A cell atlas of the developing human outflow tract of the heart and its adult 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[1], Magnus Rattray[1], Simon D Bamforth[3,4] and Nicoletta Bobola[1,4]

[1]Faculty of Biology, Medicine and Health, University of Manchester, UK

[2]Translational and Clinical Research Institute, Newcastle University.

[3]Newcastle University Biosciences Institute, Faculty of Medical Sciences, Newcastle.

* equal contribution

[4] Correspondence: Nicoletta.Bobola@manchester.ac.uk; Simon.Bamforth@newcastle.ac.uk

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 disease 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. Our findings define the cellular and molecular signatures of the OFT and its distinct cell lineages, which is critical for understanding congenital heart defects and developing cardiac tissue for regenerative medicine.

Introduction

Congenital heart disease (CHD) is a major birth defect 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 3-week period (Carnegie Stage (CS)13 – CS23). 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, 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] [8, 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 [10-21]. 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 post conception week (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 outflow tracts 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 [22].

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 7AAD dye. Nuclei were FACS sorted using a 100 µm diameter nozzle and a sheath pressure of 20PSI as per 10x protocols (document CG000375). 7-aminoactinomycin D (7-AAD) 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

The sequence files generated from the sequencer were processed using 10x Genomics custom pipeline cellranger, cellranger v.3.1.0 for developmental samples and cellranger v. 6.1.2 for adult samples. This pipeline generated the fastq files which were aligned to the hg38 custom genome using cellranger default parameters. For developmental data we used a premRNA_patched gtf file to count the intronic reads; for the adult data we enabled the include-introns option during cellranger count runs. The pipeline then identified the barcodes associated with nuclei and counted UMIs mapped to each nucleus. Cellranger uses STAR aligner to align reads to the genome, discarding counts mapping to multiple loci during counting.

Filtering

We removed low-quality nuclei 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. We adjusted the threshold of these three parameters individually for each sample to retain only the 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, we conducted the analysis in multiple stages. We first analyzed each individual sample separately. We then combined all the developmental time-points (embryonic and fetal) together, and the adult time-points together. Finally, we combined the developmental time points and the adult time-points using scanpy [23].This final aggregation gave us a total of 30,166 nuclei for downstream analysis.

Normalisation and classification of cell-cycle phase

We used the deconvolution-based method [24] to normalize the raw counts in individual sample analysis. For scanpy integration we used scanpy's default normalization method that normalizes each cell b total count over all genes to ensure that each cell has the same total count after normalization. We used the cyclone method [25] to calculate the score of G1 phase and G2M phase for each nucleus.

Visualisation & Clustering

For visualizing individual samples, we first identify the Highly Variable Genes (HVGs). We decomposed the variance of each gene expression values into technical and biological components; HVG are 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 their putative clusters using the dynamic tree cut method [26]. Dynamic tree cut method identifies the branch cutting point of a dendrogram dynamically and combines the advantage of both hierarchical and K-medoid clustering approach. This method identified 7 clusters in the aggregated population. For all datasets combined, we used scanpy's highly_variable_genes (scanpy v. 1.9.5). Scanpy's highly_variable_genes uses dispersion-based method under the hood where the genes are first grouped in to a bin based on mean expression and then the normalized dispersion is calculated by scaling with the mean and standard deviation of the genes in that specific bin. To cluster nuclei in scanpy aggregated data, we used Louvain clustering in scanpy with a resolution of 0.4.

Identification of marker genes

For individual sample analysis, we used the findMarkers function from scran 1.26.2 package that applies one sided t-test to identify marker genes for each cluster [27]. The function also returns FDR values for multiple testing correction using Benjamini-Hochberg procedure. These marker genes were then used to annotate the cell types of a cluster. For the scanpy aggregated data, we used scanpy's rank_genes_groups 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 using DAVID [28, 29] at top 100 genes, with P values limited to < 0.05.

RNA-velocity

We applied scVelo (v. 0.2.4) to identify transient cellular dynamics in the embryonic stage [30]. To calculate the spliced vs un-spliced ratio for each gene we used RNA-velocity's command line tool, velocyto 10x with default parameters [31] and used scVelo to compute velocity based on the spliced vs un-spliced ratio.

Spatial Transcriptomics

We used Spaceranger v.1.3.1 to pre-process the Visium slides. After filtering the lower count spots and genes, we normalized the data using scanpy's normalize_total function. We clustered the spots using leiden clustering and identified the marker genes for each of the spots, and spatially variable genes using Moran's notes [32]. As each of the Visium spots has more than one cell we used cell2location (v. 0.1.3) to deconvolute the spots [33]. 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. We then trained the model 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 similar expression pattern to a given gene, we identified all the spots with expression value > 1 for the given gene. We removed every gene that is not expressed in at least 50% of these spots. Any remaining genes expressed in > 25% of the spots where expression of the given gene is zero was also filtered out, resulting in a list of genes with similar expression to our gene of interest.

Lineage tracing algorithm

We developed a novel lineage tracing algorithm to infer relationships between embryo and adult cells. First, we characterized distinctive signatures in embryonic progenitors. We used Wilcoxon test to identify the top 100 differentially expressed (DE) genes between our chosen progenitor populations. Next, we identified the top 5000 marker genes for each adult cluster, by comparing each cluster with the rest of the dataset. We counted the occurrence of embryonic signature genes in the top 5000 marker genes of adult clusters; adult clusters with top hits were identified as related to the original embryonic lineage. Statistical significance was calculated using a hypergeometric test.

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 [34] 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 combination of *PDGFRA* [35] and *PDGFRB* [36], 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* [37], a long non-coding RNA (lncRNA), which do not distinguish cell types but discriminate adult from developmental tissues (Fig S1D-G).

Spatial resolution of mesenchymal nuclei in the OFT

We focused our attention on mesenchymal cells, which build critical structures in the OFT, the separation of the aorta from the pulmonary artery (Fig 2A, C, E) and the semilunar valves at the base of the aorta and pulmonary artery (Fig 2B, D, F). To dissect the heterogeneity within mesenchymal cells and identify lineage subtypes in the developing OFT, we performed nuclei sub-clustering of the embryonic and fetal datasets, along with tSNE visualization (Fig S2AB). Using this approach, we obtained a total of seven mesenchymal clusters, two embryonic and five fetal (Fig 2GH). In contrast to embryonic nuclei, which do not express cell-type markers, fetal nuclei largely express *DCN* (a fibroblast marker), with the exception of cluster 9 which contains *MYH11*-positive smooth muscle cells (Fig 2I, see also Fig S1C).

To visualize the distribution of mesenchymal clusters 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. 2J). We used Cell2location [33] 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 S2C). Mesenchymal cells largely distributed within and around the vessels, and also mapped to the valves of the pulmonary artery and the aorta (Fig S2D). Endothelial cells were primarily located in the aortic valves (Fig S2E). Immune cells were scattered around the tissue (Fig S2F), while neural cells were concentrated in a spot of the atria (Fig S2G). Next, we mapped the five fetal mesenchymal clusters to distinct structures in the OFT (Fig 2K) and used distinctive markers to confirm spatial assignments. Clusters 3 and 6 (Fig 2L-L^{I}) map to the arterial outer walls and the septum between the aorta and pulmonary arteries and largely colocalise with *DCN*-positive cells (Fig 2L^{II}). These two clusters largely overlap. Cluster 9 (Fig 2M) coincides with *MYH11* expression (Fig 2M^{I}): *MYH11* is a terminal marker of smooth

muscle differentiation [38], which identifies the aortic smooth muscle layer. Cluster 12 (Fig 2N) concentrates to both aortic and pulmonary valves and co-localize with *HAPLN1,* encoding the extracellular matrix cross-linking protein Hyaluronan and Proteoglycan Link Protein 1 (Fig 2N[I]), a marker of the semilunar valves [39]. Finally, the smallest cluster (15) is restricted to the cardiac ventricle and is *DCN*-negative, therefore was eliminated from our further analysis (Fig S2H). 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://spatialtranscriptomics-uom.bmh.manchester.ac.uk/.

