eLife

Reviewed Preprint
v1 • September 11, 2025
Not revised

Reviewed Preprint
v2 • April 21, 2026
Revised by authors

For correspondence:
Simon.Bamforth@newcastle.ac.uk
Nicoletta.Bobola@manchester.ac.uk

* equal contribution

Competing interests: No competing interests declared

Funding: See page 20

Reviewing editor: Sylvia Evans, University of California, San Diego, United States

A cell atlas of the developing human outflow tract of the heart and its adult aortic valve derivatives

Rotem Leshem[1,*], Syed Murtuza Baker[1,*], Joshua Mallen[1], Lu Wang[2], John Dark[2], Andrew D Sharrocks[1], Karen Piper Hanley[1], Neil A Hanley[3,1], Magnus Rattray[1], Simon D Bamforth[4] ✉, Nicoletta Bobola[1] ✉

[1]Faculty of Biology, Medicine and Health, University of Manchester, Manchester, United Kingdom • [2]Translational and Clinical Research Institute, Newcastle University, Newcastle, United Kingdom • [3]College of Medicine & Health, University of Birmingham, Birmingham, United Kingdom • [4]Newcastle University Biosciences Institute, Faculty of Medical Sciences, Newcastle, United Kingdom

eLife Assessment

This is a **valuable** study that presents human single nuclei RNA-seq and spatial transcriptomics data of the developing outflow tract and adult aortic valves that will facilitate research in this area. Data presented are **solid**, with bioinformatics analyses showing cell lineage and trajectory relationships, intriguingly suggesting persistence of embryonic signature in adult aortic valve cells. The latter results would be strengthened by experimental validation.

https://doi.org/10.7554/eLife.107748.2.sa4

Abstract

The outflow tract (OFT) of the heart carries blood away from the heart into the great arteries. During embryogenesis, the OFT divides to form the aorta and pulmonary trunk, creating the double circulation present in mammals. Defects in this area account for one-third of all congenital heart defect cases. Here, we present comprehensive transcriptomic data on the developing OFT at two distinct timepoints (embryonic and fetal) and its adult derivatives, the aortic valves, and use spatial transcriptomics to define the distribution of cell populations. We uncover that distinctive embryonic signatures persist in adult cells and can be used as labels to retrospectively attribute relationships between cells separated by a large time scale. Single-cell regulatory network inference identifies GATA6, a transcription factor linked to common arterial trunk and bicuspid aortic valve, as a key regulator of valve precursor cells. Its downstream network reveals candidate drivers of human cardiac defects and illuminates the molecular mechanisms of both normal and pathological valve development. Our findings define the cellular and molecular signatures of the human OFT and its distinct cell lineages, which is critical for understanding congenital heart defects and developing cardiac tissue for regenerative medicine.

Introduction

Congenital heart defects (CHD) are major birth abnormalities affecting ~1% of new born babies [1]. The outflow tract (OFT) carries blood away from the heart into the great arteries. During embryogenesis the OFT divides to provide separate aorta and pulmonary trunk vessels, which arise from the left and right ventricles respectively, giving the double circulation found in mammals [2]. Defects specifically affecting the OFT of the heart represent a third of all CHD cases [3]. In humans, the formation and remodelling of the OFT is a relatively rapid process, occurring in embryogenesis over a ~ 4-week period from Carnegie Stage (CS)13 (4 weeks) to CS23 (8 weeks). Septation of the OFT begins around the CS14 stage when a transient aortopulmonary septal complex protrudes from the dorsal wall of the aortic sac [4]. This divides the common OFT vessel

into separate aorta and pulmonary trunks. The arterial (semilunar) valves are formed in the intermediate component of the OFT, initially as endocardial cushions [5] and continue to mature after septation of the arterial vessels. The morphological changes underlying OFT formation are orchestrated by two main cell lineages, the second heart field and the cardiac neural crest. The second heart field is a population of cardiac progenitor cells, originating from the pharyngeal mesoderm, that contribute to the formation of the myocardium of the right ventricle, atria and the OFT of the heart [6]. The neural crest is a transient, pluripotent cell population, which migrate from the neural tube to multiple areas of the body. The cardiac neural crest, a sub-population of the neural crest, forms the aorticopulmonary septal complex, which separates the aorta and pulmonary trunk [7–9].

The use of model systems has substantially advanced our understanding of the cell lineages that contribute to the OFT. However, much less is known about OFT development in humans. In addition, relative to the embryonic period, adulthood is underexplored at the molecular level, and a clear relationship between cell lineages and cells found in the adult OFT is lacking. Previous studies have conducted single cell analysis of both developing and adult (whole or microdissected) hearts [10–24]. Here, we present comprehensive transcriptomic data of the developing OFT (two distinct timepoints, embryonic and fetal) and its adult derivatives, the aortic valves, providing a large reference framework of OFT cell repertoires and their gene expression profiles. Using spatial transcriptomics, we describe the distribution of cell populations and cell–cell co-localizations. Remarkably, we identify the persistence of distinctive embryonic signatures in cells separated by a large timescale and use these signatures to establish lineage relationship between embryonic and adult cells. Our study defines the cellular and molecular signatures of the developing OFT and adult valves, and highlights the distinct cell lineages that construct these structures. This is of major importance for understanding the origin of congenital heart malformations and for producing cardiac tissue for use in regenerative medicine.

Materials and methods

Sample acquisition

Male CS16-17 (embryonic) and 12 pcw (fetal) OFTs were collected after pregnancy termination and snap frozen by the Human Developmental Biology Resource (*www.hdbr.org*). Embryonic samples were staged according to appearance by the HDBR embryo staging guidelines and pooled due to their limited size, with one CS16 and one CS17 sample in each pool. For fetal (12pcw) OFT, one OFT was used for each single nuclei RNA-seq, and one OFT for spatial transcriptomics. Adult aortic valve tissue was collected from three female adult (age 55-70) hearts declined for clinical transplantation with written informed consent for research from their families. Donors did not have ECHO evidence of aortic valve pathology or past medical history of aortic valve disease and inspection of aortic valve leaflets during dissection showed no signs of calcification. The research ethical approval (REC ref 16/NE/0230) was provided by the North East - Newcastle & North Tyneside Research Ethics Committee. The hearts were retrieved in the clinical standard fashion; arrested with 1 L of cold St Thomas' cardioplegia solution at the agreed time point when both cardiothoracic and abdominal retrieval teams were ready for organ procurement. The donor hearts were then rapidly retrieved and preserved with either static cold storage or a hypothermic oxygenated perfusion device for 4 hours during which they were transported back to the laboratory at Newcastle University. After preservation, they were reanimated on a modified Langendorff system with blood-based oxygenated perfusate for 4 hours. At the end of the normothermic reperfusion, the donor hearts were dissected. Both left and right OFTs containing aortic and pulmonary valves respectively were excised, immediately snap frozen by being submerged into liquid nitrogen and stored at -80°C in The Newcastle Institute of Transplantation Tissue Biobank (17/NE/0022). The material was then obtained from The Newcastle Institute of Transplantation Tissue Biobank for analysis.

Imaging

Human samples were processed for high-resolution episcopic microscopy and micro-computed tomography techniques as previously described. Briefly, aligned serial digital sections were imported into Amira (ThermoFisher Scientific) to produce two- and three-dimensional images [25].

Nuclei extraction

Snap frozen embryonic and fetal tissue samples were minced using either a Dounce Homogeniser or a pellet pestle, then lysed for 30 min in ice cold lysis buffer (10 mM Tris HCl pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$, 0.1%Tween-20, 0.1% Igepal, 0.0005% Digitonin, 0.2 U/µl Protector RNAse inhibitor) or until no tissue pieces could be seen. Lysis was stopped with wash buffer (10 mM Tris HCl pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$, 0.1% Tween-20, 0.1% BSA, 0.2 U/µl Protector RNAse inhibitor) and filtered through a 20 µm filter. Nuclei were resuspended with PBS with Bovine Serum Albumin (0.1% BSA UltraPure™, AM2616 Invitrogen), and Protector RNAse inhibitor (0.2 U/µl, 3335399001 Roche), and processed using 10x Chromium Single Cell 3' with target recovery set at 2500 nuclei/sample. Snap frozen adult aorta samples were cryosectioned to 50 µm, morphology was assessed by Hematoxylin and Eosin (H&E) staining and the valve tissue identified. Valve tissue from 3-4 unstained frozen sections was scraped off the slides, minced using pellet pestle and lysed in ice cold Igepal lysis buffer (10 mM Tris HCl pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$, 0.05% Igepal, 1 mM DTT, 1 U/µl Protector RNAse inhibitor) for 10 min. Sample was filtered using 40 µm cell strainer. After spinning down (500 rcf, 5 min, 4 °C) supernatant was removed and gently replaced with PBS/BSA (PBS containing BSA 1% and RNAse inhibitor 1 U/µl) and incubated for 5 min without disturbing the pellet. Supernatant was then removed, and the pellet resuspended in PBS/BSA supplemented with 7-aminoactinomycin D (7-AAD) dye. Nuclei were FACS sorted using a 100 µm diameter nozzle and a sheath pressure of 20PSI as per 10x protocols (document CG000375). 7AAD allowed nuclei to be identified from debris generated during the processing procedure. 7AAD stained nuclei were excited with a 488nm blue laser and emission was collected through a 685-725nm bandpass filter, at an event rate of ~300-500 events per second. Following sorting, nuclei were permeabilized using 0.05X Lysis buffer (10 mM Tris HCl pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$, 0.01% Igepal, 0.001% Digitonin, 0.05 mM DTT, 1% BSA, 1 U/µl Protector RNAse inhibitor) for 1 minute, then washed and processed using 10x Chromium Single Cell 3' with target recovery set at 2500 nuclei/sample.

Visium spatial gene expression

One 12pcw whole fetal heart sample was embedded in OCT and cryosectioned at 5 µm according to 10x Genomics (document CG000240). A tissue optimisation was performed according to 10x Genomics protocol (document CG000238). Permeabilization time was set to 12 min. Four spatial gene expression sections were collected from the base of the aorta at regular intervals ending at the pulmonary valve. Tissue was processed and libraries created according to manufacturer instructions (document CG000239). 25% Ct value was determined by qPCR at 16 cycles.

Single nuclei isolation and library construction

Gene expression libraries were prepared from single nuclei using the Chromium Controller and Single Cell 3'Reagent Kits v3.1 (10x Genomics, Inc. Pleasanton, USA) according to the manufacturer's protocol (document CG000315). Briefly, nanoliter-scale Gel Beads-in-emulsion (GEMs) were generated by combining barcoded Gel Beads, a master mix containing nuclei, and partitioning oil onto a Chromium chip. Nuclei were delivered at a limiting dilution, such that the majority (90-99%) of generated GEMs contained no nuclei, while the remainder largely contained a single nucleus. The Gel Beads were then dissolved, primers released, and any co-partitioned nuclei lysed. Primers containing an Illumina TruSeq Read 1 sequencing primer, a 16-nucleotide 10x Barcode, a 12-nucleotide unique molecular identifier (UMI) and a 30-nucleotide poly(dT) sequence were then mixed with the nuclear lysate and a master mix containing reverse transcription (RT) reagents. Incubation of the GEMs then yielded barcoded cDNA from poly-adenylated mRNA. Following incubation, GEMs were broken and pooled fractions recovered. First-

strand cDNA was then purified from the post GEM-RT reaction mixture using silane magnetic beads and amplified via PCR to generate sufficient mass for library construction. Enzymatic fragmentation and size selection were then used to optimize the cDNA amplicon size. Illumina P5 & P7 sequences, i7 and i5 sample indexes, and TruSeq Read 2 sequence were added via end repair, A-tailing, adaptor ligation, and PCR to yield final Illumina-compatible sequencing libraries.

Sequencing

The resulting sequencing libraries comprised standard Illumina paired-end constructs flanked with P5 and P7 sequences. The 16 bp 10x Barcode and 12 bp UMI were encoded in Read 1, while Read 2 was used to sequence the cDNA fragment. i7 and i5 sample indexes were incorporated as the sample index reads. Paired-end sequencing (28:90) was performed on the Illumina NextSeq500 platform using NextSeq 500/550 High Output v2.5 (150 Cycles) reagents. The .bcl sequence data were processed for QC purposes using bcl2fastq software (v. 2.20.0.422) and the resulting .fastq files assessed using FastQC (v. 0.11.3), FastqScreen (v. 0.14.0) and FastqStrand (v. 0.0.7) prior to pre-processing with the CellRanger pipeline.

Data analysis

Data pre-processing

Sequence files generated by the sequencer were processed using the 10x Genomics Cell Ranger pipeline. Developmental samples were analyzed with Cell Ranger v3.1.0, while adult samples used v6.1.2. FASTQ files were generated and aligned to a custom hg38 reference genome using default parameters. For developmental data, a custom premRNA_patched gtf file wsa used to include intronic reads, whereas for adult data the 'include-introns' option was enabled during the cellranger count runs. The pipeline identified cell barcodes corresponding to individual nuclei and quantified unique molecular identifiers (UMIs) per nucleus. Read alignment was performed using the STAR aligner, with multimapping reads excluded from UMI counting.

Filtering

Low-quality nuclei were removed from the dataset using three commonly used parameter for cell quality evaluation, the number of UMIs per cell barcode (library size), the number of genes per cell barcode and the proportion of UMIs that are mapped to mitochondrial genes. The threshold of these three parameters were adjusted individually for each sample to retain only good quality nuclei from each of the samples. Outlier nuclei with a total read counts > 50,000 were removed as potential doublets. After filtering, the number of retained nuclei/total nuclei were as follows (all replicates are biological replicates): CS16-17_rep1 4703/7011; CS16-17_rep2 2576/7761; 12W_rep1, 10,151/12,110; 12W_rep2, 7306/16,110; AV1 2765/3008; AV2 446/609; AV3 2219/2362. To achieve a more fine-grained information for each dataset, each individual sample was first analyzed separately. Developmental time-points (embryonic and fetal) were then merged, and adult time-points were combined independently. Finally, all time points (developmental and adult) were integrated using scanpy [26], yielding a total of 30,166 nuclei for downstream analysis.

Normalization and classification of cell-cycle phase

For individual sample analysis, raw gene expression counts were normalized using a deconvolution-based method [27]. For integrated analysis in Scanpy, the default normalization approach was applied. Cell cycle phase scores (G1 and G2/M) for each nucleus were calculated using the cyclone method [28].

Visualization & Clustering

The variance of each gene expression values was decomposed into technical and biological components; Highly Variable Genes (HVGs) were identified as genes for which biological components were significantly greater than zero. HVG genes were used to reduce the dimensions of the dataset using PCA; the dimension of dataset was further reduced to 2D using t-SNE and UMAP, where 1 to 14 components of the PCA were given as input. Nuclei were grouped into seven

clusters using the dynamic tree cut method [29]. For all datasets combined, scanpy's highly_variable_genes (scanpy v. 1.9.5) were identified. Nuclei in scanpy aggregated data were clustered using Louvain clustering in scanpy with a resolution of 0.4.

Identification of marker genes

For individual sample analysis, marker genes for each cluster were identified using the findMarkers function from scran 1.26.2 package [30]. The function also returns FDR values for multiple testing correction using Benjamini-Hochberg procedure. Marker genes were then used to annotate the cell types of a cluster. For the scanpy aggregated data, scanpy's rank_genes_groups was used to rank genes in each cluster using Wilcoxon rank-sum test method, which compares each cluster to the union of the rest of the clusters. This function also uses Benjamini-Hochberg procedure for multiple test correction. Gene ontology enrichment was evaluated for the top 100 genes using DAVID [31, 32], with P values limited to < 0.05.

RNA-velocity

scVelo (v. 0.2.4) was applied to identify transient cellular dynamics in the embryonic stage [33]. The spliced vs un-spliced ratio for each gene was calculated using the RNA-velocity's command line tool, velocyto 10x with default parameters [34].

SCENIC

The SCENIC pipeline was applied to infer gene regulatory network in our developmental samples. In the first step, GRNBoost2 created a gene co-expression module potentially regulated by the same TF, called regulon, followed by cisTarget to identify TFs that directly regulate each co-expression module, based on motif enrichments. Finally AUCell was used to calculate regulon activity scores in each cell. Output from AUCell was then used for downstream analysis using pySCENIC. Regulon Specificity Score (RSS) was used to quantify the activity of each regulon and identify cluster-specific regulons.

Spatial Transcriptomics

Spaceranger v.1.3.1 was used to pre-process the Visium slides. After filtering the lower count spots and genes, data were normalized using scanpy's normalize_total function. Spots were clustered using leiden clustering, and spatially variable genes were identified using Moran's notes [35]. As each of the Visium spots has more than one cell cell2location (v. 0.1.3) was used to deconvolute the spots [36]. For cell2location genes were filtered using filter_genes() function and setting the parameter, cell_count_cutoff = 5, cell_percentage_cutoff2 = 0.03 and nonz_mean_cutoff = 1.12. The model was then trained using reference cell type signatures from snRNA-seq data, estimated using a negative binomial (NB) regression model. The Cell2location() function was then used to map the reference data to spatial spots while setting the N_cells_per_location parameter to 12 and detection_alpha to 20.

Identification of gene expression patterns in spatial transcriptomics

To identify genes with expression patterns similar to a gene of interest, all spatial spots with expression values greater than 1 for the target gene were first selected. Genes not expressed in at least 50% of these spots were excluded. From the remaining set, any gene expressed in more than 25% of the spots where the target gene had zero expression was also removed. This filtering process yielded a list of genes with expression profiles closely matching that of the gene of interest.

