University of Groningen

Transcriptional and Cellular Diversity of the Human Heart

Tucker, Nathan R.; Chaffin, Mark; Fleming, Stephen J.; Hall, Amelia W.; Parsons, Victoria A.; Bedi, Kenneth C.; Akkad, Amer-Denis; Herndon, Caroline N.; Arduini, Alessandro; Papangeli, Irinna

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Link to publication in University of Groningen/UMCG research database

Circulation

ORIGINAL RESEARCH ARTICLE

Transcriptional and Cellular Diversity of the Human Heart

BACKGROUND: The human heart requires a complex ensemble of specialized cell types to perform its essential function. A greater knowledge of the intricate cellular milieu of the heart is critical to increase our understanding of cardiac homeostasis and pathology. As recent advances in low-input RNA sequencing have allowed definitions of cellular transcriptomes at single-cell resolution at scale, we have applied these approaches to assess the cellular and transcriptional diversity of the nonfailing human heart.

METHODS: Microfluidic encapsulation and barcoding was used to perform single nuclear RNA sequencing with samples from 7 human donors, selected for their absence of overt cardiac disease. Individual nuclear transcriptomes were then clustered based on transcriptional profiles of highly variable genes. These clusters were used as the basis for between-chamber and between-sex differential gene expression analyses and intersection with genetic and pharmacologic data.

RESULTS: We sequenced the transcriptomes of 287 269 single cardiac nuclei, revealing 9 major cell types and 20 subclusters of cell types within the human heart. Cellular subclasses include 2 distinct groups of resident macrophages, 4 endothelial subtypes, and 2 fibroblast subsets. Comparisons of cellular transcriptomes by cardiac chamber or sex reveal diversity not only in cardiomyocyte transcriptional programs but also in subtypes involved in extracellular matrix remodeling and vascularization. Using genetic association data, we identified strong enrichment for the role of cell subtypes in cardiac traits and diseases. Intersection of our data set with genes on cardiac clinical testing panels and the druggable genome reveals striking patterns of cellular specificity.

CONCLUSIONS: Using large-scale single nuclei RNA sequencing, we defined the transcriptional and cellular diversity in the normal human heart. Our identification of discrete cell subtypes and differentially expressed genes within the heart will ultimately facilitate the development of new therapeutics for cardiovascular diseases.

Nathan R. Tucker, PhD*
Mark Chaffin, MS*
Stephen J. Fleming, PhD
Amelia W. Hall, PhD
Victoria A. Parsons, BS
Kenneth C. Bedi, Jr, BS
Amer-Denis Akkad, PhD
Caroline N. Herndon, PhD
Alessandro Arduini, PhD
Irinna Papangeli, PhD
Carolina Roselli, MS
François Aguet, PhD
Seung Hoan Choi, PhD
Kristin G. Ardlie, PhD
Mehrtash Babadi, PhD
Kenneth B. Margulies, MD
Christian M. Stegmann, PhD
Patrick T. Ellinor[iD], MD, PhD

Clinical Perspective

What Is New?

- Recent advances in single-cell sequencing have provided an unprecedented view of the diversity of the cell subtypes in health and disease.
- We performed large-scale single-nucleus RNA sequencing to define the cellular and transcriptional diversity in the 4 chambers of the normal human heart.
- Using data from >280 000 single nuclei, we identified 9 major and >20 subtypes of cells within the human heart.

What Are the Clinical Implications?

- Combining genetic and single-nucleus sequencing data identified the most relevant cell types for multiple common cardiovascular diseases.
- Identification of discrete cell subtypes and differentially expressed genes in the human heart will facilitate drug discovery efforts by enabling cell type–specific models of therapeutic targets.

The heart is an organ that acts without rest, ceaselessly beating >2 billion times in the average human lifetime. Given the central function of the heart as a pump, it is understandable that much of the cardiac research focus has been centered on the cell subtype most responsible for contractile functionality: the cardiomyocyte. However, cardiomyocytes do not function in isolation, instead contracting as part of a complex ensemble of specialized cell types including those responsible for tissue perfusion, remodeling of the interstitial space, and autonomic regulation. A greater understanding of the complex cellular milieu of the heart is critical to advance our understanding of cardiac homeostasis and pathology.

Analysis of transcription of RNA species, a highly dynamic process, is one method for defining cell types and states. To date, transcriptional analyses of the human heart have largely been performed in bulk tissue RNA sequencing studies. These studies have yielded important insight into regional and pathologic differences in tissue-level expression, but they are unable to resolve the cell types from which any differential expression occurs. Recent advances in single-cell RNA sequencing, particularly technologies that are centered on microfluidic encapsulation and cellular barcoding,[1,2] have made deconvolution of these expression profiles technologically feasible. Efforts are underway to define the cellular diversity in all organ systems. The Human Cell Atlas[3] and Human BioMolecular Atlas Program (https://commonfund.nih.gov/hubmap) are of particular note in humans, and the Tabula Muris project[4] has provided insight into the murine cell subtype transcriptome.

Because of challenges with tissue availability and cellular isolation, there have been relatively few studies of the cardiac system. Analyses of heart tissue from humans[5,6] and model systems[4,7] have been published recently but are limited in scope. A comprehensive analysis of cell subtype expression profiles from the nonfailing human heart has yet to be performed. The transcriptional map of the nonfailing human heart at single-cell resolution, together with an understanding of its normal interindividual variability, serves as a baseline against which one can obtain equally high-resolution and quantitative maps of cardiac pathologies.

We performed single- nucleus RNA sequencing (snRNA-seq) on 287 269 nuclei derived from the 4 chambers of the normal human heart. We identified 9 major cell types and >20 cell subtypes. We observed marked differences in cell subtype transcription by chamber, laterality, and sex. We then intersected the snRNA-seq data with the results from genome-wide association studies to prioritize cell subtypes for cardiovascular disease risk and with the druggable genome to facilitate the identification of novel therapeutic targets for cardiovascular diseases. Our data provide a methodologic framework and large-scale resource available to the broader scientific community.

METHODS

Data Availability

Raw sequence data will be made available through dbGaP (the database of Genotypes and Phenotypes) accession number phs001539.v1.p1. Processed data with interactivity for gene search functions will be available through the Broad Institute's Single Cell Portal (https://singlecell.broadinstitute.org/single_cell/study/SCP498/transcriptional-and-cellular-diversity-of-the-human-heart) under study ID SCP498.

Human Tissue Samples

Adult human myocardial samples of European ancestry were collected from deceased organ donors by MAGNet (the Myocardial Applied Genetics Network; www.med.upenn.edu/magnet). For all donors, clinical examination and medical history displayed no indications of structural heart disease. Using clinical transplantation methods, all hearts were arrested in situ with at least 1 L of ice-cold crystalloid cardioplegia solution, as reported previously.[8,9] Hearts were transported to the laboratory in ice-cold cardioplegia solution until cryopreservation (always <4 hours). Written informed consent for research use of donated tissue was obtained from next of kin in all cases. Research use of tissues was approved by the institutional review boards at the Gift-of-Life Donor Program, the University of Pennsylvania, Massachusetts General Hospital, and the Broad Institute.

Single-Nucleus RNA Sequencing

Single-nucleus suspensions were generated by a series of cellular membrane lysis, differential centrifugation, and filtration steps. Isolated nuclei were loaded into the 10× Genomics

microfluidic platform (single cell 3′ solution, v2) for an esti-mated recovery of 5000 cells per device. Processing of libraries was performed according to the manufacturer's instructions with a few modifications. After sequencing, preprocessing steps were performed using CellRanger 2.1.1 followed by postprocessing using CellBender v0.1 (https://github.com/broadinstitute/CellBender) scanpy 1.4[10] and Seurat 2.3.4.[11] Calculation of exon/intron location of reads was performed using scR-Invex (https://github.com/broadinstitute/scrinvex). Full methodologic details for reagents used, nuclear isolation, library construction, quality control, and analysis can be found in the Expanded Methods and Table I in the Data Supplement.

Statistical Methods for Differential Gene Expression Analysis

Between-chamber and between-sex differential gene expres-sion analyses were performed for the top 5 most abundant cell types in the aggregated 4-chamber map. This included cardiomyocytes (clusters 3, 4, 6, and 15), fibroblasts (clusters 1, 2, and 14), endothelial cells (clusters 9 and 10), pericytes (cluster 7), and macrophages (cluster 8). Additional subclus-ters within the cardiomyocytes and endothelial cells were removed if they had an enriched proportion of spliced tran-scripts, often accompanied by mitochondrial gene markers.

For each cell type, we fit a generalized linear mixed model to compare gene expression in different groups using the R package lme4.[12] For a given gene in a given cell type, we first assumed that the unique molecular identifier (UMI) counts in cell i from experiment j of individual k, denoted y_{ijk}, fol-lowed a negative binomial distribution,[13]

$y_{ijk} \approx NB(\lambda_{ijk}, \theta)$, where θ represents inverse overdispersion.[13] In many cases, θ approached infinity and we therefore reverted to a Poisson assumption, $y_{ijk} \approx Poisson(\lambda_{ijk})$, if $\theta > 10\,000$ for either the null or the full model. We constructed 2 generalized linear mixed models for $\log(\lambda_{ijk})$, specifically:

$$Null : \log(\lambda_{ijk}) = \beta_0 + b_k + s_{jk} + \varepsilon_{ijk} + \log(UMI_i)$$

$$Full : \log(\lambda_{ijk}) = \beta_0 + \beta_1 group + b_k + s_{jk} + \varepsilon_{ijk} + \log(UMI_i)$$

where β_0 is a global mean UMI, β_1 is the fixed effect for the group of comparison (chamber or sex), $\log(UMI_i)$ is an offset of the total UMI in cell i, and b_k, s_{jk}, and ε_{ijk} are random effects for biological sample, experiment, and residual error normally distributed with mean 0 and variances σ_b^2, σ_s^2, and σ_ε^2, respectively. Any genes where $\theta < 0.10$ from either the null or full negative binomial model were removed as very high overdispersion created problems in model convergence.

In lme4 notation, the negative binomial mixed model was fit as follows:

Null Model : glmer.nb($Y \sim 1 + (1 | I / S) + $ offset(log(UMI)))

Full Model : glmer.nb($Y \sim 1 + Group + (1 | I / S) + $ offset(log(UMI)))

The Poisson model was fit as follows:

Null Model : glmer($Y \sim 1 + (1 | I / S) + $ offset(log(UMI)),
family = 'poisson')

Full Model : glmer($Y \sim 1 + Group + (1 | I / S) + $ offset(log(UMI)),
family = 'poisson')

where Y represents UMI counts, I is a random effect of bio-logical individual, S is a random effect of experiment, UMI are the total UMI counts in the given cell, and $Group$ represents the fixed effect comparison of interest.

Significance was tested using a likelihood ratio test com-paring the full model to the null model.

Only genes expressed in at least 1% of either group in the given comparison were tested. To avoid capturing genes only present in the ambient background RNA or genes whose expression comes from cluster misclassification, only genes with a standardized positive predictive value (PPV50; Expanded Methods in the Data Supplement) >0.55 or PPV50 >0.50 for the cluster of interest were included for testing chamber comparisons and sex comparisons, respectively. To account for multiple testing in a given comparison of interest, we controlled our false discovery rate using the Benjamini-Hochberg procedure for all genes tested across the 5 consid-ered cell types. Any gene with a false discovery rate–corrected $P < 0.01$ was considered significant.

RESULTS

snRNA-Seq of the Human Adult Myocardium

We obtained cardiac tissue samples from 7 potential trans-plant donors, including 4 women and 3 men, without any clinical evidence of cardiac dysfunction (Table). Tissue sam-ples taken from the lateral aspect of the 4 cardiac cham-bers were subjected to nuclear isolation and processing for snRNA-seq (10× Genomics 3′ Single Cell Solution v2). Each sample was processed in replicate, and the second sample underwent a modification in reverse transcription that significantly increased library complexity (Expanded Methods in the Data Supplement). In total, 56 libraries were generated, which were then subjected to cell calling, background adjustment, quality control filtering, and cell alignment (Expanded Methods in the Data Supplement). The workflow for filtration steps and resultant values of samples or cells passing quality control at various phases appear in Figure Ia in the Data Supplement.

