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

Rotem Salmi Leshem[1*], Syed Murtuza Baker[2*], Joshua Mallen[1], Lu Wang[3], John Dark[3], Simon Bamforth[4,5] and Nicoletta Bobola[1,5]

[1]School of Medical Sciences, [2]School of Health Sciences, University of Manchester, UK
[3]Translational and Clinical Research Institute, Newcastle University.
[4]Newcastle University Biosciences Institute, Faculty of Medical Sciences, Newcastle.
* equal contribution

[5] 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 the aorta and pulmonary trunk, which originate from the left and right ventricles, respectively, creating the double circulation present in mammals. Defects in this area account for one-third of all congenital heart disease cases. Although model systems have greatly enhanced our understanding of the cell lineages that contribute to the OFT, our knowledge of OFT development in humans remains limited. Here, we present comprehensive transcriptomic data on the developing OFT at 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. 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 OFT carries blood away from the heart into the great arteries. Defects specifically affecting the outflow tract (OFT) of the heart represent a third of all CHD cases [2]. 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 [3]. In humans the formation and remodelling of the OFT is a relatively rapid process, occurring over a 3-week period (Carnegie Stage (CS)13 – CS22). In human embryos septation of the developing 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 distalmost region of the OFT is connected, via the aortic sac, to the pharyngeal arch arteries, which later remodel into the vessels of the mature aortic arch [6]. The morphological changes underlying OFT formation are orchestrated by two main cell lineages, second heart field (SHF)-derived myocardium and endothelium, and neural crest cells (NCC). The SHF 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 [7]. NCC are a transient, pluripotent cell population, which migrate from the neural tube to multiple areas of

the body. The cardiac NCC, a sub-population of NCC, plays a critical role in the septation of the OFT [8]. NCC form the aorticopulmonary septal complex, which separates the aorta and pulmonary trunk [9, 10].

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 significantly underexplored at the molecular level, and a clear relationship between cell lineages and cells found in the adult OFT is lacking. Here, we present comprehensive transcriptomic data on the developing OFT at two distinct timepoint (embryonic and fetal) and its adult derivatives, 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. 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 (12 pcw) 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 and retrieved in the clinical standard fashion. Hearts were arrested with 1L 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 snapped frozen by being submerged into liquid nitrogen and stored at -80 until analysis. All human hearts were obtained from the transplant donors with written

informed consent from their families. The research ethical approval (REC ref 16/NE/0230) was provided by the North East - Newcastle & North Tyneside Research Ethics Committee.

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 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 [11]. 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 lasers 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 12 pcw whole fetal heart sample was embedded in OCT and cryosectioned at 5 µm according to [12]. A tissue optimisation was performed according to 10x protocols [13] and [14]. 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 [15]. 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 [16]. 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 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 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 sample. 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: CS16-17_rep1 4703/7011; CS16-17_rep2 nuclei retained 2576/7761; 12W_rep1, 10,151/12,110 cells; 12W_rep2, 7306/16,110 cells; 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 [17].This final aggregation gave us a total of 30,166 cells for downstream analysis.

Normalisation and classification of cell-cycle phase

We used the deconvolution-based method [18] 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 have same total count after normalization. We used the cyclone method [19] to calculate the score of G1 phase and G2M phase for each cell.

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 were 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 [20]. 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'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 package to identify marker genes for each cluster [21]. 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 method which compares each cluster to the union of the rest of the clusters.

RNA-velocity

We applied scVelo to identify transient cellular dynamics in embryonic stage [22]. To calculate the spliced vs un-spliced ratio for each gene we used RNA-velocity's command line tool, velocyto 10x with default parameters [23] and used scVelo to compute velocity based on the spliced vs un-spliced ratio.

Spatial Transcriptomics

We used Spaceranger v.1.3.1 [24] 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 [25]. As each of the Visium spot has more than one cell we used cell2location to deconvolute the spots [26]. For cell2location genes were filterd 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 to12 and detection_alpha to 20.

