Science
AAAS

Supplementary Materials for

Single-nucleus cross-tissue molecular reference maps toward understanding disease gene function

Gökcen Eraslan *et al*.

Corresponding authors: Ayellet V. Segrè, ayellet_segre@meei.harvard.edu; François Aguet, francois@broadinstitute.org; Orit Rozenblatt-Rosen, orit.r.rosen@gmail.com; Kristin G. Ardlie, kardlie@broadinstitute.org; Aviv Regev, aviv.regev.sc@gmail.com

Science **376**, eabl4290 (2022)
DOI: 10.1126/science.abl4290

The PDF file includes:

Other Supplementary Material for this manuscript includes the following:

Materials and Methods

Methods

Biospecimens

Donor and sample characteristics

All samples were selected from donors that were enrolled as part of the GTEx project. As previously described (*7*), all GTEx tissue samples were derived from deceased donors, with study authorization obtained via next-of-kin consent for the collection and banking of de-identified tissue samples for scientific research.

While the vast majority of tissues collected from GTEx donors were preserved in PAXgene fixative (*180*), which is not compatible for scRNA-seq, a subset of duplicate samples were collected from 8 tissue sites (Breast – Mammary tissue, Esophagus – Mucosa, Esophagus – Muscularis, Heart – Left Ventricle, Lung, Muscle – Skeletal, Prostate, and Skin – Sun Exposed (Lower leg)) and flash frozen. We selected 3-4 samples from each of the 8 tissues that satisfied the following criteria: (1) both genders and all available ancestral/ethnic groups are represented; (2) approximately span the age range collected for GTEx (21-70 years old); (3) RNA quality from the matched PAXgene-preserved tissue fulfilled GTEx requirements for bulk RNA-Seq (RIN $\geq$ 5.5); and (4) RNA-seq (from bulk matched tissue) and a donor genotype (from Whole Genome Sequencing) were available.

All GTEx samples underwent pathology review as part of that study protocol (*180*), to validate tissue origin, content, and integrity. Tissues were also reviewed for evidence of any disease (cancer, infectious disease, inflammatory disease) to confirm that collected biospecimens were "normal" or non-diseased and were acceptable for inclusion in the GTEx inventory and study. For

the samples selected for this study, we performed a second pathology review to provide greater detail about the broad cellular makeup of each tissue sample (**table S1**). The accompanying H&E images for each sample are available on the GTEx portal Histology Viewer (http://gtexportal.org/home/histologyPage).

Sample processing

Approximately 400mg of frozen tissue was obtained from each sample to allow the application of multiple nucleus extraction methods. 20-50 mg of tissue was cut per preparation. For donor pooling, ~10-15 mg of tissue from each donor were combined. Tissue remained frozen during cutting and weighing.

Each tissue sample was prepared as previously described for the EZ, CST, NST and TST protocols (*26*). Briefly, each tissue piece was subjected to nucleus isolation and snRNA-Seq, using each of the four conditions, with either commercial EZ buffer and mechanical breakdown using douncing (*19*) (EZ), or with nucleus isolation in salt-Tris buffer with detergent, either NP-40, CHAPS, or Tween-20 (NST, CST and TST receptively), with chopping to assist mechanical breakdown of the tissue (*26*).

For the CST, NST, TST isolations: On ice, each piece of frozen tissue was placed into one well of a 6-well plate with salt-Tris buffer containing 146 mM NaCl (Cat#S6546-1L, Sigma-Aldrich), 1 mM $CaCl_2$ (Cat#97062-820, VWR), 21 mM $MgCl_2$ (Cat#M1028-10X1ML, Sigma-Aldrich), 10 mM Tris pH 8.0 (CAT#AM9855G, Thermo Fisher Scientific), supplemented with detergent: CHAPS (Cat#220201-1GM, EMD Millipore) at 0.49% (w/v), Tween20 (Cat#100216-360, VWR) at 0.03% (w/v), or Nonidet™ P40 Substitute (Cat#AAJ19628AP, Fisher Scientific) at 0.02%

(w/v). Tissue was then chopped with Tungsten Carbide Straight 11.5 cm Fine Scissors (14558-11, Fine Science Tools, Foster City, CA) for 10 min on ice. Samples were then filtered through a 40 µm Falcon cell strainer (Thermo Fisher Scientific, cat. no. 08-771-1) into a 50ml conical tube. The well and filter were washed with an additional 1 ml of detergent-buffer solution, followed by a wash with 3ml of buffer without detergent. The ~ 5 ml sample volume was then transferred to a 15 ml conical tube and centrifuged for 5 minutes, 500g at 4°C in a swinging bucket centrifuge with soft break. Following centrifugation, the sample was placed on ice, supernatant was removed carefully, and the pellet was resuspended in salt-Tris buffer without detergent. The nucleus solution was then filtered through a 35 µm Falcon cell strainer (Corning, cat. no. 352235). Nuclei were counted, and 7,000 nuclei of the single-nucleus suspension were loaded onto the Chromium Chips for the Chromium Single Cell 3′ Library (V2, PN-120233) according to the manufacturer's recommendations (10x Genomics). Mouse tissues were prepared using the CST nucleus isolation protocol.

For the EZ protocol: Tissue samples were cut into pieces <0.5 cm and homogenized using a glass Dounce tissue grinder (Sigma, cat. no. D8938). The tissue was homogenized 20 times with pestle A and 20 times with pestle B in 2 ml of ice-cold nuclei EZ lysis buffer (NUC101-1KT, Sigma-Aldrich). Then, a volume of 3 ml of cold EZ lysis buffer was added, and sample was incubated on ice for 5 min. Sample was then centrifuged at 500g for 5 min at 4 °C, washed with 5 ml ice-cold EZ lysis buffer and incubated on ice for 5 min. After additional centrifugation as in the previous step, supernatant was removed and the nucleus pellet was washed with 5 ml nuclei suspension buffer (NSB; consisting of 1X PBS, 0.01% BSA and 0.1% RNase inhibitor (Clontech, cat. no. 2313A)). Isolated nuclei were resuspended in 2 ml NSB, filtered through a 35 µm cell strainer

(Corning-Falcon, cat. no. 352235) and counted. A final concentration of 1,000 nuclei per µl was used for loading on a 10x V2 channel.

snRNA-seq library preparation and sequencing

Nuclei were partitioned into Gel Beads in Emulsion (GEMs) using the GemCode instrument. Lysis and barcoded reverse transcription of RNA occurred in GEMs, followed by amplification, shearing and adaptor and sample index attachment according to the manufacturer's protocol (10x Genomics). Libraries were sequenced on an Illumina HiSeq X or NextSeq.

snRNA-seq data pre-processing

Raw sequence files were demultiplexed into the fastq format using the *cellranger mkfastq* command (*Cell Ranger* v2.1.0, 10X Genomics). A custom pre-mRNA reference genome was generated including both introns and exons using the commands recommended by 10X's Cell Ranger pipeline (https://support.10xgenomics.com/single-cell-gene-expression/software/pipelines/latest/advanced/references). The "cellranger count" command was used to generate gene expression matrices.

To remove ambient RNA (*16*, *30*, *168*, *181*), we used *CellBender* (*30*) v2's probabilistic model to generate corrected gene expression matrices after ambient removal. *CellBender* was run on a *Terra* cloud computing environment (https://app.terra.bio) on all raw gene expression matrices using the *remove-background-v2-alpha* workflow with FPR=0.01 option. The total number of nuclei identified by CellBender was 439,772.

Following ambient RNA correction, we removed "low-quality" nucleus profiles defined as those with either <200 detected genes, >5,000 detected genes, or <400 UMIs, retaining 265,831 nuclei passing QC. Protocol- and individual-specific clusters and the clusters with high predicted doublet proportions and mixed marker sets were removed manually retaining 209,126 nuclei for all subsequent analyses.

Estimation of exon:intron ratios

We computed exon:intron ratios from read counts in exonic, intronic, and intergenic regions obtained with Scrinvex (https://github.com/getzlab/scrinvex) applied to the processed BAM-file. First, a collapsed annotation file was created using the "genes.gtf" created in the Cell Ranger v2 reference using the GTEx collapse annotation script (https://github.com/broadinstitute/gtex-pipeline/tree/master/gene_model). Then, Scrinvex was run as "scrinvex ${collapsed_gtf} ${cellranger_bam} -b ${barcodes} -o {out_file} -s {summary_file}" where cellranger_bam and barcodes are output from cellranger count.

Data harmonization using disentangled conditional VAEs

We used a conditional β-TCVAE (total correlation variational autoencoder) (*182*) to obtain disentangled representation of cells, while simultaneously factoring out unwanted sources of biological and technical variation (age, sex, self-reported ethnicity, nucleus isolation protocol, and ischemic time) from the latent representation of cells via conditioning. Python source code of β-TCVAE implemented by Yann Dubois (https://github.com/YannDubs/disentangling-vae) was adapted for tabular single-cell data by adding fully connected encoder and decoder layers. Furthermore, we added the conditioning support for continuous and categorical variables.

Conditional β-TCVAE with 64 latent dimensions was applied to log(TP10K+1)-transformed counts after subsetting the genes to 10,000 highly variable protein coding genes. Age, sex, nucleus isolation protocol, ischemic time and self-reported race/ethnicity were used for conditioning. We fitted the model on the entire dataset with beta values 1 (where β-TCVAE reduces to a standard cVAE), 2, 3, 5 and 10. Higher beta values lead to over-smoothened reconstructions and higher loss of variation in the data (**fig. S3**). To minimize the loss of structure in the UMAP embeddings of nuclei and to keep the distinctness of tissue-specific neighborhoods, we used beta=2.0 as a final hyperparameter. We used a softplus output activation to produce non-negative outputs. Mean squared error (MSE) was used as a loss function. UMAP embeddings and *k*-nearest neighbors (*k*-NN) graph-based clusters were inferred using a 15-nearest neighbor graph built with the mean VAE latent space embeddings of the nuclei profiles using sc.pp.neighbors(), sc.tl.umap() and sc.tl.leiden() functions of Scanpy (*183*).

Benchmarking integration methods

The cVAE integration was compared with Harmony (*184*), MNN (*185*) and BBKNN (*186*) by several metrics. MNN (*185*) (Python implementation by Chris Kang, https://github.com/chriscainx/mnnpy/, version 0.1.9.5) and BBKNN (*186*) (https://github.com/Teichlab/bbknn, version 1.5.1) were run on 2,000 highly variable genes. The PyTorch implementation of Harmony (*184*) (https://github.com/lilab-bcb/harmony-pytorch, version 0.1.3) was run on 50 PCs that were fitted using 2,000 highly variable genes. cVAEs with 8, 64 and 256 latent variables were trained on 5,000 highly variable genes. Each method was run on the entire dataset five times with random seeds and used "channel" as a batch variable. For MNN, Harmony and cVAE, a *k*-nearest neighbor (*k*-NN) graph was built with *k*=30 neighbors. The graph-representation output of BBKNN was used. Graph-based clustering (Leiden) was

performed with resolution of 1.0 on the uncorrected data and data corrected with four methods. The sc.pp.highly_variable_genes, sc.pp.pca, sc.pp.neighbors, sc.tl.leiden functions from Scanpy were used for the identification of highly variable genes, PCA, *k*-NN graph construction, and graph-based clustering, respectively. scib Python package (*187*) was used for quantification of the integration performance via adjusted Rand index (ARI), normalized mutual information (NMI) and kBET metrics and for performing MNN and BBKNN methods. kBET was run on the gene expression space for MNN, Harmony-corrected PC space for Harmony, k-NN representation for BBKNN and the latent space for cVAE.

Our cVAE performed (specifically the variant with 64D latent space) on par with Harmony (*184*), MNN and BBKNN in conservation of cell type-specific variation based on kBET (*188*), NMI and ARI metrics (**fig. S4A**). In the runtime comparison, cVAE performed better than all methods (**fig. S4B**) and performed better than Harmony based on NMI and ARI but not as good as MNN and BBKNN. The cVAE performed comparably to Harmony (and both better than MNN and BBKNN) in removal of protocol-, channel- and individual-specific effects based on kBET, NMI and ARI metrics (**fig. S4D**). 64D cVAE obtained higher scores than 8D and 256D cVAEs (**fig. S4A-D**).

Mouse skeletal muscle snRNA-seq data pre-processing, integration, and annotation

After generating the count matrices from the raw files with Cell Ranger (v2.1.0, 10X Genomics), we removed the "low-quality" nucleus profiles defined as those with either <200 detected genes, >5,000 detected genes, or <400 UMIs, retaining 31,904 nuclei passing QC. Processed heart and esophagus mouse snRNA-seq datasets from Drokhlyansky et al. (*26*) were downloaded from https://singlecell.broadinstitute.org/single_cell/study/SCP1038/the-human-and-mouse-enteric-nervous-system-at-single-cell-resolution. Three muscle datasets were then concatenated.

Similar to the human nuclei profiles, we used a conditional β-TCVAE (*182*) to correct for mouse- or batch-specific effects. Conditional VAE with 64 latent dimensions was applied to log(TP10K+1)-transformed counts after subsetting the genes to the 10,000 highly variable protein coding genes. We used beta=2.0 as a hyperparameter with a softplus output activation for non-negative outputs. UMAP embeddings and graph-based clusters were inferred using a 15-nearest neighbor graph built with the mean VAE latent space embeddings of the nuclei profiles using sc.pp.neighbors(), sc.tl.umap() and sc.tl.leiden() functions of Scanpy. After differential expression via the sc.tl.rank_gene_groups() function, we manually annotated the clusters by comparing highly expressed genes with the literature-based marker lists.

For muscle disease gene set enrichment, we mapped mouse genes to human genes using the ortholog gene list from NCBI HomoloGene (https://ftp.ncbi.nih.gov/pub/HomoloGene/current/homologene.data). Fisher's exact test (implemented in fisher Python package) was used for the enrichment test.

Curation of reference genes from prior studies

To curate a literature-based set of reference genes (**table S3**), we first screened primary literature for cell-types found in histology samples for each targeted region. We then curated, at minimum, 3 marker genes that were identified by a primary experimental method (i.e., FISH) in the human target tissue of interest that overlap with significant, differentially expressed genes (one *vs*. rest, Welch's t-test). For cell-types that are present in multiple tissues, such as certain immune cell-types or vascular endothelial cells, we included "pan" annotation markers that may have been

found in any one tissue to confirm a nucleus's putative cell-type. Finally, we resort to murine experimental confirmation, if needed (denoted in **table S3** with murine analog gene-name).

Data-driven identification of marker genes using differential expression

Raw counts were converted to log(TP10K+1) values prior to differential expression analysis using the "sc.pp.normalize_total" and "sc.pp.log1p" functions of Scanpy. After accounting for the protocol- and sex-specific effects using ComBat (*189*) via "sc.pp.combat" function of Scanpy, we used Welch's t-test for differential expression by running "sc.tl.rank_genes_groups" function of Scanpy separately in each tissue where each broad and granular cell type was compared against the rest.

Cell type diversity analysis

Shannon entropy was calculated for each sample or channel across all N broad cell classes, according to the formula $-\sum_i(p_i \times \log_2(p_i))$, where p_i is the proportion of cells in cell class *i*. The TST, CST and NST protocols generally had comparable cell-type diversity scores in each tissue, with higher variability in skin, breast, and prostate (variance = 0.243 (skin), 0.585 (breast), 0.38 (prostate), < 0.1 (all other tissues)). Entropy was plotted per sample using the ggplot2 and cowplot R packages, and with the "quasirandom" function from the ggbeeswarm package to avoid overlap of plotted points. We used a linear mixed-effects model ("lmer" function from the lme4 R package) to check for association of entropy with the nuclei preparation protocol.

Comparison to published heart snRNA-seq

Our snRNA-seq data for heart left ventricle tissues was compared to two published snRNA-seq dataset for the heart left ventricle (*33*, *34*). For Tucker et al. (*33*), we retrieved a final, annotated Scanpy AnnData object (health_human_4chamber_map_unnormalized_V3.h5ad) and raw counts (gene_sorted-matrix.mtx) from the Broad Institute's Single Cell Portal at https://singlecell.broadinstitute.org/single_cell/study/SCP498. *CellBender*-corrected counts for the GTEx and Tucker *et al*. datasets were converted to log(TP10K+1) values using the "sc.pp.normalize_total" and "sc.pp.log1p" functions of Scanpy. For Litviňuková et al. (*34*), we subset the normalized, profiled nuclei from their "global" Scanpy AnnData object for the "LV" region available at https://www.heartcellatlas.org. Highly variable genes were selected using the parameters in "sc.pp.highly_variable_genes" "min_mean=0.0125", "max_mean=3", and "min_disp=0.5." Data were scaled using "sc.pp.scale" and highly variable genes were used for PCA via "sc.tl.pca". Data were integrated between the three cohorts using Harmony (*184*). We identified a population of ventricular myonuclei in Litviňuková *et al.* that clustered with our dataset and Tucker *et al.*'s ventricular, cytoplasmic myonuclei. We compared sample proportions as described below in *Proportion Analysis*.

Identification of contaminant transcripts from ambient RNA

We identified contaminant transcripts (from ambient RNA) by comparing the log(TP10K+1)-transformed expression values before and after the removal of ambient RNA using *CellBender* (*30*). In the comparison, we calculated the *L*2-norm of the differences between the expression values before and after ambient RNA removal for each gene separately. We used the linalg.norm function from the NumPy Python package. Norms averaged over the genes were compared

between the protocols via two-sided t-test using the scipy.stats.ttest_ind function from SciPy Python package.

By this analysis, we detected markers of epithelial cell types in breast (*e.g.*, *KRT15, KRT7*), esophagus mucosa (e.g., *S100A9, S100A8, KRT13*), lung (e.g., *SFTPC, SFTPB, SFTPA1*), prostate (e.g., *MSMB, KLK3, KLK2*), and skin (e.g., *KRT10, DST, COL7A1*) as major contributors to spurious estimates of gene expression and misidentification of cell types in those tissues. In muscle tissues, genes that were highly expressed in myonuclei were among the top contaminants (e.g., *MYL9, TAGLN, DES* in esophagus muscularis; *MYL2, TPM1, TNNC1* in heart, and *ACTA1, TPM2, TNNT1, TTN* in skeletal muscle).

Comparison of snRNA-seq and scRNA-seq data

Lung

The Travaglini et al. scRNA-seq dataset (Human Lung Cell Atlas (*40*)) was downloaded from https://www.synapse.org/#!Synapse:syn21041850/wiki/600865 as a Seurat object as described in https://github.com/krasnowlab/HLCA. Blood samples were removed, retaining 72,976 cells from 3 patients. Original annotations given in the "free_annotation" column were mapped to broad cell types used in our study according to the mapping in **table S4**. 60,095 scRNA-seq profiles were compared with 11,983 nuclei profiles from the TST protocol matching 15 shared broad cell categories: immune (alveolar macrophage), immune (macrophage), epithelial cell (ciliated), immune (NK cell), epithelial cell (alveolar type II), epithelial cell (alveolar type I), epithelial cell (basal), fibroblast, pericyte/SMC, immune (T cell), epithelial cell (club), immune (mast cell), immune (B cell), endothelial (vascular) and endothelial (lymphatic). Serous, goblet, mesothelial,

mucous, plasmacytoid dendritic and neuroendocrine cells, platelet/megakaryocytes and lipofibroblasts were found in scRNA-seq but not snRNA-seq.

Skin

Collection and experimental processing. Samples for scRNA-seq were obtained from the abdomen, from discarded excess tissue removed during abdominoplasty (n=2, IRB 2017P001913/PHS). (Samples for snRNA-seq in this study were from the left or right leg, 2cm below the patella on the medial side as described above).

For processing fresh skin, hair and fat were removed and tissue was cut into small pieces, followed by two washes with 30 mL cold PBS in a 50 mL tube. Skin was then dissociated for single cell suspension using either the Miltenyi Biotec Whole Skin Dissociation Kit, human (cat no. 130-101-540) according to manufacturer's guidelines, or with dissociation medium containing 5 mL (2%) FBS in RPMI, 100 µg / mL Liberase TM (Sigma Aldrich, cat. no. 5401127001), 100 µg / mL Dispase (Sigma Aldrich, cat. no. 4942078001), 100 µg / mL DNase I (Sigma Aldrich, cat. no. 11284932001). For this protocol, samples were incubated in a rotating 50 mL tube, at 37°C for 3 hours, with pipetting every hour. After incubation, large undigested pieces were removed, and suspension was placed in a new tube, spun down for 3 minutes at 400g, followed by a wash with cold PBS. RBC lysis was performed using 500 µL ACK (Thermo Fisher Scientific, cat. no. A1049201) on ice for 1 minute, followed by a wash with 8 mL cold PBS. Samples were then pelleted and treated with 200 µL TrypLE (Life Technologies, cat. no. 12604013) for two minutes while pipetting and washed with 1 mL 10% FBS in RPMI to quench, followed by a wash with 1 mL of cold PBS. Pellet was then resuspended in 1 mL of 0.4% BSA (Ambion, cat. no. AM2616) in PBS, and filtered through a 70 µm strainer (Falcon, cat. no. 352350). Cells were counted by

mixing 5 µl of Trypan blue (Thermo Fisher Scientific, cat. no. T10282) with 5 µl of the sample and loaded on INCYTO C-Chip Disposable Hemocytometer, Neubauer Improved (VWR, cat. no. 82030-468). Cells were loaded onto a 10x Genomics Single-Cell Chromium Controller, using V2 3' GEX kit.

Computational analysis. Sequencing files were demultiplexed and quantified using Cellranger 2.0.1. and the human reference genome package version GRCh38-1.2.0 (https://support.10xgenomics.com/single-cell-gene-expression/software/release-notes/build#grch38_1.2.0), using commands "cellranger mkfastq" and "cellranger count", respectively. We then ran CellBender as described above to remove ambient RNA and retained high quality cells (at least 200 genes detected and at most 20% of reads mapping to mitochondrial genes per cell) for a total of 27,199 cells. We performed standard processing using the Seurat R package (v3) including total sum normalization and log transformation with a multiplicative factor of 10^5 ("NormalizeData") resulting in log(TP10K+1) units, determination of variable genes ("FindVariableFeatures" using the vst method), scaling ("ScaleData"), dimensionality reduction by principal component analysis ("RunPCA", k=50) and graph-based clustering ("FindNeighbors" for k=50 and "FindClusters" at a resolution of 1). Using differential gene expression ("FindAllMarkers") and annotation by curated marker genes, we annotated 15 cell classes. Of these, 11 were shared with this study's skin snRNA-seq broad cell types: endothelial cell (lymphatic), endothelial cell (vascular), epithelial cell (basal keratinocyte), epithelial cell (suprabasal keratinocyte), melanocyte, immune (Langerhans), sweat gland cell, fibroblast, pericyte/SMC, immune (T cell) and immune (DC/macrophage). Adipocytes, cornified keratinocytes and sebaceous gland cells were unique to the snRNA-seq data, and preadipocytes were found only in the scRNA-seq data.

Prostate

Published scRNA-seq data (Henry et al. (*38*)) from three anatomical zones from coronal sections of the whole prostate separated post-cystoprostatectomy was downloaded from GEO GSE117403 (D17, D27, Pd). (SnRNA-seq in our study was derived from any representative, non-nodular region, avoiding seminal vesicles). All samples were combined and processed as described above for skin scRNA-seq, resulting in a total of 82,822 cells annotated to 11 shared cell classes: epithelial cell (basal), epithelial cell (luminal), epithelial cell (club), epithelial cell (Hillock), endothelial cell (lymphatic), endothelial cell (vascular), pericyte/SMC, neuroendocrine, immune (DC/macrophage), fibroblast, myocyte (smooth muscle). Additional epithelial subsets identified only in the scRNA-seq data were *FOXI1*$^{+}$, *SEMG1*$^{+}$, stressed and cycling cells, and only the snRNA-seq data had lymphocytes, mast and Schwann cells. The discrepancy in the immune cells may be because the scRNA-seq data was enriched for epithelial and mesenchymal cells. TST, CST+TST, and TST protocols are used in single-nucleus single-cell correlations in lung, skin, and prostate, respectively.

Mapping scRNA-seq and snRNA-seq profiles with a random forest classifier

To compare profiles between scRNA-seq and snRNA-seq datasets, a multi-class random forest classifier was trained on nuclei (cells) profiles as the training set with the function "*randomForest"* from the *randomForest* R package, and used to predict the classes of cells (nuclei) profiles with the "predict" function, with the shared broad cell classes (15 in lung, 11 in skin, and 11 in prostate) as class labels (cells/nuclei unique to only scRNA-Seq/snRNA-Seq were removed prior to analysis). The smaller of up to 70% of the cell class size or 1,000 cells were randomly sampled per class to form a training set, and genes were selected as the intersection of the top 3,000 highly

variable genes from both nuclei and cell profiles. Highly variable genes were derived using the *FindVariableFeatures* function with "vst" as the selection method. Broad cell types were accurately predicted in scRNA-seq by a classifier trained on snRNA-seq with a median accuracy of 89%, 89%, 89.3% across skin, lung, and prostate, respectively (**Fig. 2C-E, fig. S11F-H**). There was a similarly high accuracy for predicting snRNA-seq broad cell types by a classifier trained on scRNA-Seq (median 88% in skin, 90% in lung), suggesting that cell intrinsic programs are well-preserved between fresh and frozen samples and techniques. Similarly, to assess the differences in using broad vs granular labels, RF classifiers were also trained on the snRNA-seq datasets using granular labels as described above and predicted on the scRNA-seq datasets. TST, CST+TST, and TST protocols are used in these classifications in lung, skin, and prostate, respectively.

Integration of skin, prostate and lung scRNA-seq dataset with our atlas

scRNA-seq and snRNA-seq profiles were concatenated using Scanpy (*183*). log(TP10K+1)-transformed counts were used to estimate 1,000 highly variable genes using the sc.pp.highly_variable_genes() function of Scanpy. The batch_key='orig.ident' (for prostate and lung) or the flavor='seurat_v3' (for skin) option was used in the sc.pp.highly_variable_genes() function. PCA was fitted using the HVGs and samples were integrated using the PyTorch implementation of Harmony (https://github.com/lilab-bcb/harmony-pytorch) in the PC space using samples as a batch variable. A *k*-nearest neighbors (*k*-NN, *k*=15) graph was constructed using the Harmony-corrected PCs. scRNA-seq-only embeddings were generated with the same procedure as the embeddings of integrated samples.

Correlation analysis between scRNA-seq and snRNA-seq profiles

Pseudobulk expression profiles were correlated between cells and nuclei. For each broad cell type category separately, pseudobulk profiles were computed as follows: The gene (rows) by cell (column) matrix of raw counts for each sample was first scaled per-column by the total counts in each cell (nucleus) and then multiplied by 10^6 to derive counts per million (CPM) for each cell (nucleus) profile. CPMs for each gene were then averaged across cells (row-wise) for each broad cell type. The resulting expression vector per sample is the pseudobulk profile. Spearman correlation was computed for the $\log_2$-transformed pseudobulk gene expression profiles of the cell and nucleus data of interest separately for protein-coding and long non-coding RNA genes selected as follows. The total set of human genes was downloaded from the Ensembl database (GRCh38 v103, 04/09/2021), using the BiomaRt R package. Mappability scores for GRCh38, computed using the ENCODE pipeline (*190*), were obtained from GTEx and averaged across gene bodies. The complete set was then partitioned into genes with both pre-mRNA and mRNA mappability > 90%, and then partitioned into protein coding and long non-coding RNA genes based on their Ensembl biotype classification.

To identify genes that deviate from the overall expected similarity, a linear model was fit to compute residuals using the command “resid(lm(cells-nuclei ~ 0))”, where the “cells" and “nuclei” are $\log_2$-transformed pseudobulk gene expression profiles of the scRNA-seq and snRNA-seq data, respectively for either protein-coding or long non-coding RNA genes. Divergent genes were defined as those with residuals greater than the 97.5th or lesser than the 2.5th percentile, respectively.

Poly-A content was computed using the *BSgenome.Hsapiens.UCSC.hg38* R package by searching for stretches of at least 20 consecutive adenine "A" bases to qualify as one "polyA" unit. The total polyA width for a gene was defined as the total number of As in such units. The gene length was also computed using the same package.

Comparison of tissue dissociation induced-stress signature scores

Single-cell dissociation protocols have been reported to cause dissociation-induced gene expression (*41*, *42*). We scored a published dissociation signature (*41*) in snRNA-seq and scRNA-seq profiles using the "AddModuleScore" function in Seurat for each shared cell classes across snRNA-seq and scRNA-seq datasets. The score was computed on the normalized gene expression units of log(TP10K+1).

Proportion analysis

Dirichlet regression was used for proportion analysis via DirichletReg R package version 0.7.0. For analysis of broad cell type proportions, "proportions ~ protocol + tissue + Age + Sex" formula was used with common parameterization, where lung and CST were used as references for protocol and tissue categorical variables.

We applied this to compare proportions of (**1**) broad cell types, (**2**) myeloid states, (**3**) comparison between scRNA-seq and snRNA-seq, and (**4**) comparison of heart snRNA-seq studies. For myeloid proportions, the formula "proportions ~ tissue + protocol | protocol" was used with alternative parameterization to correct both mean and precision estimates for the protocol effects. We fitted eight models, each with a different tissue used as a reference. For comparing proportions between protocols, the formula "counts ~ protocol | protocol" was used with the alternative

parameterization. For comparison of scRNA-seq *vs*. snRNA-seq, scRNA-seq was used as the reference. For comparison between snRNA-seq datasets in heart, this study's data was used as the reference. Samples with fewer than 30 nuclei profiles were excluded from the proportion models. In the comparison of proportions among heart snRNA-seq studies, data from this study's EZ protocol, Litviňuková et al. (*34*) and Tucker et al. (*33*) showed a higher proportion of muscle cells in comparison to the CST, NST and TST protocols (Dirichlet regression, LRT with CST as baseline: Benjamini-Hochberg FDR=2.5×10^{-4} (EZ), 0.002 (Tucker et al.) , 6.3×10^{-4} (Litviňuková et al.)), and lower proportions of endothelial cells (adj. P=0.002 (Tucker et al.), 4×10^{-10}, (Litviňuková et al.)). The immune and adipose compartments were comparable in proportions (Benjamini-Hochberg FDR>0.1) across protocols and studies, however, Litviňuková et al. (*34*) had lower proportions of fibroblasts (Benjamini-Hochberg FDR= 0.1). Taken together, all four protocols retain expected cell groups shared with published protocols.

Normalization of proportions

To account for the variation in total number of nuclei profiled in each tissue (*e.g.,* 5,327 in skin and 36,574 in heart) in **Figs. S2D** and **S25F,** where the proportions of nuclei from each tissue for each cell type are visualized, we normalized each proportion by the total number of nuclei profiled in that tissue. To visualize the proportion of nuclei from tissue q for the cell type p, instead of $N_{p,q}/\sum_i N_{p,i}$ where $N_{x,y}$ represents number of nuclei for cell type x in tissue y, we used $(N_{p,q}/\sum_i N_{i,q}) / (\sum_j (N_{p,j}/\sum_i N_{i,j}))$ formula. Therefore, these visualizations approximately show what the proportions would be, if the same number of nuclei were recovered in each tissue.

Co-embedding of bulk and snRNA-seq pseudobulk profiles

GTEx v8 bulk RNA-seq expression data were downloaded from the GTEx portal https://gtexportal.org, converted into an AnnData object and concatenated with the pseudobulk profiles of the nuclei profiled in this study. Channels with fewer than 30 nuclei were not used. The number of highly variable genes (calculated only on the pseudobulk data), the number of PCs, the number of neighbors, *k*, used in the *k*-NN graph and the similarity metric used in the *k*-NN graph were determined by the Tree of Parzen Estimators (TPE) random hyperparameter search method implemented in hyperopt Python package (v0.2.5) (*191*). The accuracy metric that was optimized in the search was defined as the fraction of bulk neighbors of pseudobulk data points that are from the same tissue site. Final parameters used were 3,522 HVGs, 70 PCs, 42 nearest neighbors and $L1$ as the *k*-NN distance metric. Bulk and pseudobulk samples were integrated using the PyTorch implementation of harmony (v0.1.5, https://github.com/lilab-bcb/harmony-pytorch) (*184*) in each iteration.

Sample embeddings based on cell type compositions

Broad cell type proportions in each profiled sample were used to construct an *k*-mutual nearest neighbors graph of channels in Scanpy. We used k=7 and Fruchterman-Reingold layout where connected components of the graph were considered clusters.

Preprocessing and visualization of myeloid nuclei profiles

We used log(TP10K+1)-transformed counts to estimate 2,000 highly variable genes. PCA was fitted using the HVGs and a *k*-nearest neighbors (*k*-NN, k=15) graph was constructed using 50 PCs. We integrated all myeloid nuclei profiles from eight tissues using the PyTorch implementation of Harmony (*184*) (https://github.com/lilab-bcb/harmony-pytorch) to correct for

individual- and protocol-specific variation in the PC space. A PAGA (*192*) graph was inferred based on the myeloid annotations and UMAP visualization was performed with PAGA initialization. Proportion bar plots and dotplots were generated using the Python implementation of the ggplot framework, plotnine version 0.7.

Diffusion map representation of monocyte and macrophage populations

Myeloid nuclei profiles were subset to MΦ *HLA*II^{high}, MΦ *LYVE1*^{high}, Mo/MΦ *FCGR3A*^{high}, Mo/MΦ *FCGR3A*^{low}, *CD14*$^{+}$ monocyte and *CD16*$^{+}$ monocyte states. Next, log(TP10K+1)-transformed counts were used to estimate 1,000 highly variable genes. PCA was fitted using the HVGs and a *k*-nearest neighbors (*k*-NN, *k*=15) graph was built using 50 PCs. We integrated all the nuclei profiles from eight tissues using the PyTorch implementation of Harmony (*184*) (https://github.com/lilab-bcb/harmony-pytorch) to correct for individual- and protocol-specific variation in the PC space, using protocols and individuals as batch variables. Finally, sc.tl.diffmap() function from Scanpy (*183*) was used for the diffusion map inference.

Classification of lipid-associated macrophages

We used a published annotated adipose scRNA-seq study (*51*) to train a logistic classifier that can classify profiles into three groups: LAMs, non-LAM macrophages and other cell types. We performed 5-fold stratified cross-validation that maximizes weighted F1-score, using the LogisticRegressionCV function from the scikit-learn Python package to optimize the *L*2-regularization parameter. To take highly imbalanced class frequencies, we employed a weighting scheme, where rarer classes (*e.g.*, LAMs) contributed more to the overall loss function. log(TP10K+1)-transformed expression values of 17,612 protein coding genes were used in the training.

Preprocessing of the external datasets used in LAM classification

Datasets with IDs GSE115469, GSE117403, GSE127246, GSE128518, GSE153643, GSE131685, GSE131778, GSE131886, GSE140393, GSE143380, GSE143704, GSE144085, GSE150672, GSE153760, GSE156776, GSE159677 were downloaded from Gene Expression Omnibus (GEO). Placenta and decidua datasets were downloaded from EBI Single Cell Expression Atlas (https://www.ebi.ac.uk/gxa/sc/home) using the E-MTAB-6701 accession via the "sc.datasets.ebi_expression_atlas" function of the Scanpy Python package. Human dataset of the enteric nervous system was downloaded from the Single Cell Portal (URL: https://singlecell.broadinstitute.org/single_cell/study/SCP1038). log(TP10K+1)-transformed counts were used for classification.

Gene set enrichment of LAM-like cell markers

LAM-like profiles were compared to other myeloid profiles by ranking genes by their mean z-scored log(TP10K+1) expression values. Top 80 genes were then evaluated for GO term enrichment using the "sc.queries.enrich" function of Scanpy (*183*) after excluding genes from the ferritin (*FTL, FTH1*) and metallothionein (*e.g.*, *MT1A, MT2A*) families. GO terms with more than 1,000 genes were not shown to highlight more specific terms.

Consensus LAM markers

LAMs were compared to other macrophages within each of the external datasets using sc.tl.rank_genes_groups() function of Scanpy. Genes with $\log_2$ fold-change > 0.5 and FDR < 0.1 were retained as significant markers within each dataset. Next, genes were ranked by the number of datasets where a gene was among the top 100 significant markers and the top 30 genes were

retained as consensus markers. GWAS where the genome-wide significant variants are mapped to the consensus LAM markers with Open Targets Genetics Locus-to-Gene score greater than 0.5 were shown in **fig. S17H**. Open Targets Locus-to-Gene scores of all GWAS variants were downloaded as described in the *Variant-to-gene mapping using the Open Targets Genetics API* section.

TF activity analysis of LAMs

The Python implementation of Dorothea (*193*, *194*) (https://github.com/saezlab/dorothea-py, version 1.0.5) was used to calculate the activity scores of 292 transcription factors, in the A, B and C confidence categories of Dorothea, for each cell. dorothea.run() function was run with "adata, regulons, center=True, num_perm=0, norm=True, scale=True, use_raw=False, min_size=5) arguments. After extracting TF scores using the dorothea.extract() function, LAMs and Mϕ groups were compared using Welch's t-test implemented in the sc.tl.rank_genes_groups() function of Scanpy. P-values (**fig. S17G**) were calculated by combining the p-values shown for individual studies using Fisher's method implemented in the scipy.stats.combine_pvalues() function of SciPy Python package.

Differential expression of MΦ *LYVE1*^{hi} *vs*. MΦ *HLA*II^{hi} states

To compare MΦ *LYVE1*^{hi} *vs* MΦ *HLA*II^{hi} states in heart and lung, a Wald test with a negative binomial regression model was used with the formula "~ 1 + cell_state + participant_id + protocol + log_number_of_genes'" on raw counts. In esophagus mucosa, participant_id and protocol covariates were not used since most MΦ *HLA*II^{hi} nuclei were detected in the GTEX-15SB6 CST sample. Benjamini-Hochberg FDR<0.05, $\log_2$(fold-change)>1.0, fraction of nuclei expressing a

gene (alpha)>5% cutoffs were used for significance. Top 100 markers given in **Table S4** and **S5** from Chakarov *et al.* (*47*) were used for the comparison of mouse and human macrophage markers.

Tissue-shared and tissue-enriched signatures of fibroblasts

Log(TP10K+1)-transformed counts of fibroblast nuclei profiles across eight tissues were used to estimate 2,000 highly variable genes. PCA was fitted using the HVGs and a *k*-nearest neighbors (*k*-NN, *k*=15) graph was built using 50 PCs. We integrated all the nuclei profiles from eight tissues using the PyTorch implementation of Harmony (*184*) (https://github.com/lilab-bcb/harmony-pytorch) to correct for individual-specific variation in the PC space using individuals as batch variables. UMAP representations were generated using the sc.tl.umap() function from Scanpy (*183*). Pseudobulk approach was used for the differential expression between the tissues where each channel with more than 30 nuclei profiles was collapsed into a single observation by taking the mean expression across all nuclei within each channel. Welch's t-test was used for comparing tissues to each other in a one-*vs*-rest manner via the sc.tl.rank_genes_groups() function in Scanpy (*183*). A common fibroblast signature was obtained by identifying significantly differentially expressed genes ($\log_2$FC > 1, FDR < 0.1) in fibroblasts within each tissue (as described in the *Data-driven identification of marker genes using differential expression* section). After selecting genes that are identified in at least seven tissues, 15 genes with highest mean z-score-normalized expression across tissues were picked.

The lung-enriched marker genes shown in **Fig. 4B** were used as a seed to find a larger set (module) of co-varying genes to highlight potential tissue-specific functions of lung fibroblasts. Using the same approach described in the section *Module-based GWAS enrichment*, gene modules were constructed by hierarchically clustering genes expressed by fibroblasts across eight tissues with

different hierarchy levels used as cutoffs to obtain modules of varying scopes. Instead of testing for the enrichment in GWAS genes as in **Fig. 7**, the inferred gene modules were tested for enrichment in lung-enriched fibroblast marker genes (*i.e.*, *NPNT, LIMCH1, ITGA8, FRAS1, HMCN1, PIEZO2, ADAMTS12, GPC3, FGFR4*). The most significantly enriched (FDR= 8.98×10^{-11}, $\log_2$FC: 8.45) gene module with fewer than 100 genes (n=68) was used in the gene set analysis (see **table S8** for the full list of genes). sc.queries.enrich() function in Scanpy was used for the gene set enrichment analysis only with the GO:BP, GO:CC and KEGG gene set sources.

GWAS where *NPNT* is associated with a genome-wide significant variant with Open Targets Locus-to-Gene score higher than 0.5 were shown in **Fig. 4H**. Open Targets Locus-to-Gene scores of all GWAS variants were downloaded as described in the section *Variant-to-gene mapping using the Open Targets Genetics API*.

Topic modeling of OMIM diseases

Metadata for 5,812 monogenic diseases including OMIM numbers and text descriptions were downloaded from OMIM (*116*) through the API (https://www.omim.org/api). Symbols of the disease-associated genes were downloaded from https://www.omim.org/static/omim/data/mim2gene.txt. Only the phenotype entries with a prefix of number sign (#) were retained and the rest were discarded. Text descriptions were preprocessed and tokenized, where words were converted to lowercase and split by whitespace after '?', '.', ',', ':', ';', "", """, '(', ')' characters were removed, yielding 42,654 tokens. Stochastic block modeling-based hierarchical topic modeling (*195*), (https://github.com/martingerlach/hSBM_Topicmodel) was run on the preprocessed text descriptions of the diseases with n_min=2 and and B_min=200 parameters in make_graph() and fit() functions, respectively. Topic models with four hierarchical

levels were obtained after the inference with differing number topics in each layer (n=10, n=53, n=229, n=941 in four layers), where the output consisted of two matrices: **(1)** a disease (which typically has a single gene association but in some cases more than one) by topic (P(topic | disease)) and **(2)** a topic by word (P(word | topic)) matrix in each layer. The former was used in gene set enrichment analysis (below) and the latter is used for the interpretation of topics.

For gene set enrichment analysis (**Fig. 5A**), the third layer with 229 topics was used to obtain granular topics. Level four of the hierarchy consisted of several singleton topics (with a single disease or word with probability of 1.0) and therefore not preferred. First, diseases with high loadings within each topic were identified based on a cutoff of the 99th percentile of the distribution of probabilities (P(topic | disease)) across diseases within each topic. If a disease with a probability higher or equal to the cutoff was associated with multiple genes, all genes associated with it were included in the gene sets. The size distribution of gene sets (n=229) had minimum, median and maximum of 58, 64 and 164 genes, respectively. Second, Fisher's exact test (fisher Python package, version 0.1.9) was used to test whether genes associated with diseases with high loadings in each topic were enriched in broad cell type markers (as described in the *Data-driven identification of marker genes using differential expression* section). Integer identifiers of topics (topic numbers) with a comma-separated list of five words with highest probabilities in each topic were used as topic labels in the figures (*e.g.,* 155:ciliary,arms,cilia,outer,bladder). Diseases and genes shown in **fig. S22** were selected based on z-score normalized log(TP10K+1) expression values, where the cell types with a mean normalized z-score greater than 1.0 are shown.

Receptor-ligand analysis

CellPhoneDB version v2.1.4 (*144*) was used for receptor-ligand analysis of granular cell types within each tissue. "cellphonedb method statistical_analysis" command was used for the analysis of each tissues with the arguments "meta.tsv counts.tsv --counts-data hgnc_symbol --project-name tissue --threads 30 --subsampling --subsampling-num-cells 50000 --subsampling-log false". Receptor-ligand plots (**Fig. 5)** were created with the ggraph R package using the "sugiyama" graph layout. Granular cell type names were converted to broad cell type labels in **Fig. 5** for clarity.

For the cell-cell interactions of dichotomous macrophage populations and other cell types in **fig. S16F-G**, receptor-ligand pairs included in CellPhoneDB v2.1.4 (*144*) were used as a source of potential interactions. First, scaled mean expression (z-score, log(TP10K+1)) of receptor and ligands within all myeloids were calculated for each granular myeloid subset. Second, scaled mean expression (z-score, log(TP10K+1)) of receptor and ligands within the entire dataset was calculated for each broad cell type. After ranking the receptor-ligand pairs based on the mean of *LYVE1*^{hi} (**fig. S16F**) or *HLAII*^{hi} (**fig. 16G**) and broad cell types, six and five pairs were highlighted in **fig. S16** for *LYVE1*^{hi} and *HLAII*^{hi} populations, respectively.

For the cell-cell interactions of fibroblasts and neurons in **fig. S18E**, the same receptor-ligand pairs were used. Scaled mean expression (z-score, log(TP10K+1)) of all granular fibroblast subsets across 8 tissues and scaled mean expression (z-score, log(TP10K+1)) of all broad cell types across 8 tissues were calculated. Finally, pairs were ranked by the mean expression of esophagus mucosa fibroblasts and esophagus muscularis neurons. Three pairs were highlighted in the figure.

Gene set enrichment of monogenic muscle disease genes

Monogenic muscle disease names, groups, subgroups and associated genes were downloaded from https://musclegenetable.fr (*118*). The resource included 928 diseases with associations to 605 unique genes. Disease genes shown in **fig. S24** are 64 genes associated with monogenic diseases in "Muscular dystrophies" and "Congenital muscular dystrophies" disease groups.

Next, the enrichment of data-driven broad and granular cell type markers ($\log_2$ fold-change > 1.0, FDR < 5%, see *Data-driven identification of marker genes using differential expression* section) in muscle disease groups and subgroups was tested using Fisher's exact test implemented by the "fisher" Python package.

Differential expression analysis of myonuclei subsets and myocyte subtypes

Gene expression profiles of regular and cytoplasmic myonuclei were compared using Welch's t-test via the sc.tl.rank_genes_groups() function of Scanpy. Similarly, myonuclei profiles of fast and slow myocytes were compared within regular and cytoplasmic subsets using Welch's t-test and the sc.tl.rank_genes_groups() function of Scanpy. In the slow *vs.* fast myocyte differential expression volcano plots (**fig. S9**), genes within [-10, 10] $\log_2$(fold-change) are shown, since the genes beyond this range were expressed only in a small fraction of cells (low alpha). Hereditary cardiomyopathy disease group was not used in the slow *vs.* fast myocyte differential expression analysis. Significantly differentially expressed genes ($|\log_2(\text{fold-change})|>0.5$ and FDR<5%) were tested for enrichment in muscle disease groups using Fisher's exact test implemented in fisher Python package.

