Transcription-Dependent Generation of a Specialized Chromatin Structure at the TCR β Locus

Joaquin Zacarías-Cabeza, Mohamed Belhocine, Laurent Vanhille, Pierre Cauchy, Frederic Koch, Aleksandra Pekowska, Romain Fenouil, Aurélie Bergon, Marta Gut, Ivo Gut, Dirk Eick, Jean Imbert, Pierre Ferrier, Jean-Christophe Andrau and Salvatore Spicuglia

J Immunol 2015; 194:3432-3443; Prepublished online 2 March 2015;

doi: 10.4049/jimmunol.1400789
http://www.jimmunol.org/content/194/7/3432
Transcription-Dependent Generation of a Specialized Chromatin Structure at the TCRβ Locus

Joaquin Zacarías-Cabeza,*,†,‡,†† Mohamed Belhocine,§,†,‡ Laurent Vanbille,§,† Pierre Cauchy,*,†,‡,†† Frederic Koch,*,†,‡ Aleksandra Pekowska,*,†,‡ Romain Fenouil,*,†,‡ Aurélie Bergon,§,†,‡ Marta Gut,*, Ivo Gut,*, Dirk Eick,*,‡ Jean Imbert,§,†,‡ Pierre Ferrier,*,†,‡,‖ Jean-Christophe Andrau,*,†,‡,‖ and Salvatore Spicuglia*,†,§,‖

V(D)J recombination assembles Ag receptor genes during lymphocyte development. Enhancers at AR loci are known to control V(D)J recombination at associated alleles, in part by increasing chromatin accessibility of the locus, to allow the recombination machinery to gain access to its chromosomal substrates. However, whether there is a specific mechanism to induce chromatin accessibility at AR loci is still unclear. In this article, we highlight a specialized epigenetic marking characterized by high and extended H3K4me3 levels throughout the Dβ–Jβ–Cβ gene segments. We show that extended H3K4 trimethylation at the Tcrb locus depends on RNA polymerase II (Pol II)–mediated transcription. Furthermore, we found that the genomic regions encompassing the two DJCβ clusters are highly enriched for Ser5-phosphorylated Pol II and short-RNA transcripts, two hallmarks of transcription initiation and early transcription. Of interest, these features are shared with few other tissue-specific genes. We propose that the entire DJCβ regions behave as transcription “initiation” platforms, therefore linking a specialized mechanism of Pol II transcription with extended H3K4 trimethylation and highly accessible Dβ and Jβ gene segments. The Journal of Immunology, 2015, 194: 3432–3443.

V (D)J recombination assembles Ag receptor genes from germline V, D, and J segments during lymphocyte development (1). In αβT cells, this leads to the subsequent expression of TCR β- and α-chains. For V(D)J recombination to occur, the presence of the lymphoid-specific proteins RAG1 and RAG2 and the ubiquitously expressed DNA repair factors from the nonhomologous end joining pathway are required (2). Control of V(D)J recombination is needed to ensure cell lineage specificity, dictate the temporal order of rearrangements, and allow allelic exclusion at certain AR genes (3). This regulation mainly relies on the modulation of chromatin accessibility at the AR-
(DJ) clusters is an initial activation event at all AR loci that generates a focal zone of RAG1/2 binding, termed the recombination center (6, 7). More insight into the accessibility model was provided by recent studies demonstrating that the PHD finger domain of RAG2 binds with high affinity to histone H3 trimethylated at K4 (H3K4me3) and that RAG2 is recruited to H3K4me3 domains genome-wide (6–8).

A central prediction of the accessibility model is, therefore, that transcriptional control elements and/or transcription itself are critical for allowing the recombination machinery to gain access to RSs (9). However, in most mammals’ genes, highly open chromatin structure is mainly confined to the cis-regulatory sequences themselves (10). In particular, H3K4me3 is highly enriched at promoter regions of expressed genes but is not generally found in the body of the genes (11). Thus, the question still remains as to how chromatin accessibility is established at the recombining gene segments and associated RSs, which are often located distant from the cis-regulatory elements. We and others have recently shown that a subset of tissue-specific genes might display broad epigenetic marking, including extended H3K4me2 and H3K4me3, along with elevated histone modification marks, we used 2×10^6 cells along with 3 μg of the following Abs: anti-H3K4me1 (ab8895; Abcam, Cambridge, U.K.), anti-H3K4me2 (ab32356; Abcam), anti-H3K4me3 (ab8580; Abcam), and anti-H3K36me3 (ab9050; Abcam). For Pol II ChIPs, the following Abs and cell numbers were used: anti–total-Pol II (Santa Cruz Biotechnology, Dallas, TX; sc-899 Abs) and anti–Ser 5P Pol II [rat monoclonal, clone 3E8 (32); 10 μg and 10×10^6 cells], anti-Ser2 Pol II [rat monoclonal, clone 3E8 (32); 10 μg and 30×10^6 cells]. The DNA fragments were purified and recovered using the QIAquick PCR Purification Kit (QIAGEN, Hilden, Germany). The quality of individual ChIP samples was checked at known target sites by quantitative PCR (qPCR), and DNA size was verified on a 2100 Bioanalyzer (Agilent, Santa Clara, CA). Primer sets used for qPCR are available upon request.

