



Supplementary Materials for

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

Seyhan Yazar *et al.*

Corresponding authors: Alex W. Hewitt, hewitt.alex@gmail.com; Joseph E. Powell, j.powell@garvan.org.au

Science **376**, eabf3041 (2022)
DOI: 10.1126/science.abf3041

The PDF file includes:

Materials and Methods
Figs. S1 to S23
Tables S1 to S5
References

Other Supplementary Material for this manuscript includes the following:

Tables S6 to S19
MDAR Reproducibility Checklist

Materials and Methods

Sample Collection and Processing

Donor recruitment: Samples were ascertained from patients or their relatives attending the Royal Hobart Hospital, Hobart Eye Surgeons as well as individuals residing in the retirement villages within Hobart, Australia. All participants gave informed written consent. This study was approved by the Tasmania Health and Medical Human Research Ethics Committee (H0012902).

Total of 1104 individuals were enrolled into the OneK1K study. While genotyping was done for all samples, scRNA-seq was completed for 1034. Initial quality control was completed on each dataset as described below, remaining data for 1034 individuals in genotyping and 1023 individuals in scRNA-seq. Intersection of these two datasets resulted in data for 993 individuals. During demultiplexing of pools, eight individuals were removed with low quality data and cells were assigned to remaining 985 individuals. For eQTL analysis, we removed two participants with less than 10 cells and one individual with active immune disease (giant cell arthritis). Data presented in this study represent samples from 982 individuals (**fig. S1**).

Donor characteristics: Blood samples were collected from both females (58%) and males (42%). Age of participants ranged from 19 to 97, with 73% of participants being older than 60 years old (**fig. S2**). Most of the participants identified themselves with a Northern European ancestry, however, the ancestral relationships were further investigated using genotype information (see “**Generation of Genotypes**”). All but one participant was without active immune disease (giant cell arteritis) at the time of sample collection.

Blood sample collection and processing: Peripheral blood samples were collected in vacutainer tubes containing either FICOLL™ and sodium heparin (8mL CPT™; BD Australia, North Ryde, NSW; 362753) or K₂EDTA (10mL; BD Australia, North Ryde, NSW; Catalogue: 366643). All samples were gently inverted and stored in the dark at ambient temperature until they were processed. All CPT™ tubes were processed within 4 hours of collection according to the manufacturer's recommendations. CPT tubes were centrifuged at 1650 x g for 20 minutes at room temperature. The separated plasma was removed, and the mononuclear layer was transferred to 50mL conical centrifuge tubes, before undergoing two 10-minute washes with 45mL PBS-2mM EDTA (centrifuged at 300 x g). The resultant cell pellet was resuspended in chilled fetal Bovine Serum (F9665; Sigma-Aldrich), RPMI-1640 Medium (R8758; Sigma-Aldrich) and dimethyl sulfoxide (472301; Sigma-Aldrich) (40% FCS, 50% RPMI, 10% DMSO). Cryovials were then transferred to -80°C in controlled-rate cooling containers before being stored in liquid nitrogen. EDTA samples were stored in the dark at 2-4°C until processed.

Generation of Genotypes

DNA sample collection and genotyping: DNA was extracted from EDTA tubes using the QIAamp Blood Mini kit with an elution volume of 50ul. Quality control was performed by spectral reading via Nanodrop, and SNP genotype data generated using the Illumina Infinium Global Screening Array according to the manufacturer's protocol.

Quality control of genotyping data: Genotype data were exported into PLINK data format using GenomeStudio PLINK Input Report Plug-in v2.1.4 and screened for SNP and individual call rates greater than 0.97, Hardy-Weinberg equilibrium failure ($p < 10^{-6}$),

and minor allele frequency greater than 0.01 (85). Samples with excess autosomal heterozygosity (± 3 SD from the mean) or with sex-mismatch were excluded (fig. S3). In addition, a genetic relationship matrix from all the autosomal SNPs was generated using GCTA (86), and one of any pair of individuals with estimated relatedness larger than 0.125 were removed from the analysis. Individuals with non-European ancestry were excluded with ± 6 SD from the European mean of PC1 and PC2. The PCs were calculated using a SMARTPCA (87) analysis. 1000 Genome Phase 3 (83) (02.05.2013 release) population was downloaded from 1000g directory within the EMBL-EBI public end point (<http://ftp.ebi.ac.uk/1000g/ftp/>) and used to define the axes, and the OneK1K samples projected onto those axes. After these quality control steps, we retained 483,482 autosomal SNPs on chromosome 1 to 22 from 1034 individuals. Additionally, we embedded the top four genotype principal components onto a two-dimensional space using Uniform Manifold Approximation and Projection (UMAP). We applied a cosine similarity metric to find the nearest neighbors and set local neighborhood size to 30 individuals (Fig. 1B).

Imputation of genotyping data: Imputation was performed on each autosomal chromosome using the Michigan Imputation Server with the Haplotype Reference Consortium panel (HRC r1.1 2016) (24), and run using Minimac3 and Eagle v2.3 on Michigan Imputation Server(88). Only SNPs with INFO> 0.8 were retained for further analysis. We retained 5,328,917 SNPs with imputation quality R2 > 0.8 and a minor allele frequency > 0.05 for subsequent *cis*-eQTL analyses (fig. S4).

Single-cell RNA Sequencing

Single-cell sample preparation and pooling: A 1mL aliquot per sample of cryopreserved cells was thawed in a 37°C water bath, washed with 9mL Iscove's Modified Dulbecco's Media (IMDM; Life Technologies; 12440061) + 5% Fetal Bovine Serum (FBS; Bovogen; SFBS-FR), and resuspended in 900uL IMDM + 10% FBS. Cells were counted using a Countess II Automated Cell Counter (ThermoFisher; AMQAX1000) and Trypan Blue viability stain (Life Technologies; T10282), and equal numbers of live cells were combined for 12-14 samples per pool. Dead cell removal was performed using the Dead Cell Removal Kit (Miltenyi; 130-090-101) according to the 10x Demonstrated Protocol. Briefly, cell pools were washed, resuspended in Dead Cell Removal magnetic microbeads, and incubated at room temperature for 15 minutes. Cell pools were diluted with Binding Buffer, and dead cells were depleted by running samples through a pre-washed MACS column within a magnetic field and collecting the column run-through. Columns were washed 4x with Binding Buffer, collecting run-through, and cell pools spun down and resuspended in IMDM + 5% FBS for final cell count and viability assessment using the Countess cell counter.

Single-cell RNA-Sequencing: Pooled single-cell suspensions (concentration range 1.16x10⁶ – 2.63x10⁶, viability 82-98%) were partitioned and barcoded using the 10x Genomics Chromium Controller (10x Genomics) and the Single Cell 3' Library and Gel Bead Kit (10x Genomics; PN-120237). The pooled cells were loaded onto the Chromium Single Cell Chip A (10X Genomics; PN-120236) to target the capture of 20,000 cells per well. GEM generation and barcoding, cDNA amplification, and library construction were performed according to the 10x Genomics Chromium User Guide. Reactions were performed in a C1000 Touch thermal cycler with a Deep Well Reaction Module (Bio-

Rad). 13 cDNA amplification cycles were performed, and half of the cDNA was used as input for library construction. 11-12 SI-PCR cycles were used depending on the amount of input cDNA. The resulting single cell transcriptome libraries contained unique sample indices for each pooled sample. Libraries for all samples were multiplexed and sequenced across five 2×150 cycle flow cells on an Illumina NovaSeq 6000 (26bp (Read 1), 8bp (Index), and 98 bp (Read 2)), generating a total of 49 billion reads.

Quantification and alignment of single-cell 3' RNA-seq data: The Cell Ranger Single Cell Software Suite (version 2.2.0) by 10x Genomics was used to process raw sequence data into FASTQ files. First, raw base calls from multiple flow cells were demultiplexed into separate pools of samples. Reads from each pool were then mapped to the GRCh37/hg19 genome (release 84) using STAR (76). The count data was processed and analyzed in R (89) as described below. Quality metrics for all pools are included in **Table S6**.

Demultiplexing scRNA-seq pools: Cells from each pool were assigned to individual donors using the Demuxlet method (77). Demuxlet implements a maximum likelihood-based approach to estimate the probability that a set of SNP genotypes (determined from SNP array sequencing) match the observed genotypes from the 3' RNA sequence reads assigned to a cell-droplet. If genotypes identified in a cell-droplet match that of a single donor, then the cell-droplet is assumed to be a single cell and labelled as a singlet. However, if genotypes from multiple donors are identified within the same cell-droplet, then the cell-droplet is labelled as a doublet and assigned to multiple individuals based on the likelihood ratio of the expected proportion of cells (α) from each individual. If there is insufficient number of SNP genotypes called from a cell-droplet, then the droplet is labelled as ambiguous. For the singlet cell-droplets, we assigned the most likely individual for each droplet using the genotype posterior probability estimate against 265,053 exonic SNPs (MAF>0.05).

Quality control and normalization of the scRNA-seq: For the 376,231 cell-droplets that were initially identified as doublets, a total of 234,396 also had a posterior probability of being a singlet of greater than 0.99. We performed additional analyses using Scrublet (78) to provide further evidence for these cell-droplets being either singlets or doublets. Scrublet was applied to all cell-droplets (n=1,420,567) that had been identified as a singlet or doublet in the Demuxlet analysis. Scrublet identifies doublets by generating simulated doublets through linear combinations of randomly sampled observed transcriptomic profiles of a given number of genes. Of the 1,420,567 cell droplets, Scrublet classified 1,295,408 as singlets and 125,159 as doublets. The cell-droplets identified as doublets by both Demuxlet and Scrublet were removed from further analysis.

For each pool of cells captured, the distributions of total number of UMIs, number of genes, and percentage of mitochondrial gene expression were normalized using an Ordered Quantile Transformation as follows:

$$g(x) = \Phi^{-1} \left(\frac{\text{rank}(x) - 1/2}{\text{length}(x) + 1} \right) \%0$$

The percentile of each value in the distribution was calculated and then mapped to the same percentile in a normal distribution using the inverse of the standard normal

cumulative density function. The Z-scores corresponding to three negative units were mapped back to the original distributions using a binomial generalized linear model with a logit link (90). Cells below the Z-score limits were considered as outliers and removed from further analysis. For the expression of mitochondrial genes, an additional upper threshold corresponding to two Z-scores was used to exclude apoptotic cells. After quality control, 1,272,518 cells were kept and 176,928 were filtered out. The average read depth of cells retained for further analysis was 34,000.

To account for the effect of variable sequencing depth, we stabilized the variance of the gene UMI count matrix using the SCTransform method(80). First, a generalized linear model with negative binomial error distribution and log link function was used to model the relationship between the counts as the response for each gene and the sequencing depth per cell as the explanatory variable. Then, a kernel regression was applied to the resulting parameters (intercept, slope, and dispersion of the negative binomial errors) for each gene to regularize the parameters with respect to the mean gene expression. The sequencing depth was regressed from the count matrix using the regularized parameters. To account for batch effects, we included the pool as well as the percentage of mitochondrial expression from each cell as covariates. Finally, the residuals were divided by the expected standard deviation to obtain Pearson residuals.

Cell classification methodology: Cells were classified into one of 14 cell types. First, we guided the classification of each cell in the OneK1K cohort into the major immune populations in a supervised manner using the gene expression data by Zheng et al. (26) as a reference. We built reference signature matrices for each of the levels in the cell-type hierarchy corresponding to lymphoid/myeloid, B cell/sub-lymphoid (B and plasma cells/T and NK cells), T cell/NK cell, and CD8+/CD4+ T cells (**Fig. 1D**). Signature matrices (S) were constructed by obtaining the differentially expressed genes between the cell-type populations for each of the four levels (e.g. lymphoid vs. myeloid) using a Wilcoxon Rank-Sum test (p -value < 0.05) and averaging the expression values for each gene across all cells as shown below:

$$S_n = \begin{bmatrix} x_{11} & \dots & x_{1j} \\ x_{21} & \dots & x_{2j} \\ \vdots & \ddots & \vdots \\ x_{i1} & \dots & x_{ij} \end{bmatrix} \%0$$

Where the rows correspond to the i^{th} differentially expressed gene and the columns the j^{th} cell type in a given level of the hierarchy. Each x_{ij} entry represents the mean expression of gene i in cell type j . In total, four signature matrices S_n were created. For every cell in our cohort, the cosine similarity was calculated against all cell-types in the signature matrices as follows:

$$\cos(\theta) = \frac{\mathbf{C} \cdot \mathbf{R}}{\|\mathbf{C}\| \|\mathbf{R}\|} \%0$$

Where \mathbf{C} represents a vector of gene expression values for a single cell, and \mathbf{R} is a column-vector from the signature matrix S corresponding to a cell-type. Each cell was labelled according to its highest cosine distance to each cell type using the hierarchy as a decision tree. For instance, a single-cell was first classified as myeloid or lymphoid based

on the cosine similarities to the column-vectors of the signature matrix **S₁**. If the cell was more similar to the lymphoid lineage as determined by the cosine distance, then **S₂** was used to sub-classify the cell as B cell or sub-lymphoid. All sub-lymphoid cells were further labelled as Natural Killer or T cell using **S₃**. Finally, the cytotoxic status of T cells was determined using **S₄**.

Once cells were assigned to the major immune cell-types, unsupervised classification was performed at every end of the hierarchy by naively clustering cells using the graph-based clustering method implemented in Seurat version 3.0 (79). A Shared Nearest Neighbor (SNN) graph was built using Euclidean distances from the principal component analysis embeddings using the first 50 PCs. Cells were then clustered using the Louvain algorithm (91). Discriminant genes for the major clusters were obtained using a Wilcoxon Rank-Sum test (*p*-value < 0.05) and all cells from each group were labelled according to the top differentially expressed genes in each cluster. If cells were incorrectly labeled using the supervised approach - as determined by the expression of classic markers (**table S1**) - then they were reassigned based on their marker expression. Using the Seurat package (version 3), we have visualized the relationship of 14 cell types using UMAP (**Fig. 1G**) and PC (**fig. S5**) based clustering methods.

Platelets and erythrocytes were removed based on the expression of PPBP and HBB genes respectively for each branch after clustering. After platelet and erythrocyte removal, 1,267,768 cells were retained for analysis.

Description of cell populations identified in the OneK1K cohort: We classified CD4⁺ T cells into three distinct populations. Population 1 contained naïve and central memory cells, population 2 contained effector memory and central memory T cells, and population 3 contained a transcriptionally distinct population expressing SOX4. CD4⁺ T cells with a naïve or central memory phenotype (CD4_{NC}) expressed high levels of CCR7, SELL and LRRN3. This is contrasted to CD4⁺ T cells with an effector memory or central memory phenotype (CD4_{ET}) that are characterized by high expression of KLRB1, GZMK, and TNFSF13B and low expression of CCR7 and SELL. Our classification framework identified a population of T cells (CD4_{sox4}) characterized by high expression of SOX4, ID2 and SELL, which show distinct transcriptional profiles to CD4_{NC} and CD4_{ET}. SOX4 is a transcription factor that in CD4⁺ T cells is associated with chronic inflammation and implicated in the formation of ectopic lymphoid-like follicles, such as those seen in rheumatoid arthritis (92).

Akin to CD4⁺ T cells, CD8⁺ T cells were also classified into three independent populations, with naïve and central memory cells having independent transcriptional signatures compared to effector memory cells, in addition to a third transcriptionally distinct population. The CD8⁺ T cells with an effector memory phenotype (CD8_{ET}) are characterized by high GNLY and NKG7 expression, while the CD8⁺ T cells with a naïve or central memory phenotype (CD8_{NC}) were characterized by high LTB, CCR7 and PASK expression. Our classification framework identified a population of CD8⁺ T cells (CD8_{S100B}) that were characterized by high expression of LTB, IL7R, KLRB1, GZMK, SLC4A10 and S100B expression. S100B is secreted by CD8⁺ S100B⁺ T cells and, among other functions, acts as a ligand of the receptor for advanced glycosylation end-products (RAGE), and yields activation of granulocytes and monocytes *in vitro* (93), raising the

possibility of a functionally distinct cytotoxic T-cell compartment. *SLC4A10* is a marker of MAIT cells, and was expressed in approximately 9% of CD8_{S100B} cells.

Cells within the B cell lineage were classified as plasma cells, immature and naïve B cells, or memory B cells. Immature and naïve recirculating B cells (B_{IN}) were characterized by the expression of *TCL1A*, *FCER2* (*CD23*) and *IL4R*, while memory B cells (B_{Mem}) were characterized by *TNFRSF13B* (*TACI*) and *CD27* expression. Plasma cells were demonstrated to have high expression of IgJ and *TNFRSF17* (*BCMA*). Innate immune lymphocytes were disaggregated into five independent types, representing two populations of NK cells, both classical and non-classical monocytes, and dendritic cells. NK cells (NK) were characterized by expression of *GZMA* and *GZMB* levels, while NK recruiting (NK_R) have high expression of *XCL1*, *XCL2*, and *GZMK* and were designated as NK_R as *XCL1* and *XCL2* are chemokines that will recruit DCs and subsequently activate T cells (94). These two populations closely resemble two populations previously described as hNK_{BL1} and hNK_{BL2}. The hNK_{BL2} cell type are similar to our NK_R and were proposed to resemble CD56^{bright} NK cells, whilst the NK subpopulation have an expression profile matching the previously described hNK_{BL1} cells resembling CD56^{dim} NK cells (95).

The classical monocytes (Monoc) had high expression levels of *CD14* and *LYZ*, while non-classical monocytes (MonoNC) expressed *CD16* (*FCGR3A*). Due to the number of dendritic cells (DC) captured per individual, we chose not to classify sub-types, and instead DC were characterized by *CST3*, *FCER1A*, and *SERPINF1* expression. Further details of the classification nomenclature and commonly used canonical markers is provided in **table S1**. Expression of these markers are displayed in **fig. S6**.

Investigation of potential batch effects on cell type composition: With our sample multiplexing approach cells from 12-14 donors were pooled per library capture. We generated data for 75 independent pools, with 2 pools replicated (77 libraries in total). To identify any potential batch effects, we have first evaluated cell type composition by pool as shown **fig. S7**. The proportions of cell types were consistent among different pools. Pool was also included as a covariate in the normalization of scRNA-seq data. Additionally, cell clustering was similar among pools as visualized by the UMAP projections (**fig. S8**).

Identification of covariates influencing individual-specific cell composition and gene expression

Individual-specific cell compositions: On average we have identified ~1291 cells per person (range: 62 to 3501, **Fig. 1C**). Distributions of cell type compositions varied among 14 cell types. While smaller and less common cell types had a kurtotic distribution, larger and more common cell types were platykurtic (**table S2**). Proportions of different cell types per individual were consistent throughout the cohort as depicted in **fig. S9**. Both cytotoxic and non-cytotoxic T cells accounted for greater composition of the cell types identified at the individual level.

Effect of sex on individual-specific cell compositions: We observed significant variations in the mean number of cells by gender in some cell populations (**fig. S10**). To account for this, we included gender as a covariate in our eQTL detection models.

Effect of age on individual-specific cell compositions: Age distribution of cohort is shown in **fig. S2**. We assessed the effect of age on cell counts. Cell-type specific

association estimates are shown in **fig. S11** and **table S3**. The largest effect of age was detected on CD8_{NC} cells ($\beta=-1.88, p<0.001$). We have binned the cohort into two equal groups based on age and run *cis*-eQTL analysis in CD8 NC cells for the two groups to exemplify the effect of age in eQTL analysis output. The analysis yielded a greater number of eQTLs in the younger group compared to older one (5494 eQTLs in the younger group vs 3068 in the older group), and consequently included age as a covariate in our eQTL analyses.

Effect of population structure: Population heterogeneity within the OneK1K cohort was controlled during sample collection. However, to remove any remaining global effects due to population structure on gene expression that could mask or skew the local effect of genetic variation, we calculated the genotype principal components (PCs) across 982 OneK1K cohort samples (combined with 1000G samples as a reference panel). We included the first six PCs as covariates in the eQTL analyses; these were sufficient to explain the population structure within the OneK1K cohort (**Fig. 1B**).

Effect of hidden or unmeasured factors: To identify unmeasured factors that could result in batch variation of gene expression, we applied the Probabilistic Estimation of Expression Residuals (PEER) method and generated 10 PEER factors per cell type. We have tested the effect of 1-10 PEER factors sequentially by adding one at time into a linear regression test model of CD8_{NC} cell *cis*-eQTL analysis. After addition of the first two PEER factors, there was a significant drop in the number of eGenes identified for CD8_{NC} cells (**table S4**). For the remaining eQTL analyses across cell types we included the first two PEER factors as covariates in the model.

Single cell *cis*-eQTL mapping

Within each cell type, *cis*-eQTLs were identified through Spearman's rank correlation testing using expression levels adjusted for sex, age, six genotype-based principal components, and two PEER factors. *p*-values were calculated from the *t*-distribution with *n*-2 degrees of freedom under the null hypothesis. We restricted our search to variants within +/- 1Mb of either end of a gene. Using the *qvalue* package by Storey (96), we have determined a specific nominal *p*-value (*q*-value) threshold for each gene at a false discovery rate (FDR) threshold of 5% and rank these *q*-values within each chromosome in a cell type to identify the top SNP affecting the (residual) expression of the tested gene within a specific cell type. Significance of the resulting SNP-gene pairs was determined at a false discovery rate (FDR) threshold of 5% within each chromosome. In the first round of eQTL analysis, we identified the top SNP (eSNP₁) for each eQTLs based on the smallest *p*-value. We subsequently performed conditional eQTL analysis by regressing out the effect of eSNP₁ sequentially and repeating the test to identify secondary eQTLs. In this second round of analysis significant secondary eQTLs were identified using the same criteria, and the top eSNP denoted as eSNP₂. These steps were repeated up to five rounds of conditional analysis, identifying up to five independent eQTLs per gene. The top eSNPs for these eQTLs are thus termed eSNP₁ to eSNP₅. Further details are provided below.

Overview of gene expression values and selection of genes for testing: The average expression of each gene per person across all genes available for each cell type was calculated using the normalized counts. We calculated the number of individuals with non-zero expression for each gene and filtered genes expressed in less than 10% of the

cohort within a particular cell type, meaning that the exact genes tested varied between cell types (**table S8**). All values were then log transformed ($\log x+1$). Only autosomal genes and genes expressed at least in one cell within each cell type were included.

Adjustment of gene expression values for confounders and generation of residual expression values: To account for the effect of covariates identified, we have calculated the residual gene expression for each person for each gene by linear regression modelling:

$$\hat{G}_X = \beta_0 + \beta_1 \cdot \text{sex} + \beta_2 \cdot \text{age} + \beta_3 \cdot PC1 \cdots + \beta_7 \cdot PC6 + \beta_8 \cdot PF1 + \beta_9 \cdot PF2$$

$$e_{X_0} = G_X - \hat{G}_X$$

\hat{G}_X is a matrix consisting of the average expression of gene X per individual. e_{X_0} is the matrix including residual expression of gene X after adjusting for sex, age, six genotyping principal components, and two PEER factors.

Conditional cis-eQTL analysis iteratively for five rounds: In the first round of analysis, we identified the primary (top) eQTL in cell type A:

$$\text{for each SNP and } e_{X_0} \text{ pair, } \varrho = 1 - \frac{6 \sum d^2}{n(n^2 - 1)}$$

ρ is the correlation between e_{X_0} and matrix of three genotypes coded as 0, 1 and 2 where 2 representing the assessed allele for each of 5,328,917 SNPs and q is the associated q -value. d is the difference between two rankings of residuals (e_A) and n is the number of measurements.

$$\begin{bmatrix} SNP_v & \rho_1 & q_1 \\ SNP_w & \rho_2 & q_2 \\ SNP_x & \rho_3 & q_3 \\ SNP_y & \rho_4 & q_4 \\ SNP_z & \rho_5 & q_5 \\ \vdots & & \end{bmatrix} \xrightarrow[\text{ranking}]{\text{q-value}} \begin{bmatrix} SNP_x & \rho_3 & q_3 \\ SNP_y & \rho_4 & q_4 \\ SNP_w & \rho_2 & q_2 \\ SNP_z & \rho_5 & q_5 \\ SNP_v & \rho_1 & q_1 \\ \vdots & & \end{bmatrix} \xrightarrow[\text{determined}]{\text{top SNP}} \begin{bmatrix} SNP_x = eSNP_1 \\ \vdots \end{bmatrix}$$

In the second round of analysis, we identified the secondary eQTL (first conditional eQTL) in cell type A:

$$\hat{e}_{X_0} = \beta_0 + \beta_1 \cdot eSNP_1$$

$$e_{X_1} = e_X - \hat{e}_{X_0}$$

$$\text{for each SNP } e_{X_1} \text{ pair, } \varrho = 1 - \frac{6 \sum d^2}{n(n^2 - 1)}$$

\hat{e}_{X_0} is the residual expression matrix from the first round of analysis. e_X, e_{X_1} is the matrix including residual expressions after regressing $eSNP_2$ genotype. ρ is the

correlation between e_{X_1} and matrix of three genotypes for each of 5,328,916 (5,328,917 – 1) SNPs and q is the associated q -value.

$$\begin{bmatrix} SNP_v & \rho_1 & q_1 \\ SNP_w & \rho_2 & q_2 \\ SNP_y & \rho_3 & q_3 \\ SNP_z & \rho_4 & q_4 \\ \vdots & & \end{bmatrix} \xrightarrow{\substack{q\text{-value} \\ ranking}} \begin{bmatrix} SNP_z & \rho_4 & q_4 \\ SNP_v & \rho_1 & q_1 \\ SNP_w & \rho_2 & q_2 \\ SNP_y & \rho_3 & q_3 \\ \vdots & & \end{bmatrix} \xrightarrow{\substack{\text{top SNP} \\ determined}} \begin{bmatrix} SNP_z = eSNP_2 \end{bmatrix}$$

In the third round of analysis, we identified the third eQTL (second conditional eQTL) in cell type A:

$$\begin{aligned} \hat{e}_{X_1} &= \beta_0 + \beta_1 \cdot eSNP_2 \\ e_{X_2} &= e_X - \hat{e}_{X_1} \\ \text{for each SNP } e_{X_2} \text{ pair, } \varrho &= 1 - \frac{6 \sum d^2}{n(n^2 - 1)} \end{aligned}$$

\hat{e}_{X_1} is the residual expression matrix from the second round of analysis. e_{X_2} is the matrix including residual expressions after regressing $eSNP_3$ genotype. ρ is the correlation between e_{X_2} and matrix of three genotypes for each of 5,328,915 SNPs and q is the associated q -value.

$$\begin{bmatrix} SNP_v & \rho_1 & q_1 \\ SNP_w & \rho_2 & q_2 \\ SNP_y & \rho_3 & q_3 \\ \vdots & & \end{bmatrix} \xrightarrow{\substack{q\text{-value} \\ ranking}} \begin{bmatrix} SNP_v & \rho_1 & q_1 \\ SNP_w & \rho_2 & q_2 \\ SNP_y & \rho_3 & q_3 \\ \vdots & & \end{bmatrix} \xrightarrow{\substack{\text{top SNP} \\ determined}} [SNP_v = eSNP_3]$$

In a similar way, $eSNP_4$ and $eSNP_5$ were determined in the fourth and fifth round of analyses, respectively.

Benefit of conditional cis-eQTL analysis: In many eQTL catalogs, gene expression is often strongly associated with multiple SNPs which may often simply reflect linkage disequilibrium among multiple signals. Our iterative conditional analysis procedure at cell type level allowed us to identify independent eQTL signals. An example of additional information gained through this approach is described in **fig. S12**.

Independent lead eQTL analysis – cell sharing: In instances where a gene had an eQTL tagged by different eSNPs in multiple cell types (**Fig. 2A** and **fig S12**), we undertook multi-directional conditional analyses to ascertain whether the loci associated with its gene expression were cell type specific or shared between cells. For example, if gene X has an eQTL in cell type A tagged by $eSNP_A$ and cell type B tagged by $eSNP_B$, then we tested whether the allelic effects of $eSNP_A$ and $eSNP_B$ were dependent on each other in relation to gene expression in their relevant cell types. To do so, we included $eSNP_A$ as a fixed covariate to retest the association of $eSNP_B$ in cell type B and vice versa as follows:

$$\hat{G}_A = \beta_0 + \beta_1 \cdot sex + \beta_2 \cdot age + \beta_3 \cdot PC1 + \dots + \beta_7 \cdot PC6 + \beta_8 \cdot PF1 + \beta_9 \cdot PF2 + \beta_{10} \cdot eSNP_B$$

$$e_{X_A} = G_A - \hat{G}_A$$

$$\text{for each SNP and } e_{X_A} \text{ pair, } \varrho = 1 - \frac{6 \sum d^2}{n(n^2 - 1)}$$

where d is the difference between two rankings of residuals (e_A) and n is the number of measurements. eSNPs are regarded as independent if the correlation (ρ_A) between e_A and $eSNP_A$ was significant. We then tested for the change in the allelic effect of an eSNP between the original analysis and the retest analysis which included the top eSNP from another cell type. Under scenarios where there was a significant change in the allelic effect, we conclude that the eSNPs either tag the same causal variant influencing gene expression in both cell types, and/or are in linkage disequilibrium with one another. Alternatively, if the allelic effect does not have a significant change in effect, then we conclude that the eSNPs tag independent causal variants for the same gene in different cell types (**Fig. 1D**).

We also tested to what extent the test statistics were inflated for SNPs that were cell type specific eSNPs in another cell type by using the *q*-value package (96). In this method, we extracted *p*-values for all unique SNP-gene pairs in each cell type and applied the ‘*q*-value’ function to generate the estimated proportion of non-null *p*-values by keeping the default setting ‘smoother’ option. A conservative estimate of the overall proportion of null *p*-values for every cell type pair is displayed in **fig. S13**.