Differential expression analysis of bulk RNA-seq profiles of muscle samples

RPKM-normalized bulk RNA-seq gene expression matrix of muscle disease patient samples was received from the authors (*122*). Patient metadata, including the clinical diagnoses, were obtained from **Table S1** in Cummings et al. (*122*). RPKM-normalized GTEx v6 expression profiles, which were also used in the Cummings et al. as control samples, were downloaded from the GTEx Portal (GTEx_Analysis_v6_RNA-seq_RNA-SeQCv1.1.8_gene_rpkm.gct.gz). Healthy controls given in **Table S4** in Cummings et al. were subsampled to 20. After concatenating the datasets, RPKM values were converted to log(TPM+1) using sc.pp.normalize_total(adata, target_sum=1e6) and sc.pp.log1p(adata) functions of Scanpy and z-score normalized for visualization and differential expression. Welch's t-test implemented by sc.tl.rank_genes_groups() function of Scanpy was used for the differential expression where the sample profiles from the patients in each clinical diagnosis group were compared with the healthy GTEx v6 muscle samples.

Linking snRNA-seq to genetic variation using GWAS/eQTL enrichment analysis

To test whether the expression of genes in GWAS loci associated with a given complex disease or trait is enriched in specific cell types more than expected by chance, ECLIPSER (*152*) was applied to compare the cell type specific expression of genes mapped to known GWAS loci for a complex trait of interest to a background distribution of GWAS loci. GWAS variant associations were obtained from Open Targets Genetics (*153*, *154*) from the NHGRI-EBI GWAS catalog and UK Biobank GWAS. Only GWAS with at least part of their samples from European ancestry and genome-wide significant associations ($P<5x10^{-8}$) were considered. 23 traits were selected (**table S14**) based on pathophysiology related to one of the 8 tissues in this study. Different GWAS that mapped to the same trait were considered together for the given trait (**table S14**). For each tissue, a background of GWAS loci (null set) was defined as the GWAS variants for all traits excluding

the set of traits selected for the particular tissue (on average 71,411 variants). The enrichment analysis consisted of three consecutive steps, as follows:

Gene mapping. Genes were mapped to trait-associated loci for the selected and background (null) traits, using the 95% credible set of fine-mapped *cis*-eQTLs and *cis*-sQTLs from each of the 49 GTEx tissues (v8) computed using DAP-G (*7*, *151*, *196*) and the European subset of GTEx samples (*7*, *151*, *196*) . Specifically, all variants in LD (r^2>=0.8) with each of the GWAS variants were identified using the GTEx whole genome sequencing variant calls as the reference panel (*7*) and PLINK 2.0 (Plink --bfile 1KG_chr_files --r2 --ld-snp-list variant_list_file --ld-window-kb 5000 --ld-window-r2 0.8 --ld-window 99999). If a GWAS variant was not present in the GTEx samples, LD proxies for the variant were searched in the 1000 Genomes Project panel (*197*) at r^2>0.8, and these proxies were subsequently checked for LD variants in the GTEx panel. GWAS associations whose variant or LD proxy variants were significant eVariants or sVariants (FDR<0.05) in any of the 49 GTEx tissues were assigned the corresponding eGene/s and sGene/s to their locus. We further included genes mapped to GWAS variants based on the 'bestLocus2Genes' mapping from Open Targets Genetics, which included additional omic data (e.g., Hi-C) and predicted deleterious protein coding variants in LD with the GWAS variant (*153*, *154*). To avoid inflation of enrichment due to LD between GWAS variants, GWAS variants that shared LD proxy variants, or eGenes or sGenes were collapsed into a single locus. This was done separately for the GWAS variants for each selected trait and for all null traits per tissue. On average, 40% of the null variant sets and 80% of the 23 selected traits had at least one mapped gene, and of the mapped loci, on average 2 genes mapped per locus, ranging from 1-37 for the selected traits and 1-170 for the null traits.

Locus scoring. For each combination of trait of interest, tissue, and cell type, each GWAS locus was first scored based on the fraction of $\log_2$ fold-change > 0.5 and FDR < 0.1 of all genes mapped to the locus.

Cell type specificity significance estimate. The significance of the cell type specificity scores of a GWAS locus set for each cell type was assessed against the distribution of values of the background GWAS loci using a Bayesian Fisher's exact test. Specifically, the counts of GWAS loci with their scores greater than the 95th percentile of scores from the background loci for a given cell type were modeled as Binomial distributions, with the parameters (θ_1, n_1) and (θ_2, n_2) for the GWAS locus set and background loci, respectively, where n_1 and n_2 are the total number of loci in the GWAS locus set and background, respectively. Uninformative uniform priors were specified for θ_1 and θ_2, leading to the conjugate Beta distributed posteriors. Next, fold enrichment was defined as the ratio of θ_1 and θ_2, and the 95% credible interval was constructed from 10^6 Monte Carlo draws from the posteriors. Of 23 selected traits, 21 had 5 or more GWAS loci with at least one mapped gene, which were analyzed for cell type specificity enrichment (**table S14**). Multiple hypothesis correction was applied tissue-wide (correcting for all cell types tested per tissue and trait) and experiment-wide (correcting for all traits by cell types and tissues tested) using the Benjamini-Hochberg FDR. In the significantly enriched cell types, cell type specific genes mapped to GWAS loci with cell type specificity scores above the 95th percentile of the background loci scores are suggested to be causal genes for the given trait in the enriched cell type.

Enrichment of heart-specific and tissue shared, myonuclei-specific genes that drove the enrichment signal of atrial fibrillation GWAS loci in heart myonuclei, was tested using GeneEnrich (*150*) based on the hypergeometric distribution, and Gene Ontology terms (**tables**

S15-16). Genes that were myonuclei-specific in heart and at least one of the other tissue/cell combinations (skeletal muscle/myonuclei (sk. muscle), esophagus muscularis/myonuclei (smooth muscle), or prostate/myonuclei (smooth muscle)) were considered tissue shared genes in the analysis.

Cell type specificity enrichment for Atrial fibrillation GWAS loci in GTEx heart tissue was tested for replication in two separate heart left ventricle snRNA-seq studies (*33*, *34*). Data was processed as described above and differential gene expression was computed similarly to that performed for GTEx tissues, correcting for donor in Tucker *et al.* (*33*), and sequencing center, age and sex in Litviňuková *et al.* (*34*), using Combat (*189*). Only genes expressed in at least 5% of cells in each given cell type were considered in these studies for cell type specificity scoring.

Variant-to-gene mapping using the Open Targets Genetics API

Publicly available JSON files (v20022712) and the GraphQL API (v20.02.07) of the Open Targets Genetics (OTG) portal were used to obtain genes mapped to the independent GWAS loci in **Figs. 6** and **7**. For study level information (*e.g.*, study IDs, number of individuals, number of significant loci), JSON files were downloaded from https://ftp.ebi.ac.uk//pub/databases/opentargets/genetics/20022712/lut/study-index. For variant-to-gene mapping and variant-level details, such as Locus-to-Gene scores, the manhattan() and studyInfo() functions were used through the GraphQL API endpoint https://api.genetics.opentargets.org/graphql using the sgqlc Python package. The manhattan() function provides the list of all significant and independent lead SNPs as well as the genes associated with them using the Locus-to-Gene scoring model for the studies stored in the OTG portal. Closest protein coding genes were used in cases where Locus-to-Gene score was not

available. GWAS with fewer than two significant loci or fewer than 3,000 individuals were excluded. UK BioBank (UKBB) traits containing "None of the above" were also removed. For the remaining studies, only the largest GWAS (based on the nInitial field of the study) was considered for a given phenotype, resulting in 4,062 studies.

Module-based GWAS enrichment

To infer gene modules enriched with GWAS risk genes and the cell types expressing these modules, genes were first hierarchical clustered with the complete linkage method and a correlation distance, i.e. $dist(g_1, g_2) = 1 - r(g_1, g_2)$ where r is the Pearson correlation coefficient between genes g_1 and g_2 , calculated with the scipy.cluster.hierarchy.linkage function from the scipy Python package. Models were fit separately in each tissue using all protein coding genes. To speed up the inference, clustering was performed in PC space, where gene loadings for 100 PCs were taken into account in the distance calculations. Gene modules were obtained at different resolutions by cutting the linkage tree at 100 different levels starting from only two clusters (*i.e.*, modules) to a highly granular level, where the number of clusters is equal to half of the number of genes. In a post-processing step, modules that were exactly the same or had fewer than three genes were removed. Next we scored all cells using each module as a signature to quantify average expression of the modules using the scanpy.tl.score_genes() function. Finally, gene module enrichment with the GWAS risk genes was estimated by testing all modules against all GWAS phenotypes using Fisher's exact test implemented in "fisher" Python package. Final cell type enrichment score was defined as the product of the gene set overlap metric (f-score) and the signature score (module expression). To prioritize the associations with high expression and high gene set overlap we used additional cutoffs such as at least four genes in the intersection of the GWAS risk genes and the module genes and enrichment score higher than 0.15.

Preprocessing and demultiplexing of pooled samples

For genotype-based demultiplexing and doublet detection, souporcell (*168*) was used as available in a Docker image from Cumulus (*198*) (Cumulus version 2020.03, souporcell version eeddcde). Vartrix v1.1.20 (https://github.com/10XGenomics/vartrix) was used instead of the older version included in the Docker image. Souporcell was applied to four samples with the following command line arguments "-t 32 -o outputdir --min_alt 10 --min_ref 10 --restarts 100 --common_variants common_variants_grch38.vcf -k 3". Using the unpooled lung and prostate samples from the same individuals, an expression-based multinomial logistic classifier was trained to predict the individuals. The "LogisticRegression(max_iter=500, penalty='l2', solver='liblinear', C=0.001, class_weight='balanced')" function from the scikit-learn Python package (v0.24.1) was used for the training. 0.6 probability cutoff was used to assign each nucleus to individuals. scrublet Python package for doublet detection was used to compare to the souporcell doublet detection method.

Supplementary Text

Note S1: Best practices in data integration methods

Our observations in the cross-tissue setting suggest that VAEs strike a balance between the need to successfully correct for batch effects, while preserving biological variation. Many other methods tend to overcorrect (*e.g.*, smoothen) the biological signal. For example, Harmony removes complex batch effects (*e.g.*, across many samples), but sometimes at the loss of biological information (*187*). Conversely, MNN and BBKNN are better at conserving biological signals, but less efficient at removing complex effects, such as the individual-specific effects present in our dataset. We recommend using both VAE-based integration methods and Harmony for complex integration tasks, followed by a thorough investigation of overcorrection for the final choice of the integration method. For a comprehensive comparison of methods and best practices, we also refer the reader to Lucken et al. (*187*).

Note S2: Cytoplasmic myonuclei

Cytoplasmic myonuclei were first described in Tucker et al. (*33*) and have high myoglobin (MB) expression and a high ratio of exon- *vs*. intron-mapping reads (*28*, *33*). While "cytoplasmic myonuclei" were initially speculated to be possible contaminants (*33*), we detected their profiles in three independent studies using six different snRNA-seq protocols (*33*, *34*), including after ambient RNA removal in our study (**fig. S1**). To further determine their distinguishing characteristics, we compared their expression profile to those of classical myonuclei, identifying significantly differentially expressed genes in each subset (*33*), with those from the cytoplasmic myonuclei showing high overlap with the markers from Tucker et al. (**fig. S8**). We further observed classical and cytoplasmic myonuclei in skeletal muscle and in mouse skeletal muscle. Overall, this supports the "cytoplasmic myonuclei" as a distinct subset that merits further dedicated investigation in future dedicated studies.

Note S3: Comparison of extraction solutions

In our comparisons, TST yielded the highest cell type diversity on average across tissues (**Fig. 2A**) and significantly higher proportions of T cells, fibroblasts, and vascular endothelial cells (FDR<10%, fig. S11A). EZ yielded significantly lower proportions for most cell types (**fig. S11A**). NST yielded significantly higher proportions of sweat gland cells and cardiac myonuclei (FDR<10%, **fig. S11A**). Overall, the EZ protocol performed least well for all metrics and across all tissues, while the other protocols performed comparably overall, but differently depending on the specific metric. Since protocols vary by their performance based on the desired goal (most diverse, high capture of

particular desired cell types), users should choose protocols by their detailed features based on their scientific goal, tissue type and complexity, and further protocol optimization may still be required. For additional comparisons and help with such choices, including in disease tissue, we also refer the interested reader to Slyper et al. (*29*).

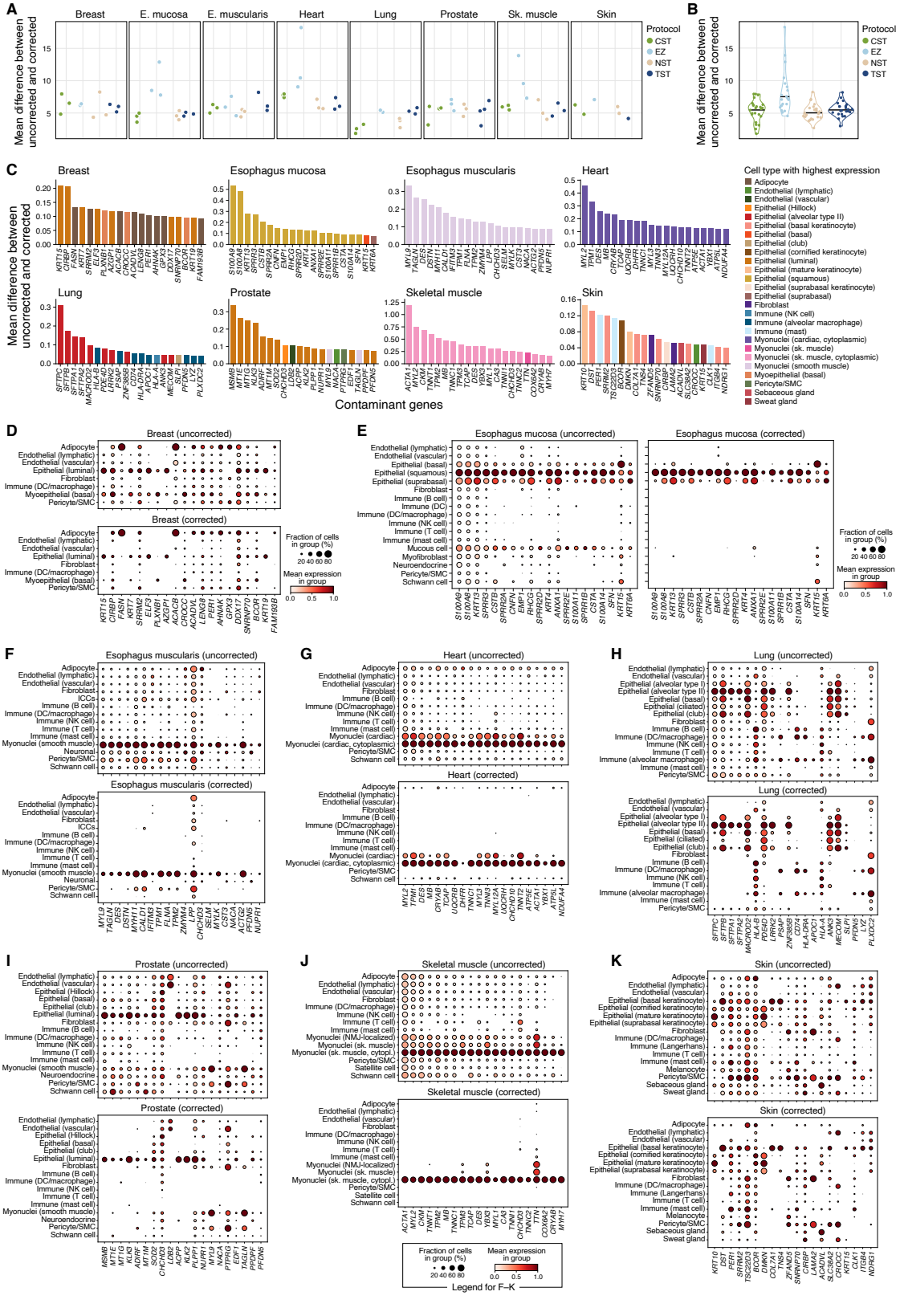

A
Breast
E. mucosa
E. muscularis
Heart
Lung
Prostate
Sk. muscle
Skin
Mean difference between uncorrected and corrected
Protocol
CST
EZ
NST
TST
B
C
Esophagus mucosa
Esophagus muscularis
Skeletal muscle
Cell type with highest expression
Adipocyte
Endothelial (lymphatic)
Endothelial (vascular)
Epithelial (Hillock)
Epithelial (alveolar type II)
Epithelial (basal keratinocyte)
Epithelial (basal)
Epithelial (club)
Epithelial (cornified keratinocyte)
Epithelial (luminal)
Epithelial (mature keratinocyte)
Epithelial (squamous)
Epithelial (suprabasal keratinocyte)
Epithelial (suprabasal)
Fibroblast
Immune (NK cell)
Immune (alveolar macrophage)
Immune (mast)
Myonuclei (cardiac, cytoplasmic)
Myonuclei (sk. muscle)
Myonuclei (sk. muscle, cytoplasmic)
Myonuclei (smooth muscle)
Myoepithelial (basal)
Pericyte/SMC
Sebaceous gland
Sweat gland
Contaminant genes
D
Breast (uncorrected)
Breast (corrected)
Fraction of cells in group (%)
Mean expression in group
E
Esophagus mucosa (uncorrected)
Esophagus mucosa (corrected)
F
Esophagus muscularis (uncorrected)
Esophagus muscularis (corrected)
G
Heart (uncorrected)
Heart (corrected)
H
Lung (uncorrected)
Lung (corrected)
I
Prostate (uncorrected)
Prostate (corrected)
J
Skeletal muscle (uncorrected)
Skeletal muscle (corrected)
K
Skin (uncorrected)
Skin (corrected)
Legend for F–K

Fig. S1. Ambient RNA correction improves cell type specificity of gene markers. (**A, B**) Ambient contamination level. Average contamination level (*y*-axis, *L*2-norm of the difference between uncorrected and corrected log(TP10K+1)-transformed expression levels averaged across genes) for each protocol and tissue **(A)** or for all samples from one protocol **(B)**. Horizontal bar in violin plots: median. **(C–K)** Potential top sources of the ambient RNA. (**C**) Mean difference between corrected and uncorrected expression values (*y*-axis) for the top genes (*x*-axis) in each tissue, colored by their cell type of highest expression in the tissue (color legends). **(D–K)** Mean expression (circle color) and fraction of expressing cells (circle size) for each of the contaminant genes from (C) (columns) in each cell type (rows) in each tissue before and after the correction for ambient RNA (as labeled on top).

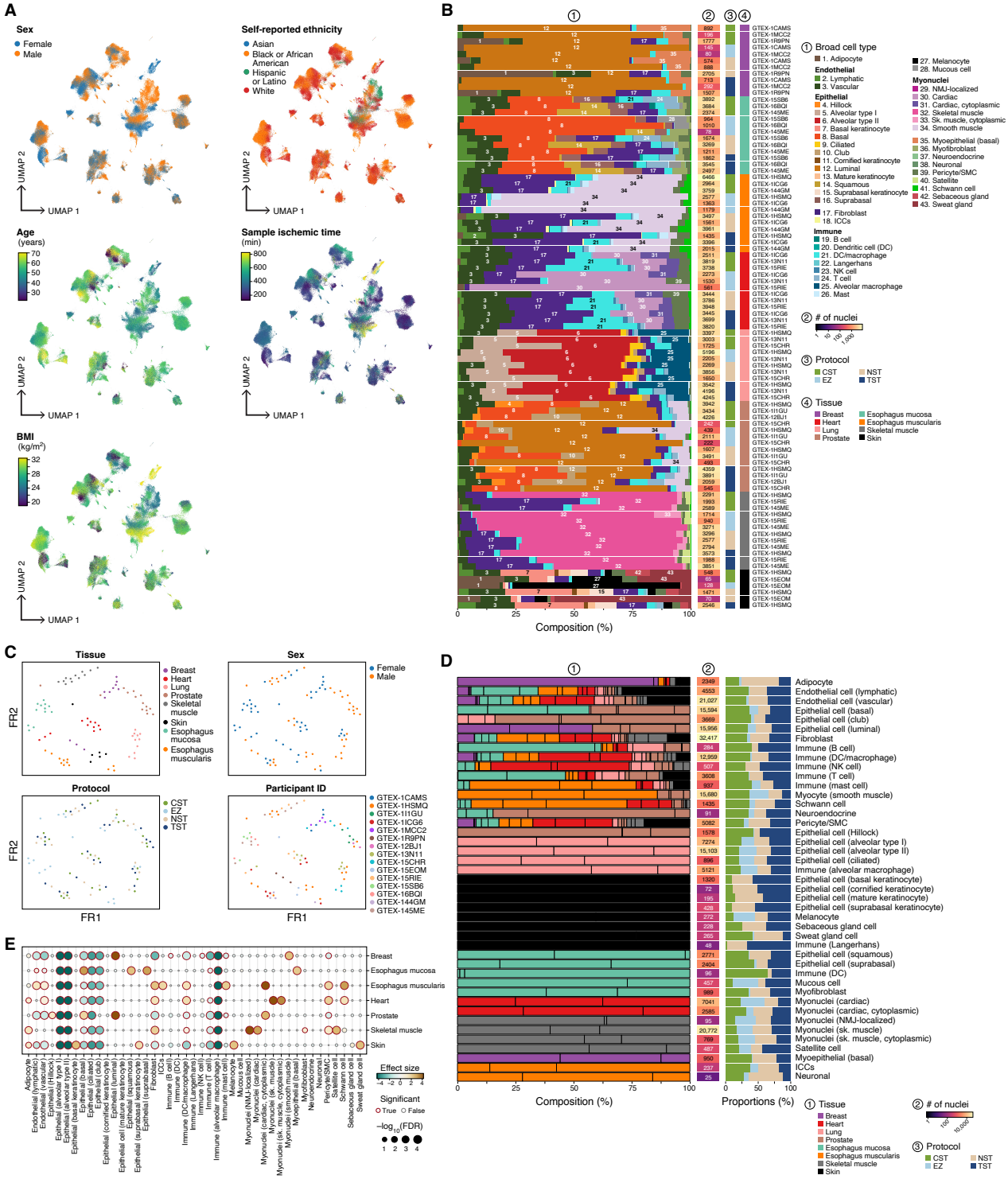

A
Sex
Female
Male
Self-reported ethnicity
Asian
Black or African American
Hispanic or Latino
White
Age (years)
Sample ischemic time (min)
BMI (kg/m²)
UMAP 1
UMAP 2
B
Composition (%)
① Broad cell type
1. Adipocyte
Endothelial
2. Lymphatic
3. Vascular
Epithelial
4. Hillock
5. Alveolar type I
6. Alveolar type II
7. Basal keratinocyte
8. Basal
9. Ciliated
10. Club
11. Cornified keratinocyte
12. Luminal
13. Mature keratinocyte
14. Squamous
15. Suprabasal keratinocyte
16. Suprabasal
17. Fibroblast
18. ICCs
Immune
19. B cell
20. Dendritic cell (DC)
21. DC/macrophage
22. Langerhans
23. NK cell
24. T cell
25. Alveolar macrophage
26. Mast
27. Melanocyte
28. Mucous cell
Myonuclei
29. NMJ-localized
30. Cardiac
31. Cardiac, cytoplasmic
32. Skeletal muscle
33. Sk. muscle, cytoplasmic
34. Smooth muscle
35. Myoepithelial (basal)
36. Myofibroblast
37. Neuroendocrine
38. Neuronal
39. Pericyte/SMC
40. Satellite
41. Schwann cell
42. Sebaceous gland
43. Sweat gland
② # of nuclei
③ Protocol
CST
EZ
NST
TST
④ Tissue
Breast
Heart
Lung
Prostate
Esophagus mucosa
Esophagus muscularis
Skeletal muscle
Skin
C
Tissue
Sex
Protocol
Participant ID
FR1
FR2
D
Composition (%)
Proportions (%)
Adipocyte
Endothelial cell (lymphatic)
Endothelial cell (vascular)
Epithelial cell (basal)
Epithelial cell (club)
Epithelial cell (luminal)
Fibroblast
Immune (B cell)
Immune (DC/macrophage)
Immune (NK cell)
Immune (T cell)
Immune (mast cell)
Myocyte (smooth muscle)
Schwann cell
Neuroendocrine
Pericyte/SMC
Epithelial cell (Hillock)
Epithelial cell (alveolar type I)
Epithelial cell (alveolar type II)
Epithelial cell (ciliated)
Immune (alveolar macrophage)
Epithelial cell (basal keratinocyte)
Epithelial cell (cornified keratinocyte)
Epithelial cell (mature keratinocyte)
Epithelial cell (suprabasal keratinocyte)
Melanocyte
Sebaceous gland cell
Sweat gland cell
Immune (Langerhans)
Epithelial cell (squamous)
Epithelial cell (suprabasal)
Immune (DC)
Mucous cell
Myofibroblast
Myonuclei (cardiac)
Myonuclei (cardiac, cytoplasmic)
Myonuclei (NMJ-localized)
Myonuclei (sk. muscle)
Myonuclei (sk. muscle, cytoplasmic)
Satellite cell
Myoepithelial (basal)
ICCs
Neuronal
E
Breast
Esophagus mucosa
Esophagus muscularis
Heart
Prostate
Skeletal muscle
Skin
Effect size
Significant
True
False
−log10(FDR)

Fig. S2. Characterization of the cross-tissue snRNA-seq atlas. (A) Sample characteristics in integrated atlas. UMAP visualization of snRNA-seq profiles (as in **Fig. 1B**), colored by donor sex, self-reported ethnicity, age, BMI, and sample ischemic time. **(B)** Distinct cell type composition of each tissue. Overall proportion of cells (%) of each type (color legend) in each experiment (rows, 1), along with number of nuclei profiled (2), lab protocol (3), tissue (4) and specimen (label on right). Numbers on bars: broad cell type numbers (color legend). **(C)** EZ isolation protocol is more distinct. Force-directed graph layout embedding of samples (dots), where sample similarity is calculated based on cell type composition (*28*). Samples are colored by tissue, sex of the donor, nuclei prep and participant ID. **(D,E)** Tissue specific and shared cell types. **(D)** Overall proportion of cells from each cell subset (bars, (1)) that are derived from each tissue (colors; normalized for the total number of nuclei profiled in each tissue (*28*)), along with total number of nuclei from that cell type (2), and the proportion of nuclei from each protocol (3). Black vertical lines: relative proportion of nuclei from each individual. **(E)** Significance ($-\log_{10}$(Benjamini–Hochberg FDR), circle size) and effect size (circle color) of the differential abundances of each cell type (columns) in each tissue (rows) compared to lung as a reference. Samples with fewer than 30 nuclei are not shown.

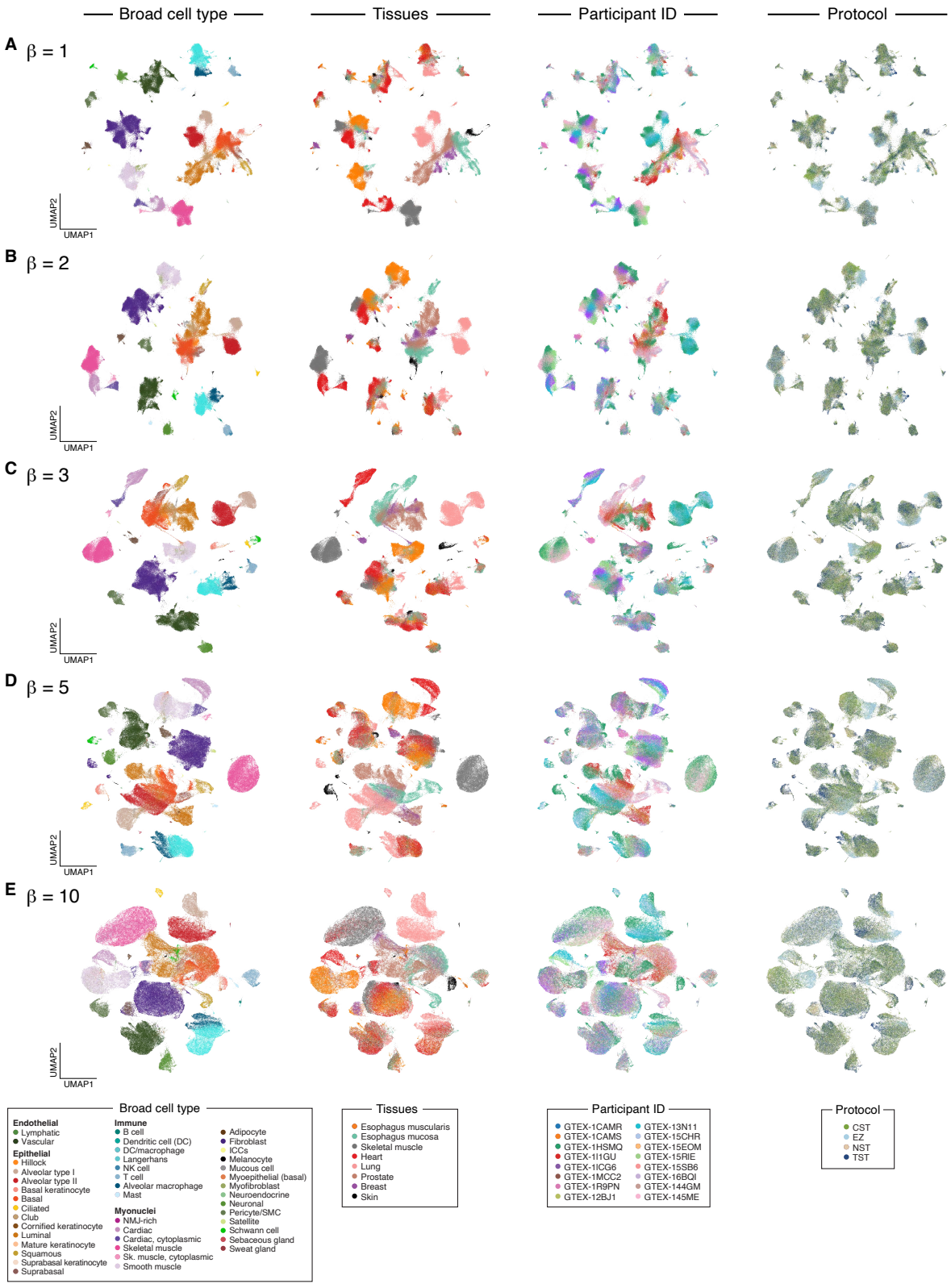

Broad cell type
Tissues
Participant ID
Protocol
A β = 1
B β = 2
C β = 3
D β = 5
E β = 10
UMAP2
UMAP1
Broad cell type
Endothelial
Lymphatic
Vascular
Epithelial
Hillock
Alveolar type I
Alveolar type II
Basal keratinocyte
Basal
Ciliated
Club
Cornified keratinocyte
Luminal
Mature keratinocyte
Squamous
Suprabasal keratinocyte
Suprabasal
Immune
B cell
Dendritic cell (DC)
DC/macrophage
Langerhans
NK cell
T cell
Alveolar macrophage
Mast
Myonuclei
NMJ-rich
Cardiac
Cardiac, cytoplasmic
Skeletal muscle
Sk. muscle, cytoplasmic
Smooth muscle
Adipocyte
Fibroblast
ICCs
Melanocyte
Mucous cell
Myoepithelial (basal)
Myofibroblast
Neuroendocrine
Neuronal
Pericyte/SMC
Satellite
Schwann cell
Sebaceous gland
Sweat gland
Tissues
Esophagus muscularis
Esophagus mucosa
Skeletal muscle
Heart
Lung
Prostate
Breast
Skin
Participant ID
GTEX-1CAMR
GTEX-1CAMS
GTEX-1HSMQ
GTEX-1I1GU
GTEX-1ICG6
GTEX-1MCC2
GTEX-1R9PN
GTEX-12BJ1
GTEX-13N11
GTEX-15CHR
GTEX-15EOM
GTEX-15RIE
GTEX-15SB6
GTEX-16BQI
GTEX-144GM
GTEX-145ME
Protocol
CST
EZ
NST
TST

Fig. S3. Hyperparameter sweep for the β parameter of the conditional variational autoencoder. (**A–E**) Higher beta values over-smoothened data representations. UMAP representation of single-nucleus profiles (dots) colored (from left to right) by broad cell type, tissue, individual, or protocol using the mean latent space coordinates from conditional β-TCVAEs trained with beta hyperparameter values of (**A**) 1.0 (**B**) 2.0, (**C**) 3.0, (**D**) 5.0, and (**E**), 10.0 (*28*). The β-TCVAE with β = 2.0 was selected to minimize loss of granularity.

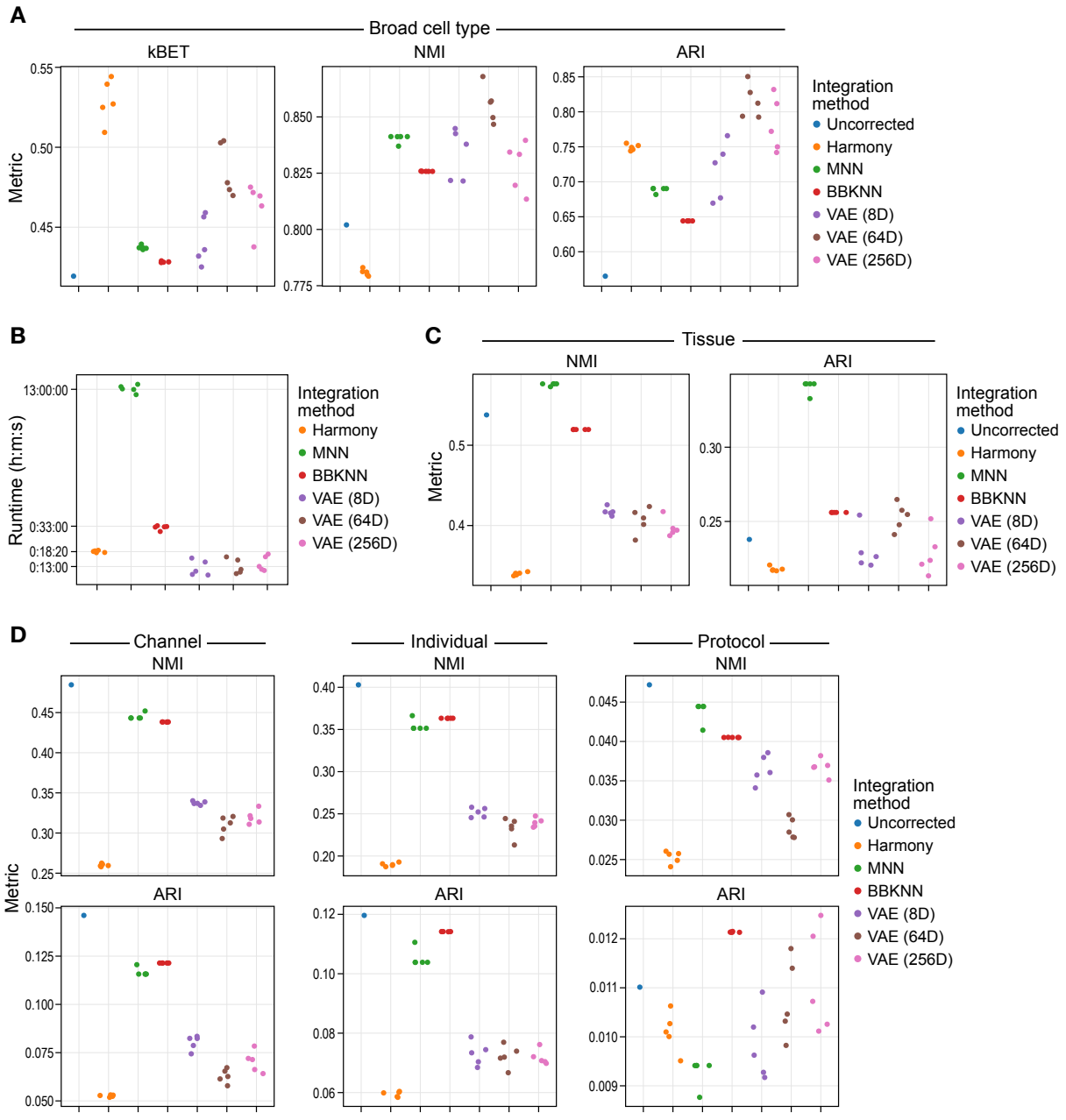


Fig. S4. Benchmarking of TCVAE to other single cell profiling integration methods. (A) Conservation of broad cell type variation. kBET (left), normalized mutual information (NMI, middle) and adjusted Rand index (ARI, right) metrics of cell type variation (*y*-axis) for uncorrected data and data integrated by each tested method (dot color, *x*-axis, (*28*)). **(B)** Runtime comparison. Runtime (*y*-axis, hour:minute:second) of each integration method (dot color, *x*-axis). **(C)** Conservation of tissue-specific variation. NMI (left) and ARI (right) metrics (*y*-axis) of tissue

specific variation in uncorrected data and data integrated by the tested methods (*x*-axis, dot color). **(D)** Correction for unwanted effects. NMI (top) and ARI (bottom) metrics (*y*-axis) assessing remaining variation due to unwanted channel (batch, left), individual (middle) and protocol (right) effects after integration with each tested method (dot color, *x*-axis). In all panels, each dot is a different random integration run with different seeds performed on the full dataset.

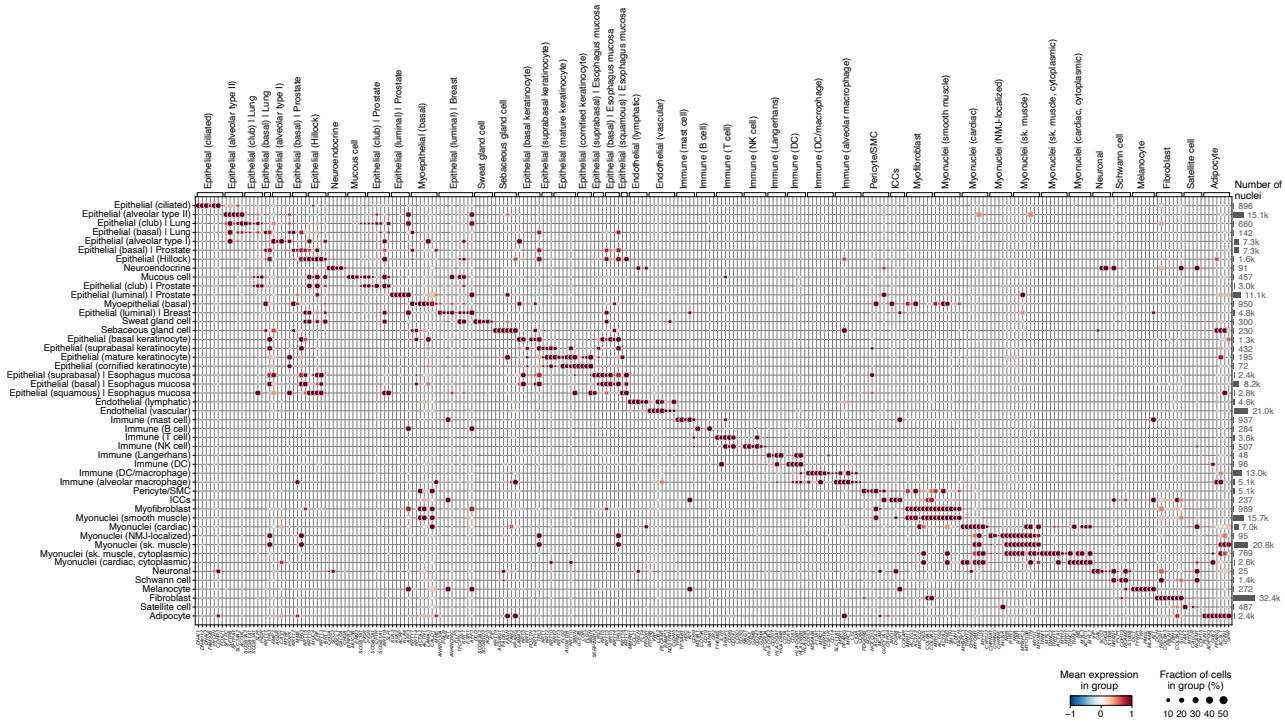
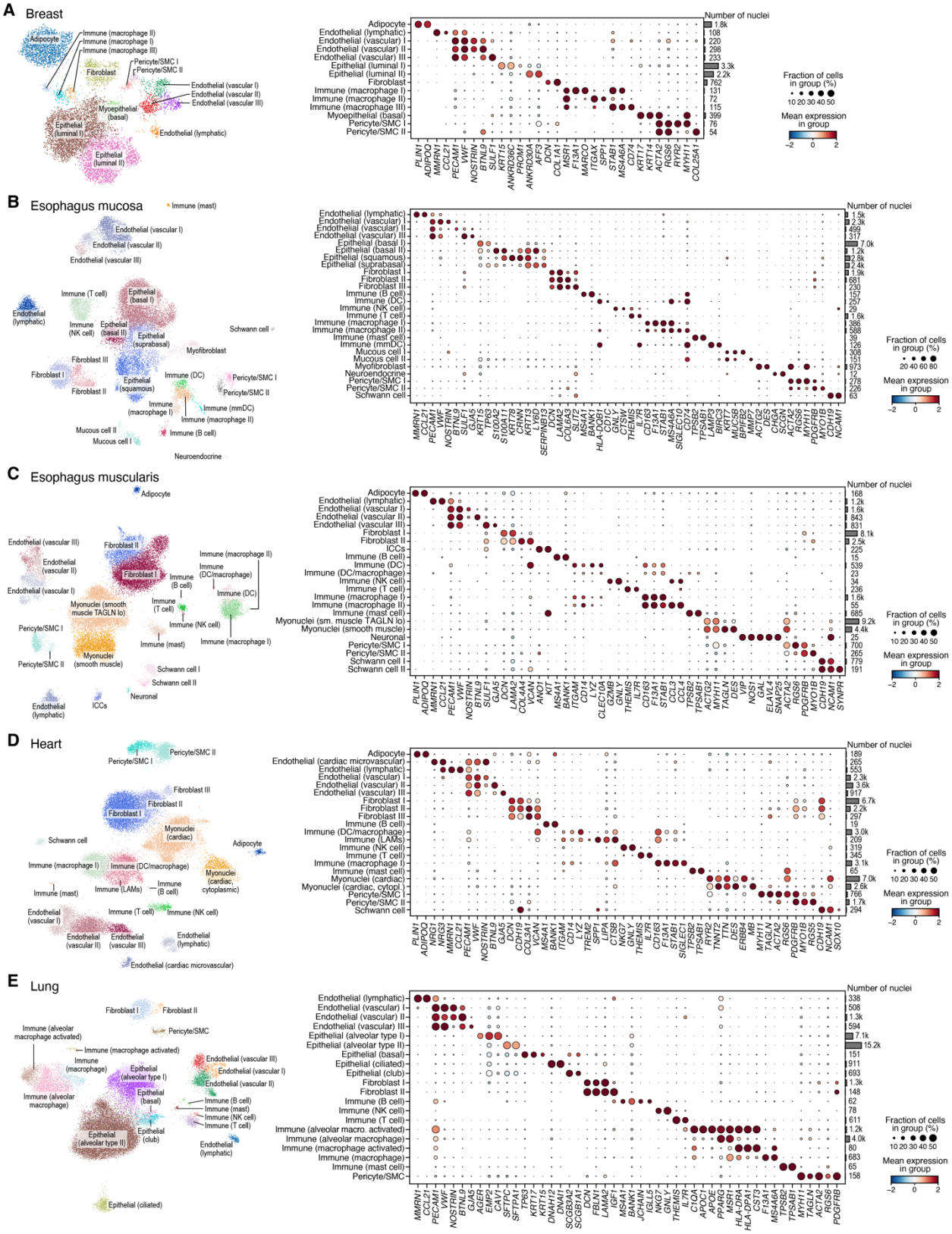


Fig. S5. Broad cell type gene markers. Scaled mean expression (*z*-score, circle color) and fraction of expressing cells (circle size) of marker genes (columns, labels at bottom) associated with each granular cell subset (rows, and labels on top) with epithelial cells with tissue-specific markers shown separately. Right: number of nuclei in each cell subset.


