MLL4 prepares the enhancer landscape for Foxp3 induction via chromatin looping

Katarzyna Placek1,6, Gangqing Hu1,6, Kairong Cui1,6, Dunfang Zhang2, Yi Ding1,3, Ji-Eun Lee4, Younghoon Jang4, Chaochen Wang1,4, Joanne Elizabeth Konkel2, Jiuzhou Song3, Chengyu Liu5, Kai Ge4, Wanjun Chen2 & Keji Zhao1

MLL4 is an essential subunit of the histone H3 Lys4 (H3K4)-methylation complexes. We found that MLL4 deficiency compromised the development of regulatory T cells (Treg cells) and resulted in a substantial decrease in monomethylated H3K4 (H3K4me1) and chromatin interaction at putative gene enhancers, a considerable portion of which were not direct targets of MLL4 but were enhancers that interacted with MLL4-bound sites. The decrease in H3K4me1 and chromatin interaction at the enhancers not bound by MLL4 correlated with MLL4 binding at distant interacting regions. Deletion of an upstream MLL4-binding site diminished the abundance of H3K4me1 at the regulatory elements of the gene encoding the transcription factor Foxp3, which is necessary for suppressive activity. We found that MLL4 catalyzed methylation of H3K4 at distant unbound enhancers via chromatin looping, which identifies a previously unknown mechanism for regulating the T cell enhancer landscape and affecting Treg cell differentiation.

Active and primed enhancers are characterized by the presence of permissive histone modifications such as histone acetylation and monomethylation of histone H3 Lys4 (H3K4me1). The activating histone marks facilitate the opening of chromatin and the recruitment of transcription factors and other regulatory machineries. The methylation of H3K4 is catalyzed by the MLL family of histone methyltransferases, which includes SETD1A, MLL1 (KMT2A)18, MLL2 (KMT2B), MLL3 (KMT2C) and MLL4 (KMT2D). MLL4 has been shown to shape enhancer patterns in mammalian cells during heart development19, myogenesis and adipogenesis20 by regulating the mono- and dimethylation of H3K4.

We found that MLL4 served a critical requirement for Treg cell development by establishing the enhancer landscape and facilitating long-range chromatin interaction. We showed that in addition to regulating the monomethylation of H3K4 at direct binding sites, MLL4 catalyzed the methylation of H3K4 in trans at distant unbound enhancers via long-distance chromatin looping. This identifies a previously unrecognized mechanism for regulating the histone modification and enhancer landscape of cells.

RESULTS

Deletion of Mll4 results in compromised Treg cell development

To investigate the function of MLL4 in T cell development, we generated mice with conditional deficiency in MLL4 in CD4+ T cells.

---

1Systems Biology Center, Division of Intramural Research, NHLBI, NIH, Bethesda, Maryland, USA. 2Mucosal Immunology Section, Division of Intramural Research, NIDCR, NIH, Bethesda, Maryland, USA. 3Department of Animal and Avian Sciences University of Maryland, College Park, Maryland, USA. 4Adipocyte Biology and Gene Regulation Section, Laboratory of Endocrinology and Receptor Biology, NIDDK, NIH, Bethesda, Maryland, USA. 5These authors contributed equally to this work. Correspondence should be addressed to W.C. (wchen@dir.nidcr.nih.gov) or K.Z. (zhaok@nhlbi.nih.gov).

Received 30 March; accepted 10 July; published online 31 July 2017; doi:10.1038/ni.3812
by breeding mice that carry loxP sites flanking exons 16–19 of MLL4 (MLL4fl/fl mice) with mice that express Cre recombinease from the T cell–specific Cd4 promoter (Cd4-Cre+ mice). We analyzed MLL4fl/fl Cd4-Cre+, MLL4fl/floxed-Cd4-Cre*, MLL4lox/+Cd4-Cre* and MLL4lox/+Cd4-Cre+ mice as controls for the effect of the Cd4-Cre construct or partial deletion of Mll4 on mouse phenotypes. We confirmed efficient deletion of the loxP-flanked Mll4 exons in Cd4+ T cells isolated from MLL4fl/fl Cd4-Cre+ mice and their partial deletion in MLL4lox/+Cd4-Cre+ cells, by RT-PCR (Supplementary Fig. 1a), RNA-seq (Supplementary Fig. 1b) and immunoblot analysis (Supplementary Fig. 1c). Deletion of exons 16–19 results in a truncated MLL4 protein and disrupts the MLL4 complex. Because we confirmed that the Cd4-Cre transgene alone had no effect on the mouse phenotype and we did not observe significant differences among MLL4fl/fl Cd4-Cre+, MLL4fl/floxed-Cd4-Cre* and MLL4lox/+Cd4-Cre+ mice in their T cell populations (Fig. 1 and Supplementary Fig. 1d–g), we call all three of these strains of mice ‘wild-type’ here (unless stated otherwise) and call MLL4fl/fl Cd4-Cre+ mice ‘Mll4-KO’ here.

Conditional deletion of Mll4 had no significant effect on T cell development in the thymus, as populations of CD4+CD8+ double-positive cells, CD4+ single-positive (CD4SP) cells and CD8+ single-positive (CD8SP) cells remained similar in all groups of mice examined (Fig. 1a,b); however, it substantially decreased the frequency and total number of CD4+Foxp3+ Treg cells in the thymus of Mll4-KO mice compared with the abundance of such cells in their wild-type littermates (Fig. 1c,d). Deletion of Mll4 also significantly reduced the number of CD4+ and CD8+ T cells in secondary lymphoid organs, including the spleen (Fig. 1e,f) and lymph nodes (Supplementary Fig. 1e,g).

Although the frequency of Foxp3+ cells within the CD4+ T cell population in spleen and lymph nodes was not substantially affected in Mll4-KO mice (Fig. 1g and Supplementary Fig. 1f), the total number of CD4+Foxp3+ T cells was reduced (Fig. 1h and Supplementary Fig. 1g) as a result of the decrease in the total CD4+ T cell population in these organs. Consistent with the preserved ratio of Treg cells to Teff cells after deletion of Mll4, we did not see an increase in the number of T cells producing interferon-γ (IFN-γ), interleukin 17A

Figure 1  Mll4 deficiency reduces the number of Treg cells in the thymus and the number of T cells in the periphery. (a) Flow cytometry identifying CD4SP, CD8SP, CD4+CD8- and CD4+CD8+ T cell populations in the thymus of MLL4fl/fl Cd4-Cre-, MLL4fl/floxed-Cd4-Cre*, MLL4lox/+Cd4-Cre* and MLL4lox/+Cd4-Cre+ mice (above plots). Numbers adjacent to outlined areas indicate percent cells in each throughout. (b) Frequency of CD4+CD8+ (double-positive (DP)), CD4SP and CD8SP T cells among total cells in the thymus of mice as in a (below plots). (c) Flow cytometry identifying CD4SP Foxp3+ cells in the thymus of mice as in a (above plots). (d) Frequency of CD4SP Foxp3+ cells in the thymus of mice as in a (below plots). (e) Flow cytometry identifying CD4+ and CD8+ T cells in the spleen of mice as in a (above plots). (f) Quantification of total CD4+ T cells (top) and CD8+ T cells (bottom) in the spleen of mice as in a (below plots). (g) Flow cytometry identifying CD4+Foxp3+ cells in the spleen of mice as in a (above plots). (h) Quantification of total CD4+Foxp3+ cells in the spleen of mice as in a (below plots). Each symbol (b,d,f,h) represents an individual mouse; small horizontal lines indicate the mean (±s.d.) ***P ≤ 0.01, ****P ≤ 0.001 and ****P ≤ 0.0001 (Kruskal-Wallis test). Data are from one experiment representative of n = 4 (a,c) or n = 5 (e.g) independent experiments or are representative of three experiments (b,d,f,h).
Figure 2. Deletion of Mll4 results in impaired T<sub>reg</sub> cell development in the gut and in vitro culture. (a) Flow cytometry identifying CD4<sup>+</sup>Foxp3<sup>+</sup> cells in the lamina propria of the small intestine of Mll4<sup>fl/fl</sup>Cd4-Cre<sup>+</sup> (WT) and Mll4-KO mice (above plots). (b) Frequency of CD4<sup>+</sup>Foxp3<sup>+</sup> cells in the lamina propria of the small intestine of mice as in a (below plots). (c) Flow cytometry identifying IFN-γ and IL-17A<sup>+</sup> cells among gated CD4<sup>+</sup> T cells from the lamina propria of the small intestine of mice as in a (above plots). (d) Frequency of IFN-γ<sup>+</sup> cells (left), IL-17A<sup>+</sup> cells (middle) and IFN-γ–IL-17A<sup>+</sup> cells (right) among CD4<sup>+</sup> T cells in small intestine lamina propria of wild-type and Mll4-KO mice (below plots). (e) RT-PCR analysis of Foxp3 mRNA in naive CD4<sup>+</sup> T cells isolated from Mll4<sup>fl/fl</sup>Cd4-Cre<sup>+</sup> (WT) and Mll4-KO mice (key) (Naive) and cells obtained from such mice and differentiated for 2–72 h (horizontal axis) under T<sub>reg</sub>-cell–inducing conditions (iT<sub>reg</sub>); results are presented relative to those of 18S ribosomal RNA and are normalized to those of wild-type naive cells. (f) Flow-cytometry analysis of the intracellular staining of Foxp3 and CD4<sup>+</sup> cells within the CD4<sup>+</sup> T cell population or aberrant cytokine production by T cells in the lungs of Mll4<sup>+/−</sup>-KO mice compared with the wild-type counterparts (data not shown), which suggested that MLL4 was not required for T<sub>reg</sub> cell development. An in vitro suppression assay demonstrated the normal functionality of T<sub>reg</sub> cells isolated from Mll4<sup>+/−</sup>Foxp3-Cre<sup>+</sup> mice (Supplementary Fig. 3j), which indicated that MLL4 had no effect on suppressive function in cells already expressing Foxp3. Together these data suggested that MLL4 regulated the development of T<sub>reg</sub> cells but not their function.

Induction of Foxp3 expression requires MLL4
To further understand the role of MLL4 in T<sub>reg</sub> cell development, we differentiated naive CD4<sup>+</sup> T cells from wild-type and Mll4-KO mice into T<sub>reg</sub> cells in vitro (to generate inducible T<sub>reg</sub> cells (iT<sub>reg</sub> cells)). We found that the expression of Foxp3 by Mll4-KO CD4<sup>+</sup> T cells was one fifth that of wild-type cells during differentiation into the iT<sub>reg</sub> cell lineage (Fig. 2e). At day 3 of differentiation, more than 70% of the CD4<sup>+</sup> T cells from wild-type mice expressed Foxp3, while only 20% of the CD4<sup>+</sup> T cells from Mll4-KO mice expressed Foxp3 (Fig. 2f). Our results obtained by staining with the division-tracking dye CFSE indicated that Mll4-KO CD4<sup>+</sup> T cells displayed a delay in proliferation and induction of Foxp3 expression compared with that of wild-type cells at early times; however, by day 4 of in vitro culture, both the proliferation and Foxp3 expression of Mll4-KO CD4<sup>+</sup> T cells were similar to that of wild-type cells (Supplementary Fig. 4a), which indicated that the compromised Foxp3 expression was not caused by a defect in cell proliferation. These observations suggested a direct role for MLL4 in the induction of Foxp3 expression.