Results

Cellular landscape of the human OFT and its adult derivatives

We conducted single nuclei (sn) RNA-seq on human OFT tissues to identify the cell types present in the human OFT. We isolated single nuclei from two pools of embryonic (CS 16-17) and two fetal (post conception week 12 (12 pcw)) OFTs and from the aortic valves from three adults (Fig.1A). The nuclei exhibit strong consistency between replicate samples (Fig 1B). After quality control and filtering, a total of 30166 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 observed the highest number of clusters in the fetal samples (n = 12) (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. 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 cell types. In addition to mesenchymal and endothelial cell types, we allocated adult and developmental clusters to immune cells. Interestingly, adult nuclei share the expression of broad markers, like *NEAT1* [27], which do not distinguish cell types but discriminate adult from developmental tissues. We confirmed cell types in the main compartments using additional verified markers (Fig S1A). The cardiac and endothelial nuclei populations appear to already express their appropriate cell type markers in the embryonic samples. 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* [28] and *PDGFRB* [29], implying these are mesenchymal cells prior to lineage differentiation (Fig S1A).

Spatial resolution of mesenchymal nuclei in the OFT

We conducted spatial gene expression analysis of the OFT to visualize cell types within this organ. We generated four sections of a 12 pcw OFT (equivalent stage to snRNA-seq), starting at the base with the aortic valves (a) and ending at the level of the pulmonary valves (d) (Fig. 2A). We transferred 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 S2A). Mesenchymal cells largely distributed within and around the vessels, and also mapped to the valves of the pulmonary artery and the aorta. Endothelial cells were primarily located in the aortic valves (Fig S2C). Immune cells were scattered around the tissue (Fig S2D), while neuronal cells were concentrated in a spot of the atria (Fig S2E).

Next, we focused our attention on mesenchymal cells, which build critical structures in the OFT, the separation of the aorta from the pulmonary artery and the semilunar valves at the base of the aorta and pulmonary artery. To dissect the heterogeneity within mesenchymal cells and identify lineage subtypes in the developing OFT, we performed nuclei sub-clustering of the whole dataset along with tSNE visualization. Using this approach, we obtained a total of seven mesenchymal clusters, two embryonic and five fetal (Fig 2BC). Embryonic nuclei do not express cell-type markers, suggesting they are still undifferentiated (Fig 2C). In contrast fetal nuclei largely express *DCN* (a fibroblast marker), with the exception of cluster 9 which contains *MYH11*-positive smooth muscle cells. We mapped the five mesenchymal clusters to

distinct structures in the OFT (Fig 2E) and used distinctive markers to confirm spatial assignments. Clusters 3 and 6 (Fig 2F-F^{I}) map to the arterial outer walls and the septum between the aorta and pulmonary arteries and largely colocalise with *DCN*-positive cells (Fig 2F^{II}). These two clusters largely overlap. Cluster 9 (Fig 2G) coincides with *MYH11* expression (Fig 2G^{I}): *MYH11* is a terminal marker of smooth muscle differentiation [30], which identifies the aortic smooth muscle layer. Cluster 12 (Fig 2H) 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 2H^{I}), a marker of the semilunar valves [31]. Finally, the smallest cluster (15) is restricted to the cardiac ventricle and is *DCN*-negative, therefore was eliminated from our further analysis (Fig. S2F). Thus, the five subtypes of mesenchymal nuclei correspond to spatially segregated cell populations.

