IMMUNOGENOMICS

Molecular diversification of regulatory T cells in nonlymphoid tissues

Joanna R. DiSpirito[1,2]*, David Zemmour[1,2]*, Deepshika Ramanan[1,2], Jun Cho[1,2], Rapolas Zilionis[3,4], Allon M. Klein[3], Christophe Benoist[1,2], Diane Mathis[1,2]†

Foxp3$^+$CD4$^+$ regulatory T cells (T_{regs}) accumulate in certain nonlymphoid tissues, where they control diverse aspects of organ homeostasis. Populations of tissue T_{regs}, 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 T_{regs} 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 T_{reg} 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 T_{regs} 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 families 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 T_{regs}. 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 (T_{regs}) 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 T_{regs}. Most T_{regs} arise as such in the thymus (tT_{regs}), but certain specialized populations derive from conventional CD4$^+$ T cells in the periphery (pT_{regs}), a process that can be partially recapitulated in vitro (iT_{regs}).

Our initial view of T_{reg} phenotype and function was largely focused on the control of effector T lymphocytes by T_{regs} circulating through lymphoid organs. However, substantially more complexity soon became apparent [reviewed in (*2*)]. T_{regs} can control other adaptive immune system cells, a variety of innate immunocytes, and even nonlymphoid cells. Moreover, T_{reg} populations with distinguishable phenotypes rein in immune responses of diverse types [for example, driven by T helper 1 (T_H1), T_H2, T_H17, or B cells], sometimes co-opting transcription factors (TFs) or signaling pathways characteristic of the target populations. T_{regs} 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 T_{reg}" 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 T_{regs} have a transcriptome strikingly different from those of lymphoid organ and other nonlymphoid-tissue T_{reg} 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 T_{regs} 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 T_{regs}, notably PPARγ ligands (*5*) and interleukin-33 (*4*, *6*). Other tissue T_{reg} 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 T_{reg} cell heterogeneity, for nonlymphoid-tissue T_{regs} 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 T_{reg} 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 T_{reg} transcriptomes.

[1]Division of Immunology, Department of Microbiology and Immunobiology, Harvard Medical School, Boston, MA 02115, USA. [2]Evergrande Center for Immunologic Diseases, Harvard Medical School and Brigham and Women's Hospital, Boston MA 02115, USA. [3]Department of Systems Biology, Harvard Medical School, Boston, MA 02115, USA. [4]Institute of Biotechnology, Vilnius University, Vilnius, LT 10257, Lithuania.
*These authors contributed equally to this work.
†Corresponding author. Email: cbdm@hms.harvard.edu

RESULTS

A more nuanced view of nonlymphoid-tissue T_{reg} transcriptomes

As a first step toward elucidating the architecture and regulation of tissue T_{reg} transcriptomes, we exploited existing microarray data sets derived from double-sorted T_{regs} from VAT (*19*), injured skeletal muscle (*7*), and the colonic lamina propria (*9*), coupled with corresponding splenic T_{reg} 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 T_{reg} 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 T_{reg} 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 T_{regs} by their tissue source, the first three principal components (PCs) accounting for 70% of the variance (Fig. 1A). Muscle T_{regs} were most closely related to splenic T_{regs}, a finding that concurs with previous functional data demonstrating that muscle (*8*), but not VAT (*4*), T_{regs} communicate extensively with the circulating T_{reg} 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 T_{regs} (*22*). Not surprisingly, genes specifically up-regulated in VAT T_{regs} were implicated in lipid metabolism processes, consistent with VAT T_{reg} accumulation of intracellular lipid droplets (*5*); whereas transcripts overrepresented primarily in muscle T_{regs} 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 T_{reg} data set is consistent with this growth factor's participation in several gut processes (*23*). Thus, transcriptional control of nonlymphoid-tissue T_{regs} appears to be layered, encompassing pan-tissue and tissue-specific elements.

Tissue-associated variability in T_{reg} chromatin accessibility

It seemed likely that the layered tissue T_{reg} 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 T_{reg} and contemporaneously harvested splenic T_{reg} populations from the same mice. Again, the data were explored as nonlymphoid tissue:lymphoid organ fold changes within

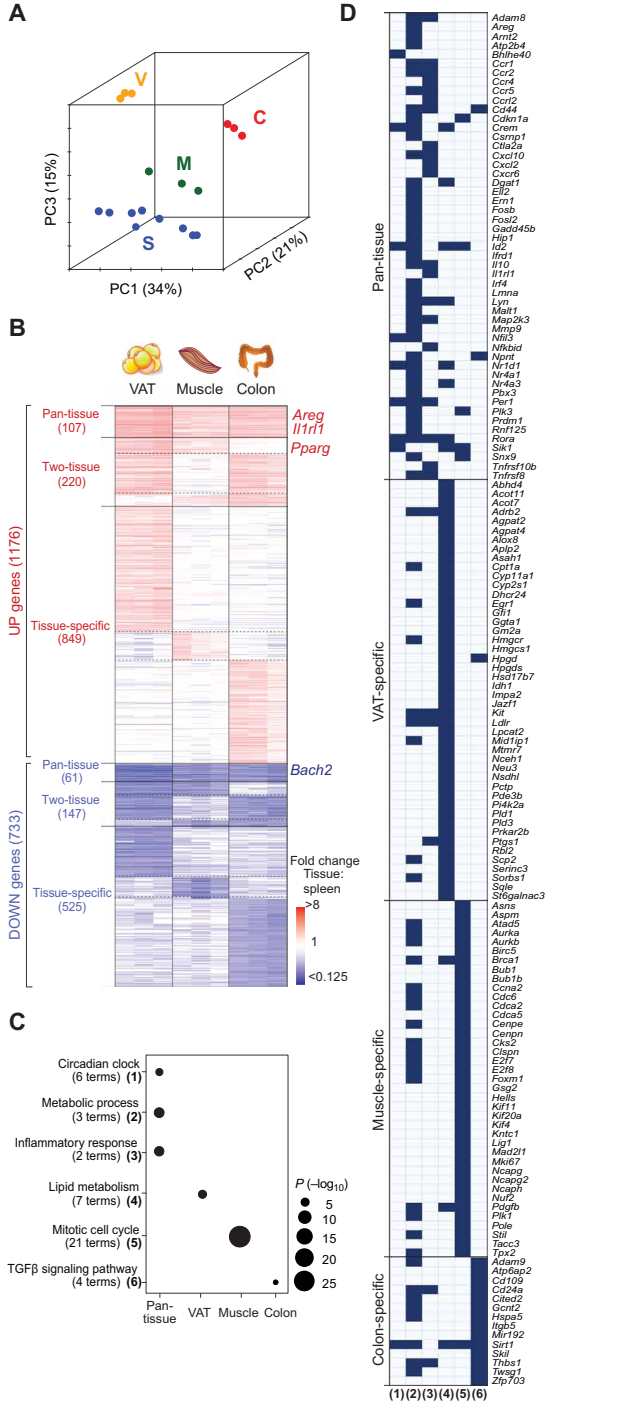


Fig. 1. Tissue T_{reg} gene sets. Microarray data sets for highly purified tissue T_{regs} from epididymal VAT (V) (*5*), injured skeletal muscle (M) (*60*), and colonic lamina propria (C) (*9*), as well as from control lymphoid organ T_{regs} ($n = 3$ for each tissue). (**A**) Principal components analysis. 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 T_{regs}: 1909 genes ordered by gene set (with numbers of genes in parentheses). (**C**) Gene ontology (GO) enrichment in the pan-tissue and tissue-specific gene sets (*61*). Numbers of terms enriched under each classification are indicated in parentheses. Fisher's exact test, *P* values. TGFβ, transforming growth factor–β. (**D**) Highlighted genes. Genes included under the GO terms enriched in the different tissues. Column numbers correspond to 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 T_{regs} (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 T_{regs} 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 T_{reg} populations confirmed that the biological replicates clustered together and that all nonlymphoid-tissue T_{reg} populations separated from their splenic T_{reg} counterparts (Fig. 2A). Again, muscle T_{regs} were the most closely related to splenic T_{regs}, and VAT and colon T_{regs} were more similar to each other than to muscle T_{regs}. 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 T_{reg} 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).

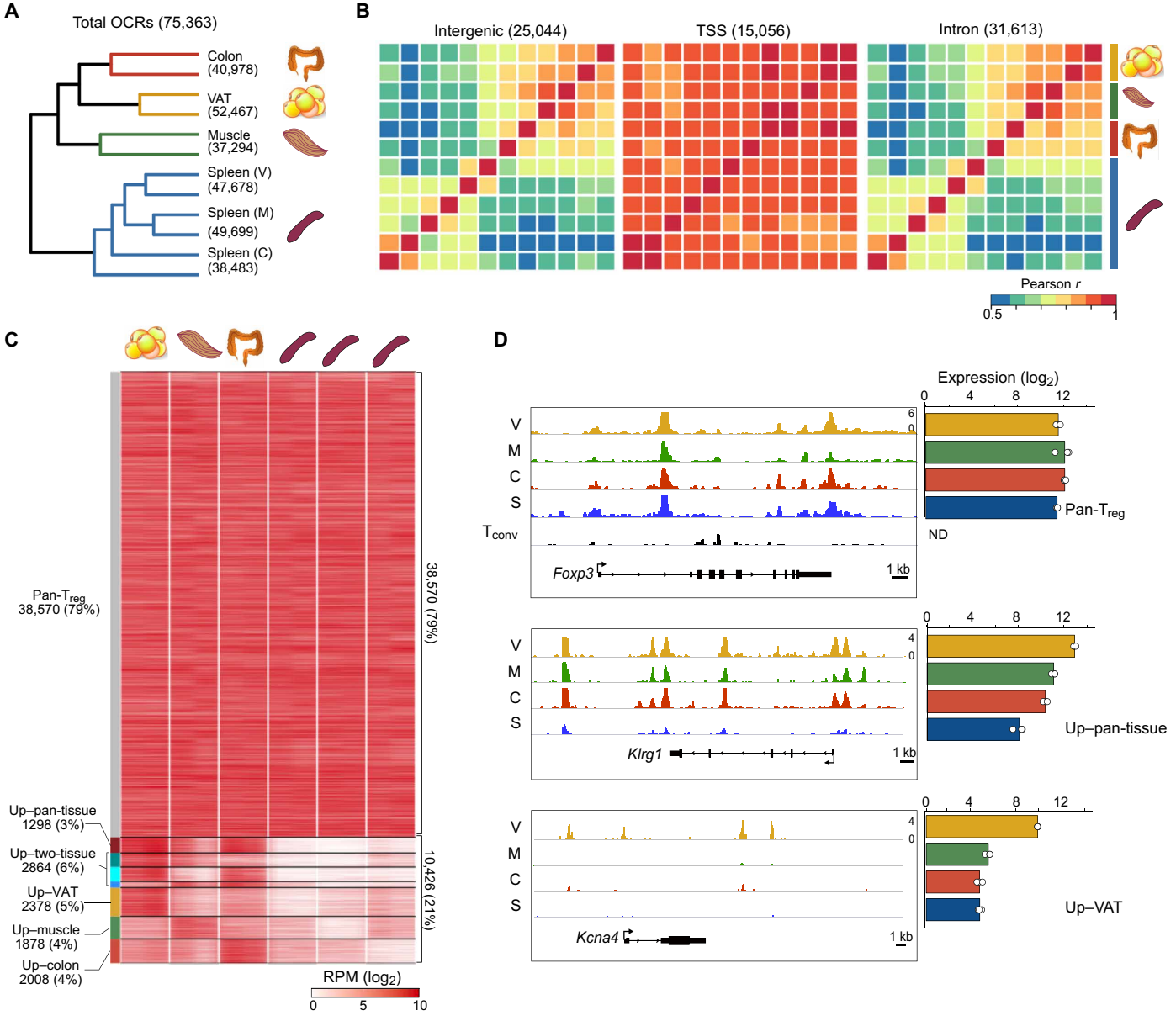


Fig. 2. Tissue-associated variation in chromatin accessibility is concentrated at TSS-distal regions. (**A**) Hierarchical clustering of T_{regs} by their pairwise Pearson correlation values for all 75,363 OCRs. (**B**) Heatmaps of Pearson pairwise correlation scores for T_{reg} OCRs located within various functional genomic elements. T_{reg} 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 T_{reg} pairing were clustered by whether their "up" status was shared across all three, two, or only one nonlymphoid tissue. Pan-T_{reg} 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 T_{reg} subsets. The T_{conv} track shows minimal accessibility at *Foxp3* in non–T_{reg} $CD4^+$ 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 tissue:spleen couples, revealing ~10 to 13,000 differentially accessible regions per tissue (fig. S1C). As performed above to define nonlymphoid-tissue T_{reg} 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-T_{reg}"; 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 T_{regs}. 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-T_{reg} OCRs constituted the majority of accessible regions around *Foxp3*, expressed in all T_{regs}, but not in conventional T (T_{conv}) cells, consistent with its lack of accessibility in T_{conv} 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, T_{regs}, and VAT-specific OCRs mirrored the VAT-specific expression of *Kcna4* (Fig. 2D). These data indicate that tissue-associated variation in chromatin accessibility, located primarily at promoter-distal regions, was a property of nonlymphoid-tissue T_{reg} populations.

OCRs shared between nonlymphoid tissue and lymphoid organ T_{regs} (i.e., pan-T_{reg} OCRs) were by far the most common (Fig. 2C). We expected to find them near genes transcribed in a pan-T_{reg} 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 T_{regs}, 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 T_{regs} 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 T_{reg} populations, it is possible that what we have termed priming of OCRs in splenic T_{regs} represented ATAC signal from a small population of nonlymphoid-tissue T_{regs} 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 T_{regs} 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 T_{regs} than in the tissue T_{reg} 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 T_{regs}, 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 T_{reg} gene expression did not merely reflect pan-hematopoietic or T lineage–specifying OCRs but instead showed a high degree of specificity for T_{regs}.