Formaldehyde-assisted isolation of regulatory elements

Formaldehyde-assisted isolation of regulatory elements (FAIRE) was performed as previously described (33), with slight modifications. Briefly, 20×10^6 thymocytes from ΔRag or ΔRagΔEβ mice were cross-linked with 1% formaldehyde for 10 min at room temperature and sonicated 14 times on an S-4000 Sonifier (Misonix, Farmingdale, NY) with 30-s pulses to give DNA fragments of length between 200 and 500 bp. The soluble chromatin of 2×10^6 thymocytes was isolated and subjected to three consecutive phenol-chloroform extractions. Samples were then incubated overnight at 65˚C to reverse cross-linking. DNA was finally purified using the MinElute PCR Purification Kit (QIAGEN). DNA concentration was measured using a Nanodrop 1000 (Thermo Scientific, Illkirch, France).

ChIP-seq data generation

Sequencing of ChIP samples was performed according to the Illumina Genome Analyzer ChIP-seq protocol and aligned against the mouse mm9 genome using integrated Eland software. As preprocessing steps, only uniquely mapped tags were used for further processing, and all duplicate tags (those with identical coordinates) were filtered out to remove possible sequencing and/or alignment artifacts. Remaining tags were processed using a custom R pipeline, employing the ShortRead library (14). Read-count intensity profiles (wiggle files) were constructed by elongating each mapped read to the estimated fragment size, and counting the elongated read overlaps within a window of 50 nucleotides after normalization of the profile by the number of mapped reads. ChIP-seq data from total-Pol II and from microcosomal-nucleus–treated H3K4me1 and H3K4me3 from ΔRag thymocytes were published previously (Ref. 34; GSE55635). Mapped reads, estimated fragment size, and Gene Expression Omnibus (GEO) accession numbers are listed in Supplemental Table I.
RNA extraction and RNA-seq experiments

Total RNA from $10 \times 10^6$ thymocytes of ΔRag mice was extracted as previously described (14). Strand-specific preparation, sequencing, and processing of short-RNA samples were carried out as explained earlier (14). RNA quantity and quality were verified using RNA Pico chips on a 2100 Bioanalyzer (Agilent). Mapped reads and GEO accession numbers are listed in Supplemental Table I. Total and polyA RNA-seq data from ΔRag thymocytes were published previously (Ref. 35; GSE44578).

ChIP-seq and RNA-seq data analyses

We first selected non-overlapping genes, harboring a single transcript annotated in the RefSeq database and longer than 8 kb (Supplemental Table II). From this set, the 300 highest expressed genes (Top-300) were selected, based on gene expression data in ΔRag thymocytes (34). To quantify the enrichment levels in H3K4me3, Ser5 of log2 ratio using IGB software (http://bioviz.org/igb/). Normalized data were displayed in the form of log2 ratio versus gene rank.

Analysis of ChIP-on-chip and FAIRE experiments

Enriched DNA fragments from ChIP or FAIRE experiments were hybridized together with input DNA to a previously described 15K array (Agilent) containing the whole Tcrb locus at 100-bp resolution (34), following the manufacturer’s instructions. The results obtained with two biological replicates were averaged and converted into SGR files using CoCAS software (37). Data from ΔRag and ΔRagΔEβ thymocytes were normalized using the overall signal on the entire microarray (excluding the transcriptional start site (TSS)) to +8 kb. In the case of the DJC1, DJC2, D2J61, J1Cy1, and J4Cy4 clusters, the region from the DJB1, DJB2, DJB3, JCy1, and JCy2 gene segments to +8 kb, respectively, was used to quantify H3K4me3, Ser5 Pol II, and short-RNAs within the gene body, the ChIP-seq signal from wiggle files was quantified within the region from the transcriptional start site (TSS) to +8 kb. In the case of the DJC1, DJC2, D2J61, J1Cy1, and J4Cy4 clusters, the region from the DJB1, DJB2, DJB3, JCy1, and JCy2 gene segments to +8 kb, respectively, was used to quantify H3K4me3, Ser5 Pol II, and short-RNAs. To directly compare expression levels between the selected RefSeq genes and the different Tcr loci gene clusters, we used polyA RNA-seq data. PolyA RNA level was estimated by counting the average number of tags at the exons of RefSeq genes and at the different Tcr gene clusters. All quantifications are shown in Supplemental Table II. The pausing index (also called traveling ratio) was calculated as previously described (36), using the selection of the Top-500 genes.