To investigate whether MLL4 has a role in maintaining Foxp3 expression, we first stimulated naive CD4<sup>+</sup> T cells from Mll4<sup>+/−</sup> mice for 24 h under T<sub>reg</sub> cell–inducing conditions and then infected the cells for 2 more days with retroviral vector expressing Cre recombinase or retrovirus expressing an empty vector. RT-PCR analysis confirmed the deletion of Mll4 exons 16–19 by the
retrovirus-expressed Cre recombinase (Fig. 3a). We observed no difference in the number of Foxp3+ cells in cultures infected with retroviral vector expressing Cre relative to that in cultures infected with retroviral expressing empty vector (Fig. 3b); this suggested that MLL4 was required only for the initiation of Foxp3 expression but was not required for its maintenance. Furthermore, MLL4 expression was much higher in naive CD4+ T cells than in effector cells of the Tq1, Tq2 and Tq17 subsets of helper T cells and iTreg cells induced in vitro (Fig. 3c,d). MLL4 expression in thymic Treg cells was comparable to that in CD4SP cells in the thymus (Fig. 3e). We also investigated whether deletion of Mll4 affected the differentiation of other effector T cells, including Tq1, Tq2 and Tq17 cells, under in vitro differentiation conditions. Only very modest changes were observed in expression of the transcription factor T-bet and IFN-γ under Tq1 conditions (Supplementary Fig. 4b,c) and in the expression of IL-17A and the transcription factor RORγ under Tq17 conditions (Supplementary Fig. 4d,e), while expression of the transcription factor GATA-3 and IL-4 was increased in Tq2 cell cultures after deletion of Mll4 (Supplementary Fig. 4f,g). These observations indicated that MLL4 was dispensable for differentiation of the Tq1, Tq2 and Tq17 cell lineages in vitro. Together these data indicated that MLL4 was required for the induction but not the maintenance of Foxp3 expression during iTreg cell development and had only minimal function in other effector T cells.

**MLL4 regulates the H3K4me1 landscape in naive CD4+ T cells**

MLL4 has been reported to regulate mainly the mono- and dimethylation of H3K4 (refs. 19,20,23), although some early reports suggested that MLL4 also regulates its trimethylation24,25. We found that deletion of Mll4 in T cells decreased the mono- or trimethylation of H3K4 only modestly, while we detected no substantial decrease in the dimethylation of H3K4 after deletion of Mll4 (Supplementary Fig. 5a); this suggested that other H3K4-methylation enzymes were present in the cells. We found that the gene encoding MLL3, which is a close homolog of MLL4 and has functions partially redundant with those of MLL4 (ref. 20), exhibited an expression pattern similar to that of Mll4 in CD4+ T cells (Supplementary Fig. 5b). Because naive CD4+ T cells expressed significantly more MLL4 than did iTreg cells and because the deletion of Mll4 at later stages of iTreg cell differentiation in vitro had no effect on Foxp3 expression, we analyzed the MLL4-binding and H3K4me1-enrichment profiles of naive CD4+ T cells by chromatin immunoprecipitation followed by deep sequencing (ChIP-Seq). Analysis of ChIP-Seq peaks by the SICER clustering algorithm26 revealed that approximately 50% of the 16,161 MLL4 peaks were located in promoter regions of genes and the other half mapped to the gene body and intergenic regions called ‘non-promoter peaks’ (Supplementary Fig. 5c). MLL4 bound predominantly to genomic regions that showed enrichment for H3K4me1, as exemplified by a 350-kilobase (kb) genomic region in the X chromosome (Fig. 4a). However, a large number of H3K4me1 peaks were not direct targets of MLL4 (Fig. 4a). Globally, while MLL4 bound to two thirds of the regions showing enrichment for H3K4me1 that overlapped the promoters, it was detected at less than 20% of the H3K4me1 peaks at other genomic regions (Fig. 4b). We relaxed the threshold by SICER26 to call MLL4-binding sites step by step until the additionally called peaks from MLL4 ChIP-Seq were indistinguishable from the input control (Supplementary Fig. 5d). Even with the most relaxed criteria, we still observed that 50–60% of the H3K4me1 peaks at non-promoter regions were not bound by MLL4 (Supplementary Fig. 5e).

Non-promoter genomic regions that showed enrichment for H3K4me1 (‘H3K4me1-enriched’) but were devoid of MLL4 binding (‘MLL4-unbound’) were linked to genes encoding products involved in T cell activation, as assessed by gene-ontology–enrichment analysis with the tool GREAT27 (Supplementary Fig. 5f), which suggested that these elements were involved in T cell differentiation and function. Genomic sequence analysis revealed that these sites were conserved in DNA sequence across mammalian species (Supplementary Fig. 5g). Those H3K4me1-enriched MLL4-unbound regions were generally less accessible than MLL4-bound regions, as revealed by genome-wide sequencing of regions sensitive to DNase (DNase-Seq)28 (Fig. 4c and Supplementary Fig. 5h). These results indicated that the chromatin environment of the H3K4me1-enriched MLL4-unbound regions were unfavorable for direct binding of MLL4 and suggested that the modulation of H3K4me1 in these regions by MLL4 might be achieved through mechanisms other than direct binding.

To identify H3K4me1 peaks affected by deletion of Mll4, we compared the H3K4me1 profiles of wild-type CD4+ T cells with those of Mll4 KO CD4+ T cells. As a specific example, deletion of Mll4 induced a substantial increase in H3K4me1 at the promoter region of Dynlt1f (which encodes a dynemin light chain) and a decrease in H3K4me1 at putative enhancers bound by MLL4 (Supplementary Fig. 6a,b). Globally, we found that 13.5% of the MLL4-bound enhancers showed
a decrease in H3K4me1 after deletion of Mll4, whereas only 1.2% showed an increase in H3K4me1 (Supplementary Fig. 6c), in support of the proposal of a direct role for MLL4 in regulating H3K4me1 at its binding sites20. Notably, deletion of Mll4 also substantially decreased H3K4me1 at enhancers without MLL4 binding, as exemplified by several putative enhancers upstream of Dynlt1f (Supplementary Fig. 6a) and Cxcr4 (which encodes a chemokine receptor) (Fig. 4d). Of all the non-promoter regions that showed enrichment for H3K4me1 but were devoid of MLL4 binding, 14.8% exhibited a significant decrease in H3K4me1 after deletion of Mll4 (Supplementary Fig. 6c).

Analysis of data obtained by RNA-based next-generation sequencing (RNA-Seq) did not reveal any substantial change in the expression of other H3K4-methylation enzymes in Mll4-KO cells (Supplementary Table 1), which suggested that the decrease in H3K4me1 at enhancers that were not directly bound by MLL4 was most probably caused by loss of MLL4 activity but not by the loss of other H3K4-methylation enzymes that resulted from the deletion of Mll4.

To investigate the effect of the deletion of Mll4 on transcription programs, we identified 576 genes and 430 genes whose expression was downregulated and upregulated, respectively, in the MLL4-deficient cells (data not shown). Gene-ontology–enrichment analysis revealed that the set of downregulated genes showed enrichment for genes encoding molecules with functions including lymphocyte activation, regulation of signal transduction, cytokine secretion, compared with the set of genes without a substantial change in expression, while the set of upregulated genes exhibited no obvious enrichment for any ontology term (Supplementary Table 2). Together our data suggested that MLL4 regulated H3K4me1 at enhancers by direct and/or indirect binding and contributed to the regulation of gene expression in naïve CD4+ T cells.

**Mll4 regulates H3K4me1 via chromatin interaction**

The results reported above raised the possibility that monomethylation of H3K4 at MLL4-unbound regions might be catalyzed in trans by MLL4 bound at different sites via chromatin looping. To test this hypothesis, we analyzed genome-wide chromatin interactions in naïve wild-type and Mll4-KO CD4+ T cells using the Hi-C chromosome-conformation–capture technique29 (statistical analysis of Hi-C libraries, Supplementary Table 3). We defined a conservative set of H3K4me1-enriched regions that were not bound by MLL4 by calling...
Figure 5 MLL4 facilitates long-range chromatin interactions. (a) ChIP-Seq signals for MLL4 in wild-type cells (top row; based on data in Fig. 4a) and interaction intensities in the genomic regions surrounding Cxcr4 (left) and Foxp3 (right) in wild-type and Mll4-KO cells (rows below) (UCSC Genome Browser image); regions showing a decrease in chromatin interaction intensity are outlined in red. (b) Frequency of bins with a decrease or increase in interaction intensity (key) after deletion of Mll4, for 2-kb genomic bins sorted on the basis of their MLL4 binding: none (Non-target), not bound but interacting with distant MLL4-binding sites (Indirect target), or directly bound (Direct target). *P < 0.01 (χ² test). (c) Frequency of bins with a decrease or increase (key) in interaction intensity after deletion of Mll4, for MLL4-bound promoter genomic bins sorted into four groups of equal size on the basis of the extent (wedge) of MLL4 binding (left), and for non-promoter genomic bins not bound by MLL4 but interacting with MLL4-bound genomic regions, sorted on the basis of the number of interacting regions (right). (d) Accumulative distribution of the change in the number of PETs linking an MLL4-unbound H3K4me1-enriched region to distant interacting regions bound by MLL4 after deletion of Mll4, sorted on the basis of their responses to deletion of Mll4 (decrease, increase or no change in H3K4me1; key). P = 0.017 (increase versus no change) and P < 2.2 × 10⁻¹⁶ (decrease versus no change) (Kolmogorov-Smirnov test). Data are pooled from n = 3 (wild-type) or n = 2 (Mll4-KO) independent experiments (error bars (c), s.d.).

