Molecular diversification of regulatory T cells in nonlymphoid tissues

Joanna R. DiSpirito1,2*, David Zemmour1,2*, Deepshika Ramanan1,2, Jun Cho1,2, Rapolas Zilionis4, Allon M. Klein3, Christophe Benoist1,2, Diane Mathis1,2†

Foxp3⁺CD4⁺ regulatory T cells (Tregs) accumulate in certain nonlymphoid tissues, where they control diverse aspects of organ homeostasis. Populations of tissue Tregs, as they have been termed, have transcriptomes distinct from those of their counterparts in lymphoid organs and other nonlymphoid tissues. We examined the diversification of Tregs in visceral adipose tissue, skeletal muscle, and the colon vis-à-vis lymphoid organs from the same individuals. The unique transcriptomes of the various tissue Treg populations resulted from layering of tissue-restricted open chromatin regions over regions already open in the spleen, the latter tagged by super-enhancers and particular histone marks. The binding motifs for a small number of transcription factor (TF) families were repeatedly enriched within the accessible chromatin stretches of Tregs in the three nonlymphoid tissues. However, a bioinformatically and experimentally validated transcriptional network, constructed by integrating chromatin accessibility and single-cell transcriptomic data, predicted reliance on different TF family members in the different tissues. The network analysis also revealed that tissue-restricted and broadly acting TFs were integrated into feed-forward loops to enforce tissue-specific gene expression in nonlymphoid-tissue Tregs. Overall, this study provides a framework for understanding the epigenetic dynamics of T cells operating in nonlymphoid tissues, which should inform strategies for specifically targeting them.

INTRODUCTION

The Foxp3⁺CD4⁺ subset of regulatory T cells (Tregs) is one of the main elements guarding against runaway immune or inflammatory responses (1). These cells’ importance is underlined by the devastating multiorgan, autoinflammatory diseases characteristic of IPEX (immune, dysregulation, polyendocrinopathy, enteropathy, X-linked) patients and scurfy mice, both of which bear Foxp3 mutations and consequently lack functional Tregs. Most Tregs arise as such in the thymus (iTregs), but certain specialized populations derive from conventional CD4⁺ T cells in the periphery (pTregs), a process that can be partially recapitulated in vitro (iTregs).

Our initial view of Treg phenotype and function was largely focused on the control of effector T lymphocytes by Tregs circulating through lymphoid organs. However, substantially more complexity soon became apparent [reviewed in (2)]. Tregs can control other adaptive immune system cells, a variety of innate immune cells, and even nonlymphoid cells. Moreover, Treg populations with distinguishable phenotypes rein in immune responses of diverse types [for example, driven by T helper 1 (TH1), TH2, TH17, or B cells], sometimes co-operating transcription factors (TFs) or signaling pathways characteristic of the target populations. Tregs, with even more distinct phenotypes can be found in nonlymphoid tissues—for example, visceral adipose tissue (VAT), skeletal muscle, skin, or the colonic lamina propria—where they influence the activities of neighboring immune and nonimmune cells.

VAT Foxp3⁺CD4⁺ T cells (3) are considered to be a paradigmatic “tissue Treg” population. These cells arise in the thymus during the first week of life, seed VAT sparsely until 10 to 15 weeks of age, and then gradually dominate the local CD4⁺ T cell compartment (4). VAT Tregs have a transcriptome strikingly different from those of lymphoid organ and other nonlymphoid-tissue Treg populations, largely driven by peroxisome proliferator–activated receptor γ (PPARγ), a nuclear receptor family member thought to be the “master regulator” of adipocyte differentiation (5). They also have a distinct, clonally expanded repertoire of antigen-specific receptors [T cell receptors (TCRs)]. VAT Tregs control local (sterile) inflammation, and thereby local and systemic metabolic indices, via influences on both immunocytes and adipocytes (3, 6). Perhaps not surprisingly, they depend on local growth and survival factors distinct from those used by lymphoid organ Tregs, notably PPARγ ligands (5) and interleukin-33 (4, 6). Other tissue Treg populations appear to be variants on the same theme, with characteristic TF profiles, transcriptomes, TCR specificities, and growth and survival factor dependencies [to the extent that these parameters have been studied [e.g., (7–13)]].

Our current understanding of Treg cell heterogeneity, for nonlymphoid-tissue Tregs in particular, is rather fuzzy. We remain ignorant of how phenotypic diversity at the population level translates to single cells, what epigenetic changes subtend the distinct transcriptomes, and what sets of TFs drive the evolution of transcriptional profiles. These issues were long refractory to analyses because of the specificity of tissue Treg populations coupled with the insensitivity of the required genomic techniques. We have now exploited existing transcriptome data sets and recent improvements in the assay for transposase-accessible chromatin with high-throughput sequencing (ATAC-seq) (14, 15) and single-cell RNA sequencing (scRNA-seq) (16–18) that permit explorations of rare populations ex vivo. Integration of ATAC-seq and scRNA-seq data culminated in the elaboration of a transcriptional network proposed to underlie the unique tissue Treg transcriptomes.
RESULTS
A more nuanced view of nonlymphoid-tissue Treg transcriptomes
As a first step toward elucidating the architecture and regulation of tissue Treg transcriptomes, we exploited existing microarray data sets derived from double-sorted Tregs from VAT (19), injured skeletal muscle (7), and the colonic lamina propria (9), coupled with corresponding splenic Treg populations from the same mice. These particular nonlymphoid tissues encompass a chronic, low-grade sterile inflammatory state; an acute, intense sterile inflammatory state; and a state of chronic, innocuous microbiota stimulation, respectively. The study design focused on nonlymphoid tissue:lymphoid organ fold changes in Treg gene expression within the same individuals, set at false discovery rates (FDRs) of <10%. This approach side-stepped potential issues related to unavoidable variations in physiological state (e.g., mouse age) or in protocol details. A total of 1909 genes were either up- or down-regulated at least twofold in at least one tissue Treg population vis-à-vis its splenic counterpart. The sensitivity of these microarrays was similar to or better than recently published RNA-seq data sets on the same tissues—for example, Delacher et al. (20) reported 3071 differentially expressed genes for VAT—whereas we found 4934 genes using the same filtering criteria (P < 0.05). Most of the transcript variation segregated Tregs by their tissue source, the first three principal components (PCs) accounting for 70% of the variance (Fig. 1A). Muscle Tregs were most closely related to splenic Tregs, a finding that concurs with previous functional data demonstrating that muscle (8), but not VAT (4), Tregs communicate extensively with the circulating Treg pool. Differentially regulated genes could be grouped into several sets according to their patterns of expression in the various nonlymphoid tissues: up- or down-regulated in all three tissues (heretofore referred to as "pan-tissue"), two of them, or just one ("tissue-specific"; Fig. 1B and table S1).