In total, 287 269 cells from 44 libraries were used in downstream analyses, including identification of cell types and states (Figure Ib and Table II in the Data Supple-ment). When constructing transcriptional maps of human donors, we used single-cell variational inference batch correction to prevent cells from segregating by individu-al donors within cell type clusters (Figure IIa in the Data Supplement). Use of CellBender remove-background al-lowed for calling of cells with lower transcriptional com-plexity, especially in the context of the relatively complex cardiomyocyte nuclei (Figure IIb in the Data Supplement), while removing contamination from ambient mRNAs. It is important to note, a 3′ capture-derived RNA sequencing library is designed to capture polyadenylated transcripts and thus does not completely identify the RNA molecules present within a ribosomal RNA-depleted, fragment-based, bulk RNA sequencing experiment.

Table.   Clinical Characteristics of Transplant Donors

File No.	Sex	Age, y	Weight, kg	Height, cm	Heart Weight, g	Left Ventricular Mass, g	Left Ventricular Mass Index, g/m²	Left Ventricular End-Diastolic Dimension, cm	Left Ventricular End-Systolic Dimension, cm	Posterior Wall Thickness, cm	Left Ventricular Ejection Fraction, %	Creatinine, mg/dL
P1221	Female	52	59	156	300	NA	NA	4.2	2.9	0.8	75	0.7
P1600	Female	51	68	163	213	134	76.5	4.2	2.8	0.7	50	0.8
P1666	Male	54	62	173	262	159	92.1	NA	NA	1.2	65	0.8
P1681	Male	39	61	170	400	232	136.7	4.5	3	0.9	60	0.7
P1702	Male	59	62	177	386	206	118.0	4.6	2.7	0.8	60	1.4
P1708	Female	60	63	160	281	159	95.0	3.9	1.8	1.1	65	0.5
P1723	Female	47	79	167	310	205	107.1	NA	NA	0.9	60	0.8

NA indicates not applicable.

A total of 17 distinct cell clusters were observed after unsupervised Louvain clustering at a resolution of 1.0. Distributions of cell clusters by chamber-specific uniform manifold approximation and projection representations are shown in Figure 1A, which are combined within a global uniform manifold approximation and projection representation in   Figure 1B and Table III in the Data Supplement. We grouped these into 9 major cell types by canonical marker and ontology analysis, followed by analyses of cell type substructure within each of these groups. Cell clusters are well represented across samples with a few notable exceptions (Figure 1C). First, cardiomyocytes derived from the atria cluster independently of those from the ventricle. Second, 1 ventricular cardiomyocyte cluster is largely found in the right ventricle of a single sample, P1708. Third, lymphocytes were preferentially found in the left ventricle of sample P1723. Two specific clusters represent cytoplasmic fragments as they are enriched for reads mapping to mature transcripts and mitochondrial genes (Figure IIc and IId in the Data Supplement). The following sections will detail the features of each cell cluster, which are described by marker genes in Figure 2 and Table IV in the Data Supplement and analyzed for gene ontology biological function terms in Figure 2. Marker genes were determined as those that display an area under the receiver operating characteristic curve (AUC) value >0.7 and an average natural log fold change >0.6 (Methods in the Data Supplement). In cases where an insufficient number of genes was identified to define a cluster, additional genes with lower levels of overall expression, but strong selectivity for the target cluster of interest, were used for cell type definitions. These were defined as genes expressed in at least 5% of target cells and with a PPV50 >0.90 (Methods). For subclustering analyses, a similar approach was used but lowering the threshold for marker genes to an AUC >0.65 and average natural log fold change >0.5. As with the clusters from the global map, for some subclusters, additional genes expressed in at least 5% of cells in the target subcluster with PPV50 >0.90 were interrogated to assign

subcluster labels (Methods in the Data Supplement). Neither AUC nor PPV50 metrics and their changes in given analyses affect clustering, which is instead governed by the resolution of the Louvain algorithm. Whereas genes having a high AUC or PPV50 value in a given cell type speak to its value as a marker, this does not imply its lack of expression elsewhere. Therefore, for follow-up studies, expression patterns of each gene of interest should be considered on a cell-type-by-cell-type basis.

Nine Major Cell Types and >20 Subclusters of Cell Types in the Human Heart

Distinct Transcriptional Profile of Atrial and Ventricular Cardiomyocytes

Cell clusters 3, 4, 5, 6, 12, and 15 comprise the most frequent major cell type of cardiomyocytes and reflect strong expression of genes involved in canonical excitation–contraction function. Clusters 5 and 12 displayed an enrichment of mature mRNAs (Figure IId in the Data Supplement), suggesting that nonnuclear regions were the source of these cells. We removed these clusters from subsequent analyses as the clear differences between cytoplasm and nuclei would further confound comparisons across chamber and sex. After this exclusion, cardiomyocytes represented 35.9% of observed cells. Cluster 3 displayed canonical markers of the atrium, including *NPPA* (AUC_3=0.91), *MYL7* (AUC_3=0.93), and *MYH6* (AUC_3=0.96; Figure 2 and Table IV in the Data Supplement). Clusters 4, 6, and 15 displayed obvious markers of mature cardiomyocytes such as *TTN* (AUC_4=0.85, AUC_6=0.86, AUC_{15}=0.79) and *MYH7* (AUC_4=0.87, AUC_{15}=0.79) but had fewer known markings of ventricular specificity in the global analysis (Figure 2 and Table IV in the Data Supplement). This is likely attributable to the splitting of ventricular cardiomyocytes among multiple clusters by the Louvain algorithm such that some subclusters are included in the reference group for marker gene identification in a given
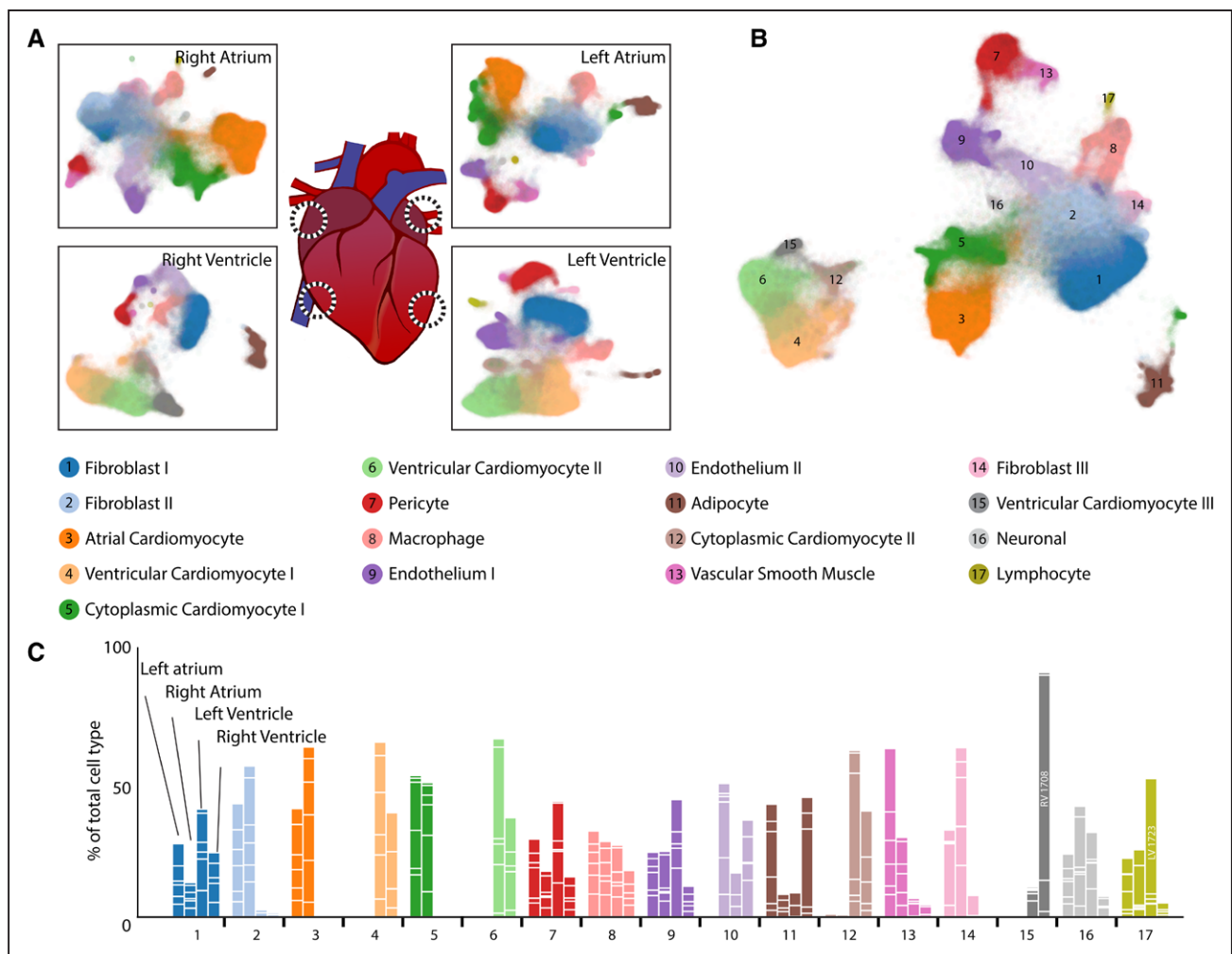
Figure 1. Observed cell types in the adult human heart.
A, Uniform manifold approximation and projection plot displaying cellular diversity present in the human heart by chamber. Each dot represents an individual cell. Colors correspond to the cell cluster labels below the panel. **B**, Combined uniform manifold approximation and projection plot containing 287 269 cells from 7 individuals. Colors and numbers correspond to the cell cluster labels as listed in the lower panel. **C**, Relative representation of cell clusters by sample. Aggregation of 4 bars for each cell cluster equals 100% for each cell type. The white lines in bars separate individual sample contributions. Colors correspond to the cell type descriptions displayed above the panel.

cluster. A separate analysis of atrial versus ventricular cardiomyocytes resolved this issue and is subsequently discussed in the cross-chamber comparisons.

Subclustering of aggregated cardiomyocytes reveals 5 subclusters (Figure IIIa in the Data Supplement). Cardiomyocyte subcluster 1 corresponds to cluster 3 from the global map and contains all atrial cardiomyocytes. In the ventricular cardiomyocytes, cardiomyocyte subcluster 5 has enrichment for mitochondrial components and an increased mature transcript proportion, suggesting that these may also be cytoplasmic contaminants. Cardiomyocyte subcluster 4 (CM-S4) correlates strongly with cluster 15 in the global map and displays increased expression of *ANKRD1* (AUC_{CM-S4}=0.82), which is thought to have a role in cardiomyopathy-associated remodeling,[14] and *KCP* ($PPV50_{CM-S4}$=0.91), a BMP (bone morphogenetic protein) modifier whose expression is associated with heart failure[15] (Figure IIIa and Table V

in the Data Supplement). These cells are most often found in the right ventricle of a single donor (73% from P1708) and may represent a marker of a subclinical cardiac pathology.

