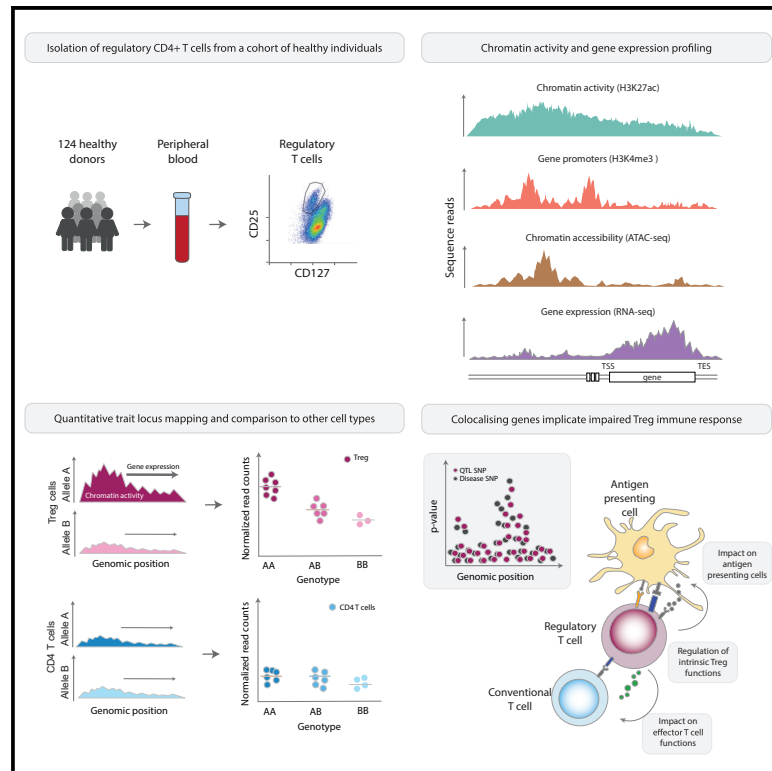


## Immune disease variants modulate gene expression in regulatory CD4<sup>+</sup> T cells

### Graphical abstract



### Authors

Lara Bossini-Castillo, Dafni A. Glinos, Natalia Kunowska, ..., Ian Dunham, David J. Roberts, Gosia Trynka

### Correspondence

gosia@sanger.ac.uk (G.T.), dafni.glinos@gmail.com (D.A.G.)

### In brief

Regulatory T cells (Treg cells) play an essential role in maintaining appropriate immune responses, but their low frequency in circulating blood has resulted in a limited number of available genomic resources. Here, Bossini-Castillo, Glinos, et al. provide a detailed map of gene expression regulation and chromatin activity in Treg cells isolated from 124 healthy individuals.

### Highlights

- A dataset of gene expression and chromatin activity in CD4<sup>+</sup> Treg cells from 124 individuals
- Map of quantitative trait loci (QTLs) for 3,685 genes and 7,195 chromatin regions
- Colocalization of Treg cell QTLs and immune disease GWAS variants to nominate drug targets



## Resource

# Immune disease variants modulate gene expression in regulatory CD4<sup>+</sup> T cells

Lara Bossini-Castillo,<sup>1,8</sup> Dafni A. Glinos,<sup>1,2,8,\*</sup> Natalia Kunowska,<sup>1</sup> Gosia Golda,<sup>1</sup> Abigail A. Lamikanra,<sup>3,4</sup> Michaela Spitzer,<sup>5,6</sup> Blagoje Soskic,<sup>1,6</sup> Eddie Cano-Gamez,<sup>1,6</sup> Deborah J. Smyth,<sup>1,6</sup> Claire Cattermole,<sup>1</sup> Kaur Alasoo,<sup>1,7</sup> Alice Mann,<sup>1</sup> Kousik Kundu,<sup>1</sup> Anna Lorenc,<sup>1</sup> Nicole Soranzo,<sup>1</sup> Ian Dunham,<sup>5,6</sup> David J. Roberts,<sup>3,4</sup> and Gosia Trynka<sup>1,6,9,\*</sup>

<sup>1</sup>Wellcome Sanger Institute, Wellcome Genome Campus, Cambridge, UK

<sup>2</sup>New York Genome Center, New York, NY, USA

<sup>3</sup>NHS Blood and Transplant, Oxford, UK

<sup>4</sup>BRC Haematology Theme, Radcliffe Department of Medicine, University of Oxford, Oxford, UK

<sup>5</sup>European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Genome Campus, Hinxton, Cambridge, UK

<sup>6</sup>Open Targets, Wellcome Genome Campus, Cambridge, UK

<sup>7</sup>Institute of Computer Science, University of Tartu, Tartu, Estonia

<sup>8</sup>These authors contributed equally

<sup>9</sup>Lead contact

\*Correspondence: [gosia@sanger.ac.uk](mailto:gosia@sanger.ac.uk) (G.T.), [dafni.glinos@gmail.com](mailto:dafni.glinos@gmail.com) (D.A.G.)

<https://doi.org/10.1016/j.xgen.2022.100117>

## SUMMARY

Identifying cellular functions dysregulated by disease-associated variants could implicate novel pathways for drug targeting or modulation in cell therapies. However, follow-up studies can be challenging if disease-relevant cell types are difficult to sample. Variants associated with immune diseases point toward the role of CD4<sup>+</sup> regulatory T cells (Treg cells). We mapped genetic regulation (quantitative trait loci [QTL]) of gene expression and chromatin activity in Treg cells, and we identified 133 colocalizing loci with immune disease variants. Colocalizations of immune disease genome-wide association study (GWAS) variants with expression QTLs (eQTLs) controlling the expression of *CD28* and *STAT5A*, involved in Treg cell activation and interleukin-2 (IL-2) signaling, support the contribution of Treg cells to the pathobiology of immune diseases. Finally, we identified seven known drug targets suitable for drug repurposing and suggested 63 targets with drug tractability evidence among the GWAS signals that colocalized with Treg cell QTLs. Our study is the first in-depth characterization of immune disease variant effects on Treg cell gene expression modulation and dysregulation of Treg cell function.

## INTRODUCTION

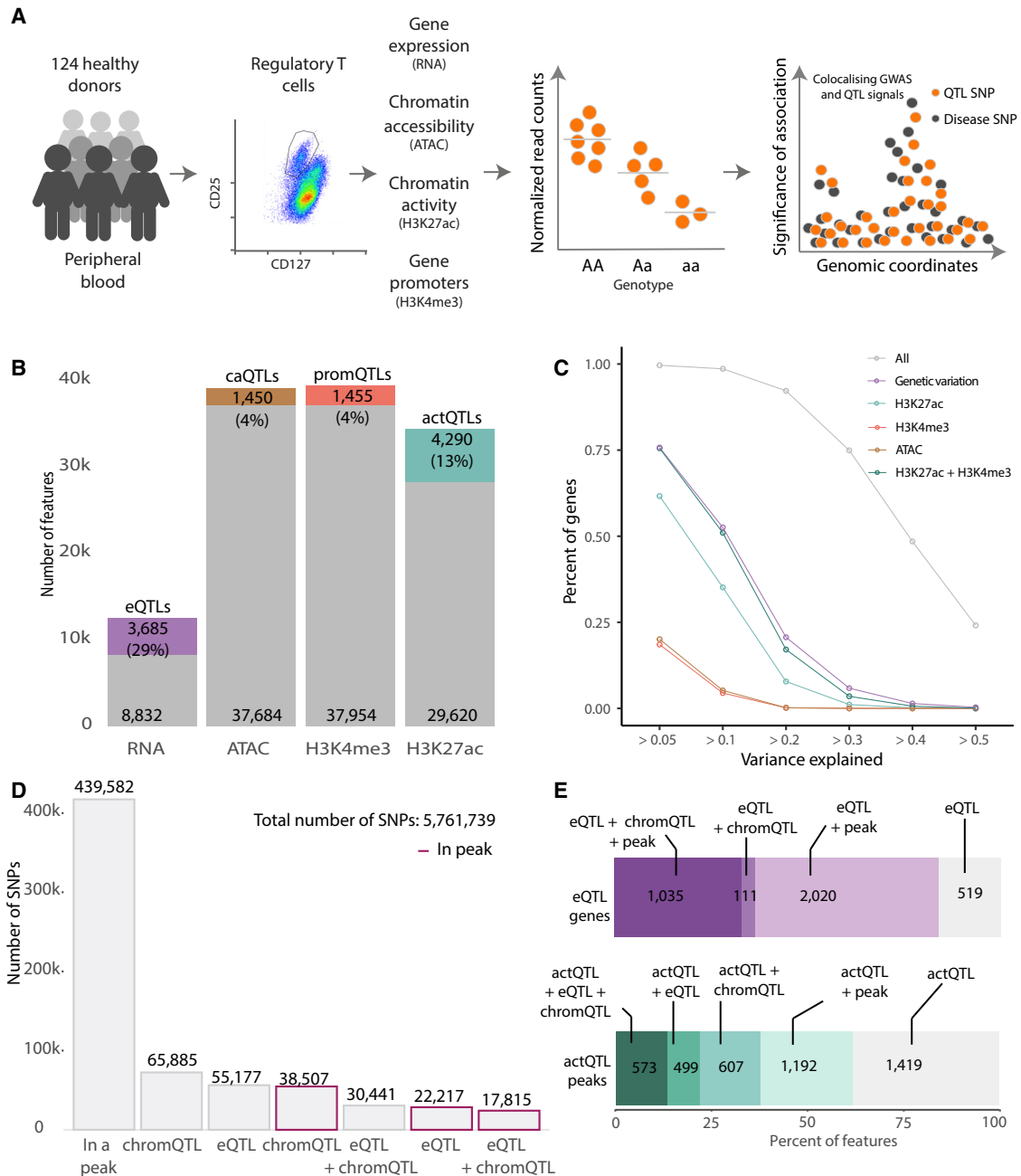
Thousands of disease variants mapped through genome-wide association studies (GWASs) provide genetic anchors to disease biology, but functional interpretation of GWAS signals has been challenging, as the vast majority of variants are non-coding. One approach for linking genetic variation to downstream effects includes expression quantitative trait locus (eQTL) mapping, in which transcript levels are correlated with genetic polymorphisms.<sup>1</sup> However, due to the linkage disequilibrium (LD) between genetic variants, the identified eQTLs often result in associations of tens to hundreds of correlated variants with gene expression levels and therefore fail to nominate the causal regulatory variants.

Prioritization of the exact regulatory variants underlying gene expression changes can be further inferred through QTL mapping of chromatin activity using chromatin accessibility or histone modifications (chromatin QTLs [chromQTLs]). In this approach, variants that modulate activity levels of chromatin marks can be physically overlapped with the chromQTL

features.<sup>2</sup> The combination of eQTLs and chromQTLs provides a powerful toolkit for linking non-coding variants to genes whose expression is modulated, for prioritizing functional variants, and for identifying mechanisms through which gene expression is regulated. Finally, colocalization<sup>3</sup> of disease GWAS signals with such QTLs can point toward causal genes and mechanisms underlying disease associations, therefore linking disease-associated variants to dysregulated pathways and new drug targets.

GWAS variants associated with common immune-mediated diseases, such as inflammatory bowel disease (IBD), type 1 diabetes (T1D), and rheumatoid arthritis (RA), are enriched in active chromatin marks that tag enhancers and promoters in the CD4<sup>+</sup> T cells, especially in regulatory T cells (Treg cells).<sup>4–6</sup> Treg cells are an infrequent yet functionally significant subset of CD4<sup>+</sup> T cells; they comprise 2%–10% of CD4<sup>+</sup> T cells and play an essential homeostatic role in the immune system by suppressing the proliferation and effector functions of conventional T cells. Immunophenotyping studies have shown that abnormal numbers of circulating Treg cells<sup>7,8</sup> and defective suppressive function of Treg cells result in a dysregulated





**Figure 1. Overview of mapped Treg cell QTLs**

(A) A schematic of our study design.

(B) Number of features defined per genomic assay and number of significant QTLs in each category.

(C) Proportion of eQTL gene expression variance explained by genetic variation and chromatin marks. We considered *cis*-regulatory elements in a  $\pm 150$ -kb window from the gene. Shown is the cumulative contribution of genes with increasing proportions of explained variance.

(D) Functional classification of tested genetic variants. Bars with purple outline indicate instances when a QTL variant maps to any chromatin peak. Categories are mutually exclusive.

(E) Classification of eQTL genes (top) and actQTL peaks (bottom). eQTL genes were classified based on the annotation of eQTL variants with chromQTLs and overlap with chromatin peaks. eQTL + chromQTL + peak, number of eQTL genes for which eQTL variants also result in a chromQTL and one of the eQTL variants mapped within a chromatin mark peak; eQTL + chromQTL, number of eQTL genes for which eQTL variants also result in a chromQTL but no variant mapped within any chromatin mark peak; eQTL + peak, number of eQTL genes for which eQTL variants map within a chromatin mark peak but no chromQTL effects were detected; eQTL, number of eQTL genes for which we were unable to map eQTL variants to chromatin mark peaks or to link them to chromQTLs. actQTL peaks

(legend continued on next page)

immune response in patients with immune diseases,<sup>9–11</sup> as well as in organ and hematopoietic stem cell transplant recipients.<sup>12,13</sup> Taken together, the genetic anchor to dysregulation of gene expression in Treg cells and immunophenotyping studies pointing toward impaired function of this cell type strongly suggest that identifying mechanisms through which genetic variants modulate Treg cell function could have important clinical implications. In addition, *ex vivo* approaches to expand Treg cell numbers and to enhance Treg cell suppressive capacity and reinforce them into patients have been successful in clinical trials for T1D<sup>14–16</sup> and Crohn's disease.<sup>17</sup>

Despite the key role of Treg cells in maintaining appropriate immune responses, their low frequency in circulating blood has resulted in a limited number of available genomic resources.<sup>18–20</sup> Consequently, immune disease variants are often interpreted in light of gene expression data from peripheral blood mononuclear cells (PBMCs), immune cell lines, or isolated major immune cell populations.<sup>2,21–26</sup> However, these datasets can either dilute or omit gene regulatory effects only present in rare cell types, therefore potentially missing biological effects meaningful to the disease.

Here, to interpret immune disease variants in the context of a cell type strongly relevant to disease biology, we generated the first detailed map of gene expression regulation in Treg cells isolated from 124 healthy individuals. We identified a total of 10,880 QTL effects (3,685 eQTLs and 7,195 chromQTLs). In comparison to closely related naive CD4 T cells as well as monocytes,<sup>24</sup> we observed 21% of the eQTLs and 29% of the active enhancer and promoter QTLs were detected only in Treg cells. By colocalizing Treg QTLs with variants associated with 14 different immune diseases, we identified 133 GWAS loci with functional relevance in Treg cells. The overlap of immune disease GWAS signals with chromQTLs functionally refined associated variants at 68 immune disease loci. We assigned Treg cell eQTL genes to 81 immune disease loci. At 52 loci, we detected colocalizations with chromQTLs that we were unable to link to downstream gene targets, indicating that the gene regulatory effects could manifest in a cell-state-specific context. Finally, we used the prioritized genes to identify drugs for repurposing and to define novel targets for validation. Our study provides a translational pathway from immune-disease-associated variants, through gene expression regulation in Treg cells, to new treatment options.

## RESULTS

### Comprehensive catalog of gene expression regulation in Treg cells

To identify genetic variants that control gene expression regulation in Treg cells isolated from healthy blood donors (Figure S1; Table S1), we profiled the transcriptome using RNA sequencing (RNA-seq) (124 individuals), chromatin accessibility

using assay for transposase-accessible chromatin using sequencing (ATAC-seq) (73 individuals), promoters using H3K4me3 (88 individuals), and active enhancer and promoter regions using H3K27ac (91 individuals; Figure 1A). We detected the expression of 12,517 genes, while chromatin profiling revealed 39,134 accessible regions, 39,409 H3K4me3 marked promoter regions, and 33,910 H3K27ac marked active chromatin regions (Figures 1B and S2). The majority of the mapped regulatory chromatin features overlapped with each other (Figure S2A). Concordant with previous studies,<sup>24,27,28</sup> we observed that H3K4me3 and chromatin accessible regions were concentrated near the transcription start sites (TSSs), while H3K27ac marked more distal gene-regulatory elements (Figure S2B). Concordantly with previously described DNA acetylation patterns, the H3K27ac peaks were wider than H3K4me3 and ATAC peaks (Figure S2C). Moreover, the fraction of reads in peaks and the correlation between replicates confirmed the quality of the defined features (Figures S2D and S2E).

Using the 62 samples for which we had complete information, including genetic variation, chromatin profiles, and whole transcriptome, we estimated the percentage of gene expression variability explained by the genetic component and by the chromatin regulatory features. We observed that the major component driving transcriptional variability was the common genetic variation; for 75% of the eQTL genes, we were able to explain 5% or more of the expression variance (Figure 1C). With the addition of the combination of chromatin marks (H3K27ac, H3K4me3, and ATAC) and the common genetic variation, we were able to explain 5% or more of the gene expression variance for all the eQTL genes. This additional gene expression variability was mainly accounted for by the combination of H3K27ac and H3K4me3. These results were in line with previous reports for other primary immune cells.<sup>24</sup> Together, the gene expression variance decomposition analysis implicated that genetic variation contributed the most toward gene expression regulation and the genetic regulation was present at both the transcriptome and the chromatin mark levels. Therefore, by connecting genetic variation to gene expression and chromatin regulatory features, we expected our Treg cell dataset to provide translational insight into immune disease GWAS loci.

Next, we performed QTL mapping to define genes and chromatin features that were under genetic control in Treg cells (STAR Methods). We detected at least one independent association for 3,685 genes (29%) and a total of 125,650 eQTL variants (eQTLs) (Figures 1B–1D; Table S2). We mapped a total of 7,195 chromQTLs, using chromatin accessibility (caQTLs, 1,450; 4%), H3K4me3 (promQTLs, 1,455; 4%), and H3K27ac (actQTLs, 4,290; 13%) histone marks, which corresponded to 9,292 non-overlapping peak regions, associated with 152,648 chromQTL variants. The majority of chromQTLs were detected in H3K27ac features (4,290 actQTLs; Figure 1B). Of all analyzed

were classified based on the annotation of actQTL variants with eQTLs and overlap with chromatin peaks. actQTL + eQTL + chromQTL, number of actQTLs that also result in an eQTL and an additional chromQTL; actQTL + eQTL, number of actQTLs that also result in an eQTL; actQTL + chromQTL, number of actQTLs that also result in an additional chromQTL; actQTL + peak, number of actQTLs that map within a chromatin mark peak without an additional chromQTL or eQTL; actQTL, number of actQTLs that we were unable to map variants to chromatin mark peaks or to link them to an additional chromQTL or eQTL.



genetic variants (5,761,739), 439,582 (7%) fall within a peak; however, only a small fraction mapped in a chromatin feature and were also linked to chromQTLs (38,507 SNPs; 0.7%), eQTLs (22,217 SNPs; 0.4%), or both (17,815 SNPs; 0.3%; [Figure 1D](#)). For 28% (1,035) of all eQTL genes, we observed that at least one eQTL variant was a chromQTL and was also physically located in a chromatin peak ([Figure 1E](#)), and for an additional 2,020 eQTL genes, we were able to link an eQTL variant to a chromatin peak, though without detecting a QTL effect on a chromatin feature. A proportion of this overlap may not be functional, as chromatin regulatory features are abundant throughout the genome and therefore likely to overlap common genetic variants by chance. Interestingly, we were unable to link the majority of actQTL variants to an eQTL ([Figure 1E](#)), implicating that these regulatory regions may modulate gene expression under a specific cellular context or through the interplay of multiple regulatory elements.

