

## RESEARCH ARTICLE SUMMARY

## IMMUNOGENOMICS

# Single-cell eQTL mapping identifies cell type-specific genetic control of autoimmune disease

Seyhan Yazar†, Jose Alquicira-Hernandez†, Kristof Wing†, Anne Senabouth, M. Grace Gordon, Stacey Andersen, Qinyi Lu, Antonia Rowson, Thomas R. P. Taylor, Linda Clarke, Katia Maccora, Christine Chen, Anthony L. Cook, Chun Jimmie Ye, Kirsten A. Fairfax, Alex W. Hewitt\*,†, Joseph E. Powell\*†

**INTRODUCTION:** The human immune system has evolved to maintain tissue homeostasis and target exogenous pathogens by regulating specialized cell populations. It displays substantial variation between individuals, defining how people vary in susceptibility to disease and respond to pathogens or cancer.

**RATIONALE:** Our knowledge of how genetic differences contribute to immune variation at the cellular level has been limited by two main challenges in the generation of data at single-cell resolution. One of these challenges is to sequence from many individuals and the other is to sequence a large number of cells from each individual. Addressing these challenges is necessary to dissect the genetic and molecular underpinnings of common, heterogeneous diseases.

**RESULTS:** We present the OneK1K cohort, which consists of single-cell RNA sequencing (scRNA-

seq) data from 1.27 million peripheral blood mononuclear cells (PBMCs) collected from 982 donors. We developed a framework for the classification of individual cells, and by combining the scRNA-seq data with genotype data, we mapped the genetic effects on gene expression in each of 14 immune cell types and identified 26,597 independent cis-expression quantitative trait loci (eQTLs). We show that most of these have an allelic effect on gene expression that is cell type-specific. Our results replicated in two independent cohorts, one of which comprises individuals with a different ancestry to our discovery cohort. Over all loci, our discovery and replication cohorts have a concordance of allelic direction ranging from 72.2 to 98.1% across cell types.

Using the top associated eQTL single-nucleotide polymorphism (eSNP) at each locus outside the major histocompatibility complex (MHC) region, we identified 990 trans-acting effects, most

(63.6%) of which were cell type-specific. We show how eQTLs have dynamic allelic effects in B cells that are transitioning from naïve to memory states. Overall, we identified a set of 1988 eSNP-eGene (a gene with an eQTL) pairs expressed across the B cell maturation landscape, of which 333 have a statistically significant change in their allelic effect as B cells differentiate. Of these, 66% were only identified from the dynamic eQTL analysis and were not observed when testing for effects independently in cell types, highlighting the importance of investigating cell state-specific effects that underlie immune cell function. We investigated how eQTLs affect the expression variation of essential immune genes in specific cell types and provided experimental support for established hypotheses of cellular mechanisms in complex autoimmune diseases.

Finally, we integrated genetic association data for seven common autoimmune diseases and identified significant enrichment of genetic effects operating in a cell type-specific manner. Through colocalization of single-cell eQTL and genome-wide association study (GWAS) loci, we found that 19% of cis-eQTLs share the same causal locus as a GWAS risk association. Using a Mendelian randomization approach, we uncovered the causal route by which 305 loci contribute to autoimmune disease through changes in gene expression in specific cell types and subsets. Of the shared causal loci, 38.4% are outside the MHC region and exhibit highly cell-specific effects. Highlighting multiple sclerosis, we identified the causal route underlying 57 risk loci. For example, we show that the loci at 3q12 causally acts through changes in *EAF2* expression, but only in immature and naïve B ( $B_{IN}$ ) and memory B ( $B_{Mem}$ ) cells, despite this gene being ubiquitously expressed in all cell types in our data.

**CONCLUSION:** This work brings together population genetics and scRNA-seq data to uncover drivers of interindividual variation in the immune system. Our results demonstrate how segregating genetic variation influences the expression of genes that encode proteins involved in critical immune regulatory and signaling pathways in a cell type-specific manner. Understanding the genetic underpinnings of immune system regulation will have broad implications in the treatment of autoimmune diseases and infections, transplantation, and cancers. ■

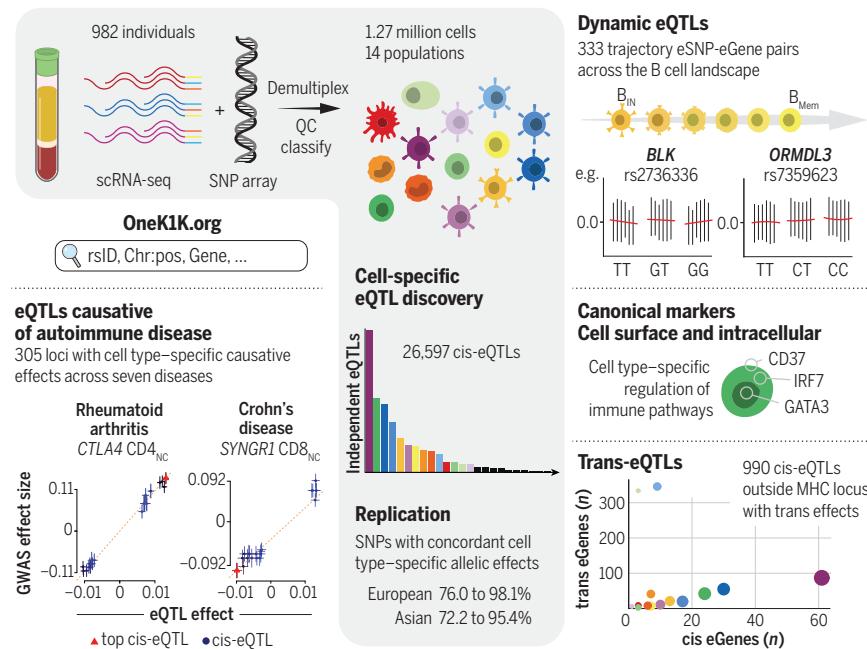
The list of author affiliations is available in the full article online.

\*Corresponding author. Email: hewitt.alex@gmail.com (A.W.H.); j.powell@garvan.org.au (J.E.P.)

†These authors contributed equally to this work.

Cite this article as S. Yazar et al., *Science* **376**, eabf3041 (2022). DOI: 10.1126/science.abf3041

**READ THE FULL ARTICLE AT**  
**S** <https://doi.org/10.1126/science.abf3041>



**Single-cell eQTL mapping and colocalization with autoimmune disease risk loci.** scRNA-seq data from 1.27 million PBMCs were used to identify 26,597 cis-eQTLs (gray box). Dynamic eQTLs were uncovered as cells move from a naïve to a memory state (top right). Genetic variation between individuals influences immune regulation in a cell type-specific manner (middle right). In this study, 990 trans-eQTL effects (bottom right) and the causal effects for 305 autoimmune disease loci were identified (bottom left). Browseable results are available at [www.onek1k.org](http://www.onek1k.org). CD4<sub>NC</sub>, CD4 naïve and central memory T cells; CD8<sub>NC</sub>, CD8 naïve and central memory T cells; QC, quality control.

## RESEARCH ARTICLE

## IMMUNOGENOMICS

# Single-cell eQTL mapping identifies cell type-specific genetic control of autoimmune disease

Seyhan Yazar<sup>1†</sup>, Jose Alquicira-Hernandez<sup>1,2†</sup>, Kristof Wing<sup>3,4†</sup>, Anne Senabouth<sup>1</sup>, M. Grace Gordon<sup>5,6,7</sup>, Stacey Andersen<sup>2</sup>, Qinyi Lu<sup>3</sup>, Antonia Rowson<sup>3,8</sup>, Thomas R. P. Taylor<sup>3</sup>, Linda Clarke<sup>9</sup>, Katia Maccora<sup>3,8</sup>, Christine Chen<sup>8</sup>, Anthony L. Cook<sup>10</sup>, Chun Jimmie Ye<sup>5,6,7,11,12,13</sup>, Kirsten A. Fairfax<sup>3</sup>, Alex W. Hewitt<sup>3,4,9\*,†</sup>, Joseph E. Powell<sup>1,14\*,†</sup>

The human immune system displays substantial variation between individuals, leading to differences in susceptibility to autoimmune disease. We present single-cell RNA sequencing (scRNA-seq) data from 1,267,758 peripheral blood mononuclear cells from 982 healthy human subjects. For 14 cell types, we identified 26,597 independent cis-expression quantitative trait loci (eQTLs) and 990 trans-eQTLs, with most showing cell type-specific effects on gene expression. We subsequently show how eQTLs have dynamic allelic effects in B cells that are transitioning from naïve to memory states and demonstrate how commonly segregating alleles lead to interindividual variation in immune function. Finally, using a Mendelian randomization approach, we identify the causal route by which 305 risk loci contribute to autoimmune disease at the cellular level. This work brings together genetic epidemiology with scRNA-seq to uncover drivers of interindividual variation in the immune system.

**T**he expression of genes in immune cells is highly variable between individuals (1–6), with this variation being both a cause and a consequence of differences in susceptibility to immune-related diseases (7, 8). Investigations into the underlying genetic contribution to immune regulation and disease development have uncovered many associated variants (9). Yet the complexity of circulating immune populations has made their mechanisms of action difficult to dissect.

Coupling transcriptional profiles with genetic variation allows the direct identification of genomic regulators of gene expression. This is important because disease-associated genetic risk variants identified through genome-wide

association studies (GWASs), including those linked to common immune-mediated diseases, are often mapped to regulatory regions of the genome (10–13). Both empirical results and theoretical models provide evidence that most common disease-associated variants act through changes in gene expression rather than directly influencing protein structure or function (14). By combining genetic information with bulk RNA sequencing (RNA-seq), the downstream effects of disease-associated genetic risk factors have been linked to expression quantitative trait loci (eQTLs). Efforts such as GTEx (15), eQTL-Gen (16), CAGE (17), and ImmVar (18) have identified eQTLs across a variety of cell types and tissues but have used bulk RNA-seq approaches, where gene expression levels represent the averaged signal over large numbers of cells. The data from these ensemble analyses are valid, but the gene expression heterogeneity between individual cells is still largely unexplored.

An important step is to define the cellular and environmental contexts in which disease-risk single-nucleotide polymorphisms (SNPs) affect gene expression levels. This will help determine the molecular and cellular mechanisms by which disease develops and inform therapeutic strategies. Beyond the ability to annotate individual disease associations, cell type-specific eQTLs are enriched for heritability across complex traits (4). This is important because many eQTL effects are tissue specific (2, 18, 19), and both fluorescence-activated cell sorting (FACS) and computational deconvolution of cell types from bulk samples have evidenced cell type-specific eQTLs (20–22). Although these studies have helped demon-

strate variation in the role of genetic loci in cell subsets, challenges remain. For example, bulk RNA-seq of FACS cell populations is biased toward known cell types that are defined by a limited set of marker genes. It does not capture the heterogeneity within a sorted population. Likewise, computational methods that deconvolve a bulk signal into cell types struggle to identify less abundant cell types and rely on approximations to estimate cell proportions (23). By contrast, single-cell RNA sequencing (scRNA-seq) enables the simultaneous, unbiased determination of cellular composition and cell type-specific gene expression, capturing intraindividual cell heterogeneity.

## Results

### The OneK1K cohort

We characterized the transcriptional variation across circulating immune cells of a large cohort (OneK1K) to explore how allelic variation is associated with changes in gene expression in a cell type-specific manner (Fig. 1A). The OneK1K cohort consists of 982 individuals of Northern European ancestry (Fig. 1B) who reported no active infection at the time of sample collection. We generated genotype data on 759,993 SNPs (figs. S1 to S3) and imputed SNPs against the Haplotype Reference Consortium panel (24). After quality control, we retained 5,328,917 SNPs with a minor allele frequency greater than 0.05 (fig. S4). We generated scRNA-seq data on 1,449,385 peripheral blood mononuclear cells (PBMCs) using a pooled multiplexing strategy. After demultiplexing, removal of doublets, and quality control, we retained 1,267,758 cells for further analysis (Fig. 1C).

### Classification of individual cells

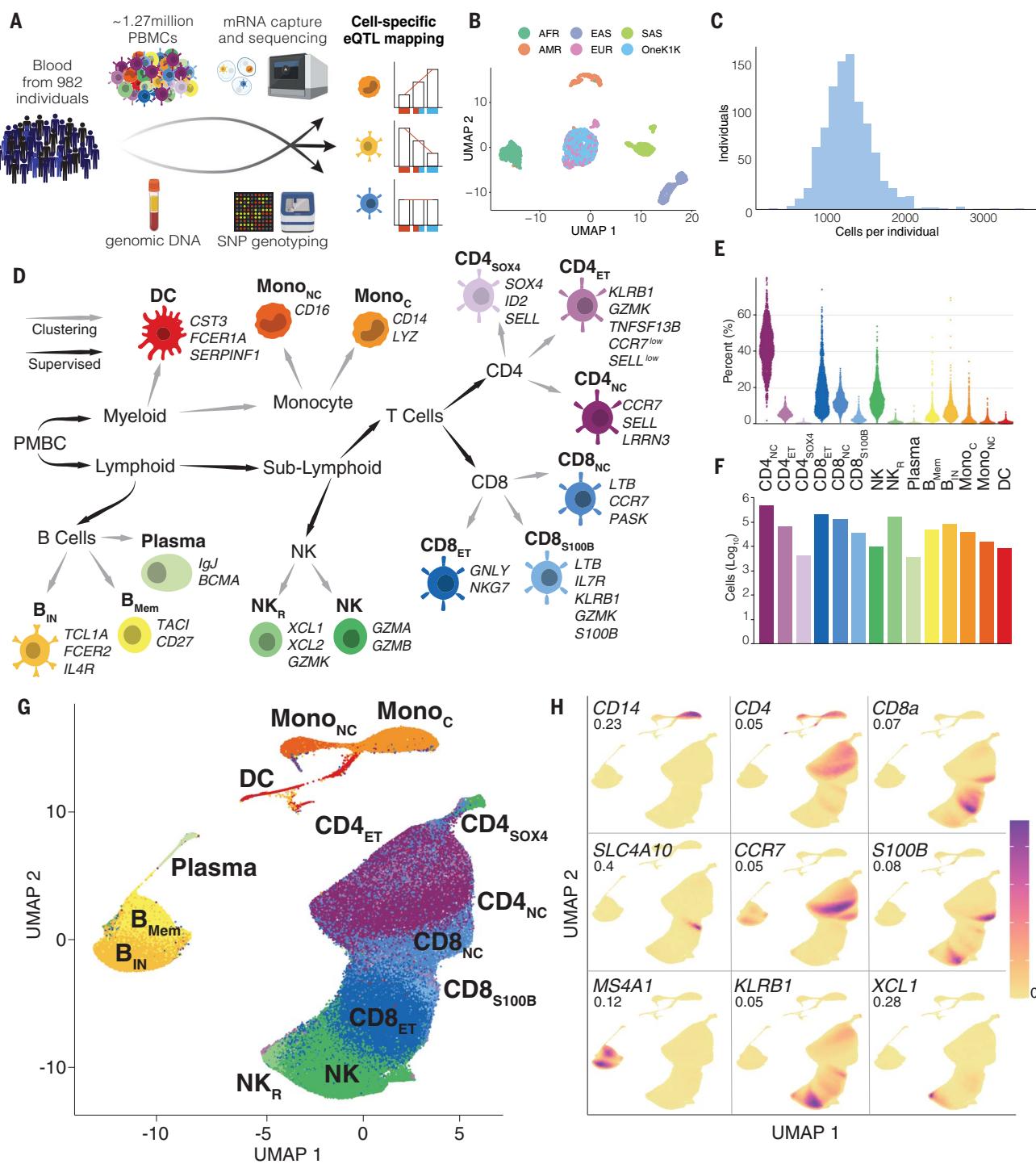
We developed a framework to independently classify each cell into one of 14 different immune cell types across the myeloid and lymphoid lineages based on their transcriptional profiles. This framework, implemented in *scPred* (25), uses a combination of hierarchical supervised and unsupervised classification methods, using FACS-sorted PBMC scRNA-seq data as a reference (26) (Fig. 1D and table S1). Cell composition ranged from 0.7% dendritic cells (DCs) to 36.6% CD4<sup>+</sup> naïve and central memory T (CD4<sub>NC</sub>) cells (Fig. 1E), with the mean and range of proportions matching those reported elsewhere (27, 28) (Fig. 1F and table S2). Visualization of cell types with uniform manifold approximation and projection (UMAP) reflects the hierarchical relationship among these cell types (Fig. 1G), which is also supported by cell coordinates across the first two principal components (fig. S5). Cells were classified using their complete transcriptional profiles. Still, to aid interpretation against other studies, we show concordance with the expression patterns of canonical markers and other