Effect of cell type cluster size on eQTL discovery: To understand the effect of cluster size on eQTL discovery, we have subsetted our largest cell type cluster CD4_{NC} into smaller clusters by taking 1, 5, 10, 25, 50 and 75% of cells per person and performed conditional eQTL analysis as described in section 5.3. As shown in **fig. S17**, the number of eQTLs discovered in CD4_{NC} cells has increased with an average of larger numbers of cells per donor.

Overlap of cis-eQTLs with GWAS of Cell Composition: We investigated whether any of the eQTL interact with cell composition by performing a GWAS of cell composition and directly overlayed these findings with the lead *cis*-eQTLs from our conditional analysis for each cell type. Cell type composition was determined for each person with each cell type using the following calculation:

$$\text{cell type proportion} = (\text{number of cell type of interest}/\text{total number of cells})$$

A linear regression model in PLINK version 1.9 was used to determine associations between SNPs and cell proportion for each cell type. The model was adjusted for age, sex, the first six principal components that accounted for the population stratification and pool number. From the identified associations, we extracted SNPs that correspond to the top *cis*-eSNPs (eSNP₁) in our conditional analysis (**table S11**). We then applied a Bonferroni correction at a significance threshold of $p < 3.0 \times 10^{-6}$.

Enrichment of *cis*-eQTLs in regulatory regions

*Fine mapping of single cell *cis*-eQTLs with scATAC-seq:* We performed quality control filtering on single cell ATAC-seq data from 9668 cells (https://cf.10xgenomics.com/samples/cell-atac/1.2.0/atac_pbmc_10k_nextgem/*) using

Signac (97). Cells with less than 200 detected features were removed, and features not expressed in at least ten cells were discarded. Next, we applied the standard quality control suggested in Signac and filtered cells with a TSS enrichment score greater than 2, a proportion of mono-nucleosomal fragments less than four. Additionally, we retained cells with a cell sequencing depth between 3000 and 20,000 UMIs, a fraction of fragments in peaks greater than 0.15, and a ratio of reads in genomic blacklist regions defined in ENCODE lower than 0.05. After quality control, we retained 8876 cells. Next, we created a gene activity matrix based on the scATAC-seq peaks. We counted the fragments within the gene body and the upstream promoter region spanning two kilobases for each gene-cell pair, resulting in 20,233 gene activities across 8876 cells. Finally, we log-normalize the activity matrix.

We integrated the scATAC-seq and the OneK1K data with Seurat (v4) by performing a canonical correlation analysis (CCA) and identifying the mutual nearest cell neighbours between both datasets. For data integration, we used only two pools (1 and 54) from the OneK1K scRNA-seq data for a comparable sample size with the scATAC-seq data. Next, cell type labels were transferred from OneK1K cells to the scATAC-seq data by computing a weights matrix between OneK1K cell anchors and the cells in the scATAC-seq data. CD4_{ET}, CD4sox4, and Plasma cells were removed from the analysis as their number of cells in the scATAC-seq was lower than 5 (**fig. S18**). Then, we called peaks for each OneK1K cell type using MACS2 (98).

The genomic distance between each *cis*-eQTL and its nearest peak across each cell type was calculated. Likewise, the distances between each SNP tested in the original analysis and its nearest peak was determined. Null distributions were determined by ten-fold bootstrap replicates sampling the same number of randomly selected SNPs as *cis*-eSNPs in each cell type. Next, we compared *cis*-eQTL distances to null distributions drawn from the same distance distribution using a *t*-test across all cell types.

*Proximity of single cell *cis*-eQTLs to gene body:* For every *cis*-eQTL in each cell type we calculated the genomic distance between the eGene and its cognate *cis*-eSNP within +/- 1Mb. To create a null distribution to compare with, we randomly sampled as many *cis*-SNPs as eQTLs were associated with each eGene in each cell type. Next, for every cell type, we compared the eQTL versus the random *cis*-SNP distances using a one-tailed Wilcoxon rank-sum test and applied FDR correction across all cell types (FDR < 0.05) (**fig. S19B**).

Replication of single cell *cis*-eQTLs

*Identification of lead *cis*-eQTLs in two replication cohorts:* To validate the eQTLs identified in the OneK1K cohort a replication analysis was performed in a separate single cell RNA-seq cohort. The cohort used contained 113 European and 89 Asian individuals, from a larger cohort containing 355 samples. Of the 202 individuals included in this analysis, 94 individuals were diagnosed with lupus but were not in an active state of the disease at the time of collection. Individuals with lupus were collected through the California Lupus Epidemiological Study (CLUES) cohort and healthy individuals were collected at the UCSF Rheumatology Clinic.

The UCSF Core Immunologic Laboratory (CIL) Ficoll separated and cryopreserved all PBMCs collected. Frozen PBMCs were profiled in 23 pools across four processing batches. PBMCs were thawed, counted using a Vi-CELL XR (Beckman Coulter) to

ensure precise and rapid cell counting, resuspended, and pooled as previously described in (77), then cells were loaded onto a 10x Chromium instrument. Following library preparation using the published 10x Chromium Single Cell 3' V2 Solution protocol, libraries were sequenced on an Illumina NovaSeq6000 at a depth of 6306-29,862 reads/cell using the recommended cycle numbers Read 1 26 cycles, i7 index 8 cycles, Read 2 98 cycles.

Individuals were genotyped on the Affymetrix World LAT Array. The Analysis Power Tools (APT) were used to process all CEL intensities, using the Axiom Best Practices Genotyping Analysis Workflow (99). CEL files were grouped by batch and DQC scores were calculated using apt-genotype-qc. CEL files with DQC scores greater or equal to 0.82 were retained. The apt-genotype-axiom command was used with the Step1 file for the corresponding to generate call rates. Array and samples with a call rate greater or equal to 97% were retained. Genotypes were called using apt-genotype-axiom with the Step2 file for the corresponding Array. Per SNP QC metrics were generated with ps-metrics and classified with ps-classification. The resulting Recommended.ps set of 800,825 probes were retained. The probes were evaluated for missingness and heterozygosity, but no individuals were removed due to these metrics. 800,825 probes were input into the Michigan Imputation Server's web portal (88), which were further filtered to 696,523 probes when accounting for invalid alleles, multiallelic sites, monomorphic sites, allele mismatches, and SNP call rates less than 90%. A total of 16,176,651 with an R2 > 0.2 were imputed from the resulting probes using the Haplotype Reference Consortium version 1.1. Further filtering of the dataset was conducted to only include SNPs with a MAF>0.1, resulting in a final 4,372,256 SNPs.

A total of 1,870,857 cell-containing droplets were obtained after sequencing. Freemuxlet was used with an error probability of 0.1 to assign each cell to a donor of origin, resulting in 1,444,450 singlets (doublet rate 22.12%, expected 22-25%). Additional removal of doublets from the same individual was performed with Scrublet (67,969 droplets, doublet rate 4.71%; doublet threshold 0.15). Contaminated platelets were removed by removing all cells with > 1 count of platelet gene markers (PF4, SDPR, GNG11, PPBP). This also removed Megakaryocytes from our analysis. Red blood cells (RBC) were removed using Leiden clustering to identify a population with very high expression of RBC markers like hemoglobin genes (HBB). In total, 1,263,676 cells remained in the final dataset after platelet and RBC removal (112,805 cells). *Scanpy* version 1.6's COMBAT (100) implementation was used to correct for batch correction and then the nUMIs and percentage of mitochondrial counts were regressed out of the dataset. After cell filtering and expression normalization using default parameters, we performed k-nearest neighbor (knn) graph construction, Leiden clustering, and projected the data onto two UMAP dimensions. Leiden communities were assigned to one of 11 cell types by comparing differentially expressed genes between clusters and canonical markers for each cell type. Each of these cell types were used to create pseudo-bulk expression profiles for each individual in each cell type. This was computed by summing all of the counts for each gene and dividing it by the total number of counts across all genes, then using the edgeR software package (101) to calculate the log2 fold change.

Replication was subsequently performed on the 26,597 lead *cis*-eQTL eSNP-eGene pairs and corresponding pseudo-bulk expression data (eight cell types) using the same eQTL mapping strategy applied to the discovery cohort. First, residual expressions are

generated using sex, age, first six genotype PCs, two PEER factors, SLE status and batch number in a linear regression model. Then, Spearman's rank correlation testing was conducted for cis eSNP-eGene pairs within each cell type. **table S5** shows the summary of this replication study. Complete list of replicated eQTLs can be found for both cohorts on **table S13 & S14**.

Replication of lead cis-eQTLs in bulk RNA-seq eQTL datasets: We used two data sources (eQTLGen and GTEx Consortium) to replicate OneK1K *cis*-eQTLs in bulk RNA-seq discoveries. The latest release of eQTLGen consortium (34) on whole blood eQTLs from more than 32,000 samples and whole blood eQTLs identified in European samples in the latest version of GTEx data (Whole_Blood.v8.EUR) (5) were selected (**fig. S20**). Of the 26,597 *cis*-eQTL we identify 16,615 associations in eQTL-Gen, and 10,759 associations in GTEx at the study specific significance threshold (**table S14**). The overlapping eQTL are more likely to be identified in multiple cell types on OneK1K, and thus would be expected to have larger allelic effects on average in bulk samples.

Dynamic eQTL mapping in B cells

We subsetted all 130,091 B_{in} and B_{mem} cells and removed 5123 outlier contaminated cells expressing CD34 and T-cell markers resulting in 124,968 B cells. Next, we computed scaled expression Pearson residuals for the 500 most variable genes applying SCTtransform(80), using pool and percentage of mitochondrial expression as covariates. Finally, we performed a principal component analysis and clustered cells using 30 principal components with the Louvain algorithm (91).

To capture the continuous cell state transitions, we built a two-dimensional space from the 500 most variable genes using PHATE (potential of heat diffusion for affinity-based transition embedding) (81). Next, we described the continual progression from immature/naive to memory B cells by ordering cells in a pseudotime curve with slingshot (82) (**Fig. 4**). Then, we grouped B cells into six quantiles across the pseudotime trajectory and identified dynamic eQTLs using both linear and nonlinear models.

Linear dynamic eQTLs in B cells: To identify linear dynamic changes of genetic effect sizes across naive-memory trajectory, we evaluated significant interaction between genotype and pseudotime on transcript levels using a similar approach proposed by Strober *et al*(102) and Nathan *et al* (103). In this method, linear dynamic eQTLs were detected using a Gaussian linear mixed model with individual specific random effects that assessed the interaction between genotype and quantile representing the pseudotime trajectory, while controlling for the linear effects of both genotype and quantile. Additionally, we included the same set of covariates from our main analysis as fixed effects (sex, age, genetic PCs, first two peer factors from each quantile).

Out of the 2343 B cell eQTLs identified in the primary *cis*-eQTL mapping, we retained eQTLs for which the eGene was expressed in at least 3 quantiles (resulting in 1988 eQTLs) and exclusively tested for this set of variant-gene pairs in order to detect pseudotime specific eQTLs. To determine whether an eQTL was dynamic, we used the 'lmer' function from lme4 R package (104) and first fitted the following linear mixed model accounting for multiple quantile observations per individual (full model):

$$\begin{aligned}
G = & \beta_0 + \beta_1 \cdot geno + \beta_2 \cdot sex + \beta_3 \cdot age + \sum_{j=1}^6 \beta_{3+j} \cdot PC_j + \sum_{k=1}^2 \beta_{9+k} \cdot PF_j \\
& + \beta_{12} \cdot pseudotime \\
& + \beta_{13} \cdot geno \cdot pseudotime \\
& + \gamma_{individual} + \varepsilon
\end{aligned}$$

Where G corresponds to the eGene expression, $geno$ is the genotype information of each individual, $pseudotime$ is a vector indicating each of the six quantiles across the trajectory, $genotype \times pseudotime$ is the interaction term quantifying the interaction between genotype and pseudotime, and $\gamma_{individual}$ is the random effect for individuals. Then, we fitted a reduced model by dropping the $genotype \times pseudotime$ interaction and computed a likelihood ratio test for the full and reduced model using R `anova'. Finally, we applied a multiple testing correction for the resulting p-values ($FDR < 0.05$) using the *quantile* R package to assess genome-wide significance based on a vector of observed p-values and a vector of permuted p-values.

Nonlinear dynamic eQTLs in B cells: Next, we aimed to identify non-linear dynamic eQTLs, where the genetic effect increases the expression levels of a gene before decreasing or disappearing completely. To test for such nonlinear changes of genetic effect sizes, we added a second order polynomial to the above linear dynamic model as shown below (full mode):

$$\begin{aligned}
G = & \beta_0 + \beta_1 \cdot geno + \beta_2 \cdot sex + \beta_3 \cdot age + \sum_{j=1}^6 \beta_{3+j} \cdot PC_j + \sum_{k=1}^2 \beta_{9+k} \cdot PF_j \\
& + \beta_{12} \cdot pseudotime + \beta_{13} \cdot pseudotime^2 \\
& + \beta_{14} \cdot geno \cdot pseudotime + \beta_{15} \cdot geno \cdot pseudotime^2 \\
& + \gamma_{individual} + \varepsilon
\end{aligned}$$

where $pseudotime^2$ is the vector of quadratic values of corresponding quantile and $genotype \times pseudotime^2$ is the quadratic interaction term between genotype and pseudotime. Similar to linear dynamic eQTL mapping, we constructed a reduced model by dropping the linear and quadratic interaction and computed a likelihood ratio test with a study-wide FDR correction ($FDR < 0.05$). We note that our study is underpowered to test other non-linear dynamic eQTLs that do not assume a continuous change in genetic effects with pseudotime.

We grouped the dynamic eQTLs into *linear* (30), *linear & quadratic* (195), and *quadratic* (108) if they were significant for any of these categories. Additionally, we performed eQTL mapping in each of the six quantiles using *Matrix eQTL* to estimate effect sizes for each of the dynamic B cell eQTLs and visualised their scaled values in each quantile using the *ComplexHeatmap* R package (105). We interrogated these lineage-dynamic eQTLs by overlapping with the GWAS hits from seven autoimmune diseases Type 1 diabetes mellitus (T1DM), ankylosing spondylitis (AS), Crohn's disease (CD), inflammatory bowel disease (IBD), multiple sclerosis (MS), rheumatoid arthritis (RA), and systemic lupus erythematosus (SLE)(**Fig. 4C&D**) using the *LDlinkR* R package (106) considering SNPs in LD ($r^2 > 0.8$) with the dynamic eQTLs.

The identification and interpretation of dynamic single cell eQTL has some limitations. One of these limitations is the inability at this stage to replicate the results in an independent dataset. There are a couple of challenges faced with the replication of results. First, defining an adequate latent space that accurately captures a transient cell state solely based on gene expression imposes some limitations. Multimodal data could provide a better assessment and definition of cell states. Additionally, unlike label transfer approaches applied to annotate discrete cell types, matching continuous cell states in a common latent space between independent datasets has not yet been resolved. Further work in reference-based trajectory transferring is necessary to conduct an accurate comparison of cell states to ultimately replicate allelic effects on continuous cell states from different datasets. Therefore, interpreting results from shared biological signatures between datasets at this stage remains a challenge. Another challenge with replicating our results is one of sample size, and the lack of additional independent cohorts of the requisite sample size to provide adequate power to identify dynamic effects. We anticipate that developments in both the cell state latent space transfer and increasingly large studies will lead to significant advances in uncovering and understanding how allelic effects vary across cell states.

Another limitation of our work on dynamic eQTL is the ability to interpret how dynamic effects are underpinned by changes in genome regulation such as chromatin accessibility. We had attempted to investigate the relationship between cell stage changes and regions of open chromatin around dynamic eQTL transcripts using scATAC-seq data. However, as above we face the dual challenge of mapping the cell states between scATAC-seq and our scRNA-seq data, and the necessity to have a large number of cells with scATAC data. The first problem will be partially solved with the arrival of 'multiomic' dual RNA+ATAC-seq assays, allowing cell states identified by RNA to be linked to ATAC variation in the sample cells. However, the development of new methods to correctly label transfer will need to account for the loss of cytoplasmic RNA in 'multiomic' data.

A final point of consideration for future studies of dynamic eQTLs is the trade-off between larger donor sample sizes versus number of cells per donor. As most studies are limited by funding availability, there is a choice to be made between the numbers of donors and the average numbers of cells per donor. Increasing donor numbers increases the statistical power to be able to estimate the allelic effects of genetic variants on gene expression. However, increasing the numbers of cells per donor enhances the ability to accurately capture and identify cell state spaces, and correspondingly the ability to determine if allelic effects vary across biological cell states.

trans-eQTL analysis of lead cis-eSNPs

To identify any SNPs that have both *cis* and *trans* associations, we mapped *trans*-eQTLs using 26,596 eSNPs from *cis*-eQTL analysis. We used the same covariates as for the *cis*-eQTL analysis (**Section 5**). Covariate-adjusted residual values were tested in a linear regression model where SNP genotype was included as the dependent variable for all gene-variant pairs. We removed *cis*-eSNP and *trans*-eGenes if the original *cis*-eGene was on the same autosomal chromosome. In total we identified 428,169 significant *trans*-eQTL at a study-wide FDR of 0.01 and filtered these to include autosomal and mitochondrial *cis*-eGenes. Next, we removed *trans* associations found in the MHC region due to low mapping quality of this region in the genotyping data. After this step, the remaining number of *trans*-eQTL associations was 990. We tested the correlation of cell type specific *trans*-eQTL discovery with cell type population and *cis* associations sizes using spearman's rank correlation test at *p*-value of 0.05 (**Fig. 6**).

Immune-related disease associations

Overlap of eQTLs and publicly available NHGRI-EBI GWAS Catalog hits: To assess the utility of lead eQTLs for functional understanding of seven common immune-related diseases, we used all genome-wide significant associations ($p < 5 \times 10^{-8}$) from the NHGRI-EBI GWAS catalog version 1.0.2, released 08/02/2020. We chose to focus on multiple sclerosis (MS), type 1 diabetes (T1DM), inflammatory bowel disease (IBD), Crohn's disease (CD), systemic lupus erythematosus (SLE), rheumatoid arthritis (RA), and ankylosing spondylitis (AS). This subset of the catalog included 2,335 associations from a total of 147 studies (**fig. S22**).

Colocalization of eQTL and GWAS associations using Mendelian randomization method: To identify the shared causal variants (eQTL and GWAS), and test if the allelic effect of a SNP on disease risk is mediated by gene expression, we applied the summary mendelian randomization (SMR) method (75) using summary statistics of seven autoimmune disease GWAS results. In order to apply this method, first we generated beta estimates and corresponding standard errors (107). We accessed the latest GWAS summary statistics for MS from International Multiple Sclerosis Genetics Consortium (<https://imsgc.net/>) (available upon request). Summary statistics for IBD, RA, AS, and T1DM GWASs were downloaded from the following repositories: (i) for IBD and CD from the IBD Genetics Consortium website (www.ibdgenetics.org), (ii) for RA from the Japanese Encyclopedia of Genetic associations by RIKEN (<http://jenger.riken.jp/en/result>), (iii) for AS and T1DM from Pan-UK Biobank (<https://pan.ukbb.broadinstitute.org/>). The SLE GWAS summary statistics was downloaded from the following link: http://urr.cat/data/GWAS_SLE_summaryStats.zip. As the SMR approach requires beta estimate for the given SNP-gene pair, we have mapped *cis*-eQTLs using Matrix eQTL and applied at FDR level of 5%. Then, using these outputs, we performed the SMR analysis with default settings only changing --peqtl-smr to 0.05.

Colocalization of eQTL and GWAS associations using Bayes factors: We used the Bayesian colocalization method (coloc)(108) to integrate *cis*-eQTLs per cell type and seven autoimmune disease (AS=ankylosing spondylitis, CD=Crohn's disease, IBD=inflammatory bowel disease, RA=rheumatoid arthritis, SLE=systemic lupus

erythematosus, T1DM=type 1 diabetes mellitus,) gwas results. This method assesses whether the gene expression and disease phenotype are based on a single shared variant by best fitting eQTL and GWAS associations into a model and summarizing the posterior probabilities. As coloc requires all the p-values within a gene region, we have used Matrix eQTL output in this analysis. We tested pairwise colocalization using GWAS and eQTL summary statistics of SNPs within a 2MB window around each gene. Then we subset these results based on the eSNP-eGene pairs from the first round of conditional analysis ($eSNP_1$) to identify colocalizations closer to the ‘truth’. Among these, a posterior probability of colocalization $>80\%$ was considered as evidence for a shared causal locus.

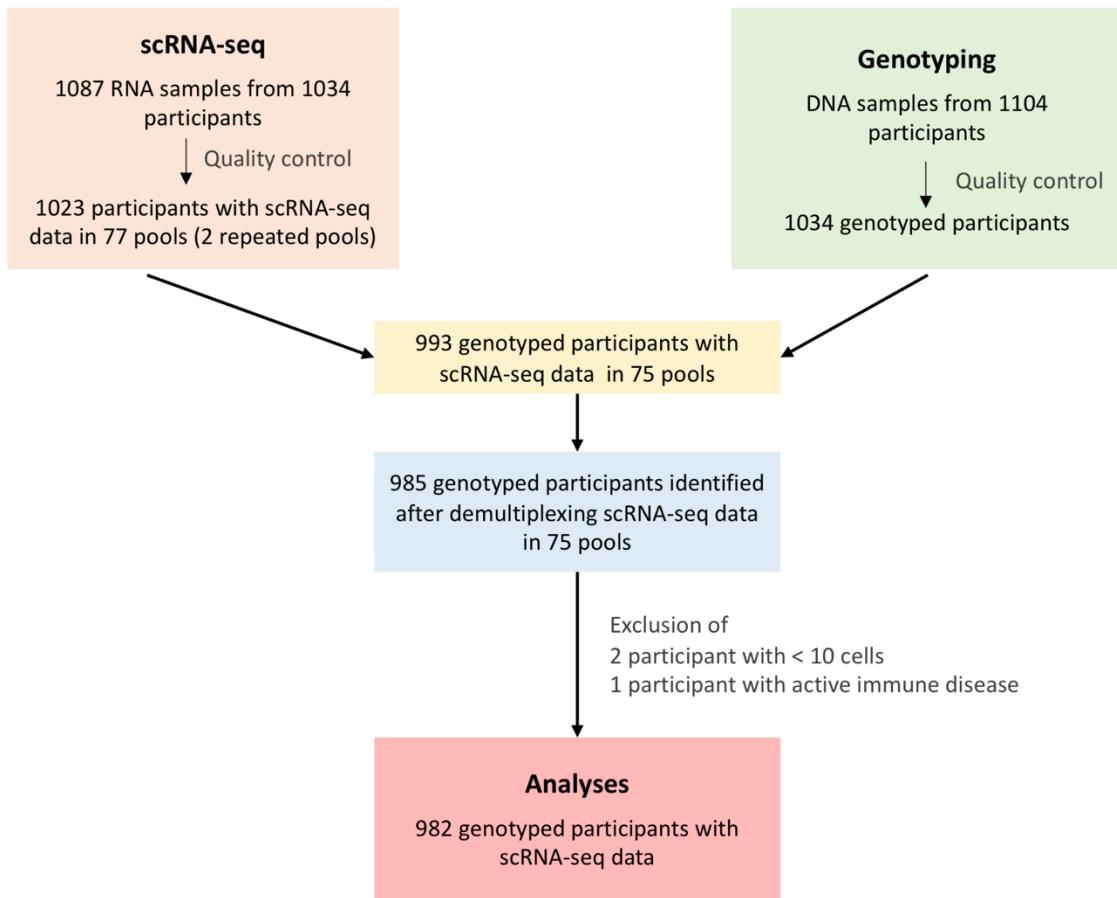


Fig. S1. Overview of sample selection and quality control for the OneK1K study.

Details of genotyping and scRNA-sequencing processing are described in sections 2 and 3 of this supplementary material respectively. Data from 982 participants were included in the single-cell eQTL mapping.

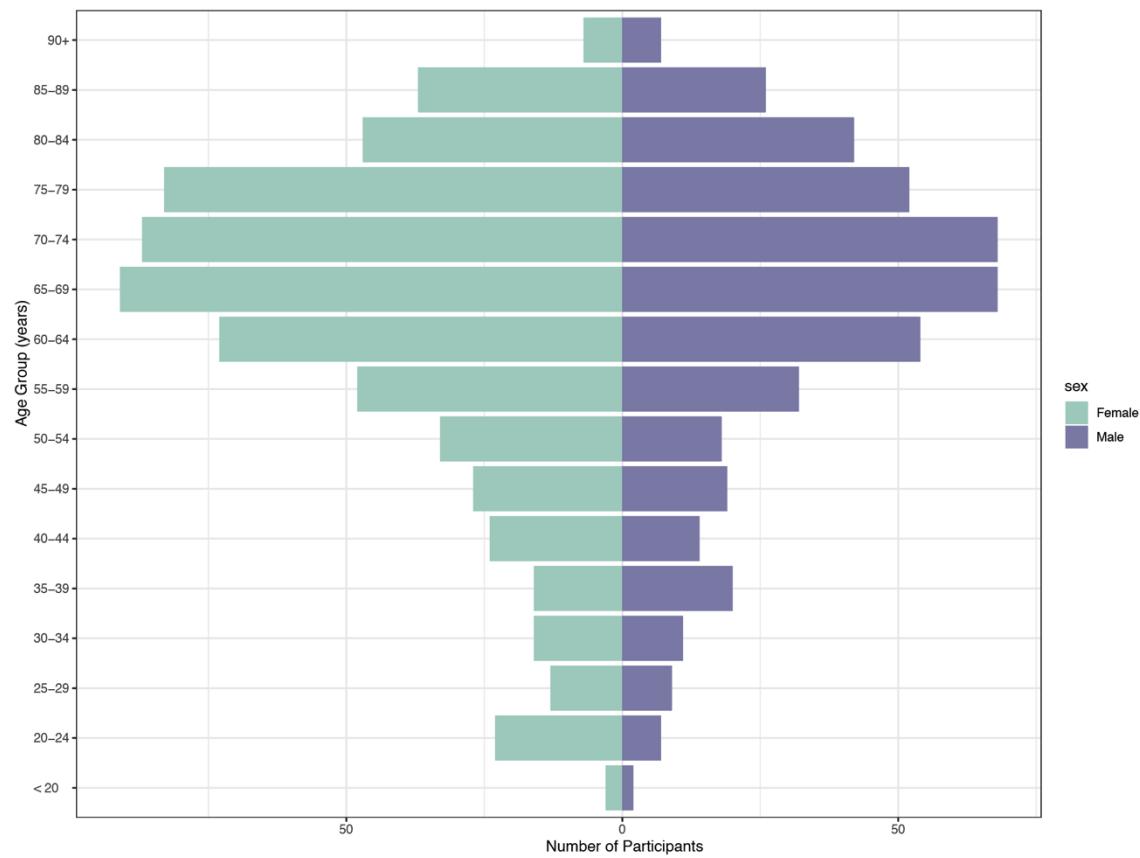


Fig. S2. Participant demographics.

Distribution of age and sex in the OneK1K cohort.

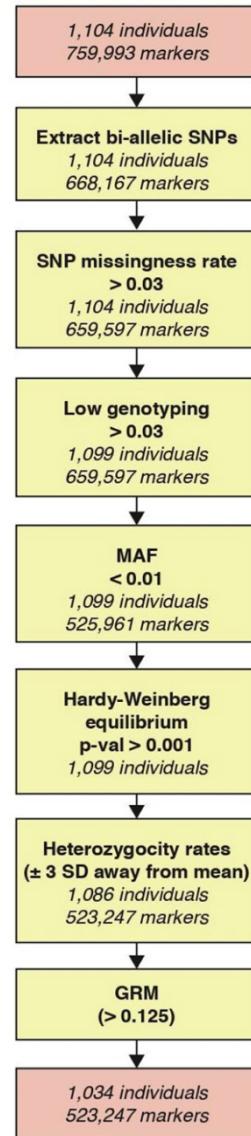


Fig. S3. Pre-imputation quality control of genotyping data.

Flowchart demonstrating the pipeline used for cleaning the genotyping data prior to imputation.

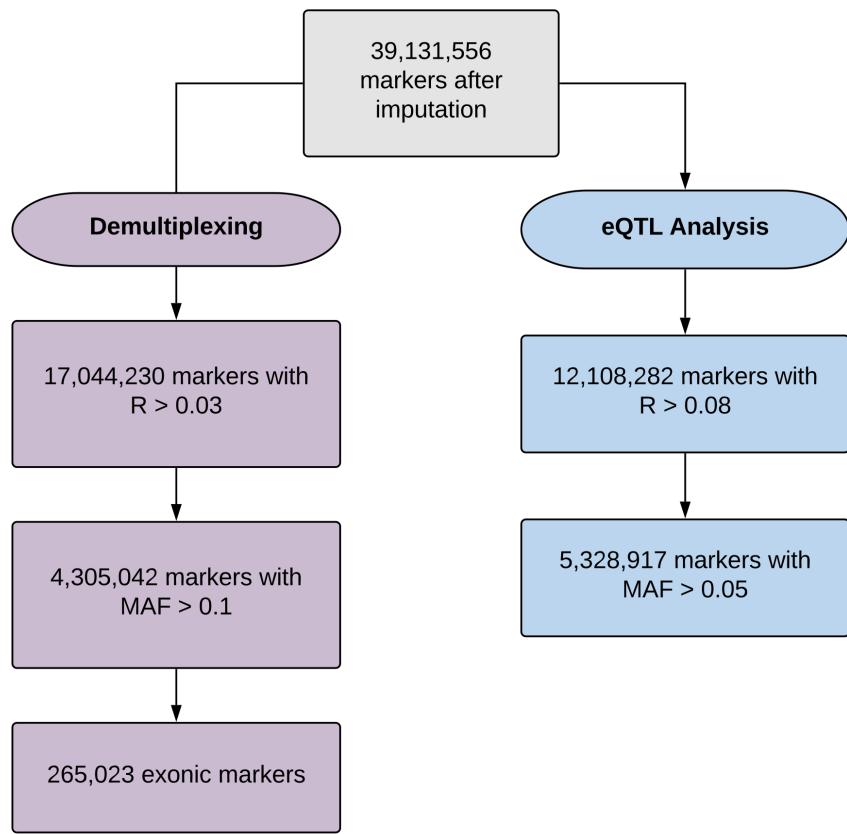


Fig. S4. Post-imputation quality control of genotyping data and selection of SNPs for demultiplexing and eQTL analysis.