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 T_{regs} (*30*), we asked whether this chromatin element might regulate geographic or functional diversification within the T_{reg} 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 T_{regs}. 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 T_{reg} 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 T_{regs} 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 T_{regs}. Therefore, we calculated the frequency of tissue-restricted genes with an associated SE in splenic T_{regs} (±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 T_{regs}. 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 T_{regs} (*30*), we quantified

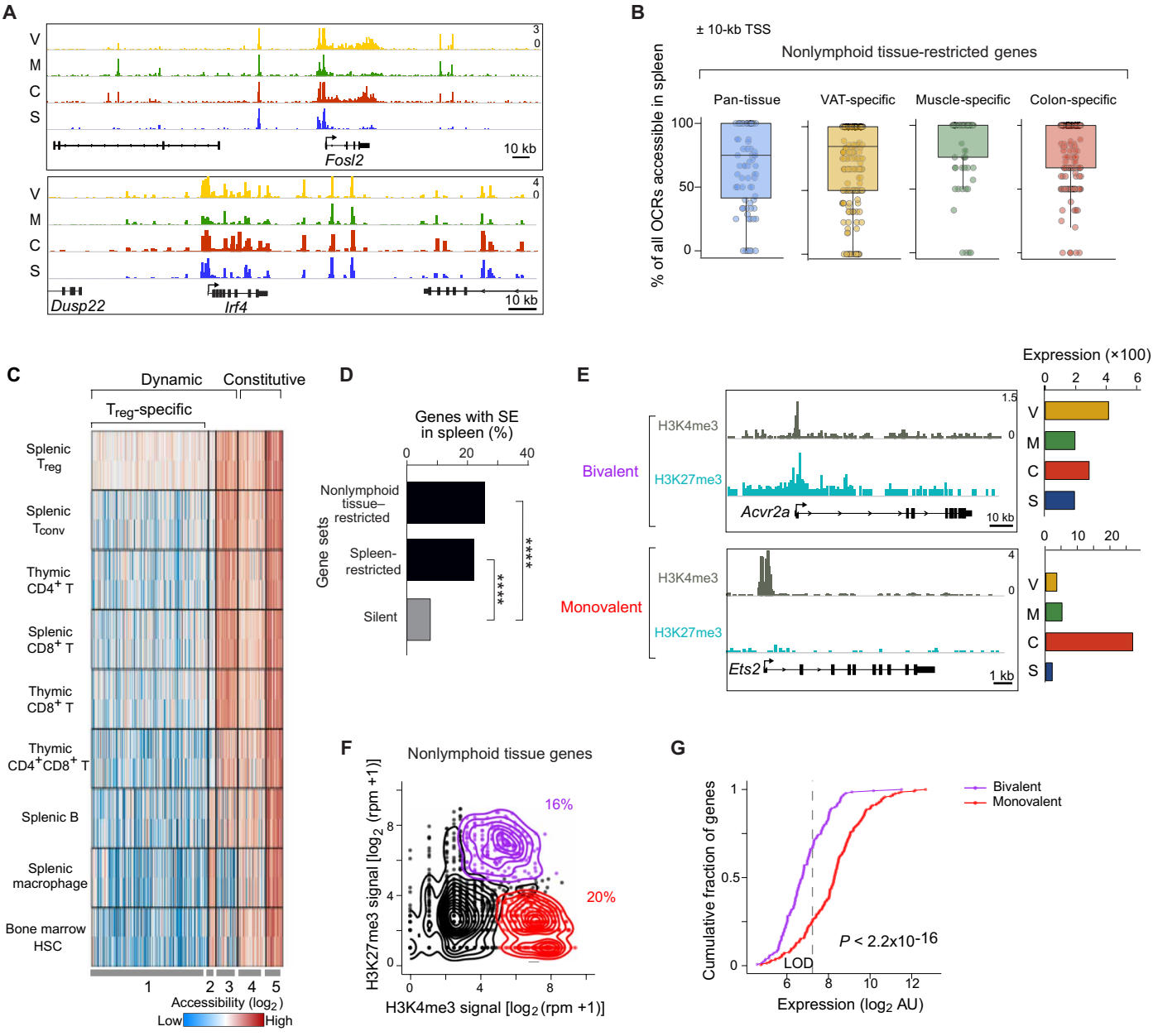


Fig. 3. Priming of tissue-restricted genes by widespread open chromatin, SEs, and bivalent TSSs. (**A**) ATAC-seq reads for two tissue–restricted genes with prominent OCRs in the spleen. (**B**) Quantification of OCR priming in the spleen for tissue-restricted genes. For each pan-tissue or tissue-specific up-regulated gene (dots), all OCRs were identified within 10 kb of its TSS, and the fraction open in the spleen are plotted. See also fig. S2. (**C**) Heatmap shows accessibility of the top quartile of primed OCRs (~2100) across T cell, myeloid, and progenitor cell types. Data are from the ImmGen ATAC-seq resource. HSC, hematopoietic stem cell. (**D**) The association of SEs in splenic T_{regs} with genes whose expression is either spleen- or tissue-restricted, compared with the background association of SEs with silent genes. ****$P < 0.0001$, binomial test. See also fig. S3. (**E**) Examples of tissue-restricted genes whose TSSs are either monovalently or bivalently marked by histone modifications in splenic T_{regs}. Note the minimal expression of these genes in splenic T_{regs} (according to microarray data). (**F**) Histone modifications at the TSSs of tissue-restricted genes in splenic T_{regs} ($n = 1135$ genes). For each gene, the levels of H3K4me3 and H3K27me3 within 1 kb of its TSS are shown, derived from reanalysis of published ChIP-seq data (*30*). Colors represent TSSs passing cutoffs for monovalency (H3K4me3 alone, red), bivalency (H3K4me3 + H3K27me3, purple), or TSSs below both cutoffs (black). (**G**) Expression differences in splenic T_{regs} between monovalently and bivalently marked tissue-restricted genes. *P* value is from a two-sided Kolmogorov-Smirnov test. LOD, level of detection; AU, arbitrary units.

bivalent histone marks associated with the various gene sets in splenic T_{regs}. 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 T_{regs}, 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 T_{regs} 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 T_{regs}, 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 T_{reg} 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 T_{reg} 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 T_{reg} 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 T_{regs}. 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 T_{regs}, 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 T_{reg} 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 T_{reg} specialization within tissues.

Therefore, we sorted T_{regs} from VAT, skeletal muscle (4 days after injury), and the colonic lamina propria plus, as controls, splenic T_{reg} and T_{conv} 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 T_{regs}. On average, 1885 unique mRNA molecules transcribed from 759 different genes were sequenced per cell, effectively covering the T_{reg} 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 dimensionality 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 T_{regs} and T_{conv} 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, T_{reg} cell space, and vice versa. This finding matches previous observations that muscle (*8*), but very rarely VAT (*4*), T_{regs} exchange with the circulating pool. *Foxp3* was expressed throughout the T_{reg} populations but not in T_{conv} cells, as expected (Fig. 5B). Various TFs have been reported to drive transcriptomic diversification of particular T_{reg} populations. Accordingly, we found *Pparg* to be expressed predominantly in VAT T_{regs} (*5*); *Tbx21* in a subset of muscle and colonic T_{regs} (*36*, *37*); *Ikzf2* and *Rorc* in distinct subpopulations of colonic T_{regs} (as well as in T_{regs} in other tissues) (*9*, *11*); and *Cxcr5*, a marker of T follicular regulatory cells (*38*, *39*), predominantly in splenic and colonic T_{regs}.

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-T_{reg} signature was up-regulated in all T_{regs} in comparison with splenic T_{conv} cells. This signature was more highly represented in nonlymphoid-tissue T_{regs} than splenic T_{regs}, consistent with the fact that the pan-T_{reg} gene set includes loci related to T_{reg} activation (*40*) and that nonlymphoid-tissue T_{regs} tend to have an activated phenotype compared with lymphoid organ T_{regs} (*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 T_{regs}; those rare cells probably correspond to the few splenic T_{regs} invading the tissue T_{reg} space in Fig. 5A. VAT T_{regs} were the most distinct population, with higher expression of the pan-tissue signature and a highly differential VAT-specific signature. Muscle T_{regs}, 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 T_{reg} population might reflect a more heterogeneous conglomeration of evolving subpopulations, as might be expected with an acute injury context, we

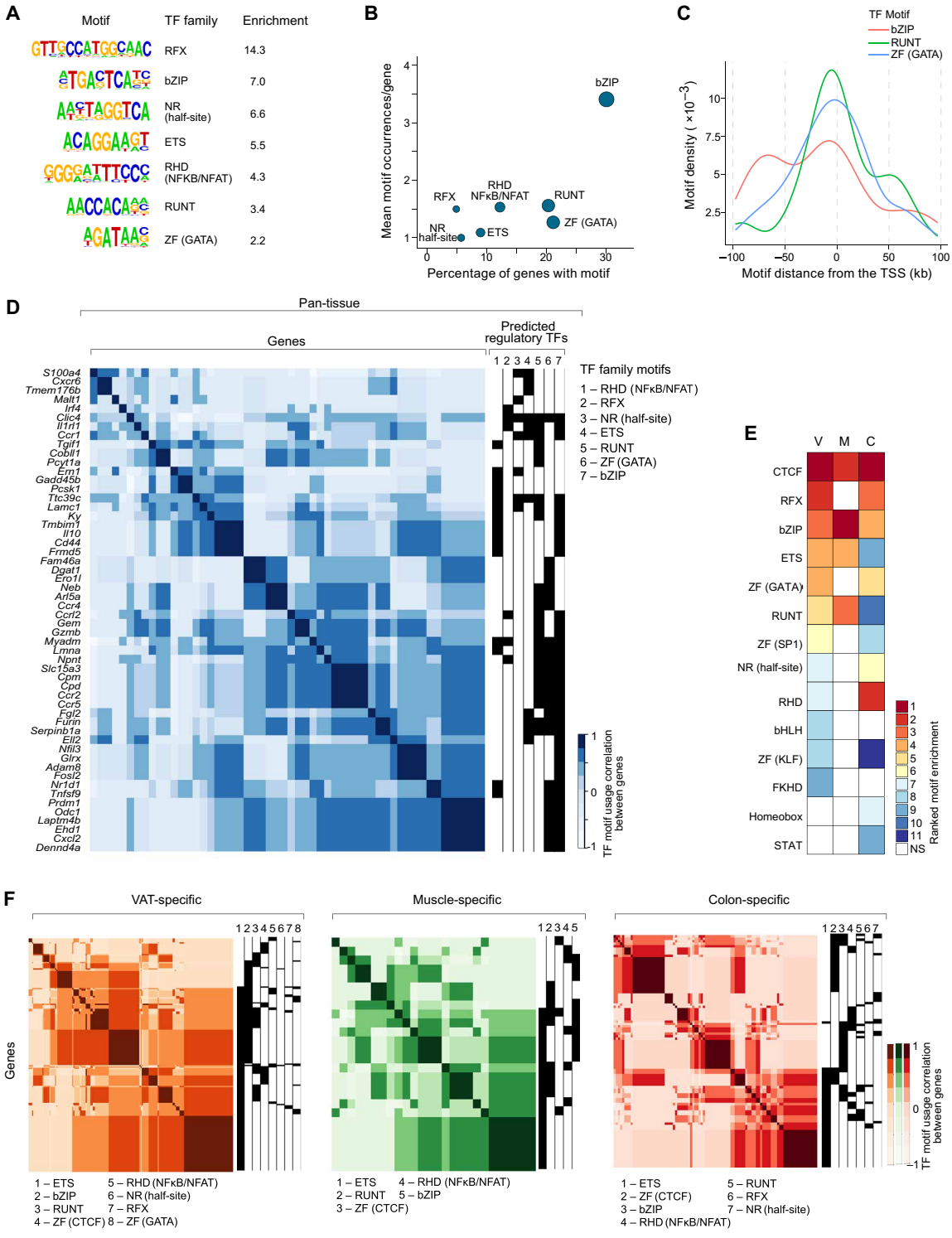


Fig. 4. A limited number of TF family motifs are associated with tissue-restricted gene signatures in T_{regs}. (**A**) TF family motifs enriched in OCRs within 100 kb of the TSS of genes up-regulated in T_{regs} from all three nonlymphoid tissues. Enrichment represents the fold change in frequency of 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 ≥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.

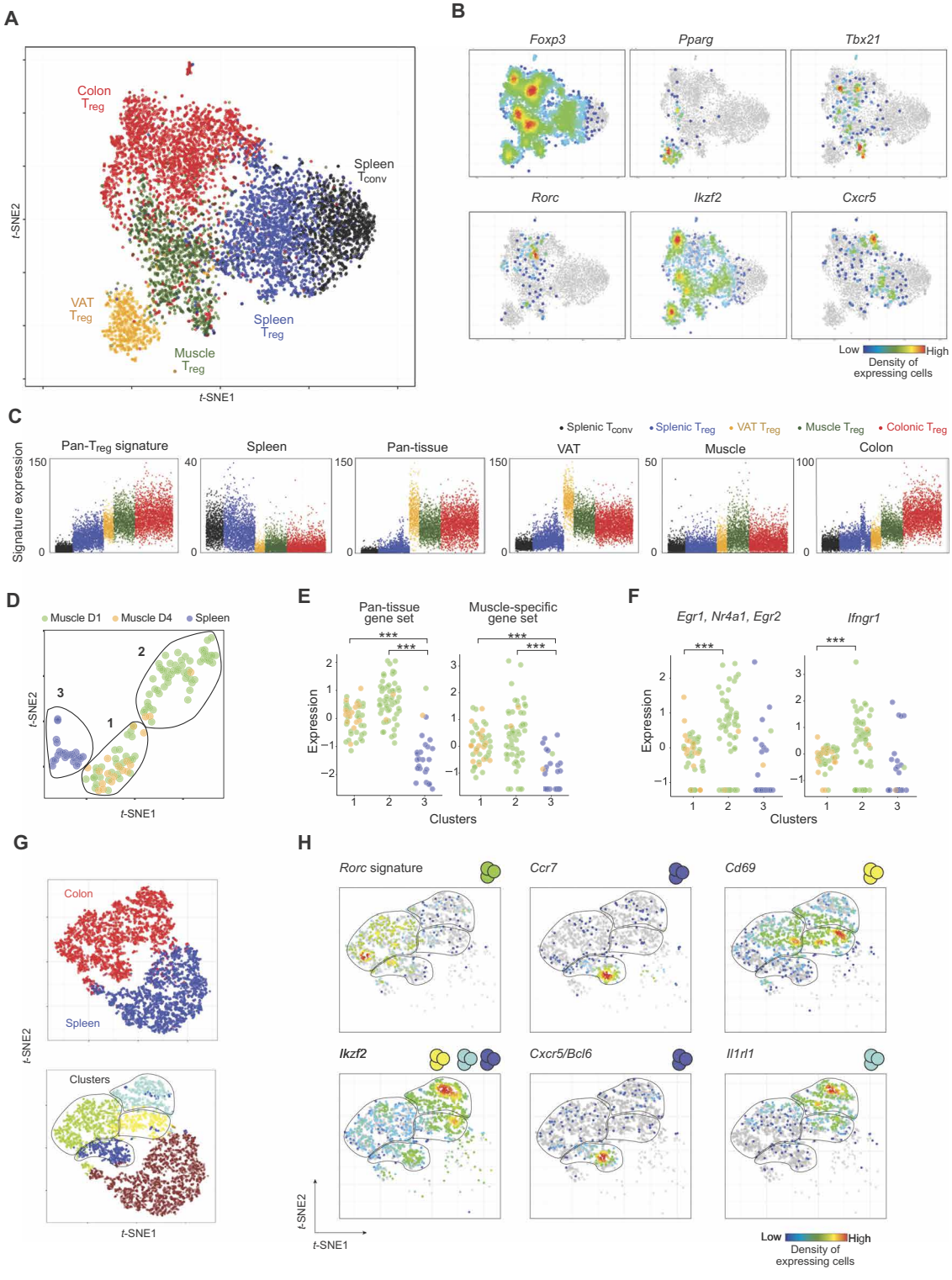


Fig. 5. scRNA-seq analysis of tissue T_{regs}. (**A**) *t*-SNE plot of the tissue and splenic T_{regs} single-cell data sets (n = 7455 cells). (**B**) Same *t*-SNE plots as in (A), highlighting cells expressing characteristic T_{reg} TFs or subpopulation markers: *Pparg* (VAT T_{regs}), *Rorc* (colonic pT_{regs}), *Ikzf2* (tT_{regs}), *Tbx21* (T_{regs} in a T_H1 milieu), and *Cxcr5* (follicular T_{regs}). Coloring signifies the density of expressing cells. (**C**) Single-cell expression of the indicated gene sets, as per Fig.1 (*z* score). Pan-T_{reg} signature from Hill *et al.* (*40*). The spleen signature is the pan-tissue down signature (as per Fig.1). (**D**) *t*-SNE plot of the CEL-Seq data sets of spleen and muscle T_{regs} 1 (D1) and 4 days (D4) after injury. Lines delimit cell clusters 1 to 3 identified by *k*-means. (**E**) Single-cell expression of the pan-tissue and muscle-specific gene sets in splenic (blue) and muscle T_{regs} isolated at days 1 (green) and 4 (yellow) after injury (*z* score). ***$P < 0.001$ by two-tailed *t* test. (**F**) As in (E), except expression of early TCR-induced genes (*Nr4a1*, *Egr1*, and *Egr2*) (left) or the *Ifgnr1* gene (right); ***$P < 0.001$. Colors relate to the indicated gene sets. See also fig. S4. (**G**) *t*-SNE plot of colonic and splenic T_{reg} single-cell transcriptomes. Top: Colonic (red) and splenic (blue) T_{regs}. Bottom: Colors and lines delimit cell clusters identified by *k*-means. (**H**) Same *t*-SNE plot as in (G) showing the density of cells expressing specific markers separating the different colonic clusters: *Ikzf2*, *Ccr7*, *Cxcr5/Bcl6*, *Cd69*, *Il1rl1*, and the *Rorc* signature.