Inflection point

We first calculated the average signal of H3K4me3 in the gene body (TSS to +8kb) for each gene. The broad H3K4me3 genes were then determined by identifying an inflection point of the average signal versus gene rank. The inflection point was computed by determining the diagonal line of the curve from endpoints, and by sliding this diagonal line to find where it is tangential. We identified 58 broad H3K4me3 genes (Supplemental Table II).

Transcription initiation platform selection

We selected promoter-associated transcription initiation platforms (TIPs), defined previously in DP thymocytes (14) and expressed in the P5424 cell line (671 TIPs). The TIPs were separated according to their size into three clusters, the region from the D and J loci. Normalized data were displayed in the form of log2 ratio using IGB software (http://bioviz.org/igb/).

CpG density

The genomic sequences ±500 bp around the TSS of each set of genes were recovered. The total number of CpG was counted for each sequence. Statistical significance between the Top-300 and the other set of genes was calculated using a Student t test.

Results

A highly open chromatin structure at the DJCβ region

To assess epigenetic features associated with chromatin remodeling of the Tcrb locus, we analyzed the three levels of histone H3K4 methylation by ChIP of thymocytes purified from Rag2-deficient mice (hereafter ΔRag), followed by high-throughput sequencing (ChIP-seq). The use of the ΔRag mouse model ensures the germline configuration of Tcrb alleles while providing an enriched and homogeneous source of T cell precursors. We concentrated our analyses on the Eβ-proximal region, including the two DJCβ clusters (Fig. 1A). We observed that H3K4 methylation marks were not exclusively localized to the known regulatory regions (i.e., the pDβ promoters and Eβ) but, instead, extended throughout the Jβ and Cβ regions. For instance, H3K4me1 and H3K4me2 covered the entire Eβ-proximal region spanning 30 kb from ~3 kb upstream of DJB1 to ~3 kb downstream of Vβ31, thus defining a domain of open chromatin that roughly corresponds to the previously described Eβ-regulated domain (22–26, 41). Intriguingly, however, H3K4me3, which has been shown to be highly enriched at promoter regions (11), but is not generally found in the body of the genes, was broadly distributed throughout the two DJCβ germline transcription units. To exclude any potential bias owing to cross-linked chromatin, we confirmed the extended profile observed for H3K4me1 and H3K4me2 at the DJCβ regions by analyzing ChIP-seq data performed with mononucleosome preparations of native chromatin from ΔRag thymocytes (34) (Fig. 1B). Moreover, Eβ-deleted alleles displayed an almost complete loss of H3K4me3 at the DJCβ regions, suggesting that this epigenetic marking depends on Eβ-mediated transcriptional activation of the locus (Fig. 1C).

We then asked whether the extended H3K4me3 profile observed at the DJCβ regions in ΔRag thymocytes could be reminiscent of a highly open chromatin. To directly determine the accessibility of the chromatin, we performed a FAIRE assay, which allowed the recovery of the soluble (i.e., nucleosome-free) fraction of the chromatin (33). As expected, FAIRE signals were highly enriched at the Eβ region in ΔRag thymocytes (Fig. 1C). In addition, we observed that regions overlapping the Dβ and Jβ gene segments also display high levels of FAIRE signal in ΔRag. We confirmed that the highly open chromatin revealed by FAIRE at the Dβ and Jβ gene segments was largely dependent on Eβ-mediated chromatin remodeling (Fig. 1C). These results confirm and extend previous observations describing extensive Eβ-dependent remodeling of the DJCβ clusters (3, 16, 22, 23, 25–27). Note, however, that residual levels of chromatin accessibility are still observed around the Dβ
gene segments in the absence of Eβ, in agreement with an Eβ-independent role of Dβ-associated promoters (23, 25). Overall, in ΔRag thymocytes, highly accessible chromatin domains at the 3′ proximal region of the Tcrb locus are not restricted to the enhancer and promoter elements, but are spread over the Dβ and Jβ gene segments, thus providing a unique chromatin signature.
The extended H3K4me3 profile is a specific feature of the Tcrb locus