Quality controlled genotype data were imputed against Haplotype Reference Consortium data using the Michigan Imputation Server. The flowchart displays how the imputed data is used in the analysis. A total of 265,023 exonic markers were used for identifying cells belonging to each participant using demuxlet (purple arm). For eQTL analysis, 5,328,917 markers with $R > 0.08$ & $MAF > 0.05$ were selected.

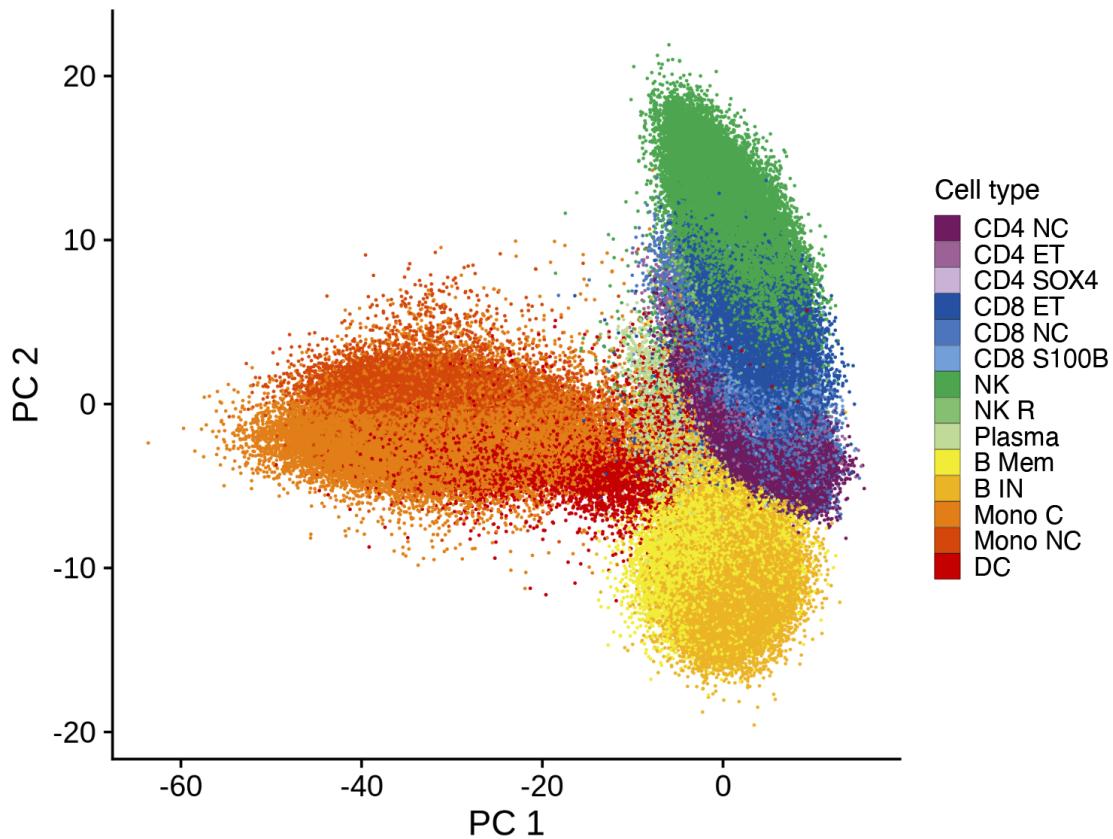


Fig. S5. Separation of 14 immune cell types using the first two transcriptional principal components.

This plot is complementary to the UMAP plot in Fig. 1G, showing 1,267,758 PBMCs plotted from all individuals, with clustering of the 14 cell types shown on the right-hand side.

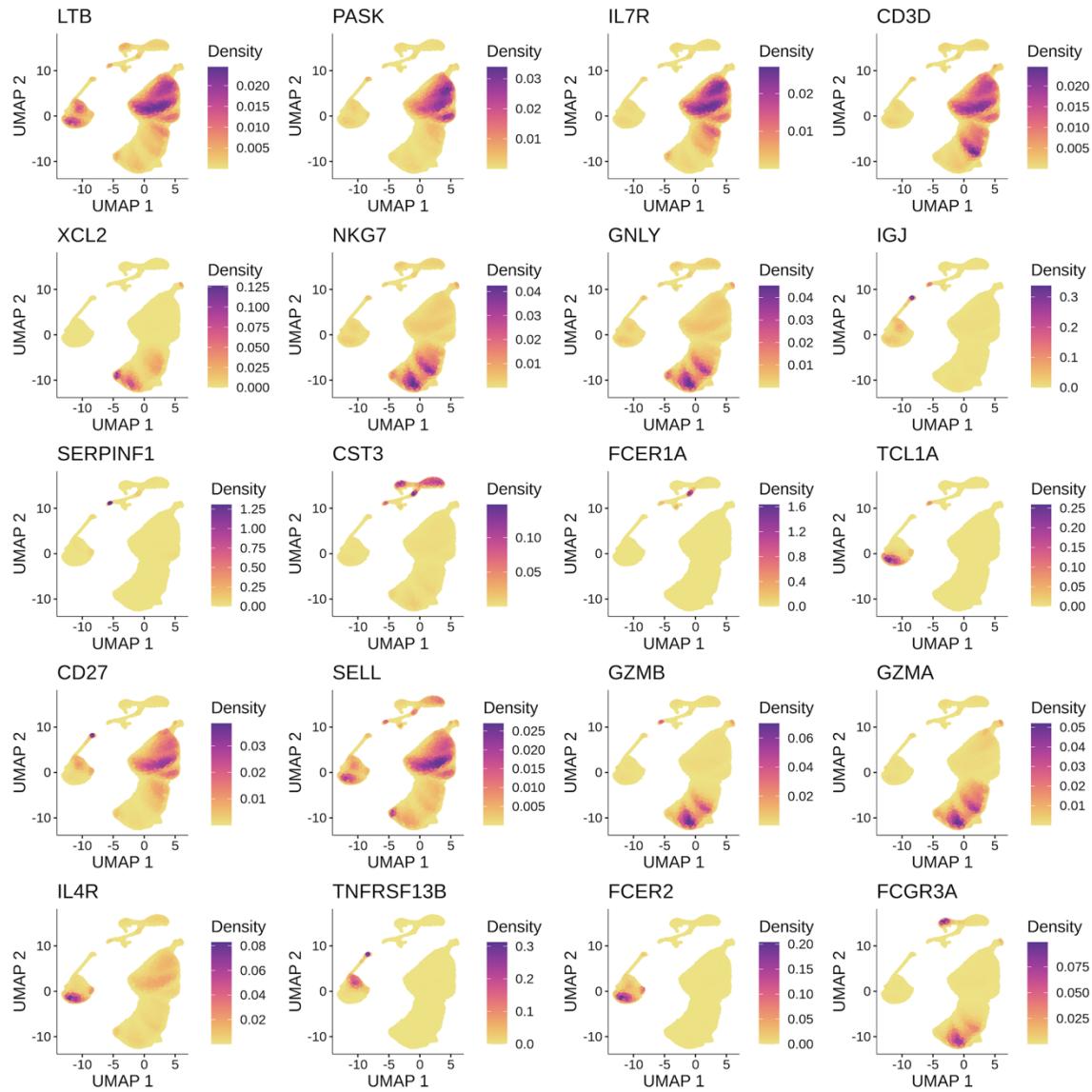


Fig. S6. Density plots of differentially expressed canonical markers of peripheral immune cells.

These figures complement those shown in **Fig. 1H**, with density corresponding to expression of the labelled gene. Panels have independent scales which are determined by the maximum expression of the transcript.

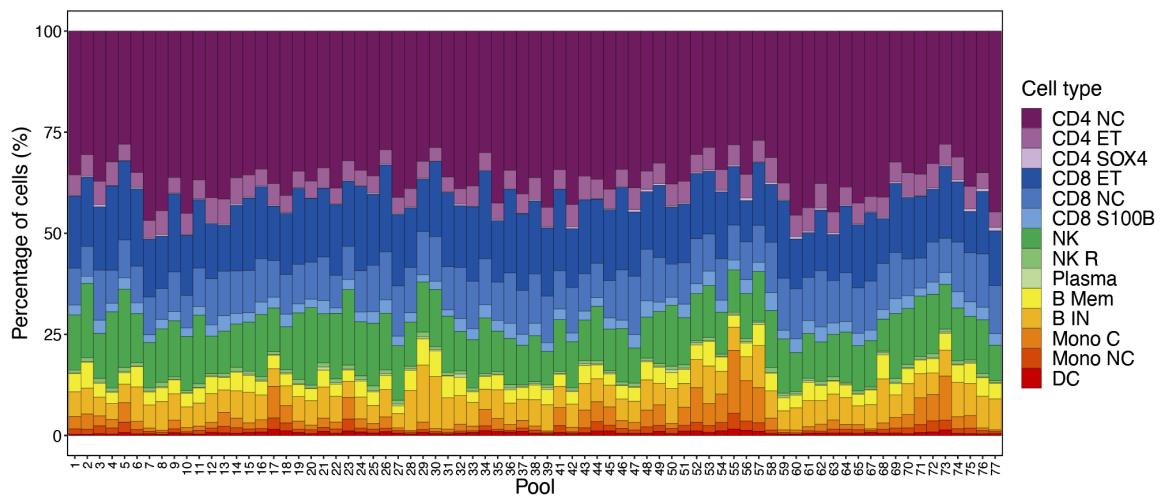


Fig. S7. Distribution of cell type proportions in 75 sequencing pools.

Each cell type is colored according to the legend on the right-hand side. Each pool contains 12-14 individual samples. Pool 40 and 66 were replicates and thus have been removed.

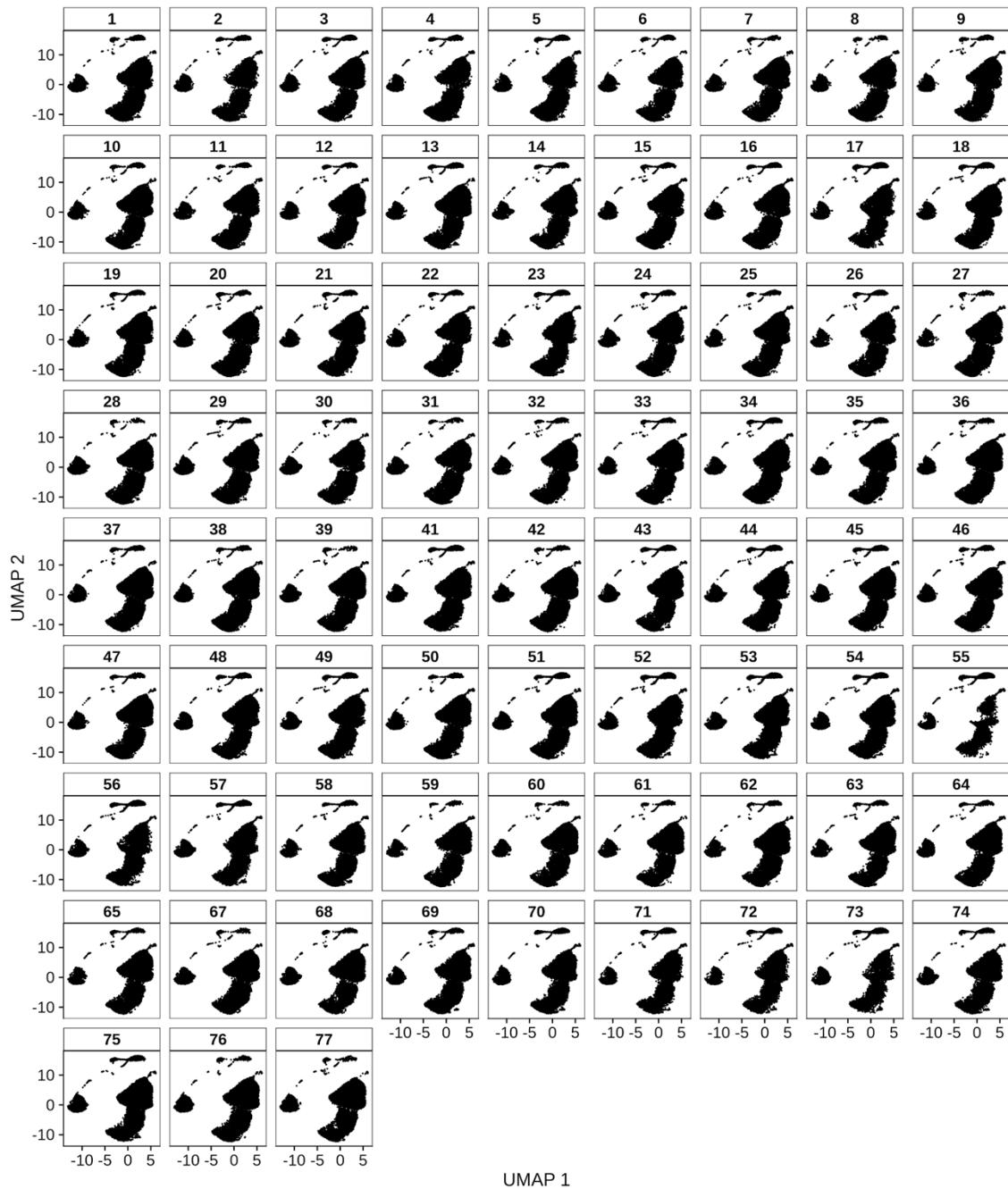


Fig. S8. UMAP plots of each pool.

UMAP plots show the architecture of the clustering of cells identified in each of the 77 independent pools (Pool 40 and 66 were replicates and thus have been removed).

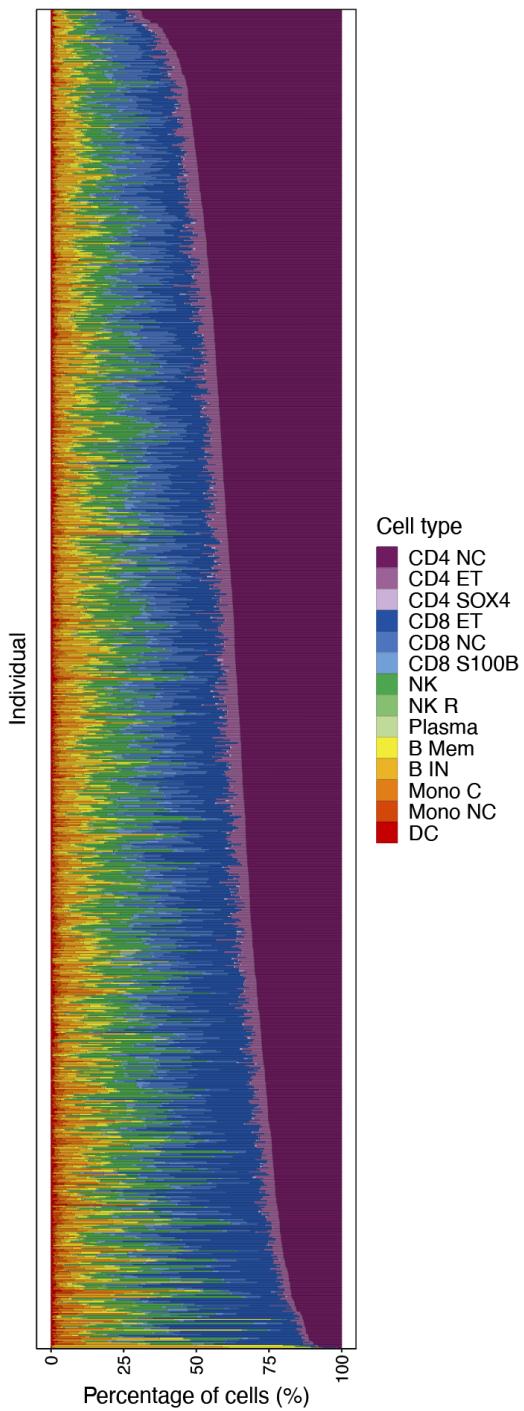


Fig. S9. Distribution of cell type proportions per participant.

Each line in the plot represents one individual's cell type composition. Each cell type is colored according to the legend on the right-hand side. The donors are ranked according to the percentage of CD4_{NC} cells.

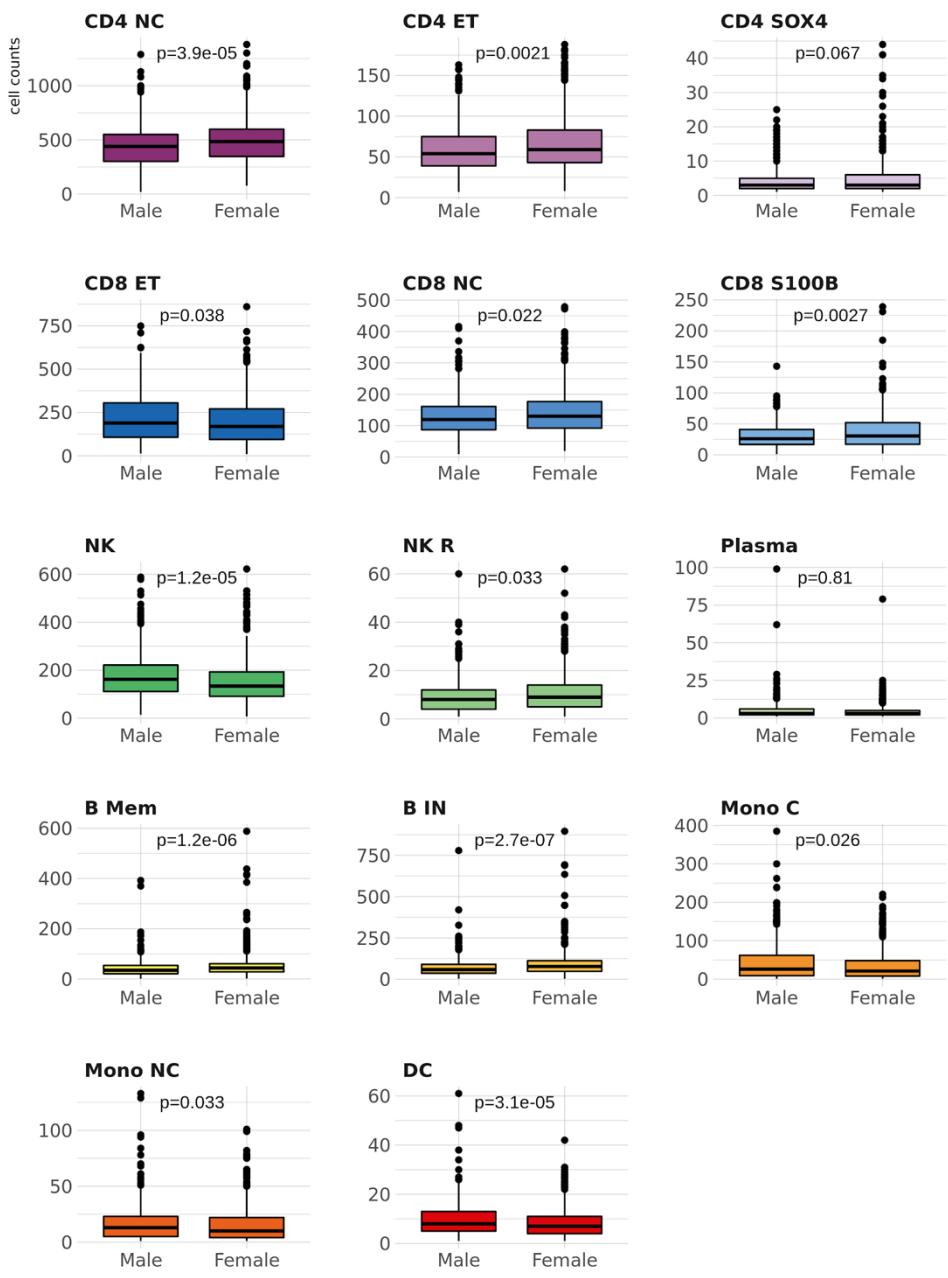


Fig. S10. Comparison of cell counts between females and males.

Each boxplot represents the distribution of cell counts for males and females as labelled. P-values represent the unpaired two samples of Wilcoxon's test results.

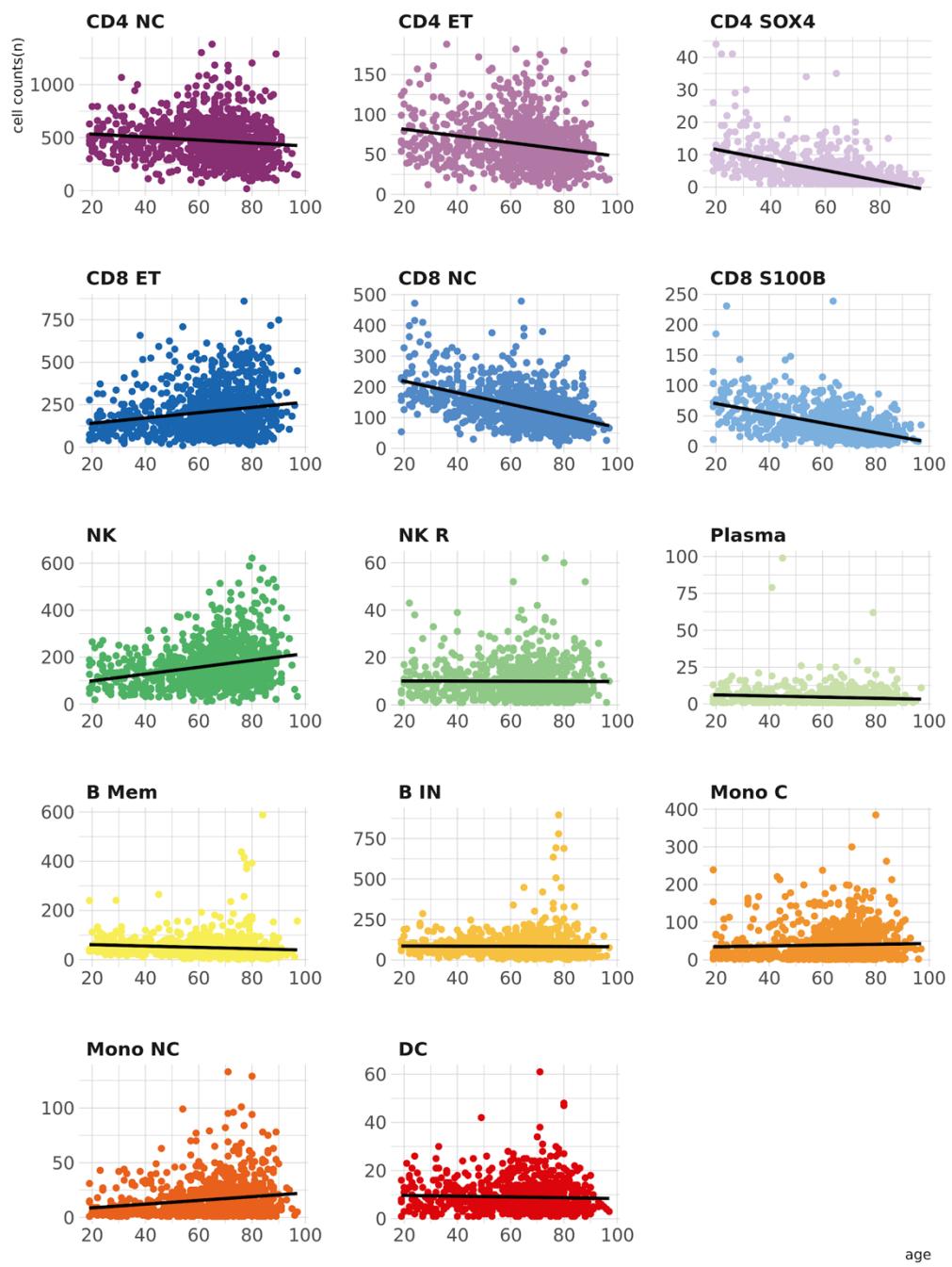


Fig. S11. Association of age with cell counts in 14 cell types.

Each scatter plot represents the correlation between age in years (*x*-axis) and cell counts (*y*-axis).

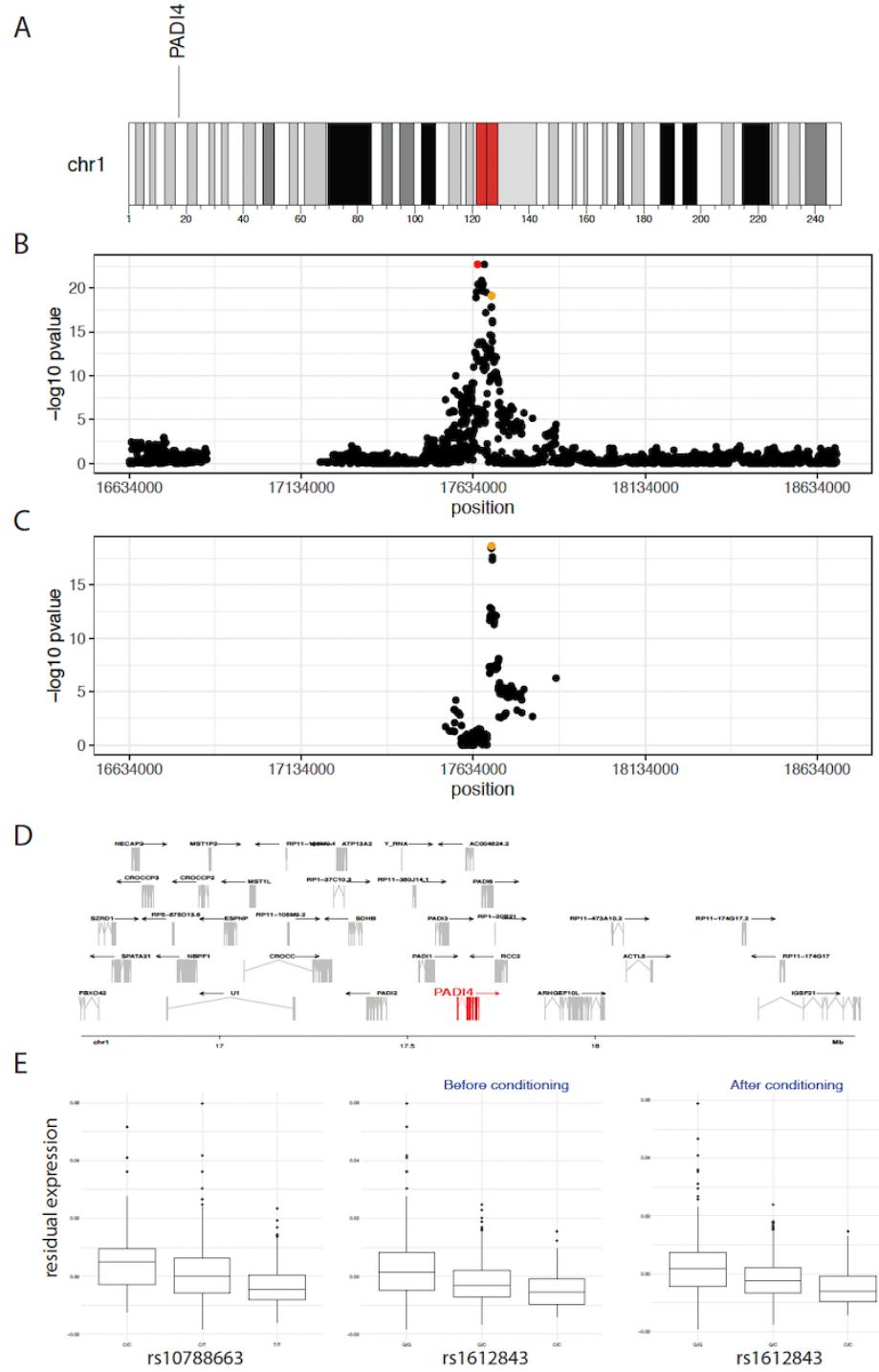


Fig. S12. Effect of eSNPs rs10788663 and rs1612843 on expression of *PADI4* gene before and after conditional analysis in CD4 Naive and CM (CD4_{NC}) cells.

Multiple independent loci could be associated with expression of a gene in a cell. Here we present an example for the *PADI4* gene. (A) Chromosomal position of *PADI4* gene (B) Locus zoom plot showing the strength and extent (y-axis) of the association signal of

lead eQTL rs10788663 and secondary eQTL rs1612843 to genomic position (x-axis) (C) Locus zoom plot displaying the association of rs1612843 after conditioning expression of *PADI4* gene on rs10788663. (D) Location of *PADI4* gene in relation to the other genes. (E) Allelic plots for rs10788663 and rs1612843 (before and after conditioning).