Developmental trajectories of mesenchymal cells in the developing OFT

At CS16-17, there are two clusters of undifferentiated mesenchymal cells. Our objective was to track their developmental trajectories to the 12pcw differentiated mesenchymal populations: fibroblasts, smooth muscle cells, and valvular interstitial cells, which occupy distinct spatial locations in the OFT (Fig 2). Connecting mesenchymal embryonic progenitors to differentiated fetal cell populations is challenging, because the embryonic nuclei are yet to express the molecular markers of differentiated lineage descendants (Fig 2I). As a first step, we characterized embryonic mesenchymal nuclei and used trajectory inference to explore their relationship with the nuclei of other cell types at the same stage. RNA velocity predicts the future state of individual cells by distinguishing between un-spliced and spliced mRNAs [31]. Because of its timescale of hours, we exclusively applied RNA velocity to embryonic samples. At the embryonic stage, we detected a movement of nuclei from cluster 7 (endothelial-like cells) specifically towards cluster 4 (mesenchymal cells) (Fig 3A). Cluster 7 is enriched in genes associated with 'aortic and pulmonary valve morphogenesis', as well as 'epithelial to mesenchymal transition' (Fig S3A; Table S3). These processes are initiated in the endocardium, suggesting that embryonic cluster 7 include an endocardial population. These observations indicate that RNA velocity detects the epithelial to mesenchymal transition (EMT) of endothelial cells in the endocardium, the cell layer adjacent to the myocardium. In mouse, this generates a mobile group of mesenchymal cells that construct the semilunar valves [40, 41]. Consistently, cluster 7 and cluster 4 populations share a similar expression signature (Fig 3B), which includes high expression of *HAPLN1*, expressed in the endocardial lining of the heart, and in the cardiac valves including the developing semilunar valves in the OFT.

Next, we tracked the developmental paths of mesenchymal cell populations and followed the descendant cells of each embryonic cluster. Trajectory inference [42] failed to detect lineage relationship between embryonic and fetal populations. We therefore used gene module scores. The rationale is that, if any distinctive developmental signature of embryonic clusters

was passed down to fetal nuclei, it would aid in tracing the trajectories of mesenchymal cell populations. We first performed pairwise differential gene expression of embryonic clusters to identify distinct developmental signatures of embryonic mesenchymal subtypes (Fig 3C). Cluster 4, which derives at least partly from endocardial cells, is enriched in cardiac-like markers (Fig 3C) and linked to 'aortic valve and endocardial cushion morphogenesis' (Fig S3B). Cluster 20 is largely associated with neural-like GOs (Fig S3C) and enriched in neural crest markers *HOXA3/B3* and *PDGFRB* (Fig 3BC). 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 transcription factors, which segregate lineages, to build two distinctive gene modules. Namely, the '*TBX20 SOX6 GATA4 PRRX1*' module is highly expressed in embryonic cluster 4, while '*MEIS1 JAG1 ROR1 PRDM6*' is highly enriched in embryonic cluster 20. This set of eight selected developmental genes was sufficient to segregate embryonic nuclei in two distinct subtypes (Fig 3D), named group 1 and group 2. Group 1 is entirely made of cluster 4 nuclei, while group 2 largely contains cluster 20 nuclei, with some contribution from cluster 4. Next, we asked if embryonic gene modules are inherited by fetal nuclei. Indeed, we found that expression of our embryonic gene modules largely separate developmental nuclei into two distinct lineages (Fig 3E): a 'blue' lineage, largely made of clusters 4-3-12, with higher expression of module 1 signature (*SOX6, TBX20*), and a 'red' lineage, containing clusters 20-6-9, with higher levels of module 2 signature (*MEIS1, PRDM6*). We combined embryonic gene modules with differentiated cell markers and obtained three distinct groups of 12pcw mesenchymal nuclei (Fig 3F). Module 1 embryonic signature is inherited by fibroblasts and valves, encompassing cluster 3 and 12 respectively (Fig 3E), suggesting a common embryonic progenitor for these cells. As module 1 defines the secondary heart field -derived embryonic cluster 4, fetal group 1 nuclei derive from the secondary heart field. Module 2 signature is inherited by two distinct cell types, smooth muscle cells (cluster 6) and fibroblasts (cluster 9), which are segregated by the expression of cell type-specific markers, *MYH11* (group 2) and *DCN* (group 3). Module 2 largely identifies cluster 20, which is enriched in neural crest markers, with a small contribution from cluster 4, suggesting fetal group 2-3 nuclei derive from the neural crest and secondary heart field. Consistently, the expression of cardiac neural crest markers *HOXA3/B3* is almost exclusively confined to 'group 2-3' fetal nuclei, suggesting that 'group 2-3' nuclei (largely clusters 6 and 9) partly derive from neural crest progenitors. This is consistent with smooth muscle cells in the aortic root deriving from neural crest and secondary heart field, as observed in mouse [43]. To confirm lineage relationships, we also developed a robust method to trace embryonic signatures in fetal cells. We largely expanded the gene module repertoires to include the top one hundred most distinctive genes of cluster 4 and 20 mesenchymal progenitors (Table S4) and looked for the populations of fetal nuclei that retained expression

of most of these genes. When applied to the entire fetal dataset (cardiac, endothelial and mesenchymal clusters in Fig S2), our method confirmed the same lineage relationship between embryonic and fetal mesenchymal nuclei already identified by gene modules (Fig 3G). In sum, our analysis indicates that the two spatially distinct mesenchymal populations in the fetal OFT, smooth muscle cells and valvular fibroblasts (Fig 2M-M^{I}; N-N^{I}) derive from separate embryonic populations (i.e. cluster 20 for smooth muscle cells and cluster 4 for valvular fibroblasts). In contrast, the *DCN*-positive, *HAPLN*-negative fibroblast cells that spatially intermingle in OFT tissues, derive from both cluster 4 and 20 embryonic progenitors (Fig S3D).

OFT defects

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 [44]. 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. *JAG1* mutations cause TOF [45]; *JAG1* transcripts largely concentrate around the aorta, in the semilunar valves and the septum (Fig 4B-B^{I}). Mutations in *GATA* family members affects OFT formation in different ways, leading to DORV (*GATA5*) [46], TOF (*GATA4, GATA6*) [47, 48] and CAT (*GATA6*) [49]. Different from *GATA5*, which shows a restricted expression to the vessels (Fig 4C, C^{I}), the distribution of *GATA6* and *GATA4* transcripts is broader. *GATA6* is detected in the vessels, the septum and cardiac tissue (Fig S4A), while *GATA4* transcripts are more prominent in cardiac tissues (Fig S4B). Mutations in *GATA4-5-6* also causes bicuspid aortic valves (BAV), a less severe condition characterized by two valves instead of three [50]. Similar to *GATA5*, *GATA6* is highly expressed in the valves, marked by *HAPLN1* (Fig S4C). As a final example, mutations in *NR2F2* cause DORV and TOF [51]. 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 4D, 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 4E). 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 [52, 53]. In addition, variants in *FOXC1* were recently identified in patients with conotruncal heart defects [10].

