Science
AAAS

Supplementary Materials for

Single-cell RNA-seq reveals cell type–specific molecular and genetic associations to lupus

Richard K. Perez *et al*.

Corresponding authors: Noah Zaitlen, nzaitlen@g.ucla.edu; Chun Jimmie Ye, jimmie.ye@ucsf.edu

Science **376**, eabf1970 (2022)
DOI: 10.1126/science.abf1970

The PDF file includes:

Other Supplementary Material for this manuscript includes the following:

Materials and Methods

Sample preparation

Informed consent was obtained from all patients sequenced in this study. 162 SLE cases were collected from the California Lupus Epidemiological Study (CLUES) cohort, including 19 samples from SLE cases in an active disease flare and 10 matched samples post-flare treatment, and 49 matching healthy controls were collected from the UCSF Rheumatology Clinic. These studies were approved by the Institutional Review Board of the University of California, San Francisco (17-22228). Peripheral blood mononuclear cells were isolated from patient donors, Ficoll separated, and cryopreserved by the UCSF Core Immunologic Laboratory (CIL). Frozen peripheral blood mononuclear cells from 46 healthy controls were obtained from consented donors and processed using published protocols from the Immune Variation project (ImmVar).

PBMC processing, RNA and antibody-derived tags (ADT) library preparation, and sequencing

Frozen PBMCs were profiled in 23 pools across four processing batches for a total of 355 samples including 94 replicates and 10 longitudinal samples for flare patients. PBMCs were thawed, counted using a Vi-CELL XR (Beckman Coulter), resuspended, and pooled (*8).* For processing batches 3 and 4, pooled cells were resuspended in 100 mL of Cell Staining Buffer (BioLegend 420201) with 5% Human TruStain FcX (BioLegend 422301) to block non-specific staining. Cells were incubated in this solution on ice for 10 minutes. Immunostaining antibodies (AbSeq, Becton Dickinson) were pooled at 2 mL per antibody, the pooled antibody was then concentrated using an Amicon Ultra 0.5 Centrifugal Filter (Millipore UFC503096) and resuspended in 100 mL volume of Cell Staining Buffer. The pool was added to the cell solution after blocking. Cells were stained for 45 minutes on ice and were then washed 3 times in Cell Staining Buffer. The cells were resuspended in DPBS with 0.04% BSA and strained using a 40-micron Flowmi strainer (Sigma BAH136800040). Stained (processing batches 3 and 4) and unstained (processing batches 1 and 2) cells were loaded onto a 10X Chromium Controller. Following library preparation using the published 10X Chromium Single Cell 3’ v2 Solution protocol, libraries were sequenced on the HiSeq4000 or NovaSeq6000 at a depth of 6,306-29,862 reads/cell using the recommended cycle numbers Read 1 26 cycles, i7 index 8 cycles, Read 2 98 cycles.

Single-cell RNA-seq data preprocessing

A total of 1,870,857 cell-containing droplets were obtained after sequencing. Freemuxlet (https://github.com/statgen/popscle/), an extension of demuxlet (*8)*, was used with an error probability of 0.1 to assign each cell to a donor of origin, preserving a total of 1,444,450 singlets (doublet rate 22.12%, expected 22-25%). Additional removal of doublets with the same genotype was performed using Scrublet (13) (67,969 droplets, doublet rate 4.71%; doublet threshold 0.15). Contaminated platelets were eliminated 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 louvain community detection to identify a population with very high expression of RBC markers like hemoglobin genes. In total, 1,263,676 cells remained in the final dataset after platelet and RBC removal (112,805 cells). Using Scanpy version 1.6 (*58*), the data was preprocessed by first adjusting for pool using COMBAT (*63*), followed by regressing total nUMIs and percentage mitochondrial UMIs. After cell filtering and expression normalization using default parameters, we performed k-nearest neighbor (KNN) graph construction, louvain, and projected the data onto two UMAP dimensions. L3 cohort refers to the cohort of cases and controls that are age matched and equal in number of cases of Asian and European ancestry in processing batch 4 and their replicates in other batches. Processing batch 4 refers to the L3 cohort samples within processing batch 4 only.

Cell type annotation, proportion, and abundance calculations

Scanpy version 1.6 (*58*) was used to cluster singlets into louvain communities with parameter settings of resolution of 3 and fixing the random state to 0. Each louvain community was assigned one of 11 cell types by comparing differentially expressed genes between communities and the most abundantly expressed genes for each community with known markers for each cell type. Cell type was confirmed using CITE-Seq dataset (**Fig. S8**). The Prolif population was broken into its component populations in the lymphocyte analysis based on high resolution clustering and gene expression of known markers for each cell type. The percentage of each cell type was calculated as the number of cells belonging to the cell type divided by the total number of cells assigned to the sample or individual. ImmVar samples were excluded from all case-control comparisons. Two samples with less than 100 cells total were excluded from the analyses. Due to computational constraints, we computed RNA velocity only on processing batch 4 myeloid samples and for consistency with the rest of myeloid analysis, myeloid lineage subpopulation composition estimates were also computed using only processing batch 4 samples. Absolute cell type abundances for SLE cases were calculated by multiplying the cell-type proportions by the total number of white blood cells per

volume for each patient reported in the electronic health record from the Complete Blood Count. For cell type percentage comparisons between cases and controls, statistical significance and effect size was computed using statsmodels python package (v0.12.0) weighted least squares models with the weights as cell counts, percentage as the endogenous variable and disease status [0,1] as the exogenous variable plus a constant. P values were adjusted for multiple comparisons using Bonferroni correction. Meta p values were calculated using Fisher's method (Fisher's combined probability test).

Electronic health record query

SLE cases with available monocyte and lymphocyte counts were selected according to the following criteria: 4,532 healthy female controls were selected according to previous work conducted with the UCSF EHR database (*64*). Adult outpatients, ages 20-90 without abnormal findings, were extracted from the EHR system at the University of California, San Francisco (UCSF) Medical Center on February 2018, spanning approximately 6 years of medical service. In patients with multiple healthy encounters, a single encounter was chosen at random. 403 SLE Cases were defined as patients ages 20-90 who have a diagnosis of an ICD-10 code M32.* appearing at least twice, 30 days apart. Lab test results for cases were taken from encounters where SLE was assigned as a primary diagnosis or principal problem diagnosis. Patients with a monocyte count less than 5 and a lymphocyte count less than 6 were excluded.

Mendelian randomization

Mendelian randomization was performed using the GSMR (Generalized Summary-data-based Mendelian Randomization) package in gcta_1.91.5beta (*16*). We applied GSMR to blood count quantitative trait loci (QTLs) in the UK Biobank (lymphocytes, monocytes, red blood cells, white blood cells, and platelets) and variants associated with SLE disease status (*4*). The 1000 Genomes Phase3 was used as the reference, a GWAS significance threshold of $5x10^{-18}$, heidi outlier threshold of 0.15, and a linkage disequilibrium threshold of 0.01 were used as parameters. Varying the GWAS significance threshold did not change our observations.

Cell-type-specific differential expression analysis

For each cell type and each individual, a pseudobulk expression profile was computed by summing all of the counts for each gene and dividing it by the total number of counts across all genes. Dividing by total counts simultaneously controlled for variability in the number of cells per individual and variability in transcript capture efficiency across pools and batches. The EdgeR package (*59*) was used to estimate the

log2 fold change and the p-value of gene expression differences between SLE and healthy donors with processing batch and age included as a covariate. Samples of European and Asian ancestries were analyzed separately. P values were adjusted for multiple comparisons for each cell type. In total, we define differentially expressed (DE) genes as |logFC| > 0.5 & $P_{adjusted}$< 0.05 in at least one cell type and ancestry. We identified 127 differentially expressed genes in total PBMC pseudobulk (|logFC| > 0.5; $P_{adjusted}$< 0.05) but not differentially expressed in any of the 11 cell types (|logFC| < 0.01; $P_{adjusted}$>0.5).

Tcyto$_{GZMH}$ signature scores

We phenotyped Tcyto$_{GZMH}$ using sets of genes to define a cytotoxic signature (*PRF1*, *GZMH*, *GZMB*), an exhaustion signature (*PDCD1*, *CTLA4*, *LAG3*, *CD160*, *TIGIT*, *HAVCR2*, *CD244*), and a type-1 ISG signature (*ISG15*, *IFI6*, *IFI44L*, *IFI44*, *RSAD2*, *CXCL10*, *IFIT2*, *IFIT3*, *IFIT1*, *IFITM3*, *OAS1*, *OAS3*, *OAS2*, *OASL*, *EPSTI1*, *RNASE1*, *RNASE2*, *IFI27*, *XAF1*, *LGALS3BP*, *SIGLEC1*, *USP18*, *APOBEC3A*, *APOBEC3B*, *MX1*). Signature scores for each cell was calculated using Scanpy's score_genes function. For TCR enrichment analysis, the highest signature score was used to annotate each cell dominant signal.

Module scores

The module score for each individual represents the mean pseudobulk expression of all module specific genes for that module. Complete gene module gene lists can be found in **Table S3**.

Tcyto$_{GZMH}$ co-expression analysis

We computed a co-expression matrix of the top 300 differentially expressed genes between L3 cases and controls for the Tcyto$_{GZMH}$ population using Scanpy's rank_genes_groups function. Then using all samples, sklearn's spectral clustering algorithm was used to perform biclustering of genes with respect to similarities in their Spearman correlations to other genes and the values were then column z scored for visualization. After recovering the gene ordering at the single-cell level, we computed an analogous co-expression matrix using the pseudobulk gene expression profile for each sample and plotted both the single-cell (lower triangle) and pseudobulk (upper triangle) co-expression patterns as a heatmap to demonstrate differences in co-expression patterns inferred by the two approaches. Figure 2E was cropped to highlight the Type-1 interferon and cytotoxic genes. Co-expression correlation between cytotoxic and Type-1 interferon scores was calculated by taking the average expression of cytotoxic genes and Type-1

interferon genes (See Tcyto$_{GZMH}$ signature scores section for details) for each cell and computed the Pearson correlation within each sample/individual. The average Spearman R across all samples was reported (Spearman R = 0.16). Computing the correlation based off a random sampling of genes of appropriate sample size returned Spearman R = 0.02.