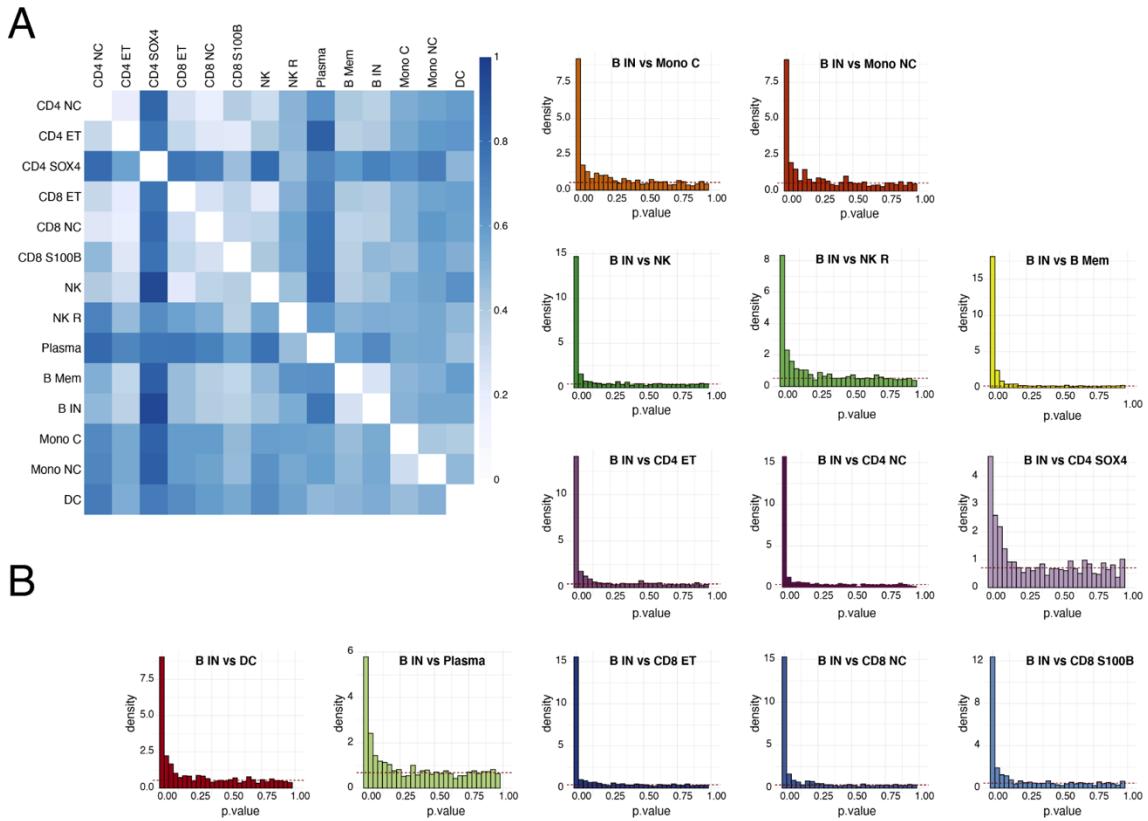


Fig. S13. Investigation of cell type specific test statistic inflation.

(A) Pairwise estimates of the overall proportion (π_1) of null p values (B) Density histogram of p values for the comparison between B_{IN} and each other individual cell type. The horizontal dashed line represents the height of our estimate of the proportion of null p values (π_1).

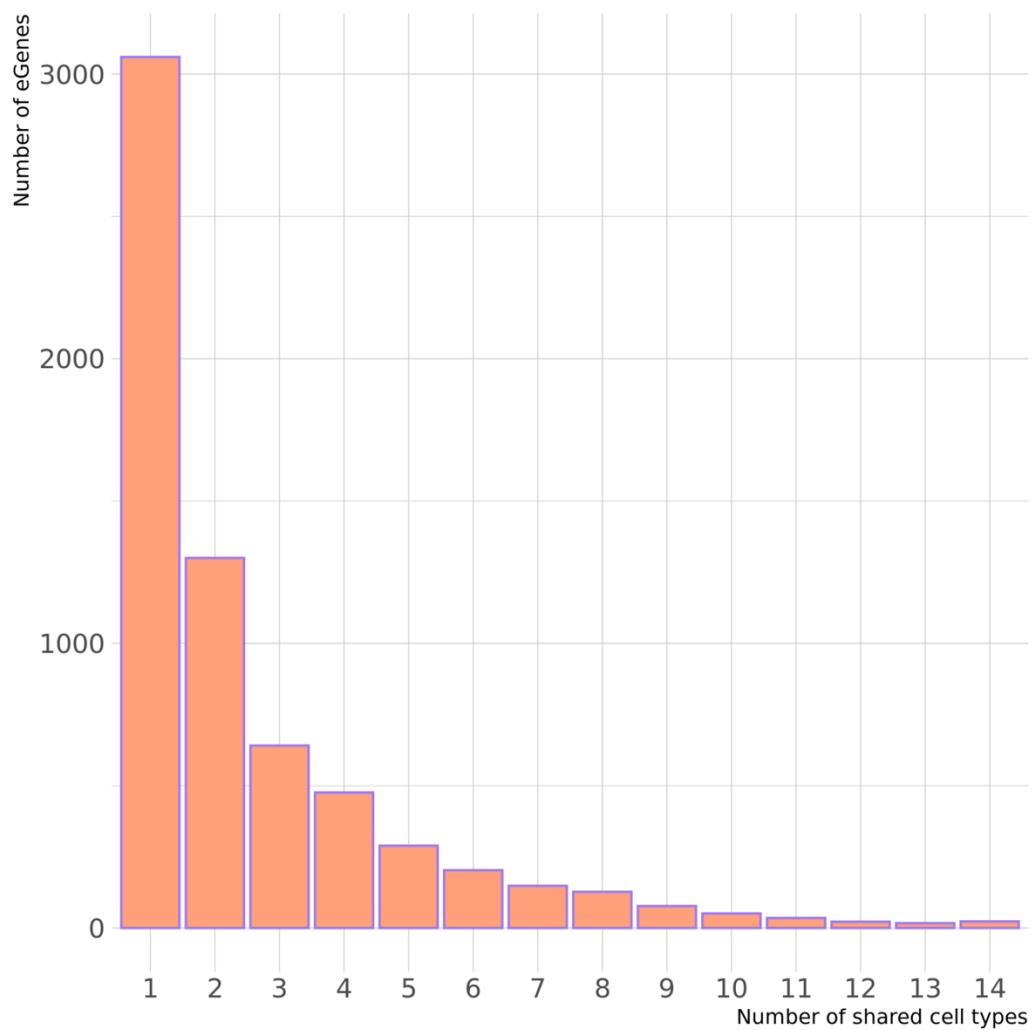


Fig. S14. Number of eGenes shared among cell types.

An eGene is defined as a gene that has an eQTL. The histogram displays the number of eGenes identified in only one cell type (3060), and the number shared amongst two or more cell types.

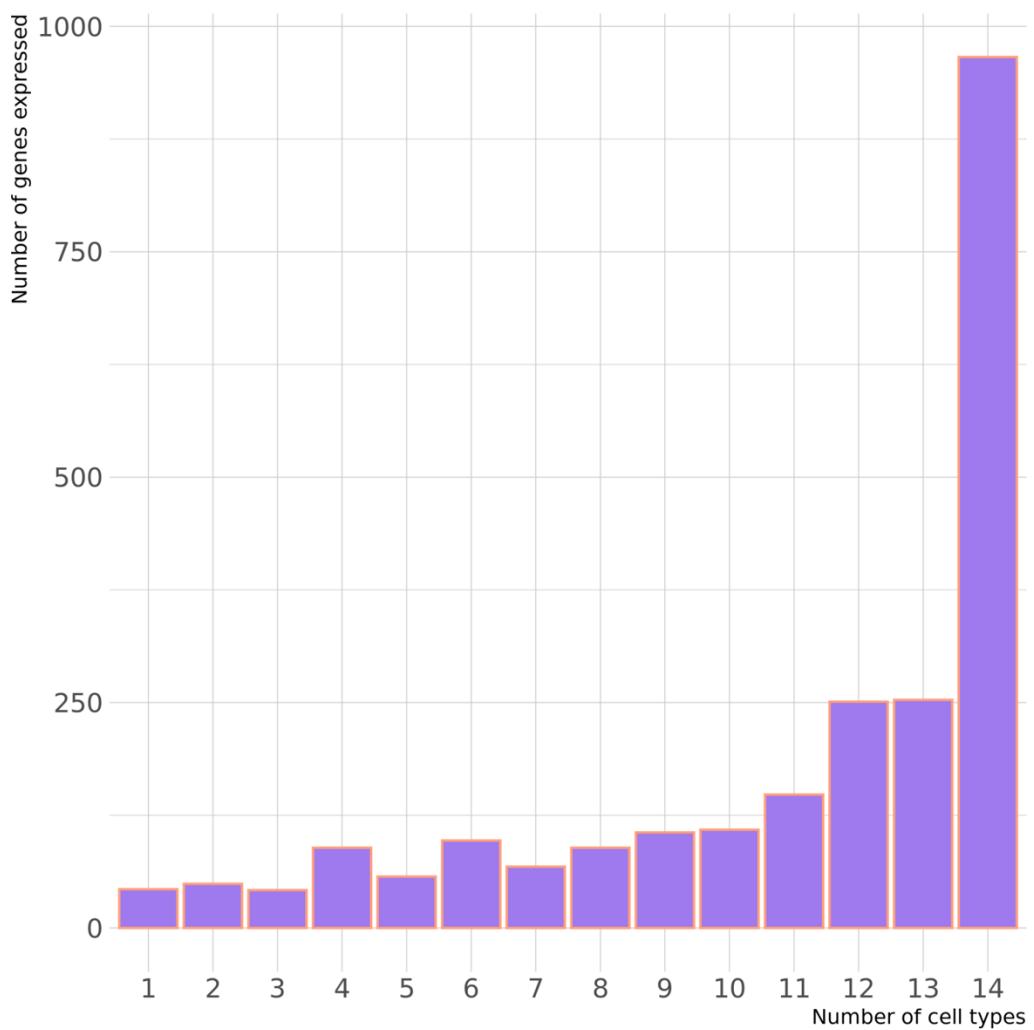
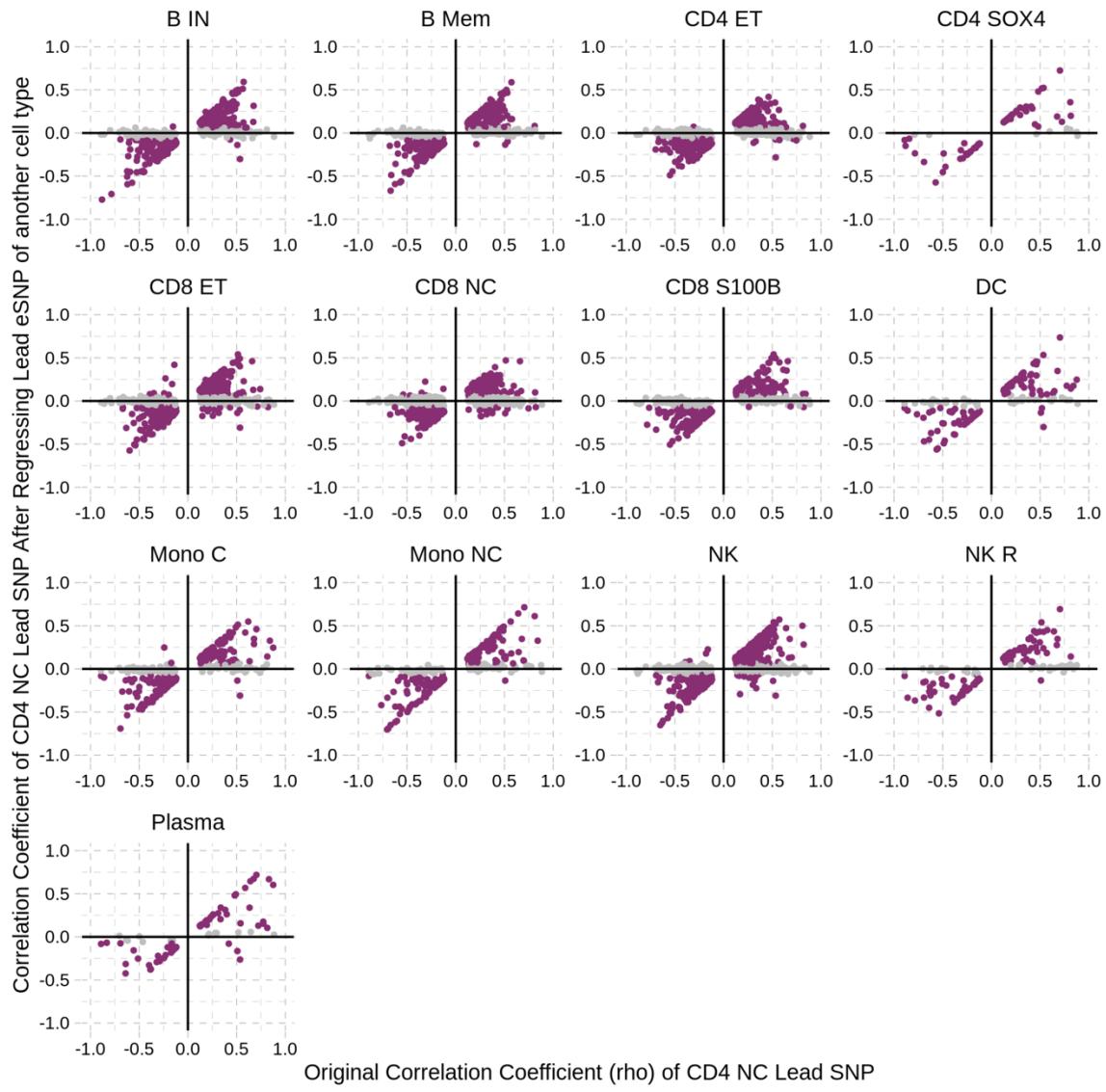
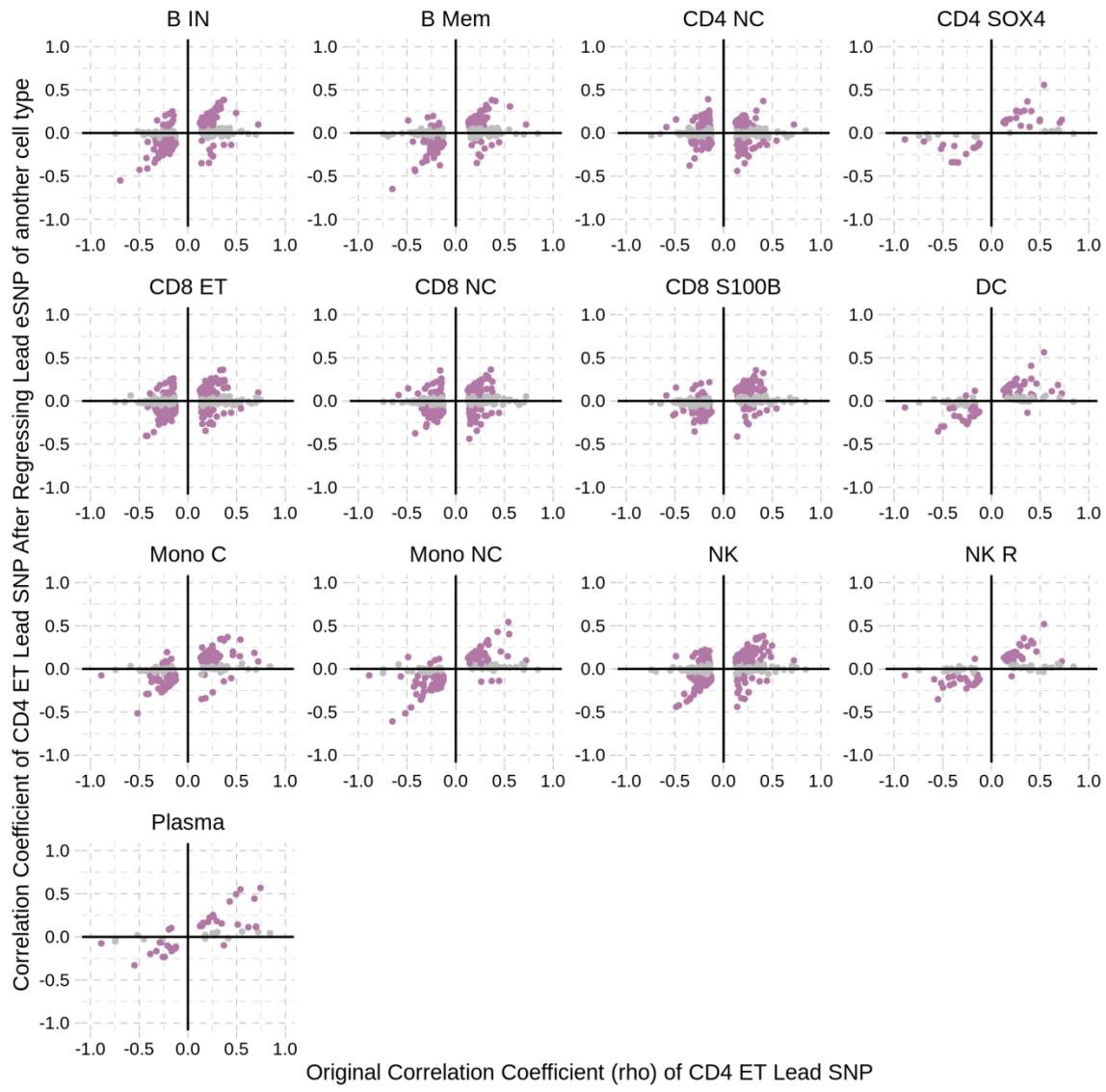
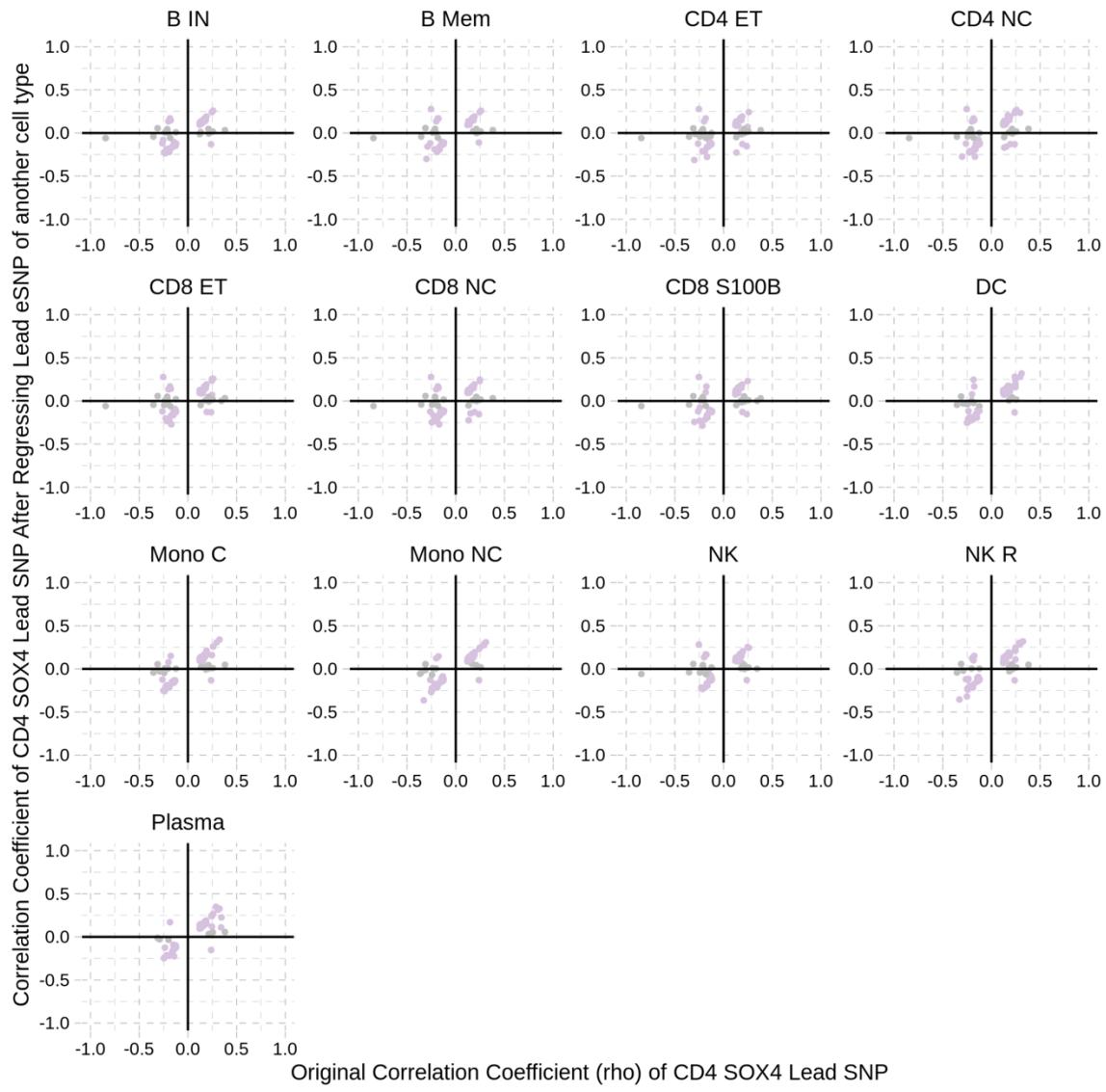


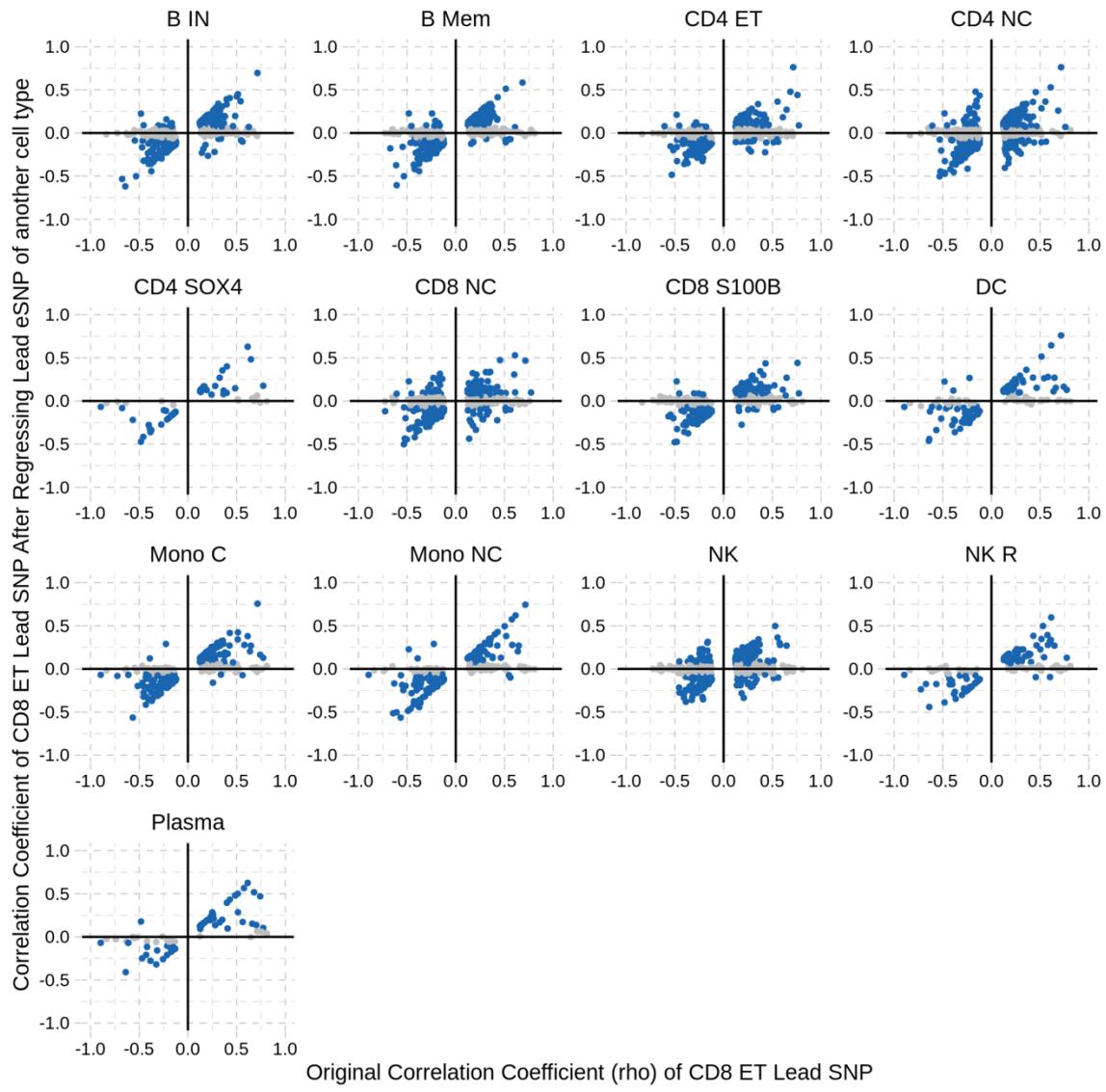
Fig. S15. Expression of eGenes in multiple cell types.

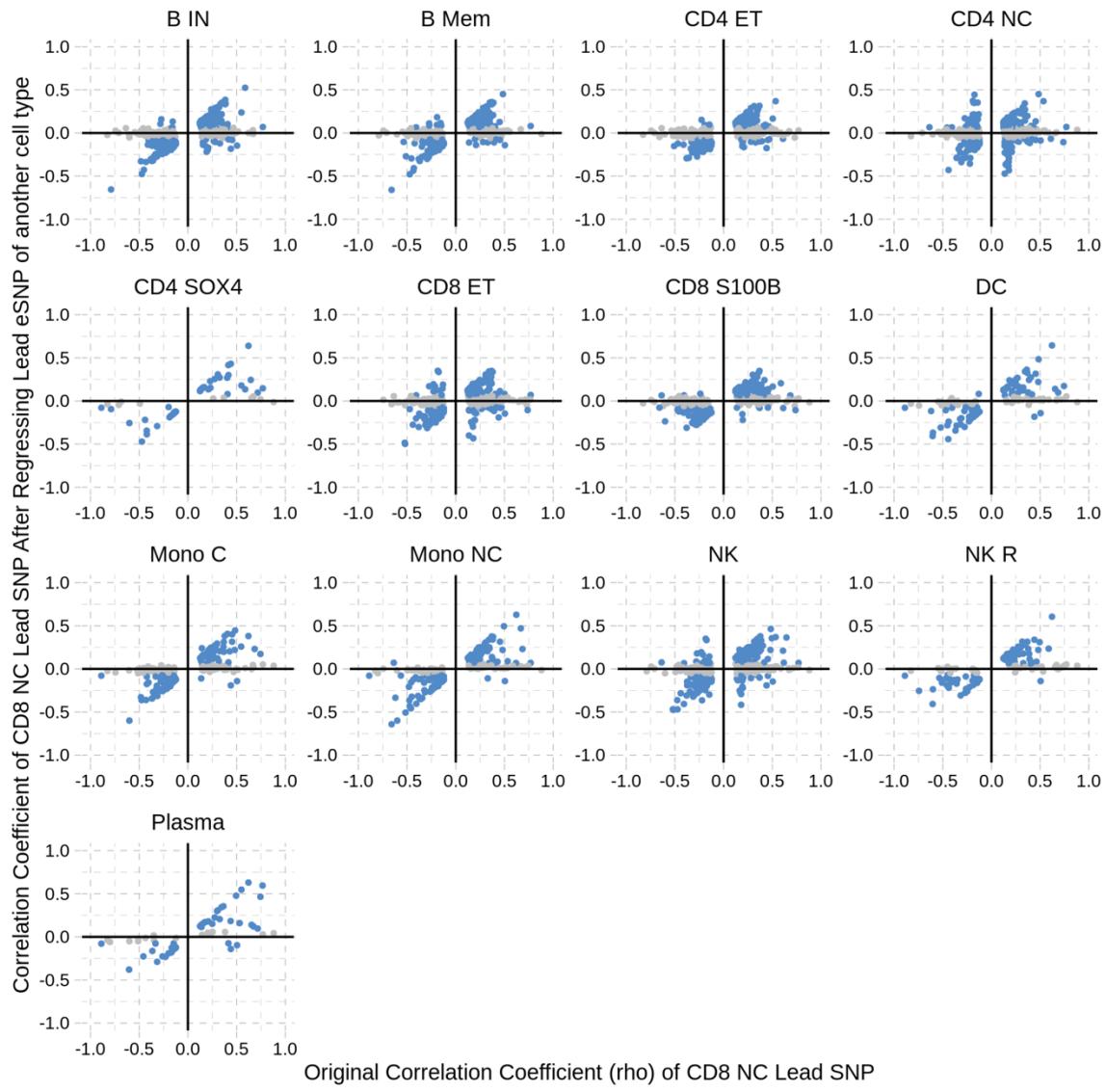
This bar plot shows the number of eGenes with at least one eQTL (y-axis) expressed in one or more cell types (x-axis).

