Molecular convergence of risk variants for congenital heart defects leveraging a regulatory map of the human fetal heart

X. Rosa Ma[1,2,*], Stephanie D. Conley[1,2,*], Michael Kosicki[3], Danila Bredikhin[1,2], Ran Cui[4,5,6,7], Steven Tran[8], Maya U. Sheth[1,2,9], Wei-Lin Qiu[7,10], Sijie Chen[1,2,11], Soumya Kundu[2,11], Helen Y. Kang[1,2,12], Dulguun Amgalan[1,2,13], Chad J. Munger[1,2], Lauren Duan[8], Katherine Dang[8], Oriane Matthys Rubio[1,8], Shinwan Kany[14,15,16,17], Siavash Zamirpour[18], John DePaolo[19], Arun Padmanabhan[20,21,22], Birth Defects Research Laboratory[23], Jeffrey Olgin[24], Scott Damrauer[25,26,27,28], Robin Andersson[7,10], Mingxia Gu[29,30], James R. Priest[8,31], Thomas Quertermous[2,32,33], Xiaojie Qiu[1,2,11,13,33,34], Marlene Rabinovitch[1,8,33,35], Axel Visel[3,36,37], Len Pennacchio[3,36,38], Anshul Kundaje[2,11], Ian A. Glass[13,39], Casey A. Gifford[1,2,8,13,33,40], James P. Pirruccello[14,24,41,42], William R. Goodyer[8,13,33,†], Jesse M. Engreitz[1,2,7,13,33,43,†]

1 Basic Science and Engineering (BASE) Initiative, Stanford Children's Health, Betty Irene Moore Children's Heart Center, Stanford, CA, USA
2 Department of Genetics, Stanford University, Stanford, CA, USA
3 Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA
4 Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, USA
5 Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, USA
6 Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA, USA
7 The Novo Nordisk Foundation Center for Genomic Mechanisms of Disease, Broad Institute of MIT and Harvard, Cambridge, MA, USA
8 Department of Pediatrics, Stanford University, Stanford, CA, USA
9 Department of Bioengineering, Stanford University, Stanford, CA, USA
10 Department of Biology, University of Copenhagen, Copenhagen, Denmark
11 Department of Computer Science, Stanford University, Stanford, CA, USA
12 Current address: PhD Program in Computational and Systems Biology, MIT, Cambridge, MA, USA
13 Maternal and Child Health Research Institute, Stanford University, Stanford, CA, USA
14 Cardiovascular Disease Initiative, Broad Institute of MIT and Harvard, Cambridge, MA, USA
15 Cardiovascular Research Center, Massachusetts General Hospital, Boston, MA, USA
16 Department of Cardiology, University Heart and Vascular Center Hamburg-Eppendorf, Hamburg, Germany
17 German Center for Cardiovascular Research (DZHK), Partner Site Hamburg/Kiel/Lübeck, Hamburg, Germany
18 School of Medicine, University of California San Francisco, San Francisco, CA, USA
19 Department of Surgery, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA
20 Gladstone Institutes, San Francisco, CA, USA
21 Department of Medicine, University of California San Francisco School of Medicine, San Francisco, CA, USA
22 Chan Zuckerberg Biohub, San Francisco, CA, USA
23 Birth Defects Research Laboratory, University of Washington, Seattle, WA, USA
24 Division of Cardiology, Department of Medicine and Cardiovascular Research Institute, University of California San Francisco, San Francisco, CA, USA
25 Division of Vascular Surgery and Endovascular Therapy, Department of Surgery, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA
26 Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA
27 Corporal Michael Crescenz VA Medical Center, Philadelphia, PA, USA
28 Penn Cardiovascular Institute, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA
29 Center for Stem Cell and Organoid Medicine, Division of Pulmonary Biology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA
30 University of Cincinnati School of Medicine, Cincinnati, OH, USA
31 Tenaya Therapeutics, South San Francisco, CA, USA
32 Department of Medicine, Stanford University, Stanford, CA, USA
33 Stanford Cardiovascular Institute, Stanford University, Stanford, CA, USA
34 Stanford Bio-X, Stanford University, Stanford, CA, USA
35 Vera Moulton Wall Center for Pulmonary Vascular Diseases, Stanford University, Stanford, CA, USA
36 US Department of Energy Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, CA, USA
37 School of Natural Sciences, University of California, Merced, Merced, CA, USA
38 Comparative Biochemistry Program, University of California, Berkeley, CA, 94720, USA
39 Department of Pediatrics and Medicine, University of Washington School of Medicine, Seattle, WA, USA
40 Institute for Stem Cell Biology and Regenerative Medicine, Stanford University, Stanford, CA, USA
41 Institute for Human Genetics, University of California, San Francisco, CA, USA
42 Bakar Computation Health Sciences Institute, University of California, San Francisco, CA, USA
43 Gene Regulation Observatory, Broad Institute of MIT and Harvard, Cambridge, MA, USA
* Co-first authors.
† Co-last authors. Correspondence to wgoodyer@stanford.edu and engreitz@stanford.edu

Abstract

Congenital heart defects (CHD) arise in part due to inherited genetic variants that alter genes and noncoding regulatory elements in the human genome. These variants are thought to act during fetal development to influence the formation of different heart structures. However, identifying the genes, pathways, and cell types that mediate these effects has been challenging due to the immense diversity of cell types involved in heart development as well as the superimposed complexities of interpreting noncoding sequences. As such, understanding the molecular functions of both noncoding and coding variants remains paramount to our fundamental understanding of cardiac development and CHD. Here, we created a gene regulation map of the healthy human fetal heart across developmental time, and applied it to interpret the functions of variants associated with CHD and quantitative cardiac traits. We collected single-cell multiomic data from 734,000 single cells sampled from 41 fetal hearts spanning post-conception weeks 6 to 22, enabling the construction of gene regulation maps in 90 cardiac cell types and states, including rare populations of cardiac conduction cells. Through an unbiased analysis of all 90 cell types, we find that both rare coding variants associated with CHD and common noncoding variants associated with valve traits converge to affect valvular interstitial cells (VICs). VICs are enriched for high expression of known CHD genes previously identified through mapping of rare coding variants. Eight CHD genes, as well as other genes in similar molecular pathways, are linked to common noncoding variants associated with other valve diseases or traits via enhancers in VICs. In addition, certain common noncoding variants impact enhancers with activities highly specific to particular subanatomic structures in the heart, illuminating how such variants can impact specific aspects of heart structure and function. Together, these results implicate new enhancers, genes, and cell types in the genetic etiology of CHD, identify molecular convergence of common noncoding and rare coding variants on VICs, and suggest a more expansive view of the cell types instrumental in genetic risk for CHD, beyond the working cardiomyocyte. This regulatory map of the human fetal heart will provide a foundational resource for understanding cardiac development, interpreting genetic variants associated with heart disease, and discovering targets for cell-type specific therapies.

Introduction

The formation of the human heart during fetal development is driven by the emergence and coordination of a vast array of different cell types[1–3]. Each of these cell types carries out distinct architectural or regulatory functions in forming the heart chambers, valves, conduction system, and blood vessels. These dynamic processes are encoded by complex regulatory programs in the human genome, involving different combinations of genes regulated by hundreds of thousands of noncoding regulatory elements and transcription factor (TFs) binding sites that are yet to be fully identified[4].

DNA variants that disrupt these regulatory programs can lead to congenital heart defects (CHD)[2,5], which are the most common birth defect and affect nearly 1% of live births[6]. Genetic studies of patients with CHD have now discovered thousands of variants in both protein-coding and noncoding regions of the genome[7,8]. Studies of such coding variants have revealed important insights into the molecular regulators of heart development, and identified hundreds of high-confidence CHD genes that currently help to guide genetic diagnosis[1,5,7–11]. Noncoding variants are far more difficult to interpret, but some have recently been associated with CHD through genome-wide association studies (GWAS)[12–15] or identified through molecular analysis of *de novo* variants[4,16–18].

Despite this progress, the genetic contributions to most CHD cases remain to be identified[2,5], and key questions about the genes, pathways and cell types, that contribute to CHD remain to be addressed:

(1) Which cell types contribute to the genetic etiology of different subtypes of CHD? CHD is a heterogeneous class of diseases, with subtypes involving different subanatomic structures of the heart, such as valve defects and malformations of the outflow tract. Early studies focused on the heart muscle cell, or cardiomyocyte, following the identification of CHD-causing mutations in transcription factors that affect myocardial specification and/or function, such as *NKX2-5*, *TBX5*, and *GATA4*[1,2,19–22]. We now appreciate that other cell types are likely to have roles in the genetic etiology of CHD, such as contributions of the cardiac neural crest lineage to malformations of the outflow tract[23,24] and an emerging role for endocardial cells in hypoplastic left heart syndrome[25]. The development of single-cell technologies, and our accompanying appreciation of the vast diversity of cell types and states in the human heart, provides a new opportunity to assess which cell types contribute to the genetic etiology of which subtypes of human CHD.

(2) How do inherited noncoding variants influence risk for CHD? A primary challenge in interpreting the functions of noncoding variants has been the lack of appropriate regulatory maps to connect these variants to their target genes and cell types. Previous studies have proposed roles for *de novo* noncoding variants linked to CHD in cardiomyocytes or endothelial cells, and identified a handful of individual variants that affect enhancer activity or alternative RNA splicing[4,16–18]. More recently, GWAS have identified inherited noncoding variants associated with CHD[12–15,26–28] and related quantitative measurements of cardiac morphology[29–38]. These recent genetic datasets provide a new opportunity to learn about genetic modifiers of CHD risk, severity, or subtype, if we could identify which variants regulate which enhancers and target genes in which of the many cell types in the developing heart.

To answer these and other outstanding questions in CHD research, we require a map of gene regulation that captures the diverse cell types and cell states that drive human fetal cardiogenesis. Recent studies have begun to collect data on gene expression or chromatin accessibility during heart development in mouse models[39–43] or from human fetal tissue[4,44–51]. However, existing datasets are still limited in cell number, sequencing depth, and developmental timepoints, precluding the application of key computational models to build a gene regulation map across cell types and states. Far deeper single-cell data, together with new computational models, is needed to interpret the functions of genetic variants associated with heart structure and function.

Here, we create such a gene regulation map of the human fetal heart, and combine it with recent human genetics datasets to reveal new insights into the genetic, molecular, and cellular etiology of CHD. First, we collect 10x Multiome (scRNA and scATAC-seq) data from 734,000 cardiac cells from 41 structurally normal hearts between 6 and 22 weeks of human fetal development. This cell atlas significantly surpasses existing datasets, both human and mouse, in terms of cell number, sequencing depth, and the developmental windows captured. We identify 90 cardiac cell types and cell states with sufficient depth, and, in each, apply predictive models to identify thousands of novel enhancers, TF binding sites, and downstream target genes. By applying this map to interpret coding and noncoding variants associated with CHD and quantitative traits, we find that high-confidence genes harboring rare coding variants in CHD patients are most strongly enriched for high expression in VICs and cardiac fibroblasts—two cell types whose roles in CHD have been underexplored. We nominate and experimentally validate functional noncoding variants that impact enhancers with restricted activities specific to particular cell types or locations in the heart. Certain genes and pathways are affected by both rare coding variants associated with CHD and common noncoding variants associated with structural traits, revealing molecular convergence on a pathway related to cell migration in VICs. Together, these results identify new cell types and pathways important for genetic regulation of heart development, and provide a framework for interpreting polygenic risk of CHD and other heart diseases. This gene regulation map will provide a foundational resource for understanding human heart development and interpreting the functions of genetic variants discovered in the genomes of patients with heart disease.

Results

A gene regulation map of the human fetal heart

We collected single-cell RNA and ATAC-seq data (10x Multiome) from 41 fetal heart samples spanning post-conception weeks 6 to 22 (**Fig. 1A; Fig. S1; Table S1**). These samples included 20 hearts from post-conception weeks (PCW) 6 to 9, a developmental stage that coincides with key events such as the formation of the outflow tract, valves, and conduction system. For PCW 6 through 9, we sampled cells from the entire heart (**Methods**), and profiled sections from the larger PCW 10 through 22 samples (**Table S2**). After quality control and filtering, we obtained a total of 734,000 single cells (range: 370-94,555 per heart), with an average of 1,682 RNA transcripts and 5,953 ATAC fragments per cell. Together, this atlas has a considerably higher cell count and sequencing depth compared to existing datasets.

Clustering and annotation yielded 13 major cell types (**Fig. 1C; Fig. S2)**, which were further annotated to a total of 101 cell populations (hereafter, "cell types") (**Fig. 1D**). We labeled cell types using canonical cell type markers as well as genes identified in recent single-cell and spatial transcriptomic studies[40,41,47,48,52] (see **Methods; Fig. 1E; Figs. S3-7**). Our atlas captures rare cell populations including neural crest cells and conduction cells, and assigns 42 cell types to likely subanatomic locations, for example distinguishing endocardial cells of the atrial septum and inflow and outflow sides of valves based on marker genes from recent spatial datasets[48] (**Fig. S4**).

Leveraging this deep dataset, we constructed a gene regulation map describing the relationships between genetic variants, TFs, enhancers, target genes, and gene programs (**Fig. 1B**). To do so, we applied a suite of predictive models, some of which we have previously developed and applied with the ENCODE Project and IGVF Consortium[53]. Importantly, these models require a depth of sequencing that is not available in previous datasets (minimum 3 million ATAC fragments per cell type)[54,55]. Here, we were able to construct cell type-specific gene regulation maps in 90 cell types, including cell populations with frequencies in our dataset as low as 0.06% (**Fig. S8**).

Here, we briefly describe each component of this gene regulation map, and in subsequent sections combine all components in integrative analyses:

Linking enhancers to target genes: To identify candidate enhancers and link them to their target genes in specific cell types, we applied scE2G — a supervised logistic regression model trained to predict the effects of CRISPR perturbations from single-cell multiome data[54]. We have previously shown that scE2G shows state-of-the-art performance at predicting the results of CRISPR enhancer perturbations, linking eQTL variants to their target genes, and linking GWAS variants to known target genes[54]. Applied to our fetal heart atlas, scE2G identified on average 58,460 enhancer-gene regulatory interactions per cell type (**Table S3**). As expected, closely related cell types shared more enhancers than distantly related cell types: for example, atrial and ventricular cardiomyocytes shared 80% of their enhancers, while lymphocytes and neural crest cells shared only 20% (**Fig. 1F**). Within each major cell type, we similarly observed a higher degree of enhancer overlap between cell types that are closely related along the differentiation trajectory. For example, over 60% of enhancers in pre-arterial endothelial cells 2 were also active in mature arterial endothelial cells (**Fig. 1G)**. For individual genes, scE2G identifies cell-type specific sets of enhancers that contribute to gene expression. For example, *MEF2A*, a core transcription factor expressed in many cell types in the developing heart, has distinct sets of enhancers in endothelial cells, VICs, and ventricular cardiomyocytes (VCMs) (**Fig. 1H**).

Identifying transcription factor binding sites and predicting effects of noncoding variants on accessibility: To identify candidate TF binding sites and predict the effects of noncoding variants, we applied ChromBPNet, DeepLIFT, and TF-MoDISCO[55–57] — a convolutional neural network method and accompanying interpretation tools. This framework predicts base pair-resolution chromatin accessibility data (pseudobulked per cell type) from DNA sequence, identifies TF motif instances that drive accessibility, and predicts the effects of noncoding variants on chromatin accessibility. We have shown

that this approach performs well at identifying TF binding sites[58] and predicting effects of CRISPR edits on gene expression[59]. Per cell type, we identified 19-39 important TF motifs and 340,000-760,000 high-confidence TF motif instances (**Table S15**, **Table S16**), with 516 consensus motifs combined across cell types (**Fig. 1I**). Predicted TF motifs included known associations such as SOX and ETS motifs in pre-arterial endothelial cells, and GATA motifs in VCMs (**Fig. 1J**). Individual regulatory elements often showed different TF motif instances predicted to contribute in different cell types (*e.g.*, **Fig. 1K**)[4].

Linking genes to gene programs: To uncover sets of co-expressed genes that may work together in similar biological pathways, we learned gene programs ("programs") by applying consensus non-negative matrix factorization (cNMF)[60] to the transcriptome data and annotating TFs with correlated motifs in the ATAC data (**Tables S4-6**; see **Methods**). cNMF identifies sets of genes that are co-expressed across single cells, and has been shown to learn diverse types of biologically coherent pathways, including metabolic pathways, developmental trajectories, and cell identity programs[60,61]. We applied cNMF to groups of major cell types, and identified 253 total gene programs representing a wide array of biological processes (**Fig. 1L,M**). For example, in endothelial cells, Endothelial program 23 is expressed in pre-arterial endothelial cells (**Fig. 1N**), includes genes previously observed to be expressed in arteriolar endothelial cells such as *SSUH2*, *GABBR2*, *HECW2*, and *SLC6A6*[41], and is regulated by TFs with known roles in the developing artery such as *SOX17*, *SOX13* and *MEF2A* (**Fig. 1O**). Endocardial program 33 ("Atrial septal endocardial cell identity") is exclusively and uniformly expressed in atrial septal endocardial cells, and includes known marker genes such as *LRRC4C*, *LSAMP*, *ALCAM*, and *NEGR1*. This program was predicted to be regulated by *GATA6*, a TF where loss-of-function mutations cause ASD[62], and *NFATC2*, a TF essential for atrial septum morphogenesis.[63] Endocardial program 29 ("WNT signaling in valvular endocardial cells") is expressed in a subset of inflow-side valvular endocardial cells ("IF valvular endocardial cells"), includes WNT signaling pathway genes such as *WNT4*, *WNT2*, and *WNT9B*, and is regulated by flow-responsive TFs known to regulate WNT signaling, KLF2 and KLF4[64] (**Tables S5-6**).

Together, this gene regulation map of the human fetal heart describes how DNA sequences in our genome regulate the expression of particular genes, and how those genes work together in cell-type specific pathways. This map provides a rich resource for understanding the regulatory logic of heart development across cell types and for interpreting DNA variants that affect heart structure and function.

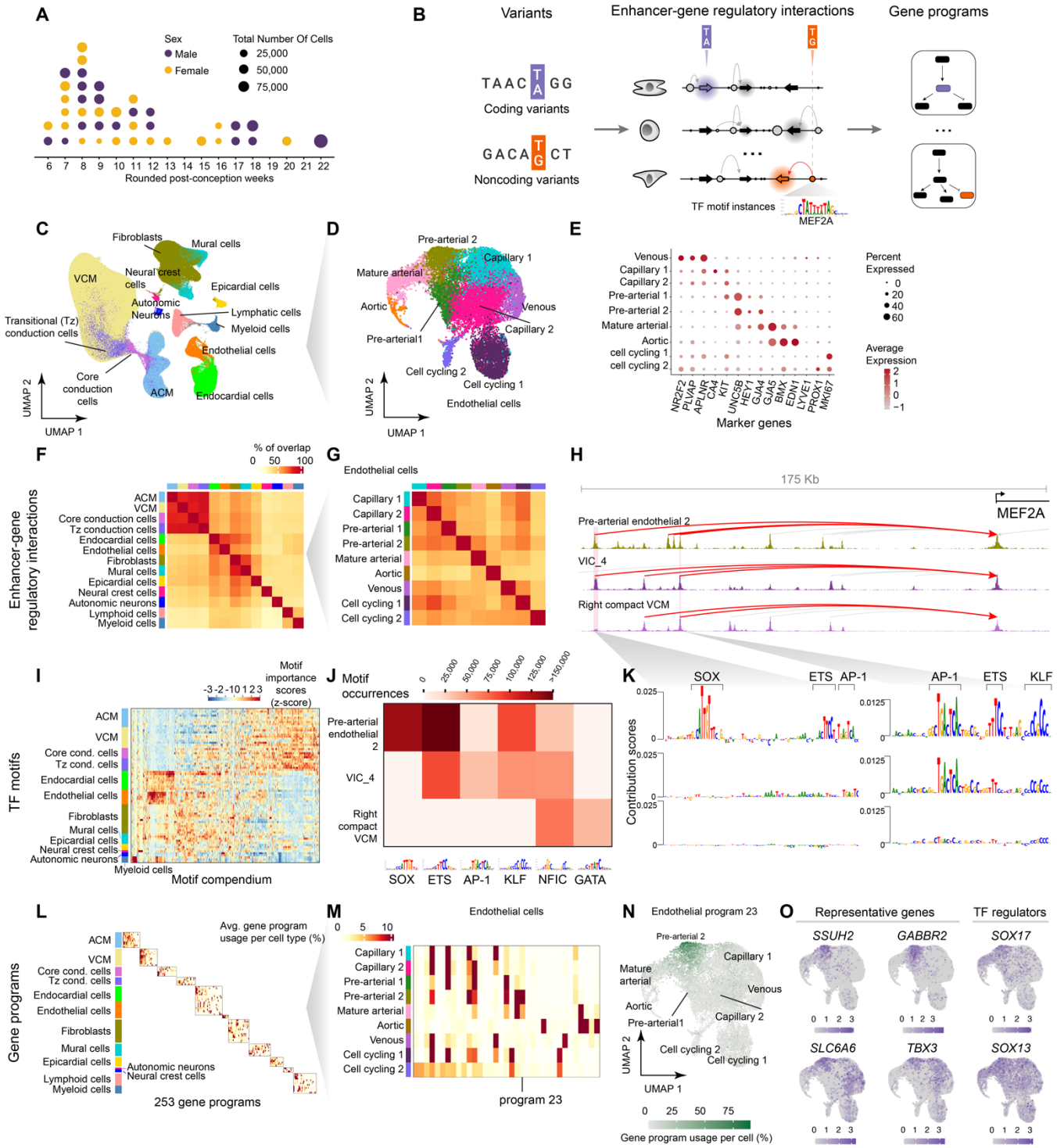


Figure 1. A gene regulation map of the human fetal heart.

A. Single-cell multiomic data (10x Multiome) was collected from 41 hearts spanning post-conception weeks 6 to 22.

B. Multiome data was used to create a gene regulation map, linking genetic variants to regulatory elements (enhancers), their target genes, and associated gene programs in each of 90 cell types with sufficient depth. The schematic illustrates the components of this map and how it can be used to interpret the molecular and cellular functions of coding (purple) and noncoding (orange) variants associated with disease. The middle panel shows cell-type-specific enhancer-gene interactions: arrows represent genes, with highlighted arrows indicating expression in a particular cell type; circles represent predicted

enhancers, with size indicating activity level, and arcs show predicted enhancer-gene interactions. Coding variants can be annotated based on the cell types in which the gene is expressed (purple). Noncoding variants can be annotated based on the cell types in which the enhancer is active, and the target gene(s) of that enhancer (orange). Right panel depicts gene programs, which are defined as sets of genes that are co-expressed across single cells that are likely to have related functions. Both coding and noncoding variants can be annotated based on how their target genes work together in gene programs (colored genes).

C. Uniform manifold approximation and projection (UMAP) embedding of scRNA-seq of all major cell types. ACM: Atrial cardiomyocytes. VCM: ventricular cardiomyocytes. Tz: transitional.

D. UMAP embedding of subset of scRNA-seq data from endothelial cells.

E. Dot plot showing the expression of marker genes of endothelial cell types. The size of the circle represents the percentage of cells within a given cell type that express the gene of interest, and the color represents the normalized expression level of that gene across all cell types shown.

F. Global similarity of predicted enhancers across major cell types. Color scale: Percentage of enhancer overlap between pairs of major cell types, with the total number of enhancers in the major cell type on the y-axis as the denominator (see **Methods**). Core cond. cells: core conduction cells. Tz cond. cells: transitional conduction cells.

G. Global similarity of predicted enhancers across endothelial cell types. Color scale: Percentage of enhancer overlap, as in F.

H. Example locus: *MEF2A* is regulated by distinct combinations of enhancers in endothelial cells, VICs, and VCMs. Signal tracks: pseudobulk scATAC-seq within each cell type. Red arcs: Elements predicted by scE2G to act as enhancers to regulate *MEF2A*. Gray arcs: Elements predicted to regulate other genes. Coordinates (hg38) of highlighted enhancers: chr15:99425061-99426373 (left), chr15:99454858-99455373 (right).

I. Compendium of TF motifs and predicted effects on chromatin accessibility across 90 cell types, as learned by ChromBPNet and TF-MoDISCO. Color scale: Z-score normalized effects of predicted motifs learned as average differences in predicted accessibility into 100 random sequences in each cell type. Core cond. cells: core conduction cells. Tz cond. cells: transitional conduction cells.

J. Predicted occurrences of selected TF motifs in cell types relevant to MEF2A enhancers (see panel K). Color scale: motif occurrences in the corresponding cell type.

K. ChromBPNet contribution scores highlight TF motif instances with cell-type-specific contributions to chromatin accessibility at two selected enhancers for MEF2A. Height of each nucleotide in the genomic sequence represents the predicted importance of that nucleotide to chromatin accessibility (by DeepLIFT). Coordinates (hg38) of the two example regions: chr15: 99425308-99425388 (left), chr15:99455071-99455122 (right).

L. Expression patterns of 253 gene programs across cell types. Gene programs were inferred using cNMF in groups of related cell types. Color scale: average gene program usage per cell type.

M. Average expression of endothelial cell programs in each of the fine-grained endothelial cell types.

N. Projection of Endothelial program 23 expression onto the endothelial cell UMAP. Color scale: gene program usage per cell.

O. Feature plots illustrating the expression of known arteriolar endothelial marker genes identified in Endothelial program 23 and the expression of predicted TF regulators of Endothelial program 23. Color scale: normalized gene expression level.

Regulatory wiring of cardiac conduction cells

The high cell number and ATAC sequencing quality allowed us to examine, for the first time, the regulatory landscape of rare but crucial cell types in the developing human heart. For example, our atlas includes 8,906 cardiac conduction cells (CCS) — specialized cell types that coordinate the heart's rhythmic contractions — and 13,103 transitional cells, which serve as a cellular bridge between conduction cells and the surrounding working myocardium, previously discovered in the murine heart[40] (**Fig. 2**; **Fig. S9**). These included cell populations from the four major substructures of the conduction system: sinoatrial node (SAN, 1,211 cells), atrioventricular node (AVN, 2795 cells), His bundle (939 cells), and Purkinje fibers (PF1 and PF2, 1,830 cells and 2,130 cells, respectively) (total cell frequencies: 0.13-0.38%; **Fig. 2A**; **Fig. S8**). Each of these populations were distinguished by established markers, such as *ISL1* and *SHOX2* for the SAN and *HS3ST3A1*[48,65] for the His bundle (**Fig. 2B,C**).

We examined our maps to identify enhancers and TFs that distinguish these CCS cell types. As expected, their global profiles of enhancer-gene regulatory interactions were distinct from other cell types (*e.g.*, 28% overlap for the SAN vs. capillary endothelial cells) (**Fig. 2D**). Among the CCS cell types, there was increased overlap in predicted enhancers within the nodes (SAN vs. AVN: 59% overlap) as compared to the ventricular conduction system (VCS) components (SAN vs. His bundle or PFs: 44-48%), consistent with distinct gene expression profiles and physiologic functions of these cell types (**Fig. 2D**). Toward understanding this molecular wiring, we examined the enhancers that regulate *ISL1*, a key TF in the development and function of the SAN (**Fig. 2E**). We identified 9 predicted enhancers for *ISL1* in SAN cells, including the one known enhancer that has been previously characterized[66]. These *ISL1* enhancers harbored predicted motif instances for other TF families involved in SAN development, including 9 for GATA family TFs, 6 for NF-Y, and 5 for TBX family TFs (**Fig. 2F**). Notably, these included several GATA motifs with stronger predicted contributions to chromatin accessibility in SAN cells versus other conduction cell types (**Fig. 2F**), and *GATA6* and *ISL1* are both predicted regulators of the gene program for SAN identity (**Fig. S10**), suggesting that *GATA6* may play a role in the up-regulation of *ISL1* observed in the SAN.

Examining the gene programs that defined each CCS cell type, we unexpectedly found many genes previously known to be involved in neural adhesion and axon development (**Fig. 2G; Fig. S11**). Neuronal cell adhesion molecules play a crucial role in mediating cell-cell interactions to achieve cell patterning and discrete cellular niches within the developing brain. Their involvement in the developing CCS, however, is not understood. By examining gene expression patterns and exploring cell-cell interactions via analysis of ligand-receptor pairs with CellChat[67] (see **Methods; Table S7**), we indeed found multiple ligand-receptor genes classically related to CNS development (*SLIT/ROBO*, *NRXN/NLGN*, *EPH/EPHRIN*) within the CCS and surrounding cell types (**Fig. 2H-M; Fig. S10B; Fig. S11**). For example, ligand *SLIT2* was expressed in PF and SAN cells, and its receptors (*ROBO1* and *ROBO2*) were expressed by surrounding cell types, including transitional SAN cells and fibroblasts (for the SAN) and VCMs (for PFs) (**Fig. 2H,K; Fig. S10B; Fig. S11A,E**). Receptors *EPHA4* and *EPHA7* were expressed in the His bundle and PF cells (**Fig. 2I; Fig. S10B; Fig. S11D,E**), and their ligand *EFNA5* was expressed in surrounding VCMs as well as transitional cells (**Fig. 2L**). Ligands NRXN1 and NRXN3 were found in AVN (**Fig. 2J; Fig. S10B; Fig. S11C**), and their receptors, including LRRTM3 and/or NLGN1, were enriched within the surrounding cells of the atrial cardiomyocytes (ACM) and VCM (**Fig. 2M**). Consistent with a possible role for these factors in CCS development, variants associated with *SLIT2* have been associated with abnormal QRS duration[68], suggestive of an abnormal VCS, and *ROBO1* variants have also been associated with abnormal EKG findings[69]. Further, systemic knockdown of *Epha4* in rats resulted in, among other phenotypes, QRS prolongation[70]. These expression patterns showed similarities but also notable differences with recent single-cell studies in the mouse fetal heart[65] and human adult CCS[71], and a study also observed upregulation of *SLIT2* and *NRXN3* in SAN and AVN, respectively, in the human fetal heart[48]. The established role of these cell adhesion molecules in neuronal differentiation and establishing cell-cell interactions hints at a similar role for the normal differentiation of the VCS and/or patterning within the developing trabecular myocardium. Together, this provides the first regulatory map of these clinically important CCS cells and illustrates the insights possible with this deep atlas.

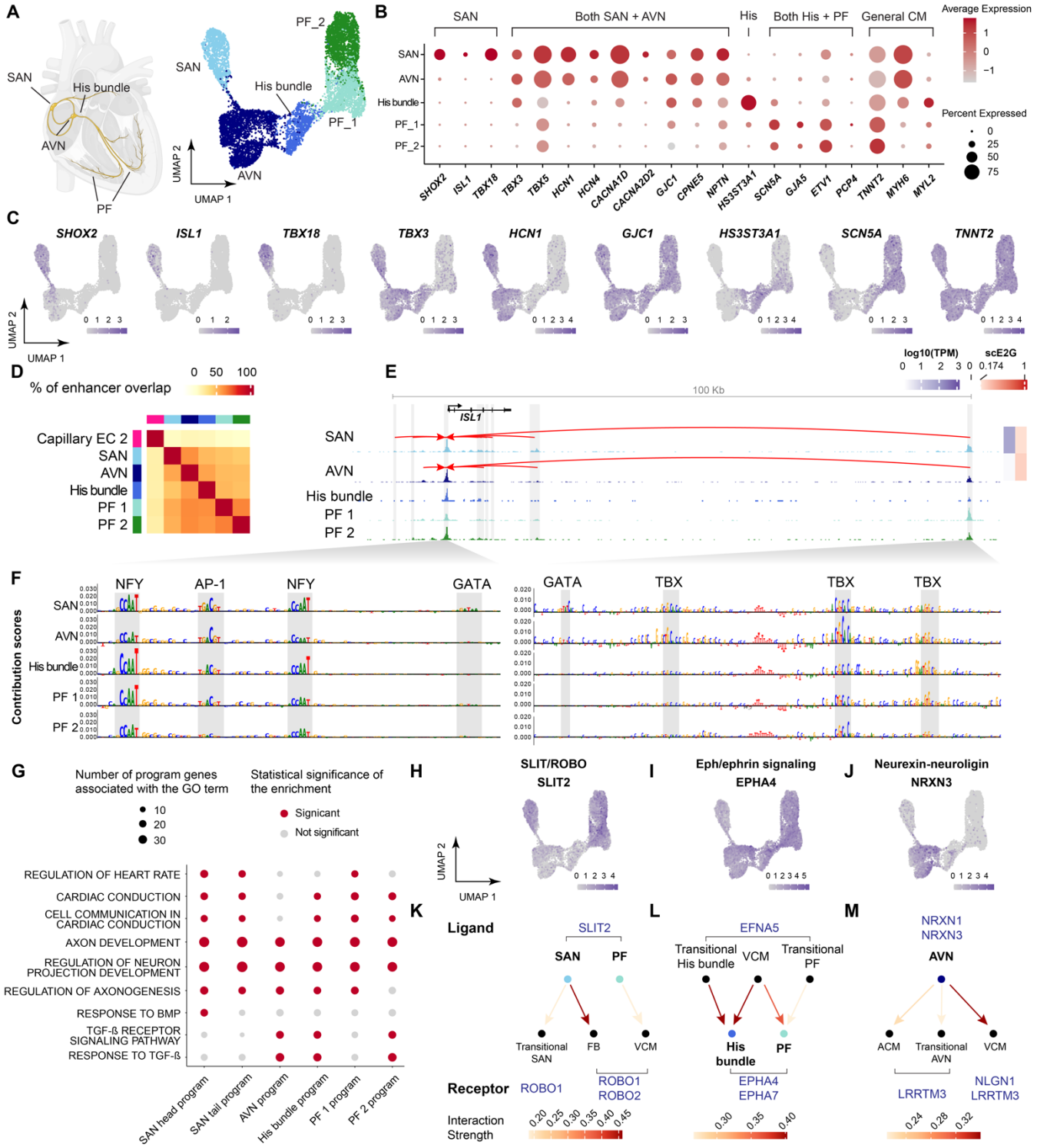


Figure 2. Regulatory wiring and neural adhesion molecules in cardiac conduction cells

A. Schematic and UMAP representations of cell types in the cardiac conduction system (CCS). SAN: Sinoatrial node. AVN: Atrioventricular node. PF: Purkinje fibers.