Normalized Shannon's entropy to measure TCR repertoire diversity

Shannon's entropy is an estimation of the average amount of information stored. In the context of TCR expansion, a lower value represents more TCR expansion. To account for variability in the number of cells per sample, Shannon's entropy was normalized by log(number of cells). The number of clones of size x is v. We controlled for age by only using cases and controls between the ages of 20 and 40.

$$p = \frac{x}{\sum(x * v)}$$

$$H = \sum(v * p * \log(p)) \Big/ log2(x * v)$$

Expression modules in SLE

Expression modules were recovered by first computing cell-type-specific pseudobulk expression profiles for each of 11 cell types or all cell types together (PBMC). Only genes that are differentially expressed in at least one cell type or PBMCs were considered. For each cell type, cases and controls were hierarchically clustered separately, then concatenated together, and gene expression for each DE gene was standardized per gene. The 11 cell-type-specific matrices and the PBMC matrix was column concatenated for visualization and clustering. Hierarchical clustering using Seaborn's Clustermap (v0.11.0 https://seaborn.pydata.org/; "complete" / Nearest Point Algorithm) was used to cluster the genes. SciPy's ((*65*); https://www.scipy.org/) fcluster was utilized to recover clusters based on the hierarchical clustering branch points with a cluster number of 15. The resulting clusters were further combined into six modules based on the expression patterns within each cell type and case and control status. Genes clusters which did not show any pattern (cell type specific or pan) were not included in modules. Genes from one cluster which showed cell type expression in both T and B cells, were added to both Up-T and Up-B modules. Module gene expression was calculated for each module gene based on the module's contributing cell types (e.g. a Mye$_{up}$ module gene value is the pseudobulk expression based only cM and ncM cells).

Pathway enrichment analysis of expression modules

Genes sets corresponding to each of six modules were submitted to ToppGene (https://toppgene.cchmc.org/enrichment.jsp) and only significantly (FDR < 0.05) enriched pathways ("Pathway" option) were retained.

Molecular clusters

Using cell-type-specific gene expression in cases from processing batch 4 (same batch used for expression modules and differential expression analysis) as features, we performed principal component analysis (PCA). Gene correlations to each principal component can be found in *Fig4_PCA_correlations.csv*. Two discernable cohorts emerged (one which resided with controls and one that did not), so we performed K-means clustering with k=2 to assign molecular clusters. We projected and assigned molecular sub-phenotypes of out-of-samples flare cases and test cases using the already trained PCA and K-means clustering models. Case replicates already found in processing batch 4, were excluded from test set.

RNA velocity for identifying the activation axis

To analyze the expression dynamics of genes that are differentially expressed between SLE and healthy donors, we first identified candidate genes by taking the genes with >1 magnitude of the log-fold change across SLE and healthy donors in each cell type. We then computed the RNA velocities using the scVelo package implemented in Python ((*24*); https://scvelo.readthedocs.io/) by using the 'dynamical' mode of the velocity estimation and default parameters. For the cM cluster, we computed the activation pseudotime axis by computing the mean velocity vector in the all-gene space and projected each cell onto the mean velocity vector. Due to computational constraints, we computed RNA velocity only on processing batch 4 myeloid cells and for consistency and comparison within the myeloid analysis, myeloid lineage subpopulation percentages were also computed using only processing batch 4 data.

Logistic Regression Predictions

We used sklearn's Logistic Regression function for all models (https://scikit-learn.org/stable/). We used elasticnet penalty with saga solver and L1/L2 ratio penalty of 0.5. Receiver Operating Curve area under the curve (AUC) was recovered using sklearn's roc_curve and roc_auc_score functions. Input features were module specific gene expression calculated with appropriate pseudobulk (e.g. B_{Up} module gene expression calculated only on B cell pseudobulk). Any scaling was computed on the train set and the test set was transformed into that space. For case/control predictions, we trained the model on samples from processing batch 4 and tested the model on samples from

processing batches 1, 2, and 3 (excluding case replicates from batch 4 found in other batches). For predicting individual clinical features, we trained a model on cases from processing batch 4 and tested the model on cases from processing batches 1, 2, and 3 (excluding case replicates from batch 4 found in other batches). Mean AUC was computed by 3-fold-cross-validation.

TCR Sequencing and Analysis

Approximately 10% of the barcoded cDNA from the 10X workflow was utilized for TCR analysis. The cDNA was further amplified with Template switch oligo (TSO) primer and the Truseq Read 1 primer with KAPA HiFi hotstart readymix (Kapa Biosystems, Thermo Fisher Scientific) for 12 cycles at the conditions 98°C for 15 seconds, 67°C for 20 seconds and 72°C for 1 minute. A pool of forward Valpha and Vbeta primers containing the Illumina read 2 primer sequence were used in conjunction with the Truseq Read 1 primer to amplify CDR3 sequences from the TCR alpha and beta loci with NEBNext Ultra II Q5 Master Mix (New England Biolabs Inc.) for 25 cycles at the conditions 98°C for 10 seconds, 68°C for 30 seconds and 72°C for 1 minute. An additional amplification step using primers containing the Illumina P5 adapter, i5 index sequences and Truseq Read 1 primer sequence with the Illumina P7 adapter, i7 index sequences and Truseq Read 2 with NEBNext Ultra II Q5 Master Mix was carried out for 8 cycles at the conditions 98°C for 15 seconds and 72°C for 1 minute and 20 seconds to create final TCR libraries for sequencing. High-throughput sequencing was done on an Illumina Hiseq 2500 Rapid run with separate lanes for the TCR alpha and TCR beta sequencing. Read 1 sequenced 26 bp containing the 10x barcode and UMIs, and Read 2 sequenced 251 bp containing the variable TCR regions plus the TCR alpha or beta CDR3 sequences.

TRA and TRB CDR3 nucleotide reads were demultiplexed by matching reads to 10X barcodes from cells with existing expression data that passed filtering in the Cell Ranger pipeline. Following demultiplexing of the TRA and TRB CDR3s, reads were aligned against known TRA/TRB CDR3 sequences then assembled into clonotype families using miXCR with similar methodologies to a previous study (*66*). For any given 10X barcode, the most abundant TRA or TRB clonotype was accepted for further analysis; if 2 TRA or TRB clonotypes were equally abundant for a given 10X barcode, the clonotype with the highest UMI reads was used for further analysis. Only cells with paired TRA and TRB were used for further downstream analysis. Paired TCR Gini coefficients were calculated based on the number of unique cells sharing a specific TRA/TRB clonotype sequence across the sample's entire cell population. Analysis

involving both TCR clonotype and function was restricted to cells with both a mapped TRA/TRB and a functional population from clustering. The TCR sequencing was performed using the singleTCR package.

Clinical Features

American College of Rheumatology (ACR) Classification Criteria (*67*) or Systemic Lupus Erythematosus Disease Activity Index (SLEDAI) criteria (*26*): ACR ANA, anti-dsDNA, anti-Smith, Leukopenia, Malar Rash, Renal involvement, Neurological involvement, Arthritis, Mucosal Ulcers, Serositis, Thrombocytopenia, Discoid Rash, Photosensitivity, Lymphopenia, anti-Phospholipid, and SLEDAI Low Complement. These clinical features were selected because they were not sparsely represented within our data and they showed mild to moderate correlation with the SLEDAI instrument.

Sample Genotyping

CLUES SLE patients were genotyped on the Affymetrix World LAT Array, while the ImmVar and flaring SLE patients were genotyped on the OmniExpressExome54 chip.

The Affymetrix .CEL intensities were processed using the Analysis Power Tools (APT) in accordance with the Axiom Best Practices Genotyping Analysis Workflow (https://assets.thermofisher.com/TFS-Assets/LSG/manuals/axiom_genotyping_solution_analysis_guide.pdf). All CEL files were grouped into a batch and DQC scores were calculated using apt-geno-qc retaining CEL files with DQC scores greater or equal to 0.82. Call rates were generated using apt-genotype-axiom with the Step1 file for the corresponding 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 SNPs were retained for the analyses. These SNPs were further evaluated for missingness and heterozygosity, but no individuals were removed due to these metrics.

Processed OmniExpressExome54 genotypes were obtained (*10*). 787,355 SNPs were retained for further analysis after evaluating for call rate, missingness and heterozygosity with plink (*69*). We input 800,825 and 787,355 SNPs into the Michigan Imputation Server's web portal (*70*) for the LatArray and OmniExpress samples, which were further filtered to 696,523 and 631,954 SNPs respectively, when accounting for invalid alleles, multiallelic sites, monomorphic sites, allele mismatches, and SNP call rates less than 90%. A total of 21,412,068 SNPs were imputed from the resulting SNPs using the Haplotype Reference Consortium version 1.1 across both genotyping cohorts

and filtered to 16,176,651 with Rsq value >0.3. All SNPs were further filtered to only include those with a minor allele frequency >10% that were imputed in all individuals, resulting in 4,372,256 SNPs.

eQTL Analysis

Approximately 3.4 million SNPs with a MAF > 10% were used to map *cis*-eQTLs within a window +/- 100kbp of each gene, and a total of 9,500 genes were tested. Gene expression for each cell type was normalized using EdgeR's (*59*) trimmed means of M values (TMM) function, and cell-type-by-cell-type *cis*-eQTLs (CBC-eQTLs) were detected using the MatrixEQTL R package (*29*) separately for Europeans and Asians from the CLUES cohort and the ImmVar cohort. Genotype PCs clearly separate out the two populations (**Fig. S6C**). Self-identified race ethnicity matched PCA determined genetic ancestry in all case. The first 10 principal components of gene expression, 7 genotype PCs, expression PCs, age, sex, SLE status, and batch were included as covariates in all of the eQTL linear models. We further mapped cell-type-shared and cell-type-specific cis-eQTLs by decomposing the single cell expression data within each individual across 8 cell-types into one cell-type-shared expression component and 8 cell-type-specific components for each cell type. Then, genotypes are regressed against each of these components using MatrixEQTL to call cell-type-shared and cell-type-specific eQTLs (*31*). Per population results from MatrixEQTL were then meta-analyzed using the METASOFT package (*28*) (http://genetics.cs.ucla.edu/meta_jemdoc/). 1000 permutations were performed for each gene tested, meta-analyzed, and empirical p-values and FDR were calculated using the qvalue R package (https://www.bioconductor.org/packages/release/bioc/html/qvalue.html). The LocusZoom standalone software was used to visualize individual loci ((*71*); https://genome.sph.umich.edu/wiki/LocusZoom).

Heritability Analysis

Heritability, *cis*-genetic correlation (rG), and environmental correlation (rE) of gene expression in each pair of cell types was calculated using a Bivariate GREML analysis from the GCTA package ((*30*) ; https://cnsgenomics.com/software/gcta/#Overview). SNPs +/-500KB of each gene were used to evaluate the heritability of that gene in each cell type. For each gene in each pair of cell types the rG and rE were calculated. Genes with poor model fits (SE<0.2) were removed and the median rG and rE for each pair of cell types were used for visualization.

reQTL Analysis

SLE cases were analyzed for reQTLs. The type-1 ISG signature score was used as an interaction term with genotype, taking into account 10 PCs of genotype, age, sex and batch as covariates. The effect size and significance of the interaction term were calculated using the modelLinearCross option in MatrixEQTL. The ISG signature was calculated as the average gene expression of the ISGs making up the Pan_{up} module as described above. Per population results from MatrixEQTL were then meta-analyzed using the METASOFT package ((*28*); http://genetics.cs.ucla.edu/meta_jemdoc/).

ATAC-seq enrichment analysis

Cell type specific ATAC-seq peaks were downloaded (https://web.stanford.edu/group/pritchardlab/dataArchive/immune_atlas_web/index.html) (*32*) and merged into major cell types (B, T, NK, Myeloid, Open). For each set of eQTLs (cell-type-by-cell-type and cell-type-specific) and peaks, a Mann-Whitney test was applied to determine the enrichment for significant SNPs residing within each set of cell-type-specific peaks.