Pathway analysis (Fig. 1, C and D) revealed that the genes up-regulated across all three nonlymphoid tissues were enriched for pathways involved in circadian rhythm, metabolism, and inflammation. Probably not coincidentally, the first two processes are linked in primary metabolic organs such as adipose tissue and the liver (21). Also noteworthy is that the pan-tissue gene set included Prdm1, proposed to be a marker of effector Tregs (22). Not surprisingly, genes specifically up-regulated in VAT Tregs were implicated in lipid metabolism processes, consistent with VAT Treg accumulation of intracellular lipid droplets (5); whereas transcripts overrepresented primarily in muscle Tregs fell within cell cycle–related pathways, reflecting the comparatively high turnover rates of this population (8). The enrichment for genes encoding proteins involved in transforming growth factor–β signaling in the colonic Treg data set is consistent with this growth factor’s participation in several gut processes (23). Thus, transcriptional control of nonlymphoid-tissue Tregs appears to be layered, encompassing pan-tissue and tissue-specific elements.

Tissue-associated variability in Treg chromatin accessibility
It seemed likely that the layered tissue Treg transcriptional profiles described above were established and/or maintained via progressive, regulated accessibility of genomic control elements. Thus, using ATAC-seq (14, 15), we performed genome-wide chromatin accessibility profiling on tissue Treg and contemporaneously harvested splenic Treg populations from the same mice. Again, the data were explored as nonlymphoid tissue:lymphoid organ fold changes within

Fig. 1. Tissue Treg gene sets. Microarray data sets for highly purified tissue Tregs from epididymal VAT (V) (5), injured skeletal muscle (M) (60), and colonic lamina propria (C) (9), as well as from control lymphoid organ Tregs (n = 3 for each tissue). (A) Principal components analysis. PC1, 2, and 3, with their proportions of explained variance. (B) Heatmap of up-regulated (UP) and down-regulated (DOWN) genes more than twofold in tissue versus lymphoid Tregs: 1909 genes ordered by the numbered terms in (C). See also table S1.

the same individuals to avoid issues related to unavoidable variations in mouse physiology or experimental protocol.

First, we examined general features of the chromatin landscapes of nonlymphoid-tissue Tregs (Fig. 2). After cataloging each sample’s peaks of sequencing reads and verifying that there were strong inter-replicate correlations (fig. S1A), we merged the “high-confidence” peaks for Tregs derived from the three nonlymphoid tissues and corresponding spleens (see Materials and Methods for details), resulting in a total of 75,363 open chromatin regions (OCRs). Comparing the normalized signal intensity in these regions across the various Treg populations confirmed that the biological replicates clustered together and that all nonlymphoid-tissue Treg populations separated from their splenic Treg counterparts (Fig. 2A). Again, muscle Tregs were the most closely related to splenic Tregs, and VAT and colon Tregs were more similar to each other than to muscle Tregs. Consistent with studies on other hematopoietic (24, 25) and nonhematopoietic (26) cell types, OCRs were most common in intergenic regions, transcriptional start sites (TSSs), and introns; and these distributions did not vary much across the different Treg populations (fig. S1B). Correlation of OCR accessibility across the populations was by far the strongest at TSS regions, with much weaker associations, hence most of the variation, occurring at intergenic and intronic regions (Fig. 2B).

Fig. 2. Tissue-associated variation in chromatin accessibility is concentrated at TSS-distal regions. (A) Hierarchical clustering of Tregs by their pairwise Pearson correlation values for all 75,363 OCRs. (B) Heatmaps of Pearson pairwise correlation scores for Treg OCRs located within various functional genomic elements. Treg populations were clustered as in (A), with the color bar at the right demarcating biological replicates. (C) Classification of OCRs. Those OCRs with ATAC signal ≥3-fold higher in each nonlymphoid tissue:splenic Treg pairing were clustered by whether their “up” status was shared across all three, two, or only one nonlymphoid tissue. Pan-Treg OCRs were not differentially accessible in any of the pairings. Heatmap shows the ATAC signal [in reads per million (RPM)] in different OCR classes. (D) ATAC-seq reads (left) for example genes whose OCR patterns matched their mRNA expression patterns (right) across the Treg subsets. The Tconv track shows minimal accessibility at Foxp3 in non–Treg CD4+ splenic T cells. ND, not determined.
To classify the regions of differential chromatin accessibility in a more discriminating manner, we first performed pairwise comparisons between the three nonlymphoid tissues: spleen couples, revealing ~10 to 13,000 differentially accessible regions per tissue (fig. S1C). As performed above to define nonlymphoid-tissue Treg gene sets (Fig. 1B), we segregated OCRs into different classes based on whether their tissue:spleen differential appeared in all nonlymphoid tissues (pan-tissue), two of them, or only one (tissue-specific; Fig. 2C). An additional class where there was no nonlymphoid-tissue spleen differential was termed “pan-Treg”; however, as discussed below, it is important to keep in mind that this designation does not mean to imply that the nearby genes are expressed only in Treg. Examination of raw ATAC-seq read distributions verified these OCR class designations and showed that, for some loci, the transcript and accessibility patterns matched. For instance, pan-Treg OCRs constituted the majority of accessible regions around Foxp3, expressed in all Treg, but not in conventional T (Tcon) cells, consistent with its lack of accessibility in Tcon cells (Fig. 2D, top); the preponderance of pan-tissue OCRs corresponded well with the high expression of KlrG1 in nonlymphoid tissue, but not in lymphoid organ, Treg, and VAT-specific OCRs mirrored the VAT-specific expression of Kenad4 (Fig. 2D). These data indicate that tissue-associated variation in chromatin accessibility, located primarily at promoter-distal regions, was a property of nonlymphoid-tissue Treg populations.

