HHS Public Access

Author manuscript

Nat Immunol. Author manuscript; available in PMC 2018 March 05.

Published in final edited form as:

Nat Immunol. 2017 September ; 18(9): 1035–1045. doi:10.1038/ni.3812.

MLL4 prepares the enhancer landscape for *Foxp3* induction via chromatin looping

Katarzyna Placek[1,6], **Gangqing Hu**[1,6], **Kairong Cui**[1,6], **Dunfang Zhang**[2], **Yi Ding**[1,3], **Ji-Eun Lee**[4], **Younghoon Jang**[4], **Chaochen Wang**[1,4], **Joanne Elizabeth Konkel**[2], **Jiuzhou Song**[3], **Chengyu Liu**[5], **Kai Ge**[4], **Wanjun Chen**[2], and **Keji Zhao**[1]

[1]Systems Biology Center, Division of Intramural Research, NHLBI, NIH, Bethesda, Maryland, USA

[2]Mucosal Immunology Section, Division of Intramural Research, NIDCR, NIH, Bethesda, Maryland, USA

[3]Department of Animal and Avian Sciences University of Maryland, College Park MD 20742-2311 USA

[4]Adipocyte Biology and Gene Regulation Section, Laboratory of Endocrinology and Receptor Biology, NIDDK, NIH, Bethesda, Maryland, USA

[5]Transgenic Core, Division of Intramural Research, NHLBI, NIH, Bethesda, Maryland, USA

Abstract

MLL4 is an essential subunit of the H3K4 methylation complexes. We report that MLL4 deficiency compromised regulatory T (T_{reg}) cell development and resulted in substantial decreases in H3K4me1 and chromatin interaction at putative enhancers, a remarkable portion of which were not direct targets of MLL4 but were enhancers that interact with MLL4-bound sites. The decrease in H3K4me1 and chromatin interaction at the MLL4-unbound enhancers correlated with MLL4 binding at distant-interacting regions. Deletion of an upstream MLL4 binding site reduced H3K4me1 at the *Foxp3* regulatory elements looped to the MLL4 binding site and compromised both thymic T_{reg} and inducible T_{reg} cell differentiation. We show that MLL4 catalyzed H3K4 methylation at distant unbound enhancers via chromatin looping, thus providing a new mechanism of regulating T cell enhancer landscape and impacting T_{reg} cell differentiation.

Regulatory T (T_{reg}) cells are central players in establishing homeostasis of the immune system by suppressing activation, proliferation and effector functions of various immune

Correspondence should be addressed to W.C. (wchen@dir.nidcr.nih.gov) or K.Z. (zhaok@nhlbi.nih.gov).

[6]These authors contributed equally to this work.

AUTHORS CONTRIBUTION

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., JE. L., Y.J., C.W., J.E.K., J.S., C.L., K.G., W.C. and K.Z. analyzed the data, K.P., G.H., K.C, W.C. and K.Z. wrote the paper.

COMPETING FINANCIAL INTEREST

The authors declare no competing financial interests.

cells[1]. They develop in the thymus from $CD4^+$ single-positive (CD4SP) cells or differentiate from naïve $CD4^+$ T cells[2]. The cytokine TGF-β drives differentiation of T_{reg} cells by up-regulating expression of Foxp3 transcription factor that is necessary for suppressive activity and serves as a marker of T_{reg} cells[3–5]. Deregulation of T_{reg} cell development and function leads to autoimmune diseases and immunopathology[1,6–8]. Because of their important roles in numerous diseases including allergy[9], autoimmunity[1,6–8], microbial infections[10] and cancer[11], T_{reg} cells have become a focus for development of various therapies aiming to treat autoimmune disorders and graft-versus-host disease[12,13]. Thus, a thorough understanding of the regulatory processes that govern T_{reg} cell differentiation is necessary.

Cell specification is under control of cell-specific enhancers. Foxp3 is the signature transcription factor that defines T_{reg} cells, which is regulated by three distal enhancer elements including conserved noncoding-sequence (CNS) 1, CNS2 and CNS3 at different stages of T_{reg} cell development[14]. The genome-wide enhancer landscape in T_{reg} cells has been recently described[15]. Foxp3 does not establish T_{reg}-specific enhancer landscape but instead exploits previously established already existing enhancers[16]. However, the mechanisms that initially establish the enhancer landscape remain unclear.

Active and primed enhancers are characterized by the presence of permissive histone modifications such as histone acetylation and histone H3 lysine 4 (H3K4) monomethylation[17]. The activating histone marks facilitate chromatin opening and recruitment of transcription factors and other regulatory machineries. H3K4 methylation is catalyzed by the MLL family of histone methyltransferases, including SETD1A, MLL1 (also called KMT2A)[18], MLL2 (also called KMT2B), MLL3 (also called KMT2C) and MLL4 (also called KMT2D). MLL4 has been shown to shape enhancer pattern in mammalian cells during heart development[19], myogenesis and adipogenesis[20] by regulating mono- and di-methylation of H3K4.

We show that MLL4 was critically required for T_{reg} cell development by establishing the enhancer landscape and facilitating long-range chromatin interaction. In addition to regulating H3K4 monomethylation at direct binding sites, we show that MLL4 catalyzed H3K4 methylation *in trans* at distant unbound enhancers via long-distance chromatin looping, thus providing a previously unrecognized mechanism of regulation of histone modification and enhancer landscape in the cells.

RESULTS

Mll4 deletion results in compromised T_{reg} development

To investigate the function of MLL4 in T cell development, we generated MLL4-conditionally deficient mice by breeding $Mll4^{fl/fl}$ mice[20] that carry *lox*P sites flanking exons 16–19 with *Cd4*-Cre^+ mice. We analyzed $Mll4^{fl/fl}$*Cd4*-Cre^-, $Mll4^{fl/fl}$*Cd4*-Cre^+, $Mll4^{fl/+}$*Cd4*-Cre^+ and $Mll4^{+/+}$*Cd4*-Cre^+ to correct for the effect of *Cd4*-Cre construct or of partial deletion of *Mll4* on mouse phenotypes. We confirmed the deletion efficiency of the floxed *Mll4* exons in $CD4^+$ T cells isolated from $Mll4^{fl/fl}$*Cd4*-Cre^+ mice and partial deletion in $Mll4^{fl/+}$*Cd4*-Cre^+cells by RT-PCR (Supplementary fig. 1a), RNA-seq (Supplementary Fig. 1b) and immunoblotting (Supplementary Fig. 1c). Deletion of exons 16–19 truncates the

MLL4 protein and disrupts the MLL4 complex[20]. Because we confirmed that the *Cd4*-Cre transgene alone has no effects on mouse phenotype and we did not observe significant differences in T cell populations between *Mll4*$^{fl/fl}$*Cd4*-Cre^{-}, *Mll4*$^{fl/+}$*Cd4*-Cre^{+} and *Mll4*$^{+/+}$*Cd4*-Cre^{+} mice (Fig. 1 and Supplementary Fig. 1d–g), we further refer the *Mll4*$^{fl/fl}$*Cd4*-Cre^{-} mice as a representative of wild-type unless stated otherwise and *Mll4*$^{fl/fl}$*Cd4*-Cre^{+} mice as *Mll4*KO.

While *Mll4* conditional deletion had no significant effects on T cell development in the thymus as CD4^{+}CD8^{+} double-positive (DP), CD4^{+} single-positive (CD4SP) and CD8^{+} single-positive (CD8SP) cell populations remained similar in all examined groups of animals (Fig. 1a, b), it substantially decreased the frequency and total number of CD4^{+}Foxp3^{+} T$_{reg}$ cells in the thymus of the *Mll4*KO mice compared to littermate wild-type control mice (Fig. 1c, d). *Mll4* deletion also significantly reduced CD4^{+} and CD8^{+} T cell numbers in secondary lymphoid organs including spleen (Fig. 1e, f) and lymph nodes (Supplementary Fig. 1e, g). Although the percentages of Foxp3^{+} cells within CD4^{+} T cell population in spleen and lymph nodes were not significantly affected in *Mll4*KO mice (Fig. 1g and Supplementary 1f), the total CD4^{+}Foxp3^{+} T cell numbers were reduced (Fig. 1h and Supplementary Fig. 1g) as a result of the decrease of total CD4^{+} T cell population in these organs. Consistent with the preserved T$_{reg}$:T$_{eff}$ cell ratio upon *Mll4* deletion, we did not see increased numbers of either interferon-γ (IFN- γ)– or interleukin 17A (IL-17A) (Supplementary Fig. 2a–d) or IL-4–producing T cells (data not shown) in the spleen and lymph nodes. We also did not see reduced Foxp3^{+} cell percentages within CD4^{+} T cell populations nor aberrant cytokine production by T cells in the lung of MLL4-deficient mice (data not shown). However, we did find a significant decrease of Foxp3^{+} cells within CD4^{+} T cells in lamina propria leukocytes of small intestine (Fig. 2a, b). Consequently, *Mll4* deletion led to increased numbers of IL-17A-producing cells, which was likely due to the reduction of T$_{reg}$ cells in small intestine (Fig. 2c, d). T$_{reg}$ cells in the *Mll4*KO mice in all examined organs showed more activated phenotype defined by increased Ki67 (Supplementary Fig. 2e and data not shown) and Helios expression[21] (Supplementary Fig. 2f and data not shown). T$_{reg}$ cells isolated from spleens and lymph nodes of wild-type and *Mll4*KO mice showed similar suppressive activities assessed by *in vitro* suppression assay (Supplementary Fig. 2g) and similar expression of T$_{reg}$ cell-associated markers such as CTLA-4, CD25 and GITR (data not shown), suggesting that MLL4 is not required for T$_{reg}$ cell function. To examine the effect of Foxp3-driven deletion of *Mll4* on T$_{reg}$ cells, we generated *Mll4*$^{fl/fl}$*Foxp3*-Cre^{+} mice by crossing *Mll4*$^{fl/fl}$ mice[20] with *Foxp3*-CreYFP^{+} mice[22]. We confirmed the specificity and efficiency of *Mll4* deletion in T$_{reg}$ cells from *Mll4*$^{fl/fl}$*Foxp3*-Cre^{+} mice (Supplementary Fig. 3a). *Mll4*$^{fl/fl}$*Foxp3*-Cre^{+} mice exhibited a slight reduction of T$_{reg}$ cell frequency in the thymus and periphery (Supplementary Fig. 3b–e, 3g, h) but no effect on cytokine production by CD4^{+} T cells (Supplementary Fig. 3f, i). *In vitro* suppression assay showed the functionality of T$_{reg}$ cells isolated from *Mll4*$^{fl/fl}$*Foxp3*-Cre^{+} mice (Supplementary Fig. 3j), indicating that MLL4 has no effect on suppressive function in cells already expressing Foxp3. All together, these data suggest that MLL4 regulates development of T$_{reg}$ cells but not T$_{reg}$ function.

MLL4 is required for induction of *Foxp3* expression

To further understand the role of MLL4 in T_{reg} development, we differentiated naïve $CD4^+$ T cells from wild-type and *Mll4*KO mice into T_{reg} (iT_{reg}) cells *in vitro*. We found expression of *Foxp3* by *Mll4*KO $CD4^+$ T cells was one-fifth that expressed by wild-type cells during differentiation into iT_{reg} lineage (Fig. 2e). At day 3 of differentiation, more than 70% of the $CD4^+$ T cells from wild-type mice expressed Foxp3 while only 20% of $CD4^+$ T cells from *Mll4*KO mice expressed Foxp3 (Fig. 2f). Our CFSE staining results indicated that *Mll4*KO $CD4^+$ T cells displayed a delay in proliferation and induction of Foxp3 expression as compared to wild-type cells at early times; however, by day 4 of *in vitro* culture both proliferation and Foxp3 expression of *Mll4*KO $CD4^+$ T cells were similar to wild-type cells (Supplementary Fig. 4a), indicating that the compromised Foxp3 expression is not caused by a defect in cell proliferation. These observations suggest a direct role of MLL4 in the induction of *Foxp3* expression.

To verify whether MLL4 plays a role in maintaining *Foxp3* expression, we first stimulated naïve $CD4^+$ T cells from *Mll4*$^{fl/fl}$ mice under T_{reg} induction conditions for 24 hours and then infected the cells with a retroviral vector carrying Cre-recombinase (pCre) or an empty vector (pRV) for two more days. RT-PCR analysis confirmed the deletion of *Mll4* exons 16–19 (Fig. 3a). We observed no difference of $Foxp3^+$ cell number in cultures infected with pCre compared to pRV viruses (Fig. 3b), suggesting that MLL4 is only required for the initiation of *Foxp3* expression but is not required for its maintenance. Furthermore, *Mll4* expression was much higher in naïve $CD4^+$ T cells as compared to *in vitro* induced effector cells T_H1, T_H2, T_H17 and iT_{reg} (Fig. 3c, d). *Mll4* expression in tT_{reg}s was comparable to CD4SP cells in the thymus (Fig. 3e) indicating another difference in developmental programs between tT_{reg} and iT_{reg}[23]. We also examined whether *Mll4* deletion affects differentiation of other effector T cells including T_H1, T_H2 and T_H17 under in vitro differentiation conditions. Only very modest changes were observed in the expression of T-bet and IFN-γ under T_H1 conditions (Supplementary Fig. 4b, c) and IL-17A and RORC under T_H17 condition (Supplementary Fig. 4d, e), while GATA-3 and IL-4 expression was increased in TH2 cultures upon *Mll4* deletion (Supplementary Fig. 4f, g). This observation indicates that MLL4 is dispensable for the differentiation of T_H1, T_H2 and T_H17 cell lineages *in vitro*. Together, these data indicate that MLL4 is required for induction but not maintenance of *Foxp3* expression during T_{reg} development and have only modest functions in other effector T cells.