### Defining gene expression regulation in Treg cells

We sought to identify QTL effects at the levels of gene expression and chromatin regulation specific to our Treg cell dataset and absent from other immune cells assayed in publicly available data. However, we recognize that such a comparison can suffer from confounders introduced by technical biases, such as differences in sample processing. Therefore, we used transcriptomics data from 91 individuals sampled by the Database of Immune Cell eQTLs Expression Epigenomics (DICE) consortium,<sup>18</sup> where different immune cell types were assayed from the same donors ([STAR Methods](#)). We retrieved data for naive T cells and memory Treg cells to directly estimate the proportion of replicable eQTL effects with our data. As a comparison, we included classical monocytes, as we expected the degree of sharing to be lower compared with Treg cells. We used pairwise  $\pi_1$  score,<sup>29</sup> which estimates the proportion of true positive associations replicating between discovery and replication cohorts. Indeed, we observed that the eQTLs detected in our Treg cell cohort replicated highly in memory Treg cells ( $\pi_1 = 0.85$ ) and naive T cells ( $\pi_1 = 0.84$ ) in the DICE data, while the sharing was lower in monocytes ( $\pi_1 = 0.71$ ; [Figures S3A](#) and [S3B](#)). The eQTLs detected in the DICE Treg cell cohort replicated more highly in our Treg cell dataset and also replicated more highly compared with DICE naive T cells, meaning that we replicated the majority of DICE eQTLs, which is likely due to differences in sample size and the average sequencing depth being greater in our dataset ([Figure S3B](#)).

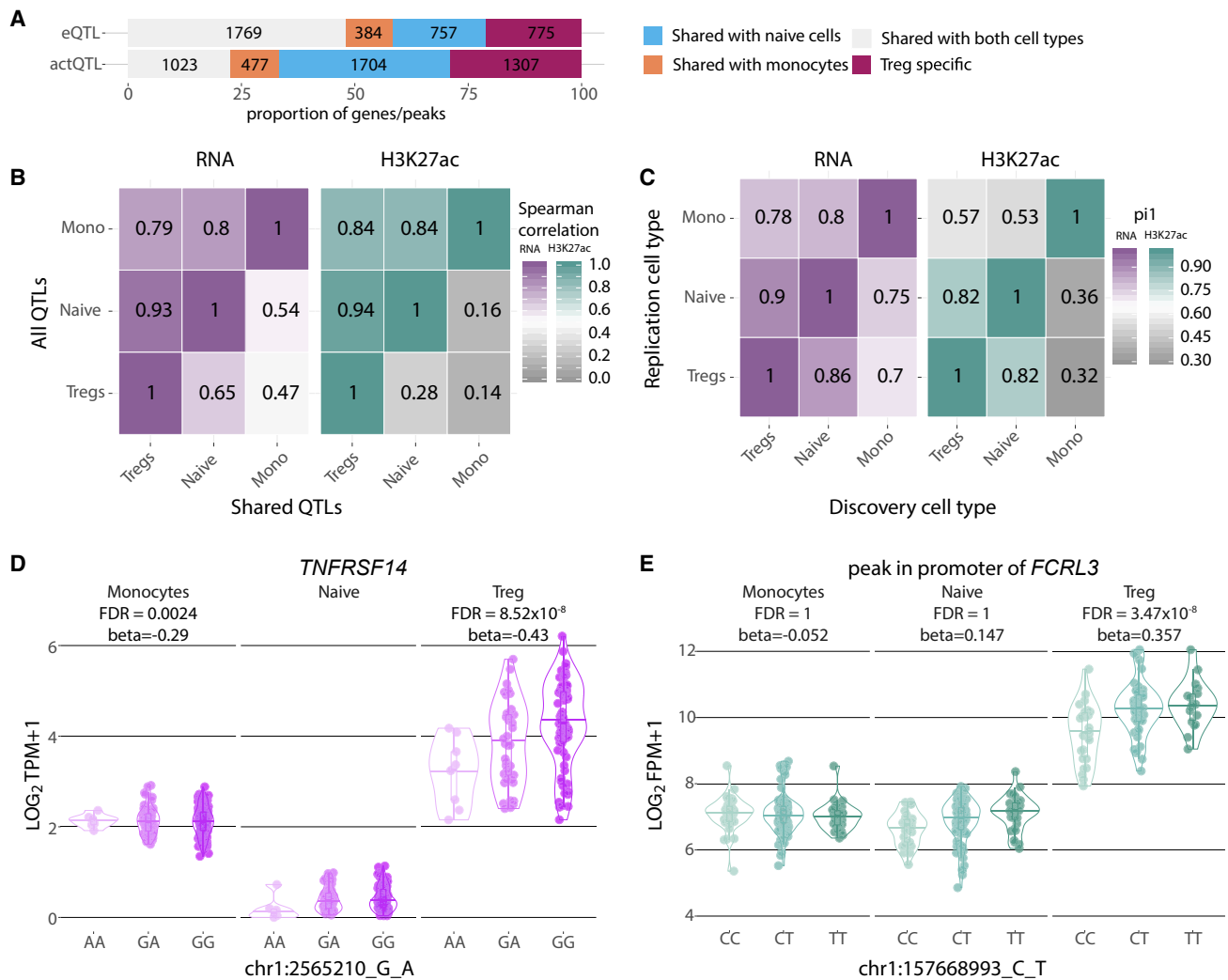
Having confirmed that our dataset was capturing effects relevant to Treg cell biology, we next used the CD4 naive T cells from BLUEPRINT project,<sup>24</sup> as it profiled both the transcriptome and H3K27ac assayed across a similar cohort to ours (197 British healthy individuals). Again, we included monocytes, as we expected lower sharing compared with naive CD4<sup>+</sup> T cells. We observed that the majority of eQTLs (69%; [Figure 2A](#); [Table S2](#)) were shared with naive CD4 T cells, with similar effect sizes and the same direction of effects ([Figures S4A](#) and [S4B](#)). A higher correlation between eQTL effects was observed between Treg cells and naive T cells (Spearman  $R^2 = 0.93$ ) than between Treg cells and monocytes

(Spearman  $R^2 = 0.76$ ), also confirmed by the  $\pi_1$  estimates ([Figures 2B](#) and [2C](#)). Despite the substantial eQTL sharing between Treg cells and the other two cell types, we classified 775 genes (21% of all eQTL genes) as specific to our Treg cell dataset, including 92 genes that were only expressed in Treg cells (intersection between 187 genes only expressed in naive T cells and 384 genes only expressed in monocytes; [STAR Methods](#); [Figures 2A](#) and [S4B](#)). Of these 775 genes, 695 were also only detected in Treg cells compared with the naive T cells and monocytes from the DICE consortium, while of the 92 genes expressed only in Treg cells, 82 were also only expressed in Treg cells compared with the DICE consortium ([Figure S4C](#)). eQTL genes specifically expressed in Treg cells, but not in the other two cell types, showed lower expression levels compared with genes expressed in all cell types ([Figure S4D](#)). Therefore, some of the eQTL effects could be shared with the other cell types and the higher sequencing depth of our study enabled capturing the transcripts of these genes while they were not detected in the BLUEPRINT datasets. Among the Treg-cell-specific eQTLs, there were many genes essential to immune function regulation, including *TNFRSF14* (false discovery rate [FDR] =  $2.86 \times 10^{-4}$ ), a chemokine that attracts lymphocytes toward epithelial cells ([Figure 2D](#)). A *TNFRSF14* eQTL is also found in monocytes, but the variants are in low LD, while in naive T cells, the gene is expressed at very low levels.

To compare the genetic effects across the same peak regions in all three cell types, we performed peak calling on reads merged from all cell types (see [STAR Methods](#)). When we compared the actQTLs across the three cell types, we observed that 1,307 (29%) actQTLs were Treg-cell-specific ([Figures 2A](#) and [S4B](#)). Although the concordance of the effect sizes across all peaks was small (Spearman  $R^2 \leq 0.28$ ), peaks with shared QTLs expressed similar effect sizes ( $R^2 \geq 0.84$ ; [Figure 2B](#)). As expected, there was a higher correlation of effect sizes between Treg cells and naive T cells (Spearman  $R^2 = 0.94$ ) than between Treg cells and monocytes (Spearman  $R^2 = 0.79$ ; [Figure S4A](#)). Results from the  $\pi_1$  analysis also reflected our observations from the RNA that the replication of Treg cell QTLs was higher in the naive dataset compared with the monocyte ([Figure 2C](#)). Among the Treg-cell-specific actQTL effects, we observed a peak in the promoter of *FCRL3* gene (chr1:157,693,404–157,705,914; FDR =  $3.47 \times 10^{-8}$ ) that is a potential negative regulator of Treg cell suppressive function ([Figure 2E](#)).<sup>30</sup>

### Treg cell QTLs colocalize with immune-disease loci

To fine-map disease-associated loci to causal genes and variants, we next integrated the Treg cell QTL results with GWAS signals from common immune diseases. We applied a Bayesian framework to test for statistical colocalization of the disease-associated variants and the Treg cell QTL signals.<sup>31</sup> Collectively, we tested 1,290 unique GWAS loci associated with 14 immune-mediated diseases: allergic diseases (ALL), ankylosing spondylitis (AS), asthma (AST), celiac disease (CEL), Crohn's disease (CD), inflammatory bowel disease (IBD), multiple sclerosis (MS), primary biliary cirrhosis (PBC), psoriasis (PS), rheumatoid arthritis (RA), systemic lupus erythematosus (SLE), type 1 diabetes (T1D), ulcerative colitis (UC), and vitiligo



**Figure 2. Comparison of eQTLs and actQTLs identified in regulatory T cells, CD4<sup>+</sup> naive cells, and monocytes**

(A) Proportion of eQTLs and actQTLs specific to Treg cells in comparison to naive T cells and monocytes.

(B) Pairwise pi1 score between the three cell types for eQTLs and actQTLs.

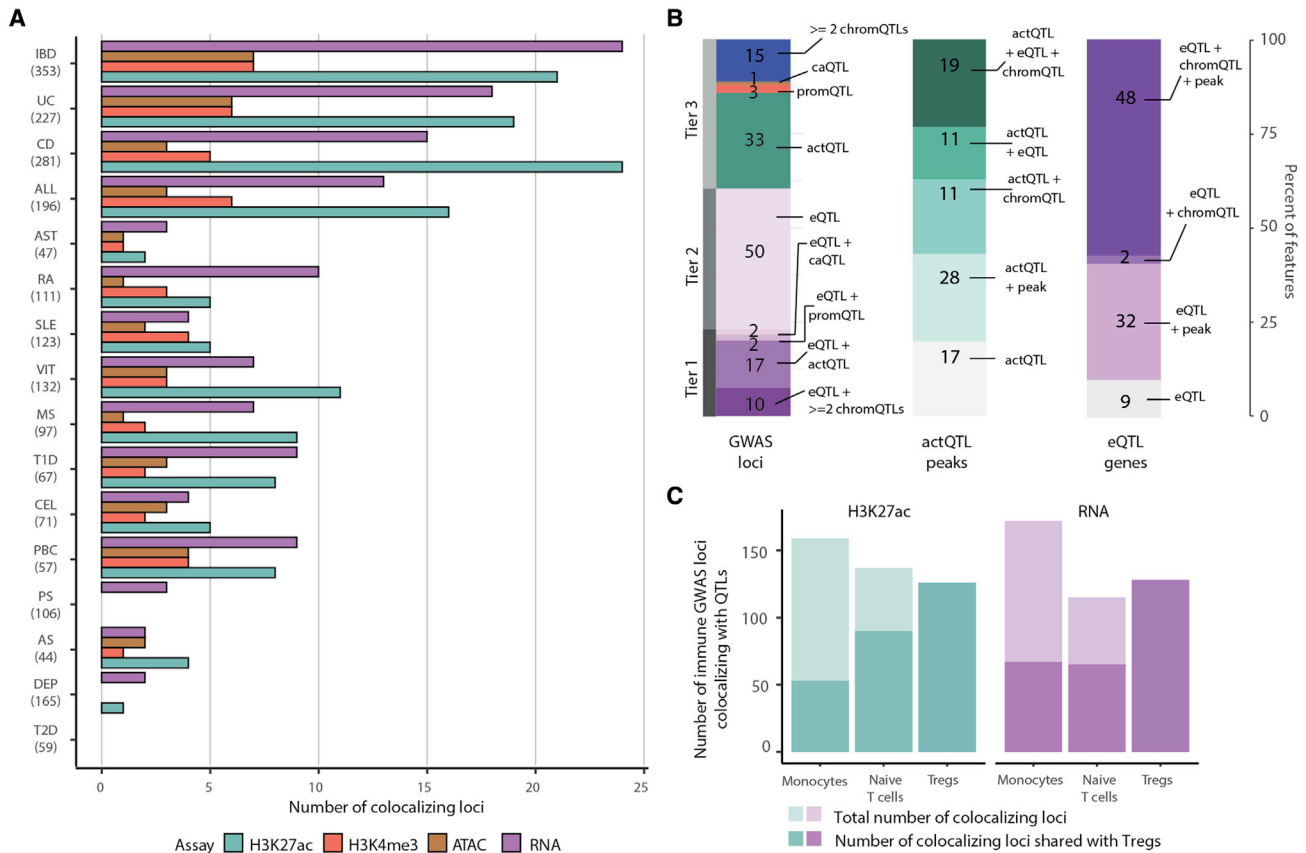
(C) Spearman correlation between the regression slopes for the same gene or peak and variant pairs of all eQTLs and actQTLs (colored) and only for the shared pairs (gray).

(D and E) Examples of Treg-cell-specific (D) eQTLs and (E) actQTLs. FPM, fragments per million; TPM, transcripts per million. FDR, false discovery rate.

(VIT) (see STAR Methods and Table S3). Diseases with the highest number of colocalizations (more than 20 colocalizing signals) included IBD, UC, CD, ALL, T1D, VIT, and PBC (Figures 3A and S5A). The high number of observed colocalizations is consistent with previous work that implicated the role of Treg cells in the pathobiology of all of these diseases<sup>5,6,32–34</sup> and to some extent also reflects the greater number of significant GWAS loci for these traits. Overall, immune-mediated diseases showed more colocalizations with Treg cell QTLs than non-immune-mediated diseases, such as type 2 diabetes or depression.

Four of the colocalizing eQTLs were shared between three or more diseases and included *BACH2* (T1D, AS, and MS), *SUOX* (ALL, VIT, and T1D), *TYK2* (PBC, RA, SLE, and T1D), and *ZFP90* (PS, ALL, and UC). The *BACH2* locus also contained an actQTL (chr6:90,264,695–90,268,560), which

colocalized with ALL, AST, MS, CEL, VIT, CD, and IBD. Similarly, *SUOX* colocalized with an actQTL (chr12:55,989,136–56,011,728), a promQTL (chr12:55,996,308–55,998,877), and a caQTL (chr12:56,041,233–56,042,198) for T1D, ALL, and VIT. We observed the largest number of colocalizations with Treg cell actQTLs. There were also chromQTLs that colocalized with multiple diseases but did not have a corresponding colocalizing eQTL. Among them, we identified (1) a region upstream of *CXCR5*, which colocalized with an actQTL (chr11:118,866,698–118,871,517), a caQTL (chr11:118,869,935–118,870,610), and a promQTL (chr11:118,869,586–118,871,234; CEL, RA, and PBC); (2) an actQTL (chr11:76,586,431–76,600,121) upstream of *LRR32* (ALL, AST, T1D, UC, CD IBD, and T1D); (3) a promQTL in an intron of *ZMIZ1* (chr10:79,240,389–79,246,577; AS, MS, and IBD); and (4) an actQTL (chr17:39,751,832–39,807,281;



**Figure 3. Colocalization of immune disease GWAS loci and Treg cell QTLs**

(A) Distribution of Treg cell eQTLs and chromatin QTLs colocalizing with different immune disease GWAS loci. Number in parentheses is state-independent loci associated with the trait. The numbers on the right side of the bars correspond to the total number of features (genes or peaks) tested for colocalization. ALL, allergic disease (asthma, hay fever, and eczema); AST, asthma; CD, Crohn's disease; CEL, celiac disease; DEP, broad depression; IBD, inflammatory bowel disease; MS, multiple sclerosis; PBC, primary biliary cirrhosis; PS, psoriasis; RA, rheumatoid arthritis; SLE, systemic lupus erythematosus; T1D, type 1 diabetes; T2D, type 2 diabetes; UC, ulcerative colitis; VIT, vitiligo.

(B) Distribution of the GWAS loci colocalizing with different types of Treg cell QTLs.

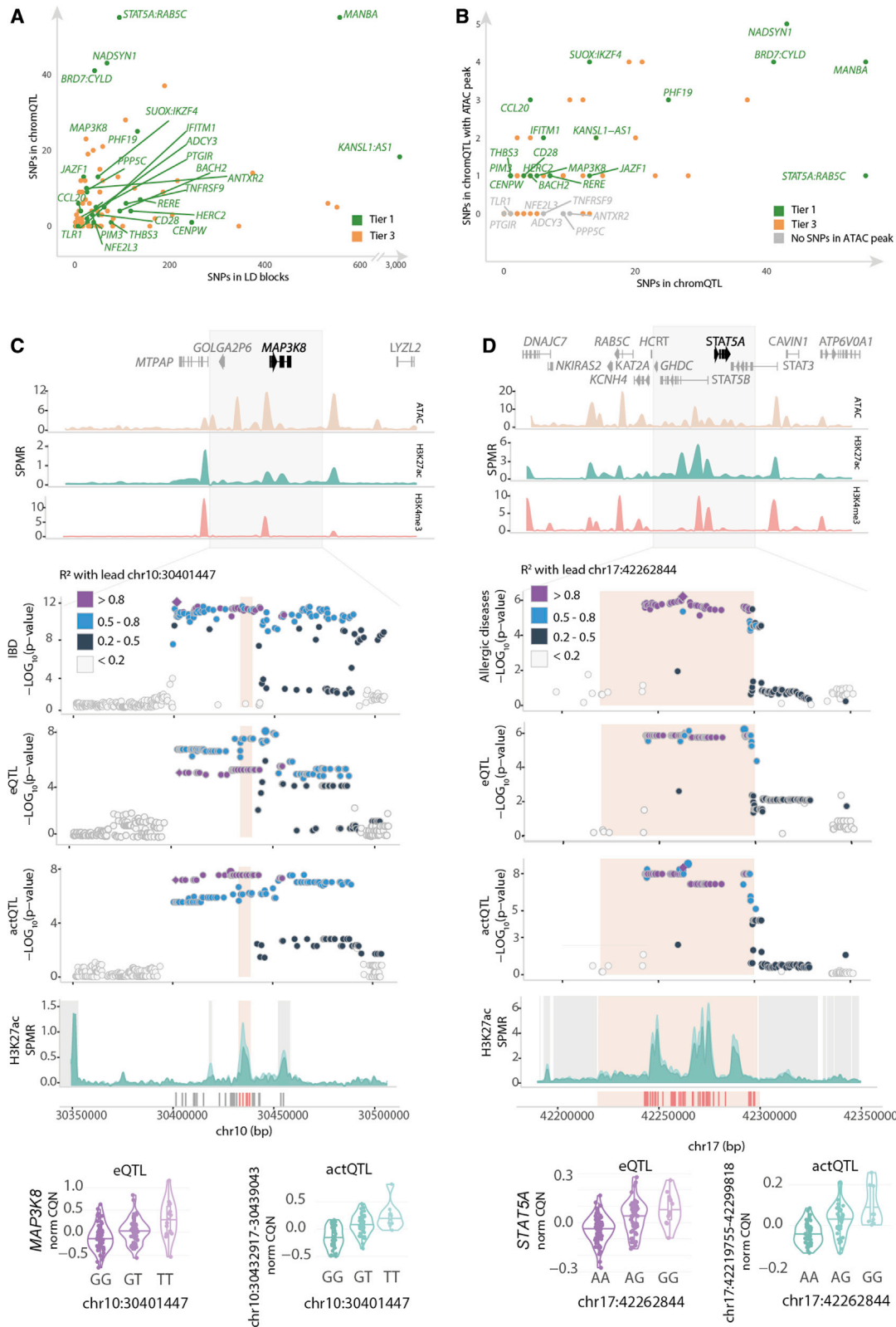
(C) Number of immune GWAS loci colocalizing with monocyte, naive T cell, and Treg cell eQTLs and actQTLs.

