The EMBO Journal (2011) 30, 4198–4210 | © 2011 European Molecular Biology Organization | All Rights Reserved 0261-4189/11
www.embojournal.org

THE EMBO JOURNAL

H3K4 tri-methylation provides an epigenetic signature of active enhancers

Aleksandra Pekowska[1,2,3,4,6,7], Touati Benoukraf[1,2,3,4,6,8], Joaquin Zacarias-Cabeza[1,2,3,4,6], Mohamed Belhocine[1,2,3,4], Frederic Koch[1,2,3,4], Hélène Holota[4,5], Jean Imbert[4,5], Jean-Christophe Andrau[1,2,3,4], Pierre Ferrier[1,2,3,4,*] and Salvatore Spicuglia[1,2,3,4,*]

[1]Centre d'Immunologie de Marseille-Luminy (CIML), Parc Scientifique de Luminy, Case 906, Marseille, France, [2]CNRS, UMR6102, Marseille, France, [3]Inserm, U631, Marseille, France, [4]Université de la Méditerranée, Marseille, France and [5]Inserm, U928, TAGC, Marseille, France

Combinations of post-translational histone modifications shape the chromatin landscape during cell development in eukaryotes. However, little is known about the modifications exactly delineating functionally engaged regulatory elements. For example, although histone H3 lysine 4 mono-methylation (H3K4me1) indicates the presence of transcriptional gene enhancers, it does not provide clear-cut information about their actual position and stage-specific activity. Histone marks were, therefore, studied here at genomic loci differentially expressed in early stages of T-lymphocyte development. The concomitant presence of the three H3K4 methylation states (H3K4me1/2/3) was found to clearly reflect the activity of *bona fide* T-cell gene enhancers. Globally, gain or loss of H3K4me2/3 at distal genomic regions correlated with, respectively, the induction or the repression of associated genes during T-cell development. In the *Tcrb* gene enhancer, the H3K4me3-to-H3K4me1 ratio decreases with the enhancer's strength. Lastly, enhancer association of RNA-polymerase II (Pol II) correlated with the presence of H3K4me3 and Pol II accumulation resulted in local increase of H3K4me3. Our results suggest the existence of functional links between Pol II occupancy, H3K4me3 enrichment and enhancer activity.

The EMBO Journal (2011) **30,** 4198–4210. doi:10.1038/emboj.2011.295; Published online 16 August 2011
Subject Categories: chromatin & transcription
Keywords: ChIP-Seq; enhancer; epigenetics; RNA-polymerase II; T lymphocyte

*Corresponding author. P Ferrier or S Spicuglia, Centre d'Immunologie de Marseille-Luminy (CIML), Parc Scientifique de Luminy, Case 906, Marseille 13009, France. Tel.: +33 49 126 9435; Fax: +33 49 126 9430; E-mail: ferrier@ciml.univ-mrs.fr or Tel.: +33 49 126 9443; Fax: +33 49 126 9430; E-mail: spicuglia@ciml.univ-mrs.fr
[6]These authors contributed equally to this work
[7]Present address: European Molecular Biology Laboratory, Heidelberg, Germany
[8]Present address: Cancer Science Institute of Singapore, National University of Singapore, Singapore

Received: 31 December 2010; accepted: 12 July 2011; published online: 16 August 2011

Introduction

Developmental cell fate events in eukaryotes are accompanied by epigenetic changes that remodel the chromatin landscape (Natoli, 2010). Notably, the wide range of post-translational covalent modifications of nucleosomal histone tails including acetylation, methylation, phosphorylation, ubiquitination and ADP ribosylation, which implies a large diversity of combinatorial patterns at genetic loci, may convey distinctive regulatory information and confer functional properties on specific genomic sites (Jenuwein and Allis, 2001; Li *et al*, 2007). For example, di-methylation of H3 lysine 9 (H3K9me2) and tri-methylation of H3 lysine 27 (H3K27me3) are mainly associated with heterochromatin and gene silencing. On the other hand, the mono-, di- and tri-methylation of histone H3 lysine 4 (H3K4me1, H3K4me2 and H3K4me3, respectively) are generally associated with euchromatin and ongoing gene expression, while another euchromatic mark, tri-methylated histone H3 lysine 36 (H3K36me3), is characteristic of transcriptional elongation.

Recently, genome-wide studies shed brighter light on the distribution of epigenetic marks and the existence of striking correlations between the local structure of the chromatin and the presence and, possibly, the functional state, of transcriptional *cis*-regulatory elements (Bernstein *et al*, 2006; Barski *et al*, 2007; Guenther *et al*, 2007; Mikkelsen *et al*, 2007; Mendenhall and Bernstein, 2008; Wang *et al*, 2008). In particular, active promoters are marked by H3K4me3, repressed promoters by H3K27me3 and poised promoters by both marks (Bernstein *et al*, 2006). These studies have also led to defining chromatin signatures characteristic of discrete types of *cis*-regulatory regions (Hon *et al*, 2009). In this respect, H3K4me3 was mostly found to be associated with promoter regions, while H3K4me1 was found to be a hallmark of promoter-distal *cis*-regulatory elements (i.e., enhancers) (Heintzman *et al*, 2007, 2009). Genome-wide detection of cell-specific H3K4me1 peaks should, therefore, make it possible to identify specific enhancer elements. However, since H3K4me1-enriched regions are generally larger than the associated *cis*-regulatory elements (Barski *et al*, 2007), it is difficult to define the exact location of the actual enhancer site(s) they contain. In addition, the presence of H3K4me1 *per se* does not strictly correlate with the functional activity of these elements (Cui *et al*, 2009; Creyghton *et al*, 2010). A more accurate definition of the chromatin signature of active enhancers is, therefore, required (Bulger and Groudine, 2011).

High-throughput studies have also demonstrated that promoter-distal regulatory regions are generally associated with RNA-polymerase II (Pol II) (Koch *et al*, 2008; De Santa *et al*, 2010; Kim *et al*, 2010). Pol II recruitment at promoter regions induces H3K4 tri-methylation (Li *et al*, 2007; Selth *et al*, 2010). Taken together, these findings suggest the possibility that H3K4me3 may also be present in enhancer regions as a consequence of local Pol II occupancy.

Although H3K4me3 enrichment has been previously reported to occur in promoter-distal genomic regions (Barski *et al*, 2007; Wang *et al*, 2008), its presence has not been linked so far to specific enhancer activity and/or function.

With a view to defining more closely the chromatin signature of active enhancers, we investigated the combinatorial dynamics of H3K4 methylation status in enhancer elements during the process of T-lymphoid cell differentiation in the adult mouse thymus. T-cell differentiation was induced by triggering pre-TCR signalling *in vivo*, which leads early thymocytes to cross the so-called β-selection checkpoint and promotes their development along the αβT-cell lineage (Hayday and Pennington, 2007). Upon studying the resulting effects on histone marks in well-defined enhancers, enhancer activity was found to be generally associated with the presence of both H3K4me2 and H3K4me3, while H3K4me1 was present regardless of their functional state. Using an engineered *Tcrb* gene enhancer mutant mouse, we further demonstrated that the strength of this enhancer directly influences the local H3K4me3/H3K4me ratio. H3K4me3 enrichment was also found to correlate with Pol II occupancy in enhancer elements. All in all, these data provide evidence that there exists a direct link between H3K4me3 chromatin mark and enhancer activity *in vivo*.

Results

Experimental design

T-lymphocyte differentiation in the thymus involves the rearrangement of T-cell receptor (*Tcr*) genes by V(D)J recombination and a series of selection events, whereby newly assembled TCR complexes signal for cell survival, proliferation and differentiation processes (Hayday and Pennington, 2007). Following productive (in frame) *Vb-DJb* gene recombination and the expression of TCRβ chains in late $CD4^{-}CD8^{-}$ double-negative (DN) thymocytes, pre-TCR complexes are formed, each consisting of the invariant pTα chain and signal-transducing CD3 proteins, along with the newly formed TCRβ chain (von Boehmer, 2005). Pre-TCR-induced signalling leads DN thymocytes to cross the so-called β-selection checkpoint, which results in massive cell proliferation and the induction of a developmental process marked by the expression of both CD4 and CD8 co-receptors, thus generating $CD4^{+}CD8^{+}$ double-positive (DP) thymocytes and the onset of *Tcra* gene rearrangement (Hayday and Pennington, 2007). We focused here on the β-selection process because of the fact that several well-known distal *cis*-regulatory elements respond to pre-TCR signalling, which makes this model system particularly suitable for investigating the interplay between chromatin dynamics and enhancer activity *in vivo* (Anderson, 2006; Krangel, 2007).

To determine the epigenetic features associated with pre-TCR signalling and β-selection processes, thymocytes purified from recombination activating gene 2-deficient mice ($Rag2^{-/-}$; hereafter ΔRag) and from ΔRag mice intraperitoneally injected with an anti-CD3ε antibody (hereafter denoted ΔRagCD3) were used. These mouse models provide convenient sources of homogeneous thymic cell populations arrested in the G1 phase of the cell cycle in the DN and DP cell compartments, respectively (Chattopadhyay *et al*, 1998; Senoo and Shinkai, 1998 and data not shown). In this context, the systemic stimulation of ΔRag thymocytes by the injected anti-CD3ε antibodies mimics pre-TCR signalling, thus promoting the transition of the cells from the DN to the DP stage. We reasoned that, by comparing histone methylation states in distal regulatory regions associated with differentially expressed genes in ΔRag versus ΔRagCD3 thymocytes, it should be possible to define a chromatin signature of enhancer activity.

Epigenetic profiles in thymocytes before and after β-selection

To assess epigenetic changes associated with β-selection, we purified ΔRag and ΔRagCD3 thymocytes and performed, in the first place, chromatin immunoprecipitation (ChIP) experiments on formaldehyde crosslinked chromatin using antibodies against six distinct histone H3 methylation marks (H3K4me1, H3K4me2, H3K4me3, H3K36me3, H3K27me3 and H3K9me2). ChIP samples were subsequently hybridized to a custom-made, high-density tiling array (see Materials and methods for details). To extent our study to the genome-wide level and exclude any potential bias due to crosslinked chromatin, we also performed ChIP-Seq experiments on native Mnase-treated chromatin for a subset of histone modifications (H3K4me1 and H3K4me3). In parallel, transcriptome DNA microarray analyses were performed to quantify gene expression levels on the same cell samples (Supplementary Table S1). To assess the quality of the ChIP-on-chip and ChIP-Seq data, histone modification profiles were first inspected at individual loci. As expected, the expressed *Cd3* loci were similarly enriched for H3K4me1/2/3 and H3K36me3 in both ΔRag and ΔRagCD3 thymocytes (Figure 1A; Supplementary Figure S1A), whereas the silent *Ebf1* locus consistently harboured the H3K9me2 and H3K27me3 marks (Supplementary Figure S1C). We also noted that T cell-specific genes (such as *Cd3* genes) were highly enriched for H3K4 methylation with a broad distribution within the gene body, in agreement with our previous findings (Pekowska *et al*, 2010).