Identification of Activated and Nonactivated Cardiac Fibroblasts

By volume, cardiomyocytes comprise the majority of heart mass; however, in the absence of structural heart disease, fibroblasts are roughly equivalent to cardiomyocytes in cell number. Given that the hearts used in this study were largely free of fibrotic remodeling (Figure IVa in the Data Supplement, for example), we expected similar representation for fibroblasts and cardiomyocyte nuclei within our data. The cells from the combination of clusters 1, 2, and 14 represent cardiac fibroblasts, constituting 32.4% of observed cells. These cells display common markers of fibroblast lineages, with enriched expression of known fibroblast genes such as *DCN*
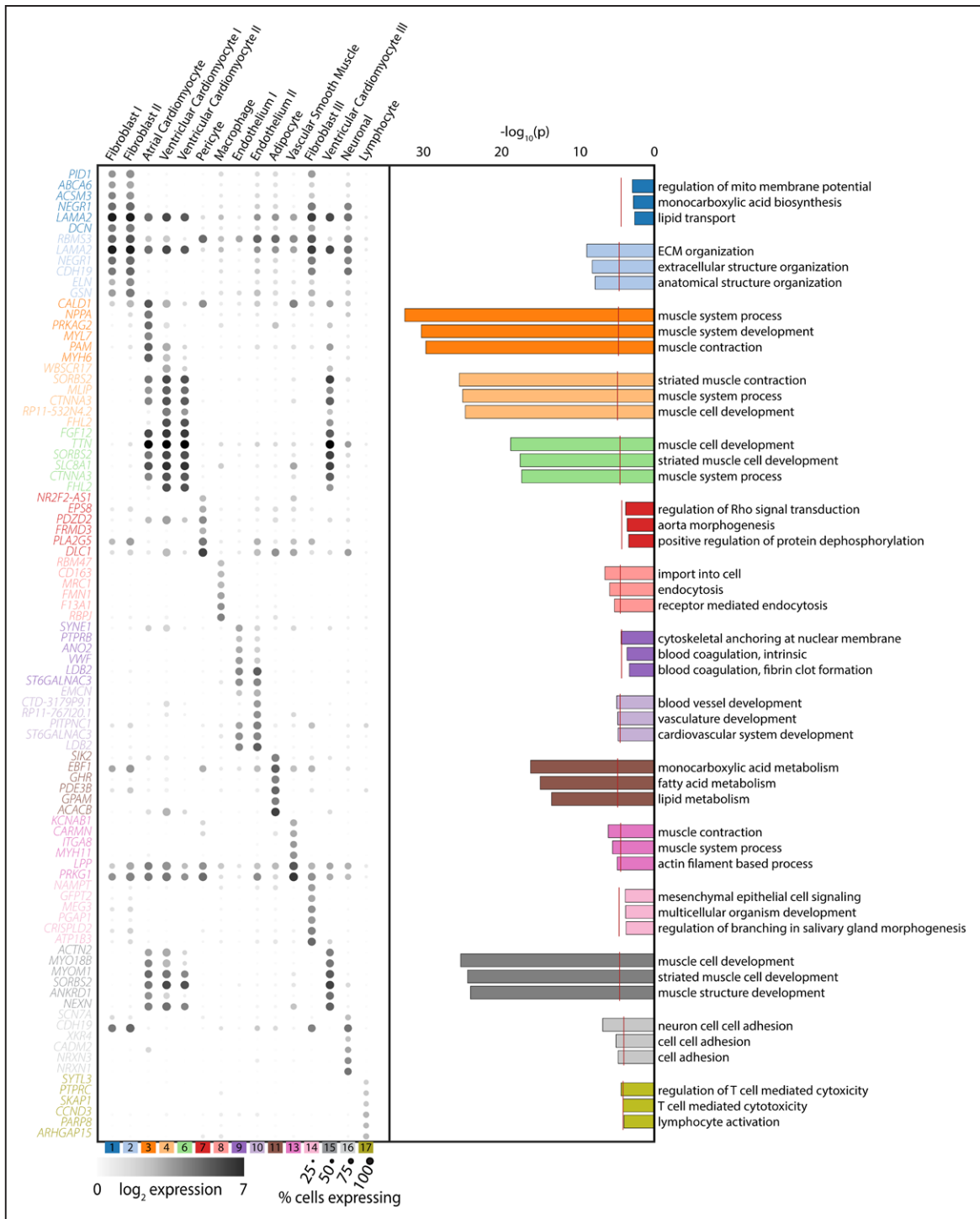
Figure 2. Gene and ontology definitions of observed cardiac cell clusters.
Dot plots display the top 6 marker genes for each supercluster as determined by area under the curve (left panel). The size of the dot represents the percentage of cells within the cluster where each marker is detected and the gradation corresponds to the mean $\log_2$ of the counts normalized by total counts per cell ×10 000. Gene ontology enrichment analysis as performed by GOStats using all genes that reach an area under the curve threshold >0.70 and an average log fold change >0.60 for the given cell cluster (right panel). The red dotted line indicates a Bonferroni statistical significance threshold. The top 3 gene ontologies are shown for each cell cluster. ECM indicates extracellular matrix.

(AUC$_1$=0.85, AUC$_2$=0.83), which encodes the proteoglycan decorin, which regulates collagen fibrillogenesis, and *ELN* (AUC$_1$=0.71, AUC$_2$=0.86), which produces elastin, a major component of the extracellular matrix (Figure 2 and Table IV in the Data Supplement). The former was

used to evaluate the distribution of the fibroblasts in our tissue samples, which exhibit the traditional interstitial localization observed in previous work (Figure IVb in the Data Supplement). In addition to extracellular matrix proteins, members of the ATP-binding cassette subfamily of

transmembrane transporters, including *ABCA6*, *ABCA8*, and *ABCA9*, were also preferentially expressed in one or more of these clusters (*ABCA6*: $AUC_1=0.80$, $AUC_2=0.77$; *ABCA8*: $AUC_1=0.79$, $AUC_2=0.79$; *ABCA9*: $AUC_1=0.75$, $AUC_2=0.74$; Table IV in the Data Supplement). Analysis of ontology for specific genes in this class displays expected terms in the realm of extracellular matrix and structural organization, with the greatest enrichment in cluster 2 (Figure 2). No terms reached significance thresholds for clusters 1 and 14, perhaps as a consequence of a lower number of genes surpassing our criteria of a marker gene within these clusters (15 and 48, respectively). This is largely a consequence of including other fibroblast clusters in the reference outgroup for marker gene testing.

To evaluate the structure in the fibroblast population, we performed local clustering of these cells, from which 4 populations were observed (Figure 3A). Fibroblast subcluster 2, which comprises a large proportion of cluster 2 in the global map, shows an enrichment for *NPPA*, a known marker of atrial cardiomyocytes (Figure 3B). Whether this is truly fibroblast-specific *NPPA* expression, an artifact derived from cardiomyocyte/fibroblast nuclear doublets, or a result of the presence of *NPPA* transcript in the extranuclear contaminant requires investigation. Fibroblast subcluster 3 (FB-S3) displays enriched expression of fibrosis-associated genes *NOX4* ($AUC_{FB-S3}=0.70$) and *IGF1* ($AUC_{FB-S3}=0.69$), and cluster FB-S4, which corresponds to cluster 14 in the main map, exhibits clear upregulation of profibrotic markers, including *ADAMTS4* ($AUC_{FB-S4}=0.69$), which encodes a profibrotic metalloprotease, *VCAN* ($AUC_{FB-S4}=0.69$), which encodes the proteoglycan versican,[16] and *AXL* ($AUC_{FB-S4}=0.69$), which encodes a receptor tyrosine kinase associated with pathologic remodeling[17] (Figure 3B and Table V in the Data Supplement). Further interrogation of these cells by RNA in situ hybridization with α-*ADAMTS4*–specific probes demonstrates an interstitial distribution throughout the tissue rather than being localized to a particular region (Figure 3C), suggesting that an organ-wide event stimulated this fibroblast state transition. To attempt to identify the lineage of these fibroblast subclusters, we intersected our data with those from fibroblast activation in mice and humans.[18,19] None of these clusters is enriched for expression of canonical markers for fibroblast activation (*POSTN*), myofibroblast transition (*MYH11*, *FAP*), or transformation to fibrocytes (*CHAD*, *COMP*; Figure 3B). Whether these cells are a previously undefined state in canonical fibroblast activation or are instead an entirely noncanonical form of fibroblast will be the focus of future work.

Vascular Support Network of Pericytes and Vascular Smooth Muscle

Defining specific markers for microvessel-associated pericytes and large vessel–associated vascular smooth muscle cells has remained difficult, because the cells derive from similar progenitors and serve similar vascular support functions. We observed a relative enrichment of pericyte-specific *PDGFRB* in cluster 7 ($AUC_7=0.75$) and the expression of smooth muscle actin (*MYH11*) in cluster 13 ($AUC_{13}=0.89$; Figure 2 and Table IV in the Data Supplement). This observation, combined with the preponderance of small vessels in our tissue samples, led us to classify the more numerous cluster 7 as pericytes and cluster 13 as vascular smooth muscle. Subcluster analyses of these cell types yielded little appreciable structure (Figure IIIb and Table V in the Data Supplement), with the exception of pericyte subcluster 2 (P-S2), which is enriched for some markers of endothelial cells (*VWF*, $AUC_{P-S2}=0.77$, for example). Whether this indicates a differentiation event, as pericytes derive from endothelial cells, potential nuclear doublets, or ambient RNA contamination within the data remains unclear.

A Complex Cardiac Immune Cell Component

Two cell clusters (8 and 17) identified in this analysis have genetic signatures consistent with immune cell types. The first, cluster 8, represents cardiac resident macrophages and can be characterized by expression of the scavenger receptors *CD163* ($AUC_8=0.84$) and *COLEC12* ($AUC_8=0.72$), the mannose receptor *MRC1* ($AUC_8=0.85$), the E3 ubiquitin ligase *MARCH1* ($AUC_8=0.72$), and natural resistance–associated macrophage protein 1 (*NRAMP1* or *SLC11A1*; $AUC_8=0.74$; Table IV in the Data Supplement). Subclustering revealed 2 populations that both express M2 polarization–associated genes, including *RBPJ* and *F13A1* in macrophage subcluster 1 (M-S1; $AUC_{M-S1}=0.85$ and 0.84, respectively) and the transmembrane collagen *COL23A1* in macrophage subcluster 2 (M-S2; $AUC_{M-S2}=0.65$; Figure 4A and Table V in the Data Supplement).