MLL4 peaks using non-stringent thresholds with SICER. On the other hand, we identified, with high confidence, a set of MLL4-bound regions using stringent thresholds with SICER. We further identified interacting regions from Hi-C pair-end tags (PETs) by comparing those with a background model generated through simulation that considered GC content, mappability and distance. Significant interactions were detected between the H3K4me1 peaks devoid of MLL4 binding and distant H3K4me1 peaks bound by MLL4, as exemplified by three putative enhancers upstream of the Cxcr4 locus (Fig. 4d). Approximately 8% of the MLL4-unbound H3K4me1 peaks were looped to distant MLL4-binding sites, which was fourfold higher than the expected frequency (Fig. 4e). The fraction of MLL4-unbound H3K4me1 peaks that showed a decrease in H3K4me1 after deletion of Mll4 was positively correlated with the number of interacting MLL4-binding sites, while the fraction that showed an increase in H3K4me1 was negatively correlated with it (Fig. 4f). Consistent with that, we found a greater decrease in H3K4me1 read density for MLL4-unbound enhancers that interacted with MLL4-binding sites than for those interacting with fewer or no MLL4-binding sites (Supplementary Fig. 6d). The enrichment for H3K4me1 at MLL4-unbound regions was positively associated with the aggregated MLL4 binding intensity from distinctly interacting sites bound by MLL4 (Fig. 4g). Furthermore, MLL4-unbound sites with the most aggregated MLL4-binding intensity from distinctly interacting sites showed the greatest decrease in H3K4me1 after deletion of Mll4 (Fig. 4h). Together these results supported the proposal that monomethylation of H3K4 at MLL4-unbound sites was catalyzed in trans by MLL4 bound to remote sites through chromatin looping (Fig. 4i). Therefore, we call the H3K4me1 peaks bound by MLL4’s direct MLL4 targets’ here and call the H3K4me1 peaks not bound by MLL4 but interacting with MLL4-bound regions 'indirect MLL4 targets’ here.

MLL4 facilitates chromatin interactions at MLL4 targets

To investigate whether MLL4 contributed to interactions between regulatory sites in naive CD4⁺ T cells, we compared the chromatin-interaction density at direct and indirect MLL4 targets in wild-type cells with that in Mll4-KO cells. Deletion of Mll4 in naive CD4⁺ T cells substantially compromised chromatin interactions at many genomic loci, as exemplified by the Cxcr4 and Foxp3 loci (Fig. 5a). To detect genome-wide changes, we divided the genome into 2-kb bins of equal size and assessed their change in interaction intensity with that of other genomic regions after deletion of Mll4. We found 23% of the direct MLL4 targets showed decreased interaction after deletion of Mll4, a proportion that was significantly greater than that of those showing increased interaction (4%) (Fig. 5b). Notably, we also found that a significantly greater fraction of the indirect MLL4 targets showed decreased interaction (10%) than increased interaction (5%) after deletion of Mll4 (Fig. 5b). In comparison, we observed no greater decrease than increase in interaction for the regions devoid of either direct or indirect MLL4 binding (Fig. 5b). The interaction decrease caused by deletion of Mll4 was associated with a decrease in the active histone mark H3K27ac (Supplementary Fig. 6e,f) and an increase in the repressive histone mark H3K27me3 (Supplementary Fig. 6e,g). Furthermore, the fraction of the direct MLL4 targets showing decreased interaction after deletion of Mll4 modestly correlated with MLL4 binding (Fig. 5c). Notably, the fraction of the indirect
Figure 5  MLL4 regulates enhancer–promoter interactions. (a) Distribution of H3K27ac ChIP-Seq signals (total reads) across active enhancers (defined as non-promoter genomic regions showing enrichment for H3K27ac signal) in naive CD4+ T cells; red vertical dashed line indicates the cut-off used to define super-enhancers. (b) Significance of the top five (most significant) gene-ontology terms for biological processes (below bars) for genes showing significant chromatin interactions with super-enhancers (Benjamini test). (c) Frequency of 2-kb genomic bins with no MLL4 binding, not bound by MLL4 but interacting with distant MLL4-binding sites, or bound directly by MLL4 (key; categories as in Fig. 5b), within super-enhancers (SE) or regular enhancers (RE) or not within any enhancer (OT) (horizontal axis). (d) Frequency of bins with a decrease or increase (key) in interaction intensity after deletion of Mll4, for 2-kb genomic bins located at enhancers defined as in c (horizontal axis). (e) Frequency of bins with a decrease in interaction intensity after deletion of Mll4, for 2-kb bins located within enhancers defined as in c (key), further sorted on the basis of their relationship to MLL4 binding as in c (horizontal axis). (f) Accumulative distribution of the interaction at Pr-SE or Pr-RE promoters or promoters showing no interaction with any enhancer (Ctrl) (key) in Mll4-KO cells relative to that in wild-type cells (Mll4-WT). P = 2.2 × 10^{-16} (Pr-SE versus Pr-RE, and Pr-RE versus Ctrl) (Kolmogorov-Smirnov test). (g) Accumulative distribution of the interacting PETs (in Mll4-WT cells relative to that in wild-type cells) with enhancers or with non-enhancer regions (Other sites) (key), for Pr-SE promoters (left) or Pr-RE promoters (right). P = 2.8 × 10^{-13} (left) and P < 2.2 × 10^{-16} (right) (Kolmogorov-Smirnov test). (h) Accumulative distribution of the gene expression in Mll4-WT cells relative to that in wild-type cells (horizontal axis) for promoters as in f (key). P = 2.5 × 10^{-8} (Pr-SE versus Pr-RE) and P < 2.2 × 10^{-16} (Pr-RE versus Ctrl) (Kolmogorov-Smirnov test). Data are from n = 1 experiment (a,b) or three experiments (c) to from two experiments (wild-type) or three experiments (Mll4-KO) (d-g; error bars (d,e), s.d.) or are pooled from two experiments (wilde-type) or three experiments (Mll4-KO) (d-g; error bars (d,e), s.d.) or are pooled from two experiments (wild-type).
Pr-SE promoters showed the most significant decrease in chromatin interactions compared with that of the other promoters (Fig. 6f and Supplementary Fig. 7a); we highlighted specific examples of super-enhancer targets for Cxcr4, Rac2 (which encodes the actin-remodeling GTPase Rac2) and Stk4 (which encodes the serine-threonine kinase Mst1) (Supplementary Fig. 7b). Furthermore, we found that deletion of MLL4 resulted in a greater decrease in Pr-SE and Pr-RE interactions than the decrease in interactions

Figure 7 Deletion of the MLL4-binding ~8.5 kb region decreases the abundance of H3K4me1 at Foxp3 regulatory elements and compromises T<sub>reg</sub> cell differentiation. (a) Distribution of MLL4 ChIP-Seq reads in wild-type CD4<sup>+</sup> T cells and of H3K4me1 ChIP-Seq reads in wild-type and MLL4-KO cells (UCSC genome browser image): blue rectangles, H3K4me1 peaks; red rectangles, MLL4 peaks; yellow, H3K4me1 peaks not bound by MLL4 but interacting with remote MLL4 binding sites by at least two PETs and showing a decrease in H3K4me1 abundance after deletion of MLL4; scissors, genomic region targeted for deletion by CRISPR; horizontal lines (bottom), H3K4me1-enriched regions linked to remote MLL4 peaks (>5 kb) by at least two PETs. (b) Distribution of ChIP-Seq reads from the Watson strand (positive values) or the Crick strand (negative values), at single-base-pair resolution, for MLL4 in wild-type cells and for H3K4me1 in wild-type and Foxp3<sup>−/−</sup> cells, around the MLL4-binding site targeted by CRISPR (scissors). (c) Distribution of H3K4me1 reads around the Foxp3 locus for control and Foxp3<sup>−/−</sup> cells (UCSC genome browser image), and H3K4me1-enriched regions linking to the MLL4-binding site by at least two PETs (Hi-C track, top): pink rectangles, genomic regions showing a significant decrease in H3K4me1 after CRISPR-mediated deletion (change in abundance, >1.5-fold; false-discovery rate, < 0.001). (d) Frequency of decreased H3K4me1 peaks after CRISPR-mediated deletion, for MLL4-unbound H3K4me1-enriched regions in chromosome X linked to the site targeted by CRISPR by at least two PETs (+) (10 regions) and those not linked to that site (−) (1,396 regions). $P = 0.006$ (γ$^2$ test). (e) RT-PCR analysis of Foxp3 in naive CD4<sup>+</sup> T cells from control and Foxp3<sup>−/−</sup> mice (key) and in T<sub>reg</sub> cells generated in vitro from those mice, assessed at days 1–3 of differentiation (horizontal axis). (f) Flow cytometry identifying Foxp3<sup>+</sup> cells among gated CD4<sup>+</sup> T cells at day 3 of differentiation into T<sub>reg</sub> cells as in e. Isotype, isotype-matched control antibody. (g) Staining intensity of Foxp3 in cells generated as in f, presented as geometric median fluorescent intensity (gMFI). (h) Flow cytometry identifying CD4<sup>+</sup>Foxp3<sup>+</sup> T cells among gated CD4<sup>+</sup> cells from the thymus of control mice ($n = 8$) and Foxp3<sup>−/−</sup> mice ($n = 8$). (i) Flow cytometry identifying CD4<sup>+</sup>Foxp3<sup>+</sup> T cells among gated CD4<sup>+</sup> cells from the spleen of control mice ($n = 7$) and Foxp3<sup>−/−</sup> mice ($n = 7$). (j) Flow cytometry identifying CD4<sup>+</sup>Foxp3<sup>+</sup> T cells among gated CD4<sup>+</sup> cells from the lymph nodes of control mice ($n = 8$) and Foxp3<sup>−/−</sup> mice ($n = 8$). (k) Frequency of Foxp3<sup>+</sup> CD4<sup>+</sup> T cells in the thymus of control and Foxp3<sup>−/−</sup> mice, gated as in h. (l) Frequency of Foxp3<sup>+</sup> CD4<sup>+</sup> T cells in the spleen of control and Foxp3<sup>−/−</sup> mice, gated as in i. (m) Frequency of Foxp3<sup>+</sup> CD4<sup>+</sup> T cells in the lymph nodes of control and Foxp3<sup>−/−</sup> mice, gated as in j. Each symbol (g–m) represents an individual mouse; small horizontal lines indicate the mean (±s.d.). Data are representative of $n = 2$ (a, top, and b–d) or $n = 3$ (a, bottom) independent experiments, are from one experiment representative of $n = 3$ independent experiments (e,f; error bars (e), s.d.), or are representative of three experiments (h–m).
between the same promoters with non-enhancer regions (Fig. 6g). In general, genes with the greatest decrease in interaction exhibited the greatest downregulation in expression in MLL4-deficient naive CD4+ T cells (Supplementary Fig. 7e). Consistent with that, genes with promoters showing fewer interactions with enhancers or super-enhancers after deletion of MLL4 exhibited a general decrease in expression (Fig. 6f,h and Supplementary Fig. 7d), which indicated that direct binding of MLL4 and/or interaction with an MLL4-bound region facilitated promoter–enhancer interactions with a role in regulating gene expression.