OCRs shared between nonlymphoid tissue and lymphoid organ Treg (i.e., pan-Treg OCRs) were by far the most common (Fig. 2C). We expected to find them near genes transcribed in a pan-Treg manner, but many were proximal to tissue-restricted genes (a designation henceforth used to encompass both the pan-tissue and tissue-specific gene sets of nonlymphoid–tissue Treg, i.e., those loci underrepresented in the spleen). Typical examples are Fosl2 and Irf4 (Fig. 3A). This observation suggested that “priming” of the chromatin landscape in splenic Treg might be a common feature of loci whose expression was specifically up-regulated in nonlymphoid tissues. As the Fosl2 and Irf4 tracings show, primed OCRs were both proximal and distal to TSSs (Fig. 3A). To estimate the degree of such priming, we calculated, for each tissue-restricted gene, the fraction of OCRs ±10 kb of its TSS that were accessible in the spleen. For both the pan-tissue and tissue-specific gene sets, the majority of OCRs for the majority of loci were already open in the spleen (Fig. 3B). Similar results were obtained using windows of ±100 or 500 kb around the TSS (fig. S2A), indicating that priming was not exclusively a TSS-proximal regulatory feature.

Because the ATAC-seq analyses were performed on bulk splenic Treg populations, it is possible that what we have termed priming of OCRs in splenic Treg represented ATAC signal from a small population of nonlymphoid-tissue Treg that recirculated back to the spleen. However, at least for VAT, trafficking between tissues and the spleen is rare (4). Moreover, even if recirculation were to occur, we would expect these “contaminating” nonlymphoid-tissue Treg to comprise a minority of cells in our spleen samples and would therefore expect the unanticipated ATAC signal at these OCRs to be lower in splenic Treg than in the tissue Treg populations. However, this was not the case (fig. S2B).

Hence, most of the loci expressed in a tissue-restricted manner, and thus turned off or down in the lymphoid organs, already had a quite accessible, or “primed,” chromatin landscape in the spleen. We wondered whether the accessibility of these elements extended back through the T cell lineage or even beyond it. So we used the ImmGen ATAC-seq resource (www.immgen.org) to analyze the most strongly primed OCRs (the top quartile based on their accessibility in the spleen) in several immune cell types. Most of these OCRs fell into one of three dynamic clusters whose accessibility exhibited variations in at least one hematopoietic lineage, whereas a minority was universally open across all of the T, B, myeloid, and progenitor cell types examined (Fig. 3C). The largest cluster of dynamic OCRs showed the highest accessibility in splenic Treg, decreasing openness within the T lineage, and even less accessibility in more distantly related lineages such as macrophages and hematopoietic stem cells. Thus, the priming of nonlymphoid-tissue Treg gene expression did not merely reflect pan-hematopoietic or T lineage–specifying OCRs but instead showed a high degree of specificity for Treg.

Priming was often not focal; rather, these OCRs occurred in clusters, encompassing both the TSS and TSS-distal regions (e.g., Irf4; Fig. 3A), similar to what has been described for peaks of Mediator occupancy and particular histone modifications at super-enhancers (SEs) in many cell types, including CD4+ T cell subsets (27, 28). Given the association between SEs and the expression of cell fate genes (29), as well as the recent identification of SEs in Treg (30), we asked whether this chromatin element might regulate geographic or functional diversification within the Treg lineage. In prelude to evaluating this hypothesis, we determined how well SEs delineated by our ATAC-seq data correlated with those previously defined using the enhancer mark H3K27ac (30), which was possible because both data sets issued from splenic Treg. Whether SEs were defined on the basis of clustered H3K27ac (fig. S3A) or ATAC (fig. S3B) peaks, there was a strong correlation between ATAC and H3K27ac signals, respectively, within the delineated stretches.

Using the well-established Rank Ordering of Super-Enhancers (ROSE) algorithm (29, 31), we found 600 to 1000 SEs in each of the Treg populations (fig. S3C), yielding a total of 1698 distinct SEs across them all. Comparison of these SEs’ average ATAC signal for nonlymphoid-tissue versus lymphoid-organ Treg revealed very few with differential accessibility (at least twofold tissue:spleen difference; fig. S3D). This strong degree of SE concordance was surprising but suggested that SEs might contribute to the priming of loci encoding tissue-restricted genes in lymphoid organ Treg. Therefore, we calculated the frequency of tissue-restricted genes with an associated SE in splenic Treg (±100 kb of the TSS) and compared this value with the frequencies of spleen-specific or silent genes with an associated SE (Fig. 3D). As expected, compared with silent loci, genes expressed in the spleen were strongly associated with an SE; nonlymphoid tissue–restricted genes were nearly as likely as spleen-specific genes to have an associated SE. These analyses yielded parallel results when SE of ±10 or 500 kb of the TSS was examined and when H3K27ac-defined SEs were examined (fig. S3E).