H3K4 tri-methylation in distal genomic regions is highly dynamic

To obtain an overall picture of the dynamics of histone H3 lysine methylation (H3Kme) signatures occurring in distal genomic regions, statistically enriched regions corresponding to each of the analysed histone marks (hereafter called 'peaks'; see Materials and methods) between ΔRag and ΔRagCD3 thymocytes were selected and their overall patterns of distribution were assessed as a function of cell-stage specificity and genomic location (Figure 1B; Supplementary Figure S1B). In the case of H3K4me1, H3K27me3 and H3K9me2, stage-specific peaks were found to be similarly distributed between promoter-proximal and promoter-distal regions. Strikingly however, stage-specific H3K4me2 and H3K4me3 peaks were more frequently detected in the distal regions. These findings are particularly noteworthy in the case of H3K4me3, as this mark has generally been reported to be associated with promoter regions (Mellor *et al*, 2008). Therefore, overall analysis of the changes in the pattern of H3Kme distribution observed upon pre-TCR signalling suggested that H3K4me3 may contribute to defining the activity of distal regulatory elements. To obtain further insights into the latter possibility,

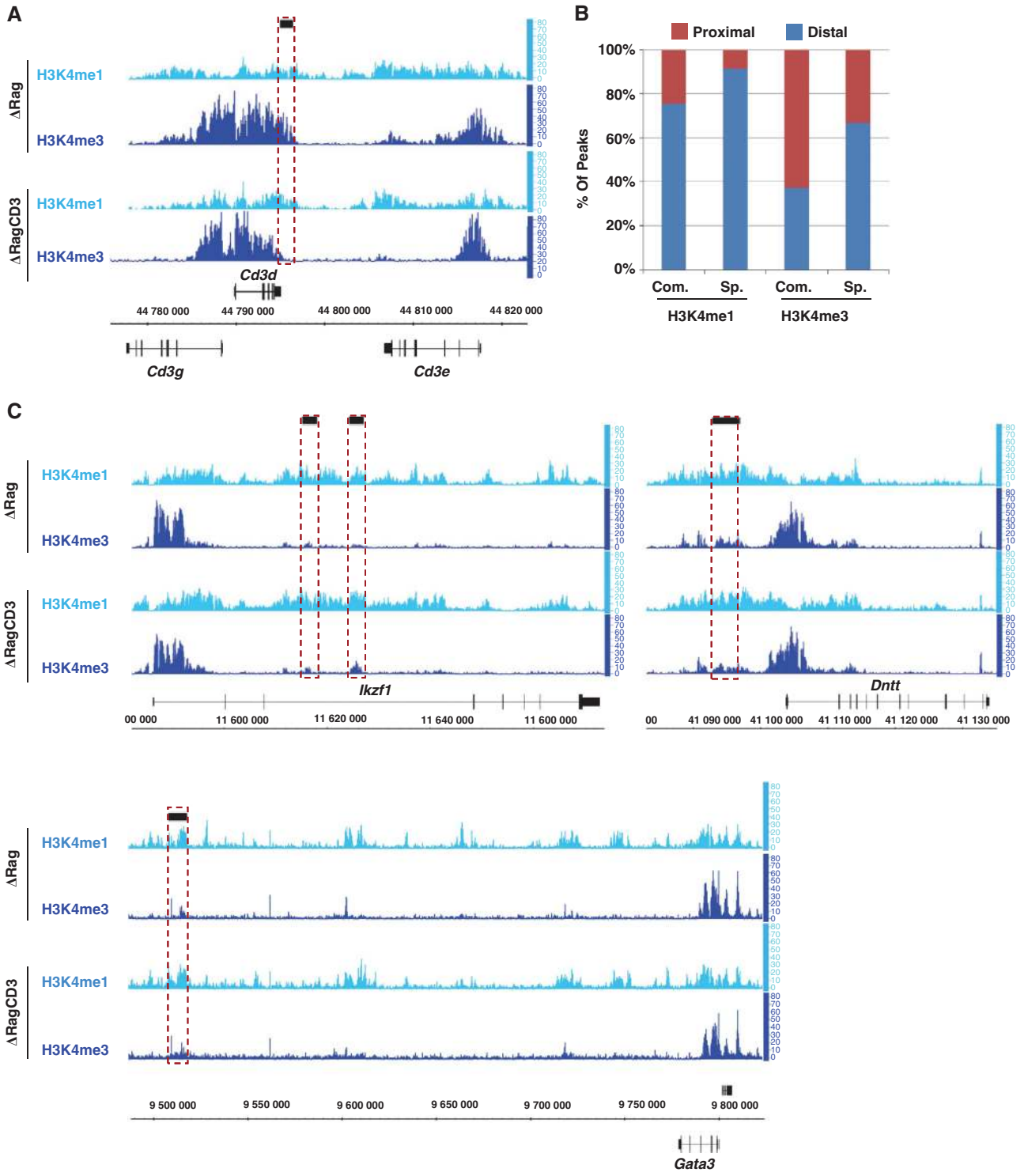

Figure 1 Comparative epigenetic profiles in ΔRag and ΔRagCD3 thymocytes. (**A**) Epigenetic profiles at the *Cd3* loci. Each track gives the results of one ChIP-Seq experiment on the histone modification specified in the ΔRag or ΔRagCD3 cell type. The genomic location and transcriptional orientation of the corresponding genes is indicated at the bottom of the panel. The positions of enhancer regions are indicated by a black rectangle and the associated epigenetic profiles are highlighted by a dotted rectangle. (**B**) Genome-wide distribution and stage specificities of enriched regions (peaks) for each epigenetic mark. Bar plots give the relative percentages of promoter-proximal and promoter-distal peaks found either in ΔRag or ΔRagCD3 thymocytes (stage specific; Sp.) or in both cell populations (common peaks; Com.). (**C**) Epigenetic profiles at three additional lymphoid-specific loci for which enhancer elements are known, including *Ikzf1* (top-left panel), *Dntt* (top-right panel) and *Gata3* (bottom panel).

we next examined the H3K4 methylation profiles more closely to distal regulatory regions, with special emphasis on enhancer elements known to play an active role in thymocyte development.

Active enhancers are generally marked by all three H3K4 methylation states

Previous large-scale studies have suggested that transcriptional enhancers may be associated with high level of

H3K4me1 (Heintzman *et al*, 2007, 2009). However, the presence of H3K4me1 *per se* does not necessarily signal the occurrence of stage-specific enhancer activity since inactive developmental genes are also associated with H3K4me1-enriched distal enhancers (Cui *et al*, 2009; Creyghton *et al*, 2010). The finding that dynamic changes of H3K4me2 and H3K4me3 peaks occur preferentially in distal genomic regions suggested the possibility that these marks might also be associated with enhancer activity. To determine whether H3K4me2 and H3K4me3 do indeed mark distal *cis*-regulatory elements, we first examined the epigenetic profiles in four separate enhancer elements known to be active in DN and DP thymocytes. These included (i) a specific T-cell enhancer located at the 3′ end of the *Cd3d* gene within the mouse *Cd3* locus (Georgopoulos *et al*, 1988); (ii) three 5′ distal DNase hypersensitive sites (DHS1-3) that synergistically activate the expression of the *Dntt* gene (encoding Terminal Deoxynucleotidyl Transferase or TdT) in T cells (Cherrier *et al*, 2008); (iii) two DHS sites (C6 and C7) located in the intronic region of the *Ikzf1* gene (coding for the transcription factor IKAROS), which also display T cell-specific enhancer activity (Kaufmann *et al*, 2003); and (iv) a lymphoid-specific enhancer located 280 kb downstream of the *Gata3* gene (Hosoya-Ohmura *et al*, 2011). As shown in Figures 1A and C, all these enhancer regions were enriched for H3K4me1 and H3K4me3 in ChIP-Seq experiments from ΔRag and ΔRagCD3 thymocytes (consistent results were also observed for H3K4me1/2/3 using the ChIP-on-chip approach, Supplementary Figure S1). As expected in the case of active regulatory elements, little or no enrichment of the repressive marks H3K27me3 and H3K9me2 was observed (Supplementary Figure S1). Authentic enhancers that are functional (or activated) in developing thymocytes are, therefore, marked by a chromatin signature consisting not only of H3K4me1, but also of H3K4me2 and H3K4me3 histone modifications.

Enhancers induced by pre-TCR signalling acquire H3K4 tri-methylation

Several loci induced by pre-TCR signalling having known distal regulatory elements were examined. We focused in particular on the transcriptional induction of the *Cd8a/b1* and *Cd4* genes, as well as on the initiation of V-to-J recombination at the *Tcra* locus, which are all crucial developmental events triggered by pre-TCR signalling and the DN to DP cell transition. The enhancer elements associated with these loci have been extensively characterized and provide appropriate models for studying the epigenetic phenomena involved in the activation of distal *cis*-regulatory elements during T-cell development (Krangel, 2007). Activating marks were found to accumulate in ΔRagCD3 thymocytes, spanning the gene bodies and surrounding intergenic regions at both the *Cd8* and *Cd4* loci (Figure 2). The *Cd8* locus harbours several enhancers (named E8I-to-E8V), which are located either within the body of the *Cd8b1* gene or in the intergenic region between the *Cd8b1* and the *Cd8a* genes, all of which are required for the proper expression of *Cd8* genes in DP thymocytes (Taniuchi *et al*, 2004 and references therein). Strikingly, all *Cd8* enhancers (except E8V) were enriched with H3K4me1 in ΔRag thymocytes, whereas they had acquired H3K4me3 mark in the case of ΔRagCD3 thymocytes (Figure 2A). In addition, the E8V enhancer appeared enriched for both H3K4me1/3 specifically in ΔRagCD3 thymocytes. The expression of the *Cd4* gene is controlled by a proximal enhancer (Ep) supporting gene expression in all subsets of thymocytes (Adlam and Siu, 2003; Taniuchi *et al*, 2004 and references therein). However, in DN thymocytes, *Cd4* expression is repressed by an intronic silencer element (Jiang *et al*, 2005). Consistent with this picture, H3K4me1/3 enrichment was observed along the Ep enhancer in ΔRag thymocytes, while the level of H3K4me3 increased significantly in ΔRagCD3 thymocytes (Figure 2B). In parallel, *Tcra V-to-J* rearrangements are known to be initiated in DP thymocytes by the activation of a unique enhancer (referred to as Eα) located at the 3′ end of the *Tcra* locus, although a pre-assembled nucleoprotein complex can readily be detected in ΔRag thymocytes (Capone *et al*, 1993; Lauzurica and Krangel, 1994; Sleckman *et al*, 1997; Hernandez-Munain *et al*, 1999; Spicuglia *et al*, 2000). Consistent with all these features, Eα was found to be associated with both H3K4me1/3 marks from the DN stage onwards (i.e., in ΔRag thymocytes; Figure 2C). However, it showed a stronger and more widespread H3K4me2/3 enrichment profile at the DP stage (i.e., in ΔRagCD3 thymocytes; Figures 2C). ChIP-on-chip data provided consistent results and extended our observations to the fact that H3K4me2 was also associated with active enhancers (Supplementary Figure S2; also note that low level of repressive marks, either H3K27me3 or H3K9me2, were observed in ΔRag thymocytes, but not in ΔRagCD3 thymocytes). These findings were further validated by independent ChIP–qPCR analyses (Figure 2D; Supplementary Figure S3). In short, the developmental regulated enhancers studied here appeared to be generally marked (i.e., primed) by H3K4me1 in ΔRag thymocytes; and displayed enhanced H3K4me2/3 levels when triggered by pre-TCR signalling.

