H3K4 tri-methylation provides an epigenetic signature of active enhancers

Aleksandra Pekowska1,2,3,4,6,7, Touati Benoukraf1,2,3,4,6,8, Joaquin Zacarias-Cabeza1,2,3,4,6, Mohamed Belhocine1,2,3,4, Frederic Koch1,2,3,4, Helène Holota4,5, Jean Imbert4,5, Jean-Christophe Andraud1,2,3,4, Pierre Ferrier1,2,3,4,* and Salvatore Spicuglia1,2,3,4,*

1Centre d’Immunologie de Marseille-Luminy (CIML), Parc Scientifique de Luminy, Case 906, Marseille, France, 2CNRS, UMR6102, Marseille, France, 3Inserm, U631, Marseille, France, 4Université de la Méditerranée, Marseille, France and 5Inserm, 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 H3K4me3/5 at distal genomic regions correlated with, respectively, the induction or the repression of associated genes during T-cell development. In the TCR 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 functional activity of these elements (Bernstein et al, 2007; Mikkelsen et al, 2007, 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).

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 H3K4me1, 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 p7α 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 Tαr 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<sup>−/−</sup>; hereafter ARag) and from ARag 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 extend our study to the genomewide 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,
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; Creighton 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., 1998); (ii) three 5' distal DNase hypersensitive sites (DHS1-3) that synergistically activate the expression of the Dnmt gene (encoding Terminal Deoxynucleotidyl Transferase or TdT) in T cells (Cerrier et al., 2008); (iii) two DHS sites (C6 and C7) located in the intronic region of the Il2f1 gene (coding for the transcription factor IRAROS), 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 Cdsb/b1 and Cd4 genes, as well as on the initiation of V-to-J recombination at the Tcrα 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 Cdsd and Cd4 loci (Figure 2). The Cdsd locus harbours several enhancers (named E8I to E8V), which are located either within the body of the Cdsb/b1 gene or in the intergenic region between the Cdsb/b1 and the Cdsd genes, all of which are required for the proper expression of Cdsd genes in DP thymocytes (Taniuchi et al., 2004 and references therein). Strikingly, all Cdsd 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, Tcrα 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 Tcrα 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 Iggl/Vpreb1, Il28ra, and Bcl7a loci were therefore examined. Remarkably, at these three loci, we likewise observed a gain in H3K4me2/me3 enrichment in the distal regulatory elements specifically in pro-B cells (Supplementary Figure S4). By contrast, the Cdsd 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
endogenous Tcrb gene enhancer (Eβ) has been replaced by a conserved core region (Eβ169, Figure 3A). Eβ169-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 Eβ169 (ΔRag;Eβ169 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;Eβ169 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 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 levels. 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
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 H3K4me3 peaks were detected in both ΔRag and ΔRagCD3 thymocytes were selected (these regions will henceforth be referred to as H3K4me3 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 trimethylation 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
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.
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. We used the T cell line P5424, which is derived from DRag thymocytes but expresses 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).

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 Tcrα and b loci provide good examples in support of this scenario. We and others have previously established that Ez is pre-assembled in the form of an inactive nucleoprotein complex in ΔRag thymocytes (Hernandez-Munain et al., 1999; Spicuglia et al., 2000). Indeed, Ez is associated with H3K4me1 at this early stage. Likewise, the truncated Eβ169 enhancer was found to be associated with higher H3K4me1 and lower H3K4me3 levels than wild-type Eβ (Figure 3). Yet, Eβ169 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 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 Cdx-intergenic region and surrounding Tca 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 S4A; 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/Flt1 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**

Rag-2<sup>−/−</sup> (ARag) (Shinkai et al., 1992) and ARag;E<sup>β</sup><sup>+/−</sup>(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 CD3c (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 RPMI medium supplemented with either 50 μM of KM05283 (Maybridge, Bridgewell, 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 ARag, ARag;E<sup>β</sup> and ARagCD3 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) for 20 s 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<sup>8</sup> 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 (so called ‘anti-ChIP’ mouse monoclonal antibodies) from Santa Cruz Biotechnology), with minor modifications. Briefly, 1 ml of chromatin extract corresponding to 10 × 10<sup>6</sup> 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<sup>6</sup> of either ARag or ARagCD3 thymocytes were digested with 10 U MNase (Roche, Switzerland) for 10 min and the soluble chromatin (mononucleosomes) from the equivalent of 5 × 10<sup>6</sup> 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

**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 ARagCD3 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<sup>6</sup>) were incubated in RPMI medium supplemented with either 50 μM of KM05283 (Maybridge, Bridgewell, 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).
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, ChIPs were prepared as described above and qPCR analyses of two biologically independent samples were carried out in triplicates. 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 200 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 tiling array at a resolution of 100 bp was used, covering the following loci: 5′-Fla, 7′-Fla, 5′-Act, 7′-Act and 5′-Acet. 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 Cy5-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 KMO5283-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 pre-processing 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/geo/) under accession numbers GSM546518, GSM546519, GSM546520, GSM546527, GSM546528 and GSM546529, which correspond to the ChIP-Seq profiles of H3K4me1, H3K4me2 and H3K4me3 in EBV knockout mice and in Kagi 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). For ChIP-on-chip analysis, ChIP samples, as well as 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.

Repoter 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× PBS and the luciferase assay (Bright-Glo, Promega) was performed according to the manufacturer’s protocol. 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/criml.

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 advices on data analysis; the Nice-Sophia-Antipolis platform for gene expression experiments; A Bergon (TAGC, Marseille) for help with Solid DNA sequencing; JZC performed the biological experiments; HH and JI performed the study; TB, JZC and SS designed the DNA microarray; AP and MB performed the luciferase reporter assay; AP, TB, MB and SS performed the bioinformatic analyses; AP, PB and SS wrote the manuscript. All authors reviewed the manuscript.

Conflict of interest
The authors declare that they have no conflict of interest.

References
Epigenetic signature of active enhancers

A Pekowska et al