B. Dot plot of known marker genes for major cell types in the CCS. The size of the circle represents the percentage of cells within a given cell type that express the gene of interest, and the color represents the normalized expression level of that gene across all cell types shown.

C. Feature plots showing the expression of marker genes. Color scale: normalized gene expression level.

D. Percentage of enhancer overlap between the cell types on the y-axis and cell types on the x-axis. Includes 5 cell types in the CCS and a population of capillary endothelial cells as an outgroup for comparison. Color scale: Percentage of enhancer overlap between pairs of cell types, with the total number of enhancers in the cell type on the y-axis as the denominator (see **Methods).**

E. scE2G predictions and normalized ATAC-seq signals for *ISL1*. Predicted enhancers and the *ISL1* promoter are highlighted in shaded regions. Arcs represent the predicted enhancer-gene regulatory interactions. Side color bars show the transcripts per million (TPM) expression of the target genes in each cell type (blue) and the corresponding scE2G scores (red, and white for scores below the threshold).

F. Per-base DeepLIFT contribution score profiles of two selected regulatory regions from subpanel E. Gray highlights indicate predicted TF motif instances, including for GATA, NF-Y, and TBX family TFs. Note that GATA motif instances have higher predicted contribution scores in SAN. The left panel shows a subregion of the *ISL1* promoter (chr5:51,383,321-51,383,428). The right panel shows a subregion of an enhancer regulating *ISL1* (chr5:51,475,542-51,475,802). All coordinates in hg38.

G. Representative enriched gene ontology (GO) terms in gene programs characterizing the CCS cell types, showing enrichment for genes related to axon development across all CCS cell types.

H. I. J. Feature plot showing the expression of selected neural adhesion ligands or receptors in the CCS.

K. L. M. Predicted cell-cell interactions between CCS components and surrounding cell types. Nodes represent cell types, with CCS components highlighted in bold text and colored according to A. Arrows point from ligand-producing cells to receptor-expressing cells. Arrow color intensity reflects the interaction strength.

Expression of genes associated with congenital heart defects

We next applied this gene regulation map of the human fetal heart to interpret the functions of genetic variants associated with CHD. Hundreds of genes have been implicated in various subtypes of CHD through analysis of rare or *de novo* coding variants in families or cohort studies[2,7,8]. However, the cell types of action of many of these genes remain unknown. With our more complete atlas of the human fetal heart, we examined the expression of genes associated with CHD in an unbiased fashion and tested whether they were preferentially expressed in particular cell types and states.

As a positive control, we first analyzed known genes for other cardiovascular diseases where the critical cell types are already established (**Table S8**). As expected, genes associated with cardiomyopathies, diseases of the heart muscle tissue, were significantly enriched for high expression within cardiomyocytes (up to 94-fold). Similarly, known genes underlying increased risk for developing aortic aneurysms were enriched in smooth muscle cells, the most abundant cell type in the aortic wall[72] (up to 92-fold; **Fig. 3A; Tables S8-9**).

For CHD, we examined a curated list of 137 known genes (hereafter, "CHD genes") for 6 subtypes (ASD, ventricular septal defects (VSD), atrioventricular septal defects, functional single ventricles, malformation of the outflow tract, and valve defects) (**Table S8**), based on CHDGene[8], which compiled genes with 3 or more case reports for syndromic or isolated CHD.

Overall, CHD genes were broadly expressed in many cell types in the heart (**Fig. S12**). Whereas most genes for cardiomyopathies or aortic aneurysms were expressed only in specific cell types, more than 60% of CHD genes were expressed (transcripts per million, TPM>10) in at least one cell type in all major cardiac cell type groups (see **Methods**). This might help to explain why many CHD mutations are found associated with many different subtypes of heart defects.

Despite that many CHD genes exhibit broad expression patterns, they showed higher expression levels in some cell types than in others. We identified 20 cell types that showed significant enrichment for at least one subtype of CHD (FDR < 0.05) (**Fig. 3A; Table S9**). For example, 10 out of 98 genes known to be associated with ASDs, including *TBX5* and *MYH6,* were more highly expressed in atrial cardiomyocytes (6-fold enrichment, $q = 0.007$). Similarly, 9 out of these 98 genes, including *NR2F2* and *FOXP1*, were highly expressed in atrial septal endocardial cells (6-fold enrichment, $q = 0.041$). Genes associated with valve defects were significantly enriched in VICs (up to 9-fold enrichment, $q = 2 \times 10^{-5}$; **Fig. 3A**). Genes associated with single ventricle disease, which involve valve lesions together with other defects, were also enriched within VICs, to an even higher degree (up to 16-fold, $q = 0.02$; **Fig. 3A; Table S9**).

Thus, the patterns of expression of high-confidence CHD genes suggest that various different cell types, beyond cardiomyocytes, may play an important role in the development of different subtypes of CHD.

CHD genes in valvular interstitial cells and cardiac fibroblasts

Out of all cell types, VICs and epicardial-derived cardiac fibroblast progenitors showed the strongest enrichments for high-confidence CHD genes. These mesenchymal cell types are known from prior animal studies to have important roles in heart development (see **Note S1**), but their direct involvement in the genetic etiology of CHD in humans remains underexplored.

Mesenchymal cells in the valves are crucial for the structural integrity and function of heart valves[73]. Our atlas includes 5 such cell populations: 3 populations of VICs (VIC_1-3) recently observed to be localized to the free segments of the valves, and two populations of valve-related mesenchymal cells recently observed to be located in intervalvular fibrous tissue towards the valve roots (VIC_4 and VIC_5). VIC_4 expresses the marker of neural crest-derived mesenchymal cells, *PENK*[48] **(Fig. 3B; Fig. S5C**), and VIC_5

expresses marker genes primarily observed around the semilunar valves[48] (**Fig. 3B; Fig. S5C**). Together, VICs were significantly enriched for genes for 5 of 6 subtypes of CHD (ASD, atrioventricular septal defects, valve lesions, malformations of the outflow tract, and single ventricle disease) (5-16-fold enrichment, q <0.021; **Fig. 3A; Table S9**).

Altogether, 23 high-confidence CHD genes were preferentially expressed in VICs, of which 12 have previously been shown to lead to heart defects in mouse models through conditional knockouts consistent with functions in VICs (**Fig. 3C; Note S1; Table S10**). For example, *LTBP2* (Latent-transforming growth factor beta-binding protein 2) is the CHD gene most specifically expressed in VICs versus other cell types (**Fig. S12**), and is particularly highly expressed in VIC_4 (**Fig. 3E**). *LTBP2* encodes an extracellular matrix protein that forms part of the latent TGFβ complex and interacts with microfibrils. Loss-of-function variants in *LTBP2* have been observed in patients with bicuspid aortic valve, transposition of the great arteries, aortic stenosis, pulmonic stenosis, and mitral valve prolapse[74,75]. Consistent with these observations, *Ltbp2* knockout in mice is embryonic lethal[76], and mouse knockin of a disease-associated missense variant leads to mitral valve prolapse[77]. In our atlas, *LTBP2* is part of a VIC gene program (VIC program 27, "TGFb signaling and cell migration") that includes 8 other high-confidence CHD genes (*GATA4, MEIS2, TBX20, TBX5, TGFBR1, TLL1, HAND2, SMAD6*) (**Fig. 3C)**. Four genes in this program carry two or more *de novo* variants in the Pediatric Cardiac Genomics Consortium (PCGC) cohort (*ANK3*, *CTNNB1*, *SAMD1, TBX5*)[7]. Consistent with the necessity of these genes in normal heart development, loss-of-function mouse models for *Gata4, Meis2, Tbx20, Smad6,* and *Tgfbr1* also result in valve defects and/of malformations of the outflow tract (see **Note S1**).

Cardiac fibroblasts include diverse mesenchymal cell populations present throughout the heart including in the septa, myocardium, and valve annuli[48,78]. The cell type in our atlas with the strongest enrichment of high-confidence CHD genes (Cardiac_fibroblast_progenitors) corresponds to an epicardial-derived fibroblast population based on expression of *TCF21, WT1 and TBX18* (**Fig. 3B** and **Fig. S5C**)[4], and was enriched for genes associated with ASD, valve defects, and malformations of the outflow tract (8-10-fold enrichment, FDR < 0.001; **Fig. 3A; Table S9**).

Of the 16 CHD genes highly expressed in cardiac fibroblast progenitors, 4 have been shown to lead to heart defects in animal models using conditional knockouts in the epicardial lineage (**Note S1; Table S10**). For example, *MEIS2* mutations have been observed in patients with ASD or VSD[79,80], coarctation of the aorta[81], Tetralogy of Fallot, hypoplastic right ventricle, and Ebstein's anomaly[82]. In mice, *Meis2* appears to play roles in both epicardial-derived fibroblasts and in VICs. In the epicardial lineage, double knockout of *Meis1/2* (using Cre recombinase driven by *Wt1* promoter) leads to a decrease in the proportion of epicardial-derived cardiac fibroblasts, and leads to VSD, malalignment of the great vessels, asymmetric septation of the outflow tract, and defects in coronary artery development[83]. Relevant to VICs, knockout of *Meis2* in the cardiac neural crest lineage (using *Sox9*-Cre) leads to severe defects in the aortic and pulmonary valves[84]. In our heart atlas, *MEIS2* is a member of VIC program 27 (see above) and also of FB program 6 ("Migration of epicardial-derived cardiac fibroblasts"), a gene program expressed in cardiac fibroblast progenitors that includes many genes involved in cell migration (*e.g.*, *SPOCK1*, *FAT4*, *SEMA6D*, *PLXNA4*, *TIAM1*, and *ENAH*). FB program 6 also includes 13 additional high-confidence CHD genes (odds ratio=6) (**Fig. 3D**), including *PBX1*, which encodes a known interaction partner of MEIS2[85], and six genes carrying *de novo* variants in the PCGC cohort (*ANK3*, *CACNA1A*, *GATA6*, *PTEN*, *SAMD11*, and *ZEB2*)[7].

In summary, many high-confidence CHD genes are highly expressed in VICs and cardiac fibroblast progenitors, some as part of co-regulated gene programs. Together with previous evidence from mouse models, this unbiased enrichment analysis suggests that these mesenchymal cell types may be important contributors to multiple subtypes of human CHD.

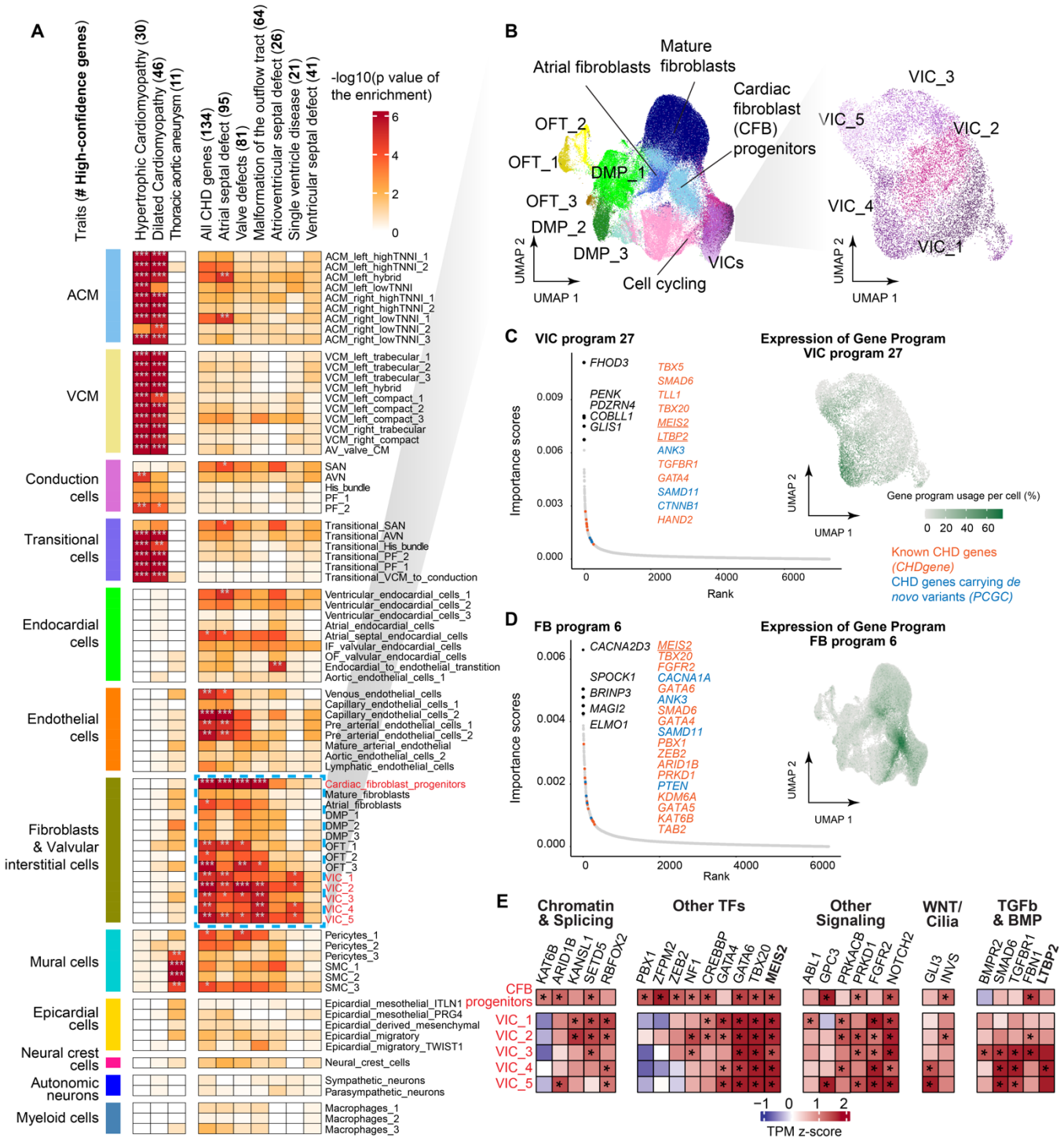


Figure 3: Congenital Heart Disease (CHD) Genes are enriched in cardiac fibroblast cells.

A. Enrichment for cardiovascular disease and CHD causal genes in cardiac cell types of the fetal heart map. Color scale: -Log_{10}-transformed p-values. One-sided fisher exact test with Benjamini-Hochberg correction (*FDR < 0.05, **FDR < 0.01, ***FDR < 0.001). # High-confidence genes: the number of CHD genes expressed in any cell type in the heatmap (TPM>1).

B. UMAP embedding of scRNA-seq of all fibroblast cells and UMAP embedding of scRNA-seq of VICs. DMP: dorsal mesenchymal protrusion; OFT: outflow tract.

C. Ranked importance scores of genes associated with VIC program 27; The top 300 genes by importance scores in VIC program 27 were considered as program genes. Black: Top 5 genes associated with VIC

program 27; Orange: high-confidence CHD genes from CHDgene in the top 300 genes[8]; Blue: CHD genes carrying two or more *de novo* variants identified in the PCGC cohort[7]. *TBX5* is both a high-confidence CHD gene and a gene carrying two or more *de novo* variants. Inset: Projection of VIC program 27 expression on the VIC UMAP, showing expression in VIC_4 and other VIC populations. Color scale: gene program usage per cell.

D. Ranked importance scores of genes associated with FB program 6. Labels denote selected program genes (among the 300 by importance score, see **Methods**). Black: Top 5 genes associated with FB program 6. Orange: high-confidence CHD genes from CHDgene[8]. Blue: CHD genes carrying two or more *de novo* variants identified in the PCGC cohort[7]. Inset: Projection of FB program 6 expression on the fibroblast UMAP, showing high expression in cardiac fibroblast progenitors and some cell cycling fibroblasts.

E. Z-score normalized TPM values of 27 CHD genes highly expressed in VICs and/or cardiac fibroblast progenitors (CFB). Bold genes (*MEIS2* and *LTBP2*) are discussed in the text. *: the gene is considered to be highly expressed in the corresponding cell types (see **Methods**).

Convergence of rare and common variation on VICs

To further explore the role of VICs in heart development and disease, we examined common noncoding variants associated with valve diseases and quantitative traits through GWAS. Previous genetic studies have noted that certain GWAS signals for acquired valve diseases are located near known CHD genes. For example, a noncoding variant near *LTBP2*, one of the CHD genes described above, is associated with mitral valve prolapse[86]. However, the specific causal enhancers or cell types for these loci have not yet been identified. Accordingly, we applied our regulatory map to test whether noncoding variants associated with valve traits might act in VICs to regulate known CHD genes and other genes in similar pathways.

We first assessed the overall enrichment of noncoding GWAS variants in VIC enhancers for 6 diseases and quantitative traits related to the valves and the aorta, including our recently conducted GWAS for measurements of aortic valve hemodynamics[87,88] (mitral valve prolapse[86], aortic stenosis[89], aortic root diameter[90], and peak velocity and mean gradient of the aortic valve[88,91], see **Methods**; **Table S11**). We indeed observed strong enrichment for genome-wide heritability in fetal VIC enhancers, using stratified linkage disequilibrium score regression (S-LDSC) (up to 12.5-fold enrichment, for aortic root diameter in predicted enhancers in VIC_3, a population of VICs in the free segments of the valve leaflets; **Fig S13; Table S12)**.

Thirteen GWAS signals for valve traits contain noncoding variants linked to 8 high-confidence CHD genes (*ELN, FBN1, GATA4, JAG1, LTBP2, NKX2-5, PIGL, TBX20*) via VIC enhancers (**Table S13)**. For example, variant rs989909 is associated with mitral valve prolapse, overlaps an enhancer predicted to regulate *LTBP2* in all VIC populations (**Fig. 4A**), and is predicted to create a composite GATA/TAL motif instance[92] (**Fig. 4C**). Variant rs6460068 is associated with aortic root diameter and aortic valve mean gradient, and overlaps a VIC enhancer predicted to regulate elastin (*ELN*)—a ubiquitous extracellular matrix protein in which loss-of-function mutations lead to supravalvular aortic stenosis and valve defects[93,94] (**Fig. S14A**). These observations suggest that common noncoding variants may tune the expression of CHD genes in fetal VICs to impact valve structure.

While not directly targeting a CHD gene, many other GWAS signals for valve traits were linked to genes that are part of the same pathways. For example, we found an enrichment of predicted target genes of valve trait GWAS signals in VIC program 27 ("VIC migration and TGFb signaling") — the same program noted above that contains *LTBP2*, *GATA4*, and 7 other CHD genes (**Fig. 4E**). A total of 17 genes linked to valve traits via GWAS variants in VIC enhancers are members of VIC program 27 (4.7-fold enrichment, $P = 6.2 \times 10^{-7}$). For example, a noncoding variant associated with mitral valve prolapse (rs165177) is predicted to create an SP-like motif instance and regulate *LCMD1*, which encodes a TF that inhibits the activity of GATA6 and whose knockdown in zebrafish leads to AV valve regurgitation[95] (**Fig. 4B,D**). Another variant, associated with aortic root diameter, is predicted to regulate *NPR3* (**Fig. S14B**), a gene where genetic misregulation has been observed in the context of valve insufficiency[96] and progressive aortic root dilatation in humans[97] .

Together, these observations suggest that noncoding variants may affect fetal VIC enhancers to regulate valve structure and function. The target genes include known CHD genes, where loss-of-function mutations lead to more severe heart defects, and other genes that are co-regulated in the same gene pathways. Thus, rare and common variants associated with valve diseases converge on shared genes and pathways in VICs. Notably, this suggests that polygenic regulation of valve structure might also influence risk for CHD (see **Discussion**).

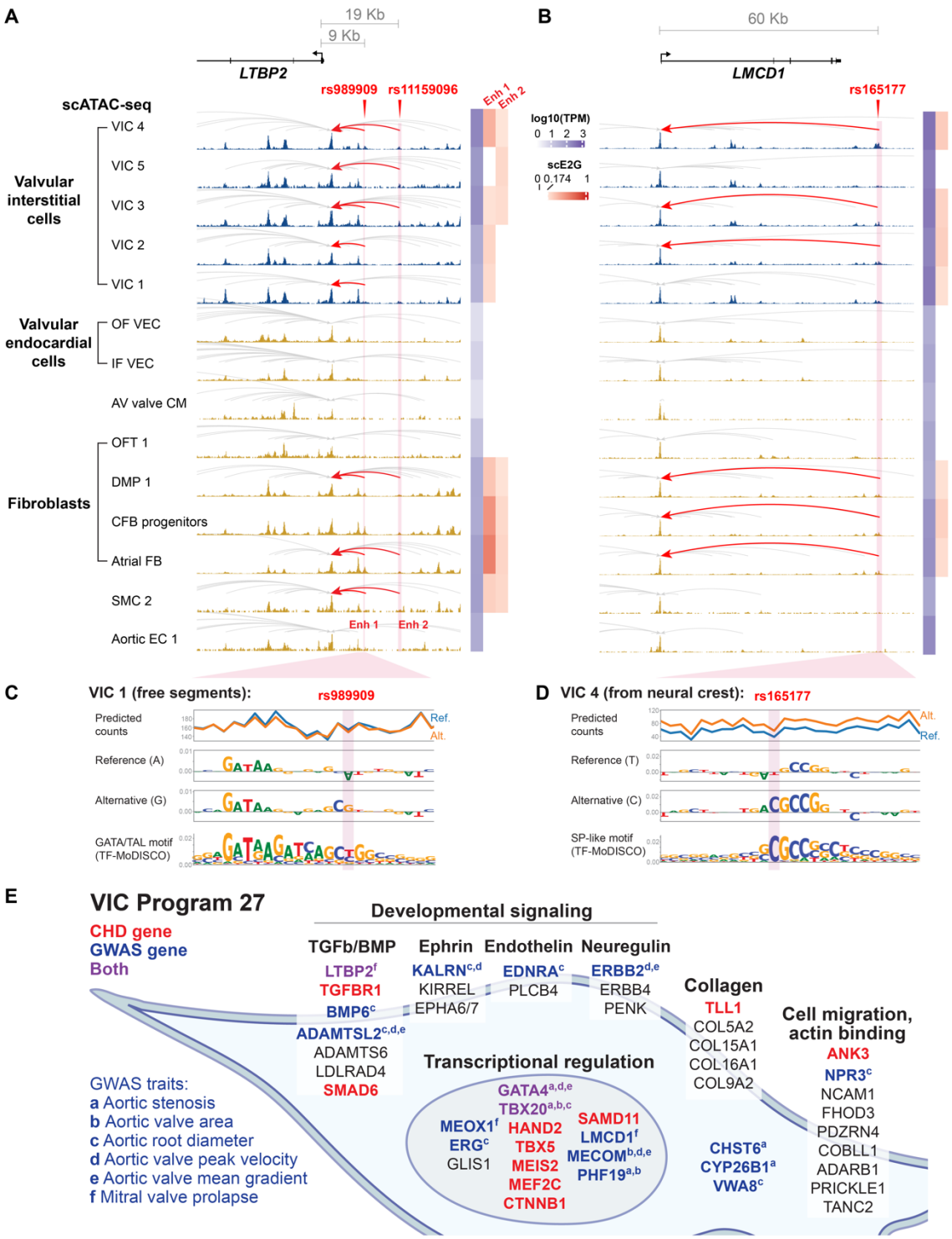

A
19 Kb
9 Kb
LTBP2
rs989909
rs11159096
Enh 1
Enh 2
scATAC-seq
Valvular interstitial cells
VIC 4
VIC 5
VIC 3
VIC 2
VIC 1
Valvular endocardial cells
OF VEC
IF VEC
AV valve CM
Fibroblasts
OFT 1
DMP 1
CFB progenitors
Atrial FB
SMC 2
Aortic EC 1
B
60 Kb
LMCD1
rs165177
log10(TPM)
0 1 2 3
scE2G
0 0.174 1
C
VIC 1 (free segments):
rs989909
Predicted counts
Ref.
Alt.
Reference (A)
Alternative (G)
GATA/TAL motif (TF-MoDISCO)
D
VIC 4 (from neural crest):
rs165177
Predicted counts
Alt.
Ref.
Reference (T)
Alternative (C)
SP-like motif (TF-MoDISCO)
E
VIC Program 27
CHD gene
GWAS gene
Both
Developmental signaling
TGFb/BMP
LTBP2f
TGFBR1
BMP6c
ADAMTSL2c,d,e
ADAMTS6
LDLRAD4
SMAD6
Ephrin
KALRNc,d
KIRREL
EPHA6/7
Endothelin
EDNRAc
PLCB4
Neuregulin
ERBB2d,e
ERBB4
PENK
Collagen
TLL1
COL5A2
COL15A1
COL16A1
COL9A2
Cell migration, actin binding
ANK3
NPR3c
NCAM1
FHOD3
PDZRN4
COBLL1
ADARB1
PRICKLE1
TANC2
Transcriptional regulation
GATA4a,d,e
TBX20a,b,c
MEOX1f
HAND2
SAMD11
ERGc
TBX5
LMCD1f
GLIS1
MEIS2
MECOMb,d,e
MEF2C
PHF19a,b
CTNNB1
CHST6a
CYP26B1a
VWA8c
GWAS traits:
a Aortic stenosis
b Aortic valve area
c Aortic root diameter
d Aortic valve peak velocity
e Aortic valve mean gradient
f Mitral valve prolapse

Figure 4: Convergence of rare and common variants on VIC pathways

A,B. scE2G predictions and normalized ATAC-seq signals are shown for *LTBP2* (A) and *LMCD1* (B), showing that GWAS variants associated with valve diseases or traits (red downward arrows) overlap enhancers linked to these genes in VICs. Enhancer predictions overlapping variants are highlighted in red. Arcs represent the predicted enhancer-gene regulatory interactions. Side color bars show the TPM expression of the target genes in each cell type (blue) and the corresponding scE2G scores (red; white for scores below the threshold). CFB: cardiac fibroblasts. The hg38 coordinates of the highlighted enhancer regions are as follows: *LTBP2* Enh 1: chr14:74621382-74622571; *LTBP2* Enh 2: chr14:74630908-74631785; *LMCD1*: chr3:8560989-8562461.

C,D. For selected variants, predicted impact on chromatin accessibility by ChromBPNet. Top: Predicted ATAC-seq counts for the reference (blue) and alternative (orange) alleles. Middle: Importance scores from ChromBPNet in VIC_1 (free segments) and VIC_4 (VICs derived from neural crest). Bottom: Motif from TF-MoDiSCO predicted to be altered by the variant. The alternative allele at rs989909 (G) creates a motif for TAL as part of a GATA/TAL heterodimer, which is predicted to have a small impact on local chromatin accessibility, consistent with previous studies of TAL factors, and is expected to have a larger effect on enhancer activity[92]. The alternative allele at rs165177 (C) creates an SP-like motif that is predicted to increase chromatin accessibility.

E. Annotation of genes in VIC Program 27, which contains many CHD genes (red), genes linked to noncoding GWAS variants for valve or aortic traits (blue), or genes with both types of evidence (purple). Black genes represent other genes in VIC Program 27 with related functions.

Noncoding risk variants affect fetal enhancers with cell type- and location-specific activities

Beyond valve traits, a handful of noncoding variants have been associated directly with subtypes of CHD through GWAS[12–15], and hundreds of noncoding variants have been associated with other quantitative traits such as functional traits of the atria and ventricles[29–38]. We applied our gene regulation map to discover the enhancers, cell types, and TFs that mediate the effects of variants associated with 45 structural heart diseases and traits (**Fig. 5A; Table S11**).

In total, our regulatory map linked noncoding variants for 770 unique GWAS signals to enhancers and target genes (**Table S13**). Heritability for each trait was enriched in a distinct set of cell types (**Fig. 5B; Table S12**). At individual loci, many of the prioritized enhancers were predicted to be active in specific cell types and spatial locations—providing new insights into the genetic etiology of heart structure and function. We present insights from 2 examples:

Variant rs6702619, near *PALMD*, is associated with a subtype of CHD (bicuspid aortic valve), as well as with calcific aortic valve stenosis and quantitative traits related to the aorta and aortic valve in adults (**Fig. 5C**)[36,89,98,99]. Previous studies have shown that rs6702619 is an eQTL for *PALMD* in the adult aortic valve[100], overlaps an enhancer in adult VIC, and affects a binding site for NFATC2[101]. The target gene, *PALMD*, has been shown to regulate endocardial-to-mesenchymal transition (EMT) and VIC remodeling[101,102]. In our map, rs6702619 overlaps an enhancer predicted to regulate *PALMD* not only in VIC but also valvular endocardial cells and aortic endothelial cells (**Fig. 5D**). ChromBPNet predicts that the disease-associated allele (G) reduces chromatin accessibility at the enhancer by 43%-68% in valvular cells, with the strongest reduction in endocardial cells on the inflow side of the valves (**Fig. 5E**), consistent with the direction of effect of the adult eQTL and NFATC2 binding[100,101]. In fetal endocardial cells, *PALMD* is part of an endocardial shear stress response gene program ("Endocardial Program 29") that is regulated by KLF2/4 and includes other factors important in valve development **(Fig. S15, Tables S5-6**). Together, these data extend previous observations by showing that this variant overlaps an enhancer active in fetal valve cells, providing a molecular explanation for its association with congenital valve defects.

At the *MTSS1* locus, the variants rs35006907, rs34866937, and rs12541595 are in strong linkage disequilibrium ($R^2 \geq 0.978$). These variants are associated with traits related to left ventricular function, including decreased end-systolic and end-diastolic volumes[35] and increased ejection fraction[35] (**Fig. 5F**). The alternative haplotype is also associated with lower expression of *MTSS1* in the left ventricle (GTEx)[103]. In our atlas, rs35006907 and rs34866937 overlap a single enhancer predicted to regulate *MTSS1* in all VCMs and a small subset of ACMs (**Fig. 5G**). ChromBPNet predicts that the alternative allele (A) at rs35006907 reduces chromatin accessibility of the enhancer by 10-30% in VCMs, with the highest reduction in a population of trabecular cardiomyocytes, by altering a motif instance that matches USF1 and MITF (**Fig. 5H**). *Mtss1* knockout in mice is embryonic lethal and leads to decreased LV end-diastolic and end-systolic dimensions[104], and knockout in medaka leads to a marked reduction in trabeculation[29]. Our results indicate that, because this enhancer is active early in fetal development, the effects of this variant on many myocardial traits could have developmental origins.

To validate the cell type specificity of these enhancers, we tested the expression patterns of these 2 enhancers in E11.5 mouse embryos using a transgenic enhancer reporter assay (**Fig. 5I; Table S14**). Both enhancers had spatially restricted activities that were consistent with the predicted cell types of action above (**Fig. 5J,K**). The *PALMD* enhancer showed specific activity in the outflow tract, consistent with the scE2G-predicted activity in aortic and valvular endocardial cells in our human atlas (**Fig. 5I**). The *MTSS1* enhancer showed specific activity in the ventricular septum (**Fig. 5K**). We further tested the effects of *PALMD* variants on enhancer activity, and found that introducing the alternative allele for rs6702619 and tightly linked rs11166276 led to undetectable enhancer activity, consistent with its strong predicted effect on accessibility (**Fig. S16**).

Together, these analyses suggest that noncoding variants can influence risk for CHD and other structural traits by regulating gene expression in specific cell types and subanatomic locations in the developing heart. Including the examples presented here, annotations of 770 GWAS signals for cardiac diseases and traits will provide a rich resource for future study (**Table S13**).