Enhancers induced during B-cell development also acquire H3K4me2/3 modifications

To establish whether a positive correlation between H3K4me2/me3 enrichment and the functional activation of developmentally regulated enhancers might also exist in another lymphoid cell lineage, we studied an independent set of H3K4me1/me2/me3 profiles obtained at two subsequent stages of early B-cell development, named pre-pro-B and pro-B cells (Lin *et al*, 2010). The changes in the levels of these marks occurring in well-known, developmentally activated distal *cis*-regulatory elements associated with the *Igll1/Vpreb1, Il28ra*, and *Bcl7a* loci were therefore examined. Remarkably, at these three loci, we likewise observed a gain in H3K4me2/3 enrichment in the distal regulatory elements specifically in pro-B cells (Supplementary Figure S4). By contrast, the *Cd3d* associated enhancer, which is inactive in B cells, was enriched with H3K4me1 and me2, but not with H3K4me3. The positive correlation observed between H3K4me2/me3 enrichment and the functional activation of enhancers in developing B lymphocytes means that this basic principle also applies to distinct lymphoid cell lineages.

Reduced activity of the Tcrb gene enhancer results in the loss of H3K4me3 and increase in H3K4me1

To further explore the links between H3K4 methylation and developmentally regulated enhancer activity, the relationships between H3K4me3 and enhancer strength were studied *in vivo*. In the mouse model used for this purpose, the

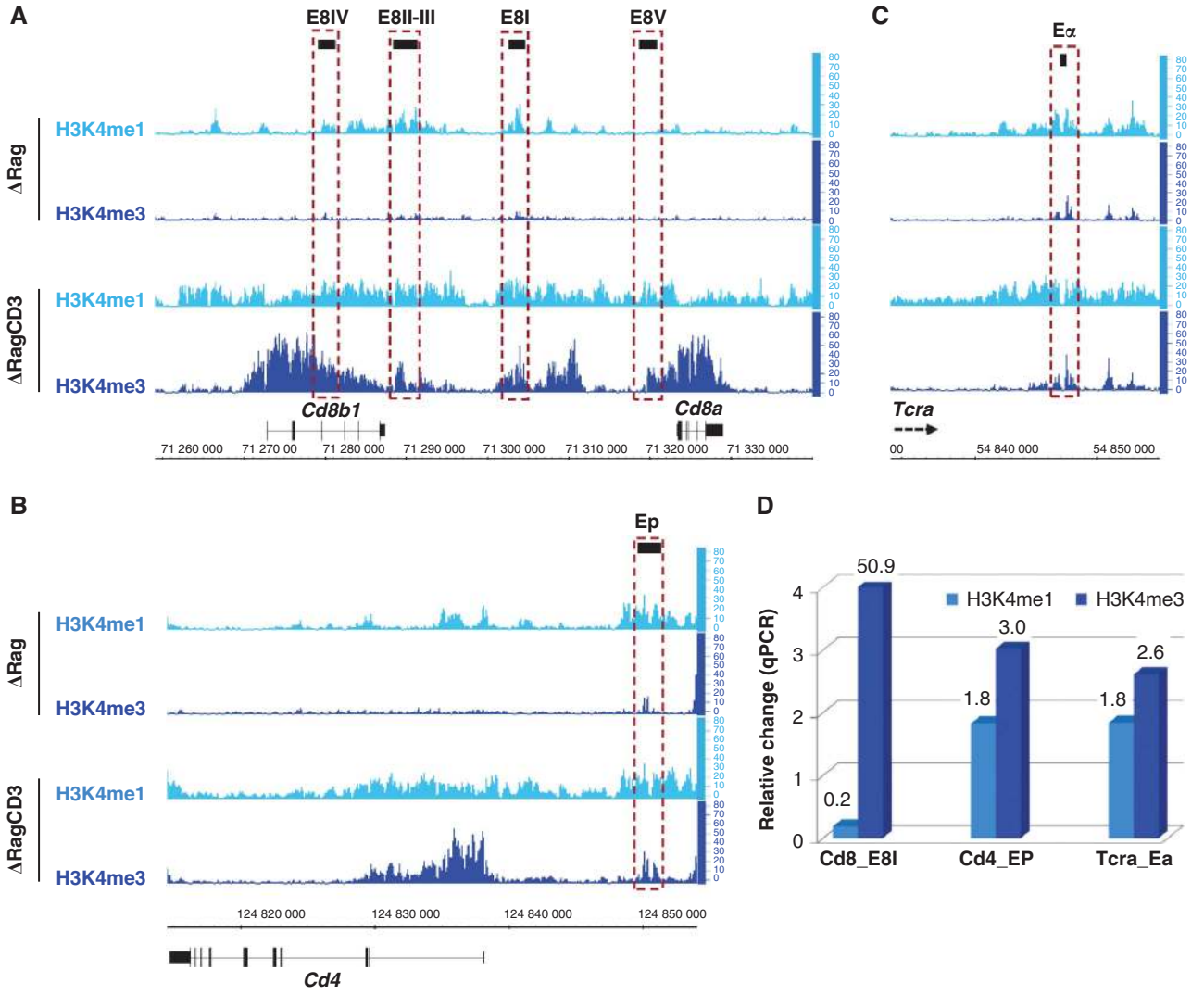

Figure 2 Epigenetic profiles of *bona fide* distal regulatory elements known to be activated *via* pre-TCR signalling. (**A**–**C**) Epigenetic profiles at the *Cd8* (**A**) and *Cd4* (**B**) loci and around the *Tcra*-associated enhancer Eα (**C**). In the later case, only the 3′ end of the *Tcra* locus is shown. The names of the associated enhancers are indicated. Legends are as in Figure 1. (**D**) Relative gain of H3K4me1 and H3K4me3 at indicated enhancer regions. The average enrichment was quantified by qPCR for each H3K4 methylation mark in ΔRag and ΔRagCD3 thymocytes (Supplementary Figure S3) and the differences between the two cell populations were plotted as a ratio of the signal obtained in ΔRag thymocytes. The relative changes are indicated at the top of each bar.

endogenous *Tcrb* gene enhancer (Eβ) has been replaced by a conserved core region (Eβ^{169}; Figure 3A). Eβ^{169}-mutated alleles display impaired enhancer functions, namely resulting in marked reduction in associated *Db-Jb* germ line transcription and recombination during early T-cell development (Bonnet *et al*, 2009). Quantitative (q)PCR assessment of ChIP samples indicated that the Eβ region is associated with significant levels of both H3K4me1 and H3K4me3 in ΔRag thymocytes, but not in ES cells (Figure 3B). Strikingly, however, replacing Eβ by Eβ^{169} (ΔRag;Eβ^{169} thymocytes) increased the levels of H3K4me1 and decreased those of the H3K4me3 histone marks.

To analyse in an extended manner the chromatin alterations induced by the loss of Eβ activity, the ChIP samples were hybridized in parallel to a dedicated high resolution array covering the *Tcrb* locus (see Materials and methods for details; note that, to prevent any bias due to genetic differences between the wild-type and mutated Eβ sequences, the signals originating from the probes overlapping the core Eβ region were removed). Comparative analyses of the ±2-kb Eβ-surrounding regions demonstrated that, in comparison with ΔRag, ΔRag;Eβ^{169} thymocytes showed a marked increase in the H3K4me1 levels, namely in the region immediately downstream of Eβ, along with a general decrease in the H3K4me3 levels (Figure 3C; main differences between the two samples are highlighted in yellow). Reducing the functional activity of the *Tcrb* gene enhancer, therefore, results in the specific loss of the H3K4me3 histone modification surrounding the Eβ area, implying that this enhancer's function is involved in local H3K4me3 enrichment.

The presence of H3K4me3 correlates with the activity of intergenic H3K4me1 domains

As described above, H3K4me1 alone marks the presence of distal enhancers, but co-enrichment with H3K4me3 give a more accurate picture of the stage-specific activity of these

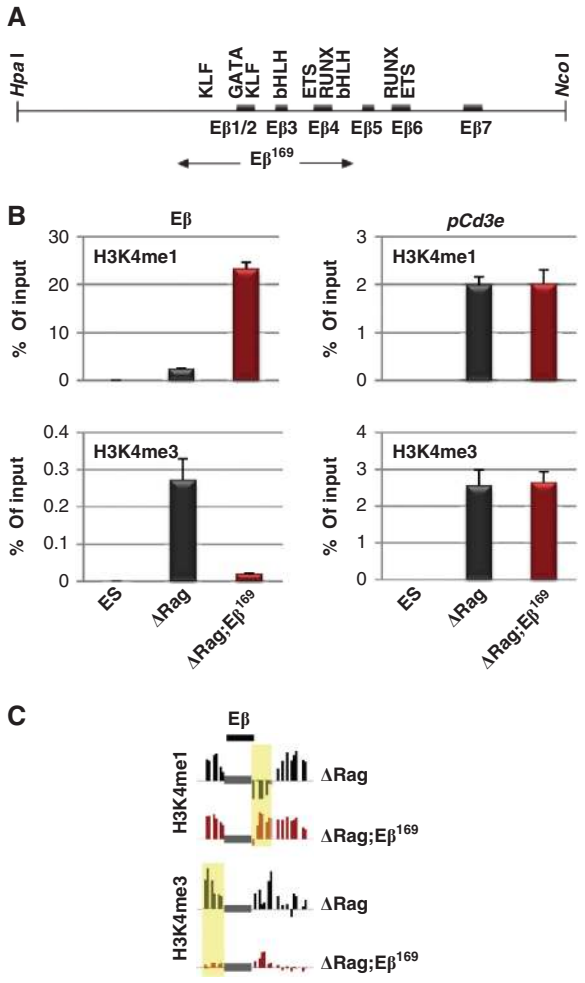

Figure 3 Loss of H3K4me3 in the Eβ^{169} knock-in mouse model. (**A**) Schematic diagram of the wild-type *Tcrb* gene enhancer (Eβ) and associated TF binding sites. The Eβ^{169} region is indicated at the bottom. (**B**) ES cell line and thymocytes from ΔRag and ΔRag mice crossed onto the Eβ$^{169/169}$ (ΔRag;Eβ^{169}) genetic background (Bonnet *et al*, 2009) were analysed in replicate by ChIP and qPCR in order to determine the level of H3K4me1 and H3K4me3 enrichment occurring within the Eβ sequences. Values were normalized by the enrichment level recorded in the *Actb* control promoter and expressed as percentages of the total input. The promoter of the *Cd3e* gene is shown as a positive control. No significant values were obtained with isotypic IgG controls. (**C**) ChIP samples were hybridized to a custom 15k array covering the *Tcrb* locus. ChIP-on-chip signals were normalized by the enrichment levels obtained in the *Actb* promoter. The Eβ-surrounding region is shown. To avoid confusion and prevent any bias due to differences between the wild-type and mutated Eβ sequences, the signals originating from the probes overlapping the core Eβ region (grey rectangles) were removed.