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 T_{reg} 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 T_{reg} population, most notably between the tT_{reg} and pT_{reg} 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 T_{reg} subsets, we carefully reexamined this issue. Unsupervised clustering of only the colonic and control splenic T_{reg} single-cell data delineated four robust cell clusters (Fig. 5G). The first level of heterogeneity distinguished pT_{regs} expressing *Rorc* transcripts (green cluster) from tT_{regs} transcribing *Ikzf2* (blue, yellow, and turquoise clusters; Fig. 5H, left). Among the tT_{regs}, 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 tT_{reg} subsets were distinguished by differential expression of a transcriptional program promoting tissue repair (*Il1rl1*, *Areg*, *Penk*, etc.; Fig. 5H, right).

A tissue T_{reg} 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 T_{regs}, 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 T_{regs} 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 issuing 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 T_{reg} genes were connected to loci encoding the AP-1 subfamily of bZIP TFs (*Bach2*/*Crem*/*Fosb*/*Jdp2*/*Junb*/*Jund*); to *Nr1d1* and *Nr4a1*, which encode nuclear receptors; and to *Gata3*. VAT- and colon-specific T_{reg} 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 *Rora* 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 T_{reg} network, genes encoding several members of the AP-1 subfamily (*Bach2*, *Crem*, *Fosb*, *Jdp2*, *Junb*, and *Jund*) 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 T_{reg} 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 *Jund*) were connected not only to tissue-specific gene targets but also to loci encoding tissue-specific TFs (*Fli1*, *Fosb*, *Maff*, etc.) that regulated the same set of gene targets. For example, the loci encoding certain broadly expressed TFs (*Jdp2*/*Jund*/*Junb*) 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 T_{reg} 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 T_{regs} (*47*). In our network, *Bach2* expression was highly anticorrelated with transcription of tissue-restricted T_{reg} 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 T_{regs} vis-à-vis their splenic counterparts (Fig. 8A). Thus, we predicted that BACH2 would bind to (and repress) tissue-restricted genes in lymphoid-organ

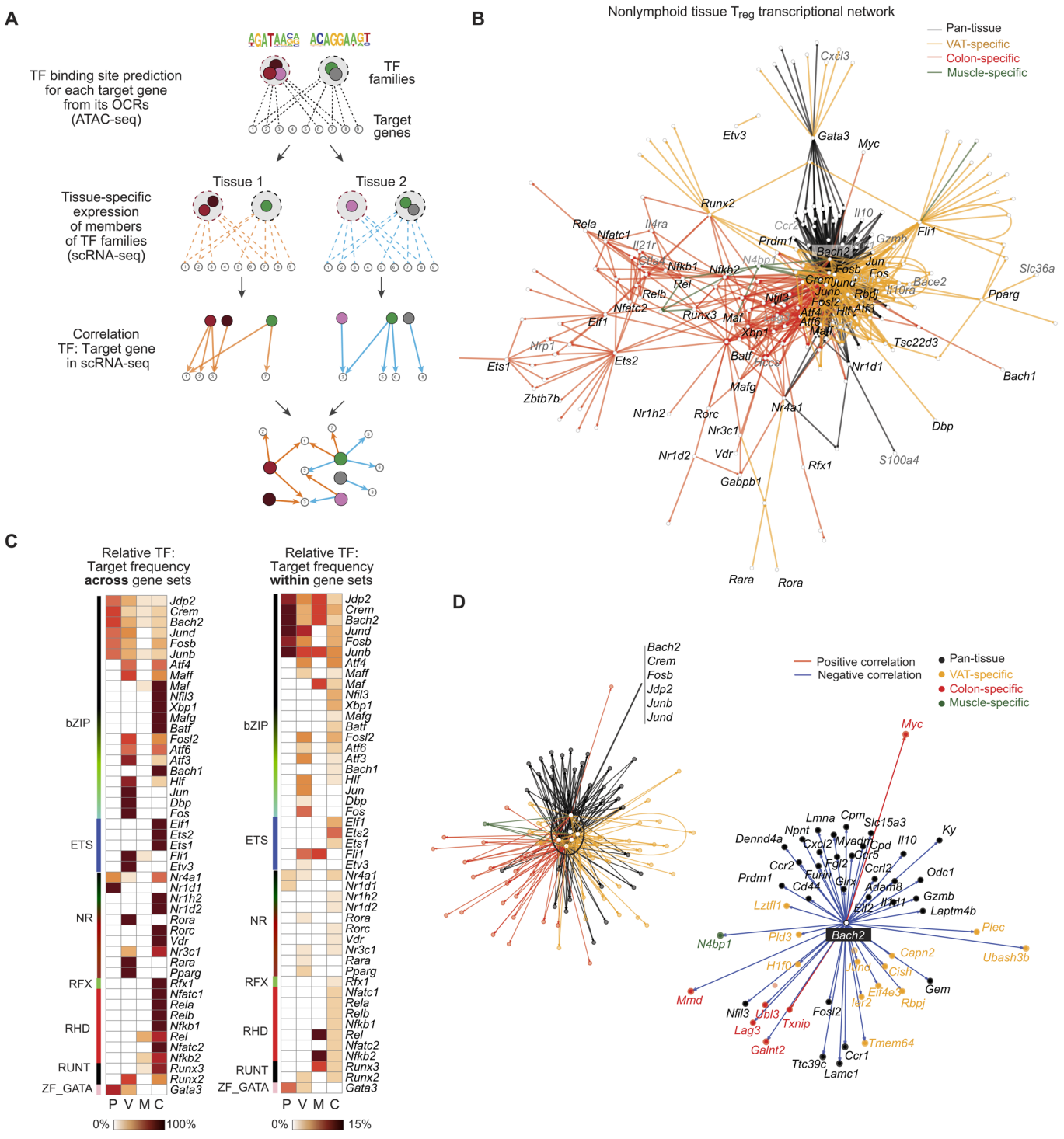


Fig. 6. Architecture of the tissue T_{reg} transcriptional network. (**A**) A schematic of network construction integrating ATAC-seq and scRNA-seq data. First, TF families are linked to target genes when their binding site is enriched in the associated OCRs (see ATAC-seq correlation heatmaps in Fig. 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 T_{reg} 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 T_{reg} 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.

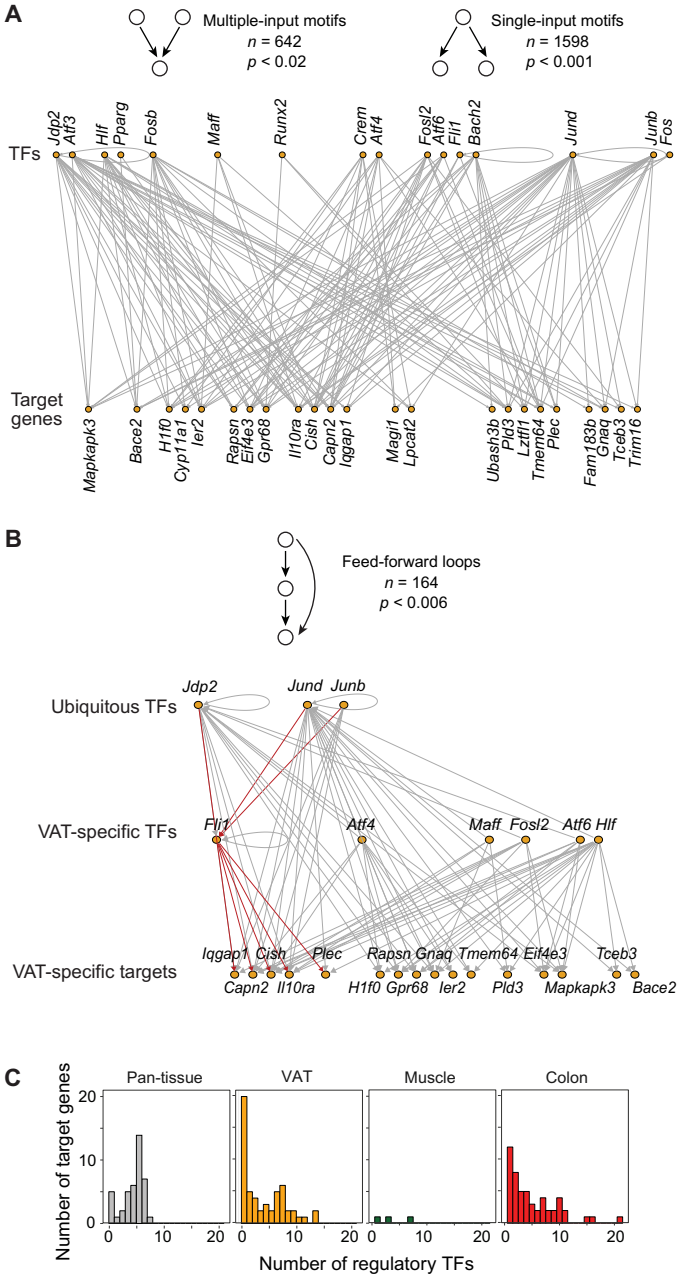


Fig. 7. VAT T_{reg} transcriptional network connectivity. (**A**) SIM and MIM instances in the VAT T_{reg} subnetwork from Fig. 6B. Top and bottom rows represent TFs and gene targets, respectively. Schematics of the motif types on top, with the number of instances and *P* values for enrichment (null distribution based on 1000 degree-preserving randomizations; see Supplementary Materials and Methods). (**B**) Feed-forward loop instances that include the ubiquitous TFs in the VAT T_{reg} subnetwork from Fig. 6B. Top, middle, and bottom rows represent the ubiquitous TFs, VAT-specific TFs, and target genes, respectively. Red lines are particular examples discussed in the text. Schematics of the motif types on top, with the number of instances and *P* values for enrichment. (**C**) Distribution of the number of connected TFs in each gene set.

T_{regs}, and that its expression must be extinguished in tissue T_{regs} for their signature genes to be turned on. We reanalyzed published BACH2 ChIP-seq data on T_{regs} generated in vitro from precursors in lymphoid organs (*47*), which revealed that BACH2's average occupancy on both pan-tissue and tissue-specific genes was higher than 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 T_{reg} gene sets in the in vitro–generated T_{regs} (Fig. 8C). Thus, the inverse correlation between levels of BACH2 binding and expression of tissue-restricted T_{reg} genes was likely to be a direct effect of its repressive activity at these loci in lymphoid organ T_{regs}, consistent with a previous report that it can repress the expression of T_H cell signature genes in T_{regs} (*47*).

As a predicted regulator of tissue-specific gene expression in T_{regs}, we revisited the role of PPARγ, a nuclear receptor family member required for the accumulation and function of bona fide VAT T_{regs} (*5*). Our network predicted functionality of PPARγ exclusively in VAT T_{regs}, although its expression was also strongly up-regulated in skeletal muscle vis-à-vis lymphoid organ T_{regs} (Fig. 8D). First, we examined mice that carry a T_{reg}-specific knockout of *Pparg* and, consequently, have a dearth of VAT T_{regs} (*5*). Examination of the T_{reg} 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 T_{regs} 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 T_{reg} population in VAT but did not change T_{reg} frequencies in the muscle, colon, or spleen (Fig. 8F).

DISCUSSION

Integrating accessible chromatin and single-cell transcriptome profiling, we explored the regulatory underpinnings of T_{reg} diversity in nonlymphoid tissues. At both the chromatin and transcription levels, adaptation of T_{regs} 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 T_{regs} 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 T_{regs}, 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 T_{regs} were distinct from splenic T_{regs} by scRNA-seq, and at day 4 the muscle T_{reg} 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 T_{reg} 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 T_{regs} 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 T_{reg}-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