To determine whether this extended profile was a general feature of highly expressed genes, we compared the H3K4me3 profiles at the two DJCβ clusters with the average H3K4me3 profiles of a set of highly expressed genes (Top-300; Fig. 1D). As predicted, expressed genes displayed an H3K4me3 enrichment around the TSS (peaks at −0.5 and +1 kb from the TSS). In comparison, the H3K4me3 profiles at the DJCβ regions extended throughout the transcribed regions with no particular enrichment at the 5′ sides. Moreover, we found that H3K4me3 levels were 3- to 4-fold higher at the DJCβ regions than the level observed around the TSS of highly expressed genes (Fig. 1D). To directly compare the H3K4me3 enrichment within the gene body of individual genes, we calculated the density of H3K4me3 at the two DJCβ clusters and within the genomic regions from the TSS to +8 kb of mRNA genes. We next plotted the H3K4me3 values in the function of mRNA levels, obtained by polyA RNA-seq (see Materials and Methods). As shown in Fig. 1E, the two DJCβ clusters displayed very high levels of H3K4me3 as compared with the rest of the genes, whereas the mRNA level of the two clusters was relatively modest. We observed that a relatively small subset of genes also displayed elevated H3K4me3 enrichment (Fig. 1E). Genes ranked in function of H3K4me3 level identified 59 genes harboring substantially higher levels of H3K4me3 (Fig. 1F; see Materials and Methods for details). These genes displayed a broad distribution of H3K4me3 within the 5′ regions of the gene body (Fig. 1D; hereafter named Broad-H3K4me3 genes), as observed for the Tcrb locus, and reminiscent of previous findings of genes associated with extended H3K4 methylation (12, 13). However, the two DJCβ clusters ranked within the top 10 of the highest H3K4me3-enriched genes in ΔRag thymocytes (Fig. 1F). Thus, the active DJCβ clusters display an unusual H3K4me3-extended chromatin structure that is larger and stronger than the one observed at the vast majority of expressed genes, without being associated with a high level of polyadenylated RNA.

Pol II–dependent chromatin remodeling

The above results raise the question of whether a specialized transcription mechanism plays a key role at the Tcrb locus, which ultimately leads to a highly accessible chromatin structure at the DJβ and JB gene segments. Chromatin accessibility at the AR loci has been generally associated with the presence of germline transcription (9). Moreover, H3K4me3 marking across the Jα segments of the Tcra locus has been shown to directly depend on germline transcription (42). More generally, functional links have been described between Pol II binding and H3K4 trimethylation at promoter regions (43, 44). Thus, we asked whether the atypical H3K4me3 profiles observed at the DJCβ regions may depend on Pol II–mediated transcription. To this end, elongating Pol II was blocked by inhibiting the CDK9 kinase with the KM05283 chemical compound (31, 34). We reasoned that if H3K4 trimethylation depends on local Pol II transcription, then its level was likely to decrease following the KM05283 treatment. In these experiments we used the pro-T cell line P5424, which is derived from Rag thymocytes and harbors a recombination-competent Tcrb locus (45). Efficient blocking of Pol II elongation upon KM05283 treatment was validated by global loss of phosphorylated serine 2 of the C-terminal domain (CTD) of Pol II (Ser2P Pol II), as assessed by Western blot (Fig. 2A), as well as complete loss at the Tcrb region of H3K36me3, a mark of transcription elongation (46; Fig. 2B). Next, we performed ChIP-seq experiments for H3K4me3 and total-Pol II, using both KM05283- and DMSO-treated chromatin. Interestingly, we observed a strong decrease in H3K4me3 within the two DJCβ transcription units in KM05283-treated P5424 cells (Fig. 2C; note that the H3K4me3 profile at the Tcrb locus was consistent between the P5424 cell line and ΔRag thymocytes). Thus, H3K4 trimethylation at the DJCβ regions is largely dependent on Pol II transcription.

In most expressed genes, transcription initiation and elongation are regulated independently. Indeed, inhibition of transcriptional elongation normally results in the loss of Pol II within the gene body and its accumulation at promoter regions (36). This result was confirmed by increased Pol-pausing index in the KM05283-treated cells (Fig. 2D) and exemplified by the average profiles of the set of 300 highly expressed genes (Fig. 2E, left panel), as well as visual inspection of several expressed genes (Fig. 2F). Strikingly, however, we observed a complete loss of Pol II binding at the two DJCβ regions after KM05283 treatment (Fig. 2C, 2E, right panel). The specific loss of Pol II binding at DJβ promoters, but not at control genes, upon inhibition of transcription elongation was further confirmed by independent ChIP-qPCR experiments (Fig. 2G). Thus, in the absence of transcription elongation, Pol II was unable to remain stably associated to the DJβ promoters, as is the case for the vast majority of expressed genes. These results suggest that, at the Tcrb locus, recruitment of Pol II is directly coupled to the elongation phase of transcription.

