

Title: An atlas of transcribed enhancers across helper T cell diversity for decoding human diseases

Authors: Akiko Oguchi^{1,2,3,†}, Akari Suzuki^{4,†}, Shuichiro Komatsu^{1,5,†}, Hiroyuki Yoshitomi^{2,6}, Shruti Bhagat², Raku Son^{1,2,3}, Raoul Jean Pierre Bonnal⁵, Shohei Kojima⁷, Masaru Koido^{8,9,10}, Kazuhiro Takeuchi^{1,2,11}, Keiko Myouzen⁴, Gyo Inoue⁴, Tomoya Hirai^{1,12}, Hiromi Sano¹, Yujiro Takegami¹³, Ai Kanemaru¹³, Itaru Yamaguchi¹³, Yuki Ishikawa¹⁰, Nao Tanaka¹⁰, Shigeki Hirabayashi^{1,14,15}, Riyo Konishi¹⁶, Sho Sekito^{2,17}, Takahiro Inoue¹⁷, Juha Kere^{18,19,20}, Shunichi Takeda^{21,22}, Akifumi Takaori-Kondo¹⁴, Itaru Endo¹², Shinpei Kawaoka^{16,23}, Hideya Kawaji^{24,25,26}, Kazuyoshi Ishigaki²⁷, Hideki Ueno^{2,6}, Yoshihide Hayashizaki^{13,26}, Massimiliano Pagani^{5,28}, Piero Carninci^{29,30}, Motoko Yanagita^{2,3}, ITEC consortium, Nicholas Parrish⁷, Chikashi Terao^{10,31,32*}, Kazuhiko Yamamoto^{4*}, Yasuhiro Murakawa^{1,2,5,11*}

Affiliations:

¹RIKEN-IFOM Joint Laboratory for Cancer Genomics, RIKEN Center for Integrative Medical Sciences; Yokohama, Japan.

²Institute for the Advanced Study of Human Biology, Kyoto University; Kyoto, Japan.

³Department of Nephrology, Graduate School of Medicine, Kyoto University; Kyoto, Japan.

⁴Laboratory for Autoimmune Diseases, RIKEN Center for Integrative Medical Sciences; Yokohama, Japan.

⁵IFOM ETS - the AIRC Institute of Molecular Oncology; Milan, Italy.

⁶Department of Immunology, Graduate School of Medicine, Kyoto University; Kyoto, Japan.

⁷Genome Immunobiology RIKEN Hakubi Research Team, RIKEN Center for Integrative Medical Sciences; Yokohama, Japan.

⁸Division of Molecular Pathology, Department of Cancer Biology, Institute of Medical Science, The University of Tokyo; Tokyo, Japan.

⁹Laboratory of Complex Trait Genomics, Department of Computational Biology and Medical Sciences, Graduate School of Frontier Sciences, The University of Tokyo; Tokyo, Japan.

¹⁰Laboratory for Statistical and Translational Genetics, RIKEN Center for Integrative Medical Sciences; Yokohama, Japan.

¹¹Department of Medical Systems Genomics, Graduate School of Medicine, Kyoto University; Kyoto, Japan.

¹²Department of Gastroenterological Surgery, Yokohama City University Graduate School of Medicine, Yokohama City University; Yokohama, Japan.

¹³K.K. DNAFORM; Yokohama, Japan.

¹⁴Department of Hematology and Oncology, Graduate School of Medicine, Kyoto University; Kyoto, Japan.

¹⁵Division of Precision Medicine, Kyushu University Graduate School of Medical Sciences, Kyushu University; Fukuoka, Japan.

¹⁶Inter-Organ Communication Research Team, Institute for Life and Medical Sciences, Kyoto University; Kyoto, Japan.

¹⁷Department of Nephro-Urologic Surgery and Andrology, Mie University Graduate School of Medicine, Mie University; Tsu, Japan.

¹⁸Department of Biosciences and Nutrition, Karolinska Institutet; Stockholm, Sweden.

¹⁹Stem Cells and Metabolism Research Program, University of Helsinki; Helsinki, Finland.

²⁰Folkhalsan Research Center; Helsinki, Finland.

²¹Department of Radiation Genetics, Graduate School of Medicine, Kyoto University; Kyoto, Japan.

²²Shenzhen University School of Medicine; Shenzhen, Guangdong, China.

²³Department of Integrative Bioanalytics, Institute of Development, Aging and Cancer, Tohoku University; Sendai, Japan.

²⁴Research Center for Genome & Medical Sciences, Tokyo Metropolitan Institute of Medical Science; Tokyo, Japan.

²⁵Preventive Medicine and Applied Genomics Unit, RIKEN Center for Integrative Medical Science; Yokohama, Japan.

²⁶RIKEN Preventive Medicine and Diagnosis Innovation Program; Wako, Japan.

²⁷Laboratory for Human Immunogenetics, RIKEN Center for Integrative Medical Sciences; Yokohama, Japan.

²⁸Department of Medical Biotechnology and Translational Medicine, Università degli Studi; Milan, Italy.

²⁹Laboratory for Transcriptome Technology, RIKEN Center for Integrative Medical Sciences; Yokohama, Japan.

³⁰Human Technopole; Milan, Italy.

³¹Clinical Research Center, Shizuoka General Hospital; Shizuoka, Japan.

³²Department of Applied Genetics, School of Pharmaceutical Sciences, University of Shizuoka; Shizuoka, Japan.

†Co-first authors

*Corresponding authors:

Chikashi Terao: chikashi.terao@riken.jp

Kazuhiko Yamamoto: kazuhiko.yamamoto@riken.jp

Yasuhiro Murakawa: yasuhiro.murakawa@riken.jp

Abstract: Transcribed enhancer maps can reveal nuclear interactions underpinning each cell type and connect specific cell types to diseases. Using a 5' single-cell RNA sequencing approach, we defined transcription start sites of enhancer RNAs and other classes of coding and non-coding RNAs in human CD4⁺ T cells, revealing cellular heterogeneity and differentiation trajectories. Integration of these datasets with single-cell chromatin profiles showed that active enhancers

with bidirectional RNA transcription are highly cell type-specific, and disease heritability is strongly enriched in these enhancers. The resulting cell type-resolved multimodal atlas of bidirectionally transcribed enhancers, which we linked with promoters using fine-scale chromatin contact maps, enabled us to systematically interpret genetic variants associated with a range of immune-mediated diseases.

One-Sentence Summary: A cell type-resolved, multi-layered atlas of heterogenous CD4⁺ T cells to pinpoint genetic causes of immune diseases.

Main Text: Profiling of active enhancers in single cells provides mechanistic insight into cell type-specific gene regulation and genetic predisposition to human diseases. Enhancers generate cell type-specific transcriptome by enhancing the expression of target genes at a distance (1–3). Various approaches have been developed to predict enhancers. The assay for transposase-accessible chromatin using sequencing (ATAC-seq) identifies accessible chromatin regions that contain active, inactive, and poised enhancers as well as other types of DNA elements such as insulators; recent single-nucleus ATAC-seq (snATAC-seq) studies have generated cell type-resolved maps of accessible chromatin regions (4–7). Chromatin signature profiling showed that histone 3 lysine 27 acetylation (H3K27ac) can distinguish active enhancers from inactive or poised enhancers containing only H3 mono-methylated at lysine 4 (H3K4me1), although histone modifications often span several kilobases (8). Moreover, 5'-end RNA sequencing methods such as cap analysis of gene expression (CAGE) and precision run-on 5'-cap sequencing (PRO-CAP) can identify active enhancers at high resolution (~150 base pair [bp]) by analyzing the transcription start sites (TSSs) of enhancer RNAs (eRNAs) (3, 9–18). Enhancers with bidirectional eRNAs are more functional than untranscribed ones (3, 11, 19, 20). However, despite its importance, we still lack comprehensive maps of eRNAs in cell types that are relevant to human diseases and traits of interest (21).

CD4⁺ T cells play a pivotal role in orchestrating the immune system, exerting a complex array of functions such as regulation of B cell antibody class switching, cytotoxic T cell activation, lymphoid follicle organization, and suppression of autoimmunity (22). The strong associations of various immune-mediated diseases with class II major histocompatibility complex (MHC), which is specifically recognized by CD4⁺ T cells, suggest their critical roles in pathogenesis of these diseases (23). Major canonical CD4⁺ T cells include T helper type 1 (Th1), Th2, Th17, T follicular helper (Tfh), and regulatory T cells (Tregs) (24), but recent single-cell transcriptomics has revealed a broader spectrum of cell types (25–35). It is increasingly recognized that rare, but functionally important, CD4⁺ T cell types exist and play key roles in human diseases (36). However, the heterogeneity of CD4⁺ T cells and how these cell types impact human diseases remain incompletely characterized.

Genetic variants associated with human diseases are concentrated within functional enhancers that are active in disease-relevant cell types (1–3, 5–7, 11, 14, 15, 21, 37–44). Therefore, integrating data on epigenetic signals at cell type-specific enhancers with large-scale human genetics data has the potential to systematically elucidate novel molecular pathways underpinning human diseases. In this study, by leveraging a “cap signature” derived from a 7-methylguanosine cap present at the bona fide 5' ends of RNAs, we developed ReapTEC (read-level pre-filtering and transcribed enhancer call) that simultaneously profiles gene expression and enhancer activity at nucleotide resolution by using 5'-end single-cell RNA-sequencing (5' scRNA-seq). We applied ReapTEC and snATAC-seq to a total of ~1 million single CD4⁺ T cells, identifying 62,803 bidirectionally transcribed candidate enhancers (btcEnh) and 218,508 accessible chromatin regions active across immensely heterogeneous CD4⁺ T cells. Furthermore, by generating fine-scale chromatin contact maps at nucleosome resolution, we reveal the links between enhancers and promoters in CD4⁺ T cells. Finally, we demonstrate the broad utility of our cell type-resolved multi-layered human CD4⁺ T cell atlas, as a publicly available resource, for pinpointing genetic and cellular causes of immune-mediated diseases (**Fig. 1A and fig. S1A and Table S1**), which will likely facilitate the further development of innovative biomarkers and drugs in the future.

5' single-cell RNA-seq of human CD4⁺ T cells

Chromium 5' scRNA-seq showed the same level of gene detection as Chromium 3' scRNA-seq (**fig. S1B**). An initial 5' scRNA-seq analysis of circulating bulk CD4⁺ T cells from 3 healthy donors (**Table S2**) identified many distinct CD4⁺ T cell clusters (**Fig. 1, B and C, and fig. S1C**). We also performed cellular indexing of transcriptomes and epitopes by sequencing (CITE-seq) (45) to profile cell surface expression of ~130 membrane proteins including CD45RA (**Fig. 1D and fig. S1, D to F**), which is generally associated with naïve T cells (24). Besides the conventional naïve, effector, and regulatory CD4⁺ T cells, clusters included rare populations of naïve-like CD4⁺ T cells highly expressing understudied *NSF* (cluster 12); CD4⁺ T cells that express the exhaustion markers *TIGIT*, *CTLA4*, and *HAVCR2* as well as the exhaustion-associated transcription factors *TOX* and *MAF* (cluster 13); and cycling CD4⁺ T cells (cluster 14) (**Fig. 1C**). We unbiasedly identified patterns of gene and protein expression across a diverse continuum of CD4⁺ T cells (**Fig. 1, C and D, and fig. S1, C to F**). All clusters were present in all 3 donors (**fig. S1G**).

We next generated a TSS map of human CD4⁺ T cells using 5' scRNA-seq. Chromium 5' scRNA-seq, which employs the template-switching method, is often used to perform immune repertoire analysis together with transcriptome profiling. Gene expression levels are typically measured by sequencing the vicinity of TSSs (but not TSSs as such) in read 2, while read 1 (26 bp) is used to obtain cellular barcodes and unique molecular identifiers, except in a few instances (46–48). Here we obtained a long read 1 (150 bp), sequencing over the template switch oligo (TSO) to reach the actual 5' ends of the cloned cDNA fragments (**fig. S2A**). Substantial background signals were present in the 5' scRNA-seq data, mainly due to strand invasion (**fig. S2B**) and premature template switching (49, 50). Nevertheless, by using only reads containing an unencoded G resulting from the 7-methylguanosine cap of bona fide 5' ends (**fig. S2, C and D**) (51), we confidently mapped exact TSSs (**Fig. 1E**), as exemplified by 5'-end signals of mRNAs at the *SOCS7* promoter (**Fig. 1F and fig. S2B, see Supplementary Text 1**).

btcEnhs identified by ReapTEC

eRNA transcription can be used as a proxy for enhancer activity and to delineate the enhancer architecture at nucleotide resolution (9–12, 14–17, 19, 20). Pseudo-bulk data from 5' scRNA-seq detected eRNAs in bulk CD4⁺ T cells at a level comparable (r : 0.67) to that of native elongating transcript–cap analysis of gene expression (NET-CAGE) (**Fig. 1G**), a method to identify eRNAs with high sensitivity (10). Although eRNAs are generally not polyadenylated, they were efficiently captured by oligo(dT) priming in internal A-rich regions, as shown by multiplexed arrays sequencing (MAS-seq) (**fig. S3**), which allows accurate long-read sequencing of full-length cDNA before fragmentation in 5' scRNA-seq (52).

We next sought to identify active enhancers *de novo* on the basis of the presence of bidirectionally transcribed eRNAs. To call btcEnhs with bidirectional eRNAs using 5' scRNA-seq data, we developed the ReapTEC approach, which starts with the identification of TSS regions. RNA transcription does not typically occur from a single nucleotide position; the 5'-end signal is observed as peaks of various width (**Fig. 1H**) (53, 54). Several peak-calling methods such as Paraclu have been developed to cluster the 5'-end signals (53, 55–57). After peak-calling, CapFilter further selects genuine peaks on the basis of the percentage of reads with an unencoded G (56); in more recent methods, a multiple logistic regression classifier was devised to minimize artifactual peaks (46, 47).

Unlike these approaches, we first selected only reads with an unencoded G before peak-calling (*i.e.*, unencoded G filtering at the level of reads, not at the level of peaks), and then simply merged them to call a high-confidence TSS peak. We next identified divergent TSS peak pairs (**Fig. 1H**), which allowed us to identify btcEnhs at high nucleotide resolution with high sensitivity (**fig. S4**). Most btcEnhs identified by ReapTEC matched DNase I hypersensitive sites (DHSs) and H3K27ac and H3K4me1 modifications in CD4⁺ T cells (**Fig. 1I**), indicating that ReapTEC captures genuine enhancer regions. Without ReapTEC filtering, the overlap rate was drastically reduced (**Fig. 1I**), showing the importance of extensive noise reduction.

Transcription and chromatin accessibility at *cis*-regulatory elements

We performed Chromium Single-Cell Multiome ATAC + Gene Expression on bulk CD4⁺ T cells to measure chromatin accessibility and 3' gene expression in parallel; then, we integrated the obtained data with 5' scRNA-seq data. Pseudo-bulk chromatin accessibility was modestly correlated with RNA transcription at gene promoters (correlation coefficient 0.59; **fig. S5A**) and distal DHSs (0.42; **fig. S5B**).

Next, we looked into the cell type specificity of RNA transcription and chromatin accessibility across diverse CD4⁺ T cell clusters. At the *CCL5* (**Fig. 1J and fig. S5C**) and *DTHD1* (**fig. S5, C and D**) promoters, chromatin accessibility was related to transcription in a cell type-specific manner across CD4⁺ T cell clusters. However, the *GZMA* transcript was strongly expressed in clusters 7, 10, and 13, whereas the *GZMA* promoter was ubiquitously open (**Fig. 1K and fig. S5C**). Cell type-specific RNA transcription was also observed at the promoter and nearby enhancers of the *GPR68* gene, whereas chromatin accessibility in these regions was similar across different CD4⁺ T cell clusters (**Fig. 1L and fig. S5C**), showing differences in RNA transcription vs. chromatin accessibility.

Highly diverse CD4⁺ T cells

We investigated *cis*-regulatory elements in heterogenous CD4⁺ T cell subsets, including barely detectable rare populations. To generate comprehensive multimodal single-cell datasets on gene expression, promoters, enhancers, chromatin accessibility, and membrane proteins (**Fig. 1A**), we performed extensive 5' scRNA-seq, Multiome (snATAC-seq + 3' snRNA-seq), and CITE-seq on cells from each of 5 representative CD4⁺ T cell groups obtained by fluorescence-activated cell sorting (58) (FACS; **Figs. 1A and 2A**): CD4⁺CD25⁺ Group A, LAG3⁺ Group B, Tfh Group C, Th1/Th2 Group D, and Th17 Group E (**Materials and Methods**). On the basis of CITE-seq protein expression, Group D was further categorized into CXCR3^{high}CCR4^{low} Th1 (clusters 0, 1, and 7), CXCR3^{low}CCR4^{high} Th2 (clusters 4 and 6), and other cell subsets (**Fig. 2A and fig. S6A to C**). Iterative clustering of Th1 and Th2 cells was further performed to identify additional clusters (**Fig. 2A and fig. S6, D and E**).

We identified a large number of rare cell populations that constitute < 0.5% of the bulk CD4⁺ T cells (**fig. S6F**). For example, CD4⁺CD25⁺ Group A cells were divided into 13 diverse clusters; 9 of them expressed the Treg master transcription factor *FOXP3* (**Fig. 2A and fig. S6, G and H**). These Treg cells were further divided into naïve (clusters 1, 5, 7) and effector Tregs (clusters 0, 3, 9, 10, 11, 12) on the basis of CITE-seq surface protein expression (**Fig. 2, A and B, and fig. S6I**). scVelo's velocity estimates (59, 60) well predicted the differentiation from naïve Tregs into effector Tregs (clusters 0, 3, 10) (**Fig. 2C**). A distinct cluster 9, which was predicted to have

originated differently from other effector Tregs by velocity estimates (**Fig. 2C**), highly expressed the Th17 transcription factor *RORC*, the immune checkpoint gene *LAG3* (61), and *GPR25*, an uncharacterized G protein-coupled receptor gene associated with ankylosing spondylitis (62) (**fig. S6H**). Group E cells were divided into 13 clusters (**Fig. 2A**), including *RORC*^{high} Th17 cells (cluster 6) and *CCR4*⁺*CCR6*⁺*CCR10*⁺ Th22 cells (clusters 0, 1, 7) (**Fig. 2, A and D, and fig. S6, G and J**). Velocity estimates predicted that these Th17 and Th22 cells might differentiate from cluster 8, which expressed relatively high levels of ribosomal protein genes (**Table S3**). Among rare cell populations in Group E, we identified uncharacterized *RORC*⁺ CD4⁺ T cells that also express *GNLY*, *GZMA*, and *GZMK* (cluster 11) (**Fig. 2A and figs. S6, G and J, and S7**). Granzyme-producing CD4⁺ T cells are rapidly emerging as key players in aging, autoimmunity, and cancer immunotherapy (63, 64). Cell clusters expressing perforin (*PRF1*), granzyme (*GZM*), or granulysin (*GNLY*) were present in all CD4⁺ T cell groups (**fig. S7**). Furthermore, CITE-seq profiling of ~130 surface proteins on Group A-E CD4⁺ T cells in single cells revealed dynamic expression of molecules involved in cytokine signaling and immunomodulatory interactions, which may indicate the potential functional roles of these rare CD4⁺ T cell clusters identified (**fig. S8 and Table S4**).

We also used publicly available 5' scRNA-seq data (with 2 × 150 bp paired-end runs to allow for TSS analysis) from a COVID-19 study that sequenced memory CD4⁺ T cells (40 COVID-19 patients and 13 healthy donors; **Table S2**) stimulated with virus-specific peptides (65) (0 h, *n* = 24,100; 6 h, *n* = 102,970; 24 h, *n* = 45,367). Upon stimulation, memory CD4⁺ T cells generally increased gene expression (**fig. S9A**), giving rise to various cell types and states (**fig. S9B**). In addition, 5' scRNA-seq and Multiome were performed on circulating bulk CD4⁺ T cells stimulated via CD3/CD28 for 48 h (*n* = 8,007) (**fig. S9C**).

Overall, 136 cell clusters with highly diverse and often continuous gene expression profiles were detected on the basis of single-cell analysis of a total of 962,944 circulating bulk, FACS-sorted, or stimulated CD4⁺ T cells (**Figs. 1A and 2A**). We visualized the hierarchical organization of the gene-expression landscape across the 136 clusters (**fig. S10A**). Gene expression matrices are provided in **Data S1** and a list of signature genes for these 136 clusters in **Table S3**. Although several of these clusters may overlap partly due to the fact that FACS sorting is not completely exclusive (**Supplementary Text 2**), analyses at the level of cell groups allowed us to reveal CD4⁺ T cell heterogeneity and differentiation in detail (**Fig. 2A, C and D, and fig. S10B, and Data S2**).

A CD4⁺ T cell type-resolved atlas of *cis*-regulatory elements

We applied ReapTEC to each of the 136 CD4⁺ T cell clusters identified by our 5' scRNA-seq; btcEnhs identified in exons were excluded when they did not overlap with CD4⁺ T cell DHS (**Materials and Methods**). In total, we identified 62,803 btcEnhs active across diverse CD4⁺ T cell populations (**Data S3 and S4**). These btcEnhs overlapped with ENCODE DHSs and H3K27ac and H3K4me1 modifications, which supports the authenticity of the btcEnhs (**Fig. 2E**). The btcEnhs were also enriched for ENCODE3 enhancer-like signatures (2) (**fig. S11A**), 34% of which were newly identified in our extensive 5' scRNA-seq analyses (**fig. S11, B to D**). About 82% of CD4⁺ T cell btcEnhs were not identified in the FANTOM5 project (**fig. S11E**); thus, our data considerably expand the catalog of transcribed enhancers. By intersecting btcEnhs with a massively parallel reporter assay (MPRA) designed to prioritize variants associated with autoimmune diseases in Jurkat immortalized human CD4⁺ T cell line (66), we found that,

overall, btcEnhs showed a significant increase in reporter signal (P value $< 2.2e-16$), indicating the functionality of btcEnhs (**fig. S11F**).

Furthermore, using snATAC-seq data from our Multiome experiments, we identified a total of 218,508 ATAC peaks in heterologous CD4⁺ T cell populations, of which approximately 88% were located distal to gene promoters (**Fig. 2F and Data S4**). We further categorized these distal ATAC (dATAC) peaks into bidirectional, unidirectional, and untranscribed regions (**Fig. 2F**) by inferring the RNA transcripts that could be transcribed from each dATAC peak on the basis of all 5' reads detected in this study.

The enrichment of transcription factor-binding motifs (TFBMs) in the btcEnhs was generally correlated with that in dATAC peaks, although a group of more strongly enriched TFBMs was observed in each group (**fig. S12A**). A much stronger enrichment of the CTCF-binding motif was observed in dATAC peaks than in btcEnhs (**fig. S12A**), suggesting that dATAC peaks contain more chromatin insulator sequences. Among bidirectional, unidirectional, and untranscribed dATAC peaks, sequence motifs recognized by specific sets of transcription factor families were enriched to a different degree, with those bound by ETS-related transcription factor families most strongly enriched in bidirectionally transcribed dATAC peaks (**Fig. 2G and Table S5**). Integration of the dATAC peak data with the chromatin immunoprecipitation sequencing (ChIP-seq) data revealed that the Mediator complex and transcription factors bound most strongly to the bidirectionally transcribed dATAC peaks, followed by unidirectionally transcribed dATAC peaks, but very weakly to untranscribed dATAC peaks (**Fig. 2H and fig. S12, B and C**). A similar trend was also observed for H3K27ac histone modification (**Fig. 2H and fig. S12C**), suggesting the functionality of bidirectionally transcribed enhancers. The intrinsic differences in the genomic sequences among these dATAC peaks were also supported by the MENTR machine learning models, which predicted these peaks from the surrounding ± 100 kb sequences (**fig. S12D**) (67). In addition, the btcEnhs and transcribed dATAC peaks were more conserved than untranscribed dATAC peaks (**Fig. 2I**).

Analysis of the dynamic pattern of RNA transcription and chromatin accessibility across the 136 cell clusters showed that eRNA expression at btcEnhs was more cell type-specific than mRNA expression at promoters (**Fig. 2J**), consistent with the role of btcEnhs in controlling gene expression in a cell type-specific manner (10, 11). Similarly, chromatin accessibility at the dATAC peaks was more cell-type specific than that at the proximal ATAC peaks (pATAC peaks) (**Fig. 2J**). As we initially observed (**Fig. 1K and L**), we found higher cell type specificity for RNA transcription than for chromatin accessibility both at promoters and enhancers (**Fig. 2J**). These results indicate the importance of capturing eRNAs for better understanding of cell-type specificity. We also analyzed dynamic eRNA expression of btcEnhs in the differentiation trajectories predicted by RNA velocity (**fig. S13**).

CD4⁺ T cell heterogeneity at the TSS level

Accurate analysis of TSSs and btcEnhs from 5' scRNA-seq data combined with ReapTEC allows us to investigate *cis*-regulatory elements such as alternative promoters and btcEnhs at nucleotide resolution (**Fig. 1, E to I**). For example, in CD4⁺ T cells activated via CD3/CD28, a sharp TSS signal was observed for the mRNA encoding the secreted form of IL-2 (*i.e.*, the isoform from which the entire signal peptide can be translated), whereas transcription of this gene starts further downstream in unstimulated cells not secreting IL-2 (**fig. S14A**). We comprehensively detected TSS peaks active across 136 diverse human CD4⁺ T cell populations

and defined 46,108 robustly expressed TSS peaks (with a cutoff of $\log_2\text{CPM} \geq 2$ in at least one cluster) (**Data S5**). A substantial fraction of these robust TSS peaks were not annotated in the GENCODE gene model (**fig. S14B**) or in FANTOM5 (**fig. S14C**). The robust TSS peaks expressed in the bulk circulating CD4⁺ T cells, including novel ones, strongly overlapped with ENCODE DHSs as well as H3K27ac and H3 tri-methylated at lysine 4 (H3K4me3) modifications (2) (**fig. S14D**).

We clustered bulk circulating CD4⁺ T cells using robustly expressed TSSs, including those transcribed from alternative promoters and btcEnhs detected by ReapTEC. This step increased the input information in comparison with conventional gene-level analysis and resulted in a total of 17 clusters (**fig. S14E**). Naïve CD4⁺ T cells were further clustered into distinct subsets characterized by specific TSS expression at *CISH*, *ID2*, *TPR*, *ARMH1*, *IFIT1*, *MX1*, and *STAT1* loci, implying the latent diversity of naïve cells (**fig. S14F**). We also examined cell-to-cell differences in TSS usage. For example, we found that the short isoform of *ARHGAP15* was predominantly expressed in Th1, Th2, and Th17 cells, whereas the long isoform was broadly expressed (**fig. S14, E to G**); the shortest isoform of the *DUT* gene was expressed only in *MKI67*⁺ cycling cells (**fig. S14H**). A candidate enhancer near *CCR7* was transcribed mainly in naïve cells, in which the *CCR7* gene is strongly expressed (**fig. S14I**). By combining 5' scRNA-seq data (TSS) with MAS-seq accurate long-read data, we comprehensively identified mRNAs, lncRNAs, and eRNAs with accurate 5' ends, including 492 *de novo* transcripts that did not overlap with the GENCODE gene models in the same directions (**Data S6**), and measured their expression levels at a single-cell level. For example, one *de novo* transcript (*MSTRG.3154*) near *AL137025.1* was specifically expressed in a subset of CD4⁺ T cells (**fig. S14J**).

Micro-C analysis of human CD4⁺ T cells

We next sought to identify target promoters of btcEnhs in CD4⁺ T cells. For this, we performed Micro-C (68–71), an enhanced chromosome conformation capture method, and resolved chromatin interactions at nucleosome resolution. We generated ultra-deep Micro-C data of a total of ~40 billion reads from resting and activated human CD4⁺ T cells (**Fig. 1A and Fig. 3, A and B, and Table S6, and Data S7**) and detected a total of 103,038 chromatin loops (**fig. S15A, and Table S7, and Data S8**). The loops identified in resting and activated cells were the same in some regions but different in others (**Fig. 3, A and B and fig. S15B**). These loops linking btcEnhs to btcEnhs (btcE-btcE), btcEnhs to promoters (btcE-P), and promoters to promoters (P-P) were highly enriched, confirming the validity of the identified loops (**Fig. 3C and fig. S15C**). We found chromatin loops connecting 11,296 btcE-P pairs (**Fig. 3D and fig. S15, D and E**). Most frequently, the loops connected btcEnhs with the most proximal genes, but some connected them with distantly located genes (**Fig. 3D**); the median distance between these btcEnhs and promoters was ~10,000 bps (**Fig. 3E**). Similarly, when Micro-C loops were intersected with ATAC peaks, dATAC peak–dATAC peak (dA-dA), dATAC peak–promoter (dA-P), and promoter–promoter (P-P) loops were highly enriched (**fig. S15F**). Enrichments of dA-P and dA-dA loops were the strongest when separately analyzed for the bi-directionally transcribed dATAC peaks (**fig. S15F**), suggesting that bidirectionally transcribed *cis*-regulatory elements have more chromatin interactions than unidirectionally transcribed or untranscribed ones.

To improve the sensitivity of the detection of fine-scale chromatin interactions centered on gene promoters, we performed promoter-capture Micro-C (72, 73) on resting CD4⁺ T cells, stimulated

CD4⁺ T cells, CD25⁺CD4⁺ T cells, LAG3⁺CD4⁺ T cells, Tfh, and Th17 cells at saturation depth (fig. S15, G and H, and Table S6, and Data S9). The btcE-P and P-P loops were highly enriched among these region-captured Micro-C (RCMC) loops (fig. S15I). Differential chromatin loops were observed in different cell types, implicating that the same gene could be driven by different sets of enhancers (fig. S15, J and K).

Substantial enrichment of disease heritability in btcEnhs

Large-scale genome-wide association studies (GWASs) have detected a number of genetic variants associated with many human immune-mediated diseases. However, functional interpretation remains challenging because the majority of the variants map to non-coding regions, where active gene regulatory elements are often poorly annotated (38).

Many genetic variants collectively contribute to pathogenesis of complex traits (74). To analyze contribution of functional annotations to heritability of human diseases, we used the linkage disequilibrium score regression method (75). Numerous immune-mediated diseases showed strong enrichment of heritability (40–120 fold) in single nucleotide polymorphisms (SNPs) mapped to CD4⁺ T cell btcEnh regions, whereas neurological diseases or physical traits did not (Fig. 4A and fig. S16A and Table S8).

Multiple sclerosis (MS), an immune-mediated disease, showed considerably higher heritability in btcEnhs than in ENCODE DHSs, histone modifications, or overall ATAC peaks (Fig. 4B and fig. S16B and Table S8), indicating the importance of active enhancers with bidirectional eRNA transcription in the pathogenesis of this disease. Among all ATAC peaks, the strongest enrichment of disease heritability was observed in the bidirectionally transcribed ones, followed by unidirectionally transcribed ones, and no enrichment was observed in the untranscribed ATAC peaks (Fig. 4B and fig. S16B), highlighting the importance of identifying bidirectionally transcribed regions.

Functional interpretation of non-coding risk variants

As GWAS loci often contain dozens of genetic variants in linkage disequilibrium with the lead SNPs, pinpointing causal variants at cell type-specific btcEnhs can potentially elucidate the context and mechanisms underlying disease pathogenesis. For this purpose, we collected trait-associated SNPs (lead SNPs and linked SNPs with $R^2 > 0.8$) from GWAS studies on immune-mediated diseases and COVID-19 and tested their overlap with the 62,803 btcEnhs annotated here. CD4⁺ T cell btcEnhs contained considerably more (3–9 fold) SNPs associated with immune-mediated diseases than randomly selected non-trait-associated SNPs of similar allele frequency and distance to coding regions (Fig. 5A). In total, we identified 606 btcEnhs that harbor SNPs associated with 18 immune-mediated diseases and COVID-19 (76), which we referred to as disease-relevant btcEnhs (Table S9).

To validate the *cis*-regulatory effects of the SNPs fine-mapped within the disease-relevant btcEnhs, we cloned enhancers carrying the reference and alternative alleles into a luciferase reporter plasmid and transfected the resulting constructs into Jurkat cells. We show that cloned DNA fragments exhibited enhancer activity that was altered by the risk allele, which might be associated with creation or disruption of TFBMs (fig. S17, A and B). Among the SNPs fine-mapped within disease-relevant btcEnhs, 96 SNPs shown in Table S10 are also annotated as expression-modulating variants (66) that show differences in expression between the reference

and alternative alleles in MPRA in Jurkat cells. *In silico* mutagenesis using MENTR predicted the effects on transcription of enhancers for around 30% of the fine-mapped risk variants (**Table S11**), which was substantially enriched over the predicted effects of unfiltered variants (2.8%) (67).

We next investigated CD4⁺ T cell subsets that may be associated with immune-mediated diseases by examining the specific patterns of eRNA transcription at disease-relevant btcEnhs. We generated a heat map of eRNA transcription levels at 606 disease-relevant btcEnhs across 136 CD4⁺ T cell clusters and found that these btcEnhs are often specifically activated in a limited subset of CD4⁺ T cells (**Fig. 5B**). Approximately two-thirds of disease-relevant btcEnhs were activated in subsets of stimulated CD4⁺ T cells (**Fig. 5B**). When investigating the activation patterns of several Treg-specific disease-relevant btcEnhs, we found that some were active mostly in a subset of effector Tregs that highly express *LAG3* and *GPR25* (Group A, clusters 9 and 12), and others in different effector Tregs (Group A, cluster 3 and 10) or naïve Tregs (Group A, cluster 1 and 5), or a combination thereof (**Fig. 5C**). These results implicate that specific CD4⁺ T cell types and states can contribute differentially to different human diseases.

Systematic analysis of immune-mediated disease pathways

We predicted the putative target genes of disease-relevant btcEnhs by combining various experimental and computational approaches. First, we used the data on chromatin loops derived from ultra-deep Micro-C and promoter-capture Micro-C in various subsets of CD4⁺ T cells. We also applied the activity-by-contact (ABC) model (77, 78) which were generated using our Micro-C, snATAC-seq, and ENCODE histone modification data (**Data S10**). We further used 25,178 btcE-P links (adjusted *P* values < 0.05) (**fig. S18A**), which were inferred by computing the co-variation of eRNA and mRNA transcription measured across the 136 CD4⁺ T cell clusters in our 5' scRNA-seq (**Table S12**). btcE-P pairs supported by Micro-C loops showed significantly higher transcriptional co-variation than all possible btcE-P pairs (*P* value < 2.2e-16) (**fig. S18B**). By integrating all these interactions because each approach has its own unique advantage, we predicted target genes for 550 of the 606 disease-relevant btcEnhs with one or more supporting lines of evidence (**Fig. 5D and Table S13**). The genes connected to disease-relevant btcEnhs were highly enriched for pathways related to immune responses and cytokine signaling (**fig. S19A**). When the predicted target genes of disease-relevant btcEnhs were analyzed separately for each disease, we found that, for example, inflammatory bowel disease (IBD) genes and MS genes were different, indicating potentially different mechanisms of disease etiology (**fig. S19, B to D**).

We then performed CRISPR activation (CRISPRa) of selected disease-relevant btcEnhs in Jurkat cells and examined changes in promoter-level expression using CAGE. We designed guide RNAs in btcEnh regions. Of the 12 targeted disease-relevant btcEnhs in which bidirectional eRNA was induced by CRISPRa, 7 had a target gene with elevated mRNA expression (FDR < 0.05) (**Fig. S20**), which was a good validation rate in light of other reports (79, 80). For example, when we activated a btcEnh containing a SNP associated with IBD, an induction of bidirectional eRNA at this locus was accompanied by specific upregulation of the *IL7R* gene (**Fig. 5E**).

Evolutionary landscape of disease-relevant btcEnhs

Using a multiple alignment of genomes of 24 species (81), we investigated the evolutionary conservation of btcEnhs. For each of the 62,803 human CD4⁺ T cell btcEnhs, we identified species that carry orthologous enhancers (**Fig. 6A and Table S14**). The rate of overlap with transposable elements was much higher among primate-specific btcEnhs than among
 5 btcEnhs conserved in all mammals (**Fig. 6B**). We also found that many primate-specific btcEnhs overlapped with long interspersed nuclear element (LINE) (**Fig. 6C**). The breakdown of transposable elements in ATAC peaks showed a slightly different pattern from that in btcEnhs, especially in human-specific ATAC peaks, where retroposons accounted for about 30% (**Fig. 6C and Table S14**).

The proportion of btcEnhs conserved in mammals was higher than that of random genomic regions (**Fig. 6A and fig. S21A**). This trend was even stronger for btcEnhs with disease-associated SNPs than for those without them (**Fig. 6D**). We found that ~25% of disease-relevant btcEnhs showed a high degree of sequence similarity between human and mouse ($\geq 90\%$) (**Fig. 6E**). For example, a conserved disease-relevant btcEnh located in an *ACOXL* intron carries SNPs
 15 associated with rheumatoid arthritis and type 1 diabetes. The genomic region surrounding this btcEnh is also conserved between human and mouse (**Fig. 6F and fig. S21B**). An enhancer knockout study in mice showed that this enhancer regulates the expression of the pro-apoptotic gene *BCL2L1* in a context-specific manner (82). We also showed that CRISPRa-mediated activation of this enhancer resulted in specific upregulation of *BCL2L1*, but not the closest gene
 20 *ACOXL*, in Jurkat cells (**Fig. 6F**). The fact that some of the disease-relevant btcEnhs were primate-specific (**Fig. 6G**) indicates the importance of examining enhancer functionality in appropriate species. A primate-specific disease-relevant btcEnh overlapping with a LINE element contains IBD-SNPs (rs11150589 and rs12598978). The CRISPRa of this enhancer identified the integrin family gene *ITGAL* as the target, which was predicted by the ABC link
 25 (**Fig. 6G and Table S13**). Integrin molecules were implicated in the pathogenesis of IBD, and a better understanding of this pathway could provide insight into potential new therapeutic strategies (83).

Discussion

Although 3' scRNA-seq is widely used as a major tool for single-cell gene expression analysis, our data demonstrated the strength of 5' scRNA-seq. Combining ReapTEC with 5' (but not 3') scRNA-seq enabled accurate TSS analysis of mRNAs, lncRNAs, and eRNAs, making it possible to study the expression of not only genes but also btcEnhs. We also performed
 35 snATAC-seq using Multiome in diverse CD4⁺ T cells to identify accessible chromatin regions. The observation that heritability of immune-mediated diseases was substantially higher in btcEnhs than in accessible chromatin regions supports the importance of active enhancers with bidirectional eRNA transcription. This is in line with previous studies showing that eRNAs are robust predictors of active enhancers and can precisely delineate functional enhancers at high nucleotide resolution (3, 11, 19, 20). It is noteworthy that cell-type specificity was higher for
 40 eRNA transcription than for chromatin accessibility, indicating the importance of capturing eRNAs to better understand cellular diversity.

Our present study provides a platform for integrating and elucidating CD4⁺ T cell diversity and genetic risk of immune-mediated diseases. The CD4⁺ T cell type-resolved atlas of TSSs, 62,803 btcEnhs, and 218,508 ATAC peaks active across 962,944 cells and 136 cell clusters provides a
 45 rich resource publicly available at the Human Cell Atlas. We identified many rare and

uncharacterized cell types, including *GPR25*-expressing Tregs. By integrating these data with large-scale human genetics data, we showed the utility of our transcribed enhancer atlas in pinpointing regulatory genetic variation associated with human diseases, and further revealed many rare and understudied cell types that contribute to a wide range of immune-mediated diseases. Thus, our multimodal atlas provides many insights into the molecular pathways underpinning human diseases and can help address variant-to-function challenges, which has hindered the translational application of GWAS studies.

There are several inherent limitations to our current study. First, eRNAs are generally expressed at low levels, limiting detection sensitivity. Therefore, btcEnhs were identified by performing pseudo-bulk analysis of a large number of single cells that fall into the same cluster. In contrast, snATAC-seq has good sensitivity in identifying accessible chromatin regions. Although our 5' scRNA-seq approach and existing bulk/single-cell approaches to predict enhancers are complementary, they are also different, with each having their own unique advantages. Hence, by combining several synergetic and complementary approaches, such as 5' scRNA-seq and snATAC-seq, we can not only identify functional enhancers with high accuracy but also map other types of gene regulatory elements and their states. Second, we profiled circulating CD4⁺ T cells from a limited number of healthy donors but established our atlas by repeatedly sampling large volumes of blood and further added publicly available 5' scRNA-seq data on stimulated effector memory CD4⁺ T cells from dozens of donors. While circulating CD4⁺ T cells were a good starting point owing to considerable prior knowledge, future studies should be expanded to profiling tissue-resident CD4⁺ T cells in health and disease as well as other cell types and tissues. Third, we did not identify the exact transcription factors that bind to disease-relevant btcEnh or how disease variants alter the binding of the transcription factors. Further studies using base editing will help reveal the molecular functions of disease variants.

We anticipate that our publicly available CD4⁺ T cell atlas will be a useful resource to facilitate further in-depth understanding of the human immune system, including its evolutionary similarities to and differences from those of other mammals. The functionality of disease-relevant enhancers conserved in mammals can be potentially tested in enhancer-knockout mice *in vivo* (82, 84), whereas primate-specific enhancers can be studied using *in vitro* model systems such as organoids. Further functional validation experiments will shed light on new clinically relevant biomarkers and therapeutic targets, paving the way towards the development of innovative medicines.

Methods Summary

All experiments using human samples were approved by the ethical review committee of RIKEN [approval no. H30-9(13)]. Written informed consent was obtained from all donors. CD4⁺ T cells were isolated by the immunomagnetic negative selection method. Stained CD4⁺ T cells were sorted using a FACSaria IIu Cell Sorter (BD Biosciences). Human CD4⁺ T cells and FACS-sorted heterogeneous populations were processed with a Chromium Next GEM Single Cell 5' kit (10x Genomics). Libraries were sequenced on an Illumina NovaSeq 6000 sequencing platform using 2 × 150 bp paired-end sequencing. Multiome assay (10x Genomics) was performed according to the manufacturer's instructions. Multiome libraries were pooled and sequenced as above with 10 cycles for i7 index and 24 cycles for i5 index. Micro-C libraries were generated using a Dovetail Micro-C Kit (Cantata Bio, Cat#21006) and were sequenced on an Illumina NovaSeq 6000 platform using 2 × 150 bp paired-end sequencing. A 130-base single-

guide RNA (sgRNA) harboring an MS2 stem-loop sequence was transfected into CRISPRa-Jurkat cells expressing dCas9-VP64 and MS2_P65_HSF1. Transfected cells were subjected to CAGE assay. Chromium scRNA-seq, snRNA-seq, and CITE-seq data were processed using Cell Ranger Software version 5.0.1 (10x Genomics) (85) and R package Seurat version 5 (4.9.9.9067) (86). Multiome data were processed by Cell Ranger ARC version 2.0.0 (10x Genomics), Seurat version 5 (4.9.9.9067) (86), and Signac version 1.10.0 (87). scRNA-seq, snRNA-seq, and Multiome 3' snRNA-seq data were integrated using canonical correlation analysis. snATAC peaks were identified from fragment files of each cluster using MACS2 version 2.2.6 with default settings as implemented in Signac version 1.10.0 (87). For ReapTEC, paired-end reads were mapped again using STAR (STARsolo) (88) to obtain reads with unencoded G, which was tagged as a soft-clipped G by STARsolo. Reads were deduplicated, and those with the barcodes of each cell type were extracted. A count file was generated for each transcription start site (TSS) using the “bamToBed” function in BEDTools version 2.30.0 (89). TSS peaks were generated by merging TSSs located within 10 bp of each other. To identify btcEnhs, TSS peak pairs were detected using scripts provided at https://github.com/anderssonrobin/enhancers/blob/master/scripts/bidir_enhancers (11) with minor modifications. Micro-C data were processed with the dovetail_tools pipeline (Cantata Bio). Chromatin loop contacts were identified by the HiCCUPS algorithm (90) using the Juicer Tools package version 2.20.0 and the scale-space representation algorithm using the Mustache package (91). Loops were called at a 1-kb resolution with SCALE-normalized contact matrices for HiCCUPS and with ICE-normalized contact matrices for Mustache, and were filtered for an FDR < 0.05. The assessment of the links between human genetics findings and information on CD4⁺ T cell btcEnhs was based on (i) the enrichment of variants significantly associated with each disease (lead SNPs and linked SNPs with $R^2 > 0.8$) in the btcEnh regions, which was calculated using GREGOR (92), and (ii) LD score regression to evaluate the enrichment of heritability of immune-mediated diseases in btcEnhs, which was computed using ldsc software (v1.0) (75).

References and Notes

1. Roadmap Epigenomics Consortium, A. Kundaje, W. Meuleman, J. Ernst, M. Bilenky, A. Yen, A. Heravi-Moussavi, P. Kheradpour, Z. Zhang, J. Wang, M. J. Ziller, V. Amin, J. W. Whitaker, M. D. Schultz, L. D. Ward, A. Sarkar, G. Quon, R. S. Sandstrom, M. L. Eaton, Y.-C. Wu, A. R. Pfenning, X. Wang, M. Claussnitzer, Y. Liu, C. Coarfa, R. A. Harris, N. Shores, C. B. Epstein, E. Gjoneska, D. Leung, W. Xie, R. D. Hawkins, R. Lister, C. Hong, P. Gascard, A. J. Mungall, R. Moore, E. Chuah, A. Tam, T. K. Canfield, R. S. Hansen, R. Kaul, P. J. Sabo, M. S. Bansal, A. Carles, J. R. Dixon, K.-H. Farh, S. Feizi, R. Karlic, A.-R. Kim, A. Kulkarni, D. Li, R. Lowdon, G. Elliott, T. R. Mercer, S. J. Neph, V. Onuchic, P. Polak, N. Rajagopal, P. Ray, R. C. Sallari, K. T. Siebenthal, N. A. Sinnott-Armstrong, M. Stevens, R. E. Thurman, J. Wu, B. Zhang, X. Zhou, A. E. Beaudet, L. A. Boyer, P. L. De Jager, P. J. Farnham, S. J. Fisher, D. Haussler, S. J. M. Jones, W. Li, M. A. Marra, M. T. McManus, S. Sunyaev, J. A. Thomson, T. D. Tlsty, L.-H. Tsai, W. Wang, R. A. Waterland, M. Q. Zhang, L. H. Chadwick, B. E. Bernstein, J. F. Costello, J. R. Ecker, M. Hirst, A. Meissner, A. Milosavljevic, B. Ren, J. A. Stamatoyannopoulos, T. Wang, M. Kellis, Integrative analysis of 111 reference human epigenomes. *Nature*. **518**, 317–30 (2015).

2. F. Abascal, R. Acosta, N. J. Addleman, J. Adrian, V. Afzal, B. Aken, J. A. Akiyama, O. Al Jammal, H. Amrhein, S. M. Anderson, G. R. Andrews, I. Antoshechkin, K. G. Ardlie, J. Armstrong, M. Astley, B. Banerjee, A. A. Barkal, I. H. A. Barnes, I. Barozzi, D. Barrell, G. Barson, D. Bates, U. K. Baymuradov, C. Bazile, M. A. Beer, S. Beik, M. A. Bender, R. Bennett, L. P. B. Bouvrette, B. E. Bernstein, A. Berry, A. Bhaskar, A. Bignell, S. M. Blue, D. M. Bodine, C. Boix, N. Boley, T. Borrman, B. Borsari, A. P. Boyle, L. A. Brandsmeier, A. Breschi, E. H. Bresnick, J. A. Brooks, M. Buckley, C. B. Burge, R. Byron, E. Cahill, L. Cai, L. Cao, M. Carty, R. G. Castanon, A. Castillo, H. Chaib, E. T. Chan, D. R. Chee, S. Chee, H. Chen, H. Chen, J. Y. Chen, S. Chen, J. M. Cherry, S. B. Chhetri, J. S. Choudhary, J. Chrast, D. Chung, D. Clarke, N. A. L. Cody, C. J. Coppola, J. Coursen, A. M. D'Ippolito, S. Dalton, C. Danyko, C. Davidson, J. Davila-Velderrain, C. A. Davis, J. Dekker, A. Deran, G. DeSalvo, G. Despacio-Reyes, C. N. Dewey, D. E. Dickel, M. Diegel, M. Diekhans, V. Dileep, B. Ding, S. Djebali, A. Dobin, D. Dominguez, S. Donaldson, J. Drenkow, T. R. Dreszer, Y. Drier, M. O. Duff, D. Dunn, C. Eastman, J. R. Ecker, M. D. Edwards, N. El-Ali, S. I. Elhajjajy, K. Elkins, A. Emili, C. B. Epstein, R. C. Evans, I. Ezkurdia, K. Fan, P. J. Farnham, N. P. Farrell, E. A. Feingold, A. M. Ferreira, K. Fisher-Aylor, S. Fitzgerald, P. Flicek, C. S. Foo, K. Fortier, A. Frankish, P. Freese, S. Fu, X. D. Fu, Y. Fu, Y. Fukuda-Yuzawa, M. Fulciniti, A. P. W. Funnell, I. Gabdank, T. Galeev, M. Gao, C. G. Giron, T. H. Garvin, C. A. Gelboin-Burkhart, G. Georgolopoulos, M. B. Gerstein, B. M. Giardine, D. K. Gifford, D. M. Gilbert, D. A. Gilchrist, S. Gillespie, T. R. Gingeras, P. Gong, A. Gonzalez, J. M. Gonzalez, P. Good, A. Goren, D. U. Gorkin, B. R. Graveley, M. Gray, J. F. Greenblatt, E. Griffiths, M. T. Groudine, F. Grubert, M. Gu, R. Guigó, H. Guo, Y. Guo, Y. Guo, G. Gursoy, M. Gutierrez-Arcelus, J. Halow, R. C. Hardison, M. Hardy, M. Hariharan, A. Harmanci, A. Harrington, J. L. Harrow, T. B. Hashimoto, R. D. Hasz, M. Hatan, E. Haugen, J. E. Hayes, P. He, Y. He, N. Heidari, D. Hendrickson, E. F. Heuston, J. A. Hilton, B. C. Hitz, A. Hochman, C. Holgren, L. Hou, S. Hou, Y. H. E. Hsiao, S. Hsu, H. Huang, T. J. Hubbard, J. Huey, T. R. Hughes, T. Hunt, S. Ibarrientos, R. Issner, M. Iwata, O. Izuogu, T. Jaakkola, N. Jameel, C. Jansen, L. Jiang, P. Jiang, A. Johnson, R. Johnson, I. Jungreis, M. Kadaba, M. Kasowski, M. Kasparian, M. Kato, R. Kaul, T. Kawli, M. Kay, J. C. Keen, S. Keles, C. A. Keller, D. Kelley, M. Kellis, P. Kheradpour, D. S. Kim, A. Kirilusha, R. J. Klein, B. Knoechel, S. Kuan, M. J. Kulik, S. Kumar, A. Kundaje, T. Kutayavin, J. Lagarde, B. R. Lajoie, N. J. Lambert, J. Lazar, A. Y. Lee, D. Lee, E. Lee, J. W. Lee, K. Lee, C. S. Leslie, S. Levy, B. Li, H. Li, N. Li, X. Li, Y. I. Li, Y. Li, Y. Li, Y. Li, J. Lian, M. W. Libbrecht, S. Lin, Y. Lin, D. Liu, J. Liu, P. Liu, T. Liu, X. S. Liu, Y. Liu, Y. Liu, M. Long, S. Lou, J. Loveland, A. Lu, Y. Lu, E. Lécuyer, L. Ma, M. Mackiewicz, B. J. Mannion, M. Mannstadt, D. Manthravadi, G. K. Marinov, F. J. Martin, E. Mattei, K. McCue, M. McEown, G. McVicker, S. K. Meadows, A. Meissner, E. M. Mendenhall, C. L. Messer, W. Meuleman, C. Meyer, S. Miller, M. G. Milton, T. Mishra, D. E. Moore, H. M. Moore, J. E. Moore, S. H. Moore, J. Moran, A. Mortazavi, J. M. Mudge, N. Munshi, R. Murad, R. M. Myers, V. Nandakumar, P. Nandi, A. M. Narasimha, A. K. Narayanan, H. Naughton, F. C. P. Navarro, P. Navas, J. Nazarovs, J. Nelson, S. Neph, F. J. Neri, J. R. Nery, A. R. Nesmith, J. S. Newberry, K. M. Newberry, V. Ngo, R. Nguyen, T. B. Nguyen, T. Nguyen, A. Nishida, W. S. Noble, C. S. Novak, E. M. Novoa, B. Nuñez, C. W. O'Donnell, S. Olson, K. C. Onate, E. Otterman, H. Ozadam, M. Pagan, T. Palden, X. Pan, Y. Park, E. C. Partridge, B. Paten, F. Pauli-Behn, M. J. Pazin, B. Pei, L. A. Pennacchio, A. R. Perez, E. H. Perry, D. D. Pervouchine, N. N. Phalke, Q. Pham, D. H. Phanstiel, I. Plajzer-Frick, G. A. Pratt, H. E. Pratt, S. Preissl, J. K. Pritchard, Y. Pritykin, M. J. Purcaro, Q. Qin, G. Quinones-Valdez, I. Rabano, E.