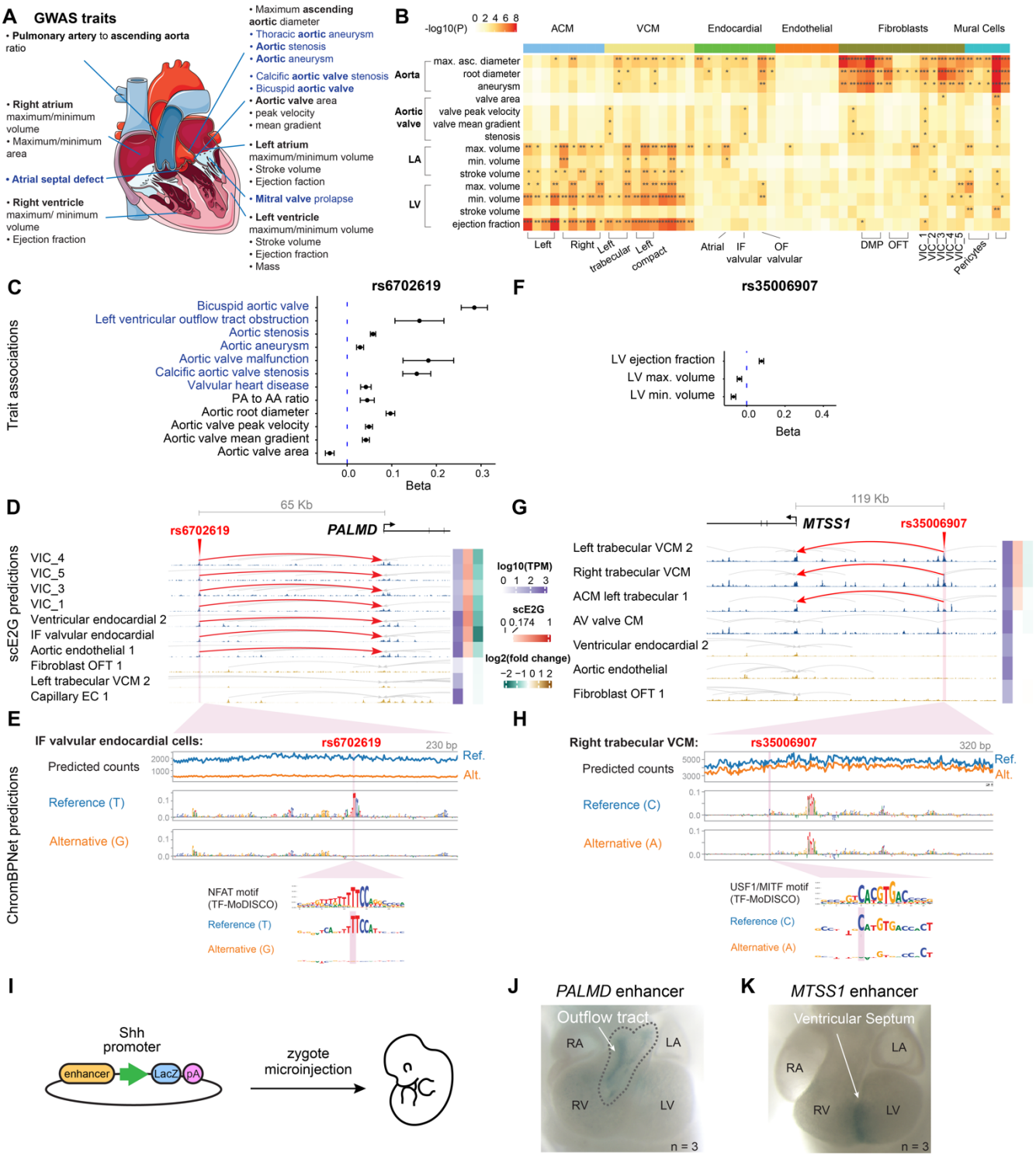


Figure 5: Noncoding variants affect enhancers with cell type and spatially restricted activities

A. Selected GWASs of quantitative measurements of the heart (gray) and cardiac diseases (blue) included in this analysis.

B. Heritability enrichment (S-LDSC) of quantitative measurements of the heart and cardiac diseases in scE2G predicted enhancers of each cell type. Representative cell types are shown. See also **Table S12**.

C. Trait associations for rs6701619. Beta: Per unit change in the outcome associated with the alternative allele. The alternative allele increases the risk for aortic and aortic valve related diseases, and increases the quantitative measurement of aortic traits except for aortic valve area.

D. rs6701619 overlaps an enhancer (hg38: chr1:99580436-99581202) predicted to regulate *PALMD* in valvular cells (red highlight and arcs). Signal tracks represent chromatin accessibility from ATAC-seq. Gray arcs represent other predicted enhancer-gene regulatory interactions in the locus. The side color bars show the TPM expression of *PALMD* in each cell type (blue), scE2G scores (red, and white for scores below the threshold), and ChromBPNet predicted log2 fold changes in chromatin accessibility upon substitution of the reference allele (T) with the alternative allele (G) (green indicates reduced accessibility, brown indicates increased accessibility). IF: inflow.

E. ChromBPNet predictions for rs6702619 in (IF) inflow valvular endocardial cells. The alternative allele is predicted to lead to a decrease in chromatin accessibility across the element (top). Middle and bottom tracks show contribution scores for the reference and alternative alleles. Inset: NFATC1 motif learned *de novo* by TF-MoDISCO, and contribution scores for the reference and alternative alleles.

F. Trait associations for rs35006907. Beta: Per unit change in the outcome associated with the alternative allele. The alternative allele increases LV ejection fraction, and decreases both the maximum LV volume and the minimum LV volume.

G. rs35006907 overlaps an enhancer (hg38: chr8:124846865-124849334) predicted to regulate *MTSS1* in VCMs (red highlight and arcs). Similar to **E**.

H. ChromBPNet predictions for rs35006907 in right trabecular VCMs. The alternative allele is predicted to lead to a decrease in chromatin accessibility across the element (top). Middle and bottom tracks show contribution scores for the reference and alternative alleles. Inset: Motif learned *de novo* by TF-MoDISCO (similar to USF1 and MITF motifs), and contribution scores for the reference and alternative alleles.

I. Schematics for the mouse transgenic assay. The predicted enhancer is cloned to drive the expression of LacZ.

J. The enhancer overlapping rs6702619 is active in the outflow tract of the developing mouse heart at E11.5.

K. The enhancer overlapping rs35006907 is active in the interventricular septum of the mouse heart at E11.5.

Discussion

Here, we illuminate the genetic and cellular etiology of CHD using a comprehensive gene regulation map of the human fetal heart. Considering 90 cell types in the human fetal heart, we implicate new cell populations in the genetic etiology of CHD subtypes, discover noncoding risk variants with cell type- and location-specific activities, and identify molecular convergence of variants associated with CHD and quantitative valve traits. These insights will help to guide future studies of CHD, provide a resource for interpreting the functions of variants observed in patient genomes, and may inform the development of therapies based on novel cell or molecular targets.

This study, by combining a comprehensive multiomic atlas of the fetal heart with known CHD genes and genetic variants, provides key evidence for the hypothesis that VICs and cardiac fibroblasts play direct roles in the genetic etiology and development of CHD[2,13,39,48,105]. In an unbiased analysis of 90 cell types in the human fetal heart, VICs and cardiac fibroblasts show the strongest enrichment for high-confidence CHD genes, including for multiple subtypes of CHD (**Fig. 3A**). Consistent with this finding, previous studies have studied individual genes with effects on mesenchymal cells and heart development in animal models (**Note S1**), noted the expression of candidate CHD genes in mesenchymal cells in other datasets[2,13,39,48], and observed differences in cardiac fibroblasts in the hearts of CHD patients after birth[106]. Our data expand on these observations by prioritizing specific genes and connecting them into molecular pathways in these cell types (**Fig. 3**), linking evidence from lineage-specific murine knockouts to unbiased analysis of human cell types (**Note S1**), and identifying noncoding variants associated with valve traits that link to known CHD genes in VICs (**Fig. 4**). Together, our results suggest that a large fraction of CHD genes, including genes such as *LTBP2*, *GATA4*, and *MEIS2*, may lead to heart defects through primary effects in mesenchymal cells. This has implications for our fundamental understanding of how heart defects arise beyond disease of the working myocardium, and motivates further development of appropriate cell and animal models to study the contributions of these cell types to CHD.

We identify, for the first time, convergence of rare variants associated with CHD and common variants associated with quantitative measurements of valve structure and function—in particular, on a gene program in neural-crest derived VICs (**Figs. 3C,4E**). This identifies a specific pathway that could mediate polygenic contribution to CHD. Indeed, recent studies have suggested the possibility of polygenic or oligogenic architectures for CHD in which different heart defects might arise from combinations of causal genetic variants in a given individual[12–15,33,107,108]. Patients or family members with rare variants in the same genes are often observed to have different types of CHD, and combining causal variants in animal models have directly confirmed certain genetic interactions[107]. Previous studies have noted individual cases where GWAS loci for quantitative cardiac traits are located near CHD genes[36], and observed genome-wide genetic correlations between certain quantitative traits and CHD risk, such as between interventricular septum structure and risk for VSDs[33]. However, specific molecular pathways that mediate such convergence have not previously been identified. Our data suggest a possible model in which common, inherited genetic variation tunes the structure of the valve by influencing VICs during fetal development. In this model, the severity or form of CHD would be influenced by the combination of rare, high-effect variants (*e.g.*, in genes such as *LTBP2*, *GATA4*, and *MEIS2*) and common, mostly noncoding variants that regulate some of these same genes or other genes in the same molecular pathways (**Fig. 4E**).

Our comprehensive atlas of noncoding regulatory elements enabled the first systematic assessment of the roles of inherited noncoding variants associated with CHD and quantitative cardiac traits during fetal development. We link hundreds of such noncoding variants to candidate enhancers, cell types, genes, and pathways (**Table S13**). Individual variants are predicted to act in specific cell types (*e.g.*, VICs and valvular endocardial cells for bicuspid aortic valve), but collectively span dozens of different cell types, highlighting the importance of considering the diverse totality of cell types present during cardiogenesis. Functional testing of two of these enhancers revealed unique, spatially constrained activity patterns, suggesting that specific spatial patterns of enhancer activity could explain how some noncoding variants

affect specific aspects of heart structure (**Fig. 5**). These specific patterns also highlight the utility of our cell type-specific enhancer catalog for identifying and testing enhancers with spatial specificity, which may be leveraged to develop expression vectors for gene therapies that target specific cell types, locations, or developmental timepoints.

Our study illustrates the importance of high-depth single-cell atlases for building gene regulation maps to interpret the impact of human genomic variation on function. Leveraging cutting-edge predictive models, we construct regulatory maps of noncoding variants (with ChromBPNet) and enhancers (with scE2G) in cell types with at least 3 million fragments. Accordingly, we are able to annotate impacts of chromatin accessibility and target genes across 90 cell types and cell states present in the fetal heart, including rare but clinically relevant cell types such as the cardiac conduction system. However, deeper data will be required to identify rarer cell types, or cell states that occur only at certain time points during development. For example, we observe rare subpopulations of migratory epicardial cells (**Fig. S5**), which are too rare even in this dataset to build accurate maps of gene regulatory elements. Our atlas also lacks hearts earlier than post-conception week 6, timepoints that are critical to the formation of the heart and that may include additional cell types and states important for the genetic etiology of CHD. As such, the variants and genes we describe here may have functions in other cell types or stages not included in our atlas. Expanding single-cell multiomic studies to include earlier time points would provide a more comprehensive understanding of how our genome encodes heart development, enabling additional discoveries that were previously only possible in model systems.

Together, this work provides new insights into the genetic etiology of CHD, and a foundational map for interpreting effects of genetic variants on heart structure and function. As genome sequencing and association studies of CHD continue to expand, we anticipate that this map will enable many additional discoveries about the genes, pathways, and cell types that contribute to heart development and disease.

Methods

Human Heart Tissue Sample Collection

Human fetal heart tissue samples were collected from three primary sources: Stanford University, Cercle Allocation Services, Inc. (IRB 20111833), and the University of Washington (UW) Birth Defects Research Laboratory (BDRL). Each collection adhered to the ethical and legal guidelines specific to the institution, including obtaining written maternal consent. The research followed the ethical and legal standards of the Stanford University Research Compliance Office, ensuring compliance with policies at Stanford and other collaborating institutions. All collected tissues were screened and confirmed to be negative for trisomies and other non-single ventricle defects based on prenatal screening. The post-conceptional age (PCW) of the tissue was determined through ultrasound assessment and by referencing the date of the last menstrual period prior to pregnancy termination. A total of 41 structurally normal human fetal hearts, ranging from 6 to 22 post-conception weeks, were collected from these sources. Samples from Cercle and BDRL were stored in 1X HBSS+/+ (Gibco 14025-092) before being flash-frozen and shipped overnight on dry ice for cryobanking at -80°C. Samples collected directly from Stanford University were flash-frozen in-house and cryobanked at -80°C. Most samples were flash-frozen within 2 hours of the procedure.

Sample Preparation for Single-Cell Multiome

Three protocols were developed to process samples, each tailored to meet specific sample conditions and experimental needs. Protocol 1 (http://dx.doi.org/10.17504/protocols.io.rm7vzjez5lx1/v1) is designed for fresh tissue, providing high-quality nuclei for downstream analyses due to its ability to preserve sample integrity. Protocol 2 (http://dx.doi.org/10.17504/protocols.io.rm7vzjez5lx1/v2) is optimized for handling small or limited tissue samples and is preferred when nuclei retention is critical, ensuring that even minimal sample inputs yield viable results. Protocol 3 (http://dx.doi.org/10.17504/protocols.io.rm7vzjez5lx1/v3), which was used for processing the majority of the samples in this study, provides consistency and scalability across larger sample sets due to its ability to maintain good quality nuclei and effective purification steps. The methods described within this subsection are based on Protocol 3, and all three protocols can be accessed via Protocols.io for detailed step-by-step guidance. Details on which samples were processed using Protocols 1, 2, or 3 can be found in **Table S2**.

Human fetal heart samples from 41 cases, ranging from 6 to 22 post-conception weeks, were processed (see **Table S1** for sample attribute information and **Table S2** for sample metadata details). These samples were flash-frozen before being processed for the experimental pipeline. Each sample was mechanically homogenized separately in a 15 mL pre-chilled Dounce tissue grinder (Millipore Sigma D9938) with 4 mL of 1x Homogenization Buffer, which consisted of 260 mM sucrose, 30 mM KCl, 10 mM MgCl2, 20 mM Tricine-KOH pH 7.8, 1 mM dithiothreitol (DTT), 0.5 mM spermidine, 0.15 mM spermine, 0.3% NP40, 60 U/mL Ribolock RNase Inhibitor, 1% BSA, and 0.5–1 tablet of cOmplete Protease Inhibitor. To begin, 20–100 mg of flash-frozen tissue was cut with a razor blade in a tissue culture dish placed on an ice block over dry ice to prevent premature thawing. The tissue was then transferred to the Dounce tissue grinder. The tissue was allowed to thaw for 5 minutes in the chilled buffer before homogenization. Homogenization was performed using 10–20 strokes with pestle A (loose), followed by 10–20 strokes with pestle B (tight), depending on the tissue amount and developmental stage (older tissue required more strokes). The homogenized solution was then filtered through a 100-µm cell strainer (PluriSelect 43-50100-51) into a pre-chilled 50 mL conical tube. To rescue any remaining nuclei, an additional 400 µL of 1x Homogenization Buffer was used to rinse the Dounce tissue grinder and passed through the same 100-µm cell strainer. The sample was further filtered by gently pipetting in 400–500 µL increments through a 70-µm Flowmi cell strainer (Bel Art H136800070) into a new 5 mL Eppendorf low-binding tube. Nuclei were pelleted by centrifugation at 350g for 5 minutes at 4°C, after which all but 50 µL of supernatant above the pellet was removed. The nuclei were then resuspended in 750 µL of 1x Homogenization Buffer, bringing the total volume to 800 µL. Further purification was achieved using an Iodixanol density gradient. Each sample was mixed with an equal volume (800 µL) of 50% Iodixanol

Solution (60% Iodixanol, 150 mM KCl, 30 mM MgCl2, 120 mM Tricine-KOH pH 7.8, nuclease-free water), resulting in a 25% nuclei mixture. The mixture was gently pipetted to ensure thorough mixing. A layer of 1000 µL of 30% Iodixanol Solution (1x Homogenization Buffer, 50% Iodixanol) was carefully added beneath the 25% nuclei mixture, followed by 1000 µL of 40% Iodixanol Solution to create a third layer. To prevent mixing between layers, the sides of the pipette tip were wiped with a kimwipe, and the tube was kept upright throughout the process. The sample was centrifuged at 3500g for 15 minutes at 4°C with the brake off (deceleration 0, acceleration 5). This process resulted in three distinct layers: the top layer (debris), the middle nuclei layer (target layer), and the bottom layer (more debris). The top layer was aspirated down to within 200–300 µL of the nuclei band. A maximum of 200 µL was collected from the target layer and transferred to a fresh 1.5 mL Eppendorf low-binding tube, avoiding excess debris. The nuclei were washed in 1 mL of 1x Homogenization Buffer, topped off to 1.5 mL, and centrifuged at 1200g for 5 minutes at 4°C. The supernatant was removed, and 700 µL of ATAC-RSB-Tween buffer (10 mM Tris-HCl pH 7.5, 10 mM NaCl, 3 mM MgCl2, nuclease-free water, 0.1% Tween-20, 1% BSA, 1 U/µL RNase Inhibitor) was added to resuspend the nuclei pellet. The solution was gently mixed by pipetting and filtered through a 40-µm Flowmi cell strainer (Bel Art H136800040) into a new 1.5 mL Eppendorf low-binding tube. Nuclei quality was assessed using Trypan Blue staining and manual counting with a hemocytometer (Hausser Scientific 1490)[49,109].

Single-Cell Multiome Library Preparation and Sequencing

The nuclei suspension was adjusted to a concentration of 10,000 nuclei/µL and processed using the 10x Genomics Chromium Controller. Nuclei were prepared for single-cell multiome sequencing following the Chromium Single Cell Multiome ATAC + Gene Expression kit (10x Genomics, CG000338 Rev F) with the Chromium Single Cell Multiome ATAC + Gene Expression Reagent Bundle (PN-1000283), and the Chromium Next GEM Chip J Single Cell Kit (PN-1000234). Libraries were uniquely indexed with Single Index Kit N Set A (PN-1000212) for snATAC-seq and Dual Index Kit TT Set A (PN-1000215) for snRNA-seq. Quality control for the cDNA and final libraries was performed using the Bioanalyzer High Sensitivity DNA Analysis Kit (Agilent, 5067-4626), and the Qubit™ 1X dsDNA High Sensitivity (HS) Assay Kit (Thermo Fisher Scientific, Q33231). Sequencing libraries were pooled and sequenced on Illumina NextSeq 550, NovaSeq, NovaSeq X and NovaSeq X Plus instruments, achieving a cumulative mean read depth of at least 20,000 total reads per nucleus for RNA and 25,000 total reads per nucleus for ATAC (or 20,000–30,000 reads per nucleus).

scRNA-seq alignment and background removal

The raw sequencing data was converted to FASTQ format using bcl2fastq. ScRNA-seq reads were aligned to GENCODE v29 (https://woldlab.caltech.edu/~diane/genome/GRCh38-V29-male-2.7.8a.tar.gz) using STARSolo[110] (v2.7.11a) following the parameters specified in starsolo.snakefile (https://github.com/detrout/woldlab-rna-seq/tree/main, commit f592629). For the downstream analysis, we used the “GeneFull_Ex50pAS” count matrix, which contains all reads overlapping genes' exons and introns. We used CellBender[111] (v0.3.0) to remove ambient RNA from the count matrices with a false positive rate (0.01). The “expected-cells” and “total-droplets-included” parameters were set based on knee plots of UMI counts per cell barcode, as recommended by the CellBender user manual. We set "--learning-rate" and "--epochs" to 0.00002 and 150, respectively. CellBender's evaluation was used to determine if reruns with different parameters were necessary. In such cases, the learning rate was halved and epochs were increased to a maximum of 300 until the performance was evaluated to be satisfying.

scATAC-seq alignment

The raw sequencing data were converted to FASTQ format using bcl2fastq. Barcode matching was performed using matcha[112] (v0.0.2) with the maximum hamming distance to tolerate between overserved and whitelisted barcodes set to 1. The scATAC-seq reads were aligned to the GRCh38 reference genome (ENCODE: GRCh38_no_alt_analysis_set_GCA_000001405.15) using bowtie2[113] (v2.4.4). The aligned reads were filtered using samtools[114] (v1.13) view with the parameters -F 524 -f 2” to only keep properly

paired mapped reads that passed platform/vendor quality checks. Read pairs mapped to more than 4 locations were discarded. Duplicate reads were removed using Picard's MarkDuplicates[115] (picard-slim v2.35.7) with lenient validation stringency. We further filtered for primary alignments and discarded reads mapped to the mitochondrial genome. Finally, the alignment was sorted by coordinates and was converted to the fragment file format, during which the reads were shifted + 4 bp and − 4 bp for positive and negative strand respectively to account for Tn5 transposase insertion during library preparation.

scATAC-seq quality control

We calculated the nucleosome signal for each cell by computing the ratio of mononucleosomal (147–294 bp) to nucleosome-free (<147 bp) fragments sequenced for each cell. Only cells called by STARSolo (see ScRNA-seq alignment and background removal) and with nucleosome signals lower than 2 were considered for further analysis. These cells were read into the ArchR[116] (v1.0.2) and filtered based on two criteria: a minimum transcription start site (TSS) enrichment score of 6 ("minTSS=6") and at least 1000 mapped ATAC-seq fragments per cell ("minFrags=1000"). We then reviewed the log-transformed unique nuclear fragments versus TSS enrichment score plots of each library. For libraries with higher sequencing depth or distinct clusters of cells with low fragment counts, we increased the minimum fragment threshold to either 2000 or 3000 fragments to exclude low-quality cells.

scRNA-seq quality control and normalization

After calling cells using STARSolo and removing cells that did not pass the ATAC quality control (see **scATAC-seq quality control)**, to account for variations in sequencing depth and library quality, we performed library quality control for each library. We examined the distribution of three key metrics:

1. Number of UMIs per cell: we filtered out cells with very low (cutoffs ranged from 200 to 2000 UMIs per cell) or very high (thresholds ranged from 20000 to 50000 UMIs per cell)
2. Number of genes detected per cell: Cells with too few detected genes were filtered out (thresholds ranged from 300 to 1000 genes per cell).
3. Percentage of UMIs originated from the mitochondrial genes: cells with more than a library-specific threshold (ranged from 2% to 30%) for mitochondrial UMI percentage were filtered out.

The count matrix of each library was then normalized by proportional fitting prior to log transformation followed by an additional proportional fitting, as recommended[117] (https://github.com/pachterlab/BHGP_2022).

Doublet removal

Four doublet removal methods were used to detect doublets, using both RNA and ATAC modalities. For ATAC-seq data, we used ArchR's "addDoubletScores()" with the default parameters to identify doublets based on their similarity to the synthetic doublets created by mixing reads from combinations of individual cells. Additionally, we used AMULET[118] (commit 9ce413f) with default parameters to detect multiplets (including homotypic ones) by identifying nuclei with significantly more loci containing more than two reads. For RNA-seq data, for each library, we initially assigned doublet scores to each cell using Scrublet with the default parameters. The automatically generated doublet score thresholds were then evaluated by plotting the assigned doublet scores in histograms and visualizing the scores on the 2D embedding of the cells. The automatically detected thresholds were retained for doublet annotation if they were supported by the visualization; otherwise, we adjusted the thresholds based on the visualization. Furthermore, for libraries containing cells from multiple donors, we detected doublets by clustering cells based on their genotype using Souporcell[119] (commit 54fd312) default parameters. This method allowed us to identify doublets originating from genetically distinct samples, regardless of their gene expression profiles. Nuclei annotated as a multiplet by any of these four methods were removed from the downstream analysis.

Sample demultiplexing

To control for technical variability, maximize flow cell utilization, and facilitate doublet detection, we used a combinatorial pooling strategy for samples with sufficient quantity. This strategy allowed us to decode the original donors of the cells downstream by the unique genotype information of each donor. Specifically, this pooling strategy involved mixing cells from multiple donors (typically 2 or 3) in various combinations (e.g., A+B, B+C, and C+B). To attribute the cells to their original donor, we leveraged the cell clustering by genotype generated by Souporcell. For each pooled library, we selected at least 15 informative loci with homozygous genotypes (reference or alternative) across the cell clusters generated by Souporcell. We then combined cell clusters from libraries with overlapping donors and clustered the cell clusters based on their genotypes at the selected loci. Finally, cell clusters from different libraries with matching genotypes were assigned to donors shared by these different libraries. For samples that were insufficient for different pooling combinations, we pooled them together by their post conception weeks.

Sex determination of the donors

We determined the sex of the donors by evaluating the expression levels of sex specific genes in the cells from the donors. Genes used for identification of male samples included *ZFY*, *DDX3Y*, *UTY*, *USP9Y*. *XIST* was used for identifying female samples.

Assignment of cells to major cell types

Following quality control and doublet removal, cells from all libraries were combined for cell type identification. The combined count matrix was normalized using the method described in **scRNA-seq quality control and normalization.** Clustering was performed at three resolutions (0.2, 0.4 and 0.6) using the original Louvain algorithm implemented in the "FindClusters()" function of Seurat[120] (v4.3.0). We visualized the relationships between clusters across these resolutions using Clustree[121] (v0.5.0) and selected 0.2 as the resolution for downstream analysis. The clusters were assigned to the following preliminary cell types based on the expression of known marker genes: ACM (*MYH6*, *GJA5*, *NPPA*), VCM (*MYH7*, *MYL2*, *FHL2*), Endothelial and endocardial cells (*PECAM1*, *CDH5*, *UNC5B*), mural cells and fibroblast-like cells (*COL1A1*, *COL1A2*, *DCN*, *LUM*, *MYH11*, *ELN*), epicardial cells (*TBX18*, *UPK3B*, *WT1*), immune cells (*CD69*, *SKAP1*, *LEF1*, *LYVE1*, *CD68*, *PTPRC*, *CD74*), and neuron and neural crest cells (*NRXN1*, *SOX10*, *TFAP2B*, *TH*, *FOXD3*). Clusters highly express genes indicative of contamination, such as *HBG1*, *TSPAN5*, *ALB*, were removed.

Clustering and cell type annotation

For all cell clustering and cell type annotation analysis, we used count-splitting[122]. Specifically, we split the count matrix of each preliminary cell type into a training set and a test set of equal sizes using negative binomial splitting implemented in countsplit[122] (v4.0.0). The gene-specific overdispersion parameters were estimated for all genes using the "vst()" function from sctransform[117,123] (v0.3.5) with the "n_genes" parameter set to NULL.

Normalization of the training matrix was performed as described previously in **scRNA-seq quality control and normalization**. To cluster the cells, we first identified the top 2000 variable genes using Seurat's "FindVariableGenes()" function. We then excluded from the list ribosomal genes ("^RPS.*|^RPL.*"), mitochondrial genes ("^MT-*"), genes associated with cell cycling (provided by cc.genes in Seurat), and sex-specific genes (identified by extracting genes unique to the X or Y chromosome from GENCODE v29 made available by ENCODE). This step aimed to prevent these genes from driving clustering while allowing them to be detected as differentially expressed genes. Clustering was performed at 10 resolutions (0.1 to 1) using the original Louvain algorithm implemented in the "FindClusters()" function of Seurat[120] (v4.3.0). The resolution was selected based on the relationships between clusters across these resolutions as visualized by Clustree[121] (v0.5.0).

For each identified cluster, we used the test matrix to identify differentially expressed genes (calculated by Wilcoxon rank-sum test) and evaluate the expression of known marker genes. Additionally, using the

original matrix, we evaluated the UMI count scaled by the corresponding sequencing library. Clusters with compromised quality were removed. These include clusters with high expression of genes indicative of contamination (e.g., *HBG1*, *TSPAN5*, *ALB*), mitochondrial or ribosomal genes, markers not consistent with the assigned preliminary cell types (e.g., cardiomyocyte markers in immune cell types), or a high proportion of cells with high UMI counts from each library (these might be doublets or multiplets). After removing these compromised clusters, normalization and clustering were repeated to identify and annotate fine-grained cell types/states. This process was repeated if any newly formed clusters showed signs of compromised quality. The markers used for each fine-grained cell type and state can be found in **Figs. S3-7**. The 2-dimensional representations of the cells were generated using uniform manifold approximation and projection (UMAP) as implemented in Seurat[120] (v4.3.0).

Of note, we opted to not correct for batch effects associated with experiment or donor due to concerns of removing biological variation by over-correction. We noted that the clustering results were largely not influenced by experiment or donor: no clusters exclusively comprised cells from a specific library or donor.

Predicting enhancer-gene regulatory interactions with scE2G

We used the scE2G model (https://github.com/EngreitzLab/sc-E2G)[54] to predict enhancer–gene regulatory interactions within each cell type. The scE2G model builds upon the previously established Activity-by-Contact (ABC) and ENCODE-rE2G models and is adapted for single-cell applications. The scATAC-seq fragments of each fine-grained cell type were pseudo-bulked. Cell types with less than 3 million fragments were excluded, affecting 11 cell types. The pseudo-bulked scATAC-seq signals ("activity") and a Hi-C contact map averaged across human cell types ("contact")[124] were used to compute the ABC score for each cell type. To incorporate the gene expression information and take advantage of the single-cell nature of the input dataset, the Kendall correlation between chromatin accessibility and gene expression across single cells within each fine-grained cell type was calculated. These correlation coefficients were subsequently integrated with corresponding ABC scores to improve calibration across varying sequencing depths.

The integrated ABC score and Kendall correlation, along with 5 additional features, were used as input logistic regression models to predict CRISPRi-validated enhancer-gene pairs in K562 cells. The additional features include 1) the number of transcription start sites (TSSs) between each enhancer and the promoter of the target gene, 2) the number of peaks between each enhancer and the promoter of the target gene, 3) chromatin accessibility at the promoter, 4) number of peaks within 5 kb of the enhancer, and 5) whether the target gene is ubiquitously expressed. The trained scE2G model was applied to score each peak-gene pair in the 90 cell types.

The resulting scE2G scores were then quantile normalized. In our analysis, excluded genes that are in the following categories: long noncoding RNAs ('^LINC'), gene isoforms ('-AS'), microRNAs ('^MIR'), small nuclear RNAs ('RNU'), genes of uncertain functions ('^LOC'), and also genes with TPM less than 1 in the respective cell type. Enhancer-gene pairs with quantile-normalized scE2G scores greater than or equal to 0.174 were considered as predicted enhancer-gene regulatory interactions.

Identification of overlapping enhancers

To identify predicted enhancers likely representing the same active chromatin regions, we grouped enhancers into enhancer groups based on location. Specifically, we sorted all predicted enhancers (including all cell types) by length and iteratively merged pairs that overlapped by at least 50% of the shorter enhancer's length. After each merge, the merged enhancer became the new candidate for subsequent merging. This process continued until no further overlaps were found. All predicted enhancers contributing to the same merged enhancers were considered to be in the same enhancer groups and were considered overlapping (**Figs. 1F,G; Fig. 2D)**.

Learning models to predict base-resolution cell type-specific scATAC-seq profiles from DNA sequence and identifying transcription factor motif instances with ChromBPNet

We applied the ChromBPNet model[58] to identify TF motif instances and predict the effects of noncoding variants on chromatin accessibility. Our application of this model to identify important TF motifs per cell type and specific TF motif instances involves four steps: (1) Training a model to predict chromatin accessibility from DNA sequence in each cell type; (2) Identifying sequences predicted to be important for chromatin accessibility by the model; (3) Identifying recurrent sequences (motifs) from the model *de novo*, and annotating them with possible TFs using existing motif databases; and (4) Identifying specific instances of these motifs across all chromatin accessible elements:

(1) To predict single nucleotide-resolution read count profiles from scATAC-seq data, we trained ChromBPNet (v0.1.7) models (https://github.com/kundajelab/chrombpnet)[58] — sequence-to-profile convolutional neural networks — for each of the cell types. ChromBPNet uses one-hot-encoded DNA sequence (A = [1, 0, 0, 0], C = [0, 1, 0, 0], G = [0, 0, 1, 0], T = [0, 0, 0, 1]) in a 2,114-bp window around scATAC-seq peaks (specifically, candidate elements defined by scE2G above) and is trained to predict scATAC-seq pseudobulk Tn5 insertion counts for the central 1,000 bp. In order to correct for the Tn5 bias, the ChromBPNet model also uses a Tn5 bias track as input, which was obtained from the ENCODE Project (ENCSR868FGK). The ChromBPNet model has two heads, which correspond to the total Tn5 insertion counts over the 1000 bp region and to the profile — a multinomial probability of Tn5 insertion counts at each position of the 1000 bp sequence, respectively. We used a 5-fold chromosome hold-out cross-validation with the training, evaluation, and test chromosomes used for each fold as follows. Test chromosomes: fold 0: [chr1], fold 1: [chr19, chr2], fold 2: [chr3, chr20], fold 3: [chr13, chr6, chr22] & fold 4: [chr5, chr16]. Validation chromosomes: fold 0: [chr10, chr8], fold 1: [chr1], fold 2: [chr19, chr2], fold 3: [chr3, chr20] & fold 4: [chr13, chr6, chr22]. Performance of the count head was evaluated based on Pearson and Spearman correlation of observed and predicted log(total counts). Performance of the profile head was evaluated using the Jensen Shannon Distance between the normalized base-resolution coverage profile (converted to a probability distribution over positions) and the predicted multinomial profile distribution from the model (**Table S17**).