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) [54] and immune (2 clusters, *ITGAM*-positive macrophages and *IKFZ1*-positive dendritic cells) [55, 56] (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* [57], endothelial clusters are distinguished by high levels of *ADAMTS9,* which encodes for a metalloproteinase implicated in aortic valve anomalies in mouse [58] (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 5E). 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 [59]. 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 S5C), with *VIM* and *VCAN* displaying a broader expression. Differently from cluster 5, clusters 4 and 7 express high levels of *CCL2*, an inflammatory

cytokine [60], suggesting nuclei in these clusters correspond to activated fibroblasts. Finally, cluster 9 displays enrichment in *DLC1* and *TLN2*, implicated in cytoskeletal changes [61, 62], and *SPARC* (osteonectin), implicated in valve calcification [63]. 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 5E; Fig S5D). 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.

Persistence of embryonic signatures in adult cells

Whilst differentiated, 12pcw nuclei maintain robust embryonic signatures (Fig 3EF). 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. We observed expression of embryonic gene modules (see Fig 3D) in distinct adult clusters. Embryonic module 2 genes were highly expressed in MYH11-positive cluster 1, while combinatorial expression of embryonic module 1 genes was distinctive of HAPLN-positive VIC in cluster 4 (Fig S6A). To address this more systematically, 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) (Fig 6A; Table S4), we looked for populations of adult nuclei enriched in the expression of most of these genes (Fig 6A). 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 6BC), while cluster 1 displays a significant match with embryonic cluster 20 (red) (Fig 6B, D). 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 3FG). Valvular fibroblasts (clusters 4,7) derive from the embryonic population in cluster 4, which is also linked to fetal valvular fibroblasts (Fig 3FG). This result is significant because our embryonic signatures derive from undifferentiated cells which are yet to express obvious differentiation markers, indicating that adult cells retain their ancestral embryonic make up. 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 S6BC); these clusters contain the majority of adult cluster 4 aggregate nuclei (Fig S6DE). 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 6EF; Fig S6F).

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 two sets of CS16/17 mesenchymal nuclei do not express obvious differentiation markers, but display distinctive neural crest and secondary heart field expression signatures. Conversely, mesenchymal nuclei at 12pcw express cell type markers and are clustered into four subgroups, which largely correspond to distinct cell populations (fibroblast-like cells, smooth muscle cells, valvular interstitial cells) and are distributed in different areas of the OFT. Using gene modules and a new lineage tracing tool, we have traced each of the four spatially separated mesenchymal populations in the fetal OFT back to their embryonic precursors, assigning valvular fibroblasts and smooth muscle cells to 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 [40, 41]. 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 [64]. Finally, both embryonic mesenchymal populations give rise to non-valvular fibroblast cells (*DCN*$^{+}$, *HAPLN1*$^{-}$*)*.

We find that distinctive embryonic signatures persist into adulthood. Previous studies have described that reactivation of fetal programs in adult heart disease [65, 66]. Crucially, however, all our adult samples are derived from healthy individuals. The phenomenon we describe seems to be a pervasive characteristic of healthy adult cells, rather than being specific to the context of heart disease or other pathological conditions. Single cell-based analyses have uncovered a high level of cell heterogeneity. Our results suggest that the inheritance of 'embryonic memories' by adult cells can be a main contributing factor to same cell type heterogeneity. In support of this, there is evidence of organ-specific developmental signatures persisting in adult fibroblasts [67]. Considering the human context, these findings offer insights into diverse susceptibility to disease and can potentially contribute to our 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 [42]. However, the significant alterations in expression profiles as cells transition from embryonic to fetal and adult cell types pose a challenge to comparing and linking cells over prolonged periods, based on their global expression profiles. The persistence of embryonic signatures into adulthood can be used as a label to retrospectively infer cellular relationships in complex tissues and provides the newfound capability to trace descendant cells. We have developed a lineage tracing tool that has the potential to substantiate extrapolated knowledge derived from model organisms to humans, as well as to uncover previously unknown, human-specific lineage relationships. We can envision its application to infer cellular relationships in different contexts, particularly for cells that are distant in terms of time (across development) or phenotype (as in 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 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 [65, 66].

Acknowledgements

We thank Andy Hayes and the other members of the Genomic Technologies Core 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) (http: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.

Data availability:

All data described have been deposited in ArrayExpress with accession numbers listed in table S3 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

Code availability:

The code used in Fig. 6 will be made publicly available through GitHub upon publication.

References

1. Hoffman, J.I. & Kaplan, S. The incidence of congenital heart disease. *J Am Coll Cardiol* **39**, 1890-1900 (2002).
2. Stefanovic, S., Etchevers, H.C. & Zaffran, S. Outflow Tract Formation-Embryonic Origins of Conotruncal Congenital Heart Disease. *J Cardiovasc Dev Dis* **8** (2021).
3. Thom, T. *et al.* Heart disease and stroke statistics--2006 update: a report from the American Heart Association Statistics Committee and Stroke Statistics Subcommittee. *Circulation* **113**, e85-151 (2006).
4. Anderson, R.H. *et al.* Normal and abnormal development of the intrapericardial arterial trunks in humans and mice. *Cardiovasc Res* **95**, 108-115 (2012).
5. Anderson, R. *et al.* The Reappraisal of Normal and Abnormal Cardiac Development, in *Hemodynamics and Cardiology Neonatology Questions and Controversies*. (eds. K. C & S. I) 391-414 (Elsevier, 2012).
6. Kelly, R.G. The second heart field. *Curr Top Dev Biol* **100**, 33-65 (2012).