SLE, ALL, and PBC) and a promQTL (chr17:39,912,458–39,929,022; PBC, AST, T1D, SLE, and ALL) in the same region but covering *IKZF3* and *ORMDL3*, respectively (Table S3).

We observed 360 significant colocalizations between the disease loci and at least one Treg cell QTL, corresponding to 133 unique GWAS loci (Figure 3B; Table S3). Of the 133 unique GWAS loci, 50 loci colocalized with eQTLs only, 52 with chromQTLs only, and 31 colocalized with both eQTL and at least one chromQTL. The colocalizations with both transcriptomic and chromatin evidence affected the expression of 37 eQTL genes, acetylation of 31 actQTL peaks, methylation of 10 promQTL peaks, and accessibility of 11 caQTL sites. Of the immune disease GWAS loci that colocalized with both Treg cell eQTLs and chromQTLs, 27 out of 31 comprised actQTLs (87%). Finally, for the vast majority (79%) of the loci where we observed disease signals colocalizing with two or more types of QTLs, the effects of the risk alleles propagated in the same direction. For example, the *CCL20* eQTL colocalized with UC variants, tagged by chr2:228,670,575, and the risk allele resulted in both reduced gene expression and

decreased H3K27 acetylation (chr2:227,804,673–227,819,6), H3K4 tri-methylation (chr2:227,805,541–227,808,260), and chromatin accessibility (chr2:227,805,505–227,805,928; Table S3). However, at 10 loci, we observed that the disease alleles resulted in opposite effects between the different types of QTLs, suggesting complex mechanisms of gene expression regulation (Table S3).

We systematically investigated all immune disease signals colocalizing with Treg cell QTLs to refine the disease-associated signals to sets of functional variants and to nominate causal genes. We classified colocalizing loci into three categories. Tier 1 loci comprised 31 signals for which the GWAS association colocalized with both eQTL and chromatin QTL (Figure 3B; Table S4). Of these, at 25 loci, the associated variants were also located within the chromQTL peaks. Loci in this category were the most informative to functionally refine disease associations, as we were able to link the GWAS signals to genes and to functional chromatin elements that regulated gene expression. Tier 2 loci contained 50 signals for which we



(legend on next page)

observed colocalization only with eQTLs. In this case, we were unable to refine the association signals to sets of functional variants, but we were able to connect the GWAS signals to candidate causal genes. Finally, the 52 loci in tier 3 included GWAS signals colocalizing with chromatin QTLs, but not eQTLs. Of these, at 40 loci, the GWAS variants overlapped a chromatin QTL peak, providing further clues to prioritize functional variants at GWAS loci (Table S4). Finally, tier 3 loci represented the majority of colocalizations. We hypothesized that gene expression effects could be manifested in a cell-state-specific context. To further nominate candidate genes regulated by the variants colocalizing with actQTLs, we used resting and activated Treg cell transcriptome data (see STAR Methods) and defined genes proximal to the QTL peaks that were differentially expressed upon cell activation (Table S5). This analysis prioritized 124 genes linked to 44 disease-colocalizing actQTLs. We went on to carry out allele-specific expression analysis for these loci and validated 36 of these genes as displaying imbalanced expression with regards to the lead GWAS variant (STAR Methods). In parallel, we used cap analysis of gene expression (CAGE) data from FANTOM5<sup>35</sup> and linked the enhancer usage of 50 of the disease-colocalizing actQTLs to the TSS expression of 374 genes (STAR Methods). Overlapping these approaches, we found 34 actQTLs connected to 56 genes differentially expressed upon Treg cell stimulation, 23 of which displayed allele-specific expression, including CD247, LRRRC32, and PRDM1. For a subset of loci, we have therefore compiled candidate gene target lists based on allelic and gene expression evidence across platforms.

Next, we assessed which of the identified eQTLs that colocalized with immune disease variants regulated gene expression specifically in Treg cells and not in naive T cells or monocytes. Out of the 81 GWAS loci that showed colocalization with Treg cell eQTLs, 31 were Treg cell exclusive and not present in naive T cells or monocytes (Figures 3C, S5C, and S5D). Similarly, 21 of 78 of actQTLs loci were Treg cell exclusive. Three of the Treg-cell-specific colocalizing eQTLs also had specific colocalization with a Treg cell actQTL: *MAP3K8*, which colocalized with UC and IBD; *IFITM1* colocalizing with PBC; and *TLR1* colocalizing with ALL. Treg-cell-colocalizing actQTLs were enriched for JUN, GATA3, and STAT6 transcription factors (Table S6).

### Colocalizing Treg cell QTLs prioritize immune disease causal variants and genes

Using the tier 1 and tier 3 loci that overlapped with chromatin QTL peaks, we refined the signals at 68 GWAS loci from a median of 48 associated variants to six functional variants per

locus (Figure 4A; Tables S4 and S7). Of the 68 loci, in 45 instances, we observed that the genetic variants additionally overlapped open chromatin peaks, allowing us to further prioritize the functional variants from an average of 13 functional variants to an average of two variants per locus, including *BACH2*, *CD28* (Figure S6), *CENPW*, *HERC2*, *JAZF1*, *MAP3K8*, *PIM3*, *RERE*, *STAT5A*, and *THBS3* loci, which colocalized with eQTLs and were refined to a single functional variant (Figure 4B; Table S4). In the case of previously statistically fine-mapped loci, in which associations have been refined to rare variants or haplotypes, such as *CD28*, *BACH2*, *CTSH*, and *TYK2*,<sup>36,37</sup> the information from Treg cell QTL colocalizations prioritized additional functional variants.

Treg-cell-exclusive colocalizations along with the Treg cell actQTL-specific colocalizations indicated regulation of pathways that were characteristic of Treg cell biology. We therefore investigated in more detail the Treg-cell-exclusive colocalization with an IBD GWAS signal, tagged by the chr10:30,401,447 (rs10826797) variant, which colocalized with an actQTL, regulating a 6-kb-large (chr10:30,432,917–30,439,043) H3K27ac peak ( $p = 9.5 \times 10^{-9}$ ) at the TSS of *MAP3K8* and an eQTL ( $p = 9.8 \times 10^{-6}$ ) for the *MAP3K8* gene (Figure 4C; Table S3). The IBD risk allele decreased the acetylation at H3K27 and downregulated the expression of *MAP3K8*. Five of the colocalizing variants overlapped this actQTL peak, of which only one SNP, chr10:30,434,664 (rs306588), overlapped a 1.5 kb ATAC peak (chr10:30,433,210–30,434,733; Tables S3 and S7). This approach refined the IBD-associated signal from 30 GWAS variants to a single functional candidate variant regulating the expression of *MAP3K8*, a kinase modulating the DNA-binding activity of FoxP3, the Treg cell hallmark transcription factor.<sup>38</sup>

In another example, we observed that a locus associated with allergies<sup>39</sup> (tagged by the index SNP chr17:42,262,844 [rs7207591]) colocalized with a *STAT5A* eQTL ( $p = 3.9 \times 10^{-6}$ ), as well as with an 80-kb actQTL (chr17:42,219,755–42,299,818;  $p = 4.2 \times 10^{-9}$ ; Figure 4D). This peak overlapped the *STAT5A* TSS. Nearly half of the LD block of allergy variants (55 out of 93 SNPs) overlapped with the regulated actQTL peak and one of the variants, chr17:42,266,938 (rs34129849), also mapped to a 629-bp open chromatin region (chr17:42,266,595–42,267,224) located in intron one of the *STAT5B* gene (Figure 4D; Table S3). Modulation of STAT5-mediated pathways could implicate broad effects on Treg cell function as *STAT5A* regulates the expression of genes downstream of the interleukin-2 (IL-2) receptor, which is critical for Treg cell development and function.<sup>40</sup>

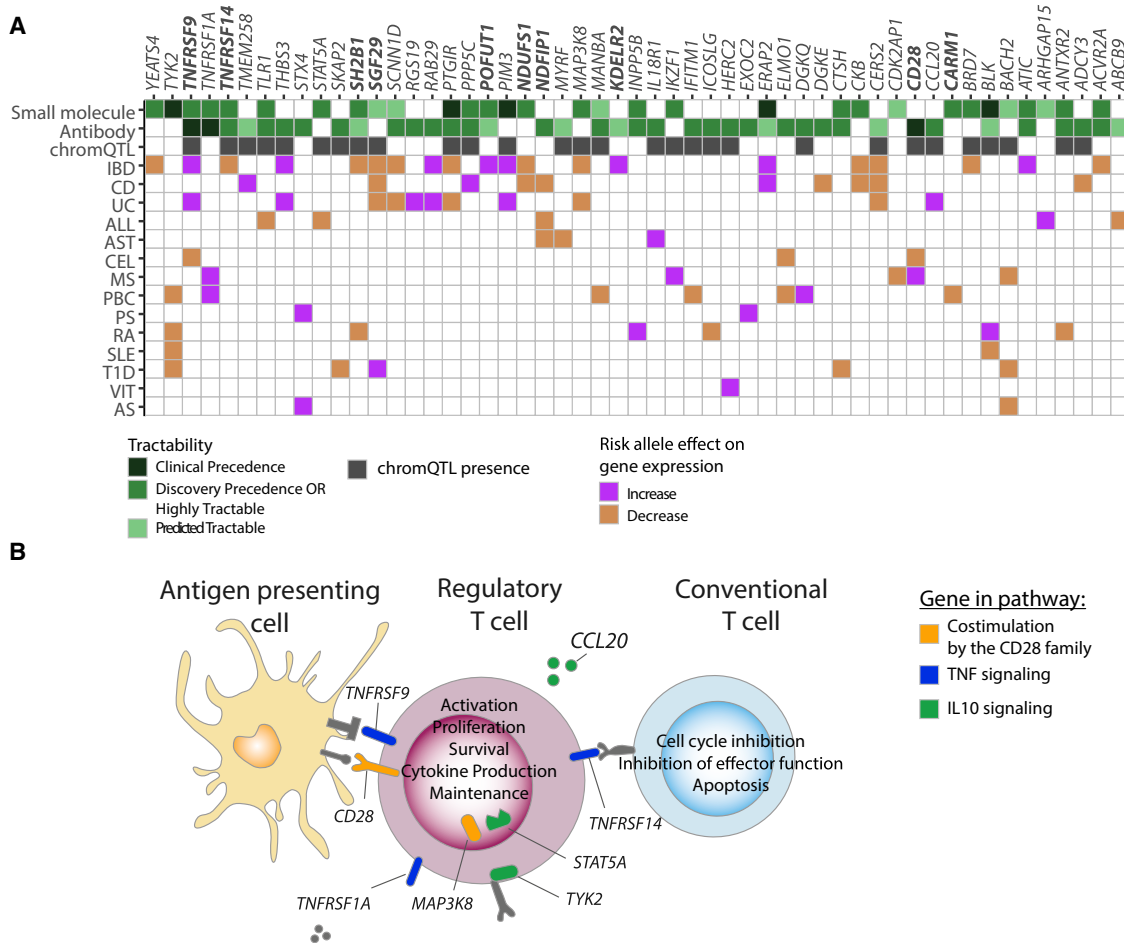
### Figure 4. Functional refinement of immune disease associations colocalizing with Treg cell QTLs

(A) The number of SNPs in LD blocks (lead GWAS signals and their proxies  $R^2 \geq 0.8$ ) on the x axis and the number of SNPs that map inside chromQTL peaks on the y axis.

(B) The number of SNPs in LD blocks that map inside chromQTL peaks on the x axis and the number of SNPs that map inside both chromQTL and an additional ATAC peak on the y axis.

(C and D) From top to bottom, the figure displays gene annotation tracks; chromatin landscape for ATAC-seq, H3K27ac, and H3K4me3 ChM-seqs; region association plots for disease; eQTL and actQTL association p values, focused on H3K27ac landscape stratified by homozygous genotypes; and genotype-stratified eQTL and actQTL violin plots. (C) Locus associated with IBD, tagged by chr10:30,401,447 (rs10826797) SNP colocalizing with *MAP3K8* eQTL and chr10:30,432,917–30,439,043 actQTL is shown. (D) Locus associated with allergies, tagged by chr17:42,262,844 (rs7207591) SNP colocalizing with *STAT5A* eQTL and chr17:42,219,755–42,299,818 actQTL is shown. CQN, conditional quantile normalized reads; SPMR, signal per million reads.





**Figure 5. Immune disease colocalizations with Treg cell QTLs inform drug targets**

(A) Tier 1 and Tier 2 loci colocalizing with immune disease GWAS variants with drug tractability evidence (green). In bold are Treg-cell-specific eQTLs. Clinical precedence, gene targeted by small molecules or antibodies approved for patient treatment or undergoing clinical trials; discovery precedence, gene product shown to bind small molecules; predicted tractable, gene predicted to be small molecule tractable; tractable high confidence, gene product with high predicted tractability as an antibody drug target; tractable medium–low confidence, gene product with predicted tractability as an antibody drug target. *NDUFS1* is not directly targeted but is part of a targeted complex.

(B) Tier 1 and tier 2 genes with tractability potential in CD28 co-stimulation (orange), TNF (blue), and anti-inflammatory IL-10 (green) pathways.

### Treg cell QTLs emphasize CD28 co-stimulation, tumor necrosis factor, and IL-10 signaling pathways for drug targeting

Despite the success of GWAS in mapping disease risk variants, the efforts to translate these findings into drug targets have been challenging. Therefore, we used the Open Targets Platform<sup>41</sup> to systematically assess whether eQTLs that colocalized with immune disease signals identify known and potential new drug targets (see STAR Methods). Of the 91 eQTL genes that colocalized with immune diseases and could be tested in the Open Targets Platform, we found nine (tier 1: *BLK*, *CD28*, *PIM3*, *PTGIR*, and *TNFRSF9* and tier 2: *ERAP2*, *NDUFS1*, *TNFRSF1A*, and *TYK2*; Figure 5A) that were already targeted by known drugs and were either used in clinical practice or undergoing clinical trials. Seven of these eQTL genes could be considered for drug repurposing: *ERAP2*, *NDUFS1*, *PIM3*, *PTGIR*, *TNFRSF1A*, *TNFRSF9*, and *TYK2*, three of which are

Treg-cell-specific eQTLs. However, most of the drugs targeting these genes are used for cancer therapies, where the desired effects include dampening the suppressive capacity of Treg cells, in contrast to immune diseases where the enhancement of Treg cell function is sought after. Nevertheless, this analysis highlighted some potential drug candidates, for example, a colocalization between a *NDUFS1* eQTL and CD, in which the disease risk allele increased gene expression, suggested repurposing metformin, which targets the NADH dehydrogenase complex (not directly *NDUFS1*). Metformin is currently used for treating type 2 diabetes; in a clinical trial for MS patients, it increased the number of Treg cells,<sup>42</sup> and in *in vitro* studies, it promoted Treg cell proliferation.<sup>43</sup>

In addition, we observed 63 genes that were not yet a part of a clinical treatment but had drug tractability evidence, of which 47 were classified as highly tractable (eight of which were specific to Treg cells; Figure 5A; Tables S8 and S9). We used Open Target's



definition of tractability (druggability), which is based on availability of a binding site in the protein that can be used for small-molecule binding, presence of an accessible epitope for antibody-based therapy, or reports of a compound in clinical trials with a modality other than small molecule or antibody. An example of a highly tractable gene was *ERAP2*, for which we observed the IBD and CD risk allele colocalizing with the *ERAP2* eQTL increased gene expression, implicating *ERAP2* as a target for validation. Collectively, we observed that genes with high tractability evidence fell into three pathways: co-stimulation by the CD28 family ( $p = 0.012$ ), tumor necrosis factor (TNF) signaling ( $p = 0.0034$ ), and IL-10 signaling ( $p = 0.01$ ; Figure 5B). These pathways play an important role in Treg cell activation, proliferation, and survival, as well as in suppression of effector T cells.

Finally, of the 91 genes that were tested, 44 had been knocked down or knocked out in mice, of which 26 had a reported immune system phenotype (Table S10). Among those, six gene knockouts resulted in an immune disease, including *Cd28*, *Ndfip1*, *Skap2*, *Tmem258*, *Tnfrsf1a*, and *Tnfrsf9*. For example, *Tnfrsf1a*<sup>-/-</sup> decreased susceptibility to experimental autoimmune encephalomyelitis, consistent with our observation that the risk allele for multiple sclerosis in Treg cells leads to increased *TNFRSF1A* gene expression levels. In addition, *Icosl*, *Ikzf1*, *Map3k8*, *Pofut1*, *Ptgirg1r*, and *Stat5a* had specific organ inflammatory phenotypes, such as reduced small intestine inflammation in *Map3k8*-deficient mice.<sup>44</sup> It is important to note that we did not observe the same direction of effects for all mouse knockouts, which could be partly due to the fact that gene perturbations are not Treg cell specific.

## DISCUSSION

Pinpointing genes that are regulated by disease-associated non-coding variants can uncover important cell pathways for drug targeting. However, leveraging information captured by GWAS variants to provide insight into disease biology and improve treatment has been challenging. Increasing availability of functional genomic resources from different cell types helps to bridge this gap. Naive and regulatory CD4<sup>+</sup> T cells are closely related, yet they play distinct functions in the immune system. Although naive CD4<sup>+</sup> T cells have been extensively characterized, Treg cells are an infrequent cell population difficult to isolate in large numbers for QTL analysis and elusive to deconvolute from bulk blood QTL data.<sup>45</sup> Therefore, mapping gene expression regulation directly in Treg cells is essential to better understand Treg cell biology.

Here, we sought to describe the role of immune-disease-associated variants on modulation of gene expression in Treg cells. We linked 133 unique immune disease loci from associated variants to functional effects in Treg cells; 50 loci were linked to gene expression, 52 loci were linked to an effect on chromatin, and 31 loci to both. Loci for which we observed colocalization with both gene expression and chromatin QTLs provide an important translational insight into mechanisms through which immune disease variants regulate Treg cell function. For example, we observed signals overlapping with Treg-cell-spe-

cific eQTLs, indicating regulation of essential Treg cell pathways, such as IL-2 signaling via *STAT5A*.