(2) We used the DeepLIFT/DeepSHAP algorithm[56] in order to interrogate ChromBPNet models and estimate the predictive contribution of each nucleotide in a query sequence to the predicted log(total counts) from the model. The base-resolution contribution scores across any query sequence are an additive decomposition of the difference of the predicted log(total counts) for the query sequence and the predicted log(total counts) for 20 dinucleotide preserving shuffled versions (neutral reference) of the query sequence i.e. a log fold change. For each cell type, we obtained DeepLIFT/DeepSHAP contribution scores for fold 0.

(3) We then used TF-MoDISco (0.5.16.4.1) (https://github.com/kundajelab/tfmodisco)[57] to identify non-redundant motifs by clustering high-confidence subsequences with high contribution scores (seqlets) across all open chromatin peak regions for each cell type (**Table S15**). The 500-bp windows surrounding the peak centers were considered for motif discovery with fifty thousand maximum seqlets per metacluster. Following identification of motifs *de novo*, we annotate these motifs to candidate TFs or TF families by comparing their contribution scores against the MEME version 4 motif database using TOMTOM[125].

(4) In order to call the predictive instances (sequence matches with high contribution) of these motifs with high sensitivity and specificity, we scanned through contribution scores in open chromatin regions with the FiNeMo (v0.25) hit caller (https://github.com/austintwang/finemo_gpu), using alpha 0.7 and 5 as a hit coefficient threshold for reporting the instances (**Table S16**).

Predicting effects of noncoding variants on chromatin accessibility with ChromBPNet

In order to score and identify variants predicted to affect chromatin accessibility, we used the ChromBPNet models for each cell type to predict base-resolution scATAC-seq coverage profiles for 1 kb

genomic sequence windows containing reference and alternate alleles of selected variants. The effect size of each variant was estimated using two measures: the log2 fold change of the total predicted coverage (total counts) over each sequence window containing the reference and alternate allele, and the Jensen Shannon distance (JSD) between the base-resolution predicted probability profiles for the reference and alternative sequences to measure a change in the profile shape.

Statistical significance was estimated for both of these scores on empirical null distributions of variant scores that were computed as follows. We shuffled the 2114 bp sequence centered at each variant multiple times while preserving the dinucleotide frequency, made 2 copies of each shuffled sequence, then inserted each allele of the variant at the center of the shuffled sequence to create a set of 100 thousand null variants in total. We scored each of these null variants with each model using the same procedure that we used to score each observed variant. Then, for each observed variant, we calculated the proportion of the null variants that had an equally high or higher (more extreme) score to generate an empirical p-value for both the log2 fold change and JSD scores. The code base for scoring variants is at https://github.com/kundajelab/variant-scorer.

Defining gene programs with consensus non-negative matrix factorization

To identify co-expressed genes (programs) within the heart atlas, we used consensus cNMF[60] (v1.3.4). To increase the resolution of the programs identified, the cells in the heart atlas were grouped into cell type groups as specified in **Table S4**.

To create the input count matrix for cNMF for each cell type group, we first excluded gene categories that are less informative for the analysis. These categories include long noncoding RNAs ('^LINC'), gene isoforms ('-AS'), microRNAs ('^MIR'), small nuclear RNAs ('RNU'), genes of uncertain functions ('^LOC'), and genes lacking Ensembl stable ID or Entrez ID (bioconductor-org.hs.eg.db, v3.17.0). Furthermore, we filtered out cells with low gene abundance (fewer than 200 expressed genes). This threshold was more lenient than standard recommendations because the cell had been through the quality control step as described in **scRNA-seq quality control and normalization**. Finally, genes expressed in less than 10 cells were removed to focus on genes with a broader expression pattern (260-8000 genes after filtering).

The implementation of cNMF is described in Kotliar et al[60]. Briefly, NMF decomposes each count matrix (cell × gene) into two lower rank matrices. One matrix represents the contribution of each gene to each program (gene contribution matrix), and the other represents the proportions in which the programs are combined within each cell (usage matrix). To overcome the non-deterministic nature of NMF and achieve robust results, cNMF ran multiple replicates of NMF on the same normalized dataset with different randomly chosen seeds while using the same K, a parameter that determines the number of programs. The results are then averaged over multiple replicates after the outlier programs are removed.

Following the recommendation outlined by Kotliar et al[60], the "prepare" step of the algorithm was run with the following parameters: 'cnmf prepare -k 15 18 21 24 27 30 33 36 39 41 43 45 47 --n-iter 100 --seed 42". This step tested 13 different K values with 100 replicates for each K. Seed 42 was used to initialize the random seeds for the replicates. The "consensus" step was run with the parameter "--local-density-threshold 0.1", where 0.1 is the Euclidean distance threshold used to identify outlier programs and was chosen based on the clustergram diagnostic plot generated by cNMF.

Finally, to determine the optimal K for each cell group, we sought the smallest K that maximized the number of unique gene ontology (GO) terms enriched in the programs calculated by Gene Set Enrichment Analysis (clusterprofiler[126], v4.10.0; r-msigdbr[127] v7.5.1: category: C5, subcategory: BP), and at the same time, minimizing the difference between the original transcripts per million (TPM) matrix and the matrix reconstructed by multiplying the gene contribution matrix (in TPM unit) with the usage matrix.

Programs representing ribosomal contamination, cardiomyocyte contamination, or expressing other markers indicative of contamination (e.g., *ABL*1, *HBG1*), as well as programs expressed in less than 5% of any cell type, were removed from further analysis. This resulted in a total of 253 programs for downstream consideration

The K chosen for each cell type group are listed below:
ACM: 39; core_conduction: 36; Endocardial_endothelial: 49; Epicardial: 39; FB: 39; MuralCells: 41; NC: 41; VCM: 43; Lymphatic_EC: 15; TzConductionCells: 30; ImmuneCells: 41.

Identifying transcription factors regulating gene programs

We projected each cell to the motif space (cisbp, chromvarmotifs[128] v0.2.0) using chromVar[129] (v1.16.0). Next, we identified TFs whose per-cell motif accessibility significantly correlated (Spearman correlation > 0.1) with the expression of the program and were among the top 300 genes most highly associated with the program.

Cell-cell interactions

We employed CellChat (v2.1.2)[67] to infer cell-cell interaction (CCI) relationships, and more precisely the cell-type to cell-type interaction in this study (**Table S7**). The log1p transformed sequencing-depth normalized gene expression counts and cell type annotations (all except for pulmonary endothelial cells) were used as primary inputs. We performed the cell-cell interaction analysis one week at a time and selected the post conception week 8 data for visualization (**Fig. 2K**) because we have the largest number of total cells collected from whole hearts at this time point.

CellChat curated a comprehensive database of ligand-receptor pairs, which serves as the knowledge foundation for analyzing the intercellular communications. Each record in this database consists of a ligand-receptor pair and their corresponding encoding genes. As a single ligand or receptor protein may have multiple subunits, it may correspond to multiple genes. These ligand-receptor pairs are systematically categorized into distinct cell-cell interaction (CCI) pathways. For instance, interactions such as DLL1-NOTCH1 and JAG1-NOTCH2 are consolidated into the "NOTCH" pathway, while pairs involving WNT ligands (e.g., WNT1-FZD1, WNT3-FZD5) are grouped into the "WNT" pathway. In our analysis, we subset the database records to the expressed ligand and receptor genes in our atlas.

CellChat implements a computational approach to analyze cell-cell interactions. The analysis generates a three-dimensional tensor of ligand-receptor interaction strengths with dimensions K × K × N, where K represents the number of cell types, and N denotes the number of ligand-receptor pairs. The interaction strength calculation involves several steps: 1) computation of ensemble average gene expression within each cell type; 2) Implementation of a Hill function-based model to quantify interaction strengths. The Hill function model integrates multiple biological factors: expression levels of ligands and receptors; cell type population size; protein subunits; presence of co-stimulatory and inhibitory receptors; effects of extracellular agonists and antagonists. This comprehensive modeling approach ensures that the calculated interaction strengths reflect the biological complexity of cell-cell communication systems.

CellChat aggregates ligand-receptor pair level activities to pathway level activities to quantify cell-cell communication through three main steps. First, it identifies significant ligand-receptor pairs through statistical tests that randomly shuffle cell type labels, keeping only pairs with p-value <0.05. Then, using the the computeCommunProbPathway() function, these significant pairs are grouped by their signaling pathways and their strengths are summed, creating two matrices (incoming and outgoing signals) where rows represent types, columns represent CCI pathways, and values show the summed strengths of all the corresponding ligand-receptor in that pathway. Note that the row sum of these two matrices are used for feature selection. Finally, these two matrices are added together to create a final signaling strength

matrix, which reveals both the primary communication modes used by cells and the similarities in their communication behaviors.

Reproducible Jupyter notebook for the CCI analyses can be found in the GitHub repository https://github.com/chansigit/heartmap-cci-reproduction.

Annotation of disease genes for CHD and adult-onset cardiac diseases

We downloaded the initial list of CHD genes and their associated CHD phenotypes and classifications from CHDGene[8] (https://chdgene.victorchang.edu.au/, accessed May 16th, 2024). In CHDGene, "Genes are only included in the database if variants in the respective gene have been reported as the monogenic cause for CHD (isolated or in the context of a syndrome) in at least 3 independent familial or sporadic cases in one or more separate publications."[8] We modified the original CHD classification as follows (**Table S8**):

- **ASD**: we merged genes linked to ASD and ASD with minor abnormalities into a single "ASD" category. We removed *LTBP2* and *TMEM260* from this category because ASD is not a listed phenotype for *LTBP2*, and *TMEM260* is associated with persistent left superior vena cava, which might cause ASD through a different mechanism from other ASD causal genes.
- **VSD**: we combined genes linked to VSD and VSD with minor abnormalities into a "VSD" category and excluded genes associated with patent foramen ovale, patent ductus arteriosus (PDA), and bicuspid aortic valve (BAV). We included *DNAH5* into the list for its VSD association.
- **Valve defects**: we renamed the "Obstructive lesions" category to "Valve defects" category. This group now includes all genes with valve lesions in their CHD phenotypes. Genes solely associated with interrupted aortic arch or coarctation of the aorta but not any of the valve lesions were moved to the malformation of the outflow tract category.
- **Malformations of the outflow tract**: genes associated with interrupted aortic arch or coarctation of the aorta were added to this category.

We also analyzed high-confidence CHD genes from whole-exome sequencing by the Pediatric Cardiac Genomics Consortium, specifically the list of genes in which two or more *de novo* variants were detected in the cohort[7].

We obtained the causal genes for adult-onset cardiac diseases from ClinVar: https://ftp.ncbi.nlm.nih.gov/pub/clinvar/gene_condition_source_id (downloaded on July 29th, 2024).

Testing disease genes for preferential expression in particular cell types

We defined preferentially expressed genes as the top 300 most highly expressed genes among those with z-score > 1 (top 16%) across all cell types, after excluding all cell cycling cell types, pulmonary endothelial cells and lymphoid cells, as well as genes that are in the following categories: long noncoding RNAs ('^LINC'), gene isoforms ('-AS'), microRNAs ('^MIR'), small nuclear RNAs ('RNU'), genes of uncertain functions ('^LOC'), and ribosomal genes ("^RPS" and "^RPL"). To test whether the disease genes are enriched in any particular cell type, we conducted a one-sided Fisher's exact test. Background genes were defined as those expressed in any of the considered cell types (TPM>1). We corrected for multiple testing across all cell types within each disease category (high-confidence CHD genes from CHDGene and adult-onset cardiac diseases) using Benjamini-Hochberg correction.

Assessing cell type heritability enrichment with stratified LD score regression

We used stratified LD score regression[130] (S-LDSC, v1.0.1) to estimate the enrichment of disease or trait heritability within cell type-specific enhancers. These enhancers were defined as all distal enhancers predicted by scE2G for each cell type. We preprocessed each GWAS summary statistics using the munge_sumstats.py provided by LDSC[131] and filtered to retain only SNPs curated by HapMap 3 as recommended[72,131]. We then performed S-LDSC with our cell type-specific scE2G enhancer annotations

along with the baseline model (v.2.2). The 1000 Genome EUR Phase 3 genotype data was used as the LD reference panel. Full S-LDSC results can be found in **Table S12.**

Fine-mapping of GWAS signals for quantitative cardiac traits

Genome-wide association studies (GWAS) were conducted for deep learning-based left atrial (LA) measurements[132] and left ventricular (LV) measurements[133]. Deep learning semantic segmentation (pixel labeling) models for UK Biobank cardiovascular magnetic resonance imaging (MRI) 2-, 3-, and 4-chamber long-axis views and short axis stacks were trained using PyTorch[134]. These were reconstructed into 3D surfaces separately for the left atrium[135] and left ventricle[36] using Poisson surface reconstruction[136], permitting volume measurements in systole and diastole. GWAS was conducted with REGENIE using data from 63,196 participants with LA and LV measurements who did not have atrial fibrillation (which alters cardiac chamber volumes due to eliminating atrial filling and contraction). Adjustment was made for covariates including age, age^2, genetic PC1-10, sex, the genotyping array, and the imaging device.

Semantic segmentation was performed to measure aortic root diameter and ascending aortic diameter from UK Biobank MRI, and GWAS was conducted with REGENIE using data from 62,936 participants with aortic root diameter measurements, 62,716 with ascending aortic diameter measurements[137]. A meta-analysis of thoracic aortic aneurysm disease GWAS was produced using the METAL software package[138] on data from three sources: Million Veteran Program[139], FinnGen release #10, and UK Biobank (effective sample size: 53,516). Multi-trait analysis of GWAS (MTAG)[140] was conducted using summary statistics from the three measurements (root diameter, ascending and descending thoracic aortic diameter) and the thoracic aortic aneurysm meta-analysis[90] .

Semantic segmentation was performed to measure aortic valve hemodynamic measurements (aortic valve area (AVA), mean gradient, and peak velocity) from UK Biobank MRI[88], and GWAS was conducted with REGENIE using data from 59,569 participants with AVA measurements and 59,571 with mean gradient and peak velocity measurements[87]. Adjustment was made for covariates including age, age^2, genetic PC1-10, sex, the genotyping array, and the imaging device. A meta-analysis of aortic stenosis disease GWAS was produced using the METAL software package configured to account for overlapping samples[138] on data from four sources: Million Veteran Program[141], FinnGen release #12, the TARGET multi-cohort GWAS[142], and UK Biobank (maximum effective sample size: 164,971). MTAG[140] was conducted using summary statistics from the hemodynamic measurements (AVA, mean gradient, and peak velocity) and the aortic stenosis meta-analysis[89].

We defined fine-mapping regions based on a 3 Mb window around each lead variant and merged regions if they overlapped[143]. LD was then computed using LDstore v2.0[144] for these regions using in-sample UK Biobank imputed genotypes. Fine-mapping was then performed using SuSiE-inf[145] (v1.2) with default settings on the phenotypes derived from UK Biobank MRI data with in-sample, covariates-adjusted LD (see sections above for covariates). Credible set coverage is set to 0.95 and purity is set to 0.5, meaning that 95% of the credible sets contain a true causal variant and that the minimal pairwise LD between all variants in each credible set is 0.5.

Preparing variants for overlapping with scE2G-predicted enhancers

We included three sets of variants for our analysis. The first set was fine-mapped using SuSiE-inf as described in Methods above. The second set was fine-mapped using SuSiE as implemented by FINNGEN[99]. The third set underwent Linkage Disequilibrium (LD) expansion. See **Table S11** for the methods used for each trait.

Specifically, for variants that were fine-mapped using SuSiE-inf, we first removed all variants that are not in a credible set (cs= -1) and have PIP values below 0.9. In the remaining variants that are in credible sets (cs != -1), the variants with the highest PIP (posterior inclusion probability) were designated as the

lead variants of their respective credible sets. Additionally, variants outside credible sets (cs= -1) but with PIP greater or equal to 0.9 were individually considered as lead variants.

For variants fine-mapped by FINNGEN[99], the summary statistics and variants within 95% credible sets were downloaded from gs://finngen-public-data-r8. Both fine-mapped datasets were filtered to include only variants with PIP over or equal to 0.1.

For GWAS results with only summary statistics or lead SNPs available, we performed LD clumping using PLINK (v1.9) with parameters "--clump-p1 5e-8 --clump-p2 0.01 --clump-r2 0.1" to identify a single most significant SNP per LD block. These SNPs were defined as the lead SNPs and subsequently expanded to include variants in LD ($r^2 \geq 0.9$) using PLINK parameters "--ld-window-kb 1000 --ld-window 99999 --ld-window-r2 0.9" and the 1000 Genome EUR Phase 3 reference panel.

Linking GWAS variants to enhancers, target genes, cell types, and gene programs

We applied our gene regulation map to interpret noncoding GWAS variants by linking them to target cell types, enhancers, genes, and gene programs. To do so, we started with the list of fine-mapped or LD-expanded variants described above. We then linked variants to genes and cell types with scE2G. Specifically, we overlapped variants with scE2G-predicted enhancer-gene regulatory interactions and filtered for the top two target genes with the highest quantile-normalized scE2G scores for each enhancer. For this step, we also included variants linked to GWAS signals that overlapped coding regions or splice sites (n=129 GWAS signals) ("AnyVariantInCSOverlapCoding" and "AnyVariantInCSOverlapSpliceSite" in **Table S13**). This approach builds on previous methods we have developed to interpret GWAS signals using enhancer-to-gene maps[54,61,124,146], extending them to enable interpreting variants across a tissue atlas including many dozens of cell types. This yielded a list of 770 unique GWAS signals linked to one or more target genes (**Table S13**).

To further annotate this set of 770 GWAS signals, we assessed whether the predicted target genes were part of a cNMF gene program in the same cell type as the scE2G prediction. We considered a gene to be part of a gene program in a given cell type if the gene was ranked among the top 300 genes in the program based on importance scores and at least 10% of cells expressing the gene program (defined as having a program usage score > 10%) belonged to the corresponding cell type.

For a subset of variants of interest, we interpreted the local effect of the variant on chromatin accessibility in that cell type using ChromBPNet. Specifically, we predicted the fold-change in chromatin accessibility at the predicted enhancer by substituting the alternative allele into the center of the ~2-Kb input sequence of ChromBPNet. We also examined contribution scores of nucleotides across the window for both the reference and alternative alleles, and applied FiNeMo to identify predicted motif instances that are affected by the variant (see **Methods** above).

Assessing enrichment of genes associated with valve traits in VIC program 27

Genes associated with valve traits are defined as ones that are predicted to be regulated by GWAS variants linked to aortic stenosis, aortic valve area, aortic root diameter, aortic valve peak velocity, aortic valve mean gradient, mitral valve prolapse. To test whether these genes are enriched in VIC program 27, we conducted a one-sided Fisher's exact test. Background genes were defined as those expressed in any of the 5 VIC cell types (TPM>1).

Enhancer reporters: Mouse transgenic assay

Transgenic E11.5 mouse embryos were generated as described previously[147]. Briefly, super-ovulating female FVB mice were mated with FVB males and fertilized embryos were collected from the oviducts.

Enhancer sequences were selected to encompass the relevant epigenetic signal and regions of high evolutionary conservation**.** The selected sequences were synthesized by Twist Biosciences and cloned into the donor plasmid containing minimal Shh promoter, lacZ reporter gene and H11 locus homology arms (Addgene, 139098) using NEBuilder HiFi DNA Assembly Mix (NEB, E2621). The sequence identity of donor plasmids was verified using long-read sequencing (Primordium). Plasmids are available upon request. A mixture of Cas9 protein (Alt-R SpCas9 Nuclease V3, IDT, Cat#1081058, final concentration 20 ng/µL), hybridized sgRNA against H11 locus (Alt-R CRISPR-Cas9 tracrRNA, IDT, cat#1072532 and Alt-R CRISPR-Cas9 locus targeting crRNA, gctgatggaacaggtaacaa, total final concentration 50 ng/µL) and donor plasmid (12.5 ng/µL) was injected into the pronucleus of donor FVB embryos. The efficiency of targeting and the gRNA selection process is described in detail in Osterwalder 2022.[147]

Embryos were cultured in M16 with amino acids at 37oC, 5% CO2 for 2 hours and implanted into pseudopregnant CD-1 mice. Embryos were collected at E11.5 for lacZ staining as described previously[147]. Briefly, embryos were dissected from the uterine horns, washed in cold PBS, fixed in 4% PFA for 30 min and washed three times in embryo wash buffer (2 mM MgCl2, 0.02% NP-40 and 0.01% deoxycholate in PBS at pH 7.3). They were subsequently stained overnight at room temperature in X-gal stain (4 mM potassium ferricyanide, 4 mM potassium ferrocyanide, 1 mg/mL X-gal and 20 mM Tris pH 7.5 in embryo wash buffer). The stained transgenic embryos were washed three times in PBS and imaged from both sides using a Leica MZ16 microscope and Leica DFC420 digital camera.

PCR using genomic DNA extracted from embryonic sacs digested with DirectPCR Lysis Reagent (Viagen, 301-C) containing Proteinase K (final concentration 6 U/mL) was used to genotype the embryos[147,148]. Specifically, a PCR spanning the 3' H11 homology arm was used to confirm successful integration of the reporter construct at the H11 locus and a PCR to detect the presence of the bacterial backbone of the reporter plasmid was used to classify the embryo as harboring a "tandem" (concatameric) rather than "single" integration. Only embryos with donor plasmid integration at H11 were used. Embryos with "tandem" integrations tend to have stronger reporter activities than those with "single" integrations, which was taken this into consideration when assessing the impact of genetic variant introduction on enhancer activity.

Data Availability

Processed, de-identified 10x Multiome data are available at https://www.synapse.org/Synapse:syn63997960. Raw sequencing data (10x Multiome) will be available via dbGaP, in compliance with the study design and consent.

Acknowledgements

We thank Salil Deshpande, Laksshman Sundaram, Kristy Red-Horse, and members of the Engreitz Lab and BASE Initiative for helpful discussions. We thank Mo Ameen, Simon Miao, Betty Liu, and William Greenleaf for technical advice on single-cell tissue dissociation methods. We thank Stanford University and the Stanford Research Computing Center for providing computational resources and support for the Sherlock high-performance compute cluster. We thank the Stanford Genomics Service Center for generating sequencing data used in this publication.

Funding for this project was provided by the Applebaum Foundation, Additional Ventures (to J.M.E. and W.G.), Stanford Maternal and Child Health Research Institute (to J.M.E.), the BASE Research Initiative

at the Lucile Packard Children's Hospital at Stanford University (to J.M.E.), and the National Institutes of Health (R24HD000836 to I.A.G.). In addition:

- J. M. Engreitz acknowledges support from the Novo Nordisk Foundation Center for Genomic Mechanisms of Disease (NNF21SA0072102); the NIH/NHGRI Impact of Genomic Variation on Consortium (UM1HG011972); and NIH/NHLBI (R01HL159176)
- W. R. Goodyer acknowledges support from NIH (K08 Mentored Clinical Scientist Research Career Development Award - 1K08HL15378501).
- X. R. Ma acknowledges support from NIHGRI training grant (2T32HG000044-21)
- M. U. Sheth acknowledges the support of an NSF Graduate Research Fellowship (DGE-1656518) and a graduate fellowship award from Knight-Hennessy Scholars at Stanford University.
- 'The Laboratory for Developmental Biology' was supported by NIH Award Number 5R24HD000836 (to I.A.G) from the Eunice Kennedy Shriver National Institute of Child Health & Human Development.
- W. Qiu and R. Andersson acknowledge support from the Novo Nordisk Foundation (NNF20OC0059796) and the Novo Nordisk Foundation Center for Genomic Mechanisms of Disease (NNF21SA0072102).
- H. Y. Kang acknowledges support from NIH Training Grant 5T32GM087237 and NIH 5R35GM141861-04.
- D. Amgalan acknowledges support from AHA Postdoctoral Fellowship (821920 and 23POSTCHF1019753) and the Stanford Maternal and Child Health Research Institute (MCHRI) Postdoctoral Support.
- T. Quertermous acknowledges support from NIH grants R01HL171045, R01 HL134817, R01HL139478, R01HL156846, R01HL151535, R01HL158525, UM1 HG011972, and the Irwin Foundation.
- M. Rabinovitch was supported by NIH/NHLBI grant R01 HL152134 and NIH/NHGRI grant HG011972, the Dwight and Vera Dunlevie Chair in Pediatric Cardiology at Stanford University, and the Gordon and Betty Moore and the BASE Research Initiative at the Lucile Packard Children's Hospital at Stanford University.
- X. Qiu acknowledges support from NIH/NHGRI R00HG012887 and NIH DP2HG014282.
- The research of M. Kosicki, A. Visel and L.A. Pennacchio was supported by National Institutes of Health (NIH) grants R01HL162304 and R01HG003988 (to A.V. and L.A.P.) and conducted at the E.O. Lawrence Berkeley National Laboratory under the US Department of Energy Contract DE-AC02-05CH11231, University of California. We thank the technical team at the Lawrence Berkeley National Laboratory for generating transgenic mouse embryos.
- A. Kundaje acknowledges support from the Impact of Genomic Variation on Function Consortium (U01HG012069).
- J. P. Pirruccello acknowledges support from NIH grant K08HL159346.
- M. Gu acknowledges support from Additional Ventures (1019125) and NIH/NHLBI (R01HL166283)
- S. Kany acknowledges support from the Deutsche Forschungsgemeinschaft (521832260)
- Arun Padmanabhan acknowledges support from NIH grant K08HL157700, the Michael Antonov Charitable Foundation, and the Frank. A. Campini Foundation.

Author Contributions

X.R.M. and J.M.E. led analysis to create the atlas and interpret genetic variants.
S.D.C., X.R.M., W.R.G., and J.M.E. led fetal tissue sample acquisition and multiome data collection.
S.D.C., B.D.R.L., and I.G. collected fetal tissue.
S.D.C., X.R.M., D.A. C.J.M., O.M.R., M.G., J.M.E, and W.R.G. developed experimental methods for multiome data collection.
S.D.C., L.D., S.T., K.D., and W.R.G. conducted multiome experiments.
X.R.M. developed and applied methods to analyze multiome data.

X.R.M., C.G., M.R., T.Q., J.M.E., and W.R.G. annotated cell types and states.
D.B. and A.K. trained ChromBPNet models to interpret variants and identify TF motif instances.
X.R.M., M.U.S., W.Q., R.A., and J.M.E. developed and applied the scE2G model.
J.P.P., J.R.P., S.K., S.Z., J.D., A.P, J.O., and S.D. contributed genome-wide association study results.
R.C. and J.P.P. conducted statistical fine-mapping and interpreted analysis of GWAS variants.
S.C. and X.Q. analyzed cell-cell interactions.
M.K., X.R.M., L.P., A.V., and J.M.E. designed, conducted, or interpreted enhancer reporter experiments.
S.D.C., J.M.E., and I.G. managed regulatory compliance.
W.R.G. and J.M.E. obtained funding and supervised the study.
X.R.M., S.D.C., W.R.G., and J.M.E. wrote the manuscript with input from all authors.

Conflict of Interest Statement

J.M.E. has received materials from 10x Genomics unrelated to this study.

Supplementary Notes

Note S1. Evidence linking CHD genes to functions in VICs and cardiac fibroblasts

We found that genes previously associated with multiple subtypes of CHD were enriched for high expression in valvular interstitial cells and cardiac fibroblast progenitors (**Fig. 3**). In total, 27 genes were highly expressed in either VICs or cardiac fibroblasts. Many of them have been studied using mouse knockouts, including via conditional knockouts in relevant cell lineages (endocardial lineages, neural crest, and epicardial for VICs[149]; and epicardial lineage for cardiac fibroblasts[78]. This supports the notion that CHD genes can have direct effects on these cell types to lead to heart defects. We describe these previous studies below, and summarize in **Table S10**.

- **BMPR2**: *BMPR2* (bone morphogenetic protein receptor type 2) plays an essential role in heart valve development, particularly through its involvement in endocardial and regulation of valvulogenesis. Conditional knockout studies using Cre-loxP lineage tracing have revealed tissue-specific functions of *BMPR2* during cardiogenesis. Specifically, conditional deletion of *Bmpr2* with Tie2-Cre, which targets endocardial cells and derived mesenchymal cells, leads to atrioventricular cushion defects, including membranous VSDs and thickened valve leaflets, but does not block EMT[150]. In addition to its function in the endocardial lineage, neural crest lineage deletion of *Bmpr2* (Wnt1-Cre) leads to abnormal positioning of the aorta[150]. In contrast, myocardial lineage knockout of *Bmpr2* did not affect heart development[150]. In humans, mutations in *BMPR2* are observed in patients with pulmonary arterial hypertension (PAH)[150,151], ASD, VSD, patent ductus arteriosus (PDA), and malformation of the outflow tract[151].

- **FGFR2** - *FGFR2* (Fibroblast Growth Factor Receptor 2) is important for cardiac development, particularly in the formation and alignment of the cardiac outflow tract (OFT). In humans, mutations in *FGFR2* are associated with CHD, including VSD, double-outlet right ventricle (DORV), and hypoplastic ventricles[152–155]. Conditional knockout studies in mice using the Cre-loxP system have shown tissue-specific roles of *FGFR2* during cardiogenesis. *Fgfr2* deletion in mesodermal cells of the second heart field (SHF) using Nkx2.5-Cre, which targets SHF-derived OFT progenitors, results in severe OFT defects, such as overriding aorta, DORV, and persistent truncus arteriosus (PTA). These defects show *FGFR2*'s role in OFT morphogenesis[155]. Notably, dual inactivation of *Fgfr1* and *Fgfr2* in the second heart field (SHF) lineage further confirms *Fgfr2*'s function in regulating endocardial epithelial-to-mesenchymal transition (EMT) and neural crest cell recruitment to the OFT cushions[155]. Germline knockout of Fgfr2b resulted in thin-walled myocardium and smaller hearts at embryonic day 17.5 (E17.5), indicating that Fgfr2b signaling is essential for proper heart development[156]. In a conditional knockout study, *Fgfr1* and *Fgfr2* were inactivated in the epicardium and epicardial-derived cells of mice using Wt1-Cre. These mice exhibited reduced epicardial Fgfr2 expression, decreased myocardial proliferation, and thinner myocardial walls[156].

- **GATA4**: *GATA4* (GATA Binding Protein 4) is a transcription factor essential for heart development, particularly in heart valve formation and maturation, through its involvement in endocardial epithelial-to-mesenchymal transition (EndoMT) and mesenchyme formation within atrioventricular cushions. Conditional knockout studies using Tie2-Cre lineage tracing, which targets endocardial cells and endocardial cell-derived VICs, show that *Gata4* deletion disrupts AV cushion development[157]. This disruption results in hypoplastic valves with few VICs, leading to defective AV septation, hypocellular cushions, and malformed valves. Specifically, *Gata4* inactivation in the endocardial lineage impairs EndoMT, reducing mesenchymal cell formation and resulting in unseptated ventricular inlets and incomplete cushion remodeling[157]. In our atlas, *GATA4* is expressed across multiple cell types, including VICs, where it is predicted to regulate gene programs involving other CHD-related genes, such as *LTBP2*, *MEIS2*, and *TBX20*. Human *GATA4* mutations are associated with various CHDs, including atrioventricular septal defects,

pulmonic stenosis, and hypoplastic right ventricle, as well as ASDs and VSDs and valve malformations[157,158].

- **GATA6**: *GATA6* (GATA Binding Protein 6) is a transcription factor involved in heart development, especially in valve formation and septation. *GATA6* mutations in humans are also associated with PTA, ASDs, and DORV, which are linked to defects in outflow tract development and valve morphogenesis[23,159]. Wnt1-Cre-mediated conditional *Gata6* knockout in neural crest-derived cells in mice results in CHD, such as PTA and interrupted aortic arch. These defects arise from disrupted morphogenetic patterning and impaired migration of cardiac neural crest cells, accompanied by downregulation of key signaling molecules like semaphorin 3C essential for cardiac neural crest cell function[159]. Additionally, deletion of *Gata6* with an Isl1-Cre mouse model targeting second heart field (SHF) progenitors leads to defects in the atrioventricular canal, including valve malformations and ASD, by impacting the proliferation and differentiation of SHF-derived cells[160]. Epicardial-specific knockout of *Gata4/Gata6* causes a thin compact myocardium and coronary artery formation defects, which are present in some patients with Tetralogy of Fallot. Furthermore, *GATA6* is highly expressed in epicardial-derived fibroblasts, regulating gene programs vital for myocardial and coronary vessel development[158,161].

- **GLI3**: Disruption in *GLI3*'s (GLI Family Zinc Finger 3) activity, often linked to ciliopathies, is associated with defects in heart valve development, such as myxomatous mitral valve and bicuspid aortic valve[162,163]. These conditions are marked by abnormal extracellular matrix composition and VIC differentiation[162]. Primary cilia, which are present on VICs during early development, play a mechanosensory and signaling role[163]. Ablation of genes critical for ciliogenesis, like *Ift88*, results in loss of cilia and subsequent valve enlargement, ECM disruption, and fibroblast-like cell differentiation[163]. Genetic models using Cre drivers like NfatC1-Cre (targeting endocardial cels and VICs) to knock out cilia-related genes show a higher prevalence of CHD, including bicuspid aortic valve and myxomatous mitral valve, with affected mice displaying altered valve morphology, increased ECM components like collagen and proteoglycans, and decreased cell density[162,163].