cis-regulatory elements. This new combinatorial signature might lead to the identification, on a large scale, of novel cell stage-specific distal regulatory elements. To check this point, the intergenic regions in which H3K4me1 peaks were detected in both ΔRag and ΔRagCD3 thymocytes were selected (these regions will henceforth be referred to as H3K4me1 domains; for details of the selection procedure, see Materials and methods). We then calculated the proportions of the H3K4me1 domains associated with H3K4me3 peaks (Figure 4A). We found that 43% of H3K4me1 domains were associated with H3K4me3, from which 13% were specifically associated with a single cell type (ΔRag or ΔRagCD3; i.e., 'dynamic domains'; Figure 4A). Importantly, intergenic H3K4me1/me3 overlapping regions were associated with higher levels of H3K4me1 as compared with the promoter elements (Supplementary Figure S5), supporting the fact that these regions correspond to putative distal regulatory elements rather than promoters of unannotated genes. To correlate stage-specific H3K4 tri-methylation events with transcriptional regulation, we selected the most differentially expressed genes between ΔRag and ΔRagCD3 thymocytes (see Materials and methods). The proportions of the dynamic H3K4me1 domains associated with either induced or repressed genes were then determined. Strikingly, H3K4me1 domains that lost H3K4me3 were found preferentially located in the vicinity of repressed genes (Figure 4B). Conversely, H3K4me1 domains that gained H3K4me3 histone mark were mainly located near genes induced by pre-TCR signalling (Figure 4B). Several examples of H3K4me1 domains showing correlated behaviour of this kind (H3K4me3 gain or loss along with activation or repression of a neighbouring gene) are shown in Figures 4C and D. They all were further validated by ChIP–qPCR (Supplementary Figure S6). We performed a similar analysis using ChIP-on-chip data taking into account the three H3K4 methylation statuses and obtained consistent results (Supplementary Figure S7). In addition, this later analysis showed that most of the changes in H3K4 methylation consisted simply of the loss or gain of a single methyl group, while no significant associations were observed with either H3K27me3 or H3K9me2 peaks (Supplementary Figure S8A). In agreement with our findings, H3K4me1 domains showing stage-specific H3K4 di- and tri-methylation were also found to be enriched with DNA motifs recognized by TFs expressed at the corresponding stage (Supplementary Figures S8B–D). Finally, we tested whether intergenic regions specifically enriched in H3K4me1/me3 overlapping peaks may be endowed with a transcriptional enhancing function. Indeed, a majority (8 out of 10) of intergenic regions harbouring H3K4me1/3 peaks demonstrated significant enhancer activity in a luciferase reporter assay opposite to those associated with H3K4me1 alone (Supplementary Figure S9). All in all, these findings strongly suggest that the functional activity of H3K4me1 domains can be assessed from the presence of additional methylation moieties, and further support the idea that distal *cis*-regulatory elements are commonly primed before the induction of gene expression.

H3K4 tri-methylation is associated with Pol II deposition in distal regulatory regions

What might the molecular links between distal *cis*-regulatory activity and H3K4me3 enrichment consist of? An obvious player in connecting these processes could be the Pol II transcriptional apparatus. Pol II transcriptional activity is regulated *via* phosphorylation of the carboxy-terminal domain (CTD; Koch *et al*, 2008). In yeast, Pol II phosphorylated at serine 5 (ser5P) of the CTD recruits the COMPASS complex to the 5′ end of genes, which in turn results in H3K4 tri-methylation at promoters (Selth *et al*, 2010). Moreover, several studies have recently reported that Pol II also binds to distal genomic regions genome-wide (De Santa *et al*, 2010; Kim *et al*, 2010). The presence of H3K4me3 in distal regulatory elements might therefore be associated with the

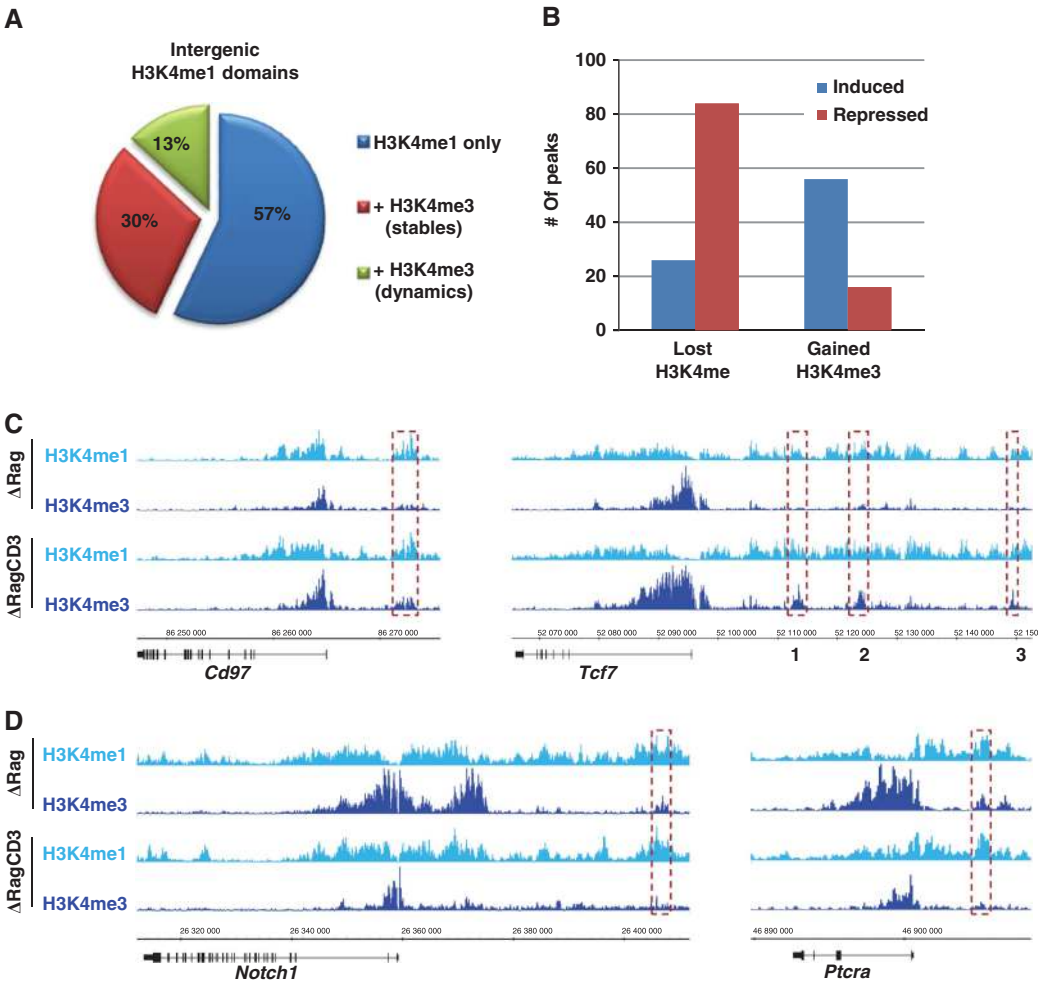

Figure 4 Epigenetic changes in distal regulatory regions are correlated with the patterns of transcriptional regulation. (**A**) Pie diagram indicating the rates of overlap between the intergenic H3K4me1 peaks observed in both ΔRag and ΔRagCD3 thymocytes and the H3K4me3 peaks found either in both thymocyte populations (stables) or in a single population (dynamics). (**B**) Percentage of dynamic H3K4me1 domains associated with either repressed (red) or induced (blue) genes, depending on whether they lost or gained H3K4me3 peaks upon pre-TCR signalling. (**C**, **D**) Examples of dynamic H3K4me1 domains (highlighted regions) associated with either induced (**C**) or repressed (**D**) genes. The legend for epigenetic profiles is as in Figure 1.

deposition of Pol II in these regions. To address this question, Pol II occupancy was analysed in ΔRag and ΔRagCD3 thymocytes by performing ChIP-Seq assays. Strikingly, all the known enhancers studied here were readily bound by Pol II when active (Figure 5A; Supplementary Figure S10A and C). In the case of the enhancers associated with pre-TCR-induced loci, Pol II occupancy was found to be either exclusively present (*Cd8* and *Tcra enhancers*) or to have increased (*Cd4 enhancer*) in ΔRagCD3 thymocytes compared with ΔRag thymocytes. In agreement with the patterns of Pol II association observed in active enhancers, stage-specific RNAs from the intergenic *Cd8* enhancer E8I were also detected (Figure 5B); and RNAs from the upstream *Dntt* enhancer were present in roughly similar levels in both ΔRag and ΔRagCD3 cells (Supplementary Figure S10B). Pol II occupancy of T cell-specific enhancers is, therefore, strongly correlated with H3K4me3 enrichment and enhancer activity. Next, we aimed to determine whether the levels of Pol II occupancy were correlated with the loss/gain of H3K4me3 within the H3K4me1 domains. Indeed, the dynamic of Pol II peaks was correlated with the loss or gain of H3K4me3 peaks within the same domains (Supplementary Figure S11A). Most strikingly, the average levels of Pol II occupancy dropped significantly in ΔRagCD3 thymocytes in the case of H3K4me1 domains that had lost H3K4me3, whereas they increased in domains that had acquired H3K4me3 (Figure 5C). In agreement with these results, we found that Pol II occupancy at intergenic H3K4me1 regions was observed only in the presence of H3K4me3 (Figure 5D; Supplementary Figure S11B), and correlated with H3K4me3 levels (Supplementary Figure S11C). Pol II recruitment and H3K4me3 enrichment are, therefore, closely linked in H3K4me1 domains modulated by pre-TCR signalling.