7. Kirby, M.L. & Hutson, M.R. Factors controlling cardiac neural crest cell migration. *Cell Adh Migr* **4**, 609-621 (2010).
8. Waldo, K., Miyagawa-Tomita, S., Kumiski, D. & Kirby, M.L. 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 (1998).
9. Jiang, X., Rowitch, D.H., Soriano, P., McMahon, A.P. & Sucov, H.M. Fate of the mammalian cardiac neural crest. *Development* **127**, 1607-1616 (2000).
10. Wei, W. *et al.* Variants in FOXC1 and FOXC2 identified in patients with conotruncal heart defects. *Genomics* **116**, 110840 (2024).
11. Xu, Y. *et al.* A single-cell transcriptome atlas profiles early organogenesis in human embryos. *Nat Cell Biol* **25**, 604-615 (2023).
12. Asp, M. *et al.* A Spatiotemporal Organ-Wide Gene Expression and Cell Atlas of the Developing Human Heart. *Cell* **179**, 1647-1660.e1619 (2019).
13. Sahara, M. *et al.* Population and Single-Cell Analysis of Human Cardiogenesis Reveals Unique LGR5 Ventricular Progenitors in Embryonic Outflow Tract. *Developmental Cell* **48**, 475-490.e477 (2019).
14. Cui, Y. *et al.* Single-Cell Transcriptome Analysis Maps the Developmental Track of the Human Heart. *Cell Rep* **26**, 1934-1950.e1935 (2019).
15. Knight-Schrijver, V.R. *et al.* A single-cell comparison of adult and fetal human epicardium defines the age-associated changes in epicardial activity. *Nature Cardiovascular Research* **1**, 1215-1229 (2022).
16. Cao, J. *et al.* A human cell atlas of fetal gene expression. *Science* **370** (2020).
17. Suryawanshi, H. *et al.* Cell atlas of the foetal human heart and implications for autoimmune-mediated congenital heart block. *Cardiovasc Res* **116**, 1446-1457 (2020).
18. Lahm, H. *et al.* Congenital heart disease risk loci identified by genome-wide association study in European patients. *J Clin Invest* **131** (2021).
19. Litviňuková, M. *et al.* Cells of the adult human heart. *Nature* **588**, 466-472 (2020).
20. Koenig, A.L. *et al.* Single-cell transcriptomics reveals cell-type-specific diversification in human heart failure. *Nature Cardiovascular Research* **1**, 263-280 (2022).
21. Tucker, N.R. *et al.* Transcriptional and Cellular Diversity of the Human Heart. *Circulation* **142**, 466-482 (2020).
22. Anderson, R.H. & Bamforth, S.D. Morphogenesis of the Mammalian Aortic Arch Arteries. *Frontiers in Cell and Developmental Biology* **10** (2022).
23. Wolf, F.A., Angerer, P. & Theis, F.J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biology* **19**, 15 (2018).
24. Lun, A., Bach, K. & Marioni, J. Pooling across cells to normalize single-cell RNA sequencing data with many zero counts. *Genome Biology* **17** (2016).
25. Scialdone, A. *et al.* Computational assignment of cell-cycle stage from single-cell transcriptome data. *Methods* **85**, 54-61 (2015).
26. Langfelder, P., Zhang, B. & Horvath, S. Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. *Bioinformatics* **24**, 719-720 (2007).
27. Lun, A.T., McCarthy, D.J. & Marioni, J.C. A step-by-step workflow for low-level analysis of single-cell RNA-seq data with Bioconductor. *F1000Res* **5**, 2122 (2016).
28. Huang, D.W., Sherman, B.T. & Lempicki, R.A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols* **4**, 44-57 (2009).
29. Sherman, B.T. *et al.* DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). *Nucleic Acids Res* **50**, W216-w221 (2022).
30. Bergen, V., Lange, M., Peidli, S., Wolf, F.A. & Theis, F.J. Generalizing RNA velocity to transient cell states through dynamical modeling. *Nat Biotechnol* **38**, 1408-1414 (2020).
31. La Manno, G. *et al.* RNA velocity of single cells. *Nature* **560**, 494-498 (2018).

32. Moran, P.A. Notes on continuous stochastic phenomena. *Biometrika* **37**, 17-23 (1950).
33. Kleshchevnikov, V. *et al.* Cell2location maps fine-grained cell types in spatial transcriptomics. *Nature Biotechnology* **40**, 661-671 (2022).
34. Korsunsky, I. *et al.* Fast, sensitive and accurate integration of single-cell data with Harmony. *Nature Methods* **16**, 1289-1296 (2019).
35. Farahani, R.M. & Xaymardan, M. Platelet-Derived Growth Factor Receptor Alpha as a Marker of Mesenchymal Stem Cells in Development and Stem Cell Biology. *Stem Cells Int* **2015**, 362753 (2015).
36. Wang, S. *et al.* 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 (2018).
37. Ge, Z. *et al.* 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 (2022).
38. Babij, P., Kelly, C. & Periasamy, M. 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 (1991).
39. Fang, M., Alfieri, C.M., Hulin, A., Conway, S.J. & Yutzey, K.E. Loss of β-catenin promotes chondrogenic differentiation of aortic valve interstitial cells. *Arterioscler Thromb Vasc Biol* **34**, 2601-2608 (2014).
40. Gittenberger-de Groot, A.C., Vrancken Peeters, M.-P.F., Mentink, M.M., Gourdie, R.G. & Poelmann, R.E. Epicardium-derived cells contribute a novel population to the myocardial wall and the atrioventricular cushions. *Circulation research* **82**, 1043-1052 (1998).
41. Kirby, M.L., Gale, T.F. & Stewart, D.E. Neural crest cells contribute to normal aorticopulmonary septation. *Science* **220**, 1059-1061 (1983).
42. Trapnell, C. *et al.* The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat Biotechnol* **32**, 381-386 (2014).
43. Sawada, H., Rateri, D.L., Moorleghen, J.J., Majesky, M.W. & Daugherty, A. 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 (2017).
44. Neeb, Z., Lajiness, J.D., Bolanis, E. & Conway, S.J. Cardiac outflow tract anomalies. *WIREs Developmental Biology* **2**, 499-530 (2013).
45. Eldadah, Z.A. *et al.* Familial Tetralogy of Fallot caused by mutation in the jagged1 gene. *Hum Mol Genet* **10**, 163-169 (2001).
46. Jiang, J.Q. *et al.* Prevalence and spectrum of GATA5 mutations associated with congenital heart disease. *Int J Cardiol* **165**, 570-573 (2013).
47. Maitra, M., Koenig, S.N., Srivastava, D. & Garg, V. Identification of GATA6 sequence variants in patients with congenital heart defects. *Pediatr Res* **68**, 281-285 (2010).
48. Tomita-Mitchell, A., Maslen, C.L., Morris, C.D., Garg, V. & Goldmuntz, E. GATA4 sequence variants in patients with congenital heart disease. *J Med Genet* **44**, 779-783 (2007).
49. Kodo, K. *et al.* GATA6 mutations cause human cardiac outflow tract defects by disrupting semaphorin-plexin signaling. *Proc Natl Acad Sci U S A* **106**, 13933-13938 (2009).
50. Yang, B. *et al.* Protein-altering and regulatory genetic variants near GATA4 implicated in bicuspid aortic valve. *Nat Commun* **8**, 15481 (2017).
51. Al Turki, S. *et al.* Rare variants in NR2F2 cause congenital heart defects in humans. *Am J Hum Genet* **94**, 574-585 (2014).
52. Sanchez, J. *et al.* Conditional inactivation of Foxc1 and Foxc2 in neural crest cells leads to cardiac abnormalities. *genesis* **58**, e23364 (2020).