- **GPC3**: To our knowledge, *GPC3* (Glypican-3) has not yet been studied in conditional knockout models within endocardial, neural crest and epicardial lineages. In humans, mutations in *GPC3* result in Simpson-Golabi-Behmel syndrome (SGBS), an X-linked disorder that presents a range of cardiac defects. These include ASDs, VSDs, PDA, patent foramen ovale, bicuspid aortic valve, dysplastic tricuspid and pulmonary valves, hypoplastic left pulmonary artery, and, in some cases, arrhythmias like supraventricular tachycardia and conduction defects[164]. In *Gpc3* knockout mouse models, the loss of *Gpc3* function leads to similar structural defects, including VSD, common atrioventricular canal, double outlet right ventricle, and coronary artery fistulas[165]. Valve abnormalities observed in *Gpc3*-deficient mice involve atrioventricular canal defects, with bridging valve leaflets and fibrous discontinuity between the aortic and mitral valves[165]. In our atlas, *GPC3* is expressed primarily in mesenchymal cells, with high expression in epicardial-derived cardiac fibroblasts, outflow tract fibroblasts, and VICs (**Fig. S12**).

- **INVS**: In mouse models, global *INVS* (inversin) mutation (insertional mutagenesis) leads to a range of CHD, including pulmonary infundibular stenosis, VSDs, and abnormalities in the right ventricular outflow tract[166]. To our knowledge, this gene has not been studied using conditional knockouts in the endocardial, neural crest, or epicardial, lineage. *INVS* (also known as *NPHP2*) mutations in humans are associated with VSDs[167,168], mitral insufficiency, and pulmonary valve stenosis[168].

- **LTBP2**: In *LTBP2* (Latent-transforming growth factor beta-binding protein 2) knockout mouse models, absence of *LTBP2* results in a high prevalence of mitral valve prolapse[77]. To our knowledge, this gene has not been studied using conditional knockouts in the endocardial, neural crest lineage, or epicardial. In humans, mutations in *LTBP2* have been linked to CHD such as

mitral valve prolapse and polyvalvular heart dysplasia[74,77].

- **MEIS2**: *MEIS2* (Meis homeobox 2) is a transcription factor involved in heart and craniofacial development, particularly influencing the morphogenesis of neural crest-derived tissues. In humans, *MEIS2* haploinsufficiency has been associated with a range of CHD, including VSD, ASD, coarctation of the aorta, Tetralogy of Fallot, hypoplastic right ventricle, and Ebstein's anomaly[82]. Conditional knockout studies in mice have shown the essential roles of MEIS2 across different cardiac lineages. Neural crest-specific deletion of Meis2 (AP2α-Cre) leads to CHD such as PTA, where the aorta and pulmonary arteries fail to separate, as well as valve malformations including defects in the aortic, pulmonary, tricuspid, and mitral valves. These defects are associated with embryonic lethality around E14 due to hemorrhaging and impaired cardiac neural crest migration[84]. Additionally, epicardial-specific deletion of both *Meis2* and its paralog, *Meis1*, using Wt1-Cre has been shown to reduce the proportion of epicardial-derived cardiac fibroblasts, leading to VSDs, misalignment of the great vessels, asymmetric septation of the outflow tract, and abnormalities in coronary artery development[84]

- **NF1**: *NF1* (neurofibromatosis type 1) plays a role in heart development, in particular by regulating EndoMT during valvulogenesis[169]. In humans, mutations in *NF1* are associated with CHD, including pulmonary valve stenosis, ASD, and VSD. Conditional knockout studies in mice using Cre-loxP lineage tracing have shown tissue-specific functions of *NF1* in cardiogenesis: epicardial-specific deletion of *Nf1* using Wt1CreERT2 results in accelerated EMT, leading to an increased number of epicardial-derived cells, such as fibroblasts, migrating into the myocardium and contributing to defects in the ventricular septum and atrioventricular valves[169]. Additionally, endothelial cell-specific knockouts show thickened endocardial cushions due to enhanced EMT, which correlates with CHD like ventricular septal defects and outflow tract abnormalities[169,170].

- **NOTCH2**: *NOTCH2* (Neurogenic locus notch homolog protein 2) plays a role in heart development, particularly through its involvement in neural crest-derived mesenchymal cells and the morphogenesis of the outflow tract. In humans, mutations in *NOTCH2* are associated with CHD such as peripheral pulmonary stenosis, aortic coarctation, and complex cardiac anomalies linked to Alagille syndrome[171]. Conditional knockout studies in mice using Cre-loxP lineage tracing have uncovered tissue-specific functions of *NOTCH2* during cardiogenesis. Specifically, deletion of *Notch2* with Pax3-Cre, which targets cardiac neural crest cells, results in narrowed aorta and pulmonary arteries, mimicking defects seen in Alagille syndrome. Similarly, Tag1-Cre-mediated inactivation of *Notch2* also leads to outflow tract defects, characterized by smaller vessel diameters due to decreased proliferation of smooth muscle cells derived from mesenchymal cells. In contrast, deletion of *Notch2* in secondary heart field derivatives using Mef2c-AHF-Cre did not cause outflow tract abnormalities, suggesting a neural crest lineage-specific function for *Notch2*[171].

- **PBX1**: *PBX1* (PBX Homeobox 1) is a TALE-class homeodomain transcription factor that contributes to the septation of the OFT, endocardial development, VSDs, conotruncal malformations, semilunar valves, and patterning of the great arteries[85,172]. In *Pbx1*-null mouse models, severe cardiovascular anomalies are observed, including PTA, VSDs, and abnormal great artery patterning. These defects primarily result from the failure of OFT septation, which leads to the formation of a single arterial trunk, as well as impaired valve development, where a single valve forms instead of distinct semilunar valves. *Pbx1* mutants also display additional vascular abnormalities, such as cervical aortic arch and aberrant subclavian artery origins, indicating a broader role in arterial patterning. Lineage tracing studies using a Wnt1-Cre mouse model, which targets neural crest cells, shows that cardiac neural crest cells can still migrate to the OFT in Pbx1-null embryos[85,172].

- **PRKD1**: In humans, *PRKD1* (Protein kinase D1) mutations are linked to CHD such as atrioventricular septal defects, bicuspid aortic valve, pulmonary stenosis, and coarctation of the

aorta[173]. In a *Prkd1* transgenic model with a CRISPR/Cas9-mediated deletion of exon 2, homozygous mice exhibited severe CHD, such as atrioventricular septal defects, bicuspid aortic valve, dysplastic pulmonary valves, and VSDs[173].

- **RBFOX2**: *RBFOX2* (RNA Binding Fox-1 Homolog 2) is an RNA-binding protein that regulates alternative splicing networks essential for cardiac development, particularly in the formation of heart chambers and valves in humans. Conditional deletion of *Rbfox2* with Nkx2.5-Cre, which affects early cardiac progenitors, disrupts chamber heart structure and valve development[174]. *RBFOX2* loss impairs cell-ECM adhesion and endocardial-to-mesenchymal transition, processes critical for valve formation and the proliferation of endocardial and myocardial cells. The model also suggests that *RBFOX2* is required across multiple cardiac cell lineages, as demonstrated by its broad expression in myocardial, endocardial, and contractile cells, which collectively contribute to normal cardiovascular morphogenesis[174].

- **SETD5**: *SETD5* (SET Domain Containing 5) haploinsufficiency is linked to CHD, including VSDs and outflow tract abnormalities in humans[175]. Deletion of *Setd5* using the Mesp1-Cre driver, which targets early mesodermal progenitors including the first and second heart fields, results in severe cardiac malformations, such as shortened outflow tracts, abnormal chamber ballooning, and embryonic lethality by E12.5[175].

- **SMAD6**: During valve development, *SMAD6* (SMAD Family Member 6) is expressed in endocardial and mesenchymal cells, contributing to regulated mesenchymal cell proliferation. In humans, disruptions in SMAD6 function have been associated with CHD, including septation abnormalities and valve malformations. In SMAD6-deficient knockout mouse models, a lack of SMAD6 leads to hyperplasia of the cardiac valves, marked by excessive mesenchymal cell proliferation and an enhanced endocardial-to-mesenchymal transformation. Additionally, mutants show OFT defects, often resulting in a narrowed ascending aorta or an enlarged pulmonary trunk[176].

- **TBX20**: *TBX20* (T-Box Transcription Factor 20) plays a role in cardiac valve development, exerting lineage-specific influences that contribute to CHD across endocardial, myocardial, and VIC lineages. *TBX20* mutations in humans are linked with a wide array of congenital heart anomalies, such as ASD, VSD, tetralogy of Fallot, total anomalous pulmonary venous connection, and cardiomyopathies with defective valvulogenesis[177–179]. In murine models, conditional knockout studies illustrate cell-specific roles of *Tbx20* during valvulogenesis. Nfatc1-Cre-mediated knockout, targeting endocardial cells and their mesenchymal derivatives, leads to valve elongation defects, shortened leaflets, and structural abnormalities in all four cardiac valves without inhibiting EMT initiation[178]. This deletion disrupts Wnt/β-catenin signaling, resulting in reduced cellular proliferation and impaired extracellular matrix organization within valve cushions. In contrast, myocardial lineage deletion of *Tbx20*, important for early cardiac patterning, shows no effect on valve formation[178]. Partial knockdown of *Tbx20* in mice also results in severe valve developmental defects[180]. Furthermore, *TBX20*'s role may extend into regulatory functions within VICs, where it could influence factors like VIC program 27.

- **TGFBR1**: *TGFBR1* (transforming growth factor-beta (TGF-β) receptor type 1) plays a role in myocardial wall formation and cardiac morphogenesis[181]. In humans, *TGFBR1* mutations are linked to VSDs, Marfan syndrome, and Loeys–Dietz syndrome.[181]. Conditional deletion of *Tgfbr1* with Tie2-Cre, targeting endocardial cells, results in hypoplastic atrioventricular cushions, impaired EMT, and thin myocardium, indicating its role in endocardial EMT and myocardial proliferation[182]. Epicardial-specific knockout using Gata5-Cre leads to detached epicardium, reduced myocardial proliferation, and defective smooth muscle formation around coronary vessels, showing its importance in epicardial-mesenchymal transformation[182]. Studies using myocardial-specific inactivation of *Tgfbr1* with cTnt-Cre mice demonstrate that while most embryos do not show overt cardiac defects due to compensation by other TGFβ receptor genes,

constitutively active *TGFBR1* expression can arrest heart development at the looping stage and lead to ventricular hypoplasia[181]. In neural crest cells, targeted deletion of *Tgfbr1* with a Wnt1-Cre model results in severe aortic arch malformations and PTA, showing its role in outflow tract development[181]. Furthermore, *Tgfbr1* expression in VICs and its regulation by miRNAs show its importance in maintaining proper signaling balance during heart development to prevent non-compaction and spongy myocardial wall defects[181].

- **ZFPM2**: *ZFPM2* (Zinc finger protein), also known as *FOG2*, regulates coronary vessel formation and morphogenesis. In humans, mutations or disruptions in *ZFPM2* have been implicated in septal defects and outflow tract malformations[183]. Knocking out *ZFPM2* in mice leads to multiple CHD, including ventricular myocardium thinning, ASD, and features of tetralogy of Fallot, such as VSDs, overriding aorta, and subpulmonic stenosis[183,184]. The absence of *ZFPM2* disrupts epithelial-to-mesenchymal transition in epicardium-derived cells, which affects the development of coronary vasculature, fibroblasts, and other mesenchymal derivatives. Rescue experiments indicate that ZFPM2 function in the myocardium is essential for inducing signals required for vascular formation[183,184].

Supplementary Figures

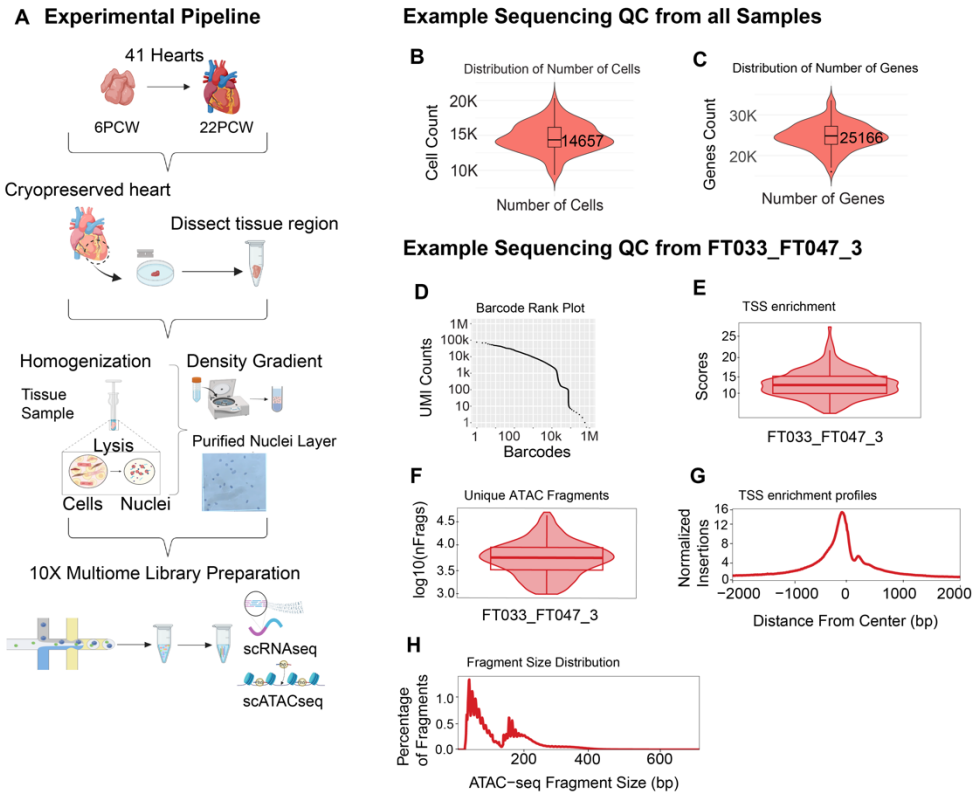


Fig. S1. Experimental pipeline and quality control metrics for single-cell 10x Multiome data

A. Experimental pipeline: Illustration of the workflow used to process 41 human fetal hearts ranging from 6 post-conception weeks (PCW) to 22 PCW. After cryopreservation, specific heart tissue regions were dissected, followed by homogenization and lysis to release nuclei. A density gradient was applied to purify nuclei. The nuclei were used for 10x Multiome library preparation, generating single-cell RNA sequencing (scRNA-seq) and single-cell ATAC sequencing (scATAC-seq) libraries[185].

B. Number of cells: Violin plot depicting the distribution of cell counts per 10x lane.

C. Number of genes: Violin plot showing the total number of genes detected per 10x lane.

D. Barcode rank plot: UMI counts across barcodes in a representative sample (FT033_FT047_3).

E. TSS enrichment scores: Violin plot illustrating the enrichment scores for transcription start site (TSS) regions in the sample FT033_FT047_3, indicating the quality of ATAC-seq data.

F. Unique ATAC fragments: Violin plot displaying the distribution of unique ATAC fragment counts per cell on a log10 scale for the sample FT033_FT047_3.

G. TSS enrichment profiles: Line graph showing the normalized insertion profile centered around transcription start sites for sample FT033_FT047_3.

H. ATAC-seq fragment size distribution: Histogram representing the distribution of ATAC-seq fragment sizes for sample FT033_FT047_3. This fragment size distribution shows that the ATAC-seq data has successfully captured a range of open chromatin regions, including nucleosome-free DNA and nucleosome-bound DNA.

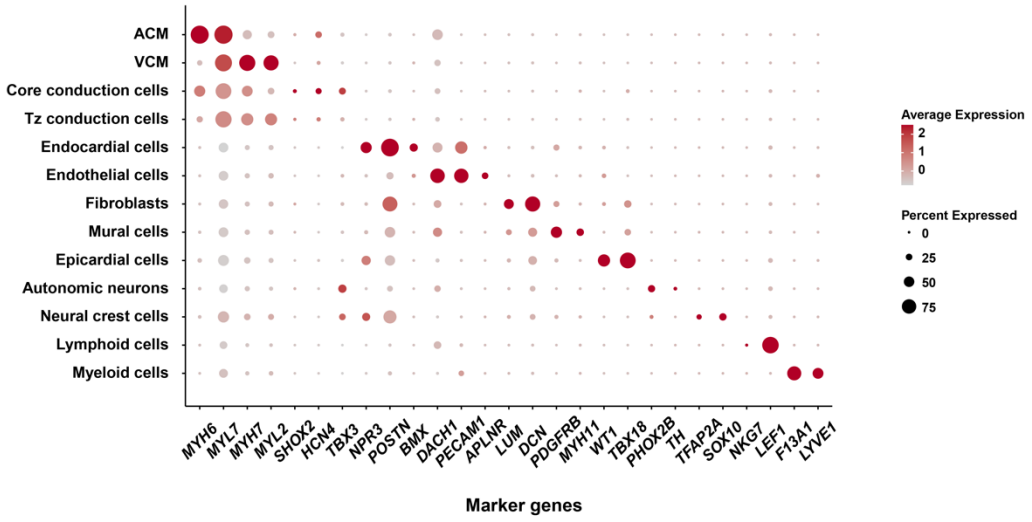


Fig. S2: Annotation of major cell types.
Dot plot of known marker genes for major cell types in the fetal heart atlas. Size of the circle represents the percentage of cells within a given cell type that express the gene of interest, and the color represents the normalized expression level of that gene across all cell types shown.

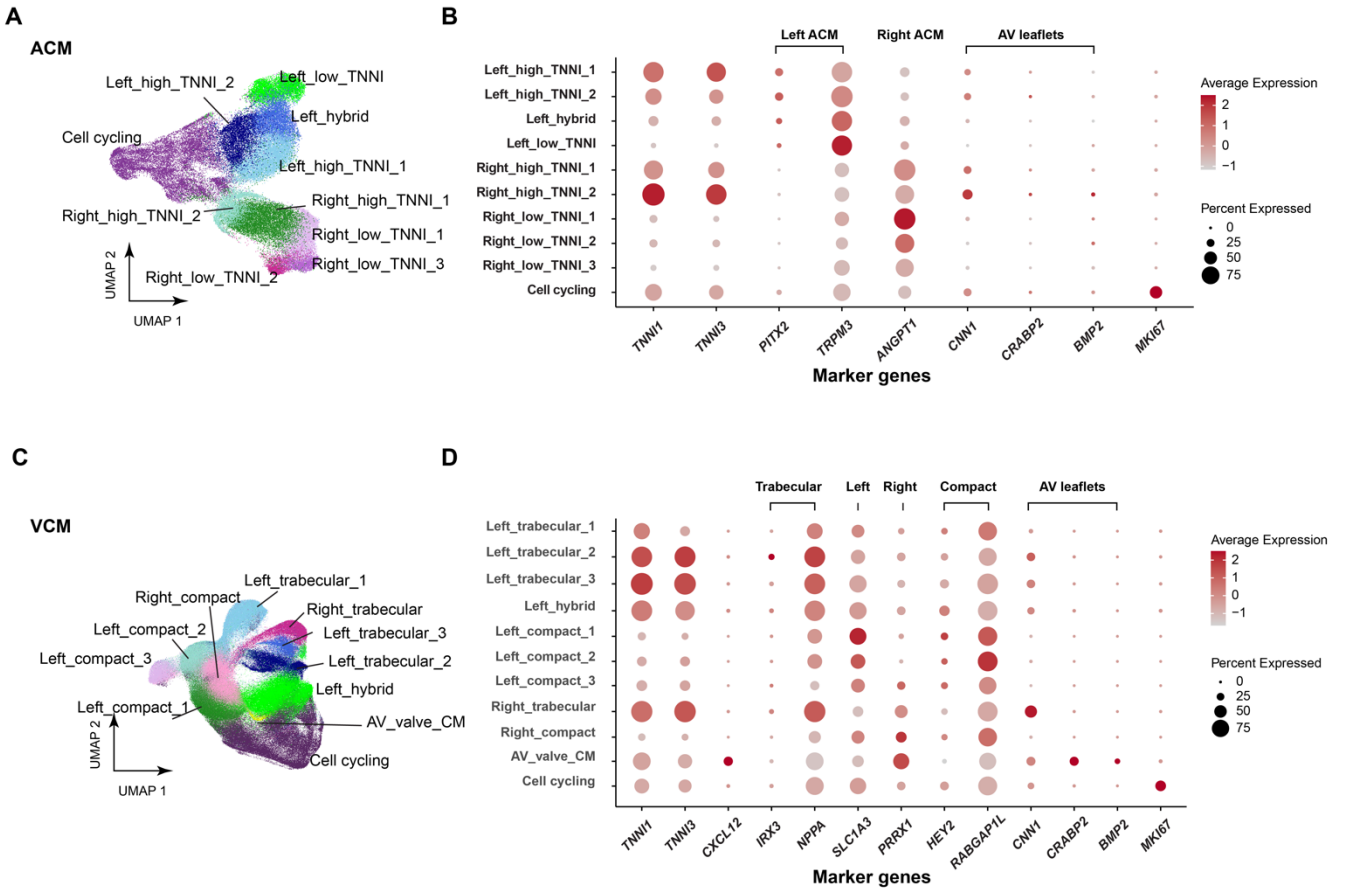


Fig. S3: Annotation of cardiomyocytes.

A. UMAP of atrial cardiomyocytes (ACM).

B. Dot plot of ACM marker genes. The size of the circle represents the percentage of cells within a given cell type that express the gene of interest, and the color represents the normalized expression level of that gene across all cell types shown. Left ACMs were identified using markers *PITX2* and *TRPM3*[47]. Right ACMs were identified using marker *ANGPT1*. As expected, ACMs with high *PITX2* and *TPRM3* expression exhibited lower *ANGPT1* expression, and vice versa. In addition to mapping ACMs to their spatial location (left vs. right atrium), ACMs can be categorized by their expression levels of Troponin (*TNNI1* and *TNNI3*). Certain ACM subtypes, termed "_high_TNNI," express Troponin at a much higher level compared to others, designated "_low_TNNI." *CNN1*, *CRABP2*, and *BMP2*, markers of cardiomyocytes in atrioventricular (AV) valve leaflets, were not significantly expressed in most ACMs, except for slightly elevated expression levels in ACM_right_high_TNNI_1. *MKI67* was used to identify cells undergoing active cell cycling.

C. UMAP of ventricular cardiomyocytes (VCM).

D. Dot plot of VCM marker genes. The size of the circle represents the percentage of cells within a given cell type that express the gene of interest, and the color represents the normalized expression level of that gene across all cell types shown. *SLC13A* and *PRRX1* were used to distinguish left and right ventricular cardiomyocytes, respectively[47]. *IRX3*[47] and *NPPA*[186] were used to label trabecular VCM, while *HEY2* and *RABGAP1L* were used to annotate compact VCM. The "VCM_left_hybrid" cluster expressed both trabecular and compact VCM markers, suggesting a transitional (hybrid) state between the two cell types. Notably, trabecular VCM cell types exhibited much higher expression levels of Troponin compared to compact VCM. Additionally, *CRABP2*, *CNN1*, and *BMP2* were markers for cardiomyocytes in atrioventricular (AV) valve leaflets, and *CXCL12* was a marker for trabecular VCM near the lumen. *MKI67* was used to identify cells undergoing active cell cycling.

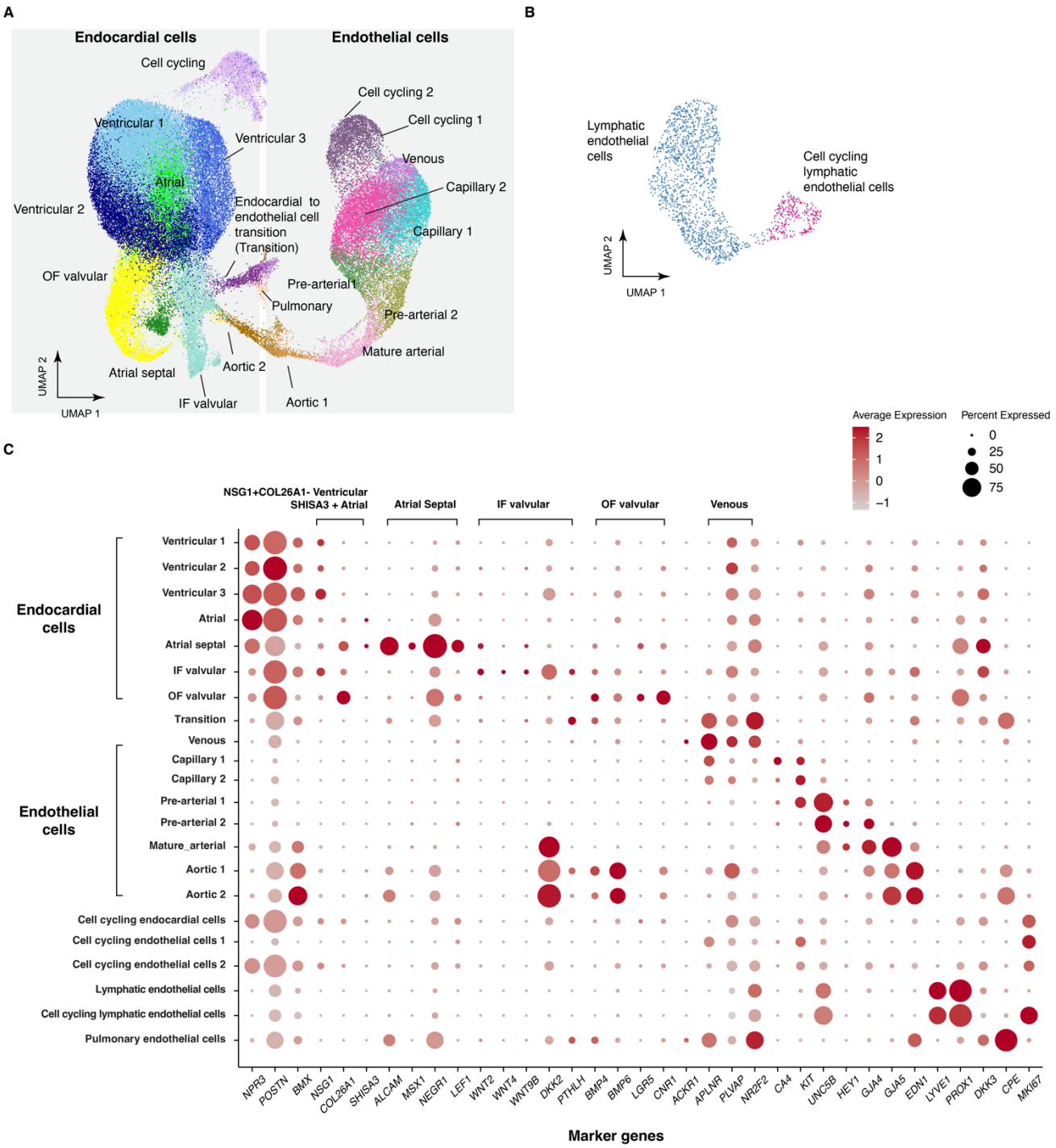


Fig. S4: Annotation of endocardial cells and endothelial cells.

A. UMAP of endocardial cells and endothelial cells (excluding lymphatic endothelial cells). IF: inflow. OF: outflow.

B. UMAP of lymphatic endothelial cells.

C. Dot plot of the marker genes for endocardial cells and endothelial cells. The size of the circle represents the percentage of cells within a given cell type that express the gene of interest, and the color represents the normalized expression level of that gene across all cell types shown. The expression of *NPR3*,

POSTN, and *BMX* were used to distinguish endocardial cells from endothelial cells. *NSG1*, *COL26A1*, and *SHISA3*[47] were used to identify chamber-specific endocardial cells. Endocardial cells highly expressed *NSG1* but have low or no expression of *COL26A1* were labeled as ventricular endocardial cells, while endocardial cells highly expressed *SHISA3* were annotated as atrial endocardial cells. Atrial septal endocardial cells were identified based on the expression levels of *ALCAM*, *MSX1*, *NEGR1*, and *LEF1*[48]. We identified endocardial cells on the inflow (IF) side of the valves based on the expression of *WNT2*, *WNT4*, *WNT9B*, *DKK2*, and *PTHLH*, while those on the outflow (OF) side were identified by *BMP4*, *BMP6*, *LGR5*, and *CNR1* expression. Finally, endocardial cells transitioning to endothelial cells ("transition") were identified by the expression of *APLNR*, *ACKR1*, and *NR2F2* [52,187]. Venous endothelial cells were identified by expression of *NR2F2*, *PLVAP*, and *APLNR*[52]. A cluster of cells likely to be pulmonary endothelial cells were labeled based on their expression of *DKK3* and *CPE*[188]. *CA4* and *KIT* were used to identify capillary endothelial cells. *UNC5B*, *HEY1*, *GJA4* and *GJA5* were used to identify arterial endothelial cells, with *GJA5* specifically marking mature arterial endothelial cells [41,52]. Both *EDN1*[189] and *BMX*[190] were used to identify aortic endothelial cells. *MKI67* was used to identify cells undergoing active cell cycling.

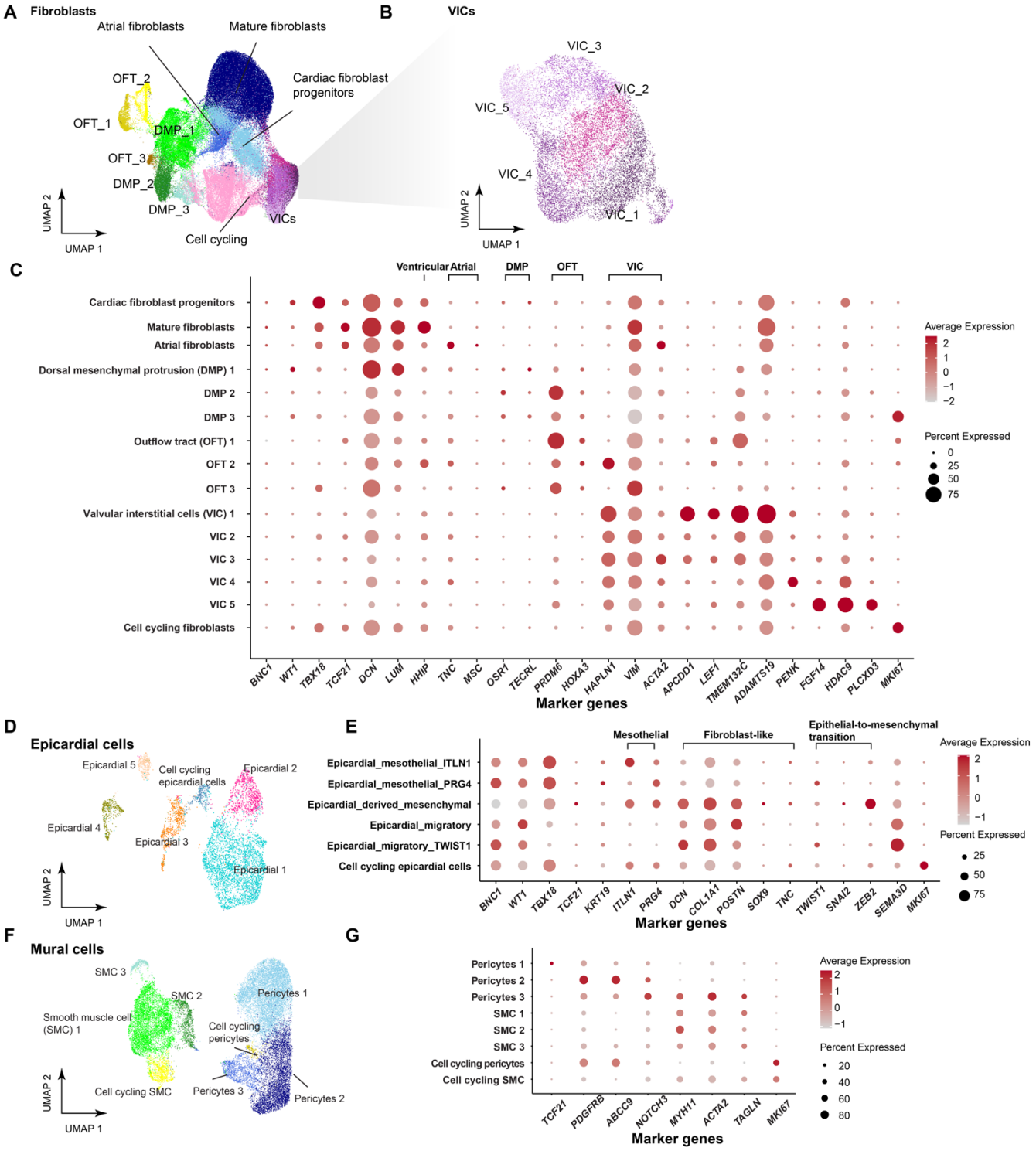


Fig. S5: Annotation of mesenchymal and epicardial cells

A. UMAP of cardiac fibroblasts and valvular interstitial cells (VICs). OFT: Outflow tract. DMP: Dorsal mesenchymal protrusion.