Given the functional link between initiating Pol II and H3K4me3 that have been described at promoter regions (Selth *et al*, 2010), we further investigated whether H3K4me3 enrichment found in enhancer regions may depend on Pol II occupancy. To this end, elongating Pol II was blocked by inhibiting the CDK9 kinase with the KM05283 chemical compound (Medlin *et al*, 2005; Tan-Wong *et al*, 2008). Normally, CDK9 phosphorylates Pol II at serine 2 in the CTD, which is required for transcriptional elongation

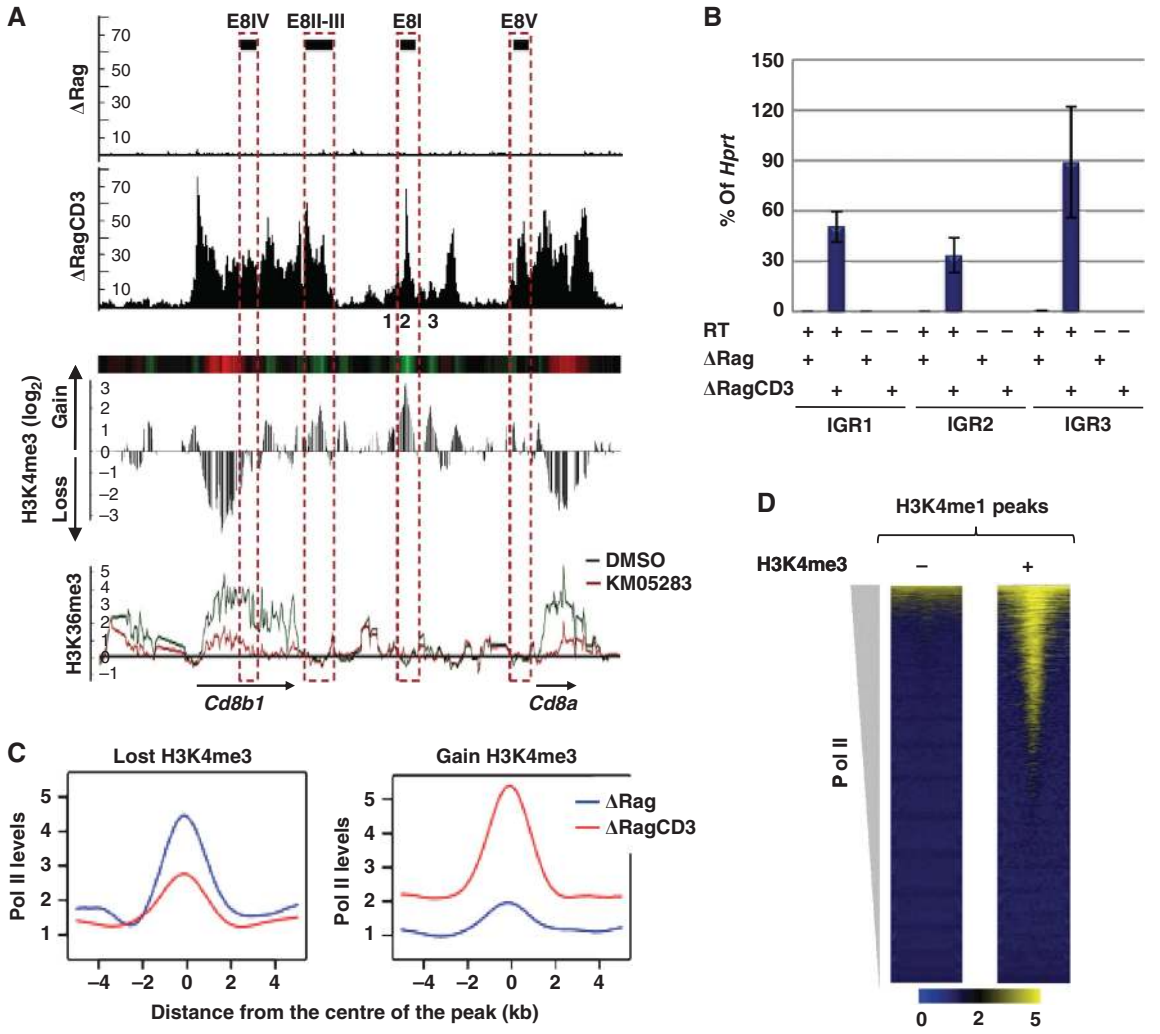

Figure 5 Pol II occupancy in distal regulatory regions is correlated with changes in H3K4me3 levels. (**A**) Pol II profiles directly assessed from ChIP-Seq data using ΔRag and ΔRagCD3 thymocytes (top panel), and relative changes in H3K4me3 and H3K36me3 levels in the *Cd8* locus following KM05283 treatment of P5424 cells (middle and bottom panels). H3K4me3 ChIPs from KM05283- and mock-treated P5424 cells were hybridized to the same 15k custom array and the experiment is shown as either bar plots (middle-bottom panel) or a colour coding bar (middle-upper panel; green and red indicates gain and loss of H3K4me3, respectively). (**B**) Results of RT–qPCR analysis showing the rates of intergenic transcription occurring at the *Cd8* locus in ΔRag and ΔRagCD3 thymocytes. Samples treated or not with reverse transcriptase (RT+ and RT−, respectively) were analysed by qPCR to study the intergenic regions (IGR1-3) indicated in (**A**). Relative transcript levels are expressed as percentages of *Hprt* expression. (**C**) Average Pol II association profiles in H3K4me1 domains overlapping Pol II peaks and that lost or acquired H3K4me3 in ΔRagCD3 thymocytes. The distance (in kb) from the middle of the H3K4me3 peak is indicated. (**D**) Heat maps of Pol II ChIP-Seq signal from ΔRagCD3 thymocytes on the intergenic H3K4me1 domains that overlap (+) or not (−) the H3K4me3 peaks. Data were ordered according to Pol II levels and centred on the middle of the H3K4me1 peak.

(Phatnani and Greenleaf, 2006). Inhibition of transcriptional elongation results in the accumulation of Pol II in both promoter and enhancer regions (Rahl *et al*, 2010 and data not shown). We reasoned that if H3K4 tri-methylation depends on the local occupancy by Pol II, then the level of H3K4me3 in enhancers is likely to increase in response to KM05283 treatment. In these experiments, we used the T cell line P5424, which is derived from ΔRag thymocytes but express markers of DP thymocytes (Mombaerts *et al*, 1995 and data not shown). The levels of H3K4me3 were compared between KM05283 and control DMSO-treated P5424 cells by performing ChIP-on-chip assays using a dedicated high-resolution array. To prevent normalization problems, the two samples were co-hybridized to the same array. Efficient inhibition of transcription elongation was validated by the significant reduction of H3K36me3 throughout the body of expressed genes (Figure 5A, bottom panel; Supplementary Figure S10D). As anticipated, after KM05283 treatment of P5424 cells, a significant increase in the H3K4me3 levels was observed in promoter and enhancer regions, concomitantly with a decrease in the H3K4me3 levels within the gene body (Figure 5A, middle panels; Supplementary Figure S10D). In conclusion, the above data suggest that H3K4me3 enrichment occurring in active enhancers is directly linked to the accumulation of Pol II in these regulatory elements.

Discussion

In the present study, we investigated the dynamics of post-translational modifications of Histone H3 in developing thymocytes and found that a specific chromatin signature involving H3K4 methylation moieties marks active enhancers

in vivo. For this purpose, a previously defined differentiation model system was used, which lends itself to studying homogenous cell populations at two critical early stages of T-lymphocyte development (Hayday and Pennington, 2007). The activity of well-defined T-cell enhancers was found to be consistently associated with the presence of all three states of H3K4 methylation (H3K4me1/2/3). Enhancers triggered by pre-TCR signalling were primed by H3K4me1 histone marks and acquired substantial levels of H3K4me2/3 during subsequent T-cell development. Importantly, H3K4me3 enrichment at enhancer regions was truly observed in ChIP-Seq experiments using native MNase-digested chromatin, ruling out the possibility of a crosslinking artefact due to enhancer–promoter interactions. Global analyses further showed that the changes in H3K4me2/3 levels occurring in intergenic H3K4me1-enriched domains closely mimic the gain or loss of expression of neighbouring genes during cell differentiation. Finally, we provided evidence for the existence of a direct link between the H3K4me3 enrichment and Pol II binding in distal regulatory regions. We conclude that a specific chromatin signature including H3K4me3 reflects the presence of active enhancers.

Genome-wide studies have led to defining chromatin signatures characteristic of discrete *cis*-regulatory elements (Hon *et al*, 2009). In particular, it has been suggested that distal regulatory elements may be enriched with H3K4me1, but not with H3K4me3 (Heintzman *et al*, 2007, 2009). Several groups subsequently used these or similar criteria to characterize enhancers in various developmental and stimulatory contexts (e.g., Birney *et al*, 2007; Visel *et al*, 2009; Ghisletti *et al*, 2010; He *et al*, 2010; Heinz *et al*, 2010; Kim *et al*, 2010). However, H3K4me1 enrichment is generally observed in enhancer regions before gene activation (Cui *et al*, 2009; Creyghton *et al*, 2010; this study). Thus, should the chromatin signature of enhancers be restricted to H3K4me1 alone, important information may be lost. Upon analysing the epigenetic patterns of well-known T-cell enhancers, we demonstrate that H3K4me2/3 specifically marks active distal regulatory elements in developing thymocytes. Although H3K4me3 mark have also been previously observed in distal genomic regions (Barski *et al*, 2007; Lupien *et al*, 2008; Robertson *et al*, 2008; Wang *et al*, 2008; Schmidl *et al*, 2009; Hoffman *et al*, 2010; Lin *et al*, 2010), this is the first time H3K4me3 enrichment have been directly correlated with enhancer activity in primary cells. However, our findings do not necessarily imply that all enhancers would carry this mark when active. Certainly, at least a fraction of H3K4me1-enriched enhancers might be active in the absence of H3K4me3. Recently, H3K27 acetylation (H3K27ac) has separately been found to be associated with developmentally activated distal regulatory regions (Creyghton *et al*, 2010; Rada-Iglesias *et al*, 2011). It will be of interest to determine whether active enhancers display both H3K27ac and H3K4me3 or whether these histone modifications specifically mark distinct types of enhancers.

Enhancer priming

Several enhancers known to respond to pre-TCR signalling were found here to be associated with H3K4me1 in ΔRag thymocytes before their activation (Figure 2). It is possible that H3K4me1 marking of developmentally regulated enhancers before transcriptional activation may constitute (or reflect) an epigenetic priming mechanism. A subset of tissue-specific enhancers was recently found to be epigenetically marked in embryonic cells, and this process seems to play an important role in the subsequent expression of the associated genes in more differentiated cells (Xu *et al*, 2007, 2009). The presence of H3K4me1 may possibly suffice to maintain an open chromatin structure and recruit at least some TFs, whereas full enhancer activation (and possibly long-range enhancer–promoter interactions) may require H3K4me2/3. The enhancers of the *Tcra* and *b* loci provide good examples in support of this scenario. We and others have previously established that Eα is pre-assembled in the form of an inactive nucleoprotein complex in ΔRag thymocytes (Hernandez-Munain *et al*, 1999; Spicuglia *et al*, 2000). Indeed, Eα is associated with H3K4me1 at this early stage. Likewise, the truncated $E\beta^{169}$ enhancer was found to be associated with higher H3K4me1 and lower H3K4me3 levels than wild-type Eβ (Figure 3). Yet, $E\beta^{169}$ has been reported to sustain the binding of a subset of Eβ-bound TFs (Bonnet *et al*, 2009).

H3K4me3 is associated with active distal regulatory elements

It is generally assumed that H3K4me3 is a hallmark of gene promoters. Based on this assumption, intergenic H3K4me3 peaks were thought to correspond to unannotated promoters of either protein-coding genes or long intergenic non-coding RNAs (LincRNA) (Mikkelsen *et al*, 2007; Guttman *et al*, 2009). Although it is difficult to discriminate between enhancers and unannotated promoters in genome-wide approaches, several of our findings make likely that a significant subset of H3K4me1 domains enriched with H3K4me3 may reflect the presence of *bona fide* enhancer regions. First, the activity of all genuine T-cell enhancers studied here was associated with the presence of H3K4me3 (Figures 1 and 2). Second, the level of H4K4me1, which is expected to be enriched at enhancers, was found to be higher at intergenic H3K4me1/me3 domains as compared with TSS-associated domains (Supplementary Figure S5). Third, several intergenic H3K4me1/me3 regions demonstrated significant enhancer activity in a luciferase reporter assays (Supplementary Figure S9). Finally, a strong correlation was found to exist between distal H3K4me1 domains that either acquired or lost H3K4me3 and, respectively, the induction or repression of neighbouring genes by pre-TCR signalling (Supplementary Figure S5B).