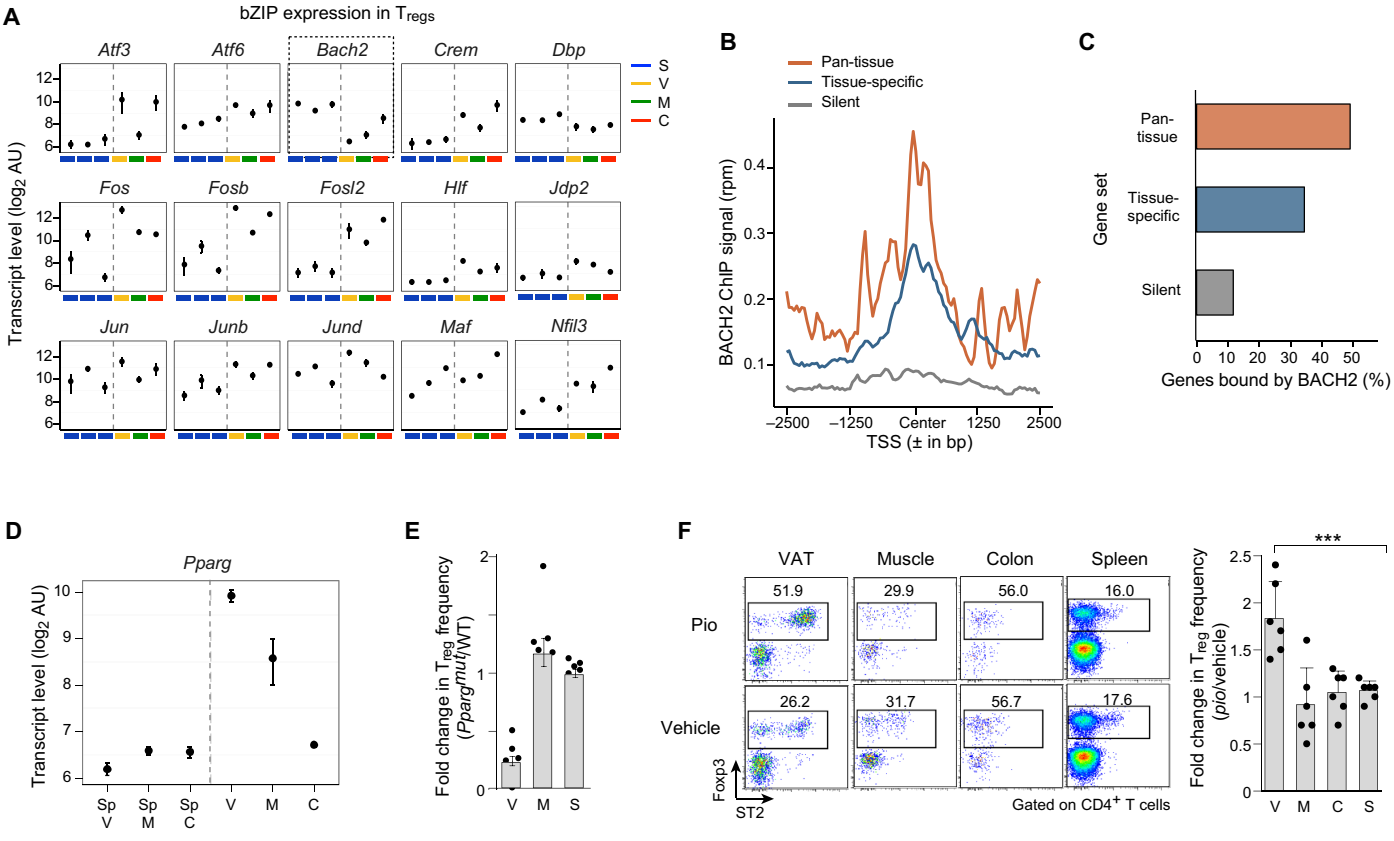


Fig. 8. Validation of tissue T_{reg} TF networks. (**A**) Expression of bZIP family TFs across splenic and nonlymphoid-tissue T_{reg} subsets. AUs represent transcript levels from microarray data. (**B**) Average genomic occupancy of BACH2 in spleen-derived induced T_{regs} at TSSs of genes up-regulated in all three tissue T_{reg} subsets (pan-tissue), only one of them (tissue-specific VAT, muscle, and colon combined into a single list for this analysis), or a control set of genes not expressed in T_{regs}. Raw ChIP-seq data from GEO GSE45975 were reanalyzed for this study (see Materials and Methods). bp, base pairs. (**C**) Percentage of genes in each set that have a BACH2 peak within 10 kb of their TSS. (**D**) Expression of *Pparg* across T_{reg} subsets, as in (A). (**E**) Effects of T_{reg}-specific PPARγ deletion on T_{reg} accumulation in skeletal muscle on day 4 after cardiotoxin injury (the peak of T_{reg} accumulation). Data are represented as the fold change in the T_{reg} frequency of each $PPARg^{mut}$ mouse versus the average frequency in wild-type (WT) mice in muscle (M) or spleen (S), within each experiment. $n = 3$ to 5 mice per group from three experiments. (**F**) Left: Frequencies of $Foxp3^+$ T_{regs} in various tissues after a 10-day treatment with pioglitazone (10 mg/kg). Dot plots are representative of $n = 2$ to 3 mice per group from two experiments. Right: Effects of pioglitazone on each T_{reg} subset measured as fold change in the T_{reg} frequency of each pioglitazone-treated mouse versus the average frequency from vehicle-treated mice within each experiment. $n = 2$ to 3 mice per group from two experiments.

within lymphoid organs, the effects of local stimuli on T_{regs}, 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 T_{regs}. 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 T_{reg} 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 T_{reg} diversification: Transcriptional programs must be established in nonlymphoid-tissue T_{regs} within the constraints of an epigenome that maintains T_{reg} identity while allowing tissue adaptation for optimum installation and function. To that end, the chromatin structure of T_{regs} in lymphoid organs is primed, through the establishment of pan-T_{reg} OCRs, to license the rapid transcriptional turn-on or up-regulation of most tissue-specific genes. The primed stretches may facilitate remodeling 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 elements and cooperate to regulate tissue T_{reg} genes via two main 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 nonlymphoid-tissue T_{regs} but facilitate activation of secondary, 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_{regs} from VAT, skeletal muscle, and colonic lamina propria (pooled from several mice) and corresponding lymphoid organ T_{regs}. 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.*Foxp3*$^{Ires\text{-}Gfp}$ (*52*), B6.*Foxp3*$^{Yfp\text{-}Cre}$ (*53*), and B6.*Pparg*$^{fl/fl}$ (*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 IS00001257). Mutant *Pparg* T_{reg} mice (T_{reg}-*Pparg*^{mut}) 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* cardiotoxin (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 identically. 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 ≥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 DNase (*58*), with an FDR of 0.001. Only OCRs occurring in both replicates were used in further 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 *Foxp3*$^{Cd90.1}$ or *Foxp3*$^{Ires\text{-}Gfp}$ 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 re-analyzed 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.sciencemag.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_{regs} 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

1. G. Plitas, A. Y. Rudensky, Regulatory T cells: Differentiation and function. *Cancer Immunol. Res.* **4**, 721–725 (2016).
2. M. Panduro, C. Benoist, D. Mathis, Tissue Tregs. *Annu. Rev. Immunol.* **34**, 609–633 (2016).
3. M. Feuerer, L. Herrero, D. Cipolletta, A. Naaz, J. Wong, A. Nayer, J. Lee, A. B. Goldfine, C. Benoist, S. Shoelson, D. Mathis, Lean, but not obese, fat is enriched for a unique population of regulatory T cells that affect metabolic parameters. *Nat. Med.* **15**, 930–939 (2009).
4. D. Kolodin, P. N. van Panhuys, C. Li, A. M. Magnuson, D. Cipolletta, C. M. Miller, A. Wagers, R. N. Germain, C. Benoist, D. Mathis, Antigen- and cytokine-driven accumulation of regulatory T cells in visceral adipose tissue of lean mice. *Cell Metab.* **21**, 543–557 (2015).
5. D. Cipolletta, M. Feuerer, A. Li, N. Kamei, J. Lee, S. E. Shoelson, C. Benoist, D. Mathis, PPAR-γ is a major driver of the accumulation and phenotype of adipose tissue T_{reg} cells. *Nature* **486**, 549–553 (2012).
6. A. Vasanthakumar, K. Moro, A. Xin, Y. Liao, R. Gloury, S. Kawamoto, S. Fagarasan, L. A. Mielke, S. Afshar-Sterle, S. L. Masters, S. Nakae, H. Saito, J. M. Wentworth, P. Li, W. Liao, W. J. Leonard, G. K. Smyth, W. Shi, S. L. Nutt, S. Koyasu, A. Kallies, The transcriptional regulators IRF4, BATF and IL-33 orchestrate development and maintenance of adipose tissue-resident regulatory T cells. *Nat. Immunol.* **16**, 276–285 (2015).

7. D. Burzyn, W. Kuswanto, D. Kolodin, J. L. Shadrach, M. Cerletti, Y. Jang, E. Sefik, T. G. Tan, A. J. Wagers, C. Benoist, D. Mathis, A special population of regulatory T cells potentiates muscle repair. *Cell* **155**, 1282–1295 (2013).
8. W. Kuswanto, D. Burzyn, M. Panduro, K. K. Wang, Y. C. Jang, A. J. Wagers, C. Benoist, D. Mathis, Poor repair of skeletal muscle in aging mice reflects a defect in local, interleukin-33-dependent accumulation of regulatory T cells. *Immunity* **44**, 355–367 (2016).
9. E. Sefik, N. Geva-Zatorsky, S. Oh, L. Konnikova, D. Zemmour, A. M. McGuire, D. Burzyn, A. Ortiz-Lopez, M. Lobera, J. Yang, S. Ghosh, A. Earl, S. B. Snapper, R. Jupp, D. Kasper, D. Mathis, C. Benoist, Individual intestinal symbionts induce a distinct population of RORγ$^+$ regulatory T cells. *Science* **349**, 993–997 (2015).
10. C. Schiering, T. Krausgruber, A. Chomka, A. Fröhlich, K. Adelmann, E. A. Wohlfert, J. Pott, T. Griseri, J. Bollrath, A. N. Hegazy, O. J. Harrison, B. M. J. Owens, M. Löhning, Y. Belkaid, P. G. Fallon, F. Powrie, The alarmin IL-33 promotes regulatory T-cell function in the intestine. *Nature* **513**, 564–568 (2014).
11. C. Ohnmacht, J.-H. Park, S. Cording, J. B. Wing, K. Atarashi, Y. Obata, V. Gaboriau-Routhiau, R. Marques, S. Dulauroy, M. Fedoseeva, M. Busslinger, N. Cerf-Bensussan, I. G. Boneca, D. Voehringer, K. Hase, K. Honda, S. Sakaguchi, G. Eberl, The microbiota regulates type 2 immunity through RORγ$^+$ T cells. *Science* **349**, 989–993 (2015).
12. T. C. Scharschmidt, K. S. Vasquez, M. L. Pauli, E. G. Leitner, K. Chu, H.-A. Truong, M. M. Lowe, R. R. Sanchez, N. Ali, Z. G. Laszik, J. L. Sonnenburg, S. E. Millar, M. D. Rosenblum, Commensal microbes and hair follicle morphogenesis coordinately drive T_{reg} migration into neonatal skin. *Cell Host Microbe* **21**, 467–477 (2017).
13. N. Ali, B. Zirak, R. S. Rodriguez, M. L. Pauli, H.-A. Truong, K. Lai, R. Ahn, K. Corbin, M. M. Lowe, T. C. Scharschmidt, K. Taravati, M. R. Tan, R. R. Ricardo-Gonzalez, A. Nosbaum, M. Bertolini, W. Liao, F. O. Nestle, R. Paus, G. Cotsarelis, A. K. Abbas, M. D. Rosenblum, Regulatory t cells in skin facilitate epithelial stem cell differentiation. *Cell* **169**, 1119–1129.e11 (2017).
14. J. D. Buenrostro, P. G. Giresi, L. C. Zaba, H. Y. Chang, W. J. Greenleaf, Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013).
15. D. Lara-Astiaso, A. Weiner, E. Lorenzo-Vivas, I. Zaretsky, D. A. Jaitin, E. David, H. Keren-Shaul, A. Mildner, D. Winter, S. Jung, N. Friedman, I. Amit, Chromatin state dynamics during blood formation. *Science* **345**, 943–949 (2014).
16. A. M. Klein, L. Mazutis, I. Akartuna, N. Tallapragada, A. Veres, V. Li, L. Peshkin, D. A. Weitz, M. W. Kirschner, Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* **161**, 1187–1201 (2015).
17. R. Zilionis, J. Nainys, A. Veres, V. Savova, D. Zemmour, A. M. Klein, L. Mazutis, Single-cell barcoding and sequencing using droplet microfluidics. *Nat. Protoc.* **12**, 44–73 (2017).
18. D. Zemmour, R. Zilionis, E. Kiner, A. M. Klein, D. Mathis, C. Benoist, Single-cell gene expression reveals a landscape of regulatory T cell phenotypes shaped by the TCR. *Nat. Immunol.* **19**, 291–301 (2018).
19. D. Cipolletta, P. Cohen, B. M. Spiegelman, C. Benoist, D. Mathis, Appearance and disappearance of the mRNA signature characteristic of T_{reg} cells in visceral adipose tissue: Age, diet, and PPARγ effects. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 482–487 (2015).
20. M. Delacher, C. D. Imbusch, D. Weichenhan, A. Breiling, A. Hotz-Wagenblatt, U. Träger, A.-C. Hofer, D. Kagebein, Q. Wang, F. Frauhammer, J.-P. Mallm, K. Bauer, C. Herrmann, P. A. Lang, B. Brors, C. Plass, M. Feuerer, Genome-wide DNA-methylation landscape defines specialization of regulatory T cells in tissues. *Nat. Immunol.* **18**, 1160–1172 (2017).
21. G. Asher, P. Sassone-Corsi, Time for food: The intimate interplay between nutrition, metabolism, and the circadian clock. *Cell* **161**, 84–92 (2015).
22. E. Cretney, A. Kallies, S. L. Nutt, Differentiation and function of Foxp3$^+$ effector regulatory T cells. *Trends Immunol.* **34**, 74–80 (2013).
23. W. Chen, W. Jin, N. Hardegen, K.-J. Lei, L. Li, N. Marinos, G. McGrady, S. M. Wahl, Conversion of peripheral CD4$^+$CD25$^-$ naive T cells to CD4$^+$CD25$^+$ regulatory T cells by TGF-β induction of transcription factor *Foxp3*. *J. Exp. Med.* **198**, 1875–1886 (2003).
24. H.-Y. Shih, G. Sciumè, Y. Mikami, L. Guo, H.-W. Sun, S. R. Brooks, J. F. Urban Jr., F. P. Davis, Y. Kanno, J. J. O'Shea, Developmental acquisition of regulomes underlies innate lymphoid cell functionality. *Cell* **165**, 1120–1133 (2016).
25. J. P. Scott-Browne, I. F. López-Moyado, S. Trifari, V. Wong, L. Chavez, A. Rao, R. M. Pereira, Dynamic changes in chromatin accessibility occur in CD8$^+$ T cells responding to viral infection. *Immunity* **45**, 1327–1340 (2016).
26. A. M. Ackermann, Z. Wang, J. Schug, A. Naji, K. H. Kaestner, Integration of ATAC-seq and RNA-seq identifies human alpha cell and beta cell signature genes. *Mol. Metab.* **5**, 233–244 (2016).
27. D. Hnisz, B. J. Abraham, T. I. Lee, A. Lau, V. Saint-André, A. A. Sigova, H. A. Hoke, R. A. Young, Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934–947 (2013).
28. G. Vahedi, Y. Kanno, Y. Furumoto, K. Jiang, S. C. Parker, M. R. Erdos, S. R. Davis, R. Roychoudhuri, N. P. Restifo, M. Gadina, Z. Tang, Y. Ruan, F. S. Collins, V. Sartorelli, J. J. O'Shea, Super-enhancers delineate disease-associated regulatory nodes in T cells. *Nature* **520**, 558–562 (2015).
29. W. A. Whyte, D. A. Orlando, D. Hnisz, B. J. Abraham, C. Y. Lin, M. H. Kagey, P. B. Rahl, T. I. Lee, R. A. Young, Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307–319 (2013).
30. Y. Kitagawa, N. Ohkura, Y. Kidani, A. Vandenbon, K. Hirota, R. Kawakami, K. Yasuda, D. Motooka, S. Nakamura, M. Kondo, I. Taniuchi, T. Kohwi-Shigematsu, S. Sakaguchi, Guidance of regulatory T cell development by Satb1-dependent super-enhancer establishment. *Nat. Immunol.* **18**, 173–183 (2017).
31. J. Lovén, H. A. Hoke, C. Y. Lin, A. Lau, D. A. Orlando, C. R. Vakoc, J. E. Bradner, T. I. Lee, R. A. Young, Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell* **153**, 320–334 (2013).
32. B. E. Bernstein, T. S. Mikkelsen, X. Xie, M. Kamal, D. J. Huebert, J. Cuff, B. Fry, A. Meissner, M. Wernig, K. Plath, R. Jaenisch, A. Wagschal, R. Feil, S. L. Schreiber, E. S. Lander, A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125**, 315–326 (2006).
33. P. Voigt, W.-W. Tee, D. Reinberg, A double take on bivalent promoters. *Genes Dev.* **27**, 1318–1338 (2013).
34. E. M. Klenova, H. C. Morse III, R. Ohlsson, V. V. Lobanenkov, The novel *BORIS+CTCF* gene family is uniquely involved in the epigenetics of normal biology and cancer. *Semin. Cancer Biol.* **12**, 399–414 (2002).
35. G.-H. Wei, G. Badis, M. F. Berger, T. Kivioja, K. Palin, M. Enge, M. Bonke, A. Jolma, M. Varjosalo, A. R. Gehrke, J. Yan, S. Talukder, M. Turunen, M. Taipale, H. G. Stunnenberg, E. Ukkonen, T. R. Hughes, M. L. Bulyk, J. Taipale, Genome-wide analysis of ETS-family DNA-binding in vitro and in vivo. *EMBO J.* **29**, 2147–2160 (2010).
36. M. A. Koch, G. Tucker-Heard, N. R. Perdue, J. R. Killebrew, K. B. Urdahl, D. J. Campbell, The transcription factor T-bet controls regulatory T cell homeostasis and function during type 1 inflammation. *Nat. Immunol.* **10**, 595–602 (2009).
37. T. G. Tan, D. Mathis, C. Benoist, Singular role for T-BET$^+$CXCR3$^+$ regulatory T cells in protection from autoimmune diabetes. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 14103–14108 (2016).
38. M. A. Linterman, W. Pierson, S. K. Lee, A. Kallies, S. Kawamoto, T. F. Rayner, M. Srivastava, D. P. Divekar, L. Beaton, J. J. Hogan, S. Fagarasan, A. Liston, K. G. C. Smith, C. G. Vinuesa, Foxp3$^+$ follicular regulatory T cells control the germinal center response. *Nat. Med.* **17**, 975–982 (2011).
39. Y. Chung, S. Tanaka, F. Chu, R. I. Nurieva, G. J. Martinez, S. Rawal, Y.-H. Wang, H. Y. Lim, J. M. Reynolds, X.-H. Zhou, H.-M. Fan, Z.-M. Liu, S. S. Neelapu, C. Dong, Follicular regulatory T cells expressing Foxp3 and Bcl-6 suppress germinal center reactions. *Nat. Med.* **17**, 983–988 (2011).
40. J. A. Hill, M. Feuerer, K. Tash, S. Haxhinasto, J. Perez, R. Melamed, D. Mathis, C. Benoist, Foxp3 transcription-factor-dependent and -independent regulation of the regulatory T cell transcriptional signature. *Immunity* **27**, 786–800 (2007).
41. T. Hashimshony, F. Wagner, N. Sher, I. Yanai, CEL-Seq: Single-cell RNA-Seq by multiplexed linear amplification. *Cell Rep.* **2**, 666–673 (2012).
42. R. J. Miragaia, T. Gomes, A. Chomka, L. Jardine, A. Riedel, A. N. Hegazy, I. Lindeman, G. Emerton, T. Krausgrube, J. Shields, M. Haniffa, F. Powrie, S. A. Teichmann, Single cell transcriptomics of regulatory T cells reveals trajectories of tissue adaptation. https://www.biorxiv.org/content/early/2017/11/22/217489 (2017).
43. R. N. Ramirez, N. C. El-Ali, M. A. Mager, D. Wyman, A. Conesa, A. Mortazavi, Dynamic gene regulatory networks of human myeloid differentiation. *Cell Syst.* **4**, 416–429.e3 (2017).
44. E. Z. Macosko, A. Basu, R. Satija, J. Nemesh, K. Shekhar, M. Goldman, I. Tirosh, A. R. Bialas, N. Kamitaki, E. M. Martersteck, J. J. Trombetta, D. A. Weitz, J. R. Sanes, A. K. Shalek, A. Regev, S. A. McCarroll, Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **161**, 1202–1214 (2015).
45. A. Tanay, A. Regev, Scaling single-cell genomics from phenomenology to mechanism. *Nature* **541**, 331–338 (2017).
46. S. Mangan, U. Alon, Structure and function of the feed-forward loop network motif. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 11980–11985 (2003).
47. R. Roychoudhuri, K. Hirahara, K. Mousavi, D. Clever, C. A. Klebanoff, M. Bonelli, G. Sciume, H. Zare, G. Vahedi, B. Dema, Z. Yu, H. Liu, H. Takahashi, M. Rao, P. Muranski, J. G. Crompton, G. Punkosdy, D. Bedognetti, E. Wang, V. Hoffmann, J. Rivera, F. M. Marincola, A. Nakamura, V. Sartorelli, Y. Kanno, L. Gattinoni, A. Muto, K. Igarashi, J. J. O'Shea, N. P. Restifo, BACH2 represses effector programs to stabilize T_{reg}-mediated immune homeostasis. *Nature* **498**, 506–510 (2013).
48. D. Gosselin, V. M. Link, C. E. Romanoski, G. J. Fonseca, D. Z. Eichenfield, N. J. Spann, J. D. Stender, H. B. Chun, H. Garner, F. Geissmann, C. K. Glass, Environment drives selection and function of enhancers controlling tissue-specific macrophage identities. *Cell* **159**, 1327–1340 (2014).
49. Y. Lavin, D. Winter, R. Blecher-Gonen, E. David, H. Keren-Shaul, M. Merad, S. Jung, I. Amit, Tissue-resident macrophage enhancer landscapes are shaped by the local microenvironment. *Cell* **159**, 1312–1326 (2014).
50. A. Vasanthakumar, Y. Liao, P. Teh, M. F. Pascutti, A. E. Oja, A. L. Garnham, R. Gloury, J. C. Tempany, T. Sidwell, E. Cuadrado, P. Tuijnenburg, T. W. Kuijpers, N. Lalaoui, L. A. Mielke, V. L. Bryant, P. D. Hodgkin, J. Silke, G. K. Smyth, M. A. Nolte, W. Shi, A. Kallies,