Developmental trajectories of mesenchymal nuclei in the developing OFT

During embryonic development, the aorta and the pulmonary artery are separated and semilunar valves are formed at the base of these vessels, resulting in the transformation of the OFT from a single tube to two distinct vessels. It is important to establish a clear relationship between embryonic and fetal clusters, as the morphological changes occurring in the OFT during development are controlled by multiple cell lineages. At 12 pcw, the OFT is composed of mesenchymal populations that have already differentiated. However, connecting fetal cell populations to their embryonic counterparts is challenging due to the absence of cell type markers in embryonic nuclei. As a first step to characterize embryonic mesenchymal nuclei, we subclustered the fetal and embryonic data (Fig S3A) 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 [23]. Because of its timescale of hours, we exclusively applied RNA velocity to embryonic samples. At the embryonic stage, we detected a movement of embryonic nuclei from cluster 7 (endothelial like cells) specifically towards cluster 4 (mesenchymal cells) (Fig 3A). Despite being different cell types, cluster 7 and cluster 4 populations share a similar expression signature, supporting a common origin (Fig 3B); this 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. These observations suggest that RNA velocity detects the epithelial to mesenchymal transition (EMT) of endothelial cells in the endocardium, the cell layer adjacent to the myocardium. This results in the formation of a mobile group of mesenchymal cells that construct the semilunar valves [32, 33].

Next, we aimed to capture the developmental trajectories of embryonic mesenchymal cells to differentiated fetal cells. Trajectory inference [34] failed to detect lineage relationship between embryonic and fetal populations. We therefore used gene module scores to track the

developmental paths of mesenchymal cell populations and follow the descendant cells of each embryonic cluster. 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). This analysis further supports that cluster 4 and cluster 20 nuclei descend from diverse embryonic progenitors. 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 cluster 4, while '*MEIS1 JAG1 ROR1 PRDM6*' is highly enriched in 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 our embryonic gene modules largely separate fetal nuclei into two distinct lineages (Fig 3E): a 'blue' lineage largely made of clusters 4-3-12 with higher expression of group 1 signature (*SOX6, TBX20*), and a 'red' lineage, containing clusters 20-6-9, with higher levels of group 2 signature (*MEIS1, PRDM6*). Consistently, when combined with the differentiated mesenchymal cell markers expressed in fetal samples, embryonic gene modules clustered 12 pcw mesenchymal cells in three distinct groups (Fig 3F). Relative to the original clustering of fetal nuclei (Fig 3E), module 1 signature defines a heterogenous group of related cell types, fibroblasts and valves, encompassing cluster 3 and 12 respectively, suggesting a common embryonic origin for these nuclei. Module 2 signature is inherited by smooth muscle cells (cluster 6) and fibroblasts (cluster 9), which are segregated by the expression of cell type-specific markers, *MYH11* and *DCN*. We also observed that cardiac neural crest markers *HOXA3/B3* [35] are almost exclusively expressed in 'group 2 and 3' nuclei, suggesting that 'group 2 and 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 [36]. It appears therefore that the embryonic module *MEIS1 JAG1 ROR1 PRDM6* identifies cells committed to become distinct cell types (non-valvular fibroblasts and smooth muscle cells) of mixed derivation. In sum, our analysis indicates that two spatially distinct mesenchymal populations in the fetal OFT (Fig 2) derive from separate embryonic population, namely group 1 (cluster 4) for valvular fibroblasts and group 2 for smooth muscle cells. In contrast, *DCN*-positive, *HAPLN*-negative mesenchymal cells derive from both group 1 and 2 embryonic nuclei. These cells appear to intermingle in OFT tissues.

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 [37]. The most severe include common arterial trunk (CAT), transposition of the great arteries (TGA), double-outlet right ventricle (DORV) and Tetralogy of Fallot (TOF). We investigated the distribution of genes whose mutations cause OFT defects. *JAG1* mutations cause TOF [38]; *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*) [39], TOF (*GATA4, GATA6*) [40, 41] and CAT (*GATA6*) *GATA6* [42]. Different from *GATA5*, which shows a restricted expression to the vessels (Fig 4C, C^{I}) the distribution of *GATA4* and *GATA6* transcripts is broad across the vessels, the septum and cardiac tissue (Fig S4AB). Mutations in *GATA4-5-6* also causes bicuspid aortic valves (BAV), a less severe condition characterized by two valves instead of three [43]. Similar to *GATA5*, *GATA6*, *GATA4* is highly expressed in the valves, marked by *HAPLN1* (Fig S4C). As a final example, mutations in *NR2F2* cause DORV and TOF [44]. 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 validate this hypothesis, we focused on *JAG1*, whose transcripts exhibit a distinct pattern around the vessels, and generated an algorithm (ref) to identify genes that display comparable expression patterns to *JAG1* (Fig 4E). Remarkably, 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 [45, 46].