Lineage tracing algorithm

A transcriptional signature-based lineage inference method was developed to infer developmental relationships between embryonic populations and their fetal and adult counterparts. For each selected embryonic progenitor population (cluster), we defined a lineage signature by identifying the top 100 differentially expressed genes relative to other embryonic mesenchymal populations. In parallel, for each fetal or adult cell cluster, we identified the top 5000 differentially expressed genes by comparing expression within each cluster to all other clusters at the same (fetal or adult)

stage. Lineage relationships were inferred by quantifying the overlap between each embryonic progenitor signature and the corresponding 5000-gene sets from later-stage clusters. Statistical significance of the observed overlaps was assessed using a hypergeometric test, enabling identification of fetal and adult populations significantly enriched for embryonic transcriptional signatures. Clusters exhibiting the strongest and most significant overlaps were interpreted as probable descendants of the corresponding embryonic progenitor populations.

All the experiments described (single nuclei RNA-seq and spatial transcriptomics) have been deposited in ArrayExpress; accession numbers are detailed in Table S1.

Results

Cellular landscape of the human OFT and its adult derivatives

To identify the cell types present in the human OFT, we conducted single nuclei (sn) RNA-seq on human OFT tissues. We isolated nuclei from 4 embryonic OFTs (CS 16-17), which were pooled into two independent pools (each from two embryos), two fetal (post conception week 12 (12pcw)) OFTs and from three adult aortic valves (Fig 1A; Table S1). The nuclei exhibit strong consistency between biological replicate samples (Fig 1B). After quality control and filtering, a total of 30,166 nuclei were segregated into 18 different clusters by unsupervised clustering (Fig 1C). These clusters were visualized by uniform manifold approximation and projection (UMAP). Each color on the chart represents a distinct cell population, arranged in order from the largest (5127 cells in cluster 0) to the smallest (355 cells in cluster 17). We decided against using batch-correction [37] to preserve the biological variability in our datasets, which reflects real changes across developmental time points (Fig S1AB). We observed the highest number of clusters (n = 12) in the fetal samples (Fig. 1D), consistent with new cell types arising from embryonic to fetal stage. The low number of clusters in the adult samples reflects sampling of the aortic valves, one of the derivatives of the entire OFT. Subsequently, we performed differential gene expression analysis to aid in the classification of each cell cluster (Table S2). Leveraging recognized markers, we were able to assign the clusters in developmental samples to three main compartments, cardiac, endothelial and mesenchymal cells (Fig 1EF). We also detected minor compartments of neuronal and immune cell types. In addition to mesenchymal (valve interstitial) and endothelial cell types, we allocated adult clusters to immune cells. We confirmed cell types in the main compartments using additional verified markers. The cardiac and endothelial nuclei populations appear to already express their appropriate cell type markers in the embryonic samples (Fig S1C). In contrast, most of the mesenchymal nuclei do not express markers of differentiated cell types (e.g. *DCN* or *MYH11*) at the earlier (embryonic) stage, but express a combination of *PDGFRA* [38] and *PDGFRB* [39], implying the embryonic stage captured mesenchymal nuclei before differentiation (Fig S1C). In addition, while embryonic and fetal cells display considerable variations in gene expression, adult cells share distinguishing features that set them apart from developing cells. This includes the expression of *NEAT1* [40], a long non-coding RNA (lncRNA), which does not distinguish cell types but discriminates adult from developmental tissues (Fig S1D-G).

Identification of a GATA6-driven mesenchymal program underlying semilunar valve development

Given their critical role in OFT development, we focused our analysis on mesenchymal cells. To explore the heterogeneity within this lineage and uncover potential subtypes, we performed nuclei sub-clustering of embryonic and fetal datasets, followed by tSNE visualization (Fig S2AB). Using this approach, we obtained seven distinct mesenchymal clusters, two embryonic and five fetal (Fig 2AB). Fetal nuclei predominantly express *DCN*, a fibroblast marker, except for cluster 9, which contains *MYH11*-positive smooth muscle cells (Fig 2C, see also Fig S1C). In contrast, embryonic mesenchymal nuclei lacked definitive lineage markers, suggesting these cells are undifferentiated or at an early progenitor stage.

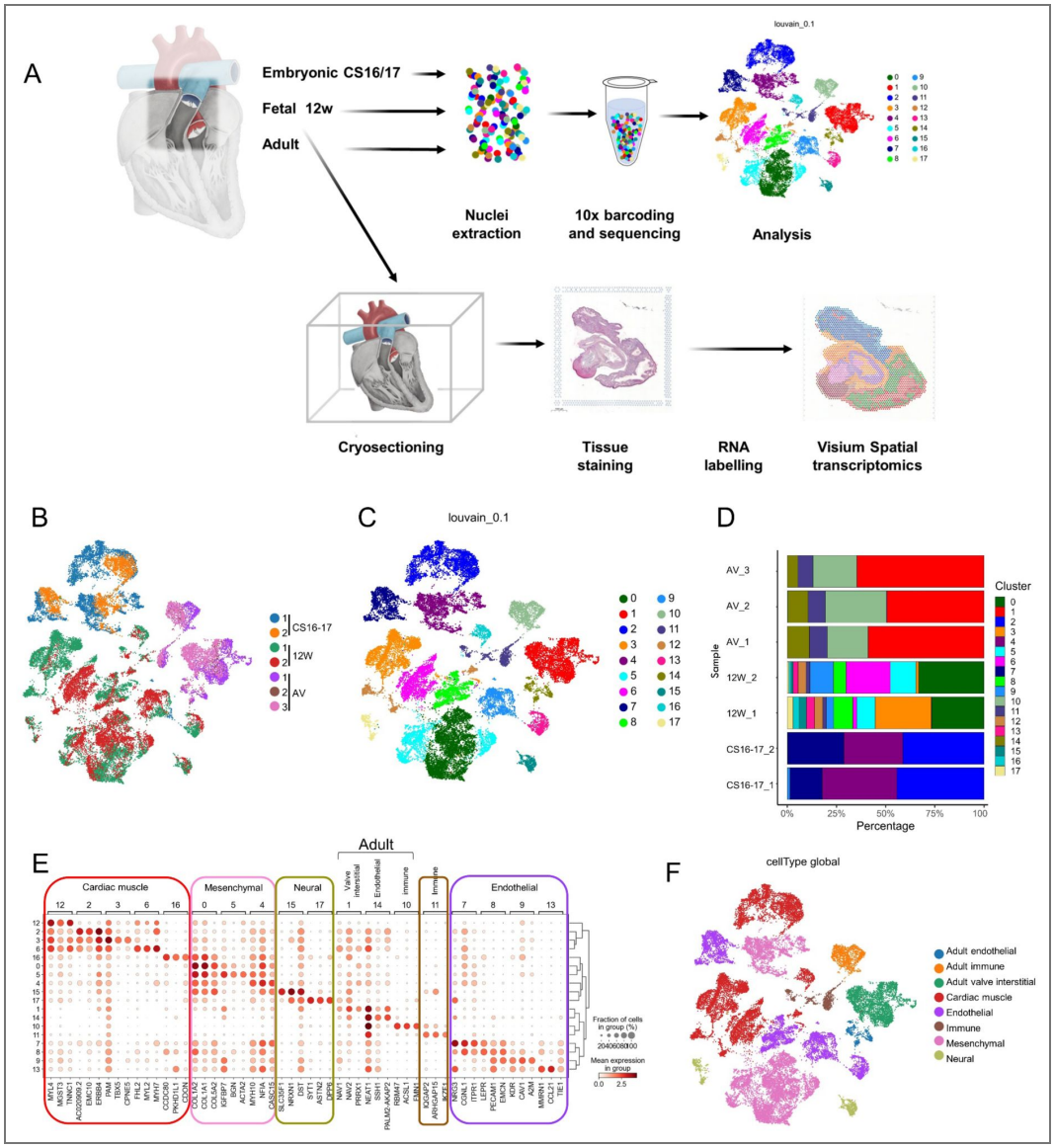

Figure 1. The cellular landscape of the developing OFT and its adult derivatives.

A. Experimental schematics. Nuclei isolated from two embryonic (CS 16-17) and two fetal (12pcw) OFTs and from three adult aortic valves (AV) were analyzed by snRNA-seq. Four cryo-sections including the OFT region of a 12pcw heart were used in spatial transcriptomics (Visium). B. Sample correlation visualized by unsupervised clustering and projected on a two-dimensional UMAP. Nuclei are colored by sample, with embryonic (blue, orange), fetal (green, red) and adult (pink, purple and brown). C. Cell clusters visualized by the same UMAP as in B. Nuclei are colored by cluster. D. Cluster composition in each sample, presented as percentage of nuclei. E. Dotplot shows the mean expression levels of top differential genes across clusters and identifies five main cell types, cardiac, endothelial, mesenchymal (including valve interstitial), neural and immune. F. Cell types in E visualised by UMAP. Nuclei are colored by cell type. See also Fig S1.

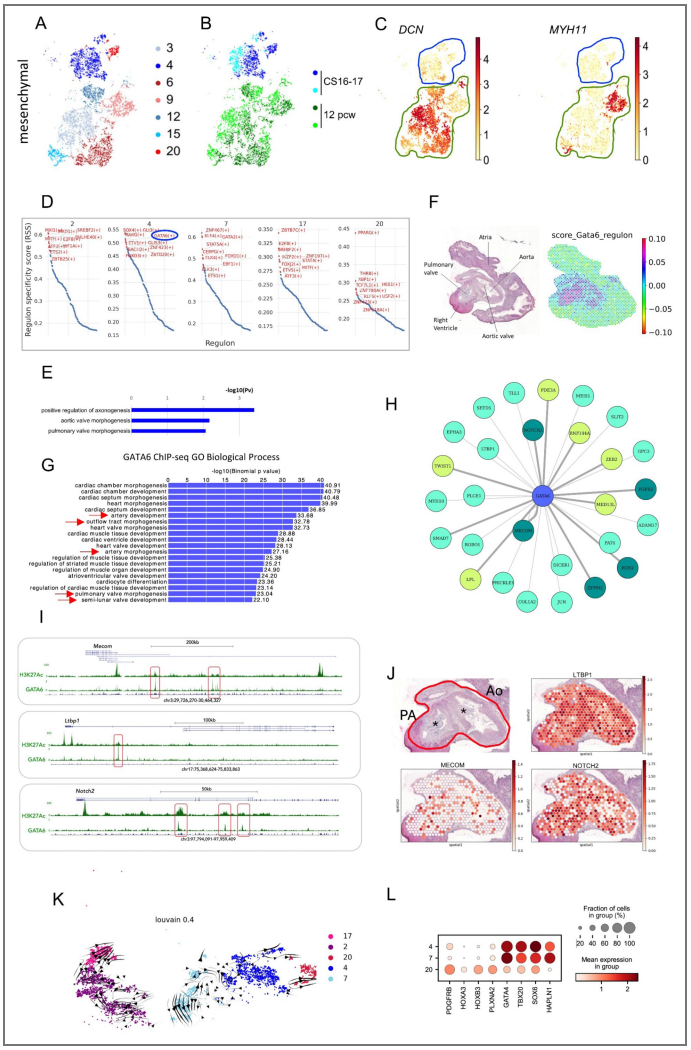

Figure 2. Characterization of embryonic mesenchymal nuclei

Mesenchymal cell clusters (A) and sample projection (B) of fetal and embryonic samples following subclustering, visualized on a two-dimensional tSNE. Nuclei are colored by cluster (A) and sample (B). C. Embryonic clusters (blue contour) do not express fibroblast (*DCN*) or smooth muscle (*MYH11*) markers, which are present in fetal nuclei (green contour). Nuclei are colored according to their scaled expression. D. Top 10 regulons in fetal clusters based on RSS score. E. Gene ontologies associated with the GATA6 regulon highlighted terms related to arterial and pulmonary valve morphogenesis. Functional annotation clustering of top 400 genes enriched GATA6 regulon was performed using DAVID and -Log10(Pv) was plotted in Excel. F. H&E-stained section showing the aortic and pulmonary arteries with their respective valves (left) and a corresponding map of the spatial expression of the GATA6 regulon (right). G. GREAT analysis of GATA6 high-confidence peaks (FE>10) in posterior pharyngeal arches and OFT at embryonic day (E) 11.5 (mouse). GATA6 peaks predominantly cluster around genes associated with cardiovascular terms, and specifically with OFT and artery development, as well as semilunar valve development (red arrows). H. Selected regulon genes associated with GATA6 binding in mouse embryo OFT and pharyngeal arches (see also Table S3) and associated with OFT-related abnormalities. Yellow genes are associated with human disease; light green genes cause mouse phenotypes; dark green genes are associated with both human and mouse defects. I. UCSC tracks of H3K27Ac ChIP-seq, GATA6 ChIP-seq (boxed in red) in posterior pharyngeal arches and OFT at E11.5 (mouse) and mammal sequence conservation at *MECOM* (top), *LTBP1* (middle) and *NOTCH2* (bottom) loci. J. Spatial Transcriptomics of Aorta (Ao) and Pulmonary Artery (PA) (clockwise): Hematoxylin and eosin (H&E) staining of the tissue area, with asterisks marking the semilunar valves; spatial distribution of *LTBP1, NOTCH2, MECOM*. K. Trajectory inference of future state of embryonic nuclei (CS16-17) showing mesenchymal (4, 20), endothelial-like (7) and cardiac (2, 17) clusters. Embryonic clusters derive from sub-clustering of aggregated fetal and embryonic nuclei shown in Fig S2A. L. Expression signatures in embryonic mesenchymal (4, 20) and endothelial-like (7) clusters. Both cluster 7 and 4 express high levels of cardiac TFs (*GATA4, TBX20*) and *HAPLN1*, a marker of semilunar valves. In contrast cluster 20 nuclei exhibit higher expression of neural crest markers, *HOXA3-B3* and *PLXNA2*.

Due to this absence of differentiated markers, which makes it challenging to define the identity of embryonic mesenchymal populations, we applied SCENIC [41] to the embryonic single-nucleus transcriptomes to infer key cell fate regulators and gain insight into their developmental potential. SCENIC links transcription factors (TFs) with their target genes based on co-expression, and it identified GATA6 as a top regulator of embryonic mesenchymal cluster 4 (Fig 2D). GATA6 is implicated in common arterial trunk (CAT) and bicuspid aortic valve (BAV) in humans [42, 43]. Pathway enrichment analysis of the 400 genes in the GATA6 regulon highlighted GO terms related to arterial and pulmonary valve morphogenesis (Fig 2E), a process known to be regulated by GATA6, supporting the idea that this regulon contains functional downstream targets involved in valve formation. Consistent with this, spatial transcriptomic analysis of a later stage (12pcw) OFT shows that GATA6 regulon is mainly restricted to the aortic and pulmonary valves (Fig 2F). To prune TF-target interactions and identify GATA6 high-confidence direct targets, we used GATA6 genomic occupancy in the mouse OFT and posterior pharyngeal arches [44]. Notably, GATA6 peaks are linked to GO terms highly specific to OFT and valve development [45], emphasizing that this dataset captures GATA6 binding to targets essential for the formation and maturation of these structures (Fig 2G). We inferred direct regulation by assigning peaks to genes [45]. The GATA6 regulon is significantly enriched for genes occupied by GATA6 ($p = 1.2 \times 10^{-33}$). This supports the interpretation that many genes within GATA6 regulon are associated with GATA6 binding, and likely to be direct GATA6 targets in the OFT. We systematically analysed direct targets for their implication in mouse phenotypes, human defects and genome-wide association studies (GWAS) using the MGI Database [46] and the EMBL GWAS database [47]. The results are summarized in the network shown in Fig 2H. The GATA6 regulon includes genes identified in GWAS studies on aortic valve calcification (*RNF144A*, *ZEB2*, *MECOM*, *LPL*, *PDE3A*, *TWIST1*) [48] as well as genes whose inactivation in mice leads to phenotypes that overlap with GATA6 loss, including aortic valve and OFT defects (Table S3).

We detect strong GATA6 peaks overlapping with H3K27Ac, a histone mark associated with active enhancers, within the *Mecom* locus (Fig 2I). In mice, mutations in *Mecom*, which encodes a histone-lysine N-methyltransferase, result in CAT, interrupted aortic arch, and ventricular septal defects [49]. In humans, GWAS have linked *MECOM* to calcified aortic stenosis [48, 50]. *MECOM* transcripts are sparsely expressed, and primarily localized within the aorta and pulmonary artery, in regions occupied by the semilunar valves (Fig 2J). The *Notch2* locus also exhibits multiple GATA6 peaks in regions of high acetylation and evolutionary conservation across mammals (Fig 2I). Conditional deletion of *Notch2* in cardiac neural crest cells leads to hypoplastic aortas and pulmonary arteries due to reduced smooth muscle content [51]. *NOTCH2* is one of the causative genes in Alagille syndrome, a multisystem disorder involving hepatic and cardiac anomalies, most commonly peripheral pulmonary stenosis and tetralogy of Fallot [52]. Like *MECOM, NOTCH2* is primarily expressed in the aorta and pulmonary artery (Fig 2J). Strong GATA6 binding is also observed at the *Ltbp1* gene, which encodes an extracellular regulator of TGF-β signalling (Fig 2I). Loss of the long isoform of *Ltbp1* in mice results in CAT and interrupted aortic arch, due to defective cardiac neural crest cell function [53]. Unlike *MECOM* and *NOTCH2*, *LTBP1* is highly expressed in the walls of the aorta and pulmonary artery, including the OFT septum (Fig 2J). Additional genes in the GATA6-regulated network include *SLIT2* and *ROBO1,* components of the SLIT-ROBO signaling pathway implicated in BAV [54], and structural genes like *MYH10* and *COL1A2* (Fig S2C). Collectively, the convergence of OFT phenotypes, OFT-specific GATA6 binding and enhancer activity (H3K27Ac enrichment), and the expression patterns of regulon genes supports a GATA6-driven transcriptional network implicated in OFT development, particularly the semilunar valves. These findings delineate molecular mechanisms underlying GATA6 function in the developing valves and highlight candidate genes that may contribute to BAV susceptibility.