TIPs cover the DJCβ transcription units

The above results suggest that 1) the broad distribution of H3K4me3 (and likely chromatin accessibility) is linked to Pol II–mediated transcription and 2) the regulation of the transcription process might differ between Tcrb locus and canonical mRNA coding genes. Pol II transcriptional activity is regulated via phosphorylation of the CTD (47). At expressed genes, phosphorylation of Ser5 of the CTD, which is associated with transcription initiation and early elongation, is found at the 5′ end of genes, whereas Ser7, which is required for productive elongation, is found to be enriched at the 3′ end of genes (43, 44). Moreover, Pol II phosphorylation at Ser5 has been shown to be required for H3K4me3 trimethylation (43, 44). To explore whether the Tcrb locus displays a distinctive Pol II profile, we first analyzed the distribution of phosphorylated and total forms of Pol II at the DJCβ1 cluster by ChIP-qPCR from ΔRag thymocytes. The expected patterns of Pol II phosphorylation were fully reproduced at control active genes Actb and Sfrs3: we found high levels of Ser5P at the TSS, low levels of phosphorylated Pol II within the gene body, and high levels of Ser5P at the 3′ end of these genes (Fig. 3A). However, in the case of the Tcrb locus, we found relatively high levels of Ser5P Pol II throughout the DJCβ1 region, whereas the Ser5P Pol II accumulated at the 3′ end of the DJCβ1 transcription unit (Fig. 3A). Indeed, although Ser5P Pol II downstream of the TSS of control genes is reduced to background levels, the enrichment at equivalent regions of the DJCβ1 cluster remains elevated.

To have a more comprehensive view of Pol II profiles at the Tcrb locus, we performed ChIP-seq experiments for both total- and Ser5P Pol II in ΔRag thymocytes. Again, we observed an accumulation of total- and Ser5P Pol II across the two DJCβ regions (Fig. 3B), whereas control genes displayed the expected patterns (Fig. 3C). Note that total- and Ser5P Pol II profiles were consistent between the ChIP-qPCR and ChIP-seq data (compare Figs. 3A with 3B, 3C). A more thorough analysis revealed that the Ser5P Pol II profiles were quantitatively and qualitatively different between the set of highly expressed genes and the DJCβ regions (Fig. 4A). Indeed, the level of Ser5P Pol II at the DJCβ regions was higher than the majority of expressed genes (Fig. 4B; the DJCβ clusters ranked in the top three of the highest Ser5P Pol II–enriched genes in ΔRag thymocytes). Thus, Pol II is found in its initiating/early elongating form throughout the entire DJCβ transcription units.
A hallmark of transcription initiation in higher eukaryotes is the presence of bidirectional short-RNAs around the TSS (hereafter short-RNA), a feature related to Pol II pausing (48). Given the above results, we hypothesized that the DJC\(\beta\) regions might be enriched in initiating short transcripts. To explore this possibility, we performed short-RNA–seq experiments from ΔRag thymocytes and compared them with strand-specific total (ribosomal-depleted) and polyA RNA-seq profiles previously generated (35). As expected, total and polyA RNA-seq signals overlapped with the DJC\(\beta\) regions and were oriented in the sense of defined transcription units (Fig. 3C). The continuous RNA-seq signal observed at the DJC\(\beta\) regions probably reflects a low splicing efficiency at this locus. Analysis of short-RNA–seq data revealed the presence of several discrete peaks of short transcripts, along with an overall enrichment of this RNA population throughout the entire DJC\(\beta\) regions, suggesting that
FIGURE 3. The DJCβ regions are enriched for transcription initiation features. (A) ChIP-qPCR assays showing the relative enrichment of total (N20), Ser2 and Ser5 phosphorylated Pol II at two active genes, at the indicated locations of the DJCβ region and at a negative control region (NC1). The genomic location of primer sets with respect to the DJβ1 gene segment or control genes is highlighted in (B) and (C). (B) Profiles of total and initiating (Ser5P) Pol II ChIP-seq experiments, as well as total polyA and short-RNAs from directional RNA-seq experiments in ΔRag thymocytes, are shown at the 3’ region of the Tcrb locus. For RNA-seq the profiles are log2 scaled, and strand orientation is indicated at the left of each panel. Other data are as in Fig. 1. (C) The ChIP-Seq profiles of total and Ser5P Pol II in ΔRag thymocytes are shown at three active genes.
Pol II pausing occurs at different places downstream of the DJB promoters (Fig. 4C, 4D). This was a specific feature of the Tcrb locus, as the overall distribution of short-RNAs was clearly different between the DJC/DJB regions and the set of highly expressed genes, for which bidirectional short-RNAs accumulate around the TSS (Fig. 4C, 4D). Previously we have identified TIPs (14), which are large genomic regions associated with Ser5P Pol II and TBP. TIPs were also associated with high levels of H3K4me3. The Tcrb might represent an extreme example of these genomic features. We concluded that the entire DJC/DJB regions behave as transcription initiating and early elongating platforms, thus providing a direct link between Pol II–mediated chromatin remodeling and H3K4 trimethylation at the DJB/JB recombination segments.