GWAS enrichment analysis

We performed stratified LDscore regression using the default parameters from the cell type specific workflow (https://github.com/bulik/ldsc/wiki/Cell-type-specific-analyses) (*33*) in a window of +/- 100kb around each gene containing a significant cell type specific *cis*-eQTL (FDR<0.5). All expressed genes were used as the background set in addition to the provided baseline scores. GWAS datasets were obtained from the Alkes' lab server (https://alkesgroup.broadinstitute.org/sumstats_formatted/), with the exception of SLE (*4*).

Modified Transcriptome Wide Association Study (TWAS) analysis

TWAS expression weights were generated using SNPs +/-100kb from each gene tested using pseudobulked expression profiles from each cell type and all PBMCs. TWAS models were generated using the FUSION software and publicly available GWAS summary statistics, mostly from TWAS hub. Secondary TWAS weights for models using the shared and cell-type-specific decomposed expression matrices were produced using the same approach as described in the eQTL discovery section. These weights were then used to produce separate models using just shared weights, just cell-type-specific weights, and both for a full model of expression. Multiple testing using TreeQTL was performed for an FDR<0.1 BY correction across all models. Genes in the MHC region of chromosome 6 were removed.

Colocalization analysis

Colocalization between eQTLs and the Bentham SLE GWAS study was performed using the COLOC package's approximant bayes factor method for summary statistics (*36*). Minor allele frequencies (MAF) were obtained from 1000 Genomes (*74*).

Single-cell ATAC-seq analysis

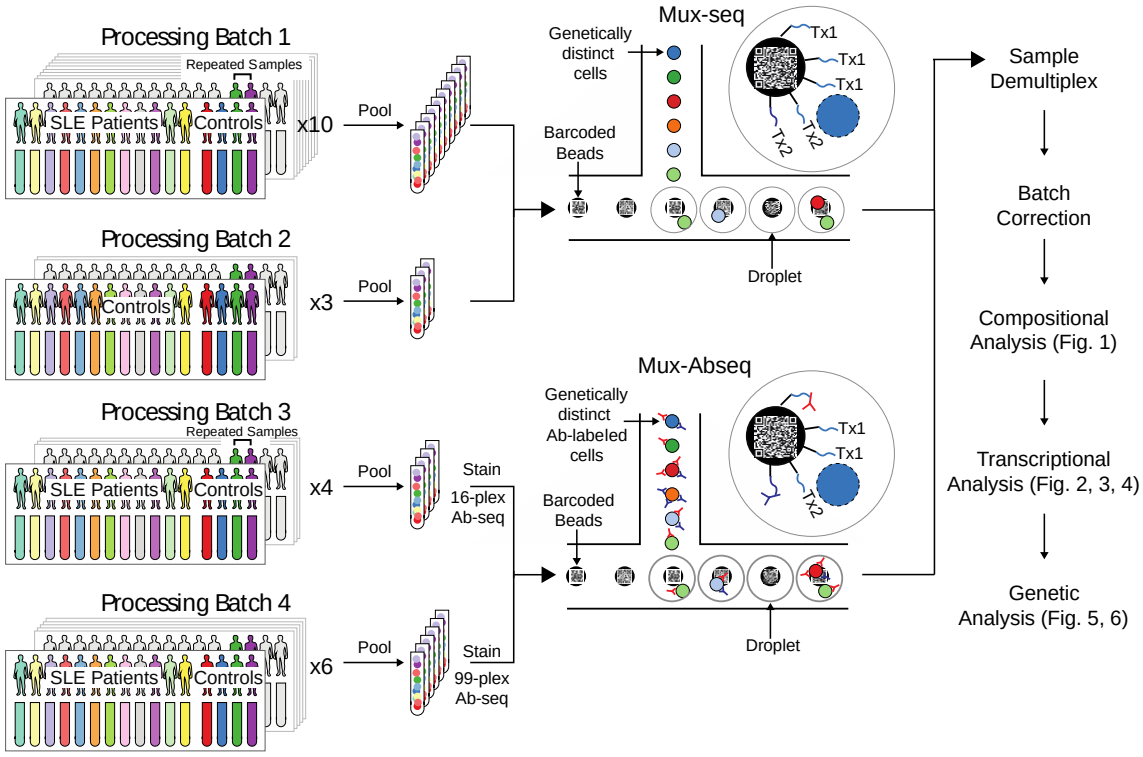
Primary PBMCs from 5 healthy individuals were incubated for 8 hours in 3 stimulation conditions (and one control condition): recombinant IFNB, recombinant IFNG, and PMA/Ionomycin. After incubation, cells were collected and mixed, and nuclei isolation was performed. Nuclei were then tagmented and assayed using 5 lanes of the 10X Chromium scATAC-seq kit according to an experimental design that allows for determination of the source sample from the individuals' SNPs. Droplet barcodes were filtered for only those containing nuclei and then demultiplexed using freemuxlet (https://github.com/statgen/popscle/). The ArchR package (*61*) and Scanpy (*58*) were then used for dimensionality reduction, clustering, and cell type annotation to produce cell type and condition gene locus plots.

A

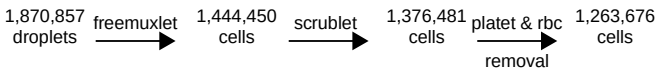
Cohort	Age	Asian	European	Hispanic	African American	F	M	**Individuals**	**Replicates**	**Samples**
CLUES Controls	44 (16.5)	23	25	1	0	48	1	**49**	**3**	**52**
CLUES Cases	45 (13.7)	76	70	0	0	132	14	**146**	**34**	**180**
ImmVar Controls	30 (7.6)	1	49	0	0	49	1	**50**	**44**	**94**
CLUES Flare	30 (5.7)	9	6	1	3	18	1	**19***	**10****	**29**
Subtotal		**107**	**149**	**2**	**3**	**244**	**17**	**264**	**91**	**355**

* 3 case replicates from CLUES Cases
** Flare/Treated Longitudinal Samples

B



C



D

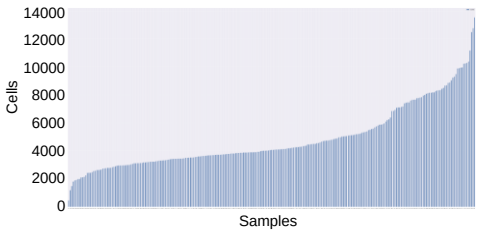


Fig. S1. Constructing a circulating census of SLE. (A) Four cohorts of SLE cases and healthy controls broken down by average age, ancestry, and gender. (n) = standard deviation. **(B)** Multiplexed single-cell RNA-sequencing applied to 23 pools of 7-19 donors per pool over four processing batches. Mux-Abseq(CITE-seq) was applied to processing batch 3 and 4 only. Cells were pooled, left unstained (processing batches one and two) or profiled using 16 and 99 DNA-conjugated antibodies (processing batch three and four), and processed using 10X Chromium 3' V2 chemistry. After sequencing, the resulting single-cell transcriptome or cell surface proteome profiles were sample demultiplexed, batch corrected by Combat, and analyzed. **(C)** Quality control and processing metrics. **(D)** Distribution of number of cells per sample

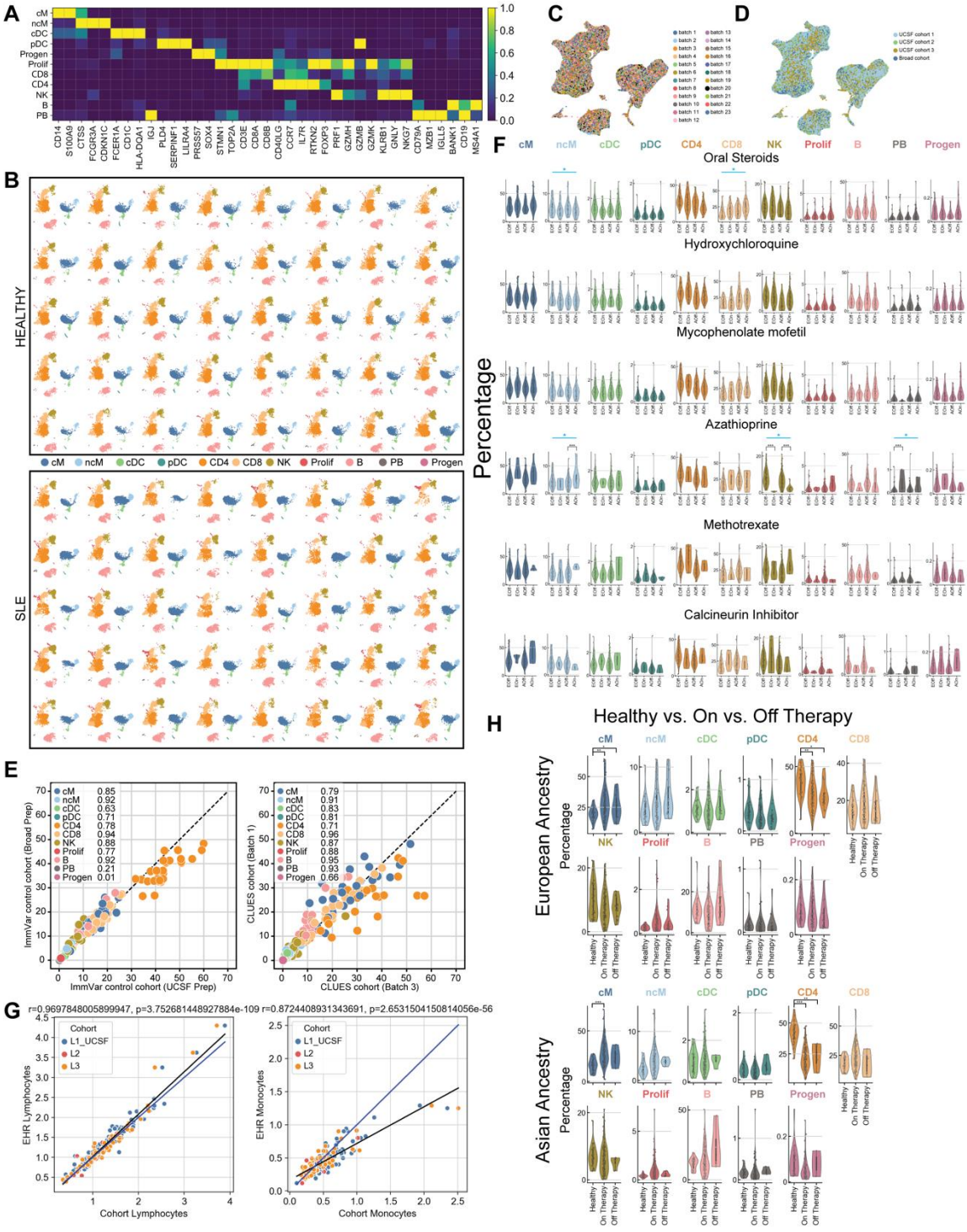


Fig. S2. Exploring effects of covariates on composition. (A) Key gene markers used to identify each of 11 cell types. **(B)** UMAP projection for 40 controls and 40 cases. **(C)** UMAP projection colored by processing pool. **(D)** UMAP projection colored by processing batch. **(E)** Correlation between cell type percentages for biological replicates

from different batches for ImmVar controls (left) and CLUES cases (right). **(F)** Percentage (y-axis) of each cell type split by cases receiving or not receiving a given therapy. (black bar and star *: WLS $P_{adjusted}$< 0.05; blue bar and star indicate significant meta-analysis by Fisher's method) **(G)** CLUES lymphocyte and monocyte abundances compared to UCSF electronic health record data colored by collection cohort**. (H)** Percentage of each cell type split by cases on therapy and cases not taking any medications at the time of sampling. WLS black * = $P_{adjusted}$<0.05, **= $P_{adjusted}$<0.01, ***= $P_{adjusted}$<0.001. Blue bar and star indicate significant Fisher's method meta P for both ancestries.