<sup>1</sup>Garvan-Weizmann Centre for Cellular Genomics, Garvan Institute of Medical Research, Sydney, NSW, Australia.

<sup>2</sup>Institute for Molecular Bioscience, University of Queensland, Brisbane, QLD, Australia. <sup>3</sup>Menzies Institute for Medical Research, University of Tasmania, Hobart, TAS, Australia.

<sup>4</sup>Department of Ophthalmology, Royal Hobart Hospital, Hobart, TAS, Australia. <sup>5</sup>Division of Rheumatology, Department of Medicine, University of California, San Francisco, San Francisco, CA, USA. <sup>6</sup>Department of Epidemiology and Biostatistics, University of California, San Francisco, San Francisco, CA, USA. <sup>7</sup>Institute for Human Genetics, University of California, San Francisco, San Francisco, CA, USA. <sup>8</sup>Department of Surgery, School of Clinical Science at Monash Health, Monash University, VIC, Australia. <sup>9</sup>Centre for Eye Research Australia, University of Melbourne, East Melbourne, VIC, Australia. <sup>10</sup>Wicking Dementia Research and Education Centre, University of Tasmania, Hobart, TAS, Australia. <sup>11</sup>Institute of Computational Health Sciences, University of California, San Francisco, San Francisco, CA, USA. <sup>12</sup>Parker Institute for Cancer Immunotherapy, San Francisco, CA, USA. <sup>13</sup>Chan Zuckerberg Biohub, San Francisco, CA, USA. <sup>14</sup>UNSW Cellular Genomics Futures Institute, University of New South Wales, Sydney, NSW, Australia.

\*Corresponding author. Email: hewitt.alex@gmail.com (A.W.H.); j.powell@garvan.org.au (J.E.P.)  
†These authors contributed equally to this work.



**Fig. 1. Population-scale scRNA-seq identifies 14 transcriptionally distinct mononuclear populations in peripheral blood.** (A) scRNA-seq data from PMBCs were generated using a pooled multiplexing strategy for 982 healthy individuals. Simultaneously, SNPs were genotyped, and data were integrated for single-cell eQTL analysis. (B) UMAP analysis shows the genetic relationship between the individuals from the OneK1K cohort and the 1000 Genomes Project (83). Individuals from the OneK1K cohort are embedded with individuals with Northern European genetic ancestry. AFR, African; AMR, ad-mixed American; EAS, East Asian; EUR, European; SAS, South Asian. (C) Mean of 1291 individual cells per donor, ranging from 62 to 3501 after scRNA-seq, demultiplexing, and quality-control filtering. (D) Hierarchical classification of cells. Each cell underwent up to four rounds of supervised clustering based on similarity to each node, as indicated by the black arrows. After that,

unsupervised clustering by Seurat (gray arrows) yielded 14 transcriptionally distinct cell types. Classification of each cell was confirmed based on cosine similarity to FACS reference data and further assessed through the interrogation of differentially expressed and prototypical genes. (E) Total percentage of each cell type as a proportion of the total sequenced population for each individual. (F) The total number of cells per cell type after sequencing, demultiplexing, and quality-control filtering. (G) UMAP of 1,267,758 PMBCs across all individuals, with 14 transcriptionally distinct populations. Color coding is the same as in (D). (H) Density plots of nine differentially expressed canonical markers of peripheral immune cells, demonstrating robust concordance with canonical markers (see fig S11 for additional markers). Values denote maximal density. The abbreviations for each cell type are displayed in table S1. The color scale is relative, ranging from no density (0) to highest density.

single-cell sequencing studies (26, 29, 30) (Fig. 1H).

After batch correction, we found no evidence for variation in cell identity, transcriptional signatures, or cell proportions across the capture pools (figs. S6 to S8). Across individuals, we sequenced an average of 1291 cells per donor (Fig. 1C). Although most of the individuals had scRNA-seq data for all 14 cell types, because of sampling variance, some cell types [predominantly CD4<sup>+</sup> T cells expressing SOX4 (CD4<sub>SOX4</sub> cells), plasma cells, and nonclassical monocytes (Mono<sub>NC</sub>)] were not sequenced for some individuals (fig. S9 and table S7). Therefore, for subsequent analyses, the sample size for eQTL analysis varied by cell type, although 12 out of the 14 populations had  $n > 930$ .

#### Single-cell eQTL analysis reveals cell-type specificity of transcriptional changes that occur because of common variants

To understand how genetic variation between individuals influences gene expression in a cell type-specific manner, we tested for the association between the genotypes of SNPs within a 1-Mb cis region of either end of a gene including the gene body and the expression of genes in each of the 14 cell types. This approach identifies eQTLs in each cell type, enabling us to assess the degree to which the genetic effects on gene expression are shared across PBMCs. Multiple SNPs within a cis region can be associated with gene expression because of the correlation between genotypes induced by linkage disequilibrium and numerous independent loci associated with the expression levels of the gene. To differentiate between these scenarios, we performed a conditional analysis for each identified eQTL, fitting the lead eQTL SNP(s) [eSNP(s)] as conditional covariates in subsequent rounds of analysis.

In total, we identified 26,597 eQTLs for 39.7% of the genes tested, with 16,597 (eSNP<sub>1</sub>) in the first round of analysis and a further 10,000 (eSNP<sub>2</sub> to eSNP<sub>5</sub>) from the four rounds of conditional tests (Fig. 2A and tables S9 and S10). The number of independent eQTLs varied between cell types, with 6473 identified in CD4<sub>NC</sub> cells and 399 in plasma cells (Fig. 2B). This variation in the number of eQTLs determined per cell type is likely a function of statistical power. There is a strong relationship between both cell proportions (Fig. 1E and fig. S17) and the number of individuals with identifiable cells (table S7). The conditional eQTL analysis identified secondary loci influencing expression in 8.1 to 19.2% of genes with an initial eQTL and more than three independent eQTLs for 10.6 to 40.6% of genes (Fig. 2A and table S9).

These conditional eQTLs identify instances where there are multiple independent loci within the cis region whose genotypes are associated with the expression levels of a gene.

For example, in CD4<sub>NC</sub> cells, we identified a primary eQTL for *PADI4*. This gene encodes an enzyme that is responsible for converting arginine residues to citrulline residues (31), thereby regulating the activity of histone H1 and consequently the maintenance of stem cells (32). *PADI4* has been implicated in the pathogenesis of rheumatoid arthritis (RA) at both a genetic and cellular level (33). The top eSNP<sub>1</sub> for this eQTL is rs10788663, where each copy of the T allele causes a decrease of an average of 0.28 mRNA transcript molecules per cell (fig. S12). In a subsequent round of conditional analysis, we fitted rs10788663 as a covariate and tested for associations again across the cis region, identifying a secondary independent eQTL marked by the top eSNP<sub>2</sub>, rs1612843. On average, individuals carrying each copy of the C allele of rs1612843 have a decrease of 0.24 mRNA transcript molecules per cell. rs10788663 is located in the first intron, whereas rs1612843 is located in the intron between exons 15 and 16 of *PADI4*, suggesting that independent transcription factors likely regulate multiple independent sites and are required for the regulation of the expression of *PADI4*. In the OneK1K cohort, the linkage disequilibrium between rs10788663 and rs1612843 is 0.0678, providing further evidence that multiple independent eQTLs influence the expression of *PADI4* in CD4<sub>NC</sub> cells. Indeed, after confirming the expected additive effect of two independent loci, we observed a mean difference of 1.04 mRNA transcripts per cell for individuals carrying homozygous T/T and C/C compared with C/C and G/G for rs10788663 and rs1612843, respectively (fig. S12). Both rs10788663 and rs1612843 associations were replicated in eQTL-Gen data (34).

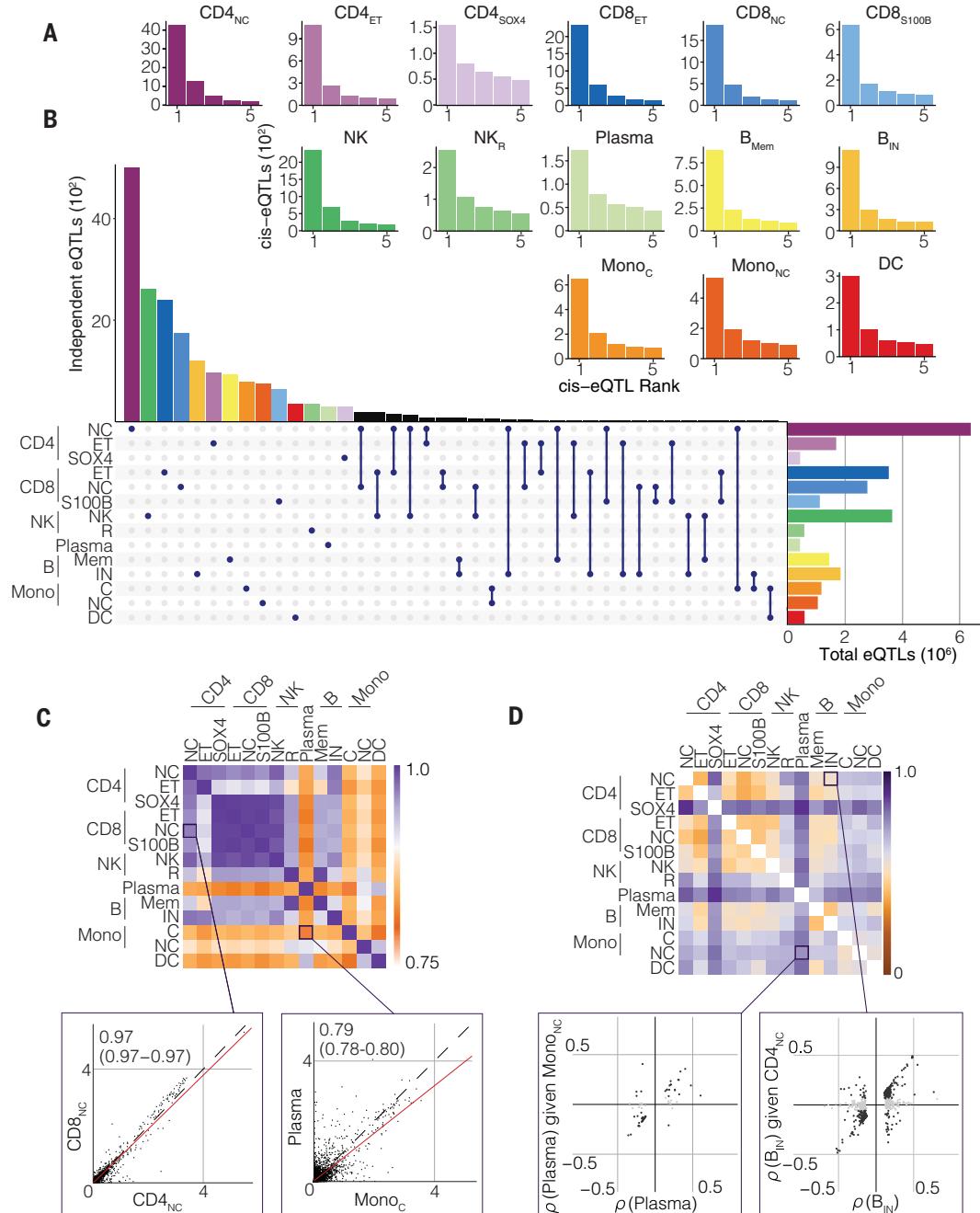
The allelic effect of genetic loci on gene expression may be distinctive to a particular cell type and absent in other cell types—a relationship we define as “cell type-specific.” We explored its prevalence by investigating the deviation of test statistics from a null distribution for cis-eQTLs in other cell types where they did not initially meet study-wide significance (Fig. 2B). The mean proportion of cis-eQTLs identified in one cell type that showed inflation of their test statistics in one other cell type was  $\pi_1 = 0.53$  (0.19 to 0.96) (fig. S13). This is evidence that with larger sample sizes, cis-eQTLs currently identified in a single cell type should reach study-wide significance in one or more other cell types. However, the magnitude of their allelic effect is likely to vary between cell types. For 3060 genes with an eQTL (eGenes) identified in only a single cell type, we do not find any evidence for allelic effects in other cell types, suggesting that these are indeed cell type-specific (fig. S14). The observation of cell type-specific eQTLs has multiple possible explanations: The gene may only be detectably expressed in one cell type,

there may be low statistical power to detect eQTLs in multiple cell types, or there is true regulatory heterogeneity across cell types.