The 52 loci for which we only detected colocalization with chromQTLs, but not eQTLs, indicate that the altered gene expression may be manifested in a specific cell state, which will require tailored functional follow-up studies. In a separate study, we demonstrated that the disease colocalization with actQTL near *LRR32* (encoding GARP) resulted in reduced GARP expression in activated Treg cells and subsequently led to reduced Treg cell suppression, which promoted development of colitis.<sup>46</sup> In addition, previous studies showed that context-specific eQTLs can be already detected in a resting state at the chromQTL level.<sup>2,22,24</sup>

On the other hand, the 50 loci colocalizing only with eQTL variants, but not chromQTLs, may be correlated with chromatin-independent gene expression regulation, such as splicing QTLs (sQTLs)<sup>47</sup> or RNA stability.<sup>48</sup> For example, we observed that *ERAP2*, an IBD- and CD-associated locus, showed an eQTL colocalization but no chromQTL effect. The lead GWAS variant chr5:96,912,106 (rs6873866) and the colocalizing lead eQTL variant chr5:96,916,728 (rs2927608) are proxies for chr5:96,900,192 (rs2248374), a sQTL present in monocyte-derived dendritic cells after influenza infection and type 1 interferon stimulation.<sup>49</sup>

By linking immune disease GWAS variants to Treg cell eQTLs, our study contributes toward building genetic evidence for the causal role of Treg cells in disease biology and supports the discovery and repurposing of drugs that modulate Treg cell function in treating immune disease patients. Validation of targets with genetic support can significantly increase the chance of clinical success.<sup>50,51</sup> Our results support the focus on modulating co-stimulatory and cytokine pathways, for example, at a CEL locus, the disease risk alleles led to decreased levels of expression of *TNFRSF9* (encodes for CD137/4-1BB). Signaling via CD137 induces cell division and proliferation,<sup>52,53</sup> however, *TNFRSF9* gene expression and protein levels increase specifically in activated Treg cells, but not in conventional T cells.<sup>54,55</sup> Furthermore, the increased expression of CD137 enhances the Treg cell capacity to suppress proliferation of effector T cells.<sup>56</sup> Therefore, the disease risk allele could result in decreased Treg cell suppressive function and promote immune imbalance. The colocalization of several TNF receptor superfamily members (*TNFRSF1A*, *TNFRSF9*, and *TNFRSF14*) further supports the development of drugs modulating TNF pathway, one of the main therapy lines for treating immune diseases.

Understanding the genetic underpinnings of immune system regulation has broad implications not only in the treatment of immune-mediated conditions but also in infections, transplantation, and cancers. For instance, in organ transplantation, numbers of Treg cells, as well as Treg cells with increased suppressive capacity, can provide a favorable environment of successful transplant tolerance.<sup>12,57,58</sup> Furthermore, in hematopoietic stem cell transplantation, high Treg cell:CD4 T cell ratios are associated with reduced acute graft-versus-host disease and reduced overall mortality.<sup>58</sup> Importantly, *in vitro* expanded Treg cells with enhanced suppressive capacity have already entered clinical trials.<sup>59</sup> Therefore, identifying genetic variants

that regulate gene expression in a specific cellular context can inform development of more effective cell therapies. Our study provides an important advancement in mapping regulation of gene expression in Treg cells, and consequently, our results can benefit a range of clinical conditions.

### Limitations of the study

There are the following limitations to our study that should be considered. For example, a subset of loci where we were unable to link chromQTLs with eQTLs could also result from long-distance gene expression regulation, the eQTL being outside of our testing window, or from combinatorial subtle effects between multiple enhancers regulating the expression of individual genes. We also recognize instances of complex regulation of gene expression that will require targeted follow-up studies to fully uncover the functional role of disease variants. For example, we observed a complex pattern of colocalization between *CD28* eQTL, nearby actQTLs, and immune disease GWAS variants. The eQTL for *CD28*, the co-stimulatory receptor found on the surface of the majority of T cells, was specific to Treg cells and absent from naive T cells. The risk alleles for CEL and MS showed reversed effects on *CD28* expression and the acetylation of the peaks, implicating complex enhancer-mediated control of *CD28* expression under cell type and cell-state-specific mechanisms. Therefore, the results we describe here form the basis for hypothesis-driven functional follow-up studies into Treg-cell-mediated development of autoimmune and inflammatory diseases.

Finally, determining cell-type-specific QTL effects is challenging due to technical confounding factors between studies, including sequencing depth, different sample sizes across studies, different protocols of sample processing, etc. Although we performed numerous analyses to demonstrate our dataset captured true Treg-cell-specific effects, we recognize that some of the Treg-cell-specific effects that we identified here could be shared with other cell types. The dropping costs of single-cell transcriptomic technologies, higher gene capture efficiency, and increasing applicability to profile transcriptome of immune cells both in circulation and in tissues will map the cell-type- and context-specific gene expression regulation with high precision.

### STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- **KEY RESOURCES TABLE**
- **RESOURCE AVAILABILITY**
  - Lead contact
  - Materials availability
  - Data and code availability
- **METHOD DETAILS**
  - Sample collection and Treg isolation
  - Sample summary
  - FACS staining
  - Culture and stimulation of isolated Tregs
  - SNP genotyping and imputation

- RNA-seq
- ATAC-seq
- H3K4me3 and H3K27ac ChIPmentation-seq
- **QUANTIFICATION AND STATISTICAL ANALYSIS**
  - RNA-seq data processing
  - Chromatin marks data processing
  - Quantitative trait locus mapping (QTLs)
  - Allele-specific expression analysis
  - Colocalization of QTL signals with immune disease GWAS
  - Gene expression variance deconvolution
  - SNP functional annotation
  - LD loci definition and classification
  - Known drug tractability evidence analysis
  - Differential gene expression analysis
  - FANTOM5 CAGE data integration
  - Transcription factor binding site (TFBS) enrichment analysis

### SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.xgen.2022.100117>.

### ACKNOWLEDGMENTS

This research was funded by the Wellcome Trust (grant number WT206194). For the purpose of open access, the author has applied a CC BY public copyright license to any Author Accepted Manuscript version arising from this submission. L.B.-C. was supported by the MRC Skills Development Fellowship (MR/N014995/1). We thank all participating blood donors and the Cambridge and Oxford NHS Blood and Transplant for the recruitment of study participants. We thank the Wellcome Sanger Institute Flow Cytometry, Sequencing, IT and Data Access facilities for their essential contribution to data generation and processing. We would like to acknowledge the enthusiasm and effort of Kiran Kumar Thurimella and Aman S. Patel in this project. We thank Emma Davenport for her feedback on the manuscript.

### AUTHOR CONTRIBUTIONS

G.T. conceived the work; L.B.-C., D.A.G., N.K., and G.T. designed the experiments; D.A.G., N.K., G.G., C.C., and D.J.S. performed fluorescence-activated cell sorting (FACS), cell culture, and DNA and RNA isolation and generated RNA-seq, ATAC-seq, and chromatin immunoprecipitation (ChIP)mentation sequencing libraries; L.B.-C. processed genotyping, ATAC-seq, and ChIPmentation sequencing data; D.A.G. processed RNA-seq data; L.B.-C. and D.A.G. analyzed and integrated the data; M.S. and I.D. performed drug tractability analyses; A.A.L. and D.J.R. recruited the blood donors; A.M. contributed to immune disease risk polymorphism enrichment analysis; K.A. contributed to the establishment of QTL analysis pipelines; B.S. and E.C.-G. helped interpret the results; K.K. contributed to genotype imputation; N.S. contributed to the comparative analysis with naive T cells; A.L. supported website development for LocusZoom plots; and L.B.-C., D.A.G., and G.T. wrote the manuscript.

### DECLARATION OF INTERESTS

The authors declare no competing interests. D.A.G. is undertaking a fellowship at Vertex Pharmaceuticals.

**INCLUSION AND DIVERSITY**

One or more of the authors of this paper self-identifies as an underrepresented ethnic minority in science. One or more of the authors of this paper self-identifies as a member of the LGBTQ+ community. One or more of the authors of this paper self-identifies as living with a disability.

Received: February 5, 2021

Revised: November 2, 2021

Accepted: March 15, 2022

Published: April 6, 2022

**REFERENCES**

- Nica, A.C., Montgomery, S.B., Dimas, A.S., Stranger, B.E., Beazley, C., Barroso, I., and Dermizakis, E.T. (2010). Candidate causal regulatory effects by integration of expression QTLs with complex trait genetic associations. *PLoS Genet.* **6**, e1000895.
- Alasoo, K., Rodrigues, J., Mukhopadhyay, S., Knights, A.J., Mann, A.L., Kundu, K., HIPSCI Consortium; Hale, C., Dougan, G., and Gaffney, D.J. (2018). Shared genetic effects on chromatin and gene expression indicate a role for enhancer priming in immune response. *Nat. Genet.* **50**, 424–431.
- Guo, H., Fortune, M.D., Burren, O.S., Schofield, E., Todd, J.A., and Wallace, C. (2015). Integration of disease association and eQTL data using a Bayesian colocalisation approach highlights six candidate causal genes in immune-mediated diseases. *Hum. Mol. Genet.* **24**, 3305–3313.
- Huang, H., Fang, M., Jostins, L., Umičević Mirkov, M., Boucher, G., Anderson, C.A., Andersen, V., Cleynen, I., Cortes, A., Crins, F., et al. (2017). Fine-mapping inflammatory bowel disease loci to single-variant resolution. *Nature* **547**, 173–178.
- Trynka, G., Sandor, C., Han, B., Xu, H., Stranger, B.E., Liu, X.S., and Raychaudhuri, S. (2013). Chromatin marks identify critical cell types for fine mapping complex trait variants. *Nat. Genet.* **45**, 124–130.
- Amariuta, T., Luo, Y., Gazal, S., Davenport, E.E., van de Geijn, B., Ishigaki, K., Westra, H.-J., Teslovich, N., Okada, Y., Yamamoto, K., et al. (2019). IMPACT: genomic annotation of cell-state-specific regulatory elements inferred from the epigenome of bound transcription factors. *Am. J. Hum. Genet.* **104**, 879–895.
- Bovenschen, H.J., van de Kerkhof, P.C., van Erp, P.E., Woestenenk, R., Joosten, I., and Koenen, H.J.P.M. (2011). Foxp3<sup>+</sup> regulatory T cells of psoriasis patients easily differentiate into IL-17A-producing cells and are found in lesional skin. *J. Invest. Dermatol.* **131**, 1853–1860.
- Mohammadnia-Afrouzi, M., Zavarani Hosseini, A., Khalili, A., Abediankenari, S., Hosseini, V., and Maleki, I. (2015). Decrease of CD4(+) CD25(+) CD127(low) FoxP3(+) regulatory T cells with impaired suppressive function in untreated ulcerative colitis patients. *Autoimmunity* **48**, 556–561.
- Ferreira, R.C., Simons, H.Z., Thompson, W.S., Rainbow, D.B., Yang, X., Cutler, A.J., Oliveira, J., Castro Dopico, X., Smyth, D.J., Savinykh, N., et al. (2017). Cells with Treg-specific FOXP3 demethylation but low CD25 are prevalent in autoimmunity. *J. Autoimmun.* **84**, 75–86.
- Fletcher, J.M., Lonergan, R., Costelloe, L., Kinsella, K., Moran, B., O’Farrelly, C., Tubridy, N., and Mills, K.H.G. (2009). CD39+Foxp3+ regulatory T Cells suppress pathogenic Th17 cells and are impaired in multiple sclerosis. *J. Immunol.* **183**, 7602–7610.
- Lindley, S., Dayan, C.M., Bishop, A., Roep, B.O., Peakman, M., and Tree, T.I.M. (2005). Defective suppressor function in CD4(+)CD25(+) T-cells from patients with type 1 diabetes. *Diabetes* **54**, 92–99.
- Fisher, S.A., Lamikanra, A., Dorée, C., Gratton, B., Tsang, P., Danby, R.D., and Roberts, D.J. (2017). Increased regulatory T cell graft content is associated with improved outcome in haematopoietic stem cell transplantation: a systematic review. *Br. J. Haematol.* **176**, 448–463.
- Sharabi, A., Tsokos, M.G., Ding, Y., Malek, T.R., Klatzmann, D., and Tsokos, G.C. (2018). Regulatory T cells in the treatment of disease. *Nat. Rev. Drug Discov.* **17**, 823–844.
- Bluestone, J.A., Buckner, J.H., Fitch, M., Gitelman, S.E., Gupta, S., Hellerstein, M.K., Herold, K.C., Lares, A., Lee, M.R., Li, K., et al. (2015). Type 1 diabetes immunotherapy using polyclonal regulatory T cells. *Sci. Transl. Med.* **7**, 315ra189.
- Rigby, M.R., Harris, K.M., Pinckney, A., DiMeglio, L.A., Rendell, M.S., Felner, E.I., Dostou, J.M., Gitelman, S.E., Griffin, K.J., Tsalikian, E., et al. (2015). Alefacept provides sustained clinical and immunological effects in new-onset type 1 diabetes patients. *J. Clin. Invest.* **125**, 3285–3296.
- Marek-Trzonkowska, N., Myśliwiec, M., Iwaszkiewicz-Grześ, D., Gliwiński, M., Derkowska, I., Żalińska, M., Zieliński, M., Grabowska, M., Zielińska, H., Piekarska, K., et al. (2016). Factors affecting long-term efficacy of T regulatory cell-based therapy in type 1 diabetes. *J. Transl. Med.* **14**, 332.
- Desreumaux, P., Foussat, A., Allez, M., Beaugerie, L., Hébuterne, X., Bouhnik, Y., Nachury, M., Brun, V., Bastian, H., Belmonte, N., et al. (2012). Safety and efficacy of antigen-specific regulatory T-cell therapy for patients with refractory Crohn’s disease. *Gastroenterology* **143**, 1207–1217.e2.
- Schmiedel, B.J., Singh, D., Madrigal, A., Valdovino-Gonzalez, A.G., White, B.M., Zapardiel-Gonzalo, J., Ha, B., Altay, G., Greenbaum, J.A., McVicker, G., et al. (2018). Impact of genetic polymorphisms on human immune cell gene expression. *Cell* **175**, 1701–1715.e16.
- Raj, T., Rothamel, K., Mostafavi, S., Ye, C., Lee, M.N., Replogle, J.M., Feng, T., Lee, M., Asinovski, N., Fröhlich, I., et al. (2014). Polarization of the effects of autoimmune and neurodegenerative risk alleles in leukocytes. *Science* **344**, 519–523.
- Ye, C.J., Feng, T., Kwon, H.-K., Raj, T., Wilson, M.T., Asinovski, N., McCabe, C., Lee, M.H., Fröhlich, I., Paik, H.-I., et al. (2014). Intersection of population variation and autoimmunity genetics in human T cell activation. *Science* **345**, 1254665.
- Westra, H.-J., Peters, M.J., Esko, T., Yaghootkar, H., Schurmann, C., Kettunen, J., Christiansen, M.W., Fairfax, B.P., Schramm, K., Powell, J.E., et al. (2013). Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat. Genet.* **45**, 1238–1243.
- Fairfax, B.P., Humburg, P., Makino, S., Naranbhai, V., Wong, D., Lau, E., Jostins, L., Plant, K., Andrews, R., McGee, C., et al. (2014). Innate immune activity conditions the effect of regulatory variants upon monocyte gene expression. *Science* **343**, 1246949.
- Wright, F.A., Sullivan, P.F., Brooks, A.I., Zou, F., Sun, W., Xia, K., Madar, V., Jansen, R., Chung, W., Zhou, Y.-H., et al. (2014). Heritability and genomics of gene expression in peripheral blood. *Nat. Genet.* **46**, 430–437.
- Chen, L., Ge, B., Casale, F.P., Vasquez, L., Kwan, T., Garrido-Martín, D., Watt, S., Yan, Y., Kundu, K., Ecker, S., et al. (2016). Genetic drivers of epigenetic and transcriptional variation in human immune cells. *Cell* **167**, 1398–1414.e24.
- Zhernakova, D.V., Deelen, P., Vermaat, M., van Iterson, M., van Galen, M., Arindrarto, W., van ’t Hof, P., Mei, H., van Dijk, F., Westra, H.-J., et al. (2017). Identification of context-dependent expression quantitative trait loci in whole blood. *Nat. Genet.* **49**, 139–145.
- Gamazon, E.R., Segrè, A.V., van de Bunt, M., Wen, X., Xi, H.S., Hormozdiari, F., Ongen, H., Konkashbaev, A., Derks, E.M., Aguet, F., et al. (2018). Using an atlas of gene regulation across 44 human tissues to inform complex disease- and trait-associated variation. *Nat. Genet.* **50**, 956–967.
- Roadmap Epigenomics Consortium; Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., et al. (2015). Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330.
- ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74.
- GTEX Consortium (2015). Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348**, 648–660.
- Kochi, Y., Myouzen, K., Yamada, R., Suzuki, A., Kurosaki, T., Nakamura, Y., and Yamamoto, K. (2009). FCRL3, an autoimmune susceptibility