The TNF receptor superfamily-NF-κB axis Is critical to maintain effector regulatory T Cells in lymphoid and non-lymphoid tissues. *Cell Rep.* **20**, 2906–2920 (2017).
51. M. Xu, M. Pokrovskii, Y. Ding, R. Yi, C. Au, O. J. Harrison, C. Galan, Y. Belkaid, R. Bonneau, D. R. Littman, c-MAF-dependent regulatory T cells mediate immunological tolerance to a gut pathobiont. *Nature* **554**, 373–377 (2018).
52. E. Bettelli, Y. Carrier, W. Gao, T. Korn, T. B. Strom, M. Oukka, H. L. Weiner, V. K. Kuchroo, Reciprocal developmental pathways for the generation of pathogenic effector TH17 and regulatory T cells. *Nature* **441**, 235–238 (2006).
53. Y. P. Rubtsov, J. P. Rasmussen, E. Y. Chi, J. Fontenot, L. Castelli, X. Ye, P. Treuting, L. Siewe, A. Roers, W. R. Henderson Jr., W. Muller, A. Y. Rudensky, Regulatory T cell-derived interleukin-10 limits inflammation at environmental interfaces. *Immunity* **28**, 546–558 (2008).
54. T. E. Akiyama, S. Sakai, G. Lambert, C. J. Nicol, K. Matsusue, S. Pimprale, Y.-H. Lee, M. Ricote, C. K. Glass, H. B. Brewer Jr., F. J. Gonzalez, Conditional disruption of the peroxisome proliferator-activated receptor gamma gene in mice results in lowered expression of ABCA1, ABCG1, and apoE in macrophages and reduced cholesterol efflux. *Mol. Cell. Biol.* **22**, 2607–2619 (2002).
55. N. A. Joshi, J. N. Fass, "Sickle: A Sliding-Window, Adaptive, Quality-Based Trimming Tool for FastQ Files," version 1.33 (2011); https://github.com/najoshi/sickle.
56. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
57. H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin; 1000 Genome Project Data Processing Subgroup, The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
58. S. Heinz, C. Benner, N. Spann, E. Bertolino, Y. C. Lin, P. Laslo, J. X. Cheng, C. Murre, H. Singh, C. K. Glass, Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).
59. D. Kim, G. Pertea, C. Trapnell, H. Pimentel, R. Kelley, S. L. Salzberg, TopHat2: Accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, R36 (2013).
60. D. Burzyn, C. Benoist, D. Mathis, Regulatory T cells in nonlymphoid tissues. *Nat. Immunol.* **14**, 1007–1013 (2013).
61. P. Langfelder, S. Horvath, WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008).
62. M. E. Ritchie, B. Phipson, D. Wu, Y. Hu, C. W. Law, W. Shi, G. K. Smyth, *limma* powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
63. R Development Core Team, "R: A Language and Environment for Statistical Computing" (The R Foundation for Statistical Computing, 2017); www.R-project.org.
64. A. R. Quinlan, I. M. Hall, BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
65. S. Neph, M. S. Kuehn, A. P. Reynolds, E. Haugen, R. E. Thurman, A. K. Johnson, E. Rynes, M. T. Maurano, J. Vierstra, S. Thomas, R. Sandstrom, R. Humbert, J. A. Stamatoyannopoulos, BEDOPS: High-performance genomic feature operations. *Bioinformatics* **28**, 1919–1920 (2012).
66. M. Meredith, D. Zemmour, D. Mathis, C. Benoist, Aire controls gene expression in the thymic epithelium with ordered stochasticity. *Nat. Immunol.* **16**, 942–949 (2015).
67. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
68. D. Grün, A. Lyubimova, L. Kester, K. Wiebrands, O. Basak, N. Sasaki, H. Clevers, A. van Oudenaarden, Single-cell messenger RNA sequencing reveals rare intestinal cell types. *Nature* **525**, 251–255 (2015).
69. L. van der Maaten, G. Hinton, Visualizing high-dimensional data using t-SNE. *J. Mach. Learn. Res.* **9**, 2579–2605 (2008).
70. H.-M. Zhang, H. Chen, W. Liu, H. Liu, J. Gong, H. Wang, A.-Y. Guo, AnimalTFDB: A comprehensive animal transcription factor database. *Nucleic Acids Res.* **40**, D144–D149 (2012).
71. J. D. Storey, R. Tibshirani, Statistical significance for genomewide studies. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 9440–9445 (2003).
72. G. Csardi, T. Nepusz, The igraph software package for complex network research. *InterJournal* **1695**, 1–9 (2006).

Acknowledgments: We thank E. Sefik and M. Panduro for help with tissue preparations and K. Hattori, G. Buruzula, C. Araneo, K. Waraska, M. Thorsen, and R. Steen for help with mice, profiling, sorting, and sequencing. **Funding:** This work was supported by NIH grants DK092541, AR070334, and AI125603 and funds from the JPB Foundation. J.R.D., D.Z., J.C., and D.R. were supported by fellowships from the American Diabetes Association (1-16-PDF-028), the Boehringer Ingelheim Fonds, the National Research Foundation of the Ministry of Science, Information, Communication and Technology of South Korea, and by a Damon Runyon Fellowship. **Author contributions:** J.R.D., D.Z., D.R., J.C., and R.Z. performed experiments, which A.M.K., C.B., and D.M. supervised. J.R.D., D.Z., C.B., and D.M. designed the study and analyzed and interpreted the data. J.R.D., D.Z., and D.M. wrote the manuscript, which all authors edited. **Competing interests:** D.M. is a co-founder and scientific advisory board member of TRex Bio; A.M.K. is a founder of 1CellBio Inc. **Data and materials availability:** Microarrays [GSM1214187-92 (muscle T_{regs} and control spleen T_{regs}); GSM1660958-60 (colonic T_{regs}) and GSM1660962-64 (control splenic T_{regs}); GSM921176-78 (VAT T_{regs}) and GSM921188-90 (control lymph node T_{regs})], ChIP-seq [GSM1120740 and GSM1120741 (spleen T_{reg} BACH2 ChIP and input); DRR061061 and DRR061062 (spleen T_{reg} H3K27ac)], scRNA-seq (GSE109742), and ATAC-seq (GSE112731).

Submitted 14 March 2018
Accepted 11 July 2018
Published 14 September 2018
10.1126/sciimmunol.aat5861

Citation: J. R. DiSpirito, D. Zemmour, D. Ramanan, J. Cho, R. Zilionis, A. M. Klein, C. Benoist, D. Mathis, Molecular diversification of regulatory T cells in nonlymphoid tissues. *Sci. Immunol.* **3**, eaat5861 (2018).

Science Immunology

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

Sci. Immunol. **3**, eaat5861.
DOI: 10.1126/sciimmunol.aat5861

Mapping T_{reg} 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 (T_{regs}) that reside in visceral adipose tissue, muscle, and the colon and compared them with the profiles generated from splenic T_{regs}. 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 T_{reg} populations.

Science Immunology
AAAS

immunology.sciencemag.org/cgi/content/full/3/27/eaat5861/DC1

Supplementary Materials for

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, Diane Mathis*

*Corresponding author. Email: cbdm@hms.harvard.edu

Published 14 September 2018, *Sci. Immunol.* **3**, eaat5861 (2018)
DOI: 10.1126/sciimmunol.aat5861

The PDF file includes:

Other Supplementary Material for this manuscript includes the following:

(available at immunology.sciencemag.org/cgi/content/full/3/27/eaat5861/DC1)

SUPPLEMENTARY METHODS

Isolation of tissue and splenic Treg populations

Isolation of Tregs

For VAT, epididymal adipose tissue was excised, cut into small pieces and digested at 37^{o}C for 30 min. with 1mg/mL collagenase II (Sigma) diluted in Dulbecco's Modified Eagle's Medium (DMEM), 2% Fetal Bovine Serum (FBS), 1mM EDTA. Cell suspensions were filtered through a sieve, and the stromovascular fraction was harvested by centrifugation at 500g for 10 min. For muscle, the quadriceps, gastrocnemius and tibialis anterior were excised from both hind limbs, cut into small pieces and digested at 37^{o}C for 30 min. with 2mg/mL collagenase II (Gibco) and 150ug/mL DNAse (Millipore Sigma) diluted in DMEM. Cell suspensions were re-suspended in 40% isotonic Percoll (GE Healthcare) underlain with 80% isotonic Percoll, and were centrifuged at 1,000g for 25 min. at room temperature. Cells were isolated from the interphase layer and stained for sorting. For colonic lamina propria, colons were excised and treated with Roswell Park Memorial Institute (RPMI) medium containing 1 mM DTT, 20 mM EDTA and 2% FBS at 37^{o}C for 15 min. to remove epithelial cells, minced and dissociated in collagenase solution (1.5 mg/mL collagenase II (Gibco), 0.5mg/mL dispase and 1% FBS in RPMI) with constant stirring at 37^{o}C for 45 min. Single-cell suspensions were then filtered and washed with DMEM, 2% Fetal FBS, 1mM EDTA.

Cell sorting

For ATAC-seq: Tregs from VAT, muscle, colon and spleen were sorted by flow cytometry using a MoFlo (Beckman-Coulter) as DAPI^{-}CD45^{+}CD8^{-}CD11b^{-}CD11c^{-}TCRβ^{+}CD4^{+}CD25^{+} cells. For microarrays and scRNA-seq: Tregs were sorted from VAT and control spleen as DAPI^{-} TCRβ^{+}CD4^{+}CD25^{+} cells (double-sorted in the case of microarrays). From the muscle, colon and their

respective control spleens, Tregs were sorted as DAPI$^-$ CD4$^+$ TCRβ$^+$ Foxp3/GFP$^+$ from Foxp3-IRES-GFP/B6 mice (*62*). Splenic Tconvs were isolated as DAPI$^-$ CD4$^+$ TCRβ$^+$ Foxp3/GFP$^-$.