MLL4 regulates H3K4me1 of Foxp3 via chromatin looping

To further understand how MLL4-dependent H3K4me1 and chromatin interactions between promoters and enhancers regulated Treg cell development, we focused our analysis on Foxp3 (ref. 4), which is regulated by its promoter and several intrinsic enhancers1,4,31,32. We found neither the Foxp3 promoter nor its known enhancers exhibited MLL4 binding in naive CD4+ T cells (Fig. 7a). Instead, strong MLL4 binding was detected at region 8.5 kb upstream of the Foxp3 transcription start site (called the −8.5 kb region here) (Fig. 7a), which exhibited a general decrease in interaction with other genomic regions after deletion of MLL4 (data not shown). This MLL4-bound region interacted with the Foxp3 promoter and several H3K4me1-enriched regions not bound by MLL4, including the 3′ untranslated region and two regions 23 kb and 30 kb upstream of the Foxp3 transcription start site (Fig. 7a). Deletion of MLL4 resulted in a substantial decrease in H3K4me1 at the Foxp3 promoter and at three putative enhancers not bound by MLL4 (Fig. 7a), which suggested that binding of MLL4 at the −8.5 kb region regulated H3K4me1 at the Foxp3 promoter and several other putative enhancers through chromatin looping.

To determine whether binding of MLL4 at the −8.5 kb region regulated H3K4me1 at the Foxp3 promoter and other regions, we deleted this MLL4-binding site in mice by CRISPR technology33 (called 'Foxp3−MLL4BS' here). We found that the CRISPR-mediated deletion (Fig. 7b) compromised H3K4me1 at four of the ten H3K4me1 peaks that interacted with the MLL4-binding site, including the Foxp3 promoter, in naive Foxp3−MLL4BS CD4+ T cells (Fig. 7c). In contrast, a significantly lower fraction of H3K4me1 peaks that did not interact with the MLL4 site from the same chromosome showed a decrease in H3K4me1 in these cells (Fig. 7d).

To investigate whether MLL4-binding site contributed to Foxp3 expression, we isolated naive CD4+ T cells from Foxp3−MLL4BS mice and mice with an intact MLL4-binding site at the −8.5 kb region (called ‘control mice’ here) and stimulated them in iTreg cell–differentiating conditions. The induction of Foxp3 was compromised in cells from the Foxp3−MLL4BS mice, as shown for mRNA (Fig. 7e) and protein (Fig. 7f,g). To determine if the −8.5 kb region contributed to Treg cell development in vivo, we analyzed Treg cells in various organs of Foxp3−MLL4BS mice. Although the reduction in the number of Treg cells in thymus was not reproducibly observed (Fig. 7h,k), we found modest but significant reduction in the number of Foxp3+ CD4+ T cells in spleen and lymph nodes in Foxp3−MLL4BS mice compared with their abundance in the control mice (Fig. 7i,j,l,m). Treg cells in Foxp3−MLL4BS mice had a more activated phenotype than that of such cells from their control littersmates, as defined by increased expression of Helios (Supplementary Fig. 8a–c). The function of Treg cells in Foxp3−MLL4BS mice remained unaffected, as cytokine production by T cells in the periphery of Foxp3−MLL4BS mice was not greater than that of their counterparts in the control mice (Supplementary Fig. 8d–g), and an in vitro suppression assay showed that Foxp3−MLL4BS and control Treg cells had similar ability to suppress the proliferation of effector T cells (Supplementary Fig. 8h). CD4+CD25+Foxp3+ thymocytes are considered precursors of mature Treg cells34–36. We observed significantly more of these precursors of Treg cells (Fig. 8a–d) and a lower intensity of Foxp3 staining in Foxp3− thymic Treg cells in Foxp3−MLL4BS mice and Mll4−KO mice than in control mice and wild-type mice, respectively (Fig. 8b,d). Together these results suggested that binding of MLL4 at the −8.5 kb region regulated H3K4me1 at the Foxp3 promoter and contributed to the induction of Foxp3 under iTreg cell conditions and to the in vivo development of Treg cells.

**Figure 8** Deletion of the MLL4-binding −8.5 kb region increases the frequency of Treg precursor cells in the thymus. (a,c) Flow cytometry of gated CD3+CD4+CD8− cells from the thymus of MLL4−/mlld−/Cre+ and Mll4fl/flCd4−Cre+ mice (n = 3 per genotype) (a) or control and Foxp3−MLL4BS mice (n = 3 per genotype) (c), stained for CD25 and Foxp3. (b,d) Staining intensity of Foxp3 in CD3+CD4+CD8− CD25+Foxp3+ thymic Treg cells (Itreg) from mice as in a (b) or c (d) (left), and frequency of the CD3+CD4+CD8− CD25+Foxp3− cell population in the thymus of mice as in a (b) or c (d) (right). Each symbol (b,d) represents an individual mouse; small horizontal lines indicate the mean (±s.d.). Data are representative of three experiments.
DISCUSSION

T cell differentiation is associated with dynamic changes in histone modifications\(^{37}\) that are regulated by histone-modifying enzymes\(^{38,39}\), suggestive of an important role for these enzymes in the development and differentiation of T cells. One major H3K4-methylation enzyme is MLL4, which regulates mainly H3K4me1 at enhancers\(^{40}\). By conditional deletion of \(\text{Mll}^4\) in CD4\(^+\) T cells, we found that MLL4 was a critical regulator of \(T_\text{reg}\) cell development. However, deletion of \(\text{Mll}^4\) after the induction of Foxp3 did not alter the expression of Foxp3 in \emph{in vitro} or \emph{in vivo}. Furthermore, \emph{in vitro} T\(_{\text{reg}}\) cell suppression assays revealed no difference in the suppressive ability of MLL4-deficient \(T_\text{reg}\) cells relative to that of wild-type \(T_\text{reg}\) cells. Consistent with that observation, MLL4 expression was rapidly downregulated after \(T_\text{reg}\) cell differentiation \emph{in vitro}. These results suggested that MLL4 was required for the establishment of chromatin structure in naive CD4\(^+\) T cells for future differentiation into the \(T_\text{reg}\) cell lineage but was not required for either sustained Foxp3 expression or \(T_\text{reg}\) cell activity.

Although MLL4 might affect a cell population by regulating genes encoding products involved in the cell cycle\(^{19}\), our CFSE-staining experiments revealed that the proliferative capacity of MLL4-KO was similar to that of wild-type cells at day 4 of differentiation, which suggested that the developmental defect of \(T_\text{reg}\) cells was caused by a specific effect of MLL4 on Foxp3 expression. Because MLL4 was not required for the maintenance of Foxp3 expression, we hypothesized that MLL4 functions to establish H3K4me1-modification patterns at the regulatory elements critical for the induction of Foxp3. Among the regulatory elements for Foxp3 expression\(^{14,40}\), we found enrichment for H3K4me1 at the promoter region and CNS3 in naive CD4\(^+\) T cells, consistent with published studies\(^{14}\). Although MLL4 did not bind to any of those known Foxp3 regulatory elements, deletion of \(\text{Mll}^4\) compromised the H3K4me1 signals at the promoter and enhancer elements of Foxp3, which suggested that MLL4 might prime Foxp3 for activation under appropriate conditions by modulating the H3K4me1 signals at those regions in naive CD4\(^+\) T cells. Indeed, MLL4 might carry out this function by binding to the \(-8.5\) kb region (upstream of the Foxp3 transcription start site). That region was looped to the promoter, CNS3 and 3′ untranslated region of Foxp3. CRISPR-medi ated deletion of this region in \(\text{Foxp3}^{3\text{MLL}4BS}\) mice led to diminished H3K4me1 signals at the promoter and 3′ untranslated regions and resulted in a defect in the induction of Foxp3, which confirmed a regulatory role for this region for Foxp3 expression. Either deletion of \(\text{Mll}^4\) or CRISPR-mediated deletion of the MLL4-binding \(-8.5\) kb region led to a significant increase in the CD4\(^+\)CD25\(^-\)Foxp3\(^-\) precursors of \(T_\text{reg}\) cells, which provided further support for our hypothesis that MLL4 binds to the \(-8.5\) kb region to prepare the enhancer landscape required for the efficient induction of Foxp3 during \(T_\text{reg}\) cell development.

We found that only about half of the H3K4me1 peaks affected by deletion of \(\text{Mll}^4\) were bound by MLL4. Since no substantial changes in the expression of \(\text{Mll}3\) or other histone methyltransferases were detected in cells in which \(\text{Mll}^4\) was deleted, we hypothesized that MLL4 might bind to one site and catalyze the methylation of H3K4 at another remote site \emph{in trans} by chromatin looping. Our data confirmed that the affected H3K4me1 peaks not bound by MLL4 were looped to chromatin regions directly bound by MLL4 and that the decrease in H3K4me1 was positively correlated to the number of loops between the unbound sites and MLL4-bound sites. That hypothesis was further confirmed by deletion of the MLL4-bound \(-8.5\) kb region of Foxp3, which decreased H3K4me1 at the MLL4-unbound Foxp3 promoter and decreased the induction of Foxp3 under \(T_\text{reg}\) cell conditions. Thus, our data support the proposal of a mechanism in which, in addition to regulating H3K4me1 at direct binding sites, MLL4 also catalyzes \emph{in trans} the monomethylation of H3K4 at unbound chromatin regions via chromatin looping to establish an enhancer landscape for cellular differentiation. Although we demonstrated this kind of activity only for MLL4 here, it is likely that a similar mechanism might be used by other chromatin-modifying enzymes within the nucleus.

METHODS

Methods, including statements of data availability and any associated accession codes and references, are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We thank the NHLBI DNA Sequencing Core facility for sequencing the libraries; the NHLBI Systems Biology Core and the NIH Biowulf High Performance Computing Systems for computing services; the NHLBI Flow Cytometry Core facility for cell sorting; the NHLBI and NIAID Animal Facilities for animal care; and X. Wang and S Gao (NHLBI Systems Biology Core) and X. Zheng (Carnegie Institution for Science) for discussions on Hi-C data analysis. Supported by the Division of Intramural Research of NHLBI (K.Z.), NIDCR (W.C.) and NIDDK (K.G.) of the US National Institutes of Health.