53. Wang, Q., Lan, Y., Cho, E.-S., Maltby, K.M. & Jiang, R. Odd-skipped related 1 (Odd1) is an essential regulator of heart and urogenital development. *Developmental Biology* **288**, 582-594 (2005).
54. Giannotta, M., Trani, M. & Dejana, E. VE-cadherin and endothelial adherens junctions: active guardians of vascular integrity. *Dev Cell* **26**, 441-454 (2013).
55. Hulin, A. *et al.* Macrophage Transitions in Heart Valve Development and Myxomatous Valve Disease. *Arteriosclerosis, Thrombosis, and Vascular Biology* **38**, 636-644 (2018).
56. Collin, M. & Bigley, V. Human dendritic cell subsets: an update. *Immunology* **154**, 3-20 (2018).
57. Bach, T.L. *et al.* VE-Cadherin Mediates Endothelial Cell Capillary Tube Formation in Fibrin and Collagen Gels1. *Experimental Cell Research* **238**, 324-334 (1998).
58. Kern, C.B. *et al.* Reduced versican cleavage due to Adamts9 haploinsufficiency is associated with cardiac and aortic anomalies. *Matrix biology* **29**, 304-316 (2010).
59. Rutkovskiy, A. *et al.* Valve Interstitial Cells: The Key to Understanding the Pathophysiology of Heart Valve Calcification. *J Am Heart Assoc* **6** (2017).
60. Kumar, S.N. & Boss, J.M. Site A of the MCP-1 distal regulatory region functions as a transcriptional modulator through the transcription factor NF1. *Molecular immunology* **37**, 623-632 (2000).
61. Yuan, B.Z. *et al.* 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 (1998).
62. Monkley, S.J., Pritchard, C.A. & Critchley, D.R. Analysis of the Mammalian Talin2 Gene TLN2. *Biochemical and Biophysical Research Communications* **286**, 880-885 (2001).
63. Bradshaw, A.D., Graves, D.C., Motamed, K. & Sage, E.H. SPARC-null mice exhibit increased adiposity without significant differences in overall body weight. *Proc Natl Acad Sci U S A* **100**, 6045-6050 (2003).
64. Sawada, H., Rateri, D.L., Moorleghen, J.J., Majesky, M.W. & Daugherty, A. 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 (2017).
65. Dirkx, E., da Costa Martins, P.A. & De Windt, L.J. Regulation of fetal gene expression in heart failure. *Biochim Biophys Acta* **1832**, 2414-2424 (2013).
66. Oka, T., Xu, J. & Molkentin, J.D. Re-employment of developmental transcription factors in adult heart disease. *Semin Cell Dev Biol* **18**, 117-131 (2007).
67. Forte, E. *et al.* Adult mouse fibroblasts retain organ-specific transcriptomic identity. *eLife* **11**, e71008 (2022).

Figure 1

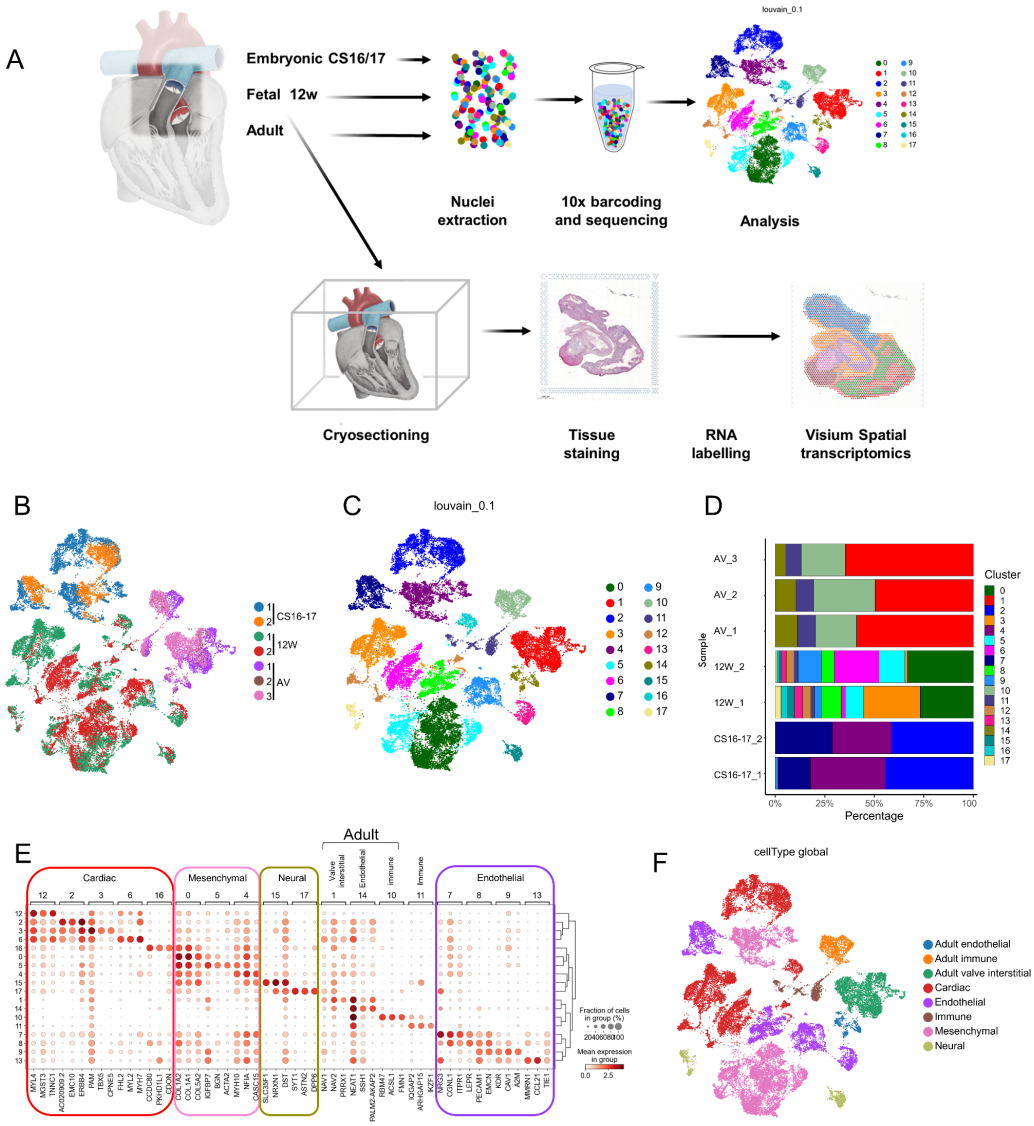

Figure 1. **The cellular landscape of the developing OFT and its adult derivatives**. A. Experimental schematics. Nuclei isolated from two embryonic (CS 16-17) and two fetal (12pcw) OFTs and from three adult aortic valves (AV) were analyzed by snRNA-seq. Four cryo-sections including the OFT region of 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 coloured 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 coloured 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 insterstitial), neural and immune. F. Cell types in E visualised by UMAP. Nuclei are coloured by cell type. See also Fig S1.