B. UMAP of VICs.

C. Dot plot of the marker genes for fibroblasts and valvular interstitial cells. The size of the circle represents the percentage of cells within a given cell type that express the gene of interest, and the color represents the normalized expression level of that gene across all cell types shown. Cardiac fibroblast progenitors

were identified by elevated expression of *WT1*, *TBX18*, and *TCF21*. *DCN* and *LUM* marked mature fibroblast[4]. To identify chamber-specific fibroblasts, *HHIP* was used for ventricular fibroblast, and *MSC* and *TNC* were used to identify atrial fibroblasts[47] . The cells located to dorsal mesenchymal protrusion (DMP; a second heart field derived mesenchyme at the venous pole of the developing heart), were identified using *OSR1* and *TECRL* [47]. Outflow tract (OFT) cells were labeled with *PRDM6* and *HOXA3*[4]. Both DMP and OFT cells have lower expression of *TBX18* and *TCF21*, consistent with second heart field origin. *HAPLN1*, *VIM,* and *ACTA2*[51,191] were used to distinguish valvular interstitial cells (VICs) from other fibroblast and fibroblast-like cells. VIC_1-3, particularly VIC_1, highly expressed *APCDD1*, *LEF1*, *TMEM132C,* and *ADAMTS19*, recently identified markers for VICs localized to the free segments of the valves[48]. Both VIC_4 are VIC_5 express recently observed markers of valve-related mesenchymal cells located in intervalvular fibrous tissue towards the valve roots. *PENK*, a marker for mesenchymal neural crest derivatives, was used to label VICs derived from neural crest (VIC_4), which were found to be located in and around the developing semilunar and atrioventricular valves. *FGF14*, *HDAC9*, *PLCXD3* are expressed in VIC_5, primarily observed around the semilunar valves. *MKI67* was used to identify cells undergoing active cell cycling.

D. UMAP of epicardial cells.

E. Dot plot of the marker genes for epicardial cells. The size of the circle represents the percentage of cells within a given cell type that express the gene of interest, and the color represents the normalized expression level of that gene across all cell types shown. Established epicardial markers such as *BNC1*, *TBX18*, *WT1*, *TCF21*, and *KRT19* were used to label epicardial cells[4,192]. We identified two distinct mesothelial epicardial populations: Epicardial_mesothelial_ITLN1 and Epicardial_mesothelial_PRG4. Epicardial_mesothelial_ITLN1 expresses *ITLN1*, a marker associated with epicardial adipose tissue; Epicardial_mesothelial_PRG4, a gene encoding lubricin secreted into pericardial fluid. Two migratory epicardial cell populations were identified: Epicardial_migratory and Epicardial_migratory_TWIST1, with the latter exhibiting higher *TWIST1* expression [192]. Both migratory epicardial cells express mesenchymal cell markers *DCN* and *POSTN*, suggesting these cells are in the process of transitioning to mesenchymal cells. Epicardial_derived_mesenchymal cells, identified by markers of epithelial-to-mesenchymal transition (*SNAI2* and *ZEB2*) and cardiac mesenchymal cell differentiation (*SOX9* and *TNC*)[193], likely represent epicardium-derived mesenchymal cells. The expression patterns of *TCF21* and *TBX18* also align with the previous observation[194] showing sequential reduction of *TCF21* and *TBX18* expression in maturing epicardium, and upregulation of *TCF21* in epicardium-derived mesenchymal cells. *MKI67* was used to identify cells undergoing active cell cycling.

F. UMAP of mural cells, including smooth muscle cells (SMCs) and pericytes.

G. Dot plot of the marker genes for mural cells. The size of the circle represents the percentage of cells within a given cell type that express the gene of interest, and the color represents the normalized expression level of that gene across all cell types shown. Expression of *PDGFRB* and *ABCC9* were used to identify pericytes[4]. Pericytes 1 exhibited elevated levels of *TCF21*, suggesting it might represent early pericytes. Pericytes 3 has the highest expression of *NOTCH3*, a gene known to be expressed at the site of arterialization[195] , and higher expression of SMC markers *ACTA2* and *MYH11*, as well as early smooth muscle marker *TAGLN*, suggesting the cells may represent the transition from pericytes to smooth muscle cells. SMC1-2 express SMC markers *ACTA2*, *MYH11*, and *TAGLN*, suggesting they are smooth muscle cells. *MKI67* was used to identify cells undergoing active cell cycling.

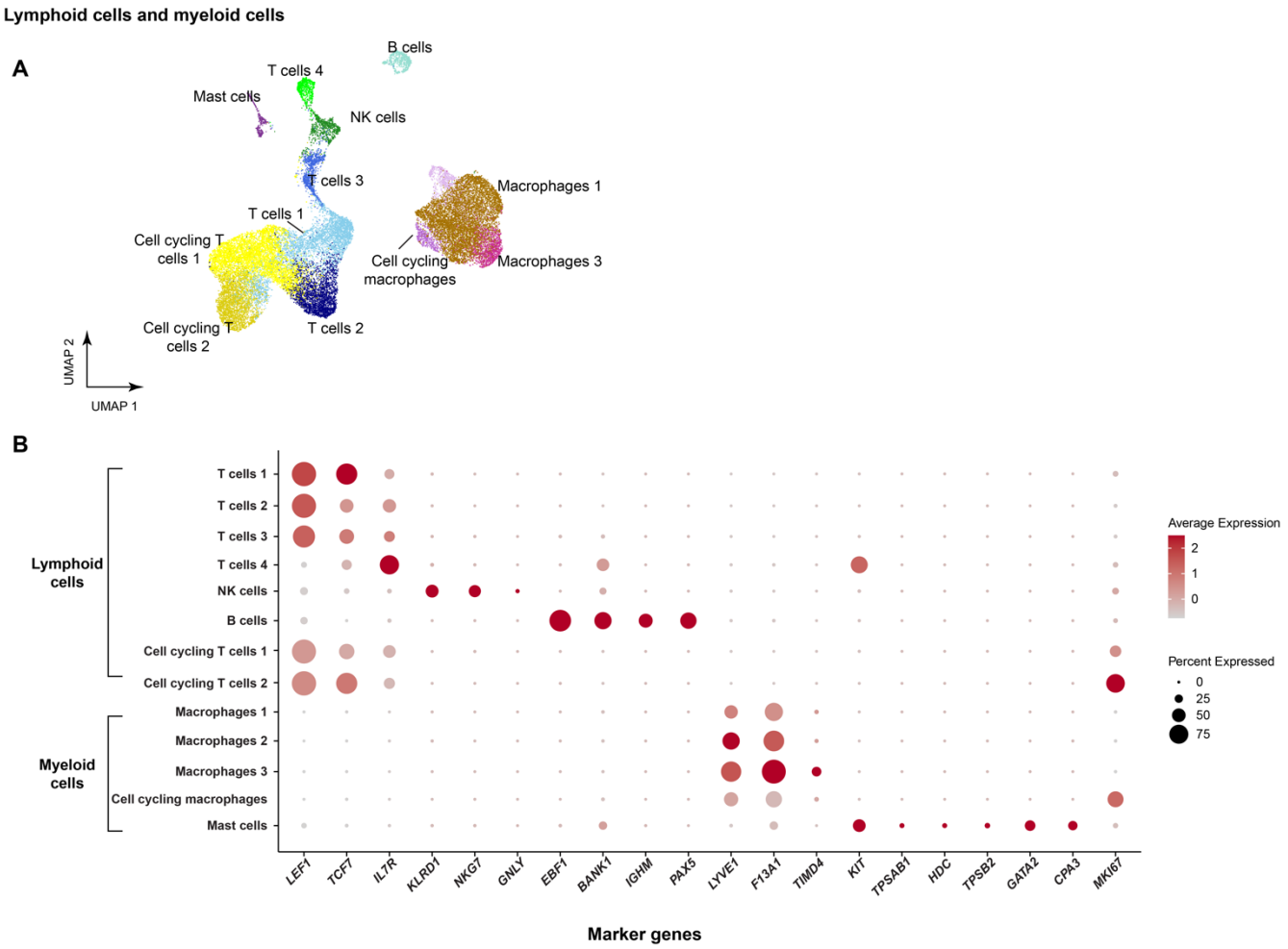


Fig. S6: Annotation of immune cells

A. UMAP of lymphoid and myeloid cells.

B. Dot plot of the marker genes for lymphoid and myeloid cells. The size of the circle represents the percentage of cells within a given cell type that express the gene of interest, and the color represents the normalized expression level of that gene across all cell types shown. Macrophages were identified using *LYVE1* and *F13A1*. Cells in "Macrophages_3" cluster are likely to be cardiac resident macrophages due to their expression of *TIMD4*[196,197]. T cells were identified using *LEF1*, *TCF7*, and *IL7R*[198]. *KLRD1*, *NKG7,* and *GNLY* were used to identify natural killer cells[199]. *EBF1*, *BANK1*, *IGHM,* and *PAX5* were used to identify B cells [200,201]. Finally, *KIT*, *TPSAB1*, *HDC*, *TPSB2*, *GATA2*, and *CPA3* were used as markers of mast cells [202–204]. *MKI67* was used to identify cells undergoing active cell cycling.

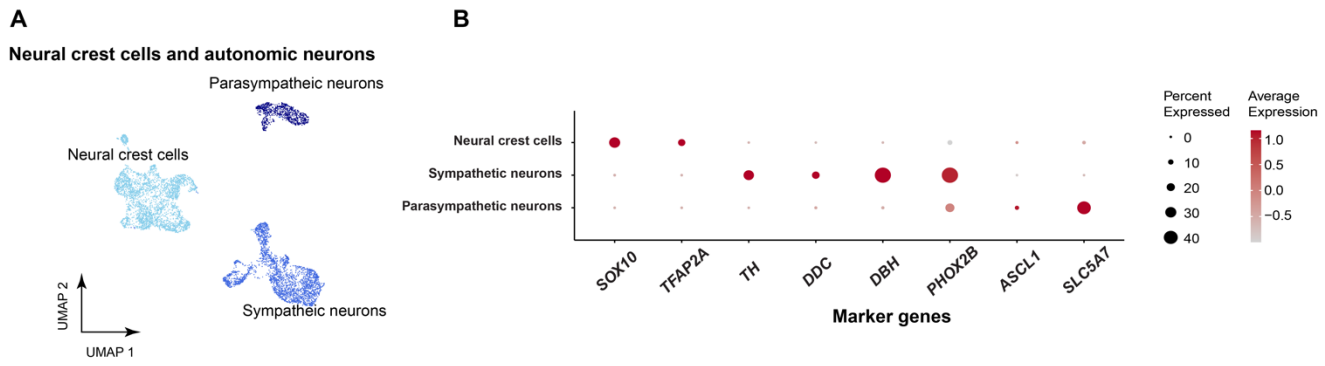


Fig. S7: Annotation of neural crest cells and autonomic neurons.

A. UMAP of neural crest cells and autonomic neurons.

B. Dot plot of the marker genes for neural crest cells and autonomic neurons. The size of the circle represents the percentage of cells within a given cell type that express the gene of interest, and the color represents the normalized expression level of that gene across all cell types shown. Neural crest cells were identified using markers *SOX10* and *TFAP2A*[205]. *TH*, *DDC*, *DHB*, *and PHOX2B* were used to label sympathetic neurons[206,207]. *ASCL1*[208] and *SLC5A7*[206] were used to label parasympathetic neurons.

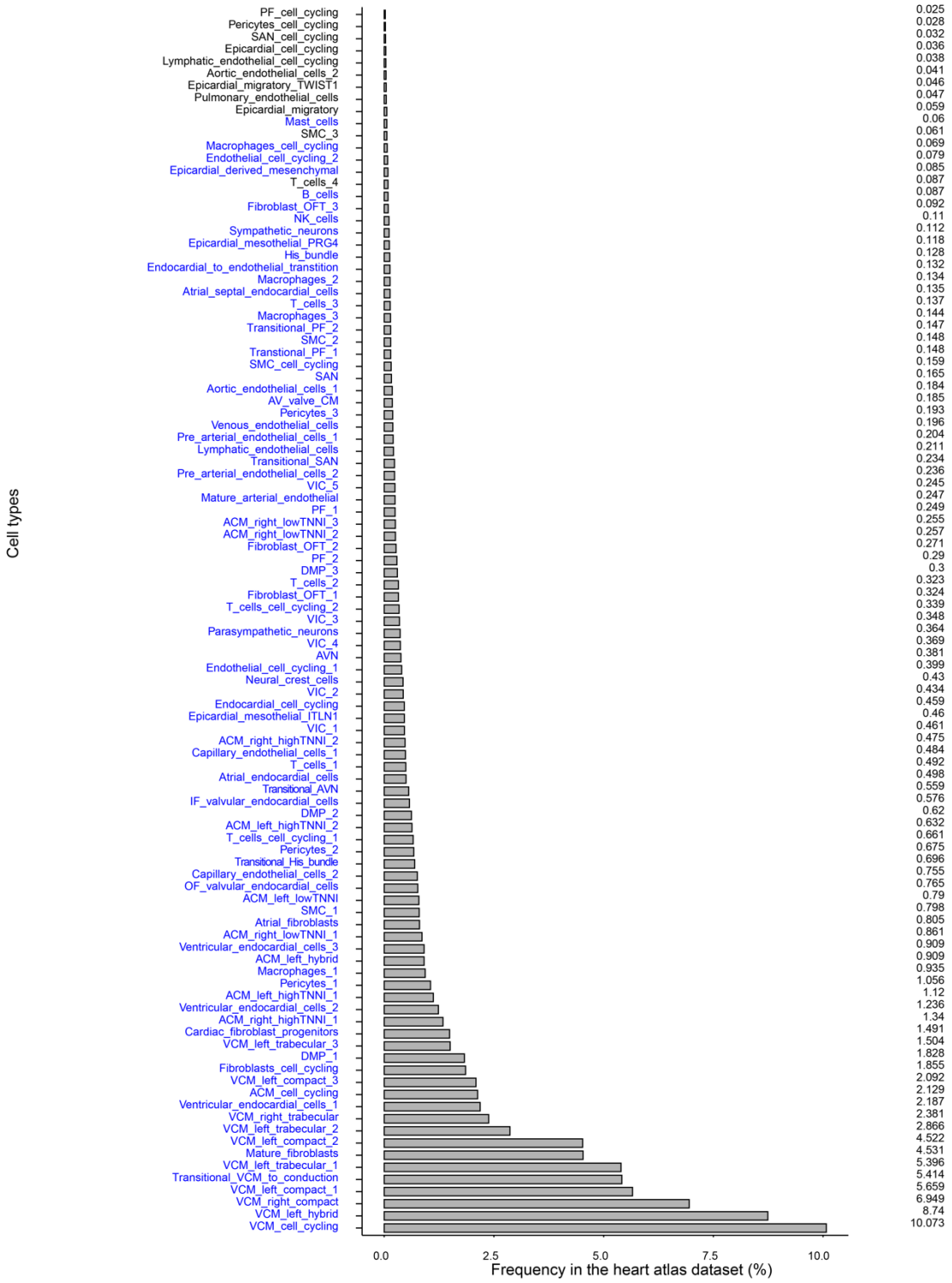
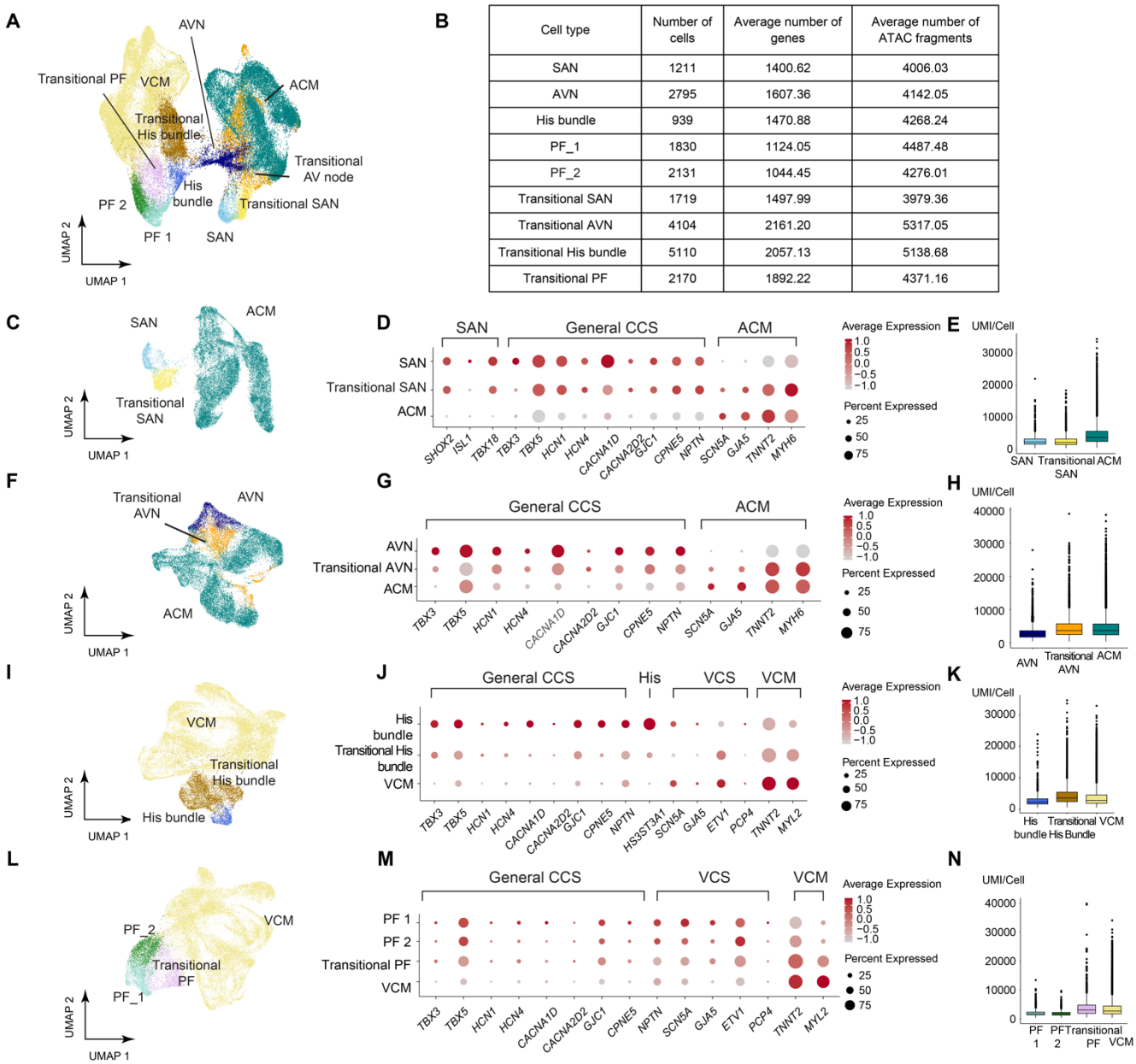


Fig. S8: Cell type frequencies.

Barplot showing the frequency of each cell type in the heart atlas. Highlighted in blue are cell types with sufficient ATAC fragment counts for scE2G and chromBPNet predictions. Due to varying sequencing depth, some less frequent cell types with deeper sequencing were included in these analyses. PF: Purkinje fibers. SAN: Sinoatrial node. SMC: Smooth muscle cells. NK: Natural killer cells. OFT: outflow tract. AV: Atrioventricular. ACM: Atrial cardiomyocytes. DMP: Dorsal mesenchymal protrusion. VIC: Valvular interstitial. AVN: Atrioventricular node. VCM: ventricular cardiomyocytes.



Cell type	Number of cells	Average number of genes	Average number of ATAC fragments
SAN	1211	1400.62	4006.03
AVN	2795	1607.36	4142.05
His bundle	939	1470.88	4268.24
PF_1	1830	1124.05	4487.48
PF_2	2131	1044.45	4276.01
Transitional SAN	1719	1497.99	3979.36
Transitional AVN	4104	2161.20	5317.05
Transitional His bundle	5110	2057.13	5138.68
Transitional PF	2170	1892.22	4371.16

Fig. S9: Transitional conduction cells in the heart atlas.

A. UMAP of the conduction cells, transitional conduction cells, and the surrounding working myocardium. ACM: Atrial cardiomyocyte. AVN: Atrioventricular node. PF: Purkinje fiber. SAN: Sinoatrial node. VCM: Ventricular cardiomyocyte.

B. Cell count, average number of genes per cell, and average number of ATAC fragments per cell for each conduction type and transitional conduction cell type.

C,F,I,L. Each UMAP represents a conduction system cell type, its corresponding transitional cell types, and the surrounding working myocardium.

D,G,J,M. Dot plots of known marker genes for conduction cells, corresponding transitional cell types, and surrounding working myocardium. CCS: Cardiac conduction system. VCS: Ventricular conduction system.

E,H,K,N. Comparison of number of UMI/cell among conduction cells, corresponding transitional cells and surrounding working myocardium. The UMI/cell of transitional cells was not the sum of the average UMI/cell of conduction cells and the average UMI/cell of surrounding working myocardium. Given this finding and that doublets were pre-emptively removed during quality control, it is unlikely that these transitional cells reflect multiplets.

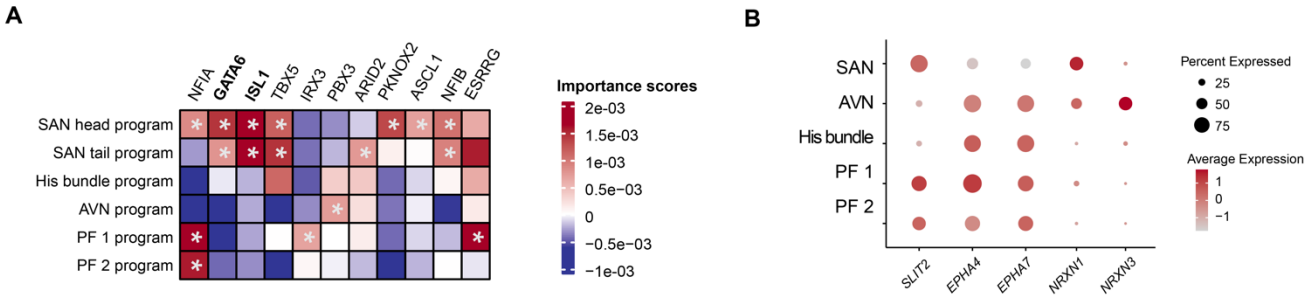


Fig. S10: Cardiac conduction cells: TF regulators of gene programs and expression of neural adhesion and axon development genes.

A. Importance scores of TFs predicted to regulate gene programs associated with the identity of each conduction system cell type. Bold TFs are discussed in the text. *: the TF is a regulator of the corresponding gene program. SAN: Sinoatrial node. AVN: Atrioventricular node. PF: Purkinje fibers.

B. Dot plot showing the expression pattern of genes involved in neural adhesion and axon development (see also **Figs. 2K-M)**. The size of the circle represents the percentage of cells within a given cell type that express the gene of interest, and the color represents the normalized expression level of that gene across all cell types shown.

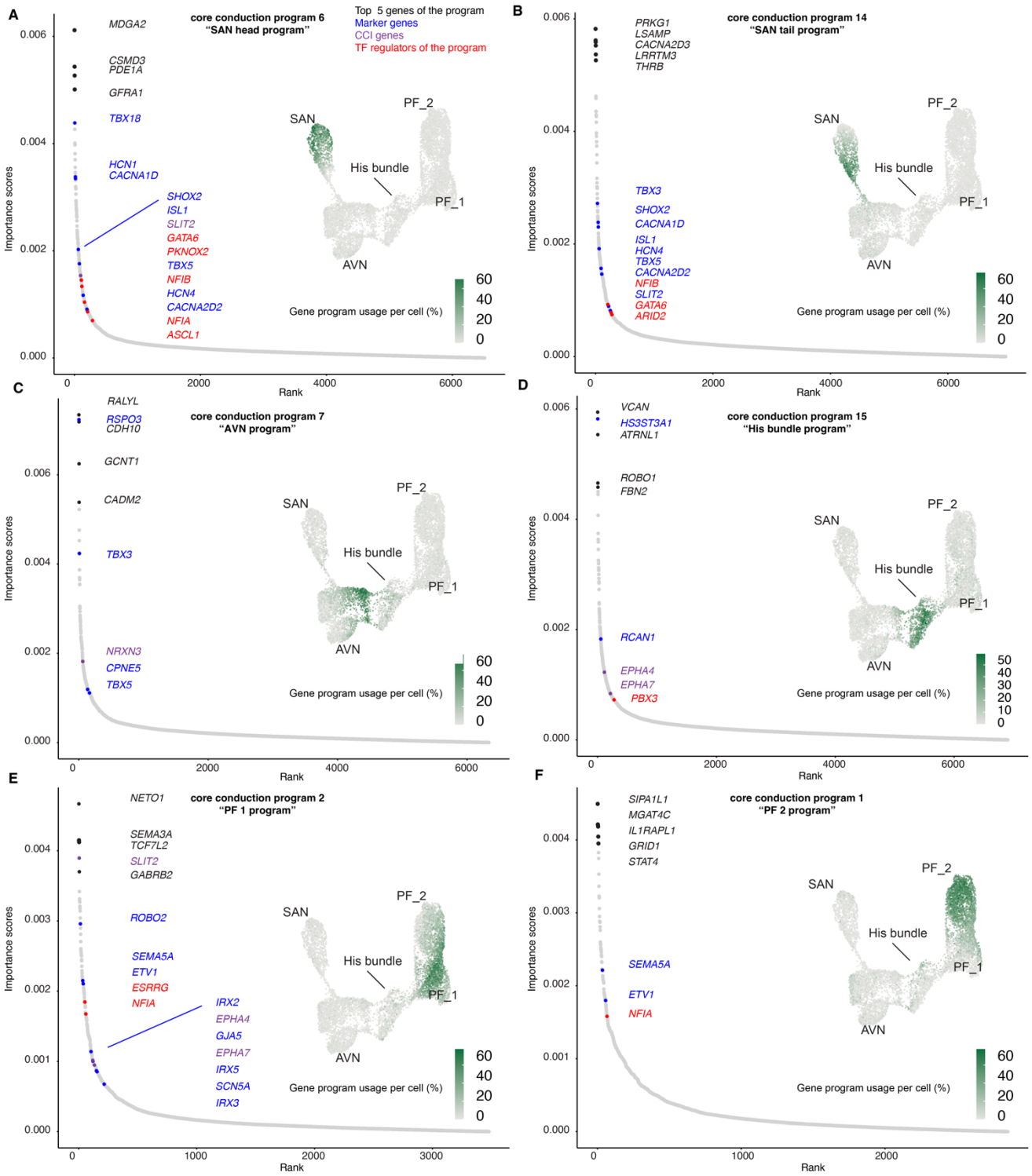


Fig. S11: Gene programs of conduction system cell types.

Annotated gene programs associated with the identity of each conduction system cell type: SAN head program (**A**); SAN tail program (**B**); AVN program (**C**); His bundle program (**D**); PF 1 program (**E**); PF 2 program (**F**). Waterfall plot shows importance scores for all considered genes, and selected genes from the 300 top program genes are colored. Black: 5 genes with the highest importance scores. Blue: Known marker genes for the corresponding cell type. Purple: Genes involved in important cell-cell interaction (CCI) pathways (from **Fig. 2**) are highlighted. Red: TFs predicted to regulate the program. Inset UMAPs: Projection of gene program expression on the cardiac conduction cell UMAP (see also **Fig. 2A**). Color scale: gene program usage per cell. SAN: Sinoatrial node. AVN: Atrioventricular node. PF: Purkinje fibers.

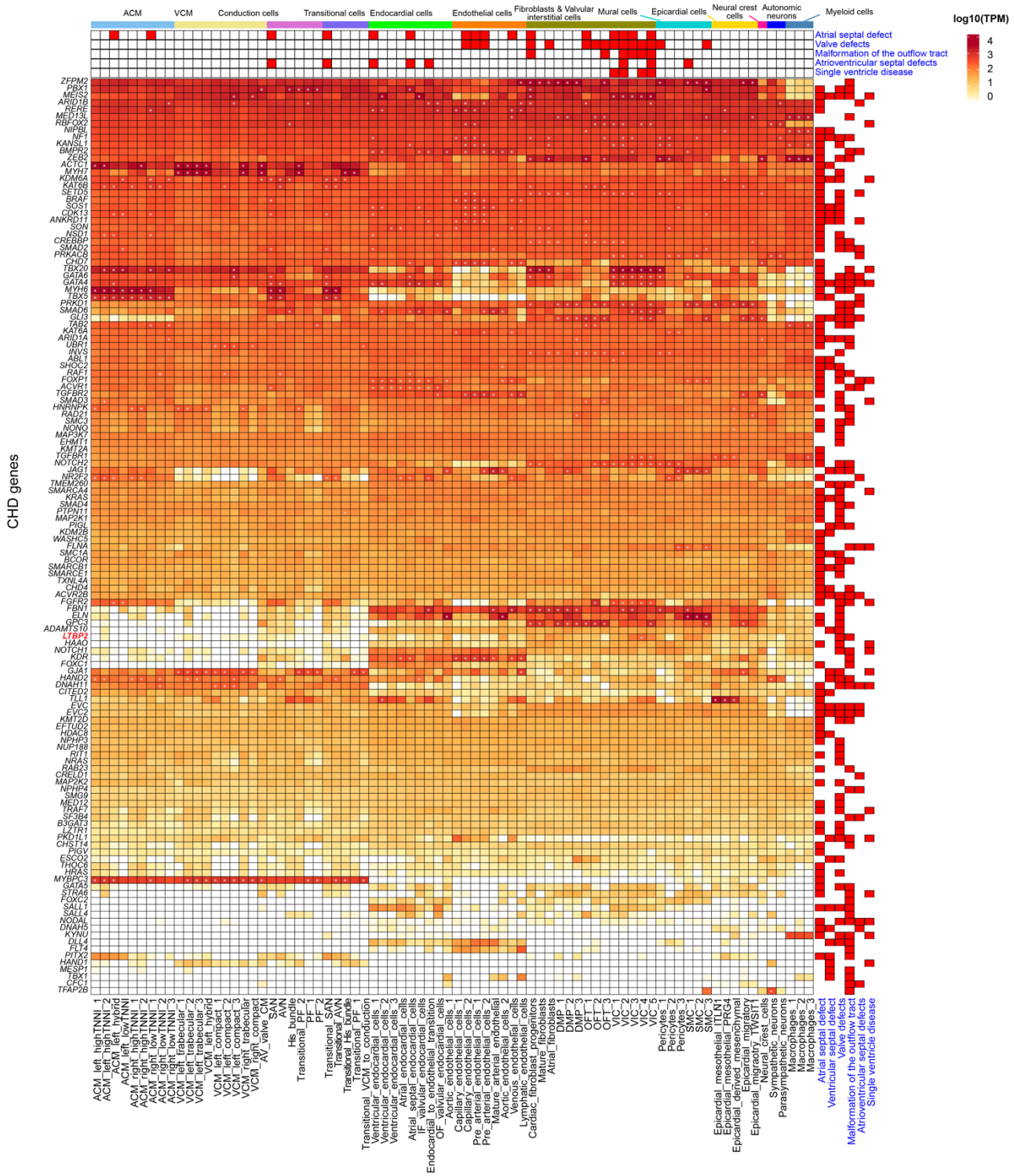


Fig. S12: Expression of high-confidence CHD genes across cardiac cell types.

Absolute expression of high-confidence CHD genes across cardiac cell types. Overall, many CHD genes are expressed in multiple cell types. However, certain cell types are enriched for having many CHD genes with particularly high expression. Main heatmap color scale: $\log_{10}$ TPM. Top annotation: red indicates that the cell type (column) is significantly enriched for high expression of the high-confidence genes associated with the corresponding CHD subtype (row) (see **Fig. 3A**). Side annotation: red indicates that the corresponding gene (row) is a high-confidence causal gene for a CHD subtype (columns). *: The gene is considered highly expressed in the corresponding cell type (see **Methods**). PF: Purkinje fibers. SAN: Sinoatrial node. SMC:

Smooth muscle cells. NK: Natural killer cells. OFT: outflow tract. AV: Atrioventricular. ACM: Atrial cardiomyocytes. DMP: Dorsal mesenchymal protrusion. VIC: Valvular interstitial. AVN: Atrioventricular node. VCM: ventricular cardiomyocytes.