A Breast
Adipocyte
Immune (macrophage II)
Immune (macrophage I)
Immune (macrophage III)
Pericyte/SMC I
Pericyte/SMC II
Fibroblast
Endothelial (vascular I)
Endothelial (vascular II)
Endothelial (vascular III)
Myoepithelial (basal)
Epithelial (luminal I)
Endothelial (lymphatic)
Epithelial (luminal II)
Number of nuclei
Fraction of cells in group (%)
Mean expression in group
B Esophagus mucosa
Immune (mast)
Endothelial (vascular I)
Endothelial (vascular II)
Endothelial (vascular III)
Immune (T cell)
Epithelial (basal I)
Endothelial (lymphatic)
Immune (NK cell)
Epithelial (basal II)
Epithelial (suprabasal)
Schwann cell
Myofibroblast
Fibroblast III
Fibroblast I
Fibroblast II
Epithelial (squamous)
Immune (DC)
Pericyte/SMC I
Pericyte/SMC II
Immune (macrophage I)
Immune (mmDC)
Immune (macrophage II)
Mucous cell II
Mucous cell I
Immune (B cell)
Neuroendocrine
C Esophagus muscularis
Adipocyte
Endothelial (vascular III)
Fibroblast II
Endothelial (vascular II)
Fibroblast I
Immune (macrophage II)
Immune (DC/macrophage)
Immune (B cell)
Endothelial (vascular I)
Myonuclei (smooth muscle TAGLN lo)
Immune (T cell)
Immune (DC)
Pericyte/SMC I
Immune (NK cell)
Immune (mast)
Immune (macrophage I)
Myonuclei (smooth muscle)
Pericyte/SMC II
Schwann cell I
Schwann cell II
Neuronal
Endothelial (lymphatic)
ICCs
D Heart
Pericyte/SMC I
Pericyte/SMC II
Fibroblast III
Fibroblast II
Fibroblast I
Myonuclei (cardiac)
Schwann cell
Adipocyte
Immune (macrophage I)
Immune (DC/macrophage)
Myonuclei (cardiac, cytoplasmic)
Immune (mast)
Immune (LAMs)
Immune (B cell)
Endothelial (vascular I)
Immune (T cell)
Immune (NK cell)
Endothelial (vascular II)
Endothelial (vascular III)
Endothelial (lymphatic)
Endothelial (cardiac microvascular)
E Lung
Fibroblast I
Fibroblast II
Immune (alveolar macrophage activated)
Pericyte/SMC
Immune (macrophage activated)
Immune (macrophage)
Endothelial (vascular III)
Endothelial (vascular I)
Epithelial (alveolar type I)
Endothelial (vascular II)
Epithelial (basal)
Immune (alveolar macrophage)
Immune (B cell)
Immune (mast)
Immune (NK cell)
Immune (T cell)
Epithelial (alveolar type II)
Epithelial (club)
Endothelial (lymphatic)
Epithelial (ciliated)

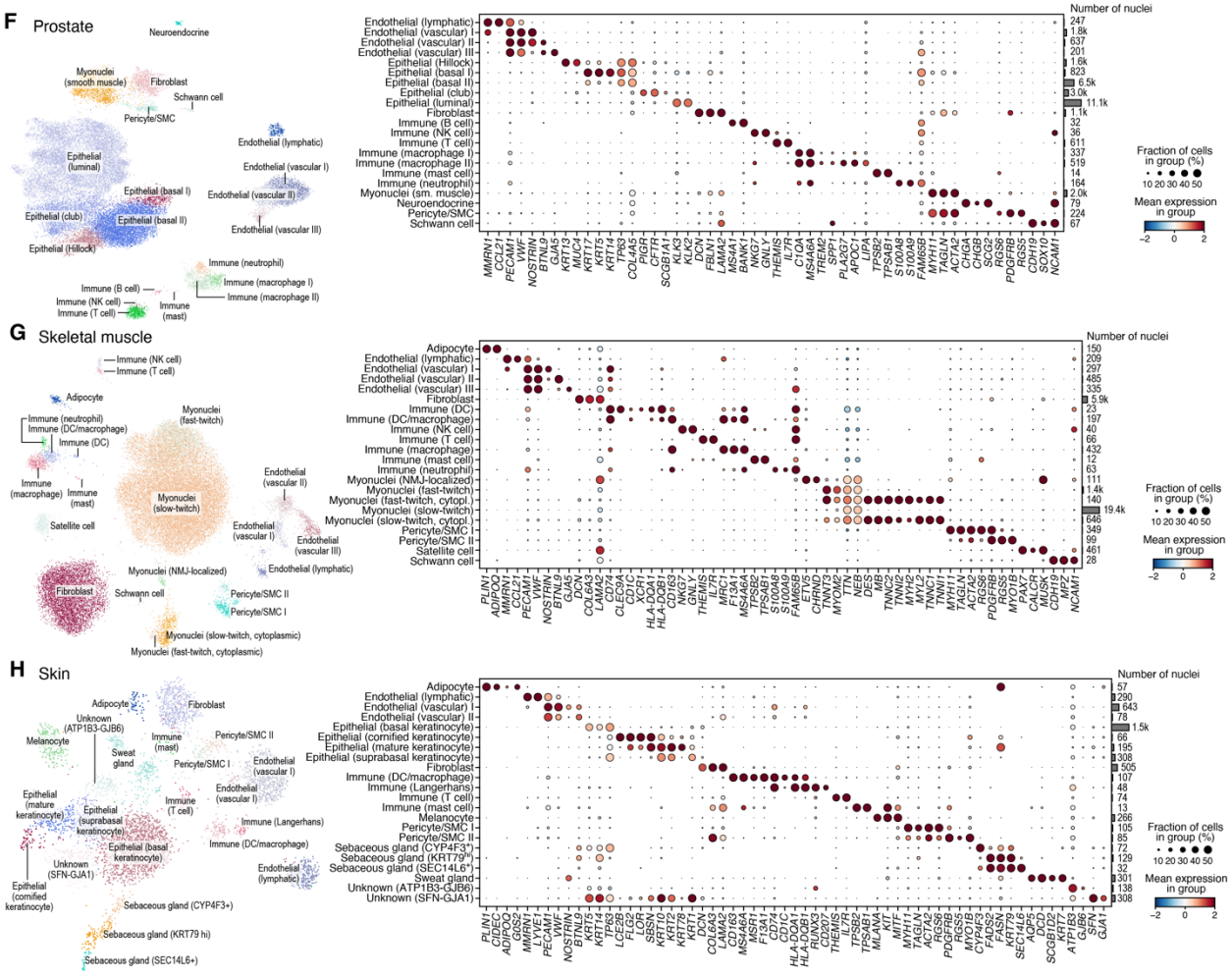


Fig. S6. Granular cell subsets in each tissue. UMAP visualization (left) of single nucleus profiles (dots) colored by granular cell type annotation and scaled mean expression (*z*-score, circle color) and fraction of expressing cells (circle size) of marker genes (columns, labels at bottom) associated with those subsets (rows, with nuclei number on right) for breast **(A)**, esophagus mucosa **(B)**, esophagus muscularis **(C)**, heart **(D)**, lung **(E)**, prostate **(F)**, skeletal muscle **(G)**, and skin **(H)**. Cell type numbering *within* each tissue (e.g., Fibroblast I) do not necessarily match across tissues.

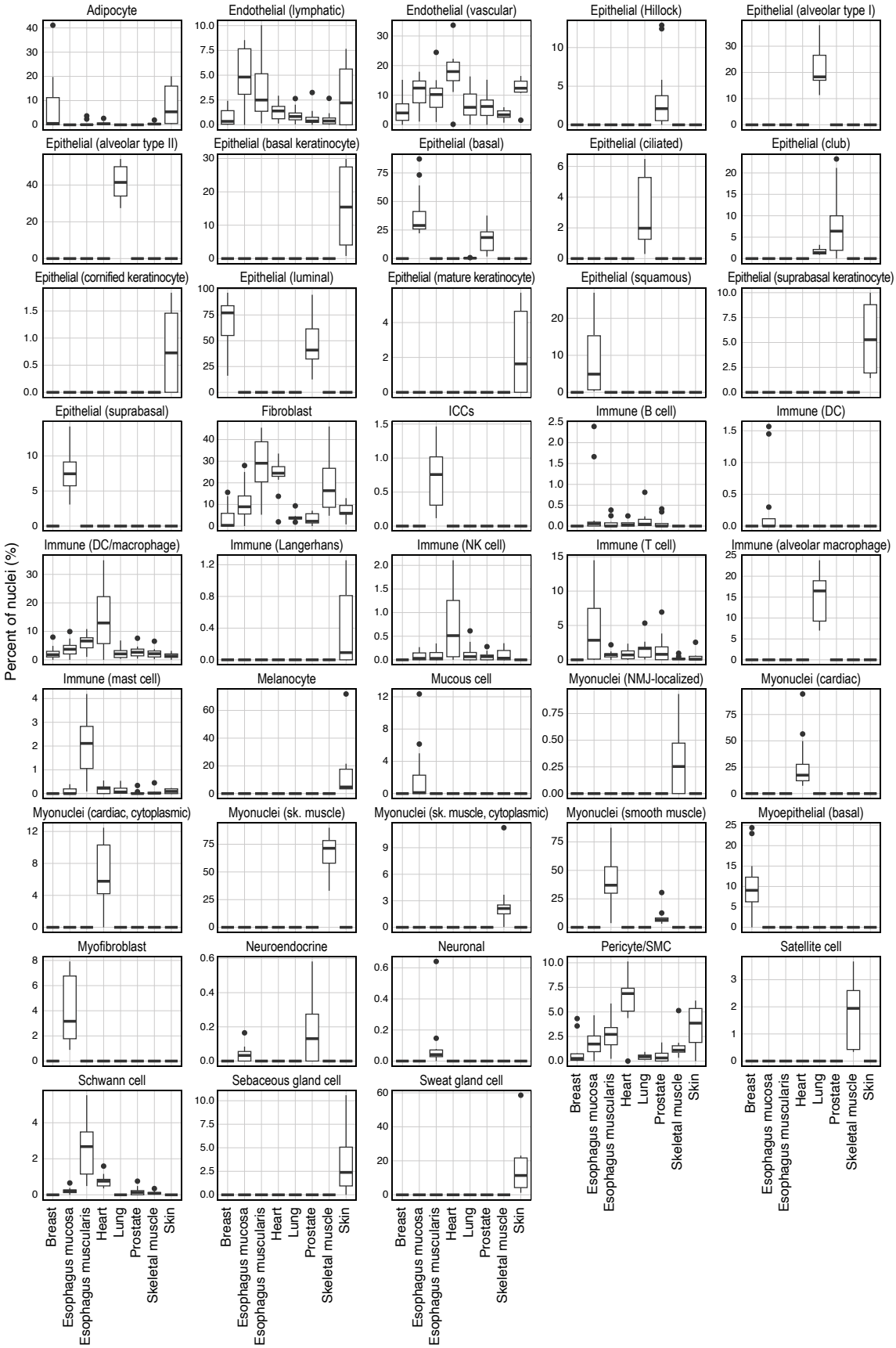

Percent of nuclei (%)
Adipocyte
Endothelial (lymphatic)
Endothelial (vascular)
Epithelial (Hillock)
Epithelial (alveolar type I)
Epithelial (alveolar type II)
Epithelial (basal keratinocyte)
Epithelial (basal)
Epithelial (ciliated)
Epithelial (club)
Epithelial (cornified keratinocyte)
Epithelial (luminal)
Epithelial (mature keratinocyte)
Epithelial (squamous)
Epithelial (suprabasal keratinocyte)
Epithelial (suprabasal)
Fibroblast
ICCs
Immune (B cell)
Immune (DC)
Immune (DC/macrophage)
Immune (Langerhans)
Immune (NK cell)
Immune (T cell)
Immune (alveolar macrophage)
Immune (mast cell)
Melanocyte
Mucous cell
Myonuclei (NMJ-localized)
Myonuclei (cardiac)
Myonuclei (cardiac, cytoplasmic)
Myonuclei (sk. muscle)
Myonuclei (sk. muscle, cytoplasmic)
Myonuclei (smooth muscle)
Myoepithelial (basal)
Myofibroblast
Neuroendocrine
Neuronal
Pericyte/SMC
Satellite cell
Schwann cell
Sebaceous gland cell
Sweat gland cell
Breast
Esophagus mucosa
Esophagus muscularis
Heart
Lung
Prostate
Skeletal muscle
Skin

Fig. S7. Distribution of broad cell type proportions within each tissue. Distribution of proportions (*y*-axis) of each broad cell type (columns) within each tissue (*x*-axis). Box plots show median, quartiles, and whiskers at 1.5 IQR (interquartile range).

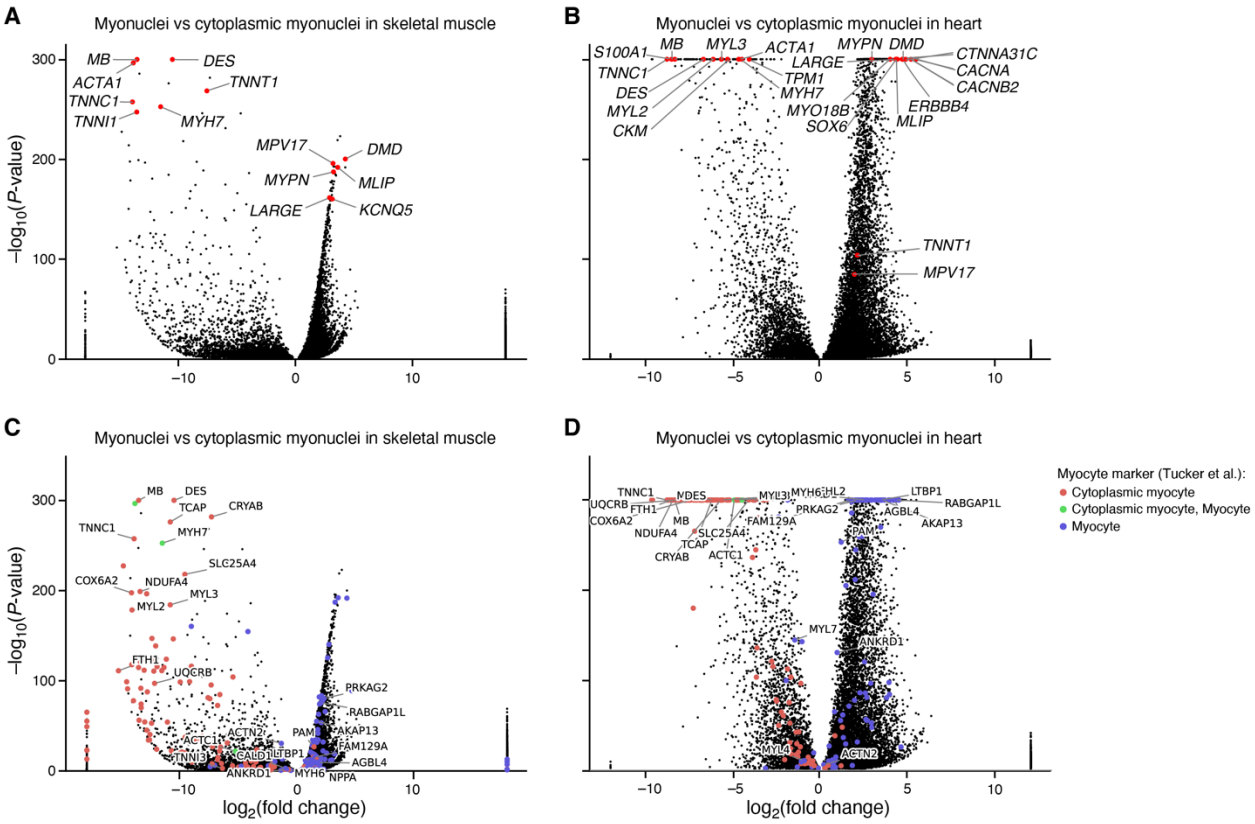


Fig. S8. Biological distinctions between classical and cytoplasmic myonuclei. Differential expression (Log_2(fold-change), *x*-axis) and associated significance ($-log_{10}$(*P*-value), *y*-axis) between classical and cytoplasmic myonuclei in skeletal muscle (**A,C**) and heart (**B,D**), highlighting key genes (A,B) or markers of the two subsets from Tucker et al. (*33*) (C,D, dot color).

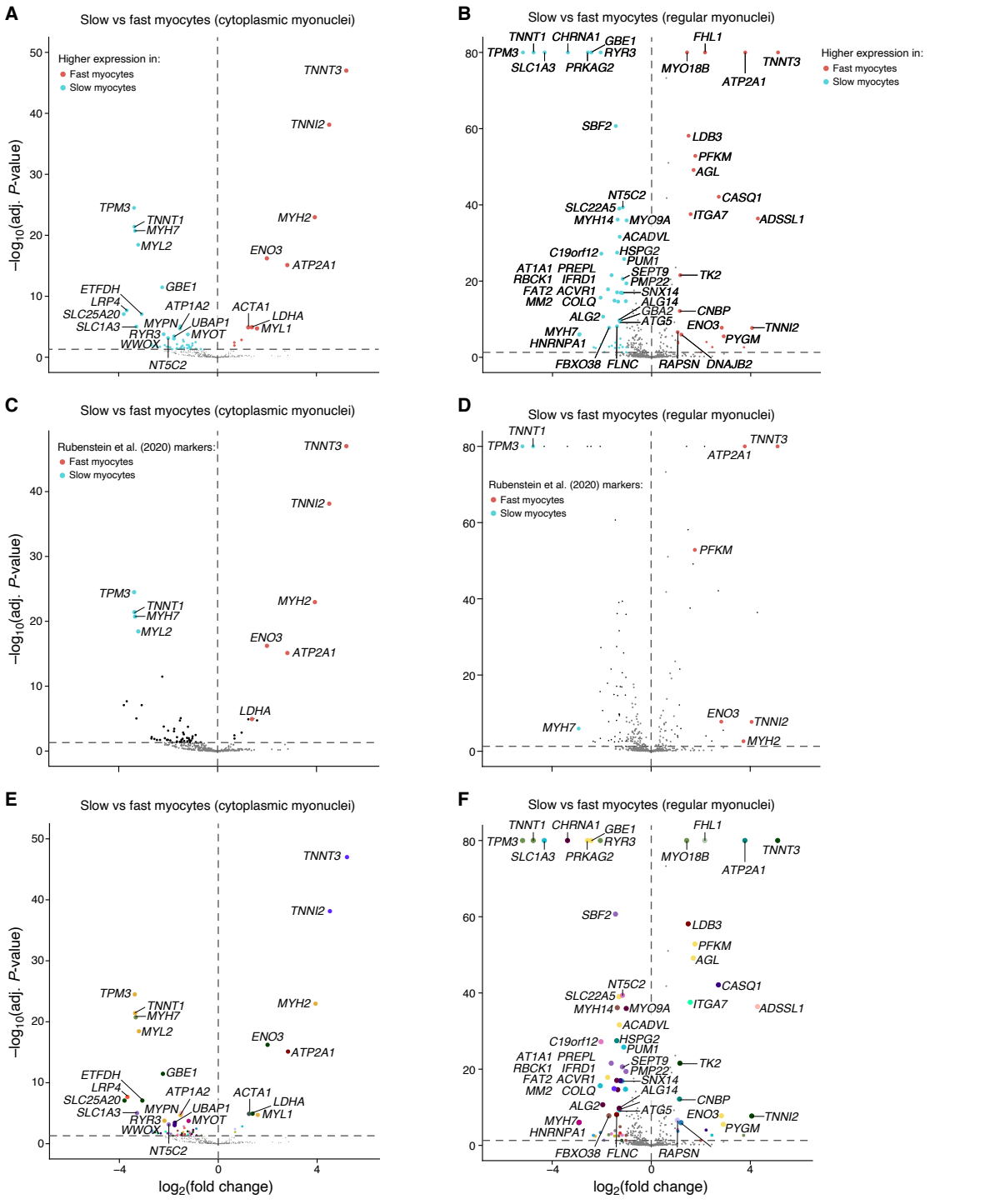

A
Slow vs fast myocytes (cytoplasmic myonuclei)
Higher expression in:
Fast myocytes
Slow myocytes
−log10(adj. P-value)
TNNT3
TNNI2
TPM3
TNNT1
MYH7
MYL2
MYH2
ENO3
ATP2A1
ETFDH
LRP4
SLC25A20
SLC1A3
GBE1
ATP1A2
MYPN
UBAP1
RYR3
MYOT
WWOX
ACTA1
LDHA
MYL1
NT5C2
B
Slow vs fast myocytes (regular myonuclei)
Higher expression in:
Fast myocytes
Slow myocytes
TNNT1
CHRNA1
GBE1
TPM3
RYR3
SLC1A3
PRKAG2
FHL1
MYO18B
TNNT3
ATP2A1
SBF2
LDB3
PFKM
AGL
NT5C2
CASQ1
SLC22A5
MYH14
MYO9A
ITGA7
ADSSL1
ACADVL
C19orf12
HSPG2
PUM1
AT1A1
PREPL
SEPT9
TK2
RBCK1
IFRD1
PMP22
FAT2
ACVR1
SNX14
MM2
COLQ
ALG14
ALG2
GBA2
CNBP
ATG5
ENO3
TNNI2
MYH7
HNRNPA1
PYGM
FBXO38
FLNC
RAPSN
DNAJB2
C
Slow vs fast myocytes (cytoplasmic myonuclei)
Rubenstein et al. (2020) markers:
Fast myocytes
Slow myocytes
−log10(adj. P-value)
TNNT3
TNNI2
TPM3
TNNT1
MYH7
MYL2
MYH2
ENO3
ATP2A1
LDHA
D
Slow vs fast myocytes (regular myonuclei)
Rubenstein et al. (2020) markers:
Fast myocytes
Slow myocytes
TNNT1
TPM3
TNNT3
ATP2A1
PFKM
ENO3
TNNI2
MYH7
MYH2
E
Slow vs fast myocytes (cytoplasmic myonuclei)
−log10(adj. P-value)
log2(fold change)
F
Slow vs fast myocytes (regular myonuclei)
log2(fold change)
Disease group, panel E
Congenital muscular dystrophies, muscular dystrophies
Congenital muscular dystrophies, congenital myopathies
Congenital myasthenic syndromes
Congenital myopathies
Congenital myopathies, distal myopathies
Congenital myopathies, ion channel muscle diseases malignant hyperthermia
Congenital myopathies, other neuromuscular disorders
Hereditary ataxias
Hereditary ataxias, motor neuron diseases
Hereditary ataxias, other neuromuscular disorders
Hereditary motor and sensory neuropathies
Hereditary motor and sensory neuropathies, hereditary paraplegias, motor neuron diseases
Hereditary motor and sensory neuropathies, motor neuron diseases
Hereditary paraplegias
Distal myopathies, muscular dystrophies, other myopathies
Metabolic myopathies
Ion channel muscle diseases
Motor neuron diseases
Myotonic syndromes
Muscular dystrophies
Other neuromuscular disorders
Muscular dystrophies, other myopathies
Disease group, panel F
Congenital muscular dystrophies, congenital myasthenic syndromes, muscular dystrophies
Congenital muscular dystrophies, congenital myopathies
Congenital muscular dystrophies, muscular dystrophies
Congenital myopathies
Congenital myopathies, muscular dystrophies, other myopathies
Congenital muscular dystrophies
Congenital myasthenic syndromes
Congenital myasthenic syndromes, other neuromuscular disorders
Congenital myopathies, distal myopathies
Congenital myopathies, motor neuron diseases
Hereditary ataxias
Hereditary motor and sensory neuropathies
Hereditary motor and sensory neuropathies, hereditary paraplegias
Hereditary motor and sensory neuropathies, motor neuron diseases
Hereditary paraplegias
Hereditary ataxias, hereditary motor and sensory neuropathies
Motor neuron diseases
Motor neuron diseases, other neuromuscular disorders
Other myopathies
Other neuromuscular disorders
Distal myopathies
Distal myopathies, other myopathies
Metabolic myopathies
Muscular dystrophies, other myopathies
Myotonic syndromes

Fig. S9. Classical and cytoplasmic skeletal muscle myonuclei from slow and fast myocytes express distinct signature and monogenic muscle disease genes. Differential expression (Log_2(fold-change), *x*-axis) and associated significance ($-log_{10}$(*P*-value), *y*-axis) between slow and fast myocyte subtypes within cytoplasmic **(A, C, E)** and classical myonuclei **(B, D, F)**, with genes highlighted (dot color) based on differential expression (A,B), known markers of fast vs. slow subsets (C,D) (*36*), and monogenic disease status (E,F) (*28*).

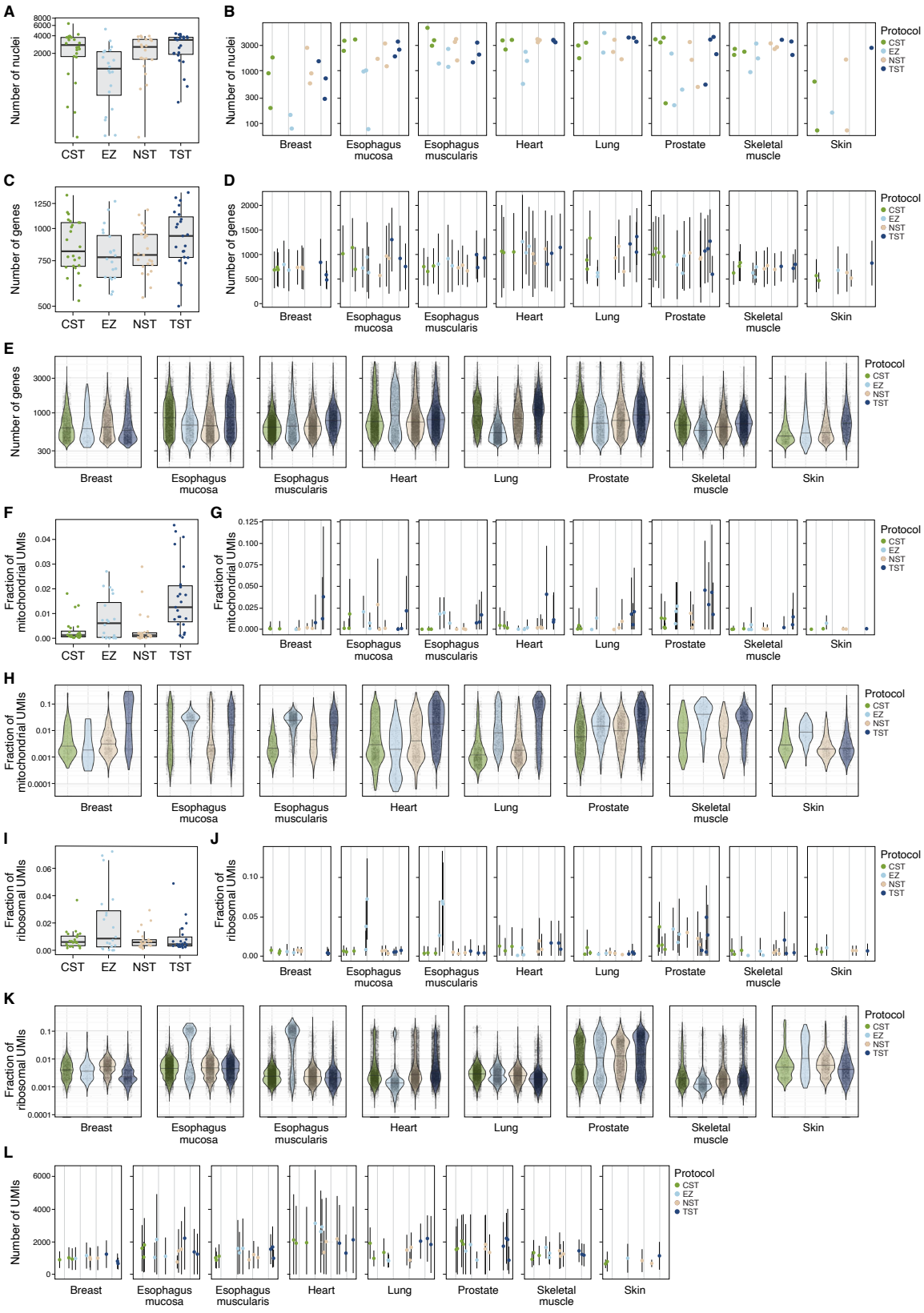

A
Number of nuclei
CST
EZ
NST
TST
B
Number of nuclei
Breast
Esophagus mucosa
Esophagus muscularis
Heart
Lung
Prostate
Skeletal muscle
Skin
Protocol
CST
EZ
NST
TST
C
Number of genes
D
Number of genes
E
Number of genes
F
Fraction of mitochondrial UMIs
G
Fraction of mitochondrial UMIs
H
Fraction of mitochondrial UMIs
I
Fraction of ribosomal UMIs
J
Fraction of ribosomal UMIs
K
Fraction of ribosomal UMIs
L
Number of UMIs

Fig. S10. Relative benchmarking of four snRNA-seq protocols on different QC measures.

(A, B) Number of recovered nuclei. Distribution (A) and number (B) of high-quality nuclei profiles (*y*-axis) recovered from each protocol in aggregate (A, *x*-axis) and for each sample in each tissue (B, *x*-axis). **(C–E)** Number of recovered genes. Distribution (C, E) and mean number (D) of genes (*y*-axis) recovered from each protocol in aggregate (C, *x*-axis), for each sample in each tissue (D, *x*-axis), and for each protocol in each tissue (E, *x*-axis). **(F–H)** Fraction of mitochondrial transcripts. Distribution (F, H) and mean (G) of the fraction of mitochondrial transcripts (*y*-axis, Unique Molecular Identifiers (UMIs)) recovered from each protocol in aggregate (F, *x*-axis), for each sample in each tissue (G, *x*-axis), and for each protocol in each tissue (H, *x*-axis). **(I–K)** Fraction of ribosomal transcripts. Distribution (I, K) and mean (J) of the fraction of ribosomal transcripts (*y*-axis, UMIs) recovered from each protocol in aggregate (F, *x*-axis), for each sample in each tissue (G, *x*-axis), and for each protocol in each tissue (H, *x*-axis). **(L)** Number of transcripts. Mean number of UMIs (*y*-axis) in each sample in each tissue. Box plots show median, quartiles, and whiskers at 1.5 IQR (interquartile range). The horizontal bar in violin plots represents the median. Error bars in (D,G,J,L) show one standard deviation above and below the mean.

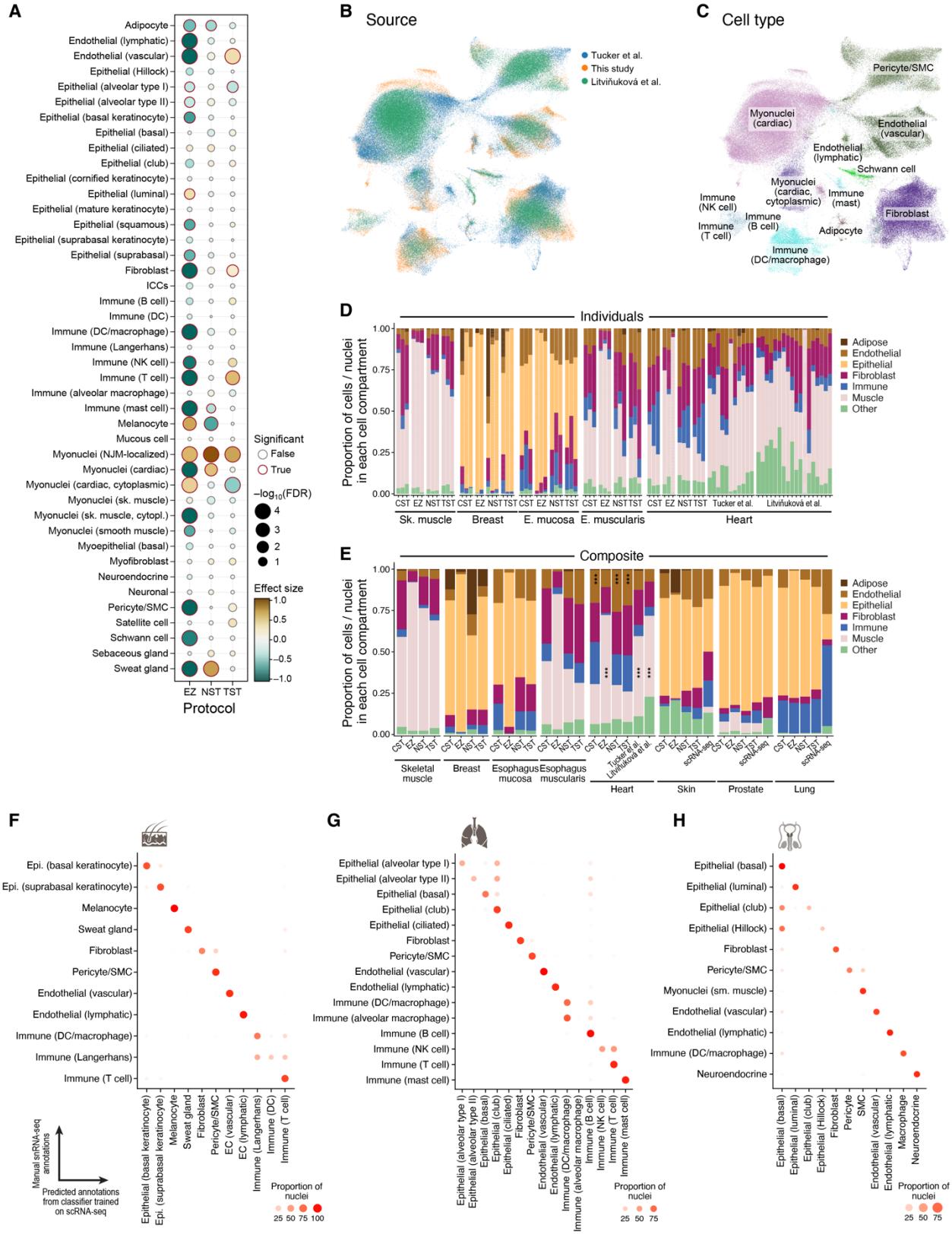

A
Adipocyte
Endothelial (lymphatic)
Endothelial (vascular)
Epithelial (Hillock)
Epithelial (alveolar type I)
Epithelial (alveolar type II)
Epithelial (basal keratinocyte)
Epithelial (basal)
Epithelial (ciliated)
Epithelial (club)
Epithelial (cornified keratinocyte)
Epithelial (luminal)
Epithelial (mature keratinocyte)
Epithelial (squamous)
Epithelial (suprabasal keratinocyte)
Epithelial (suprabasal)
Fibroblast
ICCs
Immune (B cell)
Immune (DC)
Immune (DC/macrophage)
Immune (Langerhans)
Immune (NK cell)
Immune (T cell)
Immune (alveolar macrophage)
Immune (mast cell)
Melanocyte
Mucous cell
Myonuclei (NJM-localized)
Myonuclei (cardiac)
Myonuclei (cardiac, cytoplasmic)
Myonuclei (sk. muscle)
Myonuclei (sk. muscle, cytopl.)
Myonuclei (smooth muscle)
Myoepithelial (basal)
Myofibroblast
Neuroendocrine
Neuronal
Pericyte/SMC
Satellite cell
Schwann cell
Sebaceous gland
Sweat gland
EZ NST TST
Protocol
Significant
False
True
−log10(FDR)
4
3
2
1
Effect size
1.0
0.5
0.0
−0.5
−1.0
B
Source
Tucker et al.
This study
Litviňuková et al.
C
Cell type
Pericyte/SMC
Myonuclei (cardiac)
Endothelial (vascular)
Endothelial (lymphatic)
Schwann cell
Myonuclei (cardiac, cytoplasmic)
Immune (mast)
Immune (NK cell)
Immune (B cell)
Fibroblast
Immune (T cell)
Adipocyte
Immune (DC/macrophage)
D
Individuals
Proportion of cells / nuclei in each cell compartment
1.00
0.75
0.50
0.25
0.00
CST EZ NST TST
Sk. muscle
Breast
E. mucosa
E. muscularis
Tucker et al.
Litviňuková et al.
Heart
Adipose
Endothelial
Epithelial
Fibroblast
Immune
Muscle
Other
E
Composite
Proportion of cells / nuclei in each cell compartment
scRNA-seq
Skeletal muscle
Breast
Esophagus mucosa
Esophagus muscularis
Heart
Skin
Prostate
Lung
F
Epi. (basal keratinocyte)
Epi. (suprabasal keratinocyte)
Melanocyte
Sweat gland
Fibroblast
Pericyte/SMC
Endothelial (vascular)
Endothelial (lymphatic)
Immune (DC/macrophage)
Immune (Langerhans)
Immune (T cell)
Epithelial (basal keratinocyte)
Epi. (suprabasal keratinocyte)
Melanocyte
Sweat gland
Fibroblast
Pericyte/SMC
EC (vascular)
EC (lymphatic)
Immune (Langerhans)
Immune (DC)
Immune (T cell)
Manual snRNA-seq annotations
Predicted annotations from classifier trained on scRNA-seq
Proportion of nuclei
25 50 75 100
G
Epithelial (alveolar type I)
Epithelial (alveolar type II)
Epithelial (basal)
Epithelial (club)
Epithelial (ciliated)
Fibroblast
Pericyte/SMC
Endothelial (vascular)
Endothelial (lymphatic)
Immune (DC/macrophage)
Immune (alveolar macrophage)
Immune (B cell)
Immune (NK cell)
Immune (T cell)
Immune (mast cell)
Proportion of nuclei
25 50 75
H
Epithelial (basal)
Epithelial (luminal)
Epithelial (club)
Epithelial (Hillock)
Fibroblast
Pericyte/SMC
Myonuclei (sm. muscle)
Endothelial (vascular)
Endothelial (lymphatic)
Immune (DC/macrophage)
Neuroendocrine
SMC
Macrophage
Proportion of nuclei
25 50 75

Fig. S11. Benchmarking of snRNA-seq protocols by cell composition and comparison to scRNA-seq. (A–E) Impact of snRNA-seq protocol on recovered cell composition. (**A**) Significance (Benjamini–Hochberg FDR < 0.1; circle outline color) of the enrichment of each cell type (rows) in each protocol (columns) relative to the CST protocol across all tissues and samples after correction for tissue specific effects (*28*). **(B,C)** UMAP representation of snRNA-seq profiles (after batch correction by individual with Harmony (*184*), (*28*) colored by study **(B)** and cell type **(C)**. **(D)** Proportion of cells or nuclei (*y*-axis) across broad cell groups (color legend) in each sample (*x*-axis), stratified by tissue and protocol. (**E**) Proportion of cells or nuclei (*y*-axis) across broad cell groups (color legend) in each tissue (*x*-axis), stratified by tissue and protocol. Asterisks indicate significantly higher (muscle EZ, Tucker et al. (*33*) and Litviňuková et al. (*34*)) or lower (endothelial, Tucker et al. (*33*) and Litviňuková et al. (*34*)) proportions compared to the CST protocol (Dirichlet regression, Benjamini–Hochberg FDR < 0.01, (*28*)). **(F–H)** Agreement in cell-intrinsic expression profiles between scRNA-seq and snRNA-seq. Proportion of nuclei (circle size and color) of each subset (rows) that are predicted to be in each cell class (columns) by a random forest classifier trained on cells, for skin **(F),** lung **(G)** or prostate **(H)**.

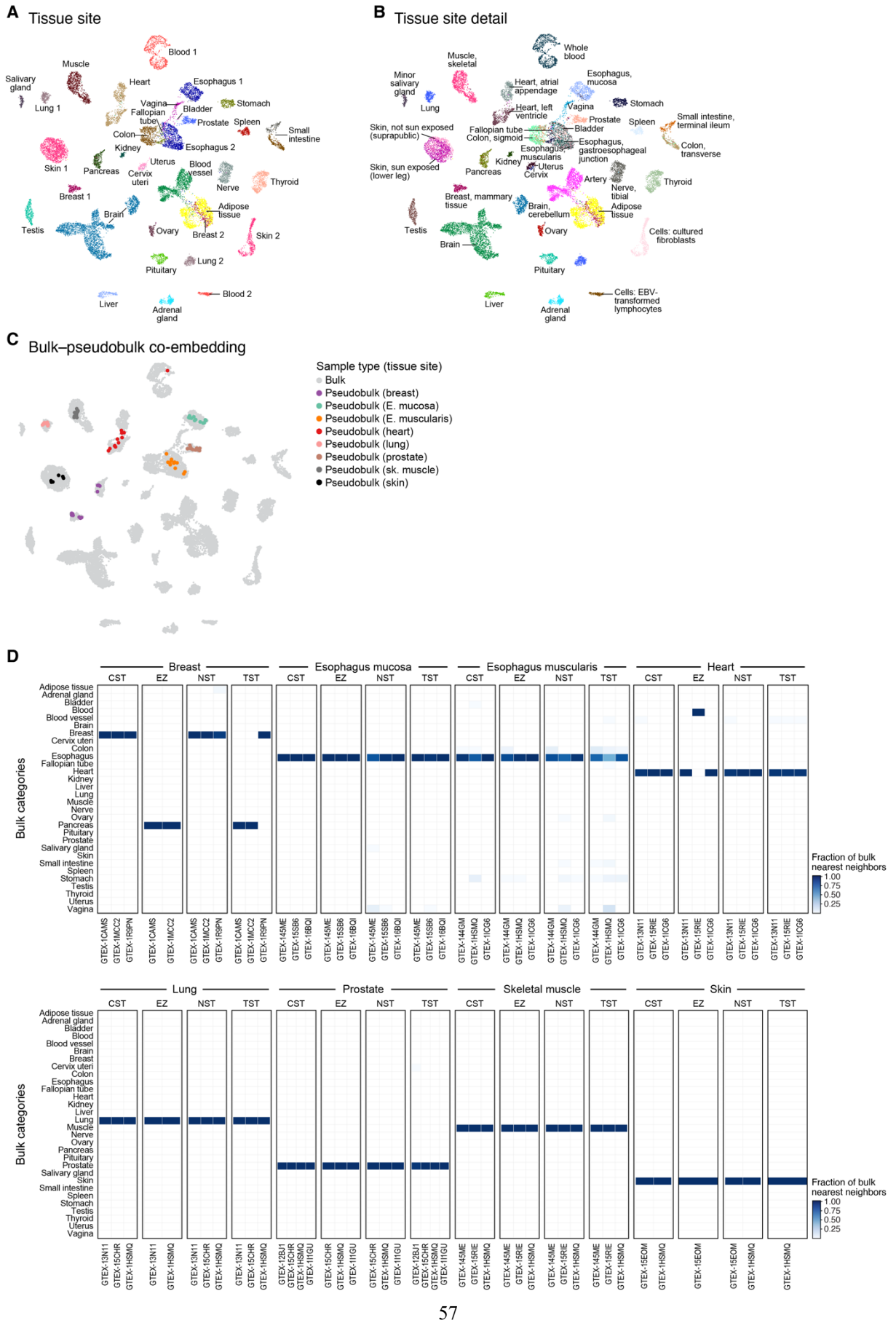
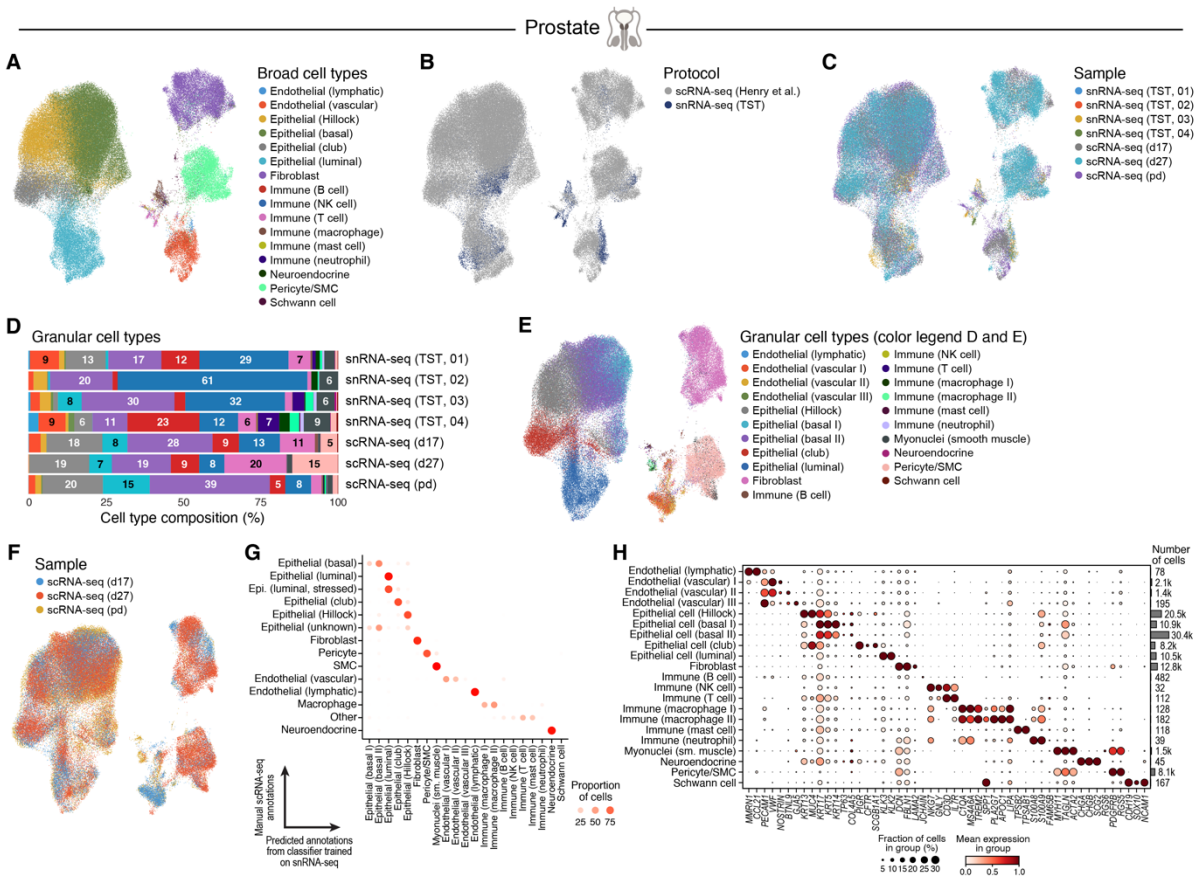

A
Tissue site
Blood 1
Muscle
Salivary gland
Heart
Esophagus 1
Lung 1
Vagina
Bladder
Stomach
Fallopian tube
Prostate
Spleen
Small intestine
Colon
Kidney
Esophagus 2
Skin 1
Pancreas
Uterus
Cervix uteri
Blood vessel
Nerve
Thyroid
Breast 1
Adipose tissue
Brain
Testis
Ovary
Breast 2
Skin 2
Lung 2
Pituitary
Liver
Blood 2
Adrenal gland
B
Tissue site detail
Whole blood
Muscle, skeletal
Minor salivary gland
Esophagus, mucosa
Heart, atrial appendage
Lung
Vagina
Stomach
Heart, left ventricle
Prostate
Spleen
Small intestine, terminal ileum
Skin, not sun exposed (suprapublic)
Fallopian tube
Bladder
Colon, sigmoid
Esophagus, gastroesophageal junction
Colon, transverse
Esophagus, muscularis
Kidney
Skin, sun exposed (lower leg)
Pancreas
Uterus
Cervix
Artery
Thyroid
Nerve, tibial
Breast, mammary tissue
Brain, cerebellum
Adipose tissue
Testis
Ovary
Cells: cultured fibroblasts
Brain
Pituitary
Liver
Adrenal gland
Cells: EBV-transformed lymphocytes
C
Bulk–pseudobulk co-embedding
Sample type (tissue site)
Bulk
Pseudobulk (breast)
Pseudobulk (E. mucosa)
Pseudobulk (E. muscularis)
Pseudobulk (heart)
Pseudobulk (lung)
Pseudobulk (prostate)
Pseudobulk (sk. muscle)
Pseudobulk (skin)
D
Breast
Esophagus mucosa
Esophagus muscularis
Heart
CST
EZ
NST
TST
Bulk categories
Adipose tissue
Adrenal gland
Bladder
Blood
Blood vessel
Brain
Breast
Cervix uteri
Colon
Esophagus
Fallopian tube
Heart
Kidney
Liver
Lung
Muscle
Nerve
Ovary
Pancreas
Pituitary
Prostate
Salivary gland
Skin
Small intestine
Spleen
Stomach
Testis
Thyroid
Uterus
Vagina
Fraction of bulk nearest neighbors
1.00
0.75
0.50
0.25
Lung
Prostate
Skeletal muscle
Skin

Fig. S12. snRNA-Seq pseudobulk matches bulk RNA-seq of matched samples. (**A–C**) Co-embedding of bulk and pseudobulk RNA-seq profiles. UMAP representation of bulk RNA-seq and pseudobulk snRNA-seq samples (dots), with each sample colored by tissue site **(A)** detailed tissue site (**B**) or highlighting only the pseudobulk samples from each site **(C)** (*28*). **(D)** Co-embedding successfully places bulk and pseudobulk samples from the same tissue in proximity, as shown by the fraction of bulk nearest neighbors (color bar) of each pseudobulk sample (columns) that are derived from each tissue site (rows).