Figure 2

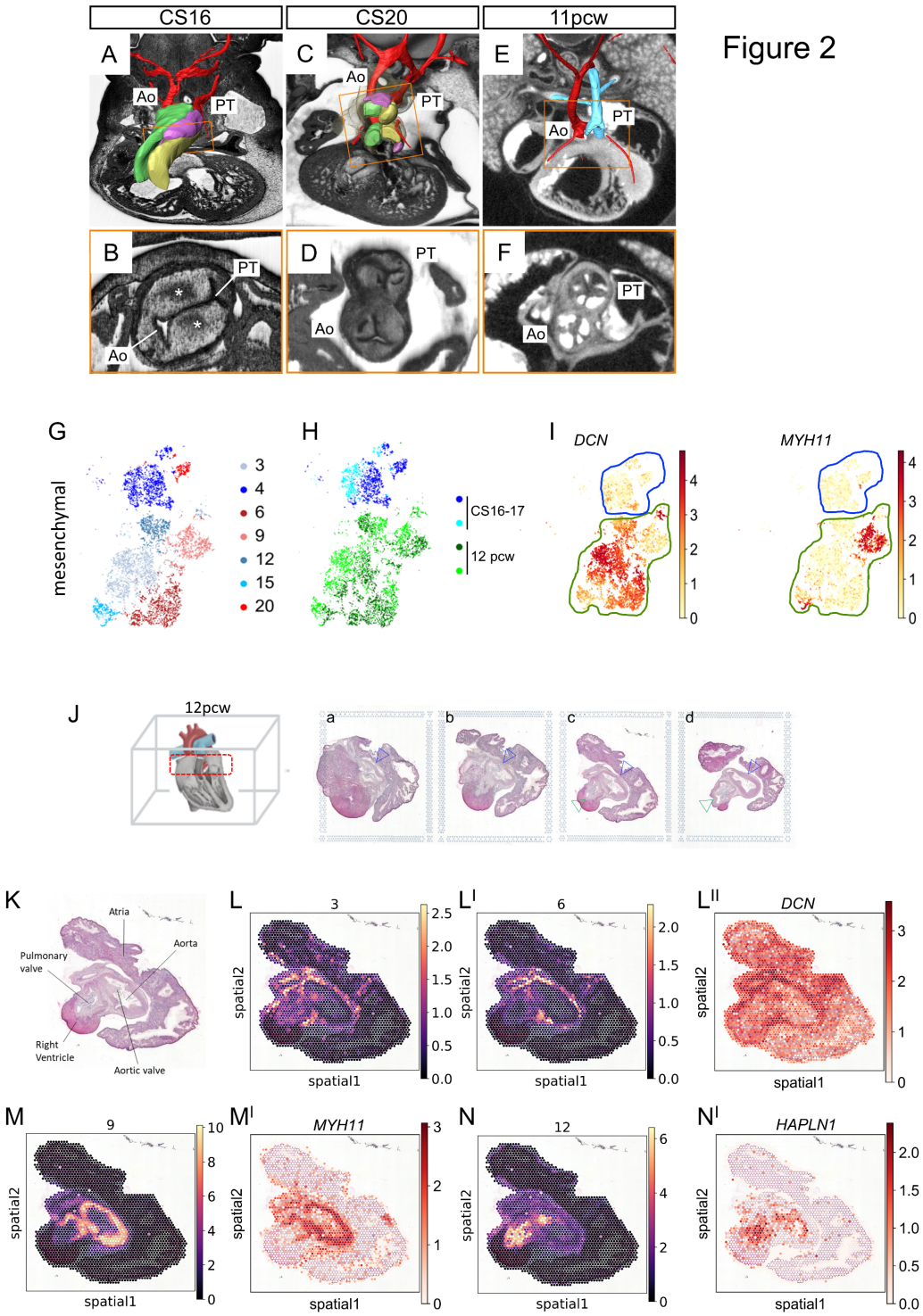
CS16
CS20
11pcw
A
B
C
D
E
F
Ao
PT
G
mesenchymal
3
4
6
9
12
15
20
H
CS16-17
12 pcw
I
DCN
MYH11
J
12pcw
a
b
c
d
K
Atria
Aorta
Pulmonary valve
Right Ventricle
Aortic valve
L
L^I
L^II
M
M^I
N
N^I
spatial1
spatial2
HAPLN1

Figure 2. Spatial distribution of mesenchymal clusters

A-F. Outflow tract 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 outflow tract has septated into the aorta (Ao) and pulmonary trunk (PT). The immature outflow tract 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, H. Mesenchymal cell clusters (G) and sample projection (H) of fetal and embryonic samples following subclustering, visualised on a two-dimensional tSNE. Nuclei are coloured by cluster (G) and sample (H). I. 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. J. 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. K. H&E section 'c' (panel J) annotated to show major structures. L, L^{I}, M, N. Spatial distribution of mesenchymal clusters (purple and yellow). L^{II}, M^{I}, N^{I}. Spatial distribution of lineage specific markers (white and red). L-L^{II}. Clusters 3 (L) and 6 (L^{I}) largely overlap with fibroblast lineage marker *DCN* (L^{II}). M-M^{I}. Cluster 9 (M) and smooth muscle lineage specific marker *MYH11* (M^{I}) map to the aortic walls as well as to the pulmonary artery. N-N^{I} Cluster 12 (N) and valve specific marker *HAPLN1* (N^{I}) are mainly found in the valves at the base of the aorta and pulmonary artery. See also Fig S3.

Figure 3

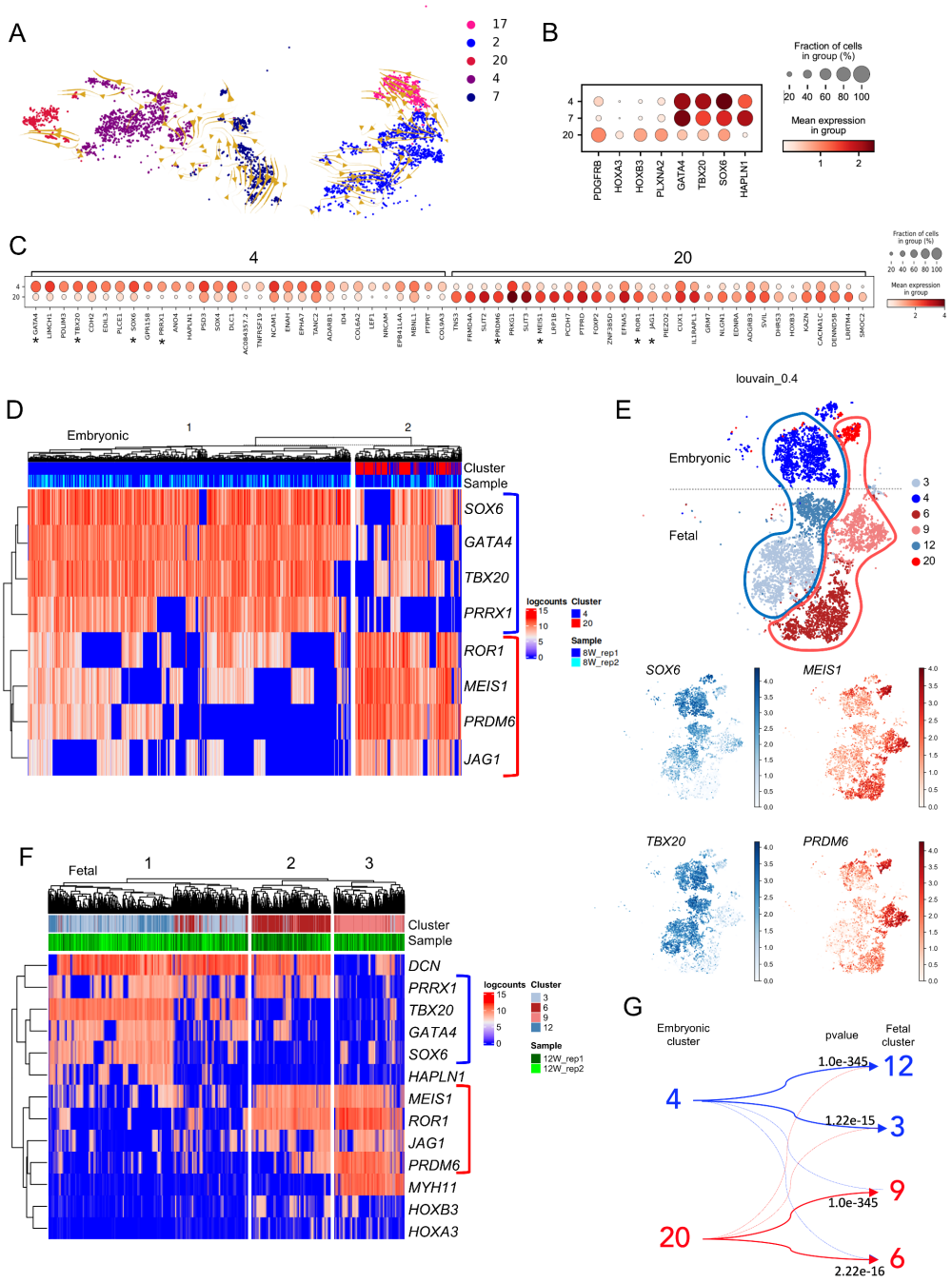
A
17
2
20
4
7
B
Fraction of cells in group (%)
20 40 60 80 100
Mean expression in group
1 2
4
7
20
PDGFRB
HOXA3
HOXB3
PLXNA2
GATA4
TBX20
SOX6
HAPLN1
C
4
20
D
Embryonic
1
2
Cluster
Sample
SOX6
GATA4
TBX20
PRRX1
ROR1
MEIS1
PRDM6
JAG1
logcounts
Cluster
4
20
Sample
8W_rep1
8W_rep2
E
louvain_0.4
Embryonic
Fetal
3
4
6
9
12
20
SOX6
MEIS1
TBX20
PRDM6
F
Fetal
1
2
3
Cluster
Sample
DCN
PRRX1
TBX20
GATA4
SOX6
HAPLN1
MEIS1
ROR1
JAG1
PRDM6
MYH11
HOXB3
HOXA3
logcounts
Cluster
3
6
9
12
Sample
12W_rep1
12W_rep2
G
Embryonic cluster
pvalue
Fetal cluster
4
20
1.0e-345
12
1.22e-15
3
1.0e-345
9
2.22e-16
6