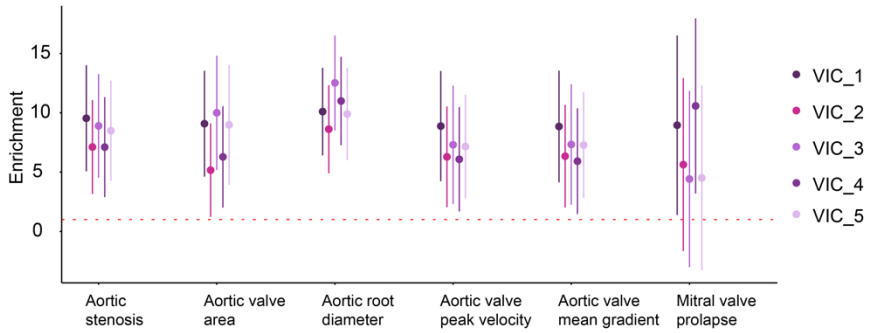


Fig. S13: Enrichment of heritability for 6 valve and aorta-related traits in VIC enhancers.
Y-axis: Genome-wide heritability enrichment of each trait in scE2G-predicted enhancers in each of the 5 subpopulations of VICs, as calculated by S-LDSC. Error bars represent 95% confidence intervals.

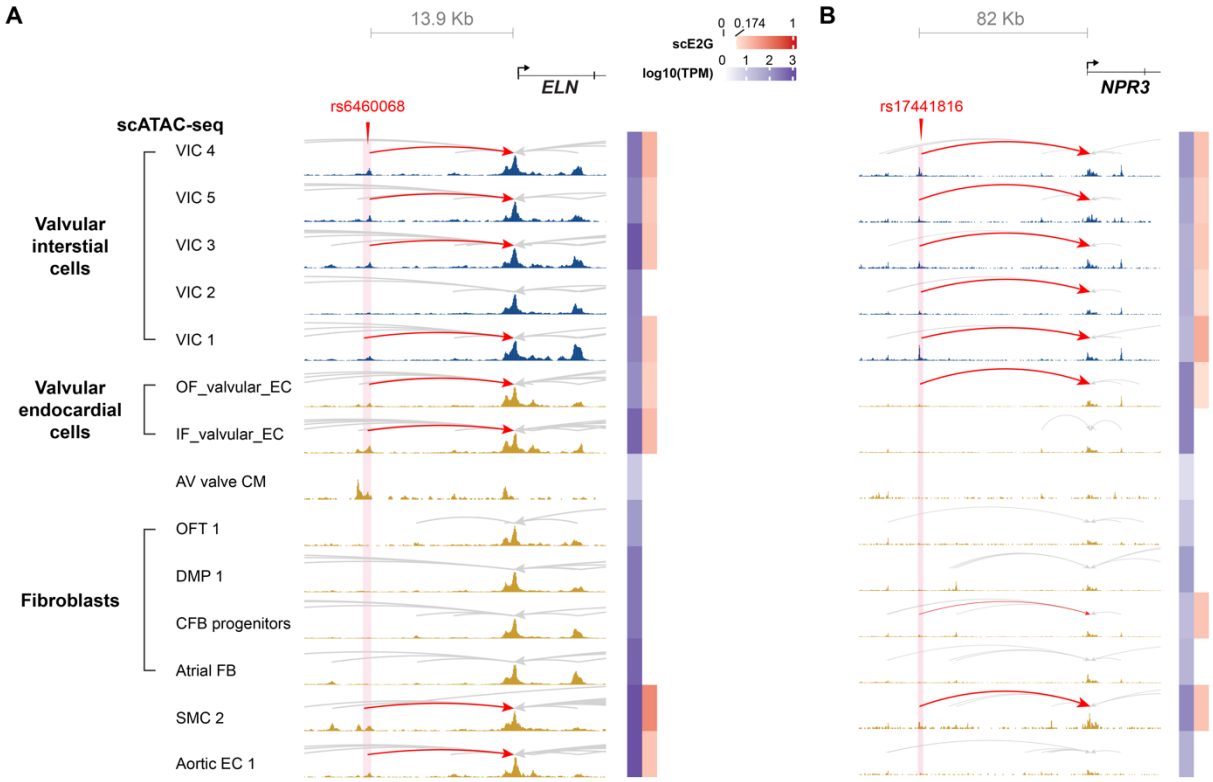


Fig. S14: Examples of common variants overlapping enhancers regulating gene expression in valvular interstitial cells.

A,B. scE2G predictions and normalized ATAC-seq signals are shown for *ELN* (**A**) and *NPR3* (**B**), showing that GWAS variants associated with valve diseases or traits (red downward arrows) overlap enhancers linked to these genes in VICs. *ELN* is also a high-confidence CHD gene, known to be associated with supravalvular aortic stenosis and aortic valve defects. Enhancer predictions overlapping variants are highlighted in red. Arcs represent the predicted enhancer-gene regulatory interactions. Side color bars show the TPM expression of the target genes in each cell type (blue) and the corresponding scE2G scores (red; white for scores below the threshold). CFB: cardiac fibroblasts. The hg38 coordinates of the highlighted enhancer regions are as follows: *ELN* chr7: 74012728-74014407; *NPR3* chr5: 32627864-32630185 (see also **Fig. 5A,B**).

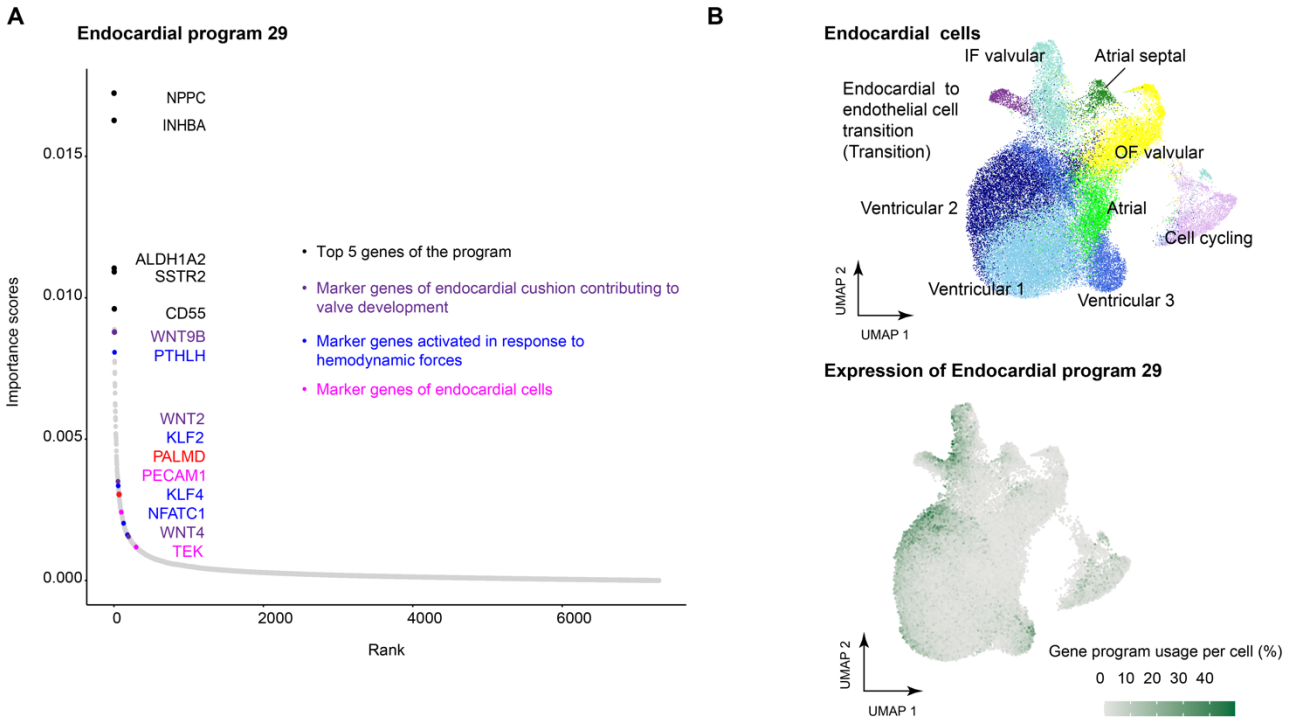


Fig. S15: Annotation of Endocardial Program 29.

A. The top genes associated with Endocardial Program 29 are involved in heart valve development, including markers for endocardial cushions, which contribute to valve formation, as well as genes activated in response to hemodynamic forces.

B. UMAP of endocardial cells and the projection of Endocardial Program 29 expression onto the endocardial cell UMAP. This program is expressed in subpopulations of multiple endocardial cell types. Color scale: gene program usage per cell.

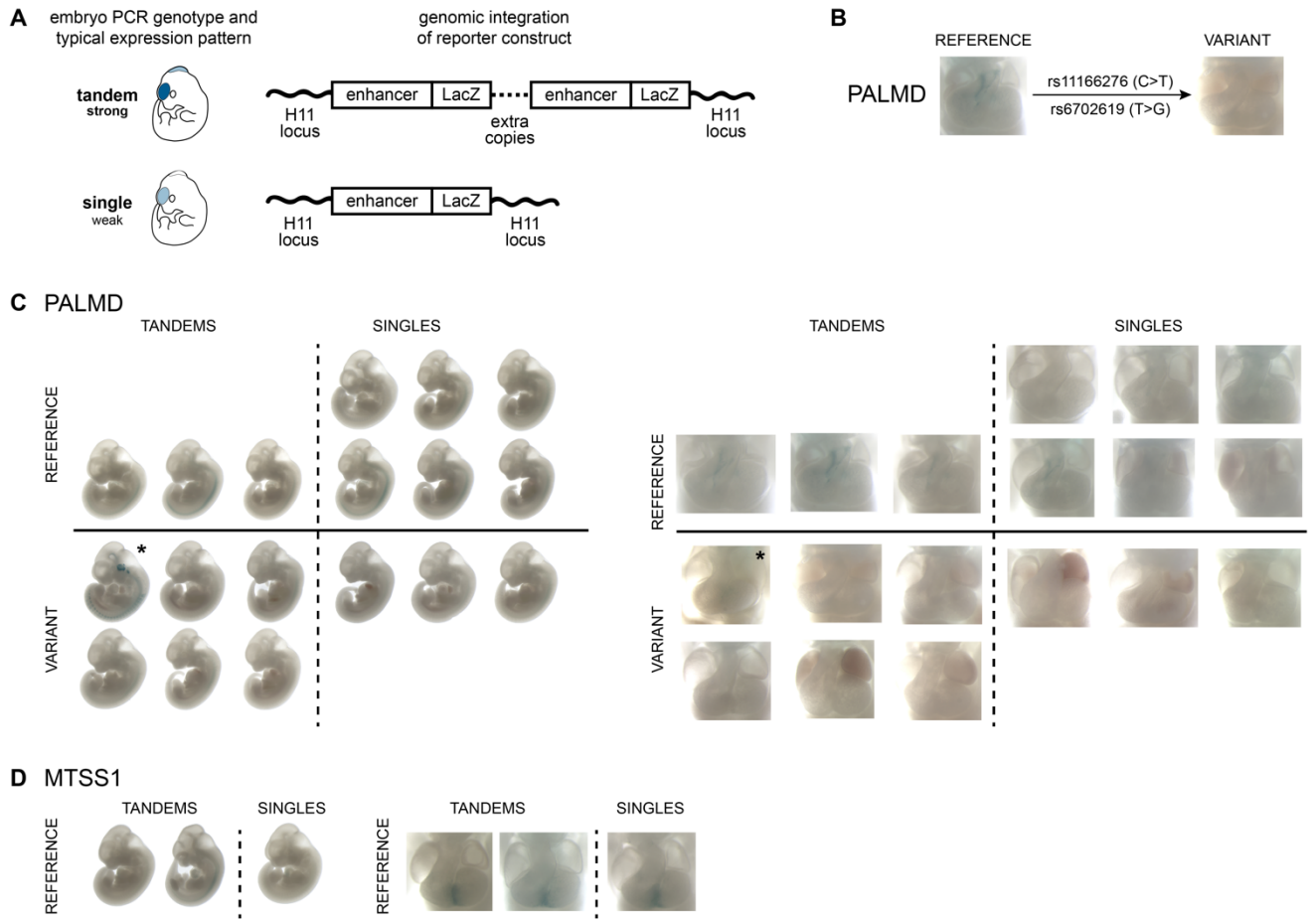


Fig. S16: Human enhancer sequences containing risk variants have location-specific activities in transgenic enhancer reporter assay in mouse embryos.

A. Mouse embryos injected with reporter constructs are genotyped using PCR to confirm successful integration at H11 safe harbor locus and establish approximate copy number of integration (single or tandem). Tandem embryos typically exhibit stronger staining patterns. See Methods for more details.

B. *PALMD* enhancer: Introduction of the risk alleles for two linked variants (rs11166276 and rs6702619) into the human enhancer sequence leads to a reduction in enhancer activity. Two representative embryos are shown. See also **Fig. 5**.

C. *PALMD* enhancer: Whole-embryo side images and frontal heart close-ups of all collected embryos with integration at H11 safe harbor locus. Variant embryos represent those with introduction of two linked alleles, as in B. *: This embryo exhibits unusual ectopic staining pattern likely due to a random genomic integration in addition to H11 integration. See also **Fig. 5**.

D. Similar to C, for *MTSS1* enhancer. See also **Fig. 5**.

Supplementary Tables

Table S1. List of human fetal heart tissue samples
This table contains information related to the sample attributes for a collection of heart tissues. Each row corresponds to a unique tissue sample, with details about the subject, sex, and age. Columns:

- **Subject ID (dbGaP):** This column contains a unique identifier for each fetal tissue sample, cross-referencing both the dbGaP dataset and the Sample Experiment Information table.
- **Sex:** Biological sex, as inferred from expression of genes on sex chromosomes (see Methods).
- **Age in Days:** Age of fetal sample in post-conception days.
- **Age in Post Conception Weeks:** Age of the heart tissue in post-conception weeks (rounded).

Table S2. 10x Multiome sample metadata
Each row corresponds to one 10x Multiome lane. Columns:

- **Subject ID (dbGaP)**: Unique Subject IDs from dbGaP that can be used to cross reference information in this table.
- **Sample ID (dbGaP) / Lane ID**: Represents a unique identifier for each 10X lane, including information about which heart samples were used to make each library. This unique identifier is the Sample ID from our dbGaP metadata.
- **Experiment Number**: Represents the unique identifier for each experiment conducted on a specific day. Each number corresponds to a distinct experimental setup.
- **10x Lane number**: Refers to the specific lane on the loading chip used during the GEM (Gel Bead-in-Emulsion) creation process in the Chromium Controller by 10X Genomics.
- **Tissue State Input**: “Flash Frozen” or “Fresh”, according to whether or not the sample was frozen between tissue collection and nuclei isolation.
- **Were samples mixed?**: This column specifies whether the samples were mixed in the 10X lane (Yes or No).
- **Mixing Ratio**: This column indicates the ratio in which samples were combined in a 10X lane. For experiments with a single sample, the ratio is represented as ‘1’. When combining two, three, or four samples, the ratios are ‘1:1’, ‘1:1:1’, and ‘1:1:1:1’ respectively, reflecting equal ratio pooling.
- **Physical Sample:** Denotes individual human fetal heart samples.
- **Physical Sample ID:** Represents the specific tissue piece used in each experiment. The format "PhysicalSample_FT###_X" is used, where "FT###" refers to the Sample ID, and "_X" denotes the unique tissue piece from that subject. In some cases, the same heart sample was profiled in multiple experiments, and, in these cases, different physical samples of tissue from the same heart were used in each experiment.
- **Region Where Tissue Was Taken**: This column specifies whether the entire heart was used or the specific location from which tissue samples were collected from human fetal hearts.
- **Protocol Version Used**: Experimental protocol version (versions 1-3) applied to each sample during the experimental portion (see Methods). The majority of our dataset was processed with version 3.
- **GEX index columns**: Barcode information for the gene expression libraries (GEX), derived from 10x Multiome Dual Index Plate TT Set A (PN 3000431).
- **ATAC i7_Index columns**: Barcode information for the ATAC-seq libraries, derived from 10x Multiome Sample Index Plate N, Set A (PN 3000427).

Table S3: Enhancer-gene regulatory interactions identified for each cell type in the heart map
In this table, each row represents the effect of one element on one nearby gene in one cell type, as predicted by scE2G. This table is available at https://www.synapse.org/Synapse:syn64012884.

- chr: chromosome for element-gene pair
- start: start coordinate for element (hg38)
- end: end coordinate for element (hg38)
- name: unique identifier for element, in the form of "class|chrN:start-end"
- class: category describing element location, one of: "promoter," "intergenic," "genic"
- TargetGene: HGNC symbol for target gene of element-gene pair

- TargetGeneTSS: transcription start site of target gene
- TargetGeneIsExpressed: boolean indicating whether the target gene is expressed based on promoter accessibility, should be "True" for all predictions
- TargetGeneEnsembl_ID: Ensembl ID corresponding to target gene
- isSelfPromoter: boolean indicating whether this element-gene pair is a promoter linked to it's own gene
- CellType: description of cell cluster for these predictions
- distance: distance between the midpoint of element and transcription start site of target gene
- normalizedATAC_prom: (model feature) normalized pseudobulk scATAC-seq signal at the promoter of the target gene
- ABC.Score: Activity-by-Contact (ABC) score corresponding to this element-gene pair, calculated using pseudobulk scATAC-seq for activity and a distance-based power law estimation for contact
- numCandidateEnhGene: (model feature) number of elements between this element and target gene transcription start site
- numTSSEnhGene: (model feature) number of other transcription start sites between this element and the target gene transcription start site
- numNearbyEnhancers: (model feature) number of elements within 5 kb of this element
- ubiqExpressed: (model feature) boolean indicating whether this gene is classified as ubiquitously expressed across cell types
- RNA_meanLogNorm: mean log-normalized read count for this gene in this cell cluster
- RNA_pseudobulkTPM: pseudobulk transcripts per million for this target gene for this cell cluster
- RNA_percentCellsDetected: percent of cells in this cell cluster in which this gene was detected
- Kendall: Kendall correlation between accessibility at this element and expression of this gene across cells in this cluster
- ARC.E2G.Score: (model feature) Activity, Responsiveness and Contact (ARC)-E2G score that integrates ABC score and Kendall correlation
- E2G.Score: raw scE2G score directly output from logistic regression model
- E2G.Score.qnorm: Final score column used for thresholding, after quantile normalization to full-depth reference scores and setting scores for genes below TPM threshold to 0
- E2G.Score.qnorm.ignoreTPM: same as E2G.Score.qnorm, but without scores for genes with TPM below threshold set to 0

Table S4: Gene programs: cell type grouping for cNMF

- CellType: cell type name.
- NMFGroup: Cell type group that the cell type was assigned to for cNMF analysis. The cell types grouped together are phenotypically similar.

Table S5: Gene programs: 300 top genes expressed in each program.

Each column corresponds to one gene program. Rows correspond to the 300 top genes ranked by importance scores in the program.

Table S6: Gene programs: TFs predicted to regulate each program by ChromVAR.

- Program: Name of the gene program
- TF: Name of the transcription factor (TF)
- CellNumber: Number of cells expressing the gene program
- P: P-value of the Spearman correlation between per-cell motif accessibility of the TF and the expression level of the gene program
- Estimate: Spearman's rank correlation coefficient
- P.adjust_bonferroni: Adjusted p-value after Bonferroni correction for multiple comparisons.
- TF_rank_in_Program: Rank of the TF in the gene program, with lower ranks indicating the expression of the TF is more highly associated with the expression of the gene program.

Table S7: Significant ligand-receptor pairs between sender and receiver cell types in cell-cell interaction analysis

This data file is available at https://www.synapse.org/Synapse:syn64108840

- **source:** Source cell type, where molecular signals are sent from.
- **target:** Target cell type, where molecular signals are sent to.

- **ligand:** Ligand gene name, note that multiple subunits of the protein are separated with an underscore. The ligand list comes from CellChat v2.1.2 database.
- **receptor:** Receptor gene name, note that multiple subunits of the protein are separated with an underscore. The receptor list comes from CellChat v2.1.2 database.
- **prob:** Cell-cell interaction strength. Calculated with R package CellChat v2.1.2.
- **pval:** Permutation-based p-value for the interaction strength. Calculated with R package CellChat v2.1.2.
- **interaction_name**: Cell-cell interaction name.
- **interaction_name_2:** Cell-cell interaction name, format 2.
- **pathway_name:** Name of the pathway that the ligand-receptor pairs belong to. The membership comes from the CellChat v2.1.2 database.
- **annotation:** Type of the ligand-receptor pair, e.g. secreted signaling, Cell-Cell Contact, etc.
- **evidence:** Supporting information for the ligand-receptor pair, e.g. PubMed ID, KEGG ID, etc. The information comes from the CellChat v2.1.2 database.
- **PCW:** Post-conception weeks. The cell-cell interaction analysis was performed one week at a time.

Table S8: Curated list of CHD genes, PCGC genes with *de novo* variants, and other cardiovascular disease genes.

- Gene: Gene name.
- Trait: Name of the disease.
- Category: The broad category of the disease genes. CHD: high-confidence genes from CHDGene; PCGC_DeNovoVariants: genes with de novo variants from PCGC; CD: other cardiovascular disease genes (see Methods).

Table S9: Enrichment of high-confidence CHD genes and other cardiovascular disease genes in cardiac cell types

- Category: The broad category of the disease genes. CHD: High-confidence CHD genes from CHDgene. PCGC_DeNovoVariants: genes with 2 or more *de novo* variants from PCGC. CD: other cardiovascular disease genes.
- Disease: Name of the disease.
- n_genes_highly_expressed_in_Celltype_and_are_Disease_genes: Number of disease genes that are highly expressed in the cell type (see Methods)
- n_genes_highly_expressed_in_Celltype_and_are_NOT_Disease_genes: Number of non-disease genes that are highly expressed in the cell type
- n_genes_NOT_highly_expressed_in_Celltype_and_are_Disease_genes: Number of disease genes that are not highly expressed in the cell type.
- n_genes_NOT_highly_expressed_in_Celltype_and_are_NOT_Disease_genes: Number of non-disease genes that are not highly expressed in the cell type.
- odds_ratio: the enrichment of disease genes in the highly expressed genes of the cell type.
- CI_lower: The low bound of the 95% confidence interval of the odds ratio.
- CI_upper: The higher bound of the 95% confidence interval of the odds ratio.
- P_value: The *p*-value of the odds ratio.
- NumberOfTrueDiseaseGenes: Total number of known causal genes for the disease.
- recall: Percentage of disease genes that are highly expressed in the cell type.
- disease_genes_in_Celltype: List of disease genes highly expressed in the cell type.
- P_fdr: Adjusted p-value after correcting for multiple testing with Benjamini–Hochberg procedure
- Celltype: Name of the cell type.

Table S10. Evidence for 27 high-confidence CHD genes that are highly expressed in VICs and/or cardiac fibroblasts

Summary of evidence presented in **Note S1** regarding 27 high-confidence CHD genes that are highly expressed in VICs and/or cardiac fibroblasts. Each row represents a unique gene.

- **Gene**: Gene symbol.
- **CHD Subtype(s)**: Subtype(s) of CHD associated with each gene (see Methods).

- **Highly expressed in VIC (Yes/No)**: Indicates whether each gene is highly expressed in any of the 5 populations of VICs
- **Highly expressed in cardiac fibroblasts (Yes/No)**: Whether the gene is highly expressed in Cardiac_fibroblast_progenitors.
- **Knockout evidence in epicardial lineage (Yes/No)**: "Yes" indicates that gene knockout in the epicardial lineage led to developmental changes or defects. "No" indicates either a lack of impact from knockout experiments or that no knockout studies were conducted for this lineage to our knowledge.
- **Knockout evidence in endocardial lineage (Yes/No)**: "Yes" indicates that gene knockout in the endocardial lineage led to developmental changes or defects. "No" indicates either a lack of impact from knockout experiments or that no knockout studies were conducted for this lineage to our knowledge.
- **Knockout evidence in neural crest lineage (Yes/No)**: "Yes" indicates that gene knockout in the neural crest lineage led to developmental changes or defects. "No" indicates either a lack of impact from knockout experiments or that no knockout studies were conducted for this lineage to our knowledge.

Table S11: List of GWAS traits and data sources

- Trait: Name of the trait in **Table S13**.
- DescriptiveName: Description of the trait.
- Citation: PMID or zenodo ID of the GWAS.
- VariantProcessingMethod: methods used for preparing variants for overlapping with scE2G-predicted enhancers (see **Methods**)

Table S12: Heritability enrichment (S-LDSC) for GWAS traits in enhancers in each cell type

- Celltype: Name of the cell type.
- major_celltype: Broad category classification of the cell type
- Enrichment: The enrichment of the heritability of the trait in the enhancers of this cell type.
- Enrichment_std_error: Standard error of the enrichment estimate
- Enrichment_p: P-value for the significance of the enrichment
- Enrichment_p_bonferroni: P-value adjusted for multiple comparisons with Bonferroni correction across all annotations, including those in the baseline model.
- Coefficient_z.score: Z-score of the regression coefficient which measures the additional contribution of this cell type's enhancers to the model
- Trait: Name of the trait.
- n_LeadSNPs: Number of lead SNPs associated with the trait.

Table S13: Prioritized variants, enhancers, genes, and gene programs for GWAS signals for heart diseases and traits.

Each row provides details about an scE2G predicted cell type-specific enhancer-gene link overlapped by a variant (listed in the "rsid" column) associated with a GWAS signal (listed in the "LeadVariant" column). Columns:

- enh_chr: Chromosome ID of the enhancer.
- enh_start: Start coordinate of the enhancer (hg38)
- enh_end: End coordinate of the enhancer (hg38)
- Trait: Name of the trait.
- DescriptiveName: Description of the trait.
- nLeadSNPs: Number of lead SNPs (credible sets) for the trait.
- CredibleSet: ID of the credible set.
- rsid: ID of the variant.
- distance: Distance between the enhancer and the promoter of the target gene.
- E2G.Score.qnorm: Quantile-normalized enhancer-to-gene (E2G) interaction score
- TargetGene: Target gene regulated by the enhancer.
- TargetGeneRank: Rank of E2G.Score.qnorm of the target gene among all target genes regulated by the same enhancer in this cell type (lower rank indicates higher E2G.Score.qnorm)
  AnyVariantInCSOverlapCoding: Whether any variant in the credible set (CS) overlaps a coding region.

- AnyVariantInCSOverlapSpliceSite: Whether any variant in the credible set (CS) overlaps a splice sites.
- ProgramsContainingTargetGene: Gene programs that the target gene is involved in
- TargetGeneRankInProgram: Rank of the target gene within the gene program (lower rank indicates stronger association)
- ProportionOfProgramExpressingInCelltype: Proportion of cells expressing the gene program that belong to this cell type.
- TFsOfProgram: Transcription factors (TFs) predicted to regulate the gene program
- TargetGeneCausesAdultCardiovascularDisease: The adult cardiovascular diseases known to be caused by the target gene.
- TargetGeneCausesCHD_CHDgene: The CHD subtypes known to be caused by the target gene.
- TargetGeneCausesCHD_PCGC: Indicates whether the target gene is a CHD gene with de novo mutations identified by PCGC.
- LeadVariant: Variant ID of the lead variant of the credible set.
- LeadVariant_P_value: P-value of the lead variant from the GWAS
- variant_P_value: P-value of the variant from the GWAS (if available)
- LeadSNP_PosteriorProb: Posterior probability of the lead SNP (if available)
- PosteriorProb: Posterior probability of the variant (if available)
- TPM: Transcripts Per Million (TPM) of the target gene
- Celltype: Cell type in which the enhancer-gene interaction occurs

Table S14: Genomic sequences tested in mouse transgenic enhancer reporter assay

- VISTA ID: ID of the enhancer in the VISTA database
- target gene: Target gene predicted to be regulated by the enhancer
- coordinates (hg38): Genomic coordinates of the human enhancer sequence
- reference sequence: Reference DNA sequence cloned for testing the activity of the enhancer.
- variant sequence: DNA sequence with the variant(s) introduced for testing their impacts on the activity of the enhancer.
- variant rsID: IDs of the variant(s)

Table S15: TF-MoDISco motifs

Each row represents one motif identified in one cell type by ChromBPNet and TF-MoDISco.

- pattern: Pattern name
- num_seqlets: Number of seqlets supporting the pattern
- match0: Best match for the pattern in the reference motif database
- qva0: q-value of the best match
- match1: Second best match in the reference motif database
- qval1: q-value for match1
- match2: Third best match
- qval2: q-value for match2
- celltype: Cell type that the patterns were identified for

Table S16: TF-MoDISco motif instances identified with FiNeMo

Each row represents one motif instance (*e.g.*, candidate TF binding site) in one cell type identified by ChromBPNet, TF-MoDISco, and FiNeMo. This data file is available at https://www.synapse.org/Synapse:syn64109173

- chr, start, end: Chromosome, start and end of the match (trimmed).
- start_untrimmed,end_untrimmed: Untrimmed start and end of the match (30 bp).
- motif_name: The name of the pattern in the TF-MoDISco table for the corresponding cell type.
- hit_coefficient: The regression coefficient for the hit, normalized per peak region.
- hit_coefficient_global: The regression coefficient for the hit, scaled by the overall importance of the region.

- hit_correlation: The correlation between the untrimmed CWM and the contribution score of the motif hit.
- hit_importance: The total absolute contribution score within the motif hit.
- strand: The orientation of the hit (+ or -).
- peak_id: The numerical index of the peak region containing the hit in the input file for the respective cell type.
- celltype: Cell type that the scanning was performed in.

Table S17: ChromBPNet performance metrics

Each row corresponds to one cell type.

- celltype: Cell type that the models were trained for
- counts_pearsonr_mean: Pearson correlation coefficient of observed and predicted log(total counts) mean across five folds.
- counts_spearmanr_mean: Spearman correlation coefficient of observed and predicted log(total counts), mean across five folds.
- median_jsd_mean: Median Jensen Shannon Distance (JSD) between the normalized base-resolution coverage profile and the predicted multinomial profile distribution, mean across five folds.
- median_norm_jsd_mean: Median of the min-max normalized JSD, mean across five folds.
- counts_pearsonr_std: Standard deviation across five folds for counts_personr.
- counts_spearmanr_std: Standard deviation across five folds for counts_spearmanr.
- median_jsd_std: Standard deviation across five folds for median_jsd.
- median_norm_jsd_std: Standard deviation across five folds for median_norm_jsd.

References

1. Bruneau, B. G. The developing heart: from The Wizard of Oz to congenital heart disease. *Development* **147**, dev194233 (2020).
2. Morton, S. U., Quiat, D., Seidman, J. G. & Seidman, C. E. Genomic frontiers in congenital heart disease. *Nat. Rev. Cardiol.* **19**, 26–42 (2022).
3. Buijtendijk, M. F. J., Barnett, P. & van den Hoff, M. J. B. Development of the human heart. *Am. J. Med. Genet. C Semin. Med. Genet.* **184**, 7–22 (2020).
4. Ameen, M. *et al.* Integrative single-cell analysis of cardiogenesis identifies developmental trajectories and non-coding mutations in congenital heart disease. *Cell* **185**, 4937–4953.e23 (2022).
5. Nees, S. N. & Chung, W. K. The genetics of isolated congenital heart disease. *Am. J. Med. Genet. C Semin. Med. Genet.* **184**, 97–106 (2020).
6. Hoffman, J. I. E. & Kaplan, S. The incidence of congenital heart disease. *J. Am. Coll. Cardiol.* **39**, 1890–1900 (2002).
7. Jin, S. C. *et al.* Contribution of rare inherited and de novo variants in 2,871 congenital heart disease probands. *Nat. Genet.* **49**, 1593–1601 (2017).
8. Yang, A. *et al.* CHDgene: A Curated Database for Congenital Heart Disease Genes. *Circ Genom Precis Med* **15**, e003539 (2022).
9. Williams, K., Carson, J. & Lo, C. Genetics of congenital heart disease. *Biomolecules* **9**, 879 (2019).
10. Richards, A. A. & Garg, V. Genetics of congenital heart disease. *Curr. Cardiol. Rev.* **6**, 91–97 (2010).
11. Pierpont, M. E. *et al.* Genetic basis for congenital heart disease: Revisited: A scientific statement from the American Heart Association. *Circulation* **138**, e653–e711 (2018).
12. Lahm, H. *et al.* Congenital heart disease risk loci identified by genome-wide association study in European patients. *J. Clin. Invest.* **131**, (2021).
13. Yu, M. *et al.* Oligogenic Architecture of Rare Noncoding Variants Distinguishes 4 Congenital Heart Disease Phenotypes. *Circ Genom Precis Med* **16**, 258–266 (2023).
14. Broberg, M. *et al.* Genome-wide association studies highlight novel risk loci for septal defects and left-sided congenital heart defects. *BMC Genomics* **25**, 256 (2024).
15. Agopian, A. J. *et al.* Genome-wide association studies and meta-analyses for congenital heart defects. *Circ. Cardiovasc. Genet.* **10**, e001449 (2017).
16. Richter, F. *et al.* Genomic analyses implicate noncoding de novo variants in congenital heart disease. *Nat. Genet.* **52**, 769–777 (2020).
17. Xiao, F. *et al.* Functional dissection of human cardiac enhancers and noncoding de novo variants in congenital heart disease. *Nat. Genet.* **56**, 420–430 (2024).
18. Jang, M. Y. *et al.* Contribution of previously unrecognized RNA splice-altering variants to congenital heart disease. *Circ. Genom. Precis. Med.* **16**, 224–231 (2023).
19. Garg, V. *et al.* GATA4 mutations cause human congenital heart defects and reveal an interaction with TBX5. *Nature* **424**, 443–447 (2003).
20. Schott, J. J. *et al.* Congenital heart disease caused by mutations in the transcription factor NKX2-5. *Science* **281**, 108–111 (1998).
21. Basson, C. T. *et al.* Mutations in human cause limb and cardiac malformation in Holt-Oram syndrome. *Nature Genetics* **15**, 30–35 (1997).
22. Yi Li, Q. *et al.* Holt-Oram syndrome is caused by mutations in TBX5, a member of the Brachyury (T) gene family. *Nature Genetics* **15**, 21–29 (1997).
23. Kodo, K. *et al.* GATA6 mutations cause human cardiac outflow tract defects by disrupting semaphorin-plexin signaling. *Proc Natl Acad Sci U S A* **106**, 13933–13938 (2009).
24. Keyte, A. & Hutson, M. R. The neural crest in cardiac congenital anomalies. *Differentiation* **84**, 25–40 (2012).
25. Miao, Y. *et al.* Single-cell RNA-seq reveals endocardial defect in hypoplastic left heart syndrome. (2019)

doi:10.1101/809731.