Prostate
A
Broad cell types
Endothelial (lymphatic)
Endothelial (vascular)
Epithelial (Hillock)
Epithelial (basal)
Epithelial (club)
Epithelial (luminal)
Fibroblast
Immune (B cell)
Immune (NK cell)
Immune (T cell)
Immune (macrophage)
Immune (mast cell)
Immune (neutrophil)
Neuroendocrine
Pericyte/SMC
Schwann cell
B
Protocol
scRNA-seq (Henry et al.)
snRNA-seq (TST)
C
Sample
snRNA-seq (TST, 01)
snRNA-seq (TST, 02)
snRNA-seq (TST, 03)
snRNA-seq (TST, 04)
scRNA-seq (d17)
scRNA-seq (d27)
scRNA-seq (pd)
D
Granular cell types
snRNA-seq (TST, 01)
snRNA-seq (TST, 02)
snRNA-seq (TST, 03)
snRNA-seq (TST, 04)
scRNA-seq (d17)
scRNA-seq (d27)
scRNA-seq (pd)
Cell type composition (%)
E
Granular cell types (color legend D and E)
Endothelial (lymphatic)
Endothelial (vascular I)
Endothelial (vascular II)
Endothelial (vascular III)
Epithelial (Hillock)
Epithelial (basal I)
Epithelial (basal II)
Epithelial (club)
Epithelial (luminal)
Fibroblast
Immune (B cell)
Immune (NK cell)
Immune (T cell)
Immune (macrophage I)
Immune (macrophage II)
Immune (mast cell)
Immune (neutrophil)
Myonuclei (smooth muscle)
Neuroendocrine
Pericyte/SMC
Schwann cell
F
Sample
scRNA-seq (d17)
scRNA-seq (d27)
scRNA-seq (pd)
G
Epithelial (basal)
Epithelial (luminal)
Epi. (luminal, stressed)
Epithelial (club)
Epithelial (Hillock)
Epithelial (unknown)
Fibroblast
Pericyte
SMC
Endothelial (vascular)
Endothelial (lymphatic)
Macrophage
Other
Neuroendocrine
Manual scRNA-seq annotations
Predicted annotations from classifier trained on snRNA-seq
Proportion of cells
H
Number of cells
Fraction of cells in group (%)
Mean expression in group
Skin
I
Broad cell types
Adipocyte
Endothelial (lymphatic)
Endothelial (vascular)
Epithelial (basal keratinocyte)
Epithelial (cornified keratinocyte)
Epithelial (mature keratinocyte)
Epithelial (suprabasal keratinocyte)
Fibroblast
Immune (DC/macrophage)
Immune (T cell)
Immune (mast cell)
Melanocyte
Pericyte/SMC
Sebaceous gland
Sweat gland
Unknown
J
Protocol
scRNA-seq (this study)
snRNA-seq (CST)
snRNA-seq (TST)
K
Sample
scRNA-seq (sample 1)
scRNA-seq (sample 2)
scRNA-seq (sample 3)
scRNA-seq (sample 4)
scRNA-seq (CST, 01)
scRNA-seq (TST, 01)
L
Granular cell types
scRNA-seq (sample 1)
scRNA-seq (sample 2)
scRNA-seq (sample 3)
scRNA-seq (sample 4)
scRNA-seq (CST, 01)
scRNA-seq (TST, 01)
Cell type composition (%)
M
Granular cell types (color legend L and M)
Adipocyte
Endothelial (lymphatic)
Endothelial (vascular I)
Endothelial (vascular II)
Epi. (basal keratinocyte)
Epi. (cornified keratinocyte)
Epi. (mature keratinocyte)
Epi. (suprabasal keratinocyte)
Fibroblast
Immune (DC/macrophage)
Immune (Langerhans)
Immune (T cell)
Immune (mast cell)
Melanocyte
Pericyte/SMC I
Pericyte/SMC II
Sebaceous gland (CYP4F3+)
Sebaceous gland (KRT79h)
Sebaceous gland (SEC14L6+)
Sweat gland
Unknown (ATP1B3-GJB6)
Unknown (SFN-GJA1)
N
Sample
scRNA-seq (sample 1)
scRNA-seq (sample 2)
scRNA-seq (sample 3)
scRNA-seq (sample 4)
O
Epi. (basal keratinocyte)
Epi. (suprabasal keratinocyte)
Melanocyte
Sweat gland
Fibroblast
Pericyte
Mesenchymal
SMC
Endothelial (vascular)
Endothelial (lymphatic)
Immune (Langerhans)
Immune (DC)
Immune (T cell)
Preadipocyte/Fibroblast
Unknown
Manual scRNA-seq annotations
Predicted annotations from classifier trained on snRNA-seq
Proportion of cells
P
Number of cells
Fraction of cells in group (%)
Mean expression in group

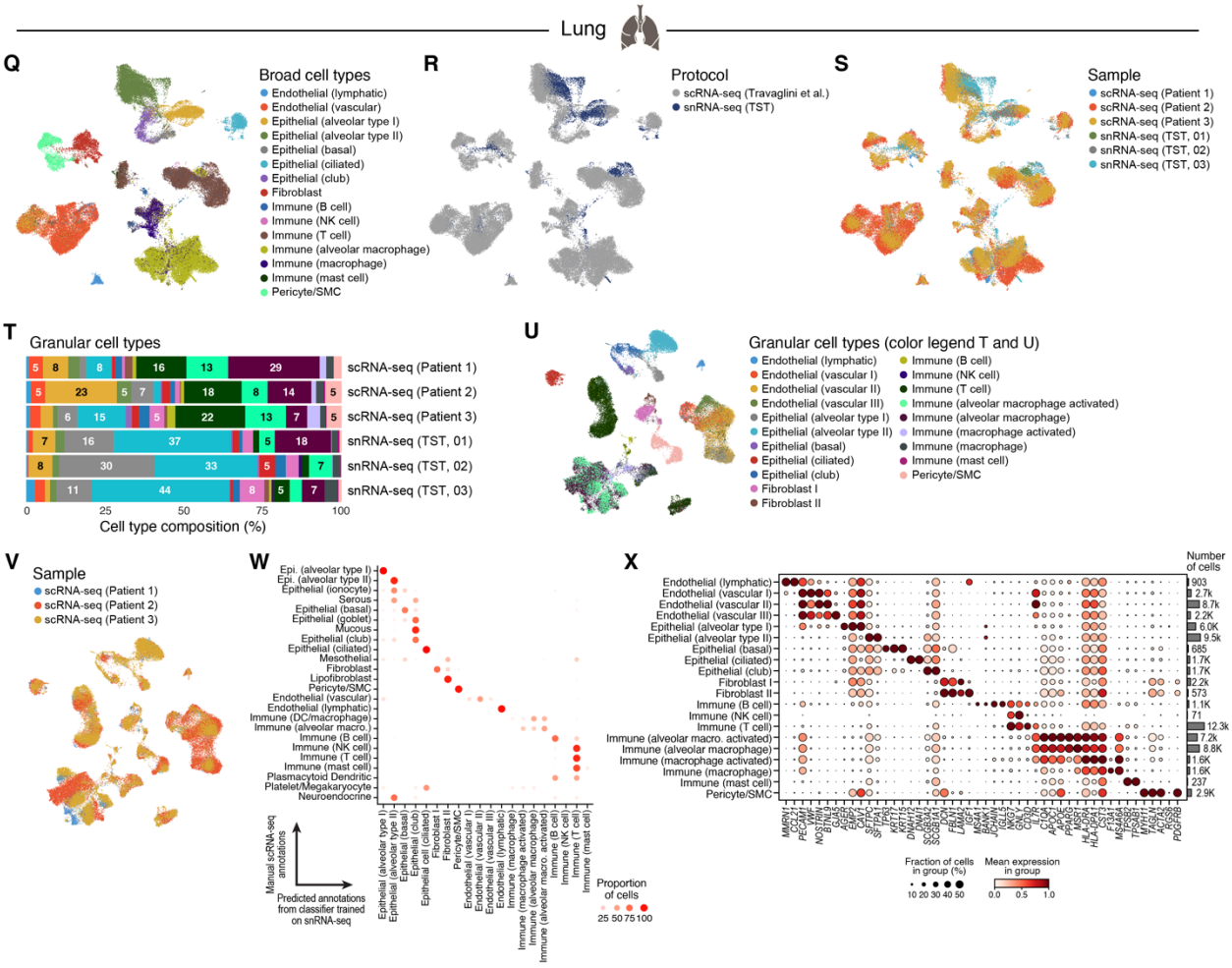


Fig. S13. Integration and annotation of single-cell and single-nucleus RNA-seq prostate, skin and lung datasets. (A–H) Integration of scRNA-seq and snRNA-seq in the prostate (*28*). (**A–C**) UMAP representation of integrated scRNA-seq (Henry et al. (*38*))) and snRNA-Seq (current study, TST) colored by predicted granular cell types collapsed into broad categories (A), protocol (B) or sample (C). (**D**) Comparison of cell composition between scRNA-seq and snRNA-seq. Proportion of predicted granular cell types (bar color) based on a classifier trained on snRNA-seq data. Numbers on bars: proportion of cell type within sample. (**E,F**) Cell type annotation in scRNA-seq. UMAP representation only of scRNA-seq profiles (Henry et al. (*38*)) colored by predicted granular cell types (E) or samples (F). (**G,H**) Agreement in cell-intrinsic expression profiles between single-

cell and single-nucleus datasets. (G) Proportion of cells (circle size and color) of each subset (rows) that are predicted to be in each nucleus class (columns) by a random forest classifier trained on nuclei for prostate. (H) Scaled mean scRNA-seq expression (*z*-score, circle color) and fraction of expressing cells (circle size) of granular cell type marker genes (as in **Fig. S6**, columns, labels at bottom) associated with those subsets (rows, with nuclei number on right).

(I–P) Integration of scRNA-seq and snRNA-seq in the skin (*28*). (**I–K**) UMAP representation of integrated scRNA-seq (current study) and snRNA-seq (current study, CST and TST) colored by predicted granular cell types collapsed into broad categories (I), protocol (J) or sample (K). (**L**) Comparison of cell composition between scRNA-seq and snRNA-seq. Proportion of predicted granular cell types (bar color) based on a classifier trained on snRNA-seq data. Numbers on bars: proportion of cell type within sample. (**M,N**) Cell type annotation in scRNA-Seq. UMAP representation only of scRNA-seq profiles (current study) colored by predicted granular cell types (M) or samples (M). (**O-P**) Agreement in cell-intrinsic expression profiles between single-cell and single-nucleus datasets. (O) Proportion of cells (circle size and color) of each subset (rows) that are predicted to be in each nucleus class (columns) by a random forest classifier trained on nuclei for prostate. (P) Scaled mean scRNA-seq expression (*z*-score, circle color) and fraction of expressing cells (circle size) of granular cell type marker genes (as in **Fig. S6**, columns, labels at bottom) associated with those subsets (rows, with nuclei number on right).

(Q–X) Integration of scRNA-seq and snRNA-seq in the lung (*28*). (**Q–S**) UMAP representation of integrated scRNA-seq (Travaglini et al.(*40*)) and snRNA-seq (current study, TST) colored by predicted granular cell types collapsed into broad categories (Q), protocol (R) or sample (S). (**T**) Comparison of cell composition between scRNA-seq and snRNA-seq. Proportion of predicted granular cell types (bar color) based on a classifier trained on snRNA-seq data. Numbers on bars: proportion of cell type within sample. (**U,V**) Cell type annotation in scRNA-Seq. UMAP

representation only of scRNA-seq profiles (Travaglini et al. (*40*)) colored by predicted granular cell types (U) or samples (V). (**W,X**) Agreement in cell-intrinsic expression profiles between single-cell and single-nucleus datasets. (W) Proportion of cells (circle size and color) of each subset (rows) that are predicted to be in each nucleus class (columns) by a random forest classifier trained on nuclei for prostate. (X) Scaled mean scRNA-seq expression (*z*-score, circle color) and fraction of expressing cells (circle size) of granular cell type marker genes (as in **Fig. S6**, columns, labels at bottom) associated with those subsets (rows, with nuclei number on right).

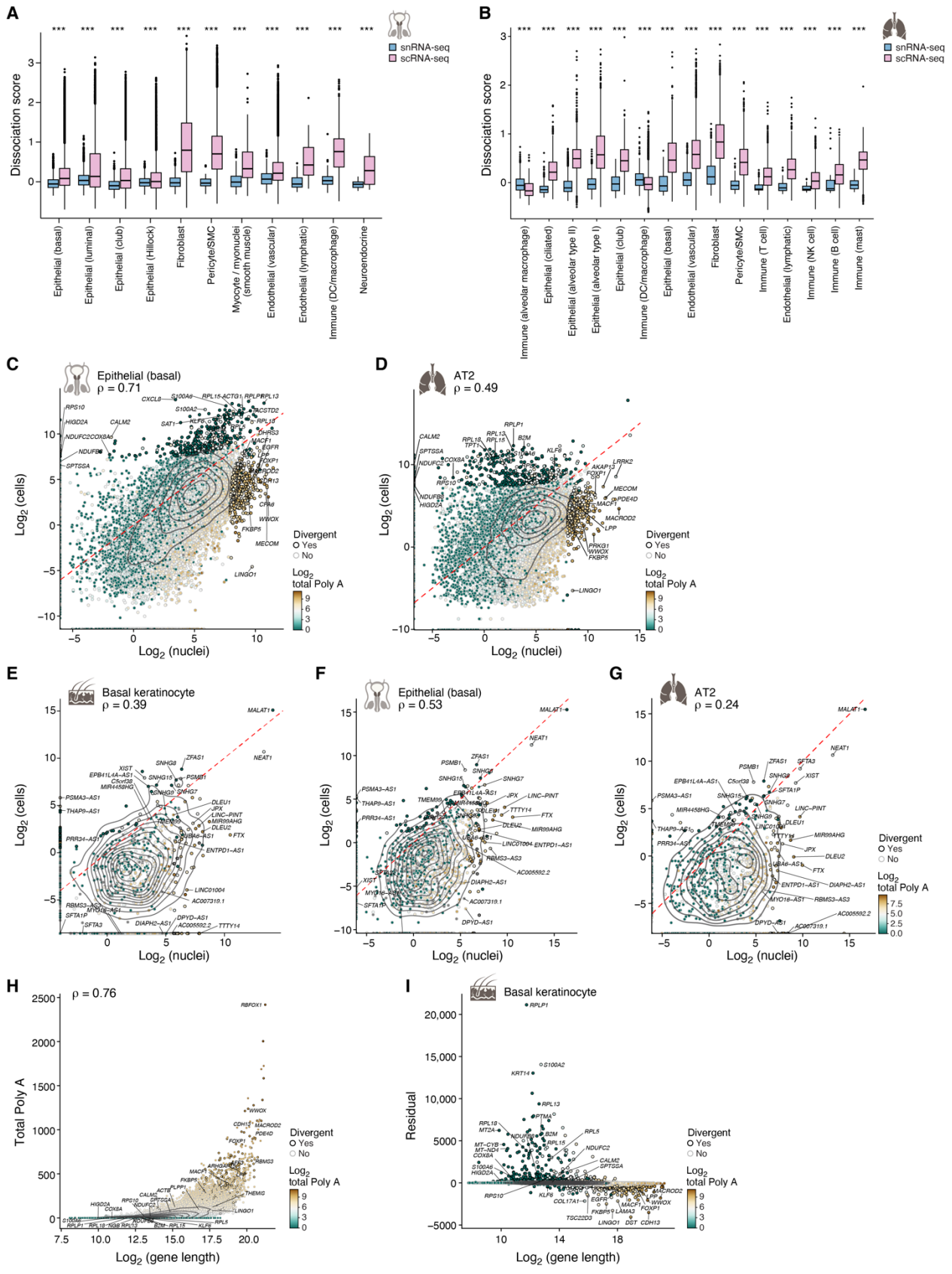

A
B
snRNA-seq
scRNA-seq
Dissociation score
Epithelial (basal)
Epithelial (luminal)
Epithelial (club)
Epithelial (Hillock)
Fibroblast
Pericyte/SMC
Myocyte / myonuclei (smooth muscle)
Endothelial (vascular)
Endothelial (lymphatic)
Immune (DC/macrophage)
Neuroendocrine
Immune (alveolar macrophage)
Epithelial (ciliated)
Epithelial (alveolar type II)
Epithelial (alveolar type I)
Immune (T cell)
Immune (NK cell)
Immune (B cell)
Immune (mast)
C
Epithelial (basal)
ρ = 0.71
D
AT2
ρ = 0.49
Log2 (cells)
Log2 (nuclei)
Divergent
Yes
No
Log2 total Poly A
E
Basal keratinocyte
ρ = 0.39
F
Epithelial (basal)
ρ = 0.53
G
AT2
ρ = 0.24
H
ρ = 0.76
Total Poly A
Log2 (gene length)
I
Basal keratinocyte
Residual

Fig. S14. Differences in gene expression between snRNA-seq and scRNA-seq

(A, B) Expression of tissue dissociation signatures are accentuated in scRNA-seq but not snRNA-seq. Distribution of the score (*y*-axis, average background corrected, log(TP10K+1)) of a dissociation signature (*41*) in scRNA-seq (pink) and snRNA-seq (blue) profiles in each major lung cell type category (*x*-axis). (*** Benjamini-Hochberg FDR < 10^{-16}, Wilcoxon rank-sum test). Box plots show median, quartiles, and whiskers at 1.5 IQR (interquartile range). (**C–G**) Limited divergence of gene expression between scRNA-seq and snRNA-Seq. Averaged pseudobulk expression values (*28*) of genes (dots) in nuclei (*x*-axis) and cells (*y*-axis), shown for protein-coding genes in (**C**) prostate epithelial (basal) cells **(D)** lung alveolar type II cells, and long non-coding (LINC) RNA genes in (**E**) skin basal keratinocytes, (**F**) prostate epithelial cells (basal), and (**G**) lung alveolar type II cells. Divergent genes (dot outline color) deviating from straight line regression fit by size of residuals (*28*). Color scale: total length of polyA stretches with at least 20 adenine bases. (**H, I**) Relation between gene expression differences in nuclei *vs*. cells, gene length and polyA stretches. **(H)** The number of polyA stretches (*y*-axis) and length (*x*-axis) of each gene. (**I**) Divergence (*y*-axis, residual of straight-line regression fit) between pseudobulk gene expression of single cell and single nucleus RNA-seq and gene length (*x*-axis) for each protein coding gene expressed in skin basal keratinocytes in both datasets.

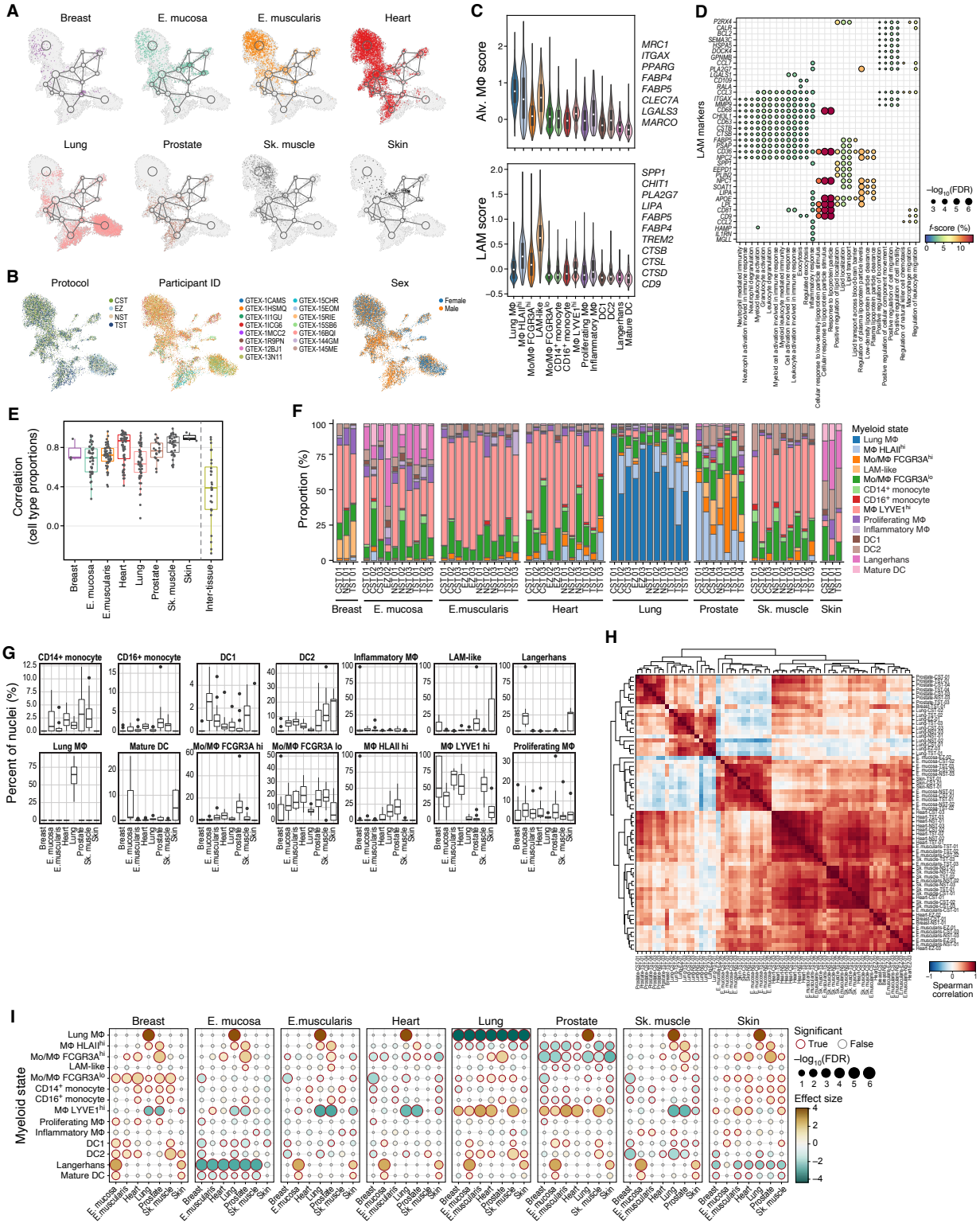

A
Breast
E. mucosa
E. muscularis
Heart
Lung
Prostate
Sk. muscle
Skin
B
Protocol
CST
EZ
NST
TST
Participant ID
Sex
Female
Male
C
Alv. MΦ score
MRC1
ITGAX
PPARG
FABP4
FABP5
CLEC7A
LGALS3
MARCO
LAM score
SPP1
CHIT1
PLA2G7
LIPA
FABP5
FABP4
TREM2
CTSB
CTSL
CTSD
CD9
Lung MΦ
MΦ HLAII hi
Mo/MΦ FCGR3A hi
Mo/MΦ FCGR3A lo
CD14+ monocyte
CD16+ monocyte
MΦ LYVE1 hi
Proliferating MΦ
Inflammatory MΦ
DC1
DC2
Langerhans
Mature DC
D
LAM markers
−log10(FDR)
f-score (%)
E
Correlation (cell type proportions)
Breast
E. mucosa
E.muscularis
Heart
Lung
Prostate
Sk. muscle
Skin
Inter-tissue
F
Proportion (%)
Myeloid state
Lung MΦ
MΦ HLAII hi
Mo/MΦ FCGR3A hi
LAM-like
Mo/MΦ FCGR3A lo
CD14+ monocyte
CD16+ monocyte
MΦ LYVE1 hi
Proliferating MΦ
Inflammatory MΦ
DC1
DC2
Langerhans
Mature DC
G
Percent of nuclei (%)
CD14+ monocyte
CD16+ monocyte
DC1
DC2
Inflammatory MΦ
LAM-like
Langerhans
Lung MΦ
Mature DC
Mo/MΦ FCGR3A hi
Mo/MΦ FCGR3A lo
MΦ HLAII hi
MΦ LYVE1 hi
Proliferating MΦ
H
Spearman correlation
I
Myeloid state
Significant
True
False
−log10(FDR)
Effect size

Fig. S15. Cross-tissue analysis of myeloid cell subsets. (A, B) Myeloid cell subsets across tissues. UMAP visualization of myeloid snRNA-seq profiles, highlighting cells from each tissue **(A)** or colored by protocol, donor, or sex **(B)**. **(C)** Alveolar and lipid-associated macrophage signatures. Distribution of signature scores (*y*-axis) of alveolar macrophages (top) and lipid-associated macrophage (LAM) (bottom) for each myeloid cell subset (*x*-axis, (*28*)). Signature genes are listed on the right. **(D)** Functional enrichment of LAM markers. Significance (-$\log_{10}$(FDR) circle size) of enrichment (F-score, circle color) of top 30 GO terms (rows) in LAM markers gene set (columns), showing LAM marker membership in each GO gene set. (**E–I**) Myeloid cell composition is similar in the same tissue and different between tissues. (**E**) Distribution of pairwise Spearman correlation coefficients (*y*-axis) of myeloid cell subset proportion profiles for samples within each tissue and between different tissues (*x*-axis). Box plots show median, quartiles, and whiskers at 1.5 IQR (interquartile range). **(F)** Proportion (%, *y*-axis) of different myeloid subsets in each sample (*x*-axis). (**G**) Distribution of proportions (*y*-axis) for each broad cell type (columns) within each tissue (*x*-axis). Box plots show median, quartiles, and whiskers at 1.5 IQR (interquartile range). **(H)** Spearman correlation coefficient (color bar) of myeloid cell subset proportion profiles of each pair of samples (columns, rows). Rows and columns were hierarchically clustered using the Euclidean distance and complete linkage. **(I)** Differences in myeloid cell distributions across tissues. Significance (circle size, $-\log_{10}$(FDR)) and effect size (circle color) of the difference in proportion in each myeloid cell subset (rows) between each tissue (columns) relative to one reference tissue (label on top).

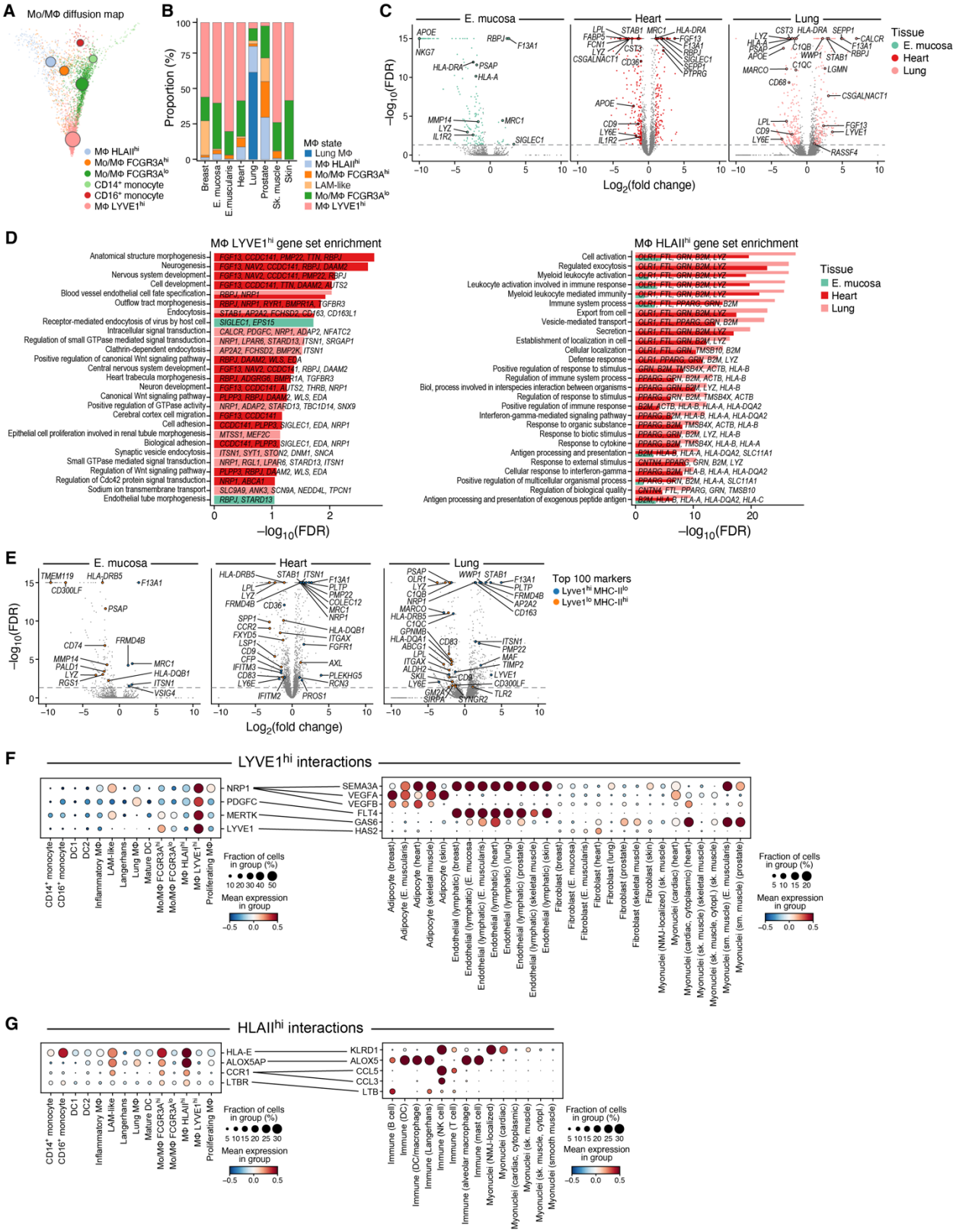

A
Mo/MΦ diffusion map
MΦ HLAII^hi
Mo/MΦ FCGR3A^hi
Mo/MΦ FCGR3A^lo
CD14+ monocyte
CD16+ monocyte
MΦ LYVE1^hi
B
Proportion (%)
Breast
E. mucosa
E.muscularis
Heart
Lung
Prostate
Sk. muscle
Skin
MΦ state
Lung MΦ
MΦ HLAII^hi
Mo/MΦ FCGR3A^hi
LAM-like
Mo/MΦ FCGR3A^lo
MΦ LYVE1^hi
C
E. mucosa
Heart
Lung
−log10(FDR)
Log2(fold change)
Tissue
E. mucosa
Heart
Lung
D
MΦ LYVE1^hi gene set enrichment
Anatomical structure morphogenesis
Neurogenesis
Nervous system development
Cell development
Blood vessel endothelial cell fate specification
Outflow tract morphogenesis
Endocytosis
Receptor-mediated endocytosis of virus by host cell
Intracellular signal transduction
Regulation of small GTPase mediated signal transduction
Clathrin-dependent endocytosis
Positive regulation of canonical Wnt signaling pathway
Central nervous system development
Heart trabecula morphogenesis
Neuron development
Canonical Wnt signaling pathway
Positive regulation of GTPase activity
Cerebral cortex cell migration
Cell adhesion
Epithelial cell proliferation involved in renal tubule morphogenesis
Biological adhesion
Synaptic vesicle endocytosis
Small GTPase mediated signal transduction
Regulation of Wnt signaling pathway
Regulation of Cdc42 protein signal transduction
Sodium ion transmembrane transport
Endothelial tube morphogenesis
−log10(FDR)
MΦ HLAII^hi gene set enrichment
Cell activation
Regulated exocytosis
Myeloid leukocyte activation
Leukocyte activation involved in immune response
Myeloid leukocyte mediated immunity
Immune system process
Export from cell
Vesicle-mediated transport
Secretion
Establishment of localization in cell
Cellular localization
Defense response
Positive regulation of response to stimulus
Regulation of immune system process
Biol. process involved in interspecies interaction between organisms
Regulation of response to stimulus
Positive regulation of immune response
Interferon-gamma-mediated signaling pathway
Response to organic substance
Response to biotic stimulus
Response to cytokine
Antigen processing and presentation
Response to external stimulus
Cellular response to interferon-gamma
Positive regulation of multicellular organismal process
Regulation of biological quality
Antigen processing and presentation of exogenous peptide antigen
−log10(FDR)
Tissue
E. mucosa
Heart
Lung
E
E. mucosa
Heart
Lung
−log10(FDR)
Log2(fold change)
Top 100 markers
Lyve1^hi MHC-II^lo
Lyve1^lo MHC-II^hi
F
LYVE1^hi interactions
NRP1
PDGFC
MERTK
LYVE1
SEMA3A
VEGFA
VEGFB
FLT4
GAS6
HAS2
CD14+ monocyte
CD16+ monocyte
DC1
DC2
Inflammatory MΦ
LAM-like
Langerhans
Lung MΦ
Mature DC
Mo/MΦ FCGR3A^hi
Mo/MΦ FCGR3A^lo
MΦ HLAII^hi
MΦ LYVE1^hi
Proliferating MΦ
Fraction of cells in group (%)
Mean expression in group
Adipocyte (breast)
Adipocyte (E. muscularis)
Adipocyte (heart)
Adipocyte (skeletal muscle)
Adipocyte (skin)
Endothelial (lymphatic) (breast)
Endothelial (lymphatic) (E. mucosa)
Endothelial (lymphatic) (E. muscularis)
Endothelial (lymphatic) (heart)
Endothelial (lymphatic) (lung)
Endothelial (lymphatic) (prostate)
Endothelial (lymphatic) (skeletal muscle)
Endothelial (lymphatic) (skin)
Fibroblast (breast)
Fibroblast (E. mucosa)
Fibroblast (E. muscularis)
Fibroblast (heart)
Fibroblast (lung)
Fibroblast (prostate)
Fibroblast (skeletal muscle)
Fibroblast (skin)
Myonuclei (NMJ-localized) (sk. muscle)
Myonuclei (cardiac) (heart)
Myonuclei (cardiac, cytoplasmic) (heart)
Myonuclei (sk. muscle) (skeletal muscle)
Myonuclei (sk. muscle, cytopl.) (sk. muscle)
Myonuclei (sm. muscle) (E. muscularis)
Myonuclei (sm. muscle) (prostate)
Fraction of cells in group (%)
Mean expression in group
G
HLAII^hi interactions
HLA-E
ALOX5AP
CCR1
LTBR
KLRD1
ALOX5
CCL5
CCL3
LTB
CD14+ monocyte
CD16+ monocyte
DC1
DC2
Inflammatory MΦ
LAM-like
Langerhans
Lung MΦ
Mature DC
Mo/MΦ FCGR3A^hi
Mo/MΦ FCGR3A^lo
MΦ HLAII^hi
MΦ LYVE1^hi
Proliferating MΦ
Fraction of cells in group (%)
Mean expression in group
Immune (B cell)
Immune (DC)
Immune (DC/macrophage)
Immune (Langerhans)
Immune (NK cell)
Immune (T cell)
Immune (alveolar macrophage)
Immune (mast cell)
Myonuclei (NMJ-localized)
Myonuclei (cardiac)
Myonuclei (cardiac, cytoplasmic)
Myonuclei (sk. muscle)
Myonuclei (sk. muscle, cytopl.)
Myonuclei (smooth muscle)
Fraction of cells in group (%)
Mean expression in group

Fig S16. Differential expression, functional enrichment, and receptor-ligand interactions of dichotomous macrophage states. (**A**) Diffusion map of monocytes, macrophages, and transitional monocyte/macrophage subset (colors) profiles (dots) in heart. Large circles: Centroids of each state (sizes proportional to the population size). (**B**) Macrophage cell proportions across tissues. Proportion (%, *y*-axis) of different macrophage subsets in each individual tissue (x-axis). (**C-E**) Mϕ *LYVE1*^{hi} and Mϕ *HLA*II^{hi} distinctive signatures. (C) Differential expression (x-axis, $\log_2$(Fold change), $x < 0$: enriched in Mϕ *HLA*II^{hi}; $x > 0$: enriched in Mϕ *LYVE1*^{hi}) and its significance ($-\log_{10}$(FDR (Benjamini-Hochberg)), Wald test) between Mϕ *LYVE1*^{hi} and Mϕ *HLA*II^{hi} profiles in the three tissues where both populations are observed. (D) Significance of enrichment (x-axis, $-\log_{10}$(Benjamini-Hochberg FDR), Fisher's exact test) of GO gene sets (y-axis) in genes differentially expressed between Mϕ *LYVE1*^{hi} and Mϕ *HLA*II^{hi} populations in the three tissues in B. (E) Differential expression plots as in (C) with marker genes of mouse *Lyve1*^{hi}MHCII^{lo} and *Lyve1*^{lo}MHCII^{hi} cells (*47*) highlighted. (**F-G**) Putative receptor-ligand interactions between dichotomous Mϕ states and cells of broad cell types. Scaled mean expression (z-score, circle color) and fraction of expressing cells (circle size) of receptor or ligand genes highly expressed by Mϕ *LYVE1*^{hi} (F, left), broad cell types that are potentially interacting with Mϕ *LYVE1*^{hi} (F, right), Mϕ *HLA*II^{hi} (G, left) and broad cell types that are potentially interacting with Mϕ *HLA*II^{hi} (G, right), (columns, labels at bottom). Lines between the genes show potential cell-cell interactions.

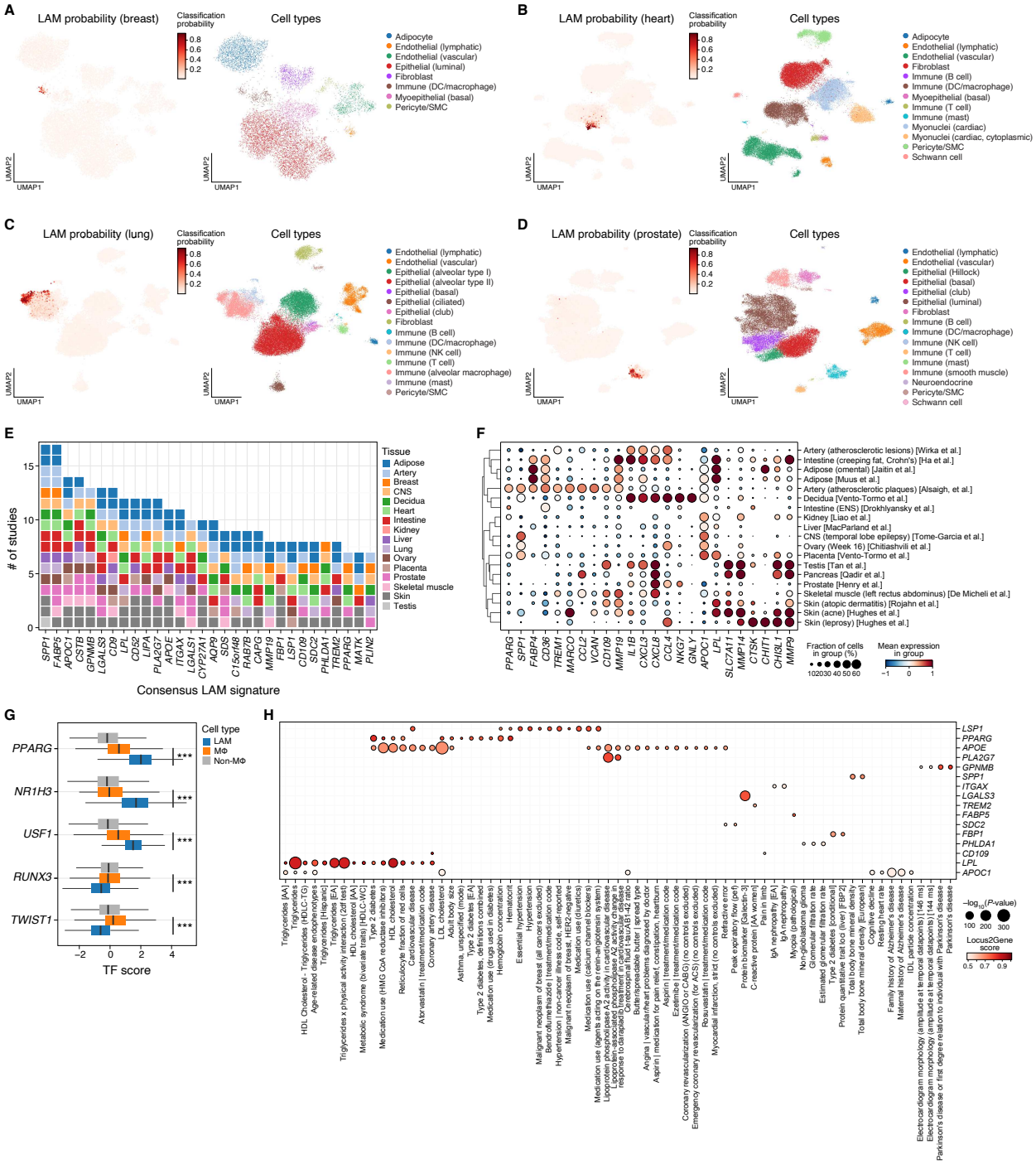

A
LAM probability (breast)
Classification probability
0.8
0.6
0.4
0.2
Cell types
Adipocyte
Endothelial (lymphatic)
Endothelial (vascular)
Epithelial (luminal)
Fibroblast
Immune (DC/macrophage)
Myoepithelial (basal)
Pericyte/SMC
UMAP1
UMAP2
B
LAM probability (heart)
Cell types
Adipocyte
Endothelial (lymphatic)
Endothelial (vascular)
Fibroblast
Immune (B cell)
Immune (DC/macrophage)
Myoepithelial (basal)
Immune (T cell)
Immune (mast)
Myonuclei (cardiac)
Myonuclei (cardiac, cytoplasmic)
Pericyte/SMC
Schwann cell
C
LAM probability (lung)
Cell types
Endothelial (lymphatic)
Endothelial (vascular)
Epithelial (alveolar type I)
Epithelial (alveolar type II)
Epithelial (basal)
Epithelial (ciliated)
Epithelial (club)
Fibroblast
Immune (B cell)
Immune (DC/macrophage)
Immune (NK cell)
Immune (T cell)
Immune (alveolar macrophage)
Immune (mast)
Pericyte/SMC
D
LAM probability (prostate)
Cell types
Endothelial (lymphatic)
Endothelial (vascular)
Epithelial (Hillock)
Epithelial (basal)
Epithelial (club)
Epithelial (luminal)
Fibroblast
Immune (B cell)
Immune (DC/macrophage)
Immune (NK cell)
Immune (T cell)
Immune (mast)
Immune (smooth muscle)
Neuroendocrine
Pericyte/SMC
Schwann cell
E
# of studies
Tissue
Adipose
Artery
Breast
CNS
Decidua
Heart
Intestine
Kidney
Liver
Lung
Ovary
Placenta
Prostate
Skeletal muscle
Skin
Testis
SPP1
FABP5
APOC1
CSTB
GPNMB
LGALS3
CD9
LPL
CD52
LIPA
PLA2G7
APOE
ITGAX
LGALS1
CYP27A1
AQP9
SDS
C15orf48
RAB7B
CAPG
MMP19
FBP1
LSP1
CD109
SDC2
PHLDA1
TREM2
PPARG
MATK
PLIN2
Consensus LAM signature
F
Artery (atherosclerotic lesions) [Wirka et al.]
Intestine (creeping fat, Crohn's) [Ha et al.]
Adipose (omental) [Jaitin et al.]
Adipose [Muus et al.]
Artery (atherosclerotic plaques) [Alsaigh, et al.]
Decidua [Vento-Tormo et al.]
Intestine (ENS) [Drokhlyansky et al.]
Kidney [Liao et al.]
Liver [MacParland et al.]
CNS (temporal lobe epilepsy) [Tome-Garcia et al.]
Ovary (Week 16) [Chitiashvili et al.]
Placenta [Vento-Tormo et al.]
Testis [Tan et al.]
Pancreas [Qadir et al.]
Prostate [Henry et al.]
Skeletal muscle (left rectus abdominus) [De Micheli et al.]
Skin (atopic dermatitis) [Rojahn et al.]
Skin (acne) [Hughes et al.]
Skin (leprosy) [Hughes et al.]
PPARG
SPP1
FABP4
CD36
TREM1
MARCO
CCL2
VCAN
CD109
MMP19
IL1B
CXCL3
CXCL8
CCL4
NKG7
GNLY
APOC1
LPL
SLC7A11
MMP14
CTSK
CHIT1
CHI3L1
MMP9
Fraction of cells in group (%)
Mean expression in group
G
Cell type
LAM
MΦ
Non-MΦ
PPARG
NR1H3
USF1
RUNX3
TWIST1
***
TF score
H
LSP1
PPARG
APOE
PLA2G7
GPNMB
SPP1
ITGAX
LGALS3
TREM2
FABP5
SDC2
FBP1
PHLDA1
CD109
LPL
APOC1
Triglycerides [AA]
Triglycerides
HDL Cholesterol - Triglycerides (HDLC:TG)
Age-related disease endophenotypes
Triglycerides [Hispanic]
Triglycerides [EA]
Triglycerides x physical activity interaction (2df test)
HDL cholesterol [AA]
Metabolic syndrome (bivariate trait) [HDLC:WC]
Type 2 diabetes
Medication use (HMG CoA reductase inhibitors)
HDL cholesterol
Reticulocyte fraction of red cells
Cardiovascular disease
Atorvastatin | treatment/medication code
Coronary artery disease
LDL cholesterol
Adult body size
Asthma, unspecified (mode)
Type 2 diabetes [EA]
Type 2 diabetes, self-report only
Medication use (drugs used in diabetes)
Hemoglobin concentration
Hematocrit
Essential hypertension
Hypertension
Malignant neoplasm of breast (all cancers excluded)
Benign neoplasm of breast
Hypertension | non-cancer illness code, self-reported
Malignant neoplasm of breast, HER2-negative
Medication use (diuretics)
Medication use (calcium channel blockers)
Medication use (agents acting on the renin-angiotensin system)
Lipoprotein-associated phospholipase A2 activity and mass
Lipoprotein-associated phospholipase A2 activity change in response to darapladib treatment in patients with CHD
Cholesterol, total
Butter/spreadable butter | spread type
Angina | vascular/heart problems diagnosed by doctor
Aspirin | medication for pain relief, constipation, heartburn
Aspirin | treatment/medication code
Ezetimibe | treatment/medication code
Coronary revascularization (ANGIO or CABG) (no controls excluded)
Emergency coronary revascularization (for ACS) (no controls excluded)
Rosuvastatin | treatment/medication code
Myocardial infarction, strict (no controls excluded)
Reflective error
Peak expiratory flow rate
Protein biomarker [Galectin-3]
C-reactive protein [AA women]
Pain in limb
IgA nephropathy
Myocardial infarction
Non-glioblastoma glioma
Glomerular filtration rate
Estimated glomerular filtration rate
Type 2 diabetes [conditional]
Protein quantitative trait loci (liver) [FB2]
Total body bone mineral density [EU consort]
Total body bone mineral density [EU consort]
Cognitive decline
Resting heart rate
Family history of Alzheimer's disease
Maternal history of Alzheimer's disease
IDL particle concentration
Electrocardiogram morphology (amplitude at temporal datapoints) [145 ms]
Electrocardiogram morphology (amplitude at temporal datapoints) [144 ms]
Parkinson's disease or first degree relation to individual with Parkinson's disease
Parkinson's disease
−log10(P-value)
Locus2Gene score

Fig. S17. Tissue distribution and GWAS associations of LAM-like cells. (A–D) LAM-like cells across tissues. UMAP visualization of nuclei profiles from breast **(A)**, heart **(B)**, lung **(C)** and prostate **(D)** (where LAM-like profiles are detected), colored by classification probabilities of LAMs (left) or by broad cell type annotations (right). **(E)** Consensus LAM signature. The number of published studies (*y*-axis) and tissue type (color) in which each LAM marker gene (*x*-axis, (*28*)) is detected as a marker. **(F)** Variable genes across LAMs. Scaled mean expression (z-score, circle color) and fraction of expressing cells (circle size) of the genes (columns) variable across the scRNA-seq datasets (rows). **(G)** LAM-associated TFs. Distribution of activity scores (*x*-axis) of TFs (*y*-axis) that are significantly high in LAMs or non-LAM macrophages. *** = $P < 10^{-20}$ (*p*-values in **Fig. 3J** are combined across studies using Fisher's method). **(H)** LAM marker genes related to GWAS genes. LAM marker genes (rows) and GWAS (columns) where the genome-wide significant variants (dot size, $-\log_{10}$(P value) of the best variant) are associated with the marker genes with high score (> 0.5, Open Targets Genetics Locus-to-gene mapping, dot color).