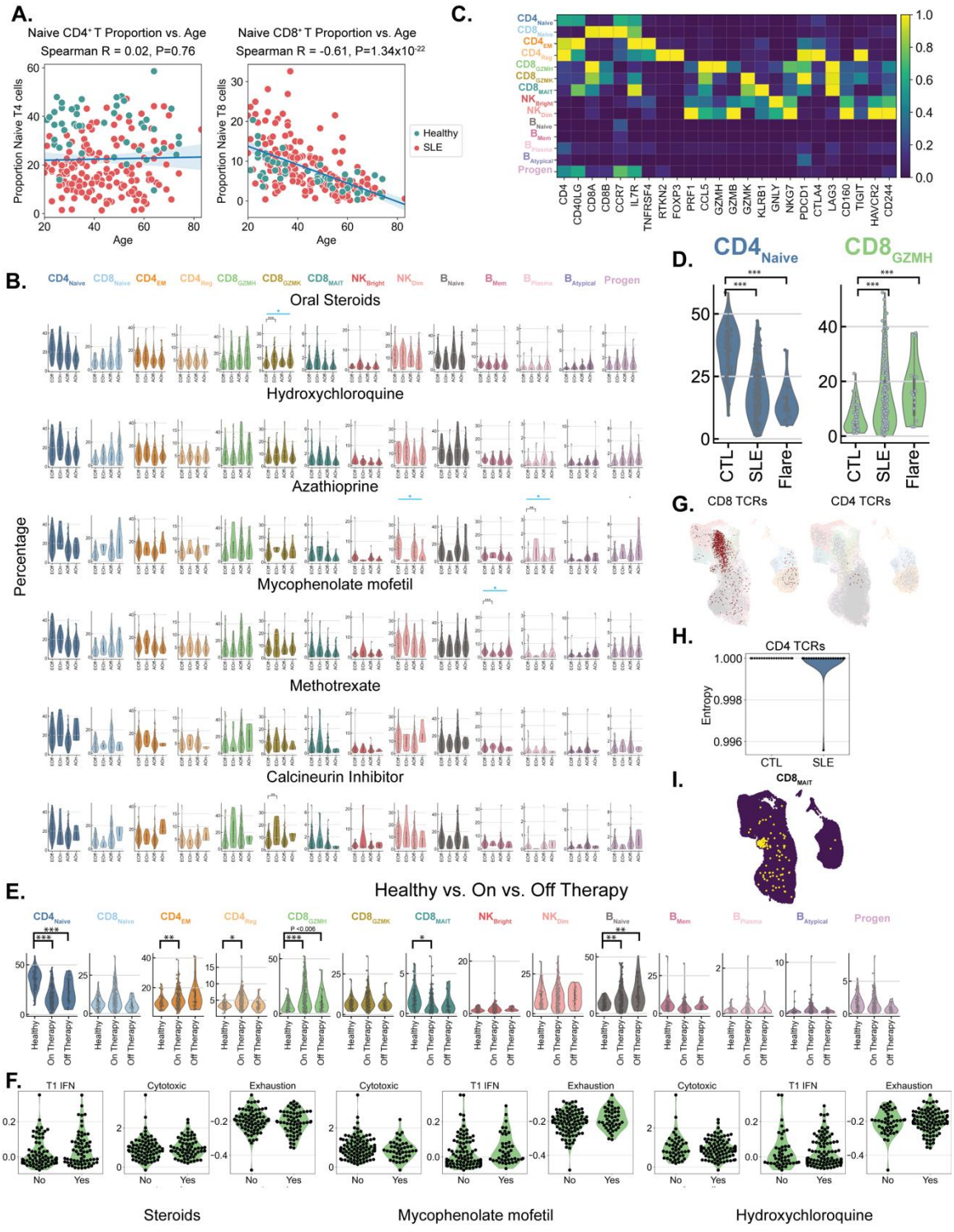


Fig. S3. Ancillary lymphocyte analysis. (A) Naïve CD4$^+$ T cells and Naïve CD8$^+$ T cell proportion vs donor age. **(B)** Percentages of each cell type split by cases receiving or not receiving a given therapy. black bar and star *: WLS $P_{adjusted} < 0.05$; blue bar and star indicate significant meta-analysis by Fisher's method) **(C)** Key gene markers used to identify each of the 14 lymphocyte populations. **(D)** Cell type percentage for naïve

$CD4^+$ T cells and *GZMH*$^+$ T cells for controls (CTL), managed cases (SLE), and cases in an active disease flare (Flare). (black bar and star *: WLS $P_{adjusted} < 0.05$) **(E)** Percentage of each cell type split by cases on therapy and cases not taking any medications at the time of sampling. (black bar and star *: WLS $P_{adjusted} < 0.05$; blue bar and star indicate significant meta-analysis by Fisher's method) **(F)** Pseudobulk expression of T1 IFN, Cytotoxic, and Exhaustion signatures within *GZMH*$^+$ T cells for cases receiving or not receiving hydrochloroquine, mycophenolate mofetil and oral steroids. **(G)** Expanded (red) and non-expanded (grey) TCRs for $CD8^+$ T cells and $CD4^+$ T cells. **(H)** Normalized Shannon Entropy for $CD4^+$ T cells**. (I)** Invariant TCR sequences associated with mucosal associated invariant T cells as a positive control of custom TCR amplification strategy. *= $P_{adjusted}<0.05$, **= $P_{adjusted}<0.01$, ***= $P_{adjusted}<0.001$. Blue bar and star indicate significant Fisher's method meta P for both ancestries.

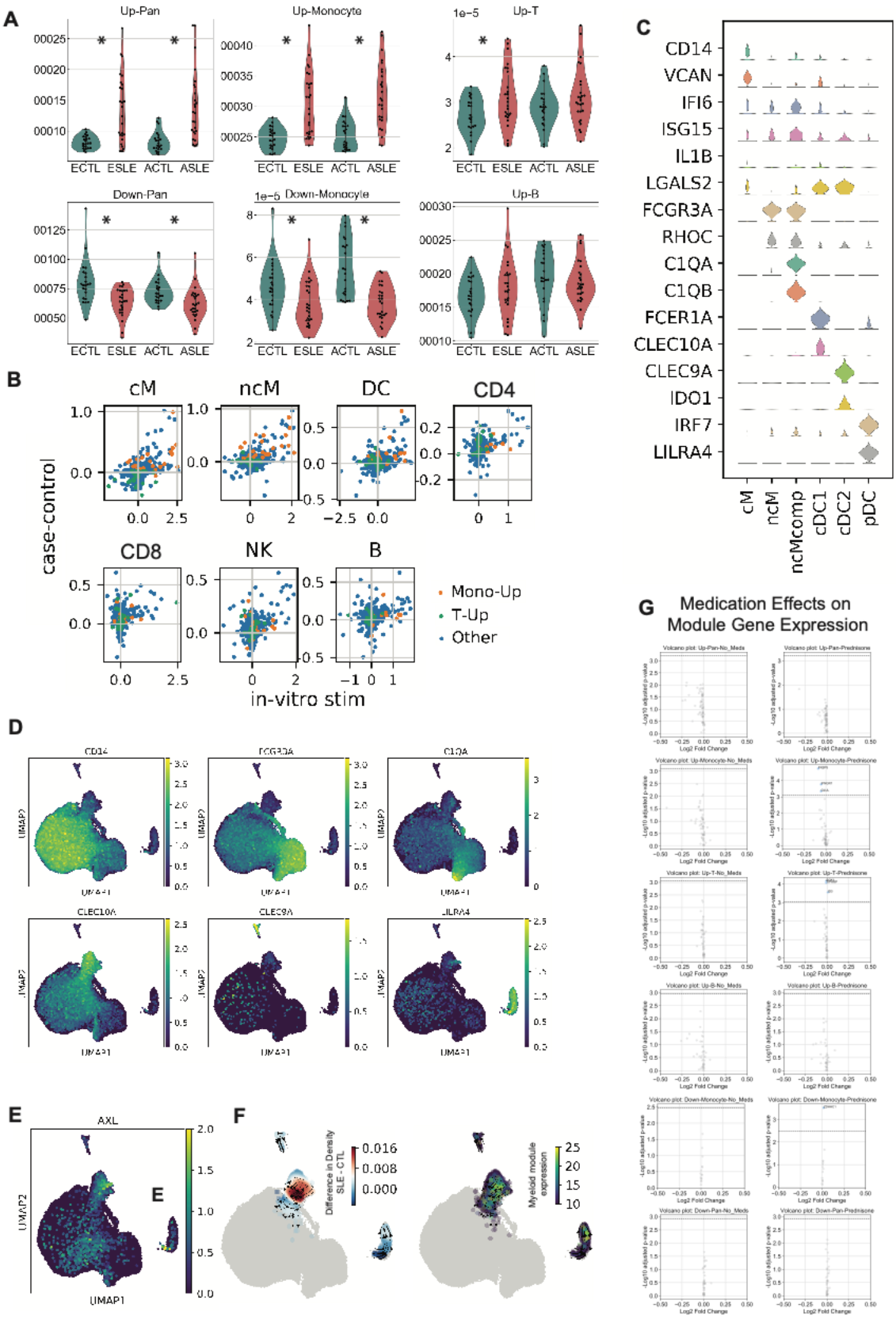

A
Up-Pan
Up-Monocyte
Up-T
Down-Pan
Down-Monocyte
Up-B
ECTL ESLE ACTL ASLE
B
cM
ncM
DC
CD4
CD8
NK
B
case-control
in-vitro stim
Mono-Up
T-Up
Other
C
CD14
VCAN
IFI6
ISG15
IL1B
LGALS2
FCGR3A
RHOC
C1QA
C1QB
FCER1A
CLEC10A
CLEC9A
IDO1
IRF7
LILRA4
cM
ncM
ncMcomp
cDC1
cDC2
pDC
D
CD14
FCGR3A
C1QA
CLEC10A
CLEC9A
LILRA4
UMAP1
UMAP2
E
AXL
F
Difference in Density SLE - CTL
Myeloid module expression
G
Medication Effects on Module Gene Expression

Fig. S4. Ancillary monocyte analysis. (A) Log10 module expression split by case (red; SLE) and control (green; CTL) and ancestry (A: Asian, E: European). * = Wilcoxon rank-sums P<0.05. **(B)** Correlation of fold change between case control (y-axis) and previously published fold change between interferon-beta stimulated versus unstimulated cells (x-axis). **(C)** Gene markers used to call cell types in myeloid lineage. **(D)** Feature plots of marker genes. **(E)** *AXL* expression in myeloid cells. **(F)** RNA velocity stream plots with SLE cell density and monocyte upregulated gene module expression. **(G)** Log2 fold change pseudobulk gene expression of module genes between cases on therapy and cases not taking any medications and cases on prednisone versus cases not on prednisone. Genes reaching significance after Bonferroni correction (dashed line) are labeled and colored blue.

