cell Lysis buffer, Sigma), CD4 T cells were enriched by negative selection using the Dynabeads untouched mouse CD4 T cells kit (ThermoFisher) according to manufacturer's instructions. Naive $CD4^+$ T cells, defined as $CD4^+CD25^-FR4^-CD62L^{high}CD44^{low}$, were then sorted from the $CD4^+$ T cell enriched fraction by fluorescence activated cell sorting (FACS Aria, BD Biosciences). The population obtained was routinely more than 98% pure.

For purification of naive $CD8^+$ T cells, lymph node (LN) cells from C57BL/6 mice (from The Jackson Laboratories (Bar Harbor, ME)) were first depleted of $CD4^+$ T cells, MHC-II^+ and Ig^+ B cells by antibody-mediated magnetic beads depletion (Biomag, QIAGEN, then electronically sorted as $CD8^+$ TCRβ+ $CD62L^{high}CD44^{low}$ cells using a FACSAria II (BD PharMingen) with a purity > 99%.

$CD4^+$ T cell cultures

Naive CD4 T cells were cultured for three days in 96-well flat bottom plates coated with 10μg/mL anti-CD3Ɛ antibody (145-2C11, NA/LE, BD) in RPMI 1640 Glutamax™ supplemented with 1 mM sodium pyruvate, non-essential amino acids, 10 mM HEPES, 100 units/mL penicillin, 100μg/mL streptomycin, 50 mM 2β-mercaptoethanol, 10% fetal calf serum (all from Thermo Fischer Scientific) and 1 μg/mL anti-CD28 antibody (37.51, NA/LE, BD). Th1 culture medium also contained 10 ng/mL recombinant mouse IL-12 (R&D Systems) and 10 μg/mL anti-IL4 neutralizing antibody (11B11, NA/LE, BD). To induce Th2 cell differentiation, culture medium was supplemented with 50 ng/mL recombinant mouse IL-4 (R&D systems) and 10 μg/mL anti-IFN-γ neutralizing antibody (XMG1.2, NA/LE, BD). Th17 culture medium contained 10 ng/mL recombinant mouse IL-1β (R&D systems), 20 ng/mL recombinant mouse IL-6 (R&D systems), 1 ng/mL recombinant human TGF-β1 (eBioscience) and 10 μg/mL anti-IL4 and anti-IFN-γ antibodies. At day 3, Th1 and Th2 cells were re-plated in the same conditioned medium but without anti-CD3Ɛ and anti-CD28 antibodies and with 30 IU/mL IL-2. Th17 cells were maintained in the same culture conditions. Cells were harvested at day 6.

METHOD DETAILS

I- Experimental Procedures

Antibodies

All antibodies used for ChIPseq in this study are commercially available. H3K4me1 (ab8895: Abcam), H3K4me3 (ab8580: Abcam), H3K27ac (ab4729: abcam), H3K27me3 (07-449: Millipore), H3K79me2 (ab3594: Abcam) and RNA Pol II (sc899x N20: Santa Cruz). Mouse antibodies used for CD4+ and CD8+ sorting (mAbs) with the following specificities were used in this study: CD4 (GK1.5), TCRb (H57-597), CD44 (1M7), CD62L (MEL-14), and I-Ab (AF6-120.1), all obtained from BD Biosciences.

T cell proliferation and differentiation analysis by flow cytometry

To analyze intracellular transcription factor expression upon T helper cell differentiation, cells were collected at the requires time points and labeled with fluorochrome-conjugated antibodies specific for T-bet (ebio4B10, Thermo Fischer Scientific), GATA-3 (TWAJ, Thermo Fischer Scientific) and RORγt (AFKJS-9, Thermo Fischer Scientific) by means of the Transcription Factor Staining Buffer Set (Thermo Fischer Scientific). For intracellular cytokine staining, cells were first stimulated at 37°C with 20 ng/mL phorbol 12-myristate 13-acetate (Millipore) and 1 μg/mL ionomycin (Millipore) for 5 hours in the presence of GolgiStop (BD Biosciences). Cells were then stained with fluorochrome-coupled antibodies specific for IL-13 (ebio13A, Thermo Fischer Scientific), IFN-γ (XMG1.2, Thermo Fischer Scientific), or IL17A (eBio17B7, Thermo Fischer Scientific) by using the Cytofix/Cytoperm Fixation/Permeabilization

Kit (BD Biosciences). Flow cytometry was performed by using a MACSQuant analyzer 10 (Myltenyi) and the data were analyzed by using FlowJo software (Tree Star).