To evaluate these different scenarios, we performed a series of analyses for each of the genes with at least one eQTL (eGene  $n = 6469$ ). Only 43 (0.7%) of these eGenes are expressed in a single cell type. The remaining 6426 are expressed in multiple cell types, with these genes expressed in an average of 11 cell types, in addition to the one with a significant eQTL (fig S15). Indeed, when we tested for the correlation in the expression levels of each of these 6426 eGenes between a pair of cell types, we identified a high overall concordance in co-expression (Fig. 2C). The pattern of average correlation in eGene expression levels between a couple of cell types followed the hematopoietic lineage relationship. For example, of the 6473 eGenes with an eQTL found only in CD4<sub>NC</sub> cells, 1392 were expressed in CD8<sup>+</sup> naïve and central memory T (CD8<sub>NC</sub>) cells and the mean correlation in gene expression between the cells was 0.97 (Fig. 2C). By contrast, in classical monocytes (Mono<sub>C</sub>), only 168 of the plasma cell eGenes were expressed, but the mean correlation of expression with plasma cells was 0.79. From these results, we can conclude that most of the eGenes with an eQTL identified in just one cell type are not due to cell type-specific expression of the eGene in most instances but rather may be due to cell type-specific expression of regulatory factors.

Having identified that these eGenes are expressed in multiple cell types, we next sought to evaluate if the observation of cell type-specific eQTLs was due to low statistical power to detect allelic effects in more than one cell type. To assess this hypothesis, we implemented an empirical framework to test the rank of the test statistics for eGene allelic effects across the nonsignificant cell types. In almost all instances, we observed none or minimal enrichment of the test statistic across cell types, suggesting that in most cases, cell type-specific eQTLs are due to specific cell regulatory mechanisms (fig. S15). In instances where we identified a marked enrichment, cell types closely related in the hematopoietic lineage existed. However, for most eGenes, we did not identify an enrichment in the test statistics, again suggesting that effects are cell type-specific. These results collectively demonstrate that most of the eQTLs identified for the 2367 eGenes are specific to just a single cell type.

For the remaining 4102 eGenes, we identified a total of 14,230 eQTLs across two or more cell types, although, for 1386 of these eGenes, we observed different lead eSNPs between cell types (Fig. 2B). Under this scenario, one hypothesis is that the same variant underlies eQTLs in multiple cell types, with differences in top eSNPs being due to variation in gene expression patterns. An alternative hypothesis is



**Fig. 2. Conditional eQTL analysis reveals cell-type specificity of transcriptional changes due to common variants.** **(A)** Up to five rounds of conditional eQTL analysis (x axis) are shown, with the number of cis-eQTLs (y-axis value times 10<sup>2</sup>) detected at a study-wide FDR less than 0.05 in each round of analysis. **(B)** Using eSNP<sub>1</sub> to eSNP<sub>5</sub>, we identify most of the eQTLs in only a single cell type. eQTLs identified in multiple cell types are connected by lines. The x axis is truncated at 30 eQTLs in a given category. The total number of eQTLs detected per cell type is shown on the right. **(C)** For each eQTL identified in a single cell type, we tested whether it was present in other cell types and compared eGene expression levels. Each unit of the heatmap (top) shows the correlation of expression levels of genes associated with an eQTL only in cell type A against cell type B. Cell type-specific eQTL genes remain consistently expressed across other cell types, with examples shown for the expression levels of CD4<sub>NC</sub> eQTL genes in CD8<sub>NC</sub> cells (bottom left).

and for  $\text{Mono}_C$  eQTL genes in plasma cells (bottom right), with  $x$ - and  $y$ -axis units representing mean UMIs per cell. The scale of the correlation coefficients between pairwise combinations of cells ranges from 0.75 to 1.0. (D) To investigate the independence of eQTLs for genes with eQTLs in more than one cell type (but tagged by different eSNPs), we tested for the change in allelic effect in cell type A, after conditioning on the eSNP from cell type B. Significant changes ( $p < 0.05$ ) imply the same eQTL in both cell types (or linkage disequilibrium between eSNPs). A lack of change provides evidence that the gene has independent eQTLs in each cell type. For example, the allelic effects of eSNPs from plasma cells after conditioning on the lead eSNPs from  $\text{Mono}_{NC}$  cells (bottom left) and of  $B_{IN}$  eSNPs after conditioning on the lead eSNPs from  $CD4_{NC}$  (bottom right) cells are shown. The heatmap (top) shows the pairwise correlations in allelic effects.  $\rho$  represents the original correlation coefficient. The abbreviations for each cell type are displayed in table S1.

that the eQTLs result from independent variants that influence expression in different cell types. To test between these hypotheses, we performed a regression strategy to evaluate the change in the test statistic of an eSNP after regressing out the effects of an eSNP from another cell type. Under this strategy, if the eSNPs tag the same causal variant for that gene or are in linkage disequilibrium with one another, then the allelic effect size of the original eSNP will decrease in the conditional analysis. Similarly, if they tag independent variants, the allelic effect will remain relatively unchanged. We performed this strategy for each pairwise combination of eQTLs where different top eSNPs were identified in different cell types.

We tested whether each eGene was tagged by two distinct variants by conditioning the lead eSNP from the first cell type on the lead eSNP from the second cell type for every pair of cell types (182 pairs). The correlation coefficients of significant independent eSNPs from shared eGenes pre- and postconditioning are shown in Fig. 2D and fig. S16. Whereas most lymphoid immune cell eQTLs had a considerable change in the correlation coefficients after conditioning, among the myeloid immune cells, the eQTL correlation coefficients remained similar (fig. S16). This finding suggests that lymphoid cell types are more likely to share genetic control of gene expression between cell types compared with myeloid cells.

#### **Evidence suggests that cell type–specific chromatin accessibility underlies a proportion of cell type–specific cis-eQTLs**

To explore the functional regulation underlying cis-eQTLs, we tested for the overlap of eSNP locations and regions of open chromatin generated from single-cell assay for transposase accessible chromatin sequencing (scATAC-seq) data from 8876 cells. Cells were classified into each of the 14 cell types, and open chromatin peaks were called for each cell type that had more than five classified cells. This filtering retained 11 cell types, comprising the most abundant populations [except CD4<sup>+</sup> T cells with an effector memory or central memory phenotype (CD4<sub>ET</sub>), CD4<sub>SOX4</sub>, and plasma cells] (fig. S18). On average, we identified 52,048 peaks per cell type, with the mean distance between an eSNP and the nearest peak ranging from 7485 to 31,383 base pairs. To determine whether the location of cis-eQTLs was significantly closer to open chromatin regions, we compared the distances between the cis-eQTLs. We randomly sampled SNPs that were selected based on the same distance distribution from the transcript to the nearest peaks per cell type using a bootstrapping technique. We observed a significant difference between the cis-eQTL distances across all cell types except CD4<sub>SOX4</sub> cells [false discovery rate (FDR) < 0.05] (fig. S19). We conclude from these results that cell

type-specific chromatin accessibility is likely to contribute to variation in allelic effects on gene expression between cell types.

#### **Single-cell eQTLs replicate in multiethnic cohorts and bulk eQTL studies**

To verify cell-specific eQTL findings, we replicated our lead eSNP results in two independent cohorts of European and Asian ancestry, consisting of 113 and 89 individuals, respectively. Of the 16,597 eSNP-eGene pairs, 10,071 were present with a minor allele frequency greater than 0.05 in both cohorts. Of these, 3198 (26%) in the European cohort and 2243 (22%) in the Asian cohort replicated at the FDR threshold of 5%, which is encouraging given the differences between the sample sizes of these cohorts and the sample size of the OneK1K discovery cohort (tables S5, S12, and S13).

Indeed, correcting the FDR distributions under the assumptions of equal sample size in the discovery and replication cohorts leads to 87 and 78% replication rates in the European and Asian cohorts, respectively. Similarly, the concordance of allelic direction over all tested loci was 76.0 to 98.1% in the European cohort and 72.2 to 95.4% in the Asian cohort. This concordance increases to 99.3 to 100% and 96.9 to 99.8%, respectively, for eQTLs replicating at an FDR less than 0.05 (Fig. 3, A and B). The discrepancy in replication rates between cohorts likely reflects differences in the allele frequencies of eSNPs between population groups. However, the results indicate that cell type–specific eQTLs are likely to be largely shared among populations. The discovery of OneK1K eQTLs was tested for replication in all cell types in the replication cohorts. At an FDR less than 0.05, replicating eQTLs and eGenes are predominantly identified in a single cell type (Fig. 3, C and D), providing further evidence for cell type–specific effects of loci on gene expression in PBMCs. The concordance of correlation coefficients between the OneK1K and replication cohorts are shown in Fig. 3E for both the European and Asian samples. We were able to replicate 62.5 and 40.4% of cis-eQTLs identified in bulk RNA-seq studies of blood samples from the eQTL-Gen Consortium (34) and GTEx Consortium (5), respectively (fig. S20 and table S14).

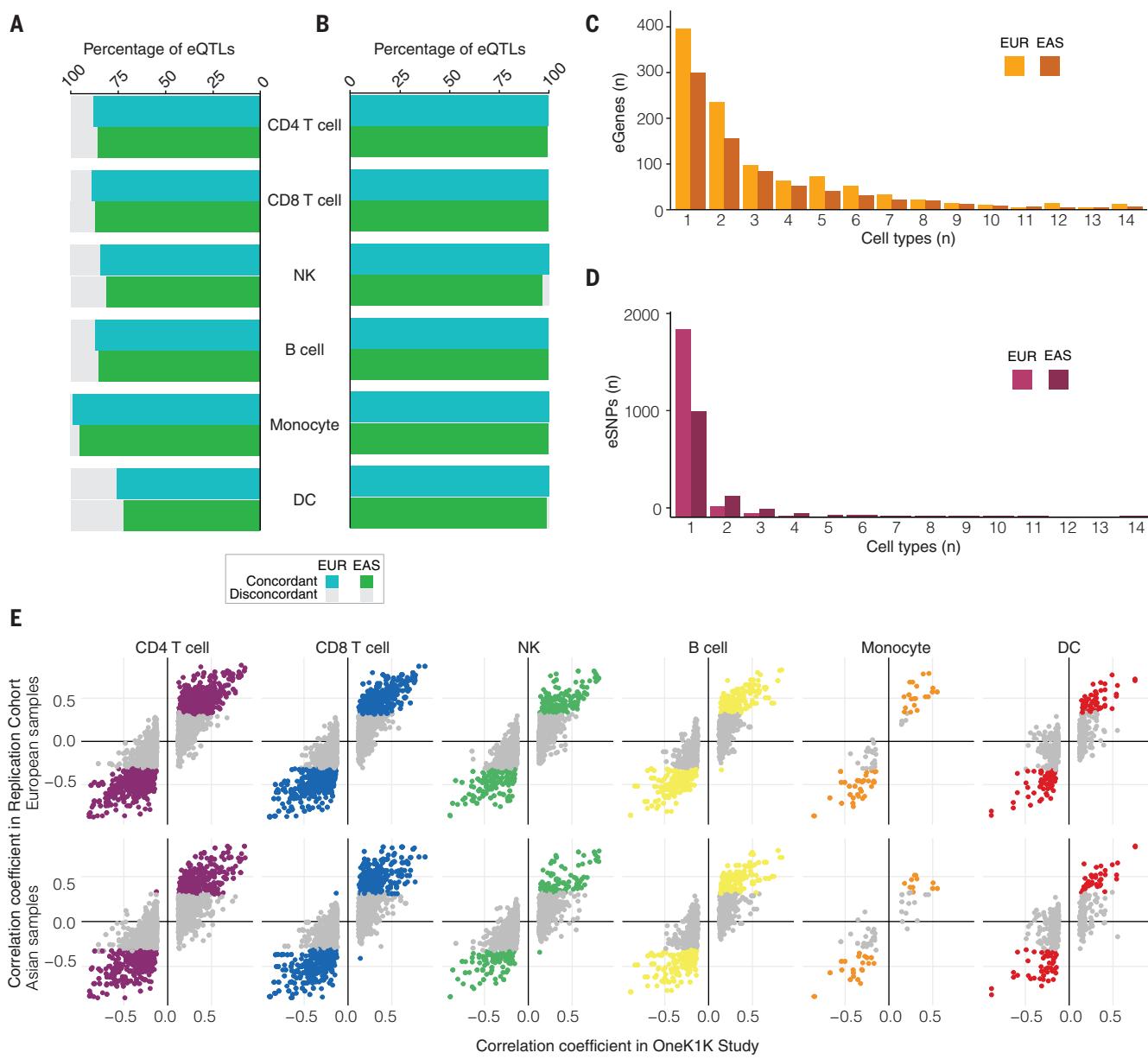
#### **Identification of dynamic eQTL allelic effects across the B cell landscape**

We investigated the dynamic effects of eQTLs across the pseudotime landscape of immature and naïve B ( $B_{IN}$ ) cells through to memory B ( $B_{Mem}$ ) cells. Cells were categorized into six quantiles (Q1 to Q6) based on their relative position on the pseudotime curve (Fig. 4, A and B). Overlaying the expression of classical markers revealed a graded change across the derived trajectory from  $B_{IN}$  (Q1) to  $B_{Mem}$  cells (Q6). For example, *TCL1A* and *IL4R* are highly

expressed in naïve B cells (35, 36) and were found to be down-regulated across the transition to  $B_{Mem}$  cells (Fig. 4C). Conversely, the expression of *CD27*, a canonical  $B_{Mem}$  cell marker (37), increased as the cells transitioned to a memory state. *IgJ* expression, a marker of immunoglobulin M (IgM) and IgA production, was up-regulated in the higher quantiles, indicating that they contain cells poised to become plasma cells (38) (Fig. 4C).

We sought to identify instances where eQTL allelic effects exhibited either linear or nonlinear changes across the trajectory of naïve to memory B cell transition. Dynamic B cell eQTLs were determined by testing the interaction between the genotype and quantile ranks using both linear and quadratic models. Of the 3074 cis-eQTLs identified in  $B_{IN}$  and  $B_{Mem}$  cells, 1988 were expressed in at least three pseudotime quantiles and tested for dynamic effects. Of these, we identified significant changes in the allelic effect across the trajectory for 333 of them (FDR < 0.05) (Fig. 4D and fig. S21).