MLL4 regulates H3K4me1 landscape in naïve $CD4^+$ T cells

MLL4 has been reported to regulate mainly H3K4 mono- and di-methylation[19,20,24], although some early reports suggested that MLL4 also regulates H3K4 tri-methylation[25,26]. We found that *Mll4* deletion in T cells only modestly decreased H3K4 monomethylation, while no significant decrease in H3K4 di- and tri-methylation was detected (Supplementary Fig. 5a), suggesting that other H3K4 methylation enzymes are present in the cells. We found that *Mll3*, which is a close homolog of *Mll4* and has partially redundant functions with *Mll4*[20], exhibited an expression pattern similar to *Mll4* in $CD4^+$ T cells (Supplementary Fig. 5b). Because naïve $CD4^+$ T cells expressed significantly more MLL4 as compared to iT_{reg} cells and that the deletion of *Mll4* at later stages of T_{reg} cell differentiation *in vitro* had no

effect on *Foxp3* expression, we analyzed MLL4 binding and H3K4me1 enrichment profiles in naïve CD4$^+$ T cells using ChIP-Seq. ChIP-Seq peak analysis using SICER[27] revealed that approximately 50% of the 16,161 MLL4 peaks were localized to promoter regions of genes and the other half mapped to the gene body and intergenic regions referred as non-promoter peaks (Supplementary Fig. 5c). MLL4 predominantly bound to genomic regions enriched with H3K4me1, as exemplified by a 350-kb genomic region in the X chromosome (Fig. 4a). However, a large number of H3K4me1 peaks were not direct targets of MLL4 (Fig. 4a, arrow heads). Globally, while MLL4 bound two-third of the H3K4me1-enriched regions overlapping with promoters, it was detected at less than 20% of the H3K4me1 peaks at other genomic regions (Fig. 4b). We relaxed the threshold by SICER[27] 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, which were enriched with H3K4me1 but devoid of MLL4 binding, were linked to genes involved in T cell activation by gene ontology enrichment analysis with GREAT[28](Supplementary Fig. 5f), suggesting that these elements are 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). These regions were generally less accessible as assessed by DNase-Seq[29] (Fig. 4c and Supplementary Fig. 5h), corresponding to a chromatin environment less favorable for direct MLL4 binding, suggesting that the modulation of H3K4me1 in these regions by MLL4 may be achieved through mechanisms other than a direct binding.

To identify H3K4me1 peaks affected by *Mll4* deletion, we compared the H3K4me1 profiles between wild-type and *Mll4*KO CD4$^+$ T cells. As a specific example, *Mll4* deletion induced substantial H3K4me1 increases at the promoter region of *Dynlt1f* (Supplementary Fig. 6a, blue box and b, top) and H3K4me1 decreases at putative enhancers bound by MLL4 (Supplementary Fig. 6a, green rectangles and b, bottom). Globally, we found 13.5% of the MLL4-bound enhancers showed decreased H3K4me1 upon *Mll4* deletion, whereas only 1.2% showed increased H3K4me1 (Supplementary Fig. 6c), supporting a direct role of MLL4 in regulating H3K4me1 at its binding sites[20]. Remarkably, *Mll4* deletion also substantially decreased H3K4me1 at enhancers without MLL4 binding, as exemplified by several putative enhancers upstream of *Dynlt1f* (Supplementary Fig. 6a, red rectangles) and *Cxcr4* (Fig. 4d, highlighted in green). Of all the non-promoter regions enriched with H3K4me1 but devoid of MLL4 binding, 14.8% exhibited a significant decrease in H3K4me1 upon *Mll4* deletion (Supplementary Fig. 6c). Analysis of RNA-Seq data did not reveal any significant changes in expression of other H3K4 methylation enzymes in *Mll4*KO cells (Supplementary Table 1), suggesting that the decrease in H3K4me1 at enhancers, which are not directly bound by MLL4, is most likely caused by loss of MLL4 activity but not by the loss of other H3K4 methylation enzymes resulting from the *Mll4* deletion.

To investigate the impact of *Mll4* deletion on transcription program, we identified 576 genes and 430 genes whose expression was down-regulated and up-regulated, respectively, in the MLL4-deficient cells. Gene ontology enrichment analysis revealed that the down-regulated

genes were enriched in functions including lymphocyte activation, regulation of signal transduction, cytokine secretion, while the up-regulated genes exhibited no obvious enrichment in any ontology term (Supplementary Table 2). Taken together, our data suggest that MLL4 regulates H3K4me1 at enhancers by direct and/or indirect binding and contributes to regulation of gene expression in naïve $CD4^+$ T cells.

MLL4 regulates H3K4me1 via chromatin interaction

The results above raised the possibility that H3K4me1 at MLL4-unbound regions may be catalyzed *in trans* by MLL4 bound at different sites via chromatin looping. To test this hypothesis, we analyzed genome-wide chromatin interactions in *Mll4*KO and wild-type naïve $CD4^+$ T cells using Hi-C[30,31] (Supplementary Table 3). We defined a conservative set of H3K4me1 regions that were not bound by MLL4 by calling MLL4 peaks using non-stringent thresholds with SICER[27]. On the other hand, we identified a highly confident set of MLL4-bound regions using stringent thresholds with SICER[27]. We identified interacting regions from Hi-C Pair-End Tags (PETs) by comparing to a background model generated through simulation, which considered GC content, mappability and distance. Interestingly, impressive 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 to the *Cxcr4* locus (Fig. 4d). Approximately 8% of the MLL4-unbound H3K4me1 peaks were looped to distant MLL4 binding sites, fourfold higher than the expectation (Fig. 4e). The fraction of MLL4-unbound H3K4me1 peaks that showed a decrease in H3K4me1 upon *Mll4* deletion was positively correlated with the number of interacting MLL4-binding sites (Fig. 4f, black bars), while the fraction that showed an increase in H3K4me1 was negatively correlated (Fig. 4f, grey bars). Consistently, we found more decrease in H3K4me1 read density for MLL4-unbound enhancers that interacted with more MLL4 binding sites than those interacting with less or none (Supplementary Fig. 6d). The enrichment of H3K4me1 at MLL4-unbound regions was positively associated with the aggregated MLL4 binding intensity from distantly interacting sites bound by MLL4 (Fig. 4g). Furthermore, MLL4-unbound sites with the most aggregated MLL4 binding intensity from distantly interacting sites showed the most decrease in H3K4me1 upon *Mll4* deletion (Fig. 4h). Together, these results support the notion that H3K4 mono-methylation at MLL4-unbound sites was *in trans* catalyzed by MLL4 bound to remote sites through chromatin looping as illustrated (Fig. 4i). Therefore, hereinafter we refer to the H3K4me1 peaks bound by MLL4 as “direct MLL4 targets” and H3K4me1 peaks not bound by MLL4 but interacting with MLL4-bound regions as “indirect MLL4 targets”.

MLL4 facilitates chromatin interactions at Foxp3 targets

To test whether MLL4 contributes to interaction between regulatory sites in naïve $CD4^+$ T cells, we compared the chromatin interaction density at direct and indirect MLL4 targets between wild-type and *Mll4*KO cells. Deletion of *Mll4* in naïve $CD4^+$ T cells substantially compromised chromatin interactions at many genomic loci as exemplified by the *Cxcr4* and *Foxp3* loci (Fig. 5a, red rectangles). To detect genome-wide changes, we divided the genome into equal-size 2-kb bins and examined their change in interaction intensity with other genomic regions upon *Mll4* deletion. We found 23% of the direct MLL4 targets showed decreased interaction upon *Mll4* deletion, which is significantly higher than those (4%)

showing increased interaction (Fig. 5b). Interestingly, we also found a significantly higher fraction (10%) of the indirect MLL4 targets showed decreased interaction than those (5%) with increased interaction upon *Mll4* deletion (Fig. 5b). In comparison, no more decrease than increase in interaction was observed for the regions devoid of either direct or indirect MLL4 binding (Fig. 5b). The interaction decrease caused by *Mll4* deletion was associated with a decrease in active histone mark H3K27ac (Supplementary Fig. 6e, f) and an increase in repressive histone mark H3K27me3 (Supplementary Fig. 6e, g). Furthermore, the fraction of the direct MLL4 targets showing decreased interaction upon *Mll4* deletion modestly correlated with MLL4 binding (Fig. 5c, left panel). Remarkably, the fraction of the indirect MLL4 targets showing decreased interaction highly correlated with the number of distantly interacting MLL4-bound sites (Fig. 5c, right panel). Furthermore, the decrease in H3K4me1 at indirect MLL4 targets upon *Mll4* deletion was associated with a decrease in interaction with remote MLL4 binding sites (Fig. 5d). These results indicate that MLL4 promotes long-distance chromatin interactions at both direct and indirect MLL4 targets.

MLL4 regulates enhancer-promoter interactions

Super-enhancers are critical regulatory elements for cell identity[32]. We identified regular and super-enhancers in naïve $CD4^+$ T cells using H3K27ac ChIP-Seq data (Fig. 6a) and analyzed chromatin interactions at these elements in wild-type and *Mll4*KO naïve $CD4^+$ T cells. Analysis of the chromatin interaction data revealed that the 467 super-enhancers in naïve $CD4^+$ T cells are linked to genes important for T cell function and differentiation (Fig. 6b). To examine the MLL4 function at enhancers, we divided non-promoter genomic regions into 2-kb bins, sorted into regular enhancers (RE), super-enhancers (SE) and other non-enhancer regions (bins not overlapping with any enhancer, OT). While only a few bins at non-enhancers were bound by MLL4, about half of the bins from super-enhancers and regular enhancers were bound by MLL4 (Fig. 6c). Meanwhile, 40% and 20% of MLL4-unbound bins within super-enhancers and regular enhancers, respectively, interacted with distant MLL4-bound regions (Fig. 6c) and thus could be indirect MLL4 targets. *Mll4* deletion resulted in significantly decreased chromatin interactions at 29% and 12% of 2-kb bins within super-enhancers and regular enhancers, respectively, while only 2% of 2-kb bins outside of enhancer regions exhibited decreased interactions (Fig. 6d). Next, we examined whether the chromatin interaction changes at regular and super-enhancers upon deletion of *Mll4* are related to MLL4 binding. Genomic bins (30%) directly bound by MLL4 within super-enhancers exhibited the greatest decrease in chromatin interactions upon *Mll4* deletion (Fig. 6e, upper green arrow head). The next most affected are the 2-kb bins (20%) that interacted with distant MLL4-bound regions within super-enhancers (Fig. 6e, lower green arrow head). The chromatin interactions at regular enhancers were similarly affected by *Mll4* deletion regardless of whether they were direct or indirect MLL4 targets (Fig. 6e, black arrow heads). These results indicate that MLL4 contributes to the chromatin interaction of enhancers with the most impact on super-enhancers upon *Mll4* deletion.

To test whether MLL4 contributes to specific promoter-enhancer interaction, we analyzed interaction changes upon *Mll4* deletion at promoters that are looped to super-enhancers (Pr-SE), regular enhancers (Pr-RE) or not looped to any enhancer (Ctrl). In general, we found that Pr-SE promoters showed the most significant decrease in chromatin interactions

compared to the others (Fig. 6f; Supplementary Fig. 7a), with specific examples of super-enhancer targets highlighted for *Cxcr4, Rac2*, and *Stk4* (known as *Mst1*) (Supplementary Fig. 7b). Further, we found that *Mll4* deletion resulted in more decrease in Pr-SE and Pr-RE interactions compared to the interactions between the same promoters with non-enhancer regions (Fig. 6g). In general, genes with the most decrease in interaction exhibited the most downregulation of expression in the MLL4-deficient naïve $CD4^+$ T cells (Supplementary Fig. 7c). Consistently, genes with promoters showing reduced interactions with enhancers/super-enhancers upon *Mll4* deletion exhibited a general decrease in expression (Fig. 6f, h and Supplementary Fig. 7d), implicating that direct MLL4 binding and/or the interaction with an MLL4-bound region facilitate promoter-enhancer interaction that plays an important role in regulating gene expression.

MLL4 regulates H3K4me1 of *Foxp3* via chromatin looping

To further understand how MLL4-dependent H3K4me1 and chromatin interaction between promoters and enhancers regulate T_{reg} cell development, we focused our analysis on $Foxp3^{REF4}$, which is regulated by its promoter and several intronic enhancers[14,33,34]. We found neither the *Foxp3* promoter nor its known enhancers exhibited MLL4 binding in naïve $CD4^+$ T cells (Fig. 7a, top track). Instead, strong MLL4 binding was detected at −8.5kb region of *Foxp3* TSS (Fig. 7a, highlighted by scissors), which exhibited a general decrease in interaction with other genomic regions upon *Mll4* deletion. This MLL4-bound region interacted with the *Foxp3* promoter and several H3K4me1-enriched regions not bound by MLL4, including the 3′UTR and two regions 23-kb and 30-kb upstream to the *Foxp3* (Fig. 7a). Deletion of *Mll4* resulted in substantial decreases in H3K4me1 in *Foxp3* promoter and three putative enhancers not bound by MLL4 (Fig. 7a, highlighted in filled yellow rectangles), suggesting that MLL4 binding at the −8.5kb region regulates H3K4me1 at *Foxp3* promoter and several other putative enhancers through chromatin looping.

To test whether MLL4 binding at the −8.5kb region regulates H3K4me1 at the *Foxp3* promoter and other regions, we deleted this MLL4 binding site in mouse using CRISPR technology[35]. We found that the CRISPR-mediated deletion (Fig. 7b) compromised H3K4me1 at four of the ten H3K4me1 peaks interacting with the MLL4-binding site, including the *Foxp3* promoter, in naïve $CD4^+$ T cells (Fig. 7c, highlighted in red rectangles). In contrast, a significantly lower fraction of H3K4me1 peaks that did not interact with the MLL4 site from the same chromosome showed decrease in H3K4me1 (Fig. 7d).