ChIP-seq

To cross-link the cells for ChIP, 1/10th volume of 10X crosslinking solution (100mM NaCl, 1mM EDTA pH 8, 0.5mM EGTA pH 8, 50mM HEPES pH 7.8 and 11% formaldehyde) was added to the cells in culture medium. After 10 minutes incubation at room temperature, glycine was added to a final concentration of 250mM to quench the remaining formaldehyde and stop cross-linking. After five minutes of quenching, cells were washed twice with cold PBS. Cells were then sonicated as described in next paragraph or snap frozen in liquid nitrogen and stored at −80°C for sonication at a later stage.

For sonication, up to 50 million cross-linked cells were lysed by re-suspending in cold 2.5mL LB1 (50mM HEPES pH 7.5, 140mM NaCl, 1mM EDTA pH 8, 10% glycerol, 0.75% NP-40, 0.25% Triton X-100) at 4°C for 20 minutes on a rotating wheel. Nuclei were pelleted down by spinning at 1000x *rcf.* in a refrigerated centrifuge and washed in 2.5mL LB2 (200mM NaCl, 1mM EDTA pH 8, 0.5mM EGTA pH 8, 10mM Tris pH 8) for 10 minutes at 4°C on a rotating wheel followed by centrifugation to collect nuclei. Nuclei were then resuspended in 1mL LB3 (1mM EDTA pH 8, 0.5mM EGTA pH 8, 10mM Tris pH 8, 100mM NaCl, 0.1% Na-Deoxycholate, 0.5% N-lauroylsarcosine) and sonicated using Bioruptor Pico (Diagenode) in 15mL tubes containing ~500μl sonication beads (Diagenode) for 25 cycles of 30 s ON and 30 s OFF pulses in 4°C water bath. All buffers (LB1, LB2 and LB3) were complemented with EDTA free Protease inhibitor cocktail (Roche), 0.2mM PMSF and 1μg/mL Pepstatin just before use. Note that sonication cycles were slightly adjusted for each cell population so to obtain a major DNA peak in input sample with a bulk at 250bp. After sonication, Triton X-100 was added to a final concentration of 1% followed by centrifugation at 20000x *rcf.* and 4°C for 10 minutes to remove particulate matter. After taking a 50μl aliquot to serve as input and to analyze fragmentation, chromatin was aliquoted and snap-frozen in liquid nitrogen and stored at −80°C until use in ChIP assays.

Input aliquots were mixed with equal volume of 2X elution buffer (100mM Tris pH 8.0, 20mM EDTA, 2%SDS) and incubated at 65°C for 12 hours for reverse-crosslinking. An equal volume of TE buffer (10mM Tris pH 8 and 1mM EDTA pH 8) was added to dilute the SDS to 0.5% followed by treatment with RNase A (0.2μg/mL) at 37°C for one hour and Proteinase K (0.2μg/mL) for two hours at 55°C. DNA was isolated by phenol:chloroform: isoamylalcohol (25:24:1 pH 8) extranction followed by Qiaquick PCR Purification (QIAGEN, Germany). Purified DNA was then analyzed on a 2% agarose gel or on Bioanalyzer (Agilent, USA) using a High Sensitivity DNA Assay.

Protein-G coated Dynabeads were incubated at 4°C in blocking solution (0.5% BSA in PBS) carrying specific antibodies to prepare beads pre-coated with specific antibody which were then used for ChIP. Sonicated chromatin was added to pre-coated beads and the mix was incubated overnight at 4°C on a rotating wheel (please refer to the Table titled "Quantity of chromatin and reagents used for each ChIP" for information on specific antibodies and quantity of chromatin used for each ChIP). After incubation with chromatin, beads were washed 7 times with Wash buffer (50mM HEPES pH 7.6, 500mM LiCl, 1mM EDTA pH 8, 1% NP-40, 0.7% Na-Deoxycholate, 1X protease inhibitor cocktail) followed by one wash with TE-NaCl buffer (10mM Tris pH 8 and 1mM EDTA pH 8, 50mM NaCl) and a final wash with TE buffer (10mM Tris pH 8 and 1mM EDTA pH 8). Immunoprecipitated chromatin was eluted by two sequential incubations with 50μl Elution buffer (50mM Tris pH 8, 10mM EDTA pH 8, 1% SDS) at 65°C for 15 minutes. The two eluates were pooled and incubated at 65°C for 12 hours to reverse-crosslink the chromatin followed by treatment with RNase A and Proteinase K and purification of DNA as described above for Input samples. Purified DNA was quantified with Qubit DS DNA HS Assay (ThermoFisher Scientific, USA).