Many of the genes with dynamic eQTL effects have a role in fine-tuning B cell migration, activation, survival, or function. For example, *SELL* is involved in integrin-mediated migration to and within tissues (39, 40). Migration to and organization of B cells within the germinal center is a critical component in generating appropriate memory and humoral outputs. The allelic effect of the intronic variant rs4987360-G on *SELL* expression is largest in immature cells, decreasing over each of the subsequent quantiles (Fig. 4E). The opposite trend is identified for SNPs that influence the expression of the Src family tyrosine kinase B lymphocyte kinase (*BLK*), a gene responsible for regulating the amplitude of signaling downstream of the B cell receptor. Both rs2736336 and rs2409780 show the greatest allelic effects in Q5 and Q6 (Fig. 4E and table S15). Interestingly, rs2736336, a variant in the promoter of *BLK*, is associated with systemic lupus erythematosus (SLE) (41), whereas rs2409780, an intronic variant, is in high linkage disequilibrium with variants associated with SLE and RA [coefficient of determination ( $R^2$ ) = 0.99, and coefficient of linkage disequilibrium ( $D'$ ) = 0.99] (13, 42). Another gene responsible for interpreting the signaling downstream of B cell surface receptors and influencing subsequent B cell proliferation and survival is c-Rel, encoded by the transcription factor *REL* (43). rs12989427 is in high linkage disequilibrium with variants associated with SLE ( $R^2$  = 0.88, and  $D'$  = 0.98), and the allelic effect follows a nonlinear relationship, peaking at the medium point of the B cell trajectory (Fig. 4E). *ORMDL3* promotes mature B cell survival by suppressing apoptosis and promoting autophagy (44). rs7359623 and rs8067378 are in high linkage disequilibrium with risk variants ( $R^2$  > 0.8, and  $D'$  > 0.9) implicated in a range of autoimmune



**Fig. 3. Replication cohort shows concordance of eQTLs across European and Asian samples.** (A) Percentage of concordance of the allelic direction of effect for all single-cell eQTLs tested for replication across cell types. Concordances for single-cell RNA of European (EUR) and Asian (EAS) populations are shown. (B) Percentage of concordance for eQTLs that replicate at a study-wide FDR significance threshold of 0.05. (C) The numbers of cell types in which the eGene is observed for eQTLs that replicate in the European or Asian cohorts. (D) The number of cell types in which eSNPs are identified in the replication cohorts. Most replicating eQTLs are cell type-specific. (E) The correlation coefficient for eQTLs in the OneK1K and replication cohorts per cell type indicated. The direction of the correlation coefficients denotes the direction of the allelic effect with respect to the reference allele. European samples are shown in the first row, and Asian samples are shown in the second row. Colored dots denote eQTLs that replicate at a study-wide FDR less than 0.05.

diseases (45–48) and have a dynamic eQTL effect on *ORMDL3* in B cells across the trajectory.

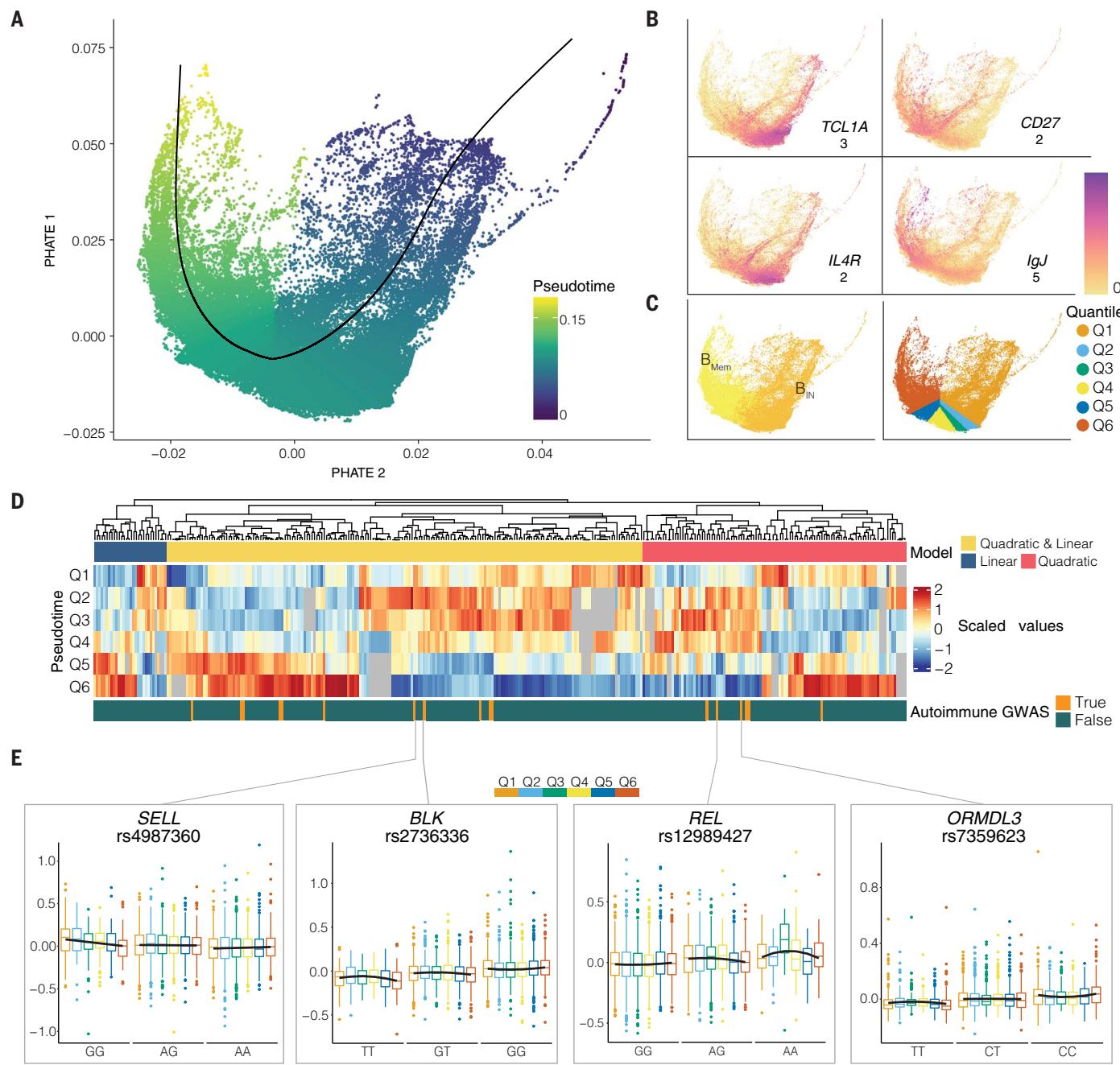
#### Genetic variation controls transcriptional regulation in a cell type-specific manner to regulate immune pathways

Although it is widely accepted that immune regulation is variable between individuals (1), the factors that cause this variation are poorly understood. By selecting genes described in the literature that affect immune regulation, we

demonstrate how genetic loci contribute to variation in the expression of immune regulatory genes in a cell type-specific manner (Fig. 5 and table S10).

Leukocyte recirculation between the blood and lymph nodes is an essential property of the immune system. It depends on the lymph node homing receptor CD62L (L-selectin) encoded by the *SELL* gene (49). We observed opposing regulation of *SELL* mRNA between the innate and adaptive immune systems under the influence of rs4987360, a common polymorphism in linkage disequilibrium with rs4987353 ( $R^2 = 1$ , and  $D' = 1$ ) that is associated with monocyte blood cell counts (50). The rs4987360-G allele decreased *SELL* mRNA in Mono<sub>c</sub> but increased *SELL* mRNA in B<sub>IN</sub> cells (Fig. 5A), illustrating how a single inherited allele can act through different cell types to influence gene expression. The dynamic eQTL analysis identified that the allelic effect of rs4987360 varied across the B cell-state landscape (Fig. 4E).

number of cell types in which eSNPs are identified in the replication cohorts. Most replicating eQTLs are cell type-specific. (E) The correlation coefficient for eQTLs in the OneK1K and replication cohorts per cell type indicated. The direction of the correlation coefficients denotes the direction of the allelic effect with respect to the reference allele. European samples are shown in the first row, and Asian samples are shown in the second row. Colored dots denote eQTLs that replicate at a study-wide FDR less than 0.05.



**Fig. 4. Dynamic eQTLs across B cell trajectories.** (A) The pseudotime projection of 124,968 B cells was derived from their progression from immature or naïve to memory cells. The pseudotime curve is represented by the solid black line. The pseudotime is represented with a color scale from 0 (the earliest pseudotime) to 1 (the latest pseudotime). (B) Mapping of  $B_{IN}$  and  $B_{Mem}$  cells and the division of landscape into six quantiles across the pseudotime trajectory. The color scale shows expression, ranging from lowest (0) to highest (maximum for each gene). (C) Density plots of canonical markers highlight B cell profile changes from immature or naïve to memory B cells across the derived pseudotime trajectory. (D) eSNP-eGene pairs with a

statistically significant difference in eQTL effect size across the B cell landscape. Both linear and quadratic models were applied to SNP-gene pairs across the pseudotime quantiles. SNPs known to be in high linkage disequilibrium ( $R^2 > 0.8$ ) with variants identified through GWAS of autoimmune diseases are displayed. Instances where the eGene was not expressed in a given quantile are shown in gray. The entire heatmap is available in fig S21.  $\beta$ , estimate of an eQTL effect size. (E) Examples of allele-specific changes in expression profiles across cell quantiles in the B cell pseudotime landscape. The scaled  $\beta$  values are shown for each eSNP-eGene pair, with the box plots colored by cell quantile with the same color coding used in (C).

The rs4987360 association replicates in bulk RNA-seq eQTL data from eQTL-Gen (34) and GTEx (5) and has allelic effects with the opposite direction in bulk B cells and monocytes (given rs2223286 with a  $R^2 = 1$  and  $D' = 1$ ) (51).

*CTLA4* is a gene-dosage sensitive, essential inhibitory receptor on T cells (52–55). In contrast to the example of *SELL*, the rs3087243-G allele downstream of *CTLA4*, which is associated with susceptibility to type 1 diabetes

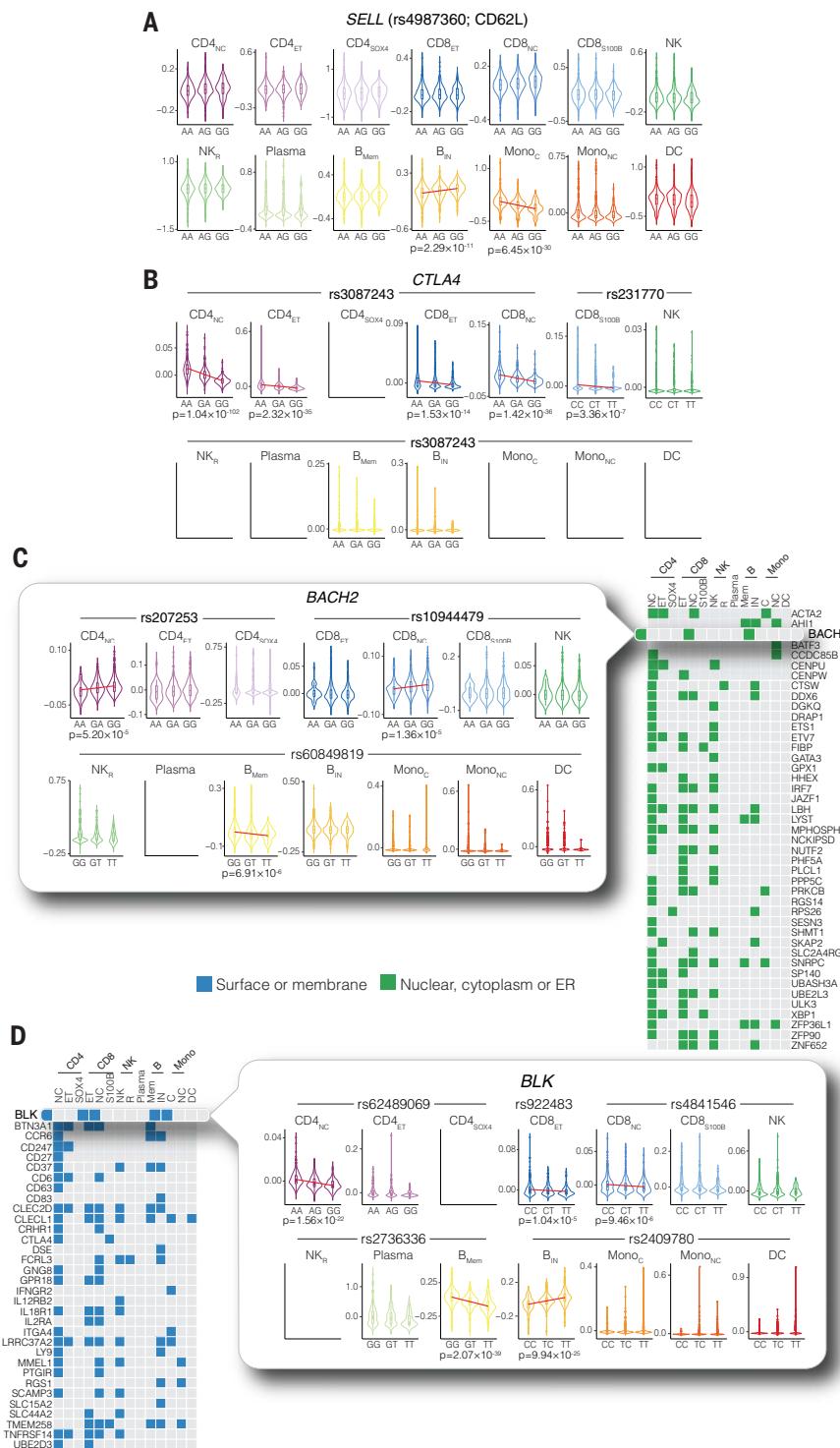
mellitus (T1DM) and RA (47, 56–58), acts in multiple cell types in the same allelic direction by decreasing *CTLA4* mRNA expression in four T cell subsets (Fig. 5B). The polymorphism rs231770 is located less than 10 kb away from

rs3087243 but is in linkage equilibrium ( $R^2 = 0.5$ ). rs231770-T is similarly associated with decreased *CTLA4* mRNA expression in CD8 $^+$  T cells with expression of S100B (CD8<sub>S100B</sub>) T cells and is associated with the autoimmune condition myasthenia gravis (59).