SCENIC analysis links mesenchymal cluster 4 to semilunar valve development. We next applied trajectory inference to examine its relationship to other embryonic clusters at the same stage. RNA velocity predicts the future state of individual cells by distinguishing between un-spliced and spliced mRNAs [34]. We detected the transition of some nuclei from cluster 7 (endothelial-like cells) specifically towards cluster 4 (mesenchymal cells) (Fig 2K). Cluster 7 exhibits a similar expression profile to cluster 4, but distinct from the other mesenchymal nuclei (cluster 20) (Fig

2L). This profile includes high expression of *HAPLN1*, encoding the extracellular matrix cross-linking protein Hyaluronan and Proteoglycan Link Protein 1, found in the endocardial lining and the developing semilunar valves in the OFT. Compared with fetal endothelial clusters, endothelial cluster 7 is enriched for genes associated with aortic and pulmonary valve morphogenesis and epithelial to mesenchymal transition (EMT) (Fig S2F; Table S4). In mice, semilunar valve formation begins with an EMT of endothelial cells in the endocardium, the cell layer adjacent to the myocardium[55, 56]. This suggests that cluster 7 likely contains an endocardial population, with RNA velocity tracking their transition into mesenchymal cells in cluster 4, thereby generating the precursors of valve interstitial cells.

To understand why GATA6 emerges as a top regulator specifically in cluster 4, we examined GATA6 expression across embryonic nuclei. Although *GATA6* is expressed in all embryonic clusters, its levels are highest in cluster 4 (Fig S2D), which may account for its restricted activity in this population. Alternatively, given that TFs typically act cooperatively, GATA6 may coregulate cluster 4 regulon in combination with additional factors.

To identify these additional factors, we compared the regulons of cluster 4 top transcriptional regulators (Fig 2D). As expected, since regulon genes are sampled from cluster 4 enriched transcripts, these regulators share many downstream targets with GATA6 (19-30% overlap). Notably, GLI3 shows substantially greater regulon overlap (56%), suggesting functional cooperation with GATA6. This is consistent with their reported cooperation in the developing mouse limb[57]. GLI3 is also enriched in cluster 4, further supporting the hypothesis that these TFs cooperate in the development and differentiation of this cell population (Fig S2E).

Spatial resolution of mesenchymal nuclei in the OFT

Mesenchymal cells build key structures in the OFT, specifically the separation of the aorta from the pulmonary artery (Fig 3A, C, E) and the semilunar valves at the base of the aorta and pulmonary artery (Fig 3B, D, F). At 12pcw, mesenchymal nuclei express markers of differentiated cell types (Fig 2C). To visualize the distribution of fetal mesenchymal cell populations within the OFT, we conducted spatial gene expression analysis. We generated four sections of a 12pcw OFT (equivalent stage to snRNA-seq), starting at the level of the aortic valves (a) and ending at the level of the pulmonary valves (d) (Fig. 3G). We used Cell2location [36] to transfer the labels from the transcriptomic data to spatial gene expression data and mapped different cell-types on each individual spatial location. Cardiac nuclei correctly mapped to the atria and the ventricle (Fig S3A). Mesenchymal cells largely distributed within and around the vessels, and also mapped to the valves of the pulmonary artery and the aorta (Fig S3B). Endothelial cells were primarily located in the aortic valves (Fig S3C). Immune cells were scattered around the tissue (Fig S3D), while neural cells were concentrated in a spot of the atria (Fig S3E). Next, we mapped the five fetal mesenchymal clusters to distinct structures in the OFT (Fig 3H) and used distinctive markers to confirm spatial assignments. Clusters 3 and 6 (Fig 3I-I^{I}) map to the arterial outer walls and the septum between the aorta and pulmonary arteries and largely colocalise with *DCN*-positive cells (Fig 3I^{II}). These two clusters largely overlap. Cluster 9 (Fig 3J) coincides with *MYH11* expression (Fig 2J^{I}): *MYH11* is a terminal marker of smooth muscle differentiation [58], which identifies the aortic smooth muscle layer. Cluster 12 (Fig 3K) concentrates to both aortic and pulmonary valves and co-localizes with *HAPLN1* (Fig 3K^{I}), a marker of the semilunar valves [59]. No molecular differences or distinguishing markers were identified between the aortic and pulmonary valves. Finally, the smallest cluster (15) is restricted to the cardiac ventricle and is *DCN*-negative, therefore was eliminated from our further analysis (Fig S3F). Thus, the five subtypes of mesenchymal nuclei, identified by snRNA-seq, largely correspond to spatially segregated cell populations of fibroblasts (cluster 3, 6), smooth muscle cells (cluster 9) and valvular interstitial cells (cluster 12). To facilitate accessibility to these data, we have created a cellxgene VIP app to visualize gene expression across four sections of the fetal heart, available at: *https://cellxgene.cziscience.com/collections/5d2077ea-7b49-45c8-b4cb-64790b698591*.

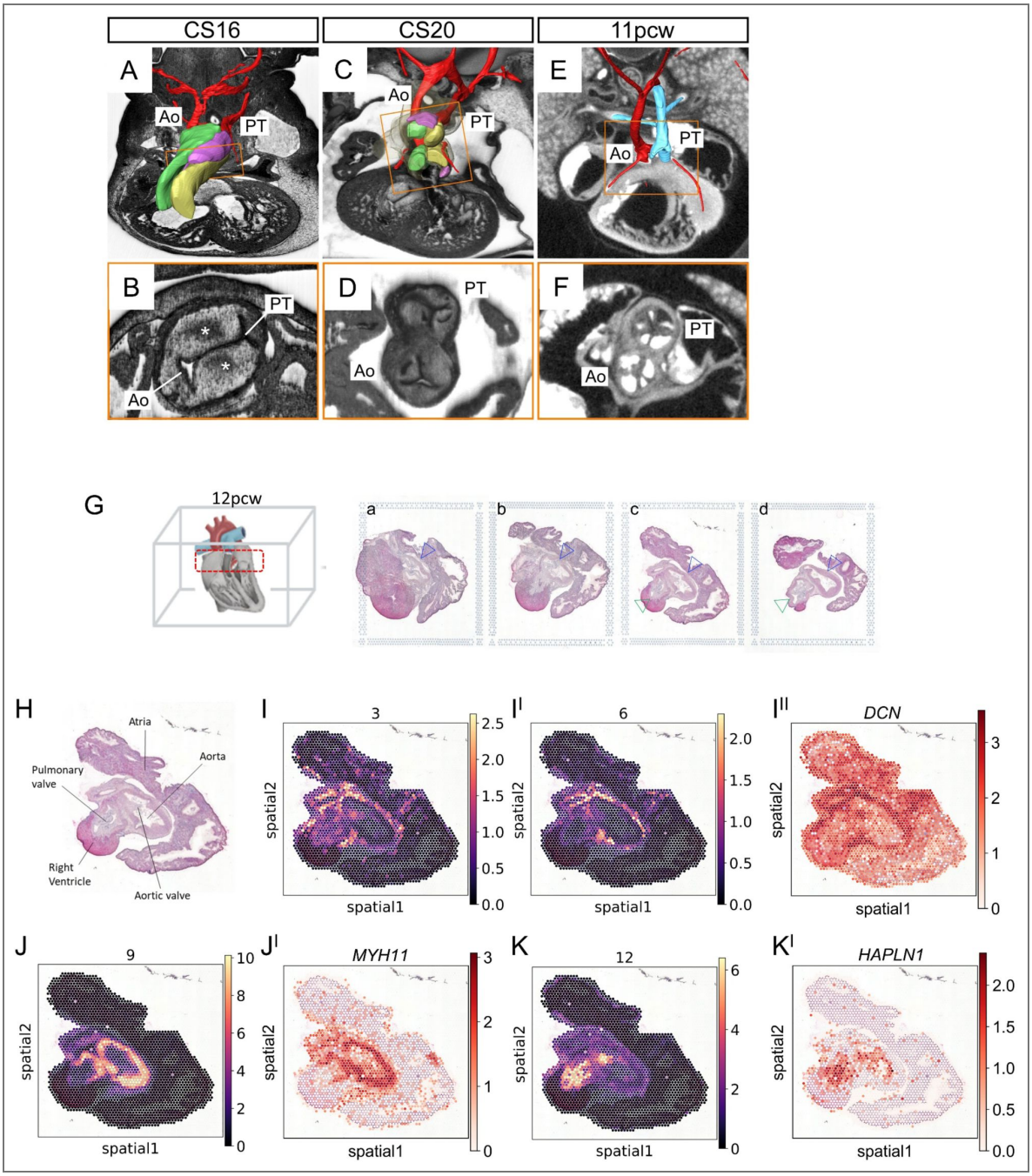

Figure 3. Spatial distribution of mesenchymal clusters

A-F. OFT valve formation and remodelling. Images were prepared using high resolution episcopic microscopy at embryonic stages (A-D) and by micro-CT at the fetal stage (EF). A. By CS16 the OFT has septated into the aorta (Ao) and pulmonary trunk (PT). The immature OFT cushions are visible: the septal (yellow), parietal (green) and intercalated (purple) cushions. B. Neural crest cells (asterisks) contribute to valve formation. CD. At CS20 the cushions have begun to remodel to form the three leaflets of the aortic and pulmonary semilunar valves. EF. At 11pcw the valves have transformed into the leaflets that control the unidirectional flow of blood from the heart. Boxed regions in A-C-E are shown at higher magnification in B-D-F. G. Heart alignment for sectioning, with the OFT region marked in red (left). H&E staining of OFT cryo-sections used for spatial transcriptomics from the base of the OFT (a) to the pulmonary valves (d). The aorta and pulmonary trunk are indicated by blue and green arrowheads, respectively. H. H&E section 'c' annotated to show major structures. I, I^{I}, J, K. Spatial distribution of mesenchymal clusters (purple and yellow). I^{II}, J^{I}, K^{I}. Spatial distribution of lineage specific markers (white and red). I-I^{II}. Clusters 3 (I) and 6 (I^{I}) largely overlap with fibroblast lineage marker *DCN* (I^{II}). J-J^{I}. Cluster 9 (J) and smooth muscle lineage specific marker *MYH11* (J^{I}) map to the aortic walls as well as to the pulmonary artery. K-K^{I} Cluster 12 (K) and valve specific marker *HAPLN1* (K^{I}) are mainly found in the valves at the base of the aorta and pulmonary artery. See also Fig S3.

Developmental trajectories of mesenchymal cells in the developing OFT

At CS16-17, we identified two clusters of undifferentiated mesenchymal cells. By 12pcw, three main types of differentiated mesenchymal populations had emerged: fibroblasts, smooth muscle cells, and valvular interstitial cells, each occupying distinct spatial locations within the OFT (Fig 2A-B). Our objective was to track the developmental trajectories of these populations and identify the 12pcw descendant cells from each embryonic cluster. Connecting mesenchymal embryonic progenitors to their differentiated fetal counterparts is challenging, because the embryonic nuclei are yet to express the molecular markers of differentiated lineage descendants (Fig 2C). Trajectory inference methods [60] failed to establish lineage relationships between embryonic and fetal populations. To overcome this, we used gene module scores. The rationale behind this approach was that any distinctive developmental signature present in the embryonic clusters would likely be retained in the fetal nuclei, thereby enabling us to trace the trajectories of mesenchymal cell populations.

We first performed pairwise differential gene expression analysis of embryonic clusters to identify distinct developmental signatures of mesenchymal subtypes. Cluster 4, which partly derives from endocardial cells, is enriched in cardiac-like markers (Fig 4A) and is linked to 'aortic valve and endocardial cushion morphogenesis' (Fig S4A). In contrast, cluster 20 is largely associated with neural-like GOs (Fig S4B) and enriched in neural crest markers *HOXA3/B3* and *PDGFRB* (Fig 4A). These observations suggest two separate embryonic origins for cluster 20 and cluster 4, neural crest and secondary heart field respectively. From the set of differentially expressed genes, we largely selected TFs, which define lineage identity, to construct two distinct gene modules. Specifically, the 'TBX20 SOX6 GATA4 PRRX1' module is highly expressed in embryonic cluster 4, while the 'MEIS1 JAG1 ROR1 PRDM6' module is enriched in embryonic cluster 20. These eight developmental genes were sufficient to segregate the embryonic nuclei into two distinct subtypes (Fig. 4B), which we designated as group 1 and group 2. Group 1 consists entirely of nuclei from cluster 4, while group 2 is primarily composed of nuclei from cluster 20, with a smaller contribution from cluster 4.

Next, we asked if embryonic gene modules are inherited by fetal nuclei. Indeed, we found that the expression of our embryonic gene modules largely segregates developmental nuclei into two distinct lineages (Fig 4C): a 'blue' lineage, composed mainly of clusters 4, 3, and 12, which exhibits higher expression of module 1 (SOX6, TBX20) and a 'red' lineage, including clusters 20-6-9, which shows higher levels of module 2 expression (*MEIS1*, *PRDM6*). We then combined embryonic gene modules with markers of differentiated cell types, obtaining three distinct groups of 12pcw mesenchymal nuclei (Fig 4D). Module 1 embryonic signature is inherited by fibroblasts and valvular cells, corresponding to cluster 3 and 12, respectively (Fig 4D), suggesting a common embryonic progenitor for these cell types. Since module 1 defines the secondary heart field–derived embryonic cluster 4, we conclude that fetal group 1 nuclei derive from the secondary heart field. Conversely, the module 2 signature was inherited by two distinct cell types: smooth muscle cells (cluster 6) and fibroblasts (cluster 9). These groups were further segregated by the expression of cell type-specific markers, *MYH11* (group 2) and *DCN* (group 3). Module 2 is primarily associated with cluster 20, which is enriched in neural crest markers, with a smaller contribution from cluster 4, suggesting that fetal group 2-3 nuclei derive from both the neural crest and secondary heart field. Consistent with this, expression of cardiac neural crest markers, *HOXA3/B3*, is almost exclusively confined to 'group 2-3' fetal nuclei, supporting the notion that these nuclei (mainly clusters 6 and 9) partially derive from neural crest progenitors. This is in line with observations in the mouse model [61] where smooth muscle cells in the aortic root originate from both neural crest and secondary heart field progenitors.

To confirm the lineage relationships, we developed a robust method to trace embryonic signatures in fetal cells. We expanded the gene module repertoires to include the top 100 most distinctive genes from the mesenchymal progenitors of clusters 4 and 20 (Table S5) and examined fetal nuclei populations that retained expression of these genes. When applied to the entire 12pcw dataset

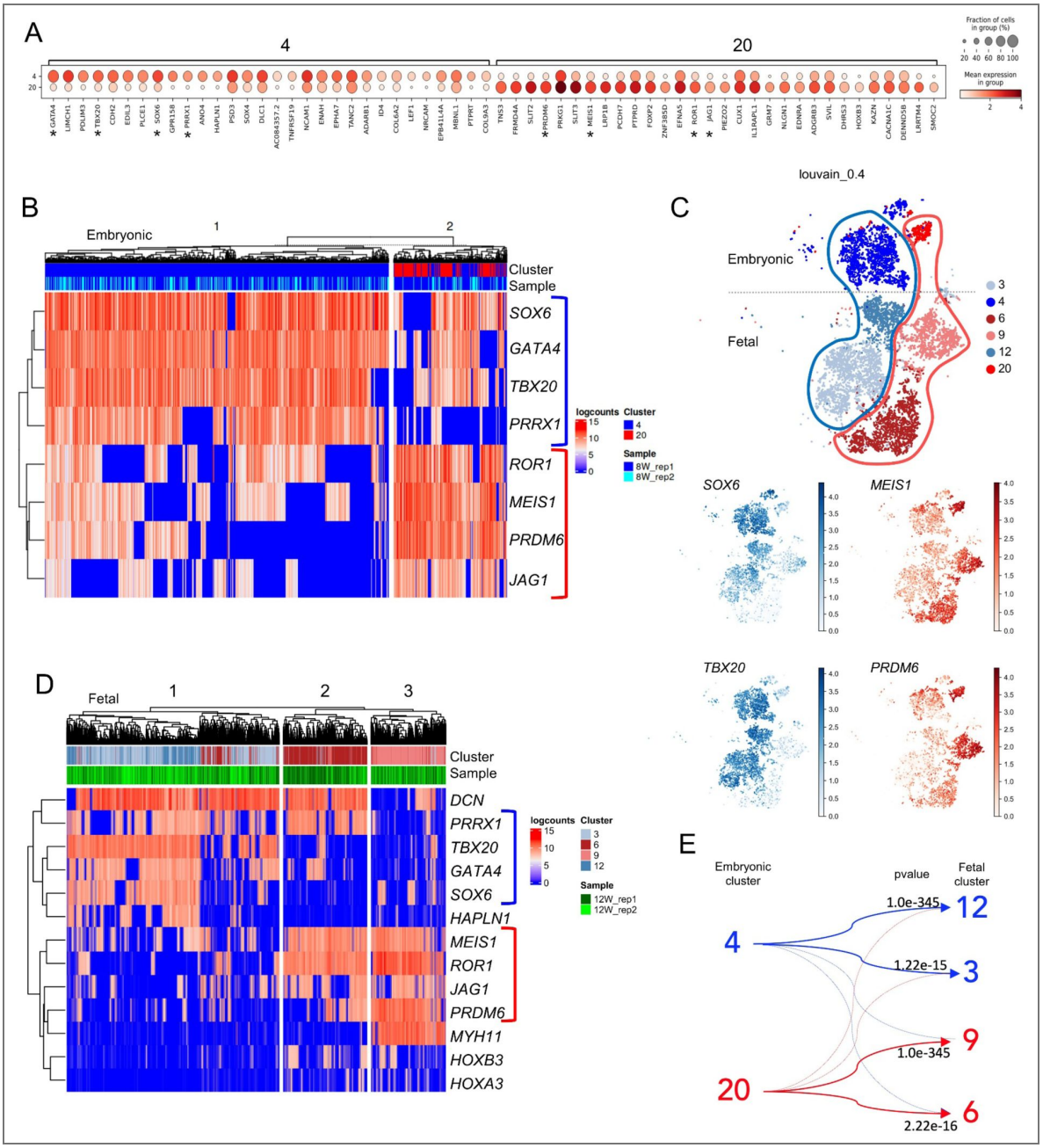

Figure 4. Lineage deconvolution of embryonic and fetal nuclei.