**Shared features between Tcrb and Broad-H3K4me3 genes**

As mentioned above, a small subset of genes was found to be associated with broad H4K4me3 marking (Fig. 1D–F). These
genes also displayed significant enrichment of Ser5P Pol II and, to a lesser extent, short initiating transcripts (Fig. 4E, 4F, Supplemental Fig. 1A, 1B). In general, genes with high levels of H3K4me3 also displayed high levels of Ser5 Pol II (Supplemental Fig. 1C). Therefore, a small subset of genes with broad H3K4me3 marking also displays features of transcriptional initiation in ΔRag

FIGURE 5. Characterization of Broad-H3K4me3 genes. (A) Enrichment scores of top 10 biological processes found enriched in the indicated gene sets. (B) Expression dynamic of Top-300 and Broad-H3K4me3 genes during T cell differentiation. Cell populations are as defined by the Immunological Genome Project and are ordered from less to most differentiated stages: ETP, early thymic progenitors; DN 1–4, ISP, immature single positive; and DPL, DP Blast, respectively. The inset shows the differential expression between the DN1-to-DN2 and DN3-to-DN4 transitions. (C) Boxplot showing the mean expression distribution in T cell samples and 69 other samples (see Materials and Methods for details). (D) Boxplot showing the distribution of the KM05283/DMSO ratio of Pol II binding at the promoters (±2 kb around the TSS) of the Top-300, Broad-H3K4me3, and TIPs-associated genes, as well as size-selected subsets of TIPs. The KM05283/DMSO ratios of the two DJC regions are also indicated. Statistical significance in (B) through (D) was calculated by the Student t test. (E) Profiles of Pol II ChIP-seq from the P5424 cell line treated with KM05283 or DMSO at the Tcrδ locus and one Broad-H3K4me3 gene. The genomic location of primer sets used in Fig. 2G are indicated by dashed arrows. The p values are as follows: **p < 10^{-2}, ***p < 10^{-3}, ****p < 10^{-4}. Proc., process; reg., regulation.
thymocytes (although Tcrb might represent an extreme example of this phenomenon).

To gain further insight into the function of Broad-H3K4me3 genes, we analyzed the functional enrichment of the biological process and found that they were specifically enriched on T cell– and immune-related functions, whereas the set of Top-300 genes were enriched for metabolic processes (Fig. 5A). Indeed, the list of Broad-H3K4me3 genes include many genes known to be involved in T cell differentiation and signaling, such as Lef1, Il2ra, Themis, Ifngr1, Fyb, RhoH, and Cd274 (Supplemental Table II). Accordingly, the set of Broad-H3K4me3 genes was highly tissue specific (Fig. 5B). Although these genes were expressed at relatively low levels in primary thymocytes, their expression was highly regulated during early T cell differentiation (Fig. 5C), namely, between DN1-to-DN2 and DN3-to-DN4 cell transitions (Fig. 4C, insets). Thus, the subset of Broad-H3K4me3 genes is reminiscent of the Tcrb locus, as they represent highly regulated genes involved in T cell function. They might represent extreme examples of genes with broad H3K4 methylation patterns described previously by us and others (12–14). To assess whether other AR genes could share the same features as the Tcrb, we analyzed, in a similar way, gene segments of the Tcrd and Tcrg locus, which are the two other AR loci in an open chromatin configuration in ΔRag thymocytes (see Materials and Methods for details). We found that gene segments from Tcrd (spanning D62-J61 gene segments) and Tcrg (Jγ1-Cγ1 and Jγ4-Cγ4 gene segments) loci also displayed high levels of H3K4me3, Ser5P Pol II, and short initiating transcripts to a similar extent as those observed for the Tcrb locus (Supplemental Figs. 1C, 2), thus suggesting that large initiating platforms might be a general feature of AR loci.