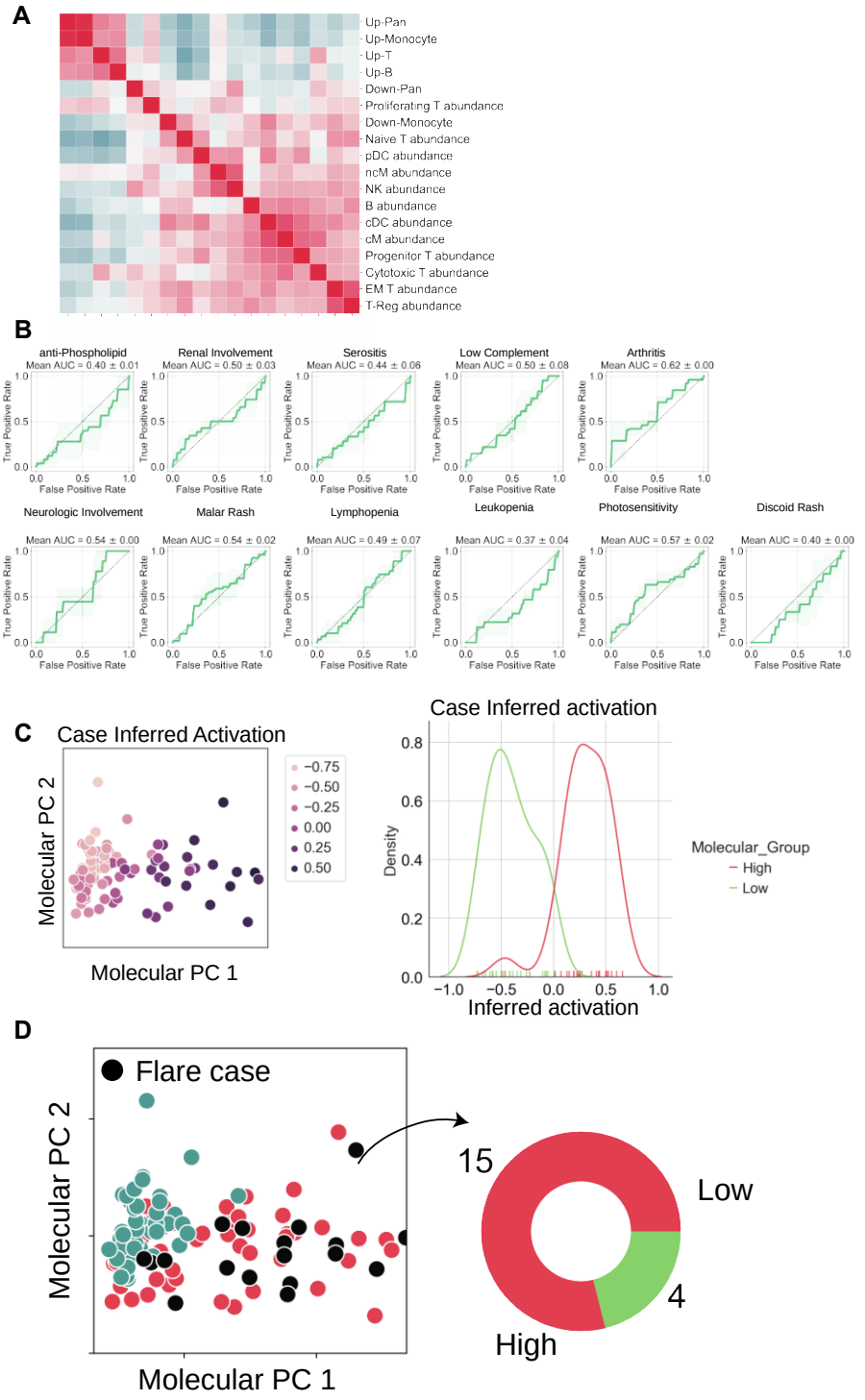


Fig. S5. Further exploration of clinical and molecular feature of SLE. (A) Correlation between $\log_{10}$ expression of five other modules (x-axis) and $\log_{10}$ abundance of cell types in processing batch 4 cases only. **(B)** Receiver Operating Curves for out-of-sample prediction using a logistic regression model based on expression module features for 11 clinical features. **(C)** Mean monocyte inferred activation for each case projected in gene

expression derived PCA space. Histogram showing the High molecular cluster cases had higher inferred activation. **(D)** Cases in active disease flare projected into train-dataset-fitted PCA space with pie chart representing predicted cluster membership. Controls: Green. Cases: Red. Flare Cases: Black.

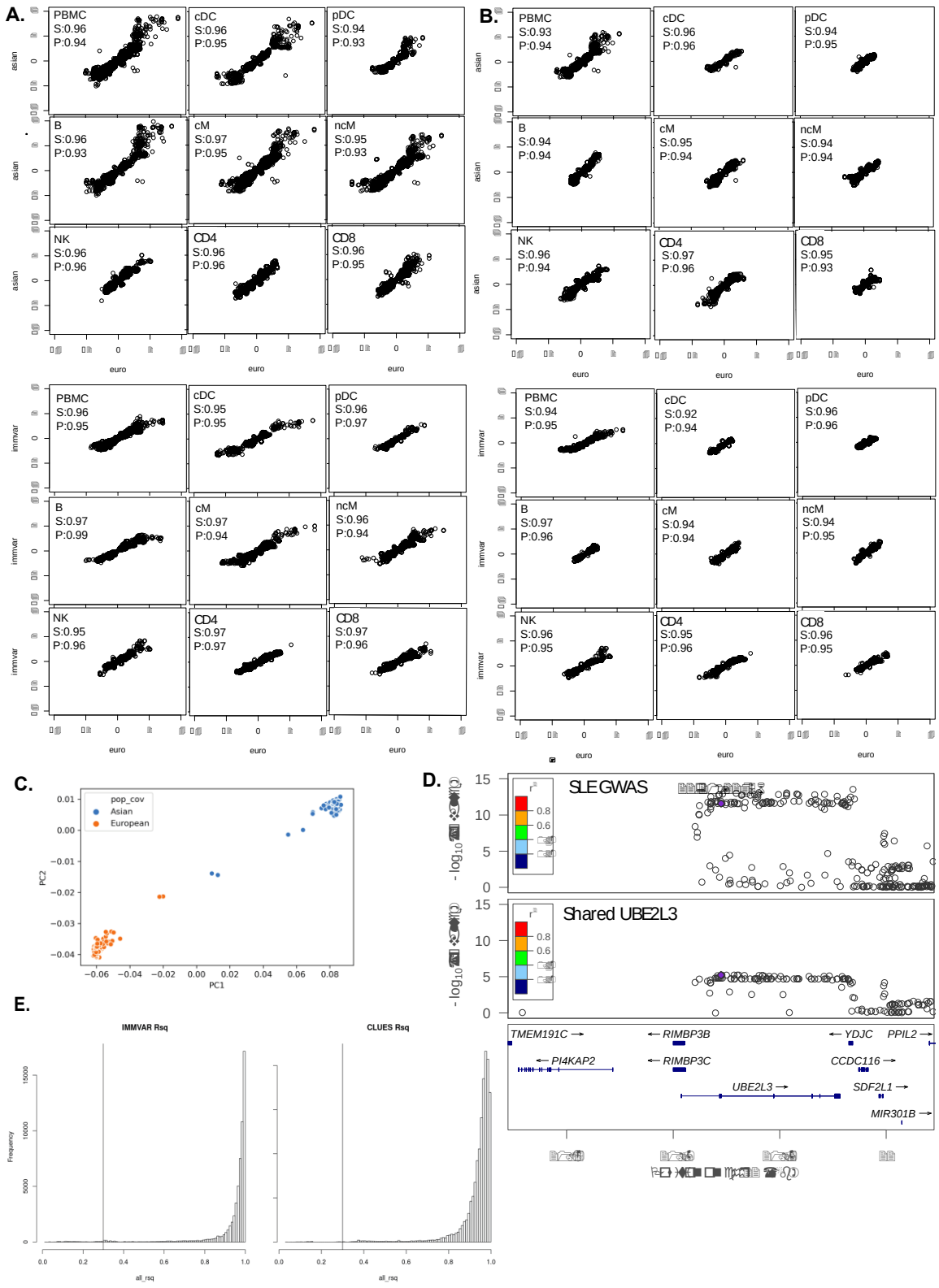


Fig. S6. Further analyses of cis-eQTLs (A) Scatterplot of effect sizes of CBC-*cis*-eQTLs per cell type between CLUES Asian and European samples as well as CLUES European and Immvar samples. **(B)** Scatterplot of effect sizes of cell-type-specific-*cis*-

eQTLs per cell type between CLUES Asian and European samples as well as CLUES European and Immvar samples. **(C)** Genotype PCs colored by population. **(D)** LocusZoom plots of shared-*cis*-eQTLs and SLE GWAS for *UBE2L3*. **(E)** Rsq distribution of imputed SNPs MAF>10%, Cutoff of 0.3 shown with vertical line on the plot.