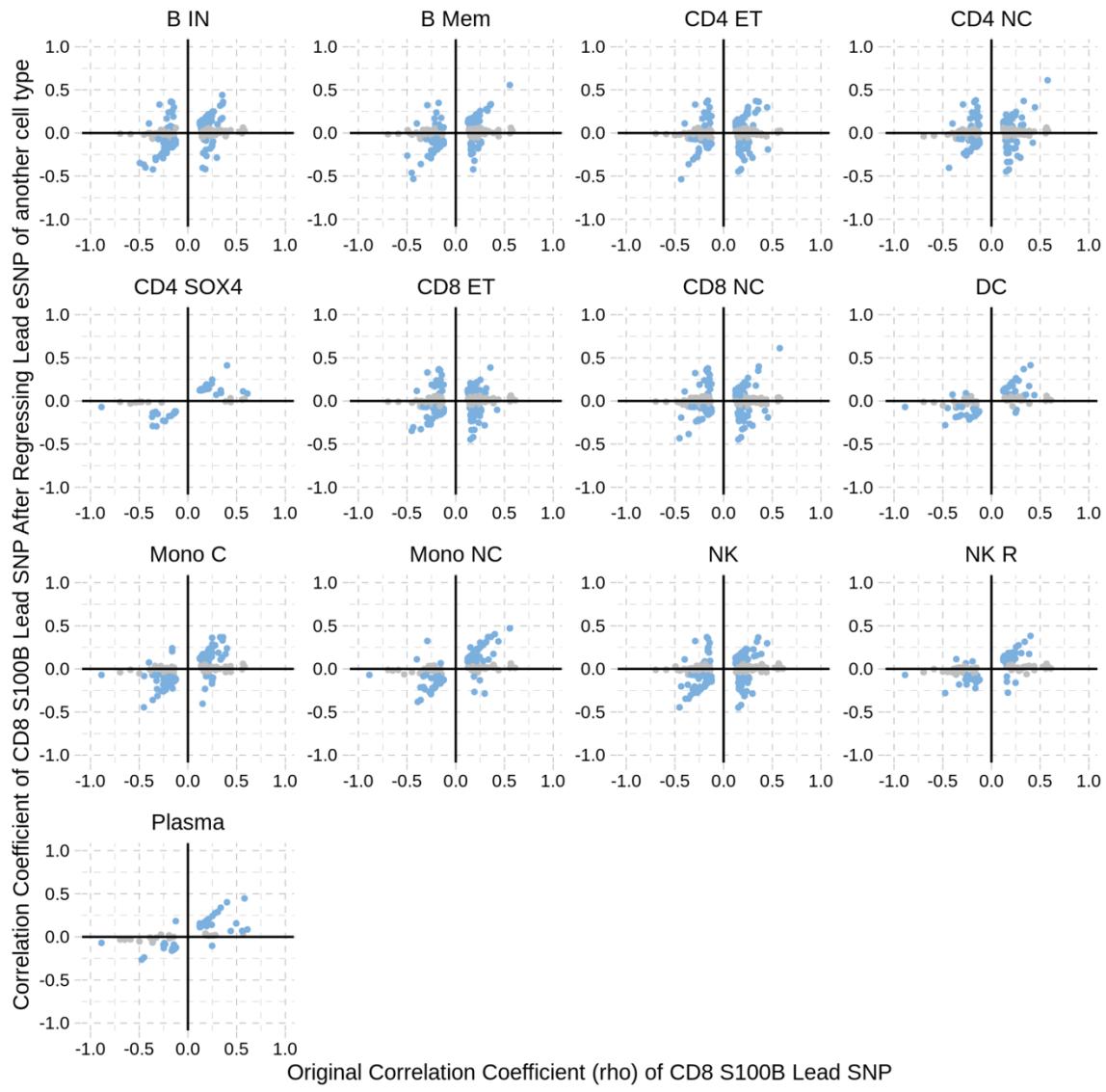


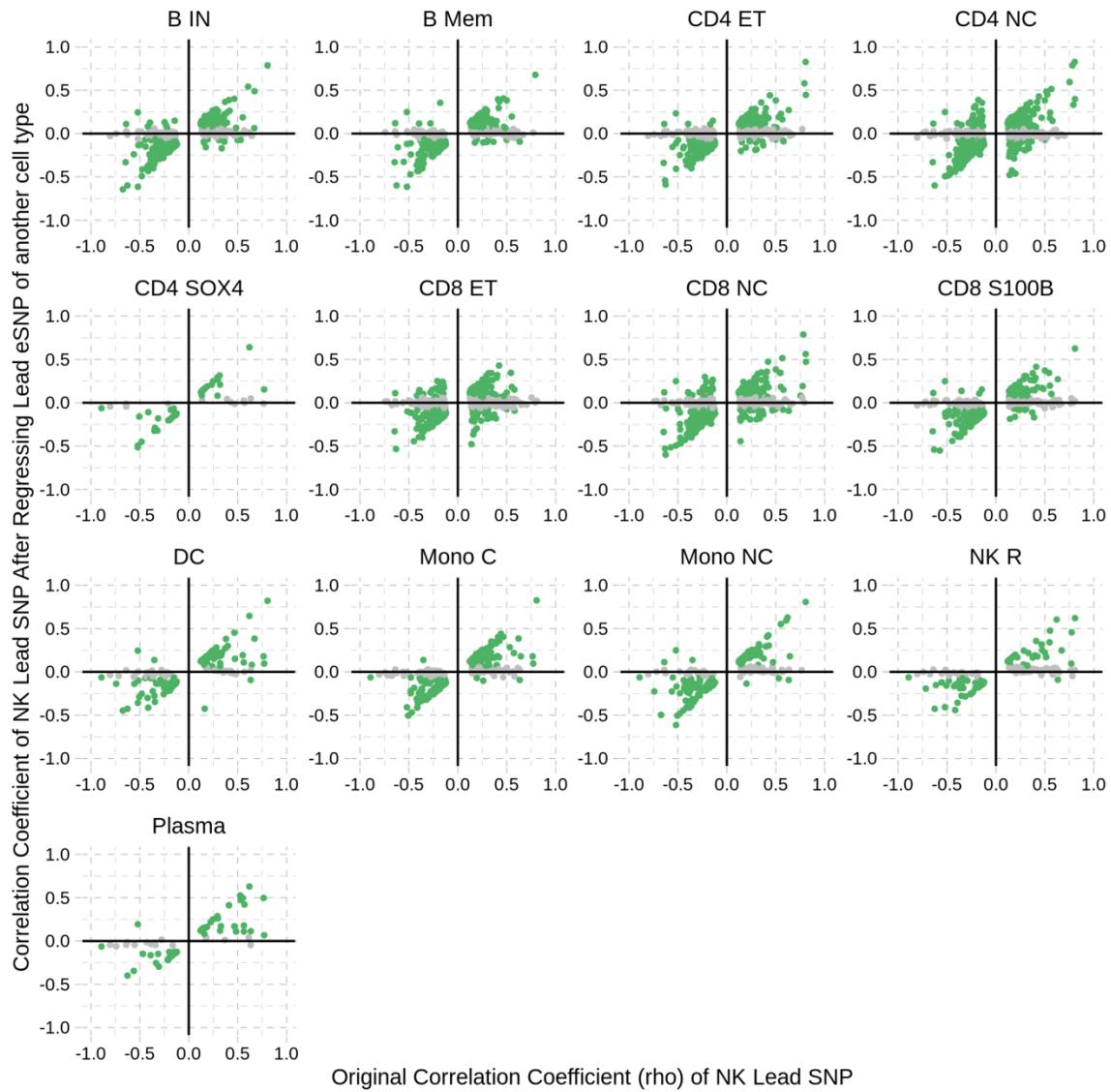


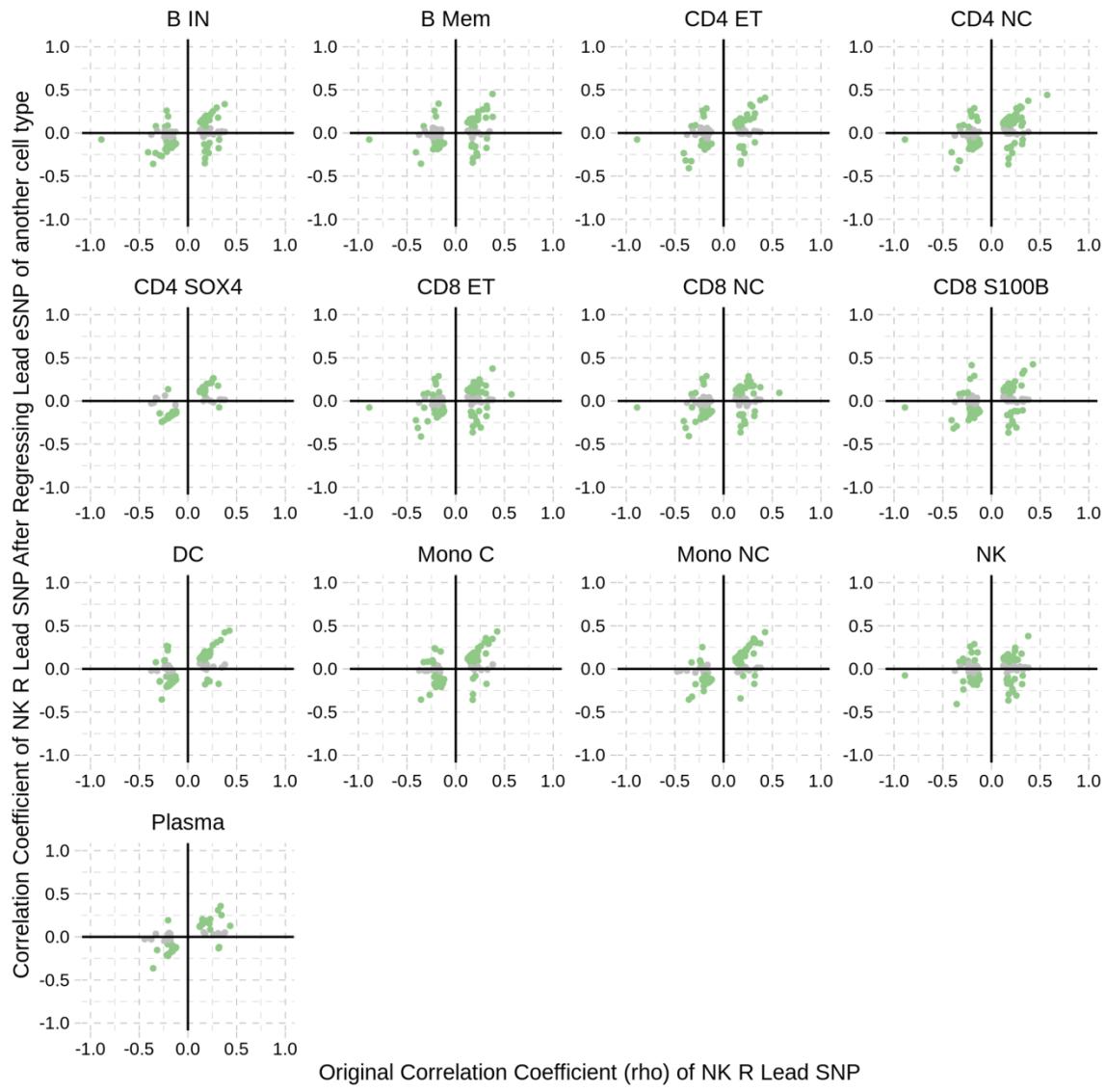


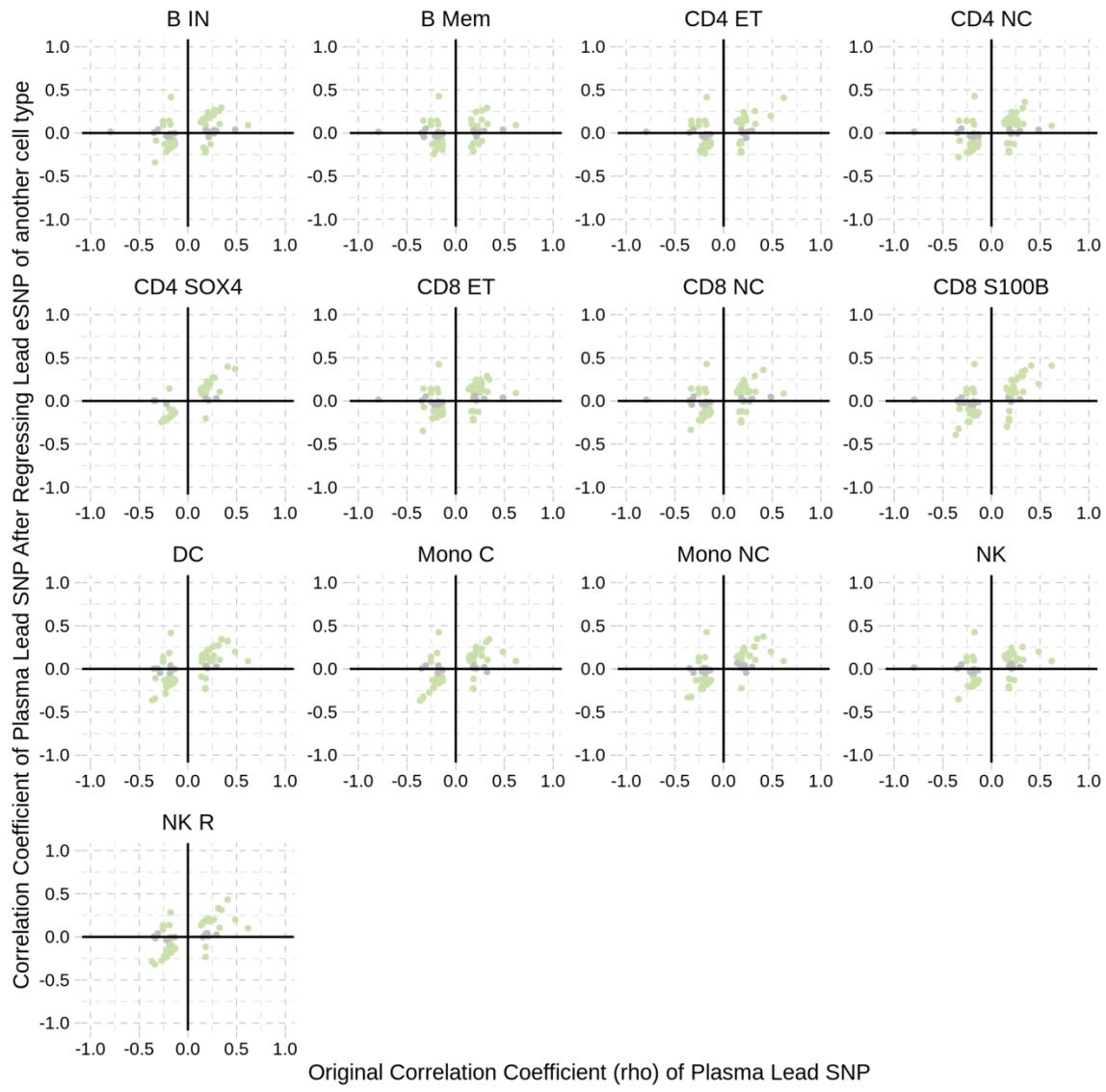


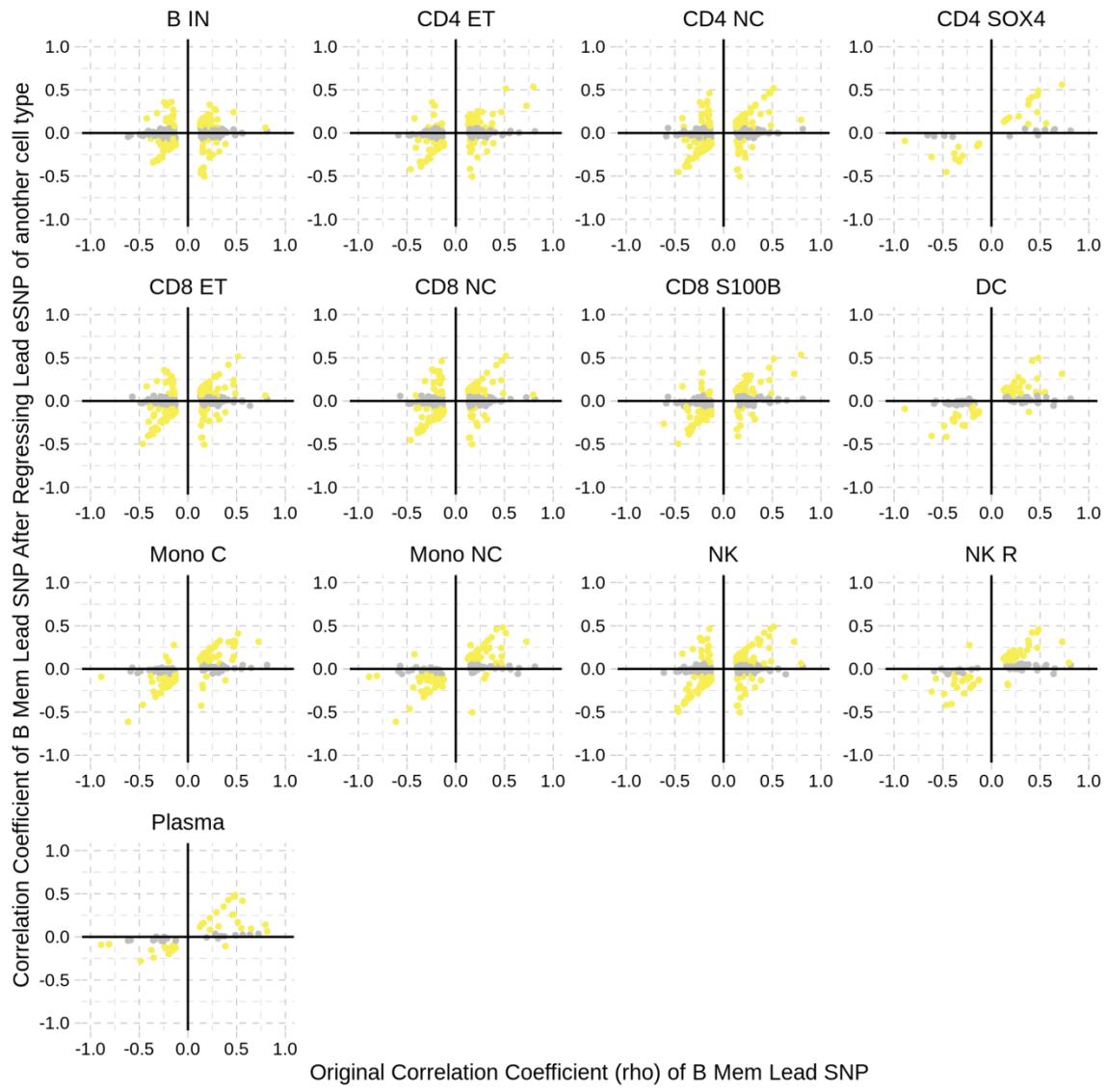


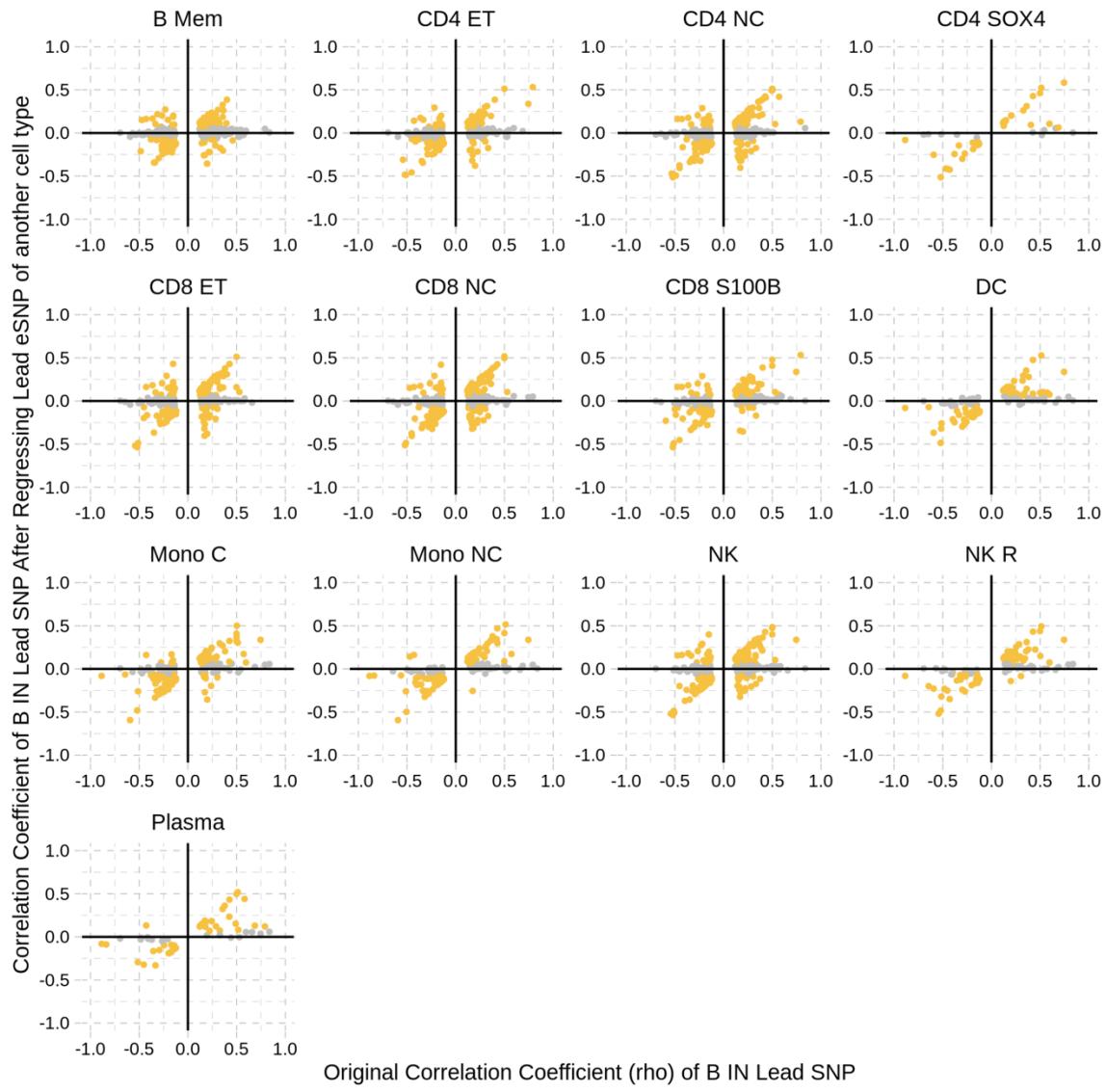


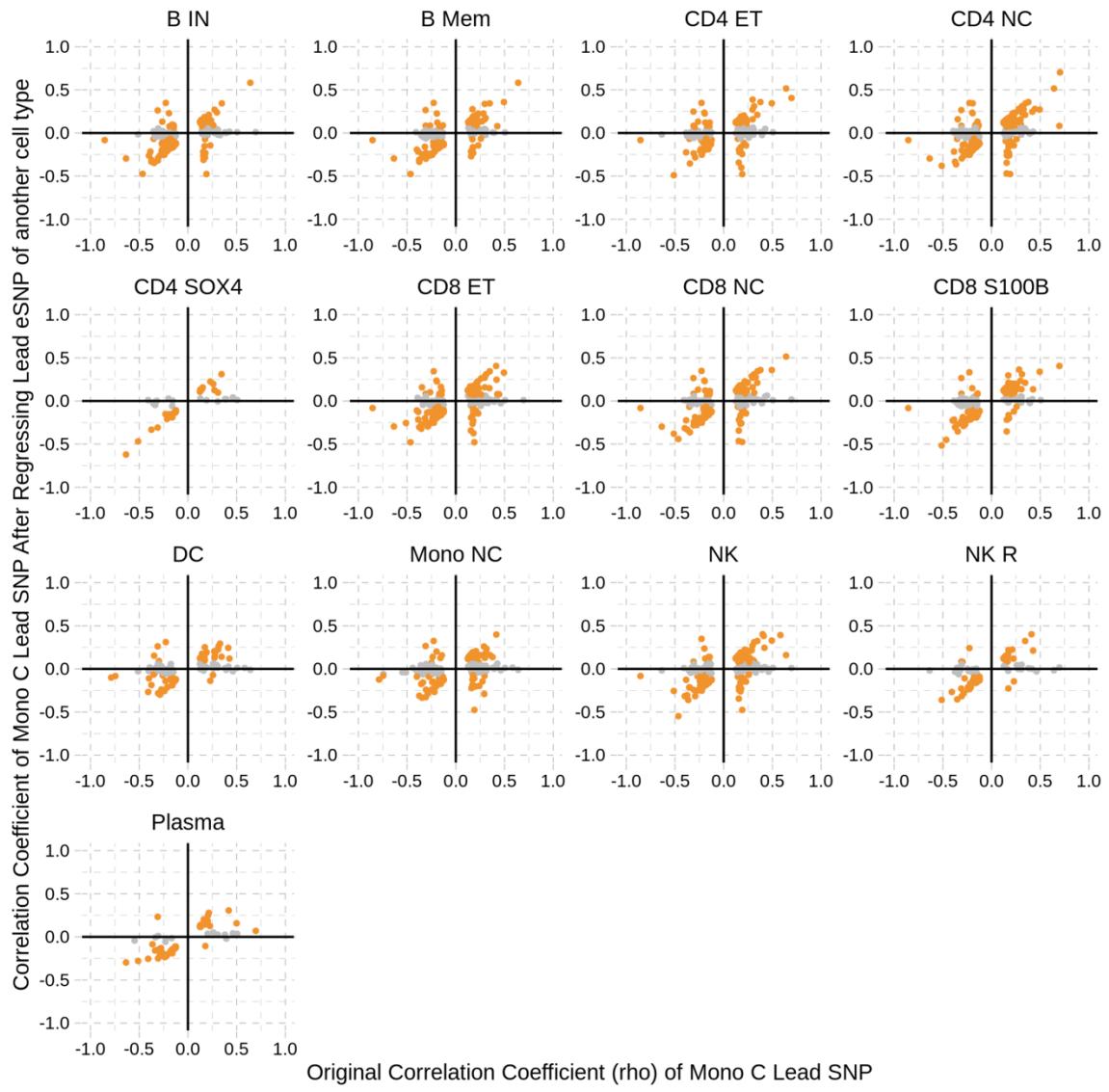


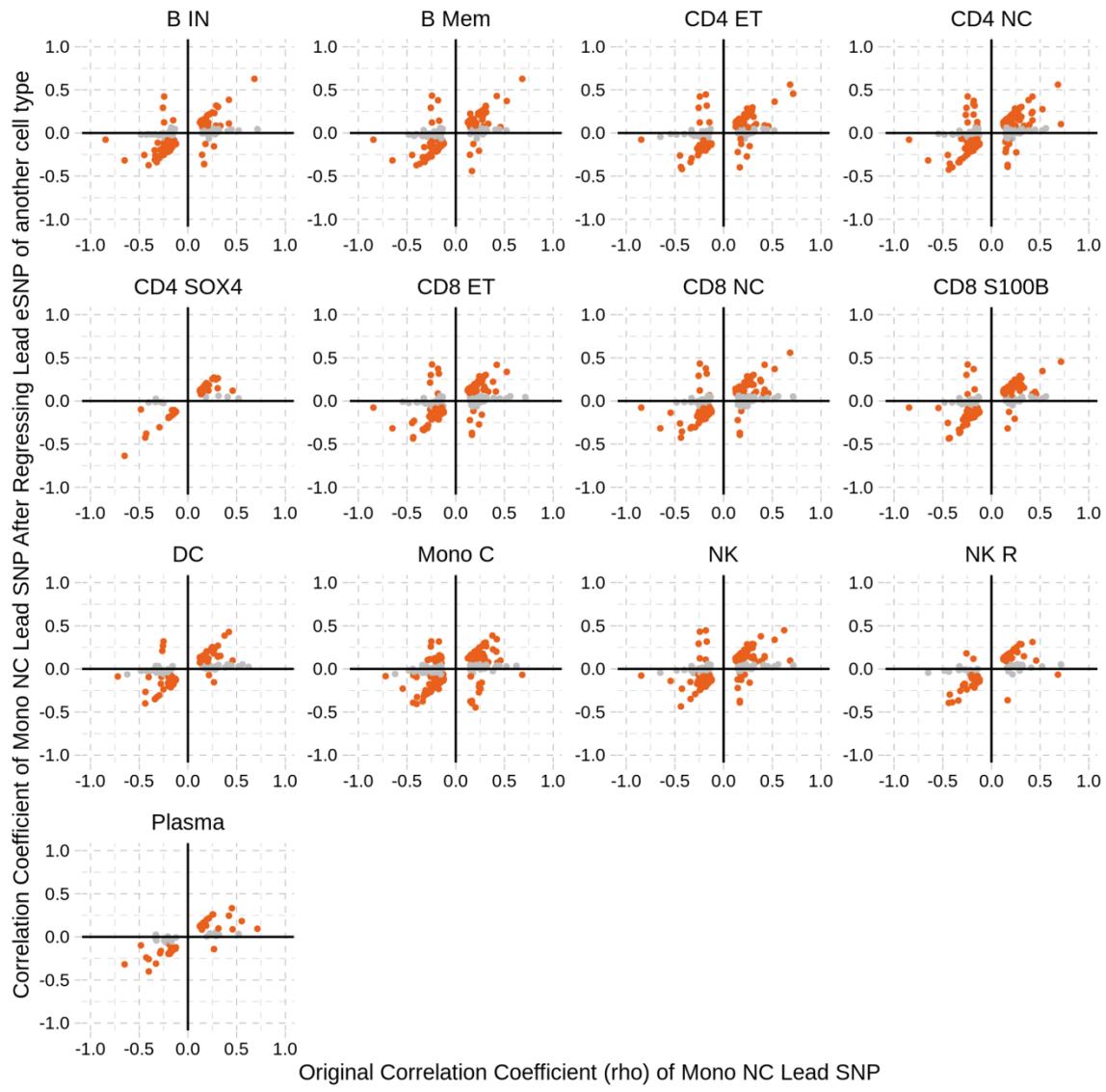












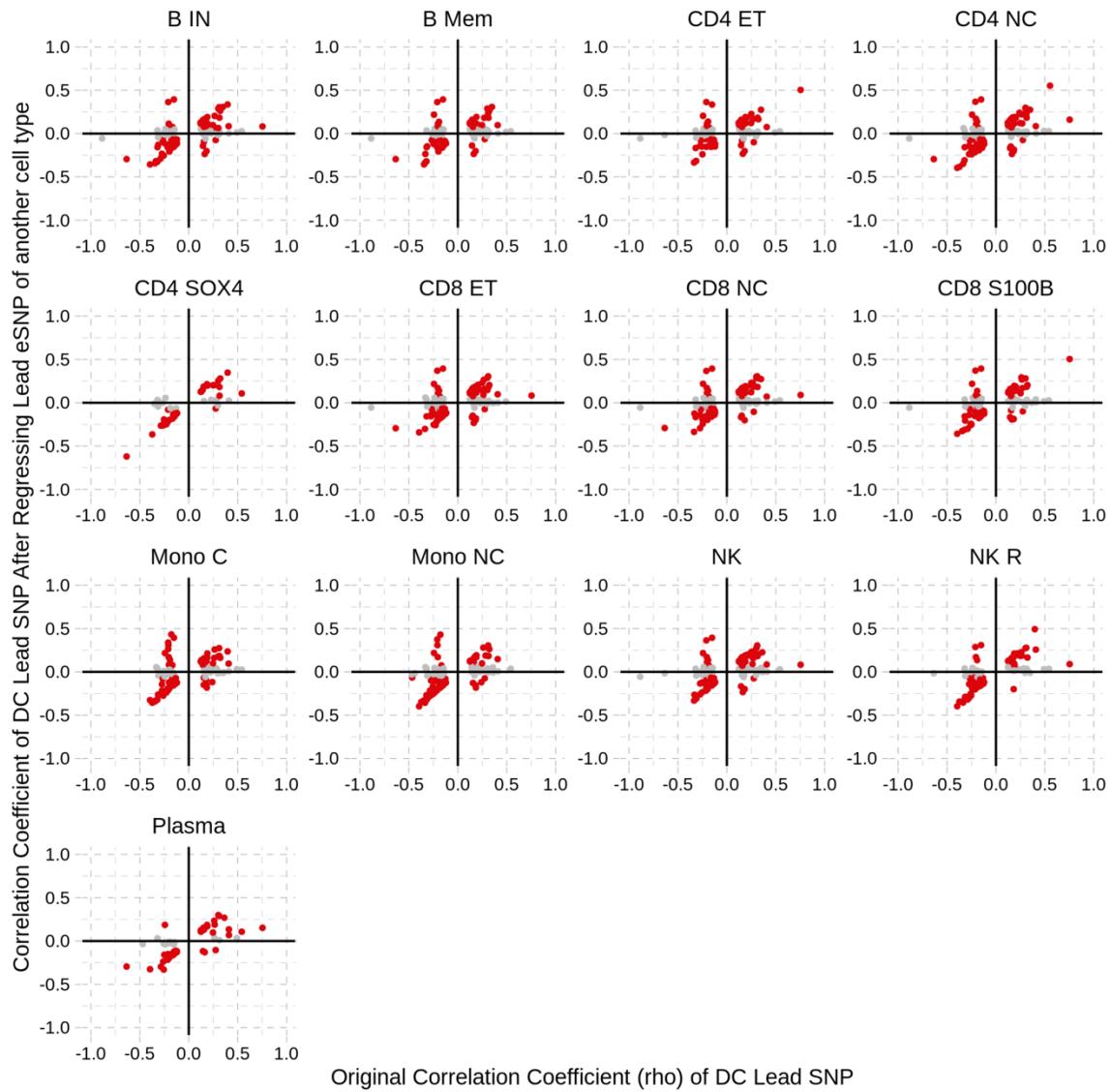


Fig. S16. Spearman's rank correlation coefficient for one cell type against another of all top *cis*-eQTLs that were identified in the subcell cluster.