- Radovani, A. Raj, N. Rajagopal, O. Ram, L. Ramirez, R. N. Ramirez, D. Rausch, S. Raychaudhuri, J. Raymond, R. Razavi, T. E. Reddy, T. M. Reimonn, B. Ren, A. Reymond, A. Reynolds, S. K. Rhie, J. Rinn, M. Rivera, J. C. Rivera-Mulia, B. S. Roberts, J. M. Rodriguez, J. Rozowsky, R. Ryan, E. Rynes, D. N. Salins, R. Sandstrom, T. Sasaki, S. Sathe, D. Savic, A. Scavelli, J. Scheiman, C. Schlaffner, J. A. Schloss, F. W. Schmitges, L. H. See, A. Sethi, M. Setty, A. Shafer, S. Shan, E. Sharon, Q. Shen, Y. Shen, R. I. Sherwood, M. Shi, S. Shin, N. Shores, K. Siebenthal, C. Sis, T. Slifer, C. A. Sloan, A. Smith, V. Snetkova, M. P. Snyder, D. V. Spacek, S. Srinivasan, R. Srivas, G. Stamatoyannopoulos, J. A. Stamatoyannopoulos, R. Stanton, D. Steffan, S. Stehling-Sun, J. S. Strattan, A. Su, B. Sundararaman, M. M. Suner, T. Syed, M. Szynekarek, F. Y. Tanaka, D. Tenen, M. Teng, J. A. Thomas, D. Toffey, M. L. Tress, D. E. Trout, G. Trynka, J. Tsuji, S. A. Upchurch, O. Ursu, B. Uszczynska-Ratajczak, M. C. Uziel, A. Valencia, B. Van Biber, A. G. van der Velde, E. L. Van Nostrand, Y. Vaydylevich, J. Vazquez, A. Victorsen, J. Vielmetter, J. Vierstra, A. Visel, A. Vlasova, C. M. Vockley, S. Volpi, S. Vong, H. Wang, M. Wang, Q. Wang, R. Wang, T. Wang, W. Wang, X. Wang, Y. Wang, N. K. Watson, X. Wei, Z. Wei, H. Weisser, S. M. Weissman, R. Welch, R. E. Welikson, Z. Weng, H. J. Westra, J. W. Whitaker, C. White, K. P. White, A. Wildberg, B. A. Williams, D. Wine, H. N. Witt, B. Wold, M. Wolf, J. Wright, R. Xiao, X. Xiao, J. Xu, J. Xu, K. K. Yan, Y. Yan, H. Yang, X. Yang, Y. W. Yang, G. G. Yardımcı, B. A. Yee, G. W. Yeo, T. Young, T. Yu, F. Yue, C. Zaleski, C. Zang, H. Zeng, W. Zeng, D. R. Zerbino, J. Zhai, L. Zhan, Y. Zhan, B. Zhang, J. Zhang, J. Zhang, K. Zhang, L. Zhang, P. Zhang, Q. Zhang, X. O. Zhang, Y. Zhang, Z. Zhang, Y. Zhao, Y. Zheng, G. Zhong, X. Q. Zhou, Y. Zhu, J. Zimmerman, J. E. Moore, M. J. Purcaro, H. E. Pratt, C. B. Epstein, N. Shores, J. Adrian, T. Kawli, C. A. Davis, A. Dobin, R. Kaul, J. Halow, E. L. Van Nostrand, P. Freese, D. U. Gorkin, Y. Shen, Y. He, M. Mackiewicz, F. Pauli-Behn, B. A. Williams, A. Mortazavi, C. A. Keller, X. O. Zhang, S. I. Elhajjajy, J. Huey, D. E. Dickel, V. Snetkova, X. Wei, X. Wang, J. C. Rivera-Mulia, J. Rozowsky, J. Zhang, S. B. Chhetri, J. Zhang, A. Victorsen, K. P. White, A. Visel, G. W. Yeo, C. B. Burge, E. Lécuyer, D. M. Gilbert, J. Dekker, J. Rinn, E. M. Mendenhall, J. R. Ecker, M. Kellis, R. J. Klein, W. S. Noble, A. Kundaje, R. Guigó, P. J. Farnham, J. M. Cherry, R. M. Myers, B. Ren, B. R. Graveley, M. B. Gerstein, L. A. Pennacchio, M. P. Snyder, B. E. Bernstein, B. Wold, R. C. Hardison, T. R. Gingeras, J. A. Stamatoyannopoulos, Z. Weng, Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature*. **583**, 699–710 (2020).
3. Y. Murakawa, M. Yoshihara, H. Kawaji, M. Nishikawa, H. Zayed, H. Suzuki, Y. Hayashizaki, Enhanced Identification of Transcriptional Enhancers Provides Mechanistic Insights into Diseases. *Trends Genet.* **32**, 76–88 (2016).
4. R. E. Thurman, E. Rynes, R. Humbert, J. Vierstra, M. T. Maurano, E. Haugen, N. C. Sheffield, A. B. Stergachis, H. Wang, B. Vernot, K. Garg, S. John, R. Sandstrom, D. Bates, L. Boatman, T. K. Canfield, M. Diegel, D. Dunn, A. K. Ebersol, T. Frum, E. Giste, A. K. Johnson, E. M. Johnson, T. Kutayavin, B. Lajoie, B. K. Lee, K. Lee, D. London, D. Lotakis, S. Neph, F. Neri, E. D. Nguyen, H. Qu, A. P. Reynolds, V. Roach, A. Safi, M. E. Sanchez, A. Sanyal, A. Shafer, J. M. Simon, L. Song, S. Vong, M. Weaver, Y. Yan, Z. Zhang, Z. Zhang, B. Lenhard, M. Tewari, M. O. Dorschner, R. S. Hansen, P. A. Navas, G. Stamatoyannopoulos, V. R. Iyer, J. D. Lieb, S. R. Sunyaev, J. M. Akey, P. J. Sabo, R. Kaul, T. S. Furey, J. Dekker, G. E. Crawford, J. A. Stamatoyannopoulos, The accessible chromatin landscape of the human genome. *Nature*. **489**, 75–82 (2012).

5. Y. E. Li, S. Preissl, X. Hou, Z. Zhang, K. Zhang, Y. Qiu, O. B. Poirion, B. Li, J. Chiou, H. Liu, A. Pinto-Duarte, N. Kubo, X. Yang, R. Fang, X. Wang, J. Y. Han, J. Lucero, Y. Yan, M. Miller, S. Kuan, D. Gorkin, K. J. Gaulton, Y. Shen, M. Nunn, E. A. Mukamel, M. M. Behrens, J. R. Ecker, B. Ren, An atlas of gene regulatory elements in adult mouse cerebrum. *Nature*. **598**, 129–136 (2021).
6. K. Zhang, J. D. Hocker, M. Miller, X. Hou, J. Chiou, O. B. Poirion, Y. Qiu, Y. E. Li, K. J. Gaulton, A. Wang, S. Preissl, B. Ren, A single-cell atlas of chromatin accessibility in the human genome. *Cell*. **184**, 5985–6001.e19 (2021).
7. Y. E. Li, S. Preissl, M. Miller, N. D. Johnson, Z. Wang, H. Jiao, C. Zhu, Z. Wang, Y. Xie, O. Poirion, C. Kern, A. Pinto-Duarte, W. Tian, K. Siletti, N. Emerson, J. Osteen, J. Lucero, L. Lin, Q. Yang, Q. Zhu, N. Zemke, S. Espinoza, A. M. Yanny, J. Nyhus, N. Dee, T. Casper, N. Shapovalova, D. Hirschstein, R. D. Hodge, S. Linnarsson, T. Bakken, B. Levi, C. D. Keene, J. Shang, E. Lein, A. Wang, M. M. Behrens, J. R. Ecker, B. Ren, A comparative atlas of single-cell chromatin accessibility in the human brain. *Science*. **382**, eadf7044 (2023).
8. M. P. Creighton, A. W. Cheng, G. G. Welstead, T. Kooistra, B. W. Carey, E. J. Steine, J. Hanna, M. A. Lodato, G. M. Frampton, P. A. Sharp, L. A. Boyer, R. A. Young, R. Jaenisch, Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 21931–21936 (2010).
9. H. Kwak, N. J. Fuda, L. J. Core, J. T. Lis, Precise Maps of RNA Polymerase Reveal How Promoters Direct Initiation and Pausing. *Science (80-.)*. **339**, 950–953 (2013).
10. S. Hirabayashi, S. Bhagat, Y. Matsuki, Y. Takegami, T. Uehata, A. Kanemaru, M. Itoh, K. Shirakawa, A. Takaori-Kondo, O. Takeuchi, P. Carninci, S. Katayama, Y. Hayashizaki, J. Kere, H. Kawaji, Y. Murakawa, NET-CAGE characterizes the dynamics and topology of human transcribed cis-regulatory elements. *Nat. Genet.* **51**, 1369–1379 (2019).
11. R. Andersson, C. Gebhard, I. Miguel-Escalada, I. Hoof, J. Bornholdt, M. Boyd, Y. Chen, X. Zhao, C. Schmidl, T. Suzuki, E. Ntini, E. Arner, E. Valen, K. Li, L. Schwarzfischer, D. Glatz, J. Raithel, B. Lilje, N. Rapin, F. O. Bagger, M. Jørgensen, P. R. Andersen, N. Bertin, O. Rackham, A. M. Burroughs, J. K. Baillie, Y. Ishizu, Y. Shimizu, E. Furuhashi, S. Maeda, Y. Negishi, C. J. Mungall, T. F. Meehan, T. Lassmann, M. Itoh, H. Kawaji, N. Kondo, J. Kawai, A. Lennartsson, C. O. Daub, P. Heutink, D. A. Hume, T. H. Jensen, H. Suzuki, Y. Hayashizaki, F. Müller, A. R. R. Forrest, P. Carninci, M. Rehli, A. Sandelin, An atlas of active enhancers across human cell types and tissues. *Nature*. **507**, 455–461 (2014).
12. E. Arner, C. O. Daub, K. Vitting-Seerup, R. Andersson, B. Lilje, F. Drabløs, A. Lennartsson, M. Rönnerblad, O. Hrydziuszko, M. Vitezic, T. C. Freeman, A. M. N. Alhendi, P. Arner, R. Axton, J. K. Baillie, A. Beckhouse, B. Bodega, J. Briggs, F. Brombacher, M. Davis, M. Detmar, A. Ehrlund, M. Endoh, A. Eslami, M. Fagiolini, L. Fairbairn, G. J. Faulkner, C. Ferrai, M. E. Fisher, L. Forrester, D. Goldowitz, R. Guler, T. Ha, M. Hara, M. Herlyn, T. Ikawa, C. Kai, H. Kawamoto, L. M. Khachigian, S. P. Klinken, S. Kojima, H. Koseki, S. Klein, N. Mejhert, K. Miyaguchi, Y. Mizuno, M. Morimoto, K. J. Morris, C. Mummery, Y. Nakachi, S. Ogishima, M. Okada-Hatakeyama, Y. Okazaki, V. Orlando, D. Ovchinnikov, R. Passier, M. Patrikakis, A. Pombo, X.-Y. Qin, S. Roy, H. Sato, S. Savvi, A. Saxena, A. Schwegmann, D. Sugiyama, R. Swoboda, H.

Tanaka, A. Tomoiu, L. N. Winteringham, E. Wolvetang, C. Yanagi-Mizuochi, M. Yoneda, S. Zabierowski, P. Zhang, I. Abugessaisa, N. Bertin, A. D. Diehl, S. Fukuda, M. Furuno, J. Harshbarger, A. Hasegawa, F. Hori, S. Ishikawa-Kato, Y. Ishizu, M. Itoh, T. Kawashima, M. Kojima, N. Kondo, M. Lizio, T. F. Meehan, C. J. Mungall, M. Murata, H. Nishiyori-Sueki, S. Sahin, S. Nagao-Sato, J. Severin, M. J. L. de Hoon, J. Kawai, T. Kasukawa, T. Lassmann, H. Suzuki, H. Kawaji, K. M. Summers, C. Wells, D. A. Hume, A. R. R. Forrest, A. Sandelin, P. Carninci, Y. Hayashizaki, Transcribed enhancers lead waves of coordinated transcription in transitioning mammalian cells. *Science* (80-.). **347**, 1010–1014 (2015).

13. T. K. Kim, M. Hemberg, J. M. Gray, A. M. Costa, D. M. Bear, J. Wu, D. A. Harmin, M. Laptewicz, K. Barbara-Haley, S. Kuersten, E. Markenscoff-Papadimitriou, D. Kuhl, H. Bito, P. F. Worley, G. Kreiman, M. E. Greenberg, Widespread transcription at neuronal activity-regulated enhancers. *Nature*. **465**, 182–187 (2010).

14. V. Sartorelli, S. M. Lauberth, Enhancer RNAs are an important regulatory layer of the epigenome. *Nat. Struct. Mol. Biol.* **27**, 521–528 (2020).

15. C. G. Danko, S. L. Hyland, L. J. Core, A. L. Martins, C. T. Waters, H. W. Lee, V. G. Cheung, W. L. Kraus, J. T. Lis, A. Siepel, Identification of active transcriptional regulatory elements from GRO-seq data. *Nat. Methods*. **12**, 433–438 (2015).

16. J. G. Azofeifa, M. A. Allen, J. R. Hendrix, T. Read, J. D. Rubin, R. D. Dowell, Enhancer RNA profiling predicts transcription factor activity. *Genome Res.* **28**, 334–344 (2018).

17. C. G. Danko, L. A. Choate, B. A. Marks, E. J. Rice, Z. Wang, T. Chu, A. L. Martins, N. Dukler, S. A. Coonrod, E. D. Tait Wojno, J. T. Lis, W. L. Kraus, A. Siepel, Dynamic evolution of regulatory element ensembles in primate CD4⁺ T cells. *Nat. Ecol. Evol.* **2**, 537–548 (2018).

18. J. M. Tome, N. D. Tippens, J. T. Lis, Single-molecule nascent RNA sequencing identifies regulatory domain architecture at promoters and enhancers. *Nat. Genet.* **50**, 1533–1541 (2018).

19. O. Mikhaylichenko, V. Bondarenko, D. Harnett, I. E. Schor, M. Males, R. R. Viales, E. E. M. Furlong, The degree of enhancer or promoter activity is reflected by the levels and directionality of eRNA transcription. *Genes Dev.* **32**, 42–57 (2018).

20. N. D. Tippens, J. Liang, A. K. Y. Leung, S. D. Wierbowski, A. Ozer, J. G. Booth, J. T. Lis, H. Yu, Transcription imparts architecture, function and logic to enhancer units. *Nat. Genet.* **52**, 1067–1075 (2020).

21. S. C. Biddie, G. Weykopf, E. F. Hird, E. T. Friman, W. A. Bickmore, DNA-binding factor footprints and enhancer RNAs identify functional non- coding genetic variants (2023), doi:<https://doi.org/10.1101/2023.11.20.567860>.

22. B. V. Kumar, T. J. Connors, D. L. Farber, Human T Cell Development, Localization, and Function throughout Life. *Immunity*. **48**, 202–213 (2018).

23. V. Matzaraki, V. Kumar, C. Wijmenga, A. Zhernakova, The MHC locus and genetic susceptibility to autoimmune and infectious diseases. *Genome Biol.* **18** (2017), doi:10.1186/s13059-017-1207-1.

24. H. T. Maecker, J. P. McCoy, R. Nussenblatt, Standardizing immunophenotyping for the Human Immunology Project. *Nat. Rev. Immunol.* **12**, 191–200 (2012).
25. M. J. T. Stubbington, O. Rozenblatt-Rosen, A. Regev, S. A. Teichmann, Single-cell transcriptomics to explore the immune system in health and disease. *Science (80-.)*. **358**, 58–63 (2017).
26. R. J. Miragaia, T. Gomes, A. Chomka, L. Jardine, A. Riedel, A. N. Hegazy, N. Whibley, A. Tucci, X. Chen, I. Lindeman, G. Emerton, T. Krausgruber, J. Shields, M. Haniffa, F. Powrie, S. A. Teichmann, Single-Cell Transcriptomics of Regulatory T Cells Reveals Trajectories of Tissue Adaptation. *Immunity*. **50**, 493–504.e7 (2019).
27. B. Soskic, E. Cano-Gamez, D. J. Smyth, K. Ambridge, Z. Ke, J. C. Matte, L. Bossini-Castillo, J. Kaplanis, L. Ramirez-Navarro, A. Lorenc, N. Nakic, J. Esparza-Gordillo, W. Rowan, D. Wille, D. F. Tough, P. G. Bronson, G. Trynka, Immune disease risk variants regulate gene expression dynamics during CD4⁺ T cell activation. *Nat. Genet.* **54**, 817–826 (2022).
28. C. Suo, E. Dann, I. Goh, L. Jardine, V. Kleshchevnikov, J. E. Park, R. A. Botting, E. Stephenson, J. Engelbert, Z. K. Tuong, K. Polanski, N. Yayon, C. Xu, O. Suchanek, R. Elmentaite, C. D. Conde, P. He, S. Pritchard, M. Miah, C. Moldovan, A. S. Steemers, P. Mazin, M. Prete, D. Horsfall, J. C. Marioni, M. R. Clatworthy, M. Haniffa, S. A. Teichmann, IMMUNOLOGY Mapping the developing human immune system across organs. *Science (80-.)*. **376** (2022), doi:10.1126/science.abo0510.
29. P. A. Szabo, H. M. Levitin, M. Miron, M. E. Snyder, T. Senda, J. Yuan, Y. L. Cheng, E. C. Bush, P. Dogra, P. Thapa, D. L. Farber, P. A. Sims, Single-cell transcriptomics of human T cells reveals tissue and activation signatures in health and disease. *Nat. Commun.* **10** (2019), doi:10.1038/s41467-019-12464-3.
30. E. Cano-Gamez, B. Soskic, T. I. Roumeliotis, E. So, D. J. Smyth, M. Baldrighi, D. Willé, N. Nakic, J. Esparza-Gordillo, C. G. C. Larminie, P. G. Bronson, D. F. Tough, W. C. Rowan, J. S. Choudhary, G. Trynka, Single-cell transcriptomics identifies an effectorness gradient shaping the response of CD4⁺ T cells to cytokines. *Nat. Commun.* **11**, 1–15 (2020).
31. J. Park, R. A. Botting, C. Domínguez Conde, D. Popescu, M. Lavaert, D. J. Kunz, I. Goh, E. Stephenson, R. Ragazzini, E. Tuck, A. Wilbrey-Clark, K. Roberts, V. R. Kedlian, J. R. Ferdinand, X. He, S. Webb, D. Maunder, N. Vandamme, K. T. Mahbubani, K. Polanski, L. Mamanova, L. Bolt, D. Crossland, F. de Rita, A. Fuller, A. Filby, G. Reynolds, D. Dixon, K. Saeb-Parsy, S. Lisgo, D. Henderson, R. Vento-Tormo, O. A. Bayraktar, R. A. Barker, K. B. Meyer, Y. Saeys, P. Bonfanti, S. Behjati, M. R. Clatworthy, T. Taghon, M. Haniffa, S. A. Teichmann, A cell atlas of human thymic development defines T cell repertoire formation. *Science (80-.)*. **367** (2020), doi:10.1126/science.aay3224.
32. A. Nathan, J. I. Beynor, Y. Baglaenko, S. Suliman, K. Ishigaki, S. Asgari, C. C. Huang, Y. Luo, Z. Zhang, K. Lopez, C. S. Lindestam Arlehamn, J. D. Ernst, J. Jimenez, R. I. Calderón, L. Lecca, I. Van Rhijn, D. B. Moody, M. B. Murray, S. Raychaudhuri, Multimodally profiling memory T cells from a tuberculosis cohort identifies cell state associations with demographics, environment and disease. *Nat. Immunol.* **22**, 781–793 (2021).

33. L. Zheng, S. Qin, W. Si, A. Wang, B. Xing, R. Gao, X. Ren, L. Wang, X. Wu, J. Zhang, N. Wu, N. Zhang, H. Zheng, H. Ouyang, K. Chen, Z. Bu, X. Hu, J. Ji, Z. Zhang, Pan-cancer single-cell landscape of tumor-infiltrating T cells. *Science* (80-.). **374** (2021), doi:10.1126/science.abe6474.
34. C. Domínguez Conde, C. Xu, L. B. Jarvis, D. B. Rainbow, S. B. Wells, T. Gomes, S. K. Howlett, O. Suchanek, K. Polanski, H. W. King, L. Mamanova, N. Huang, P. A. Szabo, L. Richardson, L. Bolt, E. S. Fasouli, K. T. Mahbubani, M. Prete, L. Tuck, N. Richoz, Z. K. Tuong, L. Campos, H. S. Mousa, E. J. Needham, S. Pritchard, T. Li, R. Elmentaite, J. Park, E. Rahmani, D. Chen, D. K. Menon, O. A. Bayraktar, L. K. James, K. B. Meyer, N. Yosef, M. R. Clatworthy, P. A. Sims, D. L. Farber, K. Saeb-Parsy, J. L. Jones, S. A. Teichmann, Cross-tissue immune cell analysis reveals tissue-specific features in humans. *Science*. **376**, eabl5197 (2022).
35. B. J. Schmiedel, C. Gonzalez-Colin, V. Fajardo, J. Rocha, A. Madrigal, C. Ramírez-Suástegui, S. Bhattacharyya, H. Simon, J. A. Greenbaum, B. Peters, G. Seumois, F. Ay, V. Chandra, P. Vijayanand, Single-cell eQTL analysis of activated T cell subsets reveals activation and cell type-dependent effects of disease-risk variants. *Sci. Immunol.* **7**, eabm2508 (2022).
36. D. A. Rao, M. F. Gurish, J. L. Marshall, K. Slowikowski, C. Y. Fonseka, Y. Liu, L. T. Donlin, L. A. Henderson, K. Wei, F. Mizoguchi, N. C. Teslovich, M. E. Weinblatt, E. M. Massarotti, J. S. Coblyn, S. M. Helfgott, Y. C. Lee, D. J. Todd, V. P. Bykerk, S. M. Goodman, A. B. Pernis, L. B. Ivashkiv, E. W. Karlson, P. A. Nigrovic, A. Filer, C. D. Buckley, J. A. Lederer, S. Raychaudhuri, M. B. Brenner, Pathologically expanded peripheral T helper cell subset drives B cells in rheumatoid arthritis. *Nature*. **542**, 110–114 (2017).
37. C. A. Boix, B. T. James, Y. P. Park, W. Meuleman, M. Kellis, Regulatory genomic circuitry of human disease loci by integrative epigenomics. *Nature*. **590**, 300–307 (2021).
38. E. Cano-Gamez, G. Trynka, From GWAS to Function: Using Functional Genomics to Identify the Mechanisms Underlying Complex Diseases. *Front. Genet.* **11**, 1–21 (2020).
39. J. Ernst, P. Kheradpour, T. S. Mikkelsen, N. Shores, L. D. Ward, C. B. Epstein, X. Zhang, L. Wang, R. Issner, M. Coyne, M. Ku, T. Durham, M. Kellis, B. E. Bernstein, Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature*. **473**, 43–49 (2011).
40. M. T. Maurano, R. Humbert, E. Rynes, R. E. Thurman, E. Haugen, H. Wang, A. P. Reynolds, R. Sandstrom, H. Qu, J. Brody, A. Shafer, F. Neri, K. Lee, T. Kuttyavin, S. Stehling-Sun, A. K. Johnson, T. K. Canfield, E. Giste, M. Diegel, D. Bates, R. S. Hansen, S. Neph, P. J. Sabo, S. Heimfeld, A. Raubitschek, S. Ziegler, C. Cotsapas, N. Sotoodehnia, I. Glass, S. R. Sunyaev, R. Kaul, J. A. Stamatoyannopoulos, Systematic localization of common disease-associated variation in regulatory DNA. *Science* (80-.). **337**, 1190–1195 (2012).
41. K. K. H. Farh, A. Marson, J. Zhu, M. Kleinewietfeld, W. J. Housley, S. Beik, N. Shores, H. Whitton, R. J. H. Ryan, A. A. Shishkin, M. Hatan, M. J. Carrasco-Alfonso, D. Mayer, C. J. Luckey, N. A. Patsopoulos, P. L. De Jager, V. K. Kuchroo, C. B. Epstein, M. J. Daly, D. A. Hafler, B. E. Bernstein, Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature*. **518**, 337–343 (2015).

42. M. Boyd, M. Thodberg, M. Vitezic, J. Bornholdt, K. Vitting-Seerup, Y. Chen, M. Coskun, Y. Li, B. Z. S. Lo, P. Klausen, P. Jan Schweiger, A. G. Pedersen, N. Rapin, K. Skovgaard, K. Dahlgaard, R. Andersson, T. B. Terkelsen, B. Lilje, J. T. Troelsen, A. M. Petersen, K. B. Jensen, I. Gögenur, P. Thielsen, J. B. Seidelin, O. H. Nielsen, J. T. Bjerrum, A. Sandelin, Characterization of the enhancer and promoter landscape of inflammatory bowel disease from human colon biopsies. *Nat. Commun.* **9**, 1661 (2018).
43. N. Alexi, R. H. Inge, G. C. Nicole, C. M. S. Johannes, Y. Miao, H. Rong, Z. H. Claudia, P. Monique, X. Jiayang, W. Yin, K. Zahara, P. P. Martina, O. Carolyn, K. N. Christian, T. S. Simon, S. Zeyang, C. K. Glass, Brain cell type – specific enhancer – promoter interactome maps and disease-risk association. *Science (80-.).* **1139**, 1134–1139 (2019).
44. W. Meuleman, A. Muratov, E. Rynes, J. Halow, K. Lee, D. Bates, M. Diegel, D. Dunn, F. Neri, A. Teodosiadis, A. Reynolds, E. Haugen, J. Nelson, A. Johnson, M. Frerker, M. Buckley, R. Sandstrom, J. Vierstra, R. Kaul, J. Stamatoyannopoulos, Index and biological spectrum of human DNase I hypersensitive sites. *Nature.* **584**, 244–251 (2020).
45. M. Stoeckius, C. Hafemeister, W. Stephenson, B. Houck-Loomis, P. K. Chattopadhyay, H. Swerdlow, R. Satija, P. Smibert, Simultaneous epitope and transcriptome measurement in single cells. *Nat. Methods.* **14** (2017), doi:10.1038/nmeth.4380.
46. J. Moody, T. Kouno, J. C. Chang, Y. Ando, P. Carninci, J. W. Shin, C. C. Hon, SCAFE: a software suite for analysis of transcribed cis-regulatory elements in single cells. *Bioinformatics.* **38**, 5126–5128 (2022).
47. R. Hou, C. C. Hon, Y. Huang, CamoTSS: analysis of alternative transcription start sites for cellular phenotypes and regulatory patterns from 5' scRNA-seq data. *Nat. Commun.* **14** (2023), doi:10.1038/s41467-023-42636-1.
48. W. Mckerrow, L. Kagermazova, N. Doudican, N. Frazzette, E. I. Kaparos, S. A. Evans, A. Rocha, J. M. Sedivy, N. Neretti, J. Carucci, J. D. Boeke, D. Fenyö, LINE-1 retrotransposon expression in cancerous, epithelial and neuronal cells revealed by 5' single-cell RNA-Seq. *Nucleic Acids Res.* **51**, 2033–2045 (2023).
49. D. T. P. Tang, C. Plessy, M. Salimullah, A. M. Suzuki, R. Calligaris, S. Gustincich, P. Carninci, Suppression of artifacts and barcode bias in high-throughput transcriptome analyses utilizing template switching. *Nucleic Acids Res.* **41** (2013), doi:10.1093/nar/gks1128.
50. X. Adiconis, A. L. Haber, S. K. Simmons, A. Levy Moonshine, Z. Ji, M. A. Busby, X. Shi, J. Jacques, M. A. Lancaster, J. Q. Pan, A. Regev, J. Z. Levin, Comprehensive comparative analysis of 5'-end RNA-sequencing methods. *Nat. Methods.* **15**, 505–511 (2018).
51. H. Ohtake, K. Ohtoko, Y. Ishimaru, S. Kato, Determination of the capped site sequence of mRNA based on the detection of cap-dependent nucleotide addition using an anchor ligation method. *DNA Res.* **11**, 305–309 (2004).
52. A. M. Al'Khafaji, J. T. Smith, K. V. Garimella, M. Babadi, V. Popic, M. Sade-Feldman, M. Gatzen, S. Sarkizova, M. A. Schwartz, E. M. Blaum, A. Day, M. Costello, T. Bowers, S. Gabriel, E. Banks, A. A. Philippakis, G. M. Boland, P. C. Blainey, N. Hacohen, High-throughput RNA isoform sequencing using programmed cDNA concatenation. *Nat. Biotechnol.* (2023), doi:10.1038/s41587-023-01815-7.

53. A. R. R. Forrest, H. Kawaji, M. Rehli, J. K. Baillie, M. J. L. De Hoon, V. Haberle, T. Lassmann, I. V. Kulakovskiy, M. Lizio, M. Itoh, R. Andersson, C. J. Mungall, T. F. Meehan, S. Schmeier, N. Bertin, M. Jørgensen, E. Dimont, E. Arner, C. Schmidl, U. Schaefer, Y. A. Medvedeva, C. Plessy, M. Vitezic, J. Severin, C. A. Semple, Y. Ishizu, R. S. Young, M. Francescato, I. A. Altschuler, D. Albanese, G. M. Altschule, T. Arakawa, J. A. C. Archer, P. Arner, M. Babina, S. Rennie, P. J. Balwierz, A. G. Beckhouse, S. Pradhan-Bhatt, J. A. Blake, A. Blumenthal, B. Bodega, A. Bonetti, J. Briggs, F. Brombacher, A. M. Burroughs, A. Califano, C. V. Cannistraci, D. Carbajo, Y. Chen, M. Chierici, Y. Ciani, H. C. Clevers, E. Dalla, C. A. Davis, M. Detmar, A. D. Diehl, T. Dohi, F. Drabløs, A. S. B. Edge, M. Edinger, K. Ekwall, M. Endoh, H. Enomoto, M. Fagiolini, L. Fairbairn, H. Fang, M. C. Farach-Carson, G. J. Faulkner, A. V. Favorov, M. E. Fisher, M. C. Frith, R. Fujita, S. Fukuda, C. Furlanello, M. Furuno, J. I. Furusawa, T. B. Geijtenbeek, A. P. Gibson, T. Gingeras, D. Goldowitz, J. Gough, S. Guhl, R. Guler, S. Gustincich, T. J. Ha, M. Hamaguchi, M. Hara, M. Harbers, J. Harshbarger, A. Hasegawa, Y. Hasegawa, T. Hashimoto, M. Herlyn, K. J. Hitchens, S. J. H. Sui, O. M. Hofmann, I. Hoof, F. Hori, L. Huminiecki, K. Iida, T. Ikawa, B. R. Jankovic, H. Jia, A. Joshi, G. Jurman, B. Kaczkowski, C. Kai, K. Kaida, A. Kaiho, K. Kajiyama, M. Kanamori-Katayama, A. S. Kasianov, T. Kasukawa, S. Katayama, S. Kato, S. Kawaguchi, H. Kawamoto, Y. I. Kawamura, T. Kawashima, J. S. Kempfle, T. J. Kenna, J. Kere, L. M. Khachigian, T. Kitamura, S. P. Klinken, A. J. Knox, M. Kojima, S. Kojima, N. Kondo, H. Koseki, S. Koyasu, S. Krampitz, A. Kubosaki, A. T. Kwon, J. F. J. Laros, W. Lee, A. Lennartsson, K. Li, B. Lilje, L. Lipovich, A. Mackay-sim, R. I. Manabe, J. C. Mar, B. Marchand, A. Mathelier, N. Mejhert, A. Meynert, Y. Mizuno, D. A. L. De Moraes, H. Morikawa, M. Morimoto, K. Moro, E. Motakis, H. Motohashi, C. L. Mummery, M. Murata, S. Nagao-Sato, Y. Nakachi, F. Nakahara, T. Nakamura, Y. Nakamura, K. Nakazato, E. Van Nimwegen, N. Ninomiya, H. Nishiyori, S. Noma, T. Nozaki, S. Ogishima, N. Ohkura, H. Ohmiya, H. Ohno, M. Ohshima, M. Okada-Hatakeyama, Y. Okazaki, V. Orlando, D. A. Ovchinnikov, A. Pain, R. Passier, M. Patrikakis, H. Persson, S. Piazza, J. G. D. Prendergast, O. J. L. Rackham, J. A. Ramilowski, M. Rashid, T. Ravasi, P. Rizzu, M. Roncador, S. Roy, M. B. Rye, E. Saijyo, A. Sajantila, A. Saka, S. Sakaguchi, M. Sakai, H. Sato, H. Satoh, S. Savvi, A. Saxena, C. Schneider, E. A. Schultes, G. G. Schulze-Tanzil, A. Schwegmann, T. Sengstag, G. Sheng, H. Shimoji, Y. Shimoni, J. W. Shin, C. Simon, D. Sugiyama, T. Sugiyama, M. Suzuki, N. Suzuki, R. K. Swoboda, P. A. C. 'T Hoen, M. Tagami, N. T. Tagami, J. Takai, H. Tanaka, H. Tatsukawa, Z. Tatum, M. Thompson, H. Toyoda, T. Toyoda, E. Valen, M. Van De Wetering, L. M. Van Den Berg, R. Verardo, D. Vijayan, I. E. Vorontsov, W. W. Wasserman, S. Watanabe, C. A. Wells, L. N. Winteringham, E. Wolvetang, E. J. Wood, Y. Yamaguchi, M. Yamamoto, M. Yoneda, Y. Yonekura, S. Yoshida, S. E. Zabierowski, P. G. Zhang, X. Zhao, S. Zucchelli, K. M. Summers, H. Suzuki, C. O. Daub, J. Kawai, P. Heutink, W. Hide, T. C. Freeman, B. Lenhard, L. V. B. Bajic, M. S. Taylor, V. J. Makeev, A. Sandelin, D. A. Hume, P. Carninci, Y. Hayashizaki, A promoter-level mammalian expression atlas. *Nature*. **507**, 462–470 (2014).
54. P. Carninci, A. Sandelin, B. Lenhard, S. Katayama, K. Shimokawa, J. Ponjavic, C. A. M. Semple, M. S. Taylor, P. G. Engström, M. C. Frith, A. R. R. Forrest, W. B. Alkema, S. L. Tan, C. Plessy, R. Kodzius, T. Ravasi, T. Kasukawa, S. Fukuda, M. Kanamori-Katayama, Y. Kitazume, H. Kawaji, C. Kai, M. Nakamura, H. Konno, K. Nakano, S. Mottagui-Tabar, P. Arner, A. Chesi, S. Gustincich, F. Persichetti, H. Suzuki, S. M. Grimmond, C. A.

Wells, V. Orlando, C. Wahlestedt, E. T. Liu, M. Harbers, J. Kawai, V. B. Bajic, D. A. Hume, Y. Hayashizaki, Genome-wide analysis of mammalian promoter architecture and evolution. *Nat. Genet.* **38**, 626–635 (2006).

55. M. C. Frith, E. Valen, A. Krogh, Y. Hayashizaki, P. Carninci, A. Sandelin, A code for transcription initiation in mammalian genomes. *Genome Res.* **18**, 1–12 (2008).
56. J. S. Cumbie, M. G. Ivanchenko, M. Megraw, NanoCAGE-XL and CapFilter: An approach to genome wide identification of high confidence transcription start sites. *BMC Genomics.* **16**, 1–13 (2015).
57. V. Haberle, A. R. R. Forrest, Y. Hayashizaki, P. Carninci, B. Lenhard, CAGER: Precise TSS data retrieval and high-resolution promoterome mining for integrative analyses. *Nucleic Acids Res.* **43** (2015), doi:10.1093/nar/gkv054.
58. M. Ota, Y. Nagafuchi, H. Hatano, K. Ishigaki, C. Terao, Y. Takeshima, H. Yanaoka, S. Kobayashi, M. Okubo, H. Shirai, Y. Sugimori, J. Maeda, M. Nakano, S. Yamada, R. Yoshida, H. Tsuchiya, Y. Tsuchida, S. Akizuki, H. Yoshifuji, K. Ohmura, T. Mimori, K. Yoshida, D. Kurosaka, M. Okada, K. Setoguchi, H. Kaneko, N. Ban, N. Yabuki, K. Matsuki, H. Mutoh, S. Oyama, M. Okazaki, H. Tsunoda, Y. Iwasaki, S. Sumitomo, H. Shoda, Y. Kochi, Y. Okada, K. Yamamoto, T. Okamura, K. Fujio, Dynamic landscape of immune cell-specific gene regulation in immune-mediated diseases. *Cell.* **184**, 3006–3021.e17 (2021).
59. G. La Manno, R. Soldatov, A. Zeisel, E. Braun, H. Hochgerner, V. Petukhov, K. Lidschreiber, M. E. Kastrioti, P. Lönnerberg, A. Furlan, J. Fan, L. E. Borm, Z. Liu, D. van Bruggen, J. Guo, X. He, R. Barker, E. Sundström, G. Castelo-Branco, P. Cramer, I. Adameyko, S. Linnarsson, P. V. Kharchenko, RNA velocity of single cells. *Nature.* **560**, 494–498 (2018).
60. V. Bergen, M. Lange, S. Peidli, F. A. Wolf, F. J. Theis, Generalizing RNA velocity to transient cell states through dynamical modeling. *Nat. Biotechnol.* **38**, 1408–1414 (2020).
61. L. T. Nguyen, P. S. Ohashi, Clinical blockade of PD1 and LAG3-potential mechanisms of action. *Nat. Rev. Immunol.* **15**, 45–56 (2015).
62. A. Cortes, J. Hadler, J. P. Pointon, P. C. Robinson, T. Karaderi, P. Leo, K. Cremin, K. Pryce, J. Harris, S. Lee, K. Bin Joo, S. C. Shim, M. Weisman, M. Ward, X. Zhou, H. J. Garchon, G. Chiocchia, J. Nossent, B. A. Lie, Ø. Førre, J. Tuomilehto, K. Laiho, L. Jiang, Y. Liu, X. Wu, L. A. Bradbury, D. Elewaut, R. Burgos-Vargas, S. Stebbings, L. Appleton, C. Farrah, J. Lau, T. J. Kenna, N. Haroon, M. A. Ferreira, J. Yang, J. Mulero, J. L. Fernandez-Sueiro, M. A. Gonzalez-Gay, C. Lopez-Larrea, P. Deloukas, P. Donnelly, P. Bowness, K. Gafney, H. Gaston, D. D. Gladman, P. Rahman, W. P. Maksymowych, H. Xu, J. B. A. Crusius, I. E. Van Der Horst-Bruinsma, C. T. Chou, R. Valle-Oñate, C. Romero-Sánchez, I. M. Hansen, F. M. Pimentel-Santos, R. D. Inman, V. Videm, J. Martin, M. Breban, J. D. Reveille, D. M. Evans, T. H. Kim, B. P. Wordsworth, M. A. Brown, Identification of multiple risk variants for ankylosing spondylitis through high-density genotyping of immune-related loci. *Nat. Genet.* **45**, 730–738 (2013).
63. K. Hashimoto, T. Kouno, T. Ikawa, N. Hayatsu, Y. Miyajima, H. Yabukami, T. Terooatea, T. Sasaki, T. Suzuki, M. Valentine, G. Pascarella, Y. Okazaki, H. Suzuki, J. W. Shin, A. Minoda, I. Taniuchi, H. Okano, Y. Arai, N. Hirose, P. Carninci, Single-cell

transcriptomics reveals expansion of cytotoxic CD4 T cells in supercentenarians. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 24242–24251 (2019).

64. A. Cachot, M. Bilous, Y. C. Liu, X. Li, M. Saillard, M. Cenerenti, G. A. Rockinger, T. Wyss, P. Guillaume, J. Schmidt, R. Genolet, G. Ercolano, M. P. Protti, W. Reith, K. Ioannidou, L. de Leval, J. A. Trapani, G. Coukos, A. Harari, D. E. Speiser, A. Mathis, D. Gfeller, H. Altug, P. Romero, C. Jandus, Tumor-specific cytolytic CD4 T cells mediate immunity against human cancer. *Sci. Adv.* **7** (2021), doi:10.1126/sciadv.abe3348.
65. B. J. Meckiff, C. Ramírez-Suástegui, V. Fajardo, S. J. Chee, A. Kusnadi, H. Simon, S. Eschweiler, A. Grifoni, E. Pelosi, D. Weiskopf, A. Sette, F. Ay, G. Seumois, C. H. Ottensmeier, P. Vijayanand, Imbalance of Regulatory and Cytotoxic SARS-CoV-2-Reactive CD4+ T Cells in COVID-19. *Cell.* **183**, 1340-1353.e16 (2020).
66. K. Mouri, M. H. Guo, C. G. de Boer, M. M. Lissner, I. A. Harten, G. A. Newby, H. A. DeBerg, W. F. Platt, M. Gentili, D. R. Liu, D. J. Campbell, N. Hacohen, R. Tewhey, J. P. Ray, Prioritization of autoimmune disease-associated genetic variants that perturb regulatory element activity in T cells. *Nat. Genet.* **54**, 603–612 (2022).
67. M. Koido, C. C. Hon, S. Koyama, H. Kawaji, Y. Murakawa, K. Ishigaki, K. Ito, J. Sese, N. F. Parrish, Y. Kamatani, P. Carninci, C. Terao, Prediction of the cell-type-specific transcription of non-coding RNAs from genome sequences via machine learning. *Nat. Biomed. Eng.* **7**, 830–844 (2023).
68. T. H. S. Hsieh, A. Weiner, B. Lajoie, J. Dekker, N. Friedman, O. J. Rando, Mapping Nucleosome Resolution Chromosome Folding in Yeast by Micro-C. *Cell.* **162**, 108–119 (2015).
69. N. Krietenstein, S. Abraham, S. V. Venev, N. Abdennur, J. Gibcus, T. H. S. Hsieh, K. M. Parsi, L. Yang, R. Maehr, L. A. Mirny, J. Dekker, O. J. Rando, Ultrastructural Details of Mammalian Chromosome Architecture. *Mol. Cell.* **78**, 554-565.e7 (2020).
70. T. H. S. Hsieh, C. Cattoglio, E. Slobodyanyuk, A. S. Hansen, O. J. Rando, R. Tjian, X. Darzacq, Resolving the 3D Landscape of Transcription-Linked Mammalian Chromatin Folding. *Mol. Cell.* **78**, 539-553.e8 (2020).
71. G. Barshad, J. J. Lewis, A. G. Chivu, A. Abuhashem, N. Krietenstein, E. J. Rice, Y. Ma, Z. Wang, O. J. Rando, A. K. Hadjantonakis, C. G. Danko, RNA polymerase II dynamics shape enhancer–promoter interactions. *Nat. Genet.* **55**, 1370–1380 (2023).
72. P. Hua, M. Badat, L. L. P. Hanssen, L. D. Hentges, N. Crump, D. J. Downes, D. M. Jeziorska, A. M. Oudelaar, R. Schwessinger, S. Taylor, T. A. Milne, J. R. Hughes, D. R. Higgs, J. O. J. Davies, Defining genome architecture at base-pair resolution. *Nature.* **595**, 125–129 (2021).
73. V. Y. Goel, M. K. Huseyin, A. S. Hansen, Region Capture Micro-C reveals coalescence of enhancers and promoters into nested microcompartments. *Nat. Genet.* **55**, 1048–1056 (2023).
74. B. Bulik-Sullivan, P. R. Loh, H. K. Finucane, S. Ripke, J. Yang, N. Patterson, M. J. Daly, A. L. Price, B. M. Neale, A. Corvin, J. T. R. Walters, K. H. Farh, P. A. Holmans, P. Lee, D. A. Collier, H. Huang, T. H. Pers, I. Agartz, E. Agerbo, M. Albus, M. Alexander, F. Amin, S. A. Bacanu, M. Begemann, R. A. Belliveau, J. Bene, S. E. Bergen, E. Bevilacqua, T. B. Bigdeli, D. W. Black, R. Bruggeman, N. G. Buccola, R. L. Buckner, W. Byerley, W.

Cahn, G. Cai, M. J. Cairns, D. Champion, R. M. Cantor, V. J. Carr, N. Carrera, S. V. Catts, K. D. Chambert, R. C. K. Chan, R. Y. L. Chen, E. Y. H. Chen, W. Cheng, E. F. C. Cheung, S. A. Chong, C. R. Cloninger, D. Cohen, N. Cohen, P. Cormican, N. Craddock, B. Crespo-Facorro, J. J. Crowley, D. Curtis, M. Davidson, K. L. Davis, F. Degenhardt, J. Del Favero, L. E. DeLisi, D. Demontis, D. Dikeos, T. Dinan, S. Djurovic, G. Donohoe, E. Drapeau, J. Duan, F. Dudbridge, N. Durmishi, P. Eichhammer, J. Eriksson, V. Escott-Price, L. Essioux, A. H. Fanous, M. S. Farrell, J. Frank, L. Franke, R. Freedman, N. B. Freimer, M. Friedl, J. I. Friedman, M. Fromer, G. Genovese, L. Georgieva, E. S. Gershon, I. Giegling, P. Giusti-Rodríguez, S. Godard, J. I. Goldstein, V. Golimbet, S. Gopal, J. Gratten, L. De Haan, C. Hammer, M. L. Hamshire, M. Hansen, T. Hansen, V. Haroutunian, A. M. Hartmann, F. A. Henskens, S. Herms, J. N. Hirschhorn, P. Hoffmann, A. Hofman, M. V. Hollegaard, D. M. Hougaard, M. Ikeda, I. Joa, A. Juliá, R. S. Kahn, L. Kalaydjieva, S. Karachanak-Yankova, J. Karjalainen, D. Kavanagh, M. C. Keller, B. J. Kelly, J. L. Kennedy, A. Khrunin, Y. Kim, J. Klovins, J. A. Knowles, B. Konte, V. Kucinskas, Z. A. Kucinskiene, H. Kuzelova-Ptackova, A. K. Kähler, C. Laurent, J. L. C. Keong, S. H. Lee, S. E. Legge, B. Lerer, M. Li, T. Li, K. Y. Liang, J. Lieberman, S. Limborska, C. M. Loughland, J. Lubinski, J. Lönnqvist, M. Macek, P. K. E. Magnusson, B. S. Maher, W. Maier, J. Mallet, S. Marsal, M. Mattheisen, M. Mattingdal, R. W. McCarley, C. McDonald, A. M. McIntosh, S. Meier, C. J. Meijer, B. Melegh, I. Melle, R. I. Meshulam-Gately, A. Metspalu, P. T. Michie, L. Milani, V. Milanova, Y. Mokrab, D. W. Morris, O. Mors, K. C. Murphy, R. M. Murray, I. Myin-Germeys, B. Müller-Myhsok, M. Nelis, I. Nenadic, D. A. Nertney, G. Nestadt, K. K. Nicodemus, L. Nikitina-Zake, L. Nisenbaum, A. Nordin, E. O'Callaghan, C. O'Dushlaine, F. A. O'Neill, S. Y. Oh, A. Olincy, L. Olsen, J. Van Os, C. Pantelis, G. N. Papadimitriou, S. Papiol, E. Parkhomenko, M. T. Pato, T. Paunio, M. Pejovic-Milovancevic, D. O. Perkins, O. Pietiläinen, J. Pimm, A. J. Pocklington, J. Powell, A. E. Pulver, S. M. Purcell, D. Quested, H. B. Rasmussen, A. Reichenberg, M. A. Reimers, A. L. Richards, J. L. Roffman, P. Roussos, D. M. Ruderfer, V. Salomaa, A. R. Sanders, U. Schall, C. R. Schubert, T. G. Schulze, S. G. Schwab, E. M. Scolnick, R. J. Scott, L. J. Seidman, J. Shi, E. Sigurdsson, T. Silagadze, J. M. Silverman, K. Sim, P. Slominsky, J. W. Smoller, H. C. So, C. C. A. Spencer, E. A. Stahl, H. Stefansson, S. Steinberg, E. Stogmann, R. E. Straub, E. Strengman, J. Strohmaier, T. S. Stroup, M. Subramaniam, J. Suvisaari, D. M. Svrakic, J. P. Szatkiewicz, E. Söderman, S. Thirumalai, D. Toncheva, P. A. Tooney, S. Tosato, J. Veijola, J. Waddington, D. Walsh, D. Wang, Q. Wang, B. T. Webb, M. Weiser, D. B. Wildenauer, N. M. Williams, S. Williams, S. H. Witt, A. R. Wolen, E. H. M. Wong, B. K. Wormley, J. Q. Wu, H. S. Xi, C. C. Zai, X. Zheng, F. Zimprich, N. R. Wray, K. Stefansson, P. M. Visscher, R. Adolfsson, O. A. Andreassen, D. H. R. Blackwood, E. Bramon, J. D. Buxbaum, A. D. Børglum, S. Cichon, A. Darvasi, E. Domenici, H. Ehrenreich, T. Esko, P. V. Gejman, M. Gill, H. Gurling, C. M. Hultman, N. Iwata, A. V. Jablensky, E. G. Jönsson, K. S. Kendler, G. Kirov, J. Knight, T. Lencz, D. F. Levinson, Q. S. Li, J. Liu, A. K. Malhotra, S. A. McCarroll, A. McQuillin, J. L. Moran, P. B. Mortensen, B. J. Mowry, M. M. Nöthen, R. A. Ophoff, M. J. Owen, A. Palotie, C. N. Pato, T. L. Petryshen, D. Posthuma, M. Rietschel, B. P. Riley, D. Rujescu, P. C. Sham, P. Sklar, D. St Clair, D. R. Weinberger, J. R. Wendland, T. Werge, P. F. Sullivan, M. C. O'Donovan, LD score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).