By linking allelic effects to changes in the expression of genes known to be implicated in autoimmune disease, we can support established hypotheses and identify previously uncharacterized examples of cellular mechanisms that underlie conditions and control immune regulation. By focusing on genes involved in autoimmune diseases, we evaluate how allelic effects vary across cell types, highlighting genes that encode membrane, nuclear, cytoplasmic, or endoplasmic reticulum (ER) proteins (Fig. 5, C and D). One example is *BACH2*, an essential transcription factor involved in differentiating memory B and T cells (60). We identify rs10944479, which was previously associated with thyroid peroxidase antibody positivity and hyperthyroidism (61) and has an eQTL effect on *BACH2* in CD8<sub>NC</sub> cells. We identify eQTLs for *BACH2* in CD4<sub>NC</sub> and B<sub>Mem</sub> cells, although the loci controlling expression in each cell type are independent of one another ( $R^2 = 0$  to 0.11; Fig. 5C). We demonstrate that rs60849819-T is associated with a significant down-regulation of *BACH2* in individuals who are homozygous for the T allele in B<sub>Mem</sub> cells, and rs207253-A has a similar effect in CD4<sub>NC</sub> cells (Fig. 5C).

Another example that provides insight into autoimmune disease is *BLK*. Five eSNPs were identified as being associated with *BLK* expression in CD4<sub>NC</sub>, CD8 $^+$  T cells with an effector memory phenotype (CD8<sub>ET</sub>), CD8<sub>NC</sub>, B<sub>Mem</sub>, and B<sub>IN</sub> cells (Fig. 5D) and are associated with RA, SLE, Sjögren's syndrome, and systemic sclerosis (41, 58, 62, 63) (Fig. 5D). One of these loci, rs2736336, results in the differential expression of *BLK* in B<sub>Mem</sub> cells. The dynamic eQTL analysis shows that allelic effects vary significantly across the B cell lineage, with the largest genetic effects observed in the quantiles of the memory B cells. rs2736336 is associated with SLE (41), and carrying copies of the autoimmune risk allele has been implicated in hyperactivation of B cells, with enhanced T cell costimulatory capacities (64). These results suggest that an allelic variation at rs2736336 contributes to interindividual variation in maintaining tolerance of B lymphocytes. Src family tyrosine kinases, such as *BLK*, are critical components of the signaling pathways that act downstream of the antigen receptor and determine the strength of the signal that a cell receives as a consequence of antigen engagement.

Finally, we sought to evaluate the impact of eQTLs on cellular composition within the OneK1K cohort. For each eSNP<sub>1</sub>, we tested for the association between an individual's



**Fig. 5. Genetic variation leads to cell type–specific immune regulation.** Cell type–specific eQTLs for genes known to play a role in immune function and common autoimmune diseases were identified. (A) The eQTL for *SELL* (CD62L) exhibits different allelic directions between the lymphoid and myeloid lineages; the effects for the eSNP rs4987360 are shown. (B) Allelic plots for the inhibitory receptor *CTLA4*. (C) With regard to nuclear, cytoplasm, or ER genes, we highlight *BACH2*, which showed significant and independent eQTLs in three cell types tagged by eSNPs rs207253, rs10944479, and rs60849819. (D) Cell type–specific eQTLs for cell surface receptors or membrane-associated proteins implicated in common autoimmune diseases. Focusing on *BLK*, we identified cell type–specific eQTLs and expression patterns across cell types. *p* values are from Spearman's rank correlation testing. Red lines indicate the allelic effect of significant eQTLs identified at an FDR less than 0.05.

genotype and cell-type proportion. At a study-wide significance threshold ( $p < 3.0 \times 10^{-6}$ ), we identified five associations, all of which affect the proportion of CD8<sub>S100B</sub> cells (table S11). The eGenes—*LSS*, *S100B*, *PRMT2*, *DIP2A*—and *PCNT* are all located within a 1-Mb region on chromosome 21q22, and the SNPs are in modest to high linkage disequilibrium with one another ( $R^2 = 0.31$  to 0.97), suggesting that a single variant influences the proportion of CD8<sub>S100B</sub> cells.

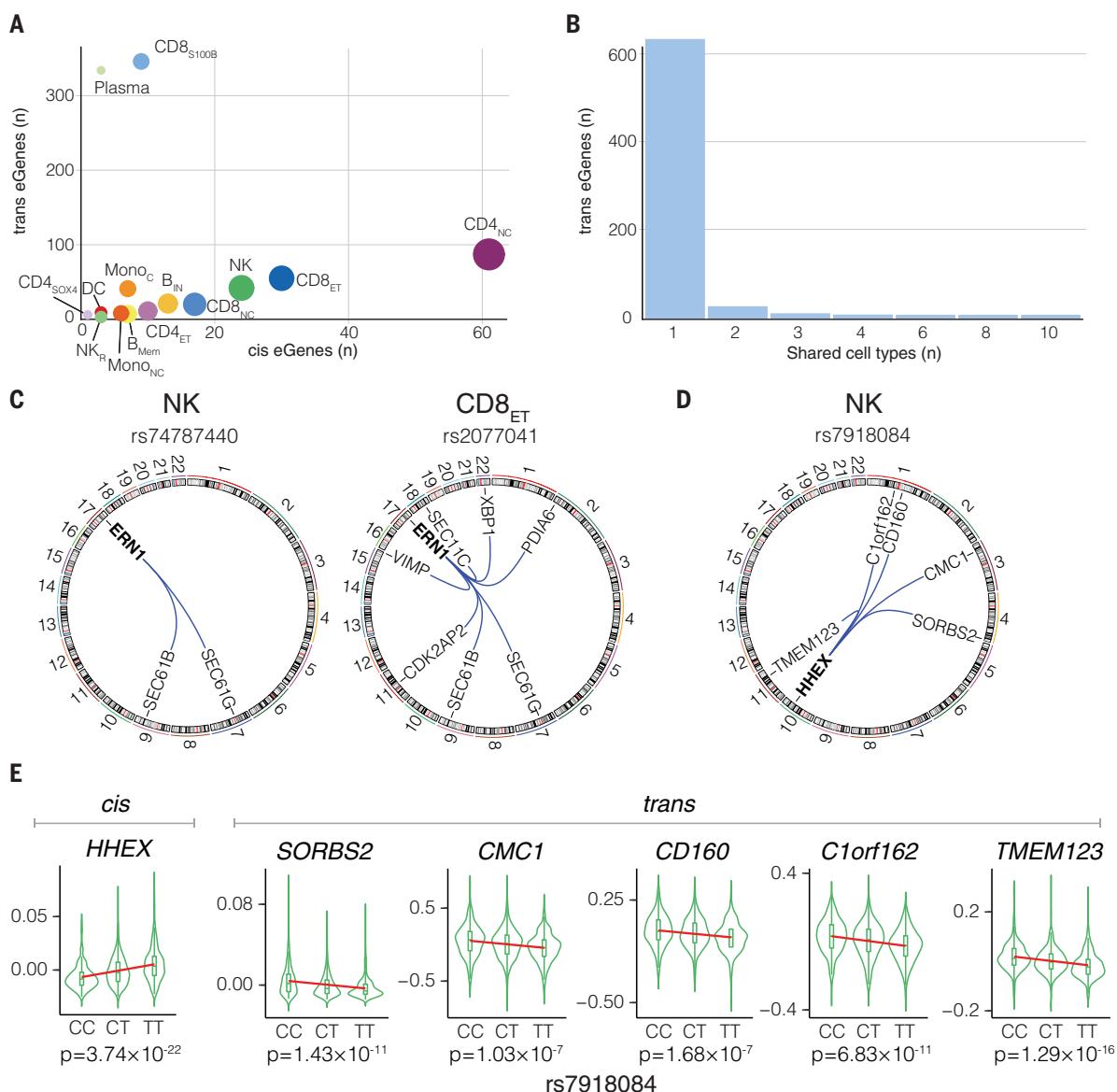
#### **Identification of cell type-specific trans-eQTLs suggests that distal genome regulation is highly cell type-specific**

We performed trans-eQTL analysis, testing the top eSNPs from each cis-eQTL against the gene

expression levels of all other genes, excluding those within  $\pm 2$  Mb of the cis-eGene and the major histocompatibility complex (MHC) locus. At a study-wide FDR of 0.01, we identified 990 trans-eQTL (median of one per cis-eSNP) (table S16). The number of trans-eGenes identified in each cell type was weakly correlated with the total number of cis-eQTLs (Spearman's  $\rho = 0.37$ ) (Fig. 6A). Compared with cis-eGenes, most trans-eGenes were specific for a cell type, and none were found ubiquitously across cell types (Fig. 6B and fig. S14).

A total of 630 cell type-specific trans-eQTL effects were identified. For example, rs2077041 has a cis effect on *ERNI* expression in CD8<sub>ET</sub> cells, with the C allele decreasing expression.

This locus has the same allelic direction of effect in seven trans-eGenes (Fig. 6C). *ERNI* is an unfolded protein response stress sensor with dual roles as a protein kinase and ribonuclease (65) and can catalyze the splicing of *XBP1* in a spliceosome-independent manner (66). Up-regulation of the master transcriptional regulator of the unfolded protein response, *XBP1*, promotes protein maturation. Individuals carrying copies of the C allele of rs2077041 have down-regulation of *XPB1* and *SEC61G*, *SEC61B*, and *SEC11C*, which are involved in the translocation, signal peptide removal, and integration of proteins across the ER membrane (67, 68). Interestingly, rs74787440 was found to also have a significant cis effect



**Fig. 6. Trans-acting eQTL mapping at single-cell resolution.** (A) The number of genes with a trans-eQTL (trans-eGenes) as a function of genes with a cis-eQTL (cis-eGenes). Bubble size corresponds to the relative number of cis-eSNPs identified as per Fig. 2A. (B) The number of trans-eGenes identified across the corresponding number of cell types. (C) eQTL associations of rs74787440 and rs2077041 exert

cis effects on *ERNI* in NK and CD8<sub>ET</sub> cells, respectively. (D) rs7918084 exerts a cis effect on *HHEX* expression and a trans effect in the opposite direction on *CD160*, *CMC1*, *SORBS2*, *TMEM123*, and *C1orf162*. (E) Change in the expression profile associated with rs7918084 genotypes on cis- and trans-genes.  $p$  values are from Spearman's rank correlation testing.

on *ERN1* expression in natural killer (NK) cells. Yet this same variant has a trans effect on *SEC61G* and *SEC61B* but not on the other genes associated with rs2077041.

When the locus on chromosome 21q22 that contains eQTLs associated with cellular composition was inspected, we identified many trans-eQTLs in this region and found that the expression levels of 118 genes throughout the genome were associated with these eSNPs (table S16). The route by which genetic variation affects CD8<sub>S100B</sub> frequency is unclear, and we find no evidence for the enrichment of functional pathways from the trans-eGenes. Across the tests, we observe a genomic inflation factor ( $\lambda$ ) of 1.05, suggesting a limited impact of single cell eQTLs on cell composition, although other significant associations would be uncovered with larger sample sizes.

Trans-eQTLs were identified at established autoimmune risk loci, including rs7918084-T (Fig. 6D), which is a cis-eQTL for *HHEX* in NK cells and is associated with atopic asthma and eosinophil counts in peripheral blood (50, 69). *HHEX* binds and represses the proapoptotic factor *BIM* (70), increasing the number of NK cells. In NK cells, rs7918084-T yields trans-eQTL effects across four chromosomes, decreasing the expression of *CD160*, *CMC1*, *SORBS2*, *TMEM123*, and *C1orf162* (Fig. 6E). *CD160* is a stimulatory receptor that is important in facilitating NK cell interferon- $\gamma$  (IFN- $\gamma$ ) production (71), with NK cell recruitment being pivotal in the development of the airway eosinophilia typical of asthma in murine models (72) and IFN- $\gamma$  secretion from NK cells in animal models of asthma being associated with reduced airway inflammation (73).

The production of IFN- $\gamma$  within airway inflammation models plays a complex role in regulating inflammation, and it has been shown that IFN- $\gamma$  acting on the airway epithelium will limit inflammation, such that lower IFN- $\gamma$  levels may lead to more asthma-related airway inflammation and obstruction (74). Mechanistically, the rs7918084-T risk allele for asthma may combine derepression of NK cell proliferation in an *HHEX*-dependent cis-acting mechanism with reduced IFN- $\gamma$  production by NK cells through *CD160* down-regulation, yielding the hallmarks of asthma.

#### Colocalization of genetic risk variants and single-cell cis-eQTLs identified cell type-specific mechanisms for autoimmune diseases

We applied an integrative approach to identify the relationship between cell type-specific eQTLs and genetic risk loci for seven common autoimmune diseases. We tested the extent to which cis-eQTLs (using eSNP<sub>i</sub>) from each cell type were enriched for 2335 trait-associated SNPs for the seven autoimmune diseases selected for cis-QTL exploration in Fig. 5, C and

D, using both colocalization and mendelian randomization approaches. Colocalization identified that 19% of cis-eQTLs have the same causal loci as GWAS risk variants (table S19). The overlap in eQTLs with GWAS loci shows significant enrichment for all diseases (Bonferroni adjusted  $p < 5.1 \times 10^{-4}$ ) and in all cell types (fig. S22). The overlap was highest in CD4<sub>NC</sub> and NK cells. Similarly, in NK recruiting (NK<sub>R</sub>) cells, there are high enrichments of overlap for inflammatory bowel disease (IBD), RA, ankylosing spondylitis (AS), and Crohn's disease (CD), which are low for multiple sclerosis (MS), SLE, and T1DM (fig. S22). These results highlight the complexity with which the polygenic effects of genetic risk for these common autoimmune diseases act at the cellular level.

Focusing on MS as an example, we identify overlapping cis-eQTL for 108 risk genes (table S17). Colocalization identified 530 gene-cell type pairs with a shared causal effect through eQTLs (Fig. 7A). The eQTL overlap for MS disease risk loci is highly cell type-specific: Of the 108 genes, 69 show eQTL overlap in just a single cell type. There are an additional 20 genes where eQTLs are identified in two cell types, 10 with eQTLs in three cell types, and five with eQTLs in four cell types. For example, for *RMT2*, which is a gene expressed in all PBMC types, we identify an overlapping eQTL and MS association in CD4<sub>NC</sub> cells only.

By contrast, for *METTL21B*, overlapping eQTLs are observed in CD4<sub>NC</sub>, CD4<sub>ET</sub>, and CD8<sub>NC</sub> cells. These results are concordant with our observations of cell type-specific eQTLs and provide further evidence for the genetic risk of common autoimmune diseases acting in a highly cell type-specific manner, where each locus contributes through changes to the function of a limited number of cell types. Still, collectively, genetic risk is endowed through the immune system.