AUTHOR CONTRIBUTIONS

K.P., G.H., K.C., C.L., W.C. and K.Z. designed experiments; K.P., G.H., K.C., D.Z., Y.D., J.-E.L., Y.J., C.W., C.L. and K.Z. conducted experiments; K.P., G.H., K.C., D.Z., Y.D., J.-E.L., Y.J., C.W., J.E.K., J.S., C.L., K.G., W.C. and K.Z. analyzed the data; and K.P., G.H., K.C., W.C. and K.Z. wrote the paper.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Reprints and permissions information is available online at http://www.nature.com/ reprints/index.htm. Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Sakaguchi, S., Yamaguchi, T., Nomura, T. & Ono, M. Regulatory T cells and immune tolerance. Cell 133, 775–787 (2008).
2. Klein, L. & Jovanovic, K. Regulatory T cell lineage commitment in the thymus. Semin. Immunol. 23, 401–409 (2011).
3. Chen, W. et al. Conversion of peripheral CD4\(^+\)CD25\(^-\) naive T cells to CD4\(^+\)CD25\(^+\) regulatory T cells by TGF-\(\beta\) induction of transcription factor Foxp3. J. Exp. Med. 198, 1875–1886 (2003).
4. Fontenot, J.D., Gavin, M.A. & Rudensky, A.Y. Effector and regulatory T-cell fate. Nature Rev. Immunol. 7, 57–63 (2007).
5. Lahl, K. et al. Selective depletion of Foxp3\(^+\) regulatory T cells induces a scurfy-like disease. J. Exp. Med. 204, 57–63 (2007).
6. Grindebacke, H. et al. Defective suppression of Th2 cytokines by CD4\(^+\)CD25\(^+\) regulatory T cells in birch allergic mice. Clin. Exp. Allergy 34, 1364–1372 (2004).
7. Kim, J.M., Rasmussen, J.P. & Rudensky, A.Y. Regulatory T cells prevent catastrophic autoimmunity throughout the lifespan of mice. Nat. Immunol. 8, 191–197 (2007).
8. Roychoudhuri, R., Eil, R.L. & Restifo, N.P. The interplay of effector and regulatory T cells, which provided further support for our hypothesis that MLL4 binds to the \(-8.5\) kb region to prepare the enhancer landscape required for the efficient induction of Foxp3 during \(T_\text{reg}\) cell development.
9. Note: Any Supplementary Information and Source Data files are available in the online version of the paper.
10. Discovers, S. & Rouse, B.T. Treg control of antimicrobial T cell responses. Curr. Opin. Immunol. 18, 344–348 (2006).
11. Raychaudhuri, R., Eil, R.L. & Restifo, N.P. The interplay of effector and regulatory T cells in cancer. Curr. Opin. Immunol. 33, 101–111 (2015).
12. Kasag, S. et al. In vivo-generated antigen-specific regulatory T cells treat autoimmunity without compromising antibacterial immune response. Sci. Transl. Med. 6, 241ra78 (2014).
13. Taylor, P.A., Lees, C.J. & Blazar, B.R. The infusion of ex vivo activated and expanded CD4\(^+\)CD25\(^+\) regulatory T cells induces a scurfy-like disease. Proc. Natl. Acad. Sci. USA 102, 3444–3449 (2005).
14. Zheng, Y. et al. Role of conserved non-coding DNA elements in the Foxp3 gene in regulatory T-cell fate. Nature 463, 808–812 (2010).
15. Schmidt, C. et al. The enhancer and promoter landscape of human regulatory and conventional T-cell subpopulations. Blood 123, 68–78 (2014).

© 2017 Nature America, Inc., part of Springer Nature. All rights reserved.
16. Samstein, R.M. et al. Foxp3 exploits a pre-existent enhancer landscape for regulatory T cell lineage specification. Cell 151, 153–166 (2012).
17. Hess, J.L. Mechanisms of transformation by MLL. Crit. Rev. Eukaryot. Gene Expr. 14, 235–254 (2004).
18. Ang, S.Y. et al. KMT2D regulates specific programs in heart development via histone H3 lysine 4 di-methylation. Development 143, 810–821 (2016).
19. Bannister, A.J. & Kouzarides, T. Regulation of chromatin by histone modifications. Cell Res. 21, 381–395 (2011).
20. Hess, J.L. Mechanisms of transformation by MLL. Crit. Rev. Eukaryot. Gene Expr. 14, 235–254 (2004).
21. Akimova, T., Beier, U.H., Wang, L., Levine, M.H. & Hancock, W.W. Helios expression is a marker of T cell activation and proliferation. PLoS One 6, e24226 (2011).
22. Rubtsov, Y.P. et al. Regulatory T cell-derived interleukin-10 limits inflammation at environmental interfaces. Immunity 28, 546–558 (2008).
23. Hu, D. et al. The MLL3/MLL4 branches of the COMPASS family function as major histone H3K4 monomethylases at enhancers. Mol. Cell. Biol. 33, 4745–4754 (2013).
24. Lee, J. et al. A tumor suppressive coactivator complex of p53 containing ASC-2 and histone H3-lysine-4 methyltransferase MLL3 or its parologue MLL4. Proc. Natl. Acad. Sci. USA 106, 8513–8518 (2009).
25. Lee, S., Lee, J., Lee, S.K. & Lee, J.W. Activating signal cointegrator-2 is an essential adaptor to recruit histone H3 lysine-4 methyltransferases MLL3 and MLL4 to the liver X receptors. Mol. Endocrinol. 22, 1312–1319 (2008).
26. Zang, C. et al. A clustering approach for identification of enriched domains from histone modification ChIP-Seq data. Bioinformatics 25, 1952–1958 (2009).
27. McLean, C.Y. et al. GREAT improves functional interpretation of cis-regulatory regions. Nat. Biotechnol. 28, 495–501 (2010).
28. Jin, W. et al. Genome-wide detection of DNase I hypersensitive sites in single cells and FFPE tissue samples. Nature 528, 142–146 (2015).
29. Rao, S.S. et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell 159, 1665–1680 (2014).
30. Hnisz, D. et al. Super-enhancers in the control of cell identity and disease. Cell 155, 934–947 (2013).
31. Feng, Y. et al. Control of the inheritance of regulatory T cell identity by a cis element in the Foxp3 locus. Cell 158, 749–763 (2014).
32. Ogawa, C. et al. TGF-β-mediated Foxp3 gene expression is cooperatively regulated by Stat5, Cebp, and AP-1 through CNS2. J. Immunol. 192, 475–483 (2014).
33. Wang, H. et al. One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. Cell 153, 910–918 (2013).
34. Burchill, M.A. et al. Linked T cell receptor and cytokine signaling govern the development of the regulatory T cell repertoire. Immunity 28, 112–121 (2008).
35. Kitagawa, Y. et al. Guidance of regulatory T cell development by Satb1-dependent super-enhancer establishment. Nat. Immunol. 18, 173–183 (2017).
36. Lio, C.W. & Hsieh, C.S. A two-step process for thymic regulatory T cell development. Immunity 28, 100–111 (2008).
37. Wei, G. et al. Global mapping of H3K4me3 and H3K27me3 reveals specificity and plasticity in lineage fate determination of differentiating CD4+ T cells. Immunity 30, 155–167 (2009).
38. Yamashita, M. et al. Crucial role of MLL for the maintenance of memory T helper type 2 cell responses. Immunity 24, 611–622 (2006).
39. Tumes, D.J. et al. The polycomb protein Ezh2 regulates differentiation and plasticity of CD4+ T helper type 1 and type 2 cells. Immunity 39, 819–832 (2013).
40. Li, X., Liang, Y., LeBlanc, M., Benner, C. & Zheng, Y. Function of a Foxp3 cis-element in protecting regulatory T cell identity. Cell 158, 734–748 (2014).
ONLINE METHODS

Mice. Mll4fl/fl mice on mixed C57BL/6 and 129 backgrounds were provided by K.G. and have been previously described. Cd4-Cre+ mice on the C57BL/6 background were purchased from Taconic. Fosp3-CreYFP mice on the C57BL/6 background were purchased from Jackson Laboratories. Male and female mice were bred and maintained in an NHLBI specific-pathogen-free animal facility. All experiments were performed on 6- to 10-week-old mice in accordance with a protocol approved by the NHLBI Animal Care and Use Committee.

Cell isolation and in vitro culture. Cd4+ T cells were purified from lymph nodes and spleen of Mll4fl/+Cd4-Cre-, Mll4fl/+Cd4-Cre-, Mll4fl/+Cd4-Cre- and Mll4fl/-Cd4-Cre- mice by magnetic selection according to the protocol provided by the manufacturer (Miltenyi Biotech, CD4+ T Cell Isolation Kit, mouse). Naïve Cd4+ T cells were purified from the lymph nodes of Mll4fl/+Cd4-Cre-, Mll4fl/+Cd4-Cre-, Mll4fl/+Cd4-Cre- and Mll4fl/-Cd4-Cre- mice, and control and Fosp3353EL4GFP mice, by magnetic selection according to the protocol provided by the manufacturer (Miltenyi Biotech, CD4+CD26+ T cell isolation kit II, mouse or Stem Cell, EasySep Mouse CD4+CD26+ T cell Isolation Kit) or were sorted by flow cytometry on a FACS Aura II cell analyzer (BD Biosciences) II cell analyzer (BD Biosciences) for CD4+CD8−/CD62L−/CD45R−T cells after pre-enrichment for CD4+ T cells by negative selection using MACS technology according to manufacturer’s instructions (as described above). The purity of naïve Cd4+ T cells was assessed by flow cytometry for CD4+, CD8− and CD62L− on a FACS Canto II (BD Biosciences) and were analyzed in FlowJo software. Over 98% purity of CD4+CD8−CD62L− cells were considered for further experiments. In vivo Treg cells were purified from lymph nodes and spleen of Mll4fl/+Cd4-Cre- and Mll4fl/+Cd4-Cre- mice by flow-cytometry sorting on a FACS Aria II cell analyzer (BD Biosciences) for CD4+CD8−CD62− after pre-enrichment for CD4+ T cells by negative selection using MACS technology (as described above) in accordance with the manufacturer’s instructions. All antibodies used for sorting and flow-cytometry analyses were purchased from eBiosciences: anti-mouse CD4 (RM4−5), anti-mouse CD8 (53−6.7), anti-mouse CD45 (30−F11, eBioscience). Cells were then washed and stained with anti-mouse Helios (Clone: 22F6, BioLegend), anti-mouse CD4 (RM4−5), anti-mouse CD8 (53−6.7), anti-mouse CD28 (Clone: 37.51, BioLegend) and flow-cytometry analyses were purchased from eBiosciences: anti-mouse CD62L (Clone: R4−6A2, 10 µg/ml, BioXcell), anti-IL-12 (Clone: R1−5D9, 10 µg/ml, BioXcell), anti-IL-12− (Clone: R1−5D9, 10 µg/mL, BioXcell), IL−2 (Cat# 212−12, 100 U/ml, Peprotech) and TGF−β (Cat# 240−B.5/5 mg/ml, R&D Systems). Cells were stained for surface markers with anti-mouse CD4 (RM4−5, eBioscience) and anti-mouse CD8 (53−6.7, eBioscience).