75. H. K. Finucane, B. Bulik-Sullivan, A. Gusev, G. Trynka, Y. Reshef, P. R. Loh, V. Anttila, H. Xu, C. Zang, K. Farh, S. Ripke, F. R. Day, S. Purcell, E. Stahl, S. Lindstrom, J. R. B. Perry, Y. Okada, S. Raychaudhuri, M. J. Daly, N. Patterson, B. M. Neale, A. L. Price, Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).
76. M. E. K. Niemi, J. Karjalainen, R. G. Liao, B. M. Neale, M. Daly, A. Ganna, G. A. Pathak, S. J. Andrews, M. Kanai, K. Veerapen, I. Fernandez-Cadenas, E. C. Schulte, P. Striano, M. Marttila, C. Minica, E. Marouli, M. A. Karim, F. R. Wendt, J. Savage, L. Sloofman, G. Butler-Laporte, H. N. Kim, S. Kanoni, Y. Okada, J. Byun, Y. Han, M. J. Uddin, G. D. Smith, C. J. Willer, J. D. Buxbaum, J. Mehtonen, H. Finucane, M. Cordioli, A. R. Martin, W. Zhou, B. Pasaniuc, H. Julienne, H. Aschard, H. Shi, L. Yengo, R. Polimanti, M. Ghoussaini, J. Schwartzentruber, I. Dunham, K. Chwialkowska, M. Francescato, A. Trankiem, M. K. Balaconis, L. Davis, S. Lee, J. Priest, A. Renieri, V. G. Sankaran, D. van Heel, P. Deelen, J. Brent Richards, T. Nakanishi, L. Biesecker, V. Eric Kerchberger, J. Kenneth Baillie, F. Mari, A. Bernasconi, S. Ceri, A. Canakoglu, B. Wolford, A. Faucon, A. K. Dutta, C. Schurmann, E. Harry, E. Birney, H. Nguyen, J. Nasir, M. Kaunisto, M. Solomonson, N. Dueker, N. Vadgama, S. Limou, S. Rahmouni, H. Mbarek, D. Darwish, M. M. Uddin, R. Albertos, J. Pérez-Tur, R. Li, L. Folkersen, I. Moltke, N. Koelling, A. Teumer, A. Kousathanas, A. Utrilla, R. A. Verdugo, R. Zárata, C. Medina-Gómez, D. Gómez-Cabrero, E. Carnero-Montoro, C. L. Cadilla, A. Moreno-Estrada, A. Garmendia, L. Moya, B. Sedaghati-Khayat, P. R. Boua, M. J. Favé, L. Francioli, A. Lemaçon, I. Migeotte, S. Patel, R. Varnai, J. L. Szentpeteri, C. Sipeky, F. Colombo, K. von Hohenstaufen, P. Lio, C. Vallerger, Q. Wang, Y. Tanigawa, H. Im, C. Han, H. Song, J. Lim, Y. Lee, S. Kim, S. Im, B. Atanasovska, H. F. Ahmad, C. Boer, P. Jansen, E. Kaja, D. Pasko, I. Kennis-Szilagyi, S. A. Kornilov, V. Prijatelj, I. Prokic, I. Sivanadhan, S. Perumal, S. Esmaeeli, N. M. Pearson, A. Auton, J. F. Shelton, A. J. Shastri, T. Filshtein-Sonmez, D. Coker, A. Symons, J. Esparza-Gordillo, S. Aslibekyan, J. O'Connell, C. Ye, C. H. Weldon, M. Perera, K. O'Leary, M. Tuck, T. O'Brien, D. Meltzer, P. O'Donnell, E. Nutescu, G. Yang, C. Alarcon, S. Herrmann, S. Mazurek, J. Banagan, Z. Hamidi, A. Barbour, N. Raffat, D. Moreno, P. Friedman, B. Ferwerda, D. van de Beek, M. C. Brouwer, A. P. J. Vlaar, W. Joost Wiersinga, D. Posthuma, E. Tissink, A. H. Koos Zwinderman, E. Uffelmann, M. van Agtmael, A. G. Algera, F. van Baarle, D. Bax, M. Beudel, H. J. Bogaard, M. Bomers, P. I. Bonta, L. Bos, M. Botta, J. de Brabander, G. de Bree, S. de Bruin, M. Bugiani, E. Bulle, O. Chouchane, A. Cloherty, D. Dongelmans, P. Elbers, L. Fleuren, S. Geerlings, B. Geerts, T. Geijtenbeek, A. Girbes, B. Goorhuis, M. P. Grobusch, F. Hafkamp, L. Hagens, J. Hamann, V. Harris, R. Hemke, S. M. Hermans, L. Heunks, M. Hollmann, J. Horn, J. W. Hovius, M. D. de Jong, R. Koning, N. van Mourik, J. Nellen, E. J. Nossent, F. Paulus, E. Peters, T. van der Poll, B. Preckel, J. M. Prins, J. Raasveld, T. Reijnders, M. Schinkel, M. J. Schultz, A. Schuurman, K. Sigaloff, M. Smit, C. S. Stijnis, W. Stilma, C. Teunissen, P. Thorat, A. Tsonas, M. van der Valk, D. Veelo, H. de Vries, M. van Vugt, D. Wouters, R. P. Minnaar, A. Kromhout, K. W. J. van Uffelen, R. A. Wolterman, G. Roberts, D. Park, C. A. Ball, M. Coignet, S. McCurdy, S. Knight, R. Partha, B. Rhead, M. Zhang, N. Berkowitz, M. Gaddis, K. Noto, L. Ruiz, M. Pavlovic, E. L. Hong, K. Rand, A. Girshick, H. Guturu, A. H. Baltzell, J. Guntz, Y. Beguin, S. Pigazzini, L. Nkambule, Y. Bouysran, A. Busson, X. Peyrassol, F. Wilkin, B. Pichon, G. Smits, I. Vandernoot, J. C. Goffard, M. Georges, M. Moutschen, B. Misset, G. Darcis, J. Guiot, L. Jadot, S. Azarzar, P. Dellot, S. Gofflot, S. Claassen, A.

Bertrand, G. Parzibut, M. Clarinval, C. Moermans, O. Malaise, K. El Kandoussi, R. Thonon, P. Huynen, A. Mesdagh, S. Melo, N. Jacques, E. Di Valentin, F. Giroule, A. Collignon, C. Radermecker, M. Lebrun, H. Perée, S. Latour, O. Barada, J. Sanchez, C. Josse, B. Boujemla, M. Meunier, E. Mariavelle, S. Anania, H. Gazon, D. Juszczak, M. Fadeur, S. Camby, C. Meuris, M. Thys, J. Jacques, M. Henket, P. Léonard, F. Fripiat, J. B. Giot, A. S. Sauvage, C. Von Frenckell, M. Mni, M. Wéry, A. Staderoli, Y. Belhaj, B. Lambermont, D. R. Morrison, V. Mooser, J. Brent Richards, V. Forgetta, R. Li, B. Ghosh, L. Laurent, A. Belisle, D. Henry, T. Abdullah, O. Adeleye, N. Mamlouk, N. Kimchi, Z. Afrasiabi, N. Rezk, B. Vulesevic, M. Bouab, C. Guzman, L. Petitjean, C. Tselios, X. Xue, J. Afilalo, M. Afilalo, M. Oliveira, B. Brenner, N. Brassard, M. Durand, E. Schurr, P. Lepage, J. Ragoussis, D. Auld, M. Chassé, D. E. Kaufmann, G. Mark Lathrop, D. Adra, L. K. Davis, N. J. Cox, J. E. Below, J. M. Sealock, A. B. Faucon, M. M. Shuey, H. G. Polikowsky, L. E. Petty, D. M. Shaw, H. H. Chen, W. Zhu, K. U. Ludwig, J. Schröder, C. Maj, S. Rolker, M. M. Nöthen, J. Fazaal, V. Keitel, B. E. O. Jensen, T. Feldt, I. Kurth, N. Marx, M. Dreher, I. Pink, M. Cornberg, T. Illig, C. Lehmann, P. Schommers, M. Augustin, L. Knopp, T. Eggermann, S. Volland, J. Altmüller, M. M. Berger, T. Brenner, A. Hinney, O. Witzke, R. Bals, C. Herr, N. Ludwig, J. Walter, C. Fuchsberger, C. Pattaro, A. De Grandi, P. Pramstaller, D. Emmert, R. Melotti, L. Foco, D. Mascalzoni, M. Gögele, F. Domingues, A. Hicks, C. R. Gignoux, S. J. Wicks, K. Crooks, K. C. Barnes, M. Daya, J. Shortt, N. Rafaels, S. Chavan, D. B. Goldstein, K. Kiryluk, S. Sengupta, W. Chung, M. P. Reilly, A. Khan, C. Wang, G. Povysil, N. Bhardwaj, A. G. Gharavi, I. Ionita-Laza, N. Shang, S. M. O'Byrne, R. Nandakumar, A. Menon, Y. S. So, E. Hod, D. Pendrick, S. K. Park, H. L. Kim, C. K. Kang, H. J. Lee, K. H. Song, K. J. Yoon, N. J. Paik, W. Seok, H. Yoon, E. J. Joo, Y. Chang, S. Ryu, W. B. Park, J. S. Park, K. U. Park, S. Y. Ham, J. Jung, E. S. Kim, H. Bin Kim, D. Ellinghaus, F. Degenhardt, M. Cáceres, S. Juzenas, T. L. Lenz, A. Albillos, A. Julià, B. Heidecker, F. Garcia, F. Kurth, F. Tran, F. Hanses, H. Zoller, J. C. Holter, J. Fernández, L. E. Sander, P. Rosenstiel, P. Koehler, R. de Cid, R. Asselta, S. Schreiber, U. Hehr, D. Prati, G. Baselli, L. Valenti, L. Bujanda, J. M. Banales, M. D'Amato, M. Romero-Gómez, M. Buti, P. Invernizzi, A. Franke, J. R. Hov, T. H. Karlsen, T. Folseraas, D. Maya-Miles, A. Teles, C. Azuure, E. M. Wacker, F. Uellendahl-Werth, H. ElAbd, J. Arora, J. Lerga-Jaso, L. Wienbrandt, M. C. Rühlemann, M. Wendorff, M. S. Vadla, O. B. Lenning, O. Özer, R. Myhre, S. Raychaudhuri, A. Tanck, C. Gassner, G. Hemmrich-Stanisak, J. Kässens, M. E. Figuera Basso, M. Schulzky, M. Wittig, N. Braun, T. Wesse, W. Albrecht, X. Yi, A. B. Ortiz, A. G. Chercoles, A. Ruiz, A. Mantovani, A. R. Holten, A. Mayer, A. Cherubini, A. Protti, A. Aghemo, A. Gerussi, A. Ramirez, A. Braun, A. Barreira, A. Lleo, A. B. Kildal, A. Glück, A. C. Nolla, A. Latiano, A. M. Dyrhol-Riise, A. Muscatello, A. Voza, A. Rando-Segura, A. Solier, B. Karina, B. Cortes, B. Mateos, B. Nafria-Jimenez, B. Schaefer, C. Bellinghausen, C. Ferrando, C. Quereda, C. Skurk, C. Thibeault, C. D. Spinner, C. Lange, C. Hu, C. Cappadona, C. Bianco, C. Sancho, D. A. L. Hoff, D. Galimberti, D. Jiménez, D. Pestaña, D. Toapanta, E. Azzolini, E. Scarpini, E. T. Helbig, E. Urrechaga, E. M. Paraboschi, E. Pontali, E. Reverter, E. Navas, E. Arana, F. G. Sánchez, F. Ceriotti, F. Malvestiti, F. Mesonero, G. Pezzoli, G. Lamorte, H. Neb, I. My, I. Hernández, I. de Rojas, I. Galván-Femenia, J. Heyckendorf, J. Rybniker, J. R. Badia, J. Schneider, J. Goikoetxea, J. Kraft, K. E. Müller, K. I. Gaede, K. Garcia-Etxebarria, K. Tonby, L. Heggelund, L. Izquierdo-Sanchez, L. Sumoy, L. J. Lippert, L. Terranova, L. Garbarino, L. Téllez, L. Roade, M. Ostadreza, M. Intxausti, M. Kogevinas, M. A. Gutiérrez-Stampa, M. J. G. T. Vehreschild, M. Marquié, M. Castoldi, M. Cecconi, M. Boada, M. J. Seilmaier, M. Mazzocco, M. Rodríguez-Gandía, N. I. Ayo, N. Blay, N.

Martínez, O. A. Cornely, O. Palmieri, P. Tentorio, P. M. Rodrigues, P. P. España, P. Hoffmann, P. Bacher, P. Suwalski, R. de Pablo, R. Nieto, S. Badalamenti, S. Ciesek, S. Bombace, S. Wilfling, S. Brunak, S. Heilmann-Heimbach, S. Ripke, T. Bahmer, U. Landmesser, U. Protzer, V. Rimoldi, V. Skogen, V. Andrade, V. Moreno, W. Poller, X. Farre, X. Wang, Y. Khodamoradi, Z. Karadeniz, A. de Salazar, A. Palom, A. E. Garcia-Fernandez, A. Blanco-Grau, A. Zanella, A. Bandera, A. Nebel, A. Biondi, A. Caballero-Garralda, A. Gori, A. Lind, A. L. Fracanzani, A. Peschuck, A. Pesenti, C. de la Horra, C. Milani, C. Paccapelo, C. Angelini, C. Cea, E. Muñoz-Díaz, E. Sandoval, E. J. Calderón, E. Solligård, F. Aziz, F. Martinelli-Boneschi, F. Peyvandi, F. Blasi, F. J. Medrano, F. Rodríguez-Frias, F. Müller, G. Grasselli, G. Costantino, G. Cardamone, G. Foti, G. Matullo, H. Kurihara, J. E. Afset, J. K. Damås, J. Ampuero, J. Martín, J. Erdmann, J. Bergan, S. Goerg, J. Ferrusquía-Acosta, J. H. Quero, J. Delgado, J. M. Guerrero, K. Risnes, L. R. Bettini, L. Moreira, L. T. Gustad, L. Santoro, L. Scudeller, M. Riveiro-Barciela, M. Schaefer, M. Carrabba, M. G. Valsecchi, M. Hernandez-Tejero, M. Acosta-Herrera, M. D'Angiò, M. Baldini, M. Cazzaniga, M. Ciccarelli, M. Bocciolone, M. Miozzo, N. Chueca, N. Montano, P. Faverio, P. Preatoni, P. Bonfanti, P. Omodei, P. Castro, R. Ferrer, R. Gualtierotti, R. Gallego-Durán, R. Morilla, S. Haider, S. Marsal, S. Aneli, S. Pelusi, S. Bosari, S. Aliberti, S. Dudman, T. Zheng, T. Pumarola, T. G. Cejudo, V. Monzani, V. Friaza, W. Peter, X. Dopazo, S. Duga, S. May, M. M. Grimsrud, D. F. Gudbjartsson, K. Stefansson, P. Sulem, G. Sveinbjornsson, P. Melsted, G. Norddahl, K. H. S. Moore, U. Thorsteinsdottir, H. Holm, M. E. Alarcón-Riquelme, D. Bernardo, M. Martínez-Bueno, S. R. Rello, R. Mägi, L. Milani, A. Metspalu, T. Laisk, K. Läll, M. Lepamets, T. Esko, E. Reimann, P. Naaber, E. Laane, J. Pesukova, P. Peterson, K. Kisand, J. Tabri, R. Allos, K. Hensen, J. Starkopf, I. Ringmets, A. Tamm, A. Kallaste, H. Alavere, K. Metsalu, M. Puusepp, K. Kristiansson, S. Koskelainen, M. Perola, K. Donner, K. Kivinen, A. Palotie, C. Rivolta, P. Y. Bochud, S. Bibert, N. Boillat, S. G. Nussle, W. Albrich, M. Quinodoz, D. Kamdar, N. Suh, D. Neofytos, V. Erard, C. Voide, P. Y. Bochud, C. Rivolta, S. Bibert, M. Quinodoz, D. Kamdar, D. Neofytos, V. Erard, C. Voide, R. Friolet, P. Vollenweider, J. L. Pagani, M. Oddo, F. Meyer zu Bentrup, A. Conen, O. Clerc, O. Marchetti, A. Guillet, C. Guyat-Jacques, S. Foucras, M. Rime, J. Chassot, M. Jaquet, R. Merlet Viollet, Y. Lannepoudenx, L. Portopena, P. Y. Bochud, P. Vollenweider, J. L. Pagani, F. Desgranges, P. Filippidis, B. Guéry, D. Haefliger, E. E. Kampouri, O. Manuel, A. Munting, M. Papadimitriou-Olivgeris, J. Regina, L. Rochat-Stettler, V. Suttels, E. Tadini, J. Tschopp, M. Van Singer, B. Viala, N. Boillat-Blanco, T. Brahier, O. Hügli, J. Y. Meuwly, O. Pantet, S. Gonseth Nussle, M. Bochud, V. D'Acremont, S. Estoppey Younes, W. C. Albrich, N. Suh, A. Cerny, L. O'Mahony, C. von Mering, P. Y. Bochud, M. Frischknecht, G. R. Kleger, M. Filipovic, C. R. Kahlert, H. Wozniak, T. Rochat Negro, J. Pugin, K. Bouras, C. Knapp, T. Egger, A. Perret, P. Montillier, C. di Bartolomeo, B. Barda, A. Carreras, I. Galván-Femenía, X. Farré, B. Cortés, J. M. Mercader, M. Guindo-Martinez, D. Torrents, J. Garcia-Aymerich, G. Castaño-Vinyals, C. Dobaño, M. Gori, M. U. Mondelli, F. Castelli, M. Vaghi, S. Rusconi, F. Montagnani, E. Bargagli, F. Franchi, M. A. Mazzei, L. Cantarini, D. Tacconi, M. Feri, R. Scala, G. Spargi, C. Nencioni, M. Bandini, G. P. Caldarelli, M. Spagnesi, A. Canaccini, A. Ognibene, A. D. Monforte, M. Girardis, A. Antinori, D. Francisci, E. Schiaroli, P. G. Scotton, S. Panese, R. Scaggiante, M. Della Monica, M. Capasso, G. Fiorentino, M. Castori, F. Aucella, A. Di Biagio, L. Masucci, S. Valente, M. Mandalà, P. Zucchi, F. Giannattasio, D. A. Coviello, C. Mussini, G. Bosio, L. Tavecchia, L. Crotti, M. Rizzi, M. T. La Rovere, S. Sarzi-Braga, M. Bussotti, S. Ravaglia, R. Artuso, A. Perrella, D. Romani,

P. Bergomi, E. Catena, A. Vincenti, C. Ferri, D. Grassi, G. Pessina, M. Tumbarello, M. Di Pietro, R. Sabrina, S. Luchi, C. Barbieri, D. Acquilini, E. Andreucci, F. Paciosi, F. V. Segala, G. Tiseo, M. Lista, M. Poscente, O. De Vivo, P. Petrocelli, A. Guarnaccia, S. Baroni, V. Perticaroli, S. Furini, S. Dei, E. Benetti, N. Picchiotti, M. Sanarico, P. Pinoli, F. Raimondi, F. Biscarini, A. Stella, M. Bergomi, K. Zguro, K. Capitani, M. Tanfoni, C. Fallerini, S. Daga, M. Baldassarri, F. Fava, E. Frullanti, F. Valentino, G. Doddato, A. Giliberti, R. Tita, S. Amitrano, M. Bruttini, S. Croci, I. Meloni, M. A. Mencarelli, C. Lo Rizzo, A. M. Pinto, G. Beligni, A. Tommasi, L. Di Sarno, M. Palmieri, M. L. Carriero, D. Alaverdian, N. Iuso, G. Inchingolo, S. Busani, R. Bruno, M. Vecchia, M. Ann Belli, S. Mantovani, S. Ludovisi, E. Quiros-Roldan, M. D. Antoni, I. Zanella, M. Siano, A. Emiliozzi, M. Fabbiani, B. Rossetti, G. Zanelli, L. Bergantini, M. D'Alessandro, P. Cameli, D. Bennet, F. Anedda, S. Marcantonio, S. Scolletta, S. Guerrini, E. Conticini, B. Frediani, C. Spertilli, A. Donati, L. Guidelli, M. Corridi, L. Croci, P. Piacentini, E. Desantis, S. Cappelli, A. Verzuri, V. Anemoli, A. Pancrazi, M. Lorubbio, E. Merlini, F. G. Miraglia, S. Venturelli, A. Cossarizza, A. Vergori, A. Gabrieli, A. Riva, F. Andretta, F. Gatti, S. G. Parisi, S. Baratti, C. Piscopo, R. Russo, I. Andolfo, A. Iolascon, M. Carella, G. Merla, G. M. Squeo, P. Raggi, C. Marciano, R. Perna, M. Bassetti, M. Sanguinetti, A. Giorli, L. Salerni, P. Parravicini, E. Menatti, T. Trotta, G. Coiro, F. Lena, E. Martinelli, S. Mancarella, C. Gabbi, F. Maggiolo, D. Ripamonti, T. Bachetti, C. Suardi, G. Parati, G. Bottà, P. Di Domenico, I. Rancan, F. Bianchi, R. Colombo, M. Falcone, K. A. Hunt, R. C. Trembath, H. C. Martin, D. Mason, B. Trivedi, J. Wright, S. Finer, C. J. Griffiths, S. Akhtar, M. Anwar, E. Arciero, S. Ashraf, G. Breen, R. Chung, C. J. Curtis, M. Chowdhury, G. Colligan, P. Deloukas, C. Durham, C. Griffiths, Q. Q. Huang, M. Hurles, S. Hussain, K. Islam, A. Khan, A. Khan, C. Lavery, S. H. Lee, R. Lerner, D. MacArthur, B. MacLaughlin, H. Martin, S. Miah, B. Newman, N. Safa, F. Tahmasebi, A. V. Smith, A. P. Boughton, K. W. Li, J. LeFaive, A. Annis, C. E. Jannes, J. E. Krieger, A. C. Pereira, M. Velho, E. Marques, I. R. Lima, M. T. Tada, K. Valino, M. McCarthy, C. Rosenberger, J. E. Lee, D. Chang, C. Hammer, J. Hunkapiller, A. Mahajan, S. Pendergrass, L. Sucheston-Campbell, B. Yaspan, H. S. Lee, E. Shin, H. Y. Jang, S. Kim, S. Kym, Y. S. Kim, H. Jeong, K. T. Kwon, S. W. Kim, J. Y. Kim, Y. R. Jang, H. Kim, J. Y. Lee, S. Lee, K. W. Choe, Y. M. Kang, S. H. Jee, K. J. Jung, V. Parikh, E. Ashley, M. Wheeler, M. Rivas, C. Bustamante, B. Pinksy, P. Febbo, K. Farh, G. P. Schroth, F. deSouza, K. Dalton, J. Christle, C. Deboever, S. Szalma, S. Rubinacci, O. Delaneau, J. Gorzynski, H. de Jong, S. Sutton, N. Youlton, R. Joshi, D. Jimenez-Morales, C. Hughes, D. Amar, A. Ioannidis, S. Hershman, A. Kirillova, K. Seo, Y. Huang, M. Shoura, N. Hammond, N. Watson, A. Raja, C. H. Huang, M. Sahoo, H. Wang, J. Zhen, A. Rakitko, V. Ilinsky, D. Yermakovich, I. Popov, A. Chernitsov, E. Kovalenko, A. Krasnenko, N. Plotnikov, I. Stetsenko, A. Kim, E. T. Cirulli, K. M. Schiabor Barrett, A. Bolze, S. White, N. L. Washington, J. T. Lu, S. Riffle, F. Tanudjaja, X. Wang, J. M. Ramirez, N. Leonetti, E. Sandoval, I. Neveux, S. Dabe, J. J. Grzymski, J. I. E. Miñano, L. A. Aguirre, E. López-Collazo, M. de la Mata Pazos, L. Cerrato, R. Lozano-Rodríguez, J. Avendaño-Ortiz, V. T. Arcos, K. M. Montalbán-Hernández, J. V. Quiroga, A. Pascual-Iglesias, C. Maroun-Eid, A. Martín-Quirós, H. Namkoong, S. Imoto, K. Katayama, K. Fukunaga, Y. Kitagawa, T. Sato, N. Hasegawa, A. Kumanogoh, A. Kimura, M. Ai, K. Tokunaga, T. Kanai, S. Miyano, S. Ogawa, R. Edahiro, K. Sonehara, Y. Shirai, M. Ishii, H. Kabata, K. Masaki, H. Kamata, S. Ikemura, S. Chubachi, S. Okamori, H. Terai, H. Tanaka, A. Morita, H. Lee, T. Asakura, J. Sasaki, H. Morisaki, Y. Uwamino, K. Nanki, Y. Mikami, K. Tomono, K. Kato, F. Matsuda, M. Takahashi, N. Hizawa, Y. Takeda, H. Hirata, T. Shiroyama, S. Miyawaki, K.

Suzuki, Y. Maeda, T. Nii, Y. Noda, T. Niitsu, Y. Adachi, T. Enomoto, S. Amiya, R. Hara, K. Takahashi, T. Anzai, T. Hasegawa, S. Ito, R. Koike, A. Endo, Y. Uchimura, Y. Miyazaki, T. Honda, T. Tateishi, S. Tohda, N. Ichimura, K. Sonobe, C. Sassa, J. Nakajima, Y. Nannya, Y. Omae, K. Takahashi, N. Harada, M. Hiki, H. Takagi, A. Nakamura, E. Tagaya, M. Kawana, K. Arimura, T. Ishiguro, N. Takayanagi, T. Isono, Y. Takaku, K. Takano, R. Anan, Y. Nakajima, Y. Nakano, K. Nishio, S. Ueda, R. Hayashi, H. Tateno, I. Hase, S. Yoshida, S. Suzuki, K. Mitamura, F. Saito, T. Ueda, M. Azuma, T. Nagasaki, Y. Yasui, Y. Hasegawa, Y. Mutoh, T. Yoshiyama, T. Shoko, M. Kojima, T. Adachi, M. Ishikawa, K. Takahashi, K. Watanabe, T. Manabe, F. Ito, T. Fukui, Y. Funatsu, H. Koh, Y. Hirai, H. Kawashima, A. Narita, K. Niwa, Y. Sekikawa, F. Saito, K. Yoshiya, T. Yoshihara, Y. Suzuki, S. Nakayama, K. Masuzawa, K. Nishi, M. Nishitsuji, M. Tani, T. Inoue, T. Hirano, K. Kobayashi, N. Miyazawa, Y. Kimura, R. Sado, T. Ogura, H. Kitamura, K. Murohashi, I. Nakachi, R. Baba, D. Arai, S. Fuke, H. Saito, N. Kuwahara, A. Fujiwara, T. Okada, T. Baba, J. Noda, S. Mashimo, K. Yagi, T. Shiomi, M. Hashiguchi, T. Odani, T. Mochimaru, Y. Oyamada, N. Mori, N. Izumi, K. Nagata, R. Taki, K. Murakami, M. Yamada, H. Sugiura, K. Hayashi, T. Shimizu, Y. Gon, S. Fujitani, T. Tsuchida, T. Yoshida, T. Kagaya, T. Kita, S. Sakagami, Y. Kimizuka, A. Kawana, Y. Nakamura, H. Ishikura, T. Takata, T. Kikuchi, D. Taniyama, M. Nakamura, N. Kodama, Y. Kaneyama, S. Maeda, Y. Nagasaki, M. Okamoto, S. Ishihara, A. Ito, Y. Chihara, M. Takeuchi, K. Onoi, N. Hashimoto, K. Wakahara, A. Ando, M. Masuda, A. Wakabayashi, H. Watanabe, H. Sageshima, T. A. Nakada, R. Abe, T. Shimada, K. Kawamura, K. Ichikado, K. Nishiyama, M. Yamasaki, S. Hashimoto, Y. Kusaka, T. Ohba, S. Isogai, M. Takada, H. Kanda, Y. Komase, F. Sano, K. Asano, T. Oguma, M. Harada, T. Takahashi, T. Shibusawa, S. Abe, Y. Kono, Y. Togashi, T. Izumo, M. Inomata, N. Awano, S. Ogawa, T. Ogata, S. Ishihara, A. Kanehiro, S. Ozaki, Y. Fuchimoto, Y. Kitagawa, S. Yoshida, S. Ogura, K. Nishiyama, K. Yoshida, S. Beppu, S. Fukuyama, Y. Eriguchi, A. Yonekawa, Y. Inoue, K. Yamagata, S. Chiba, O. Narumoto, H. Nagai, N. Ooshima, M. Motegi, H. Sagara, A. Tanaka, S. Ohta, Y. Shibata, Y. Tanino, Y. Sato, Y. Yamada, T. Hashino, M. Shinoki, H. Iwagoe, T. Imamura, A. Umeda, H. Shimada, M. Endo, S. Hayashi, M. Takahashi, S. Nakano, M. Yatomi, T. Maeno, T. Ishii, M. Utsugi, A. Ono, K. Kanaoka, S. Ihara, K. Komuta, L. Franke, M. Boezen, A. Claringbould, E. Lopera, R. Warmerdam, J. M. Vonk, I. van Blokland, P. Lanting, A. P. S. Ori, M. Obeidat, A. I. Hernández Cordero, D. D. Sin, Y. Bossé, P. Joubert, K. Hao, D. Nickle, W. Timens, M. van den Berge, Y. C. A. Feng, J. Mercader, S. T. Weiss, E. W. Karlson, J. W. Smoller, S. N. Murphy, J. B. Meigs, A. E. Woolley, R. C. Green, E. F. Perez, S. Zöllner, J. Wang, A. Beck, L. G. Sloofman, S. Ascolillo, R. P. Sebra, B. L. Collins, T. Levy, S. C. Sealfon, D. M. Jordan, R. C. Thompson, K. Gettler, K. Chaudhary, G. M. Belbin, M. Preuss, C. Hoggart, S. Choi, S. J. Underwood, I. Salib, B. Britvan, K. Keller, L. Tang, M. Peruggia, L. L. Hiester, K. Niblo, A. Aksentijevich, A. Labkowsky, A. Karp, M. Zlatopolsky, M. Zyndorf, A. W. Charney, N. D. Beckmann, E. E. Schadt, N. S. Abul-Husn, J. H. Cho, Y. Itan, E. E. Kenny, R. J. F. Loos, G. N. Nadkarni, R. Do, P. O'Reilly, L. M. Huckins, M. A. R. Ferreira, G. R. Abecasis, J. B. Leader, M. N. Cantor, A. E. Justice, D. J. Carey, G. Chittoor, N. S. Josyula, J. A. Kosmicki, J. E. Horowitz, A. Baras, M. C. Gass, A. Yadav, T. Mirshahi, J. J. Hottenga, M. Bartels, E. J. C. de Geus, M. G. Nivard, A. Verma, M. D. Ritchie, D. Rader, B. Li, S. S. Verma, A. Lucas, Y. Bradford, F. Zara, V. Salpietro, M. Scala, M. Iacomino, P. Scudieri, R. Bocciardi, C. Minetti, A. Riva, M. S. Vari, J. F. Rahier, E. Giorgio, D. Carli, E. Louis, C. M. Bulik, M. Landén, A. Brusco, G. B. Ferrero, F. Madia, B. Fundin, S. I. Ismail, C. Saad, Y. Al-Sarraj, R. M. Badji, W. Al-Muftah, A. Al

Thani, N. Afifi, J. Klovins, V. Rovite, R. Rescenko, R. Peculis, M. Ustinova, H. Zeberg, R. Frithiof, M. Hultström, M. Lipcsey, R. Johnson, D. H. Geschwind, N. Freimer, M. J. Butte, Y. Ding, A. Chiu, T. S. Chang, P. Boutros, L. Moutsianas, M. J. Caulfield, R. H. Scott, S. Walker, A. Stuckey, C. A. Odhams, D. Rhodes, T. Fowler, A. Rendon, G. Chan, P. Arumugam, K. J. Karczewski, D. J. Wilson, C. A. Spencer, D. W. Crook, D. H. Wyllie, A. M. O'Connell, J. Brent Richards, E. G. Atkinson, K. Tsuo, N. Baya, P. Turley, R. Gupta, R. K. Walters, D. S. Palmer, G. Sarma, N. Cheng, W. Lu, C. Churchhouse, J. I. Goldstein, D. King, C. Seed, S. Bryant, F. Kyle Satterstrom, G. Band, S. G. Earle, S. K. Lin, N. Arning, J. Armstrong, J. K. Rudkin, S. Callier, C. Cusick, N. Soranzo, J. H. Zhao, J. Danesh, E. Di Angelantonio, A. S. Butterworth, Y. V. Sun, J. E. Huffman, K. Cho, C. J. O'Donnell, P. Tsao, J. Michael Gaziano, G. Peloso, Y. L. Ho, M. Mian, F. Scaggiante, X. Chang, J. R. Glessner, H. Hakonarson, J. Kenneth Baillie, P. J. McGuigan, L. S. P. Moore, M. P. Vizcaychipi, K. Hall, A. Campbell, A. Nichol, G. Ward, V. J. Page, K. Adeniji, D. Agranoff, K. Agwuh, D. Ail, E. L. Aldera, A. Alegria, B. Angus, A. Ashish, D. Atkinson, S. Bari, G. Barlow, S. Barnass, N. Barrett, C. Bassford, S. Basude, D. Baxter, M. Beadsworth, J. Bernatoniene, J. Berridge, N. Best, P. Bothma, D. Chadwick, R. Brittain-Long, N. Bulteel, T. Burden, V. Caruth, D. Chambler, N. Chee, J. Child, T. Clark, P. Collini, C. Cosgrove, J. Cupitt, M. T. Cutino-Moguel, P. Dark, C. Dawson, S. Dervisevic, P. Donnison, S. Douthwaite, A. Drummond, I. DuRand, A. Dushianthan, T. Dyer, C. Evans, C. Eziefula, C. Fegan, A. Finn, D. Fullerton, S. Garg, A. Garg, E. Gkrania-Klotsas, J. Godden, A. Goldsmith, C. Graham, E. Hardy, S. Hartshorn, D. Harvey, P. Havalda, D. B. Hawcutt, M. Hobrok, M. Jacobs, S. Jain, P. Jennings, A. Kaliappan, V. Kasipandian, S. Kegg, M. Kelsey, J. Kendall, C. Kerrison, I. Kerslake, O. Koch, G. Koduri, G. Koshy, S. Laha, S. Laird, S. Larkin, T. Leiner, P. Lillie, J. Limb, V. Linnett, J. Little, M. Lyttle, M. MacMahon, E. MacNaughton, R. Mankregod, H. Masson, E. Matovu, K. McCullough, R. McEwen, M. Meda, J. Minton, K. Ward, M. Mirfenderesky, K. Mohandas, Q. Mok, J. Moon, E. Moore, P. Morgan, C. Morris, K. Mortimore, S. Moses, M. Mpenge, R. Mulla, M. Murphy, M. Nagel, T. Nagarajan, M. Nelson, M. K. O'Shea, I. Otahal, M. Ostermann, M. Pais, S. Panchatsharam, D. Papakonstantinou, H. Paraiso, B. Patel, N. Pattison, J. Pepperell, M. Peters, M. Phull, S. Pintus, J. S. Pooni, F. Post, D. Price, R. Prout, N. Rae, H. Reschreiter, T. Reynolds, N. Richardson, M. Roberts, D. Roberts, A. Rose, G. Rousseau, B. Ryan, T. Saluja, A. Shah, P. Shanmuga, A. Sharma, A. Shawcross, J. Sizer, M. Shankar-Hari, R. Smith, C. Snelson, N. Spittle, N. Staines, T. Stambach, R. Stewart, P. Subudhi, K. Tatham, J. Thomas, C. Thompson, R. Thompson, A. Tridente, D. Tupper-Carey, M. Twagira, A. Ustianowski, N. Vallotton, L. Vincent-Smith, S. Visuvanathan, A. Vuylsteke, S. Waddy, R. Wake, A. Walden, I. Welters, T. Whitehouse, P. Whittaker, A. Whittington, P. Papineni, M. Wijesinghe, M. Williams, L. Wilson, S. Cole, S. Winchester, M. Wiselka, A. Wolverson, D. G. Wooton, A. Workman, B. Yates, P. Young, J. Kenneth Baillie, R. Beale, A. D. Bretherick, M. H. Fourman, J. Furniss, E. Gountouna, G. Grimes, C. Haley, D. Harrison, C. Hayward, S. Keating, L. Klaric, P. Klenerman, A. M. Meynert, J. Millar, E. Pairo-Castineira, N. Parkinson, C. P. Ponting, D. J. Porteous, K. Rawlik, A. Richmond, C. D. Russell, X. Shen, B. Shih, A. Tenesa, V. Vitart, B. Wang, J. F. Wilson, Y. Wu, J. Yang, Z. Yang, M. Zechner, R. Zhai, C. Zheng, L. Norman, R. Pius, T. M. Drake, C. J. Fairfield, S. R. Knight, K. A. Mclean, D. Murphy, C. A. Shaw, J. Dalton, M. Girvan, E. Saviciute, S. Roberts, J. Harrison, L. Marsh, M. Connor, S. Halpin, C. Jackson, C. Gamble, G. Leeming, A. Law, M. Wham, S. Clohisey, R. Hendry, J. Scott-Brown, C. Begg, C. Hinds, A. Y. W. Ho, P. W. Horby, J. Knight, L. Ling, D. Maslove, D. McAuley, H. Montgomery, P. J. M. Openshaw, K. Rowan, M. G. Semple, C. Summers, T. Walsh, L.

Armstrong, H. Bates, E. Dooks, F. Farquhar, B. Hairsine, C. McParland, S. Packham, Z. Alldis, R. Astin-Chamberlain, F. Bibi, J. Biddle, S. Blow, M. Bolton, C. Borra, R. Bowles, M. Burton, Y. Choudhury, D. Collier, A. Cox, A. Easthope, P. Ebano, S. Fotiadis, J. Gurasashvili, R. Halls, P. Hartridge, D. Kallon, J. Kassam, I. Lancoma-Malcolm, M. Matharu, P. May, O. Mitchelmore, T. Newman, M. Patel, J. Pheby, I. Pinzuti, Z. Prime, O. Prysyazhna, J. Shiel, C. Tierney, S. Wood, A. Zak, O. Zongo, M. Forsey, A. Nicholson, J. Riches, M. Vertue, C. Wasson, S. Finn, J. Green, E. Collins, B. King, L. Grauslyte, M. Hussain, T. Pogreban, L. Rosaroso, E. Salciute, G. Franke, J. Wong, A. George, L. Akeroyd, S. Bano, M. Bromley, L. Gurr, T. Lawton, J. Morgan, K. Sellick, D. Warren, B. Wilkinson, J. McGowan, C. Ledgard, A. Stacey, K. Pye, R. Bellwood, M. Bentley, R. Loosley, H. McGuinness, H. Tench, R. Wolf-Roberts, S. Gibson, A. Lyle, F. McNeela, J. Radhakrishnan, A. Hughes, A. Ali, M. Brady, S. Dale, A. Dance, L. Gledhill, J. Greig, K. Hanson, K. Holdroyd, M. Home, D. Kelly, R. Kitson, L. Matapure, D. Melia, S. Mellor, T. Nortcliffe, J. Pinnell, M. Robinson, L. Shaw, R. Shaw, L. Thomis, A. Wilson, T. Wood, L. A. Bayo, E. Merwaha, T. Ishaq, S. Hanley, D. Antcliffe, D. Banach, S. Brett, P. Coghlan, Z. Fernandez, A. Gordon, R. Rojo, S. S. Arias, M. Templeton, R. Jha, L. Lim, R. Bi, B. Scholefield, L. Ashton, A. Williams, C. Cheyne, A. Saunderson, A. Allan, F. Anderson, C. Kaye, J. Liew, J. Medhora, T. Scott, E. Trumper, A. Botello, P. Polgarova, K. Stroud, E. Meaney, M. Jones, A. Ng, S. Agrawal, N. Pathan, D. White, E. Daubney, K. Elston, R. Parker, A. Reddy, I. Turner-Bone, L. Wilding, P. Harding, R. Jacob, C. Jones, C. Denmade, M. Croft, I. White, V. Krishnamurthy, L. Lim, D. Griffin, N. Muchenje, M. Mupudzi, R. Partridge, J. A. Conyngham, R. Thomas, M. Wright, M. A. Corral, V. Bastion, D. Clarke, B. David, H. Kent, R. Lorusso, G. Lubimbi, S. Murdoch, M. Penacerrada, A. Thomas, J. Valentine, A. Vochin, R. Wulandari, B. Djeugam, J. Dawson, S. Garrioch, M. Tolson, J. Aldridge, L. G. de Almeida Martins, J. Carungcong, S. Beavis, K. Dale, R. Gascoyne, J. Hawes, K. Pritchard, L. Stevenson, A. Whileman, R. Crawley, A. Crew, M. Cunningham, A. Daniels, L. Harrison, S. Hope, K. Inweregbu, S. Jones, N. Lancaster, J. Matthews, A. Nicholson, G. Wray, L. Benham, Z. Bradshaw, J. Brown, M. Caswell, S. Melling, S. Preston, N. Slawson, E. Stoddard, S. Warden, E. Combes, T. Joefield, S. Monnery, V. Beech, S. Trotman, B. Hopkins, L. Thrasyvoulou, H. Willis, S. Anderson, J. Birch, E. Collins, K. Hammerton, R. O'Leary, C. Abernathy, L. Foster, A. Gratrix, V. Martinson, P. Parkinson, E. Stones, L. Carbral-Ortega, R. Kapoor, D. Loader, K. Castle, C. Brandwood, L. Smith, K. Birchall, L. Kolakaluri, D. Baines, A. Sukumaran, I. Mapfunde, M. Meredith, L. Morris, L. Ryan, A. Clark, J. Sampson, C. Peters, M. Dent, M. Langley, S. Wei, A. Andrew, M. Chablani, A. Kirkby, K. Netherton, M. Bates, J. Dasgin, J. Gill, A. Nilsson, E. Apetri, C. Basikolo, B. Blackledge, L. Catlow, B. Charles, R. Doonan, J. Harris, A. Harvey, D. Horner, K. Knowles, S. Lee, D. Lomas, C. Lyons, T. Marsden, D. McLaughlan, L. McMorrow, J. Pendlebury, J. Perez, M. Poulaka, N. Proudfoot, M. Slaughter, K. Slevin, M. Taylor, V. Thomas, D. Walker, A. Michael, M. Collis, M. Clark, M. Coulding, E. Jude, J. McCormick, O. Mercer, D. Potla, H. Rehman, H. Savill, V. Turner, M. Davey, D. Golden, R. Seaman, J. Hunt, J. Dearden, E. Dobson, M. Mulcahy, S. Munt, G. O'Connor, J. Philbin, C. Rishton, R. Tully, S. Winnard, L. Cagova, A. Fofano, L. Garner, H. Holcombe, S. Mephram, A. M. Mitchell, L. Mwaura, K. Praman, A. Vuylsteke, J. Zamikula, G. Bercades, I. Hass, N. MacCallum, G. Martir, E. Raith, A. Reyes, D. Smyth, A. Taylor, R. A. Hughes, H. Thomas, A. Rees, M. Duskova, J. Phipps, S. Brooks, M. Edwards, P. Alexander, S. Allen, J. Bradley-Potts, C. Brantwood, J. Egan, T. Felton, G. Padden, L. Ward, S. Moss, S. Glasgow, K. Beesley, S. Board, A. Kubisz-Pudelko, A. Lewis, J. Perry, L. Pippard, D. Wood, C. Buckley, A. Brown, J.

Gregory, S. O'Connell, Z. Belagodu, B. Fuller, A. Gherman, O. Olufuwa, R. Paramsothy, C. Stuart, N. Oakley, C. Kamundi, D. Tyl, K. Collins, P. Silva, J. Taylor, L. King, C. Coates, M. Crowley, P. Wakefield, J. Beadle, L. Johnson, J. Sargeant, M. Anderson, C. Jardine, D. Williams, V. Parris, S. Quaid, E. Watson, J. Melville, J. Naisbitt, R. Joseph, M. Lazo, O. Walton, A. Neal, M. Hill, T. Kannan, E. Allan, K. Darlington, F. Davies, J. Easton, S. Kumar, R. Lean, D. Menzies, R. Pugh, X. Qiu, L. Davies, H. Williams, J. Scanlon, G. Davies, C. Mackay, J. Lewis, S. Rees, S. Coetzee, A. Gales, M. Raj, C. Sell, H. Langton, M. Watters, C. Novis, G. Arbane, A. Bociek, S. Campos, N. Grau, T. O. Jones, R. Lim, M. Marotti, C. Whitton, A. Barron, C. Collins, S. Kaul, H. Passmore, C. Prendergast, A. Reed, P. Rogers, R. Shokkar, M. Woodruff, H. Middleton, O. Polgar, C. Nolan, K. Mahay, C. Sri-Chandana, J. Scherewode, L. Stephenson, S. Marsh, H. Bancroft, M. Bellamy, M. Carmody, J. Daglish, F. Moore, J. Rhodes, M. Sangombe, S. Kadiri, J. Scriven, A. Ayers, W. Harrison, J. North, A. Cavazza, M. Cockrell, E. Corcoran, M. Depante, C. Finney, E. Jerome, M. McPhail, M. Nayak, H. Noble, K. O'Reilly, E. Pappa, R. Saha, S. Saha, J. Smith, A. Knighton, M. Gill, P. Paul, V. Ratnam, S. Shelton, I. Wynter, D. Baptista, R. Crowe, R. Fernandes, R. Herdman-Grant, A. Joseph, A. Loveridge, I. McKenley, E. Morino, A. Naranjo, R. Simms, K. Sollesta, A. Swain, H. Venkatesh, J. Khera, J. Fox, R. Barber, C. Hewitt, A. Hildrith, K. Jackson-Lawrence, S. Shepardson, M. Wills, S. Butler, S. Tavares, A. Cunningham, J. Hindale, S. Arif, L. George, S. Twiss, D. Wright, M. Holland, N. Keenan, M. Lyons, H. Wassall, C. Marsh, M. Mahenthiran, E. Carter, T. Kong, O. Adanini, N. Bhatia, M. Msiska, L. Mew, E. Mwaura, F. Williams, L. Wren, S. B. Sutherland, C. Battle, E. Brinkworth, R. Harford, C. Murphy, L. Newey, T. Rees, M. Williams, S. Arnold, D. Brealey, J. Hardy, H. Houlden, E. Moncur, A. Tariq, A. Tucci, K. Convery, D. Fottrell-Gould, L. Hudig, J. Keshet-Price, G. Randell, K. Stammers, M. Abdelrazik, D. Bakthavatsalam, M. Elhassan, A. Ganesan, A. Haldeos, J. Moreno-Cuesta, D. Purohit, R. Vincent, K. Xavier, K. Rohit, F. Alasdair, M. Saleem, C. David, S. Jenkins, Z. Lamond, A. Wall, J. Reynolds, H. Campbell, M. Thompsom, S. Dodds, S. Duffy, D. Butcher, S. O'Sullivan, N. Butterworth-Cowin, B. Deacon, M. Hibbert, C. Potheary, D. Tetla, C. Woodford, L. Durga, G. Kennard-Holden, L. O. R. de Gordo, E. Peasgood, C. Phillips, D. Skinner, J. Gaylard, D. Mullan, J. Newman, E. Davies, S. Sathe, L. Brimfield, Z. Daly, D. Pogson, S. Rose, A. Collins, W. Khaliq, E. T. Gude, L. Allen, E. Beranova, N. Crisp, J. Deery, T. Hazelton, A. Knight, C. Price, S. Tilbey, S. Turki, S. Turney, J. Giles, S. Booth, G. Bell, K. English, A. Katary, L. Wilcox, R. Campbell, N. Clarke, J. Whiteside, M. Mascarenhas, A. Donaldson, J. Matheson, F. Barrett, M. O'Hara, L. O'Keefe, C. Bradley, D. Collier, A. Hormis, R. Walker, V. Maynard, T. Patel, M. Smith, S. Chukkambotla, A. Kazi, J. Hartley, J. Dykes, M. Hijazi, S. Keith, M. Khan, J. Ryan-Smith, P. Springle, J. Thomas, N. Truman, S. Saad, D. Coleman, C. Fine, R. Matt, B. Gay, J. Dalziel, S. Ali, D. Goodchild, R. Harling, R. Bhattejee, W. Goddard, C. Davison, S. Duberly, J. Hargreaves, R. Bolton, M. Verlander, A. Williams, H. Blackman, B. Creagh-Brown, S. Donlon, N. Michalak-Glinska, S. Mtuwa, V. Pristopan, A. Salberg, E. Smith, S. Stone, C. Piercy, J. Verula, D. Burda, R. Montaser, L. Harden, I. Mayangao, C. Marriott, P. Bradley, C. Harris, J. Cooper, C. Finch, S. Liderth, A. Quinn, N. Waddington, K. Fidler, E. Tagliavini, K. Donnelly, L. Abel, M. Brett, B. Digby, L. Gemmell, J. Hornsby, P. MacGoey, P. O'Neil, R. Price, N. Rodden, K. Rooney, R. Sundaram, N. Thomson, R. Flanagan, G. Hughes, S. Latham, E. McKenna, J. Anderson, R. Hull, K. Rhead, D. Branney, J. Frankham, S. Pitts, N. White, D. Cristiano, N. Dormand, Z. Farzad, M. Gummadi, K. Liyanage, B. V. Patel, S. Salmi, G. Sloane, V. Thwaites, M. Varghese, A. C. Zborowski, S. Bean, K. Burt, M. Spivey, C. Eastgate-

Jackson, H. Filipe, D. Martin, A. Maharajh, S. M. Garcia, M. De Neef, C. Lynch, L. Roche, G. S. Howe, J. Singh, K. Turner, H. Ellis, N. Stroud, S. Cherian, S. Cutler, A. E. Heron, A. Roynon-Reed, T. Szakmany, G. Williams, O. Richards, Y. Cheema, N. Ahmad, J. Barker, K. Bauchmuller, S. Bird, K. Cawthron, K. Harrington, Y. Jackson, F. Kibutu, B. Lenagh, S. Masuko, G. H. Mills, A. Raithatha, M. Wiles, J. Willson, H. Newell, A. Lye, L. Nwafor, C. Jarman, S. Rowland-Jones, D. Foote, J. Cole, R. Thompson, J. Watson, L. Hesseldon, I. Macharia, L. Chetam, J. Smith, A. Ford, S. Anderson, K. Birchall, K. Housley, S. Walker, L. Milner, H. Hanratty, H. Trower, P. Phillips, S. Oxspring, B. Donne, E. Bevan, J. Martin, D. Trodd, G. Watson, C. W. Brown, L. Bunni, C. Jennings, M. Latif, R. Marshall, G. Subramanian, N. Bandla, M. Gellamucho, M. Davies, C. Thompson, F. Trim, B. Eapen, C. Ahmed, B. Baines, S. Clamp, J. Colley, R. Haq, A. Hayes, J. Hulme, S. Hussain, S. Joseph, R. Kumar, Z. Maqsood, M. Purewal, B. Chandler, K. Elliott, J. Mallinson, A. Turnbull, K. Dent, E. Horsley, M. N. Akhtar, S. Pearson, D. Potoczna, S. Spencer, H. Blakemore, B. Borislavova, B. Faulkner, E. Gendall, E. Goff, K. Hayes, M. Thomas, R. Worner, K. Smith, D. Stephens, C. C. Delgado, D. Dawson, L. Ding, G. Durrant, O. Ezeobu, S. Farnell-Ward, A. Harrison, R. Kanu, S. Leaver, E. Maccacari, S. Manna, R. P. Saluzzio, J. Queiroz, T. Samakomva, C. Sicat, J. Texeira, E. F. Da Gloria, A. Lisboa, J. Rawlins, J. Mathew, A. Kinch, W. J. Hurt, N. Shah, V. Clark, M. Thanasi, N. Yun, K. Patel, V. Crickmore, G. Debreceni, J. Wilkins, L. Nicol, I. Burn, G. Hambrook, K. Manso, R. Penn, P. Shanmugasundaram, J. Tebbutt, D. Thornton, A. Rostron, A. Roy, L. Woods, S. Cornell, F. Wakinshaw, K. Rogerson, J. Jarman, P. Anderson, K. Archer, K. Austin, C. Davis, A. Durie, O. Kelsall, J. Thrush, C. Vigurs, H. L. Wood, H. Tranter, A. Harrison, N. Cowley, M. McAlindon, A. Burtenshaw, S. Digby, E. Low, A. Morgan, N. Cother, T. Rankin, S. Clayton, A. McCurdy, S. Allibone, R. Mary-Genetu, A. Patel, A. Mac, A. Murphy, P. Mahjoob, R. Nazari, L. Worsley, A. Fagan, I. A. M. Ali, K. Beaumont, M. Blunt, Z. Coton, H. Curgenvin, M. Elsaadany, K. Fernandes, S. M. Ally, H. Rangarajan, V. Sarathy, S. Selvanayagam, D. Vedage, M. White, J. Fernandez-Roman, D. O. Hamilton, E. Johnson, B. Johnston, M. Lopez Martinez, S. Mulla, D. Shaw, A. A. C. Waite, V. Waugh, I. D. Welters, K. Williams, T. Bemand, E. Black, A. Dela Rosa, R. Howle, S. Jhanji, R. R. Baikady, K. C. Tatham, B. Thomas, M. Halkes, P. Mercer, L. Thornton, J. West, T. Baird, J. Ruddy, R. Reece-Anthony, M. Birt, A. Cowton, A. Kay, M. Kent, K. Potts, A. Wilkinson, S. Naylor, E. Brown, M. Clark, S. Purvis, J. Cole, R. Davies, D. Duffin, H. Hill, B. Player, E. Thomas, A. Williams, C. M. Beith, K. Black, S. Clements, A. Morrison, D. Strachan, M. Taylor, M. Clarkson, S. D'Sylva, K. Norman, T. Coventry, S. Fowler, A. McGregor, A. Brady, R. Chan, S. McIvor, H. Prady, H. Whittle, B. Mathew, M. Clapham, R. Harper, U. Poultney, P. Rice, T. Smith, R. Mutch, Y. Baird, A. Butler, I. Chadbourn, L. Folkes, H. Fox, A. Gardner, R. Gomez, G. Hobden, L. Hodgson, K. King, M. Margaron, T. Martindale, E. Meadows, D. Raynard, Y. Thirlwall, D. Helm, J. Margalef, S. Greer, K. Shuker, S. Smuts, J. Duffield, O. Smith, L. Mallon, W. Claire, I. Birkinshaw, J. Carter, K. Howard, J. Ingham, R. Joy, H. Pearson, S. Roche, Z. Scott, E. Knights, A. Price, A. Thomas, C. Thorpe, A. Abraheem, P. Bamford, K. Cawley, C. Dunmore, M. Faulkner, R. Girach, H. Jeffrey, R. Jones, E. London, I. Nagra, F. Nasir, H. Sainsbury, C. Smedley, R. Khade, A. Sundar, G. Tsinaslanidis, T. Behan, C. Burnett, J. Hatton, E. Heeney, A. Mitra, M. Newton, R. Pollard, R. Stead, J. Birch, L. Bough, J. Goodsell, R. Tutton, P. Williams, S. Williams, B. Winter-Goodwin, A. Cowley, J. Highgate, F. Auld, J. Donnachie, I. Edmond, L. Prentice, N. Runciman, D. Salutous, L. Symon, A. Todd, P. Turner, A. Short, L. Sweeney, E. Murdoch, D. Senaratne, K. Burns, A. Higham, T. Anderson, D. Hawcutt, L. O'Malley, L.