A second immune cell population (cluster 17) selectively expresses a number of well-known T-cell markers. This includes the T-cell surface antigen *CD2* ($PPV50_{17}=0.99$), the early T-cell activation antigen *CD69* ($PPV50_{17}=0.99$), and the T-cell receptor–associated transmembrane adaptor 1 (*TRAT1*; $PPV50_{17}=0.98$; Table II in the Data Supplement). *PTPRC/CD45* ($AUC_{17}=0.77$), an essential regulator of T- and B-cell antigen receptor signaling, the T-cell immune adaptor *SKAP1* ($AUC_{17}=0.77$), and the thymocyte selection marker *CD53* ($PPV50_{17}=0.91$) show selectivity to this cluster (Figure 2 and Table IV in the Data Supplement). This overall lymphocyte population can be further subdivided into 2 distinct subclusters (lymphocyte subcluster 1 and lymphocyte subcluster 2). Given the expression of tryptases (*TPSB2*, *TPSAB1*) and the FcεR1 subunit *MS4A2*, lymphocyte subcluster 2 exhibits canonical signatures of mast cells. An additional gene of note within lymphocyte subcluster 2 is *KIT*, which was long associated with cardiac resident stem cells but largely refuted.[20]
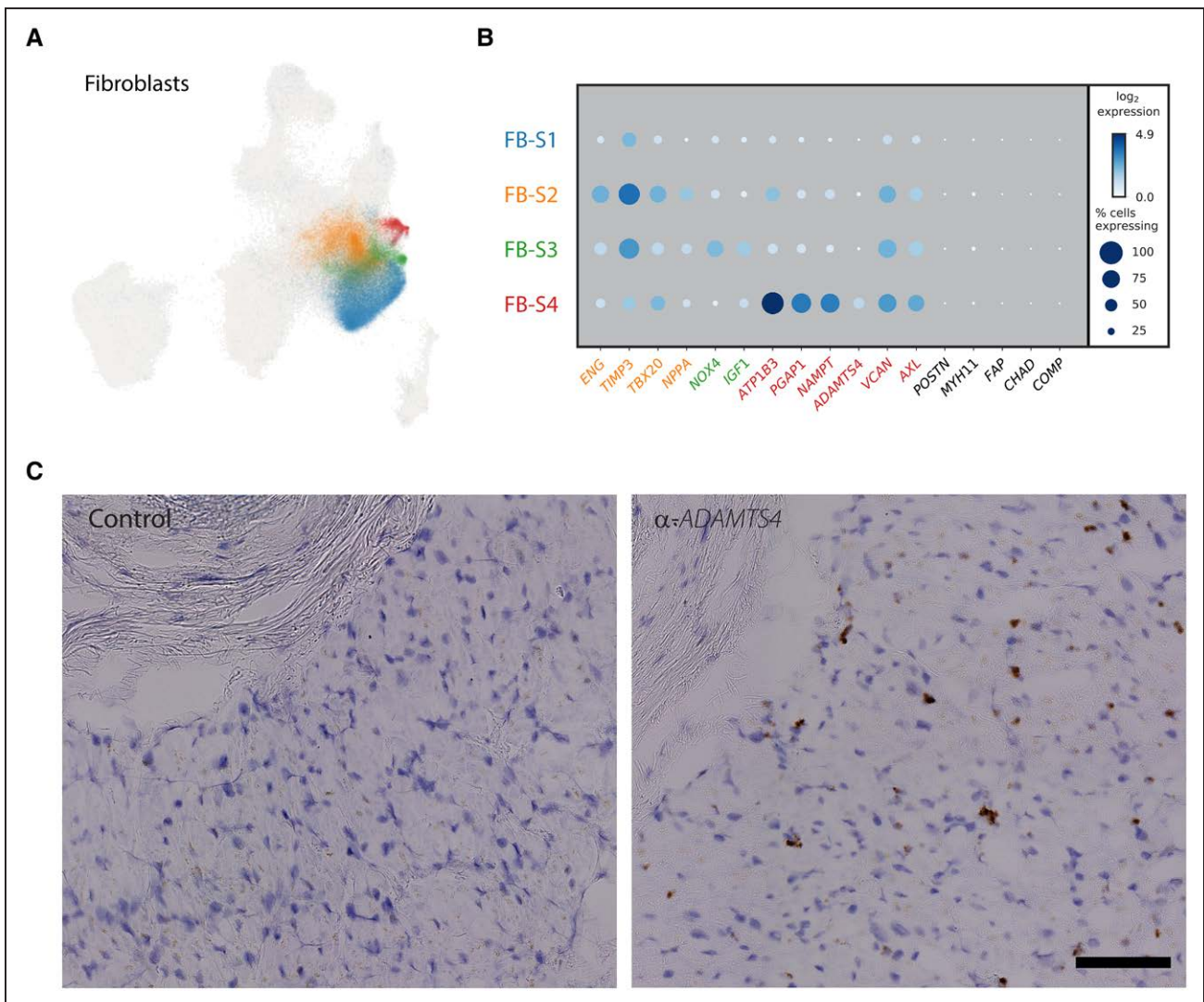
Figure 3. Subclustering fibroblasts to identify activated and quiescent fibroblasts.
A, Uniform manifold approximation and projection plot representing the 4 observed fibroblast subclusters superimposed over the global uniform manifold approximation and projection distribution. Each dot represents an individual cell and is colored by its respective subcluster. **B**, Dot plot detailing the percentage of cells where each gene is detected (dot size) and mean $\log_2$ expression (blue hue) for representative subcluster marker genes. Each row represents the cell subcluster as displayed in (A) as according to color. **C**, Representative RNA in situ hybridization showing localization of *ADAMTS4*-positive cells (brown stain) in sample LV1723 compared with a nonspecific RNA probe (control). Localization of nuclei is shown with hematoxylin (blue stain). Scale bar represents 100 μm.

We observe enriched *KIT* expression within this lymphocyte subpopulation but found no evidence for coexpression in any cell with signatures of being progenitors or precursors for cardiomyocytes (Figure 4B and Table V in the Data Supplement). Whereas we also observe detectable expression of KIT in endothelial cells (>1 transcript in 0.4% of cells), where it has been reported to be important in differentiation,[21] it is expressed to a much lower level than in lymphocyte subcluster 2 mast cells.

Identification of Vascular and Nonvascular Endothelial Cells

The endothelial cell component of the heart consists of those cells that line the large and small circulatory vessels, the lymphatics, and the endocardium. From global clustering, we identified 2 major endothelial cell clusters (clusters 9 and 10), which express canonical markers such as *VWF* (AUC_9=0.88, AUC_{10}=0.77) and *PECAM-1* (AUC_9=0.71, AUC_{10}=0.81) but were unable to resolve subtypes further before subclustering analysis (Figure 2 and Table IV in the Data Supplement).

Five subclusters were identified in combined endothelial clusters 9 and 10 (Figure 4C). We were unable to clearly resolve subclusters based on AUC markers alone, but interrogation of less abundant genes with significant selectivity proved useful in identifying subcluster populations. For example, in subcluster 4 (L-EC), we observed enrichment for cells expressing lymphatic endothelial cell markers including *PROX1*, *FLT4*, and *PDPN* ($PPV50_{L-EC}$ of 0.95, 0.91, and 0.94, respectively; Table V in the Data Supplement). A subset of cells in endothelial subcluster 2 (EC-S2) express *BMX* (AUC_{EC-S2}=0.65), an
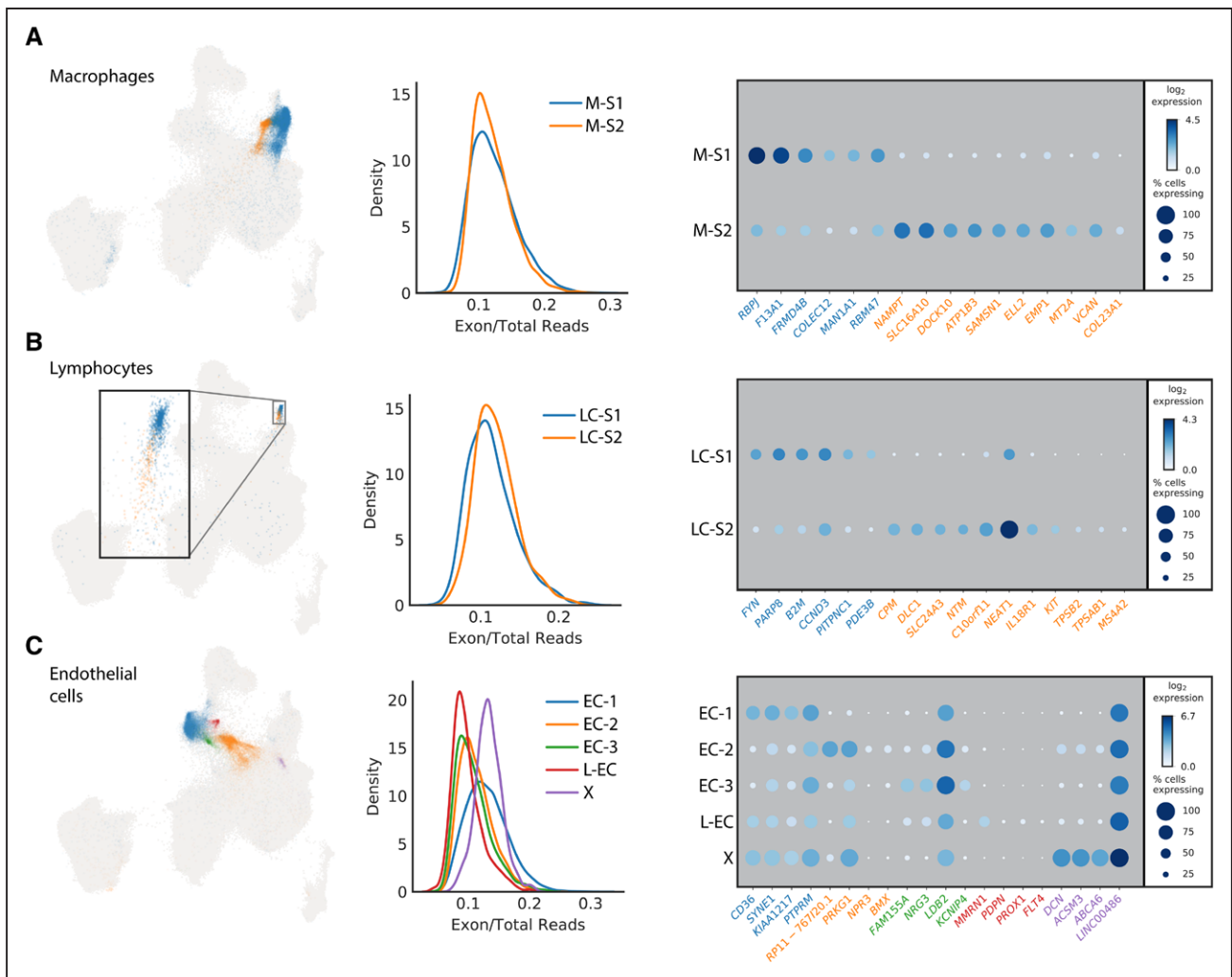
Figure 4. Subclustering to identify additional cellular diversity within macrophages, endothelial cells, and lymphocytes.
A, Left panel shows the uniform manifold approximation and projection distribution of the 2 identified macrophage (M) subclusters. Each dot represents an individual cell colored by its respective subcluster. Center panel represents the calculated proportion of exonic mapping reads for the 2 subclusters. Right panel details the top markers by area under the curve for each subcluster. The size of the dot relates to the percentage of cells within the cluster that express that markers whereas the gradation relates to the mean $\log_2$ of the counts normalized by total counts per cell $\times$ 10 000. **B**, Left panel is the distribution of the 2 subclusters for lymphocytes (LC) in the global uniform manifold approximation and projection. Each dot represents an individual cell colored by its respective subcluster. Inset is the magnification of the outlined region. Center panel displays equivalent exon mapping reads for each of the subclusters. Right panel displays the top genes defining each subcluster as defined by area under the curve. **C**, Left panel is the distribution of the 5 identified subclusters of endothelial cells (EC) within the global uniform manifold approximation and projection plot. Each dot represents an individual cell colored by its respective subcluster. Center panel details the percentage of exon mapping reads, where cluster X (purple) has enrichment for exonic reads. Right panel shows a dot plot of the top markers for each subcluster by area under the curve with the addition of those markers used for identification of the lymphatic endothelium cluster derived from the standardized positive predictive value.

artery-specific endothelial cell marker, as well as *NPR3* (AUC_{EC-S2}=0.65). In mice, *NPR3* is selectively expressed in adult endocardium,[22] suggesting the EC-S2 population may represent endocardial cells (Table V in the Data Supplement). These observations reflect the fact that the heart biopsies used did not include any large vessels, explaining in part the lack of distinct arterial and venous endothelial cell populations.

Epicardial Adipocytes Enriched in the Leukocyte Marker CD96

Epicardial adipose tissue is present in human hearts, which comprises up to 20% of its total mass.[23] Adipocytes may also be observed within the heart itself in

pathologic conditions such as obesity or cardiomyopathy. Tissues were generally free of myocardial adiposity as observed by histology in our samples with the exception of the right ventricle of P1723 (Figure IVb in the Data Supplement). Given that cells of this sample are not overly represented in the cluster, we propose that cluster 11 is composed primarily of epicardial adipocytes, with ontology analysis identifying terms such as fatty acid and lipid metabolism (Figure 2). These cells were characterized by genes whose expression ultimately regulate the size and stability of lipid droplets, such as *CIDEC* (AUC_{11}=0.72) and *PLIN5* (AUC_{11}=0.78). These data also support the view of epicardial fat as an endocrine organ. *ADIPOQ*, which modulates fatty acid

transport and increases intracellular calcium, is present in nearly 65% of adipocyte nuclei but only 0.3% of other cell types ($AUC_{11}=0.82$). *TRHDE*, which inactivates thyrotropin-releasing hormone, and *IGF-1* are also strongly enriched within this population ($AUC_{11}=0.76$ and $AUC_{11}=0.76$, respectively). *IGF-1* also has an important role in cell growth, proliferation, and resistance to death later in the life of an individual, functions that directly relate to its significant role in the development of obesity.[24] It is surprising that these cells are also enriched for *CD96*, a marker most often identified with natural killer and T cells ($AUC_{11}=0.73$; Table IV in the Data Supplement).