Figure 3. Lineage deconvolution of embryonic and fetal nuclei. A. 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 S3A. B. Expression signatures in embryonic mesenchymal (4, 20) and endothelial-like (7) clusters. Both cluster 7 and 20 express high levels of cardiac TFs (*GATA4, TBX20*) and *HAPLN1*, a marker of semilunar valves. In contrast cluster 4 nuclei exhibit higher expression of neural crest markers, *HOXA3-B3* and *PLXNA2.* C. Pairwise differential gene expression of the two embryonic mesenchymal clusters; genes chosen for gene modules are marked by asterisks. D. Heatmap using embryonic gene modules, obtained using k-means clustering, separates embryonic mesenchymal nuclei in two groups. E. 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. F. Heatmap of embryonic gene modules and cell type marker genes using k-means clustering identifies three main groups of fetal nuclei. G. 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 3F.

Figure 4

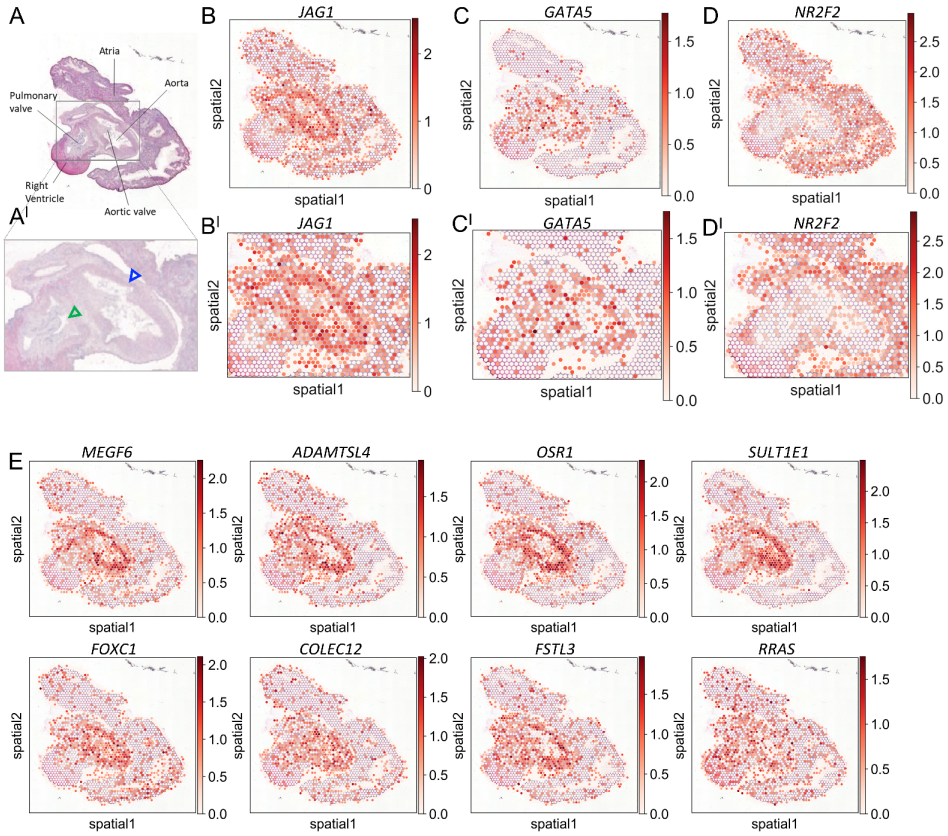

Figure 4. Genes mutated in congenital OFT defects. A. H&E staining of spatial transcriptomics section (Fig 2Jc), and magnified view of aortic and pulmonary valve area (A^I). The aorta and pulmonary trunk are indicated by blue and green arrowheads, respectively. B-D. *JAG1* (B), *GATA5* (C) and *NR2F2* (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*.