To test whether this MLL4 binding site contributes to *Foxp3* expression, we isolated naïve $CD4^+$ T cells from the CRISPR deletion mice and stimulated them in iT_{reg} differentiation conditions. *Foxp3* induction was indeed compromised in the CRISPR deletion cells as shown for mRNA (Fig. 7e) and protein (Fig. 7f, g). To examine if the −8.5kb region contributes to the T_{reg} cell development *in vivo*, we analyzed T_{reg} cells in different organs of CRISPR mice. Although the reduction in thymus was not reproducibly observed (Fig. 7h, k), we found modest but significant reduction of $CD4^+Foxp3^+$ cell numbers in spleen and lymph nodes (Fig. 7i, j, l and m). T_{reg} cells in the CRISPR mouse showed more activated phenotype than control littermates defined by increased expression of Helios (Supplementary Fig. 8a–c). The function of T_{reg} cells in CRISPR mouse remained

unaffected as cytokine production by T cells in the periphery of CRISPR mouse was not increased (Supplementary Fig. 8d–g) and *in vitro* suppression assay showed similar ability of CRISPR and control T_{reg} cells to suppress proliferation of effector cells (Supplementary Fig. 8h). The $CD4^+CD25^+Foxp3^-$ single positive thymocytes are considered precursors of mature T_{reg} cells[36–38]. *Mll4* deletion or CRISPR-mediated deletion of the −8.5kb MLL4 binding site resulted in a significant increase in these T_{reg} precursor cells (Fig. 8a–d) and decreased Foxp3 gMFI values in $Foxp3^+$ tT_{reg} cells (Fig. 8b, d). Together, these results suggest that MLL4 binding at the −8.5kb region regulates H3K4me1 of *Foxp3* promoter and contributes to *Foxp3* induction under iT_{reg} conditions and *in vivo* development of T_{reg} cells.

DISCUSSION

T cell differentiation is associated with dynamic changes of histone modifications[39], which are regulated by histone modifying enzymes[40,41], suggesting an important role of these enzymes in T cell development and differentiation. One major H3K4 methylation enzyme is MLL4, which mainly regulates H3K4me1 at enhancers[20]. Using conditional deletion of *Mll4* in $CD4^+$ T cells, we found that MLL4 is a critical regulator of T_{reg} cell development. However, deletion of *Mll*4 after *Foxp3* induction did not alter the expression of *Foxp3 in vitro* and *in vivo*. Furthermore, *in vitro* T_{reg}-suppression assays revealed no difference in suppression capacities of MLL4-deficient T_{reg}s vs. wild type T_{reg}s. Consistent with this observation *Mll4* expression is rapidly down regulated upon T_{reg} differentiation *in vitro*. These results suggested that MLL4 is required for establishing the chromatin structure in naïve $CD4^+$ T cells for future differentiation into T_{reg} lineage, but not required for either sustained *Foxp3* expression or T_{reg} activity.

Although MLL4 may affect a cell population by regulating cell cycle genes[19], our CFSE staining experiments revealed that the proliferative capacity of *Mll4*KO and wild type cells is similar at later stages of differentiation, suggesting that the T_{reg} developmental defect is 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 the H3K4me1 modification patterns at the regulatory elements critical for *Foxp3* induction. Among the regulatory elements for *Foxp3* expression[14,42], we found H3K4me1 enrichment at the promoter region and CNS3 in naïve $CD4^+$ T cells, consistent with previous studies[14]. Although MLL4 did not bind to any of these known *Foxp3* regulatory elements, *Mll4* deletion compromised the H3K4me1 signals at the promoter and enhancer elements of *Foxp3*, suggesting that MLL4 may prime *Foxp3* for activation under appropriate conditions by modulating the H3K4me1 signals at these regions in naïve $CD4^+$ T cells. Indeed, MLL4 may carry out this function by binding to the −8.5kb regions upstream of *Foxp3* TSS. This region is looped to the promoter, CNS3 and 3'UTR region of *Foxp3*. CRISPR-mediated deletion of this region in mice led to decreased H3K4me1 signals at the promoter and 3'UTR regions and resulted in defect of *Foxp3* induction, thus confirming a regulatory role of this region for *Foxp3* expression. Either *Mll4* deletion or CRISPR-mediated deletion of the −8.5kb MLL4 binding site led to a significant increase in the $CD4^+CD25^+Foxp3^-$ T_{reg} precursor cells, thus providing further support to our hypothesis that MLL4 binds to the −8.5kb region to prepare the enhancer landscape required for efficient induction of *Foxp3* during T_{reg} development.

We found that only about half of the affected H3K4me1 peaks by *Mll4* deletion were bound by MLL4. Since no significant changes in the expression of *Mll3* and other HMTs in *Mll4*-deleted cells were detected, we hypothesized that MLL4 may bind to one site and catalyze H3K4 methylation of another remote site *in trans* by chromatin looping. Our data confirmed that the affected H3K4me1 peaks unbound by MLL4 are looped to chromatin regions directly bound by MLL4 and the decrease in H3K4me1 is positively correlated to the number of loops between the unbound sites and MLL4-bound sites. The hypothesis is further confirmed by deletion of the MLL4-bound −8.5kb region of *Foxp3*, which decreased H3K4me1 at the MLL4-unbound *Foxp3* promoter and decreased *Foxp3* induction under the iT_{reg} conditions. Thus, our data support the mechanism that in addition to regulate H3K4me1 at direct binding sites, MLL4 also catalyzes *in trans* H3K4 monomethylation of unbound chromatin regions via chromatin looping to establish an enhancer landscape for cellular differentiation. Although we only demonstrate this kind of activity for MLL4 here, it is very likely that a similar mechanism can be used by other chromatin-modifying enzymes within nucleus.

ONLINE METHODS

Mice

Mll4$^{fl/fl}$ mice on mixed C57BL/6 and 129 backgrounds were kindly provided by K. Ge (NIDDK, National Institutes of Health) and have been previously described[20]. *Cd4*-Cre^{+43} mice on C57BL/6 background were purchased from Taconic. *Foxp3*-CreYFP^{+22} mice on 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 6–10-week-old mice in accordance with the 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 *Mll4*$^{fl/fl}$*Cd4*-Cre$^-$, *Mll4*$^{+/+}$*Cd4*-Cre$^+$, *Mll4*$^{fl/+}$*Cd4*-Cre$^+$ and *Mll4*$^{fl/fl}$*Cd4*-Cre$^+$ mice by magnetic selection according to the protocol provided by manufacturers (Miltenyi Biotech, $CD4^+$ T Cell Isolation Kit, mouse). Naïve $CD4^+$ T cells were purified from lymph nodes of *Mll4*$^{fl/fl}$*Cd4*-Cre$^-$, *Mll4*$^{+/+}$*Cd4*-Cre$^+$, *Mll4*$^{fl/+}$*Cd4*-Cre$^+$, *Mll4*$^{fl/fl}$*Cd4*-Cre$^+$, CRISPR control and CRISPR *Foxp3* –8.5 kb KO mice by magnetic selection according to the protocol provided by manufacturers (Miltenyi Biotech, $CD4^+CD62L^+$ T cell isolation kit II, mouse or Stem Cell, EasySep™ Mouse $CD4^+CD62L^+$ T cell Isolation Kit) or FACS sorted on FACSAria II cell analyzer (BD Biosciences) for $CD4^+CD8^-CD62L^+$ 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 were assessed by FACS analysis for $CD4^+$, $CD8^-$ and $CD62L^+$ on FACSCanto II (BD Biosciences) and analyzed in FlowJo software. Over 98% purity of $CD4^+CD8^-CD62L^+$ cells were considered for further experiments. *Ex vivo* regulatory T cells were purified from lymph nodes and spleen of *Mll4*$^{fl/fl}$*Cd4*-Cre$^-$ and *Mll4*$^{+/+}$*Cd4*-Cre$^+$ mice by flow cytometry sorting on FACSAria II cell analyzer (BD Biosciences) for $CD4^+CD8^-CD25^+$ after pre-enrichment for $CD4^+$ T cells by negative selection using MACS technology (as described above) with accordance to manufacturers'

instructions. All antibodies used for sorting and FACS analyses were purchased from eBiosciences: anti-mouse CD4 (RM4–5), anti-mouse CD8 (53-6.7), anti-mouse CD4 (RM4–5), anti-mouse CD62L (MEL-14). Dead cells were excluded by DAPI staining.

In vitro T_{reg} cell differentiation and FACS analysis

Naïve $CD4^+$ T cells were incubated in T_{reg} cell differentiation condition in the presence of the following: plate-bound anti-CD3 (2 μg/ml, eBioscience), anti-CD28 (3 μg/ml, eBioscience), anti-IL-4 (10 μg/ml, BioXCell), anti-IFN-γ (10 μg/ml, BioXCell), anti-IL-12 (10 μg/ml, BioXcell), IL-2 100 U/ml, Peprotech) and TGF-β (5 ng/ml, R&D Systems). Cells were stained for surface markers with anti-mouse CD4 (RM4–5, eBioscience) and anti-mouse CD8 (53-6.7 eBiosciences). Intracellular staining was performed using Cytofix/Cytoperm 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 (FJK-16s, eBioscience). Dead cells were excluded from analysis using LIVE/DEAD Fixable Aqua Dead Cell Stain Kit (Life Technologies). Sample were run on FACSCanto II cell analyzer (BD Biosciences) and analyzed with FlowJo software.

In vitro Mll4 deletion

A Cre-carrying pCre-GFP retroviral vector was used for deletion of *Mll4* from $Mll4^{fl/fl}$ cells. The retroviral particles were packaged in 293T cells with the pEco packaging plasmid. $Mll4^{fl/fl}$ naive $CD4^+$ T cells were cultured in T_{reg} cell differentiation condition for 24 h. Cells were infected with control pRV-GFP or Cre-carrying pCre-GTP fresh retroviruses on second day of culture. Cells were cultured in T_{reg} cell differentiation conditions for additional two days. Then GFP^+ cells were FACS sorted on FACSAria II cell sorter (BD Biosciences) and analyzed for Foxp3 expression. The efficiency of transduction was 60–90%.

In vitro T_{reg} cell suppression assay

In vitro suppression assay was performed as described previously[44]. Briefly, $CD4^+CD25^+$ T_{reg} cells from lymph nodes and spleen of wild-type and *Mll4* KO mice and $CD4^+CD25^-$ effector T (T_{eff}) cells from wild-type mice were isolated by magnetic selection according to the protocol provided by manufacturers (Miltenyi Biotech, $CD4^+$ T Cell Isolation Kit, mouse) followed by FACS sorting on FACSAria II (BD Biosciences) with following antibodies: anti-mouse CD4 (RM4–5; eBioscience) and andi-mouse CD25 (7D4, BD Biosciences). Prior to stimulation $CD4^+CD25^-$ T_{eff} cells were stained with CFSE according to manufacturer's instructions (ThermoFisher Scientific, CellTrace™ Violet Cell Proliferation Kit, for flow cytometry). $CD4^+CD25^-$ T_{eff} (4×10^4/well) were then stimulated with 0.5 μg/ml soluble anti-CD3 and irradiated APCs (2×10^4/well) in the presence or absence of freshly isolated $CD4^+CD25^+$ T_{reg} cells in U-bottom 96-well plates. At day 3 of culture the T_{eff} cell proliferation was measured by flow cytometry for CFSE dilutions and Foxp3 expression on FACSCanto II cell analyzer (BD Biosciences) and analyzed with FlowJo software. Antibody used for flow cytometry analysis: anti-mouse CD4 (RM4–5; eBioscience) and anti-mouse Foxp3 (FJK-16s, eBioscience). Dead cells were excluded from analysis using LIVE/DEAD Fixable Aqua Dead Cell Stain Kit (Life Technologies).

Lamina propria lymphocytes preparation

Lamina propria lymphocytes (LPLs) were isolated and FACS analyzed as described previously[45]. Briefly, small intestines without Peyer's patches were disrupted mechanistically. Segments of tissues were washed extensively followed by incubation for 20 min at 37 °C with vigorous shacking in pre-warmed RPM medium supplemented with 3% FBS, 5 mM EDTA and dithiothreitol at 0.145 mg/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 and cell suspensions were resuspended and analyzed by flow cytometry.

Intracellular cytokine staining

IL-17a and IFN-γ cytokine expression were assessed by flow cytometry as described previously[45]. Briefly, 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 protein transport inhibitor GolgiPlug (BD Pharmingen). Cells were stained for surface markers with following antibodies: anti-mouse TCRβ (H57–597; eBioscience), anti-mouse CD4 (RM4–5; eBioscience), anti-mouse CD45 (30-F11, eBioscience). Cells were washed and fixed using Cytofix/Cytoperm buffer (BD Pharmingen 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 (FJK-16s, eBioscience) and anti-human/mouse T-bet (eBio4B10, eBiosciences). Dead cells were excluded from analysis using Zombie Yellow Fixable Viability Kit (Biolegend) or LIVE/DEAD Fixable Aqua Dead Cell Stain Kit (Life Technologies).

ChIP-seq

ChIP-seq assays with two independent experiments if not stated otherwise on naïve $CD4^+$ T cells were performed as described previously[46,47]. Briefly, cells for ChIP-seq were fixed for 10 min in 1% formaldehyde, sonicated and chromatin immunoprecipitation was performed with anti-MLL4[19,20], anti-H3K27ac (Abcam; single experiment) and anti-H3K27me3 (07-449, Millipore) antibodies. 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 (Epicentre), indexed, amplified and sequenced on an Illumina 2G Genome Analyzer.

RNA-Seq

RNA-Seq experiments were previously described[48]. PolyA RNA was isolated from naïve $CD4^+$ T cells and T_{reg} cells with a Dynabeads mRNA Direct Kit (610.12, Invitrogen) following protocol provided by manufacturer. mRNA was reverse transcribed with the use of Super Script Double-Stranded cDNA Synthesis Kit (Invitrogen) and Random Hexamer Primers (Invitrogen). dscDNA was sonicated on Bioruptor (Diagenode), blunt-ended using an End-It DNA-Repair kit (Epicentre), indexed, amplified and sequenced on an Illumina 2G

Genome Analyzer. The RNA-Seq experiments were done with two independent experiments.

Immunoblot

$CD4^+$ 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 incubated with anti-H3K4me1 (ab8895, Abcam), anti-K4me2 (ab32356, Abcam), anti-K4me3 (17-614, Millipore), anti-MLL4 and anti-actin beta, anti-Rbbp5 and anti-panH3 as a loading control. Blots were visualized with Pierce ECL Western Blotting Substrate (Thermo Scientific).