A. Pairwise differential gene expression of the two embryonic mesenchymal clusters; genes chosen for gene modules are marked by asterisks. B. Heatmap using embryonic gene modules, obtained using k-means clustering, separates embryonic mesenchymal nuclei in two groups. C. Mesenchymal cell clusters of embryonic and fetal time points, projected on a two-dimensional tSNE and labelled using gene modules. Fetal clusters derive from separate 'blue' and 'red' embryonic lineages. D. Heatmap of embryonic gene modules and cell type marker genes using k-means clustering identifies three main groups of fetal nuclei. E. Lineage trajectories of embryonic and fetal nuclei. Using the entire fetal datasets, cluster 3 and 12 nuclei are identified as descendent of embryonic cluster 4, while cluster 6 and 9 are the most likely descendant of embryonic cluster 20, consistent with the use of gene modules in the mesenchymal subset of fetal nuclei in 3D.

(including cardiac, endothelial and mesenchymal clusters), our method confirmed the same lineage relationship between embryonic and fetal mesenchymal nuclei that were previously identified using gene modules (Fig 4E). In sum, our analysis indicates that the two spatially distinct mesenchymal populations in the fetal OFT, smooth muscle cells and valvular fibroblasts (Fig 3J-J^{I}; K-K^{I}) derive from separate embryonic populations: cluster 20 for smooth muscle cells and cluster 4 for valvular fibroblasts. In contrast, the *DCN*-positive, *HAPLN*-negative fibroblast cells, which spatially intermingle in OFT tissues, derive from both cluster 4 and 20 embryonic progenitors (Fig S4C).

Cellular constituents of adult aortic valves

Semilunar (aortic and pulmonary) valves are the only distinctive and recognizable derivatives of the OFT that are retained in the adult. For adult samples, we focussed on the aortic valves because of their frequent association with disease, both genetic (bicuspid aortic valves) and adult (valve calcification) disease. We collected female samples to mitigate individual variability and maximise the possibility to analyse healthy aortic valves, justified by the lower incidence and severity of aortic disease in females versus males. Histologically normal aortic valves were procured from healthy adult hearts collected for transplantation and subsequently rejected. A total of 5,430 single nuclei from three human aortic valve samples (Fig 5A) were segregated into 11 different clusters by unsupervised clustering (Fig 5B). Leveraging established markers, we could separate the clusters into three major compartments, interstitial (7 clusters, largely *NAV2*-positive), endothelial (2 clusters, *CDH5*-positive) [62] and immune (2 clusters, *ITGAM*-positive macrophages and *IKFZ1*-positive dendritic cells) [63, 64] (Fig 5C).

We performed differential gene expression analysis to aid in the classification of each cell cluster. Heatmaps (Fig S5A) group together immune (2,6) and endothelial cell types (8,10). In addition to *CDH5* [65], endothelial clusters are distinguished by high levels of *ADAMTS9*, which encodes for a metalloproteinase implicated in aortic valve anomalies in mouse [66] (Fig 5D). *ADAMTS9* is also expressed in cluster 11, which is *CDH5*-negative (Fig 5D). Of the remaining clusters, cluster 1 is transcriptionally distinct and contains *CALD1*-, *ACT1*-positive SMC. AV3 is highly enriched in this cluster (26% vs 5.8% and 2.9% in AV1 and AV2 respectively) (Fig S5B); we attribute this skewed enrichment to aortic wall tissue being sampled with the aortic valves in AV1, as SMC are not a main constituent of the aortic valves, rather than to intrinsic variability across samples. Indeed, when cluster 1 is removed, the three samples are consistently similar to each other (Fig S5C). Cluster 11, which accounts for a small proportion of nuclei in all three samples (ranging from 1.5% to 2.3%) was the most related cluster to cluster 1 (Fig S5B; Fig 5E).

Clusters 3,4,5,7 display expression of similar transcripts, such as *NAV2*, *LTBP1*, *FN1* and were assigned to valvular interstitial cells (VIC) (Fig S5A; Fig 5D). VICs deposit three highly organized layers (fibrosa, spongiosa and ventricularis) of extracellular matrix (ECM), that compose the mature valve structure [67]. These cells are identified by the expression of transcripts encoding for ECM proteins, most notably *COL1A1*, *VIM*, *COL3A1*, *VCAN*, *BGN* and *LUM*. ECM encoding transcripts were enriched in clusters 3,4,5,7 (VIC) across all the adult samples examined (Fig S5D), with *VIM* and *VCAN* displaying a broader expression. Differently from cluster 5, clusters 4 and 7 express high levels of *CCL2*, an inflammatory cytokine [68], suggesting nuclei in these clusters correspond to activated fibroblasts. Finally, cluster 9 displays enrichment in *DLC1* and *TLN2*, implicated in cytoskeletal changes [69, 70], and *SPARC* (osteonectin), implicated in valve calcification [71]. While the above clusters are represented across all samples, AV1 clusters 4-7-9 contain on average more nuclei relative to AV2 and AV3 (Fig 5D; Fig S5E). The increased expression of *SPARC* (cluster 9), combined with higher levels of *CCL2* (cluster 4 and 7) in AV1, indicates the possibility of inflammatory processes that could eventually lead to valve calcification in AV1 sample. We did not detect myocardial cells in the adult valve tissue, consistent with evidence that myocardium contributes to early arterial root and cushion formation but does not persist in mature valves. Myocardial gene expression is already absent from the valve leaflet cluster by CS16-19[22]. Our adult dataset therefore reflects the valve complex and adjacent arterial root region, a subset of embryonic OFT derivatives rather than the entire OFT myocardium.

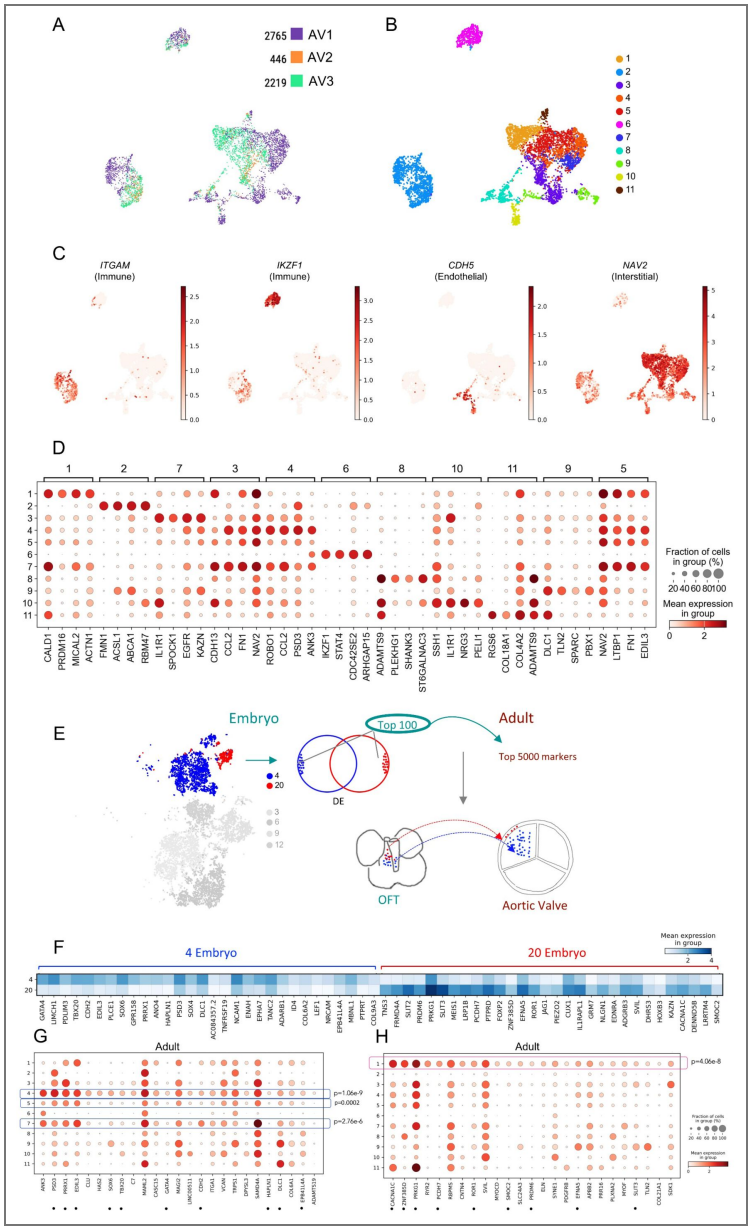

Figure 5. Cellular constituents of the mature aortic valves.

A. Aortic valve sample association projected on a two-dimensional UMAP. Nuclei are colored by sample. B. Nuclei clusters visualized by unsupervised clustering. Nuclei are colored by cluster. C. Cell lineages identified using established lineage-specific markers. Each nucleus is colored based on the scaled expression of the indicated marker. D. Top differentially expressed genes in clusters identify known lineage markers. E. Overview of the method used to trace adult descendent of embryonic nuclei. The first step is the identification of distinctive signatures in embryonic progenitors; for this, we used the top 100 differentially expressed (DE) genes in our chosen progenitor populations, cluster 4 (blue) and cluster 20 (red). The second step is the identification of the top 5000 marker genes for each adult population; this is done by comparing each cluster with the rest of the dataset. Finally, we search for the 100 DE embryonic genes in the marker genes of adult clusters. Adult clusters with top hits are identified as the descendent of the embryonic lineage; the statistical significance is calculated using a hypergeometric test. F. Dotplot displaying the 30 top DE genes (mean expression values) in embryonic cluster 4 and cluster 20, respectively. The same dot plot, previously shown in Fig 3C, has been included here to facilitate cross-comparison with Fig 6GH. G. Distribution of cluster 4 embryonic signature genes in adult nuclei clusters. Cluster 4, 7 and 5 express a highly significant fraction of embryonic cluster 4 genes. Top 30 DE genes in embryonic cluster 4 and cluster 20 (shown in F) are highlighted by dots. H. Distribution of cluster 20 embryonic signature genes in adult nuclei clusters. Cluster 1 expresses a highly significant fraction of embryonic cluster 20 genes. Top 30 DE genes in embryonic cluster 4 and cluster 20 (shown in F) are highlighted by dots.

Persistence of embryonic signatures in adult cells

Whilst differentiated, 12pcw nuclei maintain robust embryonic signatures (Fig 4DE). We asked if such signatures also persist in terminally differentiated adult cells and could be used as labels to retrospectively attribute cellular relationships between embryo and adult cells. To address this, we leveraged our method to identify embryonic signatures in adult cells. Using the top one hundred most distinctive genes of the embryonic mesenchymal progenitors (cluster 4 and 20 in the embryo) (Table S5), we looked for populations of adult nuclei enriched in the expression of most of these genes (Fig 5E). We found that adult clusters 1,4,7 retain expression of a highly significant fraction of distinctive embryonic genes, with clusters 4,7 largely maintaining embryonic cluster 4 expression signature (blue) (Fig 5FG), while cluster 1 displays a significant match with embryonic cluster 20 (red) (Fig 5FH). The finding that smooth muscle cells (adult cluster 1) are derived from embryonic cluster 20 is consistent with cluster 20 being the source of smooth muscle cells at 12pcw (Fig 4). Valvular fibroblasts (clusters 4,7) derive from the embryonic population in cluster 4, which is also linked to fetal valvular fibroblasts (Fig 4). This result is significant because our embryonic signatures derive from undifferentiated cells which are yet to express obvious differentiation markers, suggesting that adult cells retain their ancestral embryonic make up. The expression of a representative gene set from the 100-gene embryonic signatures was projected onto adult valve cells, confirming the findings shown in Fig 5F-H (Fig S7AB). As our adult clusters reflect aggregation of three individual samples, we asked if distinctive embryonic signatures can be detected despite confounding individual factors (ageing, environmental exposure, genetic background, etc). We performed the same analysis on individual samples AV1 and AV3 (AV2 was excluded due to the lower number of nuclei). We independently re-clustered each sample and applied our method to identify the descendent nuclei of embryonic cluster 4 within individual samples. Our method linked embryonic cluster 4 to cluster 0 in AV1, and to cluster 2 and 3 nuclei in AV3 (Fig S6AB); these clusters contain the majority of adult cluster 4 aggregate nuclei (Fig S6CD). This indicates that distinctive embryonic signatures can be detected over individual variability. In sum, our analysis indicates that distinctive patterns of embryonic gene expression persist in adult cells, can be consistently detected in individual adult samples, and can be used as labels to retrospectively attribute cellular relationships between embryo and adult cells. Finally, as the level of expression of a gene across different cells can provide an initial indication of its functional role, we explored the relative expression levels of distinctive embryonic signatures in embryonic nuclei and their adult descendent nuclei. Generally embryonic genes display lower expression levels in adult cells relative to their embryonic progenitors (Fig S6E-F).

Spatial profiling of OFT defect genes reveals candidates for congenital heart malformations

OFT defects occur when the vessels leaving the heart do not form or remodel correctly, resulting in problems with blood circulation and/or oxygenation in the neonate [72]. The most severe include common arterial trunk (CAT), transposition of the great arteries (TGA), double-outlet right ventricle (DORV) and Tetralogy of Fallot (TOF). Using spatial transcriptomics, we investigated the distribution of genes whose mutations cause OFT defects (Fig 6A-A^{I}). *JAG1* mutations cause TOF [73]; *JAG1* transcripts largely concentrate around the aorta, in the semilunar valves and the septum (Fig 6B-B^{I}). Mutations in *GATA* family members affects OFT formation in different ways, leading to DORV (*GATA5*) [74], TOF (*GATA4, GATA6*) [75, 76] and CAT (*GATA6*) [43]. Different from *GATA5*, which shows a restricted expression to the vessels (Fig 6C, C^{I}), the distribution of *GATA6* and *GATA4* transcripts is broader. *GATA6* is detected in the vessels, the septum and cardiac tissue (Fig S8A), while *GATA4* transcripts are more prominent in cardiac tissues (Fig S8B). Mutations in *GATA4-5-6* also causes BAV, a less severe condition characterized by two valves instead of three [77]. Similar to *GATA5*, *GATA6* is highly expressed in the valves, marked by *HAPLN1* (Fig 3K^{I}). As a final example, mutations in *NR2F2* cause DORV and TOF [78]. Differently from the previous examples, *NR2F2* transcripts are largely excluded from OFT structures and are mainly located in cardiac tissues surrounding the aorta and the pulmonary artery (Fig 6D, D^{I}).

We reasoned that genes exhibiting a distribution pattern similar to, or matching those mutated in OFT defects, could serve as potential candidates for OFT anomalies. To test this hypothesis, we focused on *JAG1*, whose transcripts exhibit a distinct pattern around the vessels, and generated an algorithm to identify genes that display comparable expression patterns to *JAG1* (Fig 6E). The algorithm filters out any gene that is not expressed in the majority of *JAG1*-positive spots (>50%) and is also expressed in *JAG1*-negative spots (see methods). Using this approach, we identified *FOXC1* and *OSR1*, whose loss of function in mouse results in semilunar valve abnormalities and aortic arch coarctation, and defects in heart septation, respectively [79, 80]. In addition, variants in *FOXC1* were recently identified in patients with conotruncal heart defects [10].

Discussion

We used snRNA-seq to analyze 30,166 nuclei, which covered two stages of OFT development (CS16-17 and 12pcw) up to the adult derivatives of the OFT, the aortic valves. In parallel, we used spatial transcriptomics to define the distribution of fetal cell populations and visualize the expression patterns of genes implicated in OFT defects, which make up one third of all cases of CHD. These datasets provide a large reference framework of OFT cell repertoires and their gene expression profiles, and constitute a valuable resource for enhancing our understanding of OFT defects and advancing the discovery of previously unknown genes implicated in these conditions. A major finding emerging from our analysis of a timeline of human development is that adult cells retain their ancestral embryonic signatures, which may have implications for adult-onset aortic valve disease.