In vitro Treg cell differentiation and flow cytometry. Naïve Cd4+ T cells were incubated in Treg cell–differentiation condition in the presence of the following: plate-bound anti-CD3 (Clone: 145−2C11, 2 µg/ml, eBioscience), anti-CD28 (Clone: 37.51, 5 µg/ml, eBioscience), anti-IL−4 (10 µg/ml, BioXcell), anti-IFN−γ (Clone: R4−6A2, 10 µg/ml, BioXcell), anti-IL−12 (Clone: R1−5D9, 10 µg/mL, BioXcell), IL−2 (Cat# 212−12, 100 U/ml, Peprotech) and TGF−β (Cat# 240−B.5/5 mg/ml, R&D Systems). Cells were cultured for surface markers with anti-mouse CD4 (RM4−5, eBioscience) and anti-mouse CD8 (53−6.7, eBioscience) Intracellular staining was performed using Cytofix/Cytperm buffer (eBioscience) to fix and permeabilize cells. Cells were then washed and stained with following antibodies: anti-mouse Helios (22F6, BioLegend), anti-mouse Ki67 (B56, BD Biosciences) and anti-mouse Foxp3 (FKJ−16s, eBioscience). Dead cells were excluded from analysis using a Zombie Y ellow Fixable Viability Kit (Biolegend) or a LIVE/DEAD Fixable Aqua Dead Cell Stain Kit (Life Technologies).

In vitro deletion of Mll4. A Cre-carrying, green fluorescent protein (GFP)-expressing retroviral vector (pCre-GFP) was used for deletion of Mll4 from Mll4fl/fl cells. The retroviral pellets were packaged in 293T cells with the pEco packaging plasmid. Naïve Mll4fl/+ Cd4+ T cells were cultured in Treg cell–differentiation conditions for 24 h. Cells were infected with fresh retrovirus expressing the empty vector (pRV-GFP) or Cre-carrying pCreeGFP on the second day of culture. Cells were cultured in Treg cell–differentiating conditions for additional 2 d. Then GFP+ cells were sorted by flow cytometry on a FACS Aria II cell sorter (BD Biosciences) and were analyzed for Foxp3 expression. The efficiency of transduction was 60−90%.

In vitro Treg cell–suppression assay. The in vitro suppression assay was performed as described previously. In brief, Cd4+CD25+ Treg cells from the lymph nodes and spleen of wild-type and Mll4-KO mice and Cd4+CD25− effector T cells (Treg cells) from wild-type mice were isolated by magnetic selection according to the protocol provided by the manufacturer (Miltenyi Biotech, CD4+ T Cell Isolation Kit, mouse) followed by flow-cytometry sorting on a FACS Aria II (BD Biosciences) with following antibodies: anti-mouse CD4 (RM4−5, eBioscience) and anti-mouse CD25 (7D4, BD Biosciences). Prior to stimulation, Cd4+CD25+ Treg cells were stained with CFSE according to the manufacturer’s instructions (ThermoFisher Scientific, CellTrace Violet (Invitrogen). Double-stranded cdna was sonicated on Bioruptor (Diagenode), blunt-ended using an End-It DNA Repair kit (Epitcine), indexed, amplified and sequenced on an Illumina 2G Genome Analyzer.

Intracellular cytokine staining. IL−17A and IFN−γ cytokine expression was assessed by flow cytometry as described previously. In brief, the cells were stimulated for 4 h in phorbol 1,2-myristate 1,3-acetate (5 ng/ml) and ionomycin (1 µg/ml) in the presence of the protein−transport inhibitor GolgiPlug (BD Pharamingen). Cells were stained for surface markers with following antibodies: anti-mouse TCRβ (H57−597, eBioscience), anti-mouse CD4 (RM4−5, eBioscience) and anti-mouse CD45 (30−F11, eBioscience). Cells were washed and fixed using Cytofix/Cytoperm buffer (BD Pharamingen or eBioscience). Intracellular staining was performed with following antibodies: anti-mouse IL−17A (TC11−18H10.1, BioLegend), anti-mouse IFN−γ (XMG1.2, eBiosciences), anti-mouse Foxp3 (FKJ−16s, eBioscience) and anti-human/mouse T−bet (eBio4B10, eBioscience). Dead cells were excluded from analysis using a Zombie Yellow Fixable Viability Kit (Biolegend) or a LIVE/DEAD Fixable Aqua Dead Cell Stain Kit (Life Technologies).

ChIp-seq. ChIP-seq assays of naïve Cd4+ T cells were performed as described previously. In brief, cells for ChIP-seq were fixed for 10 min in 1% formaldehyde and sonicated and chromatin immunoprecipitation was performed with anti-MLL4 (ref. 19,20) and anti–H3K27me3 (07–449, Millipore). ChIP-seq for H3K4me1 was performed on native (not fixed) chromatin with anti–H3K4me1 (ab8895, Abcam). ChIP DNA was end-repaired using an End-It DNA End-Repair kit (Epitcine), and was indexed, amplified and sequenced on an Illumina 2G Genome Analyzer.

RNA-Seq. RNA-Seq experiments were previously described. Poly(A) RNA was isolated from naïve Cd4+ T cells and Treg cells with a Dynabeads mRNA Direct Kit (610.12, Invitrogen) following the protocol provided by the manufacturer. mRNA was reverse-transcribed with the use of Super Script Double-Stranded cDNA Synthesis Kit (Invitrogen) and Random Hexamer Primers (Invitrogen). Double-stranded cDNA was sonicated on Bioruptor (Diagenode), blunt-ended using an End-It DNA Repair kit (Epitcine), indexed, amplified and sequenced on an Illumina 2G Genome Analyzer.

Immunoblot analysis. Cd4+ T cells were lysed in RIPA buffer and proteins were resolved by the Novex NuPage SDS-PAGE gel system (Life Technologies). Proteins were transferred to Supported Nitrocellulose Membrane (Bio-Rad) and were incubated with anti-H3K4me1 (ab8895, Abcam), anti-H3K4me2...
(ab32356, Abcam), anti-H3K4me3 (17-614, Millipore), anti-MLL4 (refs. 19,20), anti-β-actin (clone: AC-15, Sigma-Aldrich), anti-Ribbp5 Cat# A300-109A, BETHYL) and anti-panH3 (Cat# ab21054, Abcam) as a loading control. Blots were visualized with Pierce ECL Western Blotting Substrate (Thermo Scientific).

Modified Hi-C protocol. Naïve CD4+ T cells were crossed linked with 1% formaldehyde for 10 min. Cells were lysed and digested with restriction enzymes CviQ1 and CviA II and Bfa I for 20 min. The Hi-C samples were processed following the in situ Hi-C protocol29 with modifications briefly described as follows: DNA ends were marked by biotin-14-dATP with Klenow (large) for 1 h at 37 °C. Blunt-ended DNA fragments were ligated with T4 DNA Ligase overnight at 16 °C. DNA was then reverse cross-linked and purified by phenol–chloroform extraction. Biotin was removed from unligated DNA ends by T4 DNA polymerase for 2 h at 12 °C. DNA was purified by phenol–chloroform and sheared to 300-500 bp by sonication followed by DNA-end repair and the addition of adenosine residues44. Biotin-labeled DNA was precipitated by streptavidin beads, followed by Illumina adaptor ligation and PCR amplification. Hi-C experiments were done with n = 2 independent experiments for Mll4-KO naïve CD4+ T cells and n = 3 independent experiments for the control cells.

Deletion of the Foxp3 – 8.5 kb MLL4-binding site using CRISPR-Cas9 technology. We used CRISPR-Cas9 technology to delete the putative 500-bp MLL4-binding site at –8.5 kb upstream of the Foxp3 transcription start site. To increase the chances of deletion, we designed two CRISPR single guide RNAs (sgRNAs) to cut each end of this 500-bp region. The sgRNAs for cutting the upstream end were 5′-GCCATGAGGATGTAGTCCAG-3′ and 5′-CTTCTGACCCCTACCTGCAA-3′, and the sgRNAs for cutting the downstream end were 5′-TGGACGGTACTGACCCCCGA-3′ and 5′-TGAAATGCGGAGGTCTTGG-3′. These four sgRNA target sequences were cloned by OriGene Technology into the pT7-Guide-IVT vector. Injectable sgRNAs were transcribed in vitro using a MEGASHortscript T7 Kit (Life Technologies), and the Cas9 mRNA was transcribed in vitro from plasmid MLM3613 (Addgene #42251) using a MESSAGE mammachine T7 Kit (Life Technologies), as previously reported35. These four sgRNAs were co-injected with Cas9 mRNA into fertilized eggs collected from B6CBAF1/J mice (JAX) at a concentration of 100 ng/µl Cas9 mRNA and 20 ng/µl of each sgRNA. The injected zygotes were cultured overnight at 37 °C in a humidified incubator with 5% CO2. The next morning, those embryos that reached two-cell stage of development were implanted into the oviducts of pseudopregnant foster mothers. Offspring born to these foster mothers were genotyped by PCR amplification, followed by DNA sequencing.

Data analysis. Definition of genomic regions. RefSeq gene annotation was downloaded from the UCSC genome browser. Promoters were defined as genomic regions spanning 2.5 kb upstream and downstream of annotated transcription start site. Gene-body regions were defined as starting from the transcription start site to the transcription end site, with the first 2.5 kb excluded. Other genomic regions are defined as intergenic. Regular active enhancers were estimated by H3K27ac peaks located beyond promoters and super-enhancers.

ChiP-Seq data analysis. All ChiP-Seq short reads were mapped to the mouse genome (mm9) by using Bowtie2 with default parameters47. Later analysis excluded short reads mapped to multiple genomic positions and kept only one read for each genomic site when it received multiple reads. ChiP-Seq-read-enriched regions were identified with SICER with a window size of 200 bp and a gap size of 400 bp28. A conservative set of MLL4-binding sites were identified with stringent parameters and input control: E value = 1,000 and false-discovery rate (FDR) = 0.001. Here, ‘E-value’ refers to the number of expected peaks by assuming a random distribution of the ChiP-Seq reads26. To have a conservative prediction on genomic regions that were not bound by MLL4, we increased the E-value for MLL4 peak call until the ChiP-Seq signal of the additionally called peaks was indistinguishable from the input signal (Supplementary Fig. 5d); for this purpose, non-stringent parameters were set: E value = 100,000 and FDR = 0.05. ChiP-Seq peaks for histone modifications H3K4me1, H3K27me3 and H3K27ac without inputs were called with an E-value of 10. Differential H3K4me1 peaks between deletion of Mll4 and control cells were predicted by EdgeR (FDR < 0.001; change of over twofold)48. The ROSE program49 was applied to the H3K27ac ChiP-Seq data to identify super-enhancers in mouse naïve CD4+ T cells with the ‘-i’ option on to exclude the contribution of peaks from promoter regions.