A modified Hi-C protocol

The detailed procedure of the modified Hi-C protocol (called meHi-C) was described in another manuscript[31]. We applied the technique to naïve $CD4^+$ T cells which were crossed linked with formaldehyde. Cells were lysed and digested with CviQ I + CviA II + Bfa I for 20 min. The Hi-C samples were processed following the *in situ* Hi-C protocol[30] with modifications briefly described as follows: DNA ends were marked by biotin-14-dATP with Klenow (large) for 1 h at 37 °C. Blunt-end 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 addition of "A"[46]. Biotin-labeled DNA was pull-downed by streptavidin beads followed by Illumina adapter 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 Foxp3 −8.5kb MLL4 binding site using CRISPR/Cas9 technology

We used the CRISPR/Cas9 technology to delete the putative 500-bp *Mll4* binding site at ~8.5-kb upstream of the *Foxp3* gene. To increase the chance of deletion, we designed two Crispr sgRNAs to cut each end of this 500-bp region. The sgRNAs for cutting the upstream end are 5′-GCCATGAGGATGTAGTCCAG-3′ and 5′-CTTCTGACCCTACCTGCCAA-3′; and the sgRNAs for cutting the downstream end are 5′-TGGACGGTACTGACCCCCGA-3′ and 5′-TGAAATGCAGGCGATTCTGG-3′. These four sgRNA target sequences were cloned by OriGene Technology into the pT7-Guide-IVT vector. Injectable sgRNAs were *in vitro* transcribed using the MEGAshortscript T7 Kit (Life Technologies), and the Cas9 mRNA was *in vitro* transcribed from plasmid MLM3613 (Addgene #42251) using the mMESSAGE mMACHINE T7 Kit (Life Technologies), as previously reported[35]. These four sgRNAs were co-injected with Cas9 mRNA into fertilized eggs collected from B6CBAF1/J mice (JAX) at the 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% CO_2. In the next morning, those embryos that reached 2-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 (TSS). Gene body regions were defined as starting from TSS to transcription ending 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 parameters[49]. 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 bps and a gap size of 400 bps[27]. A conservative set of MLL4 binding sites were identified with stringent parameters and input control: E value = 1,000 and FDR = 0.001. Here, E-value refers to the number of expected peaks by assuming a random distribution of the ChIP-Seq reads[27]. 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 *Mll4* deletion and control cells were predicted by EdgerR (FDR < 0.001; FC > 2)[50]. The ROSE program[51] was applied to the H3K27ac ChIP-Seq data to identify super-enhancers in mouse naïve $CD4^+$ T cells with the "-t" 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 kilo-base of exon model per million reads)[52] with in-house script. The calculation of the fold change of gene expression excluded genes that showed RPKM less than one in both the KO cells and the control cells. Differential expressed genes were identified by a Fold change of > 1.5 and an expression value of > 3 for at least one of the two conditions: *Mll4*KO and control.

Hi-C data analysis—Pair-end short reads were mapped to the mouse genome (mm9) by Bowtie2 with default parameters[49]. 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 examined the significance of interaction for bin-pairs that are separated by at least L bps and are 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. Briefly, for each observed intra-chromosomal PET that linked two bins, we randomly assigned it to two bins from the same chromosome, but required that 1) the GC contents from the new positions were similar to the

original (with a fluctuation up to 2%); 2) the mappability scores of the new bins were similar to those of the original (allowing a fluctuation up to 0.05); and 3) 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 for 5000 times.

With the background models, we calculated the expected number of bin-pairs linked by at least n PETs and separated by at least L-bp and compared it to the observation. For instance, 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, corresponding to an enrichment ratio (ε) of 3 = 11,237/3,608. We used the enrichment ratio as a threshold for interaction call. As expected, a lower number of n will be needed as L increased to achieve the same ε. This led us to define a threshold step function that depended on distance L to ascertain the minimal number of PETs n required for a bin-pair interaction call under $\varepsilon = 3$. An ε of 3 means that out of three observations, one could be explained by the background model and is therefore likely false positive, while the remaining cannot and are therefore likely true positives. In other words, an ε of 3 corresponds to an estimated specificity (SP) of 67% (($\varepsilon-1$)/ ε). We noticed that the ε 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 $\varepsilon = 3$ for chromosome 1, in contrast to 882 bin-pairs from the background model, corresponding to an estimated SP = 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 $\varepsilon = 3$. For a given genomic bin of 1-kb, we estimated its interaction intensity by the number of PETs linked to any other interacted bins outsides this region, normalized by the total number of PETs from interacted bin-pairs across the genome. To generate a smoothed view (e.g., 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 number of PETs linking to any other interacted bins (Fig. 5b–d) or any other interacted bins with certain features (e.g., overlapping with enhancers; Fig.6d and 6e) from the same chromosome for each condition. The EdgeR package[50] was applied to identify regions showing differential interaction between *Mll4* KO cells and control cells (FC > 1.5 and FDR < 0.05).

To investigate the relationship between the change in H3K4me1 abundance at MLL4-unbound genomic region and the change in interaction intensity with MLL4-bound remote sites (Fig. 5e), we sorted the H3K4me1 regions based on their response to *Mll4* deletion in terms of the change in H3K4me1 abundance. For each H3K4me1 region, the numbers of PETs linking the region to interacted MLL4 binding sites were recorded for the *Mll4*-deleted

cells and control cells and were transformed into a fold change. The distributions of the fold change were then compared among the three groups of H3K4me1 regions.

Gene ontology enrichment analysis—A gene is called as 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 Resource[53] or GREAT[28]. Redundant GO terms from the DAVID output were removed by using REVIGO[54].

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 upon request.

Statistics

We applied two-sided Kolmogorov-Smirnov (KS; two-tailed throughout the manuscript) test to assess the difference in the accumulated distribution of ChIP-Seq read density (Fig. 4g, Supplementary Fig. S6b), FC of ChIP-Seq read density (Fig. 4h, Supplementary Fig. 6d, 7b and 7c), FC of interaction intensity (Fig. 5e, 6f and 6g, Supplementary Fig. 7a), and FC of gene expression (Fig. 6h, Supplementary Fig.7c and 7d) between any two groups of interests. KS-test is a nonparametric method, testing whether two probability distributions differ, and requires no prior knowledge about the distributions[55]. In all these figures (accumulative distribution), the y-axis of the accumulative distribution of a measurement across a set refers to the % within the set that show an amount less than the one specified in the x-axis. A line shifting to the left side means a general less amount in that measurement. The $\chi 2$ test was used for comparison of two portions from independent samples, presented as a percentage (Fig.5b and 7d). Statistical significance for other figures was calculated by ANOVA test or Kruskal–Wallis when sample did not meet Gaussian distribution assessed by 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 upon request.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

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 service; the NHLBI Flow Cytometry Core facility for cell sorting; the NHLBI and NIAID Animal Facilities for animal care; X. Wang and S Gao from NHLBI Systems Biology Core and X. Zheng from Carnegie Institution for Science for discussion on Hi-C data analysis. The work was supported by Division of Intramural Research of NHLBI (K.Z.), NIDCR (W.C.) and NIDDK (K.G.) of NIH.

References

1. Sakaguchi S, Yamaguchi T, Nomura T, Ono M. Regulatory T cells and immune tolerance. Cell. 2008; 133:775–787. DOI: 10.1016/j.cell.2008.05.009 [PubMed: 18510923]
2. Klein L, Jovanovic K. Regulatory T cell lineage commitment in the thymus. Seminars in immunology. 2011; 23:401–409. DOI: 10.1016/j.smim.2011.06.003 [PubMed: 21733719]
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. The Journal of experimental medicine. 2003; 198:1875–1886. DOI: 10.1084/jem.20030152 [PubMed: 14676299]
4. Fontenot JD, Gavin MA, Rudensky AY. Foxp3 programs the development and function of CD4+CD25+ regulatory T cells. Nature immunology. 2003; 4:330–336. DOI: 10.1038/ni904 [PubMed: 12612578]
5. Khattri R, Cox T, Yasayko SA, Ramsdell F. An essential role for Scurfin in CD4+CD25+ T regulatory cells. Nature immunology. 2003; 4:337–342. DOI: 10.1038/ni909 [PubMed: 12612581]
6. Bennett CL, et al. The immune dysregulation, polyendocrinopathy, enteropathy, X-linked syndrome (IPEX) is caused by mutations of FOXP3. Nature genetics. 2001; 27:20–21. DOI: 10.1038/83713 [PubMed: 11137993]
7. Kim JM, Rasmussen JP, Rudensky AY. Regulatory T cells prevent catastrophic autoimmunity throughout the lifespan of mice. Nature immunology. 2007; 8:191–197. DOI: 10.1038/ni1428 [PubMed: 17136045]
8. Lahl K, et al. Selective depletion of Foxp3+ regulatory T cells induces a scurfy-like disease. The Journal of experimental medicine. 2007; 204:57–63. DOI: 10.1084/jem.20061852 [PubMed: 17200412]
9. Grindebacke H, et al. Defective suppression of Th2 cytokines by CD4CD25 regulatory T cells in birch allergics during birch pollen season. Clinical and experimental allergy : journal of the British Society for Allergy and Clinical Immunology. 2004; 34:1364–1372. DOI: 10.1111/j.1365-2222.2004.02067.x [PubMed: 15347368]
10. Suvas S, Rouse BT. Treg control of antimicrobial T cell responses. Current opinion in immunology. 2006; 18:344–348. DOI: 10.1016/j.coi.2006.03.005 [PubMed: 16616481]
11. Roychoudhuri R, Eil RL, Restifo NP. The interplay of effector and regulatory T cells in cancer. Current opinion in immunology. 2015; 33:101–111. DOI: 10.1016/j.coi.2015.02.003 [PubMed: 25728990]
12. Kasagi S, et al. In vivo-generated antigen-specific regulatory T cells treat autoimmunity without compromising antibacterial immune response. Science translational medicine. 2014; 6:241ra278.
13. Taylor PA, Lees CJ, Blazar BR. The infusion of ex vivo activated and expanded CD4(+)CD25(+) immune regulatory cells inhibits graft-versus-host disease lethality. Blood. 2002; 99:3493–3499. [PubMed: 11986199]
14. Zheng Y, et al. Role of conserved non-coding DNA elements in the Foxp3 gene in regulatory T-cell fate. Nature. 2010; 463:808–812. DOI: 10.1038/nature08750 [PubMed: 20072126]
15. Schmidl C, et al. The enhancer and promoter landscape of human regulatory and conventional T-cell subpopulations. Blood. 2014; 123:e68–78. DOI: 10.1182/blood-2013-02-486944 [PubMed: 24671953]
16. Samstein RM, et al. Foxp3 exploits a pre-existent enhancer landscape for regulatory T cell lineage specification. Cell. 2012; 151:153–166. DOI: 10.1016/j.cell.2012.06.053 [PubMed: 23021222]
17. Bannister AJ, Kouzarides T. Regulation of chromatin by histone modifications. Cell research. 2011; 21:381–395. DOI: 10.1038/cr.2011.22 [PubMed: 21321607]
18. Hess JL. Mechanisms of transformation by MLL. Critical reviews in eukaryotic gene expression. 2004; 14:235–254. [PubMed: 15663355]
19. Ang SY, et al. KMT2D regulates specific programs in heart development via histone H3 lysine 4 di-methylation. Development. 2016; 143:810–821. DOI: 10.1242/dev.132688 [PubMed: 26932671]
20. Lee JE, et al. H3K4 mono- and di-methyltransferase MLL4 is required for enhancer activation during cell differentiation. eLife. 2013; 2:e01503. [PubMed: 24368734]

21. Akimova T, Beier UH, Wang L, Levine MH, Hancock WW. Helios expression is a marker of T cell activation and proliferation. PloS one. 2011; 6:e24226. [PubMed: 21918685]
22. Rubtsov YP, et al. Regulatory T cell-derived interleukin-10 limits inflammation at environmental interfaces. Immunity. 2008; 28:546–558. DOI: 10.1016/j.immuni.2008.02.017 [PubMed: 18387831]
23. Shevach EM, Thornton AM. tTregs, pTregs, and iTregs: similarities and differences. Immunol Rev. 2014; 259:88–102. DOI: 10.1111/imr.12160 [PubMed: 24712461]
24. Hu D, et al. The MLL3/MLL4 branches of the COMPASS family function as major histone H3K4 monomethylases at enhancers. Molecular and cellular biology. 2013; 33:4745–4754. DOI: 10.1128/MCB.01181-13 [PubMed: 24081332]
25. Lee J, et al. A tumor suppressive coactivator complex of p53 containing ASC-2 and histone H3-lysine-4 methyltransferase MLL3 or its paralogue MLL4. Proceedings of the National Academy of Sciences of the United States of America. 2009; 106:8513–8518. DOI: 10.1073/pnas.0902873106 [PubMed: 19433796]
26. Lee S, Lee J, Lee SK, Lee JW. Activating signal cointegrator-2 is an essential adaptor to recruit histone H3 lysine 4 methyltransferases MLL3 and MLL4 to the liver X receptors. Molecular endocrinology. 2008; 22:1312–1319. DOI: 10.1210/me.2008-0012 [PubMed: 18372346]
27. Zang C, et al. A clustering approach for identification of enriched domains from histone modification ChIP-Seq data. Bioinformatics. 2009; 25:1952–1958. DOI: 10.1093/bioinformatics/btp340 [PubMed: 19505939]
28. McLean CY, et al. GREAT improves functional interpretation of cis-regulatory regions. Nature biotechnology. 2010; 28:495–501. DOI: 10.1038/nbt.1630
29. Jin W, et al. Genome-wide detection of DNase I hypersensitive sites in single cells and FFPE tissue samples. Nature. 2015; 528:142–146. DOI: 10.1038/nature15740 [PubMed: 26605532]
30. Rao SS, et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell. 2014; 159:1665–1680. DOI: 10.1016/j.cell.2014.11.021 [PubMed: 25497547]
31. Ren G, et al. CTCF-mediated enhancer-promoter interaction is a critical regulator of cell-to-cell variation of gene expression. Molecular cell. 2017 minor revision.
32. Hnisz D, et al. Super-enhancers in the control of cell identity and disease. Cell. 2013; 155:934–947. DOI: 10.1016/j.cell.2013.09.053 [PubMed: 24119843]
33. Feng Y, et al. Control of the inheritance of regulatory T cell identity by a cis element in the Foxp3 locus. Cell. 2014; 158:749–763. DOI: 10.1016/j.cell.2014.07.031 [PubMed: 25126783]
34. Ogawa C, et al. TGF-beta-mediated Foxp3 gene expression is cooperatively regulated by Stat5, Creb, and AP-1 through CNS2. Journal of immunology. 2014; 192:475–483. DOI: 10.4049/jimmunol.1301892
35. Wang H, et al. One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. Cell. 2013; 153:910–918. DOI: 10.1016/j.cell.2013.04.025 [PubMed: 23643243]
36. Burchill MA, et al. Linked T cell receptor and cytokine signaling govern the development of the regulatory T cell repertoire. Immunity. 2008; 28:112–121. DOI: 10.1016/j.immuni.2007.11.022 [PubMed: 18199418]
37. Kitagawa Y, et al. Guidance of regulatory T cell development by Satb1-dependent super-enhancer establishment. Nature immunology. 2017; 18:173–183. DOI: 10.1038/ni.3646 [PubMed: 27992401]
38. Lio CW, Hsieh CS. A two-step process for thymic regulatory T cell development. Immunity. 2008; 28:100–111. DOI: 10.1016/j.immuni.2007.11.021 [PubMed: 18199417]
39. Wei G, et al. Global mapping of H3K4me3 and H3K27me3 reveals specificity and plasticity in lineage fate determination of differentiating CD4+ T cells. Immunity. 2009; 30:155–167. DOI: 10.1016/j.immuni.2008.12.009 [PubMed: 19144320]
40. Yamashita M, et al. Crucial role of MLL for the maintenance of memory T helper type 2 cell responses. Immunity. 2006; 24:611–622. DOI: 10.1016/j.immuni.2006.03.017 [PubMed: 16713978]