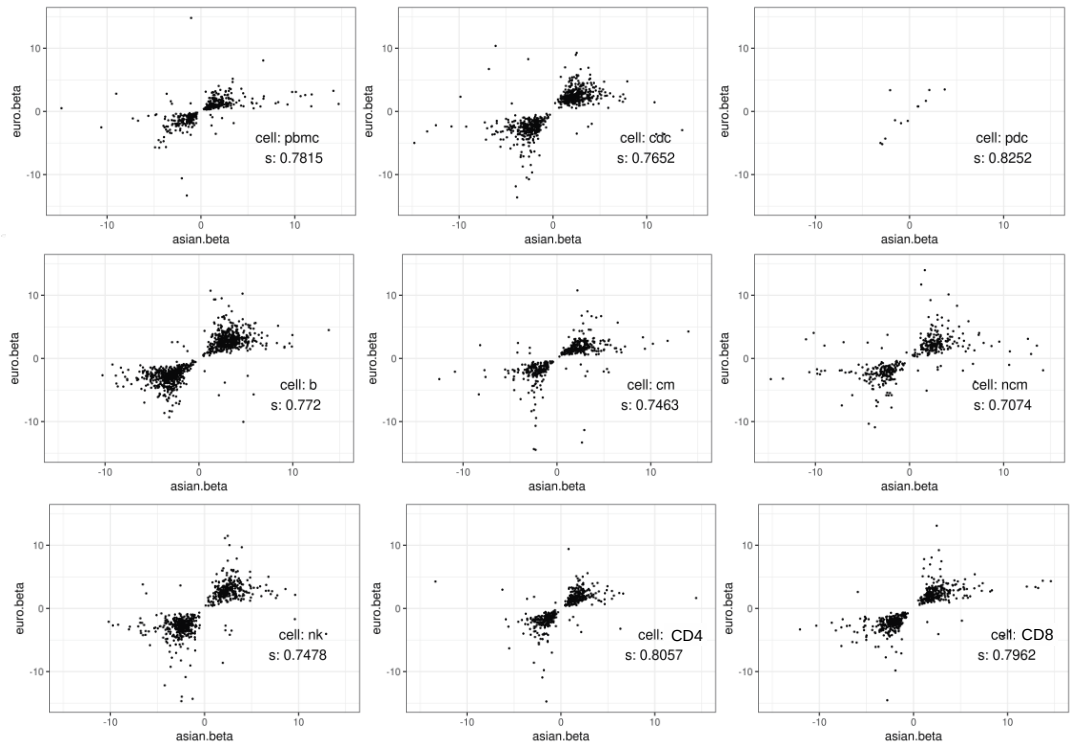


Fig. S7. Cis-reQTLs are consistent between European and Asian samples. (A) Scatterplot of effect sizes of *cis*-reQTLs per cell type between Asian and European samples.

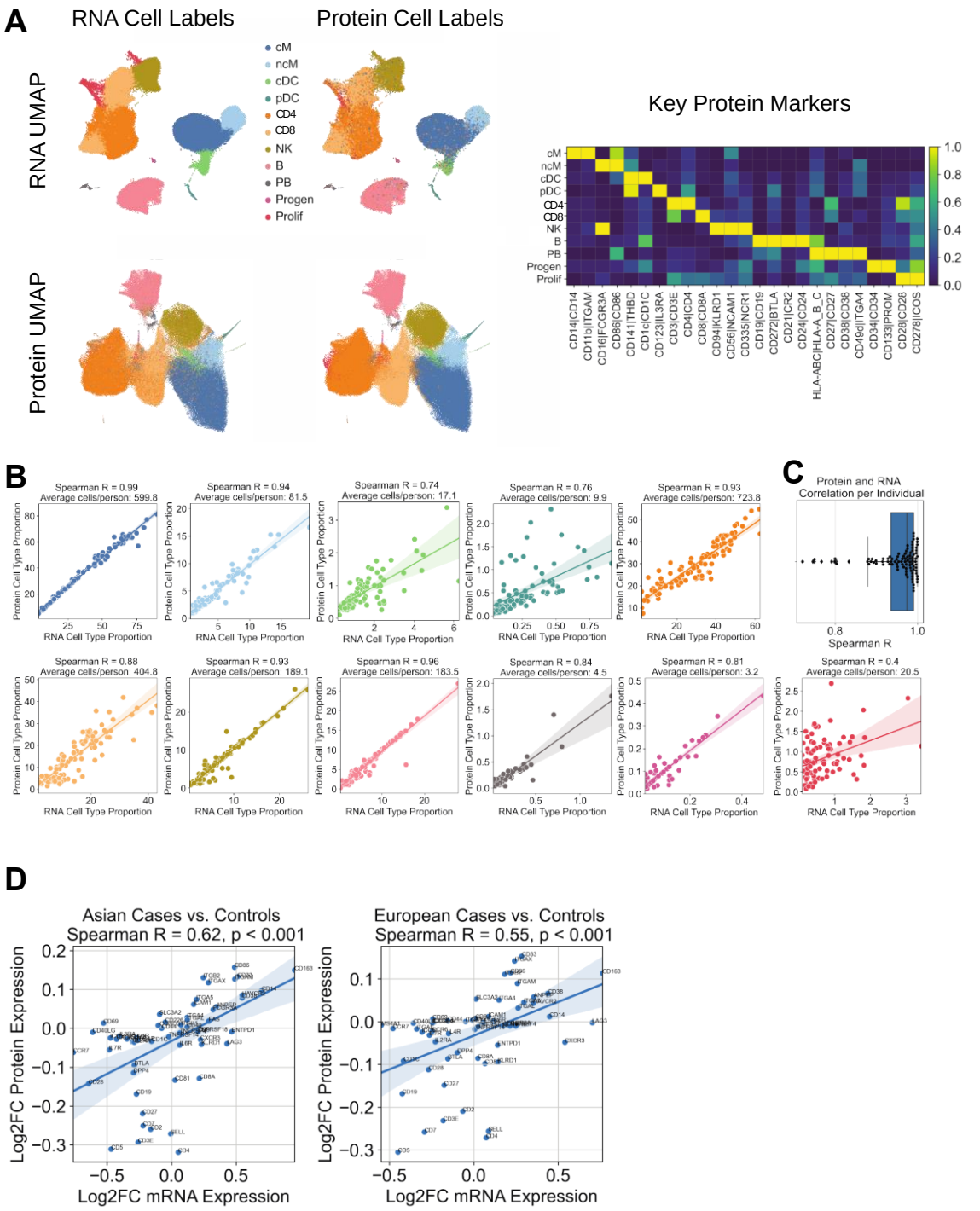


Fig. S8. CITE-Seq validation cohort. (**A**) UMAP projections based on RNA and protein features (Key Protein Markers heatmap). Cell types labeled based on RNA or protein derived clustering. (**B**) Correlation between RNA and protein based cell type proportions for each cell type. (**C**) Box and Whiskers plot demonstrating the spread of correlations between RNA and protein based cell type proportion for each individual. (**D**) Correlation

between log2 fold change expression between cases and controls based either on protein expression or RNA expression.

References and Notes

1. E. E. Carter, S. G. Barr, A. E. Clarke, The global burden of SLE: Prevalence, health disparities and socioeconomic impact. *Nat. Rev. Rheumatol.* **12**, 605–620 (2016). doi:10.1038/nrrheum.2016.137 Medline

2. A. Kaul, C. Gordon, M. K. Crow, Z. Touma, M. B. Urowitz, R. van Vollenhoven, G. Ruiz-Irastorza, G. Hughes, Systemic lupus erythematosus. *Nat. Rev. Dis. Primers* **2**, 16039 (2016). doi:10.1038/nrdp.2016.39 Medline

3. R. Banchereau, S. Hong, B. Cantarel, N. Baldwin, J. Baisch, M. Edens, A.-M. Cepika, P. Acs, J. Turner, E. Anguiano, P. Vinod, S. Kahn, G. Obermoser, D. Blankenship, E. Wakeland, L. Nassi, A. Gotte, M. Punaro, Y.-J. Liu, J. Banchereau, J. Rossello-Urgell, T. Wright, V. Pascual, Personalized Immunomonitoring Uncovers Molecular Networks that Stratify Lupus Patients. *Cell* **165**, 551–565 (2016). doi:10.1016/j.cell.2016.03.008 Medline

4. J. Bentham, D. L. Morris, D. S. C. 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 Medline

5. J. Banchereau, V. Pascual, Type I interferon in systemic lupus erythematosus and other autoimmune diseases. *Immunity* **25**, 383–392 (2006). doi:10.1016/j.immuni.2006.08.010 Medline

6. D. Nehar-Belaid, S. Hong, R. Marches, G. Chen, M. Bolisetty, J. Baisch, L. Walters, M. Punaro, R. J. Rossi, C.-H. Chung, R. P. Huynh, P. Singh, W. F. Flynn, J.-A. Tabanor-Gayle, N. Kuchipudi, A. Mejias, M. A. Collet, A. L. Lucido, K. Palucka, P. Robson, S. Lakshminarayanan, O. Ramilo, T. Wright, V. Pascual, J. F. Banchereau, Mapping systemic lupus erythematosus heterogeneity at the single-cell level. *Nat. Immunol.* **21**, 1094–1106 (2020). doi:10.1038/s41590-020-0743-0 Medline

7. S. Sharma, Z. Jin, E. Rosenzweig, S. Rao, K. Ko, T. B. Niewold, Widely divergent transcriptional patterns between SLE patients of different ancestral backgrounds in sorted

immune cell populations. *J. Autoimmun.* **60**, 51–58 (2015). doi:10.1016/j.jaut.2015.04.002 Medline

8. 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 Medline

9. C. M. Lanata, J. Nititham, K. E. Taylor, S. A. Chung, D. G. Torgerson, M. F. Seldin, B. A. Pons-Estel, T. Tusié-Luna, B. P. Tsao, E. F. Morand, M. E. Alarcón-Riquelme, L. A. Criswell, Genetic contributions to lupus nephritis in a multi-ethnic cohort of systemic lupus erythematous patients. *PLOS ONE* **13**, e0199003 (2018). doi:10.1371/journal.pone.0199003 Medline

10. T. Raj, K. Rothamel, S. Mostafavi, C. Ye, M. N. Lee, J. M. Replogle, T. Feng, M. Lee, N. Asinovski, I. Frohlich, S. Imboywa, A. Von Korff, Y. Okada, N. A. Patsopoulos, S. Davis, C. McCabe, H. I. Paik, G. P. Srivastava, S. Raychaudhuri, D. A. Hafler, D. Koller, A. Regev, N. Hacohen, D. Mathis, C. Benoist, B. E. Stranger, P. L. De Jager, Polarization of the effects of autoimmune and neurodegenerative risk alleles in leukocytes. *Science* **344**, 519–523 (2014). doi:10.1126/science.1249547 Medline

11. M. N. Lee, C. Ye, A.-C. Villani, T. Raj, W. Li, T. M. Eisenhaure, S. H. Imboywa, P. I. Chipendo, F. A. Ran, K. Slowikowski, L. D. Ward, K. Raddassi, C. McCabe, M. H. Lee, I. Y. Frohlich, D. A. Hafler, M. Kellis, S. Raychaudhuri, F. Zhang, B. E. Stranger, C. O. Benoist, P. L. De Jager, A. Regev, N. Hacohen, Common genetic variants modulate pathogen-sensing responses in human dendritic cells. *Science* **343**, 1246980 (2014). doi:10.1126/science.1246980 Medline

12. 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 Medline

13. 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 Medline

14. V. A. Traag, L. Waltman, N. J. van Eck, From Louvain to Leiden: Guaranteeing well-connected communities. *Sci. Rep.* **9**, 5233 (2019). doi:10.1038/s41598-019-41695-z Medline