Finally, we asked whether Pol II binding at Broad-H3K4me3 genes was also highly sensitive to transcription elongation, as observed for the Tcrb locus. Quantification of Pol II levels around the TSS of the Top-300 and Broad H3K4me3 genes in P542 cells treated with either DMSO or KM05283 demonstrated that Pol II binding is specifically lost at BroadH4K4me3 genes, although not to the same extent as observed around the Dβ gene segments. This finding was evidenced at several genes, including the Tcrd and Ifngr1 loci (Fig. 5E), and validated by independent ChIP-qPCR (Fig. 2G; note that the Tcrb locus could not be analyzed, as this gene was found to be inactive in the P542 cell line; data not shown). To determine whether this phenomenon was a general property of TIPs, we analyzed our previously defined selection of TIPs-associated genes in DP thymocytes (14), excluding the genes that were not expressed in the P542 cell line (see Materials and Methods for details). As a group, the TIPs-associated genes did not display a loss of Pol II binding at their promoters after inhibition of Pol II elongation (Fig. 5D). However, when TIPs were classified according to their size, we found that genes associated with large TIPs (>2.5 kb) significantly lost Pol II binding at their promoters (Fig. 5D). We concluded that a subset of Broad-H3K4me3 genes and large TIPs-associated genes display regulatory features similar to those of the Tcrb locus, including tissue-specific gene expression, the presence of a TIP, and coupled Pol II recruitment and elongation (Fig. 6).

**Discussion**

Previous work from our laboratory and other laboratories has shown a remarkable open chromatin structure encompassing the Dβ-Jβ recombination center, including chromatin accessibility and histone marking (3, 16, 22, 26–28). More specifically, H3K4me3 was found to be enriched at Dβ and Jβ segments using ChIP-qPCR (7). Similar extended H3K4me3 patterns have been shown across the Jα segments of the Tcra locus (42). In the current study, we extend these findings by showing that the distribution of H3K4 trimethylation over the DβJβ regions is both quantitatively and qualitatively different from that in the vast majority of expressed genes (Fig. 1). Although H3K4me3 generally accumulates within 2kb around the TSS of genes (11), we observed that H3K4me3 enrichment at the DJCB clusters is much broader, spanning ±8 kb downstream of the germline Dβ promoters, and including all Jβ gene segments. Moreover, the level of H3K4 trimethylation found at the DJCB clusters was exceptionally high, representing one of the most enriched domains in developing thymocytes. We show that this extended profile depends on an unusual Pol II regulation process. In the case of canonical genes, Pol II accumulates around the TSS in its initiating form (high Ser5P Pol II), which correlates with high enrichment of H3K4me3 and the presence of short-RNA transcripts. However, in the case of the Tcrb locus, the entire DJCB regions display features of transcription initiation and Pol II pausing, including high levels of Ser5P Pol II and short-RNAs. Unexpectedly, inhibition of Pol II elongation resulted in complete loss of Pol II across the DJCB clusters (Fig. 2). To our knowledge, this is the first example in mammals whereby Pol II accumulation at the promoter is strictly dependent on transcription elongation. Remarkably, this phenomenon was also observed at the Tcrd locus (Figs. 2G, 5E).

We propose that a high level of initiating Pol II throughout the entire DJCB regions targets the H3K4 histone methyltransferases, resulting in an unusual extended H3K4me3 profile, and ultimately...
leads to a highly accessible chromatin structure around the Dβ and Jβ gene segments (Fig. 6).