Enhancer activation and Pol II-dependent deposition of H3K4me3

Pol II occupancy of enhancers has been described at several loci (Marenduzzo *et al*, 2007; Koch *et al*, 2008) and seems to be a common trait of distal regulatory elements (De Santa *et al*, 2010; Kim *et al*, 2010), most probably associated with the setting of enhancers' activity. In this study, we confirmed and extended these previous results by showing that Pol II is specifically recruited by active enhancers in differentiating T cells (Figure 5A; Supplementary Figure S10). Pol II occupancy was found to be strikingly well correlated with the dynamic process of H3K4me3 enrichment occurring in distal regulatory elements (Figures 5C and D; Supplementary Figure S11). Targeting of the yeast H3K4 methyltransferase (COMPASS/Set1) complex to 5′ transcribed regions requires ser5P Pol II (Ng *et al*, 2003; Selth *et al*, 2010). This is also presumed to be

true for at least some of the homologous mammalian MLL1 and Set1 complexes (Buratowski, 2009). Interestingly, it is the ser5P Pol II form which appears to be associated with distal regulatory regions (De Santa *et al*, 2010; Kim *et al*, 2010). It is therefore likely that, as in the case of promoter regions, the H3K4me3 levels present in enhancers may actually reflect their local occupancy by initiating Pol II.

Another characteristic of the T cell-specific enhancers studied here is the high extent of H3K4me, which can spread over several kbs within a given locus. The spreading of H3K4me is clearly correlated with the induction of enhancer activity (see for example the *Cd8*-intergenic region and surrounding *Tcra* enhancer regions). This phenomenon is unlikely to be due to a technical artefact during the ChIP procedure as it was observed using either crosslinked or native preparations of chromatin. In line with these findings, we have previously reported that the bodies of tissue-specific genes are extensively covered with H3K4me and proposed the existence of promoter-proximal *cis*-regulatory platforms (Pekowska *et al*, 2010). The observation that inhibition of Pol II elongation affects H3K4me3 levels at both gene bodies and *cis*-regulatory regions (Figure 5A; Supplementary Figure S10) strongly supports a close link between H3K4 methylation, Pol II recruitment and transcription in mammalian genes. The present findings extend our previous conclusions and strongly suggest that regulatory components of highly tissue-specific genes may carry a characteristic chromatin signature involving high levels of methylated H3K4 enrichment including both the gene bodies and the intergenic regions.

In addition to the basic findings presented here on the chromatin signature of active enhancers, a number of putative distal regulatory elements were identified, including those at loci playing important roles in T cells such as *Ets1/Fli1* or *Tcf7* (Supplementary Figure S9). At a more global level, we drew up cell-specific maps featuring six histone modification profiles, along with Pol II occupancy, at two different stages of T-cell differentiation. These maps could be used in the future to study the epigenetic regulation of genes involved in normal and pathological T-cell development.

In conclusion, the enrichment of H3K4me2/3 observed here in distal regulatory regions, which was found to be highly correlated with gene expression and Pol II occupancy, suggests that H3K4 di- and tri-methylation may be involved in the functional activity of promoter-distal regulatory regions. The fact that a similar situation was found to occur in developing B lymphocytes (Supplementary Figure S4) means that this basic principle applies to other developing cells as well.

Materials and methods

Mice and cell lines

Rag2$^{-/-}$ (ΔRag) (Shinkai *et al*, 1992) and ΔRag;Eβ^{169} (Bonnet *et al*, 2009) male mice bred on a C57BL/6J background were used in this study. To obtain ΔRagCD3 thymocytes, 4-week-old mice were injected intraperitoneally with 75 μg of monoclonal antibodies (Abs) against CD3ε (Cat. No: 553058, BD Bioscience, San Diego, CA); and thymi were collected 7 days after the injection. Mice were housed under specific pathogen-free conditions and handled in keeping with European directives. The P5424 cell line has been previously described (Mombaerts *et al*, 1995).

Cell preparation and staining

Thymi from 4- to 6-week-old mice were extracted and mechanically disrupted on a 75-μm nylon mesh (BD Bioscience). To check for thymocyte differentiation after CD3 stimulation (ΔRagCD3), cells were stained with anti-CD4 and anti-CD8 Abs (Cat. No: 553048 and 553031, respectively, BD Bioscience) and analysed by FACS. Only cells from mice in which staining showed at least 95% of CD4$^+$CD8$^+$ cells were used for further experiments.

RNA preparation and quantitative RT–PCR (RT–qPCR) assays

Total RNA from thymocytes was prepared using TRIzol reagent (Invitrogen, Paisley, UK) as recommended by the manufacturer. The samples were resuspended in Rnase-free water and treated with DNase I (Ambion, Warrington, UK) for 30 min at 37°C. RNA quality was then checked using a Bioanalyzer (Agilent Technologies, Santa Clara, USA). Only RNA with RNA integrity number >9 was then used. For RT–qPCR assays, RNA was converted into cDNA using the Superscript III (Invitrogen) reverse transcriptase together with random hexamers. cDNA was subsequently amplified with specific primers (Supplementary Table S2) using a 7500 Fast instrument (Life Technologies, CA) and following the manufacturer's instruction.

Gene expression analysis

Transcriptome experiments were performed in triplicate using the Affymetrix Mouse Gene 1.0 ST platform. RNA samples were amplified, labelled and hybridized according to the manufacturer's recommendations. Expression data were RMA normalized using the Affy library from R-bioconductor (http://www.R-project.org). Assessment of individual gene expression levels was in agreement with published transcriptome data (Hoffmann *et al*, 2003; Puthier *et al*, 2004). Differential expression between ΔRag and ΔRagCD3 samples was assessed using a moderated *t*-statistic test computed by R-limma lmFit and eBayes functions (Smyth, 2004). To select the most differentially expressed genes, we applied a stringent threshold (adjusted *P*-value <0.001) and retrieved 923 repressed and 1008 induced genes (Supplementary Table S1).

KM05283 treatment

Exponentially growing P5424 cells (15×10^6) were incubated in RPMI medium supplemented with either 50 μM of KM05283 (Maybridge, Cornwall, UK) or control DMSO (Sigma-Aldrich, St Louis, MO) for 24 h at 37°C. Following incubation, cells were washed twice with 1× PBS and processed for ChIP as indicated below. Inhibition of Pol II Ser2 phosphorylation was checked by western blot, as described previously (Medlin *et al*, 2005).

Chromatin immunoprecipitation

ChIP of crosslinked chromatin from ΔRag, ΔRag;Eβ^{169}, ΔRagCD3 thymocytes, P5424 and mouse embryonic stem (ES) cell lines, was performed essentially as previously described (Spicuglia *et al*, 2002), using the ChIP and EZ ChIP kits (Upstate Lake placid, NY; Cat. No: 17-295 and 17-371, respectively). Cell extracts were sonicated eight times, using the S-4000 Sonifier (Misonix, NY) with 30 s pulses, in order to obtain DNA fragments between 200 and 500 bp in length. Precipitation of pre-cleared chromatin from 1×10^6 cells was carried out overnight at 4°C using the following Abs: anti-H3K4me1 (Abcam, Cambridge, UK; Cat. No: 8895, 4 μg), anti-H3K4me2 (Upstate; Cat. No: 07030, 5 μl), anti-H3K4me3 (Abcam; Cat. No: 8580, 3 μg), anti-H3K9me2 (Abcam; Cat. No: 1220, 2 μg), anti-H3K27me3 (Abcam; Cat. No: 05851, 4 μg) or anti-H3K36me3 (Abcam; Cat. No: 9050, 3 μg). Immune complexes were collected using either Protein A (H3K4Me1 and H3K4me2 ChIPs) or Protein G (all remaining ChIPs) sepharose beads. Pol II ChIP was carried out as described elsewhere (Boyer *et al*, 2005) using N20 Abs (sc-899X, Santa Cruz Biotechnology), with minor modifications. Briefly, 1 ml of chromatin extract corresponding to 10×10^6 cells was pre-washed with 40 μl of magnetic beads (Invitrogen) for 1 h at 4°C in a rotating wheel. The ChIP procedure was then carried out using Abs/bead complexes (4 μg of Ab/40 μl beads). After reverse crosslinking, enriched DNA fragments were extracted with phenol/chloroform and recovered using the Qiaquick PCR-purification kit (Qiagen, Germany). Mononucleosome preparation of native chromatin (MNase-ChIP) was carried out as described (Umlauf *et al*, 2004), with minor modifications. Briefly, the nuclei from 20×10^6 of either ΔRag or ΔRagCD3 thymocytes were digested with 10 U MNase (Roche, Switzerland) for 10 min and the soluble chromatin fraction (mononucleosomes) from the equivalent of 5×10^6 of cells was used for ChIP using Abs against H3K4me1 and H3K4me3. The quality of individual ChIP samples was checked at known target sites by

qPCR and DNA size was verified on a 2100 Bioanalyzer (Agilent Technologies). For ChIP-on-chip analysis, ChIP samples, as well as an aliquot of identically treated input DNA, were amplified by performing ligation-mediated PCR (LM–PCR), as previously described (Ren *et al*, 2000). For ChIP–qPCR validation of histone mark enrichment, ChIP was prepared as described above and qPCR analyses of two biologically independent samples were carried out in triplicate. Primers used for qPCR are described in Supplementary Table S2.

Design and analysis of Chip-on-chip experiments

In all, 650 highly regulated genes and 25 control genes were selected from published transcriptome data sets on purified thymocyte subsets (Hoffmann *et al*, 2003). The custom tiling array (244k) was designed to cover the $+50/-50$ kb genomic regions located upstream and downstream of the selected genes, and an additional set comprising 18 gene clusters (including *Tcr* and *Hox* loci) at a resolution of 280 bp, using the eArray web tool (http://earray.chem.agilent.com). In total, 1717 genes were covered in this array (exhaustive list of all genes is given in Supplementary Table S1). Annotations were generated according to the mouse genome assembly Build 37 by NCBI. In some experiments, a dedicated 15k tilling array at a resolution of 100 bp was used, covering the following loci: *Tcrb*, *Tcra*, *Cd8*, *Cd4*, *Cd3* and *Actb*. Regarding the ChIP-on-chip procedure, 2 μg of LM–PCR amplified material was labelled using the BioPrime Plus Array CGH Genomic Labelling System (Invitrogen) according to the manufacturer's recommendations. In the case of the dedicated array, the ChIP material was used directly for labelling without any previous amplification. Input and ChIP samples were labelled with Cy3-dUTP and Cy5-dUTP (Perkin-Elmer, Massachusetts), respectively. Hybridization and washing were performed as recommended by the manufacturer (Agilent Technologies). For Pol II inhibition experiments, ChIP samples from KM05283-treated cells were labelled with Cy5-dUTP and those from DMSO-treated cells were labelled with Cy3-dUTP, and co-hybridized to the same array. Images were scanned using a DNA microarray scanner and processed using FE v9.5.1 Software (Agilent Technologies). Microarrays were intraarray normalized using the preprocess function provided with the Ringo library (Toedling *et al*, 2007). ChIP-on-chip data obtained with the same Ab were interarray normalized using the VSN function from R-Bioconductor (Huber *et al*, 2002). The results obtained with at least two biological replicates were averaged and converted into GFF and SGR files for further analysis. ChIP-on-chip data were displayed in the form of a log2 (ChIP/Input) ratio using the IGB tool (Nicol *et al*, 2009). Significantly enriched regions (peaks) were isolated using the CoCAS software program (Benoukraf *et al*, 2009) and the Ringo threshold option (99.5 and 99 percentiles as the main and extended peak thresholds, respectively). Two peaks were granted to overlap when the smaller one had at least 70% in common with the larger region. H3K4me1 domains were defined by the overlap between H3K4me1 intergenic peaks detected in both ΔRag and ΔRagCD3 thymocytes and the other epigenetic peaks (exhaustive list of all peaks is provided in Supplementary Table S3). Single average profiles were generated by performing a linear regression among the regions of interest with a resolution of 100 bp, using the approx function provided with the R-Base library. The mean of each single profile belonging to the same group was then calculated.