During human embryo development, the OFT undergoes formation and remodelling between CS13 and CS23. This process results in the separation of the aorta and pulmonary trunks, creating the double circulation found in mammals. It also leads to the formation of the arterial valves, which continue to mature after the arterial vessels have separated. Focussing primarily on mesenchymal cells, which are responsible for building the semilunar valves and the separation of the aorta from the pulmonary artery, we identify two distinct groups of embryonic progenitors at the earliest stage (CS16/17). These mesenchymal nuclei do not yet express clear differentiation markers, but display distinctive neural crest and secondary heart field expression signatures. In contrast, by 12pcw mesenchymal nuclei express cell type markers and cluster into four subgroups, corresponding to fibroblast-like cells, smooth muscle cells, valvular interstitial cells, each localized to distinct regions within the OFT. Using gene modules and a new lineage tracing tool, we traced these four fetal mesenchymal populations back to their embryonic precursors. Our results reveal that valvular fibroblasts and smooth muscle cells have separate embryonic origins. Valvular fibroblasts, the mesenchymal cells involved in constructing the semilunar valves, appear to originate from endothelial cells. This aligns with the finding that mouse arterial valves derive from endothelial cells undergoing EMT in the endocardium - the myocardium-adjacent cell layer

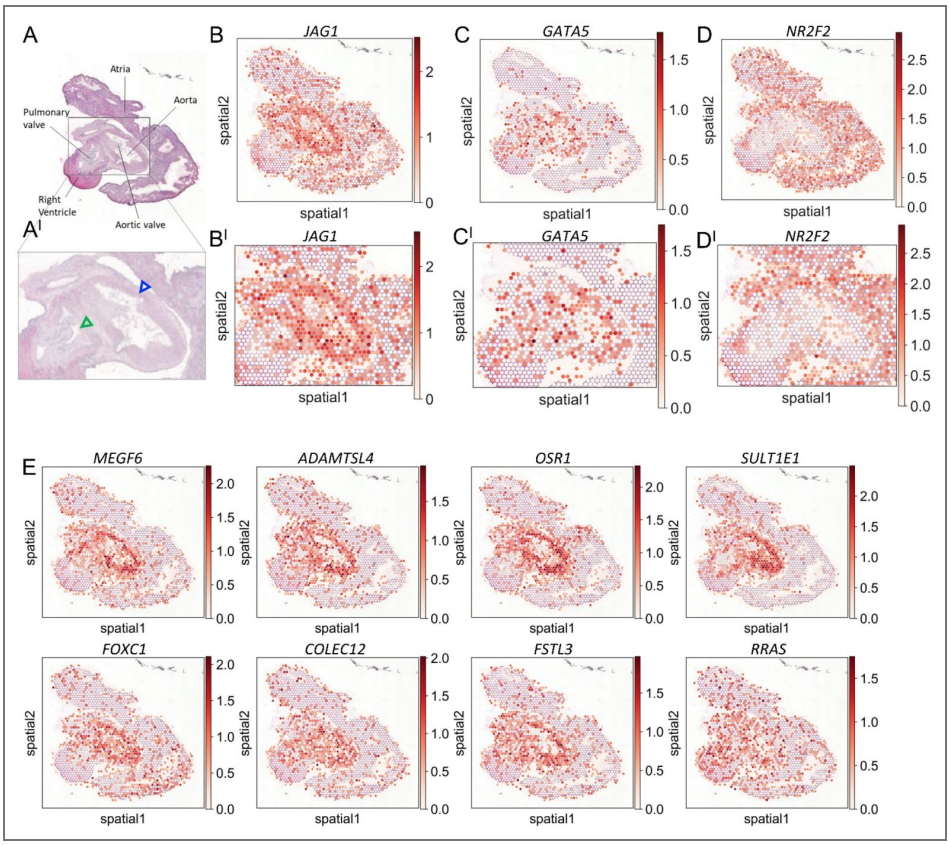

Figure 6. **Genes mutated in congenital OFT defects.**

A. H&E staining of spatial transcriptomics section (Fig 2Jc), and magnified view of aortic and pulmonary valve area (A^I). The aorta and pulmonary trunk are indicated by blue and green arrowheads, respectively. B-D. *JAG1* (B), *GATA5* (C) and *NR2F2* (D) gene expression patterns on the same section. B^I- DI, Genes as in BD with corresponding magnification of valve area (B^I, C^I, D^I). E. Genes identified as displaying spatially similar expression pattern to *JAG1*.

[55], [56]. The mesenchymal progenitors of fetal valvular fibroblasts and adult VICs are regulated by a GATA6-driven gene network. Mutations in *GATA6* are a known cause of BAV, suggesting that this regulatory program plays a critical role in semilunar valve development. The GATA6 regulon includes genes whose mutations are associated with OFT and valve defects in mice, as well as aortic valve disease in humans. Although we did not uncover a novel single-gene marker specific to humans (analogous to LRG5[13]), our identification of a GATA6 network highlights molecular mechanisms downstream of GATA6 that drive valve formation, advancing our understanding of normal valve development and informing the search for genes involved in BAV susceptibility and aortic valve disease. In contrast, smooth muscle cells derive from a distinct CS16-17 embryonic population. A consistent set of these cells displays expression of cardiac neural crest markers, in agreement with previous observations in mice, indicating that smooth muscle cells in the human aortic root originate from the neural crest and secondary heart field [81]. Finally, both embryonic mesenchymal populations also give rise to non-valvular fibroblast cells (*DCN*$^{+}$, *HAPLN1*$^{-}$).

We find that distinctive embryonic signatures persist into adulthood. While previous studies have described that reactivation of fetal programs in adult heart disease [82], [83], all our adult samples are derived from healthy individuals. This suggests that the persistence of early developmental signatures is a pervasive feature of normal adult cells, not merely a pathological response. Single-cell analyses have revealed high levels of heterogeneity within the same cell type. Our findings suggest that the inheritance of 'embryonic memories' by adult cells may be a major contributing factor to this heterogeneity. In support of this, organ-specific developmental signatures have been observed in adult fibroblasts [84]. In the human context, the persistence of distinct embryonic signatures in adult cell types may influence their susceptibility to disease, and potentially contribute to a better understanding of disease heterogeneity.

Connecting cell lineages over extended periods of time can be challenging. Existing tools typically use single-cell RNA sequencing to capture the complete expression profiles of individual cells in a single experiment, treating each cell as a unique time point on a continuum [60]. However, the significant transcriptional changes that occur as cells transition from embryonic to fetal and adult cell types pose a challenge when comparing and linking cells over prolonged periods based on their global expression profiles. The persistence of embryonic signatures into adulthood opens the possibility of using these enduring molecular "labels" to trace developmental ancestry in complex tissues, particularly in contexts where related cells are temporally distant or phenotypically divergent, such as in aging or cancer.

The observation that embryonic expression signatures persist in adult cells raise obvious questions about functional significance. We find that embryonic genes are generally expressed at lower levels in adult cells relative to embryonic cells (where they are known to have a function); their persistent expression may reflect fortuitous remnants of developmental histories. Alternatively, these retained embryonic gene expressions could serve as preserved developmental blueprints, ready to be swiftly and efficiently reactivated when the need arises, such as during injury or tissue repair. Reinforcing this perspective, heart failure is associated with the reawakening of a fetal gene program [82], [83].

In summary, our work extends beyond confirming previously reported cell types by (i) defining a GATA6-regulated human valve progenitor lineage and its derivatives, (ii) establishing distinct embryonic origins for smooth muscle and valvular fibroblasts, and (iii) demonstrating the persistence of embryonic signatures in adult valve cell populations. These conclusions are directly supported in tissue by our spatial transcriptomics data, which map these lineages and regulatory programs to defined anatomical domains within the human OFT and semilunar valves.

Data availability

All data described have been deposited in ArrayExpress with accession numbers listed in table S1 and will also be made available upon acceptance of the manuscript through the Human Cell Atlas. The spatial transcriptomics data is accessible at *https://spatialtranscriptomics-uom.bmh.manchester.ac.uk/*. Instructions on how to use the app are available at the cellxgene

VIP website: *https://interactivereport.github.io/cellxgene_VIP/tutorial/docs/how-to-use-cellxgene-vip.html*. The code used in Fig. 5 will be made publicly available through GitHub upon publication.

Acknowledgements

We thank Andy Hayes and the other members of the Genomic Technologies Core Facility, Roger Meadows of the Bioimaging facility and Gareth Howell of the Flow Sorting Core Facility at the University of Manchester. We also thank Zoulfia Darieva, Peyman Zarrineh, Rachel Jennings and Aoibheann Mullan for help and discussions. A special thanks to Jasmin Turner for help with cryosectioning and histology. This work was supported by joint funding from the Medical Research Council (MRC) (*https://www.mrc.ukri.org*) grant MR/S03613X/1 and British Heart Foundation (BHF) grant SP/18/12/34300 to NB and SDB. The Human Developmental Biology Resource (*www.hdbr.org*) is funded jointly by the Medical Research Council and the Wellcome Trust (MR/R006237/1). Adult tissue acquisition was funded by the National Institute for Health and Care Research (NIHR) Blood and Transplant Research Unit in Organ Donation and Transplantation (NIHR203332), a partnership between NHS Blood and Transplant, University of Cambridge and Newcastle University.

Additional files

Table S1. Accession numbers of single cell experiments and spatial transcriptomics.

Table S2. Marker genes (top 100) for each of the 18 clusters (0-17) in Figure 1.

Table S3. GATA6 regulon genes associated with GATA6 binding in the OFT and pharyngeal arches in E11.5 mouse embryos. Association was established using GATA6 peaks with a fold enrichment cutoff > 10 and GREAT standard association rules.

Table S4. Genes enriched in cluster 7 (embryonic endothelial) relative to clusters 9-13 (fetal endothelial).

Table S5. Embryonic cluster 4 and Cluster 20 top 100 differentially expressed genes.

Supplentary Figures.

Additional information

Funding

Funder	Grant reference number	Author
UKRI | Medical Research Council (MRC)	MR/S03613X/1	Simon Bamforth Nicoletta Bobola
British Heart Foundation (BHF)	SP/18/12/34300	Nicoletta Bobola Simon Bamforth

References

1. **Hoffman J.I.,** Kaplan S (2002) The incidence of congenital heart disease. *J Am Coll Cardiol* **39**:1890-1900 https://doi.org/10.1016/s0735-1097(02)01886-7 | PubMed
2. **Stefanovic S.,** Etchevers H.C., Zaffran S (2021) Outflow Tract Formation-Embryonic Origins of Conotruncal Congenital Heart Disease. *J Cardiovasc Dev Dis* **8** https://doi.org/10.3390/jcdd8040042 | PubMed
3. **Thom T.,** et al. (2006) Heart disease and stroke statistics--2006 update: a report from the American Heart Association Statistics Committee and Stroke Statistics Subcommittee. *Circulation* **113**:e85-151 https://doi.org/10.1161/circulationaha.105.171600 | PubMed

4. **Anderson R.H.,** et al. (2012) Normal and abnormal development of the intrapericardial arterial trunks in humans and mice. *Cardiovasc Res* **95**:108-115 https://doi.org/10.1093/cvr/cvs147 | PubMed
5. **Anderson R.,** et al. (2012) The Reappraisal of Normal and Abnormal Cardiac Development. In: Kleinman C.S., Seri K. (Eds). *Hemodynamics and Cardiology Neonatology Questions and Controversies* Elsevier. pp. 391-414 https://doi.org/10.1016/b978-1-4377-2763-0.00019-6
6. **Kelly R.G** (2012) The second heart field. *Curr Top Dev Biol* **100**:33-65 https://doi.org/10.1016/b978-0-12-387786-4.00002-6 | PubMed
7. **Waldo K.,** Miyagawa-Tomita S., Kumiski D., Kirby M.L (1998) Cardiac neural crest cells provide new insight into septation of the cardiac outflow tract: aortic sac to ventricular septal closure. *Dev Biol* **196**:129-144 https://doi.org/10.1006/dbio.1998.8860 | PubMed
8. **Jiang X.,** Rowitch D.H., Soriano P., McMahon A.P., Sucov H.M (2000) Fate of the mammalian cardiac neural crest. *Development* **127**:1607-1616 https://doi.org/10.1242/dev.127.8.1607 | PubMed
9. **Kirby M.L.,** Hutson M.R (2010) Factors controlling cardiac neural crest cell migration. *Cell Adh Migr* **4**:609-621 https://doi.org/10.4161/cam.4.4.13489 | PubMed
10. **Wei W.,** et al. (2024) Variants in FOXC1 and FOXC2 identified in patients with conotruncal heart defects. *Genomics* **116**:110840 https://doi.org/10.1016/j.ygeno.2024.110840 | PubMed
11. **Xu Y.,** et al. (2023) A single-cell transcriptome atlas profiles early organogenesis in human embryos. *Nat Cell Biol* **25**:604-615 https://doi.org/10.1038/s41556-023-01108-w | PubMed
12. **Asp M.,** et al. (2019) A Spatiotemporal Organ-Wide Gene Expression and Cell Atlas of the Developing Human Heart. *Cell* **179**:1647-1660.e1619 https://doi.org/10.1016/j.cell.2019.11.025 | PubMed
13. **Sahara M.,** et al. (2019) Population and Single-Cell Analysis of Human Cardiogenesis Reveals Unique LGR5 Ventricular Progenitors in Embryonic Outflow Tract. *Developmental Cell* **48**:475-490.e477 https://doi.org/10.1016/j.devcel.2019.01.005 | PubMed
14. **Cui Y.,** et al. (2019) Single-Cell Transcriptome Analysis Maps the Developmental Track of the Human Heart. *Cell Rep* **26**:1934-1950.e1935 https://doi.org/10.1016/j.celrep.2019.01.079 | PubMed
15. **Knight-Schrijver V.R.,** et al. (2022) A single-cell comparison of adult and fetal human epicardium defines the age-associated changes in epicardial activity. *Nature Cardiovascular Research* **1**:1215-1229 https://doi.org/10.1038/s44161-022-00183-w | PubMed
16. **Cao J.,** et al. (2020) A human cell atlas of fetal gene expression. *Science* **370** https://doi.org/10.1126/science.aba7721 | PubMed
17. **Suryawanshi H.,** et al. (2020) Cell atlas of the foetal human heart and implications for autoimmune-mediated congenital heart block. *Cardiovasc Res* **116**:1446-1457 https://doi.org/10.1093/cvr/cvz257 | PubMed
18. **Lahm H.,** et al. (2021) Congenital heart disease risk loci identified by genome-wide association study in European patients. *J Clin Invest* **131** https://doi.org/10.1172/jci141837 | PubMed
19. **Litviňuková M.,** et al. (2020) Cells of the adult human heart. *Nature* **588**:466-472 https://doi.org/10.1038/s41586-020-2797-4 | PubMed
20. **Koenig A.L.,** et al. (2022) Single-cell transcriptomics reveals cell-type-specific diversification in human heart failure. *Nature Cardiovascular Research* **1**:263-280 https://doi.org/10.1038/s44161-022-00028-6 | PubMed
21. **Tucker N.R.,** et al. (2020) Transcriptional and Cellular Diversity of the Human Heart. *Circulation* **142**:466-482 https://doi.org/10.1161/circulationaha.119.045401 | PubMed
22. **Queen R.,** et al. (2023) Spatial transcriptomics reveals novel genes during the remodelling of the embryonic human arterial valves. *PLOS Genetics* **19**:e1010777 https://doi.org/10.1371/journal.pgen.1010777 | PubMed
23. **Farah E.N.,** et al. (2024) Spatially organized cellular communities form the developing human heart. *Nature* **627**:854-864 https://doi.org/10.1038/s41586-024-07171-z | PubMed

24. **De Bono C.,** et al. (2025) Multi-modal refinement of the human heart atlas during the first gestational trimester. *Development* **152** https://doi.org/10.1242/dev.204555 | PubMed

25. **Anderson R.H.,** Bamforth S.D (2022) Morphogenesis of the Mammalian Aortic Arch Arteries. *Frontiers in Cell and Developmental Biology* **10** https://doi.org/10.3389/fcell.2022.892900 | PubMed

26. **Wolf F.A.,** Angerer P., Theis F.J (2018) SCANPY: large-scale single-cell gene expression data analysis. *Genome Biology* **19**:15 https://doi.org/10.1186/s13059-017-1382-0 | PubMed

27. **Lun A.,** Bach K., Marioni J (2016) Pooling across cells to normalize single-cell RNA sequencing data with many zero counts. *Genome Biology* **17** https://doi.org/10.1186/s13059-016-0947-7 | PubMed

28. **Scialdone A.,** et al. (2015) Computational assignment of cell-cycle stage from single-cell transcriptome data. *Methods* **85**:54-61 https://doi.org/10.1016/j.ymeth.2015.06.021 | PubMed

29. **Langfelder P.,** Zhang B., Horvath S (2007) Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. *Bioinformatics* **24**:719-720 https://doi.org/10.1093/bioinformatics/btm563 | PubMed

30. **Lun A.T.,** McCarthy D.J., Marioni J.C (2016) A step-by-step workflow for low-level analysis of single-cell RNA-seq data with Bioconductor. *F1000Res* **5**:2122 https://doi.org/10.12688/f1000research.9501.2 | PubMed

31. **Huang D.W.,** Sherman B.T., Lempicki R.A (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols* **4**:44-57 https://doi.org/10.1038/nprot.2008.211 | PubMed

32. **Sherman B.T,** et al. (2022) DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). *Nucleic Acids Res* **50**:W216-w221 https://doi.org/10.1093/nar/gkac194 | PubMed

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

34. **La Manno G.,** et al. (2018) RNA velocity of single cells. *Nature* **560**:494-498 https://doi.org/10.1038/s41586-018-0414-6 | PubMed

35. **Moran P.A** (1950) Notes on continuous stochastic phenomena. *Biometrika* **37**:17-23 https://doi.org/10.1093/biomet/37.1-2.17 | PubMed

36. **Kleshchevnikov V.,** et al. (2022) Cell2location maps fine-grained cell types in spatial transcriptomics. *Nature Biotechnology* **40**:661-671 https://doi.org/10.1038/s41587-021-01139-4 | PubMed

37. **Korsunsky I.,** et al. (2019) Fast, sensitive and accurate integration of single-cell data with Harmony. *Nature Methods* **16**:1289-1296 https://doi.org/10.1038/s41592-019-0619-0 | PubMed

38. **Farahani R.M.,** Xaymardan M (2015) Platelet-Derived Growth Factor Receptor Alpha as a Marker of Mesenchymal Stem Cells in Development and Stem Cell Biology. *Stem Cells Int* **2015**:362753 https://doi.org/10.1155/2015/362753 | PubMed