41. Tumes DJ, et al. The polycomb protein Ezh2 regulates differentiation and plasticity of CD4(+) T helper type 1 and type 2 cells. Immunity. 2013; 39:819–832. DOI: 10.1016/j.immuni.2013.09.012 [PubMed: 24238339]
42. Li X, Liang Y, LeBlanc M, Benner C, Zheng Y. Function of a Foxp3 cis-element in protecting regulatory T cell identity. Cell. 2014; 158:734–748. DOI: 10.1016/j.cell.2014.07.030 [PubMed: 25126782]
43. Lee PP, et al. A critical role for Dnmt1 and DNA methylation in T cell development, function, and survival. Immunity. 2001; 15:763–774. [PubMed: 11728338]
44. Zhang P, et al. PARP-1 controls immunosuppressive function of regulatory T cells by destabilizing Foxp3. PloS one. 2013; 8:e71590. [PubMed: 23977081]
45. Zanvit P, et al. Antibiotics in neonatal life increase murine susceptibility to experimental psoriasis. Nature communications. 2015; 6:8424.
46. Barski A, et al. High-resolution profiling of histone methylations in the human genome. Cell. 2007; 129:823–837. DOI: 10.1016/j.cell.2007.05.009 [PubMed: 17512414]
47. Wei G, et al. Genome-wide analyses of transcription factor GATA3-mediated gene regulation in distinct T cell types. Immunity. 2011; 35:299–311. DOI: 10.1016/j.immuni.2011.08.007 [PubMed: 21867929]
48. 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. 2009; 37:e106. [PubMed: 19528076]
49. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012; 9:357–359. DOI: 10.1038/nmeth.1923 [PubMed: 22388286]
50. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010; 26:139–140. DOI: 10.1093/bioinformatics/btp616 [PubMed: 19910308]
51. Whyte WA, et al. Master transcription factors and mediator establish super-enhancers at key cell identity genes. Cell. 2013; 153:307–319. DOI: 10.1016/j.cell.2013.03.035 [PubMed: 23582322]
52. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods. 2008; 5:621–628. DOI: 10.1038/nmeth.1226 [PubMed: 18516045]
53. Jiao X, et al. DAVID-WS: a stateful web service to facilitate gene/protein list analysis. Bioinformatics. 2012; 28:1805–1806. DOI: 10.1093/bioinformatics/bts251 [PubMed: 22543366]
54. Supek F, Bosnjak M, Skunca N, Smuc T. REVIGO summarizes and visualizes long lists of gene ontology terms. PloS one. 2011; 6:e21800. [PubMed: 21789182]
55. Hollander, M., Wolfe, DA. Nonparametric statistical methods. Wiley; 1973.

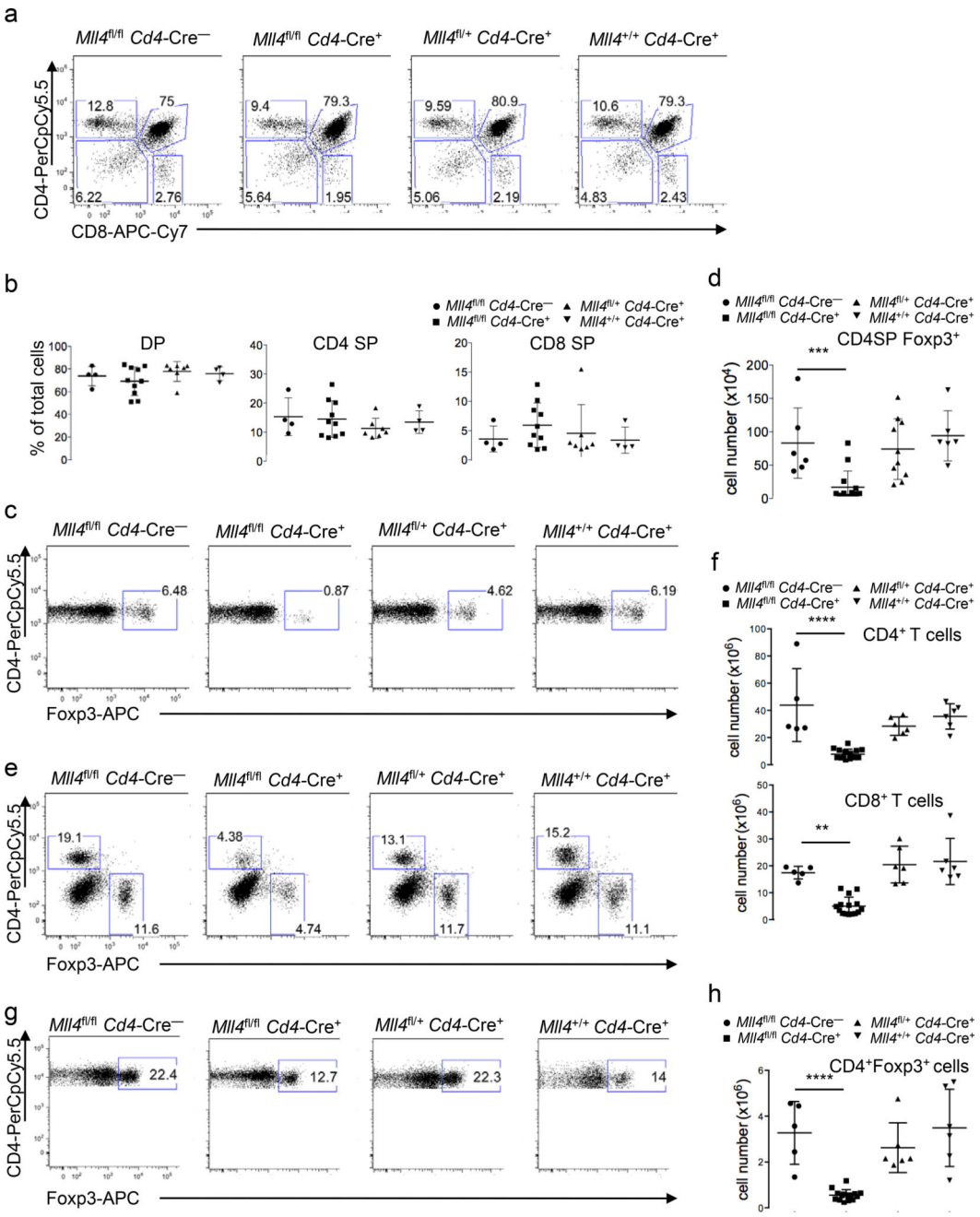

Fig. 1. *Mll4* deficiency reduces T_{reg} cell numbers in the thymus and T cell numbers in the periphery

(a) Representative flow cytometry plots of CD4 SP, CD8 SP and DP T cell populations in the thymus of *Mll4*$^{fl/fl}$*Cd4*-Cre^{-}, *Mll4*$^{fl/fl}$*Cd4*-Cre^{+}, *Mll4*$^{fl/+}$*Cd4*-Cre^{+} and *Mll4*$^{+/+}$*Cd4*-Cre^{+} mice. Shown is one representative experiment (n=4 independent experiments).

(b) Percentages of CD4 SP, CD8 SP and DP T cells in the thymus of *Mll4*$^{fl/fl}$*Cd4*-Cre^{-}, *Mll4*$^{fl/fl}$*Cd4*-Cre^{+}, *Mll4*$^{fl/+}$*Cd4*-Cre^{+} and *Mll4*$^{+/+}$*Cd4*-Cre^{+} mice. Error bars: standard deviations. Center line: mean.

(c) Representative flow cytometry plots of CD4 SP Foxp3$^+$ cells in the thymus of *Mll4*$^{fl/fl}$*Cd4*-Cre$^-$, *Mll4*$^{fl/fl}$*Cd4*-Cre$^+$, *Mll4*$^{fl/+}$*Cd4*-Cre$^+$ and *Mll4*$^{+/+}$*Cd4*-Cre$^+$ mice. Shown is one representative experiment (n=4 independent experiments).
(d) Percentages of CD4 SP Foxp3$^+$ cells in the thymus of *Mll4*$^{fl/fl}$*Cd4*-Cre$^-$, *Mll4*$^{fl/fl}$*Cd4*-Cre$^+$, *Mll4*$^{fl/+}$*Cd4*-Cre$^+$ and *Mll4*$^{+/+}$*Cd4*-Cre$^+$ mice. ***$P \leq 0.001$ (Kruskal-Wallis test). Error bars: standard deviations.
(e) Representative flow cytometry plots of CD4$^+$ and CD8$^+$ T cells in the spleen of *Mll4*$^{fl/fl}$*Cd4*-Cre$^-$, *Mll4*$^{fl/fl}$*Cd4*-Cre$^+$, *Mll4*$^{fl/+}$*Cd4*-Cre$^+$ and *Mll4*$^{+/+}$*Cd4*-Cre$^+$ mice. Shown is one representative experiment (n=5 independent experiments).
(f) Quantification of total CD4$^+$ (upper panel) and CD8$^+$ (lower panel) T cell numbers in the spleen of *Mll4*$^{fl/fl}$*Cd4*-Cre$^-$, *Mll4*$^{fl/fl}$*Cd4*-Cre$^+$, *Mll4*$^{fl/+}$*Cd4*-Cre$^+$ and *Mll4*$^{+/+}$*Cd4*-Cre$^+$ mice. Center line: mean. **$P \leq 0.01$ and ****$P \leq 0.0001$ (Kruskal-Wallis test)
(g) Representative flow cytometry plots of CD4$^+$Foxp3$^+$ cells in the spleen of *Mll4*$^{fl/fl}$*Cd4-Cre*$^-$, *Mll4*$^{fl/fl}$*Cd4*-Cre$^+$, *Mll4*$^{fl/+}$*Cd4*-Cre$^+$ and *Mll4*$^{+/+}$*Cd4*-Cre$^+$ mice. Shown is one representative experiment (n=5 independent experiments).
(h) Quantification of total CD4$^+$Foxp3$^+$ cell numbers in the spleen of *Mll4*$^{fl/fl}$*Cd4*-Cre$^-$, *Mll4*$^{fl/fl}$*Cd4*-Cre$^+$, *Mll4*$^{fl/+}$*Cd4*-Cre$^+$ and *Mll4*$^{+/+}$*Cd4*-Cre$^+$ mice. ****$P \leq 0.0001$ (Kruskal-Wallis test). Error bars: standard deviations. Center line: mean.