ChIP-Seq data generation and analysis

ChIP-Seq data from mouse B cells (Lin *et al*, 2010) were downloaded from GEO Data Sets (http://www.ncbi.nlm.nih.gov/gds) under accession numbers GSM546518, GSM546519, GSM546520, GSM546527, GSM546528 and GSM546529, which correspond to the ChIP-Seq profiles of H3K4me1, H3K4me2 and H3K4me3 in EBF knockout mice and in Rag1 knockout mice. Tags were aligned to the reference genome (NCBI37/mm9) using Bowtie (Langmead *et al*, 2009). Sequencing of Pol II ChIP and MNase-ChIP DNA samples was performed on an AB SOLiD V4.0 (Life Technologies) according to the manufacturer's protocol and mapped onto the NCBI37/mm9 reference genome using the SOLiD pre-processing pipeline. Only uniquely mapped sequence tags were used for further analysis. To reduce the risk of potential bias, we discarded all tags with exactly the same coordinates. Reads were then elongated to 245 bp (Pol II) or 146 bp (MNase-ChIP). The values were summed up in non-overlapping 100 bp bins along each chromosome. The summed values were finally normalized by dividing them by the overall mean values per bin. Enriched regions (peaks) were isolated using CoCAS software and default settings (peaks found also in Input control samples were removed). Annotated ChIP-Seq peaks are provided in Supplementary Table S4. To obtain the average Pol II profiles around the intergenic H3K4me1 peaks, the 100-bp bins were annotated with respect to the middle of the H3K4me1 domain in the regions ranging from -2 to $+2$ kb, and a linear interpolation was performed in order to obtain the same relative positions of the bins in all H3K4me1 domains. All data analyses were performed using R software program.

Reporter assay

EL4 cells were transiently transfected in 12-well plates with 0.4 μg of Renilla vector along with 4.2 μg of pGL3-promoter vector (E1910; Promega) or pGL3-promoter vectors containing the specified genomic regions cloned downstream the SV40 promoter (Supplementary Table S5). The transfection was performed by using Cell Line Nucleofector Kit L (VCA-1005; AMAXA-LONZA) according to the manufacturer's protocol. Twenty-four hours after transfection, cells were washed once with $1\times$ PBS and the luciferase assay (Bright-Glo, Promega) was performed according to the manufacturer protocols. Values are expressed as fold increase in luciferase counts over the pGL3-promoter vector and normalized by the Renilla intensities.

Data availability

Gene expression, ChIP-on-chip and ChIP-Seq files are available online at http://www.comline.fr/ciml.

Supplementary data

Supplementary data are available at *The EMBO Journal* Online (http://www.embojournal.org).

Acknowledgements

We thank Drs W Huber and J Toedling (EMBL, Heidelberg) for helpful advises on data analysis; the Nice-Sophia-Antipolis platform for gene expression experiments; A Bergon (TAGC, Marseille) for help with Solid DNA sequencing; Drs C Murre (University of California, USA) and G Natoli (European Institute of Oncology, Milan) for pGL3 constructs; and members of the PF laboratory for critical reading of this manuscript. ChIP-Seq sequencing was performed by the IBiSA Transcriptomics and Genomics Marseille-Luminy platform. Work in PF laboratory is supported by institutional grants from Inserm and CNRS, and by specific grants from the '*Fondation Princesse Grace de Monaco*', the '*Association pour la Recherche sur le Cancer*' (ARC), the '*Agence Nationale de la Recherche*' (ANR), the '*Institut National du Cancer*' (INCa) and the Commission of the European Communities. AP was supported by a Marie Curie research training fellowship (MRTN-CT-2006-035733) and the '*Fondation pour la Recherche Médicale*' (FRM); TB was supported by the ANR and the FRM.

Author contributions: AP, TB, PF and SS conceived the frame of the study; TB, JZC and SS designed the DNA microarray; AP and JZC performed the biological experiments; HH and JI performed the sequencing of ChIP-Seq experiments; JCA and FK participated in the luciferase reporter assay; AP, TB, MB and SS performed the bioinformatic analyses; AP, PF and SS wrote the manuscript. All authors reviewed the manuscript.

Conflict of interest

The authors declare that they have no conflict of interest.

References

Adlam M, Siu G (2003) Hierarchical interactions control CD4 gene expression during thymocyte development. *Immunity* **18:** 173–184

Anderson MK (2006) At the crossroads: diverse roles of early thymocyte transcriptional regulators. *Immunol Rev* **209:** 191–211

Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, Wang Z, Wei G, Chepelev I, Zhao K (2007) High-resolution profiling of histone methylations in the human genome. *Cell* **129:** 823–837

Benoukraf T, Cauchy P, Fenouil R, Jeanniard A, Koch F, Jaeger S, Thieffry D, Imbert J, Andrau JC, Spicuglia S, Ferrier P (2009) CoCAS: a ChIP-on-chip analysis suite. *Bioinformatics (Oxford, England)* **25:** 954–955

Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, Cuff J, Fry B, Meissner A, Wernig M, Plath K, Jaenisch R, Wagschal A, Feil R, Schreiber SL, Lander ES (2006) A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125:** 315–326

Birney E, Stamatoyannopoulos JA, Dutta A, Guigo R, Gingeras TR, Margulies EH, Weng Z, Snyder M, Dermitzakis ET, Thurman RE, Kuehn MS, Taylor CM, Neph S, Koch CM, Asthana S, Malhotra A, Adzhubei I, Greenbaum JA, Andrews RM, Flicek P *et al* (2007) Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447:** 799–816

Bonnet M, Huang F, Benoukraf T, Cabaud O, Verthuy C, Boucher A, Jaeger S, Ferrier P, Spicuglia S (2009) Duality of enhancer functioning mode revealed in a reduced TCRβ gene enhancer knockin mouse model. *J Immunol* **183:** 7939–7948

Boyer LA, Lee TI, Cole MF, Johnstone SE, Levine SS, Zucker JP, Guenther MG, Kumar RM, Murray HL, Jenner RG, Gifford DK, Melton DA, Jaenisch R, Young RA (2005) Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* **122:** 947–956

Bulger M, Groudine M (2011) Functional and mechanistic diversity of distal transcription enhancers. *Cell* **144:** 327–339

Buratowski S (2009) Progression through the RNA polymerase II CTD cycle. *Mol Cell* **36:** 541–546

Capone M, Watrin F, Fernex C, Horvat B, Krippl B, Scollay R, Ferrier P (1993) TCRα and TCRβ gene enhancers confer tissue- and stage-specificity on V(D)J recombination events. *EMBO J* **12:** 4335–4346

Chattopadhyay S, Whitehurst CE, Schwenk F, Chen J (1998) Biochemical and functional analyses of chromatin changes at the TCRβ gene locus during CD4−CD8− to CD4+CD8+ thymocyte differentiation. *J Immunol* **160:** 1256–1267

Cherrier M, D'Andon MF, Rougeon F, Doyen N (2008) Identification of a new cis-regulatory element of the terminal deoxynucleotidyl transferase gene in the 5′ region of the murine locus. *Mol Immunol* **45:** 1009–1017

Creyghton MP, Cheng AW, Welstead GG, Kooistra T, Carey BW, Steine EJ, Hanna J, Lodato MA, Frampton GM, Sharp PA, Boyer LA, Young RA, Jaenisch R (2010) Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci USA* **107:** 21931–21936

Cui K, Zang C, Roh TY, Schones DE, Childs RW, Peng W, Zhao K (2009) Chromatin signatures in multipotent human hematopoietic stem cells indicate the fate of bivalent genes during differentiation. *Cell Stem Cell* **4:** 80–93

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

Georgopoulos K, van den Elsen P, Bier E, Maxam A, Terhorst C (1988) A T cell-specific enhancer is located in a DNase I-hypersensitive area at the 3′ end of the CD3-delta gene. *EMBO J* **7:** 2401–2407

Ghisletti S, Barozzi I, Mietton F, Polletti S, De Santa F, Venturini E, Gregory L, Lonie L, Chew A, Wei CL, Ragoussis J, Natoli G (2010) Identification and characterization of enhancers controlling the inflammatory gene expression program in macrophages. *Immunity* **32:** 317–328

Guenther MG, Levine SS, Boyer LA, Jaenisch R, Young RA (2007) A chromatin landmark and transcription initiation at most promoters in human cells. *Cell* **130:** 77–88

Guttman M, Amit I, Garber M, French C, Lin MF, Feldser D, Huarte M, Zuk O, Carey BW, Cassady JP, Cabili MN, Jaenisch R, Mikkelsen TS, Jacks T, Hacohen N, Bernstein BE, Kellis M, Regev A, Rinn JL, Lander ES (2009) Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* **458:** 223–227

Hayday AC, Pennington DJ (2007) Key factors in the organized chaos of early T cell development. *Nat Immunol* **8:** 137–144

He HH, Meyer CA, Shin H, Bailey ST, Wei G, Wang Q, Zhang Y, Xu K, Ni M, Lupien M, Mieczkowski P, Lieb JD, Zhao K, Brown M, Liu XS (2010) Nucleosome dynamics define transcriptional enhancers. *Nat Genet* **42:** 343–347

Heintzman ND, Hon GC, Hawkins RD, Kheradpour P, Stark A, Harp LF, Ye Z, Lee LK, Stuart RK, Ching CW, Ching KA, ntosiewicz-Bourget JE, Liu H, Zhang X, Green RD, Lobanenkov VV, Stewart R, Thomson JA, Crawford GE, Kellis M *et al* (2009) Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* **459:** 108–112

Heintzman ND, Stuart RK, Hon G, Fu Y, Ching CW, Hawkins RD, Barrera LO, Van Calcar S, Qu C, Ching KA, Wang W, Weng Z, Green RD, Crawford GE, Ren B (2007) Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet* **39:** 311–318