Microarray analyses

The microarray data have been published (*5, 7, 9*) and are accessible on Gene Expression Omnibus: GSM1214187-92 (muscle Tregs and control spleen Tregs); GSM1660958-60 (colonic Tregs), GSM1660962-64 (control splenic Tregs); GSM921176-78 (VAT Tregs), GSM921188-90 (control LN Tregs).

Definition of gene-sets

Gene-expression differences between each parenchymal-tissue-Treg population and its splenic Treg control were calculated using the lmFit and eBayes functions of the limma package (*63*), based on all genes with an expression value ≥100 in at least one tissue. Up- or down-regulated genes were defined as FC > 2 or <0.5 and $p < 0.05$ for each tissue compared with the spleen. In total, 1,909 genes were called differentially expressed in at least one tissue (1883 with a false-discovery rate (FDR) < 10%).

The differentially expressed genes were then separated into pan-tissue (up and down), two-tissue (up and down) and tissue-specific (up and down) gene-sets. Differential genes can either be up-, down- or not differentially expressed in each of the three tissues, which creates 27 theoretical gene-sets. 1,887 genes matched to 14 major sets. The remaining 22 genes were assigned to these 14 sets as follows from these definitions. Up and down pan-tissue gene-sets contained up- and down-regulated genes in all three tissues. Up and down two-tissue gene-sets contained up- and down-regulated genes in two tissues (not differential in the remaining one). Tissue-specific up-

genes: up-regulated in one tissue (down-regulated or not differential in the others). Tissue- specific down-genes: down in one tissue (not differential in the others).

PCA

PCA was performed on this gene-list, after log-scaling, using prcomp from the sats package on R (*64*).

GO enrichment

GO enrichment analysis was done using the *GOenrichmentAnalysis* function from the *WGCNA* package on R (*61*). The p-values are calculated using Fisher's exact test with ontologies including at least one of the supplied genes as background. Highlighted in Fig. 1 are those pathways or processes that showed multiple enriched GO terms (e.g. circadian clock: entrainment of circadian clock by photoperiod, photoperiodism, entrainment of circadian clock, regulation of circadian rhythm, etc).

ATAC-seq

Library preparation

Every library was generated from 10 – 20,000 cells, which was determined based on preliminary experiments varying cell number and using the fraction of reads in peaks (FRIP), and total number of peaks, as quality metrics. The overall quality of a library was evaluated by its FRIP, which was 25-40% in preliminary experiments; any library where the FRIP was <25% was excluded from the study. For library generation, we adapted published protocols (*14, 15*). Cells were sorted into 400uL of MACS buffer [1x phosphate-buffered saline (PBS), 0.5% bovine serum albumin (BSA), 2mM EDTA] and pelleted by centrifugation at 500g for 15 min. (All spins used low acceleration and

brake settings, at 4°C.) Cell pellets were re-suspended in 100µl of cold hypotonic lysis buffer [10mM Tris-HCl (pH 7.5), 10mM NaCl, 3mM MgCl2 and 0.1% NP40], followed by immediate centrifugation at 550g for 30 min. The pellet was re-suspended in 5µL of transposition reaction mix [1µL of Tagment DNA Enzyme and 2.5µL of Tagment DNA Buffer from Nextera DNA Sample Prep Kit (Illumina), 1.5µL H_2O], and was incubated at 37°C for 60 min. for DNA to be fragmented and tagged. For library preparation, two sequential 7-cycle polymeras chain reaction (PCR) reactions were performed. After the first PCR, the libraries were selected for small fragments (< 600bp) using SPRI beads (Beckman Coulter) followed by a second round of PCR with the same conditions in order to obtain the final library. Libraries were sequenced on a NextSeq500 (Illumina) to generate paired-end reads (50bp, forward; 34bp, reverse).

Clustering

For hierarchical clustering, the union of OCRs was created using HOMER *mergePeaks*. Read counts in these regions were normalized as reads per million, per library, followed by quantile normalization across all libraries. Normalized read counts were used for complete hierarchical clustering based on Pearson's correlation coefficient as a distance metric (R, v3.2.1). Distributions of functional genomic elements in OCRs were found using HOMER *annotatePeaks.pl*, and these designations were used to filter out TSS regions in subsequent analyses. OCR classes were defined by tissue:splenic Treg fold-changes ≥3 in normalized ATAC-seq read counts. Individual OCRs were assigned to all genes whose TSSs were within the specified distance (10, 100 or 500 Kb, see individual figure legends) from the center of the OCR using BEDTools *window* v2.26.0 (*65*).

Super-enhancers

For super-enhancer delineation, raw H3K27ac ChIP-seq reads from splenic Tregs were obtained from the DNA Data Bank of Japan [(*30*) (DDBJ DRA005202)], processed similarly to the methods used for ATAC-seq, and peaks were called using HOMER *findPeaks*. The ROSE algorithm (*29*) was used for the identification of super-enhancers based on the clustering of ATAC-seq or H3K27ac peaks (TSS excluded) located within 12.5 Kb of one another and ranking peaks by their read densities. For ATAC-seq, ROSE was run on each biological replicate and only SEs found in both were used in further analyses. For clustering of SEs, a union of all SEs was generated using BEDOPS v2.4.14 (*66*). Read densities in these regions were calculated by finding total read counts (HOMER *annotatePeaks.pl -d*), normalizing to reads per million and the length of the SE and quantile normalizing these values across all samples. SEs were then clustered by implementation of the K-means algorithm in R.

TF motifs and target gene prediction

Enriched TF motifs were found in OCRs vs. control, random genomic regions that matched OCRs in size, TSS-proximity and GC content. The significance of a motif's enrichment in target vs. control regions was determined by the hypergeometric test (HOMER *findMotifsGenome.pl*). Enrichment scores are presented as the fold-change of a motif's frequency in target OCRs vs. its frequency in control regions, with only significantly enriched motifs shown ($p < 0.01$). For comparing relative motif enrichments across tissue-Treg subsets, ranked enrichments, rather than absolute p-values, were used to control for the different numbers of tissue-specific OCRs across the Treg subsets (i.e. VAT ~ colon > muscle), which would affect p-value calculations.

ATAC-seq predicted TF:target-gene modules: for each tissue-specific up-regulated gene defined by microarray analysis (Fig. 1), all tissue-specific OCRs ±100Kb of its TSSs were found using the BEDTools *window* function. For each gene, its OCRs were then scanned for the presence of the consensus motif representing each significantly enriched TF family (Fig. 5), using HOMER *findMotifsGenome.pl –find* and consensus motif position weight matrix (PWM) files derived from the HOMER vertebrate TF database (v4.6). A vector of motif scores was then calculated for each gene (0 or 1 marking absence or presence, respectively, of each motif in ≥1 OCR linked to that gene). A gene:motif-score matrix was constructed for each tissue, and Pearson's correlation scores were used for complete hierarchical clustering of genes (R, v3.2.1).

scRNA-seq using inDrops

Library preparation

Library preparation was performed as described in (*17, 18*). 40-80,000 Tregs or Tconvs were first sorted by flow cytometry in complete medium (RPMI 10% FBS). Just before encapsulation, cells were spun for 5 min at 4°C 500g and resuspended in PBS-15% OptiPrep Density Gradient Medium (OptiPrep; Sigma-Aldrich) at a concentration of 80,000 cells/ml. The density gradient prevents sedimentation during encapsulation and allows a uniform flow of cell to enter the microfluidic device. Encapsulation was done by flowing 4 different inputs in the device: cells, reverse transcriptase (RT) buffer, primer hydrogels, and oil (water in oil droplets). Primer hydrogels contained RT primers with a unique single-cell barcode. From 5' to 3', the primer contained a T7 promoter for amplification (in vitro transcription), the PE1 sequencing primer, a unique single-cell barcode (16-19bp) (147,456 possible barcodes), a UMI (6bp) and a polyT sequence. Around 5,000 single cells were encapsulated in droplets of 3-4 nl containing a primer hydrogel bead and the RT buffer

(SuperScriptIII, Invitrogen). Encapsulation was performed under 30 min in order to maintain cell viability and the emulsion was collected in a 500uL tube.

Reverse transcription was performed immediately after encapsulation. Firstly, the primers were released from the gels by exposing the emulsion to UV for 7 min, and then reverse transcription was done at 50°C for 2h followed by enzyme inactivation for 15 min at 70°C. Emulsions were then split into small aliquots to have no more than 3'000 cells per tube, broken down by adding 20% 1H,1H,2H,2H-perfluorooctanol (PFO) in HFE-7500 oil (3M Novec 7500 Engineered Fluid, Novec) and kept at -80°C.

Samples were thawed on ice and spun at 4°C for 5 minutes at 19,000g to pellet cell debris. Excess primers and hydrogel beads were removed by filtration through a nucleic acid purification column and enzyme digestion (ExoI, HinfI). After purification of the DNA:RNA duplex with 1.2X AMPure beads (Beckman), second-strand cDNA synthesis was performed (NEB) for 2.5 h at 16°C. The library was then amplified using T7 *in vitro* transcription for 15h at 37°C (HiScribe T7 High Yield RNA Synthesis Kit, NEB). After purification, half of the amplified RNA was used for further processing. First, the amplified RNA was fragmented for 3 min at 70°C (Magnesium RNA Fragmentation Kit, Ambion) and purified using AMPure Beads (1.2X). Then reverse transcription with random hexamers was done for 1h at 42°C (PrimeScript Reverse Transcriptase, Takara Clontec). A final PCR was done to amplify the library and add the P5-P7 and Illumina index primers (Kapa 2× HiFi HotStart PCR mix, Kapa Biosystems). The number of cycles was chosen by first running a quantitative PCR on 1/20th of the RT reaction. The optimum was between 11 and 14 cycles. Library size was analyzed with a Tapestation (High Sensitivity D1000 ScreenTape, Agilent technologies), quantified by PCR and sequenced using NextSeq 500 and custom primers (read 1: 40bp, index: 7bp, read 2: 51 bp).

Data processing

Data processing was processed as previously described (*17, 18*). Read 1 contains transcript information. Read 2 contains the UMI (6bp), single-cell barcode (16-19bp) and a conserved sequenced named W1. Fastq reads were first filtered for quality (>80% with Sanger Q>20) and on the expected structure of the reads: paired reads were kept only if read 2 contained the W1 and polyT sequence and if read 1 did not contain them (*18*). Single-cell demultiplexing was performed against the possible barcode space and only reads mapping unambiguously and with less than 2 mismatches were kept (*18*). For each single-cell barcode, the reads were mapped against the mouse mm10 transcriptome (complemented with the sequence of the Foxp3Cd90.1 or Foxp3Ires-Ggfp transgenes) using tophat2 (–library-type fr-firststrand) (*59*). Reads mapping to multiple regions or having a low alignment score (MAPQ <10) were filtered out (samtools flag 256). Duplicated reads were filtered as follows using UMIs and custom scripts (*18*): for each set of reads mapping to one gene, we kept all reads having UMIs with a pairwise string distance ≥ 2 (Levenshtein distance for taking into account indels and substitution). A final matrix with genes in rows and cells in columns was then written.

A first quality-control was performed after demultiplexing by calculating the proportion of total reads that mapped to the most abundant single-cell barcodes. Good libraries had >70% of reads mapping to a small set of single-cell barcodes (500-3,000), as expected from successful encapsulation. A second quality-control measure involved the distribution of mapping rates and sequencing coverage. Good libraries were normally distributed with an average 80% mapping rate (5% SD) and more than 90% UMI duplicated reads (indicating sequencing saturation). Single cells with more than 500 genes were kept for the analysis. Total UMI count normalization was performed and scaled to the median UMI count.

Contaminating cells or doublets were identified using the following criteria. Firstly, we compared each single-cell transcriptome to the Immgen dataset (www.immgen.org) and calculated the likelihood for each single cell to be any of the Immgen cell types. The Immgen matrix of gene expression was used to provide prior probabilities (probability to express *gene i* in *cell type j* = p_{ij}), and we calculated for each single cell *C* the likelihood to be of cell type j (L_{cj}) (multinomial model). Log posterior probabilities were derived by normalizing so that they summed to 1 for each single cell.

$$L_{cj} = \sum_i c_i * \log(p_{ij})$$

Contaminant cells were flagged when a T cell type was not amid the top 5 most likely cell types. Secondly, we performed a PCA on the dataset using the most variable genes. Contaminant cells were flagged as they clustered in the first PC. Cells matching a combination of these two criteria were removed from the analysis. Thirdly, we analyzed the distribution of UMIs and genes in each single cell. Single cells with more than 6,000 UMIs were also filtered out (possible doublets).

scRNA-seq using CEL-seq

Single-cell RNA sequencing libraries were generated with a modified CEL-seq protocol (*41*) for studying the differences between muscle Tregs at day 1 and day 4 after injury because too few cells can be isolated at day 1 to perform InDrop. The protocol is described at length in (*67*). Briefly, single muscle Tregs at day 1 or 4 after injury and control splenic Tregs were index-sorted via flow cytometry into a 96-well plate. Each well of this plate contained a lysis buffer with a uniquely barcoded reverse-transcription primer, RNAseOUT (Invitrogen) and water. Each primer contained a T7 promoter, the 5′ TruSeq Illumina adaptor, a unique molecular barcode (4–9 bp), a single-cell DNA barcode (8–16 bp) and an oligo(dT) sequence (24 bp). After mechanical shaking, plates were frozen in carbonic ice and kept at -80°C. When ready to make libraries, RNA was denaturated by incubation of the plates for 3 min at 70 °C. Reverse transcription was performed in each well independently, first

strand with ArrayScript Reverse Transcriptase (Ambion) and second-strand with the mRNA Second Strand Synthesis Module (NEBNext). Each well then contained a single-cell cDNA library with a unique barcode. The libraries containing different barcodes were pooled to perform *in vitro* transcription (MEGAshortscript T7 transcription kit, Ambion). After fragmentation, the amplified RNA was 3' ligated to an Illumina adapter. A reverse transcription and PCR was then performed to amplify the library for sequencing. Data processing was done similarly to that for the InDrop libraries.

scRNA-seq data analyses

Differential gene expression was performed using the DESeq statistical model (*68*). The distribution of counts C_{ij} for each gene i in group j is modeled by a negative binomial of parameter μ_{ij}(mean) and dispersion

$$C_{ij} = NB\left(\mu_{ij}, \theta_i\right)$$

First, the dispersion was estimated using the background model according to (*69*) in which a linear model is fitted between the $\log_2(\mu_i)$ and $\log_2(\sigma_i^2)$ for each gene *i*. For each gene *i*, the dispersion dispersion $\theta_i = \frac{\mu_i^2}{\max\left(\mu_i + 10^{-6}, s_{\mu_i}^2\right)} - \mu_i$, was estimated, $s_{\mu_i}^2$ being the fitted variance of gene *i*. A generalized linear model was then fit for each gene *i* in each group *j* [glm(formula = gene ~ groups, data = tmp, family = negative.binomial(θ_i))], and we used an analysis of variance (anova) to compare the coefficients in each model and obtain a p-value. FDRs were calculated to correct for multiple testing.

Data were visualized using the t-Distributed Stochastic Neighbor Embedding (t-SNE) dimensionality reduction algorithm (*70*). A PCA was first performed on the top 100 most variable genes that were expressed in more than 1% of the cells. The Fano factor (variance/mean) was used as a metric for variability because of its independence from the mean in Poisson distributed data.