At least 1ng of ChIP DNA was used to prepare sequencing library with Illumina ChIP Sample Library Prep Kit (Illumina, USA). Barcoded libraries from different samples were pooled together and sequenced on Illumina HiSeq2000 platform in paired-end sequencing runs.

Quantity of chromatin and reagents used for each ChIP.

Mark	No. Of cells (x 10^6)	Antibody	Supplier	Antibody quantity (μg)	Dynabeads Protein G (μl)
H3K27ac	5	ab4729	Abcam	2	20
H3K27me3	5	07-449	Millipore	2	20
H3K4me1	5	ab8895	Abcam	2	20
H3K4me3	5	ab8580	Abcam	2	20
H3K79me2	5	ab3594	Abcam	2	20
RNA Pol II	50	sc899x (N20)	Santa Cruz	10	100

Total RNA-seq

RNA was extracted from cells using TRIzol Reagent (ThermoFisher Scientific, USA) according to manufacturer's instructions. Any contaminating DNA was digested with rigorous Turbo DNase (ThermoFisher Scientific, USA) treatment according to manufacturer's instruction. Purified RNA was quantified with Nanodrop 1000 instrument and quality was assessed using RNA Nano or Pico Assay kit with Bioanalyzer (Agilent Technologies, USA). Only the RNA samples with RIN above 8 were used for sequencing.

For strand-specific sequencing, ribosomal RNA was removed from total RNA with Ribo-Zero rRNA Removal Kit (EpiCenter, USA) according to manufacturer's instructions and depletion of rRNA was confirmed by analyzing the samples on RNA Pico Assay on Bioanalyzer (Agilent Technologies, USA). Libraries were prepared with ScriptSeq Total RNA Library prep kit (EpiCenter, USA) according to manufacturer's instructions. Purified libraries were then analyzed with HS DNA Assay Kit on Bioanalyzer (Agilent Technologies, USA) and sequenced on Illumina HiSeq2000 platform.

CapStarr-seq

The principle of CapStarr-seq was described previously (Dao et al., 2017; Vanhille et al., 2015). The detailed step-by-step protocol is accessible on Protocol Exchange (Dao et al., 2017). The reporter library was generated from genomic DNA extracted from mouse thymus DNA. For target enrichment, a home-designed oligonucleotide microarray containing 964688 probes, covering 13152 genomic regions of 400 bp, was constructed using the SureSelect technology (Agilent, 1M format) and the eArray tool default settings (https://earray.chem.agilent.com/earray/). The library contained 7759 enhancers (H3K4me1 + H3K27ac + PolII), 2758 enhancers (H3K4me1 + H3K27ac), 580 enhancers (H3K4me1), 1109 active promoters (H3K4me3 + H3K27ac + PolII), 500 random TSSs and 483 negative regions. The reporter library was transfected into P5424 cells using the Neon Transfection System (Thermo Fisher Scientific; pulse voltage 1,450 V, pulse width 10, pulse number 3). For each replicate, 30 × 10^6 cells were transfected with 150 μg of library; two independent transfection replicates were performed. The transfected and non-transfected (plasmid input) libraries were single-end sequenced on the Illumina NextSeq 500 platform, and reads were mapped to the mm9 reference genome using standard procedures. The coverage of each genomic region was calculated using BEDTools (v2.17.0), and the ratio of the CapStarr-seq coverage over the input (fold change) was computed for each sample. Regions with enhancer activity were defined by determining the inflexion point of the ranked fold change (Table S4). Enhancers were defined as regions displaying enhancer activity in both replicates.