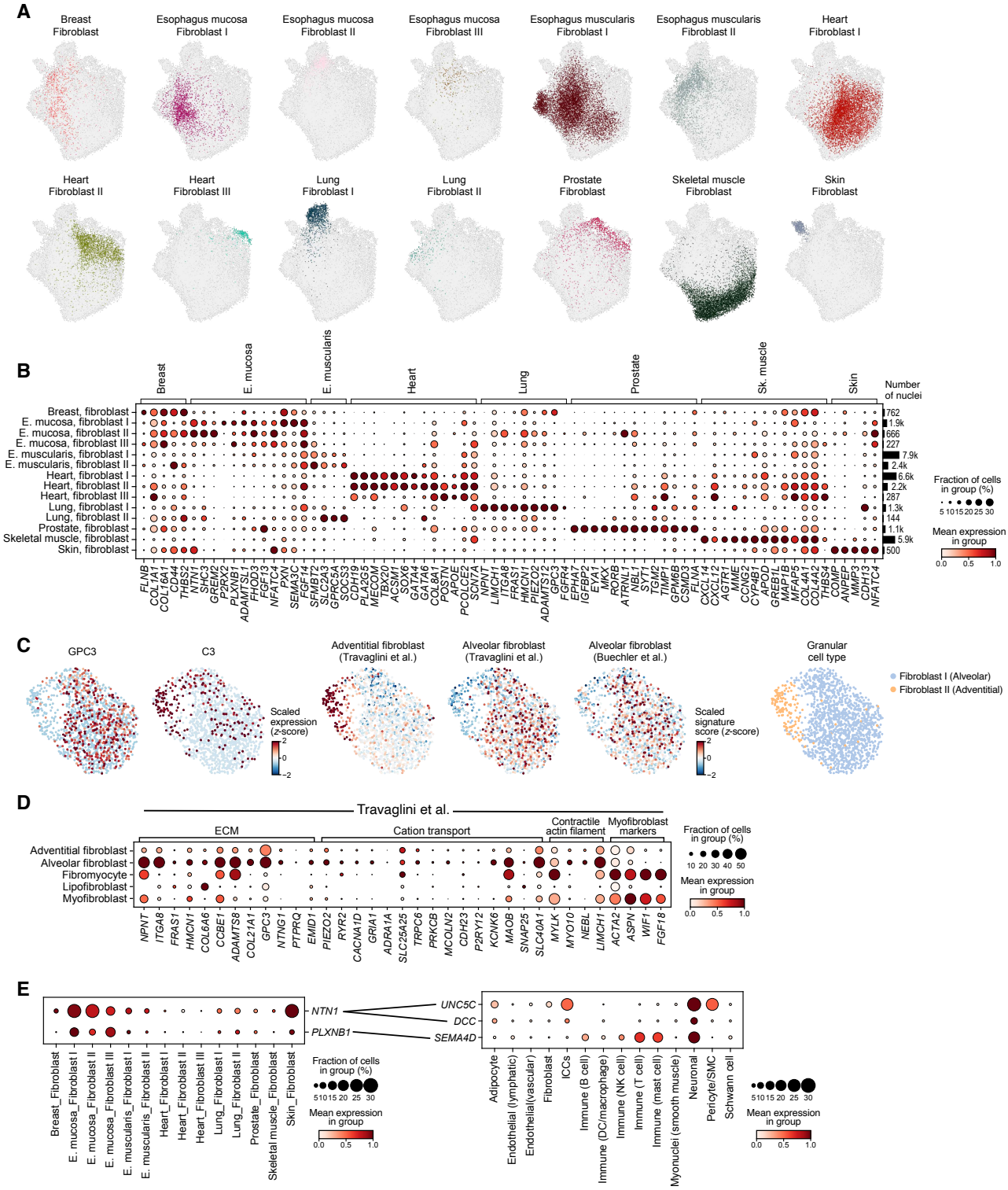
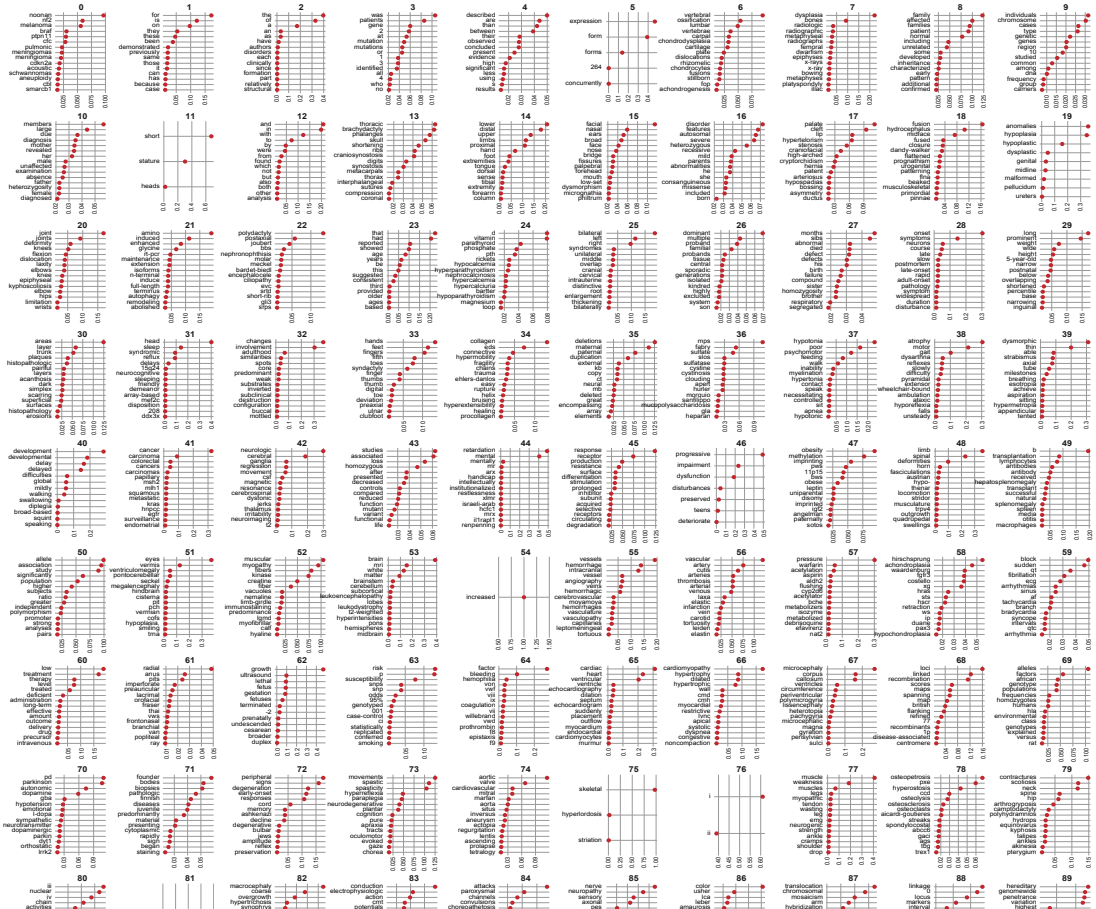

A
Breast Fibroblast
Esophagus mucosa Fibroblast I
Esophagus mucosa Fibroblast II
Esophagus mucosa Fibroblast III
Esophagus muscularis Fibroblast I
Esophagus muscularis Fibroblast II
Heart Fibroblast I
Heart Fibroblast II
Heart Fibroblast III
Lung Fibroblast I
Lung Fibroblast II
Prostate Fibroblast
Skeletal muscle Fibroblast
Skin Fibroblast
B
Breast
E. mucosa
E. muscularis
Heart
Lung
Prostate
Sk. muscle
Skin
Number of nuclei
Breast, fibroblast
E. mucosa, fibroblast I
E. mucosa, fibroblast II
E. mucosa, fibroblast III
E. muscularis, fibroblast I
E. muscularis, fibroblast II
Heart, fibroblast I
Heart, fibroblast II
Heart, fibroblast III
Lung, fibroblast I
Lung, fibroblast II
Prostate, fibroblast
Skeletal muscle, fibroblast
Skin, fibroblast
762
1.9k
666
227
7.9k
2.4k
6.6k
2.2k
287
1.3k
144
1.1k
5.9k
500
Fraction of cells in group (%)
Mean expression in group
FLNB COL1A1 COL16A1 CD44 THBS2 NTN1 SHC3 GREM2 P2RX2 PLXNB1 ADAMTSL1 FGFR3 FGF13 NFATC4 PXN SEMA3C FGF14 SFMBT2 SLC2A3 GPRC5A SOCS3 CDH19 PLCB1 MECOM TBX20 ACSM1 SOX6 GATA4 GATA6 COL8A1 POSTN APOE POLR2E SOX11 NPNT LIMCH1 ITGA8 FRAS1 HMCN1 PIEZO2 ADAMTS12 GPC3 FGFR4 EPHA7 IGFBP2 EYA1 MKX RORB ATRNL1 NBL1 SYT1 TGM2 TIMP1 GPM6B CSMD3 ELN CXCL1 CXCL12 AGTR1 MME CCNG2 CYP4B1 APOD GREB1L MAP1B MFAP5 COL4A1 COL4A2 THBS4 COMP ANPEP MMP3 CDH13 NFATC4
C
GPC3
C3
Adventitial fibroblast (Travaglini et al.)
Alveolar fibroblast (Travaglini et al.)
Alveolar fibroblast (Buechler et al.)
Granular cell type
Scaled expression (z-score)
Scaled signature score (z-score)
Fibroblast I (Alveolar)
Fibroblast II (Adventitial)
D
Travaglini et al.
ECM
Cation transport
Contractile actin filament
Myofibroblast markers
Adventitial fibroblast
Alveolar fibroblast
Fibromyocyte
Lipofibroblast
Myofibroblast
Fraction of cells in group (%)
Mean expression in group
NPNT ITGA8 FRAS1 HMCN1 COL6A6 CCBE1 ADAMTS8 COL21A1 GPC3 NTNG1 PTPRQ EMID1 PIEZO2 RYR2 CACNA1D GRIA1 ADRA1A SLC25A25 TRPC6 PRKCB MCOLN2 CDH23 P2RY12 KCNK6 MAOB SNAP25 SLC40A1 MYLK MYO10 NEBL LIMCH1 ACTA2 ASPN WIF1 FGF18
E
NTN1
PLXNB1
UNC5C
DCC
SEMA4D
Breast_Fibroblast
E. mucosa_Fibroblast I
E. mucosa_Fibroblast II
E. mucosa_Fibroblast III
E. muscularis_Fibroblast I
E. muscularis_Fibroblast II
Heart_Fibroblast I
Heart_Fibroblast II
Heart_Fibroblast III
Lung_Fibroblast I
Lung_Fibroblast II
Prostate_Fibroblast
Skeletal muscle_Fibroblast
Skin_Fibroblast
Adipocyte
Endothelial (lymphatic)
Endothelial(vascular)
Fibroblast
ICCs
Immune (B cell)
Immune (DC/macrophage)
Immune (NK cell)
Immune (T cell)
Immune (mast cell)
Myonuclei (smooth muscle)
Neuronal
Pericyte/SMC
Schwann cell
Fraction of cells in group (%)
Mean expression in group

Fig. S18. Tissue-specific fibroblast signatures. (A) Fibroblast subsets across tissues. UMAP representation of fibroblasts profiles colored by granular fibroblast types in each tissue. **(B)** Tissue-enriched markers of fibroblasts across eight tissues. Min-max scaled mean expression (log(TP10K+1), circle color) and fraction of expressing cells (circle size) of marker genes (columns) associated with granular fibroblast annotations in each tissue subset (rows, and labels on top). **(C-D)** Distinct features of lung fibroblasts. (C) UMAP representation of lung fibroblast profiles colored by scaled (*z*-score) expression of individual genes or signatures of adventitial and alveolar fibroblasts (Travaglini et al. (*40*), Buechler et al. (*108*)), or granular lung fibroblast annotations. (D) Min-max scaled mean expression (log(TP10K+1), circle color) and fraction of expressing cells (circle size) of genes from ECM, cation transport, contractile actin filament pathways and myofibroblast markers (pathway labels at top) in granular fibroblast types (rows) in lung scRNA-seq (*40*). **(E)** Putative receptor-ligand interactions between fibroblasts in esophagus mucosa and neurons in esophagus muscularis. Mean expression (min-max scaled, circle color) and fraction of expressing cells (circle size) of receptor or ligand genes highly expressed by esophagus mucosa fibroblasts (**E**, left), and cognate receptor or ligand genes highly expressed by neurons in esophagus muscularis (**E**, right). Lines between the genes show potential cell-cell interactions.


Disease-associated gene loadings of OMIM topics

Fig. S19. Relation of monogenic diseases to disease topics. Scaled loadings (*z*-score, *x*-axis) of top 15 diseases (rows) with the genes associated for each of the topics. Genes of the diseases with multiple gene associations are given as a comma-separated list.


Word loadings of OMIM topics

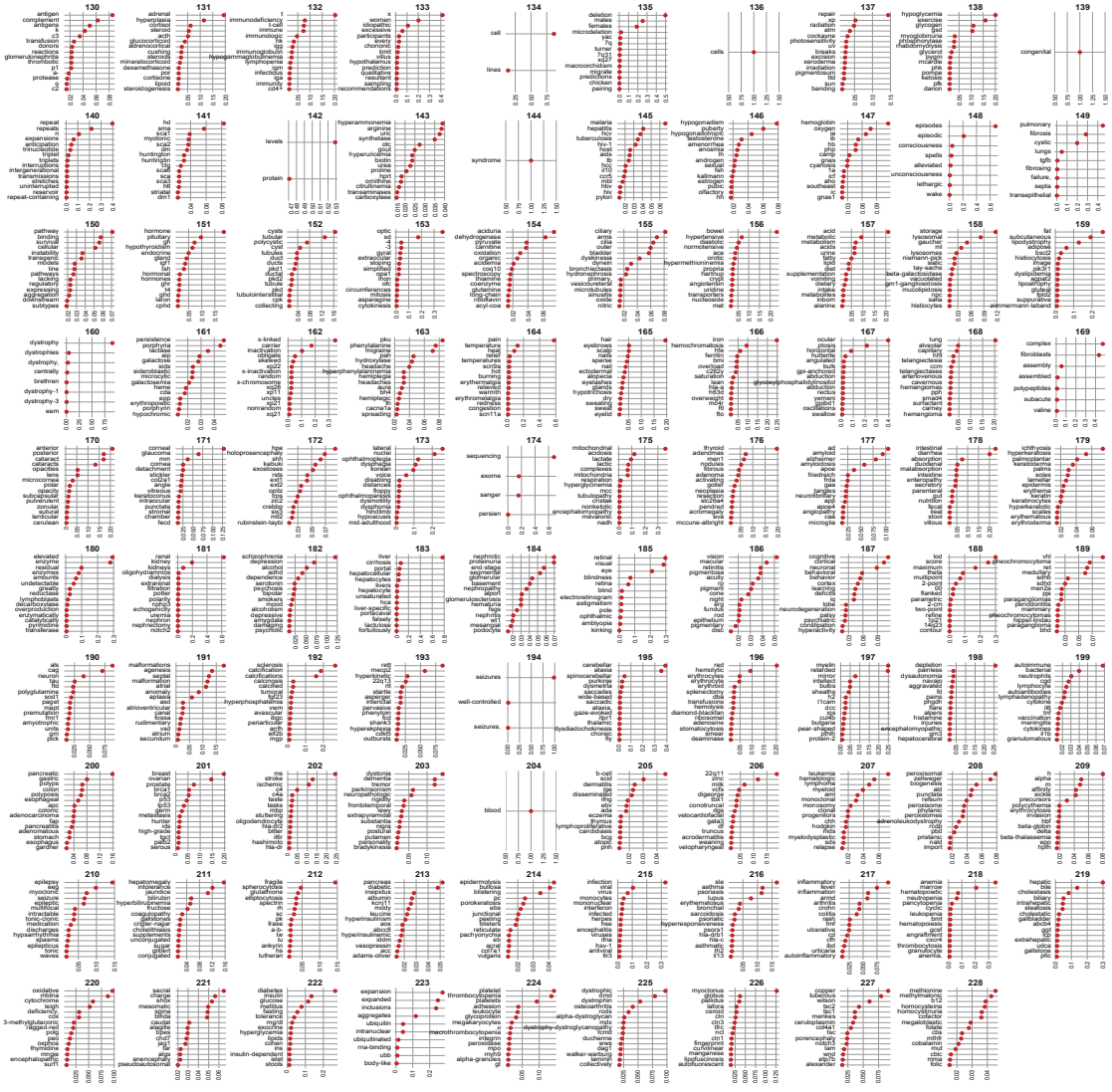


Fig. S20. Loadings of words for each monogenic disease topic. Probabilities (*x*-axis) of top 15 words (rows) for each of the monogenic disease topics.

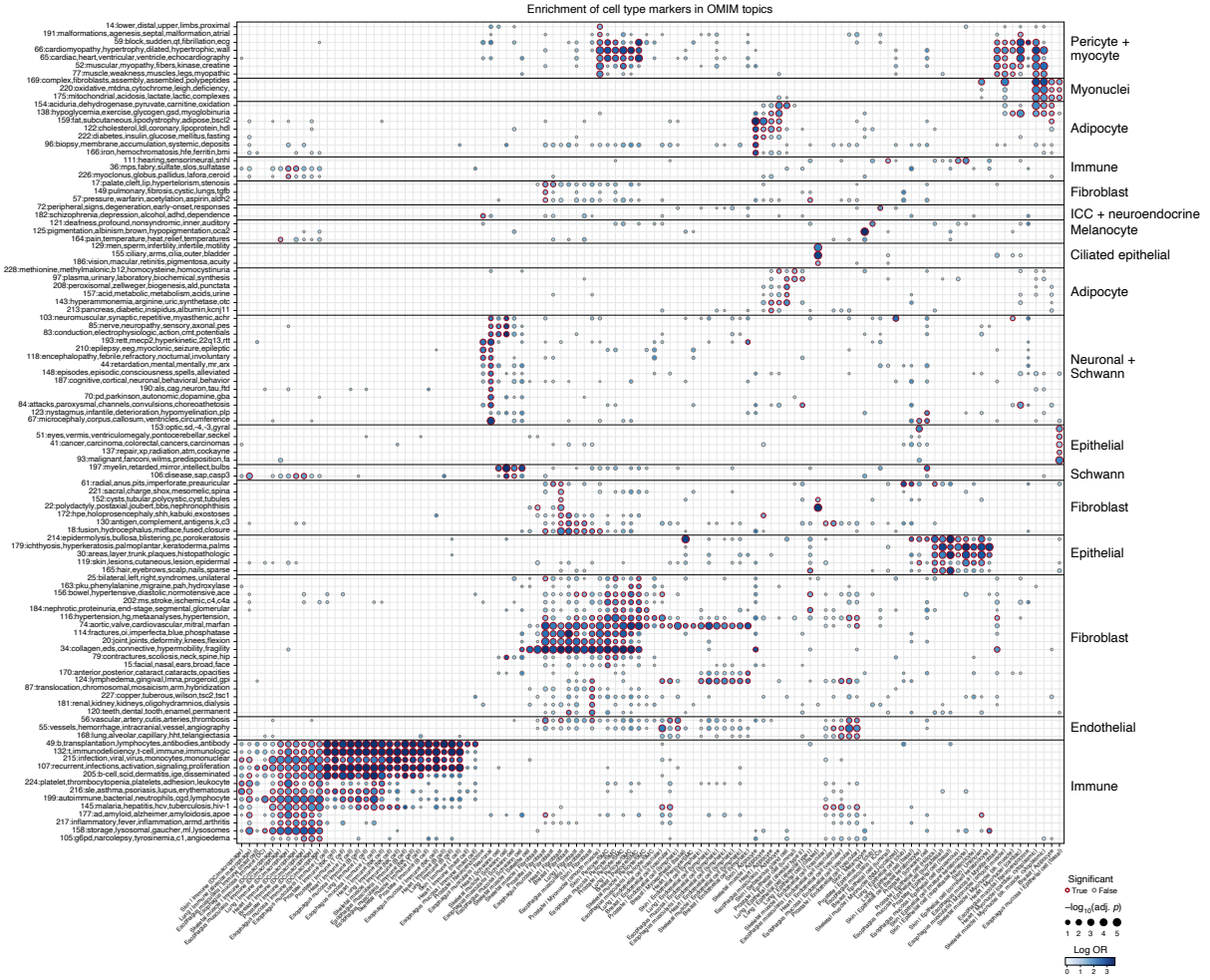


Fig. S21. Enrichment of cell type markers in OMIM topics. Enrichment of broad cell type markers with monogenic disease gene sets inferred by topic modeling (*28*). Effect size (log odds ratio, dot color) and significance ($-\log_{10}$(Benjamini–Hochberg FDR), dot size) of enrichment of genes from each monogenic disease topic (rows) for cell type markers of broad cell subsets (columns) in each tissue. Rows and columns are hierarchically clustered using the Pearson correlation metric and complete linkage method. Topic names are given as an integer identifier and comma-separated five words with highest loadings. Major topic groups inferred by hierarchical clustering are labeled with cell types associated with topics (rows, labels on right).

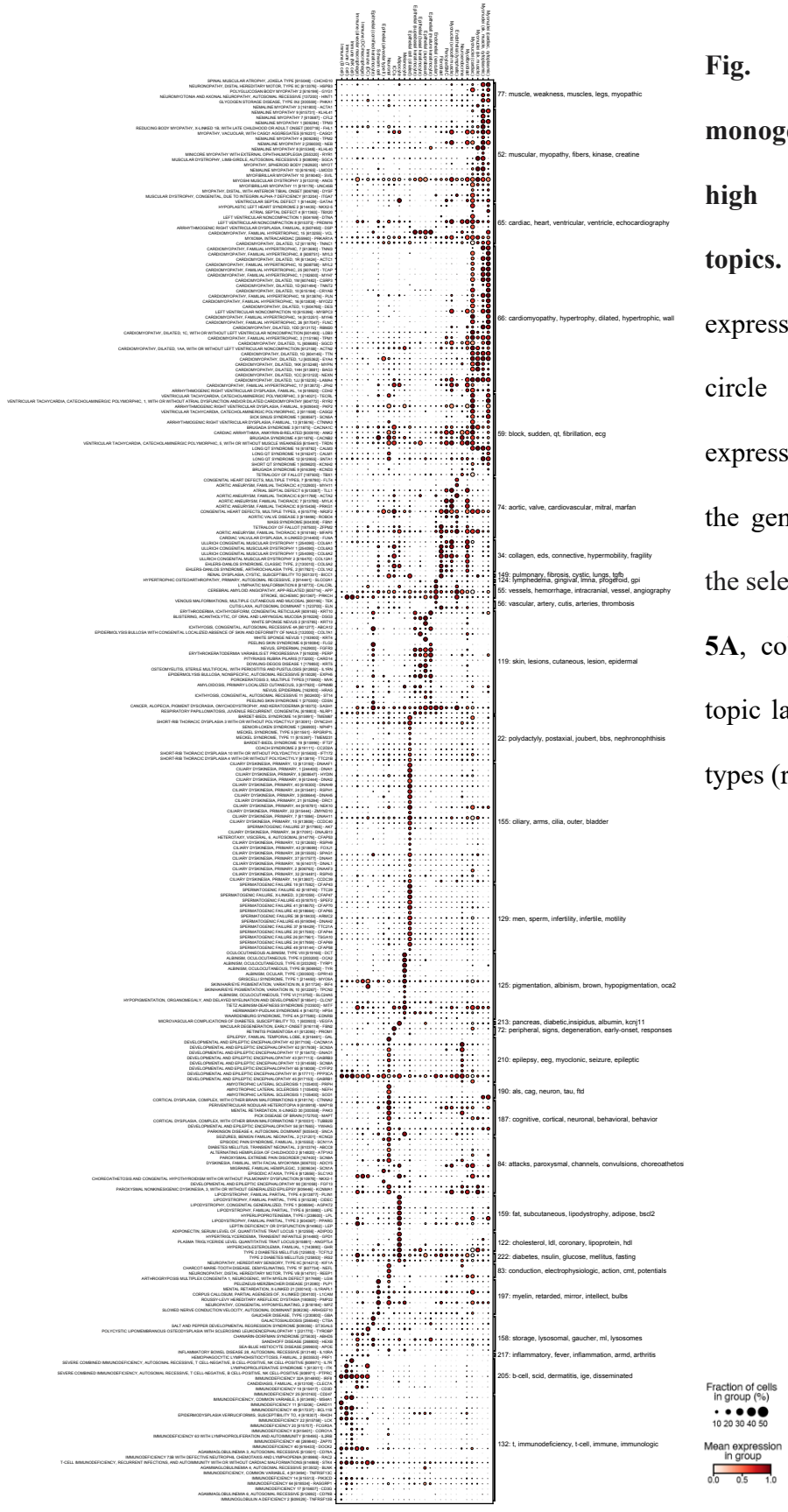
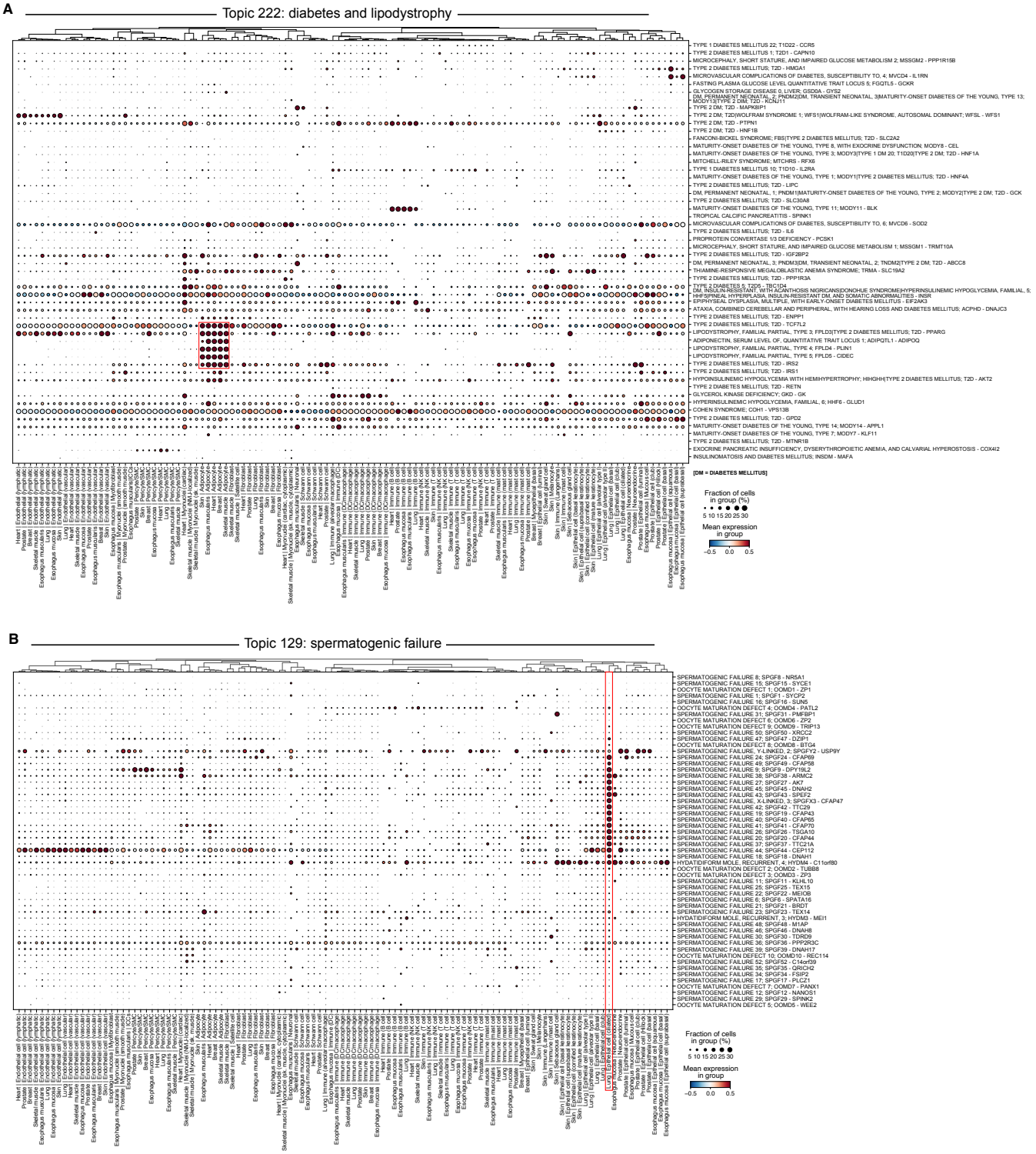


Fig. S22. Expression of monogenic disease genes with high loadings in selected topics. Min-max scaled mean expression (log(TP10K+1), circle color) and fraction of expressing cells (circle size) of the genes with high loadings in the selected topics (shown in **Fig. 5A**, columns, labels at bottom, topic labels at top) in broad cell types (rows) in all tissues.


A
Topic 222: diabetes and lipodystrophy
TYPE 1 DIABETES MELLITUS 22; T1D22 - CCR5
TYPE 2 DIABETES MELLITUS 1; T2D1 - CAPN10
MICROCEPHALY, SHORT STATURE, AND IMPAIRED GLUCOSE METABOLISM 2; MSSGM2 - PPP1R15B
TYPE 2 DIABETES MELLITUS; T2D - HMGA1
MICROVASCULAR COMPLICATIONS OF DIABETES, SUSCEPTIBILITY TO, 4; MVCD4 - IL1RN
FASTING PLASMA GLUCOSE LEVEL QUANTITATIVE TRAIT LOCUS 5; FGQTL5 - GCKR
GLYCOGEN STORAGE DISEASE 0, LIVER; GSD0A - GYS2
DM, PERMANENT NEONATAL, 2; PNDM2|DM, TRANSIENT NEONATAL, 3|MATURITY-ONSET DIABETES OF THE YOUNG, TYPE 13; MODY13|TYPE 2 DIM; T2D - KCNJ11
TYPE 2 DM; T2D - MAPKBP1
TYPE 2 DM; T2D|WOLFRAM SYNDROME 1; WFS1|WOLFRAM-LIKE SYNDROME, AUTOSOMAL DOMINANT; WFSL - WFS1
TYPE 2 DM; T2D - PTPN1
TYPE 2 DM; T2D - HNF1B
FANCONI-BICKEL SYNDROME; FBS|TYPE 2 DIABETES MELLITUS; T2D - SLC2A2
MATURITY-ONSET DIABETES OF THE YOUNG, TYPE 8, WITH EXOCRINE DYSFUNCTION; MODY8 - CEL
MATURITY-ONSET DIABETES OF THE YOUNG, TYPE 3; MODY3|TYPE 1 DM 20; T1D20|TYPE 2 DM; T2D - HNF1A
MITCHELL-RILEY SYNDROME; MTCHRS - RFX6
TYPE 1 DIABETES MELLITUS 10; T1D10 - IL2RA
MATURITY-ONSET DIABETES OF THE YOUNG, TYPE 1; MODY1|TYPE 2 DIABETES MELLITUS; T2D - HNF4A
TYPE 2 DIABETES MELLITUS; T2D - LIPC
DM, PERMANENT NEONATAL, 1; PNDM1|MATURITY-ONSET DIABETES OF THE YOUNG, TYPE 2; MODY2|TYPE 2 DM; T2D - GCK
TYPE 2 DIABETES MELLITUS; T2D - SLC30A8
MATURITY-ONSET DIABETES OF THE YOUNG, TYPE 11; MODY11 - BLK
TROPICAL CALCIFIC PANCREATITIS - SPINK1
MICROVASCULAR COMPLICATIONS OF DIABETES, SUSCEPTIBILITY TO, 6; MVCD6 - SOD2
TYPE 2 DIABETES MELLITUS; T2D - IL6
PROPROTEIN CONVERTASE 1/3 DEFICIENCY - PCSK1
MICROCEPHALY, SHORT STATURE, AND IMPAIRED GLUCOSE METABOLISM 1; MSSGM1 - TRMT10A
TYPE 2 DIABETES MELLITUS; T2D - IGF2BP2
DM, PERMANENT NEONATAL, 3; PNDM3|DM, TRANSIENT NEONATAL, 2; TNDM2|TYPE 2 DM; T2D - ABCC8
THIAMINE-RESPONSIVE MEGALOBLASTIC ANEMIA SYNDROME; TRMA - SLC19A2
TYPE 2 DIABETES MELLITUS; T2D - PPP1R3A
TYPE 2 DIABETES 6; T2D5 - TBC1D4
DM, INSULIN-RESISTANT, WITH ACANTHOSIS NIGRICANS|DONOHUE SYNDROME|HYPERINSULINEMIC HYPOGLYCEMIA, FAMILIAL, 5; HHF5
PINEAL HYPERPLASIA, INSULIN-RESISTANT DM, AND SOMATIC ABNORMALITIES - INSR
EPIPHYSEAL DYSPLASIA, MULTIPLE, WITH EARLY-ONSET DIABETES MELLITUS - EIF2AK3
ATAXIA, COMBINED CEREBELLAR AND PERIPHERAL, WITH HEARING LOSS AND DIABETES MELLITUS; ACPHD - DNAJC3
TYPE 2 DIABETES MELLITUS; T2D - ENPP1
TYPE 2 DIABETES MELLITUS; T2D - TCF7L2
LIPODYSTROPHY, FAMILIAL PARTIAL, TYPE 3; FPLD3|TYPE 2 DIABETES MELLITUS; T2D - PPARG
ADIPONECTIN, SERUM LEVEL OF, QUANTITATIVE TRAIT LOCUS 1; ADIPQTL1 - ADIPOQ
LIPODYSTROPHY, FAMILIAL PARTIAL, TYPE 4; FPLD4 - PLIN1
LIPODYSTROPHY, FAMILIAL PARTIAL, TYPE 5; FPLD5 - CIDEC
TYPE 2 DIABETES MELLITUS; T2D - IRS2
TYPE 2 DIABETES MELLITUS; T2D - IRS1
HYPOINSULINEMIC HYPOGLYCEMIA WITH HEMIHYPERTROPHY; HIHGHH|TYPE 2 DIABETES MELLITUS; T2D - AKT2
TYPE 2 DIABETES MELLITUS; T2D - RETN
GLYCEROL KINASE DEFICIENCY; GKD - GK
HYPERINSULINEMIC HYPOGLYCEMIA, FAMILIAL, 6; HHF6 - GLUD1
COHEN SYNDROME; COH1 - VPS13B
TYPE 2 DIABETES MELLITUS; T2D - GPD2
MATURITY-ONSET DIABETES OF THE YOUNG, TYPE 14; MODY14 - APPL1
MATURITY-ONSET DIABETES OF THE YOUNG, TYPE 7; MODY7 - KLF11
TYPE 2 DIABETES MELLITUS; T2D - MTNR1B
EXOCRINE PANCREATIC INSUFFICIENCY, DYSERYTHROPOIETIC ANEMIA, AND CALVARIAL HYPEROSTOSIS - COX4I2
INSULINOMATOSIS AND DIABETES MELLITUS; INSDM - MAFA
[DM = DIABETES MELLITUS]
Fraction of cells in group (%)
5 10 15 20 25 30
Mean expression in group
−0.5 0.0 0.5
B
Topic 129: spermatogenic failure
SPERMATOGENIC FAILURE 8; SPGF8 - NR5A1
SPERMATOGENIC FAILURE 15; SPGF15 - SYCE1
OOCYTE MATURATION DEFECT 1; OOMD1 - ZP1
SPERMATOGENIC FAILURE 1; SPGF1 - SYCP2
SPERMATOGENIC FAILURE 16; SPGF16 - SUN5
OOCYTE MATURATION DEFECT 4; OOMD4 - PATL2
SPERMATOGENIC FAILURE 31; SPGF31 - PMFBP1
OOCYTE MATURATION DEFECT 6; OOMD6 - ZP2
OOCYTE MATURATION DEFECT 9; OOMD9 - TRIP13
SPERMATOGENIC FAILURE 50; SPGF50 - KRCC2
SPERMATOGENIC FAILURE 47; SPGF47 - DZIP1
OOCYTE MATURATION DEFECT 8; OOMD8 - BTG4
SPERMATOGENIC FAILURE, Y-LINKED, 2; SPGFY2 - USP9Y
SPERMATOGENIC FAILURE 24; SPGF24 - CFAP69
SPERMATOGENIC FAILURE 49; SPGF49 - CFAP58
SPERMATOGENIC FAILURE 9; SPGF9 - DPY19L2
SPERMATOGENIC FAILURE 38; SPGF38 - ARMC2
SPERMATOGENIC FAILURE 27; SPGF27 - AK7
SPERMATOGENIC FAILURE 45; SPGF45 - DNAH2
SPERMATOGENIC FAILURE 43; SPGF43 - SPEF2
SPERMATOGENIC FAILURE, X-LINKED, 3; SPGFX3 - CFAP47
SPERMATOGENIC FAILURE 42; SPGF42 - TTC29
SPERMATOGENIC FAILURE 19; SPGF19 - CFAP43
SPERMATOGENIC FAILURE 40; SPGF40 - CFAP65
SPERMATOGENIC FAILURE 41; SPGF41 - CFAP70
SPERMATOGENIC FAILURE 26; SPGF26 - TSGA10
SPERMATOGENIC FAILURE 20; SPGF20 - CFAP44
SPERMATOGENIC FAILURE 37; SPGF37 - TTC21A
SPERMATOGENIC FAILURE 44; SPGF44 - CEP112
SPERMATOGENIC FAILURE 18; SPGF18 - DNAH1
HYDATIDIFORM MOLE, RECURRENT, 4; HYDM4 - C11orf80
OOCYTE MATURATION DEFECT 2; OOMD2 - TUBB8
OOCYTE MATURATION DEFECT 3; OOMD3 - ZP3
SPERMATOGENIC FAILURE 11; SPGF11 - KLHL10
SPERMATOGENIC FAILURE 25; SPGF25 - TEX15
SPERMATOGENIC FAILURE 22; SPGF22 - MEIOB
SPERMATOGENIC FAILURE 6; SPGF6 - SPATA16
SPERMATOGENIC FAILURE 21; SPGF21 - BRDT
SPERMATOGENIC FAILURE 23; SPGF23 - TEX14
HYDATIDIFORM MOLE, RECURRENT, 3; HYDM3 - MEI1
SPERMATOGENIC FAILURE 48; SPGF48 - MTAP
SPERMATOGENIC FAILURE 46; SPGF46 - DNAH8
SPERMATOGENIC FAILURE 30; SPGF30 - TDRD9
SPERMATOGENIC FAILURE 36; SPGF36 - PPP2R3C
SPERMATOGENIC FAILURE 39; SPGF39 - DNAH17
OOCYTE MATURATION DEFECT 10; OOMD10 - REC114
SPERMATOGENIC FAILURE 52; SPGF52 - C14orf39
SPERMATOGENIC FAILURE 35; SPGF35 - ORC1L2
SPERMATOGENIC FAILURE 34; SPGF34 - FSIP2
SPERMATOGENIC FAILURE 17; SPGF17 - PLCZ1
OOCYTE MATURATION DEFECT 7; OOMD7 - PANX1
SPERMATOGENIC FAILURE 12; SPGF12 - NANOS1
SPERMATOGENIC FAILURE 29; SPGF29 - SPINK2
OOCYTE MATURATION DEFECT 5; OOMD5 - WEE2
Fraction of cells in group (%)
5 10 15 20 25 30
Mean expression in group
−0.5 0.0 0.5

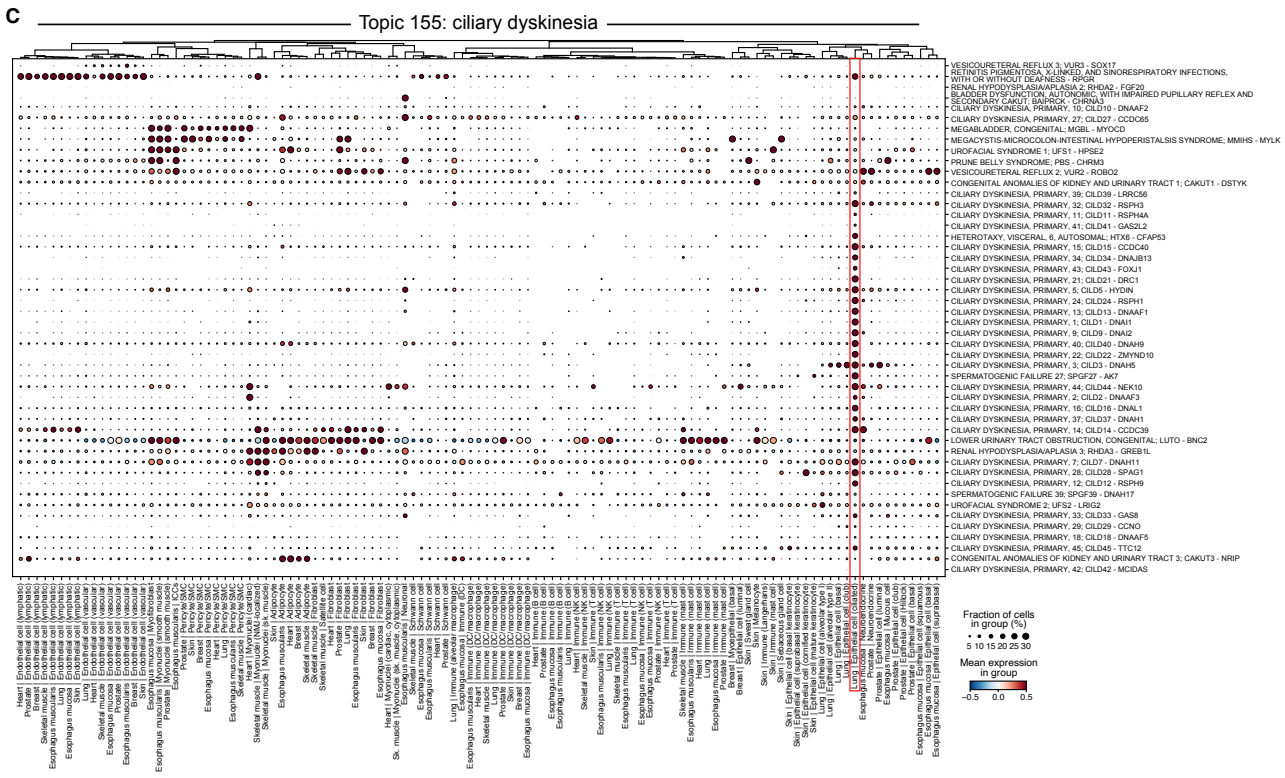


Fig. S23. Expression of genes with high loadings in three selected topics. Scaled mean expression (*z*-score, circle color) and fraction of expressing cells (circle size) of the genes with high loadings (rows, labels at right) in topics 222 **(A)**, 129 **(B)**, and 155 **(C)** in broad cell types in each tissue (columns, labels at bottom). Diseases associated with the genes are given in row labels before the gene names. Red rectangles: key associations highlighted in the text.