Autonomic Neuronal Inputs of the Intrinsic Cardiac Network

The heart is innervated by the central nervous system through the cardiac plexus, which distributes parasympathetic (vagal) and sympathetic stimulation. An intrinsic cardiac autonomic network, consisting of ganglionated plexi within epicardial fat pads, resides in all 4 chambers of the heart. We identified a subset of neuronal cells in cluster 16, defined largely by neuronal cell adhesion genes such as the neurexins (*NRXN1*, $AUC_{16}=0.91$; and *NRXN3*, $AUC_{16}=0.87$) and *NCAM2* ($AUC_{16}=0.73$) rather than by electrophysiology or secretory-associated genes. The only ion channel gene identified as a marker in this cluster is *SCN7A* ($AUC_{16}=0.74$), initially described in glia but now understood to reside in other cell types of the nervous system.[25] For signaling genes, the receptor genes *ADGRB3* ($AUC_{16}=0.72$), which acts to promote angiogenesis, and *SHISA9* ($AUC_{16}=0.72$), which modulates α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid–type glutamate receptors, were robustly expressed in this cluster (Table IV in the Data Supplement). Given the sampling location of the lateral wall and the presence of this neuronal subtype through all 4 chambers, it is likely that the neuronal cells identified in the present study are derived from the intrinsic cardiac autonomic network.

Differential Expression Analysis Uncovers Chamber- and Sex-Specific Gene Expression Profiles in Cell Subtypes

We next determined whether expression programs in the major cell types differed by cardiac chamber or sex. Before performing differential expression testing, we first removed any cluster or subcluster that was previously labeled as cytoplasmic (clusters 5 and 12 from the global map and cardiomyocyte subcluster 5 and endothelial subcluster 5), collapsed cell clusters into their respective major cell types, and removed genes with a poor PPV50 for the major cell type of interest (Methods). We then performed differential expression testing using

a generalized linear mixed model framework on the 5 most numerous major cell types (cardiomyocytes, fibroblasts, endothelial cells, pericytes, and macrophages). The smaller number of cells for other cell types coupled with the sparsity of snRNA-seq expression matrices sequencing limited our ability to confidently call differentially expressed genes in rare cell types. For all cell types, we performed gene ontology analysis on differentially expressed genes in the left versus right side of the heart, but no terms reached statistical significance (Figure V in the Data Supplement).

Cardiomyocytes Are the Most Distinct Cell Type Between Chambers

Atrial and ventricle cardiomyocytes are well known to have distinct physiologic functions, contractile properties, and electric signaling. These functional and structural differences are reflected in discrete transcriptional profiles. As anticipated, when we compared the atria with the ventricles, we observed 2300 genes that were significant at a false discovery rate of 0.05 (Figure 5A through 5C and Table VI in the Data Supplement). These differences were exemplified by an increased expression of *HEY2* and *MYH7* in the ventricles (effect size=3.75, $P=1.5\times10^{-27}$ and effect size=1.83, $P=1.07\times10^{-16}$, respectively) and *NPPA* and *MYL4* in the atria (effect size=6.89, $P=1.22\times10^{-30}$ and effect size=4.23, $P=2.94\times10^{-29}$, respectively). We identified 2058 differentially expressed genes between the left atrium and left ventricle, but only 1134 differentially expressed genes between the right atrium and right ventricle.

In contrast with the marked transcriptional patterns observed between the atria and ventricles, there were many fewer genes that were differentially expressed when comparing the left versus the right side of the heart. A comparison of the left versus right atria revealed 248 differentially expressed genes, whereas only 24 genes were differentially expressed between the left and right ventricles.

Closer inspection of the data yields noteworthy insights into chamber-specific expression programs. For example, the atrial fibrillation susceptibility gene *PITX2*[26] was observed in 2.3% of left atrial cardiomyocytes and in <0.05% of cardiomyocytes in any other chamber. It is interesting that *HCN4* is present in 4.3% of cardiomyocytes from the right atrium, in only ≈1% of cardiomyocytes from the right ventricle and left ventricle, and <0.5% of cardiomyocytes from the left atrium. The *HCN4* gene encodes the ion channel responsible for spontaneous depolarization and has also been associated with atrial fibrillation.

Other genes with limited or unexplored roles in cardiomyocyte biology also exhibit chamber preference. Among these, *HAMP*, which encodes a protein for regulating iron export, and the solute carrier gene
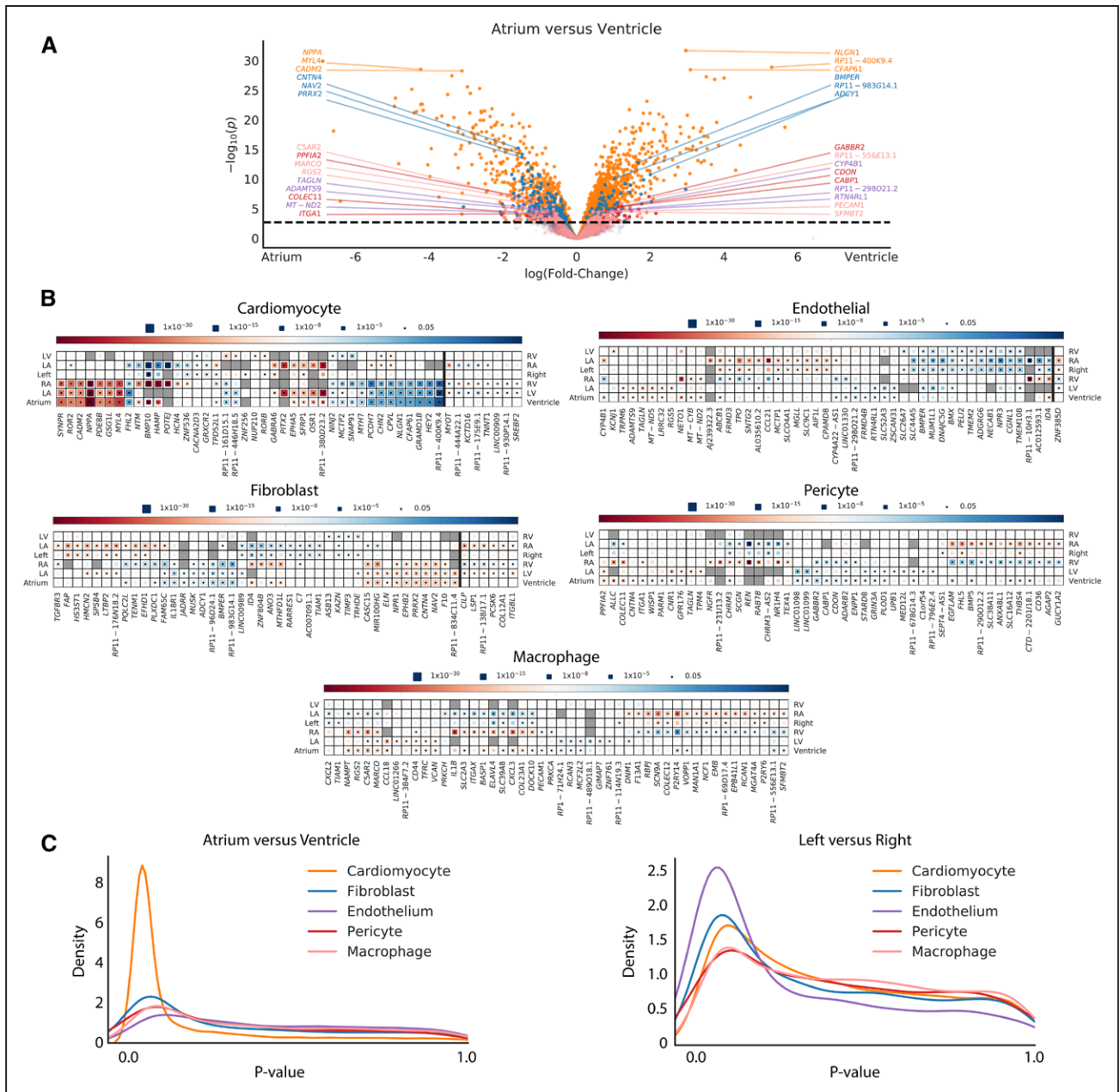
Figure 5. Differential expression analyses for chamber-specific signatures of major cell types.
A, Volcano plot detailing differential expression of genes when comparing the aggregated atrial and ventricular chambers in cardiomyocytes (orange), fibroblasts (blue), endothelial cells (purple), pericytes (red), and macrophages (pink). The X-axis represents the fixed effect from the generalized linear mixed model and the Y-axis represents the $-\log_{10}$ (P value). Dotted line indicates the false discovery rate–adjusted P value threshold for statistical significance. The top 3 genes upregulated in atrial cells and ventricle cells are highlighted for each cell major cell type. **B**, Heat maps detailing a representative selection of significantly differentially expressed genes between chambers within major cell types. Color indicates whether the gene is enriched within the chamber listed on the left (red) or right (blue). Size of the inset block indicates the P value for the comparison. Dot within the block indicates statistical significance for the given comparison. Genes to the right of the dark vertical line are those with different directionalities when comparing atria versus ventricles on the left or right side. **C**, Density plot displaying the number of genes with certain P values across the P value spectrum within each major cell type for atrium versus ventricle (left panel) and left versus right (right panel) comparisons.

SLC5A12 are found predominantly in the right atrium (present in 18.3% and 5.8% of cardiomyocytes in the right atrium, respectively, compared with <1% of cardiomyocytes in any other chamber). Eight genes display significant differences in expression in opposing directions when comparing left or right atrium with its respective ventricular partner (Figure 5B and

Table VI in the Data Supplement). Among these are *MYOT* (left: effect size=0.75, P=8.86×10⁻⁵; right: effect size=−0.93, P=2.00×10⁻⁵) and *TNNT1* (left: effect size=0.64, P=0.001; right: effect size=−0.45, P=1×10⁻⁴), which are enriched in the right ventricle and left atrium and play critical roles in sarcomeric organization and function.

Noncardiomyocytes Display Striking Chamber Specificity

Whereas differences in cardiomyocytes between chambers were expected, it was less clear from previous work whether chamber specificity exists within other cardiac resident cells. It is surprising that there were profound signatures of chamber specificity in the other cell types examined. A total of 765 genes surpassed the false discovery rate–corrected P value threshold in fibroblasts for at least 1 comparison of chamber or laterality. A total of 125 genes in pericytes, 320 genes in macrophages, and 354 genes in endothelial cells were also found to be differentially expressed (Figure 5B and Table VI in the Data Supplement).

Among fibroblasts, pericytes, and macrophages, the atrial versus ventricular comparisons account for the majority of differential expression, with the right atrial cells being consistently the most divergent. In some cases, this divergence is sufficient to drive some of the subclustering observed in Figure 4 and Figure III in the Data Supplement. The most striking example of this is in the macrophage population, where the differential expression between the right atrial macrophages and those of other chambers is strong enough to detect a second macrophage subcluster (M-S2, Figure 4A), which consists almost entirely of right atrial cells (94.0%). In contrast, endothelial cells are most distinct when comparing sidedness (220 differentially expressed genes for left versus right, 43 differentially expressed genes for atrium versus ventricle). Again, much of this is driven by the right atrium, with 217 significant genes when compared with the left atrium. This difference manifests in the subclustering, where right atrial cells make up 88.2% of endothelial cluster 3 (Figure 4C).

Similar to the cardiomyocytes, some genes display different directionalities when comparing atria versus ventricles on the left or right side (Figure 5B and Table VI in the Data Supplement). This includes 6 genes in fibroblasts, including *CILP* (left: effect size=1.13, P=4.04×10^{-5}; right: effect size=−1.55, P=8.43×10^{-7}) and *ITGBL1* (left: effect size=1.07; P=7.26×10^{-6}; right: effect size=−0.67, P=4.08×10^{-4}), which have links to the regulation of fibrosis,[27,28] and 1 gene in endothelial cells, *ZNF385D* (left: effect size=0.82, P=0.001; right: effect size=−1.14, P=1.60×10^{-8}). There are profound differences in the expression profiles of nonmyocytes across the cardiac chambers.