Each dot in a plot within a panel shows whether an eGene was tagged by two independent SNPs. The lead eSNP from the first cell type was conditioned on the lead eSNP from the second cell type for every pair of cell types (182 pairs). (Colored dots=significant, grey dots=not significant).

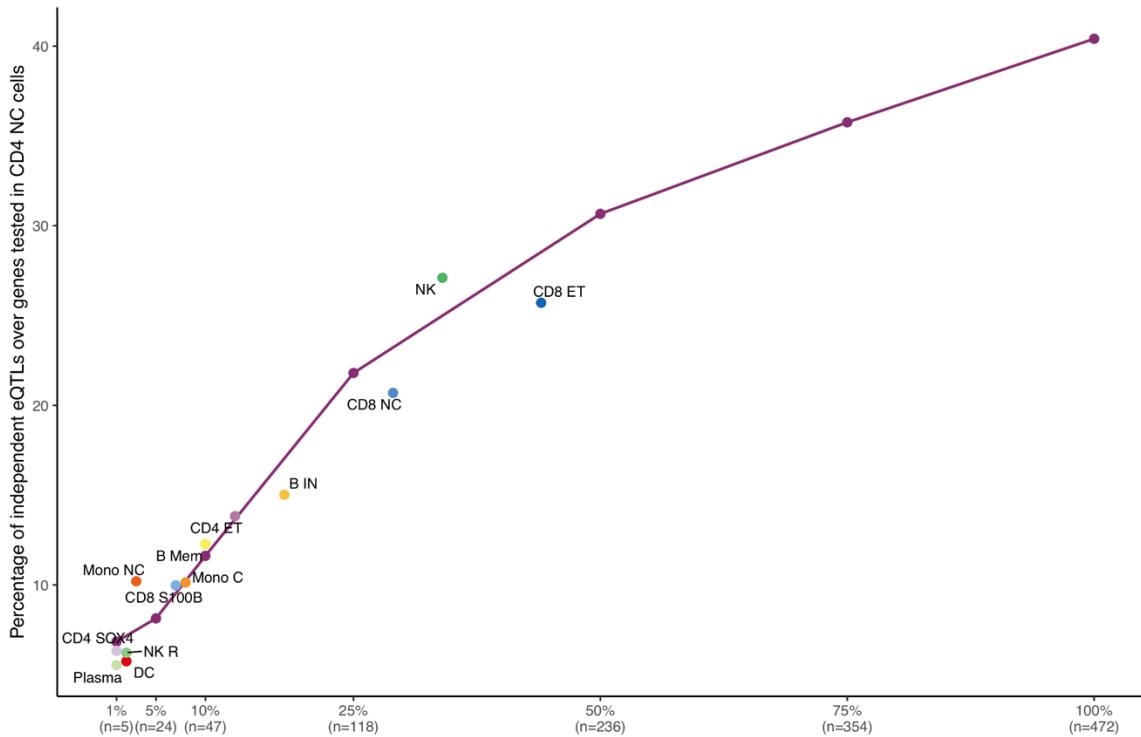


Fig. S17. Association of cell type cluster size and *cis*-eQTL discovery.

Number of independent *cis*-eQTL weighted by number of genes tested identified as a function of cell numbers. Average numbers of cells per person are indicated in parentheses in the *x*-axis. Labelled dots represent the number of *cis*-eQTLs discovered per cell type in the original analysis. Line joins data on the fractionation of CD4_{NC} population modelled.

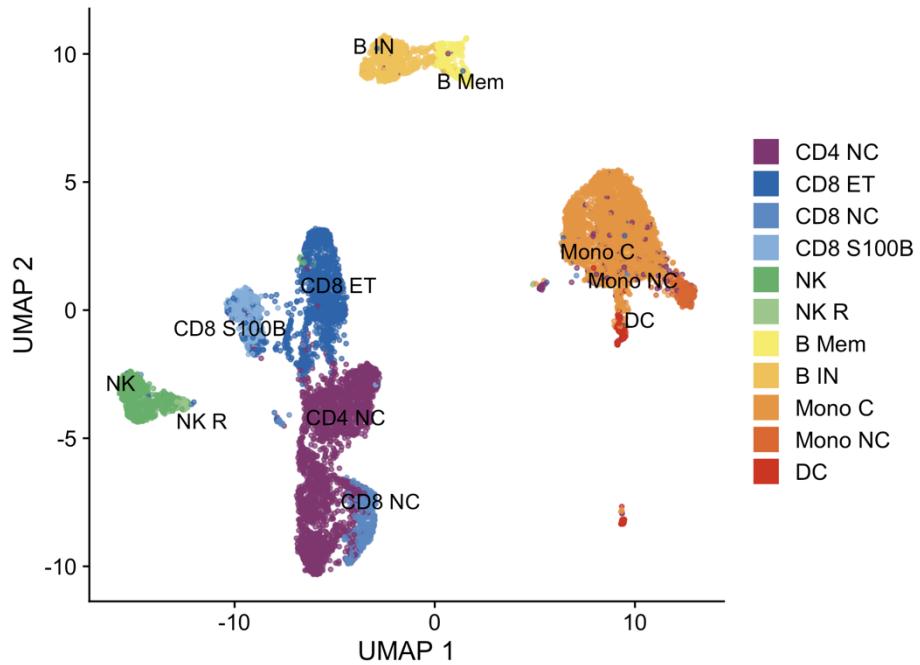


Fig. S18. UMAP visualization of 8876 cells from the scATAC-seq dataset.

Cell annotation corresponds to the cell types defined for the OneK1K dataset. Cell labels were transferred using Seurat V4. Only three cell types (Plasma, CD4_{ET}, CD4_{SOX4}) were absent after cell type classification.

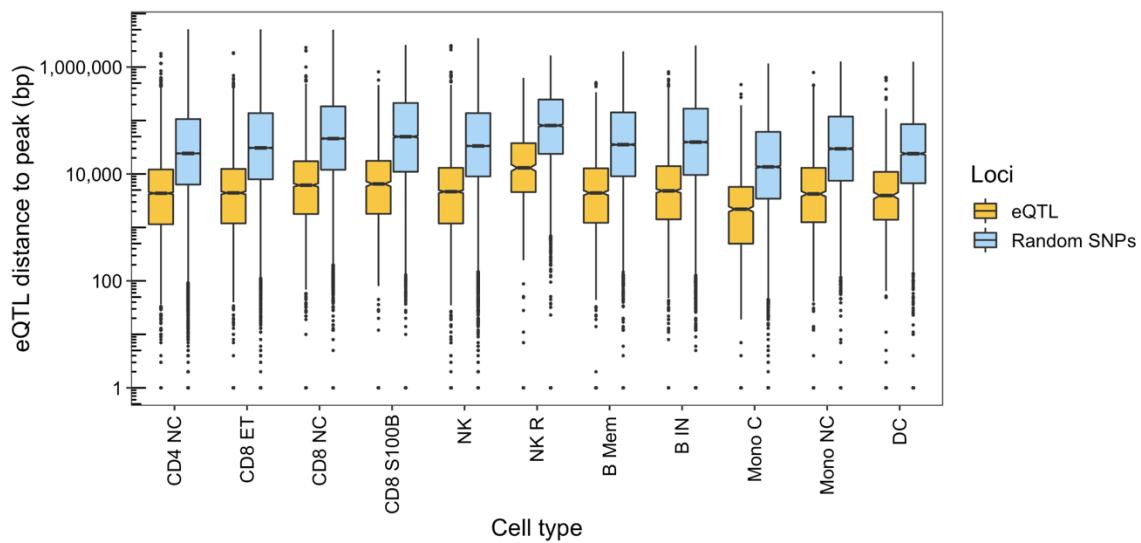
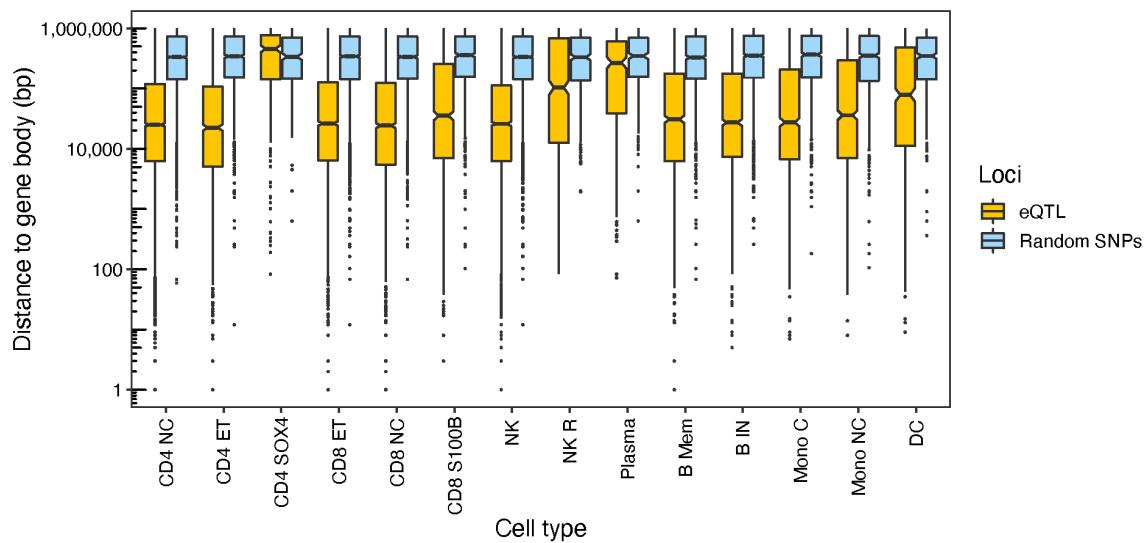
A**B**

Fig. S19. *cis*-eQTL distances to scATAC-seq peaks and gene body.

A) Genomic distances from *cis*-eQTLs identified in each population to the closest open chromatin region are shown in yellow. Random *cis*-SNPs tested for eQTL analysis are shown in blue. B) Genomic distances from the same set *cis*-eQTLs to the gene body of relevance. The y-axis in both plots shows the distances as log₁₀(base pairs + 1). Significant difference for all eQTL and random *cis*-SNPs (FDR < 0.05) across all cell types was found except for CD4sox4 (and not included in first figure due to its absence in the scATAC-seq data).

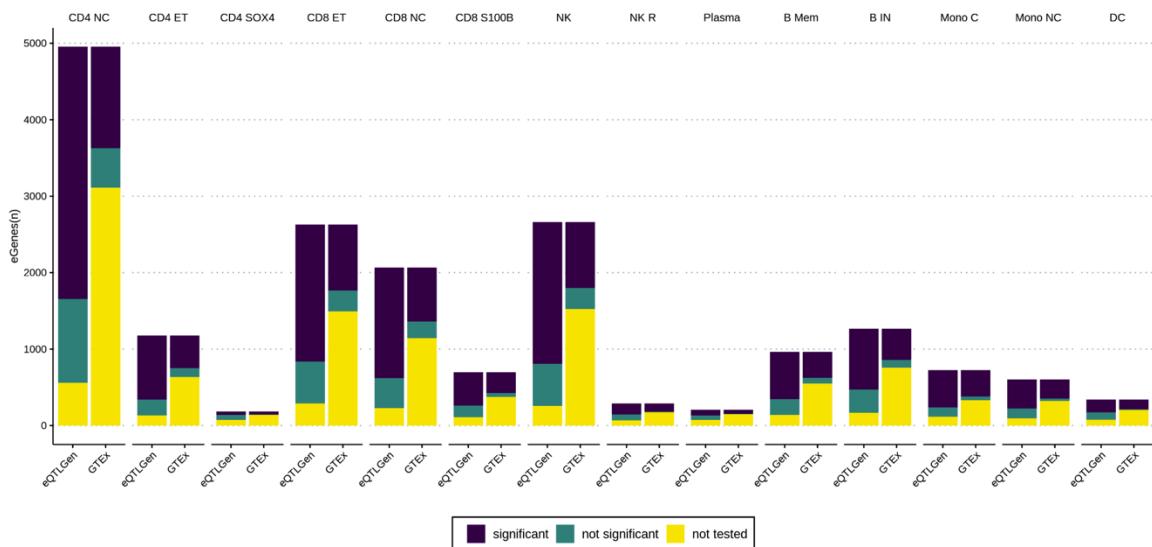


Fig. S20. Replication in eQTLGen and GTEx Consortia.

For each cell type, significant and non-significant eGenes as well as genes that are not tested (either they were not expressed or dropped off during QC step).

Pseudotime

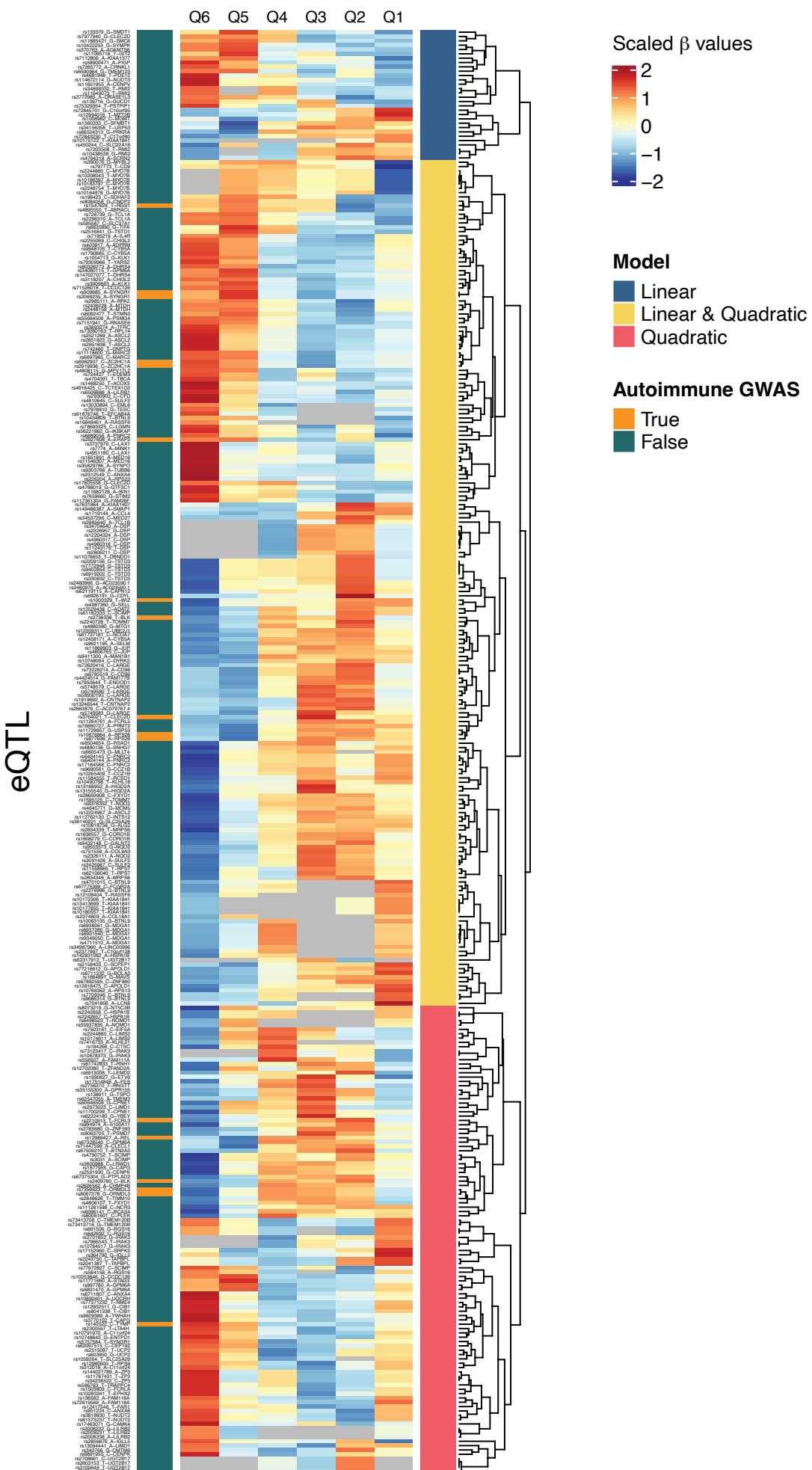


Fig. S21. eSNP-eGene pairs with a statistically significant change in *cis*-eQTL effect size across the B cell landscape.

Both linear and quadratic models were applied to SNP-gene pairs across the pseudotime quantiles, and SNPs known to be in high LD ($R^2 > 0.8$) with variants identified through GWAS of seven autoimmune diseases are displayed.

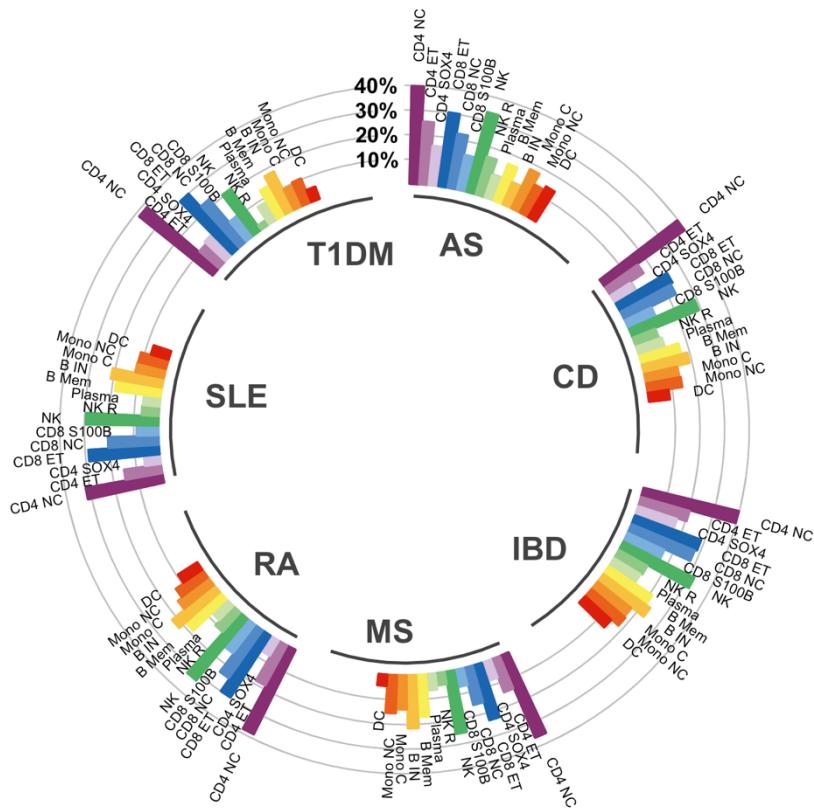


Fig. S22. Overlap of *cis*-eQTLs with GWAS catalog hits.

The percentage of overlapping cell type specific eQTLs between risk variants identified by GWAS for seven autoimmune diseases (MS=multiple sclerosis, RA=rheumatoid arthritis, SLE=systemic lupus erythematosus, AS=ankylosing spondylitis, CD=Crohn's disease, T1DM=type 1 diabetes mellitus, IBD=inflammatory bowel disease).

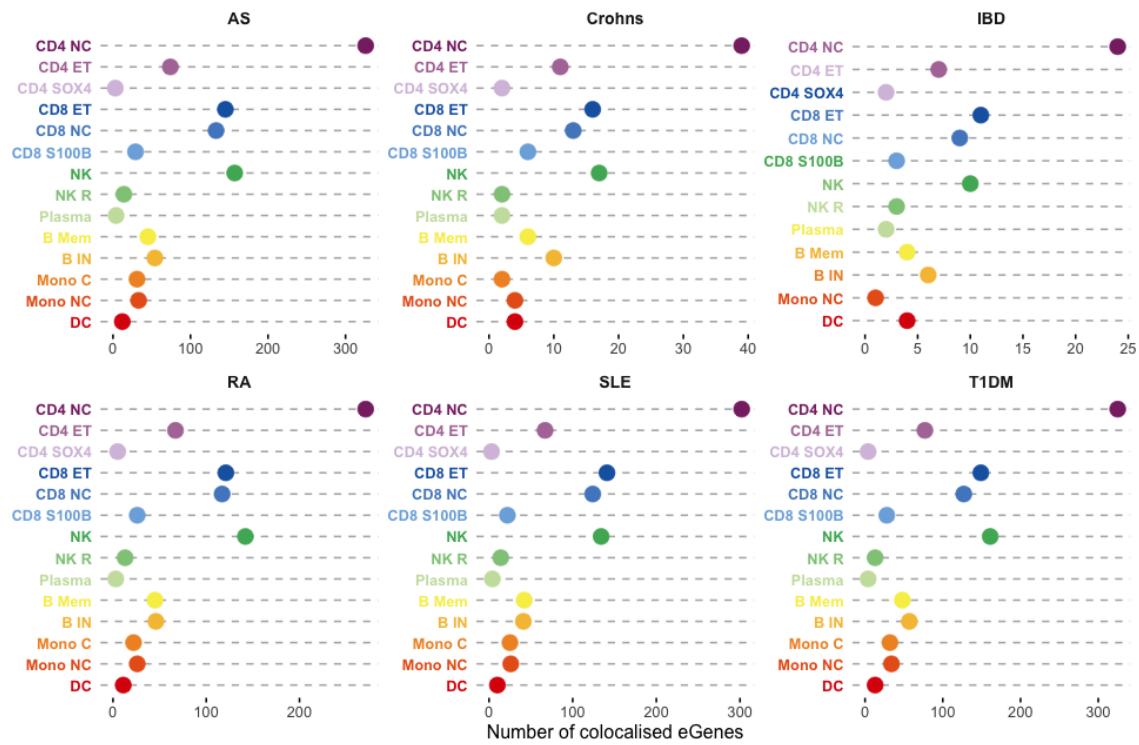


Fig. S23. Number of *cis*-eGenes colocalized with GWAS associations using Bayes factors.

(AS=ankylosing spondylitis, CD=Crohn's disease, IBD=inflammatory bowel disease, RA=rheumatoid arthritis, SLE=systemic lupus erythematosus, T1DM=type 1 diabetes mellitus)

Table S1.

Classical markers used for cell type identification.

Cell type	Abbreviation	Marker 1	Marker 2	Marker 3	Marker 4	Marker 5	Marker 6	Marker 7
CD4+ Naïve and Central Memory T cell	CD4 NC	<i>CD3D</i>	<i>CD4</i>	<i>CCR7</i>	<i>SELL</i>	<i>LRRN3</i>		
CD4+ Effector memory and central memory T cell	CD4 ET	<i>CD3D</i>	<i>CD4</i>	<i>KLRB1</i>	<i>GZMK</i>	<i>TNFSF13B</i>	<i>IL7R</i>	
CD4+ SOX4 T cell	CD4 SOX4	<i>CD3D</i>	<i>CD4</i>	<i>SOX4</i>	<i>ID2</i>	<i>SELL</i>		
CD8+ Naïve and Central memory T cell	CD8 NC	<i>CD3D</i>	<i>CD8A</i>	<i>LTB</i>	<i>CCR7</i>	<i>PASK</i>		
CD8+ Effector memory T cell	CD8 ET	<i>CD3D</i>	<i>CD8A</i>	<i>GNLY</i>	<i>NKG7</i>	<i>KLRB1</i>		
CD8+ S100B T cell	CD8 S100B	<i>CD3D</i>	<i>CD8A</i>	<i>S100B</i>	<i>KLRB1</i>	<i>LTB</i>	<i>IL7R</i>	<i>GZMK</i>
Natural killer cell	NK	<i>NCAMI</i>	<i>GZMA</i>	<i>GZMB</i>				
Natural killer cell Recruiting	NK R	<i>NCAMI</i>	<i>GZMK</i>	<i>XCL1</i>	<i>XCL2</i>			
Immature and Naïve B cell	B IN	<i>MS4A1</i>	<i>TCL1A</i>	<i>FCER2</i>	<i>IL4R</i>			
Memory B cell	B Mem	<i>MS4A1</i>	<i>CD27</i>	<i>TNFRSF13B</i>				
Plasma cell	Plasma	<i>TNFRSF17</i>	<i>IgJ</i>					
Classical Monocyte	Mono C	<i>CD14</i>	<i>LYZhi</i>					
Non-Classical Monocyte	Mono NC	<i>CD16</i>						
Dendritic cell	DC	<i>FCER1A</i>	<i>CST3</i>	<i>SERPINF1</i>				

Table S2.

Summary of cell type composition (%) in the OneK1K Study.

Cell type	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
CD4 NC	11.27	29.9	36.51	36.5	43.32	68.72
CD4 ET	0.86	3.69	4.81	5.04	6.09	12.68
CD4 SOX4	0.05	0.14	0.25	0.37	0.45	2.4
CD8 ET	0.76	8.00	13.84	15.83	21.25	53.08
CD8 NC	2.69	8.06	10.08	10.79	13.19	25.11
CD8 S100B	0.25	1.48	2.41	2.83	3.69	11.61
NK	1.68	7.72	11.37	12.39	15.67	46.18
NK R	0.06	0.40	0.67	0.81	1.03	4.95
B Mem	0.17	2.12	3.34	3.74	4.58	25.84
B IN	0.32	3.5	5.44	6.19	7.88	27.73
Plasma	0.05	0.12	0.23	0.36	0.42	8.46
B Mem	0.17	2.12	3.34	3.74	4.58	25.84
B IN	0.32	3.5	5.44	6.19	7.88	27.73
Monocyte C	0.06	0.76	1.78	3.18	4.24	23.85
Monocyte NC	0.05	0.34	0.79	1.24	1.73	11.96
DC	0.05	0.37	0.62	0.75	0.96	5.49

Table S3.

Summary statistics for association of age with cell counts.

Cell type	Beta estimate	SE	P-value
CD4 NC	-1.410	0.382	2.38E-04
CD4 ET	-0.424	0.058	7.99E-13
CD4 SOX4	-0.161	0.009	2.75E-64
CD8 ET	1.560	0.266	6.49E-09
CD8 NC	-1.880	0.114	3.24E-54
CD8 100B	-0.798	0.043	1.56E-65
NK	1.460	0.173	1.53E-16
NK R	-0.002	0.015	8.90E-01
Plasma	-0.038	0.013	3.26E-03
B IN	-0.050	0.144	7.27E-01
B Mem	-0.274	0.089	2.20E-03
Mono C	0.107	0.087	2.19E-01
Mono NC	0.171	0.033	3.00E-07
DC	-0.016	0.013	1.96E-01

Table S4.

Changes in the number of *cis*-eGenes with addition of each covariate. The final model used in the analysis is highlighted in grey.