39. **Wang S.,** et al. (2018) Platelet-derived growth factor receptor beta identifies mesenchymal stem cells with enhanced engraftment to tissue injury and pro-angiogenic property. *Cell Mol Life Sci* **75**:547-561 https://doi.org/10.1007/s00018-017-2641-7 | PubMed

40. **Ge Z.,** et al. (2022) Long noncoding RNA NEAT1 promotes cardiac fibrosis in heart failure through increased recruitment of EZH2 to the Smad7 promoter region. *Journal of Translational Medicine* **20**:7 https://doi.org/10.1186/s12967-021-03211-8 | PubMed

41. **Aibar S.,** et al. (2017) SCENIC: single-cell regulatory network inference and clustering. *Nature Methods* **14**:1083-1086 https://doi.org/10.1038/nmeth.4463 | PubMed

42. **Gharibeh L.,** et al. (2018) GATA6 Regulates Aortic Valve Remodeling, and Its Haploinsufficiency Leads to Right-Left Type Bicuspid Aortic Valve. *Circulation* **138**:1025-1038 https://doi.org/10.1161/circulationaha.117.029506 | PubMed

43. **Kodo K.,** et al. (2009) GATA6 mutations cause human cardiac outflow tract defects by disrupting semaphorin-plexin signaling. *Proc Natl Acad Sci U S A* **106**:13933-13938 https://doi.org/10.1073/pnas.0904744106 | PubMed

44. **Losa M.,** et al. (2017) A tissue-specific, Gata6-driven transcriptional program instructs remodeling of the mature arterial tree. *eLife* **6**:e31362 https://doi.org/10.7554/eLife.31362 | PubMed

45. **McLean C.Y.,** et al. (2010) GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol* **28**:495-501 https://doi.org/10.1038/nbt.1630 | PubMed

46. **Motenko H.,** Neuhauser S.B., O'Keefe M., Richardson J.E (2015) MouseMine: a new data warehouse for MGI. *Mamm Genome* **26**:325-330 https://doi.org/10.1007/s00335-015-9573-z | PubMed

47. **Cerezo M.,** et al. (2024) The NHGRI-EBI GWAS Catalog: standards for reusability, sustainability and diversity. *Nucleic Acids Research* **53**:D998-D1005 https://doi.org/10.1093/nar/gkae1070 | PubMed

48. **Thériault S.,** et al. (2024) Integrative genomic analyses identify candidate causal genes for calcific aortic valve stenosis involving tissue-specific regulation. *Nature Communications* **15**:2407 https://doi.org/10.1038/s41467-024-46639-4 | PubMed

49. **Bard-Chapeau E.A.,** et al. (2014) Mice Carrying a Hypomorphic Evi1 Allele Are Embryonic Viable but Exhibit Severe Congenital Heart Defects. *PLOS One* **9**:e89397 https://doi.org/10.1371/journal.pone.0089397 | PubMed

50. **Small A.M.,** et al. (2023) Multiancestry Genome-Wide Association Study of Aortic Stenosis Identifies Multiple Novel Loci in the Million Veteran Program. *Circulation* **147**:942-955 https://doi.org/10.1161/circulationaha.122.061451 | PubMed

51. **Varadkar P.,** et al. (2008) Notch2 is required for the proliferation of cardiac neural crest-derived smooth muscle cells. *Developmental Dynamics* **237**:1144-1152 https://doi.org/10.1002/dvdy.21502 | PubMed

52. **McDaniell R.,** et al. (2006) NOTCH2 Mutations Cause Alagille Syndrome, a Heterogeneous Disorder of the Notch Signaling Pathway. *The American Journal of Human Genetics* **79**:169-173 https://doi.org/10.1086/505332 | PubMed

53. **Todorovic V.,** et al. (2007) Long form of latent TGF-β binding protein 1 (Ltbp1L) is essential for cardiac outflow tract septation and remodeling. *Development* **134**:3723-3732 https://doi.org/10.1242/dev.008599 | PubMed

54. **Mommersteeg M.T.M.,** Yeh M.L., Parnavelas J.G., Andrews W.D (2015) Disrupted Slit-Robo signalling results in membranous ventricular septum defects and bicuspid aortic valves. *Cardiovascular Research* **106**:55-66 https://doi.org/10.1093/cvr/cvv040 | PubMed

55. **Gittenberger-de Groot A.C.,** Vrancken Peeters M.-P.F., Mentink M.M., Gourdie R.G., Poelmann R.E. (1998) Epicardium-derived cells contribute a novel population to the myocardial wall and the atrioventricular cushions. *Circulation research* **82**:1043-1052 https://doi.org/10.1161/01.res.82.10.1043 | PubMed

56. **Kirby M.L.,** Gale T.F., Stewart D.E (1983) Neural crest cells contribute to normal aorticopulmonary septation. *Science* **220**:1059-1061 https://doi.org/10.1126/science.6844926 | PubMed

57. **Hayashi S.,** et al. (2016) Gata6-Dependent GLI3 Repressor Function is Essential in Anterior Limb Progenitor Cells for Proper Limb Development. *PLOS Genetics* **12**:e1006138 https://doi.org/10.1371/journal.pgen.1006138 | PubMed

58. **Babij P.,** Kelly C., Periasamy M (1991) Characterization of a mammalian smooth muscle myosin heavy-chain gene: complete nucleotide and protein coding sequence and analysis of the 5' end of the gene. *Proc Natl Acad Sci U S A* **88**:10676-10680 https://doi.org/10.1073/pnas.88.23.10676 | PubMed

59. **Fang M.,** Alfieri C.M., Hulin A., Conway S.J., Yutzey K.E (2014) Loss of β-catenin promotes chondrogenic differentiation of aortic valve interstitial cells. *Arterioscler Thromb Vasc Biol* **34**:2601-2608 https://doi.org/10.1161/atvbaha.114.304579 | PubMed

60. **Trapnell C.,** et al. (2014) The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat Biotechnol* **32**:381-386 https://doi.org/10.1038/nbt.2859 | PubMed

61. **Sawada H.,** Rateri D.L., Moorleghen J.J., Majesky M.W., Daugherty A (2017) Smooth Muscle Cells Derived From Second Heart Field and Cardiac Neural Crest Reside in Spatially Distinct Domains in the Media of the Ascending Aorta—Brief Report. *Arteriosclerosis. Thrombosis, and Vascular Biology* **37**:1722-1726 https://doi.org/10.1161/atvbaha.117.309599 | PubMed

62. **Giannotta M.,** Trani M., Dejana E (2013) VE-cadherin and endothelial adherens junctions: active guardians of vascular integrity. *Dev Cell* **26**:441-454 https://doi.org/10.1016/j.devcel.2013.08.020 | PubMed

63. **Hulin A.,** et al. (2018) Macrophage Transitions in Heart Valve Development and Myxomatous Valve Disease. *Arteriosclerosis. Thrombosis, and Vascular Biology* **38**:636-644 https://doi.org/10.1161/atvbaha.117.310667 | PubMed

64. **Collin M.,** Bigley V (2018) Human dendritic cell subsets: an update. *Immunology* **154**:3-20 https://doi.org/10.1111/imm.12888 | PubMed

65. **Bach T.L.,** et al. (1998) VE-Cadherin Mediates Endothelial Cell Capillary Tube Formation in Fibrin and Collagen Gels1. *Experimental Cell Research* **238**:324-334 https://doi.org/10.1006/excr.1997.3844 | PubMed

66. **Kern C.B.,** et al. (2010) Reduced versican cleavage due to Adamts9 haploinsufficiency is associated with cardiac and aortic anomalies. *Matrix biology* **29**:304-316 https://doi.org/10.1016/j.matbio.2010.01.005 | PubMed

67. **Rutkovskiy A.,** et al. (2017) Valve Interstitial Cells: The Key to Understanding the Pathophysiology of Heart Valve Calcification. *J Am Heart Assoc* **6** https://doi.org/10.1161/jaha.117.006339 | PubMed

68. **Kumar S.N.,** Boss J.M (2000) Site A of the MCP-1 distal regulatory region functions as a transcriptional modulator through the transcription factor NF1. *Molecular immunology* **37**:623-632 https://doi.org/10.1016/s0161-5890(00)00097-3 | PubMed

69. **Yuan B.Z.,** et al. (1998) Cloning, characterization, and chromosomal localization of a gene frequently deleted in human liver cancer (DLC-1) homologous to rat RhoGAP. *Cancer Res* **58**:2196-2199 PubMed

70. **Monkley S.J.,** Pritchard C.A., Critchley D.R (2001) Analysis of the Mammalian Talin2 Gene TLN2. *Biochemical and Biophysical Research Communications* **286**:880-885 https://doi.org/10.1006/bbrc.2001.5497 | PubMed

71. **Bradshaw A.D.,** Graves D.C., Motamed K., Sage E.H (2003) SPARC-null mice exhibit increased adiposity without significant differences in overall body weight. *Proc Natl Acad Sci U S A* **100**:6045-6050 https://doi.org/10.1073/pnas.1030790100 | PubMed

72. **Neeb Z.,** Lajiness J.D., Bolanis E., Conway S.J (2013) Cardiac outflow tract anomalies. *WIREs Developmental Biology* **2**:499-530 https://doi.org/10.1002/wdev.98 | PubMed

73. **Eldadah Z.A.,** et al. (2001) Familial Tetralogy of Fallot caused by mutation in the jagged1 gene. *Hum Mol Genet* **10**:163-169 https://doi.org/10.1093/hmg/10.2.163 | PubMed

74. **Jiang J.Q.,** et al. (2013) Prevalence and spectrum of GATA5 mutations associated with congenital heart disease. *Int J Cardiol* **165**:570-573 https://doi.org/10.1016/j.ijcard.2012.09.039 | PubMed

75. **Maitra M.,** Koenig S.N., Srivastava D., Garg V (2010) Identification of GATA6 sequence variants in patients with congenital heart defects. *Pediatr Res* **68**:281-285 https://doi.org/10.1203/pdr.0b013e3181ed17e4 | PubMed

76. **Tomita-Mitchell A.,** Maslen C.L., Morris C.D., Garg V., Goldmuntz E (2007) GATA4 sequence variants in patients with congenital heart disease. *J Med Genet* **44**:779-783 https://doi.org/10.1136/jmg.2007.052183 | PubMed

77. **Yang B.,** et al. (2017) Protein-altering and regulatory genetic variants near GATA4 implicated in bicuspid aortic valve. *Nat Commun* **8**:15481 https://doi.org/10.1038/ncomms15481 | PubMed

78. **Al Turki S.,** et al. (2014) Rare variants in NR2F2 cause congenital heart defects in humans. *Am J Hum Genet* **94**:574-585 https://doi.org/10.1016/j.ajhg.2014.03.007 | PubMed

79. **Sanchez J.,** et al. (2020) Conditional inactivation of Foxc1 and Foxc2 in neural crest cells leads to cardiac abnormalities. *genesis* **58**:e23364 https://doi.org/10.1002/dvg.23364 | PubMed

80. **Wang Q.,** Lan Y., Cho E.-S., Maltby K.M., Jiang R (2005) Odd-skipped related 1 (Odd1) is an essential regulator of heart and urogenital development. *Developmental Biology* **288**:582-594 https://doi.org/10.1016/j.ydbio.2005.09.024 | PubMed

81. **Sawada H.,** Rateri D.L., Moorleghen J.J., Majesky M.W., Daugherty A (2017) Smooth Muscle Cells Derived From Second Heart Field and Cardiac Neural Crest Reside in Spatially Distinct Domains in the Media of the Ascending Aorta-Brief Report. *Arterioscler Thromb Vasc Biol* **37**:1722-1726 https://doi.org/10.1161/atvbaha.117.309599 | PubMed

82. **Dirkx E.,** da Costa Martins P.A., De Windt L.J (2013) Regulation of fetal gene expression in heart failure. *Biochim Biophys Acta* **1832**:2414-2424 https://doi.org/10.1016/j.bbadis.2013.07.023 | PubMed

83. **Oka T.,** Xu J., Molkentin J.D (2007) Re-employment of developmental transcription factors in adult heart disease. *Semin Cell Dev Biol* **18**:117-131 https://doi.org/10.1016/j.semcdb.2006.11.012 | PubMed

84. **Forte E.,** et al. (2022) Adult mouse fibroblasts retain organ-specific transcriptomic identity. *eLife* **11**:e71008 https://doi.org/10.7554/eLife.71008 | PubMed

Rotem Leshem, Syed Murtuza Baker, Nicoletta Bobola (2023) A cell atlas of the human outflow tract of the heart and its adult derivatives - CS16-17 Replicate 1. ArrayExpress. ID E-MTAB-13456 https://www.ebi.ac.uk/biostudies/ArrayExpress/studies/E-MTAB-13456

Rotem Leshem, Syed Murtuza Baker, Nicoletta Bobola (2023) A cell atlas of the human outflow tract of the heart and its adult derivatives. ArrayExpress. ID E-MTAB-13447 https://www.ebi.ac.uk/biostudies/ArrayExpress/studies/E-MTAB-13447

Rotem Leshem, Syed Murtuza Baker, Nicoletta Bobola (2023) A cell atlas of the human outflow tract of the heart and its adult derivatives - Adult samples. ArrayExpress. ID E-MTAB-13453 https://www.ebi.ac.uk/biostudies/ArrayExpress/studies/E-MTAB-13453

Rotem Leshem, Syed Murtuza Baker, Nicoletta Bobola (2023) A cell atlas of the human outflow tract of the heart and its adult derivatives - 10X visium sample. ArrayExpress. ID E-MTAB-13461 https://www.ebi.ac.uk/biostudies/ArrayExpress/studies/E-MTAB-13461

Peer reviews

Reviewer #1 (Public review):

Summary:

The study by Bobola et al reports single nuclear expression analysis with some supporting spatial expression data of human embryonic and fetal cardiac outflow tracts compared to adult aortic valves. The transcription factor GATA6 is identified as a top regulator of one of the mesenchymal subpopulations and potential interacting factors and downstream target genes are identified bioinformatically. Additional bioinformatic tools are used to describe cell lineage relationships and trajectories for developmental and adult cardiac cell types.

Strengths:

The strengths of the study are studies of human tissue and extensive gene expression data that will be valuable to the field.

Weaknesses:

In the revised manuscript the data remain largely correlative since functional relationships in cell lineages and gene regulatory interactions are based on coexpression data and

bioinformatic analyses that were not subjected to further validation.

https://doi.org/10.7554/eLife.107748.2.sa3

Reviewer #2 (Public review):

Summary:

The manuscript by Leshem et al. presents a transcriptomic analysis of the developing human outflow tract (OFT) at embryonic and fetal stages using snRNAseq and spatial transcriptomic. Additionally, the authors analyze transcriptomic data from the adult aortic valve to compare embryonic and adult cell population, aiming to identify persistent embryonic transcriptional signatures in adult cells. A total of 15 clusters were identified from the embryonic and fetal OFT samples, including three mesenchymal and four endothelial clusters. Using SCENIC analysis on the embryonic snRNAseq data, the authors identified GATA6 as a key regulator of valve precursor cells. Spatial transcriptomic analysis of four fetal OFT sections further revealed the spatial distribution of mesenchymal nuclei, smooth muscle cells, and valvular interstitial cells. Trajectory analysis identified two distinct developmental origins of fetal mesenchymal cells: the neural crest and the second heart field. Finally, the authors used snRNAseq data from the adult aortic valve to propose that embryonic transcriptional signatures persist in a subset of adult cells.

Strengths:

(1) The study offers a rich and detailed dataset, combining snRNA-seq and spatial transcriptomics in human embryonic and fetal OFT, which are challenging to obtain.

(2) The use of SCENIC and trajectory analysis adds mechanistic insight into cell lineage and regulatory programs during valve development.

(3) This study confirms GATA6 ss a key regulator of valve precursor cells.

(4) Comparison between embryonic/fetal and adult datasets represents a novel attempt to trace persistence of developmental transcriptional programs.

Weaknesses:

(1) A major limitation is the lack of experimental validation to support key conclusions, particularly the claim of persistent embryonic transcriptional signatures in adult cells.

(2) The manuscript would benefit from a clearer discussion of how these results advance beyond previous studies in human heart and valve development.

(3) The comparison between embryonic and adult data is interesting but would be more convincing with additional evidence supporting the proposed persistence of embryonic transcriptional signatures in adult cells

Comments on revisions:

The final section of the results concludes with the search for a distribution pattern similar to JAG1. The authors end their article by identifying the FOXC1 and OSR1 genes without providing further validation for their discovery, which is regrettable.

https://doi.org/10.7554/eLife.107748.2.sa2

Reviewer #3 (Public review):

Leshem et al have generated a transcriptional cell atlas of the human outflow tract at two developmental timepoints and its adult valvular derivatives. This carefully performed study provides a useful resource for the study of known genes implicated in outflow tract defects and potentially also to discover new disease genes. The authors reveal neural crest and mesodermal contributions to different outflow tract components and show that GATA6, known to play a role in arterial valve development, controls a set of genes expressed in endocardial derived cells during valve development. Interestingly the results reveal intersection with GLI3 and suggest lineage persistence of gene expression through to the adult timepoint, a main new finding of this study.

Comments on revisions:

The authors have carefully addressed previous comments, including the addition of new analysis pointing to potential cooperation between GATA6 and GLI3.

https://doi.org/10.7554/eLife.107748.2.sa1

Author response:

The following is the authors' response to the original reviews.