Sex-Based Differential Expression Identifies Genes Associated With Myopathy and Coronary Artery Disease

Biological sex has a profound effect on cardiac morphology, physiology, and susceptibility to cardiovascular disease, but the molecular differences of the heart between the sexes remain obscure. Given the inclusion of 4 female and 3 male donors in our data, we proceeded to separate the cells by sex and performed differential expression testing in the same 5 major cell types both globally and by chamber of origin. The number of sex-specific genes was greatly reduced when compared with those derived from chamber specificity in the previous section. This may be attributable to limited sample numbers by sex (4 female versus 3 male), a greater importance of cytoplasmic RNAs in sex-specific differences (ie, RNA half-life), or a general concordance in gene expression profiles of men and women at single-cell resolution. In total, 17 genes exhibited sex-based differential expression within cardiomyocytes, 2 within the endothelium, 10 within the fibroblasts, 3 within the macrophages, and none for the pericyte comparisons (Table VII in the Data Supplement). Approximately one-third of the genes that were differentially expressed by sex were autosomal (cardiomyocyte, n=6; fibroblast, n=4). As anticipated, several of these differentially expressed genes are related to hormonal signaling. *CRISPLD2* is induced by the progesterone receptor[29] and *UGT2B4* is involved in estrogen metabolite modification.[30] *NEB*, which encodes the sarcomeric structural protein nebulin, is enriched within the left ventricle in male hearts (effect size=1.54, P=1.73×10^{-6}); *ZNF827*, which resides proximal to a genome-wide association study (GWAS) locus for coronary artery disease,[31] is expressed at increased levels in female hearts, with the most marked upregulation in the right atrium (effect size=1.31, P=2.12×10^{-6}).

Integration of snRNA-Seq Data With Cardiovascular Genetics and the Druggable Genome

We next aimed to apply our snRNA-seq data to better understand the basis of human cardiovascular disease using 3 complementary approaches. First, we examined the cell type–specific expression of genes implicated in Mendelian forms of cardiovascular disease. Second, we related cardiac transcriptional data to the data derived from population-based GWAS for cardiovascular diseases and traits. Third, we intersected our snRNA-seq data with genes that are potentially druggable to identify novel therapeutic targets for cardiovascular diseases.

Genes Implicated in Cardiomyopathies and Arrhythmia Syndromes Are Enriched in Cardiomyocytes

Intersection of our snRNA-seq data with a panel of genes previously implicated in cardiomyopathies and arrhythmia syndromes revealed 3 general patterns. First, as anticipated, >25% of the pathogenic genes showed enriched selectivity (AUC > 0.70) in the cardiomyocyte population (Figure 6 and Figure VI in the Data Supplement; 17/75 genes for arrhythmias [P<0.0001],
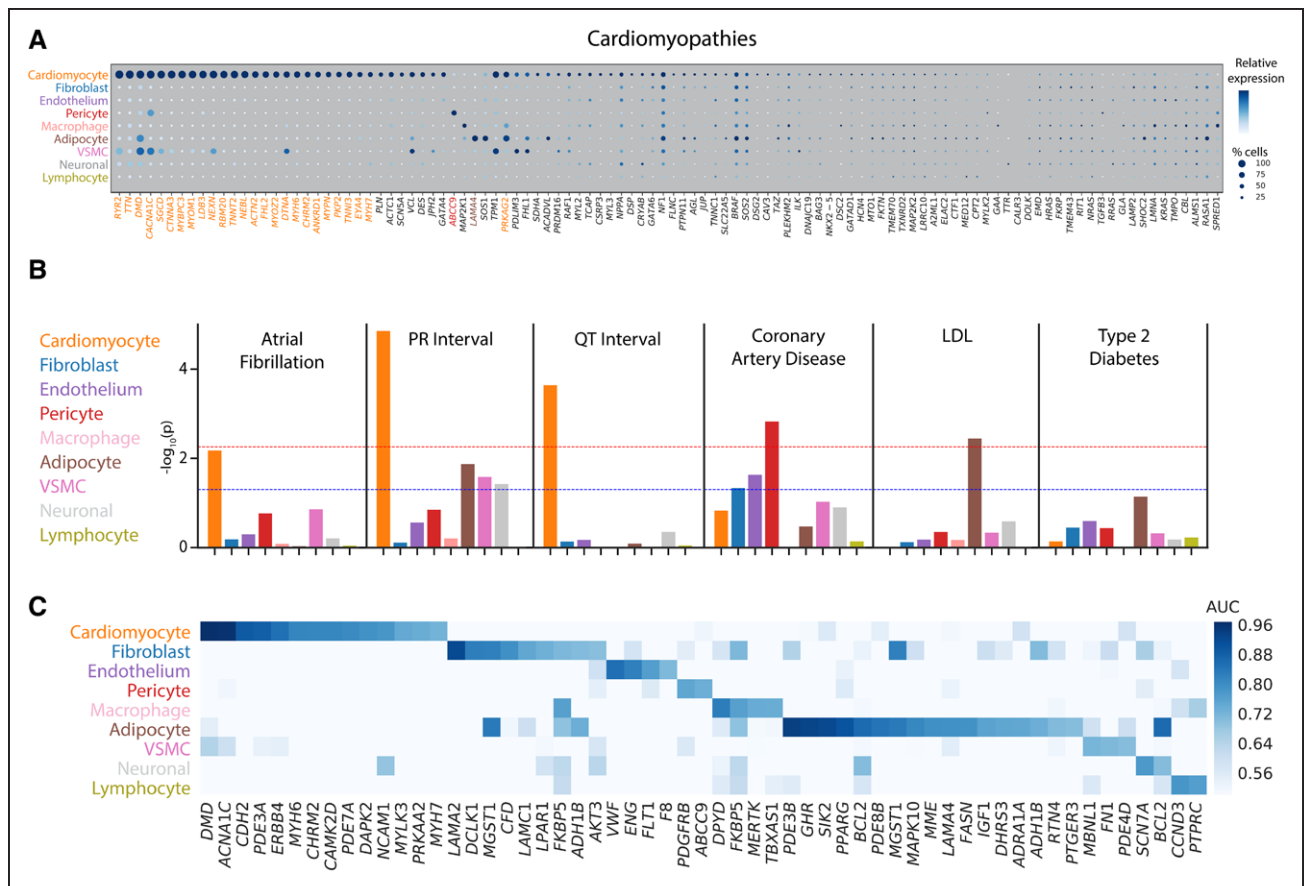
Figure 6. Integration of single-nucleus RNA sequencing with genetic associations to uncover disease biology.
A, Dot plot for genes on standard cardiomyopathy clinical testing panels. The size of each dot represents the percent of cells in which the gene of interest is detected and the shading represents the relative expression of the gene. Color of the genes corresponds to the cell type for which the area under the curve (AUC) reaches 0.70 or greater. Genes with black labels indicate no cell type that reaches this threshold. Size and shade of the dot corresponds to percentage of cells and relative expression, respectively. **B**, Results of linkage disequilibrium score regression analyses on the combined major cell types. Dotted lines display unadjusted (blue) and Bonferroni-adjusted (red) P value thresholds for statistical significance. Colors of the bars correspond to the color of the major cell type labels on the left. **C**, Heat map detailing the intersection between single-nucleus RNA sequencing data and tier 1 druggable genes. Genes with an AUC >0.70 in at least 1 cell type are shown. Shade of the color represents the AUC value for the gene within each cell type. Of note, genes that have an AUC >0.70 in multiple cell types appear multiple times in the plot. LDL indicates low-density lipoprotein.

27/106 genes for cardiomyopathies [$P<0.0001$]. Second, a smaller subset of known pathogenic genes are highly expressed in noncardiomyocyte populations. This pattern was exemplified by the *ABCC9* gene, which has been implicated in dilated cardiomyopathy and is predominantly expressed in pericytes. Similarly, *LAMA4*, which encodes a component of the extracellular matrix and has been associated with dilated cardiomyopathy, was specifically expressed in adipocytes ($AUC_{AD}=0.79$). Third, we found that approximately half of the genes implicated in Mendelian cardiovascular diseases were not highly or broadly expressed in the healthy human heart.

Combining GWAS and snRNA-Seq Data to Identify the Most Relevant Cell Types for Cardiovascular Diseases

To identify putative cell types of interest to a set of complex traits and diseases, we used linkage disequilibrium score regression to partition genetic heritability

from GWAS studies. Assuming a cis-regulatory model for single nucleotide polymorphism function, the approach partitions single nucleotide polymorphism heritability derived from GWAS across regions near genes considered to be cell type–specific in our snRNA-seq data. Should single nucleotide polymorphism–trait associations be enriched around cell type–specific genes, this suggests that heritability of the trait is driven in part by the genetic effects in that cell type. We applied this approach to a range of cardiometabolic traits, as shown in Figure 6B and Table VIII in the Data Supplement.

Integration of our single-nucleus sequencing results with GWAS data for cardiometabolic traits revealed the expected enrichment in cardiomyocytes for 2 electrocardiographic traits: the PR interval ($P=1.4\times10^{-5}$) and the QT interval ($P=2.3\times10^{-4}$). We observed similar cardiomyocyte enrichment for the most common cardiac arrhythmia, atrial fibrillation ($P=0.007$). It is interesting that we also observed a marked enrichment in pericytes for genes at the loci for myocardial infarction ($P=0.001$)

and in adipocytes for low-density lipoprotein cholesterol (P=0.004).

After examining global enrichments, we chose to use a more reductionist approach to evaluate potentially unique expression profiles of disease-associated genes. Expression quantitative trait loci (eQTL) mapping, which evaluates changes in gene expression attributable to genotype, is a common strategy for linking a GWAS locus to a particular gene. We used the intersection of known disease- or trait-associated eQTLs from GTEx[32] and our own work[33] to determine the cell type where the transcript of interest is most highly expressed. For each trait, we limited our analysis to genes from the most disease-relevant tissue; for example, the QT interval is only intersected with left ventricular eQTLs and atrial fibrillation only those from the left atrium. eQTLs are derived from tissue-level RNAseq experiments and are thus predisposed to discover signals in more prevalent cell types. It is surprising that rather than patterns that indicate cardiomyocyte centered expression, genes generally show nonspecific cell type expression, with a few interesting patterns emerging (Figure VIb in the Data Supplement). In the left ventricle, 1 of the 11 putative genes for PR interval (*PDZRN3*) shows enriched expression in cardiomyocytes (AUC_{CM}=0.88), 1 of 21 putative genes for QT interval (*SLC35F1*) shows enriched expression in neuronal cells (AUC_{NR}=0.71), and 2 of 37 putative genes for coronary artery disease show enriched expression in adipocytes (*C6orf106*, AUC_{AD}=0.70) and vascular smooth muscle cells (*LMOD1*, AUC_{VSMC}=0.74). It is interesting that in the left atrium, the putative PR interval gene *PDZRN3* shows enriched expression in adipocytes (AUC_{AD}=0.76) and 2 of 12 atrial fibrillation genes show enriched expression in cardiomyocytes (*CASQ2*, AUC_{CM}=0.74) and endothelial cells (*SYNE2*, AUC_{EN}=0.74).

Cell Type–Specific Expression of Potentially Druggable Genes

To identify potential drug targets in cardiac tissue, we aimed to identify tier 1 classified genes from the druggable genome[34] that shows selectivity toward particular cardiac cell types. This tier includes targets of both approved drugs and those in clinical development. Of the 1420 potential genes, 53 unique genes were specifically expressed in at least 1 major cell type with an AUC >0.70 (Figure 6C). Most commonly these genes were found in adipocytes (n=17), cardiomyocytes (n=14), and fibroblasts (n=9). Among these, *CACNA1C*, the receptor for calcium channel blockers that are commonly used to treat hypertension, and *PDE3A*, a known target of inamrinone for treatment of congestive heart failure,[35,36] showed selectivity toward cardiomyocytes. The selective expression of other druggable genes in cardiac cell types, and particularly in nonmyocytes, will provide new opportunities for future therapeutic development.

DISCUSSION

We have developed a comprehensive map of the transcriptional landscape in normal human heart comprising snRNA-seq for >280 000 cells. Our work provides at least 4 novel advances that will enhance our understanding of cardiovascular biology. First, we have developed the largest collection of single nuclear transcriptomes from the human heart to date. These robust data set allowed us to define 9 major clusters and at least 20 subclusters of cell types in the healthy heart. Second, we identified unexpected differences in chamber-, laterality-, and sex-specific transcriptional programs across major subtypes of cardiac cells. Third, we linked specific cell types to common and rare genetic variants underlying cardiovascular diseases. Last, we generated an analytic and statistical framework for handling the unique challenges of cardiac single nuclear data that will be of broad interest to the scientific community.