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

Hernandez-Munain C, Sleckman BP, Krangel MS (1999) A developmental switch from TCRδ enhancer to TCRα enhancer function during thymocyte maturation. *Immunity* **10:** 723–733

Hoffman BG, Robertson G, Zavaglia B, Beach M, Cullum R, Lee S, Soukhatcheva G, Li L, Wederell ED, Thiessen N, Bilenky M, Cezard T, Tam A, Kamoh B, Birol I, Dai D, Zhao Y, Hirst M, Verchere CB, Helgason CD *et al* (2010) Locus co-occupancy, nucleosome positioning, and H3K4me1 regulate the functionality of FOXA2-, HNF4A-, and PDX1-bound loci in islets and liver. *Genome Res* **20:** 1037–1051

Hoffmann R, Bruno L, Seidl T, Rolink A, Melchers F (2003) Rules for gene usage inferred from a comparison of large-scale gene expression profiles of T and B lymphocyte development. *J Immunol* **170:** 1339–1353

Hon G, Wang W, Ren B (2009) Discovery and annotation of functional chromatin signatures in the human genome. *PLoS Comput Biol* **5:** e1000566

Hosoya-Ohmura S, Lin YH, Herrmann M, Kuroha T, Rao A, Moriguchi T, Lim KC, Hosoya T, Engel JD (2011) An NK and T cell enhancer lies 280 kilobase pairs 3′ to the gata3 structural gene. *Mol Cell Biol* **31:** 1894–1904

Huber W, von Heydebreck A, Sultmann H, Poustka A, Vingron M (2002) Variance stabilization applied to microarray data calibration and to the quantification of differential expression. *Bioinformatics (Oxford, England)* **18**(Suppl 1)**:** S96–S104

Jenuwein T, Allis CD (2001) Translating the histone code. *Science (New York, NY)* **293:** 1074–1080

Jiang H, Zhang F, Kurosu T, Peterlin BM (2005) Runx1 binds positive transcription elongation factor b and represses transcriptional elongation by RNA polymerase II: possible mechanism of CD4 silencing. *Mol Cell Biol* **25:** 10675–10683

Kaufmann C, Yoshida T, Perotti EA, Landhuis E, Wu P, Georgopoulos K (2003) A complex network of regulatory elements in Ikaros and their activity during hemo-lymphopoiesis. *EMBO J* **22:** 2211–2223

Kim TK, Hemberg M, Gray JM, Costa AM, Bear DM, Wu J, Harmin DA, Laptewicz M, Barbara-Haley K, Kuersten S, Markenscoff-Papadimitriou E, Kuhl D, Bito H, Worley PF, Kreiman G, Greenberg ME (2010) Widespread transcription at neuronal activity-regulated enhancers. *Nature* **465:** 182–187

Koch F, Jourquin F, Ferrier P, Andrau JC (2008) Genome-wide RNA polymerase II: not genes only!. *Trends Biochem Sci* **33:** 265–273

Krangel MS (2007) T cell development: better living through chromatin. *Nat Immunol* **7:** 687–694

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

Lauzurica P, Krangel MS (1994) Temporal and lineage-specific control of T cell receptor α/δ gene rearrangement by T cell receptor α and δ enhancers. *J Exp Med* **179:** 1913–1921

Li B, Carey M, Workman JL (2007) The role of chromatin during transcription. *Cell* **128:** 707–719

Lin YC, Jhunjhunwala S, Benner C, Heinz S, Welinder E, Mansson R, Sigvardsson M, Hagman J, Espinoza CA, Dutkowski J, Ideker T, Glass CK, Murre C (2010) A global network of transcription factors, involving E2A, EBF1 and Foxo1, that orchestrates B cell fate. *Nat Immunol* **11:** 635–643

Lupien M, Eeckhoute J, Meyer CA, Wang Q, Zhang Y, Li W, Carroll JS, Liu XS, Brown M (2008) FoxA1 translates epigenetic signatures into enhancer-driven lineage-specific transcription. *Cell* **132:** 958–970

Marenduzzo D, Faro-Trindade I, Cook PR (2007) What are the molecular ties that maintain genomic loops? *Trends Genet* **23:** 126–133

Medlin J, Scurry A, Taylor A, Zhang F, Peterlin BM, Murphy S (2005) P-TEFb is not an essential elongation factor for the intronless human U2 snRNA and histone H2b genes. *EMBO J* **24:** 4154–4165

Mellor J, Dudek P, Clynes D (2008) A glimpse into the epigenetic landscape of gene regulation. *Curr Opin Genet Dev* **18:** 116–122

Mendenhall EM, Bernstein BE (2008) Chromatin state maps: new technologies, new insights. *Curr Opin Genet Dev* **18:** 109–115

Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G, Alvarez P, Brockman W, Kim TK, Koche RP, Lee W, Mendenhall E, O'Donovan A, Presser A, Russ C, Xie X, Meissner A, Wernig M, Jaenisch R, Nusbaum C *et al* (2007) Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448:** 553–560

Mombaerts P, Terhorst C, Jacks T, Tonegawa S, Sancho J (1995) Characterization of immature thymocyte lines derived from T-cell receptor or recombination activating gene 1 and p53 double mutant mice. *Proc Natl Acad Sci USA* **92:** 7420–7424

Natoli G (2010) Maintaining cell identity through global control of genomic organization. *Immunity* **33:** 12–24

Ng HH, Robert F, Young RA, Struhl K (2003) Targeted recruitment of Set1 histone methylase by elongating Pol II provides a localized mark and memory of recent transcriptional activity. *Mol Cell* **11:** 709–719

Nicol JW, Helt GA, Blanchard Jr SG, Raja A, Loraine AE (2009) The Integrated Genome Browser: free software for distribution and exploration of genome-scale datasets. *Bioinformatics (Oxford, England)* **25:** 2730–2731

Pekowska A, Benoukraf T, Ferrier P, Spicuglia S (2010) A unique H3K4me2 profile marks tissue-specific gene regulation. *Genome Res* **20:** 1493–1502

Phatnani HP, Greenleaf AL (2006) Phosphorylation and functions of the RNA polymerase II CTD. *Genes Dev* **20:** 2922–2936

Puthier D, Joly F, Irla M, Saade M, Victorero G, Loriod B, Nguyen C (2004) A general survey of thymocyte differentiation by transcriptional analysis of knockout mouse models. *J Immunol* **173:** 6109–6118

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

Rahl PB, Lin CY, Seila AC, Flynn RA, McCuine S, Burge CB, Sharp PA, Young RA (2010) c-Myc regulates transcriptional pause release. *Cell* **141:** 432–445

Ren B, Robert F, Wyrick JJ, Aparicio O, Jennings EG, Simon I, Zeitlinger J, Schreiber J, Hannett N, Kanin E, Volkert TL, Wilson CJ, Bell SP, Young RA (2000) Genome-wide location and function of DNA binding proteins. *Science* **290:** 2306–2309

Robertson AG, Bilenky M, Tam A, Zhao Y, Zeng T, Thiessen N, Cezard T, Fejes AP, Wederell ED, Cullum R, Euskirchen G, Krzywinski M, Birol I, Snyder M, Hoodless PA, Hirst M, Marra MA, Jones SJ (2008) Genome-wide relationship between histone H3 lysine 4 mono- and tri-methylation and transcription factor binding. *Genome Res* **18:** 1906–1917

Schmidl C, Klug M, Boeld TJ, Andreesen R, Hoffmann P, Edinger M, Rehli M (2009) Lineage-specific DNA methylation in T cells correlates with histone methylation and enhancer activity. *Genome Res* **19:** 1165–1174

Selth LA, Sigurdsson S, Svejstrup JQ (2010) Transcript elongation by RNA polymerase II. *Annu Rev Biochem* **79:** 271–293

Senoo M, Shinkai Y (1998) Regulation of Vβ germline transcription in RAG-deficient mice by the CD3epsilon-mediated signals: implication of Vβ transcriptional regulation in TCRβ allelic exclusion. *Int Immunol* **10:** 553–560

Shinkai Y, Rathbun G, Lam KP, Oltz EM, Stewart V, Mendelsohn M, Charron J, Datta M, Young K, Stall AL, Alt FW (1992) RAG-2 deficient mice lack mature lymphocytes owing to inability to initiate V(D)J rearrangement. *Cell* **68:** 855–867

Sleckman BP, Bardon CG, Ferrini R, Davidson L, Alt FW (1997) Function of the TCRα enhancer in αβ and γδ T cells. *Immunity* **7:** 505–515

Smyth GK (2004) Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol* **3:** Article3

Spicuglia S, Kumar S, Yeh JH, Vachez E, Chasson L, Gorbatch S, Cautres J, Ferrier P (2002) Promoter activation by enhancer-dependent and -independent loading of activator and coactivator complexes. *Mol Cell* **10:** 1479–1487

Spicuglia S, Payet D, Tripathi RK, Rameil P, Verthuy C, Imbert J, Ferrier P, Hempel WM (2000) TCRα enhancer activation occurs via a conformational change of a pre-assembled nucleo-protein complex. *EMBO J* **19:** 2034–2045

Taniuchi I, Ellmeier W, Littman DR (2004) The CD4/CD8 lineage choice: new insights into epigenetic regulation during T cell development. *Adv Immunol* **83:** 55–89

Tan-Wong SM, French JD, Proudfoot NJ, Brown MA (2008) Dynamic interactions between the promoter and terminator regions of the mammalian BRCA1 gene. *Proc Natl Acad Sci USA* **105:** 5160–5165

Toedling J, Skylar O, Krueger T, Fischer JJ, Sperling S, Huber W (2007) Ringo–an R/Bioconductor package for analyzing ChIP-chip readouts. *BMC Bioinformatics* **8:** 221

Umlauf D, Goto Y, Feil R (2004) Site-specific analysis of histone methylation and acetylation. *Methods Mol Biol* **287:** 99–120

Visel A, Blow MJ, Li Z, Zhang T, Akiyama JA, Holt A, Plajzer-Frick I, Shoukry M, Wright C, Chen F, Afzal V, Ren B, Rubin EM, Pennacchio LA (2009) ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* **457:** 854–858

von Boehmer H (2005) Unique features of the pre-T-cell receptor alpha-chain: not just a surrogate. *Nat Rev Immunol* **5:** 571–577

Wang Z, Zang C, Rosenfeld JA, Schones DE, Barski A, Cuddapah S, Cui K, Roh TY, Peng W, Zhang MQ, Zhao K (2008) Combinatorial patterns of histone acetylations and methylations in the human genome. *Nat Genet* **40:** 897–903

Xu J, Pope SD, Jazirehi AR, Attema JL, Papathanasiou P, Watts JA, Zaret KS, Weissman IL, Smale ST (2007) Pioneer factor interactions and unmethylated CpG dinucleotides mark silent tissue-specific enhancers in embryonic stem cells. *Proc Natl Acad Sci USA* **104:** 12377–12382

Xu J, Watts JA, Pope SD, Gadue P, Kamps M, Plath K, Zaret KS, Smale ST (2009) Transcriptional competence and the active marking of tissue-specific enhancers by defined transcription factors in embryonic and induced pluripotent stem cells. *Genes Dev* **23:** 2824–2838