MODEL	Number of eGenes
Sex	299
Sex + PC1	400
Sex + PC1 + PC2	409
Sex + PC1 + PC2 + PC3	404
Sex + PC1 + PC2 + PC3 + PC4	408
Sex + PC1 + PC2 + PC3 + PC4 + PC5	407
Sex + PC1 + PC2 + PC3 + PC4 + PC5 + PC6	409
Sex + PC1 + PC2 + PC3 + PC4 + PC5 + PC6 + Age	416
Sex + PC1 + PC2 + PC3 + PC4 + PC5 + PC6 + Age + PF1	422
Sex + PC1 + PC2 + PC3 + PC4 + PC5 + PC6 + Age + PF1 + PF2	424
Sex + PC1 + PC2 + PC3 + PC4 + PC5 + PC6 + Age + PF1 + PF2 + PF3	288
Sex + PC1 + PC2 + PC3 + PC4 + PC5 + PC6 + Age + PF1 + PF2 + PF3 + PF4	274
Sex + PC1 + PC2 + PC3 + PC4 + PC5 + PC6 + Age + PF1 + PF2 + PF3 + PF4 + PF5	278
Sex + PC1 + PC2 + PC3 + PC4 + PC5 + PC6 + Age + PF1 + PF2 + PF3 + PF4 + PF5 + PF6	279
Sex + PC1 + PC2 + PC3 + PC4 + PC5 + PC6 + Age + PF1 + PF2 + PF3 + PF4 + PF5 + PF6 + PF7	281
Sex + PC1 + PC2 + PC3 + PC4 + PC5 + PC6 + Age + PF1 + PF2 + PF3 + PF4 + PF5 + PF6 + PF7 + PF8	288
Sex + PC1 + PC2 + PC3 + PC4 + PC5 + PC6 + Age + PF1 + PF2 + PF3 + PF4 + PF5 + PF6 + PF7 + PF8 + PF9	286
Sex + PC1 + PC2 + PC3 + PC4 + PC5 + PC6 + Age + PF1 + PF2 + PF3 + PF4 + PF5 + PF6 + PF7 + PF8 + PF9 + PF10	282

Table S5.

Replication summary statistics for both cohorts.

Cell Type	Types	OneK1K Study	Total number of eSNPs	Total eQTLs (eGene- eSNP pairs) in OneK1K Study	Total number of lead eQTLs	Total number of eGene- eSNP pairs tested in Replicatio- n Cohort 1 & 2	Total number of significant eQTLs in Replica- tion Cohort 1	Total samples (at 0.05 FDR level)	Total Replicated samples	Total eQTLs in Replica- tion Cohort 1	Total Replicated samples (at 0.05 FDR level)	Total Replicated samples
			OneK1K Cell in OneK1K Study	Replicatio- n Cohort 1	Replicatio- n & 2	European samples	Cohort 1 European samples	Asian samples	Cohort 2 Asian samples	Cohort 2 Asian samples	Cohort 2 Asian samples	
B-cell	B IN and B Mem	2028	2194	1474	612	41.5	486	33.0				
CD4-T-cell	CD4 ET, CD4 NC, and CD4 SOX4	5185	5513	3732	1495	40.1	1011	27.1				
CD8-T-cell	CD8 ET, CD8 NC, and CD8 S100B	4336	4817	3320	1592	48.0	1221	36.8				
DC	DC	297	298	258	125	48.4	95	36.8				
Monocytes	Mono C, and Mono NC	1117	1181	922	594	64.4	428	46.4				
NK	NK, and NK R	2527	2594	1956	81	4.1	64	3.3				
TOTAL	-	15490	16597	11662	4499	38.6	3305	28.3				

References and Notes

1. P. Brodin, M. M. Davis, Human immune system variation. *Nat. Rev. Immunol.* **17**, 21–29 (2017). [doi:10.1038/nri.2016.125](https://doi.org/10.1038/nri.2016.125) [Medline](#)
2. L. Chen, B. Ge, F. P. Casale, L. Vasquez, T. Kwan, D. Garrido-Martín, S. Watt, Y. Yan, K. Kundu, S. Ecker, A. Datta, D. Richardson, F. Burden, D. Mead, A. L. Mann, J. M. Fernandez, S. Rowlston, S. P. Wilder, S. Farrow, X. Shao, J. J. Lambourne, A. Redensek, C. A. Albers, V. Amstislavskiy, S. Ashford, K. Berentsen, L. Bomba, G. Bourque, D. Bujold, S. Busche, M. Caron, S.-H. Chen, W. Cheung, O. Delaneau, E. T. Dermitzakis, H. Elding, I. Colgiu, F. O. Bagger, P. Flliceck, E. Habibi, V. Iotchkova, E. Janssen-Megens, B. Kim, H. Lehrach, E. Lowy, A. Mandoli, F. Matarese, M. T. Maurano, J. A. Morris, V. Pancaldi, F. Pourfarzad, K. Rehnstrom, A. Rendon, T. Risch, N. Sharifi, M.-M. Simon, M. Sultan, A. Valencia, K. Walter, S.-Y. Wang, M. Frontini, S. E. Antonarakis, L. Clarke, M.-L. Yaspo, S. Beck, R. Guigo, D. Rico, J. H. A. Martens, W. H. Ouwehand, T. W. Kuijpers, D. S. Paul, H. G. Stunnenberg, O. Stegle, K. Downes, T. Pastinen, N. Soranzo, 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](https://doi.org/10.1016/j.cell.2016.10.026) [Medline](#)
3. Y. I. Li, B. van de Geijn, A. Raj, D. A. Knowles, A. A. Petti, D. Golan, Y. Gilad, J. K. Pritchard, RNA splicing is a primary link between genetic variation and disease. *Science* **352**, 600–604 (2016). [doi:10.1126/science.aad9417](https://doi.org/10.1126/science.aad9417) [Medline](#)
4. F. Hormozdiari, S. Gazal, B. van de Geijn, H. K. Finucane, C. J.-T. Ju, P.-R. Loh, A. Schoech, Y. Reshef, X. Liu, L. O'Connor, A. Gusev, E. Eskin, A. L. Price, Leveraging molecular quantitative trait loci to understand the genetic architecture of diseases and complex traits. *Nat. Genet.* **50**, 1041–1047 (2018). [doi:10.1038/s41588-018-0148-2](https://doi.org/10.1038/s41588-018-0148-2) [Medline](#)
5. GTEx Consortium, The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318–1330 (2020). [doi:10.1126/science.aaz1776](https://doi.org/10.1126/science.aaz1776) [Medline](#)
6. W. J. Astle, H. Elding, T. Jiang, D. Allen, D. Ruklisa, A. L. Mann, D. Mead, H. Bouman, F. Riveros-Mckay, M. A. Kostadima, J. J. Lambourne, S. Sivapalaratnam, K. Downes, K. Kundu, L. Bomba, K. Berentsen, J. R. Bradley, L. C. Daugherty, O. Delaneau, K. Freson, S. F. Garner, L. Grassi, J. Guerrero, M. Haimel, E. M. Janssen-Megens, A. Kaan, M. Kamat, B. Kim, A. Mandoli, J. Marchini, J. H. A. Martens, S. Meacham, K. Megy, J. O'Connell, R. Petersen, N. Sharifi, S. M. Sheard, J. R. Staley, S. Tuna, M. van der Ent, K. Walter, S.-Y. Wang, E. Wheeler, S. P. Wilder, V. Iotchkova, C. Moore, J. Sambrook, H. G. Stunnenberg, E. Di Angelantonio, S. Kaptoge, T. W. Kuijpers, E. Carrillo-de-Santa-Pau, D. Juan, D. Rico, A. Valencia, L. Chen, B. Ge, L. Vasquez, T. Kwan, D. Garrido-Martín, S. Watt, Y. Yang, R. Guigo, S. Beck, D. S. Paul, T. Pastinen, D. Bujold, G. Bourque, M. Frontini, J. Danesh, D. J. Roberts, W. H. Ouwehand, A. S. Butterworth, N. Soranzo, 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](https://doi.org/10.1016/j.cell.2016.10.042) [Medline](#)
7. C. C. Whitacre, Sex differences in autoimmune disease. *Nat. Immunol.* **2**, 777–780 (2001). [doi:10.1038/ni0901-777](https://doi.org/10.1038/ni0901-777) [Medline](#)
8. 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](https://doi.org/10.1016/S1568-9972(03)00006-5) [Medline](#)
9. A. Buniello, J. A. L. MacArthur, M. Cerezo, L. W. Harris, J. Hayhurst, C. Malangone, A.

- McMahon, J. Morales, E. Mountjoy, E. Sollis, D. Suveges, O. Vrousgou, P. L. Whetzel, R. Amode, J. A. Guillen, H. S. Riat, S. J. Trevanion, P. Hall, H. Junkins, P. Flicek, T. Burdett, L. A. Hindorff, F. Cunningham, H. Parkinson, 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](https://doi.org/10.1093/nar/gky1120) [Medline](#)
10. L. A. Hindorff, P. Sethupathy, H. A. Junkins, E. M. Ramos, J. P. Mehta, F. S. Collins, T. A. Manolio, 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](https://doi.org/10.1073/pnas.0903103106) [Medline](#)
11. 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). [doi:10.1038/s41467-018-03766-z](https://doi.org/10.1038/s41467-018-03766-z) [Medline](#)
12. 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](https://doi.org/10.1136/annrheumdis-2018-213678) [Medline](#)
13. J. Bentham, D. L. Morris, D. S. Cunningham Graham, C. L. Pinder, P. Tombleson, T. W. Behrens, J. Martín, B. P. Fairfax, J. C. Knight, L. Chen, J. Replogle, A.-C. Syvänen, L. Rönnblom, R. R. Graham, J. E. Wither, J. D. Rioux, M. E. Alarcón-Riquelme, T. J. Vyse, 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](https://doi.org/10.1038/ng.3434) [Medline](#)
14. 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](https://doi.org/10.1016/j.tig.2020.08.009) [Medline](#)
15. GTEx Consortium, Genetic effects on gene expression across human tissues. *Nature* **550**, 204–213 (2017). [doi:10.1038/nature24277](https://doi.org/10.1038/nature24277) [Medline](#)
16. H.-J. Westra, M. J. Peters, T. Esko, H. Yaghootkar, C. Schurmann, J. Kettunen, M. W. Christiansen, B. P. Fairfax, K. Schramm, J. E. Powell, A. Zhernakova, D. V. Zhernakova, J. H. Veldink, L. H. Van den Berg, J. Karjalainen, S. Withoff, A. G. Uitterlinden, A. Hofman, F. Rivadeneira, P. A. C. Hoen, E. Reinmaa, K. Fischer, M. Nelis, L. Milani, D. Melzer, L. Ferrucci, A. B. Singleton, D. G. Hernandez, M. A. Nalls, G. Homuth, M. Nauck, D. Radke, U. Völker, M. Perola, V. Salomaa, J. Brody, A. Suchy-Dicey, S. A. Gharib, D. A. Enquobahrie, T. Lumley, G. W. Montgomery, S. Makino, H. Prokisch, C. Herder, M. Roden, H. Grallert, T. Meitinger, K. Strauch, Y. Li, R. C. Jansen, P. M. Visscher, J. C. Knight, B. M. Psaty, S. Ripatti, A. Teumer, T. M. Frayling, A. Metspalu, J. B. J. van Meurs, L. Franke, Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat. Genet.* **45**, 1238–1243 (2013). [doi:10.1038/ng.2756](https://doi.org/10.1038/ng.2756) [Medline](#)
17. L. R. Lloyd-Jones, A. Holloway, A. McRae, J. Yang, K. Small, J. Zhao, B. Zeng, A. Bakshi, A. Metspalu, M. Dermitzakis, G. Gibson, T. Spector, G. Montgomery, T. Esko, P. M. Visscher, J. E. Powell, 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](https://doi.org/10.1016/j.ajhg.2016.12.008) [Medline](#)

18. C. J. Ye, T. Feng, H.-K. Kwon, T. Raj, M. T. Wilson, N. Asinovski, C. McCabe, M. H. Lee, I. Frohlich, H.-I. Paik, N. Zaitlen, N. Hacohen, B. Stranger, P. De Jager, D. Mathis, A. Regev, C. Benoist, Intersection of population variation and autoimmunity genetics in human T cell activation. *Science* **345**, 1254665 (2014). [doi:10.1126/science.1254665](https://doi.org/10.1126/science.1254665) [Medline](#)
19. J. E. Powell, A. K. Henders, A. F. McRae, M. J. Wright, N. G. Martin, E. T. Dermitzakis, G. W. Montgomery, P. M. Visscher, 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](https://doi.org/10.1101/gr.126540.111) [Medline](#)
20. K. Ishigaki, Y. Kochi, A. Suzuki, Y. Tsuchida, H. Tsuchiya, S. Sumitomo, K. Yamaguchi, Y. Nagafuchi, S. Nakachi, R. Kato, K. Sakurai, H. Shoda, K. Ikari, A. Taniguchi, H. Yamanaka, F. Miya, T. Tsunoda, Y. Okada, Y. Momozawa, Y. Kamatani, R. Yamada, M. Kubo, K. Fujio, K. Yamamoto, Polygenic burdens on cell-specific pathways underlie the risk of rheumatoid arthritis. *Nat. Genet.* **49**, 1120–1125 (2017). [doi:10.1038/ng.3885](https://doi.org/10.1038/ng.3885) [Medline](#)
21. B. J. Schmiedel, D. Singh, A. Madrigal, A. G. Valdovino-Gonzalez, B. M. White, J. Zapardiel-Gonzalo, B. Ha, G. Altay, J. A. Greenbaum, G. McVicker, G. Seumois, A. Rao, M. Kronenberg, B. Peters, P. Vijayanand, Impact of genetic polymorphisms on human immune cell gene expression. *Cell* **175**, 1701–1715.e16 (2018). [doi:10.1016/j.cell.2018.10.022](https://doi.org/10.1016/j.cell.2018.10.022) [Medline](#)
22. S. Kim-Hellmuth, F. Aguet, M. Oliva, M. Muñoz-Aguirre, S. Kasela, V. Wucher, S. E. Castel, A. R. Hamel, A. Viñuela, A. L. Roberts, S. Mangul, X. Wen, G. Wang, A. N. Barbeira, D. Garrido-Martín, B. B. Nadel, Y. Zou, R. Bonazzola, J. Quan, A. Brown, A. Martinez-Perez, J. M. Soria, G. Getz, E. T. Dermitzakis, K. S. Small, M. Stephens, H. S. Xi, H. K. Im, R. Guigó, A. V. Segrè, B. E. Stranger, K. G. Ardlie, T. Lappalainen, GTEx Consortium, Cell type–specific genetic regulation of gene expression across human tissues. *Science* **369**, eaaz8528 (2020). [doi:10.1126/science.aaz8528](https://doi.org/10.1126/science.aaz8528) [Medline](#)
23. 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](https://doi.org/10.1038/s41467-020-19015-1) [Medline](#)
24. S. McCarthy, S. Das, W. Kretzschmar, O. Delaneau, A. R. Wood, A. Teumer, H. M. Kang, C. Fuchsberger, P. Danecek, K. Sharp, Y. Luo, C. Sidore, A. Kwong, N. Timpson, S. Koskinen, S. Vrieze, L. J. Scott, H. Zhang, A. Mahajan, J. Veldink, U. Peters, C. Pato, C. M. van Duijn, C. E. Gillies, I. Gandin, M. Mezzavilla, A. Gilly, M. Cocca, M. Traglia, A. Angius, J. C. Barrett, D. Boomsma, K. Branham, G. Breen, C. M. Brummett, F. Busonero, H. Campbell, A. Chan, S. Chen, E. Chew, F. S. Collins, L. J. Corbin, G. D. Smith, G. Dedoussis, M. Dorr, A.-E. Farmaki, L. Ferrucci, L. Forer, R. M. Fraser, S. Gabriel, S. Levy, L. Groop, T. Harrison, A. Hattersley, O. L. Holmen, K. Hveem, M. Kretzler, J. C. Lee, M. McGue, T. Meitinger, D. Melzer, J. L. Min, K. L. Mohlke, J. B. Vincent, M. Nauck, D. Nickerson, A. Palotie, M. Pato, N. Pirastu, M. McInnis, J. B. Richards, C. Sala, V. Salomaa, D. Schlessinger, S. Schoenherr, P. E. Slagboom, K. Small, T. Spector, D. Stambolian, M. Tuke, J. Tuomilehto, L. H. Van den Berg, W. Van Rheenen, U. Volker, C. Wijmenga, D. Toniolo, E. Zeggini, P. Gasparini, M. G. Sampson, J. F. Wilson, T. Frayling, P. I. W. de Bakker, M. A. Swertz, S. McCarroll, C. Kooperberg, A. Dekker, D. Altshuler, C. Willer, W. Iacono, S. Ripatti, N. Soranzo, K. Walter, A. Swaroop, F. Cucca, C. A.

- Anderson, R. M. Myers, M. Boehnke, M. I. McCarthy, R. Durbin, Haplotype Reference Consortium, A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.* **48**, 1279–1283 (2016). [doi:10.1038/ng.3643](https://doi.org/10.1038/ng.3643) [Medline](#)
25. 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](https://doi.org/10.1186/s13059-019-1862-5) [Medline](#)
26. 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**, 14049 (2017). [doi:10.1038/ncomms14049](https://doi.org/10.1038/ncomms14049) [Medline](#)
27. V. Orrù, M. Steri, G. Sole, C. Sidore, F. Virdis, M. Dei, S. Lai, M. Zoledziewska, F. Busonero, A. Mulas, M. Floris, W. I. Mentzen, S. A. M. Urru, S. Olla, M. Marongiu, M. G. Piras, M. Lobina, A. Maschio, M. Pitzalis, M. F. Urru, M. Marcelli, R. Cusano, F. Deidda, V. Serra, M. Oppo, R. Pilu, F. Reinier, R. Berutti, L. Pireddu, I. Zara, E. Porcu, A. Kwong, C. Brennan, B. Tarrier, R. Lyons, H. M. Kang, S. Uzzau, R. Atzeni, M. Valentini, D. Firinu, L. Leoni, G. Rotta, S. Naitza, A. Angius, M. Congia, M. B. Whalen, C. M. Jones, D. Schlessinger, G. R. Abecasis, E. Fiorillo, S. Sanna, F. Cucca, Genetic variants regulating immune cell levels in health and disease. *Cell* **155**, 242–256 (2013). [doi:10.1016/j.cell.2013.08.041](https://doi.org/10.1016/j.cell.2013.08.041) [Medline](#)
28. P. Brodin, V. Jojic, T. Gao, S. Bhattacharya, C. J. L. Angel, D. Furman, S. Shen-Orr, C. L. Dekker, G. E. Swan, A. J. Butte, H. T. Maecker, M. M. Davis, 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](https://doi.org/10.1016/j.cell.2014.12.020) [Medline](#)
29. E. Cano-Gamez, B. Soskic, T. I. Roumeliotis, E. So, D. J. Smyth, M. Baldridge, 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**, 1801 (2020). [doi:10.1038/s41467-020-15543-y](https://doi.org/10.1038/s41467-020-15543-y) [Medline](#)
30. A.-C. Villani, R. Satija, G. Reynolds, S. Sarkizova, K. Shekhar, J. Fletcher, M. Griesbeck, A. Butler, S. Zheng, S. Lazo, L. Jardine, D. Dixon, E. Stephenson, E. Nilsson, I. Grundberg, D. McDonald, A. Filby, W. Li, P. L. De Jager, O. Rozenblatt-Rosen, A. A. Lane, M. Haniffa, A. Regev, N. Hacohen, Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. *Science* **356**, eaah4573 (2017). [doi:10.1126/science.aah4573](https://doi.org/10.1126/science.aah4573) [Medline](#)
31. B. P. Fairfax, P. Humburg, S. Makino, V. Naranbhai, D. Wong, E. Lau, L. Jostins, K. Plant, R. Andrews, C. McGee, J. C. Knight, Innate immune activity conditions the effect of regulatory variants upon monocyte gene expression. *Science* **343**, 1246949 (2014). [doi:10.1126/science.1246949](https://doi.org/10.1126/science.1246949) [Medline](#)
32. M. A. Christophorou, G. Castelo-Branco, R. P. Halley-Stott, C. S. Oliveira, R. Loos, A. Radziszewska, K. A. Mowen, P. Bertone, J. C. R. Silva, M. Zernicka-Goetz, M. L. Nielsen, J. B. Gurdon, T. Kouzarides, Citrullination regulates pluripotency and histone H1 binding to chromatin. *Nature* **507**, 104–108 (2014). [doi:10.1038/nature12942](https://doi.org/10.1038/nature12942) [Medline](#)

33. F. Cornélis, S. Fauré, M. Martinez, J. F. Prud'homme, P. Fritz, C. Dib, H. Alves, P. Barrera, N. de Vries, A. Balsa, D. Pascual-Salcedo, K. Maenaut, R. Westhovens, P. Migliorini, T. H. Tran, A. Delaye, N. Prince, C. Lefevre, G. Thomas, M. Poirier, S. Soubigou, O. Alibert, S. Lasbleiz, S. Fouix, C. Bouchier, F. Lioté, M. N. Loste, V. Lepage, D. Charron, G. Gyapay, A. Lopes-Vaz, D. Kuntz, T. Bardin, J. Weissenbach, ECRAF, 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](https://doi.org/10.1073/pnas.95.18.10746) [Medline](#)
34. U. Võsa, A. Claringbould, H.-J. Westra, M. J. Bonder, P. Deelen, B. Zeng, H. Kirsten, A. Saha, R. Kreuzhuber, S. Yazar, H. Brugge, R. Oelen, D. H. de Vries, M. G. P. van der Wijst, S. Kasela, N. Pervjakova, I. Alves, M.-J. Favé, M. Agbessi, M. W. Christiansen, R. Jansen, I. Seppälä, L. Tong, A. Teumer, K. Schramm, G. Hemani, J. Verlouw, H. Yaghootkar, R. Sönmez Flitman, A. Brown, V. Kukushkina, A. Kalnapanakis, S. Rüeger, E. Porcu, J. Kronberg, J. Kettunen, B. Lee, F. Zhang, T. Qi, J. A. Hernandez, W. Arindrarto, F. Beutner, J. Dmitrieva, M. Elansary, B. P. Fairfax, M. Georges, B. T. Heijmans, A. W. Hewitt, M. Kähönen, Y. Kim, J. C. Knight, P. Kovacs, K. Krohn, S. Li, M. Loeffler, U. M. Marigorta, H. Mei, Y. Momozawa, M. Müller-Nurasyid, M. Nauck, M. G. Nivard, B. W. J. H. Penninx, J. K. Pritchard, O. T. Raitakari, O. Rotzschke, E. P. Slagboom, C. D. A. Stehouwer, M. Stumvoll, P. Sullivan, P. A. C. 't Hoen, J. Thiery, A. Tönjes, J. van Dongen, M. van Iterson, J. H. Veldink, U. Völker, R. Warmerdam, C. Wijmenga, M. Swertz, A. Andiappan, G. W. Montgomery, S. Ripatti, M. Perola, Z. Katalik, E. Dermitzakis, S. Bergmann, T. Frayling, J. van Meurs, H. Prokisch, H. Ahsan, B. L. Pierce, T. Lehtimäki, D. I. Boomsma, B. M. Psaty, S. A. Gharib, P. Awadalla, L. Milani, W. H. Ouwehand, K. Downes, O. Stegle, A. Battle, P. M. Visscher, J. Yang, M. Scholz, J. Powell, G. Gibson, T. Esko, L. Franke, J. Yang, BIOS Consortium, i2QTL Consortium, 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](https://doi.org/10.1038/s41588-021-00913-z) [Medline](#)
35. J. W. Said, K. K. Hoyer, S. W. French, L. Rosenfelt, M. Garcia-Lloret, P. J. Koh, T. C. Cheng, G. G. Sulur, G. S. Pinkus, W. M. Kuehl, D. J. Rawlings, R. Wall, M. A. Teitel, 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](https://doi.org/10.1038/labinvest.3780264) [Medline](#)
36. E. F. Wagner, N. Hanna, L. D. Fast, N. Kouttab, P. R. Shank, A. Vazquez, S. Sharma, Novel diversity in IL-4-mediated responses in resting human naive B cells versus germinal center/memory B cells. *J. Immunol.* **165**, 5573–5579 (2000).
[doi:10.4049/jimmunol.165.10.5573](https://doi.org/10.4049/jimmunol.165.10.5573) [Medline](#)
37. 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](https://doi.org/10.1084/jem.188.9.1691) [Medline](#)
38. A. Kallies, J. Hasbold, K. Fairfax, C. Pridans, D. Emslie, B. S. McKenzie, A. M. Lew, L. M. Corcoran, P. D. Hodgkin, D. M. Tarlinton, S. L. Nutt, 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](https://doi.org/10.1016/j.jimmuni.2007.04.007) [Medline](#)
39. 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). [Medline](#)

40. M. Togni, K. D. Swanson, S. Reimann, S. Kliche, A. C. Pearce, L. Simeoni, D. Reinhold, J. Wienands, B. G. Neel, B. Schraven, A. Gerber, 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](https://doi.org/10.1128/MCB.25.18.8052-8063.2005) [Medline](#)
41. C. D. Langefeld, H. C. Ainsworth, D. S. Cunningham Graham, J. A. Kelly, M. E. Comeau, M. C. Marion, T. D. Howard, P. S. Ramos, J. A. Croker, D. L. Morris, J. K. Sandling, J. C. Almlöf, E. M. Acevedo-Vásquez, G. S. Alarcón, A. M. Babini, V. Baca, A. A. Bengtsson, G. A. Berbotto, M. Bijl, E. E. Brown, H. I. Brunner, M. H. Cardiel, L. Catoggio, R. Cervera, J. M. Cucho-Venegas, S. R. Dahlqvist, S. D’Alfonso, B. M. Da Silva, I. de la Rúa Figueroa, A. Doria, J. C. Edberg, E. Endreffy, J. A. Esquivel-Valerio, P. R. Fortin, B. I. Freedman, J. Frostegård, M. A. García, I. G. de la Torre, G. S. Gilkeson, D. D. Gladman, I. Gunnarsson, J. M. Guthridge, J. L. Huggins, J. A. James, C. G. M. Kallenberg, D. L. Kamen, D. R. Karp, K. M. Kaufman, L. C. Kottyan, L. Kovács, H. Lastrupup, B. R. Lauwerys, Q.-Z. Li, M. A. Maradiaga-Ceceña, J. Martín, J. M. McCune, D. R. McWilliams, J. T. Merrill, P. Miranda, J. F. Moctezuma, S. K. Nath, T. B. Niewold, L. Orozco, N. Ortego-Centeno, M. Petri, C. A. Pineau, B. A. Pons-Estel, J. Pope, P. Raj, R. Ramsey-Goldman, J. D. Reveille, L. P. Russell, J. M. Sabio, C. A. Aguilar-Salinas, H. R. Scherbarth, R. Scorza, M. F. Seldin, C. Sjöwall, E. Svenungsson, S. D. Thompson, S. M. A. Toloza, L. Truedsson, T. Tusié-Luna, C. Vasconcelos, L. M. Vilá, D. J. Wallace, M. H. Weisman, J. E. Wither, T. Bhangale, J. R. Oksenberg, J. D. Rioux, P. K. Gregersen, A.-C. Syvänen, L. Rönnblom, L. A. Criswell, C. O. Jacob, K. L. Sivils, B. P. Tsao, L. E. Schanberg, T. W. Behrens, E. D. Silverman, M. E. Alarcón-Riquelme, R. P. Kimberly, J. B. Harley, E. K. Wakeland, R. R. Graham, P. M. Gaffney, T. J. Vyse, Transancestral mapping and genetic load in systemic lupus erythematosus. *Nat. Commun.* **8**, 16021 (2017). [doi:10.1038/ncomms16021](https://doi.org/10.1038/ncomms16021) [Medline](#)
42. P. K. Gregersen, C. I. Amos, A. T. Lee, Y. Lu, E. F. Remmers, D. L. Kastner, M. F. Seldin, L. A. Criswell, R. M. Plenge, V. M. Holers, T. R. Mikuls, T. Sokka, L. W. Moreland, S. L. Bridges Jr., G. Xie, A. B. Begovich, K. A. Siminovitch, 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](https://doi.org/10.1038/ng.395) [Medline](#)
43. F. Köntgen, R. J. Grumont, A. Strasser, D. Metcalf, R. Li, D. Tarlinton, S. Gerondakis, 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](https://doi.org/10.1101/gad.9.16.1965) [Medline](#)
44. J. Dang, X. Bian, X. Ma, J. Li, F. Long, S. Shan, Q. Yuan, Q. Xin, Y. Li, F. Gao, Y. Gong, Q. Liu, 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](https://doi.org/10.4049/jimmunol.1602124) [Medline](#)
45. A. H. Beecham, N. A. Patsopoulos, D. K. Xifara, M. F. Davis, A. Kemppinen, C. Cotsapas, T. S. Shah, C. Spencer, D. Booth, A. Goris, A. Oturai, J. Saarela, B. Fontaine, B. Hemmer, C. Martin, F. Zipp, S. D’Alfonso, F. Martinelli-Boneschi, B. Taylor, H. F. Harbo, I. Kockum, J. Hillert, T. Olsson, M. Ban, J. R. Oksenberg, R. Hintzen, L. F. Barcellos, C. Agliardi, L. Alfredsson, M. Alizadeh, C. Anderson, R. Andrews, H. B. Søndergaard, A. Baker, G. Band, S. E. Baranzini, N. Barizzone, J.