Rad, N. Rogers, P. Saunderson, K. S. Allison, D. Afolabi, J. Whitbread, D. Jones, R. Dore, L. Lankester, N. Nikitas, C. Wells, B. Stowe, K. Spencer, S. Cathcart, K. Duffy, A. Puxty, K. Puxty, L. Turner, J. Ireland, G. Semple, P. Barry, P. Hilltout, J. Evitts, A. Tyler, J. Waldron, V. Irvine, B. Shelley, O. Akinkugbe, A. Bamford, E. Beech, H. Belfield, M. Bell, C. Davies, G. A. L. Jones, T. McHugh, H. Meghari, L. O'Neill, M. J. Peters, S. Ray, A. L. Tomas, C. Gorman, A. Gupta, E. Timlick, R. Brady, S. Bonner, K. Hugill, J. Jones, S. Liggett, A. Bashyal, N. Davidson, P. Hutton, S. McKechnie, J. Wilson, N. Flint, P. Rekha, D. Hales, C. Cruz, S. Gopal, N. Harris, V. Lake, S. Metherell, E. Radford, I. Clement, B. Patel, A. Gulati, C. Hays, K. Webster, A. Hudson, A. Webster, E. Stephenson, L. McCormack, V. Slater, R. Nixon, H. Hanson, M. Fearby, S. Kelly, V. Bridgett, P. Robinson, C. Almaden-Boyle, P. Austin, L. Cabrelli, S. Cole, M. Casey, S. Chapman, C. Whyte, A. Brayne, E. Fisher, J. Hunt, P. Jackson, D. Kaye, N. Love, J. Parkin, V. Tuckey, L. van Koutrik, S. Carter, B. Andrew, L. Findlay, K. Adams, M. Bruce, K. Connolly, T. Duncan, H. T. Michael, G. Lindergard, S. Hey, C. Fox, J. Alfonso, L. J. Durrans, J. Guerin, M. Hruska, A. Eltayeb, T. Lamb, T. Hodgkiss, L. Cooper, J. Rothwell, C. Dennis, A. McGregor, S. Srikanan, A. Sukha, K. Davies, L. O'Brien, Z. Omar, E. Perkins, T. Lewis, I. Sutherland, H. Brooke, S. Buckley, J. C. Suarez, R. Charlesworth, K. Hansson, J. Norris, A. Poole, R. Sandhu, B. Sloan, E. Smithson, M. Thirumaran, V. Wagstaff, A. Metcalfe, J. Camsooksai, C. Humphrey, B. Wadams, Y. DeAth, C. Adams, A. Agasou, T. Arden, A. Bowes, P. Boyle, M. Beekes, H. Button, N. Capps, M. Carnahan, A. Carter, D. Childs, D. Donaldson, K. Hard, F. Hurford, Y. Hussain, A. Javaid, J. Jones, S. Jose, M. Leigh, T. Martin, H. Millward, N. Motherwell, R. Rikunenko, J. Stickley, J. Summers, L. Ting, H. Tivenan, L. Tonks, R. Wilcox, M. Bokhari, R. Lucas, W. McCormick, J. Ritzema, A. Sanderson, H. Wild, N. Baxter, S. Henderson, S. Kennedy-Hay, C. McParland, L. Rooney, M. Sim, G. McCreath, M. Brunton, J. Caterson, H. Coles, M. Frise, S. G. Rai, L. Keating, E. Tilney, S. Bartley, P. Bhuie, C. Downes, K. Holding, K. Riches, M. Hilton, M. Hayman, D. Subramanian, P. Daniel, L. Zitter, S. Benyon, S. Marriott, L. Park, S. Keenan, E. Gordon, H. Quinn, K. Baines, G. Andrew, J. Kenneth Baillie, L. Barclay, M. Callaghan, S. Clark, D. Hope, L. Marshall, C. McCulloch, K. Briton, J. Singleton, S. Birch, K. Simpson, J. Craig, C. Demetriou, C. Eckbad, S. Hierons, L. Howie, S. Mitchard, L. Ramos, A. Serrano-Ruiz, K. White, F. Kelly, V. Amin, E. Anastasescu, V. Anumakonda, K. Karthik, R. Kausar, K. Reid, J. Smith, J. Imeson-Wood, A. Bellini, J. Bryant, A. Mayer, A. Pickard, N. Roe, J. Sowter, A. Howlett, K. Criste, R. Cusack, K. Golder, H. Golding, O. Jones, S. Leggett, M. Male, M. Marani, K. Prager, T. Williams, B. Roberts, K. Salmon, P. Gondo, B. Hadebe, A. Kayani, B. Masunda, A. Ahmed, A. Morris, S. Jakkula, K. Long, S. Whiteley, E. Wilby, B. Ogg, S. Moultrie, M. Odam, J. Bewley, Z. Garland, L. Grimmer, B. Gumbrill, R. Johnson, K. Sweet, D. Webster, G. Efford, S. Bennett, E. Goodwin, M. Jackson, A. Kent, C. Tibke, W. Woodyatt, A. Zaki, A. Daniel, J. Finn, R. Saha, P. Bremmer, J. Allan, T. Geary, G. Houston, A. Meikle, P. O'Brien, D. Bell, R. Boyle, K. Douglas, L. Glass, E. Lee, L. Lennon, A. Rattray, R. Charnock, D. McFarland, D. Cosgrove, B. Attwood, P. Parsons, S. Carmody, M. Oblak, M. Popescu, M. Thankachen, R. Baruah, S. Morris, S. Ferguson, A. Shepherd, A. Altabaibeh, A. Alvaro, K. Gilbert, L. Ma, L. Mostoles, C. Parmar, K. Simpson, C. Jetha, L. Booker, A. Pratley, T. Cosier, G. Millen, N. Schumacher, H. Weston, J. Rand, B. Alex, B. Bach, W. S. Barclay, D. Bogaert, M. Chand, G. S. Cooke, A. B. Docherty, J. Dunning, A. da Silva Filipe, T. Fletcher, C. A. Green, E. M. Harrison, J. A. Hiscox, S. Ijaz, S. Khoo, W. S. Lim, A. J. Mentzer, L. Merson, M. Noursadeghi, M. Palmarini, W. A. Paxton, G. Pollakis, N. Price, A. Rambaut, D. L.

- Robertson, V. Sancho-Shimizu, J. T. Scott, T. de Silva, L. Sigfrid, T. Solomon, S. Srischandran, D. Stuart, R. S. Tedder, E. C. Thomson, A. A. Roger Thompson, R. S. Thwaites, R. K. Gupta, C. Palmieri, O. V. Swann, M. Zamboni, M. E. Dumas, J. L. Griffin, Z. Takats, K. Chechi, P. Andrikopoulos, A. Osagie, M. Olanipekun, S. Liggi, M. R. Lewis, G. dos Santos Correia, C. J. Sands, P. Takis, L. Maslen, W. Greenhalf, V. Shaw, S. E. McDonald, K. A. Ahmed, J. A. Armstrong, M. Ashworth, I. G. Asiimwe, S. Bakshi, S. L. Barlow, L. Booth, B. Brennan, K. Bullock, B. W. A. Catterall, J. J. Clark, E. A. Clarke, L. Cooper, H. Cox, C. Davis, O. Dincarslan, C. Dunn, P. Dyer, A. Elliott, A. Evans, L. Finch, L. W. S. Fisher, T. Foster, I. Garcia-Dorival, P. Gunning, C. Hartley, R. L. Jensen, C. B. Jones, T. R. Jones, S. Khandaker, K. King, R. T. Kiy, C. Koukorava, A. Lake, S. Lant, D. Latawiec, L. Lavelle-Langham, D. Lefteri, L. Lett, L. A. Livoti, M. Mancini, S. McDonald, L. McEvoy, J. McLauchlan, S. Metelmann, N. S. Miah, J. Middleton, J. Mitchell, S. C. Moore, E. G. Murphy, R. Penrice-Randal, J. Pilgrim, T. Prince, W. Reynolds, P. Matthew Ridley, D. Sales, V. E. Shaw, R. K. Shears, B. Small, K. S. Subramaniam, A. Szemiel, A. Taggart, J. Tanianis-Hughes, J. Thomas, E. Trochu, L. van Tonder, E. Wilcock, J. Eunice Zhang, L. Flaherty, N. Maziere, E. Cass, A. doce Carracedo, N. Carlucci, A. Holmes, H. Massey, N. Wrobel, K. Morrice, A. MacLean, R. Armstrong, J. Kenneth Baillie, C. Boz, A. Brown, R. Clark, A. Coutts, L. Cullum, N. Day, L. Donnelly, E. Duncan, A. Fawkes, P. Finernan, T. Gilchrist, A. Golightly, K. Hafezi, A. Law, D. Law, R. Law, S. Law, L. Macgillivray, A. Maclean, H. Mal, S. McCafferty, E. McMaster, J. Meikle, L. Murphy, W. Oosthuyzen, T. Paterson, A. Stenhouse, M. Swets, H. Szoor-McElhinney, F. Taneski, L. C. W. Turtle, T. Wackett, M. Ward, J. Weaver, J. Coyle, B. Gallagher, R. Lidstone-Scott, D. Hamilton, K. Schon, A. Furlong, H. Biggs, F. Griffiths, E. Andrews, K. Brickell, M. Smyth, L. Murphy, G. Carson, H. Hardwick, C. Donohue, Mapping the human genetic architecture of COVID-19. *Nature*. **600**, 472–477 (2021).
77. C. P. Fulco, J. Nasser, T. R. Jones, G. Munson, D. T. Bergman, V. Subramanian, S. R. Grossman, R. Anyoha, B. R. Doughty, T. A. Patwardhan, T. H. Nguyen, M. Kane, E. M. Perez, N. C. Durand, C. A. Lareau, E. K. Stamenova, E. L. Aiden, E. S. Lander, J. M. Engreitz, Activity-by-contact model of enhancer–promoter regulation from thousands of CRISPR perturbations. *Nat. Genet.* **51**, 1664–1669 (2019).
78. J. Nasser, D. T. Bergman, C. P. Fulco, P. Guckelberger, B. R. Doughty, T. A. Patwardhan, T. R. Jones, T. H. Nguyen, J. C. Ulirsch, F. Lekschas, K. Mualim, H. M. Natri, E. M. Weeks, G. Munson, M. Kane, H. Y. Kang, A. Cui, J. P. Ray, T. M. Eisenhaure, R. L. Collins, K. Dey, H. Pfister, A. L. Price, C. B. Epstein, A. Kundaje, R. J. Xavier, M. J. Daly, H. Huang, H. K. Finucane, N. Hacohen, E. S. Lander, J. M. Engreitz, Genome-wide enhancer maps link risk variants to disease genes. *Nature*. **593**, 238–243 (2021).
79. J. A. Morris, C. Caragine, Z. Daniloski, J. Domingo, T. Barry, L. Lu, K. Davis, M. Ziosi, D. A. Glinos, S. Hao, E. P. Mimitou, P. Smibert, K. Roeder, E. Katsevich, T. Lappalainen, N. E. Sanjana, Discovery of target genes and pathways at GWAS loci by pooled single-cell CRISPR screens. *Science (80-.)*. **380** (2023), doi:10.1126/science.adh7699.
80. M. Gasperini, A. J. Hill, J. L. McFaline-Figueroa, B. Martin, S. Kim, M. D. Zhang, D. Jackson, A. Leith, J. Schreiber, W. S. Noble, C. Trapnell, N. Ahituv, J. Shendure, A Genome-wide Framework for Mapping Gene Regulation via Cellular Genetic Screens. *Cell*. **176**, 377–390.e19 (2019).

81. M. Blanchette, W. J. Kent, C. Riemer, L. Elnitski, A. F. A. Smith, K. M. Roskin, R. Baertsch, K. Rosenbloom, H. Clawson, E. D. Green, D. Haussler, W. Miller, Aligning multiple genomic sequences with the threaded blockset aligner. *Genome Res.* **14**, 708–715 (2004).
- 5 82. M. A. Hojo, K. Masuda, H. Hojo, Y. Nagahata, K. Yasuda, D. Ohara, Y. Takeuchi, K. Hirota, Y. Suzuki, H. Kawamoto, S. Kawaoka, Identification of a genomic enhancer that enforces proper apoptosis induction in thymic negative selection. *Nat. Commun.* **10** (2019), doi:10.1038/s41467-019-10525-1.
- 10 83. I. Dotan, M. Allez, S. Danese, M. Keir, S. Tole, J. McBride, The role of integrins in the pathogenesis of inflammatory bowel disease: Approved and investigational anti-integrin therapies. *Med. Res. Rev.* **40**, 245–262 (2020).
- 15 84. R. Nasrallah, C. J. Imianowski, L. Bossini-Castillo, F. M. Grant, M. Dogan, L. Placek, L. Kozhaya, P. Kuo, F. Sadiyah, S. K. Whiteside, M. R. Mumbach, D. Glinos, P. Vardaka, C. E. Whyte, T. Lozano, T. Fujita, H. Fujii, A. Liston, S. Andrews, A. Cozzani, J. Yang, S. Mitra, E. Lugli, H. Y. Chang, D. Unutmaz, G. Trynka, R. Roychoudhuri, A distal enhancer at risk locus 11q13.5 promotes suppression of colitis by Treg cells. *Nature.* **583**, 447–452 (2020).
- 20 85. G. X. Y. Zheng, J. M. Terry, P. Belgrader, P. Ryvkin, Z. W. Bent, R. Wilson, S. B. Ziraldo, T. D. Wheeler, G. P. McDermott, J. Zhu, M. T. Gregory, J. Shuga, L. Montesclaros, J. G. Underwood, D. A. Masquelier, S. Y. Nishimura, M. Schnall-Levin, P. W. Wyatt, C. M. Hindson, R. Bharadwaj, A. Wong, K. D. Ness, L. W. Beppu, H. J. Deeg, C. McFarland, K. R. Loeb, W. J. Valente, N. G. Ericson, E. A. Stevens, J. P. Radich, T. S. Mikkelsen, B. J. Hindson, J. H. Bielas, Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8** (2017), doi:10.1038/ncomms14049.
- 25 86. Y. Hao, T. Stuart, M. H. Kowalski, S. Choudhary, P. Hoffman, A. Hartman, A. Srivastava, G. Molla, S. Madad, C. Fernandez-Granda, R. Satija, Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nat. Biotechnol.* (2023), doi:10.1038/s41587-023-01767-y.
- 30 87. T. Stuart, A. Srivastava, S. Madad, C. A. Lareau, R. Satija, Single-cell chromatin state analysis with Signac. *Nat. Methods.* **18**, 1333–1341 (2021).
88. A. Dobin, C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, T. R. Gingeras, STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics.* **29**, 15–21 (2013).
- 35 89. A. R. Quinlan, I. M. Hall, BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics.* **26**, 841–842 (2010).
90. S. S. P. Rao, M. H. Huntley, N. C. Durand, E. K. Stamenova, I. D. Bochkov, J. T. Robinson, A. L. Sanborn, I. Machol, A. D. Omer, E. S. Lander, E. L. Aiden, A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell.* **159**, 1665–1680 (2014).
- 40 91. A. Roayaei Ardakany, H. T. Gezer, S. Lonardi, F. Ay, Mustache: Multi-scale detection of chromatin loops from Hi-C and Micro-C maps using scale-space representation. *Genome Biol.* **21**, 1–17 (2020).

92. E. M. Schmidt, J. Zhang, W. Zhou, J. Chen, K. L. Mohlke, Y. E. Chen, C. J. Willer, GREGOR: Evaluating global enrichment of trait-associated variants in epigenomic features using a systematic, data-driven approach. *Bioinformatics*. **31**, 2601–2606 (2015).
93. MurakawaLab / ReapTEC : First release of ReapTEC pipeline . (2024),
doi:10.5281/zenodo.10972341.
94. T. Borggreffe, B. D. Giaimo, *Enhancers and Promoters* (2021;
<http://www.springer.com/series/7651><https://link.springer.com/10.1007/978-1-0716-1597-3>), vol. 2351.
95. M. Murata, H. Nishiyori-Sueki, M. Kojima-Ishiyama, P. Carninci, Y. Hayashizaki, M. Itoh, Detecting expressed genes using CAGE. *Methods Mol. Biol.* **1164**, 67–85 (2014).
96. N. Tanaka, M. Koido, A. Suzuki, N. Otomo, H. Suetsugu, Y. Kochi, K. Tomizuka, Y. Momozawa, Y. Kamatani, Biobank Japan Project, S. Ikegawa, K. Yamamoto, C. Terao, Eight novel susceptibility loci and putative causal variants in atopic dermatitis. *J. Allergy Clin. Immunol.* **148**, 1293–1306 (2021).
97. P. L. Germain, M. D. Robinson, A. Lun, C. Garcia Meixide, W. Macnair, Doublet identification in single-cell sequencing data using scDblFinder. *F1000Research*. **10**, 1–26 (2022).
98. M. D. Young, S. Behjati, SoupX removes ambient RNA contamination from droplet-based single-cell RNA sequencing data. *Gigascience*. **9**, 1–10 (2020).
99. I. Korsunsky, N. Millard, J. Fan, K. Slowikowski, F. Zhang, K. Wei, Y. Baglaenko, M. Brenner, P. ru Loh, S. Raychaudhuri, Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods*. **16**, 1289–1296 (2019).
100. J. Harrow, A. Frankish, J. M. Gonzalez, E. Tapanari, M. Diekhans, F. Kokocinski, B. L. Aken, D. Barrell, A. Zadissa, S. Searle, I. Barnes, A. Bignell, V. Boychenko, T. Hunt, M. Kay, G. Mukherjee, J. Rajan, G. Despacio-Reyes, G. Saunders, C. Steward, R. Harte, M. Lin, C. Howald, A. Tanzer, T. Derrien, J. Chrast, N. Walters, S. Balasubramanian, B. Pei, M. Tress, J. M. Rodriguez, I. Ezkurdia, J. Van Baren, M. Brent, D. Haussler, M. Kellis, A. Valencia, A. Reymond, M. Gerstein, R. Guigó, T. J. Hubbard, GENCODE: The reference human genome annotation for the ENCODE project. *Genome Res*. **22**, 1760–1774 (2012).
101. H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. **25**, 2078–2079 (2009).
102. T. Smith, A. Heger, I. Sudbery, UMI-tools: Modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. *Genome Res*. **27**, 491–499 (2017).
103. W. J. Kent, A. S. Zweig, G. Barber, A. S. Hinrichs, D. Karolchik, BigWig and BigBed: Enabling browsing of large distributed datasets. *Bioinformatics*. **26**, 2204–2207 (2010).
104. S. Kovaka, A. V. Zimin, G. M. Pertea, R. Razaghi, S. L. Salzberg, M. Pertea, Transcriptome assembly from long-read RNA-seq alignments with StringTie2. *Genome Biol.* **20**, 1–13 (2019).
105. A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*. **30**, 2114–2120 (2014).

106. M. D. Robinson, D. J. McCarthy, G. K. Smyth, edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. **26**, 139–140 (2009).
107. I. Dunham, A. Kundaje, S. F. Aldred, P. J. Collins, C. A. Davis, F. Doyle, C. B. Epstein, S. Fietze, J. Harrow, R. Kaul, J. Khatun, B. R. Lajoie, S. G. Landt, B. K. Lee, F. Pauli, K. R. Rosenbloom, P. Sabo, A. Safi, A. Sanyal, N. Shores, J. M. Simon, L. Song, N. D. Trinklein, R. C. Altshuler, E. Birney, J. B. Brown, C. Cheng, S. Djebali, X. Dong, J. Ernst, T. S. Furey, M. Gerstein, B. Giardine, M. Greven, R. C. Hardison, R. S. Harris, J. Herrero, M. M. Hoffman, S. Iyer, M. Kellis, P. Kheradpour, T. Lassmann, Q. Li, X. Lin, G. K. Marinov, A. Merkel, A. Mortazavi, S. C. J. Parker, T. E. Reddy, J. Rozowsky, F. Schlesinger, R. E. Thurman, J. Wang, L. D. Ward, T. W. Whitfield, S. P. Wilder, W. Wu, H. S. Xi, K. Y. Yip, J. Zhuang, B. E. Bernstein, E. D. Green, C. Gunter, M. Snyder, M. J. Pazin, R. F. Lowdon, L. A. L. Dillon, L. B. Adams, C. J. Kelly, J. Zhang, J. R. Wexler, P. J. Good, E. A. Feingold, G. E. Crawford, J. Dekker, L. Elnitski, P. J. Farnham, M. C. Giddings, T. R. Gingeras, R. Guigó, T. J. Hubbard, W. J. Kent, J. D. Lieb, E. H. Margulies, R. M. Myers, J. A. Stamatoyannopoulos, S. A. Tenenbaum, Z. Weng, K. P. White, B. Wold, Y. Yu, J. Wrobel, B. A. Risk, H. P. Gunawardena, H. C. Kuiper, C. W. Maier, L. Xie, X. Chen, T. S. Mikkelsen, S. Gillespie, A. Goren, O. Ram, X. Zhang, L. Wang, R. Issner, M. J. Coyne, T. Durham, M. Ku, T. Truong, M. L. Eaton, A. Dobin, A. Tanzer, J. Lagarde, W. Lin, C. Xue, B. A. Williams, C. Zaleski, M. Röder, F. Kokocinski, R. F. Abdelhamid, T. Alioto, I. Antoshechkin, M. T. Baer, P. Batut, I. Bell, K. Bell, S. Chakraborty, J. Chrast, J. Curado, T. Derrien, J. Drenkow, E. Dumais, J. Dumais, R. Duttagupta, M. Fastuca, K. Fejes-Toth, P. Ferreira, S. Foissac, M. J. Fullwood, H. Gao, D. Gonzalez, A. Gordon, C. Howald, S. Jha, R. Johnson, P. Kapranov, B. King, C. Kingswood, G. Li, O. J. Luo, E. Park, J. B. Preall, K. Presaud, P. Ribeca, D. Robyr, X. Ruan, M. Sammeth, K. S. Sandhu, L. Schaeffer, L. H. See, A. Shahab, J. Skancke, A. M. Suzuki, H. Takahashi, H. Tilgner, D. Trout, N. Walters, H. Wang, Y. Hayashizaki, A. Reymond, S. E. Antonarakis, G. J. Hannon, Y. Ruan, P. Carninci, C. A. Sloan, K. Learned, V. S. Malladi, M. C. Wong, G. P. Barber, M. S. Cline, T. R. Dreszer, S. G. Heitner, D. Karolchik, V. M. Kirkup, L. R. Meyer, J. C. Long, M. Maddren, B. J. Raney, L. L. Grassefder, P. G. Giresi, A. Battenhouse, N. C. Sheffield, K. A. Showers, D. London, A. A. Bhinge, C. Shestak, M. R. Schaner, S. K. Kim, Z. Z. Zhang, P. A. Mieczkowski, J. O. Mieczkowska, Z. Liu, R. M. McDaniel, Y. Ni, N. U. Rashid, M. J. Kim, S. Adar, Z. Zhang, T. Wang, D. Winter, D. Keefe, V. R. Iyer, M. Zheng, P. Wang, J. Gertz, J. Vielmetter, E. C. Partridge, K. E. Varley, C. Gasper, A. Bansal, S. Pepke, P. Jain, H. Amrhein, K. M. Bowling, M. Anaya, M. K. Cross, M. A. Muratet, K. M. Newberry, K. McCue, A. S. Nesmith, K. I. Fisher-Aylor, B. Pusey, G. DeSalvo, S. L. Parker, S. Balasubramanian, N. S. Davis, S. K. Meadows, T. Eggleston, J. S. Newberry, S. E. Levy, D. M. Absher, W. H. Wong, M. J. Blow, A. Visel, L. A. Pennachio, H. M. Petrykowska, A. Abyzov, B. Aken, D. Barrell, G. Barson, A. Berry, A. Bignell, V. Boychenko, G. Bussotti, C. Davidson, G. Despacio-Reyes, M. Diekhans, I. Ezkurdia, A. Frankish, J. Gilbert, J. M. Gonzalez, E. Griffiths, R. Harte, D. A. Hendrix, T. Hunt, I. Jungreis, M. Kay, E. Khurana, J. Leng, M. F. Lin, J. Loveland, Z. Lu, D. Manthravadi, M. Mariotti, J. Mudge, G. Mukherjee, C. Notredame, B. Pei, J. M. Rodriguez, G. Saunders, A. Sboner, S. Searle, C. Sisu, C. Snow, C. Steward, E. Tapanari, M. L. Tress, M. J. Van Baren, S. Washietl, L. Wilming, A. Zadissa, Z. Zhang, M. Brent, D. Haussler, A. Valencia, N. Addelman, R. P. Alexander, R. K. Auerbach, S. Balasubramanian, K. Bettinger, N.

- Bhardwaj, A. P. Boyle, A. R. Cao, P. Cayting, A. Charos, Y. Cheng, C. Eastman, G. Euskirchen, J. D. Fleming, F. Grubert, L. Habegger, M. Hariharan, A. Harmanci, S. Iyengar, V. X. Jin, K. J. Karczewski, M. Kasowski, P. Lacroute, H. Lam, N. Lamarre-Vincent, J. Lian, M. Lindahl-Allen, R. Min, B. Miotto, H. Monahan, Z. Moqtaderi, X. J. Mu, H. O'Geen, Z. Ouyang, D. Patacsil, D. Raha, L. Ramirez, B. Reed, M. Shi, T. Slifer, H. Witt, L. Wu, X. Xu, K. K. Yan, X. Yang, K. Struhl, S. M. Weissman, L. O. Penalva, S. Karmakar, R. R. Bhanvadiah, A. Choudhury, M. Domanus, L. Ma, J. Moran, A. Victorsen, T. Auer, L. Centanin, M. Eichenlaub, F. Gruhl, S. Heermann, B. Hoeckendorf, D. Inoue, T. Kellner, S. Kirchmaier, C. Mueller, R. Reinhardt, L. Schertel, S. Schneider, R. Sinn, B. Wittbrodt, J. Wittbrodt, G. Jain, G. Balasundaram, D. L. Bates, R. Byron, T. K. Canfield, M. J. Diegel, D. Dunn, A. K. Ebersol, T. Frum, K. Garg, E. Gist, R. S. Hansen, L. Boatman, E. Haugen, R. Humbert, A. K. Johnson, E. M. Johnson, T. V. Kuttyavin, K. Lee, D. Lotakis, M. T. Maurano, S. J. Neph, F. V. Neri, E. D. Nguyen, H. Qu, A. P. Reynolds, V. Roach, E. Rynes, M. E. Sanchez, R. S. Sandstrom, A. O. Shafer, A. B. Stergachis, S. Thomas, B. Vernot, J. Vierstra, S. Vong, H. Wang, M. A. Weaver, Y. Yan, M. Zhang, J. M. Akey, M. Bender, M. O. Dorschner, M. Groudine, M. J. MacCoss, P. Navas, G. Stamatoyannopoulos, K. Beal, A. Brazma, P. Flicek, N. Johnson, M. Lusk, N. M. Luscombe, D. Sobral, J. M. Vaquerizas, S. Batzoglou, A. Sidow, N. Hussami, S. Kyriazopoulou-Panagiotopoulou, M. W. Libbrecht, M. A. Schaub, W. Miller, P. J. Bickel, B. Banfai, N. P. Boley, H. Huang, J. J. Li, W. S. Noble, J. A. Bilmes, O. J. Buske, A. D. Sahu, P. V. Kharchenko, P. J. Park, D. Baker, J. Taylor, L. Lochovsky, An integrated encyclopedia of DNA elements in the human genome. *Nature*. **489**, 57–74 (2012).
108. F. Ramírez, F. Dündar, S. Diehl, B. A. Grüning, T. Manke, DeepTools: A flexible platform for exploring deep-sequencing data. *Nucleic Acids Res.* **42**, 187–191 (2014).
109. H. Li, Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. **00**, 1–3 (2013).
110. W. Xu, Q. Zhong, D. Lin, Y. Zuo, J. Dai, G. Li, G. Cao, CoolBox: a flexible toolkit for visual analysis of genomics data. *BMC Bioinformatics*. **22**, 1–9 (2021).
111. J. Cairns, P. Freire-Pritchett, S. W. Wingett, C. Várnai, A. Dimond, V. Plagnol, D. Zerbino, S. Schoenfelder, B. M. Javierre, C. Osborne, P. Fraser, M. Spivakov, CHiCAGO: Robust detection of DNA looping interactions in Capture Hi-C data. *Genome Biol.* **17**, 1–17 (2016).
112. J. Zhou, C. L. Theesfeld, K. Yao, K. M. Chen, A. K. Wong, O. G. Troyanskaya, Deep learning sequence-based ab initio prediction of variant effects on expression and disease risk. *Nat. Genet.* **50**, 1171–1179 (2018).
113. C. Terao, H. Yoshifuji, T. Matsumura, T. K. Naruse, T. Ishii, Y. Nakaoka, Y. Kirino, K. Matsuo, T. Origuchi, M. Shimizu, Y. Maejima, E. Amiya, N. Tamura, T. Kawaguchi, M. Takahashi, K. Setoh, K. Ohmura, R. Watanabe, T. Horita, T. Atsumi, M. Matsukura, T. Miyata, Y. Kochi, T. Suda, K. Tanemoto, A. Meguro, Y. Okada, A. Ogimoto, M. Yamamoto, H. Takahashi, S. Nakayamada, K. Saito, M. Kuwana, N. Mizuki, Y. Tabara, A. Ueda, I. Komuro, A. Kimura, M. Isobe, T. Mimori, F. Matsuda, Genetic determinants and an epistasis of LILRA3 and HLA-B*52 in Takayasu arteritis. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 13045–13050 (2018).

114. A. Auton, G. R. Abecasis, D. M. Altshuler, R. M. Durbin, D. R. Bentley, A. Chakravarti, A. G. Clark, P. Donnelly, E. E. Eichler, P. Flicek, S. B. Gabriel, R. A. Gibbs, E. D. Green, M. E. Hurles, B. M. Knoppers, J. O. Korbel, E. S. Lander, C. Lee, H. Lehrach, E. R. Mardis, G. T. Marth, G. A. McVean, D. A. Nickerson, J. P. Schmidt, S. T. Sherry, J. Wang, R. K. Wilson, E. Boerwinkle, H. Doddapaneni, Y. Han, V. Korchina, C. Kovar, S. Lee, D. Muzny, J. G. Reid, Y. Zhu, Y. Chang, Q. Feng, X. Fang, X. Guo, M. Jian, H. Jiang, X. Jin, T. Lan, G. Li, J. Li, Y. Li, S. Liu, X. Liu, Y. Lu, X. Ma, M. Tang, B. Wang, G. Wang, H. Wu, R. Wu, X. Xu, Y. Yin, D. Zhang, W. Zhang, J. Zhao, M. Zhao, X. Zheng, N. Gupta, N. Gharani, L. H. Toji, N. P. Gerry, A. M. Resch, J. Barker, L. Clarke, L. Gil, S. E. Hunt, G. Kelman, E. Kulesha, R. Leinonen, W. M. McLaren, R. Radhakrishnan, A. Roa, D. Smirnov, R. E. Smith, I. Streeter, A. Thormann, I. Toneva, B. Vaughan, X. Zheng-Bradley, R. Grocock, S. Humphray, T. James, Z. Kingsbury, R. Sudbrak, M. W. Albrecht, V. S. Amstislavskiy, T. A. Borodina, M. Lienhard, F. Mertes, M. Sultan, B. Timmermann, M. L. Yaspo, L. Fulton, V. Ananiev, Z. Belaia, D. Beloslyudtsev, N. Bouk, C. Chen, D. Church, R. Cohen, C. Cook, J. Garner, T. Hefferon, M. Kimelman, C. Liu, J. Lopez, P. Meric, C. O'Sullivan, Y. Ostapchuk, L. Phan, S. Ponomarov, V. Schneider, E. Shekhtman, K. Sirotkin, D. Slotta, H. Zhang, S. Balasubramaniam, J. Burton, P. Danecek, T. M. Keane, A. Kolb-Kokocinski, S. McCarthy, J. Stalker, M. Quail, C. J. Davies, J. Gollub, T. Webster, B. Wong, Y. Zhan, C. L. Campbell, Y. Kong, A. Marcketta, F. Yu, L. Antunes, M. Bainbridge, A. Sabo, Z. Huang, L. J. M. Coin, L. Fang, Q. Li, Z. Li, H. Lin, B. Liu, R. Luo, H. Shao, Y. Xie, C. Ye, C. Yu, F. Zhang, H. Zheng, H. Zhu, C. Alkan, E. Dal, F. Kahveci, E. P. Garrison, D. Kural, W. P. Lee, W. F. Leong, M. Stromberg, A. N. Ward, J. Wu, M. Zhang, M. J. Daly, M. A. DePristo, R. E. Handsaker, E. Banks, G. Bhatia, G. Del Angel, G. Genovese, H. Li, S. Kashin, S. A. McCarroll, J. C. Nemes, R. E. Poplin, S. C. Yoon, J. Lihm, V. Makarov, S. Gottipati, A. Keinan, J. L. Rodriguez-Flores, T. Rausch, M. H. Fritz, A. M. Stütz, K. Beal, A. Datta, J. Herrero, G. R. S. Ritchie, D. Zerbino, P. C. Sabeti, I. Shlyakhter, S. F. Schaffner, J. Vitti, D. N. Cooper, E. V. Ball, P. D. Stenson, B. Barnes, M. Bauer, R. K. Cheetham, A. Cox, M. Eberle, S. Kahn, L. Murray, J. Peden, R. Shaw, E. E. Kenny, M. A. Batzer, M. K. Konkel, J. A. Walker, D. G. MacArthur, M. Lek, R. Herwig, L. Ding, D. C. Koboldt, D. Larson, K. Ye, S. Gravel, A. Swaroop, E. Chew, T. Lappalainen, Y. Erlich, M. Gymrek, T. F. Willems, J. T. Simpson, M. D. Shriver, J. A. Rosenfeld, C. D. Bustamante, S. B. Montgomery, F. M. De La Vega, J. K. Byrnes, A. W. Carroll, M. K. DeGorter, P. Lacroute, B. K. Maples, A. R. Martin, A. Moreno-Estrada, S. S. Shringarpure, F. Zakharia, E. Halperin, Y. Baran, E. Cerveira, J. Hwang, A. Malhotra, D. Plewczynski, K. Radew, M. Romanovitch, C. Zhang, F. C. L. Hyland, D. W. Craig, A. Christoforides, N. Homer, T. Izatt, A. A. Kurdoglu, S. A. Sinari, K. Squire, C. Xiao, J. Sebat, D. Antaki, M. Gujral, A. Noor, K. Ye, E. G. Burchard, R. D. Hernandez, C. R. Gignoux, D. Haussler, S. J. Katzman, W. J. Kent, B. Howie, A. Ruiz-Linares, E. T. Dermitzakis, S. E. Devine, H. M. Kang, J. M. Kidd, T. Blackwell, S. Caron, W. Chen, S. Emery, L. Fritsche, C. Fuchsberger, G. Jun, B. Li, R. Lyons, C. Scheller, C. Sidore, S. Song, E. Sliwerska, D. Taliun, A. Tan, R. Welch, M. K. Wing, X. Zhan, P. Awadalla, A. Hodgkinson, Y. Li, X. Shi, A. Quitadamo, G. Lunter, J. L. Marchini, S. Myers, C. Churchhouse, O. Delaneau, A. Gupta-Hinch, W. Kretschmar, Z. Iqbal, I. Mathieson, A. Menelaou, A. Rimmer, D. K. Xifara, T. K. Oleksyk, Y. Fu, X. Liu, M. Xiong, L. Jorde, D. Witherspoon, J. Xing, B. L. Browning, S. R. Browning, F. Hormozdiari, P. H. Sudmant, E. Khurana, C. Tyler-Smith, C. A. Albers, Q. Ayub, Y. Chen, V. Colonna, L. Jostins, K. Walter, Y. Xue, M. B. Gerstein, A. Abyzov, S. Balasubramanian, J. Chen, D. Clarke, Y.

- Fu, A. O. Harmanci, M. Jin, D. Lee, J. Liu, X. J. Mu, J. Zhang, Y. Zhang, C. Hartl, K. Shakir, J. Degenhardt, S. Meiers, B. Raeder, F. P. Casale, O. Stegle, E. W. Lameijer, I. Hall, V. Bafna, J. Michaelson, E. J. Gardner, R. E. Mills, G. Dayama, K. Chen, X. Fan, Z. Chong, T. Chen, M. J. Chaisson, J. Huddleston, M. Malig, B. J. Nelson, N. F. Parrish, B. Blackburne, S. J. Lindsay, Z. Ning, Y. Zhang, H. Lam, C. Sisuu, D. Challis, U. S. Evani, J. Lu, U. Nagaswamy, J. Yu, W. Li, L. Habegger, H. Yu, F. Cunningham, I. Dunham, K. Lage, J. B. Jaspersen, H. Horn, D. Kim, R. Desalle, A. Narechania, M. A. W. Sayres, F. L. Mendez, G. D. Poznik, P. A. Underhill, D. Mittelman, R. Banerjee, M. Cerezo, T. W. Fitzgerald, S. Louzada, A. Massaia, F. Yang, D. Kalra, W. Hale, X. Dan, K. C. Barnes, C. Beiswanger, H. Cai, H. Cao, B. Henn, D. Jones, J. S. Kaye, A. Kent, A. Kerasidou, R. Mathias, P. N. Ossorio, M. Parker, C. N. Rotimi, C. D. Royal, K. Sandoval, Y. Su, Z. Tian, S. Tishkoff, M. Via, Y. Wang, H. Yang, L. Yang, J. Zhu, W. Bodmer, G. Bedoya, Z. Cai, Y. Gao, J. Chu, L. Peltonen, A. Garcia-Montero, A. Orfao, J. Dutil, J. C. Martinez-Cruzado, R. A. Mathias, A. Hennis, H. Watson, C. McKenzie, F. Qadri, R. LaRocque, X. Deng, D. Asogun, O. Folarin, C. Happi, O. Omoniwa, M. Stremlau, R. Tariyal, M. Jallow, F. S. Joof, T. Corrah, K. Rockett, D. Kwiatkowski, J. Kooner, T. T. Hien, S. J. Dunstan, N. ThuyHang, R. Fonnier, R. Garry, L. Kanneh, L. Moses, J. Schieffelin, D. S. Grant, C. Gallo, G. Poletti, D. Saleheen, A. Rasheed, L. D. Brooks, A. L. Felsenfeld, J. E. McEwen, Y. Vaydylevich, A. Duncanson, M. Dunn, J. A. Schloss, A global reference for human genetic variation. *Nature*. **526**, 68–74 (2015).
115. Z. Zhu, P. H. Lee, M. D. Chaffin, W. Chung, P. R. Loh, Q. Lu, D. C. Christiani, L. Liang, A genome-wide cross-trait analysis from UK Biobank highlights the shared genetic architecture of asthma and allergic diseases. *Nat. Genet.* **50**, 857–864 (2018).
116. Y. Han, Q. Jia, P. S. Jahani, B. P. Hurrell, C. Pan, P. Huang, J. Gukasyan, N. C. Woodward, E. Eskin, F. D. Gilliland, O. Akbari, J. A. Hartiala, H. Allayee, Genome-wide analysis highlights contribution of immune system pathways to the genetic architecture of asthma. *Nat. Commun.* **11**, 1–13 (2020).
117. J. Z. Liu, S. Van Sommeren, H. Huang, S. C. Ng, R. Alberts, A. Takahashi, S. Ripke, J. C. Lee, L. Jostins, T. Shah, S. Abedian, J. H. Cheon, J. Cho, N. E. Daryani, L. Franke, Y. Fuyuno, A. Hart, R. C. Juyal, G. Juyal, W. H. Kim, A. P. Morris, H. Poustchi, W. G. Newman, V. Midha, T. R. Orchard, H. Vahedi, A. Sood, J. J. Y. Sung, R. Malekzadeh, H. J. Westra, K. Yamazaki, S. K. Yang, J. C. Barrett, A. Franke, B. Z. Alizadeh, M. Parkes, B. K. Thelma, M. J. Daly, M. Kubo, C. A. Anderson, R. K. Weersma, Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nat. Genet.* **47**, 979–986 (2015).
118. E. López-Isac, M. Acosta-Herrera, M. Kerick, S. Assassi, A. T. Satpathy, J. Granja, M. R. Mumbach, L. Beretta, C. P. Simeón, P. Carreira, N. Ortego-Centeno, I. Castellvi, L. Bossini-Castillo, F. D. Carmona, G. Orozco, N. Hunzelmann, J. H. W. Distler, A. Franke, C. Lunardi, G. Moroncini, A. Gabrielli, J. de Vries-Bouwstra, C. Wijmenga, B. P. C. Koeleman, A. Nordin, L. Padyukov, A. M. Hoffmann-Vold, B. Lie, R. Ríos, J. L. Callejas, J. A. Vargas-Hitos, R. García-Portales, M. T. Camps, A. Fernández-Nebro, M. F. González-Escribano, F. J. García-Hernández, M. J. Castillo, M. A. Aguirre, I. Gómez-Gracia, B. Fernández-Gutiérrez, L. Rodríguez-Rodríguez, P. García de la Peña, E. Vicente, J. L. Andreu, M. Fernández de Castro, F. J. López-Longo, L. Martínez, Fonollosa, A. Guillén, G. Espinosa, C. Tolosa, A. Pros, M. Rodríguez-Carballeira, F. J. Narváez, M. Rubio-Rivas, Ortiz-Santamaría, A. B. Madroño, M. A. González-Gay, B.

Díaz, L. Trapiella, A. Sousa, M. V. Egurbide, P. Fanlo-Mateo, L. Sáez-Comet, F. Díaz, Hernández, E. Beltrán, J. A. Román-Ivorra, E. Grau, J. J. Alegre-Sancho, M. Freire, F. J. Blanco-García, N. Oreiro, T. Witte, A. Kreuter, G. Riemekasten, P. Airó, C. Magro, A. E. Voskuyl, M. C. Vonk, R. Hesselstrand, S. Proudman, W. Stevens, M. Nikpour, J. Zochling, J. Sahhar, J. Roddy, P. Nash, K. Tymms, M. Rischmueller, S. Lester, T. Vyse, A. L. Herrick, J. Worthington, C. P. Denton, Y. Allanore, M. A. Brown, T. R. D. J. Radstake, C. Fonseca, H. Y. Chang, M. D. Mayes, J. Martin, GWAS for systemic sclerosis identifies multiple risk loci and highlights fibrotic and vasculopathy pathways. *Nat. Commun.* **10**, 1–14 (2019).

119. J. Chiou, R. J. Geusz, M. L. Okino, J. Y. Han, M. Miller, R. Melton, E. Beebe, P. Benaglio, S. Huang, K. Korgaonkar, S. Heller, A. Kleger, S. Preissl, D. U. Gorkin, M. Sander, K. J. Gaulton, Interpreting type 1 diabetes risk with genetics and single-cell epigenomics. *Nature*. **594**, 398–402 (2021).
120. S. Oki, T. Ohta, G. Shioi, H. Hatanaka, O. Ogasawara, Y. Okuda, H. Kawaji, R. Nakaki, J. Sese, C. Meno, ChIP-Atlas: a data-mining suite powered by full integration of public ChIP-seq data. *EMBO Rep.* **19**, 1–10 (2018).
121. Z. Zou, T. Ohta, F. Miura, S. Oki, ChIP-Atlas 2021 update: a data-mining suite for exploring epigenomic landscapes by fully integrating ChIP-seq, ATAC-seq and Bisulfite-seq data. *Nucleic Acids Res.* **50**, W175–W182 (2022).
122. C. E. Grant, T. L. Bailey, W. S. Noble, FIMO: Scanning for occurrences of a given motif. *Bioinformatics*. **27**, 1017–1018 (2011).
123. J. A. Castro-Mondragon, R. Riudavets-Puig, I. Rauluseviciute, R. Berhanu Lemma, L. Turchi, R. Blanc-Mathieu, J. Lucas, P. Boddie, A. Khan, N. M. Perez, O. Fornes, T. Y. Leung, A. Aguirre, F. Hammal, D. Schmelter, D. Baranasic, B. Ballester, A. Sandelin, B. Lenhard, K. Vandepoele, W. W. Wasserman, F. Parcy, A. Mathelier, JASPAR 2022: The 9th release of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.* **50**, D165–D173 (2022).
124. T. L. Bailey, J. Johnson, C. E. Grant, W. S. Noble, The MEME Suite. *Nucleic Acids Res.* **43**, W39–W49 (2015).
125. A. Siepel, G. Bejerano, J. S. Pedersen, A. S. Hinrichs, M. Hou, K. Rosenbloom, H. Clawson, J. Spieth, L. D. W. Hillier, S. Richards, G. M. Weinstock, R. K. Wilson, R. A. Gibbs, W. J. Kent, W. Miller, D. Haussler, Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* **15**, 1034–1050 (2005).
126. M. Tarailo-Graovac, N. Chen, Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr. Protoc. Bioinforma.*, 1–14 (2009).
127. R. Hubley, R. D. Finn, J. Clements, S. R. Eddy, T. A. Jones, W. Bao, A. F. A. Smit, T. J. Wheeler, The Dfam database of repetitive DNA families. *Nucleic Acids Res.* **44**, D81–D89 (2016).
128. R. A. Policastro, G. E. Zentner, Global approaches for profiling transcription initiation. *Cell Reports Methods*. **1**, 100081 (2021).
129. P. Carninci, C. Kvam, A. Kitamura, T. Ohsumi, Y. Okazaki, M. Itoh, M. Kamiya, K. Shibata, N. Sasaki, M. Izawa, M. Muramatsu, Y. Hayashizaki, C. Schneider, High-

efficiency full-length cDNA cloning by biotinylated CAP trapper. *Genomics*. **37**, 327–336 (1996).