The number of significant PCs was determined by comparison with PCA over a randomized matrix as described by Klein et al. (*16*). Two-dimensional tSNE was run using the Rtsne function (*69*), on the significant PCs, setting the seed for reproducibility and using the following parameters (perplexity = 50, max_iter = 1000). Clustering was done using k-means. Density of cells expressing a specific gene was plotted using dropout-corrected counts (see below) and the densCols function from the grDevice package on R (*64*). Signature scores for each single cell (Fig. 5) were computed by summing the counts for the up-regulated genes and subtracting the counts for the down-regulated genes. Scores were Z-score normalized in the plots.

Tissue-Treg network construction and motif enrichment

A network was built around the ATAC-seq modules described above. A vector of motif scores was calculated for each pan-tissue and tissue-specific gene (number of TF family motifs present in the OCRs linked to that gene). We refined these scores using scRNA-seq, which allowed i) identification of the specific TF family member(s) expressed in each tissue; ii) an assessment of whether the presence of the TF motif in a target gene was functional by evaluating the covariation between the TF and the target gene across thousands of single tissue-Treg cells. The list of members in each TF family was obtained from (*71*). We first identified the specific TF members differentially expressed in VAT, colon or muscle in comparison with its splenic counterparts (FDR < 5%). We used this information to replace the TF family with the expressed TF members in the ATAC gene:motif-score matrices of the pan-tissue and tissue-specific gene-sets, and then calculated the Spearman correlation of each TF with its putative pan-tissue or tissue-specific target-genes (in which a motif for this TF was found). We measured the significance of the correlations with the corPvalue function from the WGCNA package in R, which calculates the Student's correlation p values effectively taking into account the number of observations (*61*). We corrected for multiple testing by

calculating the false discovery rate with the qvalue package in R (FDR) (*72*). We also assessed significance of the correlations with null distributions obtained by randomization of the single cell data and multiple testing corrections with FDR. If the TF:target correlation was significant (FDR < 2.5% with both methods) (*61*), we added to the TF:target-gene matrix a weight of 1 for a positive correlation or -1 for a negative correlation, or added a 0 for no significant correlation. We calculated these TF:target-gene matrices for the pan-tissue and tissue-specific gene-sets and used them to build a graph using the igraph package in R (*73*). In this graph, each vertex was a gene, and a directed edge was drawn between a TF and its target gene. Vertices were placed on the plane using the force-directed layout algorithm by Fruchterman and Reingold (*73*). Robustness of the networks was assessed by randomization of the single cell data. We shuffled our data by randomly redistributing read counts per gene with the sample function in R per row of the data matrix. We permuted the data and ran the above algorithm 1,000 times, storing at each iteration the number of links in the network. This was used as null distribution of the number of connections in the network. Network motifs are small connected subgraphs with specific patterns that are significantly enriched in comparison with random networks. We found graph motifs of size 3 in the VAT-specific network by using the function graph.motifs from the igraph package. Degree preserving randomizations (n=1000) of the tissue network were employed to calculate the null distribution and to get a p-value for motif enrichment in the tissue network (using the VertexSort package in R). We found instances of enriched motifs in the VAT-specific network using the subgraph_isomorphisms from the igraph package.

ChIP-seq

For mapping BACH2 ChIP-seq peaks to genes, tissue-Treg gene sets defined in Figure 1B were used as follows: “pan-tissue” represents genes up-regulated in all three non-lymphoid tissue Treg

subsets, and “tissue-specific” combines genes uniquely up-regulated in VAT, muscle, or colon. A control gene set was generated by randomly sampling genes whose mRNA was below the limit of detection in all Tregs in this study (both lymphoid and non-lymphoid tissue). The average profiles of BACH2 ChIP-seq signal ± 2.5Kb of the TSS of genes in each set were generated using NGSplot (Shen et al., 2014). For determining the percentage of genes occupied by BACH2, TSS coordinates of genes in each set were extended by 10Kb in both 5’ and 3’ directions and intersected with BACH2 ChIP-seq peaks using BEDtools window function.

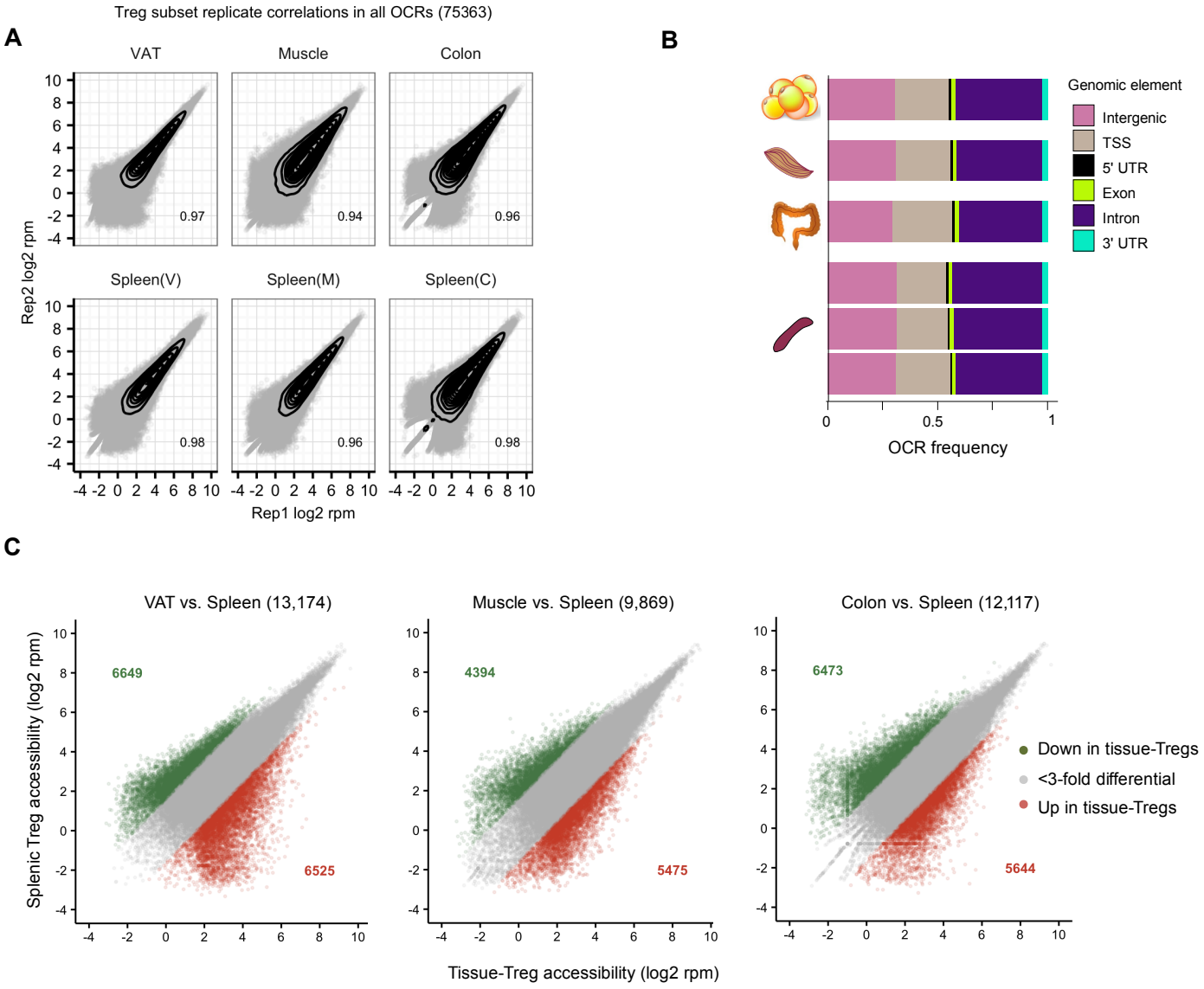
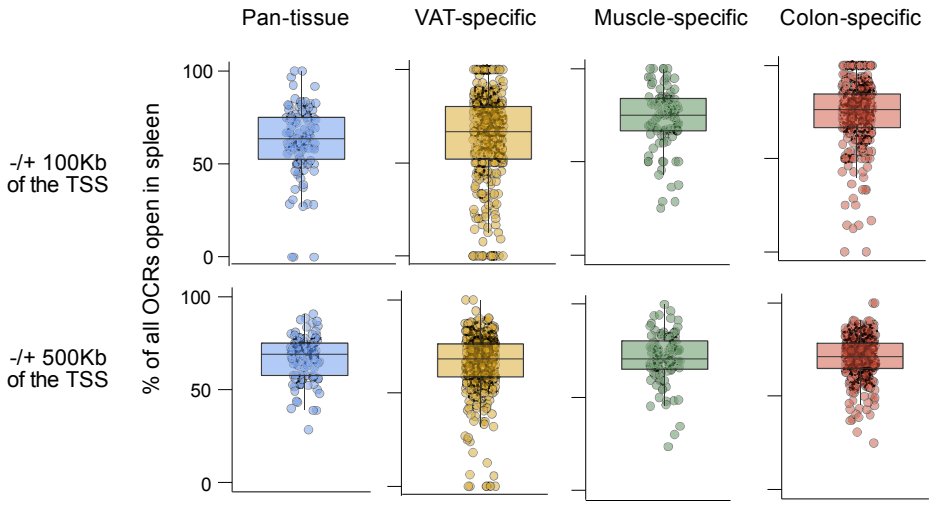


Figure S1: ATAC-seq general features.

(A) Comparison of biological replicate ATAC-seq signals (normalized reads per million mapped reads) for the OCRs found in the various Treg populations. Each plot is annotated with the Pearson *r* value. (B) Distribution of functional genomic elements within each Treg population's OCRs. (C) The OCRs with differential accessibility between each designated tissue:spleen Treg pair are highlighted (≥3-fold difference in normalized read counts). The numbers of gained (red) or lost (green) OCRs are indicated.

A



B

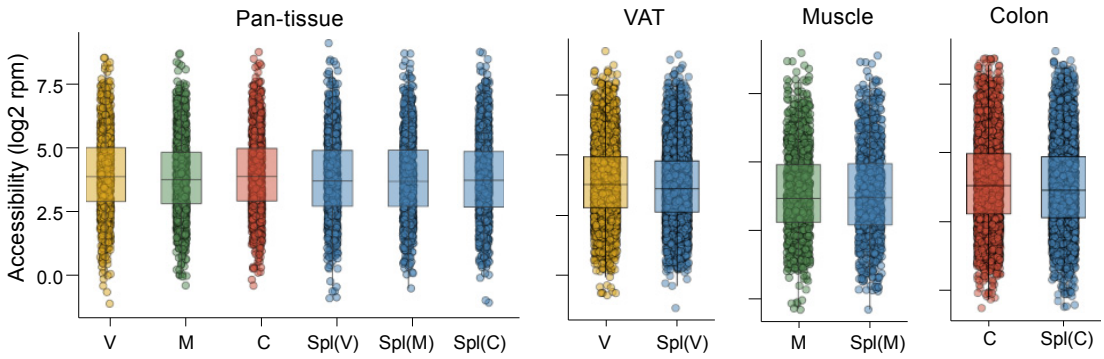


Figure S2: Priming of OCRs at tissue-restricted genes in splenic T_{regs} also occurs at TSS-distal regions and is not composed of weak OCRs.

(A) As per Fig. 3B, except OCRs +/- 100 (upper) or 500 (lower) kb from the TSS were quantified. (B) Accessibility of primed OCRs across Treg subsets. For tissue-restricted genes in each of the four gene sets, all nearby OCRs (±100Kb of the TSS) present in both tissue and spleen were identified. Boxplots show the distributions of their accessibility signals within each Treg subset, indicating that these OCRs are as strong in splenic Tregs as they are in tissue-Tregs, arguing against their likelihood of originating from a minor contaminating population of tissue-Tregs that had re-circulated back to the spleen.

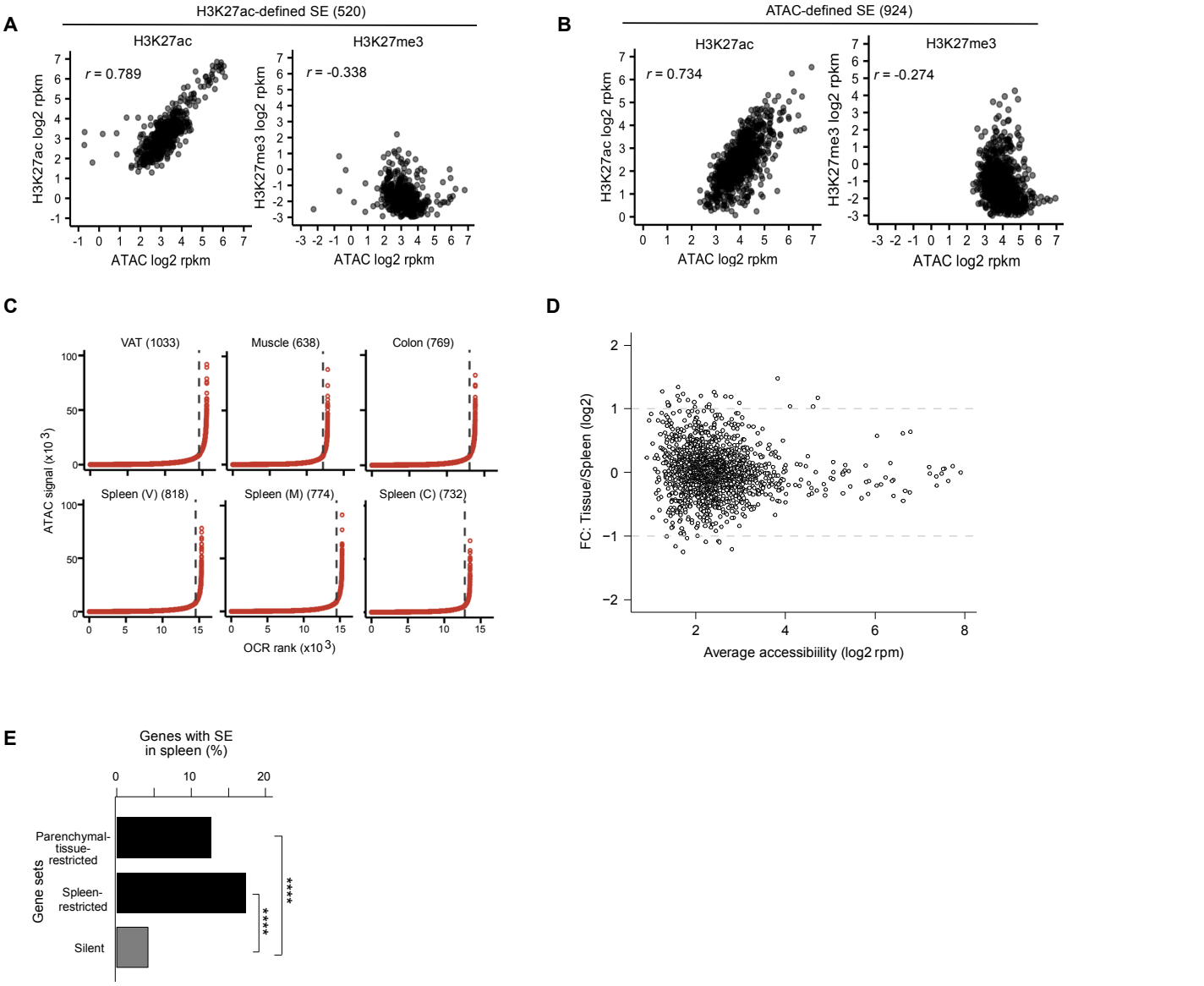


Fig. S3. Generation and analysis of ATAC-seq–derived SEs across T_{reg} subsets.