RNA-Seq data analysis. The mRNA expression of a gene is quantified by RPKM (reads per kilobase of exon model per million reads)30 with an in-house script. The calculation of the change in gene expression (‘fold’ values) excluded genes that showed an RPKM less than one in both the Mll4-KO cells and wild-type cells. Differentially expressed genes were identified by a change in expression of >1.5-fold and an expression value of >3 for at least one of the two conditions: Mll4-KO and wild-type.

Hi-C data analysis. Pair-end short reads were mapped to the mouse genome (mm9) by Bowtie2 with default parameters47. In-house scripts were used to exclude pair-end tags–reads (PETs) with each end mapped to different chromosomes or showing low mapping quality (MAPQ < 10). For PETs mapped to the same positions, only one PET was kept.

To call intra-chromosomal interaction from Hi-C PETs, we divided the genome into bins of 1 kb and assessed the significance of interaction for bin pairs that were separated by at least 1 bp and were linked by at least n PETs using background models generated through simulation, which explicitly considered the number of total PETs, the distributions of GC content, mappability and distance from the observation. In brief, for each observed intra-chromosomal PET that linked two bins, we randomly assigned it to two bins from the same chromosome, but required that the GC contents from the new positions were similar to the original (with a fluctuation up to 2%); the mappability scores of the new bins were similar to those of the original (allowing a fluctuation up to 0.05); and the distance between the two bins was the same as the original. The procedure was repeated for every intra-chromosomal PET such that the background shared the same number of PETs with the observation. To save computational time, we carried out the simulation on chromosome 1, and repeated it 5,000 times.

With the background models, we calculated the expected number of bin pairs linked by at least n PETs and separated by at least 1 bp and compared it to the observation. For example, the expected number of n = 4 under L = 11,000 was 3,608 for the Hi-C data generated for the control naïve CD4+ T cells, and the observed number was 11,237, which corresponded to an enrichment ratio (Σ of 3 = 11,237/3,608). We used the enrichment ratio as a threshold for an interaction call. As expected, a lower number of n was needed as L increased to achieve the same Σ. This led us to define a threshold step function that depended on distance L to ascertain the minimal number of PET n values required for a bin–pair interaction call under Σ = 3. An Σ of 3 meant that of three observations, one could be explained by the background model and was therefore probably a false-positive result, while the remaining two could not be explained that was and were therefore probably true positive results. In other words, an Σ of 3 corresponded to an estimated specificity of 67% (Σ/1)/Σ. We noticed that the Σ was underestimated for interaction calls made for functional genomic regions: we identified 12,413 interacted bin pairs bound by MLL4 and/or marked by H3K4me1 under Σ = 3 for chromosome 1, in contrast to 882 bin-pairs from the background model, which corresponded to an estimated specificity of 93%. The interaction among regulatory regions associated with H3K4me1 and/or MLL4 was the focus of this project.

To generate a BEDGraph presentation of interaction intensity of genomic bins, we considered PETs from interacted bins called under Σ = 3. For a given genomic bin of 1 kb, we estimated its interaction intensity by the number of PETs linked to any other interacted bins outside this region, normalized by the total number of PETs from interacted bin pairs across the genome. To generate a smoothed view (for example, Fig. 5a), we applied a sliding window of 10 bins at a step of one bin and averaged the interaction intensities of all bins within the window.

To call genomic regions with differential interaction between two conditions, we extended the bin size to 2 kb and counted the PETs that linked to any other interacted bins (Fig. 5b,c) or any other interacted bins with certain features (for example, overlapped with enhancers; Fig. 6d,e) from the same chromosome for each condition. The EdgeR package48 was applied to identify regions showing differential interaction in Mll4-KO cells relative to that in wild-type cells (a change of over 1.5-fold and FDR < 0.05).

https://doi.org/10.1038/n.3812

© 2017 Nature America, Inc., part of Springer Nature. All rights reserved.
To investigate the relationship between the change in the abundance of H3K4me1 at MLL4-unbound genomic region and the change in interaction intensity with MLL4-bound remote sites (Fig. 5d), we sorted the H3K4me1-enriched regions on the basis of their response to the deletion of MLL4 in terms of the change in the abundance of H3K4me1. For each H3K4me1-enriched region, the numbers of PETs linking the region to interacted MLL4-binding sites were recorded for the cells in which MLL4 was deleted and wild-type cells and were transformed into a change in number (‘fold’ values). The distribution of that change was then compared among the three groups of H3K4me1-enriched regions.

Gene-ontology–enrichment analysis. A gene was called a target of an enhancer if any 1-kb bin from the promoter of the gene interacted with any bin from the enhancer. Gene-ontology–enrichment analysis was carried out using the online DAVID Bioinformatics Resource51 or GREAT27. Redundant GO terms from the DAVID output were removed by using REVIGO52.

Code availability. C++ code used to generate the step function for the identification of chromatin interaction from Hi-C PETs based on simulation is available on request.

Statistics. We applied a two-sided Kolmogorov-Smirnov test (two-tailed throughout the manuscript) to assess the difference in the accumulated distribution of ChIP-Seq read density (Fig. 4g and Supplementary Fig. 6b), change in ChIP-Seq read density (‘fold’ values) (Fig. 4h and Supplementary Fig. 6d.f.g), change in interaction intensity (‘fold’ values) (Figs. 5d and 6f.g and Supplementary Fig. 7a) and change in gene expression (‘fold’ values) (Figs. 6h and Supplementary Fig. 7c.d) for any two groups of interests. The Kolmogorov-Smirnov test is a nonparametric method that tests whether two probability distributions differ and requires no prior knowledge about the distributions53. In all the figures noted above (accumulative distribution), the χ² test was used for comparison of two portions from independent samples, presented as a percentage (Figs. 5b and 7d).

Statistical significance for other figures was calculated by the ANOVA test or Kruskal–Wallis when a sample did not meet Gaussian distribution assessed by the D’Agostino and Pearson omnibus normality test.

Data availability. The sequencing data including Hi-C, ChIP-Seq data and RNA-Seq have been deposited in the Gene Expression Omnibus database with accession number GSE69162. The source data of all figures that support the findings of this study are available from the corresponding author on request.

A Life Sciences Reporting Summary for this paper is available online.

41. Lee, P.P. et al. A critical role for Dnmt1 and DNA methylation in T cell development, function, and survival. *Immunity* **15**, 763–774 (2001).
42. Zhang, P. et al. PARP-1 controls immunosuppressive function of regulatory T cells by destabilizing Foxp3. *PLoS One* **8**, e71590 (2013).
43. Zanvit, P. et al. Antibiotics in neonatal life increase murine susceptibility to experimental psoriasis. *Nat. Commun.* **6**, 8424 (2015).
44. Barski, A. et al. High-resolution profiling of histone methylations in the human genome. *Cell* **129**, 823–837 (2007).
45. Wei, G. *et al.* Genome-wide analyses of transcription factor GATA3-mediated gene regulation in distinct T cell types. *Immunity* **35**, 299–311 (2011).
46. Chepelev, I., Wei, G., Tang, Q. & Zhao, K. Detection of single nucleotide variations in expressed exons of the human genome using RNA-Seq. *Nucleic Acids Res.* **37**, e106 (2009).
47. Langmead, B. & Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
48. Robinson, M.D., McCarthy, D.J. & Smyth, G.K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
49. Whyte, W.A. *et al.* Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307–319 (2013).
50. Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcription by RNA-Seq. *Nat. Methods* **5**, 621–628 (2008).
51. Jiao, X. *et al.* DAVID-WS: a stateful web service to facilitate gene/protein list analysis. *Bioinformatics* **28**, 1805–1806 (2012).
52. Supek, F., Bošnjak, M., Škunca, N. & Šmuc, T. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One* **6**, e21800 (2011).
53. Hollander, M. & Wolfe, D.A. *Nonparametric Statistical Methods* (Wiley, 1973).
Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work we publish. This form is published with all life science papers and is intended to promote consistency and transparency in reporting. All life sciences submissions use this form; while some list items might not apply to an individual manuscript, all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

1. Sample size
   Describe how sample size was determined.
   Sample size was determined empirically, at least two independent experiments were conducted with different number of littermates.

2. Data exclusions
   Describe any data exclusions.
   n/a

3. Replication
   Describe whether the experimental findings were reliably reproduced.
   All the findings were reliably reproduced.

4. Randomization
   Describe how samples/organisms/participants were allocated into experimental groups.
   Animals used were littermates allocated to groups on the base of their genotype.

5. Blinding
   Describe whether the investigators were blinded to group allocation during data collection and/or analysis.
   Animals were identified with an ear tag number not the genotype. The genotype of animal was not known till the final analysis.

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. Statistical parameters
   For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or the Methods section if additional space is needed).

   - The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
   - A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly.
   - A statement indicating how many times each experiment was replicated.
   - The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section).
   - A description of any assumptions or corrections, such as an adjustment for multiple comparisons.
   - The test results (e.g. p values) given as exact values whenever possible and with confidence intervals noted.
   - A summary of the descriptive statistics, including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range).
   - Clearly defined error bars.

See the web collection on statistics for biologists for further resources and guidance.

7. Software
   Describe the software used to analyze the data in this study.
   Flow cytometry data were analysed on FlowJO software v 8.8.7, statistical analysis on FACS data was preformed using GraphPad Prism. WesternBlot
signal was calculated using ImageJ. ChIP-Seq peaks called by SICER V1.1. Differential ChIP-Seq peaks called by Edger 3. Super enhancer called by ROSE.

For all studies, we encourage code deposition in a community repository (e.g. GitHub). Authors must make computer code available to editors and reviewers upon request. The Nature Methods guidance for providing algorithms and software for publication may be useful for any submission.

Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).

anti-Mll4 antibody (rabbit polyclonal) was provided by Dr. Kai Ge, they were validated by Western Blotting and published previously, all the rest of antibodies were commercially purchased and validated

Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.

b. Describe the method of cell line authentication used.

c. Report whether the cell lines were tested for mycoplasma contamination.

d. If any of the cell lines used in the paper are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.

Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.

Mll4fl/fl mice on mixed C57BL/6 and 129 backgrounds were provided by Dr Kai Ge (NIDDK, National Institutes of Health) and have been previously described. CD4Cre+ mice on C57BL/6 background were purchased from Taconic. Foxp3CreYFP+ mice on C57BL/6 background were purchased from Jackson Laboratories. Foxp3 CRISP mice were generated in NHLBI Transgenic Core at NIH under the direction of Dr Chengyu Liu as described in method section. Mice were bred and maintained in an NHLBI specific-pathogen free animal facility. All experiments were performed on 6- to 10-week-old mice in accordance with the protocol approved by the NHLBI Animal Care and Use Committee.