Fig. S24. Expression of genes that are causal for muscular dystrophies in three muscle types. Scaled mean expression (dot color, min-max scale) and fraction of expressing cells (dot size) for muscular dystrophy genes (rows, gene labels on left, disease names on right) in each broad cell type in esophagus muscularis, heart and skeletal muscle (columns, labels at bottom). Major gene and cell type groups identified by hierarchical clustering (Pearson correlation metric with Ward and average linkages for cell types and genes, respectively) are split with black bars.

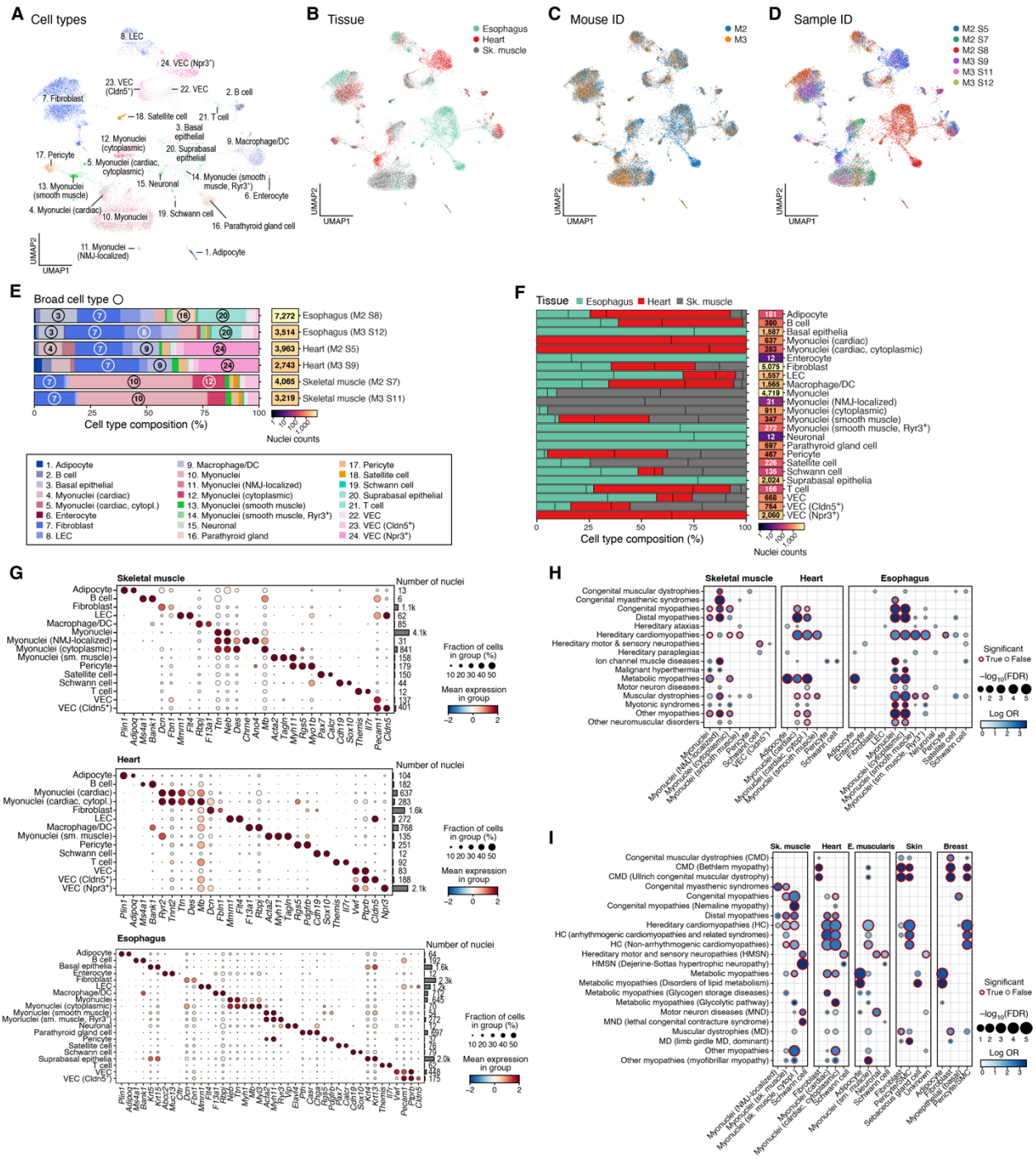


Fig. S25. Analysis of monogenic muscle disease genes in mouse muscle tissues. (**A–D**) Mouse muscle tissue atlas. UMAP visualization of snRNA-seq profiles from skeletal muscle, heart, and esophagus in mouse, colored by cell type **(A)**, tissue **(B)**, mouse (**C**) or sample IDs **(D)**. (**E, F**) Mouse muscle atlas cell type composition. Proportion of nuclei (*x*-axis) of each type (color) in

each tissue (sample) (**E**), and of each tissue in each cell type (**F,** after normalizing for the total number of nuclei profiled in each tissue, (*28*)) along with the number of profiled nuclei (right). Circled numbers refer to cell types in the color legend. Black vertical lines in bars in F: relative proportion from each mouse. **(G)** Cell type markers. Scaled mean expression (dot color, *z*-score) and fraction of expressing cells (dot size) for marker genes (columns) in each cell subset (rows). (**H, I**) Enrichment of broad cell type markers with monogenic muscle disease genes. Effect size (log odds ratio, dot color) and significance ($-\log_{10}$(Benjamini–Hochberg FDR), dot size) of enrichment of mouse orthologs (*28*) of genes from each monogenic muscle disease group (rows) for (**H**) cell type markers of broad cell subsets (columns) in each mouse tissue (**I**) or from human muscle tissues (as in **Fig. 5B**) as well as skin and breast. Red border: Benjamini–Hochberg FDR < 0.1.

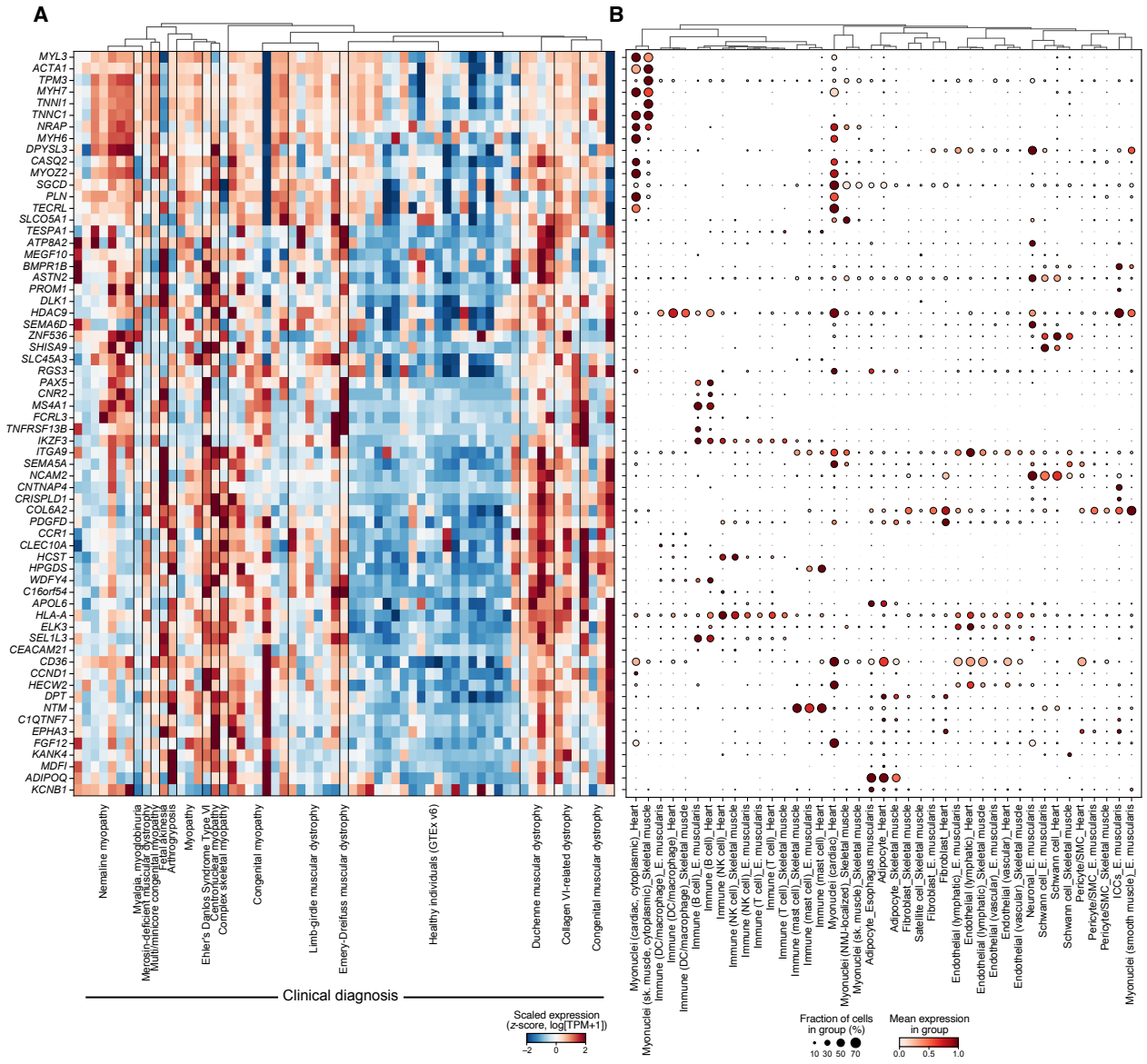


Fig. S26. Single nucleus expression of genes differentially expressed between muscle samples from healthy individuals and muscle disease patients. Scaled mean expression of genes (rows) differentially expressed between muscle samples from healthy individuals (GTEx v6, (*140*)) and muscle disease patients (*122*) in bulk RNA-seq (columns: individuals grouped and colored by clinical diagnosis, (*28*)) **(A)** and snRNA-seq (columns: broad cell types) in esophagus muscularis, heart and skeletal muscle **(B)**. Expression is *z*-score normalized (log(TPM+1)) in bulk RNA-seq,

min-max scaled (log(TP10K+1)) in snRNA-seq. Dot size and color represent mean snRNA-seq expression and the fraction of expressing nuclei in each group in (B).

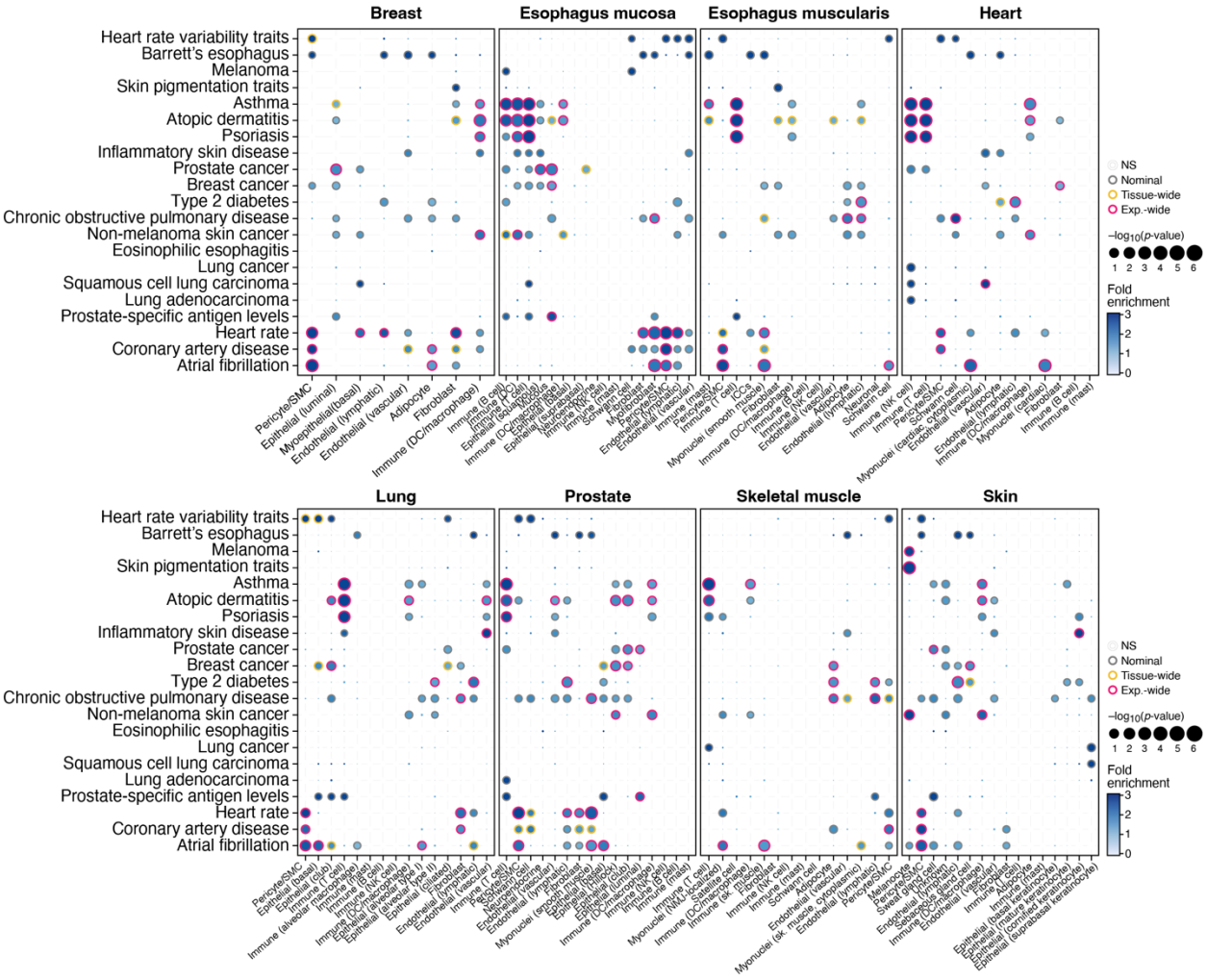


Fig. S27. Cell type enrichment of GWAS loci from 21 complex traits in 8 tissues using broad cell type annotations. Significance (circle size, $-\log_{10}(p\text{-value})$) and effect size (circle color, fold-enrichment) of enrichment of GWAS locus sets of complex traits (rows) in each broad cell type category (columns) in the eight tissues in the cross-tissue atlas (panels). Grey, orange, red borders: nominal, tissue-wide (Benjamini–Hochberg (BH) FDR < 0.05 correcting for all cell types tested per tissue and per trait) and experiment-wide (BH FDR < 0.05 correcting for all cell types tested across 8 tissues and 21 traits) significance results, respectively.

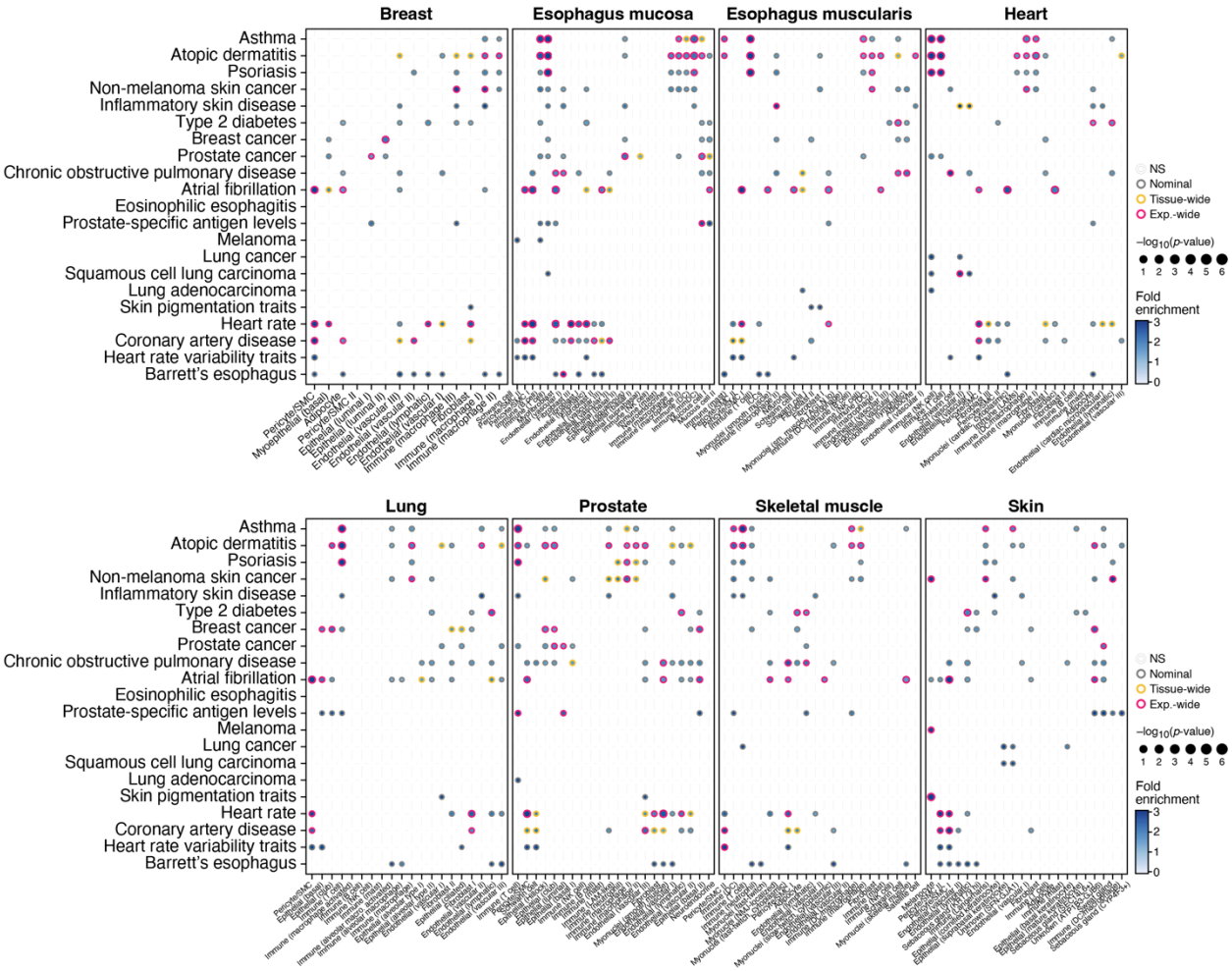


Fig. S28. Cell type enrichment of GWAS loci from 21 complex traits in 8 tissues using granular cell type annotations. Significance (circle size, –$\log_{10}$(*p*-value)) and effect size (circle color, fold-enrichment) of enrichment of GWAS locus set (rows) in each granular cell type category (columns) in the eight tissues in the cross-tissue atlas (panels). Grey, orange, red borders: nominal, tissue-wide (Benjamini–Hochberg (BH) FDR < 0.05 correcting for all cell types tested per tissue and trait) and experiment-wide (BH FDR < 0.05 correcting for all cell types tested across 8 tissues and 21 traits) significance results.

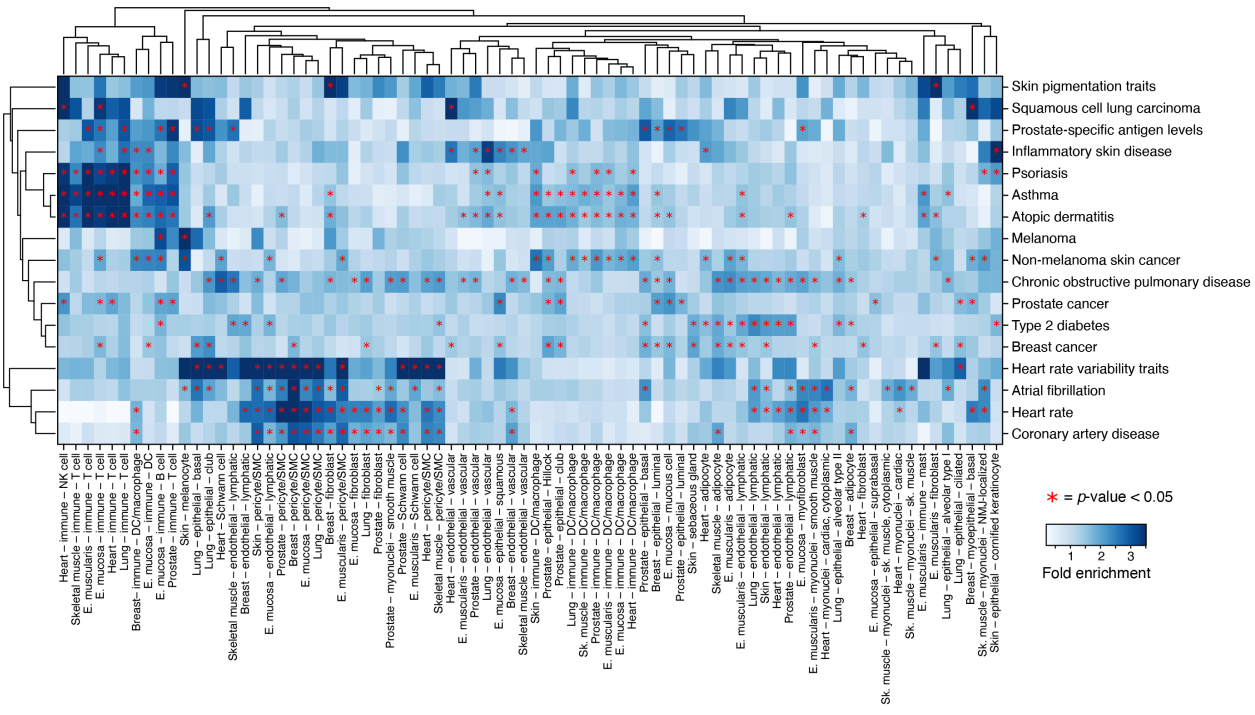


Fig. S29. Cell type specificity enrichment of 21 complex traits across broad cell types from 8 tissues. Fold-enrichment (color bar) of each GWAS locus set (rows) for specificity in each broad cell type category (columns). Red stars: nominal significance ($P < 0.05$). Only traits with at least one enrichment at tissue-wide significance (BH FDR < 0.05 correcting for number of cell types tested per tissue per trait) are shown. Rows and columns are hierarchically clustered using the Euclidean distance and average linkage method. Red stars denote nominal significance.

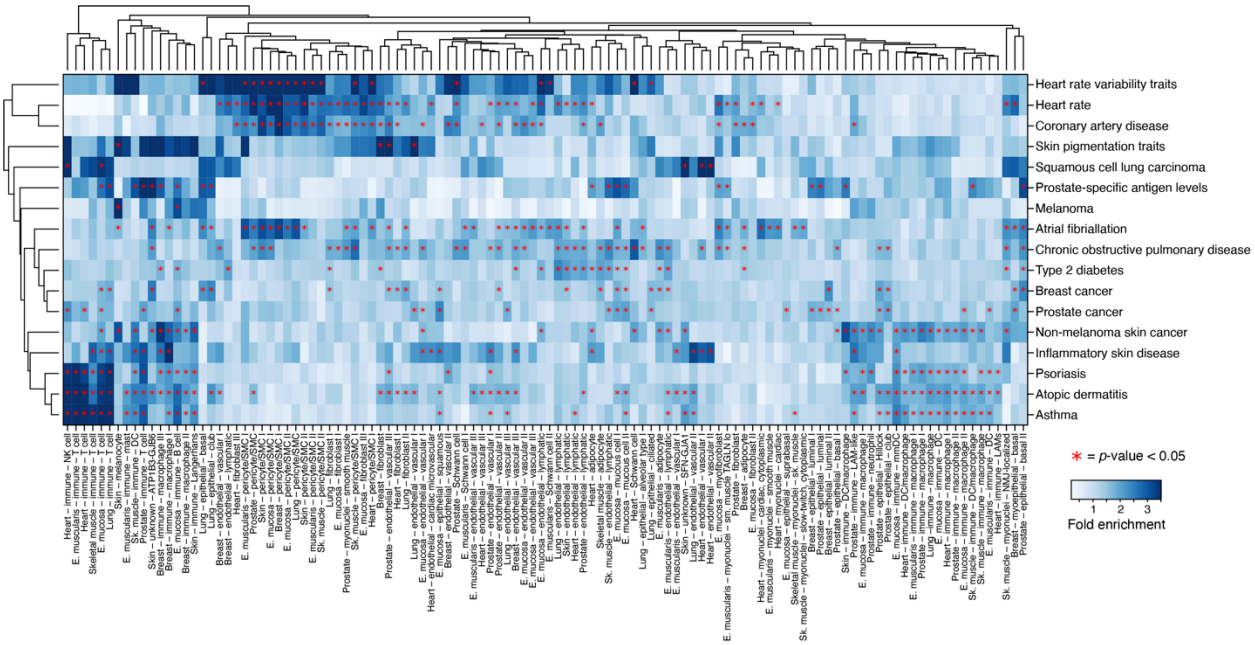


Fig. S30. Cell type specificity enrichment of 21 complex traits across granular cell types from 8 tissues. Fold-enrichment (color bar) of each GWAS locus set (rows) for specificity in each granular cell type category (columns). Red stars: nominal significance ($P < 0.05$). Only traits with at least one enrichment at tissue-wide significance (BH FDR < 0.05 correcting for number of cell types tested per tissue per trait) are shown. Rows and columns are hierarchically clustered using the Euclidean distance and average linkage method. Red stars denote nominal significance.

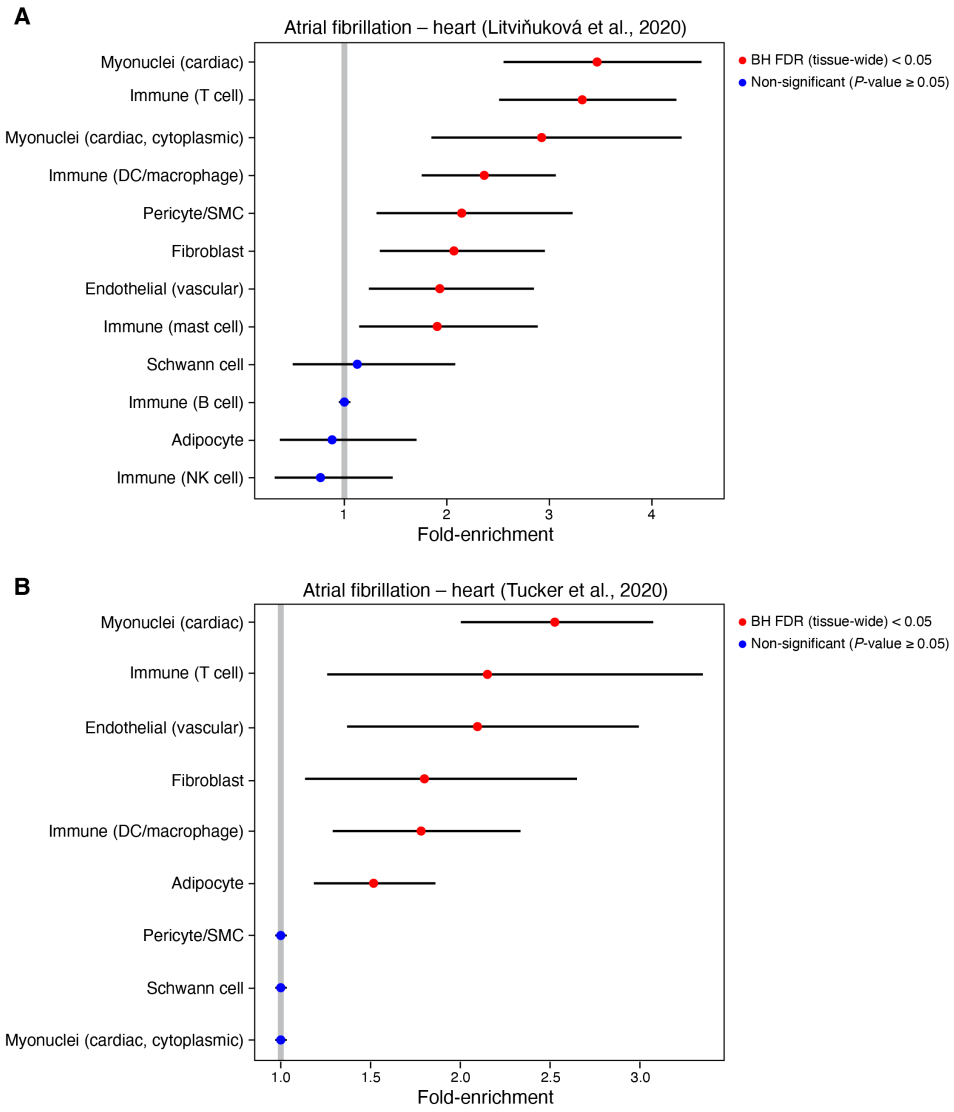


Fig. S31. Cell type specific enrichment of e/sQTL genes mapped to Atrial Fibrillation GWAS loci in two non-GTEx snRNA-seq heart studies. Fold-enrichment (*x*-axis) of cell type specific expression (*y*-axis) of genes mapped to atrial fibrillation GWAS loci in Litviňuková et al. (*34*) (**A**) and Tucker et al. (*33*) (**B**) heart snRNA-seq. Cell clusters were defined using the cell type markers defined in this study. Error bars: 95% credible intervals. Red: tissue-wide significant (Benjamini Hochberg (BH) FDR < 0.05); Orange: nominal significant ($P < 0.05$); Blue: non-significant ($P \geq 0.05$). Cell type-specific gene expression was defined as $\log_2$(fold-change) > 0.5, false discovery rate (FDR) < 0.1, and expression in at least 5% of cells in each given cell type.

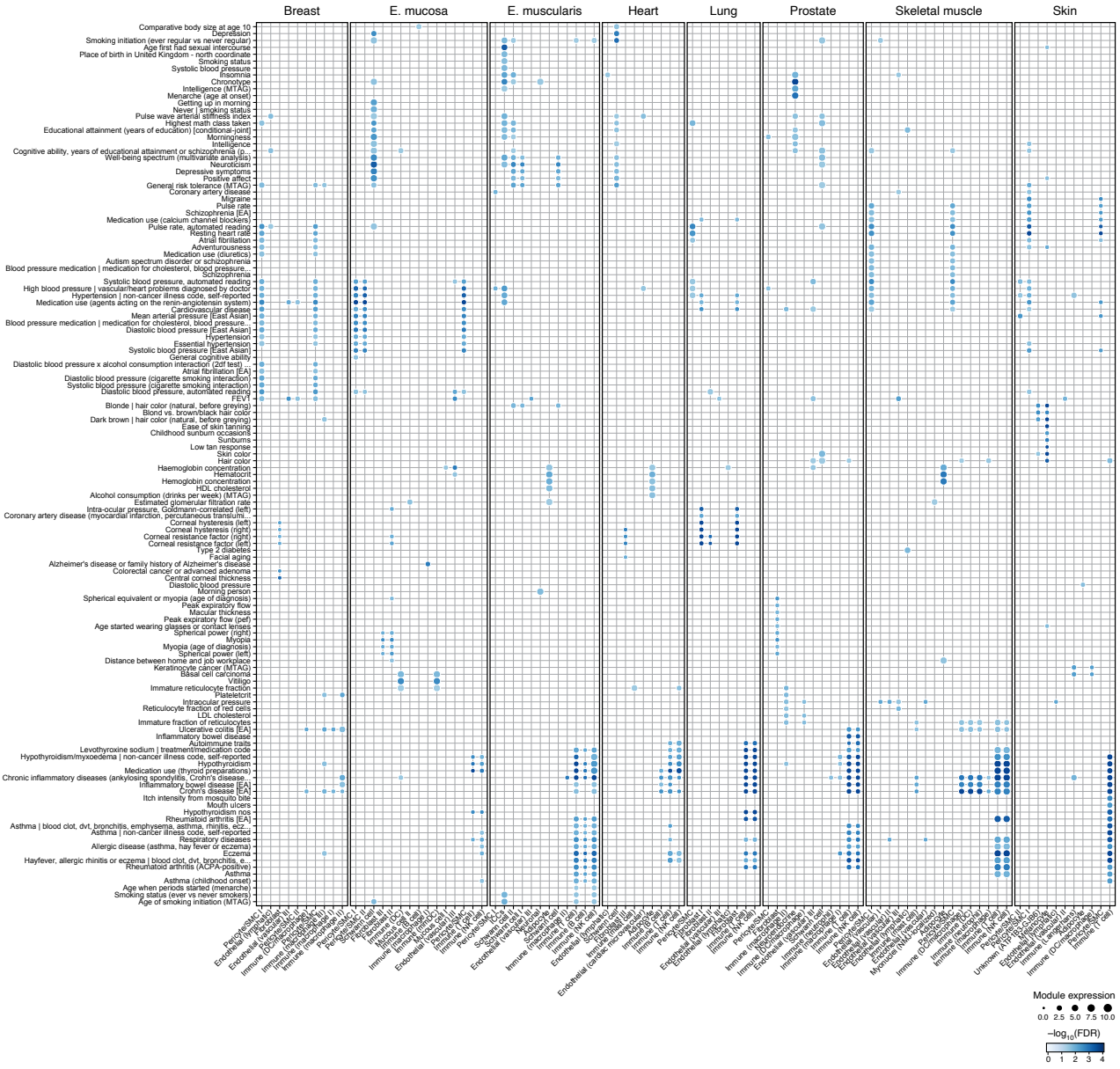


Fig. S32. Module-based GWAS gene set enrichment analysis highlights associations between cell types and traits/diseases. Significance ($-\log_{10}$(Benjamini–Hochberg FDR), dot size) of enrichment of GWAS module (rows) expression (dot color) in each cell type (columns).

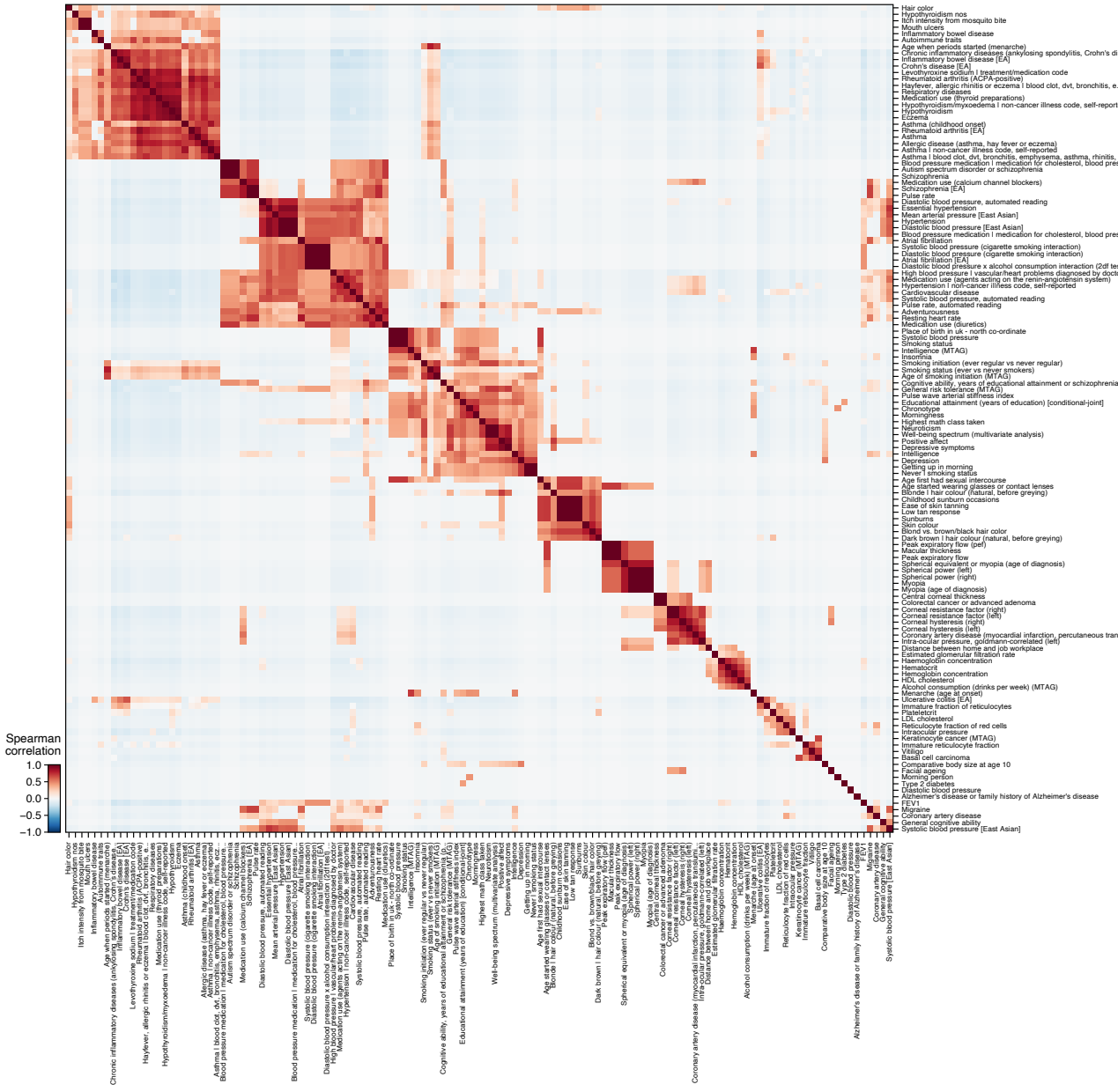


Fig. S33. Disease/trait grouping by GWAS-cell type enrichments. Similarity (Spearman correlation coefficient, color bar) between each pair of GWAS traits/diseases (rows, columns) based on their enriched cell type profiles (as in **Fig. S32**).

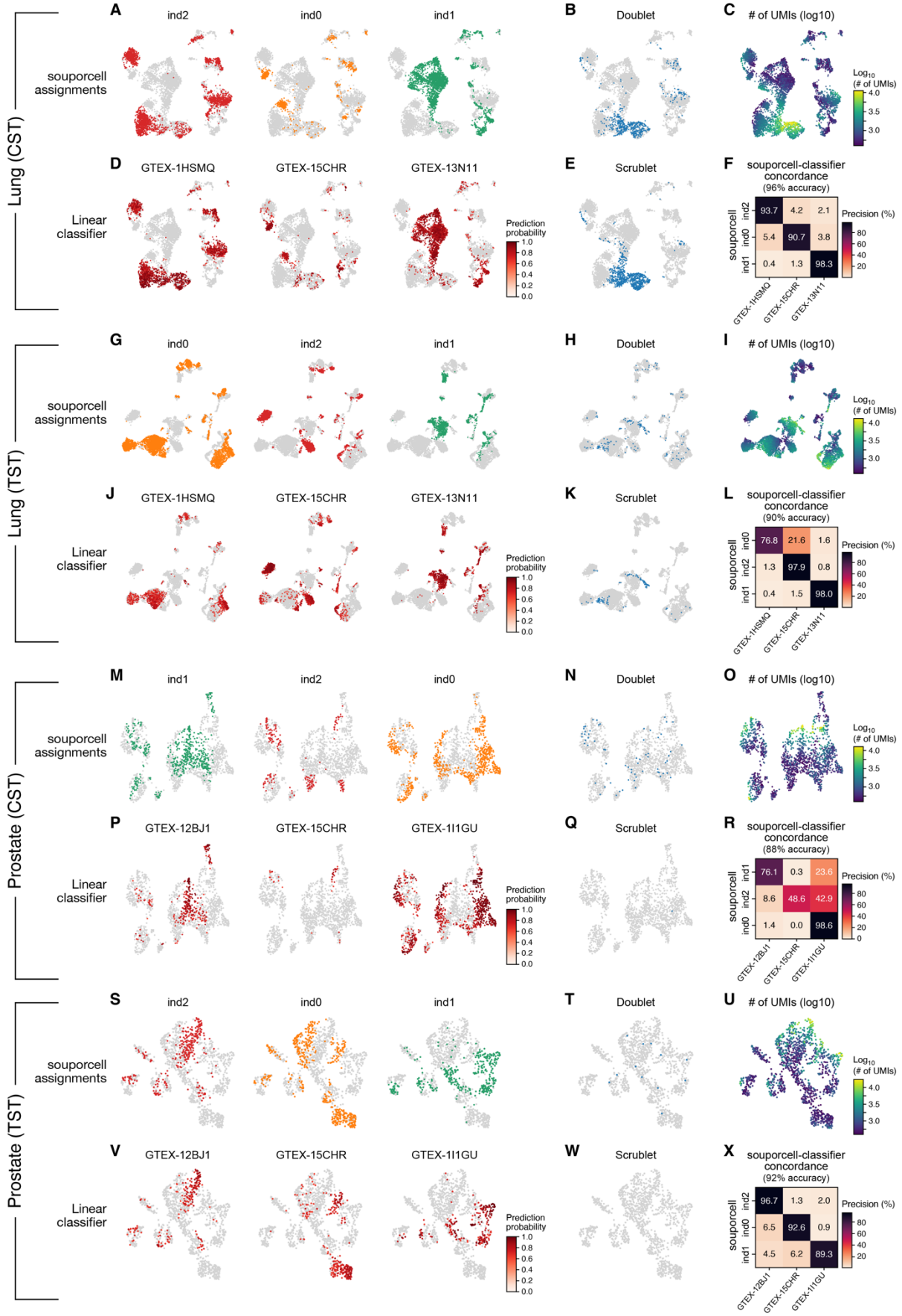

Lung (CST)
Lung (TST)
Prostate (CST)
Prostate (TST)
souporcell assignments
Linear classifier
A
ind2
ind0
ind1
B
Doublet
C
# of UMIs (log10)
Log10 (# of UMIs)
4.0
3.5
3.0
D
GTEX-1HSMQ
GTEX-15CHR
GTEX-13N11
Prediction probability
1.0
0.8
0.6
0.4
0.2
0.0
E
Scrublet
F
souporcell-classifier concordance
(96% accuracy)
souporcell
ind2
ind0
ind1
93.7
4.2
2.1
5.4
90.7
3.8
0.4
1.3
98.3
GTEX-1HSMQ
GTEX-15CHR
GTEX-13N11
Precision (%)
80
60
40
20
G
ind0
ind2
ind1
H
Doublet
I
# of UMIs (log10)
J
GTEX-1HSMQ
GTEX-15CHR
GTEX-13N11
K
Scrublet
L
souporcell-classifier concordance
(90% accuracy)
76.8
21.6
1.6
1.3
97.9
0.8
0.4
1.5
98.0
M
ind1
ind2
ind0
N
Doublet
O
# of UMIs (log10)
P
GTEX-12BJ1
GTEX-15CHR
GTEX-1I1GU
Q
Scrublet
R
souporcell-classifier concordance
(88% accuracy)
76.1
0.3
23.6
8.6
48.6
42.9
1.4
0.0
98.6
0
S
ind2
ind0
ind1
T
Doublet
U
# of UMIs (log10)
V
GTEX-12BJ1
GTEX-15CHR
GTEX-1I1GU
W
Scrublet
X
souporcell-classifier concordance
(92% accuracy)
96.7
1.3
2.0
6.5
92.6
0.9
4.5
6.2
89.3

Fig. S34. Successful processing of multiplexed samples by snRNA-seq. Analysis of snRNA-seq profiles from frozen tissue samples from 3 individuals processed in multiplex for lung (**A–L**) and prostate (**M–X**) by either the CST (**A–F, M–R**) or the TST (**G–L, S–X**) protocols. UMAP visualizations of nuclei profiles colored by the demultiplexing assignments of each nucleus by souporcell to individuals **(A, G, M, S)**, doublets **(B, H, N, T)**, $\log_{10}$-transformed total number of UMIs (**C, I, O, U**), demultiplexing prediction probabilities (unassigned dots are shown grey) of an expression-based linear classifier **(D, J, P, V)**, and doublets predicted by Scrublet **(E, K, Q, W)**. Heat maps (**F, L, R, X**) show the concordance (% precision, color bar) between souporcell (rows) and the linear classifier (columns) calls. Accuracy values are labeled on top.