Previous single-cell sequencing of the heart has focused on murine models of health and disease,[4,7,37–40] with limited forays into analyses of human tissues.[5,6] Notable examples of the latter include compelling studies of fetal development and cardiomyopathy–control comparisons. The rarity of data from humans highlights the inherent technical and logistical challenges associated with these studies. Ideal tissue harvesting requires coordination between clinical and laboratory teams to quickly isolate and preserve the metabolically active, ischemia-sensitive tissue. After tissue isolation, additional challenges emerge, including problematic cell isolation protocols combined with large disparities in cell size necessitating nuclear rather than whole-cell sequencing. Furthermore, the lysis of cells for single nuclear isolation produces significant cytoplasmic RNA contamination in the form of ambient RNA, which we remove using a probabilistic model developed by our group. In human tissue, there is also significant intersample diversity such that cell alignment across samples is required for any additional cell subtype comparisons. As batch correction with the commonly used canonical correlation analysis may remove sample-specific clusters,[11] we applied a deep neural network to correct batch effects using the single-cell variational inference tool.[41] Last, the transcriptional complexity of nuclei is not equivalent between cell types, making identification of droplets containing cells versus those that are empty more challenging than typical cell-based protocols. To overcome this challenge, we called cells using our CellBender remove-background tool, which compares each droplet to the background signature of ambient RNA to identify and retain cell types with lower average transcriptional coverage.

The result of highly collaborative effort is a large-scale map of the transcriptional diversity of the human heart that is approximately 50 times larger than previous

human studies. The scope of our study afforded us the ability to interrogate rarer cell types, perform detailed cellular subclustering, and define the signatures of cell types beyond what was previously possible. Our data will be a unique resource for the cardiovascular research community and is available for further exploration at the Broad Institute's Single Cell Portal (https://portals.broadinstitute.org/single_cell). These data will facilitate the independent evaluation of the cell types we have described and provide the opportunity for reanalyses and more liberal cellular subclustering, examination of the expression of genes of interest, and additional comparisons across and within cell groups.

Beyond analyses we have presented here, we anticipate that this work will serve as a framework for further studies, both as a reference data set of human nonfailing samples and as an analytic framework for further comparisons. We were excited to read the initial studies of human disease comparisons by single-cell sequencing, and hope that the data and approach here will facilitate further comparisons of this kind in the future. As highlighted with the discovery of ionocytes based on cystic fibrosis transmembrane conductance regulator expression in patients with cystic fibrosis,[42] we hope to identify similar rare disease-specific cellular subtypes that can be used in cardiovascular disease research. Looking forward, recent advances in noncardiovascular single-cell work using LIGER[43] and Seurat v3.0[44] have highlighted the potential for multimodal integration of transcriptome and epigenome data sets. Generation of richer data sets of this nature in these samples and others will further facilitate translational discoveries, while overcoming limitations of any single data modality. We hope that this is the first entry in a larger series of large human transcriptomes to be published by our group and others. When combined, these data can facilitate analyses that require significant sample sizes, such as eQTL analyses that link risk loci to genes; these methods are just beginning to be applied to single-cell data.[45]

Limitations

Our study was subject to several potential limitations. Although this is a much larger collection of human cardiac transcriptomes than any other study to date, these samples may not reflect the complete diversity contained within nonfailing hearts. These particular samples do not address the possibility of regional transcriptional programs within the chambers, nor do they directly address the potential contribution of undiagnosed pathology. Studies to expand the number of normal and diseased tissue comparisons are ongoing, which may prove essential in interpretation of genetic risk loci. All individuals in this study were of European descent; transcriptional profiling of samples from other races and ethnicities should be a goal in the future.

Sex-based comparisons were relatively underpowered and should not be interpreted as a comprehensive assessment of sex-based transcriptional difference in the heart. Nuclear transcriptomes represent a small percentage of the total mRNA present in a cell and differ significantly from the population of species present in the cytoplasm. Follow-up studies that examine the concordance of whole cell versus nuclear transcriptomes will clarify the differences in these 2 populations of mRNA. We did not observe a subpopulation of canonically activated fibroblasts, but this was not unexpected given that the focus of the current study was on cardiac tissue from healthy donors. Methods to remove ambient RNA, identify nuclear doublets, and perform batch correction are imperfect; even after correction, droplets are expected to retain some background signal. This should be kept in mind when interpreting the data, especially when observing the expression of genes from common cell types, such as cardiomyocytes, in other cell types.

Conclusions

Single-cell RNA sequencing is a revolutionary tool for characterizing known and novel cell types and states in health and disease. We provide a large-scale map of the transcriptional and cellular diversity in the normal human heart. Our identification of discrete cell subtypes and differentially expressed genes within the heart will ultimately facilitate the development of new therapeutics for cardiovascular diseases.

Correspondence

Patrick T. Ellinor, The Broad Institute of MIT and Harvard, 75 Ames Street, Cambridge, MA 02142. Email ellinor@mgh.harvard.edu

Affiliations

Precision Cardiology Laboratory (N.R.T., M.C., S.J.F., A.W.H., A.-D.A., C.N.H., A.A., I.P., C.R., S.H.C., M.B., C.M.S., P.T.E.) and Data Sciences Platform (S.J.F., M.B.), The Broad Institute of MIT and Harvard (F.A., K.G.A.), Cambridge, MA. Cardiovascular Research Center, Massachusetts General Hospital, Boston (N.R.T., A.W.H., V.A.P., P.T.E.). Masonic Medical Research Institute, Utica, NY (N.R.T.). Perelman School of Medicine, University of Pennsylvania, Philadelphia (K.C.B., K.B.M.). Precision Cardiology Laboratory, Bayer US LLC, Cambridge, MA (A.-D.A., I.P., C.M.S.). University Medical Center Groningen, University of Groningen, the Netherlands (C.R.).

Supplemental Materials

Expanded Methods
Data Supplement Figures I–VI
Data Supplement Tables I–VIII
Data Supplement Excel File I

REFERENCES

1. Macosko EZ, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M, Tirosh I, Bialas AR, Kamitaki N, Martersteck EM, et al. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell.* 2015;161:1202–1214. doi: 10.1016/j.cell.2015.05.002

2. Zheng GX, Terry JM, Belgrader P, Ryvkin P, Bent ZW, Wilson R, Ziraldo SB, Wheeler TD, McDermott GP, Zhu J, et al. Massively parallel digital transcriptional profiling of single cells. *Nat Commun.* 2017;8:14049. doi: 10.1038/ncomms14049

3. Regev A, Teichmann SA, Lander ES, Amit I, Benoist C, Birney E, Bodenmiller B, Campbell P, Carninci P, Clatworthy M, et al. The human cell atlas. *Elife.* 2017;6. doi: 10.7554/eLife.27041

4. Schaum N, Karkanias J, Neff NF, May AP, Quake SR, Wyss-Coray T, Darmanis S, Batson J, Botvinnik O, Chen MB, et al. Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. *Nature.* 2018;562:367–372. doi: 10.1038/s41586-018-0590-4

5. Cui Y, Zheng Y, Liu X, Yan L, Fan X, Yong J, Hu Y, Dong J, Li Q, Wu X, et al. Single-cell transcriptome analysis maps the developmental track of the human heart. *Cell Rep.* 2019;26:1934–1950.e5. doi: 10.1016/j.celrep.2019.01.079

6. Gladka MM, Molenaar B, de Ruiter H, van der Elst S, Tsui H, Versteeg D, Lacraz GPA, Huibers MMH, van Oudenaarden A, van Rooij E. Single-cell sequencing of the healthy and diseased heart reveals cytoskeleton-associated protein 4 as a new modulator of fibroblasts activation. *Circulation.* 2018;138:166–180. doi: 10.1161/CIRCULATIONAHA.117.030742

7. Skelly DA, Squiers GT, McLellan MA, Bolisetty MT, Robson P, Rosenthal NA, Pinto AR. Single-cell transcriptional profiling reveals cellular diversity and intercommunication in the mouse heart. *Cell Rep.* 2018;22:600–610. doi: 10.1016/j.celrep.2017.12.072

8. Chen CY, Caporizzo MA, Bedi K, Vite A, Bogush AI, Robison P, Heffler JG, Salomon AK, Kelly NA, Babu A, et al. Suppression of detyrosinated microtubules improves cardiomyocyte function in human heart failure. *Nat Med.* 2018;24:1225–1233. doi: 10.1038/s41591-018-0046-2

9. Dipla K, Mattiello JA, Jeevanandam V, Houser SR, Margulies KB. Myocyte recovery after mechanical circulatory support in humans with end-stage heart failure. *Circulation.* 1998;97:2316–2322. doi: 10.1161/01.cir.97.23.2316

10. Wolf FA, Angerer P, Theis FJ. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* 2018;19:15. doi: 10.1186/s13059-017-1382-0

11. Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol.* 2018;36:411–420. doi: 10.1038/nbt.4096

12. Bates D, Mächler M, Bolker BM, Walker SC. Fitting linear mixed-effects models using lme4. *J Stat Softw.* 2015;67. doi: 10.18637/jss.v067.i01

13. Chen W, Li Y, Easton J, Finkelstein D, Wu G, Chen X. UMI-count modeling and differential expression analysis for single-cell RNA sequencing. *Genome Biol.* 2018;19:70. doi: 10.1186/s13059-018-1438-9

14. Arimura T, Bos JM, Sato A, Kubo T, Okamoto H, Nishi H, Harada H, Koga Y, Moulik M, Doi YL, et al. Cardiac ankyrin repeat protein gene (ANKRD1) mutations in hypertrophic cardiomyopathy. *J Am Coll Cardiol.* 2009;54:334–342. doi: 10.1016/j.jacc.2008.12.082

15. Ye J, Wang Z, Wang M, Xu Y, Zeng T, Ye D, Liu J, Jiang H, Lin Y, Wan J. Increased kielin/chordin-like protein levels are associated with the severity of heart failure. *Clin Chim Acta.* 2018;486:381–386. doi: 10.1016/j.cca.2018.08.033

16. Vistnes M, Aronsen JM, Lunde IG, Sjaastad I, Carlson CR, Christensen G. Pentosan polysulfate decreases myocardial expression of the extracellular matrix enzyme ADAMTS4 and improves cardiac function in vivo in rats subjected to pressure overload by aortic banding. *PLoS One.* 2014;9:e89621. doi: 10.1371/journal.pone.0089621

17. Batlle M, Castillo N, Alcarraz A, Sarvari S, Sangüesa G, Cristóbal H, García de Frutos P, Sitges M, Mont L, Guasch E. Axl expression is increased in early stages of left ventricular remodeling in an animal model with pressure-overload. *PLoS One.* 2019;14:e0217926. doi: 10.1371/journal.pone.0217926

18. Fu X, Khalil H, Kanisicak O, Boyer JG, Vagnozzi RJ, Maliken BD, Sargent MA, Prasad V, Valiente-Alandi I, Blaxall BC, et al. Specialized fibroblast differentiated states underlie scar formation in the infarcted mouse heart. *J Clin Invest.* 2018;128:2127–2143. doi:10.1172/JCI98215

19. Aghajanian H, Kimura T, Rurik JG, Hancock AS, Leibowitz MS, Li L, Scholler J, Monslow J, Lo A, Han W, et al. Targeting cardiac fibrosis with engineered T cells. *Nature.* 2019;573:430–433. doi: 10.1038/s41586-019-1546-z

20. van Berlo JH, Kanisicak O, Maillet M, Vagnozzi RJ, Karch J, Lin SC, Middleton RC, Marbán E, Molkentin JD. c-Kit+ cells minimally contribute cardiomyocytes to the heart. *Nature.* 2014;509:337–341. doi: 10.1038/nature13309

21. Maliken BD, Kanisicak O, Karch J, Khalil H, Fu X, Boyer JG, Prasad V, Zheng Y, Molkentin JD. Gata4-dependent differentiation of c-Kit+-derived endothelial cells underlies artefactual cardiomyocyte regeneration in the heart. *Circulation.* 2018;138:1012–1024. doi: 10.1161/CIRCULATIONAHA.118.033703