Given the seemingly transcriptionally permissive state of tissue-restricted genes, we wondered whether additional epigenomic mechanisms layered on top of the open chromatin structure were actively restraining their expression in splenic Treg. In embryonic stem (ES) cells, the promoters of lineage-restricted genes are primed by the accumulation of “bivalent” histone marks that can both promote and inhibit transcription (H3K4me3 and H3K27me3, respectively), which is thought to facilitate the fine-tuning of gene expression during cell differentiation (32, 33). Compared with monovalent (H3K4me3–marked) genes, bivalent loci are expressed at lower levels in ES cells and are up- or down-regulated depending on the pathway of differentiation eventually followed (32). Using publically available chromatin immunoprecipitation (ChIP)–seq data sets on splenic Treg (30), we quantified
bivalent histone marks associated with the various gene sets in splenic Tregs. We found examples of tissue-restricted genes whose promoters were tagged either monovalently with the activating H3K4me3 mark (Ets2; Fig. 3E) or bivalently with H3K4me3 and the repressive H3K27me3 mark (Acvr2a; Fig. 3E). Quantifying levels of these two histone modifications within 1 kb of the TSSs revealed that most loci had nominal levels of the two tags (Fig. 3F, black dots and density lines). For those loci with appreciable occupancy by the modified histones, their TSSs were split between those bivalently and monovalently (H3K4me3 only) marked (Fig. 3F, purple and red, respectively). There was a separation between the typical expression level of bivalently versus monovalently marked tissue-restricted genes...
in splenic Tregs, the former being significantly reduced (Fig. 3G).
Moreover, transcript levels for most of the bivalently marked genes were below, or only slightly above, the limit of detection, suggesting that these genes included those that would be turned on in response to tissue-derived signals, and were perhaps kept nominally transcribed in splenic Tregs by the presence of H3K27me3 and the transcriptional silencing machinery that it recruits.

**Predominant enrichment of a small, shared set of TF family motifs in tissue-restricted OCRs**
Having cataloged OCRs near loci coregulated in Tregs, we interrogated their DNA sequences for TF family motifs. A search of the pan-tissue OCRs within 100 kb of TSSs of the ~100 genes up-regulated in all three nonlymphoid tissues revealed repeated enrichment of the motifs for a small number of TF families, including those for RFX, bZIP, nuclear receptors, and ETS (Fig. 4A). bZIP motifs were present along the highest proportion of loci (34%) and in the most instances per gene (mean, 3.5; Fig. 4B), suggesting that one or more members of this family might be major drivers of the shared tissue Treg gene expression signature. The short length (seven nucleotides) of the bZIP motif is unlikely to account for this pattern, because equally short or even shorter sequences (such as those bound by ETS, RUNX, or GATA family members) were not enriched to the same extent. Overlaying the genomic locations of the motifs most highly represented among the pan-tissue gene set (i.e., the bZIP, RUNT, and GATA motifs) revealed, in all three cases, broad peaks at ~10 kb from the TSS, hinting that they might work together! in cis (Fig. 4C). Scanning OCRs linked to the pan-tissue gene set for motifs representing each TF family and clustering genes by their TF family profiles revealed that most loci were linked to more than one TF family (Fig. 4D). The motifs of certain families showed preferential coenrichments, for example, RFX commonly with RUNT and bZIP, and NR with ETS (Fig. 4D).

A similar analysis using the three sets of tissue-specific Treg OCRs and genes implicated a few additional TF families not enriched in pan-tissue OCRs, most prominently CTCF, whose binding site was among the three most highly enriched motifs for each of the tissue Treg populations (Fig. 4E). The overrepresented CTCF family motif is bound by only two members: CTCF itself and BORIS, a related protein expressed almost exclusively in germ cells (34). Therefore, it is likely that CTCF plays a common role in establishing and/or maintaining tissue-specific OCR landscapes in nonlymphoid-tissue Tregs. To assess the association of highly ranked TFs with tissue-specific loci, we again used the chromatin accessibility data to construct predicted TF family:target gene modules for each tissue (Fig. 4F). Given the far greater number of tissue-specific genes for VAT and colon Tregs, modules for these tissues were much more robust than those for muscle. In contrast to pan-tissue loci, distinct clusters of VAT- or colon-specific genes were associated with a single TF family, most prominently ETS, bZIP, or CTCF.

**Representation of tissue Treg gene sets in single cells**
For many of the TF families highlighted above, there is a great degree of overlap between the DNA motifs recognized by individual family members [e.g., ETS1 and ETS2 (35)], preventing a rigorous delineation of tissue-specific roles for individual TFs based solely on enrichment of their predicted binding sites within ATAC-seq peaks. Unfortunately, the number of cells available per mouse (usually a few thousand) precludes obtaining robust TF ChIP-seq data for most nonlymphoid tissues. Rather, we reasoned that exploiting scRNA-seq data, correlating expression of individual TF family members with their predicted target genes in thousands of individual cells, could, first, verify (or not) the regulatory modules predicted by ATAC-seq; second, refine these modules into networks of individual TF family members and their predicted target genes; and third, determine whether, and to what extent, different members of the same TF family were deployed to drive Treg specialization within tissues.

Therefore, we sorted Tregs from VAT, skeletal muscle (4 days after injury), and the colonic lamina propria plus, as controls, splenic Treg and Tconv cells from the same mice, with independent duplicates for each population. Single cells were encapsulated in microfluidic droplets following the inDrops protocol (17). In total, we analyzed 7455 individual Tregs. On average, 1885 unique mRNA molecules transcribed from 759 different genes were sequenced per cell, effectively covering the Treg transcriptome space of 17,841 genes (Fig. S4A). We first displayed the data as a t-SNE (t-distributed stochastic neighbor embedding) plot, which permits dimensional reduction (Fig. 5A). No major batch effects were observed as the data from the various spleen controls substantially overlapped, as did the duplicate data from the various tissues (fig. S4B). The cells from each of the three nonlymphoid tissues clustered together on the t-SNE plot and separately from splenic Tregs and Tconv cells. There was very little overlap between the tissues, and no cell clusters that included participants from all three tissues were observed. The spleen hosted a small number of cells that appeared to “reach into” the muscle or colon, but not VAT, Treg cell space, and vice versa. This finding matches previous observations that muscle (8), but very rarely VAT (4), Tregs exchange with the circulating pool. Foxp3 was expressed throughout the Treg populations but not in Tconv cells, as expected (Fig. 5B). Various TFs have been reported to drive transcriptomic diversification of particular Treg populations. Accordingly, we found Pparg to be expressed predominantly in VAT Tregs (5); Tbx21 in a subset of muscle and colonic Tregs (36, 37); Ikzf2 and Rorc in distinct subpopulations of colonic Tregs (as well as in Tregs in other tissues) (9, 11); and Cxcr5, a marker of follicular regulatory cells (38, 39), predominantly in splenic and colonic Tregs.