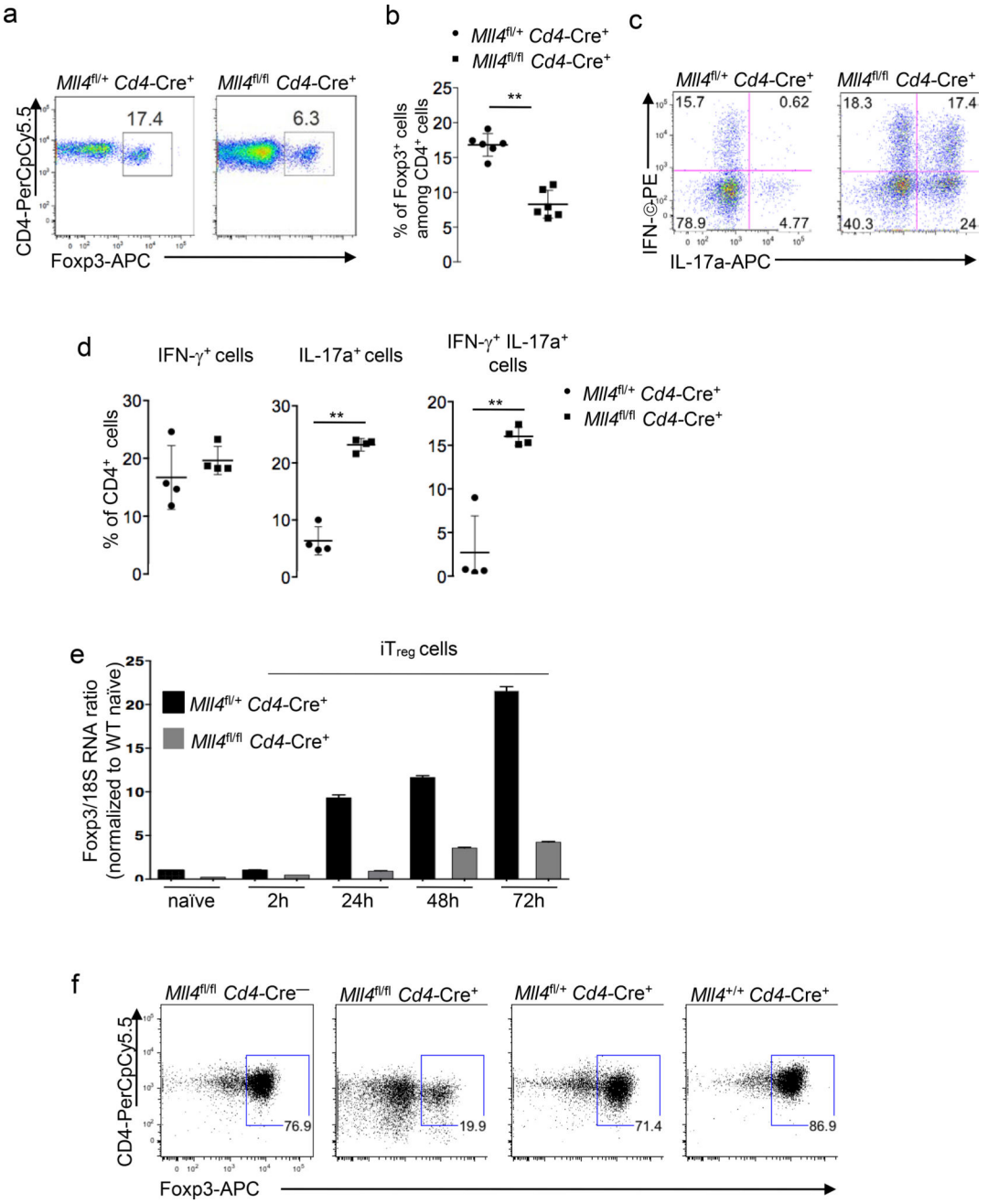

Fig. 2. *Mll4* deletion results in impaired T_{reg} development in the gut and *in vitro* culture

(a) Representative flow cytometry plots of CD4$^+$Foxp3 homozygous CRISPR deletion cells in small intestine lamina propria of *Mll4*$^{fl/+}$*Cd4*-Cre$^+$ (WT) and *Mll4*$^{fl/fl}$*Cd4*-Cre$^+$ (*Mll4*KO) mice. Shown is one representative experiment (n=4 independent experiments).

(b) Percentages of CD4$^+$Foxp3$^+$ cells in small intestine lamina propria from mice as in (a). Error bars: standard deviations. Center line: mean.

(c) Representative flow cytometry plots of IFN-γ$^+$ and IL-17a$^+$ CD4$^+$ T cells in small intestine lamina propria of *Mll4*$^{fl/+}$*Cd4*-Cre$^+$ (WT) and *Mll4*$^{fl/fl}$*Cd4*-Cre$^+$ (*Mll4*KO) mice.

Gated $CD4^+$ T cells. Shown is one representative experiment (n=4 independent experiments).

(d) Percentages of IFN-γ^+ (left column), IL-$17a^+$ (middle column) and IFN-γ^+IL-$17a^+$ (right column) $CD4^+$ T cells in small intestine lamina propria of *Mll4*$^{fl/+}$*Cd4*-Cre^+ (WT) and *Mll4*$^{fl/fl}$*Cd4*-Cre^+ (*Mll4*KO) mice. Error bars: standard deviations. Center line: mean.

(e) RT-PCR results of *Foxp3* expression relative to 18S in naive $CD4^+$ T cells isolated from *Mll4*$^{fl/fl}$*Cd4*-Cre^- (WT) and *Mll4*$^{fl/fl}$*Cd4*-Cre^+ (*Mll4*KO) mice and stimulated under T_{reg} cell-inducing conditions for indicated periods of time. Shown is one representative experiment (n=2 independent experiments). Error bars: standard deviations.

(f) Flow cytometry analysis of intracellular staining for Foxp3 and CD4 in *in vitro* generated iT_{reg} cells, for 3 days as in **e**. Shown is one representative experiment (n=4 independent experiments).

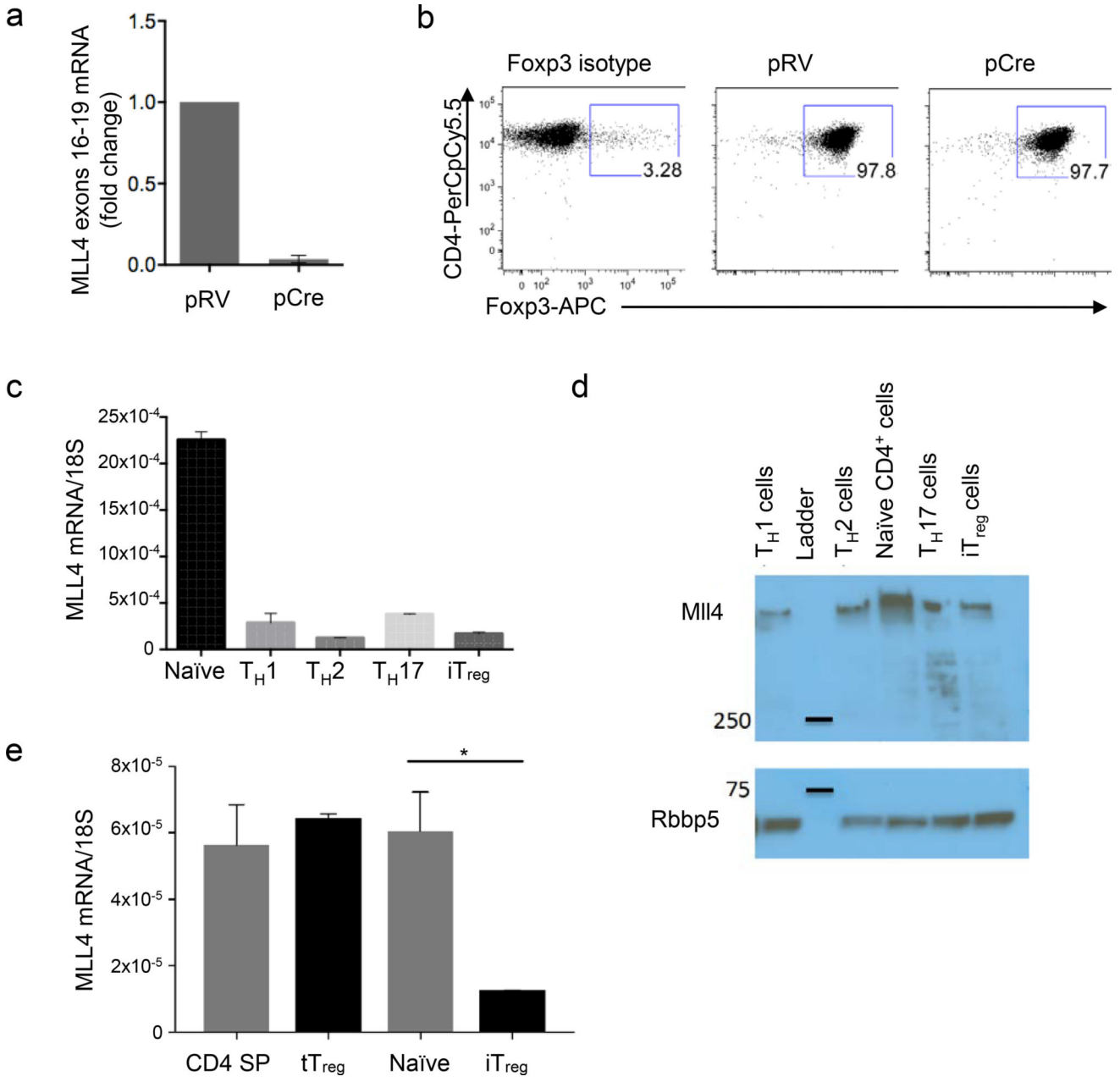

Fig. 3. *Mll4* expression pattern in different CD4$^+$ T cell types

(a) Expression of *Mll4* floxed alleles (16–19) in T_{reg} cells generated from *Mll4*$^{fl/fl}$ naive CD4$^+$ T cells infected with a Cre-carrying pCre or a control pRV retrovirus at day 2 of culture, assessed on GFP$^+$ cells sorted two days post infection. n=3, error bars: standard deviations.

(b) Representative flow cytometry plot of Foxp3 expression in cells from (**a**) two days post-infection with pRV (middle column) or pCre virus (right column). Shown is one representative experiment (n=3 independent experiments).

(c) *Mll4* expression in naïve CD4$^+$ T cells and T_H1, T_H2, T_H17 and T_{reg} cells differentiated *in vitro* from naïve CD4$^+$ T cells for 3 days. Normalized to 18S. Each *in vitro* differentiation

condition was repeated at least two times in independent experiments. Error bars: standard deviations.

(d) Western Blot for Ml4 and Rbbp5 as an endogenous control on naive $CD4^+$ T cells and *in vitro* generated T_H1, T_H2, T_H17 and iT_{reg} cell for 4 days. A representative blot of 2 independent experiments.

(e) Mll4 expression in $CD4^+CD8^-Foxp3GFP^+$, $CD4^+CD8^-Foxp3GFP^-$ and $CD4^+CD8^-Foxp3GFP^-CD62L^+$ T cells of thymus and periphery and in vitro generated Treg by RT-PCR. CD4SP: CD4 single positive cells $CD4^+CD8^-$ $Foxp3GFP^-$ from the thymus, tT_{reg}: $CD4^+CD8^-Foxp3GFP^+$ cells from the thymus, naive: $CD4^+CD8^-Foxp3GFP^-CD62L^+$ cells from lymph nodes and spleen, iT_{reg}: *in vitro* induced T_{reg} (cells were not sorted for GFP^+ at day3 but the whole culture was harvested, 80% of cell were GFP^+). * p value < 0.05 was calculated by Kruskal-Wallis test. Error bars: standard deviations.

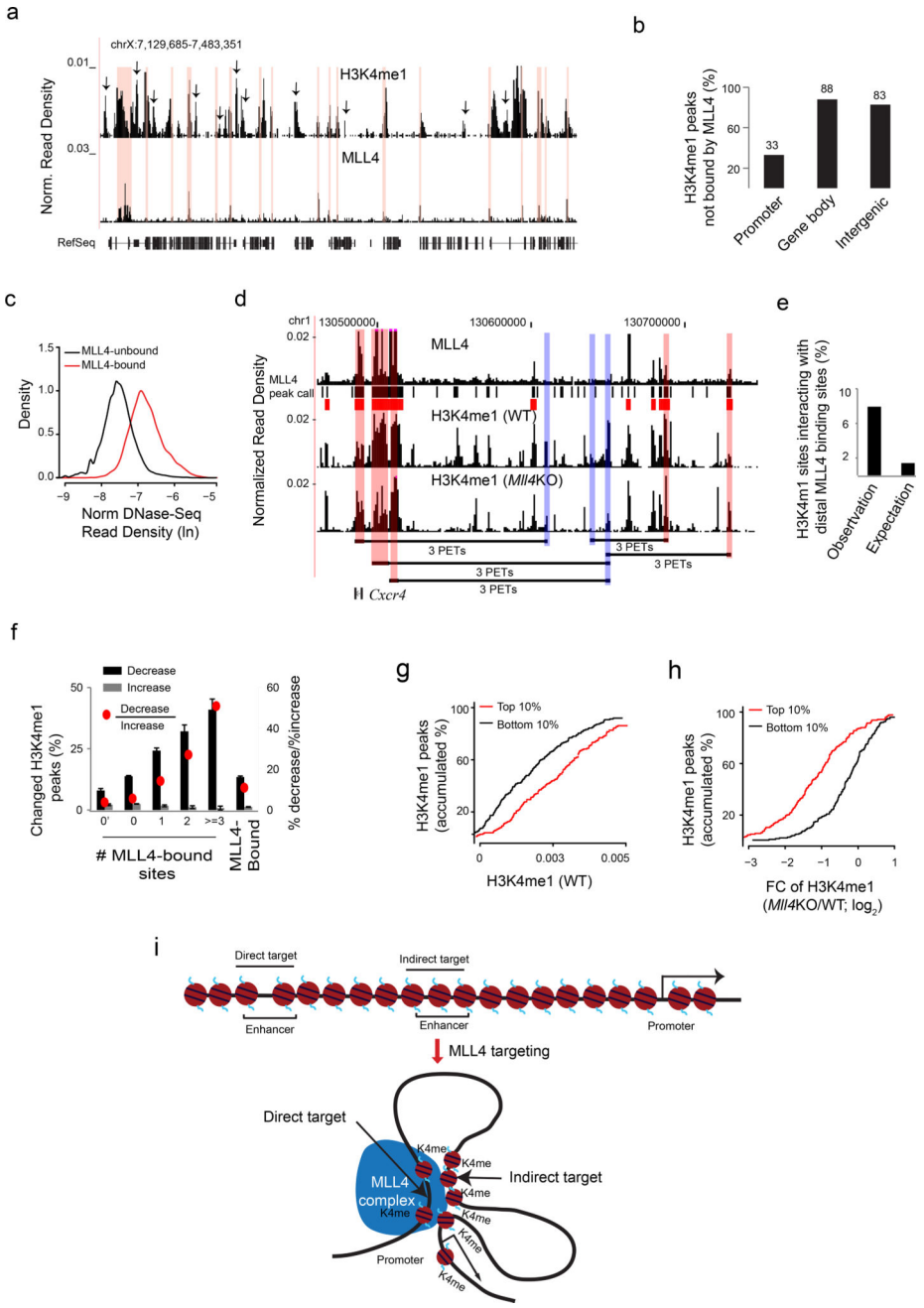

Fig. 4. *Mll4* deletion reduces H3K4me1 at MLL4-unbound regions via chromatin lopping

(a) A UCSC Genome Browser image showing the distributions of ChIP-Seq reads for H3K4me1 (upper track) and MLL4 (lower track) across a genomic region of 0.4 million base-pairs of the X chromosome in naïve CD4$^+$ T cells. The highlighted regions in pink indicate MLL4 binding peaks detected from ChIP-Seq data under stringent settings. Arrowheads indicate the regions enriched with H3K4me1 but free of MLL4 binding or with only background level of MLL4 binding. *y*-axis: number of reads per base pair per million reads. Representative of n = 2 independent experiments.