130. M. M. Guerrini, A. Oguchi, A. Suzuki, Y. Murakawa, Cap analysis of gene expression (CAGE) and noncoding regulatory elements. *Semin. Immunopathol.* **44**, 127–136 (2022).
131. M. Kazuo, S. Sumio, Oligo-capping: a simple method to replace the cap structure of eukaryotic mRNAs with oligoribonucleotides. *Gene*. **138**, 171–174 (1994).
132. Y. Y. Zhu, E. M. Machleder, A. Chenchik, R. Li, P. D. Siebert, Reverse transcriptase template switching: A SMART™ approach for full-length cDNA library construction. *Biotechniques*. **30**, 892–897 (2001).
133. C. Plessy, N. Bertin, H. Takahashi, R. Simone, M. Salimullah, T. Lassmann, M. Vitezic, J. Severin, S. Olivarius, D. Lazarevic, N. Hornig, V. Orlando, I. Bell, H. Gao, J. Dumais, P. Kapranov, H. Wang, C. A. Davis, T. R. Gingeras, J. Kawai, C. O. Daub, Y. Hayashizaki, S. Gustincich, P. Carninci, Linking promoters to functional transcripts in small samples with nanoCAGE and CAGEscan. *Nat. Methods*. **7**, 528–534 (2010).
134. T. Kouno, J. Moody, A. T. J. Kwon, Y. Shibayama, S. Kato, Y. Huang, M. Böttcher, E. Motakis, M. Mendez, J. Severin, J. Luginbühl, I. Abugessaisa, A. Hasegawa, S. Takizawa, T. Arakawa, M. Furuno, N. Ramalingam, J. West, H. Suzuki, T. Kasukawa, T. Lassmann, C. C. Hon, E. Arner, P. Carninci, C. Plessy, J. W. Shin, C1 CAGE detects transcription start sites and enhancer activity at single-cell resolution. *Nat. Commun.* **10** (2019), doi:10.1038/s41467-018-08126-5.
135. E. Z. Macosko, A. Basu, R. Satija, J. Nemesh, K. Shekhar, M. Goldman, I. Tirosh, A. R. Bialas, N. Kamitaki, E. M. Martersteck, J. J. Trombetta, D. A. Weitz, J. R. Sanes, A. K. Shalek, A. Regev, S. A. McCarroll, Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell*. **161**, 1202–1214 (2015).
136. A. M. Klein, L. Mazutis, I. Akartuna, N. Tallapragada, A. Veres, V. Li, L. Peshkin, D. A. Weitz, M. W. Kirschner, Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell*. **161**, 1187–1201 (2015).
137. Y. Furuichi, A. LaFiandra, A. J. Shatkin, 5'-Terminal structure and mRNA stability. *Nature*. **266**, 235–9 (1977).
138. D. De Rie, I. Abugessaisa, T. Alam, E. Arner, P. Arner, H. Ashoor, G. Åström, M. Babina, N. Bertin, A. M. Burroughs, A. J. Carlisle, C. O. Daub, M. Detmar, R. Deviatiiarov, A. Fort, C. Gebhard, D. Goldowitz, S. Guhl, T. J. Ha, J. Harshbarger, A. Hasegawa, K. Hashimoto, M. Herlyn, P. Heutink, K. J. Hitchens, C. C. Hon, E. Huang, Y. Ishizu, C. Kai, T. Kasukawa, P. Klinken, T. Lassmann, C. H. Lecellier, W. Lee, M. Lizio, V. Makeev, A. Mathelier, Y. A. Medvedeva, N. Mejhert, C. J. Mungall, S. Noma, M. Ohshima, M. Okada-Hatakeyama, H. Persson, P. Rizzu, F. Roudnický, P. Sætrum, H. Sato, J. Severin, J. W. Shin, R. K. Swoboda, H. Tarui, H. Toyoda, K. Vitting-Seerup, L. Winteringham, Y. Yamaguchi, K. Yasuzawa, M. Yoneda, N. Yumoto, S. Zabierowski, P. G. Zhang, C. A. Wells, K. M. Summers, H. Kawaji, A. Sandelin, M. Rehli, Y. Hayashizaki, P. Carninci, A. R. R. Forrest, M. J. L. De Hoon, An integrated expression atlas of miRNAs and their promoters in human and mouse. *Nat. Biotechnol.* **35**, 872–878 (2017).

139. M. Slyper, C. B. M. Porter, O. Ashenberg, J. Waldman, E. Drokhlyansky, I. Wakiro, C. Smillie, G. Smith-Rosario, J. Wu, D. Dionne, S. Vigneau, J. Jané-Valbuena, T. L. Tickle, S. Napolitano, M. J. Su, A. G. Patel, A. Karlstrom, S. Gritsch, M. Nomura, A. Waghray, S. H. Gohil, A. M. Tsankov, L. Jerby-Arnon, O. Cohen, J. Klughammer, Y. Rosen, J. Gould, L. Nguyen, M. Hofree, P. J. Tramontozzi, B. Li, C. J. Wu, B. Izar, R. Haq, F. S. Hodi, C. H. Yoon, A. N. Hata, S. J. Baker, M. L. Suvà, R. Bueno, E. H. Stover, M. R. Clay, M. A. Dyer, N. B. Collins, U. A. Matulonis, N. Wagle, B. E. Johnson, A. Rotem, O. Rozenblatt-Rosen, A. Regev, A single-cell and single-nucleus RNA-Seq toolbox for fresh and frozen human tumors. *Nat. Med.* **26**, 792–802 (2020).

Acknowledgments: This publication is part of the Human Cell Atlas (www.humancellatlas.org/publications). We thank the study participants who generously provided blood samples. We are grateful to all the members of the FANTOM consortium, the Single-Cell Genome Information Analysis Core at ASHBI Kyoto University, and members of the Murakawa laboratory for helpful inputs and discussions. We thank the sequencing platform at RIKEN IMS. The super-computing resource was provided by the Human Genome Center at the University of Tokyo (SHIROKANE supercomputer) and by the Information System Division at RIKEN (HOKUSAI Sailing Ship supercomputer). We thank Dr. Spyros Goulas (Kyoto University) for critical reading of the manuscript and valuable inputs. We thank Mr. Teppei Komatsu (Pacific Biosciences) and Dr. Miwako Kitazume (Tomy Digital Biology Co., Ltd.) for their helpful suggestions on the MAS-seq method. We thank Dr. Alexandra V. Andreeva and other professional editors from ELSS, Inc. for carefully editing the manuscript. We thank Dr. Shinya Oki (Kyoto University), Mr. Zhaunan Zou (Kyoto University), and Prof. Mitunori Saitou (Kyoto University) for helpful discussions. AO, RS, TH, and SH were supported by the RIKEN Junior Research Associate Program. SB was supported by the JSPS Postdoctoral Fellowship for Overseas Researchers. Lenti_MS2_P65_HSF1_2A_puro and Lenti_MS2_P65_HSF1_2A_puro are available from Addgene under a material transfer agreement with the Addgene.

Funding:

JSPS Grants-in-Aid for Scientific Research (KAKENHI) 16H06153 (YM)
 JSPS Grants-in-Aid for Scientific Research (KAKENHI) 18H03992 (YM)
 JSPS Grants-in-Aid for Scientific Research (KAKENHI) 20K21407 (YM)
 JSPS Grants-in-Aid for Scientific Research (KAKENHI) 20KK0186 (YM)
 JSPS Grant-in-Aid for JSPS Fellows (KAKENHI) 22KF0209 (YM)
 JSPS Grants-in-Aid for Scientific Research (KAKENHI) 18H05285 (KY)
 JSPS Grants-in-Aid for Scientific Research (KAKENHI) 20H00462 (CT)
 JSPS Grants-in-Aid for Scientific Research (KAKENHI) 23K14164 (MK)
 JST ERATO grant J0321001 (YM)
 AMED grant JP223fa627010 (KY)
 AMED grant JP22ama121015 (KY)
 AMED grant JP22ek0410080 (HY)

AMED grant JP21ek0109555 (CT)
 AMED grant JP21tm0424220 (CT)
 AMED grant JP21ck0106642 (CT)
 AMED grant JP23ek0410114 (CT)
 5 AMED grant JP23tm0424225 (CT)
 AMED-CREST grant JP19gm1210009 (MY)
 Grant from Takeda Science Foundation (YM)
 Grant from Japan Foundation for Applied Enzymology (YM)
 Grant from BRISTOL-MYERS SQUIBB and Macomics (MP)
 10 Grant from Mitsubishi Tanabe Pharma (MY)
 Grant from Boehringer Ingelheim (MY)
 AIRC grant n° IG2019 23826 (MP)
 Fondazione AIRC under 5 per mille 2019 22759 (MP)
 Ministero dell'Istruzione, dell'Università e della Ricerca award PNRR M4C2-
 15 Investimento 1.4-CN00000041 – NextGenerationEU (MP)

Author contributions:

Conceptualization: AO, AS, SKom, CT, KY, YM
 Methodology: AO, SKom, RS, HY, YT, YM
 20 Investigation: AO, AS, HY, KT, KM, GI, TH, HS, AK, IY, SH, RK
 Visualization: AO, YM
 Data curation: AO, AS, SKom, HY, SB, RS, RJPB, SKoj, MK, NT, SH, RK, HK, CT
 Formal analysis: AO, SKom, SB, SKoj, MK, CT
 Resources: AS, KM, GI, KY
 25 Software: AO, SKom, RS, RJPB, SKoj, MK, SS, CT
 Funding acquisition: HY, MY, CT, KY, YM
 Project administration: CT, KY, YM
 Supervision: HY, TI, JK, ST, ATK, IE, SKa, HK, KI, HU, YH, MP, PC, MY, NP, CT, KY,
 YM
 30 Writing – original draft: AO, NP, YM
 Writing – review & editing: AO, AS, SKom, HY, SB, RS, RJPB, SKoj, KT, KM, GI, NT,
 SH, SS, JK, ST, ATK, HK, KI, HU, YH, MP, PC, MY, NP, CT, KY, YM

Competing interests: SK is an employee of NTT DATA. YT, AK, IY, and YH are
 employees of K.K. DNAFORM. KT is an employee of Revorf. MY received research grants
 35 from Mitsubishi Tanabe Pharma and Boehringer Ingelheim. YM received grant funding from

Revorf and K.K. DNAFORM. YM is a co-founder of Revorf, and were a member of advisory boards without financial interests. MP is a stakeholder of CheckmAb s.r.l.. MP is on the board of Directors of CheckmAb s.r.l..

Data and materials availability: Raw sequencing data and processed data have been deposited to National Bioscience Database Center (NBDC) Human Database (accession code: hum0350). Raw sequencing data originated from human subjects have been deposited to Japanese Genotype-phenotype Archive (JGA, accession code: JGAS000689). Raw sequencing data originated from Jurkat cell line have been deposited to Sequence Read Archive (DRA, accession code: PRJDB13816). Processed data have been deposited to Genomic Expression Archive (GEA, accession code: E-GEAD-508). Supplementary Data S1 to S10 have been deposited to Dryad (<https://doi.org/10.5061/dryad.pk0p2ngwx>). The ReapTEC code and custom scripts are available at Github (<https://github.com/MurakawaLab/ReapTEC>, https://github.com/MurakawaLab/CAGE_pipeline, https://github.com/koido/MENTR_CD4T) and Zenodo (93).

Supplementary Materials

Materials and Methods

Supplementary Text 1 and 2

Figs. S1 to S21

Tables S1 to S17

References (94–139)

Data S1 to S10

ITEC Consortium Members

Matteo Guerrini¹, Hiroaki Hatano², Kazuyoshi Ishigaki², Yuki Ishikawa³, Shuichiro Komatsu⁴, Michihiro Kono², Yasuhiro Murakawa⁴, Masahiro Nakano², Akiko Oguchi⁴, Raku Son⁴, Akari Suzuki¹, Kazuhiro Takeuchi⁴, Nao Tanaka³, Chikashi Terao³, Kohei Tomizuka³, Kazuhiko Yamamoto¹, Soichiro Yoshino³ (in alphabetical order)

¹Laboratory for Autoimmune Diseases, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan

²Laboratory for Human Immunogenetics, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan

³Laboratory for Statistical and Translational Genetics, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan

⁴RIKEN-IFOM Joint Laboratory for Cancer Genomics, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan

Fig. 1. Identification of bidirectionally transcribed enhancers using 5' scRNA-seq. (A) Overview of assays performed in this study. The number of libraries generated in each assay is shown for each CD4⁺ T cell subset or condition. TSS, transcription start site. **(B)** Uniform

manifold approximation and projection (UMAP) and clustering analysis of single-cell data obtained by integrating 5' scRNA-seq, 5' snRNA-seq, and Multiome (3' snRNA-seq and snATAC-seq) data. (C and D) Violin plots of the expression of the signature genes (C) and proteins (D). (E) Pie charts for the fractions of reads (read 1) mapping to different classes of genomic regions in 5' scRNA-seq of circulating bulk CD4⁺ T cells. A TSS is defined as a ± 300 -bp region around the 5' ends of a transcript (GENCODE v39). Unencoded G reads, reads starting with an unencoded G, which is soft-clipped in read 1. (F) UCSC Genome Browser view of pseudo-bulk 5' scRNA-seq data for circulating bulk CD4⁺ T cells at the *SOCS7* locus. Red, plus strand; blue, minus strand. (G) The RNA levels of bulk CD4⁺ T cell enhancers (defined by CAGE and NET-CAGE) in pseudo-bulk 5' scRNA-seq plotted against those in NET-CAGE. (H) Schematic representation of the read-level pre-filtering and transcribed enhancer call (ReapTEC) approach. Only the reads that started with an unencoded G were retained for downstream analyses. This pre-filtering ensured a highly accurate TSS peak call. On the basis of high-confidence TSS peaks, divergent TSS peak pairs were identified to define bidirectionally transcribed candidate enhancer (btcEnh) regions. (I) Heat maps of DNase I-hypersensitive sites (DHS), histone 3 lysine 27 acetylation (H3K27ac), and histone 3 lysine 4 mono-methylation (H3K4me1) signals in CD4⁺ T cells (ENCODE datasets) ± 2 kb around the btcEnh midpoints identified by 5' scRNA-seq without (-) or with (+) ReapTEC in the Treg cluster 11 of circulating bulk CD4⁺ T cells. (J–L) Genome browser views at *CCL5* (J), *GZMA* (K), and *GPR68* (L) loci comparing chromatin accessibility (snATAC-seq) and TSS signals (5' scRNA-seq).

Fig. 2. An atlas of candidate enhancers across diverse human CD4⁺ T cells. (A) Clustering analyses of scRNA-seq data. From the Group D cell clusters, Th1 cells (clusters 0,1,7) and Th2 cells (clusters 4 and 6) were extracted and re-clustered. Gating strategy to sort each fraction is also shown. *n*, the number of cells. (B) UMAP plot based on surface protein expression and colored by assigned cell types. (C) Velocity estimates projected onto the UMAP of Group A. (D) Expression of *CCR10* (top) and velocity estimates (bottom) projected onto the UMAP of Group E. (E) Heat maps of DHS, H3K27ac, and H3K4me1 signals ± 2 kb around the btcEnh midpoints (*n* = 62,803), shown as in Fig. 1I. (F) Pie chart categorizing snATAC-seq peaks into peaks located within ± 300 bp around the 5' end of the transcript (GENCODE v39) (proximal ATAC [pATAC] peaks) and peaks located outside the region (distal ATAC [dATAC] peaks). pATAC and dATAC peaks were further categorized into transcribed and untranscribed peaks. (G) Heatmap of the scaled *z*-scores of transcription factor-binding motif (TFBM) enrichment ($-\log_{10}[E \text{ values}]$) in different categories of dATAC peaks. Each column corresponds to one sequence motif, and a total of 150 TFBMs that meet a cutoff of $-\log_{10}(E \text{ values}) \geq 100$ in at least one category are shown. (H) Enrichment of the Mediator complex and transcription factor bindings and H3K27ac histone modification in bidirectionally and unidirectionally transcribed dATAC peaks relative to untranscribed dATAC peaks. $*P < 2.2e-16$, Fisher's exact test. (I) Box plots of the distribution of phastCons score for btcEnhs (*n* = 62,803), transcribed dATAC peaks (*n* = 103,558), and untranscribed dATAC peaks (*n* = 88,637). $**P < 0.005$, $***P < 2.2e-16$, Mann-Whitney U test. (J) Cumulative distribution function of cell-type specificity scores of transcription levels at promoters and btcEnhs and those of chromatin accessibility at pATAC and dATAC peaks ($P < 2.2e-16$, Mann-Whitney U test).

Fig. 3. Chromatin contact map in human CD4⁺ T cells at single-nucleosome resolution. (A and B) Micro-C chromatin contact maps at the *CCR7* locus (A) and the *MAF* locus (B) in resting (upper right triangles) and activated CD4⁺ T cells (lower left triangles). Loops are shown under the maps. (C) Bar plots of the enrichment of Micro-C loops linking the indicated genomic

regions over the shuffled control loops. **(D)** Histogram of the distance rank of the genes linked to btcEnhs, where genes are ranked according to the closest TSS to a given btcEnh. btcE-P pairs, btcEnh-promoter pairs. **(E)** Distribution of btcE-P genomic distances.

Fig. 4. Considerable enrichment of disease heritability in btcEnhs. **(A)** Heritability enrichment estimates for 62,803 btcEnhs. Linkage disequilibrium score regression analysis was performed using GWAS summary statistics. MS, multiple sclerosis; T1D, type 1 diabetes; RA, rheumatoid arthritis; SLE, systemic lupus erythematosus; PBC, primary biliary cholangitis; AD, atopic dermatitis; AS, ankylosing spondylitis; IBD, inflammatory bowel disease; UC, ulcerative colitis; COVID_B2, COVID-19 hospitalization vs. population (9,986 vs. 1,877,672); COVID_A2, severe respiratory COVID-19 vs. population (5,101 vs. 1,383,241); COVID_C2, COVID-19 vs. population (38,984 vs. 1,644,784); COVID_B1, COVID-19 hospitalization vs. COVID-19 without hospitalization (4,829 vs 11,816); T2D, type 2 diabetes; BMI, body mass index. **(B)** Enrichment of multiple sclerosis heritability in different classes of genomic regions. Enrichment of heritability was higher in bidirectionally transcribed regions, such as btcEnhs identified by 5' scRNA-seq and bidirectionally transcribed dATAC peaks, than in the other indicated *cis*-regulatory regions. All data are mean $\pm$ s.e.

Fig. 5. Systematic interpretation of cellular and molecular functions of variants associated with immune-mediated diseases. **(A)** Fold enrichment of the number of disease-associated single nucleotide polymorphisms (GWAS-SNPs) that overlap with CD4⁺ T cell btcEnhs over that of random SNPs. SS, Sjögren's syndrome; AA, alopecia areata; AD, atopic dermatitis; PSC, primary sclerosing cholangitis; JIA, juvenile idiopathic arthritis; SSc, systemic sclerosis; Critical, COVID-19 critical illness; BD, Behçet's disease; BiPolar, bipolar disorder; TG, triglyceride; 13snps, 13 COVID-19-associated SNPs; Hospital, COVID-19 hospitalization; Education, educational attainment; HDL, high-density lipoprotein cholesterol; OA, osteoarthritis; LDL, low-density lipoprotein cholesterol; FBG, fasting blood glucose; Reported, reported COVID-19 infection; other abbreviations are as in Fig. 4A. **(B)** Heat map of scored expression of 606 btcEnhs that overlap with SNPs associated with immune-mediated diseases. Dotted-line box highlights btcEnhs activated in stimulated memory CD4⁺ T cells. **(C)** Heat map of scored expression of Treg btcEnhs highlighted in red-line boxes in (B) among the *FOXP3*⁺ clusters of Group A CD4⁺ T cells. **(D)** Number of disease-relevant btcEnhs with predicted target genes (*y* axis) per combination of tools (*x* axis). The pie chart shows the breakdown of btcEnhs by the number of lines of evidence (LE) that contributed to target gene prediction. RCMC, region-capture Micro-C; ABC model, activity-by-contact model. **(E)** UCSC Genome Browser view of CAGE TSS signals showing upregulation of *IL7R* mRNA by CRISPR activation (CRISPRa) of a disease-relevant btcEnh (chr5:35830423–35831044). RCMC loops from the *IL7R* promoter are shown. Zoomed-in view of the btcEnh region on the bottom left corner shows bidirectionally transcribed enhancer RNAs reproducibly induced by CRISPRa. Red, plus strand; blue, minus strand. Scatter plots of the global comparison of transcription at btcEnhs (*n* = 62,803) (left) and promoters (*n* = 90,565) (right) between CRISPRa and controls. Red dots indicate expression levels of the btcEnh (left) and the target *IL7R* promoter (right).

Fig. 6. Evolution of disease-relevant btcEnhs. **(A)** Pie charts of the evolutionary conservation of CD4⁺ T cell btcEnhs. **(B)** Bar plots of the fractions of mammalian-conserved btcEnhs (*n* = 9,183) and primate-specific btcEnhs (*n* = 11,181) that overlap with transposable elements. ****P* < 2.2e-16, Fisher's exact test. **(C)** Proportions of mammalian btcEnhs and ATAC peaks overlapping with various types of transposable elements. **(D)** Conservation of CDSs (*n* =

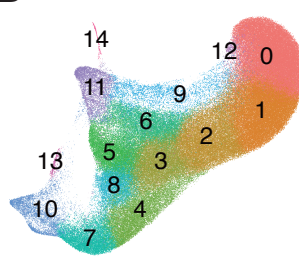
62,803), btcEnhs with ($n = 606$) or without ($n = 62,197$) GWAS-SNPs, and random genomic regions ($n = 62,803$). (E) Frequency distribution of 606 disease-associated btcEnhs categorized by sequence conservation between human and mouse. (F) UCSC Genome Browser view of the CAGE data at the *ACOXL* and *BCL2L11* loci showing activation of the *BCL2L11* promoter via CRISPRa-mediated activation of the mammalian-conserved btcEnh (chr2:110858277–110859299) that contains SNPs associated with RA and T1D. Zoomed-in view of the btcEnh region shows bidirectional enhancer RNAs reproducibly induced by CRISPRa. Red, plus strand; blue, minus strand. (G) UCSC Genome Browser view of an example of a primate-specific btcEnh overlapping with a LINE transposable element located ~1.5 kb upstream of *ITGAL*. This btcEnh harbors SNPs associated with IBD. Activation of this enhancer via CRISPRa results in the induction of the *ITGAL* promoter. Red, plus strand; blue, minus strand.

Fig. 1**A**

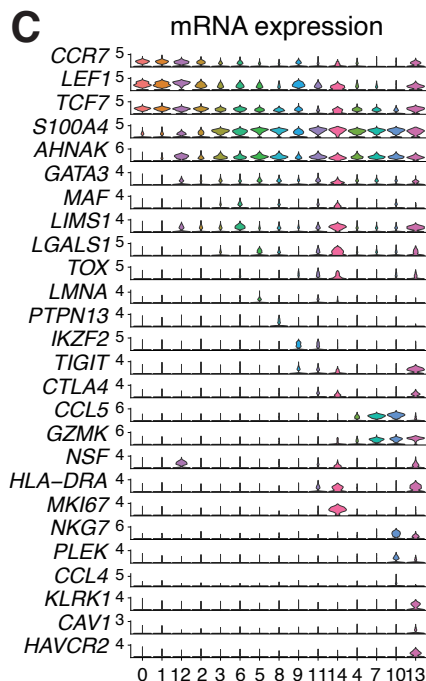
Assays in human
CD4⁺ T cells (original data)

Bulk Treg Th1/2 Th17 Tfh LAG3⁺ Activated Covid study (Meckliff et al.)

Single cell (<i>n</i> = 962,944)	TSS analysis (promoter, enhancer)								
	5' scRNA-seq	22	10	15	12	8	8	1	26
	5' snRNA-seq	3	—	—	—	—	—	—	—
	Gene expression								
	3' scRNA-seq	1	—	—	—	—	—	—	—
	3' snRNA-seq	9	2	2	2	3	1	1	—
	Gene model								
	MAS-seq	1	1	—	—	—	—	—	—
	Chromatin accessibility								
snATAC-seq	9	2	2	2	3	1	1	—	
Protein expression									
5' CITE-seq	2	3	8	2	2	1	—	—	
Bulk	Genome topology								
	Micro-C	34	—	—	—	—	—	30	—
	RMC	36	3	—	2	2	2	30	—
	TSS analysis (promoter, enhancer)								
	CAGE	3	—	—	—	—	—	—	—
NET-CAGE	3	—	—	—	—	—	—	—	

BBulk CD4⁺ T cells (n = 198,681 cells)

0 Naïve_1 8 Th17
1 Naïve_2 9 Treg_1
2 Tfh_1 10 Th1_3
3 Tfh_2 11 Treg_2
4 Th1_1 12 Naïve_3
5 Th2/Th17 13 Exhausted
6 Tfh_3 14 Cycling
7 Th1_2

C**D**

Surface protein expression

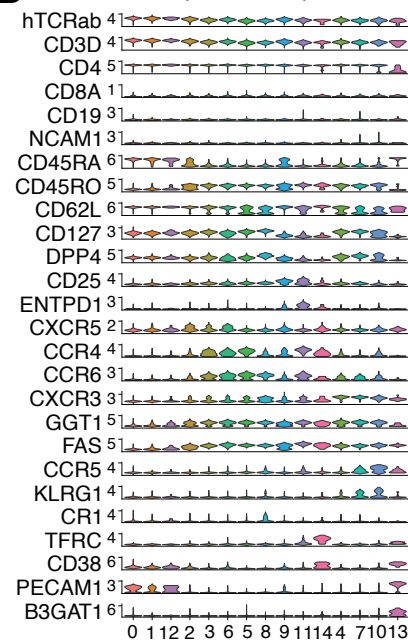
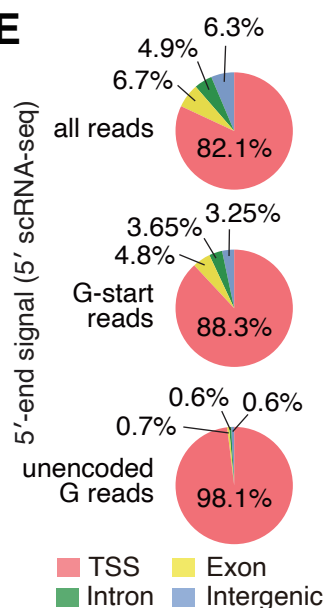
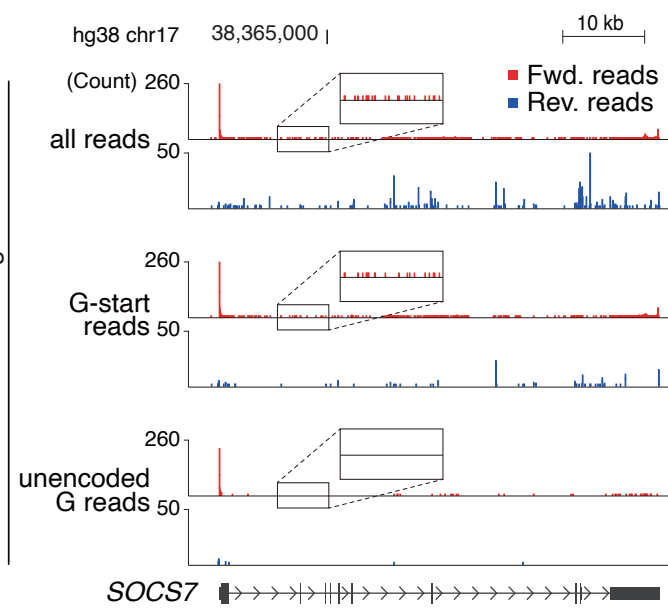
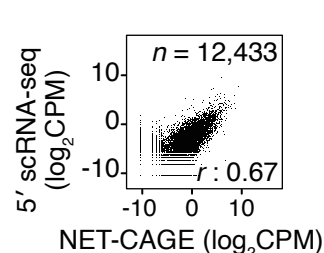
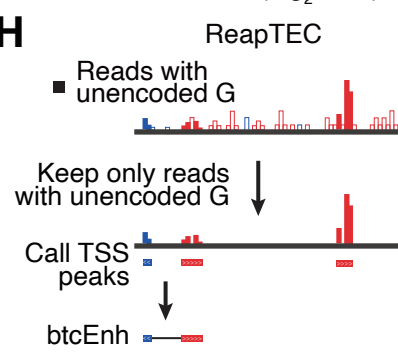
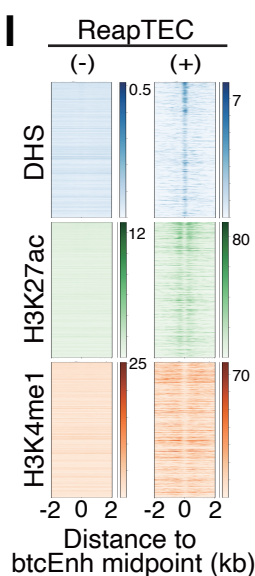
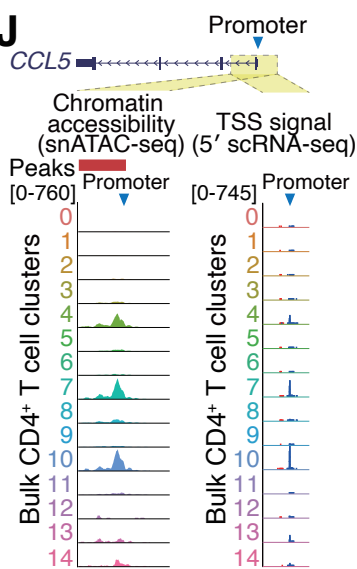
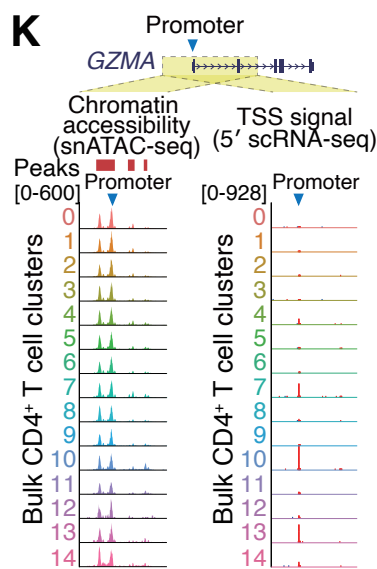
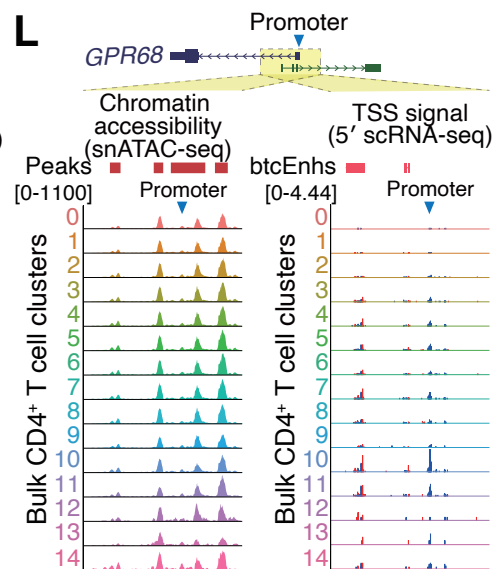
**E****F****G****H****I****J****K****L**

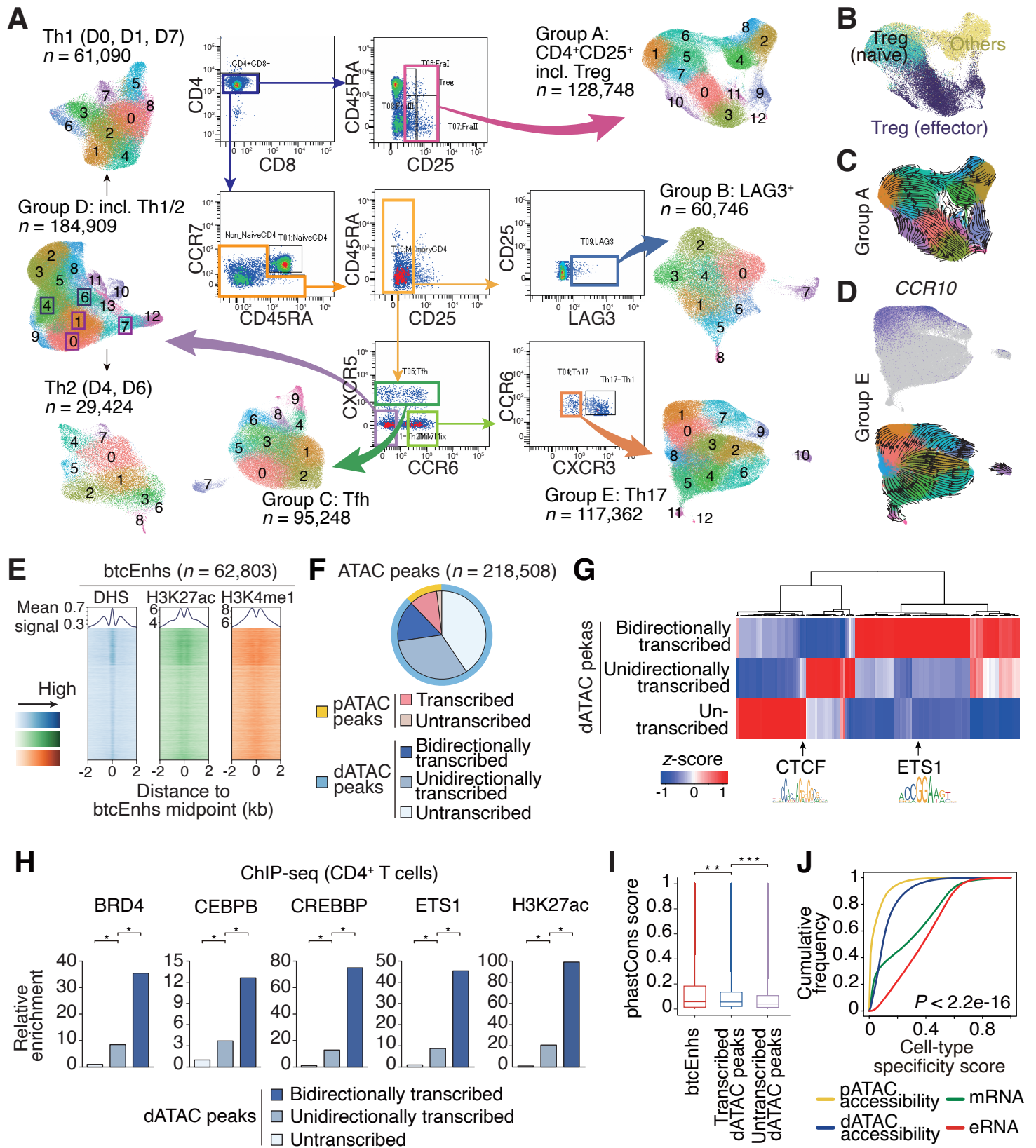
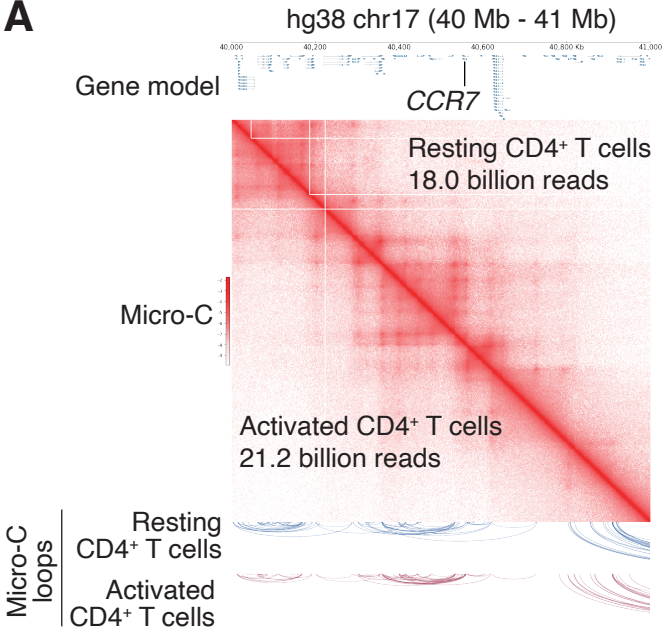
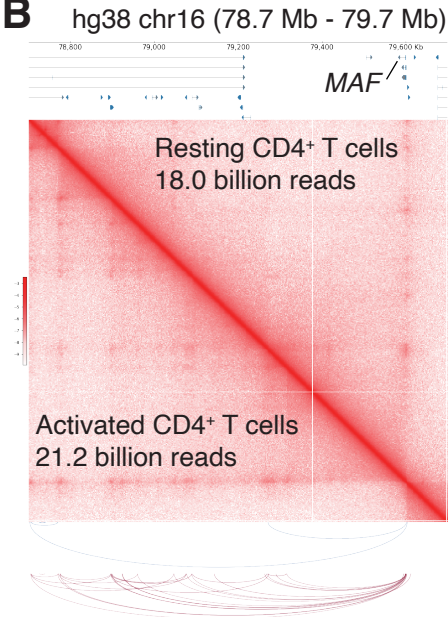
Fig. 2

Fig. 3

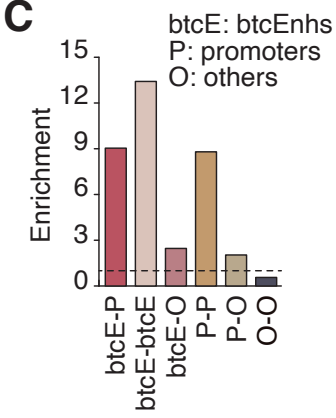
A



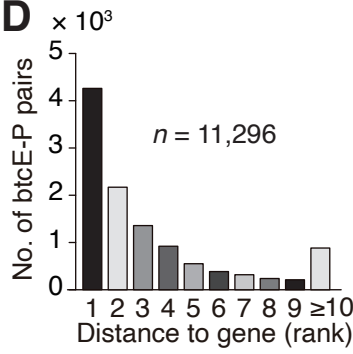
B



C



D



E

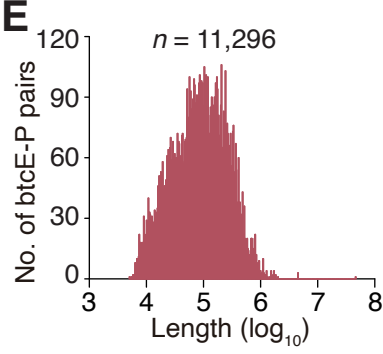


Fig. 4

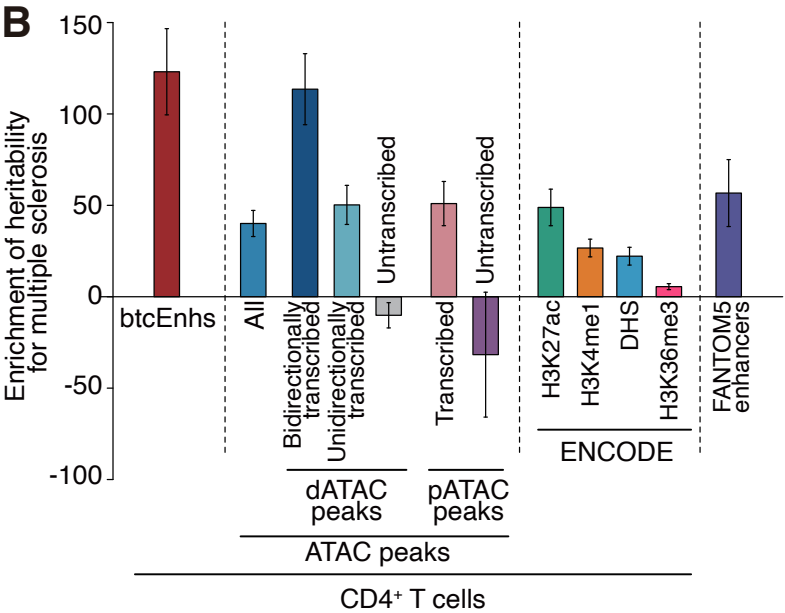
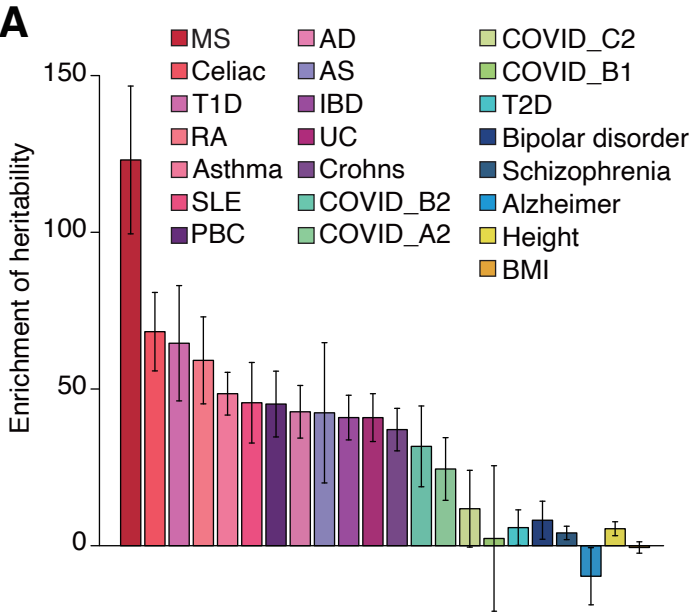


Fig. 5

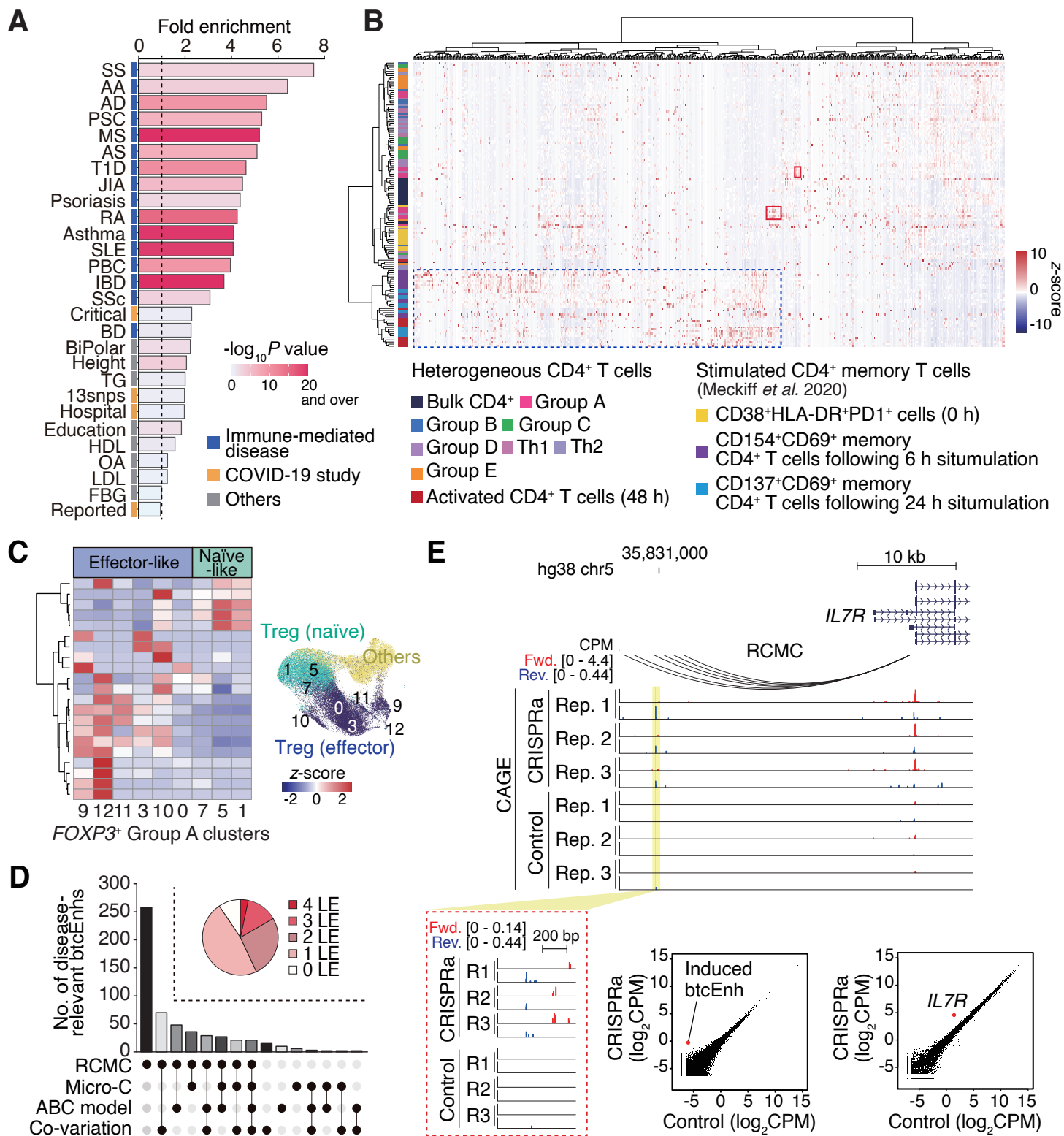
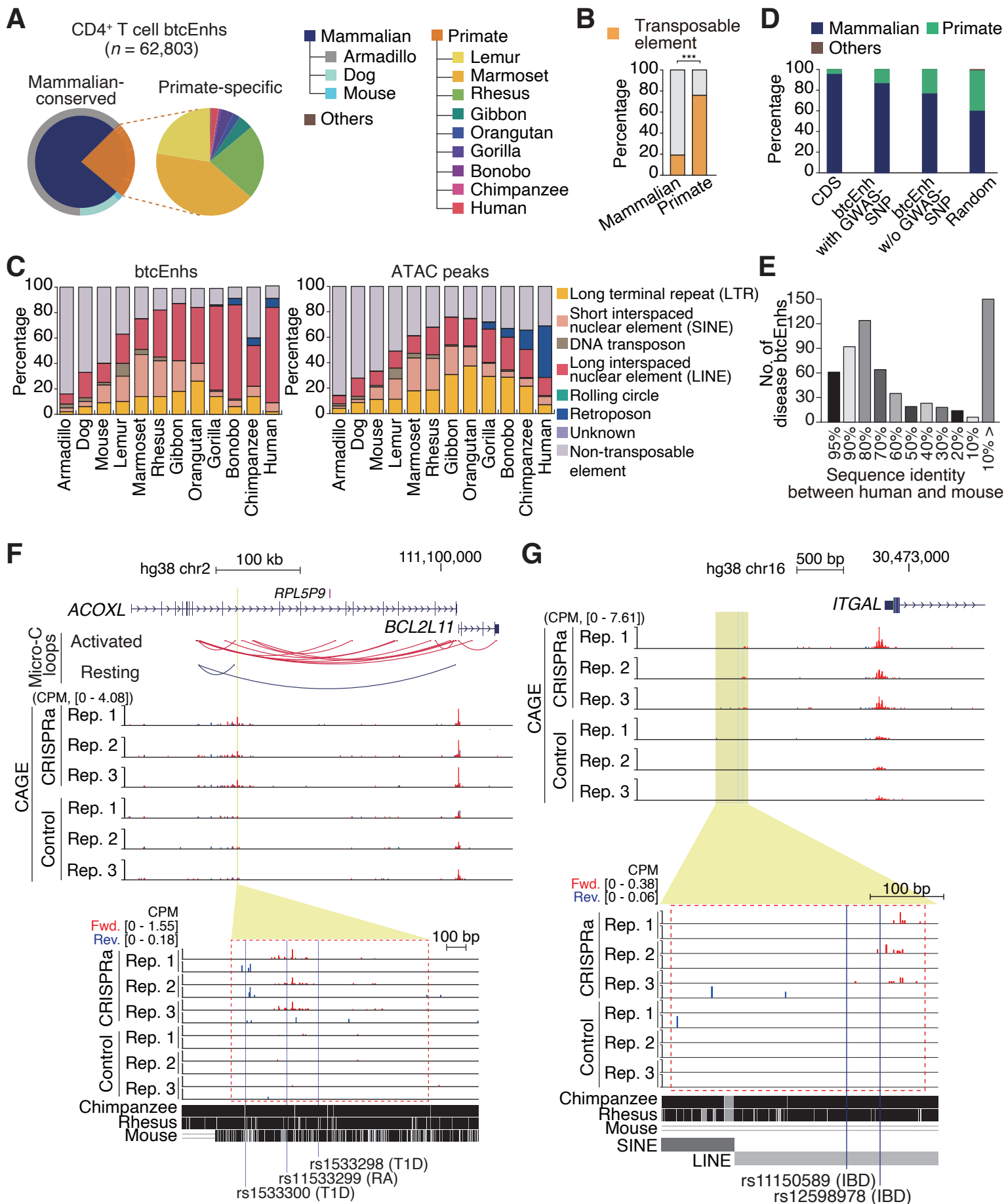


Fig. 6





Supplementary Materials for

An atlas of transcribed enhancers across helper T cell diversity for decoding human diseases

Akiko Oguchi^{1,2,3,†}, Akari Suzuki^{4,†}, Shuichiro Komatsu^{1,5,†}, Hiroyuki Yoshitomi^{2,6}, Shruti Bhagat², Raku Son^{1,2,3}, Raoul Jean Pierre Bonnal⁵, Shohei Kojima⁷, Masaru Koido^{8,9,10}, Kazuhiro Takeuchi^{1,2,11}, Keiko Myouzen⁴, Gyo Inoue⁴, Tomoya Hirai^{1,12}, Hiromi Sano¹, Yujiro Takegami¹³, Ai Kanemaru¹³, Itaru Yamaguchi¹³, Yuki Ishikawa¹⁰, Nao Tanaka¹⁰, Shigeki Hirabayashi^{1,14,15}, Riyo Konishi¹⁶, Sho Sekito^{2,17}, Takahiro Inoue¹⁷, Juha Kere^{18,19,20}, Shunichi Takeda^{21,22}, Akifumi Takaori-Kondo¹⁴, Itaru Endo¹², Shinpei Kawaoka^{16,23}, Hideya Kawaji^{24,25,26}, Kazuyoshi Ishigaki²⁷, Hideki Ueno^{2,6}, Yoshihide Hayashizaki^{13,26}, Massimiliano Pagani^{5,28}, Piero Carninci^{29,30}, Motoko Yanagita^{2,3}, ITEC consortium, Nicholas Parrish⁷, Chikashi Terao^{10,31,32*}, Kazuhiko Yamamoto^{4*}, Yasuhiro Murakawa^{1,2,5,11*}

†Co-first authors

*Corresponding author: chikashi.terao@riken.jp (C.T.); kazuhiko.yamamoto@riken.jp (K.Y.); yasuhiro.murakawa@riken.jp (Y.M.)