Deconvolution of the gene expression clusters delineated in Fig. 1B at the single-cell level maintained the distinctions originally observed at the population level, although less sharply, likely due to the reduced sensitivity of scRNA-seq (fig. S4C). The expression of each signature was mostly bimodal across the different populations (Fig. 5C). For example, the pan-Treg signature was up-regulated in all Tregs in comparison with splenic Tconv cells. This signature was more highly represented in nonlymphoid-tissue Tregs than splenic Tregs, consistent with the fact that the pan-Treg gene set includes loci related to Treg activation (40) and that nonlymphoid-tissue Tregs tend to have an activated phenotype compared with lymphoid organ Tregs (3, 7, 9). Similarly, the pan-tissue signature was only weakly expressed in the spleen, with very few cells achieving the highly elevated expression levels typical of tissue Tregs; those rare cells probably correspond to the few splenic Tregs invading the tissue Treg space in Fig. 5A. VAT Treg, were the most distinct population, with higher expression of the pan-tissue signature and a highly differential VAT-specific signature. Muscle Tregs, on the other hand, expressed the pan-tissue gene set at the lowest level and had only the weakest tissue-specific signal.

Surmising that the fainter distinction of the muscle Treg population might reflect a more heterogeneous conglomeration of evolving subpopulations, as might be expected with an acute injury context, we
Fig. 4. A limited number of TF family motifs are associated with tissue-restricted gene signatures in Tregs. (A) TF family motifs enriched in OCRs within 100 kb of the TSS of genes up-regulated in Tregs from all three nonlymphoid tissues. Enrichment represents the fold change in frequency of the motif in these OCRs versus control regions, and only significantly enriched motifs ($P < 0.01$) are shown. (B) The percentage of pan-tissue genes containing $\geq 1$ instance of each motif within 100 kb of its TSS ($x$ axis) versus the average number of motif occurrences per gene ($y$ axis). (C) The positions of bZIP, RUNT, and GATA motifs are shown for the set of pan-tissue OCRs found within 100 kb of pan-tissue genes. (D) Clustering of pan-tissue genes based on their ATAC-predicted regulatory TF families. TF motifs were linked to each gene based on their presence in an OCR assigned to that gene. Each gene then received a score of 0 or 1 for each of the seven TF families shown, and genes were then clustered according to their pairwise correlation scores. (E) TF family motifs significantly enriched in OCRs within 100 kb of the TSSs of each set of tissue-specific up-regulated genes. Enrichment values within each set are ranked to allow comparison across subsets. STAT, signal transducers and activators of transcription; NS, not significant. (F) As in (D), except tissue-specific genes were clustered.
obtained additional single-cell data using the CEL-Seq platform (41), which tends to yield more transcripts per cell but on fewer cells. The T_reg response to injury was dynamic. At day 1, muscle T_regs formed two clusters (Fig. 5D), both significantly enriched (vis-à-vis the spleen) in expression of the pan-tissue and muscle Treg gene sets (Fig. 5E). The cells of cluster 2 expressed more genes related to TCR stimulation (e.g., Nr4a1, Egr1, and Egr2) and the interferon-γ response (Ifngr1) than did the cluster 1 constituents (Fig. 5F). At day 4, muscle T_regs were more homogeneous, almost exclusively composed of cluster 1 cells.
Previous studies have documented substantial phenotypic and functional diversity within the colonic Treg population, most notably between the tTreg and pTreg subpopulations (9–11). Because a recent scRNA-seq study (42) failed to confirm this heterogeneity but, as described above, our all-tissue t-SNE plots (Fig. 5, A and B) uncovered what appeared to be distinct colonic Treg subsets, we carefully reexamined this issue. Unsupervised clustering of only the colonic and control splenic Treg single-cell data delineated four robust cell clusters (Fig. 5G). The first level of heterogeneity distinguished pTreg, expressing Rorc transcripts (green cluster) from tTreg transcribingIkzf2 (blue, yellow, and turquoise clusters; Fig. 5H, left). Among the tTreg one subset was probably made up of cells circulating through or residing in mucosal lymphoid structures as they preferentially expressed Ccr7, Cxcr5, and Bcl6 transcripts (Fig. 5H, middle). The other two tTreg subsets were distinguished by differential expression of a transcriptional program promoting tissue repair (Il1rl1, Areg, Penk, etc.; Fig. 5H, right).

**A tissue Treg transcriptional network: How to achieve specificity with common TF binding motifs**

Having validated the scRNA-seq data, we integrated it with the ATAC-seq results to identify individual TF family members most likely to drive tissue-restricted gene expression in nonlymphoid–tissue Treg, ultimately resulting in a TF-target gene regulatory network. Some previous studies have attempted to integrate profiles of chromatin accessibility and gene expression to yield insights into mechanisms of transcriptional regulation (43). However, these efforts relied on population-level RNA-seq data, which would have very limited value in uncovering meaningful gene-gene correlations in our case as it would be based on only four data sets (spleen, VAT, muscle, and colon). scRNA-seq should greatly increase the power of the analysis as it would entail evaluation of gene coexpressions across thousands of individual cells (44, 45). First, we identified particular TF family members differentially expressed in tissue Treg compared with their splenic counterparts (FDRs < 2.5%). Then, to determine whether the putative controlling factors could be functional, we calculated the covariation in expression between a TF and its putative targets. Thus, our predicted regulatory modules (Fig. 4, D and F) were refined by including a score that aggregated a TF’s motif in an OCR, its putative target genes, and to what extent expression of the TF and its target genes covaried in individual cells [Fig. 6A (for schema) and fig. S5, A to C].