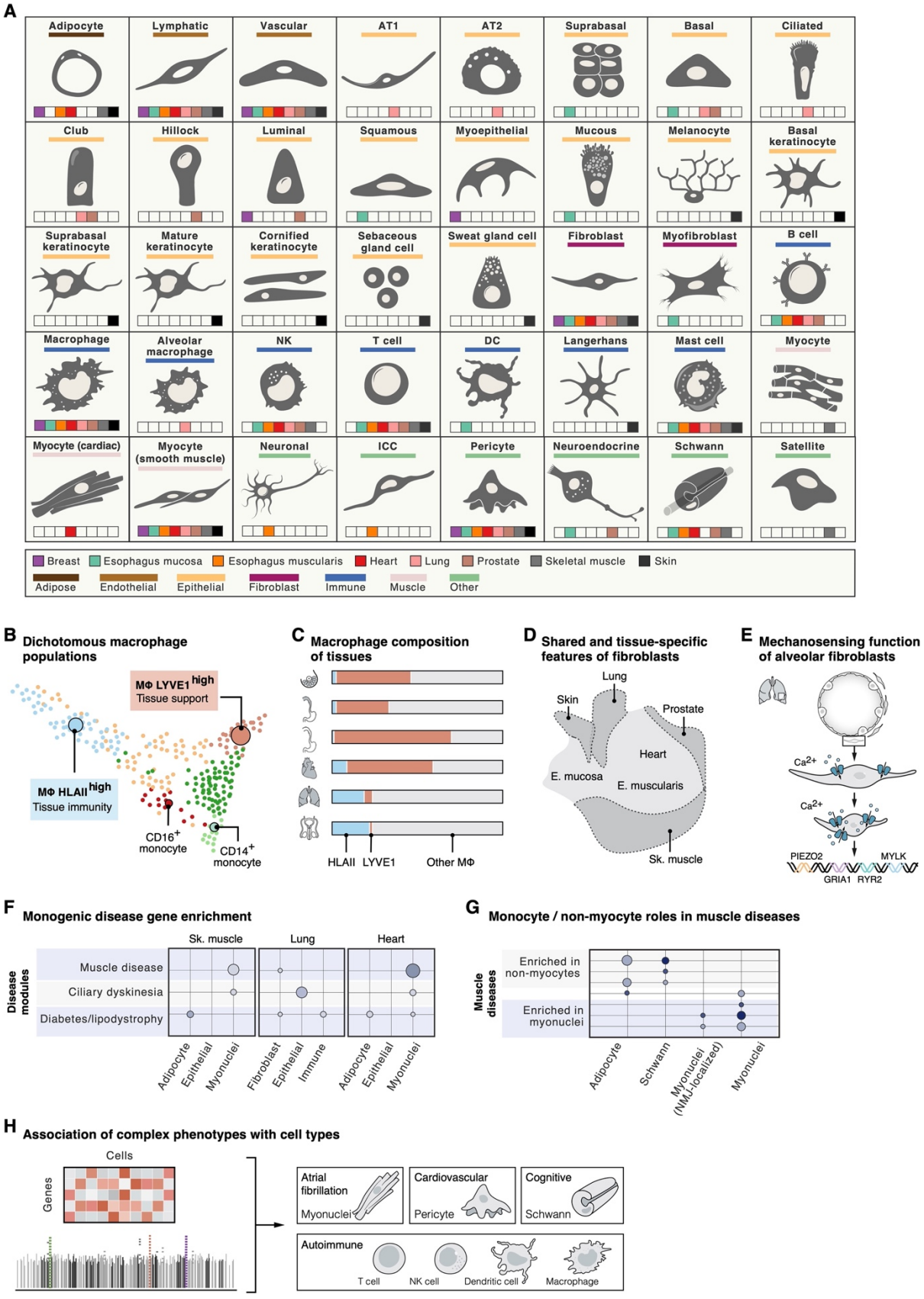

A
Adipocyte
Lymphatic
Vascular
AT1
AT2
Suprabasal
Basal
Ciliated
Club
Hillock
Luminal
Squamous
Myoepithelial
Mucous
Melanocyte
Basal keratinocyte
Suprabasal keratinocyte
Mature keratinocyte
Cornified keratinocyte
Sebaceous gland cell
Sweat gland cell
Fibroblast
Myofibroblast
B cell
Macrophage
Alveolar macrophage
NK
T cell
DC
Langerhans
Mast cell
Myocyte
Myocyte (cardiac)
Myocyte (smooth muscle)
Neuronal
ICC
Pericyte
Neuroendocrine
Schwann
Satellite
Breast
Esophagus mucosa
Esophagus muscularis
Heart
Lung
Prostate
Skeletal muscle
Skin
Adipose
Endothelial
Epithelial
Fibroblast
Immune
Muscle
Other
B
Dichotomous macrophage populations
MΦ LYVE1 high
Tissue support
MΦ HLAII high
Tissue immunity
CD16+ monocyte
CD14+ monocyte
C
Macrophage composition of tissues
HLAII
LYVE1
Other MΦ
D
Shared and tissue-specific features of fibroblasts
Lung
Skin
Prostate
Heart
E. mucosa
E. muscularis
Sk. muscle
E
Mechanosensing function of alveolar fibroblasts
Ca2+
Ca2+
PIEZO2
MYLK
GRIA1
RYR2
F
Monogenic disease gene enrichment
Sk. muscle
Lung
Heart
Disease modules
Muscle disease
Ciliary dyskinesia
Diabetes/lipodystrophy
Adipocyte
Epithelial
Myonuclei
Fibroblast
Epithelial
Immune
Adipocyte
Epithelial
Myonuclei
G
Monocyte / non-myocyte roles in muscle diseases
Muscle diseases
Enriched in non-myocytes
Enriched in myonuclei
Adipocyte
Schwann
Myonuclei (NMJ-localized)
Myonuclei
H
Association of complex phenotypes with cell types
Cells
Genes
Atrial fibrillation
Myonuclei
Cardiovascular
Pericyte
Cognitive
Schwann
Autoimmune
T cell
NK cell
Dendritic cell
Macrophage

Fig. S35. Schematic summarizing the key findings of this study. (A) Broad cell types in the cross-tissue atlas. Illustration of 40 (of 43) broad cell types in the cross-tissue atlas labeled by cellular compartments (horizontal color line) and tissues in which the cell type is detected (colored boxes). Cytoplasmic and NMJ-localized myonuclei subsets are not shown. **(B)** Dichotomy between LYVE1 and HLAII-expressing macrophages and the differentiation trajectory of these populations. **(C)** Compositional differences of macrophages across tissues. **(D)** Shared and tissue-specific features of fibroblasts. **(E)** Mechanosensing function of alveolar fibroblasts. **(F)** Intra- and cross-tissue cell type association with monogenic disorders. **(G)** Potential roles of myonuclei subsets and non-myocytes in monogenic muscle diseases. **(H)** Associations of complex phenotypes with gene programs and QTL genes mapped to GWAS loci.

Table S1. (separate file)

Study overview, sample and individual metadata, nuclei counts and extraction solution descriptions

Table S2. (separate file)

Differential expression of broad and granular cell types for each tissue

Table S3. (separate file)

Literature-curated cell type markers and references

Table S4. (separate file)

Cell type harmonization and gene correlations between this study and other snRNA-seq and scRNA-seq studies

Table S5. (separate file)

Differential expression of granular myeloid states

Table S6. (separate file)

Studies used for the investigation of LAM presence and diversity

Table S7. (separate file)

GWAS associated to HLAII-high, LYVE1-high and LAM-like macrophage markers

Table S8. (separate file)

Differential expression of fibroblasts, T cells and adipocytes across tissues

Table S9. (separate file)

Gene set enrichment results, gene and word loadings of OMIM topic model

Table S10. (separate file)

Monogenic muscle diseases and associated genes obtained from the 2021 version of Benarroch et al.

Table S11. (separate file)

Cell type marker monogenic muscle disease gene set enrichment analysis results

Table S12. (separate file)

Monogenic muscle disease genes differentially expressed in broad and granular cell types with disease information

Table S13 (separate file)

Putative cell-cell interactions across broad cell types in each tissue

Table S14. (separate file)

List of GWAS and number of loci tested for cell type enrichment analysis.

Table S15. (separate file)

GWAS cell type specificity enrichment analysis (broad annotations)

Table S16. (separate file)

GWAS cell type specificity enrichment analysis (granular annotations)

Table S17. (separate file)

Application of ECLIPSER to heart-related traits and two separate heart snRNA-seq studies

Table S18. (separate file)

Gene set enrichment analysis of tissue-shared, myonuclei-specific genes driving enrichment signal of atrial fibrillation GWAS loci in heart myonuclei

Table S19. (separate file)

Gene set enrichment analysis of heart- and myonuclei-specific genes driving enrichment signal of atrial fibrillation GWAS loci in heart myonuclei

Table S20. (separate file)

Module-based GWAS enrichment analysis

References and Notes

1. V. Tam, N. Patel, M. Turcotte, Y. Bossé, G. Paré, D. Meyre, Benefits and limitations of genome-wide association studies. *Nat. Rev. Genet.* **20**, 467–484 (2019). doi:10.1038/s41576-019-0127-1 Medline

2. M. C. Mills, C. Rahal, A scientometric review of genome-wide association studies. *Commun. Biol.* **2**, 9 (2019). doi:10.1038/s42003-018-0261-x Medline

3. E. Cano-Gamez, G. Trynka, From GWAS to function: Using functional genomics to identify the mechanisms underlying complex diseases. *Front. Genet.* **11**, 424 (2020). doi:10.3389/fgene.2020.00424 Medline

4. J. G. Camp, R. Platt, B. Treutlein, Mapping human cell phenotypes to genotypes with single-cell genomics. *Science* **365**, 1401–1405 (2019). doi:10.1126/science.aax6648 Medline

5. G. Sun, Z. Li, D. Rong, H. Zhang, X. Shi, W. Yang, W. Zheng, G. Sun, F. Wu, H. Cao, W. Tang, Y. Sun, Single-cell RNA sequencing in cancer: Applications, advances, and emerging challenges. *Mol. Ther. Oncolytics* **21**, 183–206 (2021). doi:10.1016/j.omto.2021.04.001 Medline

6. S. S. Potter, Single-cell RNA sequencing for the study of development, physiology and disease. *Nat. Rev. Nephrol.* **14**, 479–492 (2018). doi:10.1038/s41581-018-0021-7 Medline

7. GTEx Consortium, The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318–1330 (2020). doi:10.1126/science.aaz1776 Medline

8. S. Kasela, K. Kisand, L. Tserel, E. Kaleviste, A. Remm, K. Fischer, T. Esko, H.-J. Westra, B. P. Fairfax, S. Makino, J. C. Knight, L. Franke, A. Metspalu, P. Peterson, L. Milani, Pathogenic implications for autoimmune mechanisms derived by comparative eQTL analysis of $CD4^+$ versus $CD8^+$ T cells. *PLOS Genet.* **13**, e1006643 (2017). doi:10.1371/journal.pgen.1006643 Medline

9. M. G. P. van der Wijst, H. Brugge, D. H. de Vries, P. Deelen, M. A. Swertz, L. Franke, LifeLines Cohort Study, BIOS Consortium, Single-cell RNA sequencing identifies celltype-specific cis-eQTLs and co-expression QTLs. *Nat. Genet.* **50**, 493–497 (2018). doi:10.1038/s41588-018-0089-9 Medline

10. D. V. Zhernakova, P. Deelen, M. Vermaat, M. van Iterson, M. van Galen, W. Arindrarto, P. van ’t Hof, H. Mei, F. van Dijk, H.-J. Westra, M. J. Bonder, J. van Rooij, M. Verkerk, P. M. Jhamai, M. Moed, S. M. Kielbasa, J. Bot, I. Nooren, R. Pool, J. van Dongen, J. J. Hottenga, C. D. A. Stehouwer, C. J. H. van der Kallen, C. G. Schalkwijk, A. Zhernakova, Y. Li, E. F. Tigchelaar, N. de Klein, M. Beekman, J. Deelen, D. van Heemst, L. H. van den Berg, A. Hofman, A. G. Uitterlinden, M. M. J. van Greevenbroek, J. H. Veldink, D. I. Boomsma, C. M. van Duijn, C. Wijmenga, P. E. Slagboom, M. A. Swertz, A. Isaacs, J. B. J. van Meurs, R. Jansen, B. T. Heijmans, P. A. C. ’t Hoen, L. Franke, Identification of context-dependent expression quantitative trait loci in whole blood. *Nat. Genet.* **49**, 139–145 (2017). doi:10.1038/ng.3737 Medline

11. 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 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. A. S. E. Cuomo, D. D. Seaton, D. J. McCarthy, I. Martinez, M. J. Bonder, J. Garcia-Bernardo, S. Amatya, P. Madrigal, A. Isaacson, F. Buettner, A. Knights, K. N. Natarajan, L. Vallier, J. C. Marioni, M. Chhatriwala, O. Stegle, HipSci Consortium, Single-cell RNA-sequencing of differentiating iPS cells reveals dynamic genetic effects on gene expression. *Nat. Commun.* **11**, 810 (2020). doi:10.1038/s41467-020-14457-z Medline

14. B. J. Strober, R. Elorbany, 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 Medline

15. N. G. Skene, J. Bryois, T. E. Bakken, G. Breen, J. J. Crowley, H. A. Gaspar, P. Giusti-Rodriguez, R. D. Hodge, J. A. Miller, A. B. Muñoz-Manchado, M. C. O'Donovan, M. J. Owen, A. F. Pardiñas, J. Ryge, J. T. R. Walters, S. Linnarsson, E. S. Lein, P. F. Sullivan, J. Hjerling-Leffler, Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium, Genetic identification of brain cell types underlying schizophrenia. *Nat. Genet.* **50**, 825–833 (2018). doi:10.1038/s41588-018-0129-5 Medline

16. C. S. Smillie, M. Biton, J. Ordovas-Montanes, K. M. Sullivan, G. Burgin, D. B. Graham, R. H. Herbst, N. Rogel, M. Slyper, J. Waldman, M. Sud, E. Andrews, G. Velonias, A. L. Haber, K. Jagadeesh, S. Vickovic, J. Yao, C. Stevens, D. Dionne, L. T. Nguyen, A.-C. Villani, M. Hofree, E. A. Creasey, H. Huang, O. Rozenblatt-Rosen, J. J. Garber, H. Khalili, A. N. Desch, M. J. Daly, A. N. Ananthakrishnan, A. K. Shalek, R. J. Xavier, A. Regev, Intra- and inter-cellular rewiring of the human colon during ulcerative colitis. *Cell* **178**, 714–730.e22 (2019). doi:10.1016/j.cell.2019.06.029 Medline

17. E. M. Weeks, J. C. Ulirsch, N. Y. Cheng, B. L. Trippe, R. S. Fine, J. Miao, T. A. Patwardhan, M. Kanai, J. Nasser, C. P. Fulco, K. C. Tashman, F. Aguet, T. Li, J. Ordovas-Montanes, C. S. Smillie, M. Biton, A. K. Shalek, A. N. Ananthakrishnan, R. J. Xavier, A. Regev, R. M. Gupta, K. Lage, K. G. Ardlie, J. N. Hirschhorn, E. S. Lander, J. M. Engreitz, H. K. Finucane, Leveraging polygenic enrichments of gene features to predict genes underlying complex traits and diseases. medRxiv 2020.09.08.20190561 [Preprint] (2020); https://doi.org/10.1101/2020.09.08.20190561.

18. K. A. Jagadeesh, K. K. Dey, D. T. Montoro, R. Mohan, S. Gazal, J. M. Engreitz, R. J. Xavier, A. L. Price, A. Regev, Identifying disease-critical cell types and cellular processes across the human body by integration of single-cell profiles and human genetics. bioRxiv 2021.03.19.436212 [Preprint] (2021); https://doi.org/10.1101/2021.03.19.436212.

19. N. Habib, I. Avraham-Davidi, A. Basu, T. Burks, K. Shekhar, M. Hofree, S. R. Choudhury, F. Aguet, E. Gelfand, K. Ardlie, D. A. Weitz, O. Rozenblatt-Rosen, F. Zhang, A. Regev,

Massively parallel single-nucleus RNA-seq with DroNc-seq. *Nat. Methods* **14**, 955–958 (2017). doi:10.1038/nmeth.4407 Medline

20. N. Habib, Y. Li, M. Heidenreich, L. Swiech, I. Avraham-Davidi, J. J. Trombetta, C. Hession, F. Zhang, A. Regev, Div-Seq: Single-nucleus RNA-seq reveals dynamics of rare adult newborn neurons. *Science* **353**, 925–928 (2016). doi:10.1126/science.aad7038 Medline

21. M. J. Petrany, C. O. Swoboda, C. Sun, K. Chetal, X. Chen, M. T. Weirauch, N. Salomonis, D. P. Millay, Single-nucleus RNA-seq identifies transcriptional heterogeneity in multinucleated skeletal myofibers. *Nat. Commun.* **11**, 6374 (2020). doi:10.1038/s41467-020-20063-w Medline

22. H. Wu, Y. Kirita, E. L. Donnelly, B. D. Humphreys, Advantages of single-nucleus over single-cell RNA sequencing of adult kidney: Rare cell types and novel cell states revealed in fibrosis. *J. Am. Soc. Nephrol.* **30**, 23–32 (2019). doi:10.1681/ASN.2018090912 Medline

23. W. Sun, H. Dong, M. Balaz, M. Slyper, E. Drokhlyansky, G. Colleluori, A. Giordano, Z. Kovanicova, P. Stefanicka, L. Balazova, L. Ding, A. S. Husted, G. Rudofsky, J. Ukropec, S. Cinti, T. W. Schwartz, A. Regev, C. Wolfrum, snRNA-seq reveals a subpopulation of adipocytes that regulates thermogenesis. *Nature* **587**, 98–102 (2020). doi:10.1038/s41586-020-2856-x Medline

24. 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

25. M. van der Wijst, D. H. de Vries, H. E. Groot, G. Trynka, C. C. Hon, M. J. Bonder, O. Stegle, M. C. Nawijn, Y. Idaghdour, P. van der Harst, C. J. Ye, J. Powell, F. J. Theis, A. Mahfouz, M. Heinig, L. Franke, The single-cell eQTLGen consortium. *eLife* **9**, e52155 (2020). doi:10.7554/eLife.52155 Medline

26. E. Drokhlyansky, C. S. Smillie, N. Van Wittenberghe, M. Ericsson, G. K. Griffin, G. Eraslan, D. Dionne, M. S. Cuoco, M. N. Goder-Reiser, T. Sharova, O. Kuksenko, A. J. Aguirre, G. M. Boland, D. Graham, O. Rozenblatt-Rosen, R. J. Xavier, A. Regev, The human and mouse enteric nervous system at single-cell resolution. *Cell* **182**, 1606–1622.e23 (2020). doi:10.1016/j.cell.2020.08.003 Medline

27. T. M. Delorey, C. G. K. Ziegler, G. Heimberg, R. Normand, Y. Yang, Å. Segerstolpe, D. Abbondanza, S. J. Fleming, A. Subramanian, D. T. Montoro, K. A. Jagadeesh, K. K. Dey, P. Sen, M. Slyper, Y. H. Pita-Juárez, D. Phillips, J. Biermann, Z. Bloom-Ackermann, N. Barkas, A. Ganna, J. Gomez, J. C. Melms, I. Katsyv, E. Normandin, P. Naderi, Y. V. Popov, S. S. Raju, S. Niezen, L. T.-Y. Tsai, K. J. Siddle, M. Sud, V. M. Tran, S. K. Vellarikkal, Y. Wang, L. Amir-Zilberstein, D. S. Atri, J. Beechem, O. R. Brook, J. Chen, P. Divakar, P. Dorceus, J. M. Engreitz, A. Essene, D. M. Fitzgerald, R. Fropf, S. Gazal, J. Gould, J. Grzyb, T. Harvey, J. Hecht, T. Hether, J. Jané-Valbuena, M. Leney-Greene, H. Ma, C. McCabe, D. E. McLoughlin, E. M. Miller, C. Muus, M. Niemi, R. Padera, L. Pan, D. Pant, C. Pe'er, J. Pfiffner-Borges, C. J. Pinto, J. Plaisted, J. Reeves, M. Ross, M. Rudy, E. H. Rueckert, M. Siciliano, A. Sturm, E. Todres, A. Waghray, S. Warren, S. Zhang, D. R. Zollinger, L. Cosimi, R. M. Gupta, N. Hacohen, H. Hibshoosh, W. Hide, A. L. Price, J. Rajagopal, P. R. Tata, S. Riedel, G. Szabo, T. L. Tickle, P. T.

Ellinor, D. Hung, P. C. Sabeti, R. Novak, R. Rogers, D. E. Ingber, Z. G. Jiang, D. Juric, M. Babadi, S. L. Farhi, B. Izar, J. R. Stone, I. S. Vlachos, I. H. Solomon, O. Ashenberg, C. B. M. Porter, B. Li, A. K. Shalek, A.-C. Villani, O. Rozenblatt-Rosen, A. Regev, COVID-19 tissue atlases reveal SARS-CoV-2 pathology and cellular targets. *Nature* **595**, 107–113 (2021). doi:10.1038/s41586-021-03570-8 Medline

28. Materials and methods are available as supplementary materials.

29. M. Slyper, C. B. M. Porter, O. Ashenberg, J. Waldman, E. Drokhlyansky, I. Wakiro, C. Smillie, G. Smith-Rosario, J. Wu, D. Dionne, S. Vigneau, J. Jané-Valbuena, T. L. Tickle, S. Napolitano, M.-J. Su, A. G. Patel, A. Karlstrom, S. Gritsch, M. Nomura, A. Waghray, S. H. Gohil, A. M. Tsankov, L. Jerby-Arnon, O. Cohen, J. Klughammer, Y. Rosen, J. Gould, L. Nguyen, M. Hofree, P. J. Tramontozzi, B. Li, C. J. Wu, B. Izar, R. Haq, F. S. Hodi, C. H. Yoon, A. N. Hata, S. J. Baker, M. L. Suvà, R. Bueno, E. H. Stover, M. R. Clay, M. A. Dyer, N. B. Collins, U. A. Matulonis, N. Wagle, B. E. Johnson, A. Rotem, O. Rozenblatt-Rosen, A. Regev, A single-cell and single-nucleus RNA-seq toolbox for fresh and frozen human tumors. *Nat. Med.* **26**, 792–802 (2020). doi:10.1038/s41591-020-0844-1 Medline

30. S. J. Fleming, J. C. Marioni, M. Babadi, CellBender remove-background: a deep generative model for unsupervised removal of background noise from scRNA-seq datasets. bioRxiv 791699 [Preprint] (2019); https://doi.org/10.1101/791699.

31. M. Kim, V. Franke, B. Brandt, E. D. Lowenstein, V. Schöwel, S. Spuler, A. Akalin, C. Birchmeier, Single-nucleus transcriptomics reveals functional compartmentalization in syncytial skeletal muscle cells. *Nat. Commun.* **11**, 6375 (2020). doi:10.1038/s41467-020-20064-9 Medline

32. M. Wolfien, A.-M. Galow, P. Müller, M. Bartsch, R. M. Brunner, T. Goldammer, O. Wolkenhauer, A. Hoeflich, R. David, Single-nucleus sequencing of an entire mammalian heart: Cell type composition and velocity. *Cells* **9**, 318 (2020). doi:10.3390/cells9020318 Medline

33. N. R. Tucker, M. Chaffin, S. J. Fleming, A. W. Hall, V. A. Parsons, K. C. Bedi Jr., A. D. Akkad, C. N. Herndon, A. Arduini, I. Papangeli, C. Roselli, F. Aguet, S. H. Choi, K. G. Ardlie, M. Babadi, K. B. Margulies, C. M. Stegmann, P. T. Ellinor, Transcriptional and cellular diversity of the human heart. *Circulation* **142**, 466–482 (2020). doi:10.1161/CIRCULATIONAHA.119.045401 Medline

34. M. Litviňuková, C. Talavera-López, H. Maatz, D. Reichart, C. L. Worth, E. L. Lindberg, M. Kanda, K. Polanski, M. Heinig, M. Lee, E. R. Nadelmann, K. Roberts, L. Tuck, E. S. Fasouli, D. M. DeLaughter, B. McDonough, H. Wakimoto, J. M. Gorham, S. Samari, K. T. Mahbubani, K. Saeb-Parsy, G. Patone, J. J. Boyle, H. Zhang, H. Zhang, A. Viveiros, G. Y. Oudit, O. A. Bayraktar, J. G. Seidman, C. E. Seidman, M. Noseda, N. Hubner, S. A. Teichmann, Cells of the adult human heart. *Nature* **588**, 466–472 (2020). doi:10.1038/s41586-020-2797-4 Medline

35. R. X. Verma, S. Kannan, B. L. Lin, K. M. Fomchenko, T. O. Nieuwenhuis, A. H. Patil, C. Lukban, X. Yang, K. Fox-Talbot, M. N. McCall, C. Kwon, D. A. Kass, A. Z. Rosenberg, M. K. Halushka, Single cell RNA-seq analysis of the flexor digitorum brevis mouse myofibers. *Skelet. Muscle* **11**, 13 (2021). doi:10.1186/s13395-021-00269-2 Medline

36. A. B. Rubenstein, G. R. Smith, U. Raue, G. Begue, K. Minchev, F. Ruf-Zamojski, V. D.

Nair, X. Wang, L. Zhou, E. Zaslavsky, T. A. Trappe, S. Trappe, S. C. Sealfon, Single-cell transcriptional profiles in human skeletal muscle. *Sci. Rep.* **10**, 229 (2020). doi:10.1038/s41598-019-57110-6 Medline

37. K. R. Jessen, R. Mirsky, The success and failure of the Schwann cell response to nerve injury. *Front. Cell. Neurosci.* **13**, 33 (2019). doi:10.3389/fncel.2019.00033 Medline

38. G. H. Henry, A. Malewska, D. B. Joseph, V. S. Malladi, J. Lee, J. Torrealba, R. J. Mauck, J. C. Gahan, G. V. Raj, C. G. Roehrborn, G. C. Hon, M. P. MacConmara, J. C. Reese, R. C. Hutchinson, C. M. Vezina, D. W. Strand, A cellular anatomy of the normal adult human prostate and prostatic urethra. *Cell Rep.* **25**, 3530–3542.e5 (2018). doi:10.1016/j.celrep.2018.11.086 Medline

39. M. D. Luecken, F. J. Theis, Current best practices in single-cell RNA-seq analysis: A tutorial. *Mol. Syst. Biol.* **15**, e8746 (2019). doi:10.15252/msb.20188746 Medline

40. K. J. Travaglini, A. N. Nabhan, L. Penland, R. Sinha, A. Gillich, R. V. Sit, S. Chang, S. D. Conley, Y. Mori, J. Seita, G. J. Berry, J. B. Shrager, R. J. Metzger, C. S. Kuo, N. Neff, I. L. Weissman, S. R. Quake, M. A. Krasnow, A molecular cell atlas of the human lung from single-cell RNA sequencing. *Nature* **587**, 619–625 (2020). doi:10.1038/s41586-020-2922-4 Medline

41. S. C. van den Brink, F. Sage, Á. Vértesy, B. Spanjaard, J. Peterson-Maduro, C. S. Baron, C. Robin, A. van Oudenaarden, Single-cell sequencing reveals dissociation-induced gene expression in tissue subpopulations. *Nat. Methods* **14**, 935–936 (2017). doi:10.1038/nmeth.4437 Medline

42. E. Denisenko, B. B. Guo, M. Jones, R. Hou, L. de Kock, T. Lassmann, D. Poppe, O. Clément, R. K. Simmons, R. Lister, A. R. R. Forrest, Systematic assessment of tissue dissociation and storage biases in single-cell and single-nucleus RNA-seq workflows. *Genome Biol.* **21**, 130 (2020). doi:10.1186/s13059-020-02048-6 Medline

43. A. Zaghlool, A. Niazi, Å. K. Björklund, J. O. Westholm, A. Ameur, L. Feuk, Characterization of the nuclear and cytosolic transcriptomes in human brain tissue reveals new insights into the subcellular distribution of RNA transcripts. *Sci. Rep.* **11**, 4076 (2021). doi:10.1038/s41598-021-83541-1 Medline

44. M. Rabani, R. Raychowdhury, M. Jovanovic, M. Rooney, D. J. Stumpo, A. Pauli, N. Hacohen, A. F. Schier, P. J. Blackshear, N. Friedman, I. Amit, A. Regev, High-resolution sequencing and modeling identifies distinct dynamic RNA regulatory strategies. *Cell* **159**, 1698–1710 (2014). doi:10.1016/j.cell.2014.11.015 Medline

45. B. W. Solnestam, H. Stranneheim, J. Hällman, M. Käller, E. Lundberg, J. Lundeberg, P. Akan, Comparison of total and cytoplasmic mRNA reveals global regulation by nuclear retention and miRNAs. *BMC Genomics* **13**, 574 (2012). doi:10.1186/1471-2164-13-574 Medline

46. W. L. Hwang, K. A. Jagadeesh, J. A. Guo, H. I. Hoffman, P. Yadollahpour, R. Mohan, E. Drokhlyansky, N. Van Wittenberghe, O. Ashenberg, S. Farhi, D. Schapiro, J. Reeves, D. R. Zollinger, G. Eng, J. M. Schenkel, W. A. Freed-Pastor, C. Rodrigues, J. Gould, C. Lambden, C. Porter, A. Tsankov, D. Dionne, D. Abbondanza, J. Waldman, M. Cuoco, L. Nguyen, T. Delorey, D. Phillips, D. Ciprani, M. Kern, A. Mehta, K. Fuhrman, R. Fropf, J. Beechem, J. S. Loeffler, D. P. Ryan, C. D. Weekes, D. T. Ting, C. R. Ferrone, J. Y.

Wo, T. S. Hong, A. J. Aguirre, O. Rozenblatt-Rosen, M. Mino-Kenudson, C. F. Castillo, A. S. Liss, T. Jacks, A. Regev, Single-nucleus and spatial transcriptomics of archival pancreatic cancer reveals multi-compartment reprogramming after neoadjuvant treatment. bioRxiv 2020.08.25.267336 [Preprint] (2020); https://doi.org/10.1101/2020.08.25.267336.

47. S. Chakarov, H. Y. Lim, L. Tan, S. Y. Lim, P. See, J. Lum, X.-M. Zhang, S. Foo, S. Nakamizo, K. Duan, W. T. Kong, R. Gentek, A. Balachander, D. Carbajo, C. Bleriot, B. Malleret, J. K. C. Tam, S. Baig, M. Shabeer, S. E. S. Toh, A. Schlitzer, A. Larbi, T. Marichal, B. Malissen, J. Chen, M. Poidinger, K. Kabashima, M. Bajenoff, L. G. Ng, V. Angeli, F. Ginhoux, Two distinct interstitial macrophage populations coexist across tissues in specific subtissular niches. *Science* **363**, eaau0964 (2019). doi:10.1126/science.aau0964 Medline

48. C. Varol, A. Mildner, S. Jung, Macrophages: Development and tissue specialization. *Annu. Rev. Immunol.* **33**, 643–675 (2015). doi:10.1146/annurev-immunol-032414-112220 Medline

49. 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

50. P. A. Reyfman, J. M. Walter, N. Joshi, K. R. Anekalla, A. C. McQuattie-Pimentel, S. Chiu, R. Fernandez, M. Akbarpour, C.-I. Chen, Z. Ren, R. Verma, H. Abdala-Valencia, K. Nam, M. Chi, S. Han, F. J. Gonzalez-Gonzalez, S. Soberanes, S. Watanabe, K. J. N. Williams, A. S. Flozak, T. T. Nicholson, V. K. Morgan, D. R. Winter, M. Hinchcliff, C. L. Hrusch, R. D. Guzy, C. A. Bonham, A. I. Sperling, R. Bag, R. B. Hamanaka, G. M. Mutlu, A. V. Yeldandi, S. A. Marshall, A. Shilatifard, L. A. N. Amaral, H. Perlman, J. I. Sznajder, A. C. Argento, C. T. Gillespie, J. Dematte, M. Jain, B. D. Singer, K. M. Ridge, A. P. Lam, A. Bharat, S. M. Bhorade, C. J. Gottardi, G. R. S. Budinger, A. V. Misharin, Single-cell transcriptomic analysis of human lung provides insights into the pathobiology of pulmonary fibrosis. *Am. J. Respir. Crit. Care Med.* **199**, 1517–1536 (2019). doi:10.1164/rccm.201712-2410OC Medline

51. D. A. Jaitin, L. Adlung, C. A. Thaiss, A. Weiner, B. Li, H. Descamps, P. Lundgren, C. Bleriot, Z. Liu, A. Deczkowska, H. Keren-Shaul, E. David, N. Zmora, S. M. Eldar, N. Lubezky, O. Shibolet, D. A. Hill, M. A. Lazar, M. Colonna, F. Ginhoux, H. Shapiro, E. Elinav, I. Amit, Lipid-associated macrophages control metabolic homeostasis in a Trem2-dependent manner. *Cell* **178**, 686–698.e14 (2019). doi:10.1016/j.cell.2019.05.054 Medline

52. T. Capucha, G. Mizraji, H. Segev, R. Blecher-Gonen, D. Winter, A. Khalaileh, Y. Tabib, T. Attal, M. Nassar, K. Zelentsova, H. Kisos, M. Zenke, K. Seré, T. Hieronymus, T. Burstyn-Cohen, I. Amit, A. Wilensky, A.-H. Hovav, Distinct murine mucosal langerhans cell subsets develop from pre-dendritic cells and monocytes. *Immunity* **43**, 369–381 (2015). doi:10.1016/j.immuni.2015.06.017 Medline

53. J. Deckers, H. Hammad, E. Hoste, Langerhans cells: Sensing the environment in health and

disease. *Front. Immunol.* **9**, 93 (2018). doi:10.3389/fimmu.2018.00093 Medline

54. E. Evren, E. Ringqvist, K. P. Tripathi, N. Sleiers, I. C. Rives, A. Alisjahbana, Y. Gao, D. Sarhan, T. Halle, C. Sorini, R. Lepzien, N. Marquardt, J. Michaëlsson, A. Smed-Sörensen, J. Botling, M. C. I. Karlsson, E. J. Villablanca, T. Willinger, Distinct developmental pathways from blood monocytes generate human lung macrophage diversity. *Immunity* **54**, 259–275.e7 (2021). doi:10.1016/j.immuni.2020.12.003 Medline

55. A. A. Patel, Y. Zhang, J. N. Fullerton, L. Boelen, A. Rongvaux, A. A. Maini, V. Bigley, R. A. Flavell, D. W. Gilroy, B. Asquith, D. Macallan, S. Yona, The fate and lifespan of human monocyte subsets in steady state and systemic inflammation. *J. Exp. Med.* **214**, 1913–1923 (2017). doi:10.1084/jem.20170355 Medline

56. H. Y. Lim, S. Y. Lim, C. K. Tan, C. H. Thiam, C. C. Goh, D. Carbajo, S. H. S. Chew, P. See, S. Chakarov, X. N. Wang, L. H. Lim, L. A. Johnson, J. Lum, C. Y. Fong, A. Bongso, A. Biswas, C. Goh, M. Evrard, K. P. Yeo, R. Basu, J. K. Wang, Y. Tan, R. Jain, S. Tikoo, C. Choong, W. Weninger, M. Poidinger, R. E. Stanley, M. Collin, N. S. Tan, L. G. Ng, D. G. Jackson, F. Ginhoux, V. Angeli, Hyaluronan receptor LYVE-1-expressing macrophages maintain arterial tone through hyaluronan-mediated regulation of smooth muscle cell collagen. *Immunity* **49**, 326–341.e7 (2018). doi:10.1016/j.immuni.2018.06.008 Medline

57. M. Bizou, R. Itier, M. Majdoubi, D. Abbadi, E. Pichery, M. Dutaur, D. Marsal, D. Calise, B. Garmy-Susini, V. Douin-Echinard, J. Roncalli, A. Parini, N. Pizzinat, Cardiac macrophage subsets differentially regulate lymphatic network remodeling during pressure overload. *Sci. Rep.* **11**, 16801 (2021). doi:10.1038/s41598-021-95723-y Medline

58. P. Ramachandran, R. Dobie, J. R. Wilson-Kanamori, E. F. Dora, B. E. P. Henderson, N. T. Luu, J. R. Portman, K. P. Matchett, M. Brice, J. A. Marwick, R. S. Taylor, M. Efremova, R. Vento-Tormo, N. O. Carragher, T. J. Kendall, J. A. Fallowfield, E. M. Harrison, D. J. Mole, S. J. Wigmore, P. N. Newsome, C. J. Weston, J. P. Iredale, F. Tacke, J. W. Pollard, C. P. Ponting, J. C. Marioni, S. A. Teichmann, N. C. Henderson, Resolving the fibrotic niche of human liver cirrhosis at single-cell level. *Nature* **575**, 512–518 (2019). doi:10.1038/s41586-019-1631-3 Medline

59. M. J. Perugorria, A. Esparza-Baquer, F. Oakley, I. Labiano, A. Korosec, A. Jais, J. Mann, D. Tiniakos, A. Santos-Laso, A. Arbelaiz, R. Gawish, A. Sampedro, A. Fontanellas, E. Hijona, R. Jimenez-Agüero, H. Esterbauer, D. Stoiber, L. Bujanda, J. M. Banales, S. Knapp, O. Sharif, D. A. Mann, Non-parenchymal TREM-2 protects the liver from immune-mediated hepatocellular damage. *Gut* **68**, 533–546 (2019). doi:10.1136/gutjnl-2017-314107 Medline

60. X. Xiong, H. Kuang, S. Ansari, T. Liu, J. Gong, S. Wang, X.-Y. Zhao, Y. Ji, C. Li, L. Guo, L. Zhou, Z. Chen, P. Leon-Mimila, M. T. Chung, K. Kurabayashi, J. Opp, F. Campos-Pérez, H. Villamil-Ramírez, S. Canizales-Quinteros, R. Lyons, C. N. Lumeng, B. Zhou, L. Qi, A. Huertas-Vazquez, A. J. Lusis, X. Z. S. Xu, S. Li, Y. Yu, J. Z. Li, J. D. Lin, Landscape of intercellular crosstalk in healthy and NASH liver revealed by single-cell secretome gene analysis. *Mol. Cell* **75**, 644–660.e5 (2019). doi:10.1016/j.molcel.2019.07.028 Medline

61. M. Guilliams, J. Bonnardel, B. Haest, B. Vanderborght, C. Wagner, A. Remmerie, A. Bujko, L. Martens, T. Thoné, R. Browaeys, F. F. De Ponti, B. Vanneste, C. Zwicker, F. R.

Svedberg, T. Vanhalewyn, A. Gonçalves, S. Lippens, B. Devriendt, E. Cox, G. Ferrero, V. Wittamer, A. Willaert, S. J. F. Kaptein, J. Neyts, K. Dallmeier, P. Geldhof, S. Casaert, B. Deplancke, P. Ten Dijke, A. Hoorens, A. Vanlander, F. Berrevoet, Y. Van Nieuwenhove, Y. Saeys, W. Saelens, H. Van Vlierberghe, L. Devisscher, C. L. Scott, Spatial proteogenomics reveals distinct and evolutionarily conserved hepatic macrophage niches. *Cell* **185**, 379–396.e38 (2022). doi:10.1016/j.cell.2021.12.018 Medline

62. E. A. Ayaub, S. Poli, J. Ng, T. Adams, J. Schupp, L. Quesada-Arias, F. Poli, C. Cosme, M. Robertson, J. Martinez-Manzano, X. Liang, J. Villalba, J. Lederer, S. G. Chu, B. A. Raby, G. Washko, C. Coarfa, M. A. Perrella, S. El-Chemaly, N. Kaminski, I. O. Rosas, Single cell RNA-seq and mass cytometry reveals a novel and a targetable population of macrophages in idiopathic pulmonary fibrosis. bioRxiv 2021.01.04.425268 [Preprint] (2021); https://doi.org/10.1101/2021.01.04.425268.

63. C. Cochain, E. Vafadarnejad, P. Arampatzi, J. Pelisek, H. Winkels, K. Ley, D. Wolf, A.-E. Saliba, A. Zernecke, Single-cell RNA-seq reveals the transcriptional landscape and heterogeneity of aortic macrophages in murine atherosclerosis. *Circ. Res.* **122**, 1661–1674 (2018). doi:10.1161/CIRCRESAHA.117.312509 Medline

64. L. Willemsen, M. P. J. de Winther, Macrophage subsets in atherosclerosis as defined by single-cell technologies. *J. Pathol.* **250**, 705–714 (2020). doi:10.1002/path.5392 Medline

65. F. Ma, T. K. Hughes, R. M. B. Teles, P. R. Andrade, B. J. de Andrade Silva, O. Plazyo, L. C. Tsoi, T. Do, M. H. Wadsworth II, A. Oulee, M. T. Ochoa, E. N. Sarno, M. L. Iruela-Arispe, E. Klechevsky, B. Bryson, A. K. Shalek, B. R. Bloom, J. E. Gudjonsson, M. Pellegrini, R. L. Modlin, The cellular architecture of the antimicrobial response network in human leprosy granulomas. *Nat. Immunol.* **22**, 839–850 (2021). doi:10.1038/s41590-021-00956-8 Medline

66. H. Keren-Shaul, A. Spinrad, A. Weiner, O. Matcovitch-Natan, R. Dvir-Szternfeld, T. K. Ulland, E. David, K. Baruch, D. Lara-Astaiso, B. Toth, S. Itzkovitz, M. Colonna, M. Schwartz, I. Amit, A unique microglia type associated with restricting development of Alzheimer's disease. *Cell* **169**, 1276–1290.e17 (2017). doi:10.1016/j.cell.2017.05.018 Medline

67. A. Grubman, G. Chew, J. F. Ouyang, G. Sun, X. Y. Choo, C. McLean, R. K. Simmons, S. Buckberry, D. B. Vargas-Landin, D. Poppe, J. Pflueger, R. Lister, O. J. L. Rackham, E. Petretto, J. M. Polo, A single-cell atlas of entorhinal cortex from individuals with Alzheimer's disease reveals cell-type-specific gene expression regulation. *Nat. Neurosci.* **22**, 2087–2097 (2019). doi:10.1038/s41593-019-0539-4 Medline

68. H. Mathys, J. Davila-Velderrain, Z. Peng, F. Gao, S. Mohammadi, J. Z. Young, M. Menon, L. He, F. Abdurrob, X. Jiang, A. J. Martorell, R. M. Ransohoff, B. P. Hafler, D. A. Bennett, M. Kellis, L.-H. Tsai, Single-cell transcriptomic analysis of Alzheimer's disease. *Nature* **570**, 332–337 (2019). doi:10.1038/s41586-019-1195-2 Medline

69. C. Muus, M. D. Luecken, G. Eraslan, L. Sikkema, A. Waghray, G. Heimberg, Y. Kobayashi, E. D. Vaishnav, A. Subramanian, C. Smillie, K. A. Jagadeesh, E. T. Duong, E. Fiskin, E. Torlai Triglia, M. Ansari, P. Cai, B. Lin, J. Buchanan, S. Chen, J. Shu, A. L. Haber, H. Chung, D. T. Montoro, T. Adams, H. Aliee, S. J. Allon, Z. Andrusivova, I. Angelidis, O. Ashenberg, K. Bassler, C. Bécavin, I. Benhar, J. Bergenstråhle, L. Bergenstråhle, L. Bolt, E. Braun, L. T. Bui, S. Callori, M. Chaffin, E. Chichelnitskiy, J. Chiou, T. M. Conlon, M.

S. Cuoco, A. S. E. Cuomo, M. Deprez, G. Duclos, D. Fine, D. S. Fischer, S. Ghazanfar, A. Gillich, B. Giotti, J. Gould, M. Guo, A. J. Gutierrez, A. C. Habermann, T. Harvey, P. He, X. Hou, L. Hu, Y. Hu, A. Jaiswal, L. Ji, P. Jiang, T. S. Kapellos, C. S. Kuo, L. Larsson, M. A. Leney-Greene, K. Lim, M. Litviňuková, L. S. Ludwig, S. Lukassen, W. Luo, H. Maatz, E. Madissoon, L. Mamanova, K. Manakongtreecheep, S. Leroy, C. H. Mayr, I. M. Mbano, A. M. McAdams, A. N. Nabhan, S. K. Nyquist, L. Penland, O. B. Poirion, S. Poli, C. Qi, R. Queen, D. Reichart, I. Rosas, J. C. Schupp, C. V. Shea, X. Shi, R. Sinha, R. V. Sit, K. Slowikowski, M. Slyper, N. P. Smith, A. Sountoulidis, M. Strunz, T. B. Sullivan, D. Sun, C. Talavera-López, P. Tan, J. Tantivit, K. J. Travaglini, N. R. Tucker, K. A. Vernon, M. H. Wadsworth, J. Waldman, X. Wang, K. Xu, W. Yan, W. Zhao, C. G. K. Ziegler, NHLBI LungMap Consortium, Human Cell Atlas Lung Biological Network, Single-cell meta-analysis of SARS-CoV-2 entry genes across tissues and demographics. *Nat. Med.* **27**, 546–559 (2021). doi:10.1038/s41591-020-01227-z Medline

70. R. C. Wirka, D. Wagh, D. T. Paik, M. Pjanic, T. Nguyen, C. L. Miller, R. Kundu, M. Nagao, J. Coller, T. K. Koyano, R. Fong, Y. J. Woo, B. Liu, S. B. Montgomery, J. C. Wu, K. Zhu, R. Chang, M. Alamprese, M. D. Tallquist, J. B. Kim, T. Quertermous, Atheroprotective roles of smooth muscle cell phenotypic modulation and the *TCF21* disease gene as revealed by single-cell analysis. *Nat. Med.* **25**, 1280–1289 (2019). doi:10.1038/s41591-019-0512-5 Medline

71. T. Alsaigh, D. Evans, D. Frankel, A. Torkamani, Decoding the transcriptome of atherosclerotic plaque at single-cell resolution. bioRxiv 2020.03.03.968123 [Preprint] (2020); https://doi.org/10.1101/2020.03.03.968123.

72. R. Vento-Tormo, M. Efremova, R. A. Botting, M. Y. Turco, M. Vento-Tormo, K. B. Meyer, J.-E. Park, E. Stephenson, K. Polański, A. Goncalves, L. Gardner, S. Holmqvist, J. Henriksson, A. Zou, A. M. Sharkey, B. Millar, B. Innes, L. Wood, A. Wilbrey-Clark, R. P. Payne, M. A. Ivarsson, S. Lisgo, A. Filby, D. H. Rowitch, J. N. Bulmer, G. J. Wright, M. J. T. Stubbington, M. Haniffa, A. Moffett, S. A. Teichmann, Single-cell reconstruction of the early maternal-fetal interface in humans. *Nature* **563**, 347–353 (2018). doi:10.1038/s41586-018-0698-6 Medline

73. K. Tan, H.-W. Song, M. Thompson, S. Munyoki, M. Sukhwani, T.-C. Hsieh, K. E. Orwig, M. F. Wilkinson, Transcriptome profiling reveals signaling conditions dictating human spermatogonia fate in vitro. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 17832–17841 (2020). doi:10.1073/pnas.2000362117 Medline

74. J. Liao, Z. Yu, Y. Chen, M. Bao, C. Zou, H. Zhang, D. Liu, T. Li, Q. Zhang, J. Li, J. Cheng, Z. Mo, Single-cell RNA sequencing of human kidney. *Sci. Data* **7**, 4 (2020). doi:10.1038/s41597-019-0351-8 Medline

75. M. M. F. Qadir, S. Álvarez-Cubela, D. Klein, J. van Dijk, R. Muñiz-Anquela, Y. B. Moreno-Hernández, G. Lanzoni, S. Sadiq, B. Navarro-Rubio, M. T. García, Á. Díaz, K. Johnson, D. Sant, C. Ricordi, A. Griswold, R. L. Pastori, J. Domínguez-Bendala, Single-cell resolution analysis of the human pancreatic ductal progenitor cell niche. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 10876–10887 (2020). doi:10.1073/pnas.1918314117 Medline

76. S. A. MacParland, J. C. Liu, X.-Z. Ma, B. T. Innes, A. M. Bartczak, B. K. Gage, J. Manuel, N. Khuu, J. Echeverri, I. Linares, R. Gupta, M. L. Cheng, L. Y. Liu, D. Camat, S. W.