Figure 5

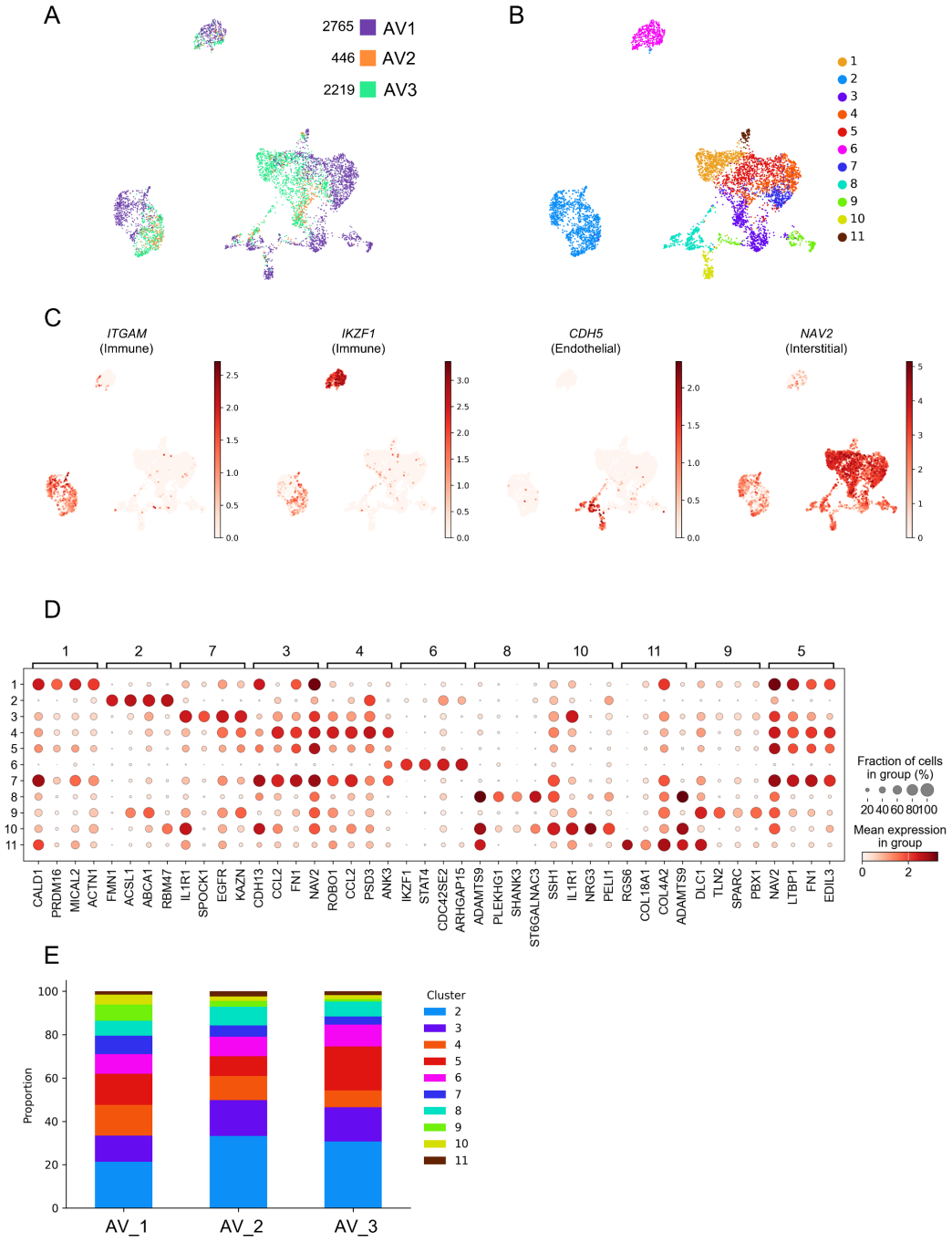

Figure 5. Cellular constituents of the mature aortic valves. A. Aortic valve sample association projected on a two-dimensional UMAP. Nuclei are coloured by sample. B. Nuclei clusters visualized by unsupervised clustering. Nuclei are coloured by cluster. C. Cell lineages identified using established lineage-specific 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. Cluster composition in each sample, presented as percentage of nuclei, with cluster 1 removed.

Figure 6

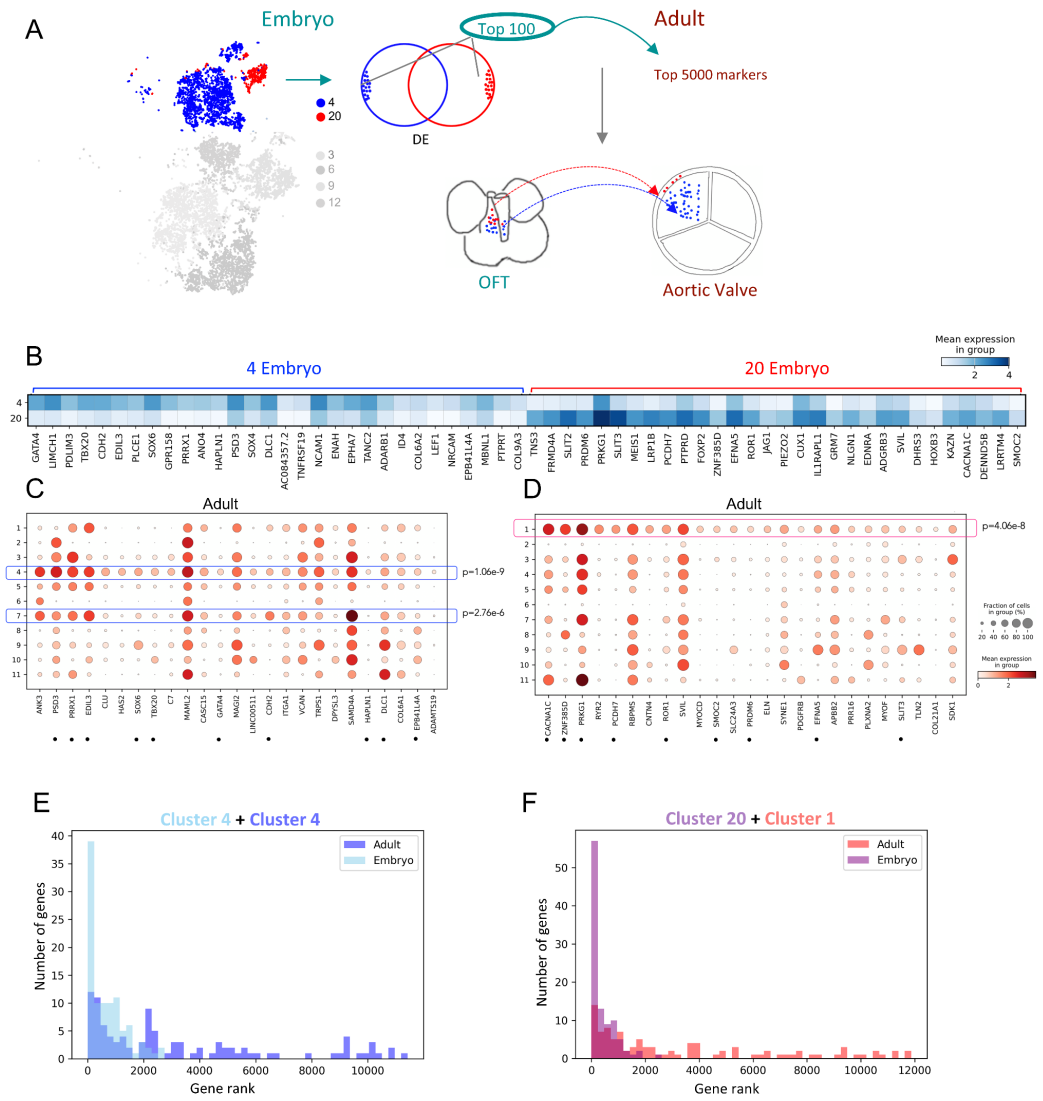

Figure 6. **Tracing lineage relationships of embryonic and adult cells**. A. Overview of the method used to trace adult descendent of embryonic nuclei. The first step is the identification of distinctive signatures in embryonic progenitors; for this, we used the top 100 differentially expressed (DE) 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. B. 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 6C-D. C. 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 B) are highlighted by dots. D. 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 B) are highlighted by dots. EF. Relative expression of embryonic signature genes in embryonic (E, 4; F, 20) and adult clusters (E, 4; F, 1), which are lineage related. Genes were ranked by expression values, relative to all the genes expressed in each cluster. In descendent adult nuclei, most embryonic genes (65% in adult cluster 4) are not represented in the top 1000 expressed genes. In contrast embryonic genes, are among the top 1000 expressed genes in embryonic nuclei (73% in embryo cluster 4).