We have previously shown that tissue-specific genes expressed in T cells generally display high levels of H3K4 methylation within the 5' region of the gene body (12). Along the same line, a recent study has shown that H3K4me3 domains that spread more broadly over genes in a given cell type preferentially mark genes that are essential for the identity and function of that cell type (13). Besides, we also described TIPs at proximal and distal sites, which were characterized by the presence of Ser5P Pol II, TBP, and epigenetic marks H3K4me1 and H3K4me3 (14). In this article, we show that genes with broad H3K4me3 domains display features related to large initiation platforms (including accumulation of Ser5P Pol II and short initiating transcripts) similar to the TIP genomic domains. However, TIP domains as defined previously in DP thymocytes (14) display a wide range of size, varying from 0.45 kb to 10 kb (80% of TIPs are <2 kb). Whether broad H3K4me3 and TIPs define the same type of genes remains to be precisely investigated, but our results suggest that common features are shared by both types of structures. Genes marked by the broadest H3K4me3 domains exhibit enhanced transcriptional consistency rather than increased transcriptional levels (13). Moreover, Pol II accumulation at the promoter of Broad-H3K4me3 genes tends to be dependent on transcription elongation, a phenomenon also observed at the promoters of genes associated with large TIPs (Fig. 5D). Thus, it is likely that the broad H3K4me3 domains defined in this article (in particular, those found at the Tcr loci) might represent a subset of larger TIPs. Indeed, larger TIPs also have a tendency to be more tissue specific (14). All in all, our results suggest the existence of a specialized transcriptional regulation mechanism restricted to a subset of tissue-specific genes. In this context, the Tcrb locus might represent an extreme example of this phenomenon. Our finding has implications not only for regulatory strategies used by AR loci but also for the epigenetic mechanisms that control gene expression of cell identity genes.

Are intrinsic genomic features responsible for the highly open and H3K4me3-enriched chromatin structure observed at the Tcrb locus? In mammals, Pol II accumulation and enrichment for active histone marks at promoters are generally linked to their high CpG content (49). We have previously shown that TIPs overlapped with CpG density, although larger TIPs displayed lower or more disperse CpG content (14). Consistently, we found that promoters of Broad-H3K4me3 and large TIPs-associated genes display significantly lower CpG density as compared with the set of Top-300 genes ($p < 0.01$ and $p < 0.0001$, respectively; Student t test; see Materials and Methods). The DJCβ regions do not contain any CpG island and also display relatively low G and C nucleotide content (data not shown). It is, therefore, plausible that in the absence of CpG islands, the Pol II molecules recruited at the Dβ associated promoters are immediately engaged in the elongation process while still harboring the transcription initiation mark (i.e., Ser5P) and therefore remain associated with H3K4 methyltransferases (43, 44)(Fig. 6). This hypothesis would be consistent with the complete loss of Pol II at the Tcrb locus after inhibition of transcription elongation (Fig. 2). Another intriguing, but not mutually exclusive, possibility is that the extended H3K4me3 profile is related to the unusual structure of the Tcrb locus, which contains several J segments, each harboring a 5' splicing site. A recent study has shown that H3K4 trimethylation at the 5' border of mammalian genes is directly linked to the length of the first exon of genes (average size is 250 nt) (50). However, in the case of the DJCβ transcription units, the first splicing donors are located at the end of each Jβ segment, ranging between 641 nt and 2.5 kb from the Dβ segments, which make the first exons considerably longer than the average size. Moreover, the Jβ-associated splicing sites appear to be relatively inefficient, as judged by the high level of RNA-seq signal observed downstream of the Jβ gene segments (Fig. 3B). As described previously (50), the first exon length >500 nt results in a flat H3K4me3 profile extending to the 3' end of the first exon, as well as increasing Pol II pausing, both features reminiscent of what is observed at the Tcrb locus. Thus, it is plausible that the location of Jβ gene segments, each behaving as a first exon, will result in the distinctive chromatin structure observed at the DJCβ clusters.

It has been recently demonstrated that RAG1 and RAG2 bind in vivo to focal regions, termed “recombination centers,” covering mainly the J segments of AR genes and within which V(D)J recombination has been suggested to take place (7). The formation of these recombination centers depends on the AR enhancers and promoters (6), and correlates with the presence of H3K4me3 (7). Thus, given the specific requirements for chromatin accessibility and H3K4me3 enrichment at J segments to ensure efficient V(D)J recombination (3, 5), we propose that the Tcrb locus (and likely other AR loci) has evolved in such a way that a specialized regulation of the transcription process confers a unique long-range epigenetic marking, ultimately allowing the establishment of a highly accessible chromatin structure at the recombining Dβ/Jβ gene segments.

Data access

ChIP-seq and RNA-seq data obtained in this study have been submitted to the National Center for Biotechnology Information’s GEO (http://www.ncbi.nlm.nih.gov/geo) under the following accession numbers: GSE63416 (www.ncbi.nlm.nih.gov/projects/geo/query/acc.cgi?acc=GSE63416), GSE46709 (www.ncbi.nlm.nih.gov/projects/geo/query/acc.cgi?acc=GSE46709), and from GSE1360722 to GSE1360727 and GSE1359828 (www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE56395). Details are available in Supplemental Table I.

Acknowledgments

We thank Dr. Eugene Oltz (Washington University, St. Louis, MO) for donating the P5424 cell line.

Disclosures

The authors have no financial conflicts of interest.

References