15. L. McInnes, J. Healy, J. Melville, UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. arXiv 1802.03426 (2018).

16. Z. Zhu, Z. Zheng, F. Zhang, Y. Wu, M. Trzaskowski, R. Maier, M. R. Robinson, J. J. McGrath, P. M. Visscher, N. R. Wray, J. Yang, Causal associations between risk factors and common diseases inferred from GWAS summary data. *Nat. Commun.* **9**, 224 (2018). doi:10.1038/s41467-017-02317-2 Medline

17. C. Bycroft, C. Freeman, D. Petkova, G. Band, L. T. Elliott, K. Sharp, A. Motyer, D. Vukcevic, O. Delaneau, J. O'Connell, A. Cortes, S. Welsh, A. Young, M. Effingham, G. McVean, S. Leslie, N. Allen, P. Donnelly, J. Marchini, The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018). doi:10.1038/s41586-018-0579-z Medline

18. V. R. Moulton, A. Suarez-Fueyo, E. Meidan, H. Li, M. Mizui, G. C. Tsokos, Pathogenesis of human systemic lupus erythematosus: A cellular perspective. *Trends Mol. Med.* **23**, 615–635 (2017). doi:10.1016/j.molmed.2017.05.006 Medline

19. K. Rubtsova, A. V. Rubtsov, M. P. Cancro, P. Marrack, Age-Associated B Cells: A T-bet-Dependent Effector with Roles in Protective and Pathogenic Immunity. *J. Immunol.* **195**, 1933–1937 (2015). doi:10.4049/jimmunol.1501209 Medline

20. A. Ferraro, A. M. D'Alise, T. Raj, N. Asinovski, R. Phillips, A. Ergun, J. M. Replogle, A. Bernier, L. Laffel, B. E. Stranger, P. L. De Jager, D. Mathis, C. Benoist, Interindividual variation in human T regulatory cells. *Proc. Natl. Acad. Sci. U.S.A.* **111**, E1111–E1120 (2014). doi:10.1073/pnas.1401343111 Medline

21. Y. Kotliarov, R. Sparks, A. J. Martins, M. P. Mulè, Y. Lu, M. Goswami, L. Kardava, R. Banchereau, V. Pascual, A. Biancotto, J. Chen, P. L. Schwartzberg, N. Bansal, C. C. Liu,

F. Cheung, S. Moir, J. S. Tsang, Broad immune activation underlies shared set point signatures for vaccine responsiveness in healthy individuals and disease activity in patients with lupus. *Nat. Med.* **26**, 618–629 (2020). doi:10.1038/s41591-020-0769-8 Medline

22. H. Shigematsu, B. Reizis, H. Iwasaki, S. Mizuno, D. Hu, D. Traver, P. Leder, N. Sakaguchi, K. Akashi, Plasmacytoid dendritic cells activate lymphoid-specific genetic programs irrespective of their cellular origin. *Immunity* **21**, 43–53 (2004). doi:10.1016/j.immuni.2004.06.011 Medline

23. 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 Medline

24. V. Bergen, M. Lange, S. Peidli, F. A. Wolf, F. J. Theis, Generalizing RNA velocity to transient cell states through dynamical modeling. *Nat. Biotechnol.* **38**, 1408–1414 (2020). doi:10.1038/s41587-020-0591-3 Medline

25. G. La Manno, R. Soldatov, A. Zeisel, E. Braun, H. Hochgerner, V. Petukhov, K. Lidschreiber, M. E. Kastriti, P. Lönnerberg, A. Furlan, J. Fan, L. E. Borm, Z. Liu, D. van Bruggen, J. Guo, X. He, R. Barker, E. Sundström, G. Castelo-Branco, P. Cramer, I. Adameyko, S. Linnarsson, P. V. Kharchenko, RNA velocity of single cells. *Nature* **560**, 494–498 (2018). doi:10.1038/s41586-018-0414-6 Medline

26. J. P. Buyon, M. A. Petri, M. Y. Kim, K. C. Kalunian, J. Grossman, B. H. Hahn, J. T. Merrill, L. Sammaritano, M. Lockshin, G. S. Alarcón, S. Manzi, H. M. Belmont, A. D. Askanase, L. Sigler, M. A. Dooley, J. Von Feldt, W. J. McCune, A. Friedman, J. Wachs, M. Cronin, M. Hearth-Holmes, M. Tan, F. Licciardi, The effect of combined estrogen and progesterone hormone replacement therapy on disease activity in systemic lupus erythematosus: A randomized trial. *Ann. Intern. Med.* **142**, 953–962 (2005). doi:10.7326/0003-4819-142-12_Part_1-200506210-00004 Medline

27. L. R. Shiow, D. B. Rosen, N. Brdicková, Y. Xu, J. An, L. L. Lanier, J. G. Cyster, M. Matloubian, CD69 acts downstream of interferon-α/β to inhibit S1P1 and lymphocyte egress from lymphoid organs. *Nature* **440**, 540–544 (2006). doi:10.1038/nature04606 Medline

28. B. Han, E. Eskin, Random-effects model aimed at discovering associations in meta-analysis of genome-wide association studies. *Am. J. Hum. Genet.* **88**, 586–598 (2011). doi:10.1016/j.ajhg.2011.04.014 Medline

29. A. A. Shabalin, Matrix eQTL: Ultra fast eQTL analysis via large matrix operations. *Bioinformatics* **28**, 1353–1358 (2012). doi:10.1093/bioinformatics/bts163 Medline

30. 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 Medline

31. A. Lu *et al*., Fast and powerful statistical method for context-specific QTL mapping in multi-context genomic studies. bioRxiv 448889 (2021). doi:10.1101/2021.06.17.448889

32. D. Calderon, M. L. T. Nguyen, A. Mezger, A. Kathiria, F. Müller, V. Nguyen, N. Lescano, B. Wu, J. Trombetta, J. V. Ribado, D. A. Knowles, Z. Gao, F. Blaeschke, A. V. Parent, T. D. Burt, M. S. Anderson, L. A. Criswell, W. J. Greenleaf, A. Marson, J. K. Pritchard, Landscape of stimulation-responsive chromatin across diverse human immune cells. *Nat. Genet.* **51**, 1494–1505 (2019). doi:10.1038/s41588-019-0505-9 Medline

33. H. K. Finucane, Y. A. Reshef, V. Anttila, K. Slowikowski, A. Gusev, A. Byrnes, S. Gazal, P.-R. Loh, C. Lareau, N. Shoresh, G. Genovese, A. Saunders, E. Macosko, S. Pollack, J. R. B. Perry, J. D. Buenrostro, B. E. Bernstein, S. Raychaudhuri, S. McCarroll, B. M. Neale, A. L. Price, Brainstorm Consortium, Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nat. Genet.* **50**, 621–629 (2018). doi:10.1038/s41588-018-0081-4 Medline

34. E. Nashi, Y. Wang, B. Diamond, The role of B cells in lupus pathogenesis. *Int. J. Biochem. Cell Biol.* **42**, 543–550 (2010). doi:10.1016/j.biocel.2009.10.011 Medline

35. X. Hu, H. Kim, E. Stahl, R. Plenge, M. Daly, S. Raychaudhuri, Integrating autoimmune risk loci with gene-expression data identifies specific pathogenic immune cell subsets. *Am. J. Hum. Genet.* **89**, 496–506 (2011). doi:10.1016/j.ajhg.2011.09.002 Medline

36. C. Giambartolomei, J. Z. Liu, W. Zhang, M. Hauberg, H. Shi, J. Boocock, J. Pickrell, A. E. Jaffe, B. Pasaniuc, P. Roussos,0 CommonMind Consortium, A Bayesian framework for multiple trait colocalization from summary association statistics. *Bioinformatics* **34**, 2538–2545 (2018). doi:10.1093/bioinformatics/bty147 Medline

37. M. F. Moffatt, M. Kabesch, L. Liang, A. L. Dixon, D. Strachan, S. Heath, M. Depner, A. von Berg, A. Bufe, E. Rietschel, A. Heinzmann, B. Simma, T. Frischer, S. A. G. Willis-Owen, K. C. C. Wong, T. Illig, C. Vogelberg, S. K. Weiland, E. von Mutius, G. R. Abecasis, M. Farrall, I. G. Gut, G. M. Lathrop, W. O. C. Cookson, Genetic variants regulating ORMDL3 expression contribute to the risk of childhood asthma. *Nature* **448**, 470–473 (2007). doi:10.1038/nature06014 Medline

38. L. Jostins, S. Ripke, R. K. Weersma, R. H. Duerr, D. P. McGovern, K. Y. Hui, J. C. Lee, L. P. Schumm, Y. Sharma, C. A. Anderson, J. Essers, M. Mitrovic, K. Ning, I. Cleynen, E. Theatre, S. L. Spain, S. Raychaudhuri, P. Goyette, Z. Wei, C. Abraham, J.-P. Achkar, T. Ahmad, L. Amininejad, A. N. Ananthakrishnan, V. Andersen, J. M. Andrews, L. Baidoo, T. Balschun, P. A. Bampton, A. Bitton, G. Boucher, S. Brand, C. Büning, A. Cohain, S. Cichon, M. D'Amato, D. De Jong, K. L. Devaney, M. Dubinsky, C. Edwards, D. Ellinghaus, L. R. Ferguson, D. Franchimont, K. Fransen, R. Gearry, M. Georges, C. Gieger, J. Glas, T. Haritunians, A. Hart, C. Hawkey, M. Hedl, X. Hu, T. H. Karlsen, L. Kupcinskas, S. Kugathasan, A. Latiano, D. Laukens, I. C. Lawrance, C. W. Lees, E. Louis, G. Mahy, J. Mansfield, A. R. Morgan, C. Mowat, W. Newman, O. Palmieri, C. Y. Ponsioen, U. Potocnik, N. J. Prescott, M. Regueiro, J. I. Rotter, R. K. Russell, J. D. Sanderson, M. Sans, J. Satsangi, S. Schreiber, L. A. Simms, J. Sventoraityte, S. R. Targan, K. D. Taylor, M. Tremelling, H. W. Verspaget, M. De Vos, C. Wijmenga, D. C. Wilson, J. Winkelmann, R. J. Xavier, S. Zeissig, B. Zhang, C. K. Zhang, H. Zhao, M. S. Silverberg, V. Annese, H. Hakonarson, S. R. Brant, G. Radford-Smith, C. G. Mathew, J. D. Rioux, E. E. Schadt, M. J. Daly, A. Franke, M. Parkes, S. Vermeire, J. C. Barrett, J. H. Cho, International IBD Genetics Consortium, Host-microbe interactions have shaped the

genetic architecture of inflammatory bowel disease. *Nature* **491**, 119–124 (2012). doi:10.1038/nature11582 Medline

39. 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 Medline

40. B. Morgan, L. Sun, N. Avitahl, K. Andrikopoulos, T. Ikeda, E. Gonzales, P. Wu, S. Neben, K. Georgopoulos, Aiolos, a lymphoid restricted transcription factor that interacts with Ikaros to regulate lymphocyte differentiation. *EMBO J.* **16**, 2004–2013 (1997). doi:10.1093/emboj/16.8.2004 Medline