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.

n/a
Flow Cytometry Reporting Summary

Form fields will expand as needed. Please do not leave fields blank.

Data presentation

For all flow cytometry data, confirm that:

1. The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
2. The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a ‘group’ is an analysis of identical markers).
3. All plots are contour plots with outliers or pseudocolor plots.
4. A numerical value for number of cells or percentage (with statistics) is provided.

Methodological details

5. Describe the sample preparation.

Spleens, lymph nodes and thymi were harvested from 6–10 weeks old mice, smashed on a cell strainer and resuspended in a FACS staining buffer, spleens were treated with ACK lysing buffer for 5 min at room temperature prior to staining. Lymphocytes from lamina propria were obtained by mechanically disrupting small intestines. Segments of tissues were washed extensively followed by incubation for 20 min at 37°C with vigorous shaking in pre-warmed RPM medium supplemented with 3% FBS, 5mM EDTA and dithiothreitol at 0.145mg/ml to remove intraepithelial lymphocytes. Next remaining tissues were digested with Liberase TL (Roche) at 0.2 mg/ml and 0.05% DNase (Sigma) in RPMI medium for 20 min at 37°C with continuous stirring. Digested tissues were minced and passed through 70- and 40-μm cell strainer. Lymphocytes were enriched by Percoll density gradient centrifugation.

Single cell suspensions were resuspended and cells were stained for surface markers according to standard protocols (30 min in dark at 4 degrees Celsius) with following antibodies: anti-mouse TCRβ (H57–597; eBioscience), anti-mouse CD4 (RM4–5; eBioscience), anti-mouse CD45 (30-F11, eBioscience) and anti-mouse CD8a (53-6.7, eBioscience).

For the cytokine expression analysis the cells were stimulated for 4h in phorbol 1,2-myristate 1,3-acetate (5ng/ml) and ionomycin (1μg/ml) in the presence of protein transport inhibitor GolgiPlug (BD Pharmingen). Next cells were stained for surface markers then washed and fixed using Cytofix/Cytoperm buffer (BD Pharmingen or eBioscience). Intracellular staining was performed with following antibodies: anti-mouse IL17a (TC11-18H10.1; Biolegend), anti-mouse IFN-γ (XMG1.2, eBiosciences), anti-mouse Foxp3 (FJK-16s, eBioscience) and anti-human/mouse Tbet (eBio4B10, eBiosciences).

Intracellular staining on not stimulated cells was performed using Cytofix/Cytoperm buffer (eBioscience). Cells were then washed and stained with following antibodies: anti-mouse Helios (22F6, BioLegend), anti-mouse Ki67 (B56, BD Biosciences) and anti-mouse FoxP3 (FJK-16s, eBioscience).
| Step | Description |
|------|-------------|
| 6.   | Identify the instrument used for data collection. | FACS data were acquired on BD FACSCanto II (BD Biosciences) or BD LSRII (BD Biosciences) |
| 7.   | Describe the software used to collect and analyze the flow cytometry data. | FACS data were analyzed with FlowJo software version 8.8.7 and v.10 |
| 8.   | Describe the abundance of the relevant cell populations within post-sort fractions. | Cells sorted for Treg suppression assays (regulatory and effector cells) showed approx. 99% purity in post-sort analysis. |
| 9.   | Describe the gating strategy used. | Dead cells were excluded from analysis using Zombie Yellow Fixable Viability Kit (Biolegend) or LIVE/DEAD Fixable Aqua Dead Cell Stain Kit (Life Technologies). |

Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information. ☑
ChIP-seq Reporting Summary

Form fields will expand as needed. Please do not leave fields blank.

Data deposition

1. For all ChIP-seq data:
   a. Confirm that both raw and final processed data have been deposited in a public database such as GEO.
   b. Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

2. Provide all necessary reviewer access links. (The entry may remain private before publication.)

   The following secure token has been created to allow review of record GSE69162 while it remains in private status:
   wfidsoyydhyhhm

3. Provide a list of all files available in the database submission.

   GA4747-1-ATCACG-naive-CD4-MLL4flflCreNeg-H3K27ac-m1_0_0_noDup-GA4748-2-CGATGT-naive-CD4-MLL4flflCrePos-H3K27ac-m1_0_0_noDup-W200-G400-E10.scoreisland.sicerdiff.gz
   GA4874-rpkm.gz
   GA4876-rpkm.gz
   GA5043-rpkm.gz
   GA5044-rpkm.gz
   GA6866-GCCAAT-ChIP-seq-mouse-WT-H3K27me3.sam_noDup-GA6872-CTTGTGA-ChIP-seq-mouse-MLL4-KO-H3K27me3.sam_noDup-W200-G400-E10.scoreisland.sicerdiff.gz
   GA6875_CD4T_Hic.bin1Kmin1Kmax2M.III.gz
   GA6876_CD4T_Hic.bin1Kmin1Kmax2M.III.gz
   GA7211-GGCTAC-ChIP-seq-mouse-Naive-WT-H3K27me3_noDup-GA7212-CTTGTGA-ChIP-seq-mouse-Naive-MLL4-KO-H3K27me3_noDup-W200-G400-E10.scoreisland.sicerdiff.gz
   GA7827-ATCAGC-ChIP-seq-mouse-8723-H3K4me1-1.sam_noDup-GA7828-CGATGT-ChIP-seq-mouse-8723-H3K4me1-1.sam_noDup-W200-G400-E10.scoreisland.sicerdiff.gz
   GA7832-GCCAAT-ChIP-seq-mouse-8725-H3K4me1-2.sam_noDup-GA7828-CGATGT-ChIP-seq-mouse-8723-H3K4me1-2.sam_noDup-W200-G400-E10.scoreisland.sicerdiff.gz
   GA5963-TGACCA-ChIPSeq-mouse-MLL3ppMLL4FF-Naiv-CD4-MLL4_sam_noDup-W200-G400-islands-summary-E100000FDR0.05.txt.gz
   GA5963-TGACCA-ChIPSeq-mouse-MLL3ppMLL4FF-Naiv-CD4-MLL4_sam_noDup-W200-G400-islands-summary-E100000FDR0.001.txt.gz
   GA6843-CTAGGC-ChIPSeq-mouse-WT-MLL4-1-kz775kz746_noDup-W200-G400-E1000.scoreisland.txt.gz
   K2852_GA7713_NNNNNN_L3456_R1.fastq.bedpe.iii.gz
   mll4_wt_pool_GA7714r799r818r909.bedpe.iii.gz
   mll4_ko_pool_GA7715r799r818r909.bedpe.iii.gz
   GA4747-1-ATCACG-naive-CD4-MLL4flflCre-H3K27ac-200-500-mouse_Gateway_SuperEnhancers.bed.gz
   GA6861GA6867_GA7201GA7202_me1Cnt.edger3.gz
   KZ344_GA4747_ATCACG_L008_R1.fastq.gz
   KZ344_GA4748_CGATGT_L008_R1.fastq.gz

Nature Immunology: doi:10.1038/ni.3812
4. If available, provide a link to an anonymized genome browser session (e.g. UCSC).

Methodological details

5. Describe the experimental replicates.

- 2, independent experiments (GA5963 & GA6843), MLL4 ChIP-Seq for naive CD4+ T cells, 73.4 (% peaks shared)
- 2, independent experiments (GA6861 & GA7201), H3K4me1 ChIP-Seq for naive CD4+ T cells control for Mll4-KO, 79.3-98.8 (% peaks shared)
- 2, independent experiments (GA6867 & GA7202), H3K4me1 ChIP-Seq for Mll4-KO naive CD4+ T cells, 82.2-98.5 (% peaks shared)
- 2, independent experiments (GA7831 & GA7832), H3K4me1 ChIP-Seq for naive CD4+ T cells control for CRISPR KO, 94.0-95.2 (% peaks shared)
- 2, independent experiments (GA7827 & GA7828), H3K4me1 ChIP-Seq for naive CD4+ T cells CRISPR KO, 92.9-94.4 (% peaks shared)
- 2, independent experiments (GA7211 & GA6866), H3K27me3 ChIP-Seq for naive CD4+ T cells control for Mll4-KO, 69.5-88.2 (% peaks shared)
- 2, independent experiments (GA7212 & GA6872), H3K27me3 ChIP-Seq for Mll4-KO naive CD4+ T cells, 78.6-82.7 (% peaks shared)
- 1, single experiment (GA4747), H3K27ac ChIP-Seq for naive CD4+ T cells control for Mll4-KO, NA
- 1, single experiment (GA4748), H3K27ac ChIP-Seq for naive Mll4-KO CD4+ T cells, NA

6. Describe the sequencing depth for each experiment.

PCR Condition:
Denature at 98°C for 30 sec.
98°C, 10”, 65°C, 30”; 72°C, 30”; 18 cycles; 72°C, 5”; 4°C, forever

all chip-seq sequencing data in this study are 50-bp single-end reads

FASTQ Total_read_num Num_of_unique_reads_(MAPQ>10)
KZ344_GA4747_ATCACG_L008_R1.fastq 17,863,913 5,266,116.00
KZ344_GA4747_CGATGTC_L008_R1.fastq 65,818,757 16,799,470.00
KZ565_GA5960_ATCACG_L006_R1.fastq 15,924,609 10,008,272.00
KZ565_GA5963_TGACCA_L006_R1.fastq 19,464,089 8,991,647.00
7. Describe the antibodies used for the ChIP-seq experiments. Mll4 antibody was provided by K. Ge., anti-H3K4me1 (ab8895) antibody and anti-H3K27ac (ab4729) antibody were purchased from Abcam, anti-H3K27me3 antibody (07-449) was purchased from Millipore.

8. Describe the peak calling parameters. ChIP-Seq read enriched regions were identified by SICER (PMID: 19505939). Settings for calling peaks for H3K4me1, H3K27me3 and H3K27ac are: window size = 200 bp, gap size = 400 bp, E value = 10.

Settings for calling peaks for MLL4 with input control are: window size = 200 bp, gap size = 400 bp (E value = 1000 & FDR < 0.001) OR (E value = 10000 & FDR < 0.05).

9. Describe the methods used to ensure data quality. With input control, which is the case for the MLL4 ChIP-Seq, under FDR < 0.05, the number of peaks predicted from SICER is 16K. Without input control, the SICER program uses E-value to control of the significance of called peaks. E-value refers to the number of expected peaks by assuming a random distribution of the ChIP-Seq reads. An empirical FDR could be estimated as the E-value divided by the number of called peaks. Giving that the number of called peaks exceeds 27K for all ChIP-Seq libraries without input, the empirical FDR is less than 0.0004.

10. Describe the software used to collect and analyze the ChIP-seq data. SICER V1.1 (PMID: 19505939)