22. Tang J, Zhang H, He L, Huang X, Li Y, Pu W, Yu W, Zhang L, Cai D, Lui KO, et al. Genetic fate mapping defines the vascular potential of endocardial cells in the adult heart. *Circ Res.* 2018;122:984–993. doi: 10.1161/CIRCRESAHA.117.312354

23. Corradi D, Maestri R, Callegari S, Pastori P, Goldoni M, Luong TV, Bordi C. The ventricular epicardial fat is related to the myocardial mass in normal, ischemic and hypertrophic hearts. *Cardiovasc Pathol.* 2004;13:313–316. doi: 10.1016/j.carpath.2004.08.005

24. Lewitt MS, Dent MS, Hall K. The insulin-like growth factor system in obesity, insulin resistance and type 2 diabetes mellitus. *J Clin Med.* 2014;3:1561–1574. doi: 10.3390/jcm3041561

25. Gorter JA, Zurolo E, Iyer A, Fluiter K, van Vliet EA, Baayen JC, Aronica E. Induction of sodium channel Na(x) (SCN7A) expression in rat and human hippocampus in temporal lobe epilepsy. *Epilepsia.* 2010;51:1791–1800. doi: 10.1111/j.1528-1167.2010.02678.x

26. Gudbjartsson DF, Arnar DO, Helgadottir A, Gretarsdottir S, Holm H, Sigurdsson A, Jonasdottir A, Baker A, Thorleifsson G, Kristjansson K, et al. Variants conferring risk of atrial fibrillation on chromosome 4q25. *Nature.* 2007;448:353–357. doi: 10.1038/nature06007

27. Wang M, Gong Q, Zhang J, Chen L, Zhang Z, Lu L, Yu D, Han Y, Zhang D, Chen P, et al. Characterization of gene expression profiles in HBV-related liver fibrosis patients and identification of ITGBL1 as a key regulator of fibrogenesis. *Sci Rep.* 2017;7. doi: 10.1038/srep43446

28. Zhang CL, Zhao Q, Liang H, Qiao X, Wang JY, Wu D, Wu LL, Li L. Cartilage intermediate layer protein-1 alleviates pressure overload-induced cardiac fibrosis via interfering TGF-β1 signaling. *J Mol Cell Cardiol.* 2018;116:135–144. doi: 10.1016/j.yjmcc.2018.02.006

29. Yoo JN, Shin H, Kim TH, Choi WS, Ferguson SD, Fazleabas AT, Young SL, Lessey BA, Ha UH, Jeong JW. CRISPLD2 is a target of progesterone receptor and its expression is decreased in women with endometriosis. *PLoS One.* 2014;9:e100481. doi:10.1371/journal.pone.0100481

30. Barre L, Fournel-Gigleux S, Finel M, Netter P, Magdalou J, Ouzzine M. Substrate specificity of the human UDP-glucuronosyltransferase UGT2B4 and UGT2B7: identification of a critical aromatic amino acid residue at position 33. *FEBS J.* 2007;274:1256–1264. doi: 10.1111/j.1742-4658.2007.05670.x

31. Verweij N, Eppinga RN, Hagemeijer Y, van der Harst P. Identification of 15 novel risk loci for coronary artery disease and genetic risk of recurrent events, atrial fibrillation and heart failure. *Sci Rep.* 2017;7:2761. doi: 10.1038/s41598-017-03062-8

32. Aguet F, Brown AA, Castel SE, Davis JR, He Y, Jo B, Mohammadi P, Park YS, Parsana P, Segrè AV, et al. Genetic effects on gene expression across human tissues. *Nature.* 2017;550:204–213. doi: 10.1038/nature24277

33. Roselli C, Chaffin MD, Weng LC, Aeschbacher S, Ahlberg G, Albert CM, Almgren P, Alonso A, Anderson CD, Aragam KG, et al. Multi-ethnic genome-wide association study for atrial fibrillation. *Nat Genet.* 2018;50:1225–1233. doi: 10.1038/s41588-018-0133-9

34. Finan C, Gaulton A, Kruger FA, Lumbers RT, Shah T, Engmann J, Galver L, Kelley R, Karlsson A, Santos R, et al. The druggable genome and support for target identification and validation in drug development. *Sci Transl Med*. 2017;9. doi: 10.1126/scitranslmed.aag1166

35. Chen X. TTD: therapeutic target database. *Nucleic Acids Res*. 2002;30:412–415. doi: 10.1093/nar/30.1.412

36. Ko Y, Morita K, Nagahori R, Kinouchi K, Shinohara G, Kagawa H, Hashimoto K. Myocardial cyclic AMP augmentation with high-dose PDEIII inhibitor in terminal warm blood cardioplegia. *Ann Thorac Cardiovasc Surg*. 2009;15:311–317.

37. See K, Tan WLW, Lim EH, Tiang Z, Lee LT, Li PYQ, Luu TDA, Ackers-Johnson M, Foo RS. Single cardiomyocyte nuclear transcriptomes reveal a lincRNA-regulated de-differentiation and cell cycle stress-response in vivo. *Nat Commun*. 2017;8:225. doi: 10.1038/s41467-017-00319-8

38. Lescroart F, Wang X, Lin X, Swedlund B, Gargouri S, Sànchez-Dànes A, Moignard V, Dubois C, Paulissen C, Kinston S, et al. Defining the earliest step of cardiovascular lineage segregation by single-cell RNA-seq. *Science*. 2018;359:1177–1181. doi: 10.1126/science.aao4174

39. Hu P, Liu J, Zhao J, Wilkins BJ, Lupino K, Wu H, Pei L. Single-nucleus transcriptomic survey of cell diversity and functional maturation in postnatal mammalian hearts. *Genes Dev*. 2018;32:1344–1357. doi: 10.1101/gad.316802.118

40. Nomura S, Satoh M, Fujita T, Higo T, Sumida T, Ko T, Yamaguchi T, Tobita T, Naito AT, Ito M, et al. Cardiomyocyte gene programs encoding morphological and functional signatures in cardiac hypertrophy and failure. *Nat Commun*. 2018;9:4435. doi: 10.1038/s41467-018-06639-7

41. Lopez R, Regier J, Cole MB, Jordan MI, Yosef N. Deep generative modeling for single-cell transcriptomics. *Nat Methods*. 2018;15:1053–1058. doi: 10.1038/s41592-018-0229-2

42. Montoro DT, Haber AL, Biton M, Vinarsky V, Lin B, Birket SE, Yuan F, Chen S, Leung HM, Villoria J, et al. A revised airway epithelial hierarchy includes CFTR-expressing ionocytes. *Nature*. 2018;560:319–324. doi: 10.1038/s41586-018-0393-7

43. Welch JD, Kozareva V, Ferreira A, Vanderburg C, Martin C, Macosko EZ. Single-cell multi-omic integration compares and contrasts features of brain cell identity. *Cell*. 2019;177:1873–1887.e17. doi: 10.1016/j.cell.2019.05.006

44. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM III, Hao Y, Stoeckius M, Smibert P, Satija R. Comprehensive integration of single-cell data. *Cell*. 2019;177:1888–1902.e21. doi: 10.1016/j.cell.2019.05.031

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

46. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010;20:1297–1303. doi: 10.1101/gr.107524.110

47. Becht E, McInnes L, Healy J, Dutertre CA, Kwok IWH, Ng LG, Ginhoux F, Newell EW. Dimensionality reduction for visualizing single-cell data using UMAP. *Nat Biotechnol*. 2019;37:38–47. doi: 10.1038/nbt.4314

48. Boufaied N, Nash C, Rochette A, Smith A, Orr B, Grace OC, Wang YC, Badescu D, Ragoussis J, Thomson AA. Identification of genes expressed in a mesenchymal subset regulating prostate organogenesis using tissue and single cell transcriptomics. *Sci Rep*. 2017;7:16385. doi: 10.1038/s41598-017-16685-8

49. Villani AC, Satija R, Reynolds G, Sarkizova S, Shekhar K, Fletcher J, Griesbeck M, Butler A, Zheng S, Lazo S, et al. Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. *Science*. 2017;356:eaah4573. doi: 10.1126/science.aah4573

50. Eisenberg MJ. Accuracy and predictive values in clinical decision-making. *Cleve Clin J Med*. 1995;62:311–316. doi: 10.3949/ccjm.62.5.311

51. Heston TF. Standardizing predictive values in diagnostic imaging research. *J Magn Reson Imaging*. 2011;33:505; author reply 506–505; author reply 507. doi: 10.1002/jmri.22466

52. Mooney MA, Wilmot B. Gene set analysis: a step-by-step guide. *Am J Med Genet B Neuropsychiatr Genet*. 2015;168:517–527. doi: 10.1002/ajmg.b.32328

53. Bulik-Sullivan B, Loh PR, Finucane HK, Ripke S, Yang J, Patterson N, Daly MJ, Price AL, Neale BM, Corvin A, et al. LD score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet*. 2015;47:291–295. doi: 10.1038/ng.3211

54. Finucane HK, Reshef YA, Anttila V, Slowikowski K, Gusev A, Byrnes A, Gazal S, Loh PR, Lareau C, Shoresh N, et al; Brainstorm Consortium. Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nat Genet*. 2018;50:621–629. doi: 10.1038/s41588-018-0081-4

55. Kuhn RM, Haussler D, Kent WJ. The UCSC genome browser and associated tools. *Brief Bioinform*. 2013;14:144–161. doi: 10.1093/bib/bbs038

56. Finucane HK, Bulik-Sullivan B, Gusev A, Trynka G, Reshef Y, Loh PR, Anttila V, Xu H, Zang C, Farh K, et al; 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*. 2015;47:1228–1235. doi: 10.1038/ng.3404

57. van Setten J, Brody JA, Jamshidi Y, Swenson BR, Butler AM, Campbell H, Del Greco FM, Evans DS, Gibson Q, Gudbjartsson DF, et al. PR interval genome-wide association meta-analysis identifies 50 loci associated with atrial and atrioventricular electrical activity. *Nat Commun*. 2018;9:2904. doi: 10.1038/s41467-018-04766-9

58. Arking DE, Pulit SL, Crotti L, van der Harst P, Munroe PB, Koopmann TT, Sotoodehnia N, Rossin EJ, Morley M, Wang X, et al; CARe Consortium; COGENT Consortium; DCCT/EDIC; eMERGE Consortium; HRGEN Consortium. Genetic association study of QT interval highlights role for calcium signaling pathways in myocardial repolarization. *Nat Genet*. 2014;46:826–836. doi: 10.1038/ng.3014

59. van der Harst P, Verweij N. Identification of 64 novel genetic loci provides an expanded view on the genetic architecture of coronary artery disease. *Circ Res*. 2018;122:433–443. doi: 10.1161/CIRCRESAHA.117.312086

60. Willer CJ, Schmidt EM, Sengupta S, Peloso GM, Gustafsson S, Kanoni S, Ganna A, Chen J, Buchkovich ML, Mora S, et al; Global Lipids Genetics Consortium. Discovery and refinement of loci associated with lipid levels. *Nat Genet*. 2013;45:1274–1283. doi: 10.1038/ng.2797

61. Mahajan A, Taliun D, Thurner M, Robertson NR, Torres JM, Rayner NW, Payne AJ, Steinthorsdottir V, Scott RA, Grarup N, et al. Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. *Nat Genet*. 2018;50:1505–1513. doi: 10.1038/s41588-018-0241-6

62. Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C, Plagnol V. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet*. 2014;10:e1004383. doi: 10.1371/journal.pgen.1004383

63. Auton A, Abecasis GR, Altshuler DM, Durbin RM, Bentley DR, Chakravarti A, Clark AG, Donnelly P, Eichler EE, Flicek P, et al. A global reference for human genetic variation. *Nature*. 2015. doi: 10.1038/nature15393

64. Fleming SJ, Marioni JC, Babadi M. CellBender remove-background: a deep generative model for unsupervised removal of background noise from scRNA-seq datasets. *BioRxiv*. 2019:791699. doi: 10.1101/791699

65. Falcon S, Gentleman R. Using GOstats to test gene lists for GO term association. *Bioinformatics*. 2007;23:257–258. doi: 10.1093/bioinformatics/btl567