The PDF file includes:

Materials and Methods
Supplementary Text 1 and 2
Figs. S1 to S21
Captions for Tables S1 to S17
Captions for Data S1 to S10

Other Supplementary Materials for this manuscript include the following:

Tables S1 to S17
Data S1 to S10

Materials and Methods

Subhead

Ethics statement

All experiments using human samples were approved by the ethical review committee of RIKEN [approval no. H30-9(13)], and were performed in accordance with the relevant guidelines and regulations. Written informed consent was obtained from all healthy donors.

Human CD4⁺ T cell isolation and flow cytometry

Blood samples were collected into vacuum tubes (Venoject II, Terumo, Cat#VP-H100K). CD4⁺ T cells were isolated by the immunomagnetic negative selection method using the MACSxpress® Whole Blood CD4 T Cell Isolation Kit, human (Miltenyi Biotec, Cat#130-098-195). Contaminating red blood cells were lysed with ACK Lysing Buffer (Thermo Fisher Scientific, Cat#A1049201). Cells were pelleted and washed twice with washing buffer: PBS (Invitrogen, Cat#10010049) supplemented with 1% FBS (Biowest, Cat#S1400-500) and 1 mM EDTA (Invitrogen, Cat#15575020). Isolated CD4⁺ T cells were resuspended in 100 µl of washing buffer, then 20 µl of human Fc receptor binding inhibitor (Thermo Fisher Scientific, Cat#14-9161-73) was added and the cells were incubated on ice for 10 min. Cells were pelleted and washed once with 1 ml of washing buffer and resuspended in 103.25 µl of an antibody cocktail and incubated for 20 min on ice. The cocktail contained 50 µl of Brilliant Stain Buffer (BD Biosciences, Cat#563794) supplemented with fluorochrome-conjugated antibodies: 5 µl CD25 (clone BC96, AF488, BioLegend, Cat#302616), 20 µl LAG-3 (polyclonal, PE, R&D, Cat#FAB2319P), 5 µl CXCR5 (CD185) (clone RF8B2, PerCP-Cy5.5, BD Biosciences, Cat#562781), 0.75 µl CD3 (clone UCHT1, PE-Cy7, BioLegend, Cat#300420), 2.5 µl CD196 (CCR6) (clone 11A9, APC, BD Biosciences, Cat#560619), 1 µl CD45RA (clone HI100, APC-Cy7, BioLegend, Cat#304128), 5 µl CD183 (clone 1C6/CXCR3, BV421, BD Biosciences, Cat#562558), 1 µl CD4 (clone RPA-T4, BV510, BD Biosciences, Cat#740161), 5 µl CD194 (CCR4) (clone L291H4, BV605, BioLegend, Cat#359418), 7.5 µl CD197 (CCR7) (clone G043H7, BV711, BioLegend, Cat#353228), and 0.5 µl CD8 (clone RPA-T8, PE-CF594, BD Biosciences, Cat#562282). Stained cells were washed twice with 1 ml of washing buffer and were resuspended at 1.0×10^7 cells/ml in basic sort buffer (1× HBSS[-] without Phenol Red [Fujifilm Wako Pure Chemicals Corporation, Cat#085-09355], 1% FBS, 1 mM EDTA [Invitrogen, Cat#15575020], 25 mM HEPES [Invitrogen, Cat#15630080]). Cells were then passed through a 35-µm strainer (Falcon, Cat#352235) to remove debris. Fluorescence data were acquired by using a FACS Aria IIu Cell Sorter (BD Biosciences). CD3⁺CD4⁺CD8⁻CCR7⁺CD45RA⁺ T cells were defined as naïve T cells. Sorted cells were collected in 1 ml of basic sort buffer, washed twice with 500 µl of 1× PBS supplemented with 0.1% BSA (Invitrogen, Cat#AM2618), and finally resuspended in the same buffer at 1.0×10^6 cells/ml.

CD4 ⁺ T cell groups	Surface marker staining for flow cytometry	The number of cells analyzed
Group A	CD3 ⁺ CD4 ⁺ CD8 ⁻ CD25 ⁺ cells, including Tregs	128,748
Group B	CD3 ⁺ CD4 ⁺ CD8 ⁻ CD25 ⁻ LAG3 ⁺ non-naïve T cells	60,746
Group C	CD3 ⁺ CD4 ⁺ CD8 ⁻ CD25 ⁻ CXCR5 ⁺ non-naïve T cells, including Tfh cells	95,248
Group D	CD3 ⁺ CD4 ⁺ CD8 ⁻ CD25 ⁻ CCR6 ⁻ CXCR5 ⁻ non-naïve T cells,	184,909

	including Th1 and Th2 cells	
Group E	CD3 ⁺ CD4 ⁺ CD8 ⁻ CD25 ⁻ CCR6 ⁺ CXCR5 ⁻ CXCR3 ⁻ non-naïve T cells, including Th17 cells	117,362

Isolation of nuclei for Multiome assay and 5' single-nucleus RNA-seq

Nuclei were isolated according to the manufacturer's instructions for the "Low Cell Input Nuclei Isolation" protocol (https://cdn.10xgenomics.com/image/upload/v1660261285/support-documents/CG000365_DemonstratedProtocol_NucleiIsolation_ATAC_GEX_Sequencing_RevC.pdf) with minor modifications. Nuclei were prepared either on ice or in a centrifuge at 4°C. Each centrifugation step was carried out for 5 min at 500 ×g. Approximately 100,000 cells were washed in ice-cold 1% FBS in PBS, centrifuged, and then resuspended in 100 µl of lysis buffer (0.01% digitonin [Thermo Fisher Scientific, Cat#BN2006], 0.1% NP-40 [Sigma-Aldrich, Cat#NP40S-100ML], 0.1% Tween-20 [Bio-Rad, Cat#1662404], 10 mM Tris-HCl, pH 7.4 [Sigma-Aldrich, Cat#T2194], 10 mM NaCl [Nacalai Tesque, Cat#06900-14], 3 mM MgCl₂ [Nippon Gene, Cat#310-90361], 1% BSA [Miltenyi Biotec, Cat#130-091-376], 1 mM DTT [Invitrogen, Cat#Y00147], and 1 U/µl RNase inhibitor [Sigma-Aldrich, Cat#3335402001] in nuclease-free water). The cells were incubated on ice for 1 min, and 400 µl of wash buffer (0.1% Tween-20, 10 mM Tris-HCl [pH 7.4], 10 mM NaCl, 3 mM MgCl₂, 1% BSA, 1 mM DTT, and 1 U/µl RNase inhibitor in nuclease-free water) was added. The cells were centrifuged, the supernatant (495 µl) was removed, and 45 µl of 1× Nuclei Buffer supplemented with 1 mM DTT and 1 U/µl RNase inhibitor in nuclease-free water was added to the pellet without pipetting. The sample was centrifuged, and the nuclei were resuspended in 1× Nuclei Buffer, and were proceed to Multiome assays or 5' snRNA-seq assays.

Single-cell RNA sequencing and single-nucleus RNA sequencing

Human CD4⁺ T cells and FACS-sorted heterogenous populations were processed with Chromium Next GEM Single Cell 5' and 3' kits (10x Genomics). In brief, cell suspensions were loaded onto a Chromium single-cell controller (10x Genomics) to generate single-cell gel bead-in-emulsions (GEMs). Reverse transcription was performed in GEMs, generating cDNA tagged with a cell barcode and a unique molecular index (UMI), followed by barcoded cDNA amplification and library construction. The version of the kit for each sample is listed in **Table S1**. Libraries were sequenced on an Illumina NovaSeq 6000 sequencing platform using paired-end, dual-index sequencing with 150 cycles for read 1 and 150 cycles for read 2.

Multiome assay

Human CD4⁺ T cells and FACS-sorted heterogenous populations were collected as described in the 'Human CD4⁺ T cell isolation and flow cytometry' subsection. Part of resting CD4⁺ T cells were kept in culture without any beads for 24 h and then stimulated with Immunocult Human CD3/CD28 T cell activator Dynabeads (Veritas, Cat#10991) for 48 h. Multiome assay was performed according to the manufacturer's instructions (https://cdn.10xgenomics.com/image/upload/v1666737555/support-documents/CG000338_ChromiumNextGEM_Multiome_ATAC_GEX_User_Guide_RevF.pdf). In brief, transposition was performed in a 10 µl reaction mixture at 37°C for 60 min, targeting up to 10,000 nuclei. Subsequently, nuclei were loaded onto a Chromium Chip J (10x Genomics, Cat# 2000264) for GEM generation and barcoding. The post-GEM cleanup was followed by a

pre-amplification step. Subsequently, each sample was divided into two fractions: one was allocated for the generation of a 3' snRNA-seq library, and the other for a snATAC-seq library. To enable multiplexing, both libraries were indexed using the Chromium i7 Sample Index N from the Set A kit (10x Genomics, Cat#3000262) for 3' snRNA-seq and the Chromium i7 Sample Index TT from the Set A kit (10x Genomics, Cat#3000431) for snATAC-seq. Finally, both libraries were pooled and sequenced on an Illumina NovaSeq 6000 platform using paired-end, dual-index sequencing with 150 cycles for read 1, 10 cycles for i7 index, 24 cycles for i5 index, and 150 cycles for read 2.

CITE-seq

We stained 5.0×10^5 circulating bulk CD4⁺ T cells and FACS-sorted heterogeneous subpopulations with TotalSeqTM-C Universal Cocktail, V1.0 (BioLegend, Cat#399905). In brief, cells were blocked with 2.5 μ l of Human TruStain FcXTM blocking reagent (BioLegend, Cat#422301) diluted in 22.5 μ l of Cell Staining Buffer (BioLegend, Cat#420201) for 10 min at 4°C. The antibody cocktail (25 μ l) was added, and cells were incubated for 30 min at 4°C. Cells were washed 3 times with Cell Staining Buffer, filtered through a 40- μ m FlowmiTM Cell Strainer (Bel-Art, H-B Instrument, Cat#H13680-0040), and resuspended at 1.0×10^6 cells/ml in 1 $\times$ PBS with 0.04% BSA (Invitrogen, Cat#AM2618). Cells were loaded onto a Chromium Next GEM Chip K or Chip N (10x Genomics, Cat#1000286 or Cat#1000375) using the V(D)J kit (v2) (10x Genomics, Cat#1000263 or Cat#1000356). Libraries were constructed according to the manufacturer's instructions and were sequenced on an Illumina NovaSeq 6000 sequencing platform using paired-end, dual-index sequencing with 150 cycles for read 1, 10 cycles for i7 index, 10 cycles for i5 index, and 150 cycles for read 2.

5' MAS-seq assay

MAS-seq library was prepared according to the manufacturer's instructions (<https://www.pacb.com/wp-content/uploads/Procedure-checklist-preparing-MAS-Seq-libraries-using-MAS-Seq-for-10x-single-cell-3-kit.pdf>) with minor modifications to adapt it to 5' scRNA-seq cDNA. In brief, 5' scRNA-seq cDNAs (15 ng each) from bulk CD4⁺ T cells (average length, 1,473 bp) or CD4⁺CD25⁺ T cells (1,432 bp) were used as inputs. To remove TSO priming artifacts, TSO PCR was performed with Fwd Primer (AAO272, CTACACGACGCTCTTCCGAT*C*T) and Rev Primer (AAO273, /5Biosg/UAUAAGCAGTGGTATCAACGCAG*A*G) (* within the DNA sequence means an internal phosphorothioate bond). These primers were adopted from the original manuscript (52) instead of the MAS capture primers from the kit. After cleanup with 1.5 $\times$ SMRTbell beads, and DNA fragments containing TSO artifacts were removed using MAS capture beads. Purified cDNA (50 μ g) was used in 16 parallel amplification reactions with MAS primers to generate cDNAs containing orientation-specific MAS segmentation adapter sequences. After cleanup with 1.5 $\times$ SMRTbell beads, amplified cDNAs were treated with MAS enzyme to create single-stranded extensions. MAS adapters and MAS ligase were added to concatenate cDNAs into a directional-linear array. After cleanup, DNA damage repair and nuclease treatment were performed, followed by another cleanup (both with 1.2 $\times$ SMRTbell beads), and SMRTbell libraries were eluted in elution buffer. MAS-seq libraries were subjected to standard CCS library preparation and were subsequently sequenced on a PacBio Sequel IIe system.

Micro-C and region capture Micro-C assays

Resting CD4⁺ T cells were collected as described in the ‘Human CD4⁺ T cell isolation and flow cytometry’ subsection. Part of these cells were kept in culture without any beads for 24 h and then stimulated with Immunocult Human CD3/CD28 T cell activator Dynabeads (Veritas, Cat#10991) for 24 h. Activated CD4⁺ T cells (1.0×10^6) were then collected as a pellet and snap-frozen at -80°C . Micro-C libraries were prepared using a Dovetail Micro-C Kit (Cantata Bio, Cat#21006) according to the manufacturer’s instructions (<https://dovetailgenomics.com/dovetail-micro-c-kit-user-guide-version-1-0/>). In brief, chromatin was crosslinked using a combination of disuccinimidyl glutarate (Thermo Fisher Scientific, Cat#A35392) and 37% formaldehyde (Sigma–Aldrich, Cat#252549). The crosslinked chromatin was then digested *in situ* with micrococcal nuclease (MNase), whose concentration was determined by titration for each batch of crosslinked cells. After digestion, chromatin fragments were extracted by lysing the cells with SDS and were then bound to Chromatin Capture Beads. The chromatin ends were repaired, ligated using biotinylated bridge adapters, and the ends containing the adapters were proximity ligated. The crosslinks were then reversed and associated proteins were degraded. The resulting DNA was purified and converted to a sequencing library using adaptors compatible with Illumina sequencing. Micro-C libraries were sequenced on an Illumina NovaSeq 6000 platform using paired-end, single-index sequencing with 150 cycles for read 1, 8 cycles for i7 index, and 150 cycles for read 2. RCMC libraries were prepared using a Dovetail Target Enrichment Kit (Cantata Bio, human pan promoter panel, Cat#25013) according to the manufacturer’s instructions (Cantata Bio, <https://dovetailgenomics.com/wp-content/uploads/securepdfs/2022/07/Dovetail@-Target-Enrichment-User-Guide-Version-1.0.pdf>). This step included hybridization of a biotin-labeled promoter probe panel with the Micro-C library pool, conducted at 70°C for 16 h. Hybridized targets were captured on and enriched with streptavidin beads, and a post-capture PCR was performed to assess the quality of the library. RCMC libraries were sequenced on an Illumina NovaSeq 6000 platform using paired-end, single-index sequencing with 250 cycles for read1, 8 cycles for i7 index, and 250 cycles for read 2.

CRISPR activation (CRISPRa) analysis

Jurkat cells (a human T lymphocyte cell line, RCB0806) were obtained from the RIKEN BRC. dCas9-VP64_GFP (plasmid 61422; Addgene) and Lenti_MS2_P65_HSF1_2A_puro (plasmid 118699; Addgene) were gifts from F. Zhang and C. Vakoc, respectively. CRISPRa-Jurkat cell line stably expressing dCas9-VP64 and MS2_P65_HSF1 was established by lentiviral transduction and subsequent selection of GFP-expressing puromycin-resistant cells. Protospacer adjacent motif (PAM) sequence for each btcEnh region was designed using Custom Alt-R™ CRISPR-Cas9 guide RNA design tool (https://sg.idtdna.com/site/order/designtool/index/CRISPR_SEQUENCE) and are listed in **Table S15**. A 130-base single-guide RNA (sgRNA) harboring an MS2 stem–loop sequence was synthesized by IDT with 2'-O-methyl-modified phosphorothioate modifications at three bases at each 5' and 3' end. Non-targeting control sgRNA or targeting sgRNA (100 nM) was transfected to CRISPRa-Jurkat cells using DharmaFECT 4 (Horizon Discovery, Cat#T-2004). Transfected cells were incubated for 48 h and subjected to CAGE assay with 3 technical replicates per experiment. Total RNA was extracted with an miRNeasy Mini kit (Qiagen, Cat#217004) and digested on column with DNase I from an RNase-free DNase set (Qiagen, Cat#79254). CAGE libraries were generated from total RNA (1 μg) as previously described (LQ-ssCAGE, (94)) with minor modifications, and were sequenced on an Illumina NovaSeq 6000 platform using paired-

end, dual-index sequencing with 150 cycles for read 1, 8 cycles for i7 index, 8 cycles for i5 index, and 150 cycles for read 2.

Total and nascent RNA purification from CD4⁺ T cells

We used the procedure described previously (10) with minor modifications. Cells ($\sim 1.0 \times 10^7$) were pelleted and lysed with 1.6 ml of Nuclei EZ Lysis Buffer (Sigma–Aldrich, Cat#NUC101) supplemented with 25 μ M α -amanitin (Fujifilm Wako Pure Chemical Corporation, Cat#010-22961), 1 $\times$ cOmplete Protease Inhibitor Cocktail (Sigma–Aldrich, Cat#04693116001), and SUPERase•IN RNase Inhibitor (200 units; Thermo Fisher Scientific, Cat#AM2696) and were incubated on ice for 5 min. Lysate (200 μ l) was mixed with 700 μ l of QIAzol (Qiagen, Cat#79306) for total RNA extraction. The remaining lysate (1.4 ml) was centrifuged at 2,000 $\times g$ for 5 min at 4°C. The nuclear pellet was rinsed with 1 ml of ice-cold 1 $\times$ D-PBS (Fujifilm Wako Pure Chemical Corporation, Cat#045-29795), resuspended in 200 μ l of nuclei lysis buffer (20 mM HEPES, pH 7.5 [Nacalai Tesque, Cat#17557-94], 300 mM NaCl [Nacalai Tesque, Cat#06900-14], 1% TERGITOL solution Type NP-40 [Sigma–Aldrich, Cat#NP40S-100ML], 2 M urea [Fujifilm Wako Pure Chemical Corporation, Cat#219-00175], 0.2 mM EDTA [Fujifilm Wako Pure Chemical Corporation, Cat#346-01971], 1 mM dithiothreitol [Fujifilm Wako Pure Chemical Corporation, Cat#048-29224], 25 μ M α -amanitin, 1 $\times$ cOmplete Protease Inhibitor Cocktail and SUPERase•IN RNase Inhibitor [200 units]), and incubated for 10 min on ice. The suspension was centrifuged at 3,000 $\times g$ for 4 min at 4°C. The insoluble nuclear fraction was rinsed once with 100 μ l of nuclei lysis buffer and was centrifuged at 3,000 $\times g$ for 5 min at 4°C. The pellets were treated with 50 μ l (50 units) of DNase I (Thermo Fisher Scientific, Cat#AM2224), 1 $\times$ DNase I Buffer (Thermo Fisher Scientific, Cat#AM2224), and SUPERase•IN RNase Inhibitor (20 units) for 30 min at 37°C with pipetting up and down several times. The lysate was mixed with 700 μ l of QIAzol and was used for nascent RNA extraction. Total and nascent RNA were extracted with an miRNeasy Mini kit (Qiagen, Cat#217004). On-column DNase I digestion was carried out with an RNase-free DNase set (Qiagen, Cat#79254).

CAGE and NET-CAGE library preparation and sequencing for bulk CD4⁺ T cells

Using total or nascent RNA (70–450 ng) obtained from bulk CD4⁺ T cells, CAGE and NET-CAGE libraries were generated as previously described (95) with the following modifications for low-input analysis. After cDNA library construction, cDNAs derived from rRNAs were removed with ZapR (SMARTer Stranded Total RNA-Seq Kit v2 - Pico Input Mammalian, Takara Bio, Cat#634413). The samples were cleaned up with 1.2 $\times$ AMPure XP Beads (Beckman Coulter, Cat#A63881) and eluted with 52 μ l of H₂O. Amplification was performed in a reaction mixture (50 μ l) containing 22 μ l of the library, 1 μ l of Tks Gflex DNA polymerase (1.25 units, Takara Bio, Cat#R060A), 25 μ l of 2 $\times$ Gflex PCR Buffer (Mg²⁺/dNTR plus), 1 μ l of forward PCR primer (10 μ M), and 1 μ l of reverse PCR index primer (10 μ M). PCR conditions were as follows: 94°C for 1 min, 8 or 9 cycles of (98°C for 10 s, 55°C for 30 s), and 68°C for 2 min. The sample was cleaned up with 1.8 $\times$ and then with 0.9 $\times$ AMPure XP Beads. CAGE and NET-CAGE libraries were sequenced on an Illumina NextSeq 500 platform using single-end, single-index sequencing with 75 cycles for read 1 and 8 cycles for i7 index.

Luciferase assay

Luciferase reporter assays were performed as previously described (96). DNA fragments (100 bp) containing reference or alternative alleles of single nucleotide polymorphisms (SNPs) rs7441808 and rs117605225 were inserted into the multiple cloning sites of the pGL4.24 vector plasmid (Promega, Cat#E8421) containing the *luc2P* gene under the control of a minimal promoter. The pGL4.74 vector (Promega, Cat#6921) containing *hRluc* gene was used as an internal control to normalize for transfection efficiency. The plasmids were transfected into Jurkat E6-1 cells (ATCC) using P3000 reagent (2 μ l/ μ g DNA) and Lipofectamine 3000 reagent (1.5 μ l/ μ g DNA; Invitrogen, Cat#L3000001). Luciferase activity was measured 24 h after transfection with a Dual-Luciferase Reporter Assay System (Promega, Cat#E1910) using a Tristar 3 plate reader (Berthold Technologies). The list shows oligos used to construct plasmids.

SNP (Alt./Ref.)	Name of oligo	Sequence (5'→3')
rs7441808 (G/A)	rs7441808_A	CTAGGGCAATTAGGCATTGAAATTTGACTATTTACCTTAGGGACTCCTCTGGGTATGT GAAGAATTCCCCTGTTTTGCTCAGGATGTGGGCAGATGAGGAAGTA
	rs7441808_A_complementary	AGCTTACTTCCTCATCTGCCACATCCTGAGCAAAACAGGGGAATTCTTCACATACC CAGAGGAGTCCCTAAGGTAAATAGTCAAATTTCAATGCCTAATTGCC
	rs7441808_G	CTAGGGCAATTAGGCATTGAAATTTGACTATTTACCTTAGGGACTCCTCTGGGTGTGT GAAGAATTCCCCTGTTTTGCTCAGGATGTGGGCAGATGAGGAAGTA
	rs7441808_G_complementary	AGCTTACTTCCTCATCTGCCACATCCTGAGCAAAACAGGGGAATTCTTCACACACC CAGAGGAGTCCCTAAGGTAAATAGTCAAATTTCAATGCCTAATTGCC
rs117605225 (G/T)	rs117605225_T	CTAGTGACAGATGGTGTGAAACCACACGGTAGCTTGGCAGTATGGGGCGGCTGCTGC TTAGTCACTCCTAACTCAGGAACCTAGTATTGAAACTTGAACAGCCT
	rs117605225_T_complementary	AGCTAGGCTGTTCAAGTTTCAATACTAGGTTCTGAGTTAGGAGTGACTAAGCAGCA GCCGCCCCATACTGCCAAGCTACCGTGTGGTTTCACACCATCTGTCA
	rs117605225_G	CTAGTGACAGATGGTGTGAAACCACACGGTAGCTTGGCAGTATGGGGCGGCTGCGG CTTAGTCACTCCTAACTCAGGAACCTAGTATTGAAACTTGAACAGCCT
	rs117605225_G_complementary	AGCTAGGCTGTTCAAGTTTCAATACTAGGTTCTGAGTTAGGAGTGACTAAGCCGCA GCCGCCCCATACTGCCAAGCTACCGTGTGGTTTCACACCATCTGTCA

Quality control and primary processing of 5' and 3' scRNA-seq data

Cell Ranger Software version 5.0.1 (10x Genomics) was used to process 5' and 3' Chromium scRNA-seq data (85). Reads were aligned to the Cell Ranger GRCh38 reference genome (refdata-gex-GRCh38-2020-A, 10x Genomics), and raw gene expression matrices were generated using the “cellranger count” function without the “--include-introns” option. The 5' scRNA-seq data were analyzed using R package Seurat version 5 (4.9.9.9067) (86) as follows. Genes expressed in three or more cells were retained. Doublets were predicted and filtered using R package scDblFinder version 1.14.0 (97). Singlets were retained for downstream analysis. Cells expressing fewer than 200 genes and those with more than 10% of all transcripts derived from the mitochondrial genome were excluded.

Initial processing of 5' snRNA-seq data

Reads were aligned to the Cell Ranger GRCh38 reference genome (refdata-gex-GRCh38-2020-A, 10x Genomics), and raw gene expression matrices were generated using the “cellranger count” function with the “--include-introns” option in Cell Ranger version 5.0.1. Ambient RNA prediction and filtering were performed using R package SoupX version 1.6.2 (98) with an automatic estimation of the contamination rate implemented in “autoEstCout”. When the automatic estimation failed, the contamination rate was estimated from the results of replicate or identical donor samples. The contamination rate in each analysis is listed in **Table S1**. Other

primary processes were performed as in the ‘Quality control and primary processing of 5’ and 3’ scRNA-seq data’ subsection.

Initial processing of Multiome data

Raw sequencing data were converted to FASTQ files using the “cellranger-arc mkfast” function in Cell Ranger ARC version 2.0.0 (10x Genomics) and bcl2fastq v2.19.0.316 (Illumina) for snATAC-seq and for 3’ snRNA-seq, respectively. Reads were aligned to the Cell Ranger GRCh38 reference genome (refdata-cellranger-arc-GRCh38-2020-A-2.0.0, 10x Genomics) and quantified using the “cellranger-arc count” function following the 10x Genomics guidelines. Ambient RNA was removed using SoupX version 1.6.2. Seurat version 5 (4.9.9.9067) (86) and Signac version 1.10.0 (87) were used to perform quality control on the data according to the tutorial (https://stuartlab.org/signac/articles/pbmc_multiomic). Doublets, apoptotic cells, and low-count cells were identified and excluded from the downstream analysis. Doublets were predicted and filtered using scDblFinder version 1.14.0 (97). Apoptotic cells were identified as those having $\geq 20\%$ of mitochondrial gene transcripts. Low-count cells were identified as those having gene counts ≤ 200 or ATAC read counts $\leq 1,000$. Cells with low-quality ATAC reads—defined as those with a nucleosome signal ≥ 2 or transcription start site (TSS) enrichment ≤ 1 —were also excluded.

Integration of 5’ scRNA-seq, snRNA-seq, and Multiome data across donors

The filtered gene-barcode matrix for each sample was aligned into an integrated and unbatched dataset across all donors using Seurat version 5 (4.9.9.9067) (86). In brief, data were normalized using the “NormalizeData” function: the UMI count for each gene from each cell was divided by the total UMI counts for the cell, multiplied by 10,000 (a scale factor) and log-transformed. Highly variable genes were identified using the “FindVariableGenes” function with default parameters. A linear transformation was applied as a standard pre-processing step using the “ScaleData” function. Principal component analysis for dimensionality reduction was performed using the “RunPCA” function. The scRNA-seq, snRNA-seq, and Multiome 3’ snRNA-seq data were integrated using canonical correlation analysis with the “IntegrateLayers” function. Finally, cells were clustered using the “FindNeighbors” and “FindClusters” functions. Uniform manifold approximation and projection (UMAP) visualizations were generated using the “RunUMAP” function. Differential expression was analyzed using the “FindAllMarkers” function, and statistical significance was estimated using the Wilcoxon rank sum test. Differentially expressed signature genes were defined as those expressed in at least 25% of the cells within one cluster at a minimum log fold change of 0.25. Signature genes for each cluster are listed in **Table S3**.

snATAC peak calling

snATAC peaks were identified from fragment files of each cluster identified as described in the ‘Integration of 5’ scRNA-seq, snRNA-seq, and Multiome data across donors’ subsection using MACS2 version 2.2.6 with default settings as implemented in the “CallPeaks” function in Signac version 1.10.0 (87).

Processing of publicly available 5’ scRNA-seq data

We retrieved 5’ scRNA-seq data for human memory CD4⁺ T cells stimulated with viral antigens (65) from the Gene Expression Omnibus database (accession number GSE152522). For

each FASTQ file, quality control and processing were performed as described above. Among human memory CD4⁺ T cells (65), we analyzed activation marker (CD38, HLA-DR, PD-1)-positive cells sorted *ex vivo* (0 h), CD154⁺CD69⁺ cells sorted following 6 h stimulation with virus-specific peptide megapools (SARS-CoV-2, coronavirus, human parainfluenza, human metapneumovirus, or influenza), and CD137⁺CD69⁺ cells sorted following 24 h stimulation with SARS-CoV-2-specific peptide megapools. 5' scRNA-seq data at each time point (0, 6, and 24 h) were integrated using Harmony version 1.0.3 (99) in the “IntegrateLayers” function in Seurat version 5 (4.9.9.9067) (86).

Processing 5' CITE-seq data

Reads were aligned to the Cell Ranger GRCh38 reference genome (refdata-gex-GRCh38-2020-A, 10x Genomics), and raw gene expression matrices and protein expression matrices were generated using the “cellranger count” function without the “--include-introns” option in Cell Ranger version 5.0.1. Seurat objects were generated according to the Seurat tutorial (https://satijalab.org/seurat/archive/v3.1/multimodal_vignette.html). Protein levels in single cells were normalized using the centered-log ratio transform implemented in Seurat. Data were integrated using corresponding 5' scRNA-seq data as described in the ‘Integration of 5' scRNA-seq, snRNA-seq, and Multiome data across donors’ subsection.

Read-level pre-filtering and transcribed enhancer call (ReapTEC)

The ReapTEC method can be applied to 5' RNA sequencing data. ReapTEC for 5' scRNA-seq and 5' snRNA-seq data consists of the following 13 steps.

1. A count matrix was generated using Cell Ranger version 5.0.1 as described above.
2. A list of filtered cell barcodes was created.

```
$ zcat filtered_feature_bc_matrix/barcodes.tsv.gz | sed -e 's/-1//g' > Sample1_whitelist.txt
```
3. The FASTQ files were mapped using STARsolo implemented in STAR version 2.7.10b (88) (hg38; GENCODE version 39 [‘primary assembly’] (100)).

```
$ STAR --runThreadN 12 --genomeDir index/ --readFilesIn Sample1_S1_R1_001.fastq.gz  
Sample1_S1_R2_001.fastq.gz --soloCBwhitelist Sample1_whitelist.txt --soloBarcodeMate 1  
--clip5pNbases 39 0 --readFilesCommand zcat --soloType CB_UMI_Simple --soloCBstart 1  
--soloCBlen 16 --soloUMIstart 17 --soloUMIlen 10 --soloStrand Reverse --  
outFileNamePrefix Sample1_ --outSAMtype BAM SortedByCoordinate --  
soloCBmatchWLtype 1MM_multi_Nbase_pseudocounts --soloUMIdedup  
1MM_Directional_UMItools \
```

```
--outSAMattributes NH HI nM AS CR UR CB UB GX GN sS sQ sM
```
4. Only read 1 reads were extracted from uniquely mapped reads using SAMtools version 1.16.1 (101).

```
$ samtools view -@ 12 -hbf 64 -q 255 Sample1_Aligned.sortedByCoord.out.bam >  
Sample1_unique_R1.bam
```
5. Read 1 5' scRNA-seq reads with an unencoded G (which is tagged as a soft-clipped G by STARsolo (88)) were extracted from BAM files using SAMtools version 1.16.1 (101) as follows.

```
$ samtools view -@ 12 -H Sample1_unique_R1.bam > Sample1_header.sam  
$ samtools view -@ 12 -F 16 Sample1_unique_R1.bam | awk -F '\t' 'BEGIN {OFS="\t"}  
{BASE = substr($10, 40, 1); if ($6 ~ /^40S[0-9]/ && BASE == "G") {print $0}}' >  
Sample1_SoftclipG_F.sam
```

- ```
$ samtools view -@ 12 -f 16 Sample1_unique_R1.bam | awk -F '\t' 'BEGIN {OFS="\t"}
{ ALT = substr($10, length($10)-39, 1); if ($6 ~ /[0-9]M40SS/ && ALT == "C") {print
$0}}' > Sample1_SoftclipG_R.sam
$ cat Sample1_header.sam Sample1_SoftclipG_F.sam Sample1_SoftclipG_R.sam | samtools
sort -@ 12 -O bam -o SoftclipG_Sample1.bam
```
6. Duplicated reads were removed using UMI-tools version 1.1.2 (102).

```
$ umi_tools dedup --per-cell -I SoftclipG_Sample1.bam --extract-umi-method=tag --umi-
tag=UR --cell-tag=CR -S SoftclipG_Sample1_deduplicated.bam
```
  7. Reads with cell barcodes corresponding to the list created in step 2 were extracted.

```
$ awk '{print "CB:Z:"$1}' Sample1_whitelist.txt > Sample1_cell_barcode.txt
$ samtools view -@ 12 -H SoftclipG_Sample1_deduplicated.bam > SAM_header
$ samtools view -@ 12 SoftclipG_Sample1_deduplicated.bam | LC_ALL=C grep -F -f
Sample1_cell_barcode.txt > filtered_SAM_body
$ cat SAM_header filtered_SAM_body > filtered.sam
$ samtools view -@ 12 -b filtered.sam > SoftclipG_Sample1_filtered.bam
```
  8. A count file was generated for each TSS using the “bamToBed” function in BEDTools
version 2.30.0 (89).

```
$ samtools view -@ 12 SoftclipG_Sample1_filtered.bam | awk 'BEGIN{OFS="\t"} {print
$23}' > Sample1_cell_barcode_tmp.txt
$ bamToBed -i SoftclipG_Sample1_filtered.bam | paste - Sample1_cell_barcode_tmp.txt |
awk 'BEGIN {OFS="\t"} { if($6=="+") {print $1, $2, $2+1, ".", $7, $6} else {print $1, $3-1,
$3, ".", $7, $6}}' | sort -k1,1 -k2,2n -k6,6 | bedtools groupby -g 1,2,3,4,5,6 -c 1 -o count | awk
'BEGIN{OFS="\t"} {if ($1 ~ /chr/) print $1, $2, $3, $5, $7, $6}' >
Sample1.CTSS.fwd.rev.cell.barcode.bed
```
  9. Bidirectionally transcribed candidate enhancers (btcEnhs) were identified.

TSS peaks were generated by merging TSSs located within 10 bp of each other. To identify btcEnhs, TSS peak pairs were detected using scripts provided at [https://github.com/anderssonrobin/enhancers/blob/master/scripts/bidir\\_enhancers](https://github.com/anderssonrobin/enhancers/blob/master/scripts/bidir_enhancers) (11) with minor modifications. The definition of bidirectionally transcribed regions was based on TSS peaks supported by at least one read. The directionality score was calculated as previously described (11) and was set to a default of 0.8. The TSS regions ( $\pm 300$  bp) of protein-coding transcripts (GRCh38; GENCODE version 39 [‘primary assembly’] (100)) were masked for negative filtering of bidirectionally transcribed regions. btcEnhs located in exonic regions and not overlapping with ENCODE CD4<sup>+</sup> T cell DNase I hypersensitive sites (DHSs) listed in **Table S16** were excluded.
  10. (Optional) To identify btcEnhs from clusters defined by Seurat or other methods, the cell barcodes for each cluster were first obtained by using R.

```
> Cluster0_object <- subset(Seurat_object, idsents = 0)
> Cluster0_cell_barcode <- colnames(Cluster0_object@assays$RNA@counts)
> write.table(Cluster0_cell_barcode, "Cluster0_cell_barcode.txt", sep="\t", col.names=F,
row.names=F, quote=F)
```
  11. (Optional) A count file for each TSS in each cluster was generated, and step 9 was repeated.

```
$ grep Sample1.CTSS.fwd.rev.cell.barcode.bed -f Cluster0_cell_barcode.txt >
Sample1_cluster0.CTSS.fwd.rev.cell.barcode.bed
```
  12. (Optional) Coverage at single-base-pair resolution was calculated with the “genomecov” function in BEDTools version 2.30.0 (89) using the count file for each TSS generated in step

8. The resultant forward and reverse bedGraph files were converted into bigwig files using the UCSC software bedGraphToBigWig (103).

Promoters were defined as TSSs of protein-coding transcripts  $\pm 300$  bp (hg38; GENCODE version 39 [‘primary assembly’] (100)). Reads mapped to the promoters, btcEnhs, and ATAC peaks were counted in a strand-specific manner using the UCSC software bigWigAverageOverBed (103). Count normalization was based on counts per million (CPM). When count was converted to  $\log_2$ CPM, a prior count of 0.25 was added to the raw counts. Directionality of TSS at ATAC peaks was defined as follows:  $D = (F - R)/(F + R)$ ; F, forward strand transcript count; R, reverse strand transcript count;  $|D| < 0.8$  was defined as an indicator of bidirectional transcription.

13. (Optional) TSS peak-based single-cell analysis. Using the count file created in step 8, the 5' ends of transcripts were counted at the single-cell level for each robust TSS peak generated in step 9 (with a cutoff of  $\log_2$ CPM  $\geq 2$  in at least one cluster) and a TSS peak-based count matrix was created. Robust TSS peaks were named according to the nearest known transcript and numbered in ascending order from upstream to downstream of the transcript. TSS peak-based count matrices were used as input files for Seurat and were pre-processed as in ‘Quality control and primary processing of 5' and 3' scRNA-seq data’ and ‘Primary processing of 5' snRNA-seq data’ sections. TSS peaks of T cell receptor-related transcripts were excluded in this analysis. Data integration and batch collection were performed using Harmony version 1.0.3 (99) in the “IntegrateLayers” function implemented in Seurat version 5 (4.9.9.9067) (86).

The ReapTEC script to identify btcEnhs from 5' RNA sequencing data is available at <https://github.com/MurakawaLab/ReapTEC>.

### RNA velocity

Cell Ranger BAM files were processed with Velocity version 0.17.17 (59) using the Cell Ranger GRCh38 reference genome (refdata-gex-GRCh38-2020-A, 10x Genomics). Output files were loaded into scVelo version 0.2.5 (<https://scvelo.readthedocs.io>) (60) in the Python 3.9.13 environment by using the default parameters and merged with the gene-expression anndata object from Seurat.

### 5' MAS-seq data processing

High-Fidelity (HiFi) reads were generated using CCS software version 6.3.0 (<https://ccs.how>) with default parameters. HiFi BAM files were processed using SMRT Link version 12 according to the guide ([https://www.pacb.com/wp-content/uploads/SMRT\\_Link\\_MAS-Seq\\_troubleshooting\\_v12.0.pdf](https://www.pacb.com/wp-content/uploads/SMRT_Link_MAS-Seq_troubleshooting_v12.0.pdf)) to adapt to the 10x Genomics 5' kit. We had the UMI+TSO trimmed together as a 23-bp component with the 16B-23U-T design. Reads with unencoded G at the 5' ends were extracted from the scisoseq.mapped.bam file. Transcripts were assembled *de novo* using filtered scisoseq.mapped.bam files of circulating bulk CD4<sup>+</sup> T cells and CD4<sup>+</sup>CD25<sup>+</sup> T cells. Stringtie2 version 2.2.1 (104) was used to assemble and merge transcripts for individual samples from each dataset with the --fr, -G, and -L options. Transcripts not overlapping with the GENCODE gene models (hg38; GENCODE version 39 [‘primary assembly’] (100)) were defined as *de novo* transcripts.

### CAGE and NET-CAGE data analysis

Adapters were trimmed with Trimmomatic version 0.39 (105). Reads were aligned to human rDNA complete repeating unit (U13369.1) using STAR version 2.7.10b (88). Reads that did not align to human rDNA were mapped to the human reference genome (hg38; GENCODE version 39 [‘primary assembly’] (100)). Uniquely mapped read 1 reads with an unencoded G were extracted from BAM files using SAMtools version 1.16.1 (101). A count file for each TSS was generated from the filtered BAM file. Coverage at single-base-pair resolution was calculated with the “genomecov” function in BEDTools version 2.30.0 (89). The resultant forward and reverse bedGraph files were converted into bigwig files using the UCSC software bedGraphToBigWig (103). Promoters were defined as TSSs of protein-coding transcripts  $\pm 300$  bp (hg38; GENCODE version 39 [‘primary assembly’] (100)), as in the mask file used in the enhancer analysis. Reads mapped to the promoters and btcEnhs identified in this study were counted in a strand-specific manner using the UCSC software bigWigAverageOverBed (103). Count normalization was based on CPM. When count was converted to  $\log_2$ CPM, a prior count of 0.25 was added to the raw counts. Differential expression was analyzed using the “exactTest” function in R package edgeR version 3.42.4 (106). Differential promoters were analyzed with the “filterByExpr” function in edgeR.

### Cell-type specificity score

The cell-type specificity score of promoters and btcEnhs was calculated from the transcript level (CPM) using entropy as previously described (10). Promoters and btcEnh with less than 10 mapped reads were excluded. The cell-type specificity score of ATAC peaks was calculated from chromatin accessibility (fragments per million, FPM). FPM was calculated by counting the number of fragments overlapping with each ATAC peak across the 136 CD4<sup>+</sup> T cell clusters and normalized to the total number of fragments in each cluster.

$$\text{Specificity score}(x) = 1 - \frac{\text{entropy}(x)}{\log_2[n]}$$

$$\text{Entropy}(x) = - \sum_{i=1}^n p_i * \log_2[p_i]; p_i = \frac{x_i}{\sum_{i=1}^n x_i}$$

$x_i$  represents the transcript level (CPM) at the promoter or btcEnh or represents the chromatin accessibility level (FPM) at the ATAC peaks in the CD4<sup>+</sup> T cell cluster  $i$ , and  $n$  represents the number of CD4<sup>+</sup> T cell clusters. The specificity score was defined to range between 0 and 1, where 0 means non-specific and 1 means cell type-specific.

### Metagene analysis of the chromatin state of btcEnhs

Publicly available data for DHS (accession number ENCFF121ZZH), H3K27ac (ENCFF018YIU), H3K4me1 (ENCFF860QIV), and H3K4me3 (ENCFF610SPR) histone modifications were retrieved from the ENCODE database ([https://www.encodeproject.org/search/?type=Experiment&control\\_type!=\\*&status=released&perturbed=false](https://www.encodeproject.org/search/?type=Experiment&control_type!=*&status=released&perturbed=false)) (2, 107). Bigwig files containing read-depth normalized signals and signal  $P$  values were used for the analysis of DHS and histone modifications, respectively. Metagene analysis was performed using deepTools version 3.5.2 (108). DHS and histone modification values per btcEnh region were aggregated using the “computeMatrix” function: computeMatrix reference-point --referencePoint center -S <bigwig file> -R <bed file> -a 2000 -b 2000 -o <output file> -p 8 -bs 10. Heatmaps of aggregated values were created using the “plotHeatmap” function: plotHeatmap -m <matrix file from the computeMatrix tool> -out <output file> --

colorMap <color> --whatToShow 'heatmap and colorbar' --dpi 100 --sortRegions keep --plotFileFormat pdf.

### **Mapping, pairing Micro-C contacts, and building contact maps**

Micro-C raw data were processed with the dovetail\_tools pipeline (<https://micro-c.readthedocs.io/en/latest/>), which was provided for the analysis of Micro-C libraries using the Dovetail Micro-C Kit. In brief, reads were mapped to the human reference assembly hg38 using BWA-MEM with flags -5SP (<http://bio-bwa.sourceforge.net/>) (109). Alignments were parsed and pairs were classified using the pairtools package (<https://github.com/mirnylab/pairtools>) to generate 4DN-compliant pairs files. Pairs with multiple hits, mapping quality score (MAPQ)  $\leq 30$ , singleton, dangling end, or self-circle, and PCR duplicates were removed. Pairs classified as uniquely mapped or rescued chimeras with MAPQ  $> 30$  on both sides were aggregated into contact matrices in the cooler format using the cooler package (<https://github.com/open2c/cooler>) at 1 kb. Output files containing all valid pairs were used for downstream analyses such as loop calling. 4DN-compliant pairs files were also converted to HIC files using the Juicer Tools package (<https://github.com/aidenlab/juicer>) (90). Contact matrices were balanced by using SCALE normalization in HIC files. Contact maps were visualized along with genomic annotations using CoolBox version 0.3.7 (110) ([https://gangcaolab.github.io/CoolBox/quick\\_start\\_API.html](https://gangcaolab.github.io/CoolBox/quick_start_API.html)).

### **Detection of chromatin loop contacts**

Chromatin loop contacts in this study were identified by the HiCCUPS algorithm (90) using the Juicer Tools package version 2.20.0 (<https://github.com/aidenlab/juicer>) and the scale-space representation algorithm (91) using the Mustache package (<https://github.com/ay-lab/mustache>). Loops were called at a 1-kb resolution with SCALE-normalized contact matrices for HiCCUPS and with ICE-normalized contact matrices for Mustache, and were filtered for an FDR  $< 0.05$ . Micro-C loops were merged using the “mergebedpe” function in R package hictoolsr version 1.1.2 (<https://github.com/EricSDavis/hictoolsr>). To classify anchors as promoters, btcEnhs, or ATAC peaks, anchor locations were compared with the corresponding chromatin features as follows. Promoters were defined as TSSs  $\pm 800$  bp of protein-coding transcripts (hg38; GENCODE version 39 [‘primary assembly’] (86)), which were longer regions than those in the mask file used in the enhancer analysis. btcEnhs and ATAC peaks were also defined as regions extended by  $\pm 500$  bp from the original regions used for anchor classification. Enrichment was calculated by comparison to shuffled loops.

### **Chromatin interaction analysis for region capture Micro-C (RCMC)**

Interaction calling and significance thresholding for RCMC was based on the workflow developed by Dovetail Genomics (<https://dovetail-capture.readthedocs.io/en/latest/interactions.html>) and the CHiCAGO tool (111). The input files for the CHiCAGO tool were the pairs files, which were produced according to the same method as for the Micro-C contact map. Replicates were merged. CHiCAGO requires two sets of genomic intervals as inputs: Bait Fragments (BF) - fragments overlapping with baits, and Other End Fragments (OEF) - all other genomic fragments to be included in the analysis. The BF was provided by Dovetail Genomics ([https://s3.amazonaws.com/dovetail.pub/capture/human/h\\_baits\\_v1.0.bed](https://s3.amazonaws.com/dovetail.pub/capture/human/h_baits_v1.0.bed)), and the OEF was generated by binning the genome with a window size of 1 kb outside the bait regions. Significant



interactions were called using CHiCAGO (111) with default parameters at 1-kb resolution and were defined as those with a CHiCAGO score of  $\geq 5$  per condition. RCMC interactions were merged using makePeakMatrix.R implemented in the CHiCAGO tool.

### **Activity-by-contact (ABC) models**

We calculated the ABC scores (77, 78) using the pipeline provided by the authors (version 0.2.2, <https://github.com/broadinstitute/ABC-Enhancer-Gene-Prediction>) to predict enhancer–gene connections in resting and activated CD4<sup>+</sup> T cells on the basis of chromatin accessibility (original pseudo-bulk snATAC-seq data from single-cell experiments), histone modifications (public H3K27ac ChIP–seq, ENCODE ENCFF484WXU for resting CD4<sup>+</sup> T cells and ENCFF878VZI and ENCFF056OHJ for activated CD4<sup>+</sup> T cells), gene expression profile (original pseudo-bulk RNA-seq data from single-cell experiments), and chromatin contact map (original Micro-C data). We conducted ABC modeling with default parameters provided by the authors (77). First, we called peaks in the chromatin accessibility dataset using MACS2 with a  $P$  value cut-off of 0.1. Second, we counted chromatin accessibility reads in each peak and retained the top 150,000 peaks with the highest read counts for the genome-wide peak calls, and we named the resulting set of regions candidate elements. Third, we calculated candidate element–promoter contacts using a Micro-C contact map with 1-kb resolution and SCALE normalization. Finally, we computed the ABC score for each candidate element–gene pair.

### **Co-variation of btcEnhs and promoters**

Spearman correlation tests were performed between the log<sub>2</sub>CPM expression of eRNA and mRNA in btcEnh–promoter pairs across 136 CD4<sup>+</sup> T cell clusters. btcEnhs with a cutoff of log<sub>2</sub>CPM  $\geq 0$  in at least one cluster and promoters with a cutoff of log<sub>2</sub>CPM  $\geq 2$  in at least one cluster were used. Promoters were defined as protein-coding gene transcript TSSs  $\pm 300$  bp and were the same as in the mask file used in the enhancer analysis. All btcEnh–promoter pairs within 1 Mb from the midpoint of each btcEnhs were calculated. The  $P$  values were corrected using the Holm’s method and the adjusted  $P$  value cut-off ( $< 0.05$ ) was set to define significantly correlated btcEnh–promoter pairs.

### **Prediction of CD4<sup>+</sup> T cell btcEnhs and chromatin accessibilities based on DNA sequence patterns**

We trained MENTR machine learning models (67) to assess the dependence of CD4<sup>+</sup> T cell btcEnhs and accessibilities on DNA sequences. Briefly, the machine learning models predict the enhancer on-off or chromatin accessibility patterns from the  $\pm 100$  kb DNA sequences (GRCh38). The models leveraged 2,002 epigenetic features predicted from DNA sequences by the pre-trained deep learning models (112). The prediction tasks are listed in **Table S11**. We trained these models using all autosomes except chromosome 8 (20% was used for early stopping) and tested the accuracy using chromosome 8. The detailed procedures and parameters are described in our previous paper (67). After training, we predicted the effects of mutations in the variants on overlapping *cis*-regulatory elements. On the basis of the previous evaluation (67), we set a permissive threshold (absolute *in silico* mutation effects of 0.05) achieving 80% concordance, and a robust threshold (absolute *in silico* mutation effects of 0.1) achieving 90% concordance.