Although overlapping GWAS SNPs and eQTLs imply that altered gene expression is involved in disease pathogenesis, there are two alternative hypotheses. One is that both the GWAS loci and eQTL have the same causal variant, but the effects on the two phenotypes are independent—that is, pleiotropy. A second explanation is that there are two independent causal loci, one for the GWAS association and the other for the eQTL. Still, they are in linkage disequilibrium with one another. To distinguish between these two hypotheses, we implemented a Mendelian randomization approach to identify evidence for the direction of causation by which risk loci for autoimmune diseases act (75). We tested for the causal relationship between all disease-associated variants ( $p < 1 \times 10^{-8}$ ) and OneKIK eQTLs across each of the 14 cell types using GWAS data from the seven autoimmune diseases previously introduced. In total, we identified

305 loci (study-wide FDR < 0.05) where the associated risk loci are identified as having a causal effect of disease risk through changes in the expression of a specific gene in one or more cell types, ranging from 4 (T1DM) to 47 (IBD) (Table 1 and table S18). Of the 305 loci, 188 are located in the MHC region, where causal effects display largely ubiquitous effects across cell types. The remaining 117 loci show patterns of highly cell type-specific causal effects, with 76 loci identified as having a causal effect in only one cell type (Table 1).

Again, using MS as an example, we evaluated the causal genes and the cell types in which they act for 90 risk loci (13). Of these, we were able to test for the causal direction of 57 risk loci based on the overlap of eQTLs in one or more cell types in OneKIK data. Our analysis identified significant (study-wide FDR < 0.05) effects for 39 genes (Fig. 7B and table S18). In the MHC region, we identified 73 loci whose causal effects on MS risk predominantly act through changes in the expression of genes in multiple cell types. For example, rs9264579 is identified as working through changes in human lymphocyte antigen class B (HLA-B) expression in all 14 cell types, whereas rs9501393 has a causal effect by changing the expression levels of *SKIV2L* in CD4<sub>NC</sub> cells only. Outside of the MHC region, we identified an additional 17 loci with causal effects that act in a more cell type-specific manner. For example, SNPs in the 1q23 region have previously been identified as associated with MS, with *FCRL3* tagged by rs7528684 ( $p = 8.9 \times 10^{-9}$ ) located within a promoter element. Our analysis identified the proximally located *FCRL3* as the causal gene for MS risk in CD8<sub>ET</sub> ( $p = 5.0 \times 10^{-7}$ ) and B<sub>IN</sub> ( $p = 6.6 \times 10^{-7}$ ) cells (Fig. 7C).

Another example is the MS risk locus at 3q12, which is tagged by lead SNP rs9882971 ( $p = 6.5 \times 10^{-9}$ ), where Mendelian randomization analysis identified *EAF2* as the causal gene in B<sub>IN</sub> ( $p = 1.7 \times 10^{-8}$ ) and B<sub>Mem</sub> ( $p = 2.8 \times 10^{-8}$ ) cells. Because *EAF2* is universally expressed, our results provide a clear example of the ability to identify cell-type genetic effects on gene expression and pinpoint the cells in which genetic risk factors are acting. A final example is the risk locus at 19p13, which is tagged by top SNP rs12984330 ( $p = 2.8 \times 10^{-9}$ ) located in the intronic region of *PIK3R2*. Our analyses identify the causal gene as *MAST3* in CD8<sub>ET</sub> and NK cells, which is located about 65 kb from the lead SNP. *MAST3* is also universally expressed, although there is known evidence of the risk locus overlapping with regulatory elements, which presents an interesting case for further exploration.

#### Discussion

This study reveals the allelic architecture of cell type-specific eQTLs in circulating immune cells. We mapped genetic effects of 14 cell types

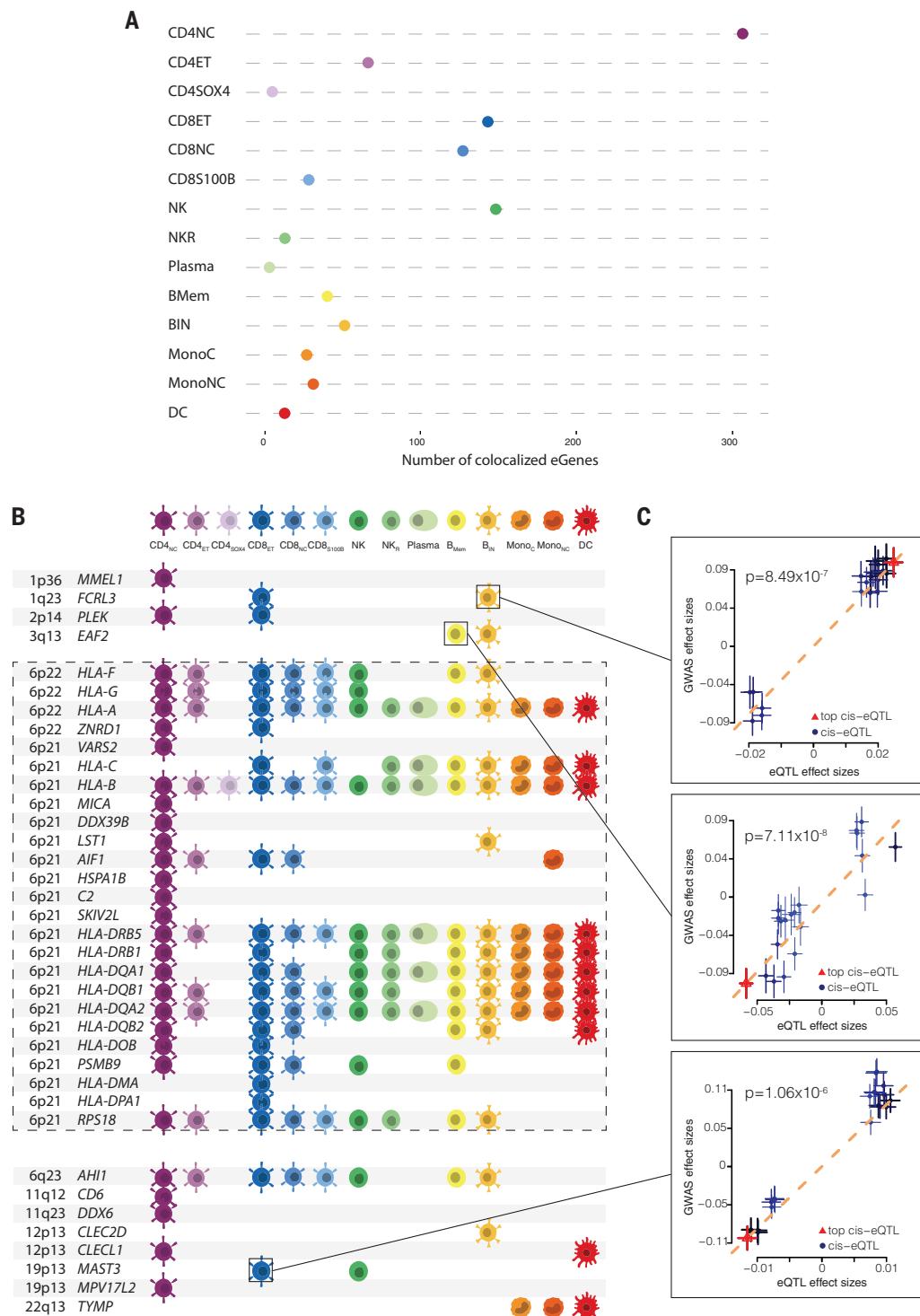
on gene expression and identified more than 26,000 independent cis-acting eQTLs and 990 trans-eQTLs outside the MHC locus. Summary statistics for our cis-eQTLs can be found in table S10 and are available for browsing from [www.onek1k.org/](http://www.onek1k.org/). We show that most of these eQTLs have an allelic effect on gene expression, which is largely cell type-specific, yet replicate in two independent cohorts. We identify ex-

amples of how genetic loci contribute to key immune function pathways. Lineage-dynamic analyses applied to B cells demonstrated expected changes in markers of B cell maturation. They facilitated the identification of dynamic eQTLs, many of which had not been identified through our primary cis-eQTL analysis. By integrating scRNA-seq eQTL data with autoimmune risk loci identified through

GWASs, we uncovered both the causal gene at these loci and resolved the cells through which these genes exert their pathogenetic effects. Mendelian randomization and colocalization analysis of our eQTLs and disease-associated SNPs were performed, providing complementary insights into the relationship between eQTLs and disease risk loci. The colocalization analysis provides evidence that the same

### Fig. 7. Dissection of autoimmune disease loci using eQTL mapping at single-cell resolution.

**(A)** Breakdown of cis-eGenes colocalized with GWAS associations for MS using Bayes factors. **(B)** Mendelian randomization was used to establish causation between overlapping GWAS loci for MS and identified eSNPs. Significant results were identified for 39 MS-related genes ( $FDR < 0.05$ ), with the 12 outside the MHC locus (dashed box) displaying highly cell type-specific effects. Colored symbols depict cell types. Differences were identified between the direct NHGRI-EBI GWAS Catalog (9) overlap and Mendelian randomization analysis for eGene and cell-specific profiles. **(C)** The effect sizes of OneK1K eQTL SNPs plotted against the allelic effects from the MS GWAS for *FCRL3*, *EAF2*, and *MAST3* in  $B_{IN}$ ,  $B_{Mem}$ , and  $CD8_{ET}$  cells, respectively, are displayed.  $p$  values are from the heterogeneity in dependent instruments (HEIDI) test.



**Table 1. A summary of significant evidence of causation between overlapping GWAS loci and identified eQTLs for autoimmune diseases.** The number of cell types in which causal effect is identified is given in parentheses. For loci with causal effects acting in multiple cell types, multiple independent eQTLs are often present (table S18). Significance threshold of FDR < 0.5.

Disease	Loci	Genes
Systemic lupus erythematosus	19	AIF1 (3), BLK (3), BTN2A1 (3), BTN2A2 (1), BTN3A2 (9), C6orf48 (4), FAM167A (1), HLA-B (2), HLA-C (4), HLA-DMA (1), HLA-DQA1 (2), HLA-DQB1 (6), HLA-DQB2 (2), HLA-DRB1 (3), HLA-DRB5 (9), MICB (1), UBE2L3 (3), XXbac-BPG181B23.7 (3), ZFP57 (5)
Rheumatoid arthritis	37	AIF1 (3), ANKRD55 (3), B3GALT4 (1), C2 (1), C6orf48 (1), CTLA4 (5), DDX6 (3), HLA-A (9), HLA-B (6), HLA-C (12), HLA-DMA (1), HLA-DOB (2), HLA-DPA1 (2), HLA-DPB1 (5), HLA-DQA1 (11), HLA-DQA2 (13), HLA-DQB1 (12), HLA-DQB2 (4), HLA-DRB1 (9), HLA-DRB5 (12), HSD17B8 (1), HSPA1B (1), IL6ST (1), LST1 (4), MDC1 (1), MICA (1), MMEL1 (1), RP11-279F6.3 (1), RP11-97H7.4 (1), SKIV2L (2), SYNGR1 (6), TAPI1 (1), TAPBP (4), UQCRC2 (1), XXbac-BPG181B23.7 (2), XXbac-BPG299F13.17 (14), ZFP57 (5)
Crohn's disease	18	ADCY7 (2), BRD7 (3), C5orf56 (1), CCDC101 (1), CTD-2260A17.2 (1), CYLD (1), ERAP2 (13), GSDMB (1), IP6K2 (1), IRF1 (4), ORMDL3 (1), RNASET2 (9), SLC22A5 (1), SLC2A4RG (1), SNX20 (1), SPNS1 (3), TUFM (2), UQCRCO (1)
Inflammatory bowel disease	47	ADCY7 (3), BRD7 (3), C5orf56 (1), CCDC101 (2), CYLD (1), EGR2 (1), ETS2 (1), FCGR3B (1), FYB (1), GMEMB2 (1), GPANK1 (1), GPX1 (2), GSDMB (1), HCG23 (1), HLA-DOB (3), HLA-DQA1 (8), HLA-DQA2 (13), HLA-DQB1 (7), HLA-DQB2 (3), HLA-DRB1 (4), HLA-DRB5 (8), LAMB1 (8), LST1 (4), MICB (1), NDUFS2 (1), ORMDL3 (5), PAPD5 (1), PEX13 (1), PNMT (1), RBM6 (1), RNASET2 (7), RP11-229P13.20 (1), RP11-324I22.4 (1), RP11-94L15.2 (1), SLC22A5 (1), SLC2A4RG (2), STMM3 (2), TCTA (1), TNFRSF9 (1), TUFM (8), UBE2L3 (4), UQCRCO (1)
Multiple sclerosis	39	AHI1 (8), AIF1 (5), C2 (1), CD6 (1), CLEC2D (1), CLECL1 (2), DDX39B (1), DDX6 (1), EAF2 (2), FCRL3 (2), HLA-A (13), HLA-B (14), HLA-C (9), HLA-DMA (1), HLA-DOB (2), HLA-DPA1 (1), HLA-DQA1 (11), HLA-DQA2 (13), HLA-DQB1 (13), HLA-DQB2 (6), HLA-DRB1 (9), HLA-DRB5 (13), HLA-F (8), HLA-G (6), HSPA1B (1), LST1 (2), MAST3 (2), MICA (1), MMEL1 (1), MPV17L2 (1), PLEK (2), PSMB9 (5), RPS18 (9), SKIV2L (1), TYMP (3), VARS2 (1), XXbac-BPG181B23.7 (3), XXbac-BPG299F13.17 (9), ZNRD1 (2)
Ankylosing spondylitis	10	AIF1 (2), C6orf48 (3), HLA-A (1), HLA-B (9), HLA-C (9), HLA-DQA1 (1), LST1 (3), MICB (1), NCR3 (2), XXbac-BPG181B23.7 (4)
Type 1 diabetes mellitus	4	HLA-DQA1 (9), HLA-DQA2 (13), HLA-DQB1 (9), HLA-DQB2 (3)

causal loci are shared between an eQTL and GWAS risk loci. This observation can be explained by either a causal effect or pleiotropy. Mendelian randomization takes this one step further, addressing these alternative hypotheses to provide evidence of the direction of causal effect (i.e., DNA to RNA to disease).

Single-cell eQTL analyses have several advantages over alternative methods that are used to map the allelic architecture of transcriptional regulation, such as cellular deconvolution from bulk RNA-seq data. For example, scRNA-seq-based approaches can identify previously uncharacterized and rare cell types, which are challenging to detect using deconvolution methods (22, 76). scRNA-seq also accurately quantifies transcriptional abundance, because amplified libraries can be collapsed back to the level of individual transcript molecules using unique molecular identifier (UMI) barcodes. Nevertheless, ongoing work investigating trans-acting variants and gene-

environment interactions at single-cell resolution is required, particularly in the immune system, where exposure to antigens or cytokines can trigger changes in the transcriptional profile of cells.