- gene, has inhibitory potential on B-cell receptor-mediated signaling. *J. Immunol.* **183**, 5502–5510.
31. Giambartolomei, C., Vukcevic, D., Schadt, E.E., Franke, L., Hingorani, A.D., Wallace, C., and Plagnol, V. (2014). Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet.* **10**, e1004383.
  32. Mottet, C., Uhlig, H.H., and Powrie, F. (2003). Cutting edge: cure of colitis by CD4+CD25+ regulatory T cells. *J. Immunol.* **170**, 3939–3943.
  33. Akdis, M., Verhagen, J., Taylor, A., Karamloo, F., Karagiannidis, C., Cramer, R., Thunberg, S., Deniz, G., Valenta, R., Fiebig, H., et al. (2004). Immune responses in healthy and allergic individuals are characterized by a fine balance between allergen-specific T regulatory 1 and T helper 2 cells. *J. Exp. Med.* **199**, 1567–1575.
  34. Ling, E.M., Smith, T., Nguyen, X.D., Pridgeon, C., Dallman, M., Arbery, J., Carr, V.A., and Robinson, D.S. (2004). Relation of CD4+CD25+ regulatory T-cell suppression of allergen-driven T-cell activation to atopic status and expression of allergic disease. *Lancet* **363**, 608–615.
  35. Andersson, R., Gebhard, C., Miguel-Escalada, I., Hoof, I., Bornholdt, J., Boyd, M., Chen, Y., Zhao, X., Schmid, C., Suzuki, T., et al. (2014). An atlas of active enhancers across human cell types and tissues. *Nature* **507**, 455–461.
  36. Diogo, D., Bastarache, L., Liao, K.P., Graham, R.R., Fulton, R.S., Greenberg, J.D., Eyre, S., Bowes, J., Cui, J., Lee, A., et al. (2015). TYK2 protein-coding variants protect against rheumatoid arthritis and autoimmunity, with no evidence of major pleiotropic effects on non-autoimmune complex traits. *PLoS One* **10**, e0122271.
  37. Westra, H.-J., Martínez-Bonet, M., Onengut-Gumuscu, S., Lee, A., Luo, Y., Teslovich, N., Worthington, J., Martin, J., Huizinga, T., Klareskog, L., et al. (2018). Fine-mapping and functional studies highlight potential causal variants for rheumatoid arthritis and type 1 diabetes. *Nat. Genet.* **50**, 1366–1374.
  38. Guo, J., Zhang, J., Zhang, X., Zhang, Z., Wei, X., and Zhou, X. (2014). Constitutive activation of MEK1 promotes Treg cell instability in vivo. *J. Biol. Chem.* **289**, 35139–35148.
  39. Ferreira, M.A., Vonk, J.M., Baurecht, H., Marenholz, I., Tian, C., Hoffman, J.D., Helmer, Q., Tillander, A., Ullemar, V., van Dongen, J., et al. (2017). Shared genetic origin of asthma, hay fever and eczema elucidates allergic disease biology. *Nat. Genet.* **49**, 1752–1757.
  40. Burchill, M.A., Yang, J., Vogtenhuber, C., Blazar, B.R., and Farrar, M.A. (2007). IL-2 receptor beta-dependent STAT5 activation is required for the development of Foxp3+ regulatory T cells. *J. Immunol.* **178**, 280–290.
  41. Ochoa, D., Hercules, A., Carmona, M., Suveges, D., Gonzalez-Urriarte, A., Malangone, C., et al. (2021). Open Targets Platform: supporting systematic drug–target identification and prioritisation. *Nucleic Acids Res.* **49**, D1302–D1310.
  42. Negrotto, L., Farez, M.F., and Correale, J. (2016). Immunologic effects of metformin and Pioglitazone treatment on metabolic syndrome and multiple sclerosis. *JAMA Neurol.* **73**, 520–528.
  43. Duan, W., Ding, Y., Yu, X., Ma, D., Yang, B., Li, Y., Huang, L., Chen, Z., Zheng, J., and Yang, C. (2019). Metformin mitigates autoimmune insulinitis by inhibiting Th1 and Th17 responses while promoting Treg production. *Am. J. Transl. Res.* **11**, 2393–2402.
  44. Kontoyiannis, D., Boulougouris, G., Manoloukos, M., Armaka, M., Apostolaki, M., Pizarro, T., Kotlyarov, A., Forster, I., Flavell, R., Gaestel, M., et al. (2002). Genetic dissection of the cellular pathways and signaling mechanisms in modeled tumor necrosis factor-induced Crohn’s-like inflammatory bowel disease. *J. Exp. Med.* **196**, 1563–1574.
  45. Aguirre-Gamboa, R., de Klein, N., di Tommaso, J., Claringbould, A., van der Wijst, M.G., de Vries, D., Brugge, H., Oelen, R., Vösa, U., Zorro, M.M., et al. (2020). Deconvolution of bulk blood eQTL effects into immune cell subpopulations. *BMC Bioinformatics* **21**, 243.
  46. Nasrallah, R., Imianowski, C.J., Bossini-Castillo, L., Grant, F.M., Dogan, M., Placek, L., Kozhaya, L., Kuo, P., Sadiyeh, F., Whiteside, S.K., et al. (2020). A distal enhancer at risk locus 11q13.5 promotes suppression of colitis by Treg cells. *Nature* **583**, 447–452.
  47. Ongen, H., and Dermitzakis, E.T. (2015). Alternative splicing QTLs in European and African populations. *Am. J. Hum. Genet.* **97**, 567–575.
  48. Pai, A.A., Cain, C.E., Mizrahi-Man, O., De Leon, S., Lewellen, N., Veyrieras, J.-B., Degner, J.F., Gaffney, D.J., Pickrell, J.K., Stephens, M., et al. (2012). The contribution of RNA decay quantitative trait loci to inter-individual variation in steady-state gene expression levels. *PLoS Genet.* **8**, e1003000.
  49. Ye, C.J., Chen, J., Villani, A.-C., Gate, R.E., Subramaniam, M., Bhangale, T., Lee, M.N., Raj, T., Raychowdhury, R., Li, W., et al. (2018). Genetic analysis of isoform usage in the human anti-viral response reveals influenza-specific regulation of ERAP2 transcripts under balancing selection. *Genome Res.* **28**, 1812–1825.
  50. Nelson, M.R., Tipney, H., Painter, J.L., Shen, J., Nicoletti, P., Shen, Y., Floratos, A., Sham, P.C., Li, M.J., Wang, J., et al. (2015). The support of human genetic evidence for approved drug indications. *Nat. Genet.* **47**, 856–860.
  51. Finan, C., Gaulton, A., Kruger, F.A., Lumbers, R.T., Shah, T., Engmann, J., Galver, L., Kelley, R., Karlsson, A., Santos, R., et al. (2017). The druggable genome and support for target identification and validation in drug development. *Sci. Transl. Med.* **9**, eaag1166.
  52. Cannons, J.L., Lau, P., Ghumman, B., DeBenedette, M.A., Yagita, H., Okumura, K., and Watts, T.H. (2001). 4-1BB ligand induces cell division, sustains survival, and enhances effector function of CD4 and CD8 T cells with similar efficacy. *J. Immunol.* **167**, 1313–1324.
  53. DeBenedette, M.A., Shahinian, A., Mak, T.W., and Watts, T.H. (1997). Costimulation of CD28- T lymphocytes by 4-1BB ligand. *J. Immunol.* **158**, 551–559.
  54. Nagar, M., Jacob-Hirsch, J., Vernitsky, H., Berkun, Y., Ben-Horin, S., Amariglio, N., Bank, I., Kloog, Y., Rechavi, G., and Goldstein, I. (2010). TNF activates a NF-kappaB-regulated cellular program in human CD45RA- regulatory T cells that modulates their suppressive function. *J. Immunol.* **184**, 3570–3581.
  55. Marson, A., Kretschmer, K., Frampton, G.M., Jacobsen, E.S., Polansky, J.K., MacIsaac, K.D., Levine, S.S., Fraenkel, E., von Boehmer, H., and Young, R.A. (2007). Foxp3 occupancy and regulation of key target genes during T-cell stimulation. *Nature* **445**, 931–935.
  56. Schoenbrunn, A., Frensch, M., Kohler, S., Keye, J., Dooms, H., Moewes, B., Dong, J., Loddenkemper, C., Sieper, J., Wu, P., et al. (2012). A converse 4-1BB and CD40 ligand expression pattern delineates activated regulatory T cells (Treg) and conventional T cells enabling direct isolation of alloantigen-reactive natural Foxp3+ Treg. *J. Immunol.* **189**, 5985–5994.
  57. Lee, K., Nguyen, V., Lee, K.-M., Kang, S.-M., and Tang, Q. (2014). Attenuation of donor-reactive T cells allows effective control of allograft rejection using regulatory T cell therapy. *Am. J. Transpl.* **14**, 27–38.
  58. Danby, R.D., Zhang, W., Medd, P., Littlewood, T.J., Peniket, A., Rocha, V., and Roberts, D.J. (2016). High proportions of regulatory T cells in PBSC grafts predict improved survival after allogeneic haematopoietic SCT. *Bone Marrow Transpl.* **51**, 110–118.
  59. Trzonkowski, P., Bieniaszewska, M., Juścińska, J., Dobyszek, A., Krzystyniak, A., Marek, N., Myśliwska, J., and Hellmann, A. (2009). First-in-man clinical results of the treatment of patients with graft versus host disease with human ex vivo expanded CD4+CD25+CD127- T regulatory cells. *Clin. Immunol.* **133**, 22–26.
  60. Browning, B.L., Zhou, Y., and Browning, S.R. (2018). A one-Penny imputed genome from next-generation reference panels. *Am. J. Hum. Genet.* **103**, 338–348.
  61. Jun, G., Flickinger, M., Hetrick, K.N., Romm, J.M., Doheny, K.F., Abecasis, G.R., Boehnke, M., and Kang, H.M. (2012). Detecting and estimating contamination of human DNA samples in sequencing and array-based genotype data. *Am. J. Hum. Genet.* **91**, 839–848.
  62. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21.



63. Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30, 923–930.
64. Jiang, H., Lei, R., Ding, S.-W., and Zhu, S. (2014). Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads. *BMC Bioinformatics* 15, 182.
65. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760.
66. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome project data processing subgroup (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079.
67. Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., et al. (2008). Model-based analysis of ChIP-seq (MACS). *Genome Biol.* 9, R137.
68. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842.
69. Delaneau, O., Ongen, H., Brown, A.A., Fort, A., Panousis, N.I., and Dermizakis, E.T. (2017). A complete tool set for molecular QTL discovery and analysis. *Nat. Commun.* 8, 15452.
70. Leporcq, C., Spill, Y., Balaramane, D., Toussaint, C., Weber, M., and Bardet, A.F. (2020). TFmotifView: a webserver for the visualization of transcription factor motifs in genomic regions. *Nucleic Acids Res.* 48, W208–W217.
71. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 550.
72. Castel, S.E., Levy-Moonshine, A., Mohammadi, P., Banks, E., and Lappalainen, T. (2015). Tools and best practices for data processing in allelic expression analysis. *Genome Biol.* 16, 195.
73. UK10K Consortium; Walter, K., Min, J.L., Huang, J., Crooks, L., Memari, Y., McCarthy, S., Perry, J.R.B., Xu, C., Futema, M., et al. (2015). The UK10K project identifies rare variants in health and disease. *Nature* 526, 82–90.
74. Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* 10, 1213–1218.
75. Schmid, C., Rendeiro, A.F., Sheffield, N.C., and Bock, C. (2015). ChIPmentation: fast, robust, low-input ChIP-seq for histones and transcription factors. *Nat. Methods* 12, 963.
76. Hansen, K.D., Irizarry, R.A., and Wu, Z. (2012). Removing technical variability in RNA-seq data using conditional quantile normalization. *Biostatistics* 13, 204–216.
77. Kent, W.J., Zweig, A.S., Barber, G., Hinrichs, A.S., and Karolchik, D. (2010). BigWig and BigBed: enabling browsing of large distributed datasets. *Bioinformatics* 26, 2204–2207.
78. Alasoo, K. (2017). Wiggleplotr: Make Read Coverage Plots from Bigwig Files.
79. Alasoo, K., Rodrigues, J., Danesh, J., Freitag, D.F., Paul, D.S., and Gaffney, D.J. (2019). Genetic effects on promoter usage are highly context-specific and contribute to complex traits. *Elife* 8, e41673.
80. Davis, J.R., Fresard, L., Knowles, D.A., Pala, M., Bustamante, C.D., Battle, A., and Montgomery, S.B. (2016). An efficient multiple-testing adjustment for eQTL studies that accounts for linkage disequilibrium between variants. *Am. J. Hum. Genet.* 98, 216–224.
81. International Genetics of Ankylosing Spondylitis Consortium (IGAS); Cortes, A., Hadler, J., Pointon, J.P., Robinson, P.C., Karaderi, T., Leo, P., Cremin, K., Pryce, K., Harris, J., et al. (2013). Identification of multiple risk variants for ankylosing spondylitis through high-density genotyping of immune-related loci. *Nat. Genet.* 45, 730–738.
82. Demenais, F., Margritte-Jeannin, P., Barnes, K.C., Cookson, W.O.C., Altmüller, J., Ang, W., Barr, R.G., Beaty, T.H., Becker, A.B., Beilby, J., et al. (2018). Multiancestry association study identifies new asthma risk loci that colocalize with immune-cell enhancer marks. *Nat. Genet.* 50, 42–53.
83. Trynka, G., Hunt, K.A., Bockett, N.A., Romanos, J., Mistry, V., Szperl, A., Bakker, S.F., Bardella, M.T., Bhaw-Rosun, L., Castillejo, G., et al. (2011). Dense genotyping identifies and localizes multiple common and rare variant association signals in celiac disease. *Nat. Genet.* 43, 1193–1201.
84. International Multiple Sclerosis Genetics Consortium (IMSGC); Beecham, A.H., Patsopoulos, N.A., Xifara, D.K., Davis, M.F., Kempainen, A., Cotsapas, C., Shah, T.S., Spencer, C., Booth, D., et al. (2013). Analysis of immune-related loci identifies 48 new susceptibility variants for multiple sclerosis. *Nat. Genet.* 45, 1353–1360.
85. Cordell, H.J., Han, Y., Mells, G.F., Li, Y., Hirschfield, G.M., Greene, C.S., Xie, G., Juran, B.D., Zhu, D., Qian, D.C., et al. (2015). International genome-wide meta-analysis identifies new primary biliary cirrhosis risk loci and targetable pathogenic pathways. *Nat. Commun.* 6, 8019.
86. Tsoi, L.C., Spain, S.L., Knight, J., Ellinghaus, E., Stuart, P.E., Capon, F., Ding, J., Li, Y., Tejasvi, T., Gudjonsson, J.E., et al. (2012). Identification of 15 new psoriasis susceptibility loci highlights the role of innate immunity. *Nat. Genet.* 44, 1341–1348.
87. Okada, Y., Wu, D., Trynka, G., Raj, T., Terao, C., Ikari, K., Kochi, Y., Ohmura, K., Suzuki, A., Yoshida, S., et al. (2014). Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature* 506, 376–381.
88. Bentham, J., Morris, D.L., Graham, D.S.C., Pinder, C.L., Tomblinson, P., Behrens, T.W., Martín, J., Fairfax, B.P., Knight, J.C., Chen, L., et al. (2015). Genetic association analyses implicate aberrant regulation of innate and adaptive immunity genes in the pathogenesis of systemic lupus erythematosus. *Nat. Genet.* 47, 1457–1464.
89. Onengut-Gumuscu, S., Chen, W.-M., Burren, O., Cooper, N.J., Quinlan, A.R., Mychaleckyj, J.C., Farber, E., Bonnie, J.K., Szpak, M., Schofield, E., et al. (2015). Fine mapping of type 1 diabetes susceptibility loci and evidence for colocalization of causal variants with lymphoid gene enhancers. *Nat. Genet.* 47, 381–386.
90. Jin, Y., Andersen, G., Yorgov, D., Ferrara, T.M., Ben, S., Brownson, K.M., Holland, P.J., Birlea, S.A., Siebert, J., Hartmann, A., et al. (2016). Genome-wide association studies of autoimmune vitiligo identify 23 new risk loci and highlight key pathways and regulatory variants. *Nat. Genet.* 48, 1418–1424.
91. de Lange, K.M., Moutsianas, L., Lee, J.C., Lamb, C.A., Luo, Y., Kennedy, N.A., Jostins, L., Rice, D.L., Gutierrez-Achury, J., Ji, S.-G., et al. (2017). Genome-wide association study implicates immune activation of multiple integrin genes in inflammatory bowel disease. *Nat. Genet.* 49, 256–261.
92. Morris, A.P., Voight, B.F., Teslovich, T.M., Ferreira, T., Segre, A.V., Steinthorsdottir, V., Strawbridge, R.J., Khan, H., Grallert, H., Mahajan, A., et al. (2012). Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat. Genet.* 44, 981–990.
93. Howard, D.M., Adams, M.J., Shirali, M., Clarke, T.-K., Marioni, R.E., Davies, G., Coleman, J.R.I., Alloza, C., Shen, X., Barbu, M.C., et al. (2018). Genome-wide association study of depression phenotypes in UK Biobank identifies variants in excitatory synaptic pathways. *Nat. Commun.* 9, 1470.
94. Bakker, O.B., Aguirre-Gamboa, R., Sanna, S., Oosting, M., Smeekens, S.P., Jaeger, M., Zorro, M., Vösa, U., Withoff, S., Netea-Maier, R.T., et al. (2018). Integration of multi-omics data and deep phenotyping enables prediction of cytokine responses. *Nat. Immunol.* 19, 776–786.
95. Raudvere, U., Kolberg, L., Kuzmin, I., Arak, T., Adler, P., Peterson, H., and Vilo, J. (2019). g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res.* 47, W191–W198.

## STAR★METHODS

### KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<b>Antibodies</b>		
anti-CD4-APC, clone OKT4	BioLegend, San Diego, U.S.	Cat. no. 317416; RRID:AB_571945
anti-CD127-FITC, clone eBioRDR5	Thermo Fisher Scientific, Waltham, U. S.	Cat. no. 11-1278-42; RRID:AB_1907342
anti-CD25-PE, clone M-A251	BioLegend, San Diego, U.S.	Cat. no. 356104; RRID:AB_2561861
anti-FOXP3-BV421, clone 206D	BioLegend, San Diego, U.S.	Cat. no. 320123; RRID:AB_2561338
H3K4me3	Active Motif, Carlsbad, U.S.	Cat. no. 39915; RRID:AB_2687512
H3K27ac	Diagenode	Cat. no. C15410196; RRID:AB_2637079
<b>Biological samples</b>		
Lymphocyte cones were obtained with informed consent from healthy adults of Caucasian origin.	NHS Blood and Transplant, Cambridge and from the NHS Blood and Transplant, Oxford	REC 15/NW/0282, REC 15/NS/0060
<b>Chemicals, peptides, and recombinant proteins</b>		
TRIzol	Thermo Fisher Scientific	15596026
NEBNext® High-Fidelity 2X PCR Master Mix	New England Biolabs, Ipswich, U.S.	M0541L
Tn5 enzyme	Nextera	TDE1
EvaGreen dye	Biotium, Fremont, U.S.	#31000
<b>Critical commercial assays</b>		
EasySep® Human CD4 <sup>+</sup> T Cell Enrichment Kit	StemCell Technologies, Vancouver, Canada	Cat. no. 19052
iDeal ChIP-seq Kit for Histones	Diagenode, Liege, Belgium	C01010059
RNeasy Mini Kit	QIAgen, Hilden, Germany	74106
KAPA RNA HyperPrep Kit	Roche, Basel, Switzerland	KK8541
Nextera DNA Library Prep Kit	Illumina, U.S.	FC-131-1096
MinElute PCR Purification Kit	QIAgen, Hilden, Germany	28006
Nextera Index Kit	Illumina, U.S.	TG-131-2001
<b>Deposited data</b>		
Raw data generated in this study	EGA	<a href="https://www.ebi.ac.uk/ega/studies/EGAS00001003516">https://www.ebi.ac.uk/ega/studies/EGAS00001003516</a>
BLUEPRINT consortium CD4 <sup>+</sup> T cell and monocyte RNA-seq and ChIP-seq datasets	EGA	EGAD00001002671, EGAD00001002674, EGAD00001002673, EGAD00001002674
DICE project data	DICE project: Linking immune disease GWAS variants to genes and cell types, Date of approval: 2019-08-23	<a href="https://dice-database.org/">https://dice-database.org/</a>
FANTOM5	Predefined enhancer-TSS bed sets	<a href="http://enhancer.binf.ku.dk/presets/enhancer_tss_associations.bed">http://enhancer.binf.ku.dk/presets/enhancer_tss_associations.bed</a>
Custom scripts and pipelines repository:	Treg_Multiomic	<a href="https://github.com/trynkaLab/">https://github.com/trynkaLab/</a>
<b>Software and algorithms</b>		
GitHub (original codes supporting this work)	<a href="https://doi.org/10.5281/zenodo.6335757">https://doi.org/10.5281/zenodo.6335757</a>	<a href="https://github.com/trynkaLab/Treg_Multiomics/tree/v1.0.1">https://github.com/trynkaLab/Treg_Multiomics/tree/v1.0.1</a>
BEAGLE 4.1	Browning et al. <sup>60</sup>	<a href="http://faculty.washington.edu/browning/beagle/beagle.html">http://faculty.washington.edu/browning/beagle/beagle.html</a>
VerifyBamID v1.0.0	Jun et al. <sup>61</sup>	<a href="https://github.com/statgen/verifyBamID/releases">https://github.com/statgen/verifyBamID/releases</a>

(Continued on next page)



**Continued**

REAGENT or RESOURCE	SOURCE	IDENTIFIER
STAR	Dobin et al. <sup>62</sup>	<a href="https://github.com/alexdobin/STAR/releases">https://github.com/alexdobin/STAR/releases</a>
subread package v1.5.1	Liao et al. <sup>63</sup>	<a href="http://subread.sourceforge.net/">http://subread.sourceforge.net/</a>
skewer	Jiang et al. <sup>64</sup>	<a href="https://github.com/relipmoc/skewer">https://github.com/relipmoc/skewer</a>
bwa	Li and Durbin. <sup>65</sup>	<a href="http://bio-bwa.sourceforge.net/">http://bio-bwa.sourceforge.net/</a>
samtools	Li et al. <sup>66</sup>	<a href="http://samtools.sourceforge.net/">http://samtools.sourceforge.net/</a>
MACS2	Zhang et al. <sup>67</sup>	<a href="https://github.com/macs3-project/MACS">https://github.com/macs3-project/MACS</a>
BEDTOOLS	Quinlan and Hall <sup>68</sup>	<a href="https://bedtools.readthedocs.io/en/latest/">https://bedtools.readthedocs.io/en/latest/</a>
QTLtools	Delaneau et al. <sup>69</sup>	<a href="https://qtltools.github.io/qtltools/">https://qtltools.github.io/qtltools/</a>
coloc v2.3-1	Giambartolomei et al. <sup>31</sup>	<a href="https://github.com/chr1swallace/coloc">https://github.com/chr1swallace/coloc</a>
TFmotifView	Leporcq et al. <sup>70</sup>	<a href="http://bardet.u-strasbg.fr/tfmotifview/">http://bardet.u-strasbg.fr/tfmotifview/</a>
g:Profiler	Raudvere et al. <sup>66</sup>	<a href="https://biit.cs.ut.ee/gprofiler/gost">https://biit.cs.ut.ee/gprofiler/gost</a>
DESeq2_1.14	Love et al. <sup>71</sup>	<a href="https://bioconductor.org/packages/release/bioc/html/DESeq2.html">https://bioconductor.org/packages/release/bioc/html/DESeq2.html</a>
ASEReadCounter (4.0.1.1)	Castel et al. <sup>72</sup>	<a href="https://gatk.broadinstitute.org/hc/en-us/articles/360037054312-ASEReadCounter">https://gatk.broadinstitute.org/hc/en-us/articles/360037054312-ASEReadCounter</a>
<b>Other</b>		
Lympholyte-H density gradient centrifugation.	(Cedarlane Labs, Burlington, Canada)	CL5020
Infinium® CoreExome-24 v1.1 BeadChip	Illumina	WG-331-1101

**RESOURCE AVAILABILITY**

**Lead contact**

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Gosia Trynka ([gosia@sanger.ac.uk](mailto:gosia@sanger.ac.uk)).