Chung, R. K. Seliga, Z. Shao, E. Lee, S. Ogawa, M. Ogawa, M. D. Wilson, J. E. Fish, M. Selzner, A. Ghanekar, D. Grant, P. Greig, G. Sapisochin, N. Selzner, N. Winegarden, O. Adeyi, G. Keller, G. D. Bader, I. D. McGilvray, Single cell RNA sequencing of human liver reveals distinct intrahepatic macrophage populations. *Nat. Commun.* **9**, 4383 (2018). doi:10.1038/s41467-018-06318-7 Medline

77. T. Chitiashvili, I. Dror, R. Kim, F.-M. Hsu, R. Chaudhari, E. Pandolfi, D. Chen, S. Liebscher, K. Schenke-Layland, K. Plath, A. Clark, Female human primordial germ cells display X-chromosome dosage compensation despite the absence of X-inactivation. *Nat. Cell Biol.* **22**, 1436–1446 (2020). doi:10.1038/s41556-020-00607-4 Medline

78. A. J. De Micheli, J. A. Spector, O. Elemento, B. D. Cosgrove, A reference single-cell transcriptomic atlas of human skeletal muscle tissue reveals bifurcated muscle stem cell populations. *Skelet. Muscle* **10**, 19 (2020). doi:10.1186/s13395-020-00236-3 Medline

79. T. K. Hughes, M. H. Wadsworth II, T. M. Gierahn, T. Do, D. Weiss, P. R. Andrade, F. Ma, B. J. de Andrade Silva, S. Shao, L. C. Tsoi, J. Ordovas-Montanes, J. E. Gudjonsson, R. L. Modlin, J. C. Love, A. K. Shalek, Second-strand synthesis-based massively parallel scRNA-Seq reveals cellular states and molecular features of human inflammatory skin pathologies. *Immunity* **53**, 878–894.e7 (2020). doi:10.1016/j.immuni.2020.09.015 Medline

80. T. B. Rojahn, V. Vorstandlechner, T. Krausgruber, W. M. Bauer, N. Alkon, C. Bangert, F. M. Thaler, F. Sadeghyar, N. Fortelny, V. Gernedl, K. Rindler, A. Elbe-Bürger, C. Bock, M. Mildner, P. M. Brunner, Single-cell transcriptomics combined with interstitial fluid proteomics defines cell type-specific immune regulation in atopic dermatitis. *J. Allergy Clin. Immunol.* **146**, 1056–1069 (2020). doi:10.1016/j.jaci.2020.03.041 Medline

81. K. Rindler, T. Krausgruber, F. M. Thaler, N. Alkon, C. Bangert, H. Kurz, N. Fortelny, T. B. Rojahn, C. Jonak, J. Griss, C. Bock, P. M. Brunner, Spontaneously resolved atopic dermatitis shows melanocyte and immune cell activation distinct from healthy control skin. *Front. Immunol.* **12**, 630892 (2021). doi:10.3389/fimmu.2021.630892 Medline

82. C. W. Y. Ha, A. Martin, G. D. Sepich-Poore, B. Shi, Y. Wang, K. Gouin, G. Humphrey, K. Sanders, Y. Ratnayake, K. S. L. Chan, G. Hendrick, J. R. Caldera, C. Arias, J. E. Moskowitz, S. J. Ho Sui, S. Yang, D. Underhill, M. J. Brady, S. Knott, K. Kaihara, M. J. Steinbaugh, H. Li, D. P. B. McGovern, R. Knight, P. Fleshner, S. Devkota, Translocation of viable gut microbiota to mesenteric adipose drives formation of creeping fat in humans. *Cell* **183**, 666–683.e17 (2020). doi:10.1016/j.cell.2020.09.009 Medline

83. J. Tome-Garcia, G. Nudelman, Z. Mussa, E. Caballero, Y. Jiang, K. G. Beaumont, Y.-C. Wang, R. Sebra, S. Akbarian, D. Pinto, E. Zaslavsky, N. M. Tsankova, Cell type-specific isolation and transcriptomic profiling informs glial pathology in human temporal lobe epilepsy. bioRxiv 2020.12.11.421370 [Preprint] (2020); https://doi.org/10.1101/2020.12.11.421370.

84. A. Iyer, M. van Eijk, E. Silva, M. Hatta, W. Faber, J. M. F. G. Aerts, P. K. Das, Increased chitotriosidase activity in serum of leprosy patients: Association with bacillary leprosy. *Clin. Immunol.* **131**, 501–509 (2009). doi:10.1016/j.clim.2009.02.003 Medline

85. A. Deczkowska, A. Weiner, I. Amit, The physiology, pathology, and potential therapeutic applications of the TREM2 signaling pathway. *Cell* **181**, 1207–1217 (2020). doi:10.1016/j.cell.2020.05.003 Medline

86. Y. Kidani, S. J. Bensinger, Liver X receptor and peroxisome proliferator-activated receptor as integrators of lipid homeostasis and immunity. *Immunol. Rev.* **249**, 72–83 (2012). doi:10.1111/j.1600-065X.2012.01153.x Medline

87. J. C. Savage, T. Jay, E. Goduni, C. Quigley, M. M. Mariani, T. Malm, R. M. Ransohoff, B. T. Lamb, G. E. Landreth, Nuclear receptors license phagocytosis by Trem2^{+} myeloid cells in mouse models of Alzheimer's disease. *J. Neurosci.* **35**, 6532–6543 (2015). doi:10.1523/JNEUROSCI.4586-14.2015 Medline

88. M. V. Plikus, X. Wang, S. Sinha, E. Forte, S. M. Thompson, E. L. Herzog, R. R. Driskell, N. Rosenthal, J. Biernaskie, V. Horsley, Fibroblasts: Origins, definitions, and functions in health and disease. *Cell* **184**, 3852–3872 (2021). doi:10.1016/j.cell.2021.06.024 Medline

89. S. Pikkarainen, H. Tokola, R. Kerkelä, H. Ruskoaho, GATA transcription factors in the developing and adult heart. *Cardiovasc. Res.* **63**, 196–207 (2004). doi:10.1016/j.cardiores.2004.03.025 Medline

90. G. M. Dittrich, N. Froese, X. Wang, H. Kroeger, H. Wang, M. Szaroszyk, M. Malek-Mohammadi, J. Cordero, M. Keles, M. Korf-Klingebiel, K. C. Wollert, R. Geffers, M. Mayr, S. J. Conway, G. Dobreva, J. Bauersachs, J. Heineke, Fibroblast GATA-4 and GATA-6 promote myocardial adaptation to pressure overload by enhancing cardiac angiogenesis. *Basic Res. Cardiol.* **116**, 26 (2021). doi:10.1007/s00395-021-00862-y Medline

91. E. Forte, M. Ramialison, H. T. Nim, M. Mara, R. Cohn, S. L. Daigle, S. Boyd, J. Travis Hinson, M. W. Costa, N. A. Rosenthal, M. B. Furtado, Adult fibroblasts retain organ-specific transcriptomic identity. bioRxiv 2021.06.03.446915 [Preprint] (2021); https://doi.org/10.1101/2021.06.03.446915.

92. M. B. Furtado, M. W. Costa, E. A. Pranoto, E. Salimova, A. R. Pinto, N. T. Lam, A. Park, P. Snider, A. Chandran, R. P. Harvey, R. Boyd, S. J. Conway, J. Pearson, D. M. Kaye, N. A. Rosenthal, Cardiogenic genes expressed in cardiac fibroblasts contribute to heart development and repair. *Circ. Res.* **114**, 1422–1434 (2014). doi:10.1161/CIRCRESAHA.114.302530 Medline

93. S. J. Forrester, G. W. Booz, C. D. Sigmund, T. M. Coffman, T. Kawai, V. Rizzo, R. Scalia, S. Eguchi, Angiotensin II signal transduction: An update on mechanisms of physiology and pathophysiology. *Physiol. Rev.* **98**, 1627–1738 (2018). doi:10.1152/physrev.00038.2017 Medline

94. S. K. Powers, A. B. Morton, H. Hyatt, M. J. Hinkley, The renin-angiotensin system and skeletal muscle. *Exerc. Sport Sci. Rev.* **46**, 205–214 (2018). doi:10.1249/JES.0000000000000158 Medline

95. D. Kiyozumi, M. Takeichi, I. Nakano, Y. Sato, T. Fukuda, K. Sekiguchi, Basement membrane assembly of the integrin α8β1 ligand nephronectin requires Fraser syndrome-associated proteins. *J. Cell Biol.* **197**, 677–689 (2012). doi:10.1083/jcb.201203065 Medline

96. D. Kiyozumi, N. Sugimoto, K. Sekiguchi, Breakdown of the reciprocal stabilization of QBRICK/Frem1, Fras1, and Frem2 at the basement membrane provokes Fraser syndrome-like defects. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 11981–11986 (2006). doi:10.1073/pnas.0601011103 Medline

97. R. Brandenberger, A. Schmidt, J. Linton, D. Wang, C. Backus, S. Denda, U. Müller, L. F. Reichardt, Identification and characterization of a novel extracellular matrix protein nephronectin that is associated with integrin α8β1 in the embryonic kidney. *J. Cell Biol.* **154**, 447–458 (2001). doi:10.1083/jcb.200103069 Medline

98. Y. Sato, T. Uemura, K. Morimitsu, R. Sato-Nishiuchi, R. Manabe, J. Takagi, M. Yamada, K. Sekiguchi, Molecular basis of the recognition of nephronectin by integrin α8β1. *J. Biol. Chem.* **284**, 14524–14536 (2009). doi:10.1074/jbc.M900200200 Medline

99. U. Müller, D. Wang, S. Denda, J. J. Meneses, R. A. Pedersen, L. F. Reichardt, Integrin α8β1 is critically important for epithelial-mesenchymal interactions during kidney morphogenesis. *Cell* **88**, 603–613 (1997). doi:10.1016/S0092-8674(00)81903-0 Medline

100. J. T. Benjamin, D. C. Gaston, B. A. Halloran, L. M. Schnapp, R. Zent, L. S. Prince, The role of integrin α8β1 in fetal lung morphogenesis and injury. *Dev. Biol.* **335**, 407–417 (2009). doi:10.1016/j.ydbio.2009.09.021 Medline

101. S. Vrontou, P. Petrou, B. I. Meyer, V. K. Galanopoulos, K. Imai, M. Yanagi, K. Chowdhury, P. J. Scambler, G. Chalepakis, *Fras1* deficiency results in cryptophthalmos, renal agenesis and blebbed phenotype in mice. *Nat. Genet.* **34**, 209–214 (2003). doi:10.1038/ng1168 Medline

102. P. Petrou, E. Pavlakis, Y. Dalezios, V. K. Galanopoulos, G. Chalepakis, Basement membrane distortions impair lung lobation and capillary organization in the mouse model for Fraser syndrome. *J. Biol. Chem.* **280**, 10350–10356 (2005). doi:10.1074/jbc.M412368200 Medline

103. J. M. Linton, G. R. Martin, L. F. Reichardt, The ECM protein nephronectin promotes kidney development via integrin α8β1-mediated stimulation of *Gdnf* expression. *Development* **134**, 2501–2509 (2007). doi:10.1242/dev.005033 Medline

104. L. McGregor, V. Makela, S. M. Darling, S. Vrontou, G. Chalepakis, C. Roberts, N. Smart, P. Rutland, N. Prescott, J. Hopkins, E. Bentley, A. Shaw, E. Roberts, R. Mueller, S. Jadeja, N. Philip, J. Nelson, C. Francannet, A. Perez-Aytes, A. Megarbane, B. Kerr, B. Wainwright, A. S. Woolf, R. M. Winter, P. J. Scambler, Fraser syndrome and mouse blebbed phenotype caused by mutations in *FRAS1*/*Fras1* encoding a putative extracellular matrix protein. *Nat. Genet.* **34**, 203–208 (2003). doi:10.1038/ng1142 Medline

105. Y. Shafeghati, A. Kniepert, G. Vakili, M. Zenker, Fraser syndrome due to homozygosity for a splice site mutation of *FREM2*. *Am. J. Med. Genet. A.* **146A**, 529–531 (2008). doi:10.1002/ajmg.a.32091 Medline

106. A. M. Alazami, R. Shaheen, F. Alzahrani, K. Snape, A. Saggar, B. Brinkmann, P. Bavi, L. I. Al-Gazali, F. S. Alkuraya, *FREM1* mutations cause bifid nose, renal agenesis, and anorectal malformations syndrome. *Am. J. Hum. Genet.* **85**, 414–418 (2009). doi:10.1016/j.ajhg.2009.08.010 Medline

107. T. J. Carney, N. M. Feitosa, C. Sonntag, K. Slanchev, J. Kluger, D. Kiyozumi, J. M. Gebauer, J. Coffin Talbot, C. B. Kimmel, K. Sekiguchi, R. Wagener, H. Schwarz, P. W. Ingham, M. Hammerschmidt, Genetic analysis of fin development in zebrafish identifies Furin and Hemicentin1 as potential novel Fraser syndrome disease genes. *PLOS Genet.* **6**, e1000907 (2010). doi:10.1371/journal.pgen.1000907 Medline

108. M. B. Buechler, R. N. Pradhan, A. T. Krishnamurty, C. Cox, A. K. Calviello, A. W. Wang, Y. A. Yang, L. Tam, R. Caothien, M. Roose-Girma, Z. Modrusan, J. R. Arron, R. Bourgon, S. Müller, S. J. Turley, Cross-tissue organization of the fibroblast lineage. *Nature* **593**, 575–579 (2021). doi:10.1038/s41586-021-03549-5 Medline

109. M. Weinstein, X. Xu, K. Ohyama, C. X. Deng, FGFR-3 and FGFR-4 function cooperatively to direct alveogenesis in the murine lung. *Development* **125**, 3615–3623 (1998). doi:10.1242/dev.125.18.3615 Medline

110. M. Rezvani, J. Wilde, P. Vitt, B. Mailaparambil, R. Grychtol, M. Krueger, A. Heinzmann, Association of a FGFR-4 gene polymorphism with bronchopulmonary dysplasia and neonatal respiratory distress. *Dis. Markers* **35**, 633–640 (2013). doi:10.1155/2013/932356 Medline

111. K. Nonomura, S.-H. Woo, R. B. Chang, A. Gillich, Z. Qiu, A. G. Francisco, S. S. Ranade, S. D. Liberles, A. Patapoutian, Piezo2 senses airway stretch and mediates lung inflation-induced apnoea. *Nature* **541**, 176–181 (2017). doi:10.1038/nature20793 Medline

112. A. Delle Vedove, M. Storbeck, R. Heller, I. Hölker, M. Hebbar, A. Shukla, O. Magnusson, S. Cirak, K. M. Girisha, M. O'Driscoll, B. Loeys, B. Wirth, Biallelic loss of proprioception-related PIEZO2 causes muscular atrophy with perinatal respiratory distress, arthrogryposis, and scoliosis. *Am. J. Hum. Genet.* **99**, 1206–1216 (2016). doi:10.1016/j.ajhg.2016.09.019 Medline

113. I. Y. Kuo, B. E. Ehrlich, Signaling in muscle contraction. *Cold Spring Harb. Perspect. Biol.* **7**, a006023 (2015). doi:10.1101/cshperspect.a006023 Medline

114. D. T. Montoro, A. L. Haber, M. Biton, V. Vinarsky, B. Lin, S. E. Birket, F. Yuan, S. Chen, H. M. Leung, J. Villoria, N. Rogel, G. Burgin, A. M. Tsankov, A. Waghray, M. Slyper, J. Waldman, L. Nguyen, D. Dionne, O. Rozenblatt-Rosen, P. R. Tata, H. Mou, M. Shivaraju, H. Bihler, M. Mense, G. J. Tearney, S. M. Rowe, J. F. Engelhardt, A. Regev, J. Rajagopal, A revised airway epithelial hierarchy includes CFTR-expressing ionocytes. *Nature* **560**, 319–324 (2018). doi:10.1038/s41586-018-0393-7 Medline

115. L. W. Plasschaert, R. Žilionis, R. Choo-Wing, V. Savova, J. Knehr, G. Roma, A. M. Klein, A. B. Jaffe, A single-cell atlas of the airway epithelium reveals the CFTR-rich pulmonary ionocyte. *Nature* **560**, 377–381 (2018). doi:10.1038/s41586-018-0394-6 Medline

116. J. S. Amberger, C. A. Bocchini, A. F. Scott, A. Hamosh, OMIM.org: Leveraging knowledge across phenotype-gene relationships. *Nucleic Acids Res.* **47**, D1038–D1043 (2019). doi:10.1093/nar/gky1151 Medline

117. K. Inaba, K. Mizuno, Sperm dysfunction and ciliopathy. *Reprod. Med. Biol.* **15**, 77–94 (2015). doi:10.1007/s12522-015-0225-5 Medline

118. L. Benarroch, G. Bonne, F. Rivier, D. Hamroun, The 2021 version of the gene table of neuromuscular disorders (nuclear genome). *Neuromuscul. Disord.* **30**, 1008–1048 (2020). doi:10.1016/j.nmd.2020.11.009 Medline

119. R. Csapo, M. Gumpenberger, B. Wessner, Skeletal muscle extracellular matrix – What do we know about its composition, regulation, and physiological roles? A narrative review. *Front. Physiol.* **11**, 253 (2020). doi:10.3389/fphys.2020.00253 Medline

120. K. Ohlendieck, D. Swandulla, Complexity of skeletal muscle degeneration: Multi-systems

pathophysiology and organ crosstalk in dystrophinopathy. *Pflugers Arch.* **473**, 1813–1839 (2021). doi:10.1007/s00424-021-02623-1 Medline

121. J. Verschuuren, E. Strijbos, A. Vincent, Neuromuscular junction disorders. *Handb. Clin. Neurol.* **133**, 447–466 (2016). doi:10.1016/B978-0-444-63432-0.00024-4 Medline

122. B. B. Cummings, J. L. Marshall, T. Tukiainen, M. Lek, S. Donkervoort, A. R. Foley, V. Bolduc, L. B. Waddell, S. A. Sandaradura, G. L. O'Grady, E. Estrella, H. M. Reddy, F. Zhao, B. Weisburd, K. J. Karczewski, A. H. O'Donnell-Luria, D. Birnbaum, A. Sarkozy, Y. Hu, H. Gonorazky, K. Claeys, H. Joshi, A. Bournazos, E. C. Oates, R. Ghaoui, M. R. Davis, N. G. Laing, A. Topf, P. B. Kang, A. H. Beggs, K. N. North, V. Straub, J. J. Dowling, F. Muntoni, N. F. Clarke, S. T. Cooper, C. G. Bönnemann, D. G. MacArthur, Genotype-Tissue Expression Consortium, Improving genetic diagnosis in Mendelian disease with transcriptome sequencing. *Sci. Transl. Med.* **9**, eaal5209 (2017). doi:10.1126/scitranslmed.aal5209 Medline

123. H. D. Gonorazky, S. Naumenko, A. K. Ramani, V. Nelakuditi, P. Mashouri, P. Wang, D. Kao, K. Ohri, S. Viththiyapaskaran, M. A. Tarnopolsky, K. D. Mathews, S. A. Moore, A. N. Osorio, D. Villanova, D. U. Kemaladewi, R. D. Cohn, M. Brudno, J. J. Dowling, Expanding the boundaries of RNA sequencing as a diagnostic tool for rare mendelian disease. *Am. J. Hum. Genet.* **104**, 466–483 (2019). doi:10.1016/j.ajhg.2019.01.012 Medline

124. A. E. Vanhaesebrouck, D. Beeson, The congenital myasthenic syndromes: Expanding genetic and phenotypic spectrums and refining treatment strategies. *Curr. Opin. Neurol.* **32**, 696–703 (2019). doi:10.1097/WCO.0000000000000736 Medline

125. A. M. Maggs, C. Huxley, S. M. Hughes, Nerve-dependent changes in skeletal muscle myosin heavy chain after experimental denervation and cross-reinnervation and in a demyelinating mouse model of Charcot-Marie-Tooth disease type 1A. *Muscle Nerve* **38**, 1572–1584 (2008). doi:10.1002/mus.21106 Medline

126. C. G. Bönnemann, The collagen VI-related myopathies: Muscle meets its matrix. *Nat. Rev. Neurol.* **7**, 379–390 (2011). doi:10.1038/nrneurol.2011.81 Medline

127. A. de Morrée, B. Flix, I. Bagaric, J. Wang, M. van den Boogaard, L. Grand Moursel, R. R. Frants, I. Illa, E. Gallardo, R. Toes, S. M. van der Maarel, Dysferlin regulates cell adhesion in human monocytes. *J. Biol. Chem.* **288**, 14147–14157 (2013). doi:10.1074/jbc.M112.448589 Medline

128. J. Lin, H. Wu, P. T. Tarr, C.-Y. Zhang, Z. Wu, O. Boss, L. F. Michael, P. Puigserver, E. Isotani, E. N. Olson, B. B. Lowell, R. Bassel-Duby, B. M. Spiegelman, Transcriptional co-activator PGC-1α drives the formation of slow-twitch muscle fibres. *Nature* **418**, 797–801 (2002). doi:10.1038/nature00904 Medline

129. H. Tajsharghi, S. Hammans, C. Lindberg, A. Lossos, N. F. Clarke, I. Mazanti, L. B. Waddell, Y. Fellig, N. Foulds, H. Katifi, R. Webster, O. Raheem, B. Udd, Z. Argov, A. Oldfors, Recessive myosin myopathy with external ophthalmoplegia associated with *MYH2* mutations. *Eur. J. Hum. Genet.* **22**, 801–808 (2014). doi:10.1038/ejhg.2013.250 Medline

130. V. Pisani, M. B. Panico, C. Terracciano, E. Bonifazi, G. Meola, G. Novelli, G. Bernardi, C. Angelini, R. Massa, Preferential central nucleation of type 2 myofibers is an invariable

feature of myotonic dystrophy type 2. *Muscle Nerve* **38**, 1405–1411 (2008). doi:10.1002/mus.21122 Medline

131. A. Vihola, G. Bassez, G. Meola, S. Zhang, H. Haapasalo, A. Paetau, E. Mancinelli, A. Rouche, J. Y. Hogrel, P. Laforêt, T. Maisonobe, J. F. Pellissier, R. Krahe, B. Eymard, B. Udd, Histopathological differences of myotonic dystrophy type 1 (DM1) and PROMM/DM2. *Neurology* **60**, 1854–1857 (2003). doi:10.1212/01.WNL.0000065898.61358.09 Medline

132. L. S. Krivickas, T. Ansved, D. Suh, W. R. Frontera, Contractile properties of single muscle fibers in myotonic dystrophy. *Muscle Nerve* **23**, 529–537 (2000). doi:10.1002/(SICI)1097-4598(200004)23:4 < 529:AID-MUS11>3.0.CO;2-Y Medline

133. N. F. Clarke, Congenital fiber-type disproportion. *Semin. Pediatr. Neurol.* **18**, 264–271 (2011). doi:10.1016/j.spen.2011.10.008 Medline

134. J. Strakova, F. Kamdar, D. Kulhanek, M. Razzoli, D. J. Garry, J. M. Ervasti, A. Bartolomucci, D. Townsend, Integrative effects of dystrophin loss on metabolic function of the mdx mouse. *Sci. Rep.* **8**, 13624 (2018). doi:10.1038/s41598-018-31753-3 Medline

135. D. Duan, N. Goemans, S. Takeda, E. Mercuri, A. Aartsma-Rus, Duchenne muscular dystrophy. *Nat. Rev. Dis. Primers* **7**, 13 (2021). doi:10.1038/s41572-021-00248-3 Medline

136. M. Sander, B. Chavoshan, S. A. Harris, S. T. Iannaccone, J. T. Stull, G. D. Thomas, R. G. Victor, Functional muscle ischemia in neuronal nitric oxide synthase-deficient skeletal muscle of children with Duchenne muscular dystrophy. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 13818–13823 (2000). doi:10.1073/pnas.250379497 Medline

137. P. A. C. 't Hoen, E. J. de Meijer, J. M. Boer, R. H. A. M. Vossen, R. Turk, R. G. H. J. Maatman, K. E. Davies, G.-J. B. van Ommen, J. C. T. van Deutekom, J. T. den Dunnen, Generation and characterization of transgenic mice with the full-length human DMD gene. *J. Biol. Chem.* **283**, 5899–5907 (2008). doi:10.1074/jbc.M709410200 Medline

138. M. W. Hogarth, A. Defour, C. Lazarski, E. Gallardo, J. Diaz Manera, T. A. Partridge, K. Nagaraju, J. K. Jaiswal, Fibroadipogenic progenitors are responsible for muscle loss in limb girdle muscular dystrophy 2B. *Nat. Commun.* **10**, 2430 (2019). doi:10.1038/s41467-019-10438-z Medline

139. D. R. Lemos, F. Babaeijandaghi, M. Low, C.-K. Chang, S. T. Lee, D. Fiore, R.-H. Zhang, A. Natarajan, S. A. Nedospasov, F. M. V. Rossi, Nilotinib reduces muscle fibrosis in chronic muscle injury by promoting TNF-mediated apoptosis of fibro/adipogenic progenitors. *Nat. Med.* **21**, 786–794 (2015). doi:10.1038/nm.3869 Medline

140. GTEx Consortium, Genetic effects on gene expression across human tissues. *Nature* **550**, 204–213 (2017). doi:10.1038/nature24277 Medline

141. X. Unamuno, J. Gómez-Ambrosi, B. Ramírez, A. Rodríguez, S. Becerril, V. Valentí, R. Moncada, C. Silva, J. Salvador, G. Frühbeck, V. Catalán, Dermatopontin, a novel adipokine promoting adipose tissue extracellular matrix remodelling and inflammation in obesity. *J. Clin. Med.* **9**, 1069 (2020). doi:10.3390/jcm9041069 Medline

142. S.-H. Heo, J.-Y. Cho, ELK3 suppresses angiogenesis by inhibiting the transcriptional activity of ETS-1 on MT1-MMP. *Int. J. Biol. Sci.* **10**, 438–447 (2014).

doi:10.7150/ijbs.8095 Medline

143. P. Podkalicka, O. Mucha, J. Dulak, A. Loboda, Targeting angiogenesis in Duchenne muscular dystrophy. *Cell. Mol. Life Sci.* **76**, 1507–1528 (2019). doi:10.1007/s00018-019-03006-7 Medline

144. M. Efremova, M. Vento-Tormo, S. A. Teichmann, R. Vento-Tormo, CellPhoneDB: Inferring cell-cell communication from combined expression of multi-subunit ligand-receptor complexes. *Nat. Protoc.* **15**, 1484–1506 (2020). doi:10.1038/s41596-020-0292-x Medline

145. G. Narkis, R. Ofir, E. Manor, D. Landau, K. Elbedour, O. S. Birk, Lethal congenital contractural syndrome type 2 (LCCS2) is caused by a mutation in *ERBB3* (*Her3*), a modulator of the phosphatidylinositol-3-kinase/Akt pathway. *Am. J. Hum. Genet.* **81**, 589–595 (2007). doi:10.1086/520770 Medline

146. O. Ibraghimov-Beskrovnaya, J. M. Ervasti, C. J. Leveille, C. A. Slaughter, S. W. Sernett, K. P. Campbell, Primary structure of dystrophin-associated glycoproteins linking dystrophin to the extracellular matrix. *Nature* **355**, 696–702 (1992). doi:10.1038/355696a0 Medline

147. A. R. Gillies, R. L. Lieber, Structure and function of the skeletal muscle extracellular matrix. *Muscle Nerve* **44**, 318–331 (2011). doi:10.1002/mus.22094 Medline

148. S.-Y. Dai, R. Nakagawa, A. Itoh, H. Murakami, Y. Kashio, H. Abe, S. Katoh, K. Kontani, M. Kihara, S.-L. Zhang, T. Hata, T. Nakamura, A. Yamauchi, M. Hirashima, Galectin-9 induces maturation of human monocyte-derived dendritic cells. *J. Immunol.* **175**, 2974–2981 (2005). doi:10.4049/jimmunol.175.5.2974 Medline

149. H. K. Finucane, B. Bulik-Sullivan, A. Gusev, G. Trynka, Y. Reshef, P.-R. Loh, V. Anttila, H. Xu, C. Zang, K. Farh, S. Ripke, F. R. Day, S. Purcell, E. Stahl, S. Lindstrom, J. R. B. Perry, Y. Okada, S. Raychaudhuri, M. J. Daly, N. Patterson, B. M. Neale, A. L. Price, ReproGen Consortium, Schizophrenia Working Group of the Psychiatric Genomics Consortium, RACI Consortium, Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015). doi:10.1038/ng.3404 Medline

150. E. R. Gamazon, A. V. Segrè, M. van de Bunt, X. Wen, H. S. Xi, F. Hormozdiari, H. Ongen, A. Konkashbaev, E. M. Derks, F. Aguet, J. Quan, D. L. Nicolae, E. Eskin, M. Kellis, G. Getz, M. I. McCarthy, E. T. Dermitzakis, N. J. Cox, K. G. Ardlie, GTEx Consortium, Using an atlas of gene regulation across 44 human tissues to inform complex disease- and trait-associated variation. *Nat. Genet.* **50**, 956–967 (2018). doi:10.1038/s41588-018-0154-4 Medline

151. A. N. Barbeira, R. Bonazzola, E. R. Gamazon, Y. Liang, Y. Park, S. Kim-Hellmuth, G. Wang, Z. Jiang, D. Zhou, F. Hormozdiari, B. Liu, A. Rao, A. R. Hamel, M. D. Pividori, F. Aguet, L. Bastarache, D. M. Jordan, M. Verbanck, R. Do, M. Stephens, K. Ardlie, M. McCarthy, S. B. Montgomery, A. V. Segrè, C. D. Brown, T. Lappalainen, X. Wen, H. K. Im, GTEx GWAS Working Group, GTEx Consortium, Exploiting the GTEx resources to decipher the mechanisms at GWAS loci. *Genome Biol.* **22**, 49 (2021). doi:10.1186/s13059-020-02252-4 Medline

152. J. M. Rouhana, J. Wang, G. Eraslan, S. Anand, A. R. Hamel, B. Cole, A. Regev, F. Aguet, K. G. Ardlie, A. V. Segre, ECLIPSER: identifying causal cell types and genes for

complex traits through single cell enrichment of e/sQTL-mapped genes in GWAS loci. bioRxiv 2021.11.24.469720 [Preprint] (2021); https://doi.org/10.1101/2021.11.24.469720.

153. M. Ghoussaini, E. Mountjoy, M. Carmona, G. Peat, E. M. Schmidt, A. Hercules, L. Fumis, A. Miranda, D. Carvalho-Silva, A. Buniello, T. Burdett, J. Hayhurst, J. Baker, J. Ferrer, A. Gonzalez-Uriarte, S. Jupp, M. A. Karim, G. Koscielny, S. Machlitt-Northen, C. Malangone, Z. M. Pendlington, P. Roncaglia, D. Suveges, D. Wright, O. Vrousgou, E. Papa, H. Parkinson, J. A. L. MacArthur, J. A. Todd, J. C. Barrett, J. Schwartzentruber, D. G. Hulcoop, D. Ochoa, E. M. McDonagh, I. Dunham, Open Targets Genetics: Systematic identification of trait-associated genes using large-scale genetics and functional genomics. *Nucleic Acids Res.* **49**, D1311–D1320 (2021). doi:10.1093/nar/gkaa840 Medline

154. E. Mountjoy, E. M. Schmidt, M. Carmona, J. Schwartzentruber, G. Peat, A. Miranda, L. Fumis, J. Hayhurst, A. Buniello, M. A. Karim, D. Wright, A. Hercules, E. Papa, E. B. Fauman, J. C. Barrett, J. A. Todd, D. Ochoa, I. Dunham, M. Ghoussaini, An open approach to systematically prioritize causal variants and genes at all published human GWAS trait-associated loci. *Nat. Genet.* **53**, 1527–1533 (2021). doi:10.1038/s41588-021-00945-5 Medline

155. E. Brakenhielm, K. Alitalo, Cardiac lymphatics in health and disease. *Nat. Rev. Cardiol.* **16**, 56–68 (2019). doi:10.1038/s41569-018-0087-8 Medline

156. F. Paneni, J. A. Beckman, M. A. Creager, F. Cosentino, Diabetes and vascular disease: pathophysiology, clinical consequences, and medical therapy: Part I. *Eur. Heart J.* **34**, 2436–2443 (2013). doi:10.1093/eurheartj/eht149 Medline

157. T. R. Einarson, A. Acs, C. Ludwig, U. H. Panton, Prevalence of cardiovascular disease in type 2 diabetes: A systematic literature review of scientific evidence from across the world in 2007–2017. *Cardiovasc. Diabetol.* **17**, 83 (2018). doi:10.1186/s12933-018-0728-6 Medline

158. C. Kothari, C. Diorio, F. Durocher, The importance of breast adipose tissue in breast cancer. *Int. J. Mol. Sci.* **21**, 5760 (2020). doi:10.3390/ijms21165760 Medline

159. A. B. Kay, Allergy and allergic diseases. First of two parts. *N. Engl. J. Med.* **344**, 30–37 (2001). doi:10.1056/NEJM200101043440106 Medline

160. S. J. Galli, M. Tsai, A. M. Piliponsky, The development of allergic inflammation. *Nature* **454**, 445–454 (2008). doi:10.1038/nature07204 Medline

161. H. Yilmaz, M. Cakmak, B. Ceydilek, C. Demir, A. Aktas, Role of interlekin-35 as a biomarker in patients with newly diagnosed Hashimoto's thyroiditis. *Endocr. Regul.* **50**, 55–61 (2016). doi:10.1515/enr-2016-0009 Medline

162. Y. Li, Y. Wang, Y. Liu, Y. Wang, X. Zuo, Y. Li, X. Lu, The possible role of the novel cytokines IL-35 and IL-37 in inflammatory bowel disease. *Mediators Inflamm.* **2014**, 136329 (2014). doi:10.1155/2014/136329 Medline

163. M.-H. Saeed, K. Kurosh, A. Zahra, D. M. Hossein, R. Davood, M. R. Ataollahi, Decreased serum levels of IL-27 and IL-35 in patients with Graves' disease. *Arch. Endocrinol. Metab.* **64**, 521–527 (2021). doi:10.20945/2359-3997000000227 Medline

164. A. Pinggera, L. Mackenroth, A. Rump, J. Schallner, F. Beleggia, B. Wollnik, J. Striessnig, New gain-of-function mutation shows CACNA1D as recurrently mutated gene in autism spectrum disorders and epilepsy. *Hum. Mol. Genet.* **26**, 2923–2932 (2017). doi:10.1093/hmg/ddx175 Medline

165. R. Mysore, G. Liebisch, Y. Zhou, V. M. Olkkonen, P. A. Nidhina Haridas, Angiopoietin-like 8 (Angptl8) controls adipocyte lipolysis and phospholipid composition. *Chem. Phys. Lipids* **207**, 246–252 (2017). doi:10.1016/j.chemphyslip.2017.05.002 Medline

166. A. Yang, E. P. Mottillo, Adipocyte lipolysis: From molecular mechanisms of regulation to disease and therapeutics. *Biochem. J.* **477**, 985–1008 (2020). doi:10.1042/BCJ20190468 Medline

167. J. Xu, C. Falconer, Q. Nguyen, J. Crawford, B. D. McKinnon, S. Mortlock, A. Senabouth, S. Andersen, H. S. Chiu, L. Jiang, N. J. Palpant, J. Yang, M. D. Mueller, A. W. Hewitt, A. Pébay, G. W. Montgomery, J. E. Powell, L. J. M. Coin, Genotype-free demultiplexing of pooled single-cell RNA-seq. *Genome Biol.* **20**, 290 (2019). doi:10.1186/s13059-019-1852-7 Medline

168. H. Heaton, A. M. Talman, A. Knights, M. Imaz, D. J. Gaffney, R. Durbin, M. Hemberg, M. K. N. Lawniczak, Souporcell: Robust clustering of single-cell RNA-seq data by genotype without reference genotypes. *Nat. Methods* **17**, 615–620 (2020). doi:10.1038/s41592-020-0820-1 Medline

169. J. Z. Kechagia, J. Ivaska, P. Roca-Cusachs, Integrins as biomechanical sensors of the microenvironment. *Nat. Rev. Mol. Cell Biol.* **20**, 457–473 (2019). doi:10.1038/s41580-019-0134-2 Medline

170. B. B. Lake, S. Chen, M. Hoshi, N. Plongthongkum, D. Salamon, A. Knoten, A. Vijayan, R. Venkatesh, E. H. Kim, D. Gao, J. Gaut, K. Zhang, S. Jain, A single-nucleus RNA-sequencing pipeline to decipher the molecular anatomy and pathophysiology of human kidneys. *Nat. Commun.* **10**, 2832 (2019). doi:10.1038/s41467-019-10861-2 Medline

171. B. B. Lake, R. Menon, S. Winfree, Q. Hu, R. M. Ferreira, K. Kalhor, D. Barwinska, E. A. Otto, M. Ferkowicz, D. Diep, N. Plongthongkum, A. Knoten, S. Urata, A. S. Naik, S. Eddy, B. Zhang, Y. Wu, D. Salamon, J. C. Williams, X. Wang, K. S. Balderrama, P. Hoover, E. Murray, A. Vijayan, F. Chen, S. S. Waikar, S. Rosas, F. P. Wilson, P. M. Palevsky, K. Kiryluk, J. R. Sedor, R. D. Toto, C. Parikh, E. H. Kim, E. Z. Macosko, P. V. Kharchenko, J. P. Gaut, J. B. Hodgin, M. T. Eadon, P. C. Dagher, T. M. El-Achkar, K. Zhang, M. Kretzler, S. Jain, An atlas of healthy and injured cell states and niches in the human kidney. bioRxiv 2021.07.28.454201 [Preprint] (2021); https://doi.org/10.1101/2021.07.28.454201.

172. I. Marek, K. F. Hilgers, W. Rascher, J. Woelfle, A. Hartner, A role for the alpha-8 integrin chain (itga8) in glomerular homeostasis of the kidney. *Mol. Cell Pediatr.* **7**, 13 (2020). doi:10.1186/s40348-020-00105-5 Medline

173. H. Usui, M. Shibayama, N. Ohbayashi, M. Konishi, S. Takada, N. Itoh, Fgf18 is required for embryonic lung alveolar development. *Biochem. Biophys. Res. Commun.* **322**, 887–892 (2004). doi:10.1016/j.bbrc.2004.07.198 Medline

174. J. Zammouri, C. Vatier, E. Capel, M. Auclair, C. Storey-London, E. Bismuth, H. Mosbah, B. Donadille, S. Janmaat, B. Fève, I. Jéru, C. Vigouroux, Molecular and cellular bases of

lipodystrophy syndromes. *Front. Endocrinol.* **12**, 803189 (2022). doi:10.3389/fendo.2021.803189 Medline

175. G. Kelsey, O. Stegle, W. Reik, Single-cell epigenomics: Recording the past and predicting the future. *Science* **358**, 69–75 (2017). doi:10.1126/science.aan6826 Medline

176. S. Ma, B. Zhang, L. M. LaFave, A. S. Earl, Z. Chiang, Y. Hu, J. Ding, A. Brack, V. K. Kartha, T. Tay, T. Law, C. Lareau, Y.-C. Hsu, A. Regev, J. D. Buenrostro, Chromatin potential identified by shared single-cell profiling of RNA and chromatin. *Cell* **183**, 1103–1116.e20 (2020). doi:10.1016/j.cell.2020.09.056 Medline

177. E. P. Mimitou, C. A. Lareau, K. Y. Chen, A. L. Zorzetto-Fernandes, Y. Hao, Y. Takeshima, W. Luo, T.-S. Huang, B. Z. Yeung, E. Papalexi, P. I. Thakore, T. Kibayashi, J. B. Wing, M. Hata, R. Satija, K. L. Nazor, S. Sakaguchi, L. S. Ludwig, V. G. Sankaran, A. Regev, P. Smibert, Scalable, multimodal profiling of chromatin accessibility, gene expression and protein levels in single cells. *Nat. Biotechnol.* **39**, 1246–1258 (2021). doi:10.1038/s41587-021-00927-2 Medline

178. E. Fiskin, C. A. Lareau, L. S. Ludwig, G. Eraslan, F. Liu, A. M. Ring, R. J. Xavier, A. Regev, Single-cell profiling of proteins and chromatin accessibility using PHAGE-ATAC. *Nat. Biotechnol.* **40**, 374–381 (2022). doi:10.1038/s41587-021-01065-5 Medline

179. G. Eraslan, The code used in Single-nucleus cross-tissue molecular reference maps toward understanding disease gene function. Zenodo (2022); https://doi.org/10.5281/zenodo.5775379.

180. L. J. Carithers, K. Ardlie, M. Barcus, P. A. Branton, A. Britton, S. A. Buia, C. C. Compton, D. S. DeLuca, J. Peter-Demchok, E. T. Gelfand, P. Guan, G. E. Korzeniewski, N. C. Lockhart, C. A. Rabiner, A. K. Rao, K. L. Robinson, N. V. Roche, S. J. Sawyer, A. V. Segrè, C. E. Shive, A. M. Smith, L. H. Sobin, A. H. Undale, K. M. Valentino, J. Vaught, T. R. Young, H. M. Moore, GTEx Consortium, A novel approach to high-quality postmortem tissue procurement: The GTEx project. *Biopreserv. Biobank.* **13**, 311–319 (2015). doi:10.1089/bio.2015.0032 Medline

181. M. D. Young, S. Behjati, SoupX removes ambient RNA contamination from droplet-based single-cell RNA sequencing data. *Gigascience* **9**, giaa151 (2020). doi:10.1093/gigascience/giaa151 Medline

182. R. T. Q. Chen, X. Li, R. Grosse, D. Duvenaud, Isolating sources of disentanglement in variational autoencoders. arXiv:1802.04942 [cs.LG] (2018).

183. 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

184. I. Korsunsky, N. Millard, J. Fan, K. Slowikowski, F. Zhang, K. Wei, Y. Baglaenko, M. Brenner, P.-R. Loh, S. Raychaudhuri, Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* **16**, 1289–1296 (2019). doi:10.1038/s41592-019-0619-0 Medline

185. L. Haghverdi, A. T. L. Lun, M. D. Morgan, J. C. Marioni, Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. *Nat. Biotechnol.* **36**, 421–427 (2018). doi:10.1038/nbt.4091 Medline

186. K. Polański, M. D. Young, Z. Miao, K. B. Meyer, S. A. Teichmann, J.-E. Park, BBKNN:

Fast batch alignment of single cell transcriptomes. *Bioinformatics* **36**, 964–965 (2020). Medline

187. M. D. Luecken, M. Büttner, K. Chaichoompu, A. Danese, M. Interlandi, M. F. Mueller, D. C. Strobl, L. Zappia, M. Dugas, M. Colomé-Tatché, F. J. Theis, Benchmarking atlas-level data integration in single-cell genomics. *Nat. Methods* **19**, 41–50 (2022). doi:10.1038/s41592-021-01336-8 Medline

188. M. Büttner, Z. Miao, F. A. Wolf, S. A. Teichmann, F. J. Theis, A test metric for assessing single-cell RNA-seq batch correction. *Nat. Methods* **16**, 43–49 (2019). doi:10.1038/s41592-018-0254-1 Medline

189. J. T. Leek, W. E. Johnson, H. S. Parker, A. E. Jaffe, J. D. Storey, The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* **28**, 882–883 (2012). doi:10.1093/bioinformatics/bts034 Medline

190. T. Derrien, J. Estellé, S. Marco Sola, D. G. Knowles, E. Raineri, R. Guigó, P. Ribeca, Fast computation and applications of genome mappability. *PLOS ONE* **7**, e30377 (2012). doi:10.1371/journal.pone.0030377 Medline

191. J. Bergstra, D. Yamins, D. Cox, "Making a science of model search: Hyperparameter optimization in hundreds of dimensions for vision architectures" in *Proceedings of the 30th International Conference on Machine Learning* (JMLR, 2013), pp. 115–123.

192. F. A. Wolf, F. Hamey, M. Plass, J. Solana, J. S. Dahlin, B. Göttgens, N. Rajewsky, L. Simon, F. J. Theis, PAGA: Graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. *Genome Biol.* **20**, 59 (2019). doi:10.1186/s13059-019-1663-x Medline

193. L. Garcia-Alonso, C. H. Holland, M. M. Ibrahim, D. Turei, J. Saez-Rodriguez, Benchmark and integration of resources for the estimation of human transcription factor activities. *Genome Res.* **29**, 1363–1375 (2019). doi:10.1101/gr.240663.118 Medline

194. C. H. Holland, J. Tanevski, J. Perales-Patón, J. Gleixner, M. P. Kumar, E. Mereu, B. A. Joughin, O. Stegle, D. A. Lauffenburger, H. Heyn, B. Szalai, J. Saez-Rodriguez, Robustness and applicability of transcription factor and pathway analysis tools on single-cell RNA-seq data. *Genome Biol.* **21**, 36 (2020). doi:10.1186/s13059-020-1949-z Medline

195. M. Gerlach, T. P. Peixoto, E. G. Altmann, A network approach to topic models. *Sci. Adv.* **4**, eaaq1360 (2018). doi:10.1126/sciadv.aaq1360 Medline

196. X. Wen, R. Pique-Regi, F. Luca, Integrating molecular QTL data into genome-wide genetic association analysis: Probabilistic assessment of enrichment and colocalization. *PLOS Genet.* **13**, e1006646 (2017). doi:10.1371/journal.pgen.1006646 Medline

197. X. Zheng-Bradley, I. Streeter, S. Fairley, D. Richardson, L. Clarke, P. Flicek, 1000 Genomes Project Consortium, Alignment of 1000 Genomes Project reads to reference assembly GRCh38. *Gigascience* **6**, 1–8 (2017). Medline

198. B. Li, J. Gould, Y. Yang, S. Sarkizova, M. Tabaka, O. Ashenberg, Y. Rosen, M. Slyper, M. S. Kowalczyk, A.-C. Villani, T. Tickle, N. Hacohen, O. Rozenblatt-Rosen, A. Regev, Cumulus provides cloud-based data analysis for large-scale single-cell and single-nucleus RNA-seq. *Nat. Methods* **17**, 793–798 (2020). doi:10.1038/s41592-020-0905-x Medline