This work brings together genetic epidemiology with scRNA-seq to uncover drivers of interindividual variation in the immune system. Our results demonstrate how segregating genetic variation influences the expression of genes that encode proteins involved in critical immune regulatory and signaling pathways in a cell type–specific manner. Understanding the genetic underpinnings of immune system regulation will have broad implications in the treatment of autoimmune diseases and infections, transplantation, and cancer.

#### Materials and methods summary

We collected peripheral blood from 1104 individuals. After DNA extraction, samples were genotyped using the Illumina Infinium Global

Screening Array. Poor genotyping quality, cryptic relatedness, and ethnic outliers were removed, yielding 1034 participants. Imputation was performed using the Michigan Imputation Server (24). PMBCs were isolated through density-gradient centrifugation from heparinized whole blood (8-ml cell preparation tubes; BD Biosciences Australia; catalog no. 362753), with live cells isolated with the Miltenyi Dead Cell Removal Kit (Miltenyi; catalog no. 130-090-101). Live cells were subsequently pooled with 12 to 14 participant samples per pool, which underwent single-cell RNA capture and barcoding with the Single Cell 3' Library and Gel Bead Kit (10x Genomics) to target the capture of 20,000 cells per well. Library preparation and multiplex sequencing using an Illumina NovaSeq 2000 generated 49 billion reads. Reads underwent processing using the Cell Ranger Single Cell Software Suite (v 2.2.0; 10x Genomics) into FASTQ files, followed by demultiplexing into their respective pools, and

were mapped to GRCh37/hg19 (release 84) using STAR (77).

Cells were assigned using genotype data to individual participants using Dexmuxlet (78), with droplets containing two or more cells excluded using Demuxlet and Scrublet (79), yielding 982 individuals in the final cohort. Cells were classified using supervised clustering into major immune populations using reference data from Zheng *et al.* (26) and then underwent unsupervised clustering using Seurat v3.0 (80). Expression values for genes were first normalized by the pool for the distribution of the total number of UMIs, the number of genes, and the percentage of mitochondrial gene expression and were subsequently adjusted for sex, age, six genotyping principal components, and two probabilistic estimation of expression residuals (PEER) factors. Subsequent single-cell cis-eQTL mapping was undertaken through five rounds of iterative conditional analysis to yield cell type-specific eSNP<sub>1</sub> to eSNP<sub>5</sub>. Lead cis-eQTLs were replicated in two independent cohorts of participants by creating pseudo-bulk populations, and trans-eQTL mapping was performed. Lineage-dynamic analysis was undertaken using SCTtransform (81) to identify 500 differentially expressed genes and filter out contaminating cells. A two-dimensional space was created using PHATE (82) and slingshot (83). Six quantiles were analyzed for the presence of dynamic eQTLs.

## REFERENCES AND NOTES

- P. Brodin, M. M. Davis, Human immune system variation. *Nat. Rev. Immunol.* **17**, 21–29 (2017). doi: 10.1038/nri.2016.125; pmid: 27916977
- L. Chen *et al.*, Genetic drivers of epigenetic and transcriptional variation in human immune cells. *Cell* **167**, 1398–1414.e24 (2016). doi: 10.1016/j.cell.2016.10.026; pmid: 27863251
- Y. Li *et al.*, RNA splicing is a primary link between genetic variation and disease. *Science* **352**, 600–604 (2016). doi: 10.1126/science.aad9417; pmid: 27126046
- F. Hormozdiari *et al.*, Leveraging molecular quantitative trait loci to understand the genetic architecture of diseases and complex traits. *Nat. Genet.* **50**, 1041–1047 (2018). doi: 10.1038/s41588-018-0148-z; pmid: 29942083
- GTEX Consortium, The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318–1330 (2020). doi: 10.1126/science.aaz1776; pmid: 32913098
- W. J. Astle *et al.*, The allelic landscape of human blood cell trait variation and links to common complex disease. *Cell* **167**, 1415–1429.e19 (2016). doi: 10.1016/j.cell.2016.10.042; pmid: 27863252
- C. C. Whitacre, Sex differences in autoimmune disease. *Nat. Immunol.* **2**, 777–780 (2001). doi: 10.1038/ni0901-777; pmid: 11526384
- G. S. Cooper, B. C. Stroehla, The epidemiology of autoimmune diseases. *Autoimmun. Rev.* **2**, 119–125 (2003). doi: 10.1016/S1568-9972(03)00006-5; pmid: 12848952
- A. Buniello *et al.*, The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.* **47**, D1005–D1012 (2019). doi: 10.1093/nar/gky1120; pmid: 30445434
- L. A. Hindorf *et al.*, Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 9362–9367 (2009). doi: 10.1073/pnas.0903103106; pmid: 19474294
- M. Boyd *et al.*, Characterization of the enhancer and promoter landscape of inflammatory bowel disease from human colon biopsies. *Nat. Commun.* **9**, 1661 (2018). doi: 10.1038/s41467-018-03766-z; pmid: 29695774
- Y. Okada, S. Eyre, A. Suzuki, Y. Kochi, K. Yamamoto, Genetics of rheumatoid arthritis: 2018 status. *Ann. Rheum. Dis.* **78**, 446–453 (2019). doi: 10.1136/annrheumdis-2018-213678; pmid: 30530827
- J. Bentham *et al.*, Genetic association analyses implicate aberrant regulation of innate and adaptive immunity genes in the pathogenesis of systemic lupus erythematosus. *Nat. Genet.* **47**, 1457–1464 (2015). doi: 10.1038/ng.3434; pmid: 26502338
- B. D. Umans, A. Battle, Y. Gilad, Where are the disease-associated eQTLs? *Trends Genet.* **37**, 109–124 (2021). doi: 10.1016/j.tig.2020.08.009; pmid: 32912663
- GTEX Consortium, Genetic effects on gene expression across human tissues. *Nature* **550**, 204–213 (2017). doi: 10.1038/nature24277; pmid: 29022597
- H.-J. Westra *et al.*, Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat. Genet.* **45**, 1238–1243 (2013). doi: 10.1038/ng.2756; pmid: 24013639
- L. R. Lloyd-Jones *et al.*, The genetic architecture of gene expression in peripheral blood. *Am. J. Hum. Genet.* **100**, 228–237 (2017). doi: 10.1016/j.ajhg.2016.12.008; pmid: 28065468
- C. J. Ye *et al.*, Intersection of population variation and autoimmunity genetics in human T cell activation. *Science* **345**, 1254665 (2014). doi: 10.1126/science.1254665; pmid: 25214635
- J. E. Powell *et al.*, Genetic control of gene expression in whole blood and lymphoblastoid cell lines is largely independent. *Genome Res.* **22**, 456–466 (2012). doi: 10.1101/gr.126540.111; pmid: 22183966
- K. Ishigaki *et al.*, Polygenic burdens on cell-specific pathways underlie the risk of rheumatoid arthritis. *Nat. Genet.* **49**, 1120–1125 (2017). doi: 10.1038/ng.3885; pmid: 28553958
- B. J. Schmiedel *et al.*, Impact of genetic polymorphisms on human immune cell gene expression. *Cell* **175**, 1701–1715.e16 (2018). doi: 10.1016/j.cell.2018.10.022; pmid: 30449622
- S. Kim-Hellmuth *et al.*, Cell type-specific genetic regulation of gene expression across human tissues. *Science* **369**, eaaz8528 (2020). doi: 10.1126/science.aaz8528; pmid: 32913075
- F. Avila Cobos, J. Alquicira-Hernandez, J. E. Powell, P. Mestdagh, K. De Preter, Benchmarking of cell type deconvolution pipelines for transcriptomics data. *Nat. Commun.* **11**, 5650 (2020). doi: 10.1038/s41467-020-19015-1; pmid: 33159064
- S. McCarthy *et al.*, A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.* **48**, 1279–1283 (2016). doi: 10.1038/ng.3643; pmid: 27548312
- J. Alquicira-Hernandez, A. Sathe, H. P. Ji, Q. Nguyen, J. E. Powell, scPred: Accurate supervised method for cell-type classification from single-cell RNA-seq data. *Genome Biol.* **20**, 264 (2019). doi: 10.1186/s13059-019-1862-5; pmid: 31829268
- G. X. Zheng *et al.*, Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8**, 14049 (2017). doi: 10.1038/ncomms14049; pmid: 28091601
- V. Orrù *et al.*, Genetic variants regulating immune cell levels in health and disease. *Cell* **155**, 242–256 (2013). doi: 10.1016/j.cell.2013.08.041; pmid: 24074872
- P. Brodin *et al.*, Variation in the human immune system is largely driven by non-heritable influences. *Cell* **160**, 37–47 (2015). doi: 10.1016/j.cell.2014.12.020; pmid: 25594173
- E. Cano-Gamez *et al.*, Single-cell transcriptomics identifies an effectors gradient shaping the response of CD4<sup>+</sup> T cells to cytokines. *Nat. Commun.* **11**, 1801 (2020). doi: 10.1038/s41467-020-15543-y; pmid: 32286271
- A.-C. Villani *et al.*, Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. *Science* **356**, eaah4573 (2017). doi: 10.1126/science.aah4573; pmid: 28428369
- B. P. Fairfax *et al.*, Innate immune activity conditions the effect of regulatory variants upon monocyte gene expression. *Science* **343**, 1246949 (2014). doi: 10.1126/science.1246949; pmid: 24604202
- M. A. Christopheros *et al.*, Citrullination regulates pluripotency and histone H1 binding to chromatin. *Nature* **507**, 104–108 (2014). doi: 10.1038/nature12942; pmid: 24463520
- F. Cornélis *et al.*, New susceptibility locus for rheumatoid arthritis suggested by a genome-wide linkage study. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 10746–10750 (1998). doi: 10.1073/pnas.95.18.10746; pmid: 9724775
- U. Vösa *et al.*, Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. *Nat. Genet.* **53**, 1300–1310 (2021). doi: 10.1038/s41588-021-00913-z; pmid: 34475573
- J. W. Said *et al.*, TCL1 oncogene expression in B cell subsets from lymphoid hyperplasia and distinct classes of B cell lymphoma. *Lab. Invest.* **81**, 555–564 (2001). doi: 10.1038/labinvest.3780264; pmid: 11304575
- E. F. Wagner *et al.*, Novel diversity in IL-4-mediated responses in resting human naïve B cells versus germinal center/memory B cells. *J. Immunol.* **165**, 5573–5579 (2000). doi: 10.4049/jimmunol.165.10.5573; pmid: 11067912
- S. G. Tangye, Y. J. Liu, G. Aversa, J. H. Phillips, J. E. de Vries, Identification of functional human splenic memory B cells by expression of CD148 and CD27. *J. Exp. Med.* **188**, 1691–1703 (1998). doi: 10.1084/jem.188.9.1691; pmid: 9802981
- A. Kallies *et al.*, Initiation of plasma-cell differentiation is independent of the transcription factor Blimp-1. *Immunity* **26**, 555–566 (2007). doi: 10.1016/j.jimmuni.2007.04.007; pmid: 17509907
- M. L. Tang, D. A. Steeber, X. Q. Zhang, T. F. Tedder, Intrinsic differences in L-selectin expression levels affect T and B lymphocyte subset-specific recirculation pathways. *J. Immunol.* **160**, 5113–5121 (1998). pmid: 9590263
- M. Togni *et al.*, Regulation of in vitro and in vivo immune functions by the cytosolic adaptor protein SKAP-HOM. *Mol. Cell. Biol.* **25**, 8052–8063 (2005). doi: 10.1128/MCB.25.18.8052-8063.2005; pmid: 16135797
- C. D. Langefeld *et al.*, Transancestral mapping and genetic load in systemic lupus erythematosus. *Nat. Commun.* **8**, 16021 (2017). doi: 10.1038/ncomms16021; pmid: 28714469
- P. K. Gregersen *et al.*, REL, encoding a member of the NF-κB family of transcription factors, is a newly defined risk locus for rheumatoid arthritis. *Nat. Genet.* **41**, 820–823 (2009). doi: 10.1038/ng.395; pmid: 19503088
- F. Köntgen *et al.*, Mice lacking the c-rel proto-oncogene exhibit defects in lymphocyte proliferation, humoral immunity, and interleukin-2 expression. *Genes Dev.* **9**, 1965–1977 (1995). doi: 10.1101/gad.9.16.1965; pmid: 7649478
- J. Dang *et al.*, ORMDL3 facilitates the survival of splenic B cells via an ATF6α-endoplasmic reticulum stress-Beclin1 autophagy regulatory pathway. *J. Immunol.* **199**, 1647–1659 (2017). doi: 10.4049/jimmunol.1602124; pmid: 28747345
- A. H. Beecham *et al.*, Analysis of immune-related loci identifies 48 new susceptibility variants for multiple sclerosis. *Nat. Genet.* **45**, 1353–1360 (2013). doi: 10.1038/ng.2770; pmid: 24076602
- K. M. de Lange *et al.*, Genome-wide association study implicates immune activation of multiple integrin genes in inflammatory bowel disease. *Nat. Genet.* **49**, 256–261 (2017). doi: 10.1038/ng.3760; pmid: 28067908
- S. Onengut-Gumuscu *et al.*, Fine mapping of type 1 diabetes susceptibility loci and evidence for colocalization of causal variants with lymphoid gene enhancers. *Nat. Genet.* **47**, 381–386 (2015). doi: 10.1038/ng.3245; pmid: 2571624
- J. C. Barrett *et al.*, Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. *Nat. Genet.* **41**, 703–707 (2009). doi: 10.1038/ng.381; pmid: 19430480
- W. M. Gallatin, I. L. Weissman, E. C. Butcher, A cell-surface molecule involved in organ-specific homing of lymphocytes. *Nature* **304**, 30–34 (1983). doi: 10.1038/304030a0; pmid: 6866086
- M.-H. Chen *et al.*, Trans-ethnic and ancestry-specific blood-cell genetics in 746,667 individuals from 5 global populations. *Cell* **182**, 1198–1213.e14 (2020). doi: 10.1016/j.cell.2020.06.045; pmid: 32888493
- B. P. Fairfax *et al.*, Genetics of gene expression in primary immune cells identifies cell type-specific master regulators and roles of HLA alleles. *Nat. Genet.* **44**, 502–510 (2012). doi: 10.1038/ng.2205; pmid: 22446964
- P. Waterhouse *et al.*, Lymphoproliferative disorders with early lethality in mice deficient in Ctla-4. *Science* **270**, 985–988 (1995). doi: 10.1126/science.270.5238.985; pmid: 7481803
- E. A. Tivol *et al.*, Loss of CTLA-4 leads to massive lymphoproliferation and fatal multiorgan tissue destruction, revealing a critical negative regulatory role of CTLA-4. *Immunity* **3**, 541–547 (1995). doi: 10.1016/1074-7613(95)90125-6; pmid: 7584144
- D. Schubert *et al.*, Autosomal dominant immune dysregulation syndrome in humans with CTLA4 mutations. *Nat. Med.* **20**, 1410–1416 (2014). doi: 10.1038/nm.3746; pmid: 25239292
- H. S. Kuehn *et al.*, Immune dysregulation in human subjects with heterozygous germline mutations in CTLA4. *Science* **345**, 1623–1627 (2014). doi: 10.1126/science.1255904; pmid: 25213377