**Materials availability**

This study did not generate new unique reagents.

**Data and code availability**

- All raw data produced here and existing data from multiple sources have the accession numbers listed in the [key resources table](#). Access to DICE project data was authorized via dbGaP to Dr. Gosia Trynka (Project: Linking immune disease GWAS variants to genes and cell types, Date of approval: 2019-08-23). We used FANTOM5 predefined enhancer-TSS bed sets. The gene expression dataset is integrated into the eQTL catalogue (<https://www.ebi.ac.uk/eqtl/>). All colocalization results can be browsed via this website: <https://www.sanger.ac.uk/science/tools/treg-colocalisation/treg-colocalisation/>
- All original code has been deposited at Zenodo and is publicly available as of the date of publication. DOIs are listed in the [key resources table](#).
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

**METHOD DETAILS**

**Sample collection and Treg isolation**

Lymphocyte cones were obtained with informed consent from donors at the NHS Blood and Transplant, Cambridge (REC 15/NW/0282) and from the NHS Blood and Transplant, Oxford (REC 15/NS/0060).

Leukodepletion cones were obtained from healthy adults of Caucasian origin. PBMCs were isolated using Lympholyte-H (Cedarlane Labs, Burlington, Canada) density gradient centrifugation. CD4<sup>+</sup> T cells fraction of the PBMCs was obtained by negative selection using EasySep® Human CD4<sup>+</sup> T Cell Enrichment Kit (Cat. no. 19052, StemCell Technologies, Vancouver, Canada), following the manufacturer's instructions. Next, the CD4<sup>+</sup> T cells were resuspended in the FACS staining buffer (2 mM EDTA and 0.5% FCS in PBS) at 10<sup>8</sup> cells per mL. The cells were stained with the following antibody cocktail: anti-CD4-APC (30 μL/mL final volume, clone OKT4, Cat. no. 317416, BioLegend, San Diego, U.S.), anti-CD127-FITC and (30 μL/mL, clone eBioRDR5, Cat.

no.11-1278-42, Thermo Fisher Scientific, Waltham, U. S.) and anti-CD25-PE (80  $\mu\text{L}/\text{mL}$ , clone M-A251, Cat. no. 356104, BioLegend) for at least 30 min at RT in the darkness. The cells were washed copiously with FACS buffer and resuspended at  $10^8$  cells per mL in full medium (IMDM, 10% FCS) and kept overnight at  $4^\circ\text{C}$ . Immediately before sorting, the cells were stained with DAPI, to discriminate between live and dead cells (Figure S1G). The  $\text{CD4}^+$ ,  $\text{CD25}^{\text{high}}$ ,  $\text{CD127}^{\text{neg}}$  population corresponding to Treg lymphocytes was sorted out for the downstream assays (Figures S1A and S1B). We sorted up to 3 million cells in order to carry out all of the downstream assays. In instances where this number was not reached we prioritised RNA-seq, followed by H3K27ac and H3K4me3 ChIP-seq, and finally ATAC-seq.

### Sample summary

For all donors we were able to extract their sex based on their genotype and for 113 of the 124 donors we had access to their age (Figure S1D). The majority of the donors (78%) were genetically assigned males and were aged over 57 years of age ( $\pm 11$ ).

### FACS staining

To verify the FOXP3 expression in the sorted Treg populations after sorting, the cells were stained for expression of CD4, CD25 and CD127 surface markers, and then stained with anti-FOXP3-BV421 antibody (5  $\mu\text{L}/10^6$  cells, clone 206D, BioLegend) using the eBioscience™ Foxp3/Transcription Factor Staining Buffer Set (Thermo Fisher Scientific), according to the manufacturer's instructions. We observed that the sorted cells were on average 80% FOXP3 positive (Figures S1B and S1E).

To define the proportions of memory and naive cells in the  $\text{CD4}^+$  population, an aliquot of  $10^6$  cells after the CD4-enrichment were resuspended in 100  $\mu\text{L}$  FACS buffer and stained with a cocktail of anti-CD4-APC and anti-CD127-FITC antibodies (3  $\mu\text{L}$  each), anti-CD25-PE (8  $\mu\text{L}$ ) and anti-CD45RA-BV785 (4  $\mu\text{L}$ , clone HI100, Cat. no. 304140, BioLegend), incubated at RT in the dark for at least 30 min, washed copiously with FACS buffer and analysed on BD Fortessa. The majority of the isolated Tregs were memory Tregs (median = 79%) (Figure S1F).

### Culture and stimulation of isolated Tregs

Whole blood samples were obtained from ten healthy adults, aged from 22 to 39 years. Live regulatory T cells ( $\text{CD4}^+$   $\text{CD25}^{\text{high}}$   $\text{CD127}^{\text{low}}$ ) were isolated as described in [Sample collection and Treg isolation](#). Cells were grown in Iscove's Modified Dulbecco's Media (IMDM) (Life Technologies, Paisley, UK), supplemented with 10% human serum (HS), 50 U/mL penicillin and streptomycin (Life Technologies) and 100 U/mL recombinant human IL-2 and incubated at  $37^\circ\text{C}$  in a humidified atmosphere of 5%  $\text{CO}_2$ . Cells were activated using PMA (5-10  $\text{ng}/\mu\text{L}$ ) with ionomycin (200  $\text{ng}/\mu\text{L}$ ) (Sigma-Aldrich) overnight (18 hours).

### SNP genotyping and imputation

A total of 551,839 genetic markers were genotyped using the Infinium® CoreExome-24 v1.1 BeadChip by Illumina. After SNP QC (MAF >10%, SNP call rate >95%, Hardy-Weinberg equilibrium (HWE) p value < 0.001) we retained 243,820 variants in our dataset. Samples with call rate <95% were removed from the analysis. After quality control per individual, the total genotyping call rate reached >99%. We performed imputation using BEAGLE 4.1 with a reference panel comprising the 1000 Genomes Phase 3<sup>60</sup> and the UK10K<sup>73</sup> samples (modelscale parameter = 2). Following imputation we required allelic R-squared ( $\text{AR}2$ )  $\geq 0.8$ , HWE p value < 0.001, and MAF >5% in both the analysed cohort and in the reference panel. We excluded 1,934 multiallelic polymorphisms from further analysis which resulted in 5,761,739 variants in our final dataset. Of those, 617,318 were insertion-deletions (INDELs). All genetic variant coordinates were lifted over to GRCh38.

Our samples clustered with the European populations included in the 1000 Genomes project (Figure S2F). We removed 1 sample due to high relatedness (identity by state,  $\text{pi}_{\text{hat}} > 0.2$ ). We used VerifyBamID v1.0.0<sup>61</sup> with the genotype information along with all the functional genomics sequencing assays (see below) to verify no sample swaps were present in the final dataset.

### RNA-seq

For RNA-seq experiments,  $0.5 \times 10^6$  sorted Treg cells were washed with ice-cold PBS and resuspended in TRIzol (Thermo Fisher Scientific). After a standard phenol/chloroform isolation step, the total RNA contained in the upper, aqueous phase was further purified with RNeasy Mini Kit (QIAGEN, Hilden, Germany), according to the manufacturer's instructions. The RNA libraries were constructed using KAPA RNA Hyper-Prep Kit (Roche, Basel, Switzerland), following a standard automated protocol. The libraries were multiplexed and sequenced at 75 bp PE on an Illumina HiSeq V4 to yield on average 57 million reads per sample.

### ATAC-seq

ATAC-seq was performed according to protocol,<sup>74</sup> with the following modifications. After sorting, the T cells were washed with ice-cold PBS and resuspended in sucrose buffer (10 mM Tris pH 8, 3 mM  $\text{CaCl}_2$ , 2 mM  $\text{MgOAc}$ , 1 mM DTT, 0.32 M sucrose, 0.5 mM EDTA, 0.25% TritonX-100), followed by 5 min incubation on ice to isolate the nuclei. Isolated nuclei were washed once with 1x TD buffer (Tagment DNA Buffer, Nextera DNA Library Prep Kit, Illumina, U.S) and resuspended in 50  $\mu\text{L}$  1x TD buffer containing 2.5  $\mu\text{L}$  of Tn5 enzyme (TDE1, Nextera). The reaction was carried out at  $37^\circ\text{C}$ , mixing and then stopped by addition of 250  $\mu\text{L}$  of buffer PB (MinElute PCR Purification Kit, QIAGEN, Hilden, Germany). The DNA was then purified on MinElute columns according to the manufacturer's instructions and eluted in 10  $\mu\text{L}$  sterile  $\text{ddH}_2\text{O}$ . The libraries were amplified using the NPM mix (Nextera

PCR Master Mix from Nextera DNA Library Prep Kit) and Index adapters i7 and i5 (Nextera Index Kit, Illumina, U.S.), according to the manufacturer's instructions. The number of amplification PCR cycles for each sample was determined individually by performing a qPCR reaction of 7.5  $\mu$ L aliquote of the mix with an addition of the EvaGreen dye (Biotium, Fremont, U.S.). The amplified libraries were SPRI purified (upper cut 0.5x, lower cut 1.8 x) on a Zephyr G3 SPE Workstation (PerkinElmer, Waltham, U.S.), multiplexed and sequenced at 75 bp PE on an Illumina HiSeq V4 to yield on average 112 million reads per sample.

### H3K4me3 and H3K27ac ChIPmentation-seq

The ChIPmentation-seq (ChM-seq) protocol was performed on 100,000 sonicated cells according to the protocol presented in Schmidl et al.<sup>75</sup> and adapted to work with the iDeal ChIP-seq Kit for Histones (Diagenode, Liege, Belgium).

After sorting, the cells were resuspended in pre-warmed full medium (IMDM, 10% FCS) at 1-2 million cells per mL and allowed to recover in the incubator (37°C, 5% CO<sub>2</sub>) for at least 30 min. The cells were then fixed by addition of formaldehyde to medium to a final concentration of 1% and 5 min incubation at 37°C, followed by quenching with glycine for 5 min at a final concentration of 125 mM min at RT with mixing. The cross-linked cells were subsequently washed twice with ice-cold PBS and snap-frozen by immersion in liquid nitrogen.

0.5  $\times$  10<sup>9</sup> frozen cells were resuspended in 250  $\mu$ L buffer iL1 with proteinase inhibitors cocktail (iDeal ChIP-seq Kit for Histones, Diagenode) and incubated for 10 min at 4°C on the Bohemian wheel. The samples were then spun down, and resuspended first in buffer iL2 with proteinase inhibitors, then in iS1 with proteinase inhibitors, in both cases also for 10 min at 4°C. The cells were then sonicated in buffer iS1 using the Bioruptor® Pico sonication device (Diagenode) to achieve fragment sizes distribution below 3 kb.

Sonicated chromatin from 100,000 cells was used for an overnight immunoprecipitation reaction with 1  $\mu$ g of antibody, either against H3K4me3 (Catalog No: 39915, Active Motif, Carlsbad, U.S.) or H3K27ac (Cat. no. C15410196, Diagenode).

The samples in deep-well plates were then washed twice for two minutes with 150  $\mu$ L of each of the buffers: iW1, iW2, iW3 (iDeal ChIP-seq Kit for Histones, Diagenode) and then with 10 mM Tris pH 8. All the washes in this protocol were performed using an Agilent Bravo Automated Liquid Handling Platform (Agilent, Santa Clara, U.S.). After the second Tris wash, a ChIPmentation reaction on the beads was conducted following the protocol outlined in Schmidl et al. Briefly, a mix containing 1  $\mu$ L Tn5 from the Nextera kit was added to the beads and incubated for 10 minutes with vigorous mixing at 37°C. Next, the reaction mix was removed using Bravo, and additional washes were performed, two with buffer iW3, followed by two washes with buffer iW4. The enriched DNA was eluted from the beads by incubation with 67  $\mu$ L buffer iE1 (1 h, RT, vigorous shaking). 3  $\mu$ L of iE2 buffer were then added to each sample and the cross-linking was reversed by an overnight incubation at 65°C in a thermocycler.

The DNA was then purified twice using SPRI beads at 1.6x ratio using a Zephyr G3 SPE Workstation. The libraries were amplified following the ATAC-seq library amplification protocol, but using NEBNext® High-Fidelity 2X PCR Master Mix (New England Biolabs, Ipswich, U.S.). Finally, the ChIPmentation libraries were sequenced to a depth of at least 13 million reads per sample and an average of 75 million reads per sample.

## QUANTIFICATION AND STATISTICAL ANALYSIS

### RNA-seq data processing

Reads were aligned to the GRCh38 human reference genome using STAR<sup>62</sup> and the Ensembl reference transcriptome (version 87). Gene counts were performed using featureCounts tools from the subread package v1.5.1<sup>63</sup> and only assigned reads were used for further processing (59.26% of reads were assigned; Figure S2D). We excluded short RNAs and pseudogenes from the analysis. We quantile normalised the gene expression values and corrected for GC-content using the CQN method.<sup>76</sup> We kept 12,059 genes with average count per gene across all donors greater than 25.

### Chromatin marks data processing

Reads were trimmed using skewer<sup>64</sup> and aligned to the GrCh38 assembly of the human genome using bwa<sup>65</sup> and employing the mem algorithm. Multi-mapping and duplicated reads were removed using samtools.<sup>66</sup> For ATAC-seq data, reads aligning to the mitochondrial chromosome were also removed. Only reads mapping to autosomes were maintained. A median of 30, 27 and 40 million reads in the ATAC-seq, H3K4me3 and H3K27ac passed this QC, respectively.

Peak calling was performed using MACS2<sup>67</sup> independently on each donor for quality control purposes. For ATAC-seq peaks were called using the standard MACS2 model and specifying `-nomodel -shift -25 -extsize 50` on fragment BED files (this is, both reads of a pair were merged into a single fragment). We generated a combined treatment set per histone mark by merging an equal number of reads per donor to reach the combined merged input size of 223 million reads. H3K4me3 peaks were called using the standard narrow peak MACS2 model, specifying `-f BAMPE -keep-dup all`, then we selected only the peaks with q-value < 0.01 and fold-change greater than 2. H3K27ac broad peaks were called using the standard broad peaks macs2 model, specifying `-f BAMPE -broad -nomodel -extsize 146 -keep-dup all`, then we selected only the peaks with q-value < 0.001 and fold-change greater than 2.

Samples with less than 10,000 peaks (median: ATAC 36,331, H3K4me3 22,815, H3K27ac 68,626), fraction of reads in peaks (FRiP) lower than 10% (median: ATAC 23%, H3K4me3 52.64%, H3K27ac 63.9%) (Figure S2D), or, for ATAC-seq, an abnormal insert profile (defined as a ratio of short inserts (<150 bp) over long inserts (>150) smaller than 1.5; average 2.03) were discarded. Additionally, the samples that did not cluster with the corresponding group in principal component analysis (considering log2 transformed number of

reads in genomic bins of 10,000 bp, after normalization by library length) were discarded from further analysis. Finally, a total of 73 (62%), 88 (79%) and 91 (78%) individuals passed these filters for ATAC, H3K4me3 and H3K27ac samples, respectively. Sixty-two donors passed QC steps for all the tested genomic assays (RNA, ATAC, H3K4me3 and H3K27ac).

In order to define a consensus set of peaks per chromatin assay, we performed a merged peak calling combining reads from all the donors. We downsampled each donor sample using samtools to 2 million fragments per ATAC-seq assay, 1.87 million read pairs per H3K4me3 and 1.86 million read pairs per H3K27ac assay in order to reach similar read counts to the sequenced inputs. We used the MACS2 parameters described above and specified `-keep-dup all`. Then, to ensure a sufficient number of reads per peak, only ATAC-seq peaks with at least 10 reads in 80% of the samples, and ChM-seq peaks with fold enrichment  $\geq 2$  and adjusted p value  $< 0.001$ , were maintained in the final set. The consensus sets were 39,642 ATAC-seq narrow peaks, 40,285 H3K4me3 ChM-seq narrow peaks and 34,457 H3K27ac broad peaks. The peak overlap between the assays as calculated using bedtools intersect, the distance to the closest transcription start site (TSS) is shown in Figures S2A and S2B). The median length was 523 bp, 794 bp and 4501.5 bp for the ATAC-seq peaks, H3K4me3 peaks and H3K27ac peaks, respectively. The median number of read pairs in each peak (calculated using featureCounts `-p -C -D 5000 -d 50`) per sample amounted 37.95 in ATAC-seq, 16.34 in H3K4me3 ChM-seq and 123.38 in H3K27ac ChM-seq (Figure S2C).

We assayed 17, 8, 15 and 15 samples twice (same donors were recruited at two different time points) for RNA, ATAC-seq, H3K27ac ChM-seq and H3K4me3 ChM-seq, respectively. We observed high correlation between both technical (same donors different times) and biological replicates (different donors) ( $R^2 > 0.8$ ) in all assays. We observed greater correlation between technical than between biological replicates, as expected (Figure 2E).

In order to compare QTL effects between the Tregs and the naive CD4 T cells and monocytes in the BLUEPRINT dataset, we combined the H3K27ac broad peaks called independently in each cell type into a consensus set of peaks. Overlapping peaks were merged using the merge option implemented in bedtools.<sup>68</sup>

Genome browser data was constructed using the MACS2 `-B` flag and reads were normalised to signal per million. The fold-enrichment was calculated using the input background and finally bigwigs were constructed using bedGraphToBigWig command from the UCSC suite of tools.<sup>77</sup> Coverage plots were generated using an adapted version of the wiggleplot R Bioconductor package.<sup>78</sup>

### Quantitative trait locus mapping (QTLs)

Prior to the QTL analysis we removed genes and peaks mapping to the MHC region (chr6: 20,000,000-40,000,000) and only kept the autosomal chromosomes. We used linear regression implemented in the QTLtools<sup>69</sup> software to map cis QTLs. For the gene expression we used a 500 kbp cis-window around the gene, while for the three chromatin mark assays we used a 100 kbp cis-window around the defined peak. In chromQTL mapping we were directly assaying the chromatin features and we were focusing specifically on those loci with chromQTL SNPs located in the controlled peaks, therefore we reasoned that a smaller window was more appropriate.<sup>79</sup> As covariates we used the top 13, 30, 22 and 33 principal components that each explained up to 1% of the observed variance in the RNA, ATAC, H3K27ac and H3K4me3, respectively. We used the `"-permute 10000"` to obtain permutation p-values for the top most significantly associated variant for each gene or peak. We then used eigenMT<sup>80</sup> to correct for the number of genes or peaks tested and used a cut-off of 5% FDR, as determined by power analysis (Figure S7).

To perform comparative analysis between Treg and monocytes and naive T cell eQTLs and actQTLs we downloaded the RNA-seq and ChIP-seq datasets generated by the BLUEPRINT consortium<sup>24</sup> from EGA (EGA: EGAD00001002671, EGAD00001002674, EGAD00001002673, EGAD00001002674) and processed the data using the same workflow as described above. We included the top 16 and 14 PCs for these datasets in the monocytes and naive T cells eQTL analyses respectively. We included the top 13 and 17 PCs for these datasets in the monocytes and naive T cells actQTL analyses, respectively. We chose this dataset because Tregs and naive T cells are closely related cells of adaptive immunity, while monocytes fall into a more distant cell type of the innate immune arm. Furthermore, the datasets are of similar size (197 and 169 individuals) to the Treg dataset and all the individuals were of British origin.