4C-seq

Purified DN, DP, CD4 and CD8 cells were crosslinked after dissociation with 2% formaldehyde in DMEM for 10 min at 23°C. The fixation was quenched with cold glycine at a final concentration of 125 mM, then cells were washed with PBS and permeabilized on ice for 1 h with 10 mM Tris-HCl, pH 8, 100 mM NaCl, 0.1% NP-40 and protease inhibitors. Nuclei were resuspended in DpnII restriction buffer at 10 million nuclei/mL concentration, and 5 million nuclei aliquots were further permeabilized by treatment for 20 min with 0.7% SDS at 65°C, then for 40 min at 37°C. The SDS was then neutralized by incubating for a further 1h with 3.3% Triton X-100 at 37°C. Nuclei were digested overnight with 1000 U DpnII at 37°C, then washed twice by centrifuging and resuspending in T4 DNA ligase buffer. *In situ* ligation was performed in 400 μL T4 DNA ligase buffer with 20,000 U T4 DNA ligase overnight at 16°C. DNA was purified by reverse cross-linking with an overnight incubation at 65°C with proteinase K, followed by RNase A digestion, phenol/chloroform extraction and isopropanol precipitation. The DNA was digested with 5 U/μg Csp6I at 37°C overnight, then re-purified by phenol/chloroform extraction and isopropanol precipitation. The DNA was then circularized by ligation with 200 U/μg T4 DNA ligase under dilute conditions (5 ng/μL DNA), and purified by phenol/chloroform extraction and isopropanol precipitation. 50 ng aliquots of this DNA were used as template for PCR with bait-specific primers containing Illumina adaptor termini (primer sequences and optimal PCR conditions available on request). PCR reactions were pooled, primers removed by washing with 1.8x AMPure XP beads, then quantified on a Bioanalyzer (Agilent) before sequencing with a HiSeq 4000 (Illumina). 4C primer sequences are available on request. Each of the 32 4C experiments presented in this study was performed in biological replicate.

II- Bioinformatic Procedures

Experiments replicate correlations

All experiments, including ChIP-seq, RNA-seq, CapStarr-seq, presented in this article were performed in biological replicates and replicate correlation shown in Figure S5 was considered as a quality control for further processing the data. ChIP-seq correlation scores were calculated by comparing the mean ChIP-seq signal for each replicate for peaks called from merged ChIP-seq replicate data. Correlation scores for RNA-seq experiments were calculated by comparing the rpkm values for each gene between replicate experiments. CapStarr-seq replicate correlation score was calculated by comparing fold enrichment over input scores for all assayed regions between replicate experiments. Similarly 4C-seq was performed in replicate for each locus and docking point and their correlation is available in Table S4.

ChIP-seq Data Processing

For ChIP-seq, raw sequencing reads were aligned to mouse genome (mm9) using Bowtie2 (Langmead and Salzberg, 2012). Sequence reads that aligned multiple times in genome with equal alignment score, were discarded as well as the duplicate reads with identical coordinates (sequencing depth taken into account) were discarded to remove potential sequencing and alignment artifacts. Aligned reads were elongated *in silico* using the DNA fragment size inferred from paired-reads or an estimated optimal fragment size for orphan reads using an in-house developed R pipeline named PASHA (Fenouil et al., 2016). These elongated reads were then used to calculate the number of fragments that overlapped at a given nucleotide thus representing an enrichment score for each nucleotide in the genome. Wiggle files representing average enrichment score every 50bp or 10bp were generated. Sequencing data from Input samples were treated in the same way to generate Input wiggle files. All wiggle files were then rescaled to normalize the enrichment scores to reads per million. Enrichment scores from Input sample wiggle files were then subtracted from ChIP sample wiggle files. This allowed us to remove/reduce the over-representation of certain genomic regions due to biased sonication, local

duplications, and DNA sequencing. To reduce the noise in IGB snapshots, we performed smoothing of the signal by replacing each 10bp bin by the average of the 5 surrounding bins on each side. Besides this, input subtraction also improves the signal/noise ratio especially for ChIPs with low enrichment. Rescaled and Input subtracted wiggle files from biological replicate experiments were then used to generate a wiggle file that represents the average signal from several biological replicates. These wiggle files can be viewed on UCSC genome browser (http://genome-euro.ucsc.edu/cgi-bin/hgHubConnect?redirect=manual&source=genome.ucsc.edu).

The Track Hub can be accessed using this link (https://ndownloader.figshare.com/files/22633064?private_link=3df2f43ff40a9ae78875). The Y axes in all figures representing ChIP-seq examples or global profiling/boxplot refer to arbitrary units calculated out of the merged replicates data as scaled normalized signal.