> ***Public Reviews:***
>
> ***Reviewer #1 (Public review):***
>
> *Summary:*
>
> *The study by Bobola et al reports single-nucleus expression analysis with some supporting spatial expression data of human embryonic and fetal cardiac outflow tracts compared to adult aortic valves. The transcription factor GATA6 is identified as a top regulator of one of the mesenchymal subpopulations, and potential interacting factors and downstream target genes are identified bioinformatically. Additional bioinformatic tools are used to describe cell lineage relationships and trajectories for developmental and adult cardiac cell types.*
>
> *Strengths:*
>
> *The studies of human tissue and extensive gene expression data will be valuable to the field.*
>
> *Weaknesses:*
>
> *(1) The expression data are largely confirmatory of previous studies in humans and mice. Thus, it is not clear what novel biological insights are being reported. While there is some novelty and impact in using human tissue, there are extensive existing publications and data sets in this area.*
>
> *(2) Major conclusions regarding spatial localization, differential gene expression, or cell lineage relationships based on bioinformatic data are not validated in the context of intact tissues.*
>
> *(3) The conclusions regarding lineage relationships are based on common gene expression in the current study and may not reflect cellular origins or lineage relationships that have previously been reported in genetic mouse models.*

(4) An additional limitation is the exclusive examination of adult aortic valve leaflets that represent only a subset of outflow tract derivatives in the mature heart. The conclusion, as stated in the title regarding adult derivatives of the outflow tract, is not accurate based on the limited adult tissue evaluated, exclusive bioinformatic approach, and lack of experimental lineage analysis of cell origins.

Reviewer #2 (Public review):

Summary:

The manuscript by Leshem et al. presents a transcriptomic analysis of the developing human outflow tract (OFT) at embryonic and fetal stages using snRNAseq and spatial transcriptomics. Additionally, the authors analyze transcriptomic data from the adult aortic valve to compare embryonic and adult cell populations, aiming to identify persistent embryonic transcriptional signatures in adult cells. A total of 15 clusters were identified from the embryonic and fetal OFT samples, including three mesenchymal and four endothelial clusters. Using SCENIC analysis on the embryonic snRNAseq data, the authors identified GATA6 as a key regulator of valve precursor cells. Spatial transcriptomic analysis of four fetal OFT sections further revealed the spatial distribution of mesenchymal nuclei, smooth muscle cells, and valvular interstitial cells. Trajectory analysis identified two distinct developmental origins of fetal mesenchymal cells: the neural crest and the second heart field. Finally, the authors used snRNAseq data from the adult aortic valve to propose that embryonic transcriptional signatures persist in a subset of adult cells.

Strengths:

(1) The study offers a rich and detailed dataset, combining snRNA-seq and spatial transcriptomics in human embryonic and fetal OFT, which are challenging to obtain.

(2) The use of SCENIC and trajectory analysis adds mechanistic insight into cell lineage and regulatory programs during valve development.

(3) This study confirms GATA6 as a key regulator of valve precursor cells.

(4) Comparison between embryonic/fetal and adult datasets represents a novel attempt to trace persistence of developmental transcriptional programs.

Weaknesses:

(1) A major limitation is the lack of experimental validation to support key conclusions, particularly the claim of persistent embryonic transcriptional signatures in adult cells.

(2) The manuscript would benefit from a clearer discussion of how these results advance beyond previous studies in human heart and valve development.

(3) The comparison between embryonic and adult data is interesting, but would be more convincing with additional evidence supporting the proposed persistence of embryonic transcriptional signatures in adult cells.

Reviewer #3 (Public review):

Leshem et al have generated a transcriptional cell atlas of the human outflow tract at two developmental timepoints and its adult valvular derivatives. This carefully performed study provides a useful resource for the study of known genes implicated in outflow tract defects and potentially also for discovering new disease genes. The authors reveal neural crest and mesodermal contributions to different outflow tract components and show that GATA6, known to play a role in arterial valve development, controls a set of genes

> *expressed in endocardium-derived cells during valve development. Interestingly, the results suggest lineage persistence of expression of certain genes through to the adult timepoint, a main new finding of this study.*
>
> *The following points should be addressed to reinforce the conclusions and emphasize the novel features of this study.*
>
> *(1) It would be helpful to clarify how these new findings confirm or diverge from what is known from analysis of neural crest and mesodermal lineage contributions to different cell populations in the mouse heart. Did the authors identify any human-specific populations of cells, such as the LGR5 population reported by Sahara et al?*
>
> *(2) The authors should clarify in the introduction and results that they consider the endocardium to be on the SHF trajectory as indicated in Figure S4C. Please add a reference for this point.*
>
> *(3) The GATA6 results are interesting and support this experimental approach. The paper would be reinforced if the authors could provide any functional validation (in addition to their GATA6 genomic occupancy data) that the designated target genes are regulated by GATA6. This might involve looking at mutant mouse embryos or cultured cells. Do the authors consider that GATA6 may regulate the endocardial to mesenchymal transition during the early stages of valve development? Or the valve interstitial cell versus fibroblast fate choice?*
>
> *(4) Do the new findings reveal whether human valves have a direct SHF to VIC trajectory (ie, without transiting through endocardium) as has been recently shown in the murine non-coronary valve leaflet? Relevant to this point, Figure 5E appears to show contributions to a single adult aortic valve leaflet - this should be explained, or corrected.*

We sincerely thank the Editor and the Reviewers for their constructive and insightful comments. We have carefully addressed the majority of the points raised and believe the revisions have substantially strengthened the manuscript.

> ***Recommendations for the authors:***
>
> ***Reviewing Editor Comments:***
>
> *Overall, the reviewers felt that integrating these datasets with prior snRNAseq datasets on human OFT (de Bono et al, 2025) would enhance analyses and provide broader context.*

Several human fetal heart single-cell datasets have been published, including De Bono et al, 2025. We carefully considered whether integrative analyses with these datasets would further strengthen our study. However, there are substantial differences in anatomical scope: most published datasets encompass broad cardiac regions, whereas our study specifically targets the OFT, enabling higher-resolution characterization of OFT-specific cell states. Integration across datasets with markedly different regional compositions would likely be driven by largescale anatomical differences rather than yield additional OFT-specific insight. In addition, cross-study integration requires batch correction. When datasets differ in anatomical scope, as well as developmental timing, and experimental protocols, stronger correction may be needed, increasing the risk of overcorrection and potential loss of biologically meaningful OFTspecific signals.

Importantly, our dataset has been deposited in the Human Cell Atlas and is fully available for future comparative analyses. We therefore believe that broader cross-dataset integration is best undertaken within such harmonized frameworks as more closely matched datasets become available.

> *Overall, cluster annotations should be more rigorous, which may be facilitated by comparisons with earlier studies.*

We have clarified all the points raised by the reviewer regarding cluster annotation. Specifically: (1) the "cardiac" cluster has been renamed "cardiac muscle" to more accurately reflect its transcriptional identity; and (2) we now explicitly state that mesenchymal populations not resolved in the initial global analysis (across all samples) were subsequently defined through dedicated sub clustering analyses performed separately for the adult and developmental datasets. These clarifications have been incorporated into the revised manuscript.

> *Citation of other spatial transcriptomics studies on human OFT would be useful.*

We apologise for missing these contributions. They have now been added to the text.

> *Can the authors identify a human-specific population of cells, such as the LGR5 population reported by Sahara et al?*

While our dataset does not reveal a novel single-gene marker comparable to the human specific LGR5 marker described for the LGR5-positive population by Sahara et al., it does identify a distinct GATA6-enriched embryonic mesenchymal population that functions as a human valve progenitor lineage. Using regulatory network analysis, RNA velocity, lineage tracing and spatial transcriptomics, we show that this GATA6-driven program is specifically associated with semilunar valve morphogenesis and that its transcriptional signature persists in fetal and adult VIC populations. Thus, the novelty of our study lies in defining this human GATA6-regulated valve progenitor population and its lineage trajectory, rather than in the identification of previously unreported single marker genes.

"….Although we have not defined a novel single-gene marker (analogous to LRG5 [Sahara et al]), our identification of a GATA6 network highlights….."

> *Further investigation of the specific role of GATA6 would strengthen findings.*
>
> *FISH studies would indicate whether GATA6 is involved in EMT or fibroblast versus valve interstitial cell fate choice.*

We have added a panel to Fig. S2 (D), showing that GATA6 expression is not restricted to specific outflow tract populations. In CS16-17 embryos, GATA6-expressing nuclei are detected across all embryonic clusters. Given this broad expression pattern, FISH analysis would not distinguish whether GATA6 functions in EMT or in fibroblast versus valve interstitial cell fate specification. While we cannot exclude the possibility that GATA6 contributes to EMT, we observe that its expression levels are highest in cluster 4 (post-EMT) cells. This suggests that GATA6 activation is more likely a consequence of the transition rather than its initiating cause (shown in Fig. S2D).

> *Functional validation of some proposed GATA6 targets would strengthen findings.*

To our knowledge, there are currently no publicly available datasets defining the GATA6 regulatory network in human OFT cells or valvular fibroblast progenitors. Existing datasets focus primarily on cardiomyocytes, which arise from a distinct developmental lineage. Given the well-established cell-type and context dependence of transcription factor activity, these datasets are unlikely to provide meaningful insight into regulatory relationships within the valvular lineage examined here.

As noted in the original submission, we previously leveraged published mouse GATA6 ChIPseq data from E11.5 OFT (DOI: *https://doi.org/10.7554/eLife.31362*) as independent support for the GATA6 regulon identified in our human dataset. In this revised version, we

have now extended this analysis by formally quantifying the overlap between the cluster 4 GATA6 regulon and genes bound by GATA6 in the mouse OFT dataset. Using a hypergeometric enrichment test, we found that the observed overlap is approximately two-fold greater than expected by chance and highly significant ($p = 1.2 \times 10^{-33}$). This statistical analysis strengthens our original interpretation and provides quantitative support that the identified regulon is strongly enriched for bona fide GATA6-bound targets in a closely related developmental context.

In addition, we examined the spatial expression pattern of the GATA6 regulon gene set and found that it specifically localizes to the semilunar valves (OFT derivatives), consistent with GATA6 activity in this developmental context. This new analysis has been incorporated into Figure 2F of the revised manuscript.

Collectively, the cross-species binding enrichment and valve-specific expression pattern provide orthogonal support for the biological relevance of the identified GATA6 regulon and strengthen the mechanistic interpretation of GATA6 function in OFT and valve development.

> *As GATA6 has been previously identified in mouse studies, can the authors identify novel transcription factors potentially involved in OFT development?*

To identify additional transcription factors potentially involved in OFT development and to define regulators that may confer specificity to GATA6 activity, we compared the GATA6 regulon with the regulons of other cluster 4 transcription factors identified by SCENIC (SOX4, GLI3, RARG, ETV1, GLIS3, BACH2, ZNF423, FOXO3, ZBTB20).

While all cluster 4 regulators share some downstream targets, GLI3 regulon showed approximately twice the degree of overlap with the GATA6 regulon compared to the other factors. This suggests a potential functional interaction between GATA6 and GLI3 in OFT associated mesenchyme. Consistent with this, cooperation between GATA6 and GLI3 has been reported in mouse limb development. These findings have now been incorporated into the Results section, and co-expression of GATA6 and GLI3 in CS16-17 populations is shown in Figure S2DE.

Although GATA6 has previously been implicated in OFT development, SCENIC analysis provides mechanistic insight by defining the downstream gene programs active in specific human embryonic lineages. Thus, the novelty of our findings lies not in re-identifying GATA6, but in characterizing its regulon in human OFT- and valve-associated mesenchyme and identifying potential cooperating regulators such as GLI3.

> *Embryonic signatures in adult valve cells are an interesting finding, that should be further explored by pseudotime trajectories, which may also indicate whether SHF cells have a direct trajectory to VIC (without transiting endocardium), as recently shown in mice.*

We included all embryonic populations, including cardiac progenitor cells (SHF), in the pseudotime trajectory analysis. However, we did not observe evidence of a direct trajectory from SHF cells toward VIC. In contrast, the same analysis consistently identified a trajectory linking endocardial cells to VIC, supporting an endocardial origin in our dataset.

> ***Reviewer #1 (Recommendations for the authors):***
>
> *(1) Major conclusions regarding cell lineages and derivatives are based on common gene expression patterns and bioinformatic tools. Thus, these conclusions are not based on empirical data, and assumptions regarding lineages based on gene expression may not be accurate. The language related to lineage analysis, derivative, and longitudinal gene expression is not supported by data. For example, studies in mice have shown that aortic valve interstitial cells from endocardial cushions and neural crest-derived lineages have*

> *overlapping patterns of ECM gene expression and cannot be easily distinguished in adults. Thus, it is not possible to determine derivation and cell origins based on gene expression alone.*

While we fully acknowledge that gene expression-based analyses provide correlative rather than direct lineage-tracing evidence, the Reviewer's statement that "it is not possible to determine derivation and cell origins based on gene expression alone," and the example cited in support, appear to equate global transcriptional similarity with the distinct embryonic transcriptional signatures that underpin our analysis.

As the Reviewer notes, a given differentiated cell type can derive from different embryonic progenitors. Due to functional convergence, differentiated cells often exhibit highly similar expression profiles that reflect their shared function rather than developmental origin. Consequently, discriminating embryonic origins based on global expression profiles, or even for highly distinctive genes of differentiated cells, is very challenging. The example cited by the Reviewer - overlapping ECM gene expression in aortic valve interstitial cells derived from endocardial cushions and neural crest - illustrates precisely this point.

However, our analysis does not rely on global transcriptional similarity or on markers of mature differentiated cells. Instead, we specifically identified gene sets that are highly distinctive of embryonic clusters prior to the onset of differentiation. These signatures are enriched for transcription factors and signaling molecules that define developmental identity, rather than functional effector genes associated with mature cell states. We have shown that these embryonic signatures persist in fetal cells (which already express differentiated markers but are developmentally closer to the embryonic stage relative to adult cells) and remain detectable, albeit attenuated, in adult cells. It is these distinctive embryonic transcriptional signatures, rather than global or shared functional gene expression, that we have used to infer potential lineage relationships.

We fully acknowledge that this constitutes correlative evidence rather than direct lineage tracing, which is not feasible in human studies. However, the persistence of embryonic regulatory signatures into fetal and adult stages provides a biologically plausible link to developmental origin. This persistence most plausibly reflects partial retention of ancestral embryonic transcriptional programs in descendant cells, rather than de novo activation later in life of embryonic genes that were never previously expressed in that cell's lineage.

> *(2) Most of the findings related to cell composition, gene expression, and cell lineages seem to be largely confirmatory of previous reports. Novel findings should be emphasized and validated in the tissues.*

We agree that several aspects of our dataset reproduce and extend findings from previous human and animal studies, which we regard as an important validation of the atlas. However, our study also provides multiple novel insights that are directly supported by our spatial data. Specifically, we (i) identify a GATA6-enriched embryonic mesenchymal valve progenitor population, (ii) delineate its GATA6 transcriptional regulon and direct targets implicated in OFT and valve disease, and (iii) trace its embryonic transcriptional signature into fetal and adult valve interstitial cell populations. These findings are strengthened by our spatial transcriptomic data, which maps the GATA6 regulon and key targets to the semilunar valves and adjacent arterial root, providing in situ validation of both cell identity and gene expression patterns (see Fig. 3 and the newly added Fig. 2F). We have revised the Discussion to more explicitly highlight these novel aspects and their spatial validation in the final

"In summary, our work goes beyond confirming previously reported cell types by (i) defining a GATA6-regulated human valve progenitor lineage and its descendants, (ii) establishing distinct embryonic origins for smooth muscle and valvular fibroblasts, and (iii) demonstrating persistence of embryonic signatures in adult valve cell populations. These

findings are directly supported in tissue by our spatial transcriptomics data, which map these lineages and regulatory programs to defined anatomical domains within the human OFT and semilunar valves."

(3) The developing outflow tract of the heart contributes to more than just the aortic valve leaflets in adults. Additional conotruncal structures need to be evaluated in order to define adult derivatives of the developing outflow tract as described in the title.

The title has been changed to reflect that only adult aortic valves were examined.

> *(4) Major conclusions regarding the GATA6 regulatory network and downstream target genes are not validated in the context of the developing outflow tract or adult valves. Is GATA6 expression restricted to specific outflow tract populations? Is GATA6 binding or responsive gene expression detected for the indicated target genes?*

We performed additional analyses that further reinforce the relationship between GATA6 and its target genes and support the biological relevance of GATA6 downstream targets in arterial valve development. Below, we address the specific questions raised by the reviewer.

> *(1) Is GATA6 expression restricted to specific outflow tract populations?*

GATA6 expression is not restricted to specific outflow tract populations. In CS16-17 embryos, GATA6-expressing cells are detected across all embryonic clusters; however, expression levels are highest in cluster 4 (valve precursor cells).

Despite this broad expression pattern, SCENIC identifies GATA6 activity (i.e., a GATA6 regulon) specifically in cluster 4. This apparent restriction of GATA6 regulatory activity to cluster 4 may be explained, at least in part, by its elevated expression levels within this cluster. Alternatively, given that transcription factors often act in a combinatorial manner, GATA6 may co-regulate its target genes in cluster 4 together with additional cluster-specific regulators. To explore this possibility, we compared the GATA6 regulon with the regulons of other cluster 4 transcription factors identified by SCENIC (namely SOX4, GLI3, RARG, ETV1, GLIS3, BACH2, ZNF423, FOXO3, ZBTB20) in order to identify potential co-regulatory modules. As expected, since these regulons are sampled from the subset of genes enriched in cluster 4, all regulators share a substantial proportion of downstream targets with GATA6. However, GLI3 stands out, showing approximately twice the degree of overlap compared to the other factors. This suggests a functional interaction between GATA6 and GLI3, consistent with previously reported cooperation in mouse limb development. These results have been incorporated into the Results section, and the expression of GATA6 and GLI3 in CS16-17 cell populations is shown in Fig. S2DE.