We used the following three criteria to define an eQTL and actQTL as cell type specific when comparing monocytes, naive and regulatory T cells: (i) the gene was expressed in one cell type only or the peak was only present in one cell type, (ii) the gene or peak was a significant QTL in one cell type ( $FDR \leq 0.05$ ) and not in the other ( $FDR > 0.2$ ) and (iii) if the same gene or peak was a QTL in both cell types and the LD between the top QTL variant in regulatory T cells and any of the significant associated signals in naive T cells was lower than  $R^2 < 0.2$ .

### Allele-specific expression analysis

We used ASEReadCounter<sup>72</sup> from the Genome Analysis ToolKit (GATK) to count the number of allele-specific fragments overlapping each variant in the RNA-seq data. We used `'-U ALLOW_N_CIGAR_READS -dt NONE -minMappingQuality 10 -rf MateSameStrand'`. We filtered out variants covered by less than 8 reads, as well as variants that fell within regions with low mappability or that displayed mapping bias using simulated data, as outlined in Castel et al.<sup>72</sup> annotated each variant to its overlapping gene. Significant allele-specific events were calculated using a binomial test, where the null was defined by taking the average reference ratio across all heterozygous sites for each sample.

### Colocalization of QTL signals with immune disease GWAS

We used *coloc* v2.3-1<sup>31</sup> with the default priors to test for colocalization between molecular QTLs and GWAS SNPs listed in [Tables S2](#) and [S3](#). We included in our analysis the summary stats for the genome-wide association studies of 14 immune related diseases: allergic diseases (ALL),<sup>39</sup> ankylosing spondylitis (AS),<sup>81</sup> asthma (AST),<sup>82</sup> celiac disease (CEL),<sup>83</sup> multiple sclerosis (MS),<sup>84</sup> primary biliary cirrhosis (PBC),<sup>85</sup> psoriasis (PS),<sup>86</sup> rheumatoid arthritis (RA),<sup>87</sup> systemic lupus erythematosus (SLE),<sup>88</sup> type 1 diabetes (T1D),<sup>89</sup> vitiligo (VIT),<sup>90</sup> inflammatory bowel disease (IBD), Crohn's disease (CD) and ulcerative colitis (UC).<sup>91</sup> We selected these diseases because they had more than 40 GWAS associated independent loci at  $p$ -value  $< 10^{-5}$ . As controls we tested two non-immunological traits with a similar number of loci, type-2 diabetes (T2D)<sup>92</sup> and depression (DEP).<sup>93</sup> CD, UC and IBD were counted as a single disease when counting for number of colocalizing diseases per gene or peak. Similarly for ALL and AST. *Coloc* tests five hypotheses for colocalization. Hypothesis zero (PP.H0) tests whether there is any association at all, PP.H1 and PP.H2 test whether there is an association with just one or the other study, PP.H3 tests whether the signal from GWAS and QTL is due to two independent SNPs, and PP.H4, test if the association between GWAS and QTL is due to a shared causal variant.

Prior to colocalization, we repeated the QTL mapping in chromatin features using a 500 kbp window, to run *coloc* at a larger window. We ran *coloc* on a 400-kb region centered on each lead eQTL and chromQTL variant that was less than 100 kb away from a GWAS variant (nominal  $p$  value  $< 10^{-5}$ ). We only kept the colocalizations between QTLs and non-HLA GWAS loci if there were more than 50 SNPs tested. To claim a true colocalizing signal we required PP.H4 to be equal or greater than 0.8<sup>3</sup>. In order to decrease the number of false positive findings in our Treg dataset, we focused on the colocalization results with common immune disease variants (MAF  $> 10\%$ ). Colocalizations between Treg QTLs and disease GWAS signals were the lead QTL variants and the lead GWAS variant had  $R^2$  LD  $< 0.5$  were discarded. For GWAS loci with  $10^{-5} > p$  value  $> 5 \times 10^{-8}$ , and colocalizing with Treg QTLs we verified in the original publications that there was a replication cohort and the final GWAS  $p$  value was lower than genome wide significance.

### Gene expression variance deconvolution

To estimate the contribution of the genetic component and chromatin marks to the transcriptome variance we fitted a multivariate linear model to the expression of each gene. Therefore, the dependent variable in the model was the gene expression, while the independent variables were the genetic variants and/or chromatin mark. We regressed out the PCs included in the analysis of each chromatin mark previous to model calculation. As described in de Bakker et al.,<sup>94</sup> we performed an initial variable selection step by identifying the genetic variants or chromatin features in a  $\pm 150$  kb window that significantly correlated with gene expression (Spearman  $p$  value  $< 0.05$ ). In order to keep only independent variables in the set of predictors, in the instances where pairs of genetic variants or chromatin features were correlated (Spearman correlation  $> 0.4$ ), we removed the variable with a lower correlation with gene expression. To determine the total variance explained we used the adjusted  $R^2$  of the model where we included all the independent variables from genetic variants and chromatin features. The individual contribution of genetic variants or each chromatin mark, or combination of the genetic variants and chromatin marks, was obtained by subtracting the  $R^2$  estimates of the models that excluded genetic variants or individual chromatin marks, or their combinations, from the  $R^2$  of the model with the total variance explained by the combination of genetic variants and all chromatin marks.

### SNP functional annotation

All the lead variants (and their proxies,  $R^2 > 0.8$ ) of every significant caQTLs, actQTLs, and promQTLs were annotated if a variant was overlapping an ATAC, H3K27ac or H3K4me3 peak ([Figure 1D](#) shows the number of polymorphisms included in the different categories, categories are mutually exclusive and a variant is assigned to the category with most functional support).

### LD loci definition and classification

A LD locus comprises a  $\pm 150$  kb window around the region defined by the lead and proxy variants ( $R^2 > 0.8$ ) for each GWAS signal that colocalizes with a Treg QTL. Loci with significant colocalization signals were classified as: i) Tier 1, at least one colocalization with an eQTL and at least one colocalization with a chromQTL; ii) Tier 2, at least one colocalization with an eQTL, no colocalization with chromQTL; and iii) Tier 3, at least one colocalization with a chromQTL ([Figure 3](#); [Tables S2](#) and [S3](#)). Loci in Tier 1 and Tier 3 in which variants overlapped with the regulated (i.e. colocalizing) chromatin QTL peak were prioritized as functional.

### Known drug tractability evidence analysis

We used the Open Targets Platform to extract information on drugs that support target-disease associations provided by ChEMBL.<sup>41</sup> We retrieved the 'known\_drug' evidence for all genes via the Open Targets API using the python client (Open Targets data release February 2020). When gathering data from OpenTargets we summarized the extracted data based on the website's recommendations (<https://platform-docs.opentargets.org/target/tractability>).

### Differential gene expression analysis

RNA-seq reads were obtained and processed as described in [RNA-seq](#) and in [RNA-seq data processing](#). Genes with at least 25 copies in at least three samples were kept, for a final table of 16,645 genes. Differential expression analysis was performed using



DESeq2\_1.14<sup>71</sup> Wald test, setting alpha at 0.05 and lfc at 1. Differentially expressed genes that mapped within a +/- 150 kb window from a chromQTL in Tier 3 loci were annotated as candidate stimulation specific eQTL genes.

#### **FANTOM5 CAGE data integration**

The CAGE data in FANTOM5<sup>35</sup> provides predefined human enhancer-TSS pair sets (note that the use of one enhancer can be correlated with the TSS expression levels of several genes and vice-versa). Using “bedtools intersect” we defined the enhancers that overlapped Treg colocalizing chromQTL peaks. Then, we linked the Treg chromQTLs to the corresponding TSS.

#### **Transcription factor binding site (TFBS) enrichment analysis**

We used TFmotifView,<sup>70</sup> which enables the user to input a set of chromosomal regions and perform a TFBS enrichment analysis on all TFs included in the JASPAR2020 database. This tool uses a custom set of regions as background. We used all Treg colocalizing actQTLs and compared them to actQTLs shared with naive CD4<sup>+</sup> T cells. We followed-up on the set of enriched TFs by using g:Profiler,<sup>95</sup> to identify enriched pathways.

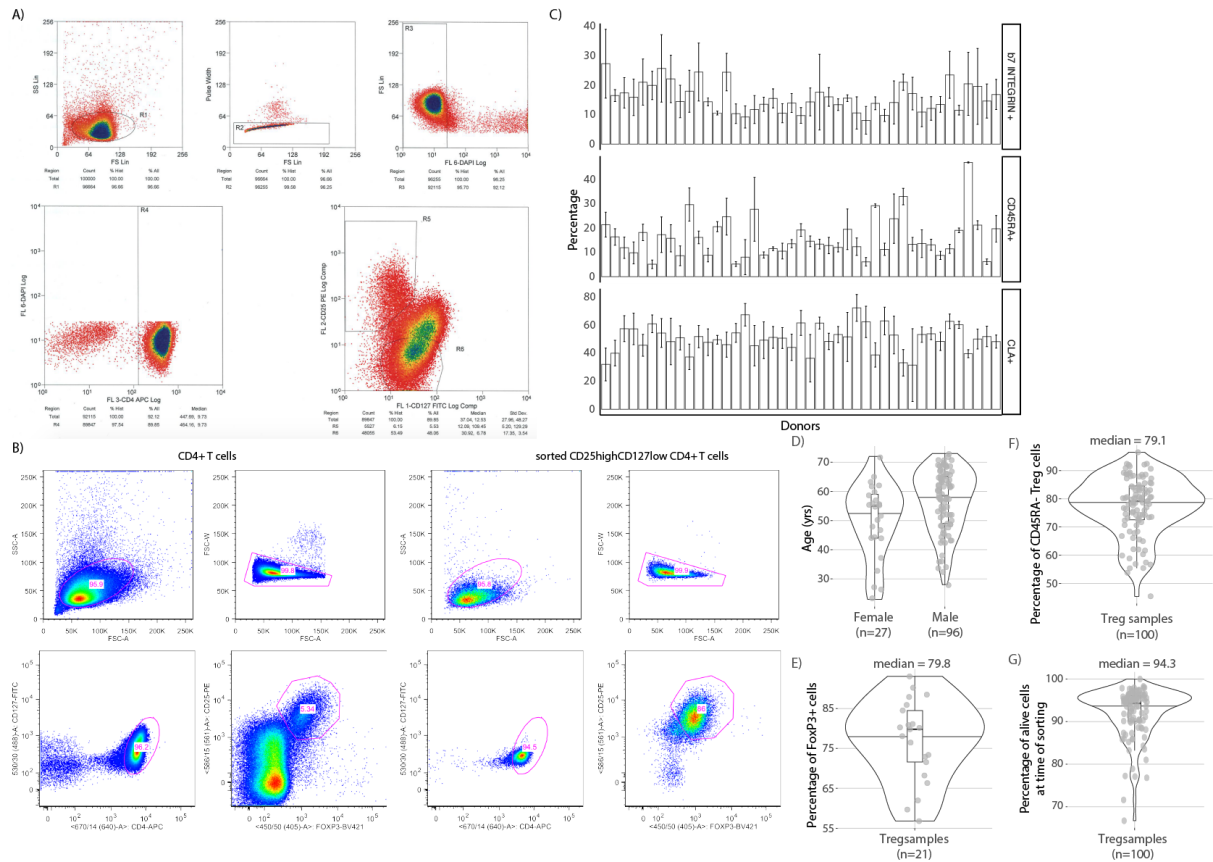


**Supplemental information**

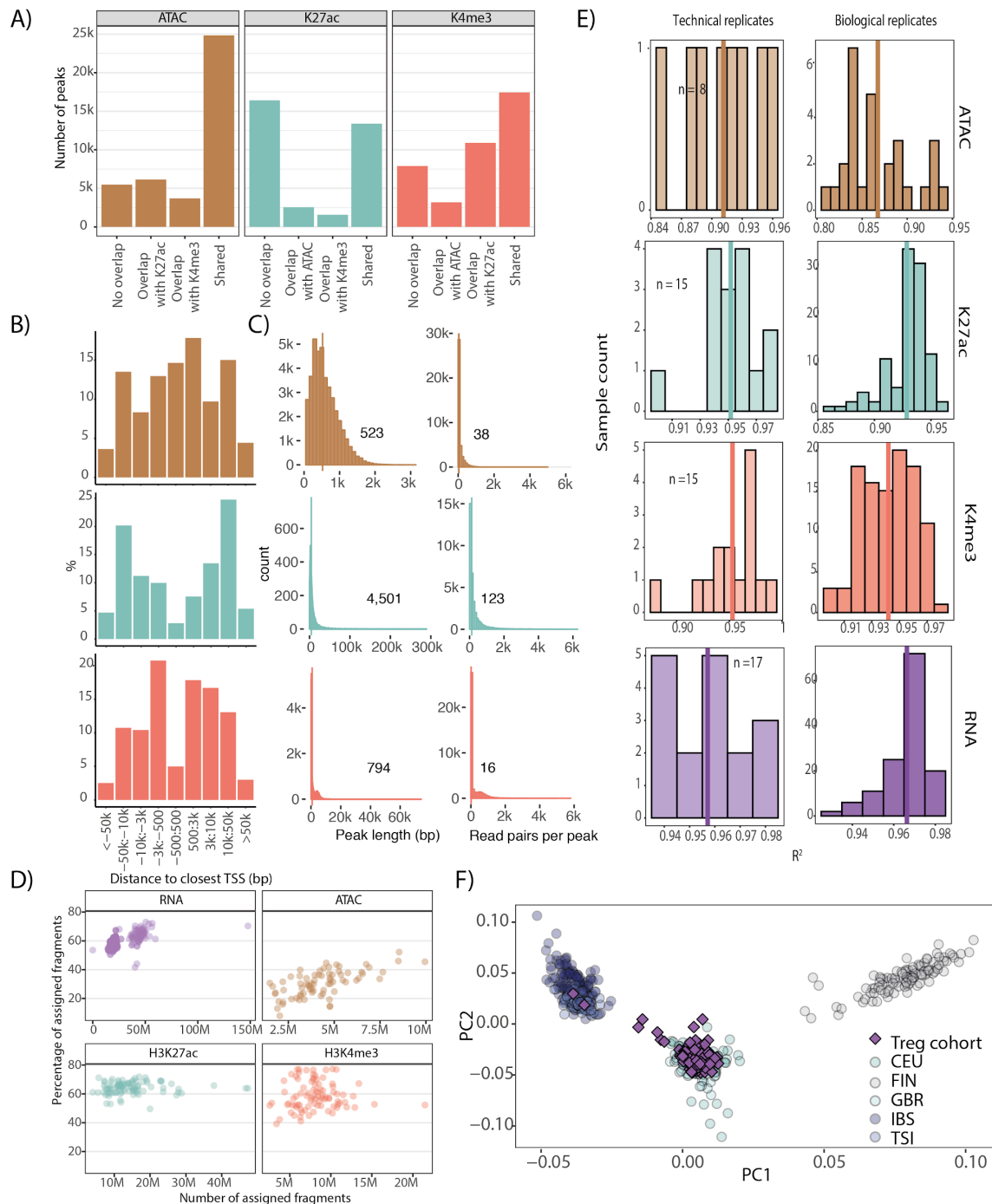
**Immune disease variants modulate gene  
expression in regulatory CD4<sup>+</sup> T cells**

**Lara Bossini-Castillo, Dafni A. Glinos, Natalia Kunowska, Gosia Golda, Abigail A. Lamikanra, Michaela Spitzer, Blagoje Soskic, Eddie Cano-Gamez, Deborah J. Smyth, Claire Cattermole, Kaur Alasoo, Alice Mann, Kousik Kundu, Anna Lorenc, Nicole Soranzo, Ian Dunham, David J. Roberts, and Gosia Trynka**

## Supplementary Figures

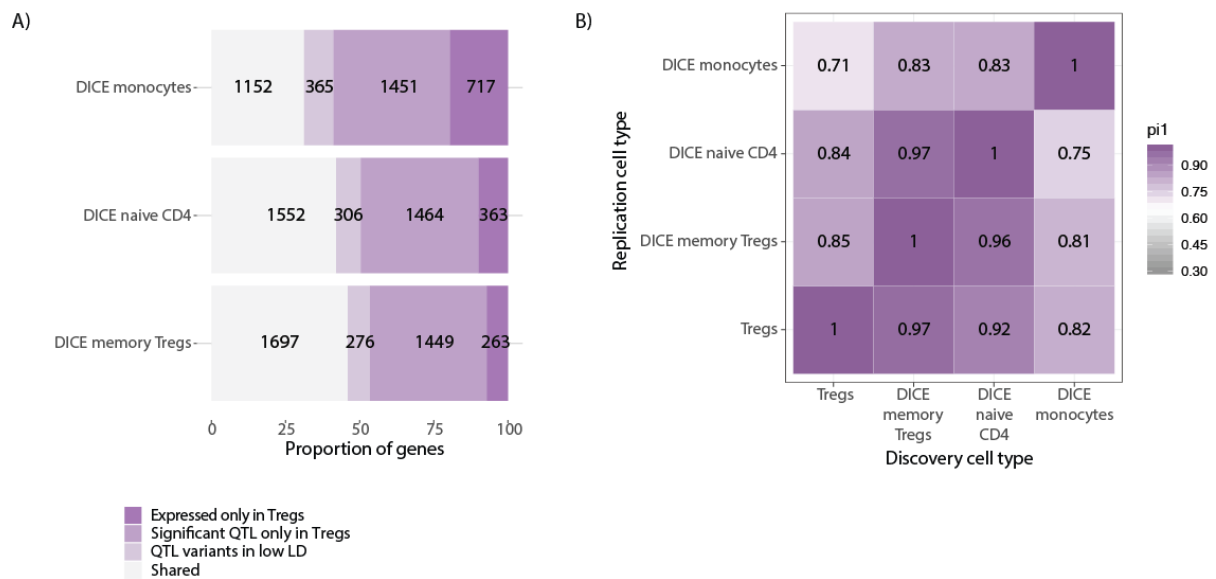


**Supplementary Figure 1. Gating strategy for Treg isolation and validation. A)** Gating used during the Treg cell sorting step. **B)** Gating strategy for the validation of Treg identity via FoxP3 staining. **C)** Percentage of regulatory T cells expressing b7 integrin, CD45RA and CLA directly after a blood draw. The same donors were assayed up to 3 times across multiple months. **D)** Donor age and sex distribution. **E)** Percentage of FOXP3+ cells per sample. **F)** Percentage of CD45RA- cells per sample. **G)** Percentage of alive cells as determined by DAPI at time of sorting. Related to STAR Methods: FACS staining