The overall TF network arising from these analyses is represented in Fig. 6B, whereas Fig. 6C illustrates the relative target gene frequencies across and within the various gene sets. Robustness and significance of the network were assessed by randomization of the single-cell data (shuffling expression levels per gene between cells). One thousand random permutations of the data failed to engender comparable network connectivity (for example, fig. S5D for the VAT-specific modules). We identified 730 connections between TFs and their target genes, well above the 18 expected links by chance alone with an FDR of 2.5% (fig. S5D). Immediately evident was that the different TF families seemed to have variable importance in regulating expression of the different gene sets. Pan-Tissue Treg genes were connected to loci encoding the AP-1 subfamily of bZIP TFs (Bach2, Crem, Fosb, Jdp2, Junb, and Jun) andNr1d1 and Nr4a1, which encode nuclear receptors; and to Gata3. VAT- and colon-specific Treg loci had the strongest links to loci encoding bZIP, ETS, and nuclear receptor family members, whereas muscle-specific genes were connected above all to Nfkb2 and Rel, specifying members of the Rel homology domain family. At a first level, tissue-restricted transcription was mediated by tissue-specific expression of TFs, as represented in the peripheral nodes of the network (Fig. 6B). For each TF family highlighted in the ATAC-seq analyses, different members were expressed in the different tissues, which correlated with tissue-specific expression of their gene targets (Fig. 6C). For example, within the nuclear receptor family, Pparg [as expected (5)], Rara, and Ror were specifically associated with VAT-specific genes, whereas Rorc [as expected (9, 11)], Vdr, and several others were linked specifically to colon-specific genes. Within the ETS family, Fli1 and Etv3 partitioned to VAT-specific and Ets1, Ets2, and Elf1 to colon-specific loci. The bZIP and RFX families showed similar examples of partitioning.

At a second level, and in an apparent paradox, tissue-specific transcription was also regulated by broadly expressed TFs. In a dense subnetwork at the center of the tissue Treg network, genes encoding several members of the AP-1 subfamily (Bach2, Crem, Fosb, Jdp2, Junb, and Jun) were connected to numerous pan-tissue and tissue-specific loci (Fig. 6, C and D, left). BACH2 was special in the sense that it was the only TF whose gene expression correlated negatively with its gene targets (Fig. 6D, right). To understand how tissue-specific transcription can result from regulation by broadly expressed TFs, we delved further into tissue Treg network architecture, focusing on three-node motif representation (nonrandom subnetworks of three vertices) in the VAT subnetwork. Single-input motifs (SIMs) and multiple-input motifs (MIMs) were very frequent (Fig. 7A), as previously reported for biological networks (46). SIMs and MIMs highlight the fact that most TFs regulated many different gene targets and, reciprocally, most genes were regulated by multiple factors, allowing coordinate gene regulation. However, more importantly, we also found frequent feed-forward loops (Fig. 7B). In the first layer of these loops, loci encoding ubiquitous TFs (Jdp2, Junb, and Jun) were connected not only to tissue-specific gene targets but also to loci encoding tissue-specific TFs (Fli1, Fosb, Maf, etc.) that regulated the same set of gene targets. For example, the loci encoding certain broadly expressed TFs (Jdp2, Junb, and Jun) connected to Fli1 to regulate VAT–specific targets such as Il10ra and Cish (and others) (represented in red). On the other hand, across all tissues, transcription of the pan-Tissue gene set connected, unsurprisingly, almost exclusively to the ubiquitously expressed AP-1 subfamily of TFs (GATA3 being the exception) and, overall, was more multiconnected than was the tissue-specific Treg genes (Fig. 7C).

**Validation of TFs underpinning tissue-restricted and tissue-specific network nodes**

As detailed above, the transcriptional network we constructed was bioinformatically robust; however, experimental validation was also imperative. To that end, we focused on two predicted network nodes, one entailing the sets of genes preferentially expressed in all of the nonlymphoid tissues and a second anticipated to have a tissue-specific impact. BACH2 was an interesting candidate as a broad regulator of transcription in nonlymphoid tissues because it is a repressor known to be present in lymphoid organ Treg (47). In our network, Bach2 expression was highly anticorrelated with transcription of tissue-restricted Treg genes, whether pan-tissue or tissue-specific (Fig. 6D). In addition, it was the only bZIP family member whose expression decreased rather than increased (or stayed constant) in tissue Treg vis-à-vis their splenic counterparts (Fig. 8A). Thus, we predicted that BACH2 would bind to (and repress) tissue-restricted genes in lymphoid-organ...
Fig. 6. Architecture of the tissue Treg transcriptional network. (A) A schematic of network construction integrating ATAC-seq and scRNA-seq data. First, TF families are linked to target genes when their binding site is enriched in the associated OCRs (see ATAC-seq correlation heatmaps in Fig. 4, D and F). The network is then refined (i) by identifying TF members within each family that are expressed in each tissue (using scRNA-seq) and (ii) by keeping connections with significant correlations between the TFs and target genes in the scRNA-seq data. (B) Directed graph of the tissue Treg transcriptional network. Each node represents a gene. A directed edge connects a TF and its target gene if the target has the TF family’s motif in an associated OCR (ATAC-seq) and if their expression is correlated in the scRNA-seq data. Node and edge colors reflect the gene set to which the target gene belongs. Gene names are those of key TFs (black) and target genes (gray) in the network. (C) TF:target gene correlations. Left: The percentage of target genes that are linked to each TF (rows) falling in the four gene sets (columns: pan-tissue, VAT, muscle, and colon). Row values sum to one. Right: The percentage of genes in each set (columns) linked to each TF (rows). Columns sum to one. (D) Network blow-ups. Left: Central hub in the tissue Treg network. Subgraph of (B) with the ubiquitous TFs and their connection to pan-tissue and tissue-specific target genes. Edge color as in (B). Right: Subgraph of (D) with Bach2 and its connections to pan-tissue and tissue-specific target genes. Node colors relate to the gene sets and edge colors to positive (red) or negative (blue) correlations with Bach2. See also fig. S5.
its occupancy on control loci (Fig. 8B). This increased BACH2 signal did not merely reflect strong binding at a few genes because it bound to the promoters of 30 to 50% of the tissue-restricted Treg gene sets in the in vitro–generated Treg (Fig. 8C). Thus, the inverse correlation between levels of BACH2 binding and expression of tissue-restricted Treg genes was likely to be a direct effect of its repressive activity at these loci in lymphoid organ Tregs, consistent with a previous report that it can repress the expression of Th cell signature genes in Treg (47).