56. Y.-F. Wang *et al.*, Identification of 38 novel loci for systemic lupus erythematosus and genetic heterogeneity between ancestral groups. *Nat. Commun.* **12**, 772 (2021). doi: [10.1038/s41467-021-21049-y](https://doi.org/10.1038/s41467-021-21049-y); pmid: [33536424](https://pubmed.ncbi.nlm.nih.gov/33536424/)
57. J. D. Cooper *et al.*, Meta-analysis of genome-wide association study data identifies additional type 1 diabetes risk loci. *Nat. Genet.* **40**, 1399–1401 (2008). doi: [10.1038/ng.249](https://doi.org/10.1038/ng.249); pmid: [18978792](https://pubmed.ncbi.nlm.nih.gov/18978792/)
58. K. Kim *et al.*, High-density genotyping of immune loci in Koreans and Europeans identifies eight new rheumatoid arthritis risk loci. *Ann. Rheum. Dis.* **74**, e13 (2015). doi: [10.1136/annrheumdis-2013-204749](https://doi.org/10.1136/annrheumdis-2013-204749); pmid: [24532676](https://pubmed.ncbi.nlm.nih.gov/24532676/)
59. A. E. Renton *et al.*, A genome-wide association study of myasthenia gravis. *JAMA Neurol.* **72**, 396–404 (2015). doi: [10.1001/jamaneurol.2014.4103](https://doi.org/10.1001/jamaneurol.2014.4103); pmid: [25643325](https://pubmed.ncbi.nlm.nih.gov/25643325/)
60. R. Roychoudhuri *et al.*, BACH2 represses effector programmes to stabilize Treg-mediated immune homeostasis - a new target in tumor immunotherapy? *J. Immunother. Cancer* **1**, O14 (2013). doi: [10.1186/2051-1426-1-S1-O14](https://doi.org/10.1186/2051-1426-1-S1-O14)
61. M. Medici *et al.*, Identification of novel genetic loci associated with thyroid peroxidase antibodies and clinical thyroid disease. *PLOS Genet.* **10**, e1004123 (2014). doi: [10.1371/journal.pgen.1004123](https://doi.org/10.1371/journal.pgen.1004123); pmid: [24536183](https://pubmed.ncbi.nlm.nih.gov/24536183/)
62. C. J. Lessard *et al.*, Variants at multiple loci implicated in both innate and adaptive immune responses are associated with Sjögren's syndrome. *Nat. Genet.* **45**, 1284–1292 (2013). doi: [10.1038/ng.2792](https://doi.org/10.1038/ng.2792); pmid: [24097067](https://pubmed.ncbi.nlm.nih.gov/24097067/)
63. A. Márquez *et al.*, Meta-analysis of Immunochip data of four autoimmune diseases reveals novel single-disease and cross-phenotype associations. *Genome Med.* **10**, 97 (2018). doi: [10.1186/s13073-018-0604-8](https://doi.org/10.1186/s13073-018-0604-8); pmid: [30572963](https://pubmed.ncbi.nlm.nih.gov/30572963/)
64. K. R. Simpfendorfer *et al.*, Autoimmune disease-associated haplotypes of BLK exhibit lowered thresholds for B cell activation and expansion of Ig class-switched B cells. *Arthritis Rheumatol.* **67**, 2866–2876 (2015). doi: [10.1002/art.39301](https://doi.org/10.1002/art.39301); pmid: [26246128](https://pubmed.ncbi.nlm.nih.gov/26246128/)
65. A. V. Korennykh *et al.*, The unfolded protein response signals through high-order assembly of Ire1. *Nature* **457**, 687–693 (2009). doi: [10.1038/nature07661](https://doi.org/10.1038/nature07661); pmid: [19079236](https://pubmed.ncbi.nlm.nih.gov/19079236/)
66. K. P. K. Lee *et al.*, Structure of the dual enzyme Ire1 reveals the basis for catalysis and regulation in nonconventional RNA splicing. *Cell* **132**, 89–100 (2008). doi: [10.1016/j.cell.2007.10.057](https://doi.org/10.1016/j.cell.2007.10.057); pmid: [18191223](https://pubmed.ncbi.nlm.nih.gov/18191223/)
67. T. Becker *et al.*, Structure of monomeric yeast and mammalian Sec61 complexes interacting with the translating ribosome. *Science* **326**, 1369–1373 (2009). doi: [10.1126/science.1178535](https://doi.org/10.1126/science.1178535); pmid: [19933108](https://pubmed.ncbi.nlm.nih.gov/19933108/)
68. K. U. Kalies, E. Hartmann, Membrane topology of the 12- and the 25-kDa subunits of the mammalian signal peptidase complex. *J. Biol. Chem.* **271**, 3925–3929 (1996). doi: [10.1074/jbc.271.7.3925](https://doi.org/10.1074/jbc.271.7.3925); pmid: [8632014](https://pubmed.ncbi.nlm.nih.gov/8632014/)
69. Z. Zhu *et al.*, Shared genetic and experimental links between obesity-related traits and asthma subtypes in UK Biobank. *J. Allergy Clin. Immunol.* **145**, 537–549 (2020). doi: [10.1016/j.jaci.2019.09.035](https://doi.org/10.1016/j.jaci.2019.09.035); pmid: [31669095](https://pubmed.ncbi.nlm.nih.gov/31669095/)
70. S. W. Jang *et al.*, Homeobox protein Hhex negatively regulates Treg cells by inhibiting Foxp3 expression and function. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 25790–25799 (2019). doi: [10.1073/pnas.1907224116](https://doi.org/10.1073/pnas.1907224116); pmid: [31792183](https://pubmed.ncbi.nlm.nih.gov/31792183/)
71. T. C. Tu *et al.*, CD160 is essential for NK-mediated IFN- $\gamma$  production. *J. Exp. Med.* **212**, 415–429 (2015). doi: [10.1084/jem.20131601](https://doi.org/10.1084/jem.20131601); pmid: [25712123](https://pubmed.ncbi.nlm.nih.gov/25712123/)
72. M. Korsgren *et al.*, Natural killer cells determine development of allergen-induced eosinophilic airway inflammation in mice. *J. Exp. Med.* **189**, 553–562 (1999). doi: [10.1084/jem.189.3.553](https://doi.org/10.1084/jem.189.3.553); pmid: [9927517](https://pubmed.ncbi.nlm.nih.gov/9927517/)
73. S. Matsubara *et al.*, IL-2 and IL-18 attenuation of airway hyperresponsiveness requires STAT4, IFN- $\gamma$ , and natural killer cells. *Am. J. Respir. Cell Mol. Biol.* **36**, 324–332 (2007). doi: [10.1165/rccm.2006-0231OC](https://doi.org/10.1165/rccm.2006-0231OC); pmid: [17038663](https://pubmed.ncbi.nlm.nih.gov/17038663/)
74. C. Mitchell, K. Provost, N. Niu, R. Homer, L. Cohn, IFN- $\gamma$  acts on the airway epithelium to inhibit local and systemic pathology in allergic airway disease. *J. Immunol.* **187**, 3815–3820 (2011). doi: [10.4049/jimmunol.1100436](https://doi.org/10.4049/jimmunol.1100436); pmid: [21873527](https://pubmed.ncbi.nlm.nih.gov/21873527/)
75. Z. Zhu *et al.*, Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat. Genet.* **48**, 481–487 (2016). doi: [10.1038/ng.3538](https://doi.org/10.1038/ng.3538); pmid: [27019110](https://pubmed.ncbi.nlm.nih.gov/27019110/)
76. A. Dobin *et al.*, STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 19–21 (2013). doi: [10.1093/bioinformatics/bts635](https://doi.org/10.1093/bioinformatics/bts635); pmid: [23104886](https://pubmed.ncbi.nlm.nih.gov/23104886/)
77. H. M. Kang *et al.*, Multiplexed droplet single-cell RNA-sequencing using natural genetic variation. *Nat. Biotechnol.* **36**, 89–94 (2018). doi: [10.1038/nbt.4042](https://doi.org/10.1038/nbt.4042); pmid: [29227470](https://pubmed.ncbi.nlm.nih.gov/29227470/)
78. S. L. Wolock, R. Lopez, A. M. Klein, Scrublet: Computational identification of cell doublets in single-cell transcriptomic data. *Cell Syst.* **8**, 281–291.e9 (2019). doi: [10.1016/j.cels.2018.11.005](https://doi.org/10.1016/j.cels.2018.11.005); pmid: [30954476](https://pubmed.ncbi.nlm.nih.gov/30954476/)
79. T. Stuart *et al.*, Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902.e21 (2019). doi: [10.1016/j.cell.2019.05.031](https://doi.org/10.1016/j.cell.2019.05.031); pmid: [31178118](https://pubmed.ncbi.nlm.nih.gov/31178118/)
80. C. Hafemeister, R. Satija, Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol.* **20**, 296 (2019). doi: [10.1186/s13059-019-1874-1](https://doi.org/10.1186/s13059-019-1874-1); pmid: [31870423](https://pubmed.ncbi.nlm.nih.gov/31870423/)
81. K. R. Moon *et al.*, Visualizing structure and transitions in high-dimensional biological data. *Nat. Biotechnol.* **37**, 1482–1492 (2019). doi: [10.1038/s41587-019-0336-3](https://doi.org/10.1038/s41587-019-0336-3); pmid: [31796933](https://pubmed.ncbi.nlm.nih.gov/31796933/)
82. K. Street *et al.*, Slingshot: Cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics* **19**, 477 (2018). doi: [10.1186/s12864-018-4772-0](https://doi.org/10.1186/s12864-018-4772-0); pmid: [29914354](https://pubmed.ncbi.nlm.nih.gov/29914354/)
83. A. Auton *et al.*, A global reference for human genetic variation. *Nature* **526**, 68–74 (2015). doi: [10.1038/nature15393](https://doi.org/10.1038/nature15393); pmid: [26432245](https://pubmed.ncbi.nlm.nih.gov/26432245/)
84. S. Yazar, J. A. Hernandez, Population-scale single-cell eQTL mapping identifies cell type specific genetic control of autoimmune disease. Zenodo (2021); doi: [10.5281/zenodo.5084395](https://doi.org/10.5281/zenodo.5084395)

**ACKNOWLEDGMENTS**

We thank the participants of this study and are grateful to D. MacArthur and C. Goodnow, who provided feedback about this work. **Funding:** This research was supported by a National Health and Medical Research Council Research Fellowship (S.Y.), Practitioner Fellowship (A.W.H.), Career Development Fellowship (J.E.P., 1107599), and Investigator Fellowship (J.E.P., 1175781). K.A.F. is supported by the Alex Gadomski Fellowship, funded by Maddie Riewoldt's Vision. Additional grant support was provided by the National Health and Medical Research Council (1150144 and 1143163), the Australian Research Council (180101405), and the Royal Hobart Hospital Research Foundation. **Author contributions:** Conceptualization: J.E.P., A.W.H.; Methodology: S.Y., J.A.-H., K.W., A.S., K.A.F., A.W.H., J.E.P.; Resources: K.W., M.G.G., Q.L., A.R., T.R.P.T., L.C., K.M., C.C., A.L.C., C.J.Y., K.A.F., A.W.H., J.E.P.; Data analysis: S.Y., J.A.-H., K.W., A.S., M.G.G., K.A.F., A.W.H., J.E.P.; Writing – original draft: S.Y., J.A.-H., K.W., K.A.F., A.W.H., J.E.P.; Writing – review and editing: All other authors; Supervision and project administration: A.W.H., J.E.P.; Funding acquisition: C.J.Y., A.W.H., J.E.P. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** OneK1K single-cell gene expression and genotype data are available via Gene Expression Omnibus (GSE19680). The cell by gene data are available at Human Cell Atlas (HCA) (<https://celxgene.czscience.com/collections/dde06e0f-ab3b-46be-96a2-a8082383c4a1>). GWAS summary statistics for MS are available from the Multiple Sclerosis Genetics Consortium (<https://imsgc.net/>). 1000 Genome Phase 3 (02/05/2013 release) data are available from (<http://ftp.ebi.ac.uk/1000g/ftp/>). Summary statistics for IBD, CD, RA, AS, and T1DM GWASs are available for public access in the following repositories: (i) for IBD and CD from the IBD Genetics Consortium website ([www.ibdgenetics.org](http://www.ibdgenetics.org)), (ii) for RA from the Japanese Encyclopedia of Genetic associations by RIKEN (<http://jenger.riken.jp/en/result>), and (iii) for AS and T1DM from Pan-UK Biobank (<https://pan.ukbb.broadinstitute.org/>). The SLE GWAS summary statistics data can be downloaded from the following link: [http://urr.cat/data/GWAS\\_SLE\\_summaryStats.zip](http://urr.cat/data/GWAS_SLE_summaryStats.zip). There are no known limitations or data transfer agreements regarding access to the summary data described above. Data used to replicate the single-cell eQTLs are available via Gene Expression Omnibus (GSE174188), and genotype data are available via dbGap (46198). Cell by gene data are available at HCA (<https://celxgene.czscience.com/collections/436154da-bcf1-4130-9c8b-120ff9a88812>). Analysis code is available on Zenodo ([484](https://zenodo.org/record/484)).

**SUPPLEMENTARY MATERIALS**[science.org/doi/10.1126/science.abf3041](https://science.org/doi/10.1126/science.abf3041)

Materials and Methods

Figs. S1 to S23

Tables S1 to S19

References (85–108)

MDAR Reproducibility Checklist

[View/request a protocol for this paper from Bio-protocol](https://science.org/view/request-a-protocol-for-this-paper-from-Bio-protocol).

16 October 2020; accepted 22 February 2022

10.1126/science.abf3041