RNA-seq Data Processing

Raw sequencing reads were aligned to mouse genome (mm9) using TopHat2 (Kim et al., 2013). Sequence reads that aligned multiple times in genome with equal alignment score, were discarded. Using strand-specific library prep of RNA samples, we could infer the strand from which the RNA was originally transcribed hence we separated the reads that align to Watson and Crick strands and processed them separately using PASHA (Fenouil et al., 2016) pipeline to generate strand-specific wiggle files. All wiggle files were then rescaled to normalize the enrichment scores to reads per million. Rescaled wiggle files from biological replicate experiments were then used to generate a wiggle file that represents the average strand-specific RNA signal from several biological replicates.

HTseq-count program from the HTSeq framework (Anders et al., 2015) was used to count the sequence reads mapping to gene annotations. Read counts were normalized to RPKM values. Gene expression values (RPKM) are provided in Table S1. Genes with values > 1 RPKM were considered to be expressed for Figure S1F. Genes with values > 0.5 RPKM were considered to be expressed for Figures 6 and S4 to keep genes with low expression.

Transcription Factor Binding Site Analysis

This analysis has been performed by HOMER algorithm (Heinz et al., 2010), with default parameters except for background sequences, where we selected inactive or non-repressed features of the same type, and motif size, with −100 to 100bp around the center.

Gene Ontology Analysis

Ontology analysis of potential target genes of active enhancers was performed with genomic regions enrichment analysis tool (GREAT v3.0.0, http://great.stanford.edu/public/html/) (McLean et al., 2010) using the default parameters (basal plus extension including curated regulatory domains).

Ontology analysis of genes with multiple promoters as well as control set of genes was performed with DAVID (https://david.ncifcrf.gov/) using mouse genome as background. Only the top non-redundant terms are shown.

Peak calling

We used wiggle files to detect the genomic regions with enrichment signals beyond background signal. For this purpose, we used *Thresholding* function of the Integrated Genome Browser (IGB) (https://bioviz.org/) to determine the Threshold above which we considered a genomic region to be enriched relative to background noise. We set the parameters with the minimum number of consecutively bins (Min Run) to be considered an enriched region and the maximum gap (Max Gap) between 2 regions to be merged (see Table titled "Parameters used for peak calling" for parameters used for each mark). *Min Run* and *Max Gap* were kept constant for each mark but the *Threshold* was adjusted after user inspection of several random regions to account for differences in Signal to Noise ratio between experiments. These parameters are then fed to an in-house script (wig peak caller) that performs automated peak-calling by using algorithm employed by Thresholding function of IGB. Genomic coordinates (BED files) of peaks called for each mark in DN, DP, nCD4, Th1, Th2 and Th17 cells are provided in Table S2 and in the UCSC Track Hub.

Parameters used for peak calling

Mark	Threshold	Min Run	Max Gap
H3K4me1 (DN, DP, nCD4)	80	150	500
H3K4me1 (Th1, Th2, Th17)	45	150	500
H3K4me3 (DN, DP, nCD4, Th2)	100	100	1400
H3K4me3 (Th1, Th17)	50	100	1400
H3K27ac (DN)	85	250	1400
H3K27ac (DP, nCD4, Th1, Th2, Th17)	65	250	1400
H3K27me3 (DN)	70	250	5000
H3K27me3 (DP)	150	250	5000
H3K27me3 (nCD4, Th1, Th2, Th17)	50	250	5000
H3K79me2 (All stages)	50	250	5000
Pol II (DN, DP, Th1)	70	150	3000
Pol II (nCD4, Th2)	50	150	3000
Pol II (Th17)	90	150	3000

Fraction of ChIP-seq signal that falls within the called peaks for each mark in each cell stage is shown in table below.

Percent of signal within called ChIP-seq peaks

	DN	DP	nCD4	Th1	Th2	Th17
H3K4me1	1.3	1.6	2	2.8	2.6	2.8
H3K4me3	0.8	0.6	0.6	0.7	0.6	0.8
H3K27ac	1.3	1.1	1.4	1.4	1.1	1.3
H3K27me3	1	0.8	1.6	0.8	1.1	1.6
H3K79me2	2.1	2.3	3	2.8	2.5	2.9
Pol II	1	0.5	0.9	1.1	1.7	1.4

Identification of Active Enhancers

Genomic regions that show simultaneous enrichment with H3K4me1, H3K27ac and Pol II and are at least ± 2000bp away from any annotated promoter were considered to be putative active enhancers. To remove any unannotated promoters from our enhancer selection, we filtered out any regions that were more enriched with H3K4me3 as compared to H3K4me1. Method used for this filter has been described previously (Descostes et al., 2014). In identified enhancers, position of the minimum signal of H3K27ac (nucleosome depleted region – NDR) that was closest to location of maximum signal of Pol II was defined as center of the region. The genomic coordinates of all identified enhancers at each stage (centered on NDR and resized to 500bp) are provided in Table S3.