(b) Percentages of H3K4me1-enriched genomic regions (pooled from n = 2 independent experiments, also applied to panels **e–h**) not bound by MLL4, of which the binding sites were predicted from ChIP-Seq data with stringent settings.

(c) Distributions of the normalized DNase-Seq read density for H3K4m1-enriched regions that are bound by MLL4 (red line) and are not bound by MLL4 (black line). n = 1 experiment.

(d) A UCSC Genome Browser image showing putative enhancers marked by H3K4me1 but not bound by MLL4 (highlighted in blue rectangles), located upstream to *Cxcr4*. The ChIP-Seq data for MLL4, H3K4me1 (WT cells) and H3K4me1 (*Mll4*KO cells) are shown in the three upper tracks. High-confidence chromatin interactions between the enhancers and MLL4-bound regions are indicated as a horizontal line linking two filled rectangles below the ChIP-Seq tracks. The numbers below the lines indicate the numbers of Hi-C PETs linking two regions. Black bars and red bars under the MLL4 track represent MLL4 peaks called under non-stringent settings and under stringent settings, respectively. WT: $Mll4^{fl/fl}$-*Cd4*-Cre^{-} and *Mll4*KO: $Mll4^{fl/fl}$-*Cd4*-Cre^{+} (throughout the figure); Representative of n = 2 independent experiments.

(e) Percentages of MLL4-unbound H3K4me1-enriched regions that showed interaction with distant genomic sites directly bound by MLL4, with expectation estimated from random interactions corrected for biases from GC contents, mappability and distance. $P < 0.01$ by χ^2-test.

(f) The left *y*-axis shows the percentage of MLL4-unbound H3K4me1 regions that show a decrease (black bars) or increase (grey bars) in H3K4me1 content upon *Mll4* deletion, sorted into four groups (marked by from “0” to “>=3”) based on the number of interacting distant regions directly bound by MLL4. Group “0’”: a subset of MLL4-unbound regions that show no sign of interaction (0 PET) with any other regions. The right *y*-axis (in red) shows the ratio of % decrease to % increase for each group. MLL4-bound H3K4me1 regions included for comparison. Error bars: standard deviations estimated by bootstrapping.

(g) Accumulative distribution of H3K4me1 abundance in WT cells for the top 10% and bottom 10% groups of MLL4-unbound H3K4me1-enriched regions that interacted with distant MLL4-bound regions, sorted based on the aggregated MLL4 binding from the distant sites. P-value = 5e-5 by Kolmogorov-Smirnov test.

(h) Accumulative changes of H3K4me1 abundance upon *Mll4* deletion for the two groups of MLL4-unbound H3K4me1 peaks defined in panel **(g)**. P-value = 4e-16 by Kolmogorov-Smirnov test.

(i) Schematic representation of remote regulation of H3K4me1 by MLL4. MLL4 not only generates H3K4me1 at its direct target sites but also at indirect target sites by methylation *in trans* via chromatin looping.

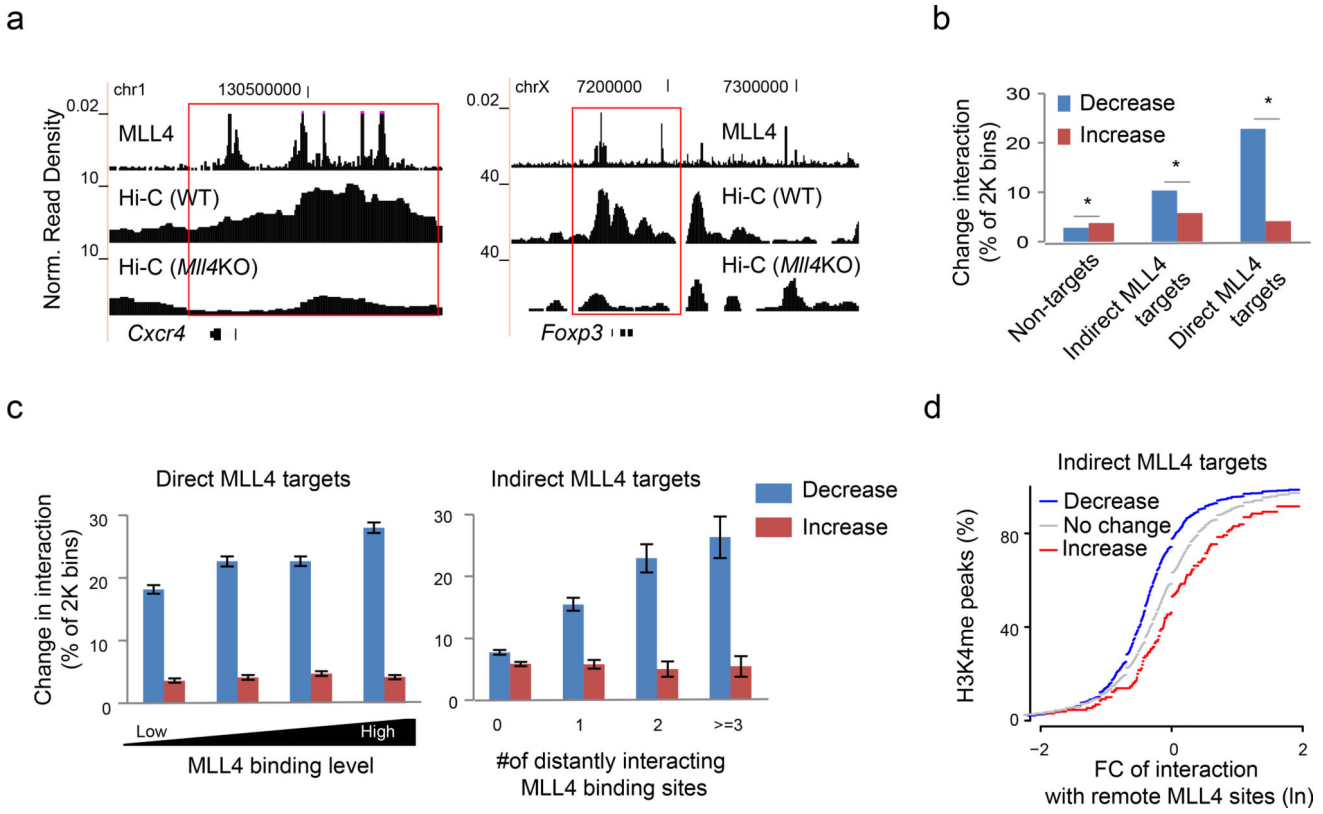

Fig. 5. MLL4 facilitates long-range chromatin interactions

(a) A UCSC Genome Browser image showing the ChIP-Seq signals for MLL4 in WT cells (based on data in Fig 4a) and interaction intensities (pooled n = 3 independent experiments for WT and n = 2 for *Mll4*KO, also applied to other panels) in the genomic regions surrounding *Cxcr4* (left) and *Foxp3* (right) in WT and *Mll4*KO cells. Red rectangles indicate regions showing a decrease in chromatin interaction intensity. WT: $Mll4^{fl/fl}$-*Cd4*-Cre^{-} and *Mll4*KO: $Mll4^{fl/fl}$-*Cd4*-Cre^{+}.

(b) Percentage of decrease (blue) or increase (red) in interaction intensity upon *Mll4*KO for 2-kb genomic bins, sorted based on their relationship to MLL4 binding: directly bound by MLL4 (Direct target), not bound by MLL4 but interacting with distant MLL4-binding sites (Indirect target), or free of MLL4 binding (Non-target). $*P < 0.01$ (χ^2-test).

(c) Percentage of decrease (blue) or increase (red) in interaction intensity upon *Mll4*KO for MLL4-bound non-promoter genomic bins sorted into four equal-sized groups based on the extent of MLL4 binding (left), and for non-promoter genomic bins not bound by MLL4 but interacting with MLL4-bound genomic regions sorted based on the number of interacting regions (right). Error bars: standard deviations.

(d) Accumulative distributions of the fold change of the number of PETs linking an MLL4-unbound H3K4me1 enriched region to distant interacting regions bound by MLL4 upon *Mll4* deletion, sorted based on their responses to *Mll4* deletion (decrease, increase or no change in H3K4me1). *P*-values (0.017 for “increase” vs “No change” and <2.2E-16 for “decrease” vs “No change”) by Kolmogorov-Smirnov test.

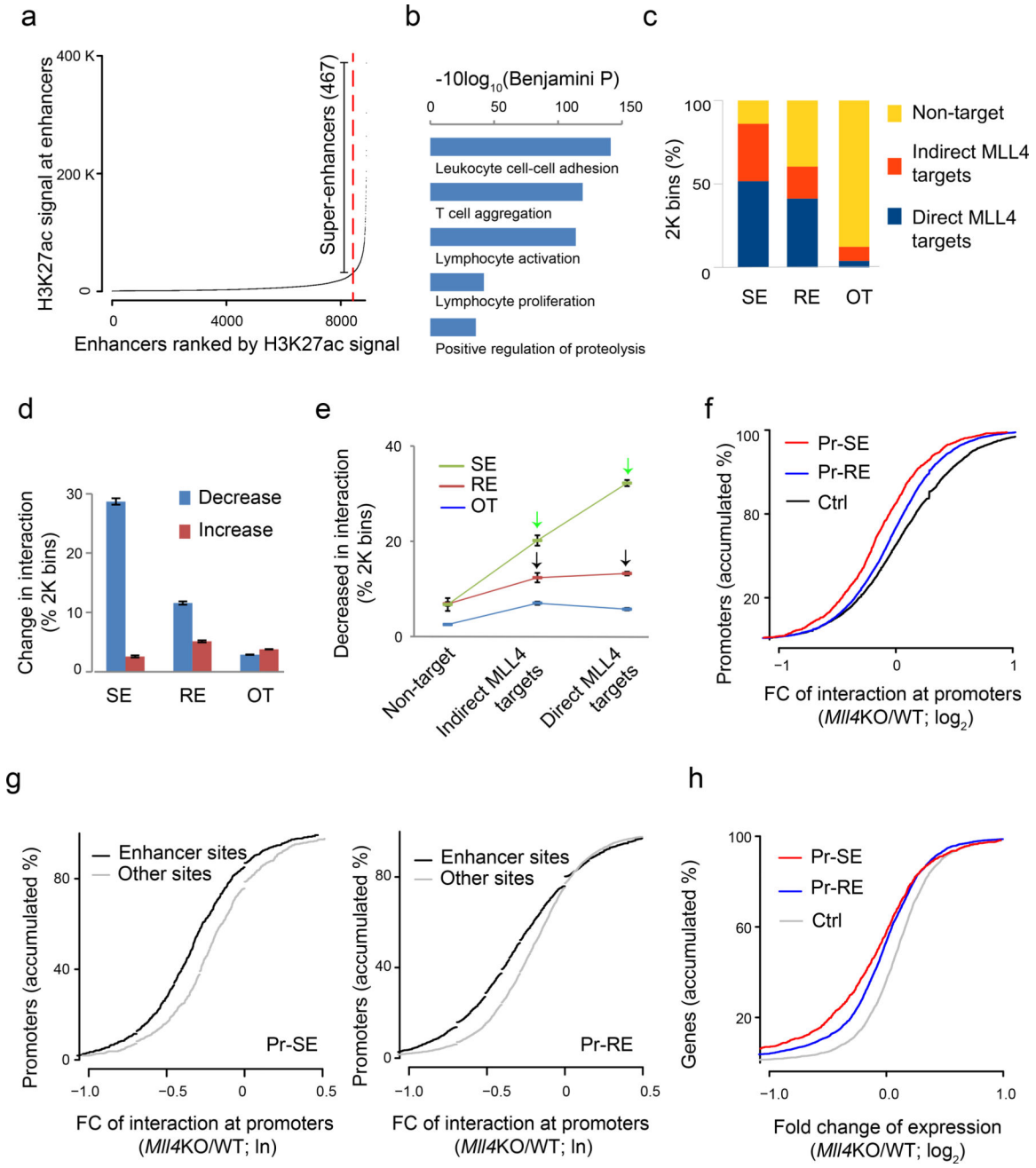

Fig. 6. MLL4 facilitates enhancer-promoter interactions

(a) Distribution of H3K27ac ChIP-Seq signal (total reads; n = 1 experiment) across active enhancers (defined as non-promoter genomic regions enriched with H3K27ac signal) in naïve $CD4^+$ T cells. Red line: cut-off used to define super-enhancer.

(b) Top five GO (gene ontology) terms in biology processes for genes that showing significant chromatin interactions with super-enhancers.

(c) Percentage of 2-kb genomic bins that are directly bound by MLL4, distantly interacted with MLL binding sites or free of MLL4 binding within each category of enhancers: super-enhancers (SE), regular enhancers (RE), or not within any enhancer (OT).

(d) Percentage of decrease (blue) or increase (red) in interaction intensity upon *Mll4* deletion for 2-kb genomic bins located at super enhancers (SE), regular enhancers (RE) or non-enhancers (OT). Error bars: standard deviations.

(e) Percentage of decrease in interaction intensity upon Mll4 deletion for 2-kb bins located within super-enhancers (SE), regular enhancers (RE) or non-enhancers (OT), further sorted based on their relationship to MLL4 binding as illustrated in panel **c**. Error bars: standard deviations.

(f) Accumulative distribution of the fold-change of interaction at promoters that interact with super-enhancers (Pr-SE, red line), regular enhancers (Pr-RE, blue line) or showing no interaction with any enhancer (Ctrl, black line) upon *Mll4* deletion. WT: *Mll4*$^{fl/fl}$-*Cd4*-Cre^{-} and *Mll4*KO: *Mll4*$^{fl/fl}$-*Cd4*-Cre^{+} (throughout the figure); *P*-values (<2.2E-16 for Pr-SE vs Pr-RE and for Pr-RE vs Ctrl) by Kolmogorov-Smirnov test.

(g) Accumulative distributions of the fold-change of interaction with enhancers (black line) or with non-enhancer regions (gray line) for gene promoters targeted by super-enhancer (left panel) or for gene promoters targeted by regular enhancers (right panel). *P*-values (2.8E-13 for left panel and <2.2E-16 for right panel) by Kolmogorov-Smirnov test.