URLs:

[https://www.encodeproject.org/files/GRCh38\\_no\\_alt\\_analysis\\_set\\_GCA\\_000001405.15/@download/GRCh38\\_no\\_alt\\_analysis\\_set\\_GCA\\_000001405.15.fasta.gz](https://www.encodeproject.org/files/GRCh38_no_alt_analysis_set_GCA_000001405.15/@download/GRCh38_no_alt_analysis_set_GCA_000001405.15.fasta.gz) (fasta file for MENTR training) (67)

[https://github.com/FunctionLab/ExPecto/blob/447737793d8d21e50e82379feca44fb9465fdc79/download\\_resources.sh](https://github.com/FunctionLab/ExPecto/blob/447737793d8d21e50e82379feca44fb9465fdc79/download_resources.sh) (pre-trained deep learning model to predict 2,002 epigenetic features) (112)

### Integration of CD4<sup>+</sup> T cell btcEnhs with human genetics

The links between human genetics findings and information on CD4<sup>+</sup> T cell btcEnhs were assessed on the basis of (i) significant variants from genome-wide association studies (GWASs) and (ii) non-significant polygenic signals (113). In approach (i), the enrichment of variants associated with each disease in the btcEnh regions identified here was evaluated using GREGOR (92). The lead SNPs with the highest association significance were obtained, and then SNPs in strong linkage disequilibrium (LD) ( $R^2 > 0.8$ ) were added using the 1000 Genomes CEU reference population (114). The European LD structure was used because most of the GWAS data on immune-mediated diseases and COVID-19 were generated in European populations. Allele frequencies of variants and distance between variants and coding regions were considered to compute the target and background overlaps of variants in the given btcEnh regions. In approach (ii), LD score regression was performed by using ldsc software (v1.0) (75) to evaluate the enrichment of heritability of immune-mediated diseases in btcEnhs. The degree of polygenic signal enrichment was computed by conditioning on the European full baseline model described in Finucane *et al.* (75), which contains coding, intronic, enhancer (defined differently here), and untranslated regions, promoters, histone marks, open chromatin regions. Summary statistics (62, 75, 115–119) are listed in **Table S17**.

### ChIP-seq enrichment and motif analysis

ChIP-seq enrichment analyses were performed using ChIP-Atlas ([https://chip-atlas.org/enrichment\\_analysis](https://chip-atlas.org/enrichment_analysis)) (120, 121). The regions obtained by extending  $\pm 500$  bp from the midpoint of each dATAC peak were used as inputs to ChIP-Atlas. The data used to generate the bar graphs for ChIP-seq enrichment of BRD4, CEBPB, CREBBP, ETS1, and H3K27ac in Figure 2H are SRX6913099, SRX10212070, SRX017698, SRX7894794, and SRX1127401, respectively. Motif scanning analysis for btcEnhs harboring SNPs associated with immune-mediated diseases was performed using FIMO version 5.5.4 (122). The motifs (*Homo sapiens*) were obtained from JASPAR 2022 (123). Motif enrichment analysis for btcEnhs and ATAC peaks was performed using AME implemented in MEME Suite version 5.5.4 (124). These regions were compared to shuffled input sequences.

```
$ ame --verbose 1 --oc . --scoring avg --method fisher --hit-lo-fraction 0.25 --evaluate-report-threshold 949.0 --control --shuffle-- --kmer 2 <sequence file> <motif file>
```

### Conservation analysis

The conservation of btcEnhs and ATAC peaks was estimated from a multiple genome alignment, “Multiz30way” (<https://hgdownload.soe.ucsc.edu/goldenPath/hg38/multiz30way/maf/>). Multiz30 contains genomes of 27 primates, mouse, dog, and armadillo. To use only the genomes with syntenic alignment, we excluded the genomes of 6 primates (proboscis monkey, golden snub, black snub, tarsier, black lemur, and Sclater’s lemur) represented by “reciprocal best” alignments. To

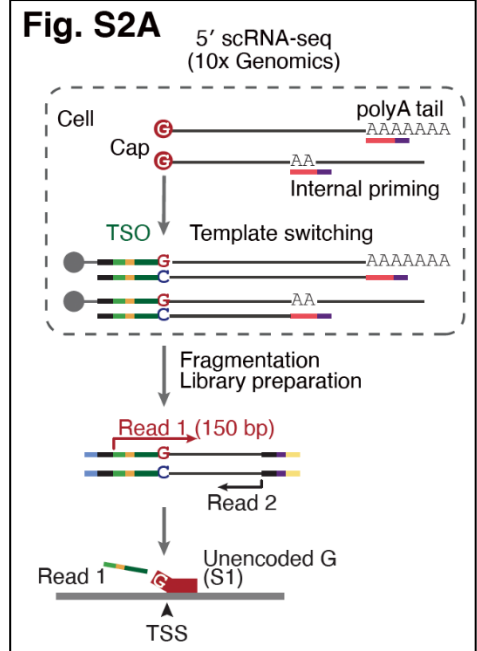
evaluate sequence conservation in the genomes of 24 species, the lengths of the regions aligned to each human btcEnh or ATAC peak were calculated. A species was postulated to carry an enhancer orthologous to a human btcEnh or ATAC peak if the aligned length of the non-human sequence was more than 50% of the length of the human sequence. PhastCons conserved elements (*125*) were downloaded from the UCSC Genome Browser (<http://hgdownload.cse.ucsc.edu/goldenPath/hg38/phastCons100way/>). The evolutionary conservation analysis also included randomly selected protein-coding sequences (CDSs) and genomic regions (62,803 regions of each type), both with the same length distribution as the btcEnhs. Using BEDTools shuffle (*89*), CDSs were selected from among those annotated with GENCODE version 39 (*100*), and random genomic regions were selected from the whole genomes. Promoter regions masked in the enhancer analysis were excluded from both selections; telomeric and centromeric regions (UCSC Table Browser) were also excluded from the selection of random genomic regions.

### **Intersection between enhancers and transposable genetic elements**

To investigate the intersection between btcEnhs (or ATAC peaks) and transposable genetic elements (transposons), the human genome assembly, GRCh38, was masked using RepeatMasker 4.1.1 (*126*) with the '-s -no\_is' option. As an input, we used the repeat library provided in Dfam (version 3.2) (*127*). A btcEnh (or ATAC peak) was considered to intersect with an annotated repeat if more than 50% of its length overlapped with the repeat.

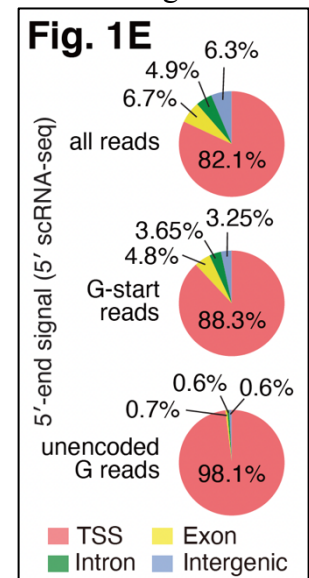
## Supplementary Text 1

To analyze the 5' ends of RNAs, various approaches have been developed (128) on the basis of the cap trapper (129, 130), oligo-capping (131) and template switching (132). For example, cap analysis of gene expression (CAGE) employs the cap trapper; Adiconis *et al.* (50) showed that this is the best method for identifying the 5' ends of mRNAs. Native elongating transcript-CAGE (NET-CAGE) (10) (relies on cap trapper) and PRO-cap (9) (relies on oligo-capping) capture the 5' ends of nascent RNAs and thus ensure sensitive detection of unstable transcripts including enhancer RNAs (eRNAs). Low-input CAGE (nanoCAGE) relies on the template switching method (133). C1 CAGE is a single-cell technology to detect transcript 5' ends by using the C1<sup>TM</sup> microfluidic system (Fluidigm), although with limited throughput (134); the development of more scalable methods is warranted to dissect many cell types and states under diverse conditions. Droplet barcoding of single cells enables single-cell transcriptomics of a large number of cells in a high-throughput manner (135, 136). 10x Genomics Chromium is one of the most widely used single-cell expression profiling platforms that use droplet-based single-cell RNA-sequencing (scRNA-seq).

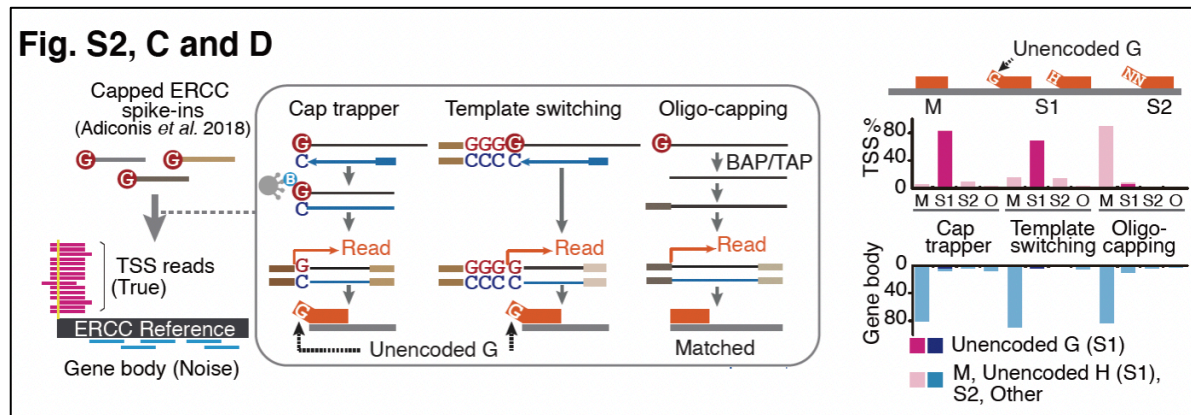


The 10x 5' GEM scRNA-seq (10x Genomics) used in this study employs the template switching method (**fig. S2A**). Any 5'-end RNA sequencing data contain background noise reads that do not originate from the actual transcription start sites (TSSs), and removal of such noise signals is necessary to detect, with high accuracy, eRNAs expressed at very low levels (see **Main Text**).

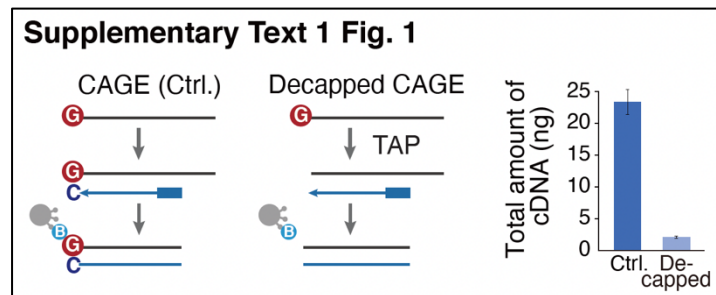
We found that 82.1% of the 5' ends of read 1 in 5' scRNA-seq were mapped to TSS regions (**Fig. 1E**). This indicates the presence of background signals in the 5' scRNA-seq data due to premature template switching and strand invasion (49, 50). To overcome this issue, we took advantage of an unencoded guanosine (G) derived from a 5' cap added to the beginning of RNA transcripts post-transcriptionally. Generally, eukaryotic RNAs transcribed by RNA polymerase II possess a 5'-end N7 methylguanosine (m7G) cap (137). This cap is reverse transcribed to a cytosine (C), which is then converted to an unencoded G in the final sequenced reads (51, 54, 56, 138) (**fig. S2A**). Reanalysis of 5' RNA sequencing data for the capped ERCC (External RNA Controls Consortium) RNA molecules (50), which are often used as spike-in mixes to accurately estimate transcript abundances, showed that most genuine TSS reads had a single unencoded G, whereas noise reads



derived from non-TSS regions perfectly matched the reference sequences in both the template switching–based method and CAGE (fig. S2, C and D).



The use of only 5' scRNA-seq reads with an unencoded G for read 1 increased the proportion of reads mapped to TSS regions from 82.1% to 98.1% (Fig. 1E). Merely taking reads that started with a G (regardless of its presence in the genome) resulted in insufficient enrichment (88.3%) of reads mapped to TSS regions (Fig. 1E). Because the oligo-capping method removes caps with tobacco acid pyrophosphatase (TAP), cap structures are absent during reverse transcription. Accordingly, most reads produced by this method did not start with unencoded G (fig. S2, C and D). When we used RNAs treated with TAP as an input for CAGE, the amount of cDNA was reduced to 10% (Supplementary Text 1 Fig. 1). Overall, filtering based on unencoded G information allowed accurate detection of the 5' ends of transcripts by removing false positives in the template switching method and CAGE.



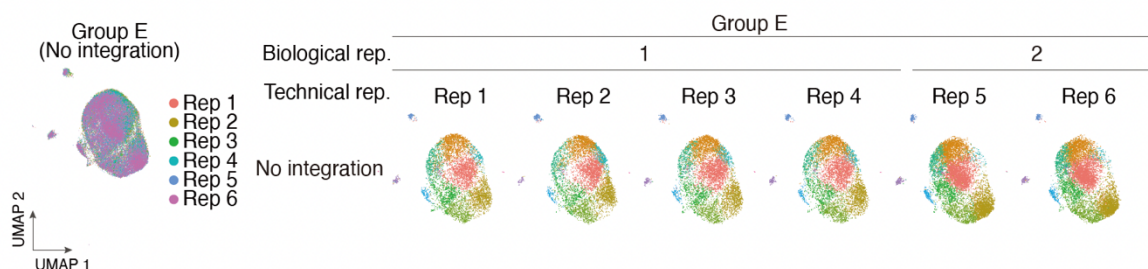


## Supplementary Text 2

### Evaluation of batch effects in our 5' scRNA-seq data

To evaluate the extent of batch effects in our experiments, we compared 5' scRNA-seq data between technical replicates prepared using different 10x Chromium chips and wells for Group E CD4<sup>+</sup> T cells isolated from the same donor on the same day. The data overlapped well without prior use of Harmony (“unintegrated” analysis), indicating that the batch effects were almost negligible (**Supplementary Text 2 Fig. 1**). The same trend was also observed for Groups A to D (**Supplementary Text 2 Fig. 2**). Biological replicates of 5' scRNA-seq data generated for Group E CD4<sup>+</sup> T cells isolated from the same donor on different days also showed good overlap without the use of Harmony (**Supplementary Text 2 Fig. 1**). These results suggest that batch effects were virtually negligible for 5' scRNA-seq data from the same donor, even if the experiments were performed on different days.

**Supplementary Text 2 Fig. 1**



### Merging CD4<sup>+</sup> T cell data from Groups A to E

To investigate how the data on CD4<sup>+</sup> T cells from different groups can be merged, we merged 5' scRNA-seq data on Groups C to E of CD4<sup>+</sup> T cells isolated from the same donor on the same day without prior use of Harmony, as batch effects were expected to be negligible, as described above. Most cells in Groups C to E were plotted in different spaces, but some were plotted in overlapping spaces (**Supplementary Text 2 Fig. 3**).

It is noteworthy that Groups C to E contained a fraction of CD4<sup>+</sup> T cells that were overlapping at the transcriptome level. This may be because sorting CD4<sup>+</sup> T cells into Groups C to E by expression of the membrane proteins CCR6, CXCR5, and CXCR3 did not completely separate them into distinct populations. This may be, at least in part, related to the continuum of helper T cells that has been well documented in many studies (25–35).

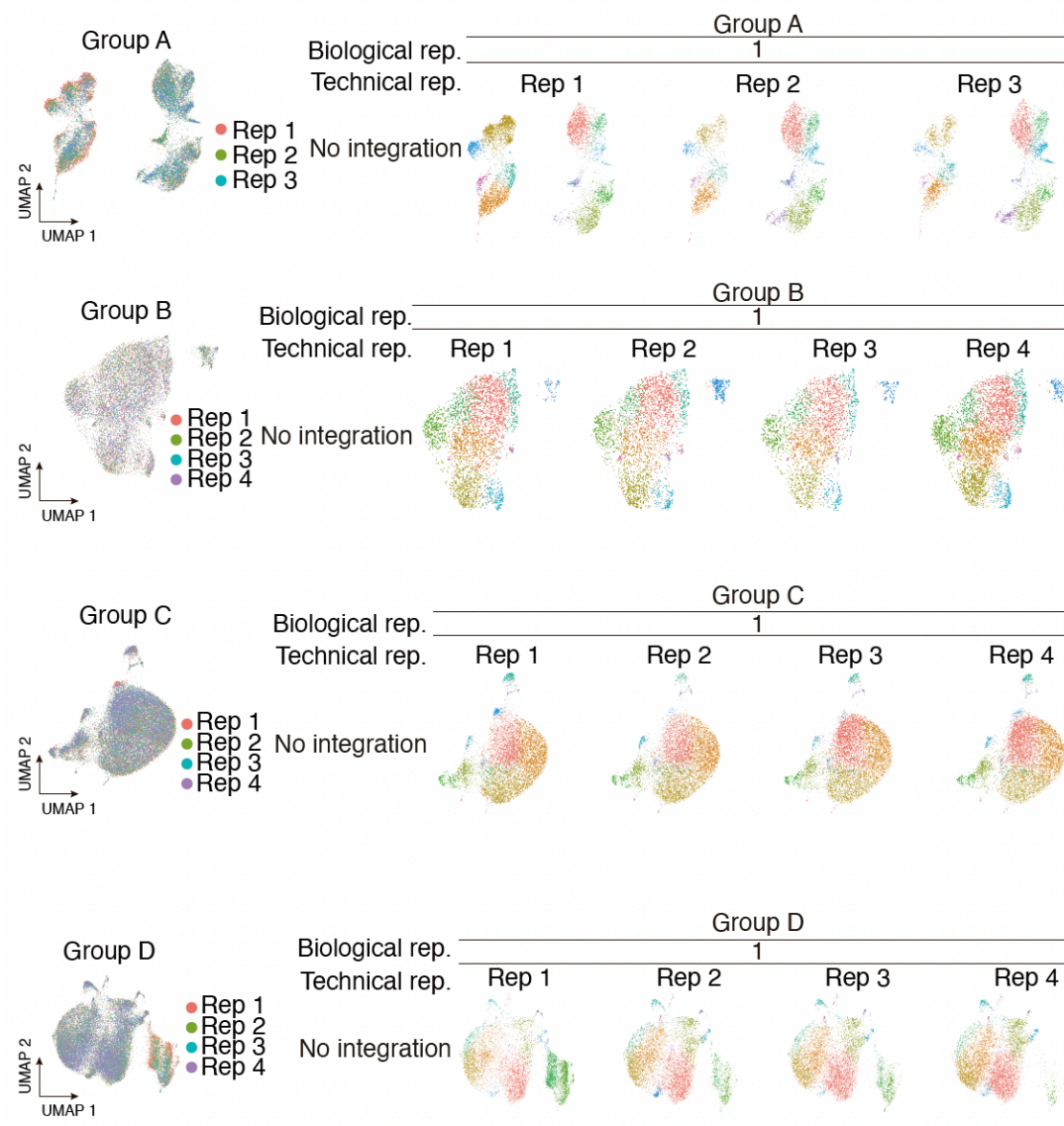
We then evaluated the performance of Harmony in our datasets on heterogeneous CD4<sup>+</sup> T cells. When the data were merged using Harmony (“Harmony integration”), CD4<sup>+</sup> T cells in Groups C to E were almost perfectly aligned (**Supplementary Text 2 Fig. 4**), even though the majority of the cells were functionally distinct. This result indicates that merging all data from different groups using Harmony can remove even biological differences in this study. We assume that this complication is likely due to the fact that we were distinguishing between very similar cell types within CD4<sup>+</sup> T cells rather than between different cell types such as neurons, hepatocytes and T cells.

### Integration of 5' scRNA-seq and Multiome data

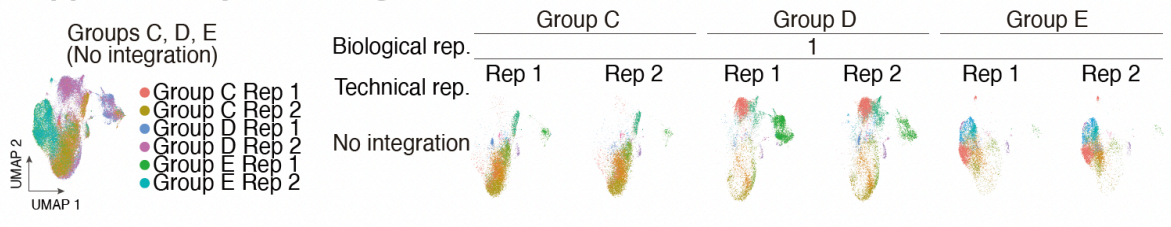
To directly compare eRNA transcription and chromatin accessibility, we merged 5' scRNA-seq data from different donors and Multiome (3' snRNA-seq) data. Harmony failed to

fully align the 5' scRNA-seq and Multiome data, but they were successfully aligned by CCA (Supplementary Text 2 Fig. 5) as previously reported by the Regev group (139). Thus, we used CCA integration to merge 5' scRNA-seq and Multiome data on CD4<sup>+</sup> T cells in each group.

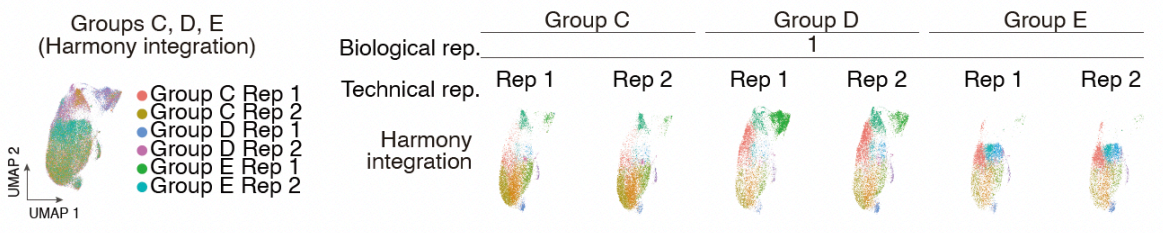
## Supplementary Text 2 Fig. 2



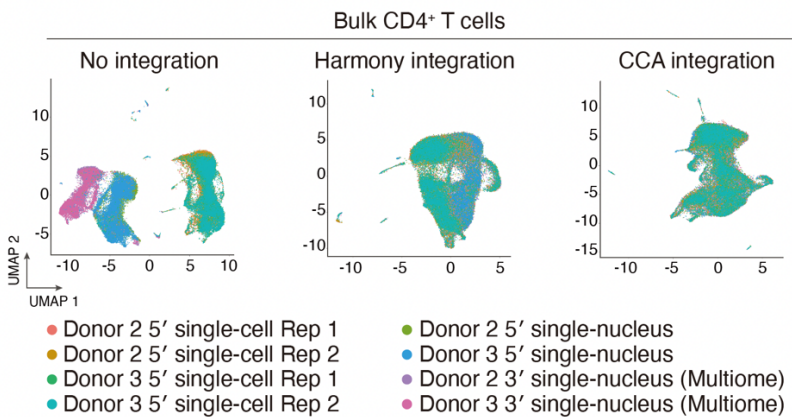
**Supplementary Text 2 Fig. 3**

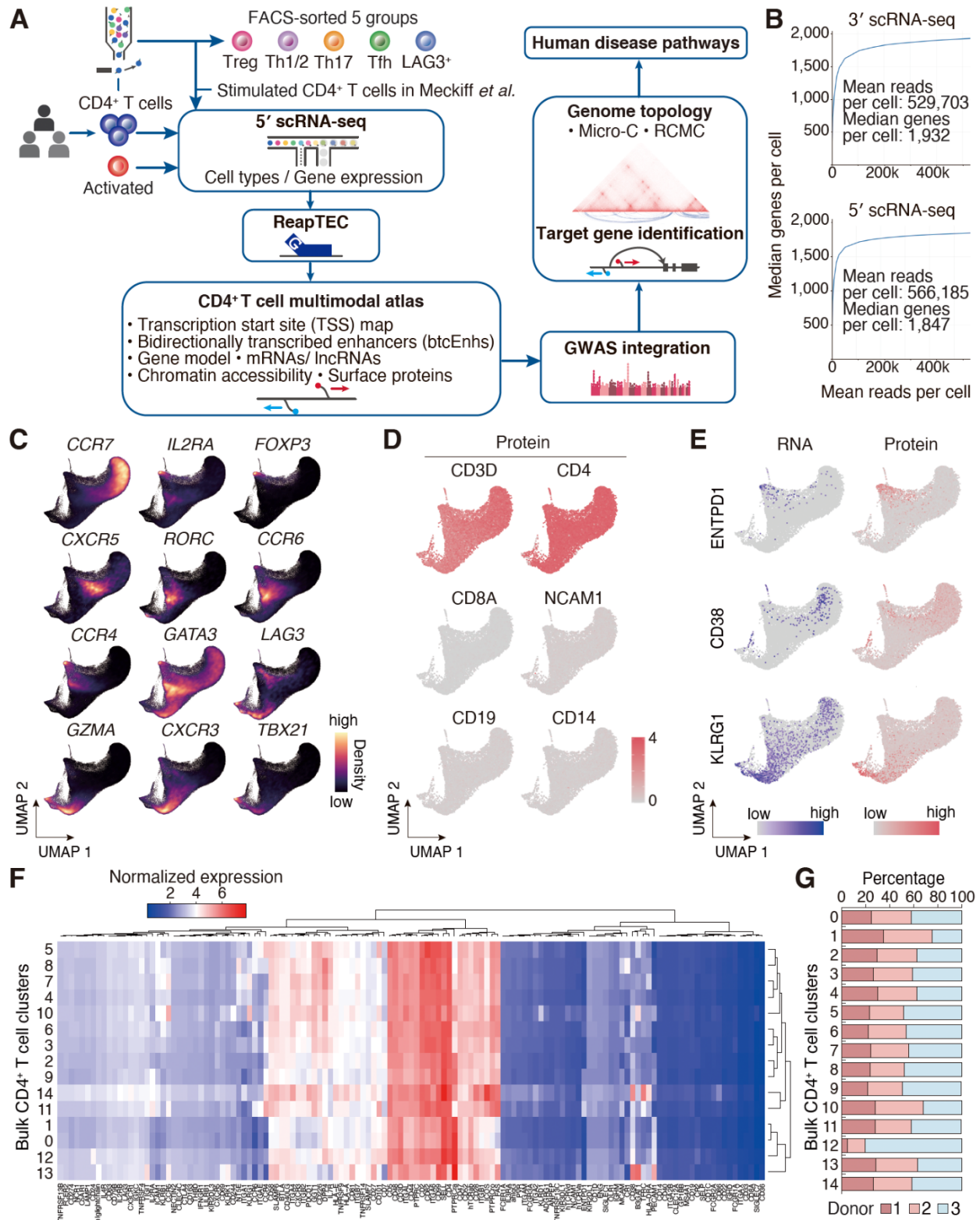


**Supplementary Text 2 Fig. 4**



**Supplementary Text 2 Fig. 5**





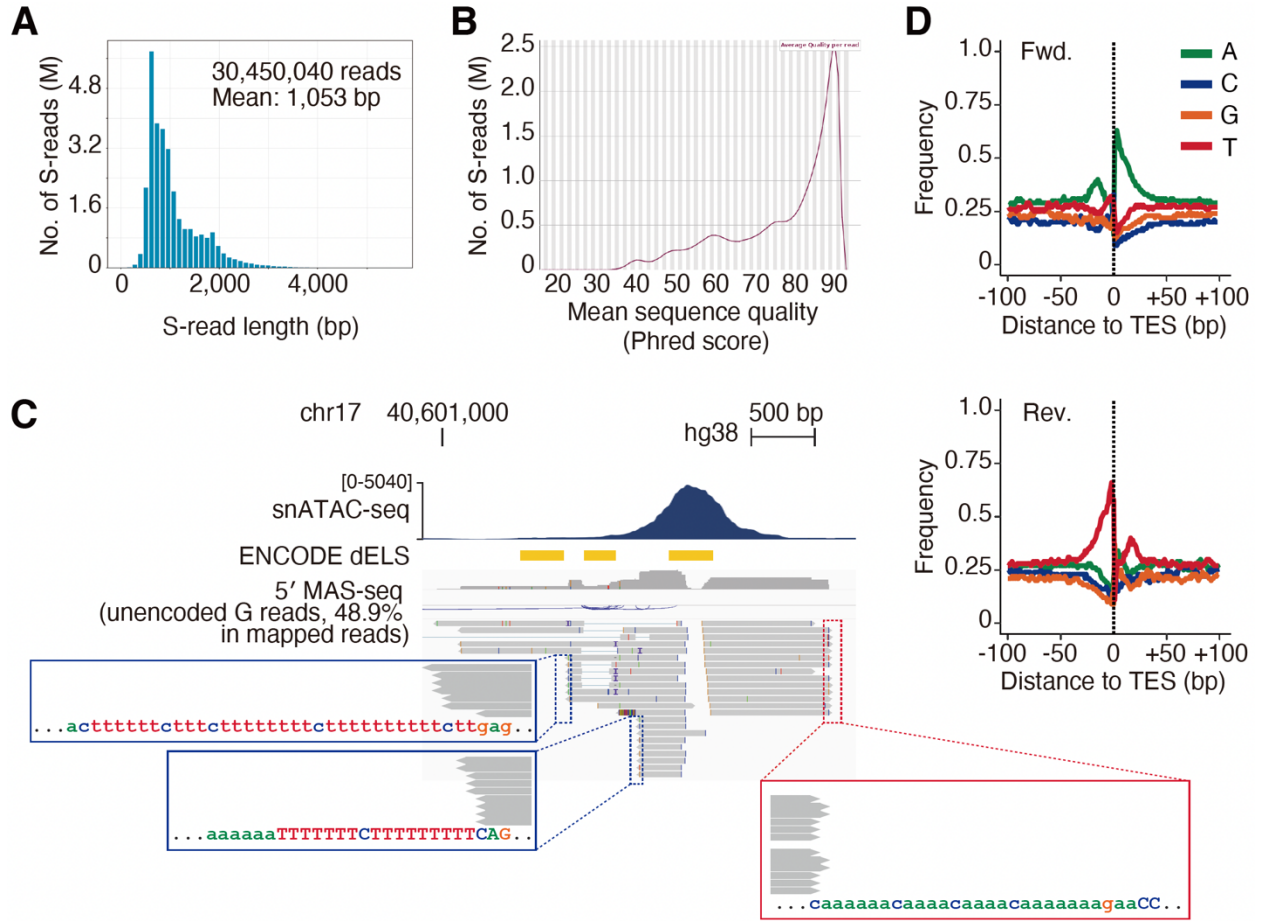
**Fig. S1. Transcriptomic and proteomic characterization of distinct cell clusters identified in circulating bulk CD4<sup>+</sup> T cells.** (A) Schematic overview of the study design. scRNA-seq, single-cell RNA sequencing; ReapTEC, read-level pre-filtering and transcribed enhancer call; FACS,

fluorescence-activated cell sorting; Treg, regulatory T cell; Th, T helper; Tfh, T follicular helper; mRNA, messenger RNA; lncRNA, long non-coding RNA; TSS, transcription start site; btcEnh, bidirectionally transcribed candidate enhancer; GWAS, genome-wide association study; RCMC, region-capture Micro-C. **(B)** Plots of the median number of genes per cell analyzed by 3' and 5' scRNA-seq in circulating bulk CD4<sup>+</sup> T cells from the same donor. **(C)** Estimated density of expression overlaid on the uniform manifold approximation and projections (UMAPs) for signature genes in circulating bulk CD4<sup>+</sup> T cells. **(D)** Cell surface protein expression projected on the UMAPs. **(E)** mRNA levels and corresponding surface protein levels projected on the UMAPs. **(F)** Heat map of normalized expression levels of surface proteins for the 15 clusters in circulating bulk CD4<sup>+</sup> T cells. Each column corresponds to one cellular indexing of transcriptomes and epitopes by sequencing (CITE-seq) antibody ( $n = 131$ ). **(G)** Proportion of cells from each donor in each cluster.

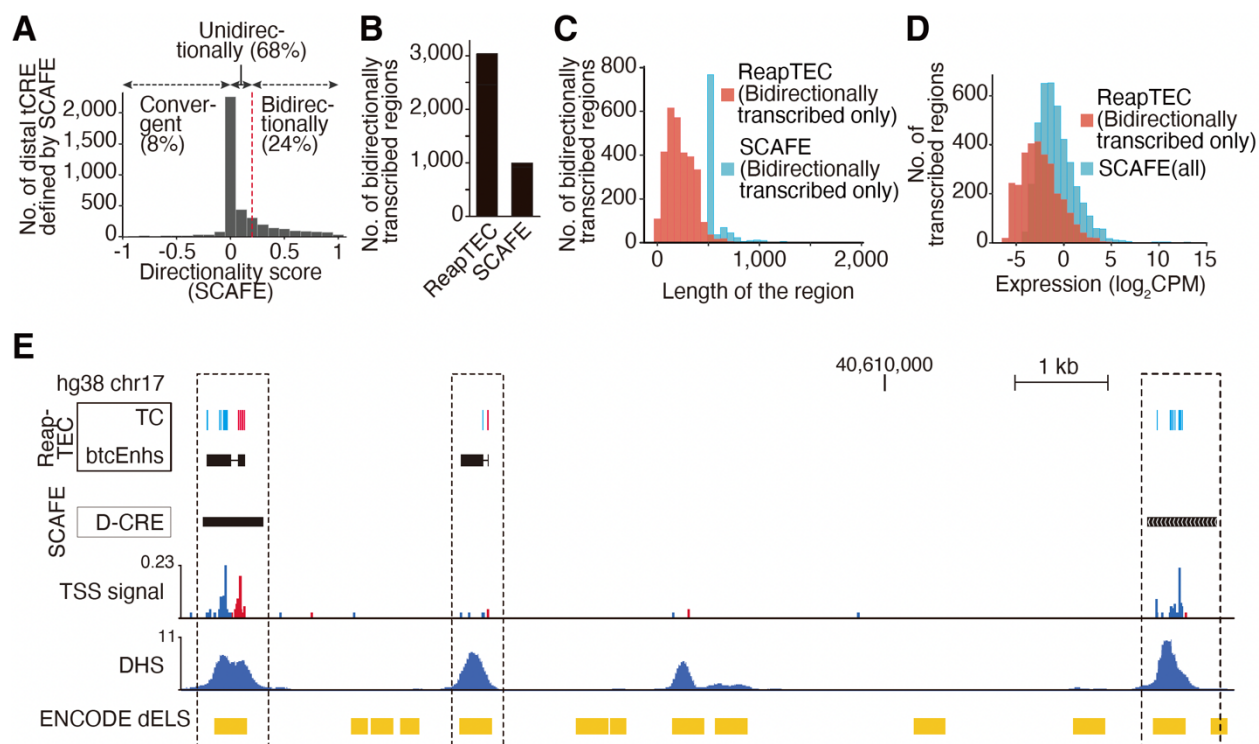




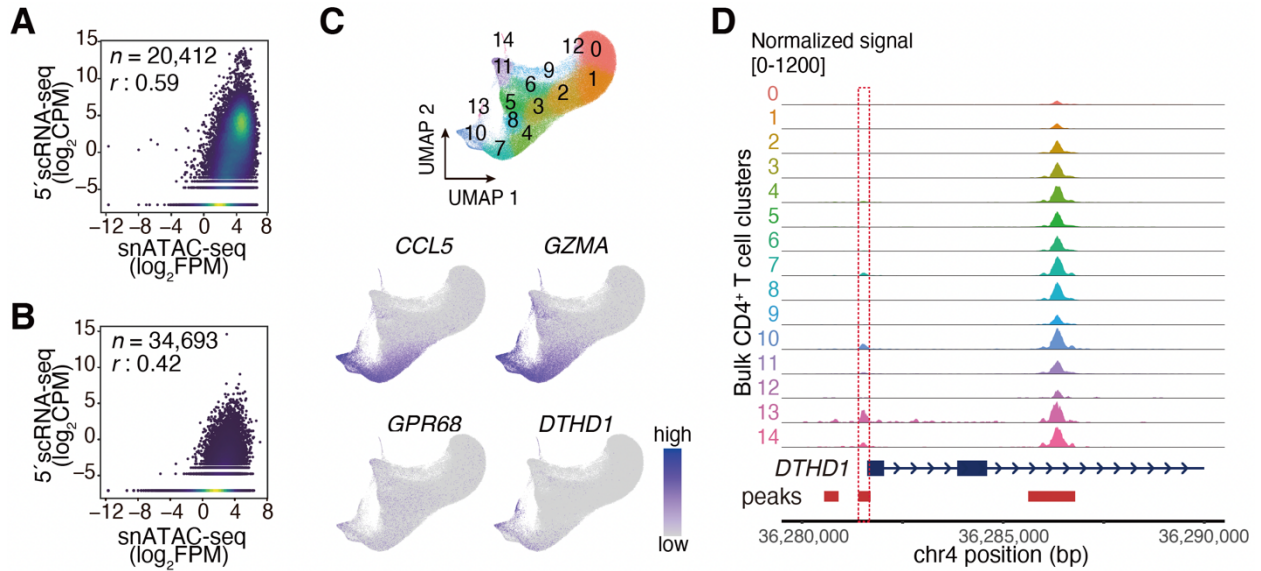
blue, reads derived from the gene body and not mapping to the 5' ends of ERCC RNA molecules. **(D)** Illustration of the mapping statuses of the read 1 5' ends (top). M, perfect match to the reference sequence; S1, one soft-clipped G or H (representing A, T, or C); S2, two soft-clipped bases. Bar graphs show the proportions of reads (read 1) with different mapping statuses of 5' ends and are based on capped ERCC data (cap trapper: GSM2772304, GSM2772306; template switching: GSM2772309, GSM2772310; oligo-capping: GSM2772313). Number of reads mapped (from left to right): 140,451; 177,615; 182,153 to TSS regions, and 3,190; 5,946; 35,947 to gene bodies. Most reads mapped to TSS regions start with a single unencoded G in the cap trapper and template-switching methods, but not in the oligo-capping method, where the cap is removed prior to library construction.



**Fig. S3. MAS-seq indicates extensive oligo(dT) priming in internal A-rich regions.** (A) Length distribution of the segmented reads (S-reads) obtained by MAS-seq. (B) Distribution of sequence quality (phred) scores for S-reads. (C) Visualization of 5' MAS-seq and snATAC-seq data at ENCODE candidate enhancers located in the intergenic regions of chromosome 17. 5' MAS-seq full-length reads with an unencoded G are shown. Uppercase, 3' ends of aligned reads; lowercase, the reference genome. dELS, distal enhancer-like signatures. (D) Nucleotide preferences near the 3' ends of aligned reads. TES, transcription end site.

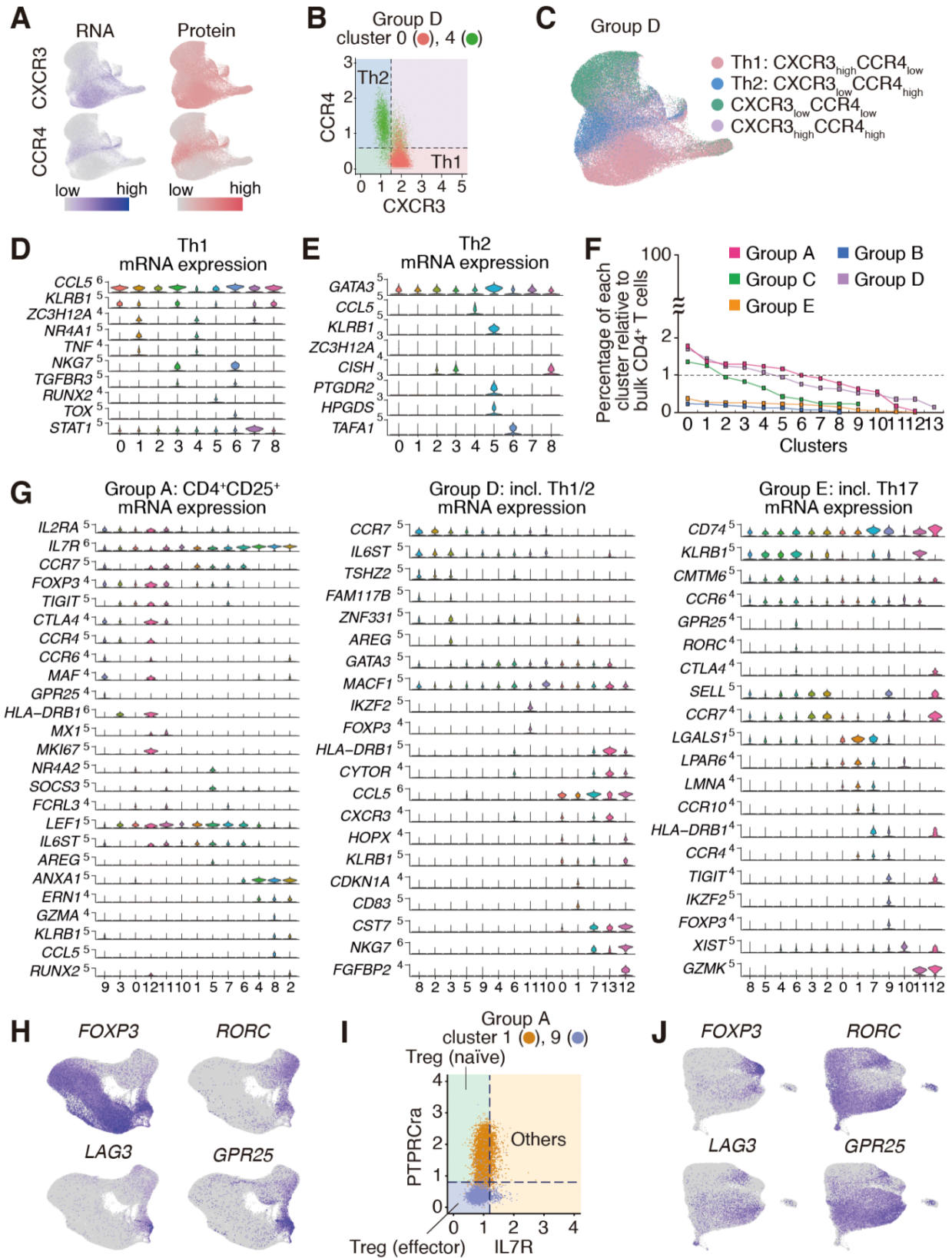


**Fig. S4. ReapTEC identifies bidirectionally transcribed enhancers with high sensitivity and at high nucleotide resolution.** (A) Distribution of directionality scores of distal transcribed *cis*-regulatory elements (tCREs) identified in circulating bulk CD4<sup>+</sup> T cells by SCAFE (46). (B) Bar plots of the number of bidirectionally transcribed regions detected in circulating bulk CD4<sup>+</sup> T cells by ReapTEC and SCAFE. (C) Length distribution of bidirectionally transcribed candidate *cis*-regulatory regions detected by ReapTEC (median length of 202 bp) and SCAFE (median length of 501 bp). (D) Distribution of the expression levels of transcribed regions detected by ReapTEC (btcEnhs, red) and SCAFE (all distal tCREs, blue). ReapTEC identified even btcEnhs expressed at low levels with high sensitivity. (E) Examples of tag clusters (TC) and btcEnhs identified by ReapTEC and distal tCREs (D-CREs) identified by SCAFE at an intergenic region of chromosome 17. Red, plus strand; blue, minus strand.

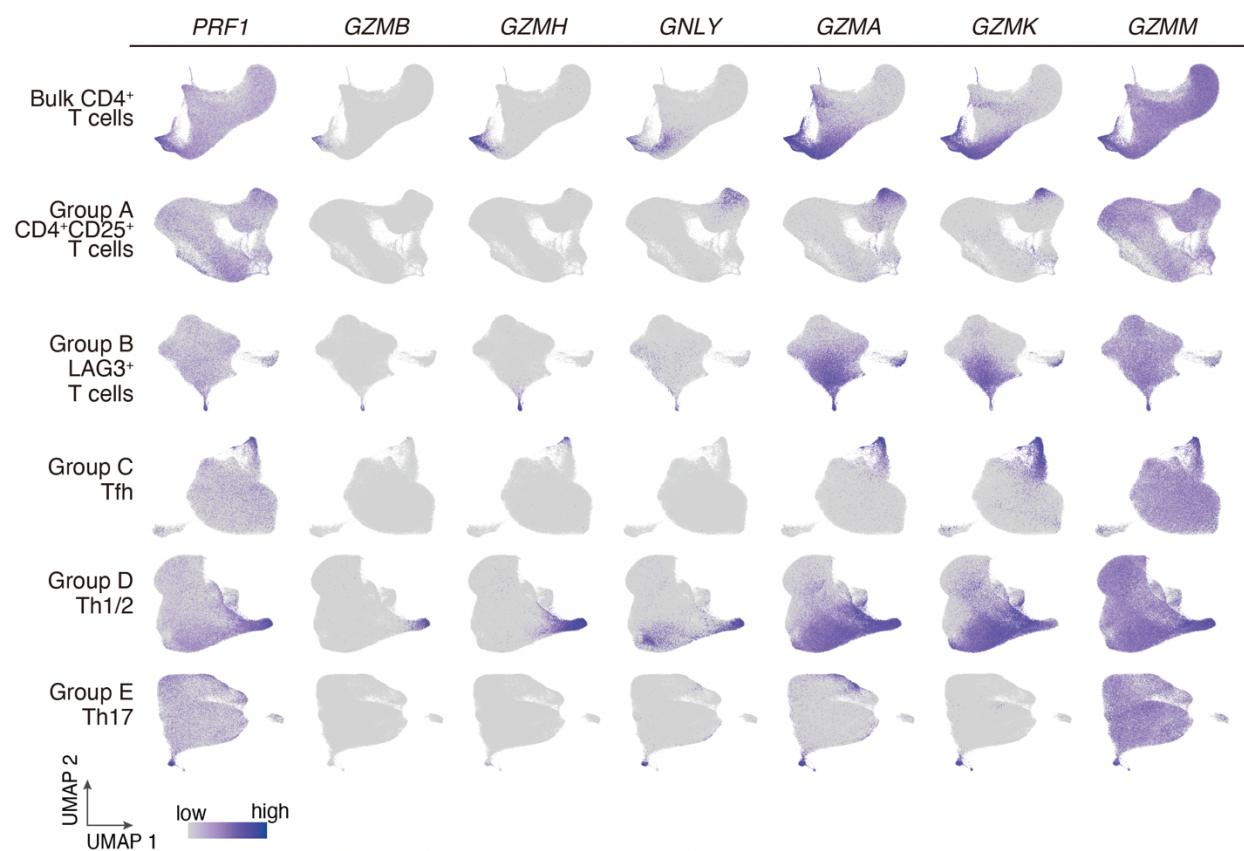


**Fig. S5. snATAC-seq in human circulating CD4<sup>+</sup> T cells.** (A and B) Density scatter plots of chromatin accessibility in snATAC-seq (x axis) and expression level in 5'scRNA-seq (y axis) at DNase hypersensitive sites (ENCODE accession: ENCFF707MRS) overlapping with  $\pm 300$ -bp regions around the TSSs of coding transcripts (GENCODE v39) (A) and at distal DNase hypersensitive sites outside the TSS regions (B). Cells from the same donor were used for snATAC-seq and 5' scRNA-seq. FPM, fragment per million; CPM, count per million. (C) UMAPs of the levels of gene expression (*CCL5*, *GZMA*, *GPR68*, and *DTHD1*) in circulating bulk CD4<sup>+</sup> T cells. Distinct clusters detected by the Seurat package are projected on the UMAP (top) in different colors. (D) Visualization of snATAC-seq peaks at the *DTHD1* locus showing the cluster-specific chromatin accessibility in circulating bulk CD4<sup>+</sup> T cells. Distinct clusters detected by the Seurat (86) and Signac packages (87) are shown in different colors.





**Fig. S6. 5'scRNA-seq analysis of various CD4<sup>+</sup> T cell subsets.** (A) mRNA levels and corresponding cell surface protein levels projected on UMAPs. *CXCR3* and *CCR4* are signature genes of Th1 and Th2 cells, respectively. (B) Scatter plot of *CXCR3* and *CCR4* surface protein expression levels for cells in clusters 0 and 4 in Group D. Dashed lines, the cutoff to classify the cells into Th1 and Th2. (C) UMAPs of cells classified by *CXCR3* and *CCR4* surface protein levels (using cutoffs defined in (B)). (D and E) Violin plots of mRNA levels of the signature genes in each cluster of Th1-like cells (D) (clusters 0, 1, and 7 in Group D) and Th2-like cells (E) (clusters 4 and 6 in Group D). (F) Distribution of the percentage of each cluster relative to the bulk CD4<sup>+</sup> T cells. (G) Violin plots of the mRNA levels of the signature genes in each cluster in Groups A, D, and E. (H) UMAPs of the levels of the *FOXP3*, *RORC*, *LAG3*, and *GPR25* genes in Group A CD4<sup>+</sup> T cells. (I) Scatter plot of IL7R and PTPRCra surface protein expression levels used to classify cells in clusters 1 and 9 of Group A into naïve and effector Tregs. Dashed lines, the cutoff used classify the cells. (J) UMAPs of the levels of the *FOXP3*, *RORC*, *LAG3*, and *GPR25* genes in Group E CD4<sup>+</sup> T cells.

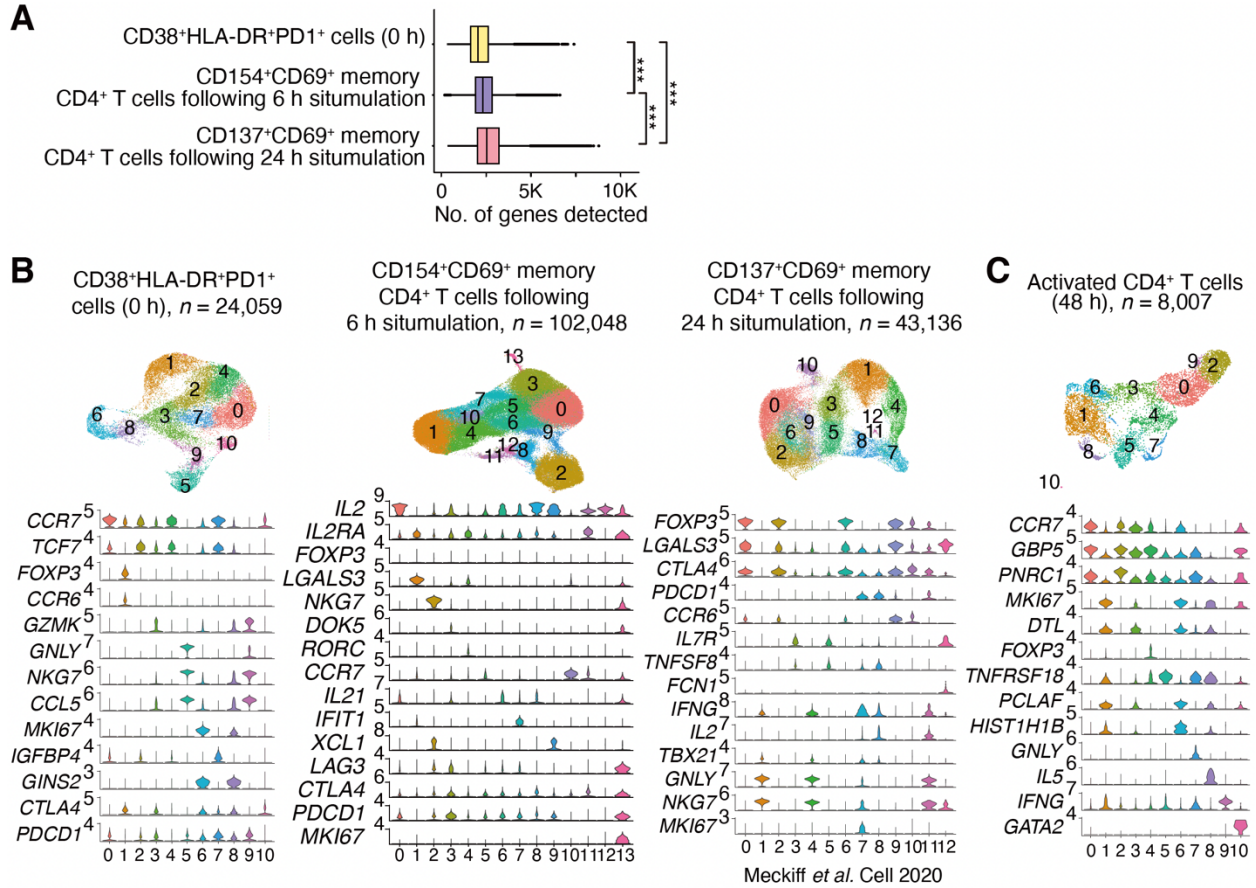


**Fig. S7. Single-cell expression profiling of granzyme-related genes in heterogeneous  $CD4^+$  T cells.** mRNA levels of perforin (*PRF1*), granulysin (*GNLY*), and granzymes (all other genes) in circulating bulk  $CD4^+$  T cells and 5 groups of  $CD4^+$  T cells.