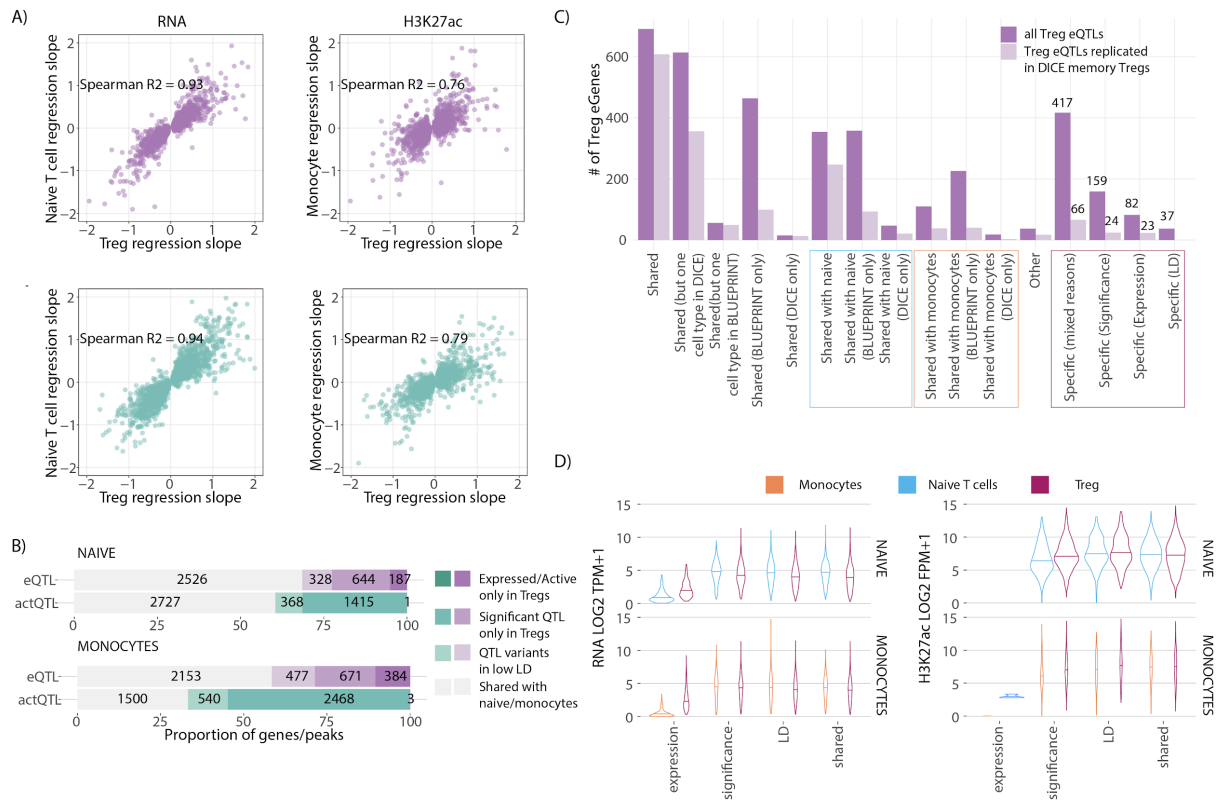


**Supplementary Figure 2. Summary statistics. A)** Peak numbers and peak overlaps across chromatin assays. In different colours are the numbers of peak overlaps per chromatin mark. **B)** Distribution of chromatin features in the proximity of the closest TSS. **C)** Distribution of peak length (left) and number of read pairs per peak (right). Median values are shown. **D)** Relationship between the total number of fragments per assay and the percentage of assigned fragments (sequenced read pairs) **E)** Correlation between technical and biological replicates in the different assays. **F)** Projection of the genotyped samples from our Treg study onto the European cohorts included in the 1000 Genomes Project. CEU: Utah Residents (CEPH) with Northern and Western European Ancestry; FIN: Finnish in Finland; GBR: British in

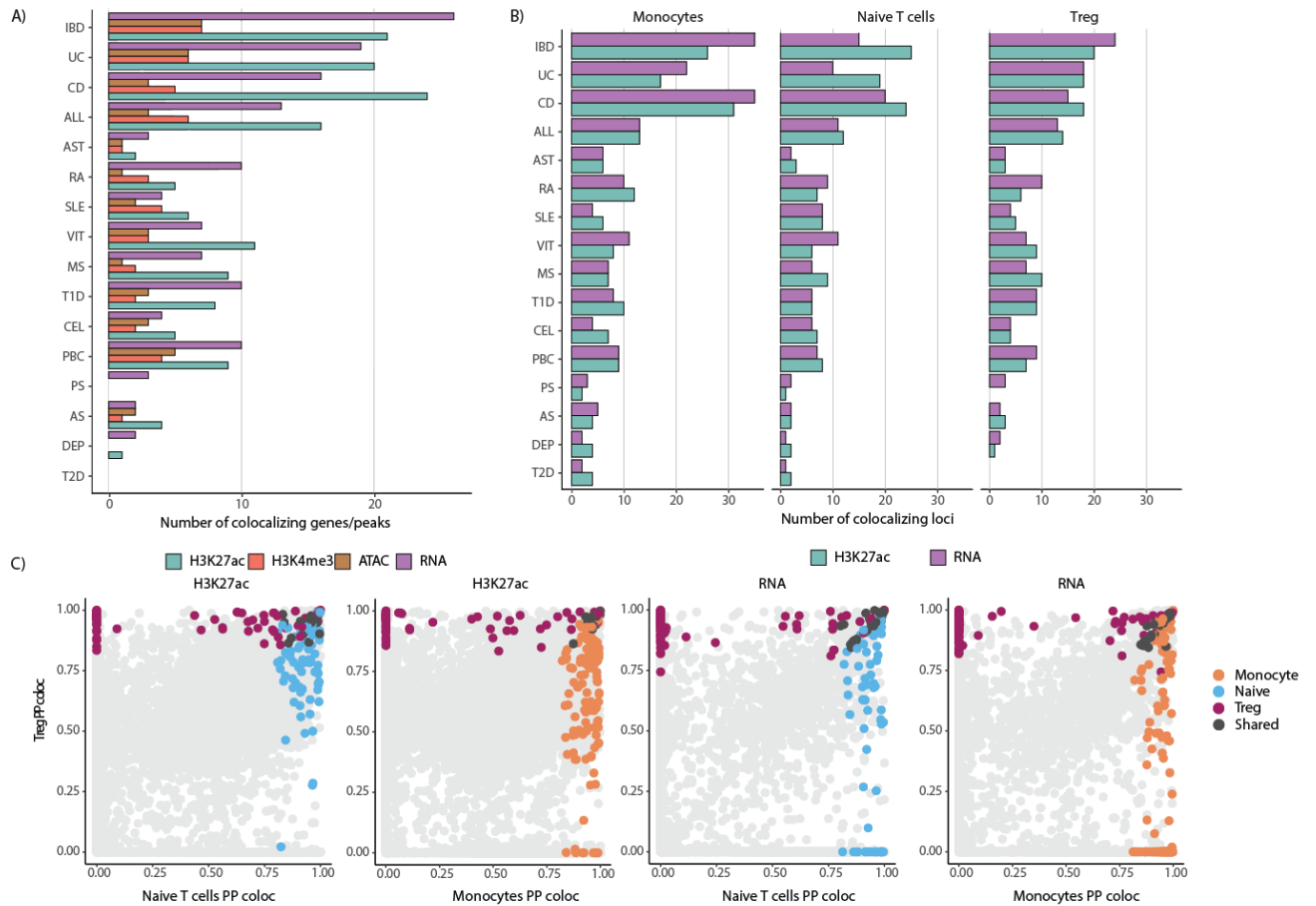
England and Scotland; IBS: Iberian Population in Spain; TSI: Toscani in Italia. Related to STAR  
 Methods: SNP genotyping and imputation



**Supplementary Figure 3. Treg eQTLs comparisons with immune cell types assayed in the DICE consortium. A)** Classification of Treg eQTLs specific to Tregs in relation to different immune cell types assayed in DICE. **B)** Pi1 scores between pairwise samples selected from DICE, Blueprint and the Treg eQTL dataset produced here. Related to Figure 2.

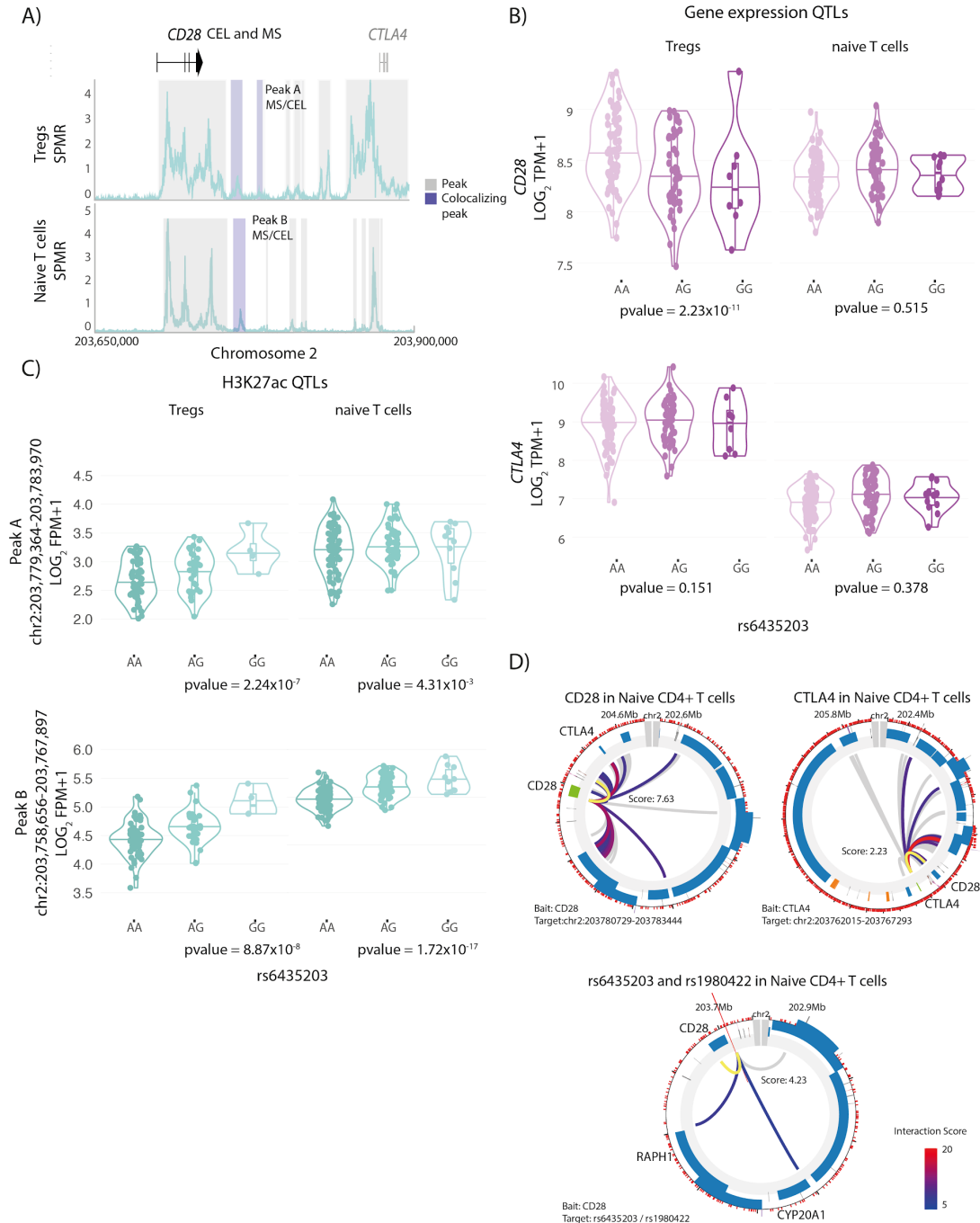


**Supplementary Figure 4. Treg eQTLs and actQTLs comparisons with CD4+ naive cells and monocytes assayed in the BLUEPRINT. A)** Correlation between the regression slopes for the top eQTL and actQTL variants discovered in CD4+ naive and regulatory T cells. Regression slope of the top eQTL and actQTL variant in regulatory T cells plotted against the slope for the same variant-gene pair in naive T cells. **B)** Classification of Treg eQTLs and actQTLs specific to Tregs in relation to naive T cells and monocytes. **C)** Classification of Treg eQTLs and actQTLs specific to Tregs in relation to naive T cells and monocytes from both the BLUEPRINT and the DICE consortia. **D)** Levels of gene expression and peak height across the QTL classifications in the different cell types. Related to Figure 2.



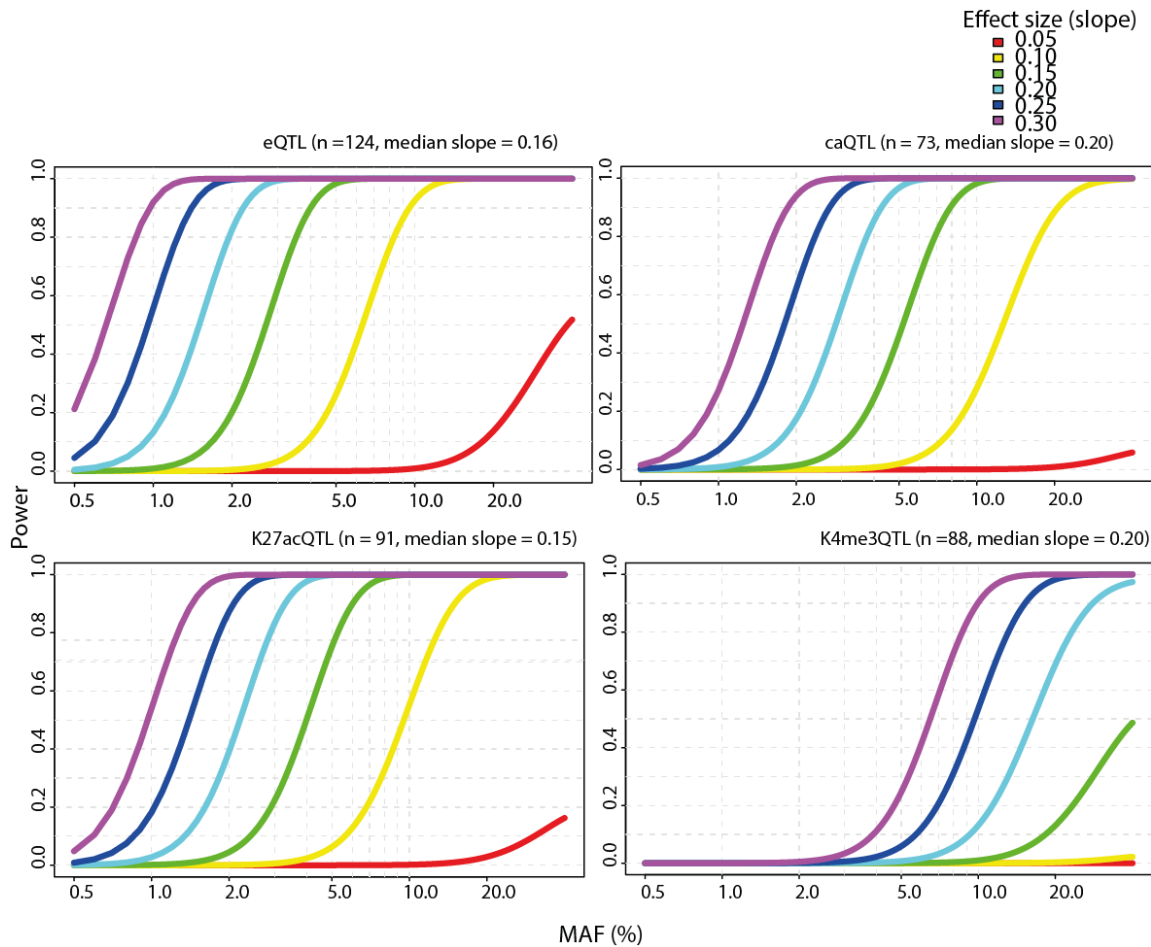
**Supplementary Figure 5. Colocalization analysis between Treg QTLs and immune GWAS studies. A)** Number of genes or peaks colocalizing with selected immune GWAS studies across the four genomic assays in Tregs. **B)** Comparison of number of GWAS loci colocalizing with eQTLs and actQTLs detected in naive T cells, monocytes and Tregs. **C)** Posterior probability (PP) of coloc for the same genes (eQTLs) or peaks (actQTLs) across all immune GWAS studies between Tregs and naive T cells or monocytes. Related to Figure 3.





**Supplementary Figure 6. Complex colocalization pattern of celiac disease and multiple sclerosis variants with cell type specific eQTLs and chromQTLs in naive and regulatory T cells at CD28/CTLA4 locus. A)** Variants associated with CEL and MS, tagged by chr2:203,745,673 (rs1980422) and chr2:203,746,472 (rs6435203) respectively, are colocalizing with a *CD28* eQTL, with Peak A (an actQTL chr2:203,779,364-203,783,970) only present in Tregs and Peak B (an actQTL chr2:203,758,282-203,767,897) only present in naive T cells. **B)** *CD28* and *CTLA4* eQTLs in Tregs and naive T cells stratified by the MS associated rs6435203 genotype. The disease risk allele for MS, (rs6435203-A, major allele), resulted in increased levels of *CD28* expression, while the risk allele for CEL, chr2:203,746,472 (rs1980422-C, minor allele) resulted in increased the levels of *CD28* mRNA (both variants are

highly linked,  $R^2 = 0.88$ ). The levels of *CTLA4* mRNA were not affected by genotype. **C)** Peak A and Peak B actQTLs in Tregs and naive T cells stratified by MS associated rs6435203 genotype. The risk allele for MS decreased the H3K27ac levels while the risk allele for CEL resulted in increased acetylation. Rs6435203 was a significant actQTL for peaks A and B in Tregs, but this variant colocalized only with peak B in naive T cells (a similar colocalization pattern was observed for the CEL associated variant, rs1980422). **D)** Promoter capture Hi-C plots of the interaction scores for the *CD28*, *CTLA4* and rs1980422 or rs6435203 loci in naive CD4<sup>+</sup> T cells published in Javierre *et al.* 2016 <sup>1</sup> (generated using Capture HiC Plotter online tool, <https://www.chicp.org/chicp/>). The Peak A region interacts with the *CD28* promoter, while the Peak B region interacts with the *CTLA4* promoter. Both SNPs interacted with the promoter of *CD28* in naive and total CD4<sup>+</sup> T cells. CEL variant (rs1980422) was also an eQTL for *CTLA4* in CD4<sup>+</sup>, CD8<sup>+</sup> cells and in testis <sup>2,3</sup>, and interacted with the promoter of *CTLA4* in total CD4<sup>+</sup> cells <sup>1</sup>. *CTLA4*, which inhibits T cell mediated immune responses by outcompeting CD28 for ligand binding <sup>4</sup>. Coordinates correspond to the GRCh38 build. Bait and target coordinates and interaction scores for specific connections (yellow). Related to Figure 4.



**Supplementary Figure 7.** Power calculations for QTL discovery using different functional genomic assays. Power was calculated for each assay using the powerEQTL package. Calculations are adjusted by sample size, number of tested SNPs, average reads per feature per assay. Power (y-axis) is plotted against minor allele frequency (MAF; x-axis). Different coloured lines correspond to different linear model slopes. Related to STAR Methods: Quantitative trait locus mapping (QTLs)

### Supplementary References

1. Javierre BM, Burren OS, Wilder SP, Kreuzhuber R, Hill SM, Sewitz S, et al. Lineage-Specific Genome Architecture Links Enhancers and Non-coding Disease Variants to Target Gene Promoters. *Cell*. 2016 Nov 17;167(5):1369–84.e19.
2. Kasela S, Kisand K, Tserel L, Kaleviste E, Remm A, Fischer K, et al. Pathogenic implications for autoimmune mechanisms derived by comparative eQTL analysis of CD4+ versus CD8+ T cells. *PLoS Genet*. 2017 Mar;13(3):e1006643.
3. Gamazon ER, Segrè AV, van de Bunt M, Wen X, Xi HS, Hormozdiari F, et al. Using an atlas of gene regulation across 44 human tissues to inform complex disease- and trait-associated variation. *Nat Genet*. 2018 Jul 1;50(7):956–67.
4. Qureshi OS, Zheng Y, Nakamura K, Attridge K, Manzotti C, Schmidt EM, et al. Trans-endocytosis of CD80 and CD86: a molecular basis for the cell-extrinsic function of

CTLA-4. Science. 2011 Apr 29;332(6029):600-3.