Identification of active promoters

Annotated TSS regions that show simultaneous enrichment with H3K4me3, H3K27ac and Pol II were considered to be active promoters (related to Figures S1F, 2D, S2B, and S2D).

Analysis of dynamic enhancers/promoters

BEDTools (v2.17.0) (Quinlan and Hall, 2010) was used to identify the enhancers/promoters that were overlapping between two cell types. Any enhancers/promoters identified in two separate cell stages but overlapped each other were considered to be active in both cell types. For active promoters, only the genes active (defined above) at a given stage that are putatively controlled (each enhancer is assigned to the closest gene) by the enhancers active (defined above) at the same stage, were considered.

Selection of groups with one or multiple promoters (Figure 4, 6, S3, and S4)

Analysis to compare genes with one or multiple TSS were performed using gene sets described in Figure S3A for Figures 4 and S3 panels. In Figures 6 and S4, an additional expression filter was added (rpkm > 0.5 at all stages) for both groups and the randomized sets for the -one promoter group- shown are representative of the several groups. The numbers of genes at each differentiation stage are DN 597, DP 500, nCD4 515, Th1 561, Th2 539, Th17 548.

ChIP-seq Average Metagene Profiles

ChIP-seq values from wiggle files were retrieved with in-house R scripts for selected genes and enhancer regions. Then we used an algorithm as described previously (Koch et al., 2011) to rescale the genes to same length by interpolating the values on 1000 points and build a matrix on which each column is averaged and resulting values are used to plot average metagene profiles. To generate average signal profiles, we selected the mm9 genes or identified enhancer regions that do not have any other annotation around boundaries within the profiled distance (Figures 2 and S2). Removal of the annotations too close to each other is necessary to avoid mixing signals from close-by annotations that can cause misinterpretation of the results. Profiles around TSS shown in Figures 6 and S4 were performed on the most distal promoters for genes with multiple TSS.

4C-seq data processing and call of long-distance interactions

All bait sequence (including and downstream of the primer sequence, up to but not including the GATC *Dpn*II site) are trimmed by the demultiplexing Sabre tool (https://github.com/najoshi/sabre), allowing two mismatches, before mapping to the mm9 genome with Bowtie (Langmead et al., 2009) and processing with the utility functions of the 4See browser (Ben Zouari et al., 2020). 4C profiles are given as the running mean values over sliding windows of 21 fragments. Table below shows the Spearman correlation scores for replicate 4C-seq experiments calculated with data within 1 Mb of the bait.

Spearman correlation scores of replicate 4C-seq experiments

Dataset	Spearman correlation	Dataset	Spearman correlation
Runx3-E2DP	0.70	Runx3-E2DN3	0.83
Runx3-E3DP	0.71	DN3_Runx3P1	0.83
DP_Runx3P1	0.73	DN3_Runx3P2	0.84
Runx3-E1DP	0.74	CD4_Runx3P1	0.84

(*Continued on next page*)

Continued

Spearman correlation scores of replicate 4C-seq experiments

Dataset	Spearman correlation	Dataset	Spearman correlation
Runx3-E1CD8	0.75	DP_S1pr1	0.84
CD8_Runx3P2	0.76	CD4_Runx3P2	0.85
DP_Runx3P2	0.77	Runx3-E2CD4	0.86
Runx3-E3DN3	0.77	DN3_mir181	0.87
Runx3-E3CD8	0.78	DP_mir181	0.87
Runx3-E3CD4	0.78	CD4_Nfatc1	0.88
Runx3-E1CD4	0.79	CD4_mir181	0.89
DN3_S1pr1	0.79	DP_Lef1	0.91
Runx3-E1DN3	0.80	CD4_S1pr1	0.92
CD8_Runx3P1	0.80	DN3_Lef1	0.93
Runx3-E2CD8	0.82	CD4_Lef1	0.94
DP_Nfatc1	0.83		

QUANTIFICATION AND STATISTICAL ANALYSIS

All ChIP-seq, RNA-seq, CapSTARR-seq and 4C-seq experiments were performed in at least two biological replicates. p values and nature of statistical tests associated to the number of asterisks in figures are described in figure legends.