(A) Applying the ROSE algorithm (see Methods) to H3K27ac ChIP-seq data yielded 520 SEs in splenic Tregs. Scatter-plots show pair-wise comparisons of the read density in these SEs between H3K27ac and ATAC-seq or between H3K27me3 (a mark not enriched at SE) and ATAC-seq. Raw ChIP-seq data were from the DNA Data Bank of Japan DRA003955 *(30)* and analyzed for this study. (B) Applying the ROSE algorithm to ATAC-seq data yielded 924 SE in splenic Tregs. Scatter-plots show pair-wise comparisons of the read density in these SEs between H3K27ac and ATAC-seq or H3K27me3 and ATAC-seq. All plots are annotated with the Pearson's *r* value. (C) Identification of SEs. SEs were delineated using the ROSE algorithm, by stitching together OCRs (≤ 12Kb apart) and ranking the stitched regions by their ATAC signal. Only the SEs observed in both replicates of the different tissues are shown. (D) MvA plot of SEs. Fold-change between the average accessibility of each SE in tissue-Tregs vs. its average accessibility in splenic Tregs, plotted against the mean SE signal across all samples. (E) As in Figure 3D, except H3K27ac-defined SEs were used. The association between H3K27ac-defined super-enhancers in splenic Tregs with genes whose expression is either spleen- or tissue-restricted, compared with the background association of SEs with silent genes. Binomal test **** p<0.0001.

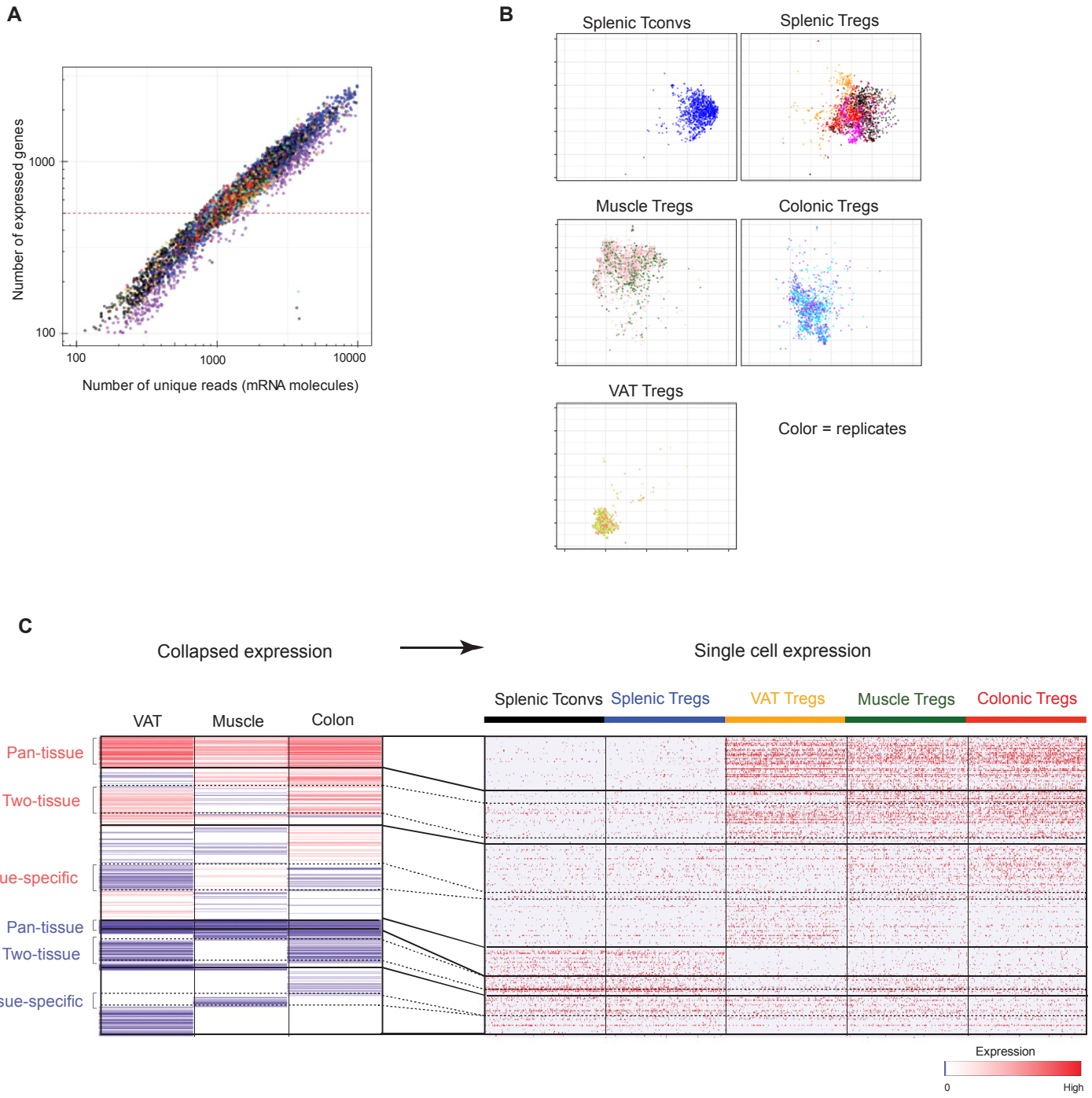


Fig. S4. Tissue T_{reg} scRNA-seq data set validation.

(A) Single-cell transcriptome coverage. Scatterplot showing the number of uniquely identified mRNAs sequenced along with the number of expressed genes per single cell, in all datasets (color coded). Single cells with less than 500 transcripts sequenced were excluded from the analysis (below the horizontal dashed line). (B) Same tSNE plot as in Fig. 5A split into individual panels for each tissue and different colors for each replicate. n = 2 for the VAT, muscle and colonic Tregs, n = 1 for Tconvs, n = 4 for splenic Tregs (1 replicate control performed for each tissue Treg and Tconv cell data-set). (C) Pan-tissue, two-tissue and tissue-specific gene-set expression in the tissue Treg scRNA-seq dataset. Left: Heatmap representing gene-expression fold-changes (log2) for each tissue (VAT, muscle, colon) compared with the paired spleen (using collapsed average expression from the scRNA-seq). Right: Biclustering heatmap deconvoluting the expression of each gene-set per single cell (normalized counts). Genes ordered per gene-set. Cells ordered per tissue.

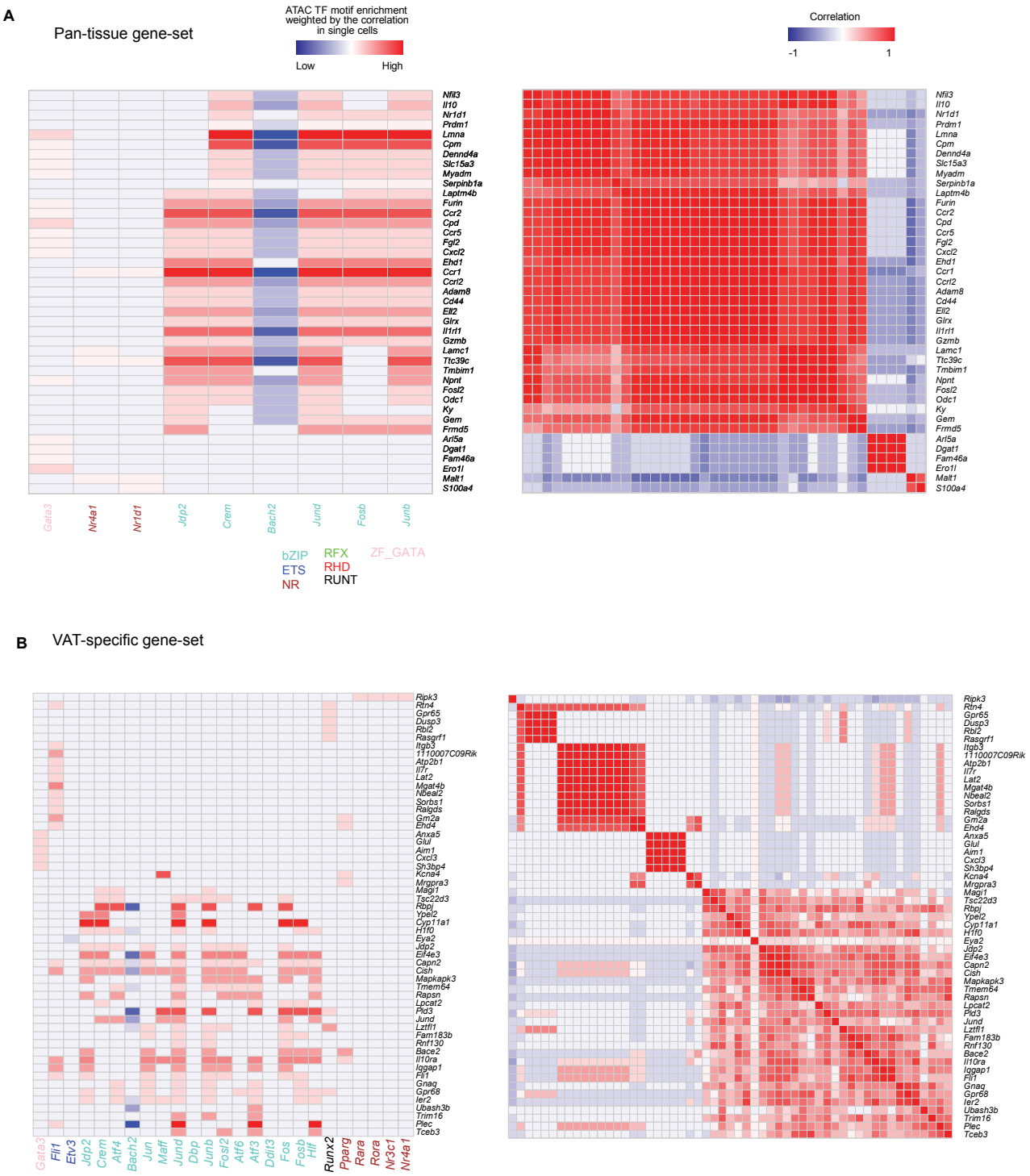
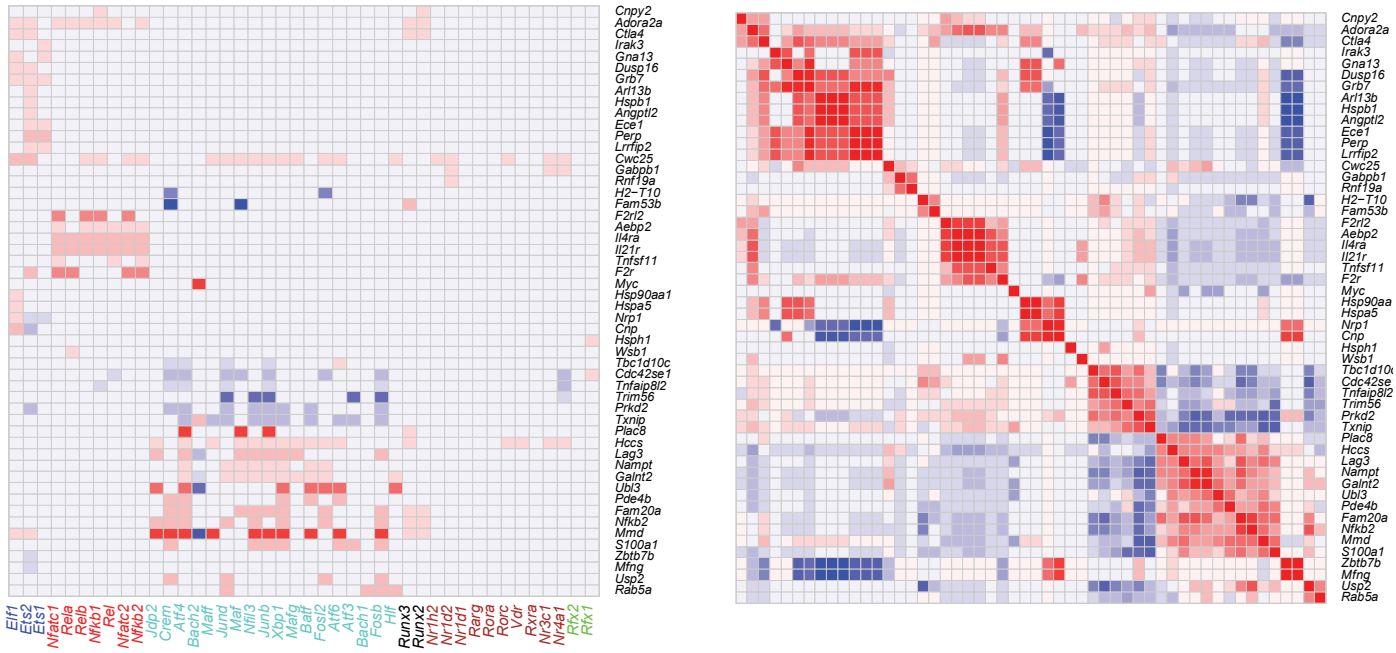


Fig. S5. Pan-tissue and tissue-specific T_{reg} modules identified by combining ATAC-seq and scRNA-seq data.

(A) Pan-tissue modules. Left: Heatmap representing the TF-family motif enrichment weighted by the correlation in the scRNA-seq data-set between each TF and the pan-tissue gene-set. In columns are individual TF family members (color-coded) expressed in all non-lymphoid tissues and enriched in pan-tissue genes. Pan-tissue genes are in rows. Right: Gene-gene correlation heatmap of pan-tissue genes based on the heatmap at the left. (B) VAT-specific modules, (C) Colon-specific modules, all plotted as per A. (D) Results of 1,000 random permutations of the scRNA-seq data for the VAT-specific module, presented as the number of significant connections in each iteration. This serves as a null distribution of the number of edges expected by chance in the network in comparison with the real network (red arrow).

C

Colon-specific gene-set



D

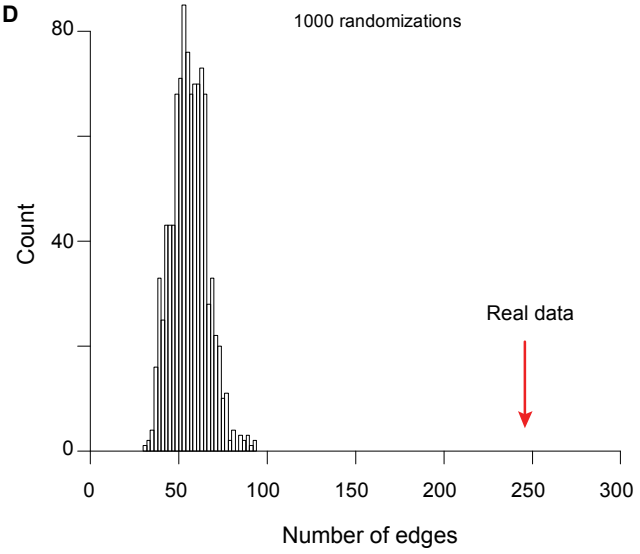


Fig. S5. Pan-tissue and tissue-specific T_{reg} modules identified by combining ATAC-seq and scRNA-seq data.