As a predicted regulator of tissue-specific gene expression in Treg, we revisited the role of PPARγ, a nuclear receptor family member required for the accumulation and function of bona fide VAT Tregs (5). Our network predicted functionality of PPARγ exclusively in VAT Treg, although its expression was also strongly up-regulated in skeletal muscle vis-à-vis lymphoid organ Tregs (Fig. 8D). First, we examined mice that carry a Treg-specific knockout of Pparg and, consequently, have a dearth of VAT Tregs (5). Examination of the Treg populations in muscle and spleen 4 days after injuring muscle of mutant mice and their wild-type littermates with cardiotoxin revealed no significant effect on the accumulation of muscle Treg in the mutants (Fig. 8E). Second, we treated standard B6 mice with the PPARγ agonist, pioglitazone coincident with cardiotoxin-induced skeletal muscle injury. As expected (5), pioglitazone robustly induced the Treg population in VAT but did not change Treg frequencies in the muscle, colon, or spleen (Fig. 8F).

**DISCUSSION**

Integrating accessible chromatin and single-cell transcriptome profiling, we explored the regulatory underpinnings of Treg diversity in nonlymphoid tissues. At both the chromatin and transcription levels, adaptation of Treg to VAT, muscle, or the colon was achieved by a combination of tissue-shared and tissue-specific modulations. However, the picture that emerged was not simply dichotomous. For the tissue-restricted gene set (both pan-tissue and tissue-specific), chromatin accessibility and gene expression patterns often did not match, with many OCRs already accessible in the spleen, suggesting some priming event before transiting of Treg from lymphoid to nonlymphoid organs. A disconnect between chromatin accessibility and transcript abundance has many parallels in other settings, including tissue adaptation. For example, macrophages from multiple nonlymphoid tissues share a set of common, epigenomically permissive enhancers that are selectively active in distinct tissues, correlating with tissue-specific gene expression (48, 49). In nonlymphoid-tissue Treg, this feature may permit or even promote rapid changes in gene expression, given that the dynamics of their differentiation within tissues can be quite rapid: A day after injury, muscle Treg were distinct from splenic Treg by scRNA-seq, and at day 4 the muscle Treg subpopulation structure had already been remodeled.

What tissue-adaptive properties are conferred by priming loci as opposed to more immediate epigenomic and transcriptional changes in response to signals received locally? We speculate that, from the perspective of modulating Treg cell function, tissue-derived signals are likely to be nonspecific—that is, they evolved to facilitate proper functioning of a heterogeneous mix of cell types, not Treg in particular. Therefore, making tissue adaptation reliant on local signals alone may activate (or repress) gene regulatory elements whose function is either not compatible with Treg-specific signals (such as those derived from the TCR or cytokine receptors) or is destabilizing to the suppressive phenotype. In contrast, by relying on a priming event...
within lymphoid organs, the effects of local stimuli on Tregs, including modulation of TF and coregulator expression or activity, could be constrained and directed to nondeleterious genomic sites.

Harnessing the natural variation in gene expression among thousands of single cells allowed us to build pan-tissue and tissue-specific regulatory networks for nonlymphoid-tissue Tregs. Our data suggest that the expression (and likely activity) of individual members of a restricted number of TF families was modulated to facilitate context-specific transcription. For instance, within the ETS family, the loci encoding ETS1, ETS2, and ELF1 were linked to colon-specific gene expression, whereas the Fli1 and Etv3 loci were linked to VAT-specific transcription.

The tissue Treg network we constructed was bioinformatically robust. Two key nodes, one pan-tissue and the other tissue-specific, were experimentally validated. In addition, several published findings provide anecdotal support (9, 11, 50, 51). Briefly, our findings argue for the following model of tissue Treg diversification: Transcriptional programs must be established in nonlymphoid-tissue Tregs within the constraints of an epigenome that maintains Treg identity while allowing tissue adaptation for optimum installation and function. To that end, the chromatin structure of Tregs in lymphoid organs is primed, through the establishment of pan-Treg OCRs, to license the rapid transcriptional turn-on or up-regulation of most tissue-specific genes. The primed stretches may facilitate reprogramming of additional regulatory elements, for example, pan-tissue or tissue-specific OCRs, by cues received once the cell arrives at a particular nonlymphoid site. TFs belonging predominantly to a limited set of families act on these tissue-restricted regulatory organizational modules. The first specifies a common “tissue” phenotype (including circadian and metabolic genes, among others) and is largely under the control of bZIP family TFs, particularly members of the AP-1 subfamily. These primary TFs are expressed in all tissue-specific modules via induction of distinct TFs (such as members of the ETS, nuclear receptor, or RUNX families) in each tissue. Primary and secondary TFs are integrated into feed-forward loops because both can activate tissue-specific target genes.
MATERIALS AND METHODS

Study design
We double-sorted 10 to 20,000 $T_{reg}$ from VAT, skeletal muscle, and colonic lamina propria (pooled from several mice) and corresponding lymphoid organ $T_{reg}$. Microarray studies were done as biological triplicates, ATAC-seq as duplicates, and scRNA-seq (inDrops) as duplicates ($n = 7955$ cells).

Mice
Male C57BL/6J mice, 25 weeks old (VAT) or 8 to 10 weeks old (muscle and colon), were purchased from the Jackson Laboratory. B6. Foxp3Ires-Gfp (52), B6. Foxp3Ires-Gfp-Cre (53), and B6. Ppargfl/fl (54) mice were bred in our specific pathogen-free facility at Harvard Medical School’s Center for Animal Resources and Comparative Medicine (Institutional Animal Care and Use Committee protocol ISO0001257). Mutant Pparg $T_{reg}$ mice ($T_{reg}$-Ppargfl/fl) were generated as in (5), with male littermate-matched experimental and control mice used for all experiments.