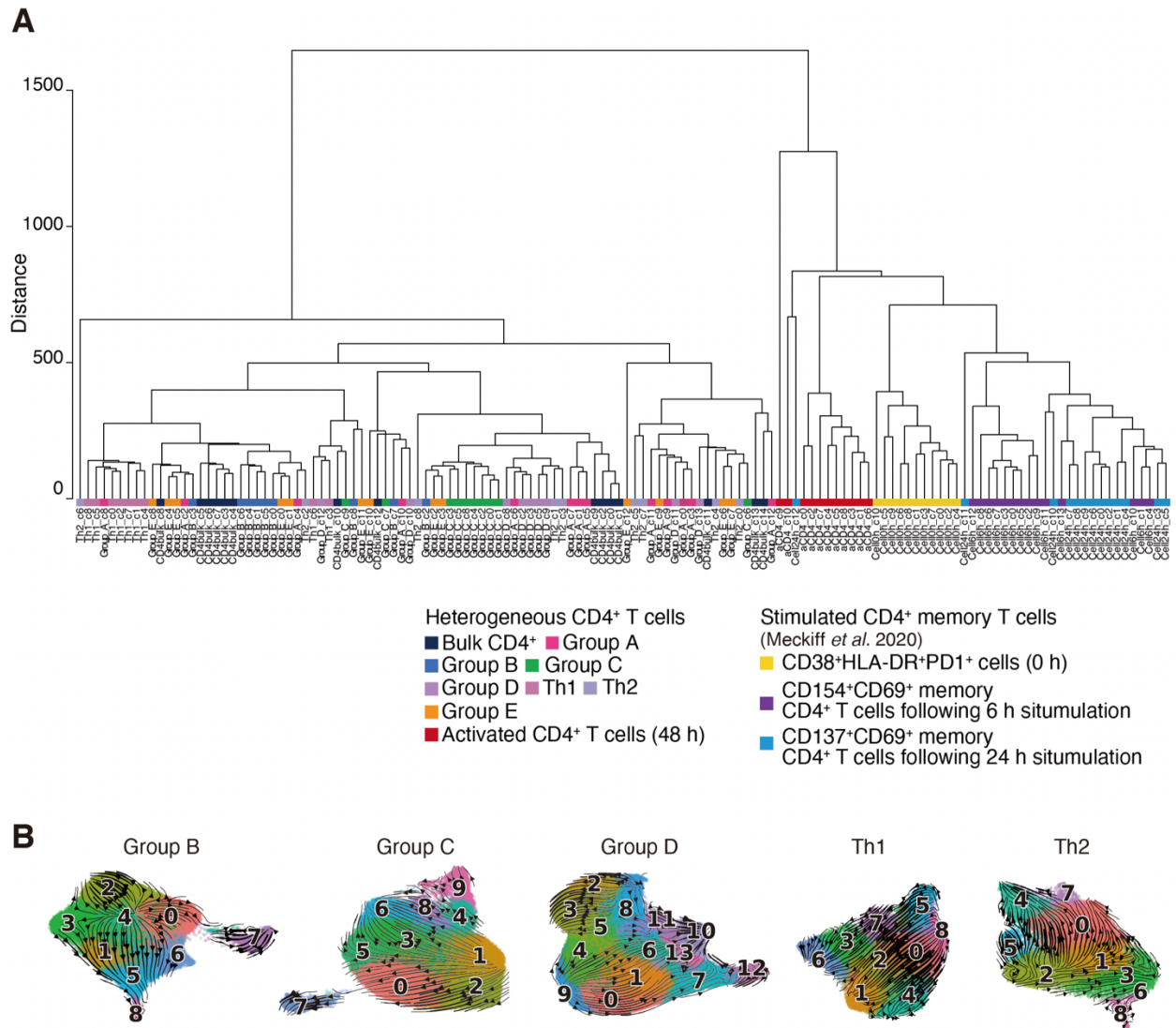


per each subset) associated with the R-HAS-198933 (Immunoregulatory interactions between a Lymphoid and a non-Lymphoid cell), hsa04060 (Cytokine-cytokine receptor interaction), and R-HAS-1280215 (Cytokine Signaling in Immune system) pathways.

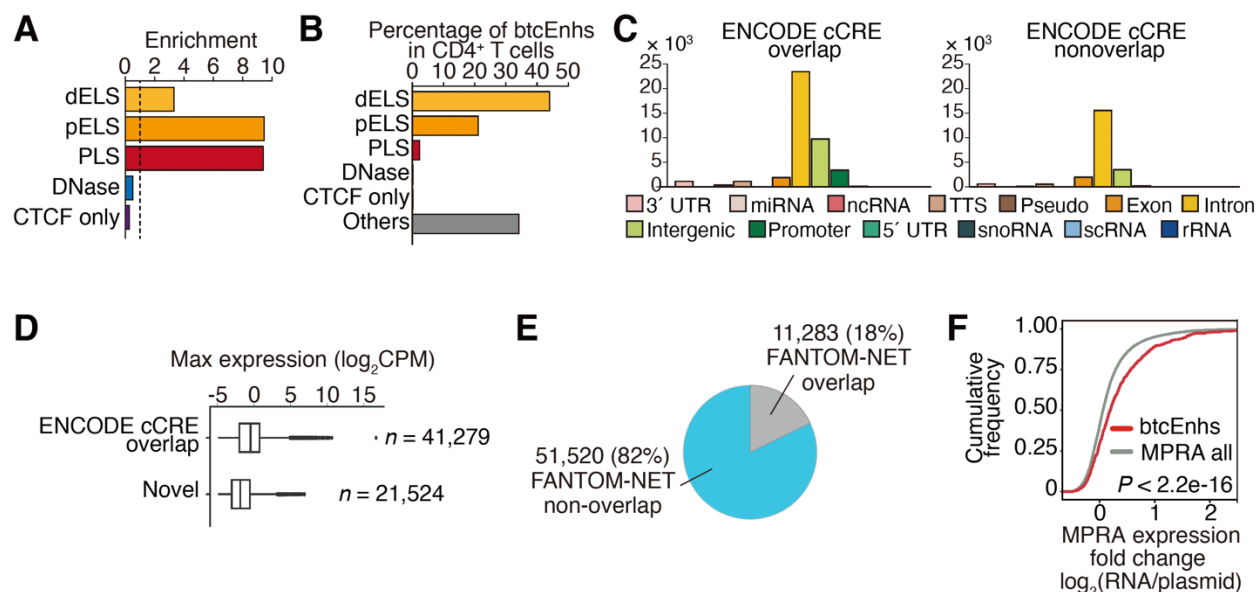




**Fig. S9. Transcriptional diversity in stimulated CD4<sup>+</sup> T cells.** (A) Distribution of the number of genes detected in 5' scRNA-seq analysis (65) of memory CD4<sup>+</sup> T cells before and after stimulation. \*\*\* $P < 2.2e-16$ , Wilcoxon rank sum test. (B and C) Top, UMAPs of distinct clusters detected by the Seurat package. Bottom, violin plots of the mRNA levels of the signature genes in each cluster.

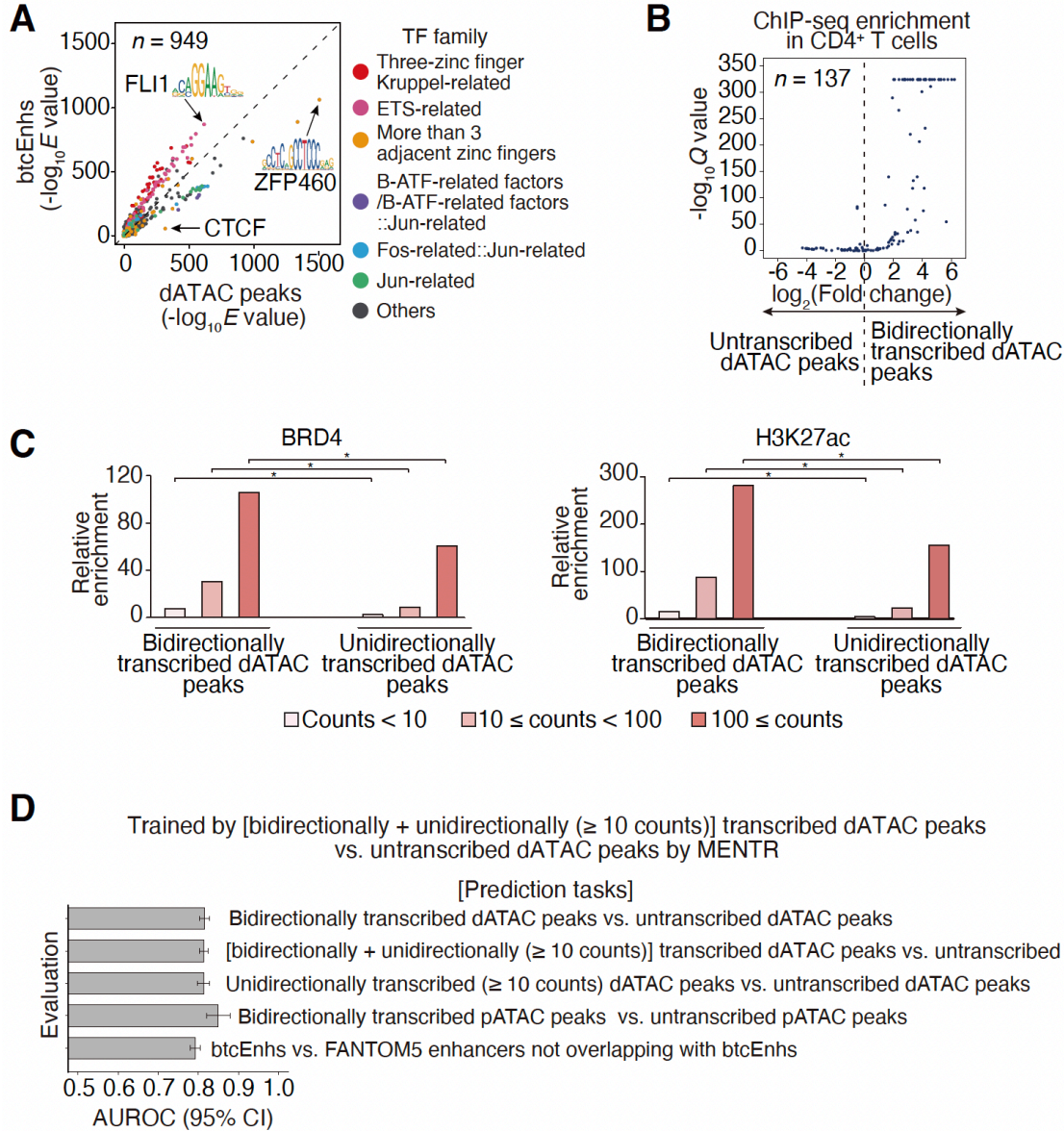


**Fig. S10. Hierarchical organization of the transcriptomic landscape across 136 CD4<sup>+</sup> T cell clusters. (A) Dendrogram of the hierarchical relationship among the 136 clusters based on promoter expression levels ( $n = 90,565$ ). (B) Velocity estimates projected onto UMAPs.**

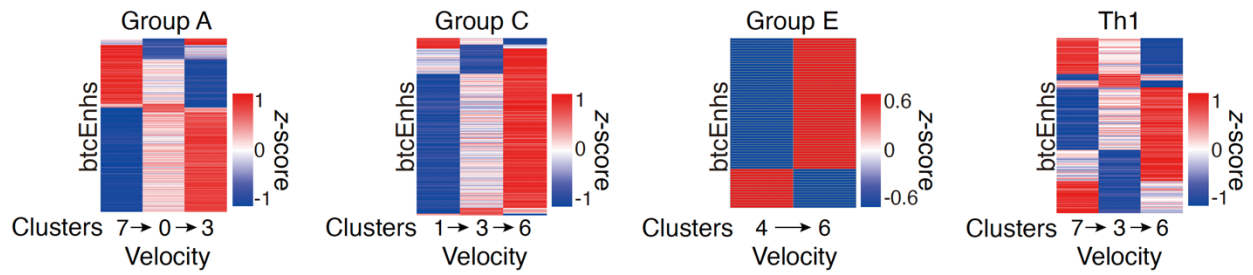


**Fig. S11. Genomic features of btcEnhs in CD4<sup>+</sup> T cells.** (A) Bar plots of the enrichment of ENCODE3 candidate *cis*-regulatory elements (cCREs) in btcEnhs compared to shuffled control genomic regions of the same size distribution as that of the btcEnh regions. (B) Bar plots of the fraction of btcEnhs that overlap with different classes of ENCODE3 cCREs (2). dELS, distal enhancer-like signatures; pELS, proximal enhancer-like signatures; PLS, promoter-like signatures; DNase, DNase I-hypersensitive sites. (C) Bar plots showing the number of known or novel btcEnhs that overlap with annotated sequences of different classes. The btcEnhs that overlap with ENCODE3 cCREs and the novel btcEnhs that do not overlap with ENCODE3 cCREs showed a similar genomic distribution. (D) Box plots of the distribution of maximum expression levels at btcEnhs among all 136 CD4<sup>+</sup> T cell clusters. Novel btcEnhs that do not overlap with ENCODE3 cCREs were also shown to express eRNAs in our 5' scRNA-seq analysis of heterogeneous CD4<sup>+</sup> T cells, albeit slightly weaker in general than btcEnhs that overlap with ENCODE3 cCREs. (E) Pie chart of the btcEnhs included and not included in the FANTOM5 enhancer atlas. (F) Cumulative distribution function of MPRA expression fold change (log<sub>2</sub>) against normalized plasmid tag counts for btcEnhs (red) and all fragments tested in the previous study (black) (66).



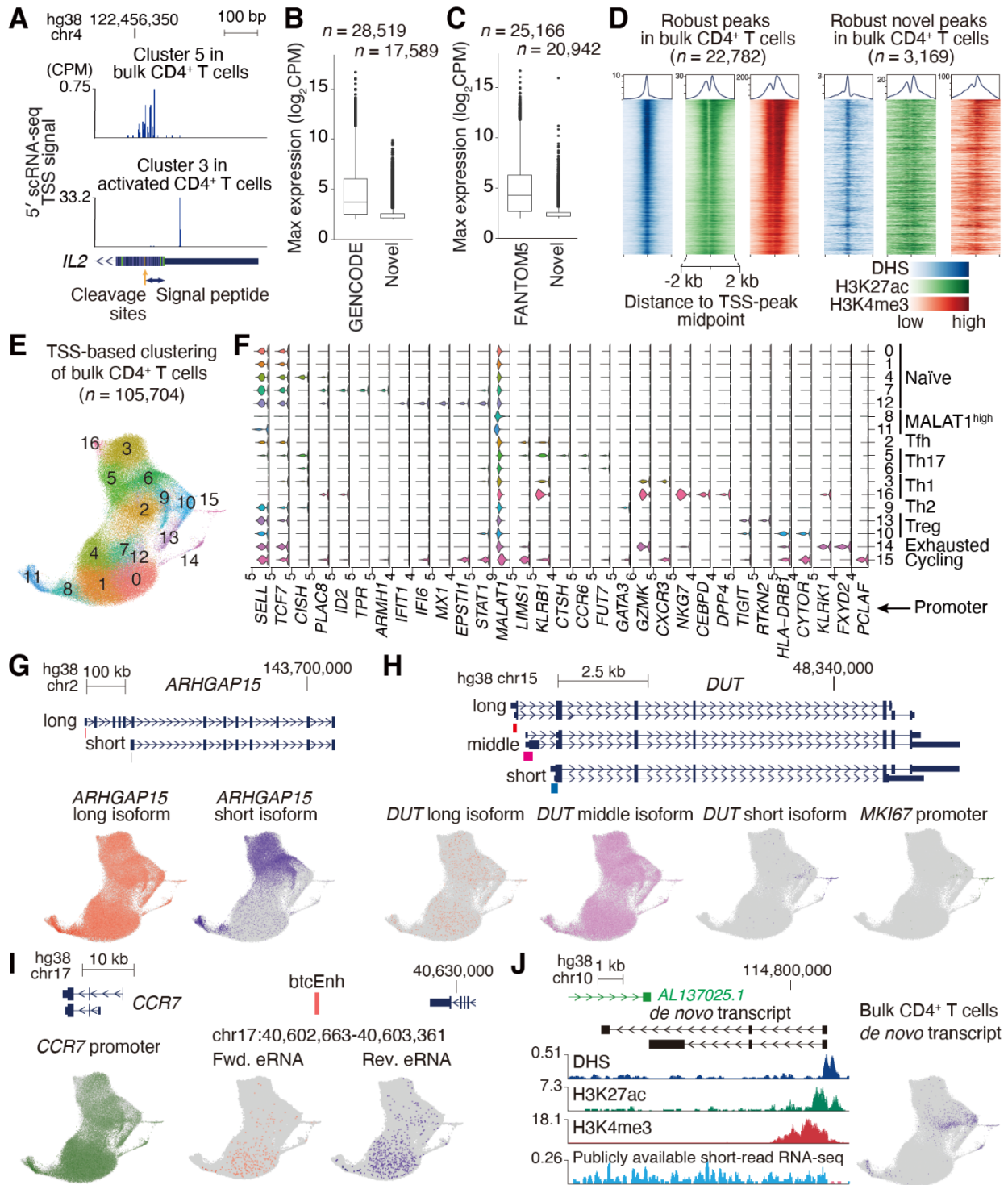


**Fig. S12. Sequence features in different categories of distal ATAC peaks.** (A) Scatter plot of the logarithms of  $E$  values for enrichment of transcription factor-binding motifs (TFBMs) in dATAC peaks and btcEnhs. TFBMs recognized by different transcription factor families are shown in different colors. (B) Volcano plot showing higher enrichment of transcription factor bindings in bidirectionally transcribed dATAC peaks compared to untranscribed dATAC peaks. (C) Enrichment of BRD4 binding and H3K27ac histone modification in bidirectionally or unidirectionally transcribed dATAC peaks (classified by different transcription levels) relative to untranscribed dATAC peaks.  $*P < 2.2e-16$ , Fisher's exact test. (D) Bar plots of the prediction accuracy (AUROC) of MENTR machine learning model (67) trained with bidirectionally or unidirectionally transcribed dATAC peaks (latter with at least 10 read counts) vs. untranscribed dATAC peaks. AUROC, area under receiver operating characteristic curve; CI, confidence interval.



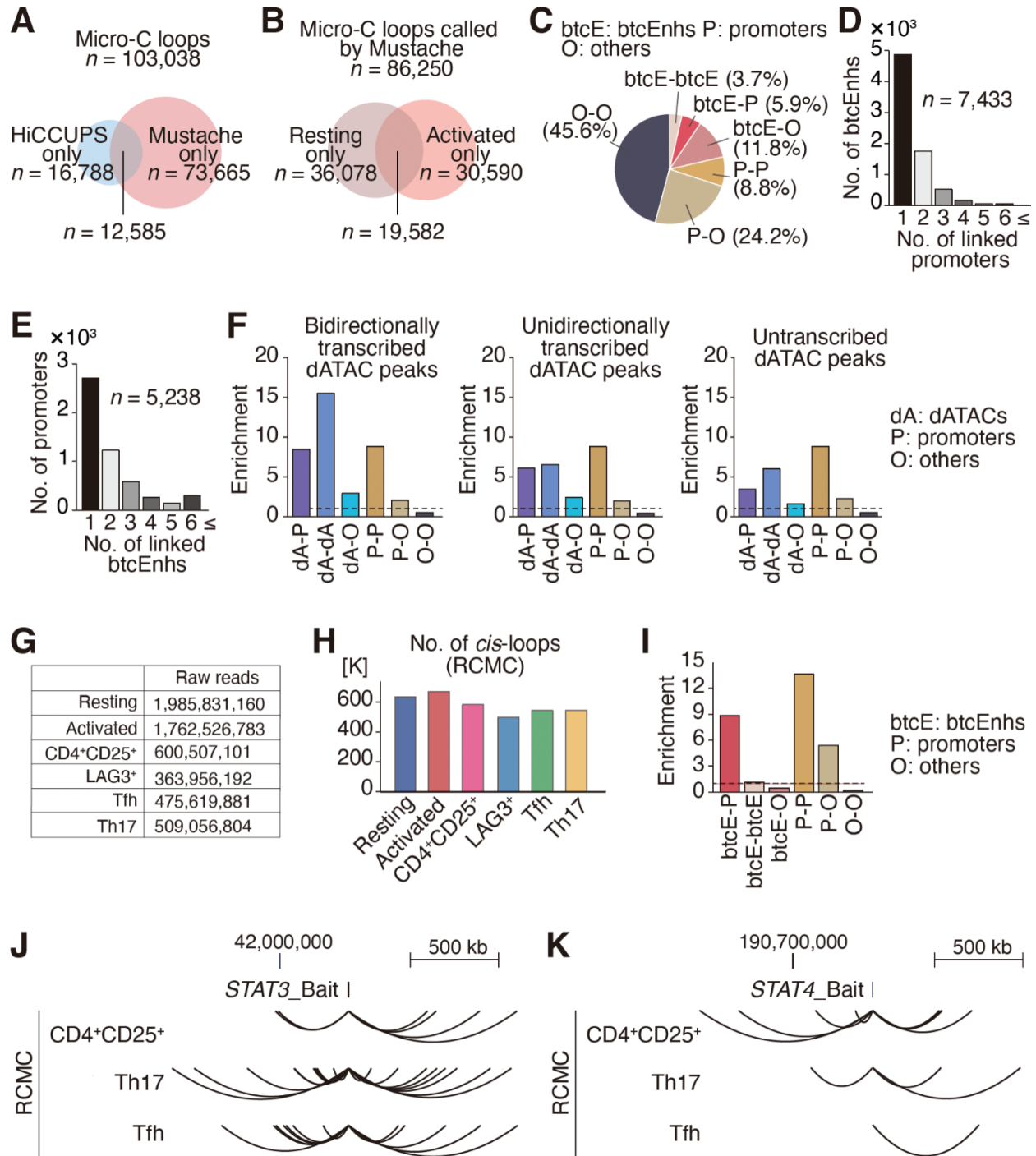
**Fig. S13. eRNA expression dynamics in the differentiation trajectories.** Heatmaps of scored eRNA expression of dynamically activated btcEnhs at a minimum log fold change of 3 and with a minimum expression of  $\log_2\text{CPM} \geq -5$ .





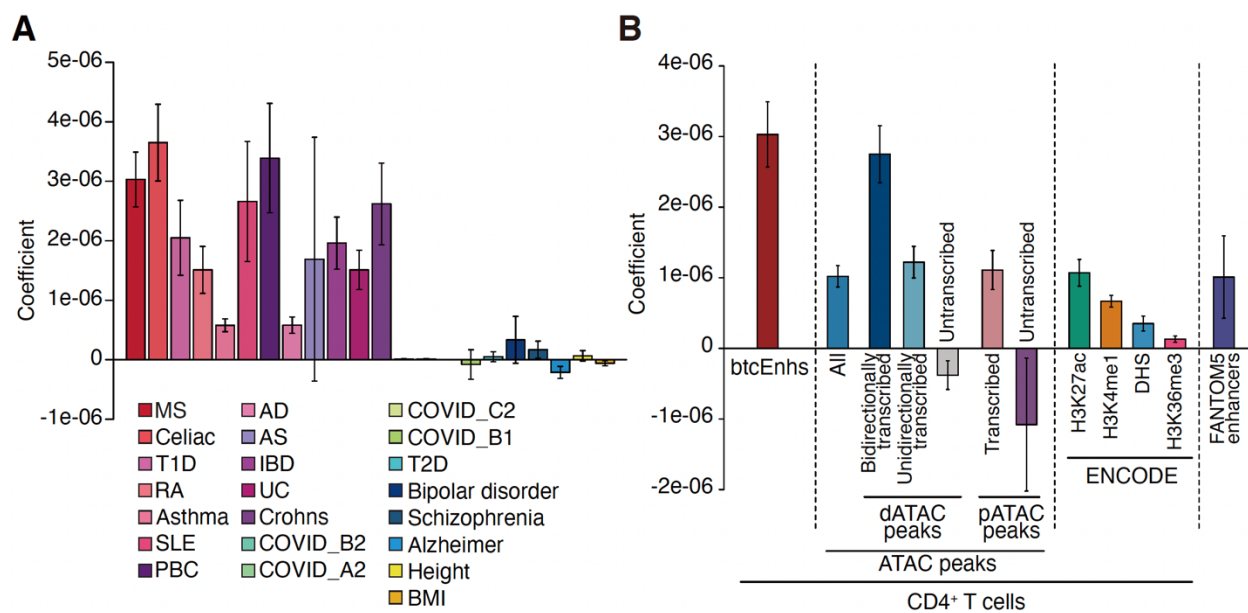
**Fig. S14. CD4<sup>+</sup> T cell heterogeneity at the TSS level.** (A) UCSC Genome Browser view of 5' scRNA-seq data at the *IL2* locus. CPM, counts per million. (B and C) Box plots of the distribution of maximum expression levels at robust TSS peaks among all 136 CD4<sup>+</sup> T cell clusters. (B) Robust TSS peaks overlapping (FANTOM5) or not (Novel) with the FANTOM5 peaks. (C) Robust TSS peaks located within  $\pm 300$ -bp regions around the 5' ends of transcripts (GENCODE) and robust TSS peaks not overlapping with the annotated TSS regions (Novel). (D)

Heat maps of DHS, histone 3 lysine 27 acetylation (H3K27ac), and histone 3 lysine 4 trimethylation (H3K4me3) signals (ENCODE CD4<sup>+</sup> T cell datasets)  $\pm 2$  kb around the robust TSS peaks or robust novel peaks not overlapping with the annotated TSS regions in bulk CD4<sup>+</sup> T cells. **(E)** UMAP and TSS-based clustering analysis of a single-cell dataset combining 5' scRNA-seq and 5' snRNA-seq datasets. **(F)** Violin plots of the expression of the signature promoters in 17 clusters identified in (E). **(G and H)** Examples of alternative promoter usage in circulating bulk CD4<sup>+</sup> T cells at the *ARHGAP15* (G) and *DUT* loci (H). UMAPs show the levels of mRNAs transcribed from each promoter. **(I)** UMAPs of levels of mRNAs from the *CCR7* promoter (left) and of enhancer RNA (eRNA) in btcEnhs located upstream of *CCR7*. **(J)** Left, UCSC Genome Browser views of the assembly of *de novo* transcripts defined by MAS-seq. The promoter region for these transcripts has a DHS peak and the H3K27ac and H3K4me3 histone modifications. These transcripts were also detected by publicly available RNA-seq data (ENCODE: ENCSR033XWU), but it was difficult to identify transcript models using only this data. *Y* axes: read-depth normalized signal for DHS, signal *P* values for modifications, and counts for RNA-seq. Right, UMAPs of the levels of *de novo* transcripts in circulating bulk CD4<sup>+</sup> T cells.



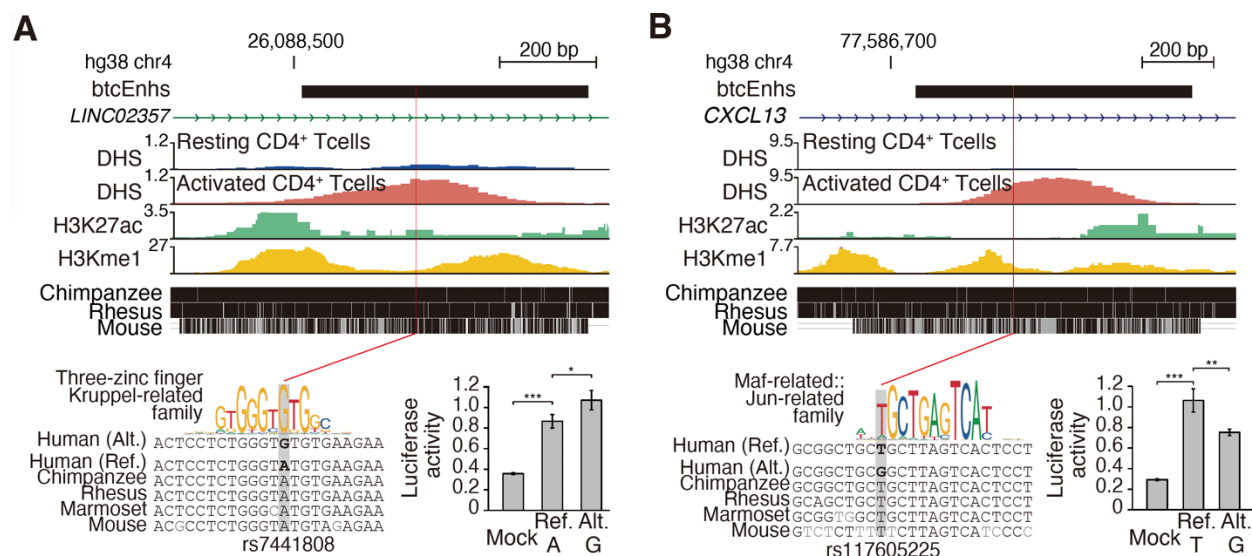
**Fig. S15. Fine-scale Micro-C and region-capture Micro-C chromatin contact maps.** (A) Venn diagram of the number of loops called by HiCCUPS and Mustache. In total, 103,038 Micro-C loops were detected in this study. (B) Venn diagram of the number of loops detected by Mustache in resting and activated CD4<sup>+</sup> T cells. (C) Pie chart of the Micro-C loop categories. btcE, btcEnh; P, promoter; O, others. (D) Histogram of the number of linked promoters per btcEnh. (E) Histogram of the number of linked btcEnhs per promoter. (F) Enrichment of Micro-C loops was analyzed as in Fig. 3C but separately for bidirectionally and unidirectionally

transcribed distal ATAC (dATAC) peaks, and for untranscribed dATAC peaks. **(G)** Number of raw reads for each region capture Micro-C (RCMC) experiment. **(H)** Bar plots of the number of *cis*-loops within 1 Mb genomic distance. **(I)** Bar plots of the enrichment of RCMC loops linking the indicated genomic regions over the shuffled control loops. btcE, btcEnh; P, promoter; O, others. **(J and K)** Examples of RCMC loops at the *STAT3* (J) and *STAT4* (K) loci showing the differences in chromatin contacts among CD4<sup>+</sup> T cell populations.

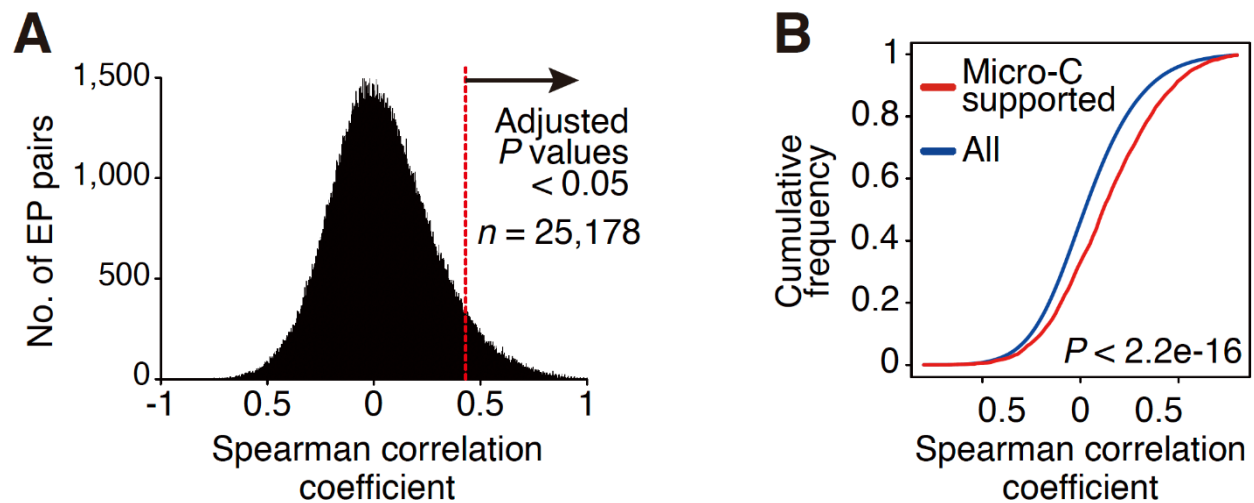


**Fig. S16. Enrichment of disease heritability in bidirectionally transcribed enhancers. (A)** Heritability coefficients for 62,803 btcEnhs in CD4<sup>+</sup> T cells. Linkage disequilibrium score regression analysis was performed using GWAS summary statistics for complex traits and immune-mediated diseases. MS, multiple sclerosis; T1D, type 1 diabetes; RA, rheumatoid arthritis; SLE, systemic lupus erythematosus; PBC, primary biliary cholangitis; AD, atopic dermatitis; AS, ankylosing spondylitis; IBD, inflammatory bowel disease; UC, ulcerative colitis; Crohns, Crohn's disease; COVID\_B2, COVID-19 hospitalization vs. population (9,986 vs. 1,877,672); COVID\_A2, severe respiratory COVID-19 vs. population (5,101 vs. 1,383,241); COVID\_C2, COVID-19 vs. population (38,984 vs. 1,644,784); COVID\_B1, COVID-19 hospitalization vs. COVID-19 without hospitalization (4,829 vs. 11,816); T2D, type 2 diabetes; BMI, body mass index. **(B)** Heritability coefficients for multiple sclerosis in different classes of genomic regions. The coefficients were higher in bidirectionally transcribed regions, such as the btcEnhs identified by 5' scRNA-seq and bidirectionally transcribed dATAC peaks, than in the other indicated *cis*-regulatory regions. All data are mean  $\pm$  s.e.

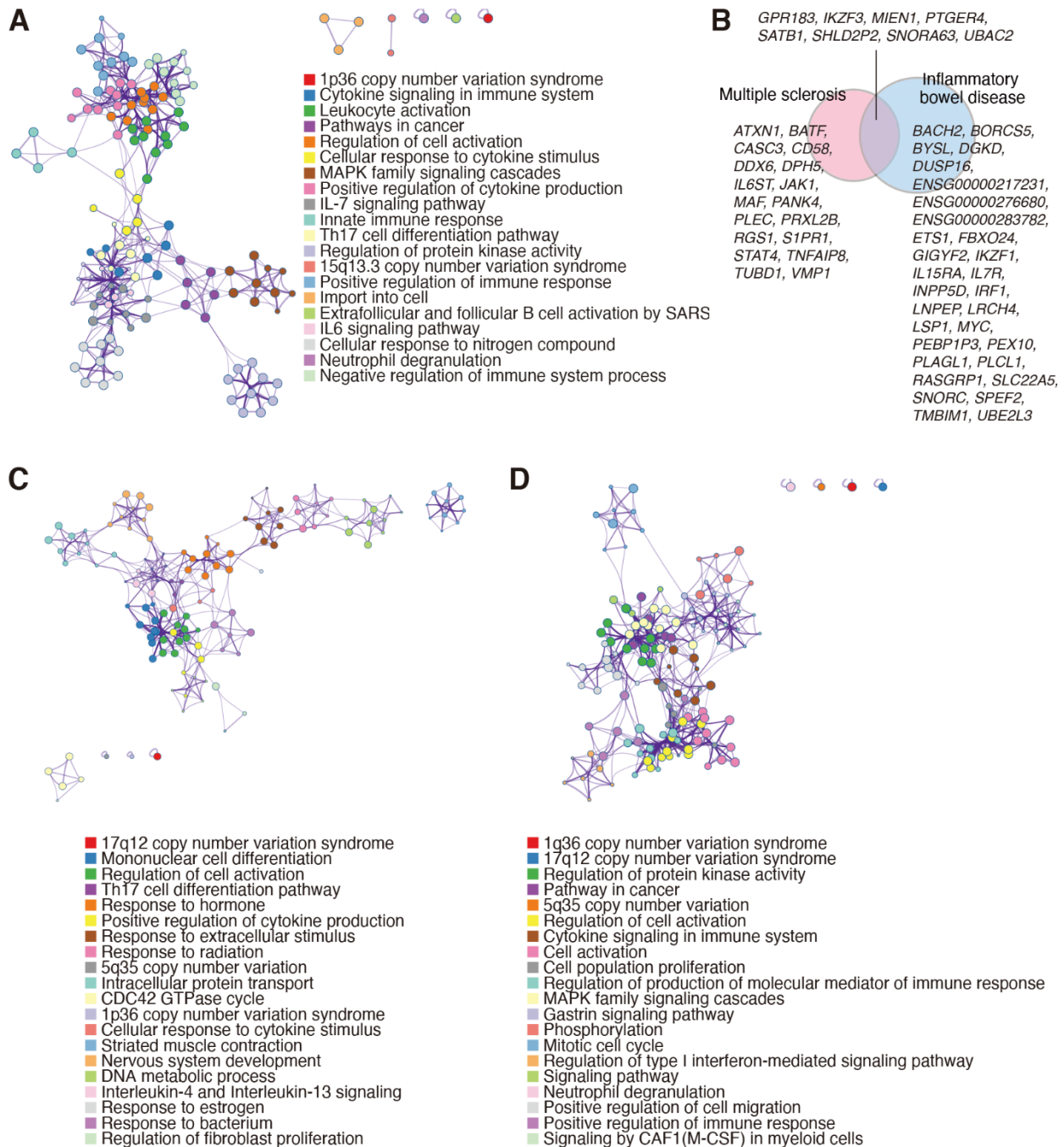




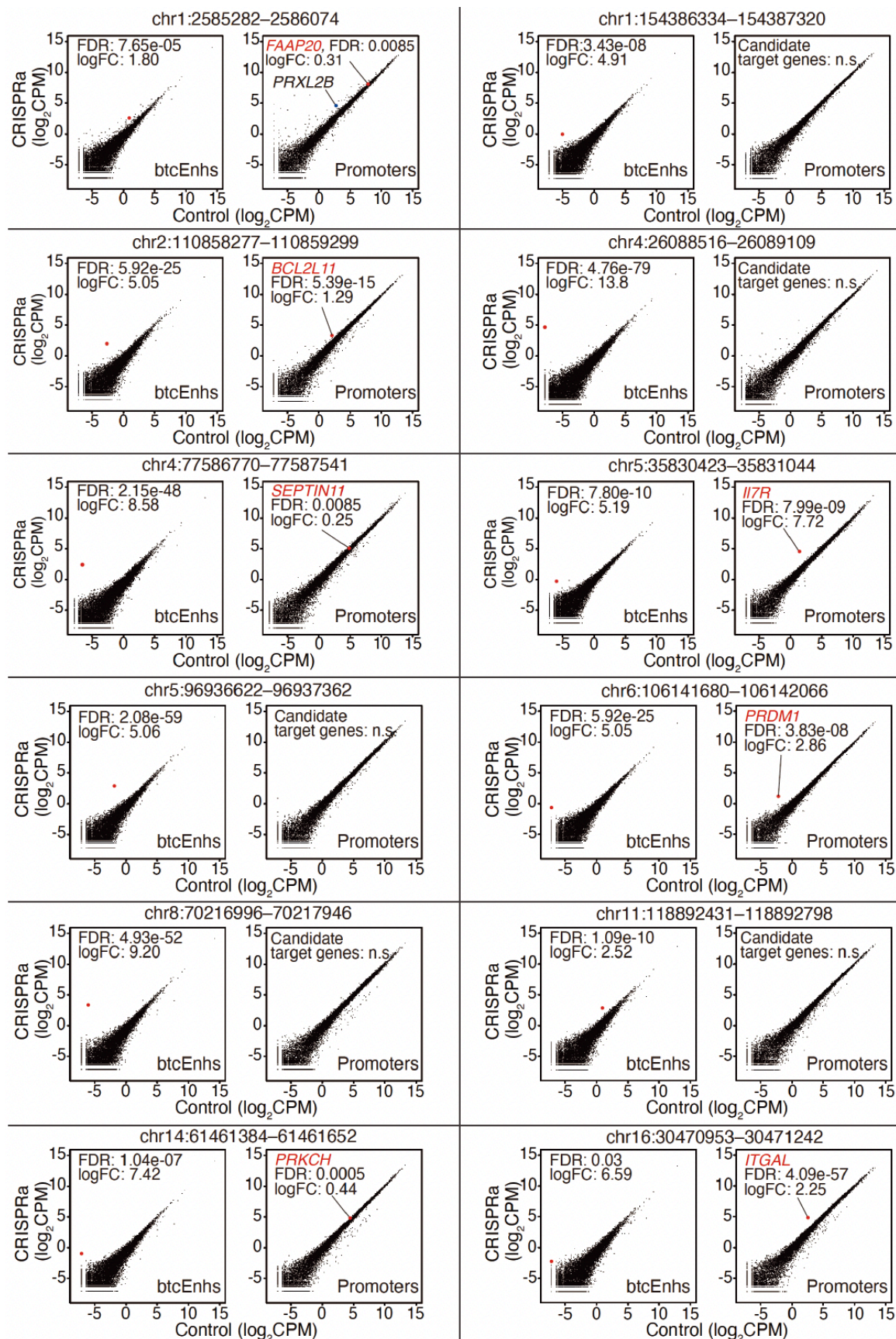
**Fig. S17. Examples of disease SNPs fine-mapped within btcEnh.** UCSC Genome Browser views of (A) a mammalian-conserved btcEnh (chr4:26088516–26089113) harboring the RA and T1D SNP rs7441808 and (B) a btcEnh (chr4:77586770–77587541) harboring the RA SNP rs117605225. Y axes: read depth–normalized signal for DHSs, and signal  $P$  values for histone modifications. The alternative (risk) G allele of rs7441808 creates a stronger transcription factor motif (A). The alternative (protective) G allele of rs117605225 disrupts the transcription factor motif (B). Bar plots of luciferase activity in Jurkat cells transfected with pGL4.24 vectors carrying enhancers (Ref., reference alleles; Alt., alternative alleles); mock transfection was used for normalization. Data are mean  $\pm$  s.d. from 12 replicates. \* $P < 0.1$ , \*\* $P < 0.05$ , \*\*\* $P < 0.001$ , Welch's  $t$ -test.



**Fig. S18. Prediction of target genes of btcEnh.** (A) Histogram of Spearman correlation coefficients for btcEnh-promoter (EP) co-variation. In total, 25,178 pairs of positively correlated btcEnh and promoters were identified (adjusted  $P$  values  $< 0.05$ ). The Holm's method was used to correct for multiple comparisons. Promoters are defined as  $\pm 300$ -bp regions around the TSSs of coding transcripts (GENCODE v39). (B) Cumulative distribution function of Spearman correlation coefficients of all btcE-P pairs (blue) and btcE-P pairs supported by Micro-C (red).  $P < 2.2e-16$



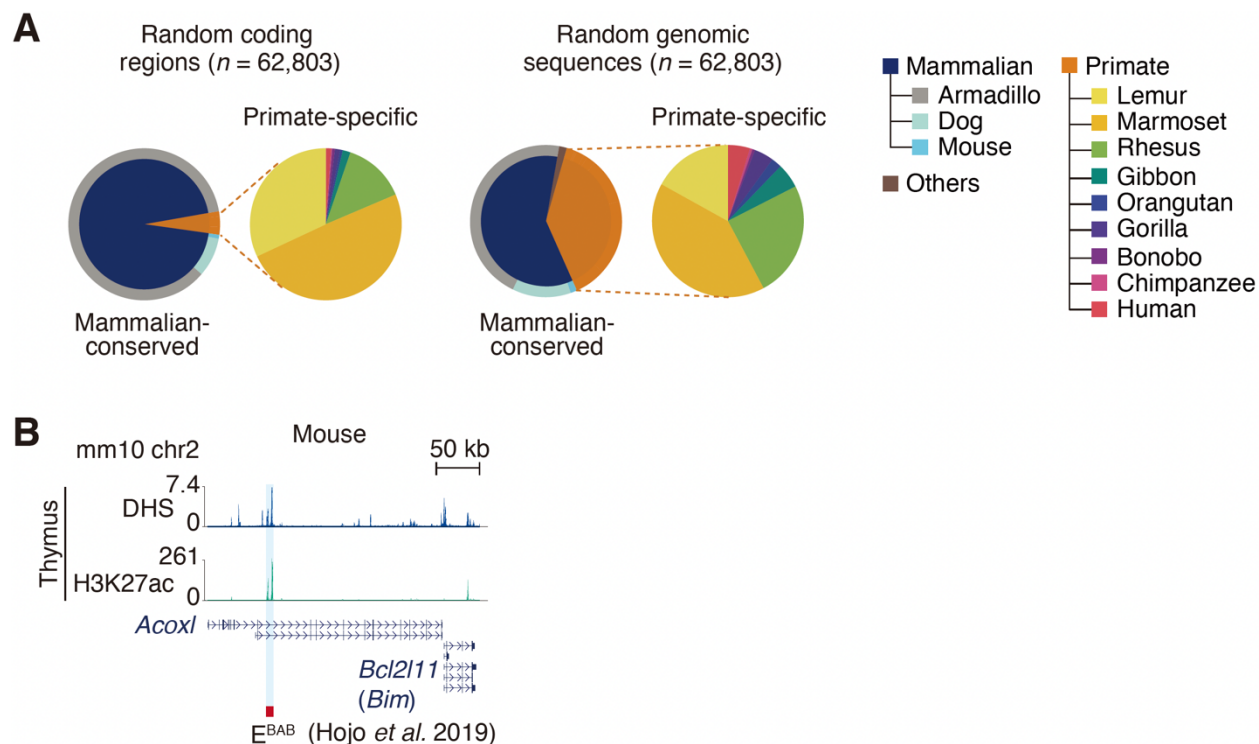
**Fig. S19. Human disease pathways.** (A) Metascape enrichment network. Visualization depicts the intra-cluster and inter-cluster similarities of enriched terms for the predicted target genes of disease-relevant btcEnhs. (B) Venn diagram showing target genes of multiple sclerosis (MS)-relevant btcEnhs and inflammatory bowel disease (IBD)-relevant btcEnhs predicted by Micro-C. (C and D) Metascape enrichment network for target genes of MS-relevant btcEnhs (C) and IBD-relevant btcEnhs (D) predicted by Micro-C, RCMC, ABC model or transcriptional co-variation.





**Fig. S20. CRISPR/Cas9-based target gene identification and validation.** Scatter plots of the global comparison of expression at btcEnhs ( $n = 62,803$ ) (left) and promoters ( $n = 90,565$ ) (right) between CRISPR activation (CRISPRa) and controls. Data for 10 disease-relevant btcEnhs targeted by CRISPRa are shown. Red dots, significant activation of btcEnhs (left) mediated by CRISPR and their target promoters (right) ( $\text{FDR} < 0.05$ ).





**Fig. S21. Evolution of CD4<sup>+</sup> T cell btcEnhs.** (A) Pie charts of the evolutionary conserved random protein-coding sequences and random genomic sequences. (B) UCSC Genome Browser view of the mouse *Acox1* and *Bcl2l11* loci corresponding to the human region shown in Fig. 6F. The genomic region containing the btcEnh is syntenic between human and mouse. In an enhancer knock-out mouse, this region has been shown to be a *Bcl2l11* (*Bim*) enhancer; it enforces proper apoptosis induction in thymic negative selection (82).

**Table S1. (separate file)**

Summary of single-cell experiments in this study.

**Table S2. (separate file)**

Summary of human subject details.

**Table S3. (separate file)**

List of signature genes.

**Table S4. (separate file)**

Seurat-based average protein expression levels.

**Table S5. (separate file)**

Motif enrichment in btcEnhs and dATAC peaks.

**Table S6. (separate file)**

Library QC of Micro-C and region Capture Micro-C.

**Table S7. (separate file)**

List of Micro-C loops.

**Table S8. (separate file)**

The results of LDSC analyses.

**Table S9. (separate file)**

List of disease-relevant btcEnhs overlapping with expression modulating variants.

**Table S10. (separate file)**

List of expression-modulating variants (emVar) within disease-relevant btcEnhs.

**Table S11. (separate file)**

*In silico* metagenesis on different categories of *cis*-regulatory elements.

**Table S12. (separate file)**

Summary of btcEnh-promoter co-variation.

**Table S13. (separate file)**

List of disease-relevant btcEnhs and their candidate target genes.

**Table S14. (separate file)**

Evolution of btcEnhs and ATAC peaks.

**Table S15. (separate file)**

Oligos for CRISPR activation experiments.

**Table S16. (separate file)**

List of ENCODE DHS bed files used for filtering btcEnhs.

**Table S17. (separate file)**

List of summary statistics and references for LDSC analysis.

**Data S1. (separate file)**

Promoter-level gene expression ( $\log_2\text{CPM}$ ) across 136  $\text{CD4}^+$  T cell clusters.

**Data S2. (separate file)**

RDS files of single-cell analyses.

**Data S3. (separate file)**

btcEnh expression levels ( $\log_2\text{CPM}$ ) across 136  $\text{CD4}^+$  T cell clusters.

**Data S4. (separate file)**

BED files of bidirectionally transcribed candidate enhancers (btcEnhs) and ATAC peaks.

**Data S5. (separate file)**

Expression matrix ( $\log_2\text{CPM}$ ) at robust TSS peaks across 136  $\text{CD4}^+$  T cell clusters.

**Data S6. (separate file)**

GTF file of *de novo* transcripts identified by 5' MAS-seq.

**Data S7. (separate file)**

Contact maps of Micro-C data.

**Data S8. (separate file)**

Results of Micro-C loops.

**Data S9. (separate file)**

Output files of CHiCAGO.

**Data S10. (separate file)**

Results of ABC models.