Barrett, C. Bellenguez, L. Bergamaschi, L. Bernardinelli, A. Berthele, V. Biberacher, T. M. Binder, H. Blackburn, I. L. Bomfim, P. Brambilla, S. Broadley, B. Brochet, L. Brundin, D. Buck, H. Butzkueven, S. J. Caillier, W. Camu, W. Carpentier, P. Cavalla, E. G. Celius, I. Coman, G. Comi, L. Corrado, L. Cosemans, I. Cournu-Rebeix, B. A. Cree, D. Cusi, V. Damotte, G. Defer, S. R. Delgado, P. Deloukas, A. di Sapiro, A. T. Dilthey, P. Donnelly, B. Dubois, M. Duddy, S. Edkins, I. Elovaara, F. Esposito, N. Evangelou, B. Fiddes, J. Field, A. Franke, C. Freeman, I. Y. Frohlich, D. Galimberti, C. Gieger, P. A. Gourraud, C. Graetz, A. Graham, V. Grummel, C. Guaschino, A. Hadjixenofontos, H. Hakonarson, C. Halfpenny, G. Hall, P. Hall, A. Hamsten, J. Harley, T. Harrower, C. Hawkins, G. Hellenthal, C. Hillier, J. Hobart, M. Hoshi, S. E. Hunt, M. Jagodic, I. Jelčić, A. Jochim, B. Kendall, A. Kermode, T. Kilpatrick, K. Koivisto, I. Konidari, T. Korn, H. Kronsbein, C. Langford, M. Larsson, M. Lathrop, C. Lebrun-Frenay, J. Lechner-Scott, M. H. Lee, M. A. Leone, V. Leppä, G. Liberatore, B. A. Lie, C. M. Lill, M. Lindén, J. Link, F. Luessi, J. Lycke, F. Macciardi, S. Männistö, C. P. Manrique, R. Martin, V. Martinelli, D. Mason, G. Mazibrada, C. McCabe, I. L. Mero, J. Mescheriakova, L. Moutsianas, K. M. Myhr, G. Nagels, R. Nicholas, P. Nilsson, F. Piehl, M. Pirinen, S. E. Price, H. Quach, M. Reunanen, W. Robberecht, N. P. Robertson, M. Rodegher, D. Rog, M. Salvetti, N. C. Schnetz-Boutaud, F. Sellebjerg, R. C. Selter, C. Schaefer, S. Shaunak, L. Shen, S. Shields, V. Siffrin, M. Slee, P. S. Sorensen, M. Sorosina, M. Sospedra, A. Spurkland, A. Strange, E. Sundqvist, V. Thijs, J. Thorpe, A. Ticca, P. Tienari, C. van Duijn, E. M. Visser, S. Vucic, H. Westerlind, J. S. Wiley, A. Wilkins, J. F. Wilson, J. Winkelmann, J. Zajicek, E. Zindler, J. L. Haines, M. A. Pericak-Vance, A. J. Ivinson, G. Stewart, D. Hafler, S. L. Hauser, A. Compston, G. McVean, P. De Jager, S. J. Sawcer, J. L. McCauley, International Multiple Sclerosis Genetics Consortium (IMSGC), Wellcome Trust Case Control Consortium 2 (WTCCC2), International IBD Genetics Consortium (IIBDGC), Analysis of immune-related loci identifies 48 new susceptibility variants for multiple sclerosis. *Nat. Genet.* **45**, 1353–1360 (2013). [doi:10.1038/ng.2770](https://doi.org/10.1038/ng.2770) [Medline](#)

46. K. M. de Lange, L. Moutsianas, J. C. Lee, C. A. Lamb, Y. Luo, N. A. Kennedy, L. Jostins, D. L. Rice, J. Gutierrez-Achury, S.-G. Ji, G. Heap, E. R. Nimmo, C. Edwards, P. Henderson, C. Mowat, J. Sanderson, J. Satsangi, A. Simmons, D. C. Wilson, M. Tremelling, A. Hart, C. G. Mathew, W. G. Newman, M. Parkes, C. W. Lees, H. Uhlig, C. Hawkey, N. J. Prescott, T. Ahmad, J. C. Mansfield, C. A. Anderson, J. C. Barrett, 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](https://doi.org/10.1038/ng.3760) [Medline](#)
47. S. Onengut-Gumuscu, W.-M. Chen, O. Burren, N. J. Cooper, A. R. Quinlan, J. C. Mychaleckyj, E. Farber, J. K. Bonnie, M. Szpak, E. Schofield, P. Achuthan, H. Guo, M. D. Fortune, H. Stevens, N. M. Walker, L. D. Ward, A. Kundaje, M. Kellis, M. J. Daly, J. C. Barrett, J. D. Cooper, P. Deloukas, J. A. Todd, C. Wallace, P. Concannon, S. S. Rich, Type 1 Diabetes Genetics Consortium, 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](https://doi.org/10.1038/ng.3245) [Medline](#)
48. J. C. Barrett, D. G. Clayton, P. Concannon, B. Akolkar, J. D. Cooper, H. A. Erlich, C. Julier, G. Morahan, J. Nerup, C. Nierras, V. Plagnol, F. Pociot, H. Schuilenburg, D. J. Smyth, H. Stevens, J. A. Todd, N. M. Walker, S. S. Rich, Type 1 Diabetes Genetics Consortium, 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](https://doi.org/10.1038/ng.381)

Medline

49. 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](https://doi.org/10.1038/304030a0) [Medline](#)
50. M.-H. Chen, L. M. Raffield, A. Mousas, S. Sakaue, J. E. Huffman, A. Moscati, B. Trivedi, T. Jiang, P. Akbari, D. Vuckovic, E. L. Bao, X. Zhong, R. Manansala, V. Laplante, M. Chen, K. S. Lo, H. Qian, C. A. Lareau, M. Beaudoin, K. A. Hunt, M. Akiyama, T. M. Bartz, Y. Ben-Shlomo, A. Beswick, J. Bork-Jensen, E. P. Bottinger, J. A. Brody, F. J. A. van Rooij, K. Chitrala, K. Cho, H. Choquet, A. Correa, J. Danesh, E. Di Angelantonio, N. Dimou, J. Ding, P. Elliott, T. Esko, M. K. Evans, J. S. Floyd, L. Broer, N. Grarup, M. H. Guo, A. Greinacher, J. Haessler, T. Hansen, J. M. M. Howson, Q. Q. Huang, W. Huang, E. Jorgenson, T. Kacprowski, M. Kähönen, Y. Kamatani, M. Kanai, S. Karthikeyan, F. Koskeridis, L. A. Lange, T. Lehtimäki, M. M. Lerch, A. Linneberg, Y. Liu, L.-P. Lyttikäinen, A. Manichaikul, H. C. Martin, K. Matsuda, K. L. Mohlke, N. Mononen, Y. Murakami, G. N. Nadkarni, M. Nauck, K. Nikus, W. H. Ouwehand, N. Pankratz, O. Pedersen, M. Preuss, B. M. Psaty, O. T. Raitakari, D. J. Roberts, S. S. Rich, B. A. T. Rodriguez, J. D. Rosen, J. I. Rotter, P. Schubert, C. N. Spracklen, P. Surendran, H. Tang, J.-C. Tardif, R. C. Trembath, M. Ghanbari, U. Völker, H. Völzke, N. A. Watkins, A. B. Zonderman, P. W. F. Wilson, Y. Li, A. S. Butterworth, J.-F. Gauchat, C. W. K. Chiang, B. Li, R. J. F. Loos, W. J. Astle, E. Evangelou, D. A. van Heel, V. G. Sankaran, Y. Okada, N. Soranzo, A. D. Johnson, A. P. Reiner, P. L. Auer, G. Lettre, VA Million Veteran Program, 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](https://doi.org/10.1016/j.cell.2020.06.045) [Medline](#)
51. B. P. Fairfax, S. Makino, J. Radhakrishnan, K. Plant, S. Leslie, A. Dilthey, P. Ellis, C. Langford, F. O. Vannberg, J. C. Knight, 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](https://doi.org/10.1038/ng.2205) [Medline](#)
52. P. Waterhouse, J. M. Penninger, E. Timms, A. Wakeham, A. Shahinian, K. P. Lee, C. B. Thompson, H. Griesser, T. W. Mak, Lymphoproliferative disorders with early lethality in mice deficient in Ctla-4. *Science* **270**, 985–988 (1995).
[doi:10.1126/science.270.5238.985](https://doi.org/10.1126/science.270.5238.985) [Medline](#)
53. E. A. Tivol, F. Borriello, A. N. Schweitzer, W. P. Lynch, J. A. Bluestone, A. H. Sharpe, 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](https://doi.org/10.1016/1074-7613(95)90125-6) [Medline](#)
54. D. Schubert, C. Bode, R. Kenefech, T. Z. Hou, J. B. Wing, A. Kennedy, A. Bulashevska, B.-S. Petersen, A. A. Schäffer, B. A. Grüning, S. Unger, N. Frede, U. Baumann, T. Witte, R. E. Schmidt, G. Dueckers, T. Niehues, S. Seneviratne, M. Kanariou, C. Speckmann, S. Ehl, A. Rensing-Ehl, K. Warnatz, M. Rakhmanov, R. Thimme, P. Hasselblatt, F. Emmerich, T. Cathomen, R. Backofen, P. Fisch, M. Seidl, A. May, A. Schmitt-Graeff, S. Ikemizu, U. Salzer, A. Franke, S. Sakaguchi, L. S. K. Walker, D. M. Sansom, B. Grimbacher, Autosomal dominant immune dysregulation syndrome in humans with *CTLA4* mutations. *Nat. Med.* **20**, 1410–1416 (2014).
[doi:10.1038/nm.3746](https://doi.org/10.1038/nm.3746) [Medline](#)
55. H. S. Kuehn, W. Ouyang, B. Lo, E. K. Deenick, J. E. Niemela, D. T. Avery, J.-N. Schickel, D. Q. Tran, J. Stoddard, Y. Zhang, D. M. Frucht, B. Dumitriu, P.

- Scheinberg, L. R. Folio, C. A. Frein, S. Price, C. Koh, T. Heller, C. M. Seroogy, A. Huttenlocher, V. K. Rao, H. C. Su, D. Kleiner, L. D. Notarangelo, Y. Rampertaap, K. N. Olivier, J. McElwee, J. Hughes, S. Pittaluga, J. B. Oliveira, E. Meffre, T. A. Fleisher, S. M. Holland, M. J. Lenardo, S. G. Tangye, G. Uzel, Immune dysregulation in human subjects with heterozygous germline mutations in *CTLA4*. *Science* **345**, 1623–1627 (2014). [doi:10.1126/science.1255904](https://doi.org/10.1126/science.1255904) [Medline](#)
56. Y.-F. Wang, Y. Zhang, Z. Lin, H. Zhang, T.-Y. Wang, Y. Cao, D. L. Morris, Y. Sheng, X. Yin, S.-L. Zhong, X. Gu, Y. Lei, J. He, Q. Wu, J. J. Shen, J. Yang, T.-H. Lam, J.-H. Lin, Z.-M. Mai, M. Guo, Y. Tang, Y. Chen, Q. Song, B. Ban, C. C. Mok, Y. Cui, L. Lu, N. Shen, P. C. Sham, C. S. Lau, D. K. Smith, T. J. Vyse, X. Zhang, Y. L. Lau, W. Yang, 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) [Medline](#)
57. J. D. Cooper, D. J. Smyth, A. M. Smiles, V. Plagnol, N. M. Walker, J. E. Allen, K. Downes, J. C. Barrett, B. C. Healy, J. C. Mychaleckyj, J. H. Warram, J. A. Todd, 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) [Medline](#)
58. K. Kim, S.-Y. Bang, H.-S. Lee, S.-K. Cho, C.-B. Choi, Y.-K. Sung, T.-H. Kim, J.-B. Jun, D. H. Yoo, Y. M. Kang, S.-K. Kim, C.-H. Suh, S.-C. Shim, S.-S. Lee, J. Lee, W. T. Chung, J.-Y. Choe, H. D. Shin, J.-Y. Lee, B.-G. Han, S. K. Nath, S. Eyre, J. Bowes, D. A. Pappas, J. M. Kremer, M. A. Gonzalez-Gay, L. Rodriguez-Rodriguez, L. Ärlestig, Y. Okada, D. Diogo, K. P. Liao, E. W. Karlson, S. Raychaudhuri, S. Rantapää-Dahlqvist, J. Martin, L. Klareskog, L. Padyukov, P. K. Gregersen, J. Worthington, J. D. Greenberg, R. M. Plenge, S.-C. Bae, 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) [Medline](#)
59. A. E. Renton, H. A. Pliner, C. Provenzano, A. Evoli, R. Ricciardi, M. A. Nalls, G. Marangi, Y. Abramzon, S. Arepalli, S. Chong, D. G. Hernandez, J. O. Johnson, E. Bartoccioni, F. Scuderi, M. Maestri, J. R. Gibbs, E. Errichiello, A. Chiò, G. Restagno, M. Sabatelli, M. Macek, S. W. Scholz, A. Corse, V. Chaudhry, M. Benatar, R. J. Barohn, A. McVey, M. Pasnoor, M. M. Dimachkie, J. Rowin, J. Kissel, M. Freimer, H. J. Kaminski, D. B. Sanders, B. Lipscomb, J. M. Massey, M. Chopra, J. F. Howard Jr., W. J. Koopman, M. W. Nicolle, R. M. Pascuzzi, A. Pestronk, C. Wulf, J. Florence, D. Blackmore, A. Soloway, Z. Siddiqi, S. Muppudi, G. Wolfe, D. Richman, M. M. Mezei, T. Jiwa, J. Oger, D. B. Drachman, B. J. Traynor, 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) [Medline](#)
60. R. Roychoudhuri, K. Hirahara, K. Mousavi, D. Clever, M. Bonelli, C. Klebanoff, V. Sartorelli, Y. Kanno, L. Gattinoni, E. Wang, H. Liu, F. Marincola, I. Kazuhiko, J. O'Shea, N. P. Restifo, 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, E. Porcu, G. Pistis, A. Teumer, S. J. Brown, R. A. Jensen, R. Rawal, G. L. Roef, T. S. Plantinga, S. H. Vermeulen, J. Lahti, M. J. Simmonds, L. L. Husemoen, R. M. Freathy, B. M. Shields, D. Pietzner, R. Nagy, L. Broer, L. Chaker, T. I. Korevaar, M. G. Plia, C. Sala, U. Völker, J. B. Richards, F. C. Sweep, C. Gieger, T. Corre, E.

Kajantie, B. Thuesen, Y. E. Taes, W. E. Visser, A. T. Hattersley, J. Kratzsch, A. Hamilton, W. Li, G. Homuth, M. Lobina, S. Mariotti, N. Soranzo, M. Cocca, M. Nauck, C. Spielhagen, A. Ross, A. Arnold, M. van de Bunt, S. Liyanarachchi, M. Heier, H. J. Grabe, C. Masciullo, T. E. Galesloot, E. M. Lim, E. Reischl, P. J. Leedman, S. Lai, A. Delitala, A. P. Bremner, D. I. Philips, J. P. Beilby, A. Mulas, M. Vocale, G. Abecasis, T. Forseen, A. James, E. Widen, J. Hui, H. Prokisch, E. E. Rietzschel, A. Palotie, P. Feddema, S. J. Fletcher, K. Schramm, J. I. Rotter, A. Klutigg, D. Radke, M. Traglia, G. L. Surdulescu, H. He, J. A. Franklyn, D. Tiller, B. Vaidya, T. de Meyer, T. Jørgensen, J. G. Eriksson, P. C. O'Leary, E. Wichmann, A. R. Hermus, B. M. Psaty, T. Ittermann, A. Hofman, E. Bosi, D. Schlessinger, H. Wallaschofski, N. Pirastu, Y. S. Aulchenko, A. de la Chapelle, R. T. Netea-Maier, S. C. Gough, H. Meyer Zu Schwabedissen, T. M. Frayling, J. M. Kaufman, A. Linneberg, K. Räikkönen, J. W. Smit, L. A. Kiemeney, F. Rivadeneira, A. G. Uitterlinden, J. P. Walsh, C. Meisinger, M. den Heijer, T. J. Visser, T. D. Spector, S. G. Wilson, H. Völzke, A. Cappola, D. Toniolo, S. Sanna, S. Naitza, R. P. Peeters, 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) [Medline](#)

62. C. J. Lessard, H. Li, I. Adrianto, J. A. Ice, A. Rasmussen, K. M. Grundahl, J. A. Kelly, M. G. Dozmorov, C. Miceli-Richard, S. Bowman, S. Lester, P. Eriksson, M.-L. Eloranta, J. G. Brun, L. G. Gøransson, E. Harboe, J. M. Guthridge, K. M. Kaufman, M. Kvarnström, H. Jazebi, D. S. Cunningham-Graham, M. E. Grandits, A. N. M. Nazmul-Hossain, K. Patel, A. J. Adler, J. S. Maier-Moore, A. D. Farris, M. T. Brennan, J. A. Lessard, J. Chodosh, R. Gopalakrishnan, K. S. Hefner, G. D. Houston, A. J. W. Huang, P. J. Hughes, D. M. Lewis, L. Radfar, M. D. Rohrer, D. U. Stone, J. D. Wren, T. J. Vyse, P. M. Gaffney, J. A. James, R. Omdal, M. Wahren-Herlenius, G. G. Illei, T. Witte, R. Jonsson, M. Rischmueller, L. Rönnblom, G. Nordmark, W.-F. Ng, X. Mariette, J.-M. Anaya, N. L. Rhodus, B. M. Segal, R. H. Scofield, C. G. Montgomery, J. B. Harley, K. L. Sivils, UK Primary Sjögren's Syndrome Registry, 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) [Medline](#)
63. A. Márquez, M. Kerick, A. Zhernakova, J. Gutierrez-Achury, W.-M. Chen, S. Onengut-Gumuscu, I. González-Álvaro, L. Rodriguez-Rodriguez, R. Rios-Fernández, M. A. González-Gay, M. D. Mayes, S. Raychaudhuri, S. S. Rich, C. Wijmenga, J. Martín, Coeliac Disease Immunochip Consortium, Rheumatoid Arthritis Consortium International for Immunochip (RACI), International Scleroderma Group, Type 1 Diabetes Genetics Consortium, 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) [Medline](#)
64. K. R. Simpfendorfer, B. E. Armstead, A. Shih, W. Li, M. Curran, N. Manjarrez-Orduño, A. T. Lee, B. Diamond, P. K. Gregersen, 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) [Medline](#)
65. A. V. Korenykh, P. F. Egea, A. A. Korostelev, J. Finer-Moore, C. Zhang, K. M. Shokat, R. M. Stroud, P. Walter, 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) [Medline](#)

66. K. P. K. Lee, M. Dey, D. Neculai, C. Cao, T. E. Dever, F. Sicheri, 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) [Medline](#)
67. T. Becker, S. Bhushan, A. Jarasch, J.-P. Armache, S. Funes, F. Jossinet, J. Gumbart, T. Mielke, O. Berninghausen, K. Schulten, E. Westhof, R. Gilmore, E. C. Mandon, R. Beckmann, 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) [Medline](#)
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) [Medline](#)
69. Z. Zhu, Y. Guo, H. Shi, C.-L. Liu, R. A. Panganiban, W. Chung, L. J. O'Connor, B. E. Himes, S. Gazal, K. Hasegawa, C. A. Camargo Jr., L. Qi, M. F. Moffatt, F. B. Hu, Q. Lu, W. O. C. Cookson, L. Liang, 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) [Medline](#)
70. S. W. Jang, S. S. Hwang, H. S. Kim, M. K. Kim, W. H. Lee, S. U. Hwang, J. Gwak, S. K. Yew, R. A. Flavell, G. R. Lee, 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) [Medline](#)
71. T. C. Tu, N. K. Brown, T.-J. Kim, J. Wroblewska, X. Yang, X. Guo, S. H. Lee, V. Kumar, K.-M. Lee, Y.-X. Fu, CD160 is essential for NK-mediated IFN- γ production. *J. Exp. Med.* **212**, 415–429 (2015). [doi:10.1084/jem.20131601](https://doi.org/10.1084/jem.20131601) [Medline](#)
72. M. Korsgren, C. G. Persson, F. Sundler, T. Bjerke, T. Hansson, B. J. Chambers, S. Hong, L. Van Kaer, H. G. Ljunggren, O. Korsgren, 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) [Medline](#)
73. S. Matsubara, K. Takeda, T. Kodama, A. Joetham, N. Miyahara, T. Koya, C. H. Swasey, M. Okamoto, A. Dakhama, E. W. Gelfand, IL-2 and IL-18 attenuation of airway hyperresponsiveness requires STAT4, IFN- γ , and natural killer cells. *Am. J. Respir. Cell Mol. Biol.* **36**, 324–332 (2007). [doi:10.1165/rcmb.2006-0231OC](https://doi.org/10.1165/rcmb.2006-0231OC) [Medline](#)
74. C. Mitchell, K. Provost, N. Niu, R. Homer, L. Cohn, IFN- γ 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) [Medline](#)
75. Z. Zhu, F. Zhang, H. Hu, A. Bakshi, M. R. Robinson, J. E. Powell, G. W. Montgomery, M. E. Goddard, N. R. Wray, P. M. Visscher, J. Yang, 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) [Medline](#)
76. 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). [doi:10.1093/bioinformatics/bts635](https://doi.org/10.1093/bioinformatics/bts635) [Medline](#)
77. H. M. Kang, M. Subramaniam, S. Targ, M. Nguyen, L. Maliskova, E. McCarthy, E. Wan, S. Wong, L. Byrnes, C. M. Lanata, R. E. Gate, S. Mostafavi, A. Marson, N. Zaitlen, L. A. Criswell, C. J. Ye, 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)

Medline

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) [Medline](#)
79. T. Stuart, A. Butler, P. Hoffman, C. Hafemeister, E. Papalexi, W. M. Mauck III, Y. Hao, M. Stoeckius, P. Smibert, R. Satija, 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) [Medline](#)
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) [Medline](#)
81. K. R. Moon, D. van Dijk, Z. Wang, S. Gigante, D. B. Burkhardt, W. S. Chen, K. Yim, A. van den Elzen, M. J. Hirn, R. R. Coifman, N. B. Ivanova, G. Wolf, S. Krishnaswamy, 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) [Medline](#)
82. K. Street, D. Risso, R. B. Fletcher, D. Das, J. Ngai, N. Yosef, E. Purdom, S. Dudoit, 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) [Medline](#)
83. A. Auton, L. D. Brooks, R. M. Durbin, E. P. Garrison, H. M. Kang, J. O. Korbel, J. L. Marchini, S. McCarthy, G. A. McVean, G. R. Abecasis, 1000 Genomes Project Consortium, A global reference for human genetic variation. *Nature* **526**, 68–74 (2015). [doi:10.1038/nature15393](https://doi.org/10.1038/nature15393) [Medline](#)
84. S. Yazar, J. A. Hernandez, Population-scale single-cell eQTL mapping identifies cell type specific genetic control of autoimmune disease. Zenodo (2021); <https://doi.org/10.5281/zenodo.5084395>.
85. C. C. Chang, C. C. Chow, L. C. Tellier, S. Vattikuti, S. M. Purcell, J. J. Lee, Second-generation PLINK: Rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015). [doi:10.1186/s13742-015-0047-8](https://doi.org/10.1186/s13742-015-0047-8) [Medline](#)
86. J. Yang, S. H. Lee, M. E. Goddard, P. M. Visscher, GCTA: A tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011). [doi:10.1016/j.ajhg.2010.11.011](https://doi.org/10.1016/j.ajhg.2010.11.011) [Medline](#)
87. A. L. Price, N. J. Patterson, R. M. Plenge, M. E. Weinblatt, N. A. Shadick, D. Reich, Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904–909 (2006). [doi:10.1038/ng1847](https://doi.org/10.1038/ng1847) [Medline](#)
88. P.-R. Loh, P. Danecek, P. F. Palamara, C. Fuchsberger, Y. A. Reshef, H. K. Finucane, S. Schoenherr, L. Forer, S. McCarthy, G. R. Abecasis, R. Durbin, A. L Price, Reference-based phasing using the Haplotype Reference Consortium panel. *Nat. Genet.* **48**, 1443–1448 (2016). [doi:10.1038/ng.3679](https://doi.org/10.1038/ng.3679) [Medline](#)
89. R Core Team, R: A language and environment for statistical computing (2017); <https://www.r-project.org/>.
90. R. A. Peterson, bestNormalize: Normalizing Transformation Functions (2019); <https://cran.r-project.org/web/packages/bestNormalize/index.html>.
91. V. D. Blondel, J.-L. Guillaume, R. Lambiotte, E. Lefebvre, Fast unfolding of communities in large networks. *J. Stat. Mech.* **2008**, P10008 (2008). [doi:10.1088/1742-5468/2008/10/P10008](https://doi.org/10.1088/1742-5468/2008/10/P10008)

92. H. Yoshitomi, S. Kobayashi, A. Miyagawa-Hayashino, A. Okahata, K. Doi, K. Nishitani, K. Murata, H. Ito, T. Tsuruyama, H. Haga, S. Matsuda, J. Toguchida, Human Sox4 facilitates the development of CXCL13-producing helper T cells in inflammatory environments. *Nat. Commun.* **9**, 3762 (2018). [doi:10.1038/s41467-018-06187-0](https://doi.org/10.1038/s41467-018-06187-0) [Medline](#)
93. J. Steiner, N. Marquardt, I. Pauls, K. Schiltz, H. Rahmoune, S. Bahn, B. Bogerts, R. E. Schmidt, R. Jacobs, Human CD8⁺ T cells and NK cells express and secrete S100B upon stimulation. *Brain Behav. Immun.* **25**, 1233–1241 (2011). [doi:10.1016/j.bbi.2011.03.015](https://doi.org/10.1016/j.bbi.2011.03.015) [Medline](#)
94. J. P. Böttcher, E. Bonavita, P. Chakravarty, H. Blees, M. Cabeza-Cabrero, S. Sammicheli, N. C. Rogers, E. Sahai, S. Zelenay, C. Reis e Sousa, NK cells stimulate recruitment of cDC1 into the tumor microenvironment promoting cancer immune control. *Cell* **172**, 1022–1037.e14 (2018). [doi:10.1016/j.cell.2018.01.004](https://doi.org/10.1016/j.cell.2018.01.004) [Medline](#)
95. A. Crinier, P. Milpied, B. Escalière, C. Piperoglou, J. Galluso, A. Balsamo, L. Spinelli, I. Cervera-Marzal, M. Ebbo, M. Girard-Madoux, S. Jaeger, E. Bollon, S. Hamed, J. Hardwicke, S. Ugolini, F. Vély, E. Narni-Mancinelli, E. Vivier, High-Dimensional single-cell analysis identifies organ-specific signatures and conserved NK cell subsets in humans and mice. *Immunity* **49**, 971–986.e5 (2018). [doi:10.1016/j.jimmuni.2018.09.009](https://doi.org/10.1016/j.jimmuni.2018.09.009) [Medline](#)
96. J. D. Storey, A. J. Bass, A. Dabney, D. Robinson, qvalue: Q-value estimation for false discovery rate control (R package version 2.22.0) (2020); <https://github.com/jdstorey/qvalue>.
97. T. Stuart, A. Srivastava, C. Lareau, R. Satija, Multimodal single-cell chromatin analysis with Signac. bioRxiv 2020.11.09.373613 (2020) [Preprint]; <https://doi.org/10.1101/2020.11.09.373613>.
98. Y. Zhang, T. Liu, C. A. Meyer, J. Eeckhoute, D. S. Johnson, B. E. Bernstein, C. Nusbaum, R. M. Myers, M. Brown, W. Li, X. S. Liu, Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008). [doi:10.1186/gb-2008-9-9-r137](https://doi.org/10.1186/gb-2008-9-9-r137) [Medline](#)
99. ThermoFisher Scientific, “Axiom genotyping solution data analysis: User guide” (Publication MAN0018363, 2020); https://assets.thermofisher.com/TFS-Assets/LSG/manuals/axiom_genotyping_solution_analysis_guide.pdf.
100. W. E. Johnson, C. Li, A. Rabinovic, Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8**, 118–127 (2007). [doi:10.1093/biostatistics/kxj037](https://doi.org/10.1093/biostatistics/kxj037) [Medline](#)
101. 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 (2010). [doi:10.1093/bioinformatics/btp616](https://doi.org/10.1093/bioinformatics/btp616) [Medline](#)
102. B. J. Strober, R. Elorban, K. Rhodes, N. Krishnan, K. Tayeb, A. Battle, Y. Gilad, Dynamic genetic regulation of gene expression during cellular differentiation. *Science* **364**, 1287–1290 (2019). [doi:10.1126/science.aaw0040](https://doi.org/10.1126/science.aaw0040) [Medline](#)
103. A. Nathan, S. Asgari, K. Ishigaki, T. Amariuta, Y. Luo, Modeling memory T cell states at single-cell resolution identifies in vivo state-dependence of eQTLs influencing disease. bioRxiv 2021.07.29.454316 [Preprint] (2021); <https://doi.org/10.1101/2021.07.29.454316>.
104. D. Bates, M. Mächler, B. Bolker, S. Walker, Fitting linear mixed-effects models using

- lme4. *J. Stat. Softw.* **67**, 1–48 (2015). [doi:10.18637/jss.v067.i01](https://doi.org/10.18637/jss.v067.i01)
105. Z. Gu, R. Eils, M. Schlesner, Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**, 2847–2849 (2016). [doi:10.1093/bioinformatics/btw313](https://doi.org/10.1093/bioinformatics/btw313) [Medline](#)
106. T. A. Myers, S. J. Chanock, M. J. Machiela, LDlinkR: An R package for rapidly calculating linkage disequilibrium statistics in diverse populations. *Front. Genet.* **11**, 157 (2020). [doi:10.3389/fgene.2020.00157](https://doi.org/10.3389/fgene.2020.00157) [Medline](#)
107. A. A. Shabalin, Matrix eQTL: Ultra fast eQTL analysis via large matrix operations. *Bioinformatics* **28**, 1353–1358 (2012). [doi:10.1093/bioinformatics/bts163](https://doi.org/10.1093/bioinformatics/bts163) [Medline](#)
108. G. Wang, A. Sarkar, P. Carbonetto, M. Stephens, A simple new approach to variable selection in regression, with application to genetic fine mapping. *J. R. Stat. Soc. Series B Stat. Methodol.* **82**, 1273–1300 (2020). [doi:10.1111/rssb.12388](https://doi.org/10.1111/rssb.12388)