Treatments
Anesthetized mice were injected with 0.03 ml per muscle of Naja mossambica mossambica cardotoxin (0.03 mg/ml; Sigma) in all hind limb muscles. Male B6 mice received intraperitoneal injections of pioglitazone (10 mg/kg; Cayman Chemical) or dimethyl sulfoxide once per day for 10 consecutive days.

Isolation of $T_{reg}$ populations
Single-cell suspensions were obtained from VAT, skeletal muscle, and colonic lamina propria (see Supplementary Materials and Methods). During each harvest, spleens were isolated from the same mouse and processed independently. Single-cell suspensions were stained in flow cytometry buffer (2% fetal bovine serum and 1 mM EDTA in Dulbecco’s modified Eagle’s medium) at 4°C for 30 min as detailed in Supplementary Materials and Methods.

Microarray analyses
The microarray data have been published (5, 7, 9) and are accessible on Gene Expression Omnibus (GEO; see Data and materials availability). Genes differentially expressed between each nonlymphoid-tissue $T_{reg}$ population and its splenic counterpart were defined as having an expression value of $\geq 100$ in at least one tissue, a fold change of $>2$ or $<0.5$, and $P < 0.05$. A total of 1909 genes were differentially expressed in at least one tissue (1883 with FDRs $< 10\%$). Additional details are in Supplementary Materials and Methods.

ATAC-seq
ATAC-seq reads from two biological replicates per sample were filtered for quality [sickle, default settings for paired-end reads; see (55)], were adapter-trimmed (cutadapt, --e 0.1, --m 20]) and mapped to the mouse genome (mm9) using Bowtie v2.2.4 (56), keeping only mate pairs that mapped to a single best location using SAMtools v1.3 (57), and were nonduplicates using Picard v1.130 (http://broadinstitute.github.io/picard). OCRs were identified by HOMER v4.6 using findPeaks-style histone (58), with an FDR of 0.001. Only OCRs occurring in both replicates were further analyzed in analyses. Principal components analysis, cumulative frequency distributions, generation of OCR class, and SE heatmaps, as well as all statistical tests, were performed in R (v3.2.1). Additional details are in Supplementary Materials and Methods.

scRNA-seq
scRNA-seq was performed following the inDrops protocol (17, 18). PolyA+ mRNAs from thousands of single cells are sequenced rapidly (<30 min) and efficiently (>70 to 80% of input cells). Besides unique single-cell barcodes, unique molecular identifiers were added to minimize noise introduced during amplification. Only the 3’ ends are sequenced, keeping strand information. The sequencing data were processed as described (18). Reads were mapped against the mouse mm10 transcriptome (complemented with sequences of the Foxp3Ires-Gfp or Foxp3Ires-Gfp transgenes) using TopHat2 (59) and filtered to retain reads with unique molecular barcodes and mapping to unique genomic regions. Further details are in Supplementary Materials and Methods.

ChIP-seq
FASTQ files representing splenic $T_{reg}$ H3K27ac ChIP-seq were reanalyzed from the DNA Data Bank of Japan (DRA003955). Reads were processed and aligned to mm9 as for ATAC-seq, except software options that were changed from paired-end to single read. Peaks were called by HOMER v4.6 using findPeaks-style histone (58), with an FDR of 0.001. H3K27ac peaks were used to identify SEs using the ROSE algorithm (29). FASTQ files representing spleen-derived iT$_{reg}$ cell BACH2 ChIP-seq were from National Center for Biotechnology Information GEO (GSE45975) and processed for the H3K27ac data. Peaks were called by HOMER v4.6 using findPeaks-style factor, with an FDR of 0.001. Further details are in Supplementary Materials and Methods.

Statistical analyses
Significance was assessed by Student’s t test, the binomial test, Wilcoxon rank sum test, Kolmogorov-Smirnov test, or null distributions, as specified in individual figure legends and Materials and Methods sections.

SUPPLEMENTARY MATERIALS
immunology.scienmag.org/cgi/content/full/3/27/eaat5861/DC1
Materials and Methods
Fig. S1. ATAC-seq general features.
Fig. S2. Priming of OCRs at tissue-restricted genes in splenic $T_{reg}$ also occurs at TSS-distal regions and is not composed of weak OCRs.
Fig. S3. Generation and analysis of ATAC-seq-derived SEs across $T_{reg}$ subsets.
Fig. S4. Tissue $T_{reg}$ scRNA-seq data set validation.
Fig. S5. Pan-tissue and tissue-specific $T_{reg}$ modules identified by combining ATAC-seq and scRNA-seq data.
Table S1. Tissue $T_{reg}$ gene sets.
Table S2. Edge list of the tissue $T_{reg}$ transcriptional network.
References (62–72)

REFERENCES AND NOTES


T. D. Immunology 27, 786–800 (2007).


Molecular diversification of regulatory T cells in nonlymphoid tissues

Joanna R. DiSpirito, David Zemmour, Deepshika Ramanan, Jun Cho, Rapolas Zilionis, Allon M. Klein, Christophe Benoist and Diane Mathis

DOI: 10.1126/sciimmunol.aat5861

Mapping Treg regulomes

Technological advances are allowing immunologists to study rare populations of immune cells that take residence in various tissues including adipose tissue, skin, and the lung. Here, DiSpirito et al. have generated transcriptomes and chromatin accessibility maps of mouse regulatory T cells (Treg) that reside in visceral adipose tissue, muscle, and the colon and compared them with the profiles generated from splenic Treg. They have used these data sets to define transcriptional networks that are shared by all these populations and to identify networks that are unique to one or more tissue-resident Treg populations.

ARTICLE TOOLS
http://immunology.sciencemag.org/content/3/27/eaat5861

SUPPLEMENTARY MATERIALS
http://immunology.sciencemag.org/content/suppl/2018/09/10/3.27.eaat5861.DC1

REFERENCES
This article cites 69 articles, 12 of which you can access for free
http://immunology.sciencemag.org/content/3/27/eaat5861#BIBL

Use of this article is subject to the Terms of Service