26. Cordell, H. J. *et al.* Genome-wide association study of multiple congenital heart disease phenotypes identifies a susceptibility locus for atrial septal defect at chromosome 4p16. *Nat. Genet.* **45**, 822–824 (2013).
27. Hu, Z. *et al.* A genome-wide association study identifies two risk loci for congenital heart malformations in Han Chinese populations. *Nat. Genet.* **45**, 818–821 (2013).
28. Cordell, H. J. *et al.* Genome-wide association study identifies loci on 12q24 and 13q32 associated with Tetralogy of Fallot. *Hum. Mol. Genet.* **22**, 1473–1481 (2013).
29. Meyer, H. V. *et al.* Genetic and functional insights into the fractal structure of the heart. *Nature* **584**, 589–594 (2020).
30. Vasan, R. S. *et al.* Genetic variants associated with cardiac structure and function: a meta-analysis and replication of genome-wide association data. *JAMA* **302**, 168–178 (2009).
31. Fox, E. R. *et al.* Genome-wide association study of cardiac structure and systolic function in African Americans: the Candidate Gene Association Resource (CARe) study: The Candidate gene Association Resource (CARe) study. *Circ. Cardiovasc. Genet.* **6**, 37–46 (2013).
32. Aung, N. *et al.* Genome-wide analysis of left ventricular image-derived phenotypes identifies fourteen loci associated with cardiac morphogenesis and heart failure development. *Circulation* **140**, 1318–1330 (2019).
33. Yu, M. *et al.* Genetic determinants of the interventricular septum are linked to ventricular septal defects and hypertrophic cardiomyopathy. *Circ. Genom. Precis. Med.* **16**, 207–215 (2023).
34. Yu, M. *et al.* Computational estimates of annular diameter reveal genetic determinants of mitral valve function and disease. *JCI Insight* **7**, (2022).
35. Pirruccello, J. P. *et al.* Analysis of cardiac magnetic resonance imaging in 36,000 individuals yields genetic insights into dilated cardiomyopathy. *Nat. Commun.* **11**, 2254 (2020).
36. Pirruccello, J. P. *et al.* Genetic analysis of right heart structure and function in 40,000 people. *Nat. Genet.* **54**, 792–803 (2022).
37. Lu, C. *et al.* Genetic architecture of cardiac structure and function. *medRxiv* (2024).
38. Wild, P. S. *et al.* Large-scale genome-wide analysis identifies genetic variants associated with cardiac structure and function. *J. Clin. Invest.* **127**, 1798–1812 (2017).
39. Feng, W. *et al.* Single-cell transcriptomic analysis identifies murine heart molecular features at embryonic and neonatal stages. *Nat. Commun.* **13**, 1–19 (2022).
40. Goodyer, W. R. *et al.* Transcriptomic Profiling of the Developing Cardiac Conduction System at Single-Cell Resolution. *Circulation research* **125**, (2019).
41. Phansalkar, R. *et al.* Coronary blood vessels from distinct origins converge to equivalent states during mouse and human development. *Elife* **10**, (2021).
42. Li, G. *et al.* Transcriptomic Profiling Maps Anatomically Patterned Subpopulations among Single Embryonic Cardiac Cells. *Developmental cell* **39**, (2016).
43. Jia, G. *et al.* Single cell RNA-seq and ATAC-seq analysis of cardiac progenitor cell transition states and lineage settlement. *Nature Communications* **9**, 1–17 (2018).
44. Domcke, S. *et al.* A human cell atlas of fetal chromatin accessibility. *Science* **370**, eaba7612 (2020).
45. Cranley, J. *et al.* Multiomic analysis reveals developmental dynamics of the human heart in health and disease. *bioRxiv* (2024).
46. Bayraktar, S. *et al.* High-resolution atlas of the developing human heart and the great vessels. *bioRxiv* (2024).
47. Farah, E. N. *et al.* Spatially organized cellular communities form the developing human heart. *Nature* **627**, 854–864 (2024).
48. Lázár, E. *et al.* Spatial Dynamics of the Developing Human Heart. *bioRxiv* (2024).
49. Miao, Y. *et al.* Intrinsic endocardial defects contribute to hypoplastic left heart syndrome. *Cell Stem Cell* **27**, 574–589.e8 (2020).
50. Liu, Z. *et al.* APOE-NOTCH axis governs elastogenesis during human cardiac valve remodeling. *Nat.*

Cardiovasc. Res. **3**, 933–950 (2024).

51. Leshem, R. *et al.* A cell atlas of the developing human outflow tract of the heart and its adult derivatives. *bioRxiv* (2023).
52. McCracken, I. R. *et al.* Mapping the developing human cardiac endothelium at single-cell resolution identifies MECOM as a regulator of arteriovenous gene expression. *Cardiovascular research* **118**, (2022).
53. IGVF Consortium. Deciphering the impact of genomic variation on function. *Nature* **633**, 47–57 (2024).
54. Sheth, M. U. *et al.* Mapping enhancer-gene regulatory interactions from single-cell data. *In preparation*.
55. Pampari, A., Shcherbina, A. & Kundaje, A. Bias factorized, base-resolution deep learning models of chromatin accessibility reveal cis-regulatory sequence syntax, transcription factor footprints and regulatory variants. *https://github.com/kundajelab/chrombpnet* https://github.com/kundajelab/chrombpnet.
56. Shrikumar, A., Greenside, P. & Kundaje, A. Learning Important Features Through Propagating Activation Differences. in *Proceedings of the 34th International Conference on Machine Learning* (eds. Precup, D. & Teh, Y. W.) vol. 70 3145–3153 (PMLR, 06--11 Aug 2017).
57. Shrikumar, A. *et al.* Technical Note on Transcription Factor Motif Discovery from Importance Scores (TF-MoDISco) version 0.5.6.5. *arXiv [cs.LG]* (2018).
58. Avsec, Ž. *et al.* Base-resolution models of transcription-factor binding reveal soft motif syntax. *Nat. Genet.* **53**, 354–366 (2021).
59. Martyn, G. E. *et al.* Rewriting regulatory DNA to dissect and reprogram gene expression. *bioRxiv* 2023.12.20.572268 (2023) doi:10.1101/2023.12.20.572268.
60. Kotliar, D. *et al.* Identifying gene expression programs of cell-type identity and cellular activity with single-cell RNA-Seq. *Elife* **8**, (2019).
61. Schnitzler, G. R. *et al.* Convergence of coronary artery disease genes onto endothelial cell programs. *Nature* **626**, 799–807 (2024).
62. Maitra, M., Koenig, S. N., Srivastava, D. & Garg, V. Identification of GATA6 sequence variants in patients with congenital heart defects. *Pediatr. Res.* **68**, 281–285 (2010).
63. de la Pompa, J. L. *et al.* Role of the NF-ATc transcription factor in morphogenesis of cardiac valves and septum. *Nature* **392**, 182–186 (1998).
64. Goddard, L. M. *et al.* Hemodynamic forces sculpt developing heart valves through a KLF2-WNT9B paracrine signaling axis. *Dev. Cell* **43**, 274–289.e5 (2017).
65. Oh, Y. *et al.* Transcriptional regulation of the postnatal cardiac conduction system heterogeneity. *Nat. Commun.* **15**, 6550 (2024).
66. Galang, G. *et al.* ATAC-seq reveals an Isl1 enhancer that regulates sinoatrial node development and function. *Circ. Res.* **127**, 1502–1518 (2020).
67. Jin, S. *et al.* Inference and analysis of cell-cell communication using CellChat. *Nature Communications* **12**, 1–20 (2021).
68. van der Harst, P. *et al.* 52 genetic loci influencing myocardial mass. *J. Am. Coll. Cardiol.* **68**, 1435–1448 (2016).
69. Libiseller-Egger, J. *et al.* Deep learning-derived cardiovascular age shares a genetic basis with other cardiac phenotypes. *Sci. Rep.* **12**, 22625 (2022).
70. Li, J. *et al.* EphA4 is highly expressed in the atria of heart and its deletion leads to atrial hypertrophy and electrocardiographic abnormalities in rats. *Life Sci.* **278**, 119595 (2021).
71. Kanemaru, K. *et al.* Spatially resolved multiomics of human cardiac niches. *Nature* **619**, 801–810 (2023).
72. Rombouts, K. B. *et al.* The role of vascular smooth muscle cells in the development of aortic aneurysms and dissections. *Eur J Clin Invest* **52**, e13697 (2022).
73. O'Donnell, A. & Yutzey, K. E. Mechanisms of heart valve development and disease. *Development* **147**, (2020).
74. Morlino, S. *et al.* LTBP2-related 'Marfan-like' phenotype in two Roma/Gypsy subjects with the LTBP2 homozygous p.R299X variant. *Am. J. Med. Genet. A* **179**, 104–112 (2019).

75. Haji-Seyed-Javadi, R. *et al.* LTBP2 mutations cause Weill-Marchesani and Weill-Marchesani-like syndrome and affect disruptions in the extracellular matrix. *Hum. Mutat.* **33**, 1182–1187 (2012).

76. Shipley, J. M. *et al.* Developmental expression of latent transforming growth factor beta binding protein 2 and its requirement early in mouse development. *Mol. Cell. Biol.* **20**, 4879–4887 (2000).

77. Shpitzen, S. *et al.* Characterization of LTBP2 mutation causing Mitral Valve Prolapse. *medRxiv* 2024.07.21.24302849 (2024) doi:10.1101/2024.07.21.24302849.

78. Tallquist, M. D. Developmental Pathways of Cardiac Fibroblasts. *Cold Spring Harb Perspect Biol* **12**, (2020).

79. Douglas, G. *et al.* De novo missense variants in MEIS2 recapitulate the microdeletion phenotype of cardiac and palate abnormalities, developmental delay, intellectual disability and dysmorphic features. *Am. J. Med. Genet. A* **176**, 1845–1851 (2018).

80. Fujita, A. *et al.* De novo MEIS2 mutation causes syndromic developmental delay with persistent gastro-esophageal reflux. *J. Hum. Genet.* **61**, 835–838 (2016).

81. Louw, J. J. *et al.* MEIS2 involvement in cardiac development, cleft palate, and intellectual disability. *Am. J. Med. Genet. A* **167A**, 1142–1146 (2015).

82. Verheije, R. *et al.* Heterozygous loss-of-function variants of MEIS2 cause a triad of palatal defects, congenital heart defects, and intellectual disability. *Eur. J. Hum. Genet.* **27**, 278–290 (2019).

83. De la Cruz Crespillo, M. E. The role of Meis transcription factors in the epicardium. (SciELO Espana/Repisalud, 2021). doi:10.4321/repisalud.13255.

84. Machon, O., Masek, J., Machonova, O., Krauss, S. & Kozmik, Z. Meis2 is essential for cranial and cardiac neural crest development. *BMC Developmental Biology* **15**, 40 (2015).

85. Stankunas, K. *et al.* Pbx/Meis deficiencies demonstrate multigenetic origins of congenital heart disease. *Circ. Res.* **103**, 702–709 (2008).

86. Roselli, C. *et al.* Genome-wide association study reveals novel genetic loci: a new polygenic risk score for mitral valve prolapse. *Eur. Heart J.* **43**, 1668–1680 (2022).

87. Kany, S. & Pirruccello, J. P. *In preparation*.

88. Kany, S. *et al.* Assessment of valvular function in over 47,000 people using deep learning-based flow measurements. *medRxiv* 2023.04.29.23289299 (2023) doi:10.1101/2023.04.29.23289299.

89. Kany, S. & Pirruccello, J. 2024 AV summary statistics. Zenodo https://doi.org/10.5281/ZENODO.14025285 (2024).

90. Pirruccello, J., Damrauer, S., Zamirpour, S. & DePaolo, J. Thoracic aorta GWAS summary statistics. Zenodo https://doi.org/10.5281/ZENODO.14025251 (2024).

91. 2024 AV summary statistics. doi:10.5281/zenodo.14025286.

92. Yun, W. J. *et al.* The hematopoietic regulator TAL1 is required for chromatin looping between the β-globin LCR and human γ-globin genes to activate transcription. *Nucleic Acids Res.* **42**, 4283–4293 (2014).

93. Hinton, R. B. *et al.* Elastin haploinsufficiency results in progressive aortic valve malformation and latent valve disease in a mouse model. *Circ. Res.* **107**, 549–557 (2010).

94. Duque Lasio, M. L. & Kozel, B. A. Elastin-driven genetic diseases. *Matrix Biol.* **71-72**, 144–160 (2018).

95. Dina, C. *et al.* Genetic association analyses highlight biological pathways underlying mitral valve prolapse. *Nat. Genet.* **47**, 1206–1211 (2015).

96. Bocciardi, R. *et al.* Overexpression of the C-type natriuretic peptide (CNP) is associated with overgrowth and bone anomalies in an individual with balanced t(2;7) translocation. *Hum. Mutat.* **28**, 724–731 (2007).

97. Boudin, E. *et al.* Bi-allelic Loss-of-Function Mutations in the NPR-C Receptor Result in Enhanced Growth and Connective Tissue Abnormalities. *Am. J. Hum. Genet.* **103**, 288–295 (2018).

98. Gehlen, J. *et al.* Elucidation of the genetic causes of bicuspid aortic valve disease. *Cardiovasc Res* **119**, 857–866 (2023).

99. Kurki, M. I. *et al.* FinnGen provides genetic insights from a well-phenotyped isolated population. *Nature* **613**, 508–518 (2023).

100. Sébastien Thériault *et al.* A transcriptome-wide association study identifies PALMD as a susceptibility gene

for calcific aortic valve stenosis. *Nat. Commun.* **9**, (2018).

101. Chignon, A. *et al.* Enhancer-associated aortic valve stenosis risk locus 1p21.2 alters NFATC2 binding site and promotes fibrogenesis. *iScience* **24**, 102241 (2021).
102. Han, Y. *et al.* Palmdelphin deficiency evokes NF-κB signaling in valvular endothelial cells and aggravates aortic valvular remodeling. *JACC Basic Transl. Sci.* **8**, 1457–1472 (2023).
103. Karlebach, G. *et al.* Alternative splicing is coupled to gene expression in a subset of variably expressed genes. *NPJ Genom Med* **9**, 54 (2024).
104. Morley, M. P. *et al.* Cardioprotective effects of MTSS1 enhancer variants. *Circulation* **139**, 2073–2076 (2019).
105. Lie-Venema, H. *et al.* Origin, fate, and function of epicardium-derived cells (EPDCs) in normal and abnormal cardiac development. *ScientificWorldJournal* **7**, 1777–1798 (2007).
106. Hill, M. C. *et al.* Integrated multi-omic characterization of congenital heart disease. *Nature* **608**, 181–191 (2022).
107. Gifford, C. A. *et al.* Oligogenic inheritance of a human heart disease involving a genetic modifier. *Science* **364**, 865–870 (2019).
108. Spendlove, S. J., Bondhus, L., Lluri, G., Sul, J. H. & Arboleda, V. A. Polygenic risk scores of endo-phenotypes identify the effect of genetic background in congenital heart disease. *HGG Adv.* **3**, 100112 (2022).
109. Corces, M. R. *et al.* The chromatin accessibility landscape of primary human cancers. *Science* **362**, (2018).
110. Kaminow, B., Yunusov, D. & Dobin, A. STARsolo: accurate, fast and versatile mapping/quantification of single-cell and single-nucleus RNA-seq data. *bioRxiv* (2021).
111. Fleming, S. J. *et al.* Unsupervised removal of systematic background noise from droplet-based single-cell experiments using CellBender. *Nat Methods* **20**, 1323–1335 (2023).
112. GitHub - GreenleafLab/matcha: Python library for fast barcode matching in sequencing data. *GitHub* https://github.com/GreenleafLab/matcha.
113. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357–359 (2012).
114. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
115. Picard. http://broadinstitute.github.io/picard/.
116. Granja, J. M. *et al.* ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. *Nat Genet* **53**, 403–411 (2021).
117. GitHub - pachterlab/BHGP_2022. *GitHub* https://github.com/pachterlab/BHGP_2022.
118. Thibodeau, A. *et al.* AMULET: a novel read count-based method for effective multiplet detection from single nucleus ATAC-seq data. *Genome Biol* **22**, 252 (2021).
119. Heaton, H. *et al.* Souporcell: robust clustering of single-cell RNA-seq data by genotype without reference genotypes. *Nat Methods* **17**, 615–620 (2020).
120. Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587.e29 (2021).
121. Zappia, L. & Oshlack, A. Clustering trees: a visualization for evaluating clusterings at multiple resolutions. *Gigascience* **7**, (2018).s
122. Neufeld, A. *et al.* Negative binomial count splitting for single-cell RNA sequencing data. *arxiv* (2023) doi:10.48550/arXiv.2307.12985.
123. GitHub - satijalab/sctransform: R package for modeling single cell UMI expression data using regularized negative binomial regression. *GitHub* https://github.com/satijalab/sctransform.
124. Gschwind, A. R. *et al.* An encyclopedia of enhancer-gene regulatory interactions in the human genome. *bioRxiv* (2023) doi:10.1101/2023.11.09.563812.
125. Bailey, T. L., Johnson, J., Grant, C. E. & Noble, W. S. The MEME suite. *Nucleic Acids Res.* **43**, W39–49 (2015).
126. Yu, G., Wang, L.-G., Han, Y. & He, Q.-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* **16**, 284–287 (2012).

127. MSigDB gene sets R package. https://igordot.github.io/msigdbr/.

128. GitHub - GreenleafLab/chromVARmotifs: R package with motifs for use with chromVAR. *GitHub* https://github.com/GreenleafLab/chromVARmotifs.

129. Schep, A. N., Wu, B., Buenrostro, J. D. & Greenleaf, W. J. chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. *Nat Methods* **14**, 975–978 (2017).

130. Finucane, H. K. *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).

131. GitHub - bulik/ldsc: LD Score Regression (LDSC). *GitHub* https://github.com/bulik/ldsc.

132. Pirruccello, J. 2024 LA summary statistics. Zenodo https://doi.org/10.5281/ZENODO.14025187 (2024).

133. Pirruccello, J. & Padmanabhan, A. 2024 LV summary statistics. Zenodo https://doi.org/10.5281/ZENODO.14025158 (2024).

134. Paszke, A. *et al.* PyTorch: An Imperative Style, High-Performance Deep Learning Library. (2019).

135. Pirruccello, J. P. *et al.* Deep learning of left atrial structure and function provides link to atrial fibrillation risk. *Nat Commun* **15**, 4304 (2024).

136. Screened poisson surface reconstruction. *ACM Transactions on Graphics (TOG)* (2013) doi:10.1145/2487228.2487237.

137. Pirruccello, J. P. *et al.* Deep learning enables genetic analysis of the human thoracic aorta. *Nat. Genet.* **54**, 40–51 (2022).

138. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).

139. Klarin, D. *et al.* Genome-wide association study of thoracic aortic aneurysm and dissection in the Million Veteran Program. *Nat Genet* **55**, 1106–1115 (2023).

140. Turley, P. *et al.* Multi-trait analysis of genome-wide association summary statistics using MTAG. *Nat Genet* **50**, 229–237 (2018).

141. Small, A. M. *et al.* Multiancestry Genome-Wide Association Study of Aortic Stenosis Identifies Multiple Novel Loci in the Million Veteran Program. *Circulation* **147**, 942–955 (2023).

142. Yu Chen, H. *et al.* Dyslipidemia, inflammation, calcification, and adiposity in aortic stenosis: a genome-wide study. *Eur Heart J* **44**, 1927–1939 (2023).

143. Kanai, M. *et al.* Insights from complex trait fine-mapping across diverse populations. *medRxiv* 2021.09.03.21262975 (2021) doi:10.1101/2021.09.03.21262975.

144. Benner, C. *et al.* Prospects of Fine-Mapping Trait-Associated Genomic Regions by Using Summary Statistics from Genome-wide Association Studies. *Am J Hum Genet* **101**, 539–551 (2017).

145. Cui, R. *et al.* Improving fine-mapping by modeling infinitesimal effects. *Nat Genet* **56**, 162–169 (2024).

146. Nasser, J. *et al.* Genome-wide enhancer maps link risk variants to disease genes. *Nature* **593**, 238–243 (2021).

147. Osterwalder, M. *et al.* Characterization of mammalian in vivo enhancers using mouse transgenesis and CRISPR genome editing. *Methods Mol. Biol.* **2403**, 147–186 (2022).

148. Kosicki, M. *et al.* VISTA Enhancer browser: an updated database of tissue-specific developmental enhancers. *Nucleic Acids Res* (2024) doi:10.1093/nar/gkae940.

149. Dye, B. & Lincoln, J. The Endocardium and Heart Valves. *Cold Spring Harb Perspect Biol* **12**, (2020).

150. Beppu, H. *et al.* BMP type II receptor regulates positioning of outflow tract and remodeling of atrioventricular cushion during cardiogenesis. *Dev Biol* **331**, 167–175 (2009).

151. Roberts, K. E. *et al.* BMPR2 mutations in pulmonary arterial hypertension with congenital heart disease. *Eur Respir J* **24**, 371–374 (2004).

152. Itoh, N., Ohta, H., Nakayama, Y. & Konishi, M. Roles of FGF Signals in Heart Development, Health, and Disease. *Front Cell Dev Biol* **4**, 110 (2016).

153. Lavine, K. J. *et al.* Endocardial and epicardial derived FGF signals regulate myocardial proliferation and differentiation in vivo. *Dev Cell* **8**, 85–95 (2005).

154. Marguerie, A. *et al.* Congenital heart defects in Fgfr2-IIIb and Fgf10 mutant mice. *Cardiovasc Res* **71**, 50–60 (2006).

155. Zhang, J. *et al.* Frs2alpha-deficiency in cardiac progenitors disrupts a subset of FGF signals required for outflow tract morphogenesis. *Development* **135**, 3611–3622 (2008).

156. Vega-Hernández, M., Kovacs, A., De Langhe, S. & Ornitz, D. M. FGF10/FGFR2b signaling is essential for cardiac fibroblast development and growth of the myocardium. *Development* **138**, 3331–3340 (2011).

157. Rivera-Feliciano, J. *et al.* Development of heart valves requires Gata4 expression in endothelial-derived cells. *Development* **133**, 3607 (2006).

158. Kolander, K. D., Holtz, M. L., Cossette, S. M., Duncan, S. A. & Misra, R. P. Epicardial GATA factors regulate early coronary vascular plexus formation. *Dev. Biol.* **386**, 204 (2014).

159. Lepore, J. J. *et al.* GATA-6 regulates semaphorin 3C and is required in cardiac neural crest for cardiovascular morphogenesis. *J Clin Invest* **116**, 929–939 (2006).

160. Tian, Y. *et al.* Characterization and in vivo pharmacological rescue of a Wnt2-Gata6 pathway required for cardiac inflow tract development. *Dev Cell* **18**, 275–287 (2010).

161. Zhou, B. *et al.* Fog2 is critical for cardiac function and maintenance of coronary vasculature in the adult mouse heart. *J. Clin. Invest.* **119**, 1462 (2009).

162. Fulmer, D. *et al.* Desert hedgehog-primary cilia cross talk shapes mitral valve tissue by organizing smooth muscle actin. *Dev Biol* **463**, 26–38 (2020).

163. Toomer, K. A. *et al.* A role for primary cilia in aortic valve development and disease. *Dev Dyn* **246**, 625–634 (2017).

164. Cottereau, E. *et al.* Phenotypic spectrum of Simpson-Golabi-Behmel syndrome in a series of 42 cases with a mutation in GPC3 and review of the literature. *Am J Med Genet C Semin Med Genet* **163C**, 92–105 (2013).

165. Ng, A. *et al.* Loss of glypican-3 function causes growth factor-dependent defects in cardiac and coronary vascular development. *Dev Biol* **335**, 208–215 (2009).

166. McQuinn, T. C., Miga, D. E., Mjaatvedt, C. H., Phelps, A. L. & Wessels, A. Cardiopulmonary malformations in the inv/inv mouse. *Anat Rec* **263**, 62–71 (2001).

167. Otto, E. A. *et al.* Mutations in INVS encoding inversin cause nephronophthisis type 2, linking renal cystic disease to the function of primary cilia and left-right axis determination. *Nat Genet* **34**, 413–420 (2003).

168. Tory, K. *et al.* Mutations of NPHP2 and NPHP3 in infantile nephronophthisis. *Kidney Int* **75**, 839–847 (2009).

169. Baek, S. T. & Tallquist, M. D. Nf1 limits epicardial derivative expansion by regulating epithelial to mesenchymal transition and proliferation. *Development* **139**, 2040–2049 (2012).

170. Wang, D. *et al.* in heart development: a potential causative gene for congenital heart disease: a narrative review. *Physiol Genomics* **55**, 415–426 (2023).

171. Varadkar, P. *et al.* Notch2 is required for the proliferation of cardiac neural crest-derived smooth muscle cells. *Dev Dyn* **237**, 1144–1152 (2008).

172. Chang, C.-P. *et al.* Pbx1 functions in distinct regulatory networks to pattern the great arteries and cardiac outflow tract. *Development* **135**, 3577–3586 (2008).

173. Waheed-Ullah, Q. *et al.* Effect of deletion of the protein kinase PRKD1 on development of the mouse embryonic heart. *J Anat* **245**, 70–83 (2024).

174. Verma, S. K. *et al.* RBFOX2 is required for establishing RNA regulatory networks essential for heart development. *Nucleic Acids Res* **50**, 2270–2286 (2022).

175. Cheung, M. Y.-Q., Roberts, C., Scambler, P. & Stathopoulou, A. Setd5 is required in cardiopharyngeal mesoderm for heart development and its haploinsufficiency is associated with outflow tract defects in mouse. *Genesis* **59**, e23421 (2021).

176. Galvin, K. M. *et al.* A role for smad6 in development and homeostasis of the cardiovascular system. *Nat. Genet.* **24**, (2000).

177. Boogerd, C. J. *et al.* Probing chromatin landscape reveals roles of endocardial TBX20 in septation. *J Clin Invest* **126**, 3023–3035 (2016).

178. Cai, X. *et al.* Tbx20 acts upstream of Wnt signaling to regulate endocardial cushion formation and valve

remodeling during mouse cardiogenesis. *Development* **140**, 3176 (2013).

179. Kirk, E. P. *et al.* Mutations in cardiac T-box factor gene TBX20 are associated with diverse cardiac pathologies, including defects of septation and valvulogenesis and cardiomyopathy. *Am J Hum Genet* **81**, 280–291 (2007).
180. Takeuchi, J. K. *et al.* Tbx20 dose-dependently regulates transcription factor networks required for mouse heart and motoneuron development. *Development* **132**, 2463–2474 (2005).
181. Peng, Y. *et al.* Critical roles of miRNA-mediated regulation of TGFβ signalling during mouse cardiogenesis. *Cardiovasc Res* **103**, 258–267 (2014).
182. Sridurongrit, S., Larsson, J., Schwartz, R., Ruiz-Lozano, P. & Kaartinen, V. Signaling via the Tgf-beta type I receptor Alk5 in heart development. *Dev Biol* **322**, 208–218 (2008).
183. Tevosian, S. G. *et al.* FOG-2, a cofactor for GATA transcription factors, is essential for heart morphogenesis and development of coronary vessels from epicardium. *Cell* **101**, 729–739 (2000).
184. Svensson, E. C. *et al.* A syndrome of tricuspid atresia in mice with a targeted mutation of the gene encoding Fog-2. *Nat Genet* **25**, 353–356 (2000).
185. Conley, S. D. Figure S1. Experimental Pipeline. *Created in BioRender* https://BioRender.com/j97f964 (2024).
186. Tian, X. *et al.* Identification of a hybrid myocardial zone in the mammalian heart after birth. *Nat Commun* **8**, 87 (2017).
187. Phansalkar, R. & Red-Horse, K. A new resource for human coronary vessel development. *Cardiovasc Res* **118**, 2875–2876 (2022).
188. Schupp, J. C. *et al.* Integrated Single-Cell Atlas of Endothelial Cells of the Human Lung. *Circulation* **144**, 286–302 (2021).
189. Boulanger, C. & Lüscher, T. F. Release of endothelin from the porcine aorta. Inhibition by endothelium-derived nitric oxide. *J Clin Invest* **85**, 587–590 (1990).
190. Ekman, N. *et al.* Bmx tyrosine kinase is specifically expressed in the endocardium and the endothelium of large arteries. *Circulation* **96**, 1729–1732 (1997).
191. Rutkovskiy, A. *et al.* Valve Interstitial Cells: The Key to Understanding the Pathophysiology of Heart Valve Calcification. *J Am Heart Assoc* **6**, (2017).
192. Knight-Schrijver, V. R. *et al.* A single-cell comparison of adult and fetal human epicardium defines the age-associated changes in epicardial activity. *Nat Cardiovasc Res* **1**, 1215–1229 (2022).
193. Wong, D., Martinez, J. & Quijada, P. Exploring the Function of Epicardial Cells Beyond the Surface. *Circ Res* **135**, 353–371 (2024).
194. Lupu, I.-E., Redpath, A. N. & Smart, N. Spatiotemporal Analysis Reveals Overlap of Key Proepicardial Markers in the Developing Murine Heart. *Stem Cell Reports* **14**, 770–787 (2020).
195. Volz, K. S. *et al.* Pericytes are progenitors for coronary artery smooth muscle. *Elife* **4**, (2015).
196. Chen, R. *et al.* Macrophages in cardiovascular diseases: molecular mechanisms and therapeutic targets. *Signal Transduct Target Ther* **9**, 130 (2024).
197. Suku, M., Forrester, L., Biggs, M. & Monaghan, M. G. Resident Macrophages and Their Potential in Cardiac Tissue Engineering. *Tissue Eng Part B Rev* **28**, 579–591 (2022).
198. Szabo, P. A. *et al.* Single-cell transcriptomics of human T cells reveals tissue and activation signatures in health and disease. *Nat Commun* **10**, 4706 (2019).
199. Yang, C. *et al.* Heterogeneity of human bone marrow and blood natural killer cells defined by single-cell transcriptome. *Nat Commun* **10**, 3931 (2019).
200. Gómez Hernández, G., Morell, M. & Alarcón-Riquelme, M. E. The Role of BANK1 in B Cell Signaling and Disease. *Cells* **10**, (2021).
201. Somasundaram, R. *et al.* EBF1 and PAX5 control pro-B cell expansion via opposing regulation of the Myc gene. *Blood* **137**, 3037–3049 (2021).
202. Idelman, G. *et al.* Inducible pluripotent stem cells to study human mast cell trajectories. *Mucosal Immunol* **17**, 1029–1044 (2024).

203. Li, Y. *et al.* GATA2 regulates mast cell identity and responsiveness to antigenic stimulation by promoting chromatin remodeling at super-enhancers. *Nat Commun* **12**, 494 (2021).

204. Prieto-García, A. *et al.* Mast cell restricted mouse and human tryptase·heparin complexes hinder thrombin-induced coagulation of plasma and the generation of fibrin by proteolytically destroying fibrinogen. *J Biol Chem* **287**, 7834–7844 (2012).

205. Simões-Costa, M. & Bronner, M. E. Establishing neural crest identity: a gene regulatory recipe. *Development* **142**, 242–257 (2015).

206. Takayama, Y. *et al.* Selective Induction of Human Autonomic Neurons Enables Precise Control of Cardiomyocyte Beating. *Sci Rep* **10**, 9464 (2020).

207. Pattyn, A., Morin, X., Cremer, H., Goridis, C. & Brunet, J. F. The homeobox gene Phox2b is essential for the development of autonomic neural crest derivatives. *Nature* **399**, 366–370 (1999).

208. Dyachuk, V. *et al.* Neurodevelopment. Parasympathetic neurons originate from nerve-associated peripheral glial progenitors. *Science* **345**, 82–87 (2014).