> *(2) Is GATA6 binding or responsive gene expression detected for the indicated target genes?*

We were unable to find public data describing the GATA6 regulatory network or its downstream targets in the specific human cell types examined here (OFT cells; valvular fibroblast progenitors). Available datasets focus primarily on cardiomyocytes, which arise from a distinct lineage, and because transcription factor function is highly cell-type and context dependent, these datasets are unlikely to be helpful in inferring regulatory relationships in the valvular lineage.

The strongest validation for the GATA6 regulon identified in this study comes from the mouse GATA6 occupancy data (this was included in the original manuscript). Although derived from a different species, GATA6 binding has been profiled in a highly related developmental context, the OFT. To assess the relevance of these data to our human findings, we performed a hypergeometric test comparing the GATA6 regulon identified in cluster 4 (this study) with genes bound by GATA6 in E11.5 mouse OFT ChIP-seq data (DOI:

https://doi.org/10.7554/eLife.31362). The observed overlap is substantially greater than expected by chance: it is approximately twice the expected value, and the enrichment is highly significant ($p = 1.2 \times 10^{-33}$). Biologically, this strongly supports the interpretation that many genes within GATA6 regulon are likely to be direct GATA6 targets, or at minimum are strongly associated with GATA6 binding, rather than representing a random gene set. This analysis has been added to the revised manuscript.

In this revised version of the manuscript, we also overlapped the expression of GATA6 regulon genes to our fetal spatial transcriptomics data. GATA6 regulon was identified in embryonic cluster 4, whose expected trajectory is fetal valvular fibroblasts (cluster 12). Remarkably, GATA6 regulon genes are expressed in both the aortic and pulmonary valves, and their expression pattern aligns closely with HAPLN1-positive valvular fibroblasts (cluster 12), further supporting the biological relevance of this gene set. This new data has been added to Fig 2(F).

Together, the strong enrichment of GATA6 regulon genes among GATA6-bound targets in the OFT, and the specific expression of this gene set within the arterial valves (cluster 4 descendant cells), support the biological relevance of GATA6 downstream targets in arterial valve development and disease. In addition, we identify GLI3 as a potential GATA6 co-binding partner.

> *(5) What are "cardiac" cell types in the embryonic single cell clustering? Are these cardiomyocytes? Cardiac is an ambiguous term if the cells being analyzed are all in the heart.*

Thank you for highlighting this ambiguity. The “cardiac” population refers specifically to cardiac muscle cells. We have updated the labels in Fig. 1E, 1F, and Fig. S3A to make this explicit.

> *(6) The methods and analytical tools seem fairly standard for single nuclear gene expression and spatial genomics studies. What are the new tools and resources being reported? The "novel lineage tracing algorithm" mentioned in the methods is not well described. A Cellxgene VIP app is mentioned, but is not described in detail. Also, it seems to be housed on a local server, which is not optimal.*

The description of the lineage tracing algorithm has been expanded in the method’s section of the paper.

The data has been submitted to the Human Cell Atlas, a coordinated global effort to systematically map human cell types using standardized, interoperable formats. Public access via cell x gene enables interactive visualization, gene-level queries, and cross-dataset comparisons without requiring advanced computational expertise. This broad accessibility enhances reproducibility, facilitates integration with complementary single-cell and spatial datasets, and maximizes the visibility, transparency, and long-term impact of our work.

> *(7) Only adult aortic valves from females were included in the study.*

The rationale for using female tissues has been explained in the result section:

We collected female samples to mitigate individual variability and maximise the possibility to analyse healthy aortic valves, justified by the lower incidence and severity of aortic disease in females versus males.

> *(8) In many of the figures, the font size of the text is too small to read.*

We have increased the font size in all figures where this was compatible with the layout. For the larger plots, additional enlargement would necessitate scaling the panels beyond the allowable page dimensions, and therefore could not be implemented.

> *(9) "CAT" is not a commonly used abbreviation for congenital heart anomalies related to persistent truncus arteriosus.*

CAT is now the preferred term for PTA as latinised terms are no longer used.

> ***Reviewer #2 (Recommendations for the authors):***
>
> *Overall, this study is thoughtfully conducted and offers valuable observations that contribute to our understanding of valve morphogenesis. However, my main concern is the lack of experimental validation to support the findings, particularly the conclusion regarding the persistence of transcriptional signatures in adult cells, which is not sufficiently substantiated or clearly argued. It is unclear how this study advances beyond previous research in humans.*
>
> *Major points:*
>
> *(1) Several recent studies have applied spatial transcriptomics to human embryonic and fetal hearts, including OFT (Asp et al., 2019; Queen et al., 2023; Farah et al., 2024; De Bono et al., 2025). It is disappointing that the authors did not acknowledge these important contributions.*

We apologise for missing these contributions. They have now been added to the text.

> *(2) The present study used snRNAseq to explore the transcriptional signature of the fetal OFT. A similar approach was used by De Bono et al. (2025) to analyze fetal hearts. Integrating these complementary snRNAseq datasets could enhance the current analysis and provide broader context for the findings.*

The reviewers suggested that integrating our datasets with prior snRNA-seq datasets on human OFT (de Bono et al., 2025) could enhance the analyses and provide broader context. While several fetal heart datasets have been published (e.g., Sahara et al.), our study focuses specifically on the OFT. These other studies do not perform cross-dataset comparisons. We therefore do not see a strong rationale for integrating ours, especially given that those datasets cover much larger regions of the heart.

> *(3) Figure 1 presents 18 distinct clusters identified through unsupervised clustering. The authors classify three of these clusters broadly as mesenchymal cells. However, the term "mesenchymal cells" lacks precision. The authors should clarify why these clusters were not more specifically defined as fibroblasts or myofibroblasts based on marker expression.*

Clustering of the full dataset does not provide sufficient resolution to distinguish all mesenchymal cell types. The clusters broadly annotated as mesenchymal comprise heterogeneous populations, including both undifferentiated embryonic mesenchymal cells and more differentiated fetal mesenchymal cells. These mesenchymal clusters were therefore further subclustered, and the resulting cell identities are described in detail in the Results sections corresponding to Fig. 2 and Fig. 3.

> *(4) The authors used SCENIC on their snRNAseq datasets to infer key cell fate regulators and identified GATA6 as a top regulator of embryonic mesenchymal cluster 4. However, the rationale for focusing on GATA6, which is already known to be associated with CHD in humans, is not fully convincing. Why not investigate a transcription factor whose role in valve development remains unexplored?*

There are two key outcomes from a SCENIC analysis: (1) the identification of major transcriptional regulators driving the differentiation of a given cluster, and (2) the identification of their regulons (the downstream gene programs they control). While GATA6 is indeed already known to be associated with CHD in humans, including valve malformations and major OFT defects, its downstream targets in the relevant human developmental lineages have not been defined. Understanding these targets is essential for clarifying the molecular basis of GATA6-mediated CHD. Thus, the significance of our result does not lie in the rediscovery of GATA6 as a CHD-related factor, but in identifying the genes it regulates in embryonic OFT- and valve-associated mesenchyme. These GATA6-controlled genes in the OFT and valves represent biologically plausible candidate genes for human OFT defects, as disruption of GATA6 targets could similarly contribute to CHD.

In this revised version we have performed a hypergeometric test showing that GATA6 regulon genes are significantly enriched among genes bound by GATA6 in the OFT. Biologically, this strongly supports the interpretation that many genes within the GATA6 regulon are likely to be direct GATA6 targets, or at minimum are strongly associated with GATA6 binding in the OFT, rather than representing a random gene set.

We have also mapped the expression of GATA6 regulon to the semilunar valves. Collectively, these analyses demonstrate that the GATA6 regulon captures a biologically coherent and developmentally relevant program, offering new mechanistic insight into how GATA6 influences OFT and valve formation and how its disruption may contribute to CHD.

> *(5) Several studies have already suggested a role for GATA6 in EMT. Do the authors propose that GATA6 regulates this process during embryonic valve development? Once again, validation using FISH would be important to support these findings.*

We do not propose that GATA6 directly regulates EMT during embryonic valve development. We rather make two independent observations: (1) cluster 4 derives from cluster 7 (likely through EMT); (2) GATA6 regulates cluster4-specific genes.

The first observation is supported by RNA velocity, which links cluster 7 to cluster 4. Supporting this interpretation, endothelial cluster 7 is enriched for genes associated with arterial valve development, and mesenchymal cluster 4 cells are identified as progenitors of fetal valve fibroblasts. Because cluster 7 is endothelial and cluster 4 is mesenchymal, this trajectory suggests an endothelial-to-mesenchymal transition.

Second, SCENIC analysis identifies GATA6 as a regulator of cluster 4 genes. Additionally, the GATA6 regulon shows distinct localization to the formed valves in fetal cells (new data added to Fig 2F). Together these findings support the notion that GATA6 regulates a gene program specific to the cell populations that will give rise to the valves and that these genes remain selectively expressed in valve cells once the arterial valves have formed.

While we cannot exclude the possibility that GATA6 contributes to EMT, we observe that GATA6 expression levels are highest in cluster 4 (post-EMT) cells, suggesting that its activation may be a consequence of the transition rather than its initiating cause (now shown in Fig S2D).

For validation using FISH, please see response to point 6 below

> *(6) I found it curious that the ST section was used to validate MECOM expression (Figure 2I), while ST had not yet been introduced at this point in the manuscript. Validation using FISH would have been a more appropriate approach.*

Thank you for drawing attention to this discrepancy. Spatial transcriptomics is now introduced before MECOM analysis, in the Results section pertaining to Figure 2F

"...spatial transcriptomic analysis of a later stage (12pcw) OFT shows that GATA6 regulon is mainly restricted to the aortic and pulmonary valves (Fig 2F)".

With regard to this and the above comment concerning FISH, while RNA FISH/RNAscope would provide an additional orthogonal approach, the Visium-based spatial transcriptomics platform directly measures MECOM transcripts in tissue sections and, in our view, represents an appropriate and sufficiently sensitive method for validating its spatial distribution in the human OFT. We have therefore relied on the spatial transcriptomics dataset to confirm and validate gene expression patterns, rather than performing additional FISH experiments. We now explicitly state that this approach serves as an independent in situ validation of gene expression, including MECOM.

> *(7) "Spatial resolution of mesenchymal nuclei in the OFT" section: It is unclear which cluster the authors are referring to in this section.*

As mentioned in the text, we "mapped the five fetal mesenchymal clusters to distinct structures in the OFT" and used distinctive markers to confirm spatial assignments.

> *(8) The authors should justify their choice to use Cell2location instead of a deconvolution method.*

We selected cell2location because it provides a probabilistic, hierarchical Bayesian framework that explicitly models technical variability across both single-cell reference data and spatial transcriptomics platforms. Rather than relying on predefined marker genes or simple linear regression, cell2location leverages the full transcriptomic profile of reference single-cell data and incorporates a factor analysis-based framework to model shared transcriptional signatures and latent structure across cell types. This approach improves discrimination between closely related cell states and reduces sensitivity to gene selection bias. Additionally, the probabilistic formulation yields uncertainty estimates for inferred cell abundances, enhancing interpretability and statistical rigor. Together, these features make cell2location particularly well suited for resolving complex cellular composition in our fetal human tissue spatial transcriptomics data.

> *(9) Figure 3: Cluster 9 is identified as endothelial, yet it includes markers such as MYH11 among its top genes, a gene more commonly associated with cells at the base of the aorta. This raises questions about the accuracy of the cluster annotation.*

We could not find the definition of cluster 9 as endothelial to which the reviewer refers to. In Fig 3, both in the result text and in the figure legend, cluster 9 is identified as smooth muscle, which is consistent with *MYH11* expression. The endothelial cluster is shown in Fig S3C.

> *(10) The approach used to trace embryonic signatures in adult cells, based on overlap with the top 100 genes in embryonic clusters, relies largely on gene expression similarity, without incorporating lineage inference tools such as RNA velocity or pseudotime analysis. This limits the ability to distinguish true developmental relationships from shared functional programs. I believe that the use of aggregated adult samples may mask individual variability. Validation in separate samples (AV1 and AV3) lacks statistical rigor. The observed lower expression of embryonic genes in adult cells further complicates interpretation, raising the possibility that these signatures reflect residual expression rather than persistent lineage markers.*

We thank the reviewer for the opportunity to clarify our approach.

We fully agree that tools such as RNA velocity and pseudotime are powerful for capturing short-term dynamic transcriptional changes and inferring lineage trajectories within continuous developmental processes. Indeed, we applied RNA velocity and identified a

transition between clusters 7 and 4 in embryonic cells (Fig 2). However, as noted in the Results section, “trajectory inference methods failed to establish lineage relationships between embryonic and fetal populations”. These methods assume temporal continuity and comparable transcriptional kinetics between cells. When comparing samples separated by large developmental intervals (e.g., embryonic versus adult tissues), these assumptions do not hold: RNA velocity vectors become unreliable and may even yield biologically meaningless directions. Therefore, rather than forcing a continuous trajectory across temporally distant datasets, we employed an anchoring approach designed to identify conserved transcriptional programs and potential lineage correspondences between embryonic and adult cell types.

To address the concern about individual variability, we performed analyses both on aggregated adult samples and on individual replicates (AV1 and AV3). The results were highly consistent across both levels of analysis, and statistical significance was supported by very low p-values, indicating that the observed patterns are robust and reproducible. We therefore believe our analysis in independent samples is statistically sound.

Finally, we agree that adult cells display lower expression of embryonic genes, and we acknowledge that these signatures may represent residual rather than persistent expression. This observation aligns with our intended interpretation: our goal was not to demonstrate enduring embryonic marker expression, but to highlight that adult cells retain transcriptional traces that connect them to their developmental origins.

> ***Reviewer #3 (Recommendations for the authors):***
>
> *(1) Please clarify if MEIS1, JAG1, ROR1, PRDM6 have been previously implicated in neural crest cell development. Are these then new potential regulators of neural crest cells? The same applies to SOX6 for the mesodermal population.*

The main reason for selecting these genes (MEIS1, JAG1, ROR1, and PRDM6 in cluster 20, and SOX6 in cluster 4) is that they serve as distinctive markers of specific embryonic clusters. Because their expression remains restricted at later developmental stages, they allow reliable tracing of bona fide descendant cells originating from cluster 20 and cluster 4 into fetal and adult tissues. Importantly, MEIS1, JAG1, ROR1, and PRDM6 were not chosen as new potential regulators of neural crest (NC) cells, but rather because their expression is enriched in cluster 20 and remains restricted at later developmental stages, allowing reliable tracing of bona fide descendant cells originating from cluster 20. Since cluster 20 is, based on transcriptional profiles, the embryonic mesenchymal cluster most closely related to the NC lineage, these markers enable lineage tracing of NC-descendent cells. Nonetheless, these genes have all been linked to neural crest biology, either through known functional roles or through specific expression patterns associated with NC development.

Similarly, SOX6 was selected for its restricted expression in cluster 4, a pattern that is preserved in its descendant populations, making it a suitable marker for tracking the mesoderm-derived lineage.

> *(2) Please comment in the text whether any regional transcriptional differences (rather than cell type differences) were detected between the aortic and pulmonary regions.*

We have added the following text to the result section related to Fig 3: “No molecular differences or distinguishing markers were identified between the aortic and pulmonary valves.”

> *(3) There appear to be no myocardial cells in the adult valve tissue - the authors could discuss what the fate of myocardium is in the embryonic OFT. Are they only looking at a subset of derivatives of the embryonic OFT?*

Our adult dataset represents the aortic valve complex and adjacent arterial root tissue (a subset of outflow tract derivatives) rather than the entire outflow tract (this has now been specified in the title). Spatial transcriptomic analysis identified myocardial gene expression within the ventricular and outflow tract walls at CS16-19, but not within the valve leaflet cluster (Queen et al., 2023). This is consistent with previous observations that myocardium contributes to the arterial root and supports early cushion formation, but does not persist in mature valve tissue, which becomes predominantly fibrous and populated by valve interstitial cells. This explanation has been added to the analysis of cell populations in the valves.

> *(4) Please equate Carnegie stages 13-23 to embryonic days or weeks of gestation in the first paragraph to help the general reader.*

We have added the suggested clarification and noted that this period spans four weeks of human development, rather than the three weeks previously indicated. The text has been updated accordingly.

> *(5) I suggest rewriting the first sentence of the introduction using the plural, as there are many different types of CHD.*

The sentence has been changed accordingly.

> *(6) It would be helpful to add the persistence of embryonic signatures into adult valve cell types in Figure 4E.*

We thank the reviewer for this helpful suggestion. To address this point, we have now added an analysis of the persistence of embryonic signatures in adult valve cell types to Figure 4E. Specifically, we selected 10 representative genes from the 100-gene embryonic signature lists of cluster 4 and cluster 20 and projected their expression onto the t-SNE shown in Figure 4E. The combined (module) expression of these 10 genes is now shown in Figure S6E, and the expression of the individual genes is presented in the newly added Figure S7.

We would like to clarify that our statistical framework identifies potential descendant populations based on significant enrichment of an embryonic gene signature. Therefore, individual embryonic genes are not necessarily expected to be expressed exclusively or uniformly within a single adult population.

> *(7) Please explain how the 2-dimensional plot in 2J relates to the other plots.*

The plot originally shown in Fig 2J (now Fig 2K) was generated by applying RNA velocity exclusively to CS16-17 nuclei. Developmental nuclei (excluding adult samples) were subclustered as shown in Fig S2AB, resulting in the 5 clusters of embryonic nuclei analysed in Fig 2J: cardiac muscle (2, 17), endothelial (7), and mesenchymal (4, 20).

https://doi.org/10.7554/eLife.107748.2.sa0