41. L. Li, Y. Li, Y. Bai, Role of GSDMB in Pyroptosis and Cancer. *Cancer Manag. Res.* **12**, 3033–3043 (2020). doi:10.2147/CMAR.S246948 Medline

42. Y. Zhang, S. A. G. Willis-Owen, S. Spiegel, C. M. Lloyd, M. F. Moffatt, W. O. C. M. Cookson, The ORMDL3 Asthma Gene Regulates ICAM1 and Has Multiple Effects on Cellular Inflammation. *Am. J. Respir. Crit. Care Med.* **199**, 478–488 (2019). doi:10.1164/rccm.201803-0438OC Medline

43. B. James, S. Milstien, S. Spiegel, ORMDL3 and allergic asthma: From physiology to pathology. *J. Allergy Clin. Immunol.* **144**, 634–640 (2019). doi:10.1016/j.jaci.2019.07.023 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 Medline

45. J. Yang, T. Ferreira, A. P. Morris, S. E. Medland, P. A. F. Madden, A. C. Heath, N. G. Martin, G. W. Montgomery, M. N. Weedon, R. J. Loos, T. M. Frayling, M. I. McCarthy, J. N. Hirschhorn, M. E. Goddard, P. M. Visscher, Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat. Genet.* **44**, 369–375 (2012). doi:10.1038/ng.2213 Medline

46. B. J. Schmiedel, G. Seumois, D. Samaniego-Castruita, J. Cayford, V. Schulten, L. Chavez, F. Ay, A. Sette, B. Peters, P. Vijayanand, 17q21 asthma-risk variants switch CTCF binding and regulate IL-2 production by T cells. *Nat. Commun.* **7**, 13426 (2016). doi:10.1038/ncomms13426 Medline

47. M. Thompson *et al*., Multi-context genetic modeling of transcriptional regulation resolves novel disease loci. bioRxiv 461579 (2021). doi:10.1101/2021.09.23.461579

48. C. J. Ye, J. Chen, A.-C. Villani, R. E. Gate, M. Subramaniam, T. Bhangale, M. N. Lee, T. Raj, R. Raychowdhury, W. Li, N. Rogel, S. Simmons, S. H. Imboywa, P. I. Chipendo, C. McCabe, M. H. Lee, I. Y. Frohlich, B. E. Stranger, P. L. De Jager, A. Regev, T. Behrens, N. Hacohen, Genetic analysis of isoform usage in the human anti-viral response reveals influenza-specific regulation of *ERAP2* transcripts under balancing selection. *Genome Res.* **28**, 1812–1825 (2018). doi:10.1101/gr.240390.118 Medline

49. J. F. Degner, A. A. Pai, R. Pique-Regi, J.-B. Veyrieras, D. J. Gaffney, J. K. Pickrell, S. De Leon, K. Michelini, N. Lewellen, G. E. Crawford, M. Stephens, Y. Gilad, J. K. Pritchard, DNase I sensitivity QTLs are a major determinant of human expression variation. *Nature* **482**, 390–394 (2012). doi:10.1038/nature10808 Medline

50. S. J. Rivero, E. Díaz-Jouanen, D. Alarcón-Segovia, Lymphopenia in systemic lupus erythematosus: Clinical, diagnostic, and prognostic significance. *Arthritis Rheum.* **21**, 295–305 (1978). doi:10.1002/art.1780210302 Medline

51. E. F. Morand, R. Furie, Y. Tanaka, I. N. Bruce, A. D. Askanase, C. Richez, S.-C. Bae, P. Z. Brohawn, L. Pineda, A. Berglind, R. Tummala, TULIP-2 Trial Investigators, Trial of anifrolumab in active systemic lupus erythematosus. *N. Engl. J. Med.* **382**, 211–221 (2020). doi:10.1056/NEJMoa1912196 Medline

52. M. Cella, D. Jarrossay, F. Facchetti, O. Alebardi, H. Nakajima, A. Lanzavecchia, M. Colonna, Plasmacytoid monocytes migrate to inflamed lymph nodes and produce large amounts of type I interferon. *Nat. Med.* **5**, 919–923 (1999). doi:10.1038/11360 Medline

53. T. B. Niewold, J. E. Adler, S. B. Glenn, T. J. A. Lehman, J. B. Harley, M. K. Crow, Age- and sex-related patterns of serum interferon-α activity in lupus families. *Arthritis Rheum.* **58**, 2113–2119 (2008). doi:10.1002/art.23619 Medline

54. P. Blanco, V. Pitard, J.-F. Viallard, J.-L. Taupin, J.-L. Pellegrin, J.-F. Moreau, Increase in activated $CD8^+$ T lymphocytes expressing perforin and granzyme B correlates with disease activity in patients with systemic lupus erythematosus. *Arthritis Rheum.* **52**, 201–211 (2005). doi:10.1002/art.20745 Medline

55. L. Casciola-Rosen, F. Andrade, D. Ulanet, W. B. Wong, A. Rosen, Cleavage by granzyme B is strongly predictive of autoantigen status: Implications for initiation of autoimmunity. *J. Exp. Med.* **190**, 815–826 (1999). doi:10.1084/jem.190.6.815 Medline

56. M. Faroudi, C. Utzny, M. Salio, V. Cerundolo, M. Guiraud, S. Müller, S. Valitutti, Lytic versus stimulatory synapse in cytotoxic T lymphocyte/target cell interaction: Manifestation of a dual activation threshold. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 14145–14150 (2003). doi:10.1073/pnas.2334336100 Medline

57. C. L. Vanderlugt, S. D. Miller, Epitope spreading in immune-mediated diseases: Implications for immunotherapy. *Nat. Rev. Immunol.* **2**, 85–95 (2002). doi:10.1038/nri724 Medline

58. F. A. Wolf, P. Angerer, F. J. Theis, SCANPY: Large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018). doi:10.1186/s13059-017-1382-0 Medline

59. 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 Medline

60. J. Chen, E. E. Bardes, B. J. Aronow, A. G. Jegga, ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. *Nucleic Acids Res.* **37**, W305–W311 (2009). doi:10.1093/nar/gkp427 Medline

61. J. M. Granja, M. R. Corces, S. E. Pierce, S. T. Bagdatli, H. Choudhry, H. Y. Chang, W. J. Greenleaf, Author Correction: ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. *Nat. Genet.* **53**, 935 (2021). doi:10.1038/s41588-021-00850-x Medline

62. R. K. Perez *et al.*, Multiplexed scRNA-seq reveals the cellular and genetic correlates of systemic lupus erythematosus Analysis Code (2021), doi:10.5281/zenodo.4724043.

63. 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 Medline

64. N. Rappoport, H. Paik, B. Oskotsky, R. Tor, E. Ziv, N. Zaitlen, A. J. Butte, Comparing Ethnicity-Specific Reference Intervals for Clinical Laboratory Tests from EHR Data. *J. Appl. Lab. Med.* **3**, 366–377 (2018). doi:10.1373/jalm.2018.026492 Medline

65. P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, SciPy 1.0 Contributors, SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nat. Methods* **17**, 261–272 (2020). doi:10.1038/s41592-019-0686-2 Medline

66. D. Zemmour, R. Zilionis, E. Kiner, A. M. Klein, D. Mathis, C. Benoist, Single-cell gene expression reveals a landscape of regulatory T cell phenotypes shaped by the TCR. *Nat. Immunol.* **19**, 291–301 (2018). doi:10.1038/s41590-018-0051-0 Medline

67. M. Aringer, K. Costenbader, D. Daikh, R. Brinks, M. Mosca, R. Ramsey-Goldman, J. S. Smolen, D. Wofsy, D. T. Boumpas, D. L. Kamen, D. Jayne, R. Cervera, N. Costedoat-Chalumeau, B. Diamond, D. D. Gladman, B. Hahn, F. Hiepe, S. Jacobsen, D. Khanna, K. Lerstrøm, E. Massarotti, J. McCune, G. Ruiz-Irastorza, J. Sanchez-Guerrero, M. Schneider, M. Urowitz, G. Bertsias, B. F. Hoyer, N. Leuchten, C. Tani, S. K. Tedeschi, Z. Touma, G. Schmajuk, B. Anic, F. Assan, T. M. Chan, A. E. Clarke, M. K. Crow, L. Czirják, A. Doria, W. Graninger, B. Halda-Kiss, S. Hasni, P. M. Izmirly, M. Jung, G. Kumánovics, X. Mariette, I. Padjen, J. M. Pego-Reigosa, J. Romero-Diaz, Í. Rúa-Figueroa Fernández, R. Seror, G. H. Stummvoll, Y. Tanaka, M. G. Tektonidou, C. Vasconcelos, E. M. Vital, D. J. Wallace, S. Yavuz, P. L. Meroni, M. J. Fritzler, R. Naden, T. Dörner, S. R. Johnson, 2019 European League Against Rheumatism/American College of Rheumatology classification criteria for systemic lupus erythematosus. *Arthritis Rheumatol.* **71**, 1400–1412 (2019). doi:10.1002/art.40930 Medline

68. D. D. Gladman, D. Ibañez, M. B. Urowitz, Systemic lupus erythematosus disease activity index 2000. *J. Rheumatol.* **29**, 288–291 (2002). Medline

69. S. Purcell, B. Neale, K. Todd-Brown, L. Thomas, M. A. R. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. W. de Bakker, M. J. Daly, P. C. Sham, PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007). doi:10.1086/519795 Medline

70. S. Das, L. Forer, S. Schönherr, C. Sidore, A. E. Locke, A. Kwong, S. I. Vrieze, E. Y. Chew, S. Levy, M. McGue, D. Schlessinger, D. Stambolian, P.-R. Loh, W. G. Iacono, A. Swaroop, L. J. Scott, F. Cucca, F. Kronenberg, M. Boehnke, G. R. Abecasis, C. Fuchsberger, Next-generation genotype imputation service and methods. *Nat. Genet.* **48**, 1284–1287 (2016). doi:10.1038/ng.3656 Medline

71. R. J. Pruim, R. P. Welch, S. Sanna, T. M. Teslovich, P. S. Chines, T. P. Gliedt, M. Boehnke, G. R. Abecasis, C. J. Willer, LocusZoom: Regional visualization of genome-wide association scan results. *Bioinformatics* **26**, 2336–2337 (2010). doi:10.1093/bioinformatics/btq419 Medline

72. N. Mancuso, H. Shi, P. Goddard, G. Kichaev, A. Gusev, B. Pasaniuc, Integrating Gene Expression with Summary Association Statistics to Identify Genes Associated with 30 Complex Traits. *Am. J. Hum. Genet.* **100**, 473–487 (2017). doi:10.1016/j.ajhg.2017.01.031 Medline

73. C. Wallace, Statistical testing of shared genetic control for potentially related traits. *Genet. Epidemiol.* **37**, 802–813 (2013). doi:10.1002/gepi.21765 Medline

74. 1000 Genomes Project Consortium, A global reference for human genetic variation. *Nature* **526**, 68–74 (2015). doi:10.1038/nature15393