(h) Accumulative distribution of the fold-change of gene expression upon *Mll4* deletion for promoters based on their interaction with super-enhancers (Pr-SE), regular enhancers (Pr-RE) or showing no interaction with any enhancer (Ctrl). *P-values* (2.5E-8 for Pr-SE vs Pr-RE and < 2.2E-16 for Pr-RE vs Ctrl) by Kolmogorov-Smirnov test.

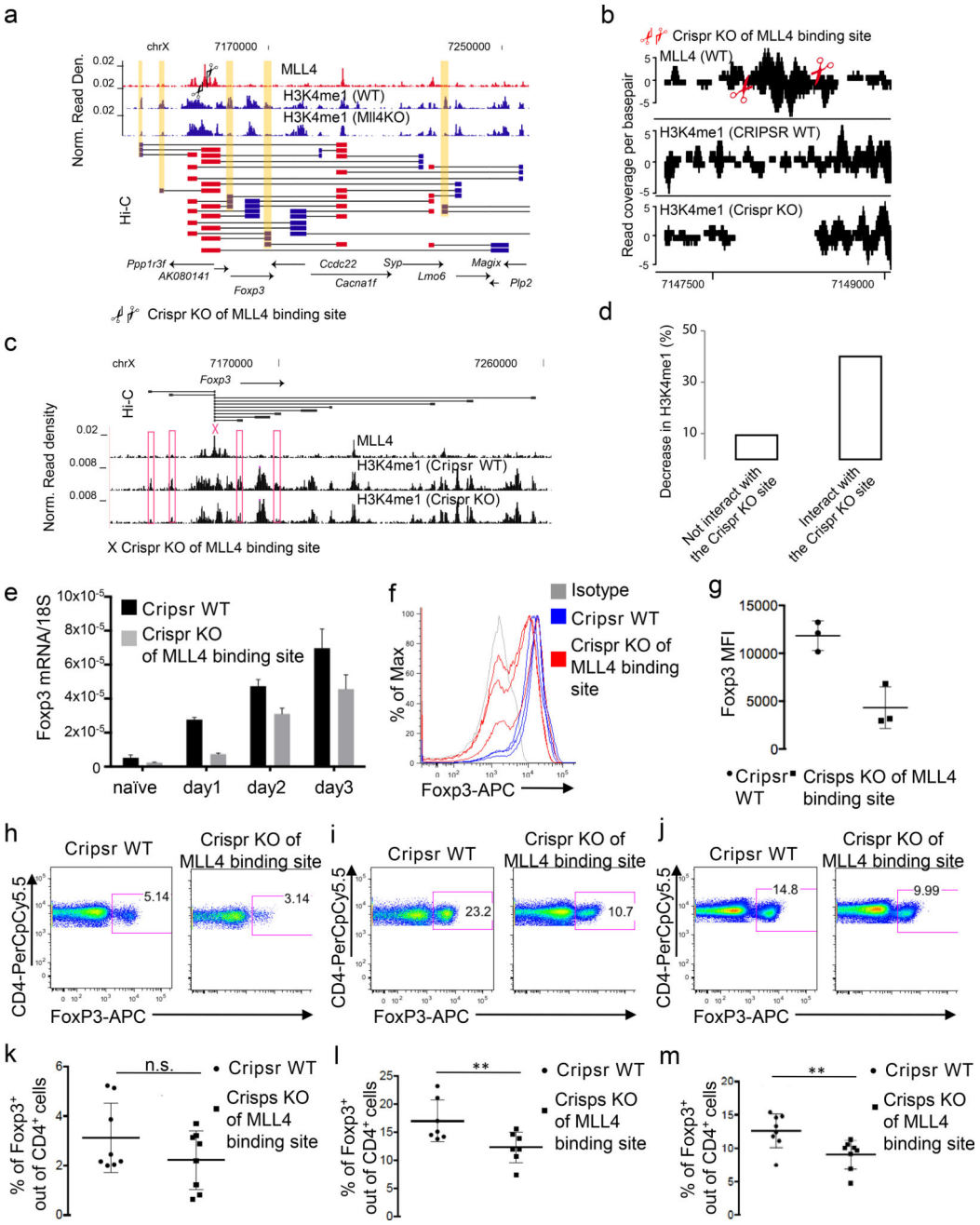

Fig. 7. Deletion of −8.5kb MLL4 binding site decreases H3K4me1 at *Foxp3* regulatory elements and compromises T_{reg} differentiation

(a) UCSC genome browser image showing the distributions of MLL4 ChIP-Seq reads in WT $CD4^+$ T cells, and H3K4me1 ChIP-Seq reads in *Mll4*KO and WT cells. WT: $Mll4^{fl/fl}$-*Cd4*-Cre^- and *Mll4*KO: $Mll4^{fl/fl}$-*Cd4*-Cre^+; Horizontal lines below the ChIP-Seq tracks: H3K4me1-enriched regions linked to remote MLL4 peaks (>5 kb) by at least two PETs; Blue rectangles: H3K4me1 peaks; Red rectangles: MLL4 peaks; Regions highlighted in yellow: H3K4me1 peaks not bound by MLL4 but interacting with remote MLL4 binding sites by at least two PETs and showing decreases in H3K4me1 abundance upon *Mll4* deletion; Scissors: genomic region targeted for deletion by CRISPR. Representative from N

= 2 independent experiments for MLL4 and H3K4me1 ChIP-Seq (also applied to panels **b–d**) and n = 3 for Hi-C.

(b) Distributions of ChIP-Seq reads from Watson strand (shown in positives) or Crick strand (shown in negatives) at a single base-pair resolution for MLL4 in WT cells (*Mll4*$^{fl/fl}$-*Cd4*-Cre^{-}), and H3K4me1 in CRISPR KO (−/ −) and Crispr WT (+/+) control cells around the MLL4 binding site targeted by CRISPR (indicated by scissors).

(c) UCSC genome browser image showing the distribution of H3K4me1 reads around the *Foxp3* locus for the cells with CRISPER deletion of MLL4-binding site as in **b** and for the Crispr WT control cells. Hi-C track: H3K4me1-enriched regions linking to the MLL4 binding site by at least two PETs; Pink rectangles: genomic regions showing significant decreases (FC > 1.5 & FDR < 0.001) in H3K4me1 upon CRISPR deletion.

(d) Percentage of decreased H3K4me1 peaks upon CRISPR deletion for two groups of MLL4-unbound H3K4me1 regions in chromosome X: those linked to the CRISPR deletion site by at least two PETs (10) and those not linked to the site (1,396). P-value = 0.006 by χ2-test.

(e) RT-PCR analysis of *Foxp3* expression in Crispr WT control and Crispr KO naive CD4^{+} T cells at indicated time points of in vitro generated T$_{reg}$ cells. Shown is one representative experiment (n=3 independent experiments). Error bars: standard deviations.

(f) Flow cytometry analysis of Foxp3^{+} expression at day 3 of T$_{reg}$ differentiation as described in panel e. Gated CD4^{+} T cells. Shown is one representative experiment (n=3 independent experiments).

(g) Geometric median fluorescent intensity gMFI of Foxp3 in cells generated as in (f). Error bars: standard deviations. Center line: mean.

(h) Representative flow cytometry plot of CD4^{+}Foxp3^{+} T cells in the thymus of Crispr WT control (n=8) and Crispr KO mice (n=8) with deletion of the MLL4-occupied region upstream of the *Foxp3* gene. Gated CD4SP cells.

(i) Representative flow cytometry plot of CD4^{+}Foxp3^{+} T cells in the spleen of Crispr WT control (n=7) and Crispr KO mice (n=7). Gated CD4^{+} cells.

(j) Representative flow cytometry plot of CD4^{+}Foxp3^{+} T cells in lymph nodes of Crispr WT control (n=8) and Crispr KO mice (n=8). Gated CD4^{+} cells.

(k) Percentages of CD4^{+}Foxp3^{+} in the thymus of Crispr WT control and Crispr KO mice. Gated as in (h). Error bars: standard deviations. Center line: mean.

(l) Percentages of CD4^{+}Foxp3^{+} in the spleen of Crispr WT control and Crispr KO mice. Gated as in (i). Error bars: standard deviations. Center line: mean.

(m) Percentages of CD4^{+}Foxp3^{+} in lymph nodes of Crispr WT control and Crispr KO mice. Gated as in (j). Error bars: standard deviations. Center line: mean.

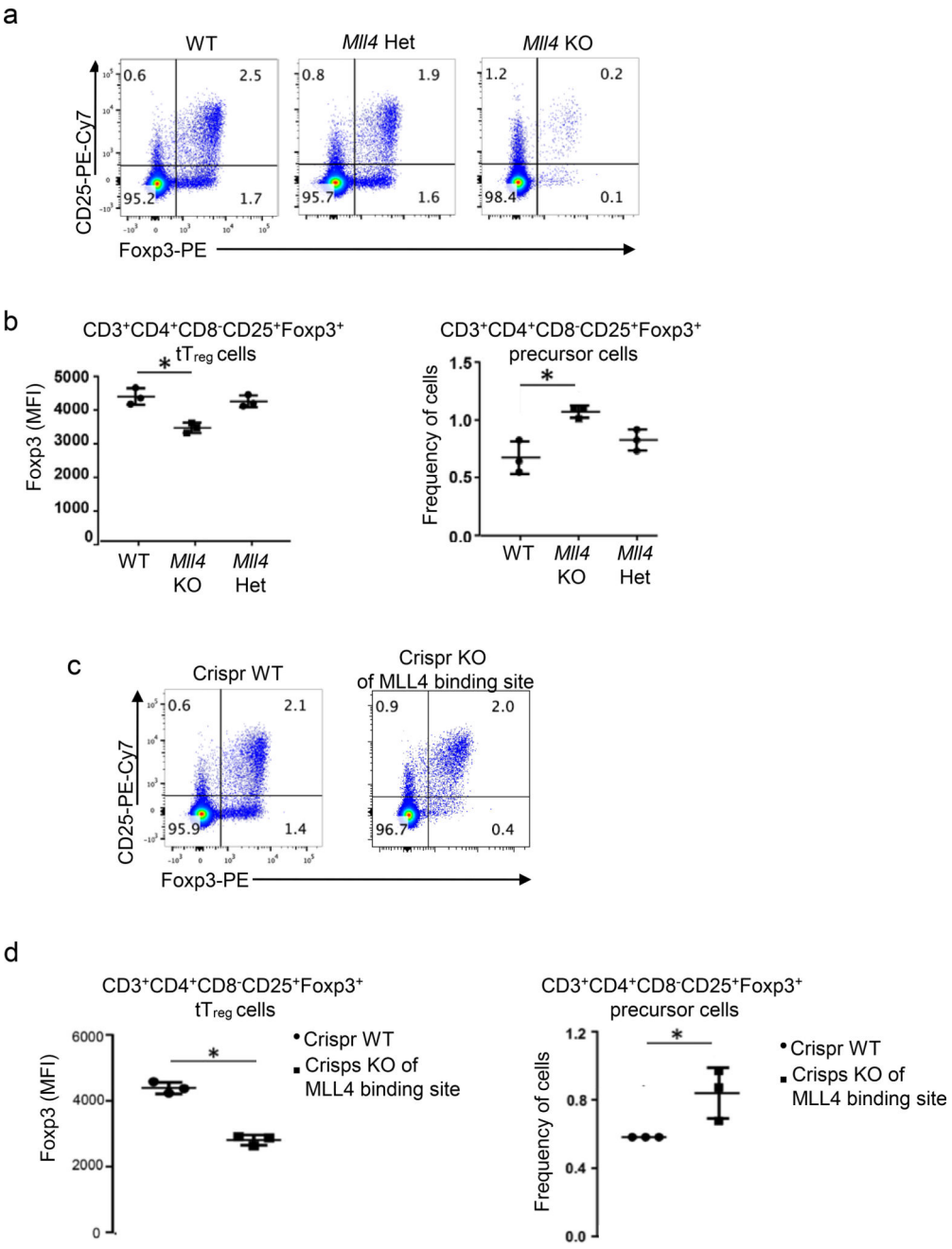

Fig. 8. Deletion of −8,5kb MLL4 binding site increases frequency of T_{reg} precursor cells in the thymus

(a) Flow cytometry analysis of thymi from *Mll4*$^{+/+}$*Cd4*-Cre^{+} (WT), *Mll4*$^{fl/+}$*Cd4*-Cre^{+} (HET) and *Mll4*$^{fl/fl}$*Cd4*-Cre^{+} (*Mll4*KO) mice. Thymi were stained for CD3, CD4, CD8, CD25 and Foxp3. Shown are gated CD3^{+}CD4^{+}CD8^{-} cells. n=3 animals analyzed

(b) MFI of Foxp3 in CD3^{+}CD4^{+}CD8^{-}CD25^{+}Foxp3^{+} tT$_{reg}$s of *Mll4*$^{+/+}$*Cd4*-Cre^{+} (WT), *Mll4*$^{fl/+}$*Cd4*-Cre^{+} (HET) and *Mll4*$^{fl/fl}$*Cd4*-Cre^{+} (*Mll4*KO) mice (left panel). Frequency of CD3^{+}CD4^{+}CD8^{-}CD25^{+}Foxp3^{-} cell population in thymi of *Mll4*$^{+/+}$*Cd4*-Cre^{+} (WT), *Mll4*$^{fl/+}$*Cd4*-Cre^{+} (HET) and *Mll4*$^{fl/fl}$*Cd4*-Cre^{+} (*Mll4*KO) mice (right panel). Error bars: standard deviations.

(c) Flow cytometry analysis of thymi from Crispr WT and Crispr KO mice. Thymi were stained for CD3, CD4, CD8, CD25 and Foxp3. Shown are gated $CD3^+CD4^+CD8^-$ cells. n=3 animals analyzed

(d) MFI of Foxp3 in $CD3^+CD4^+CD8^-CD25^+Foxp3^+$ tT_{reg}s of mice as in (**c**) (left panel). Frequency of $CD3^+CD4^+CD8^-CD25^+Foxp3^-$ cell population in thymi of mice as in (**c**) (right panel). Error bars: standard deviations.