Cellular constituents of mature aortic valves

Histologically normal aortic valves were procured from hearts collected for transplantation and subsequently rejected. A total of 5,430 single nuclei from three different 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) [47] and immune (2 clusters, *ITGAM*-positive macrophages and *IKFZ1*-positive dendritic cells) [48, 49] (Fig 5C).

With the exception of cluster 1, which was predominantly composed of cells from sample AV3 (Fig S5A), patient samples were evenly spread out across the UMAP (Fig 5C). We performed differential gene expression analysis to aid in the classification of each cell cluster. Heat maps (Fig 5SB) group together immune (2,6) and endothelial cell types (8,10). In addition to *CDH5* (ref), endothelial clusters are distinguished by high levels of *ADAMTS9,*

which encodes for a metalloproteinase implicated in aortic valve anomalies in mouse [50]. *ADAMTS9* is also expressed in cluster 11, which is *CDH5*-negative (Fig 5D). The remaining clusters (1,3,4,5,7,9,11) share expression of similar transcripts, such as *NAV2*, *ROBO1*, *LTBP1*, *FN1*, *COL4A2* and were assigned to interstitial cells. 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 S5C); as SMC are not a main constituent of the aortic valves, we attribute this skewed enrichment to aortic wall tissue being sampled with the aortic valves in AV1. The most related cluster to cluster 1, was cluster 11 which accounts for a small proportion of nuclei in all three samples (ranging from 1.5% to 2.3%).

The mature valve structure is composed of three highly organized layers of extracellular matrix (fibrosa, spongiosa and ventricularis), deposited by valvular interstitial cells (VICs) and enriched in collagen fibers, proteoglycans and elastin fibers, respectively. VICs are identified by the expression of transcripts encoding for ECM proteins, most notably *COL1A1*, *VIM*, *COL3A1*, *VCAN*, *BGN* and *LUM*; distinct subpopulations of VICs producing specific types of ECM in the developing valve can account for VIC heterogeneity [51]. We found only a scattered expression of *COL1A1* and *COL3A1* and, with the exception of *VIM* and *VCAN* (Fig S5B), a modest enrichment of other ECM encoding transcripts in nuclei of interstitial cells across all the adult samples examined (Fig S5C), possibly reflecting a decreased deposition of ECM in adult valves. Among interstitial nuclei, clusters 4,5 and 7 were the most related. Differently from cluster 5, clusters 4 and 7 express high levels of *CCL2*, an inflammatory cytokine [52], suggesting nuclei in these clusters correspond to activated fibroblasts. While AV3 was more enriched in cluster 5 (20.3% nuclei in AV3, relative to 14.3% and 9% in AV2 and AV3, respectively) we observed opposite proportions for cluster 4 (14.2% in AV1 relative to 7% in AV3) and cluster 7 (8.5% in AV1 relative to 3.7% in AV3) (Fig 5E; Fig S5D). Finally, cluster 9 displays enrichment in *DLC1* and *TLN2*, implicated in cytoskeletal changes [53, 54], and *SPARC* (osteonectin), implicated in valve calcification [55]. Cluster 9 is overrepresented in AV1 (7% relative to 2.6% AV2 and 1.1% 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 potentially lead to valve calcification in this AV sample.

Discussion

The OFT transports blood away from the heart and into the great arteries. During human embryo development, the OFT undergoes formation and remodelling between CS13 and CS22. 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. We used snRNA-seq to

analyze 30166 OFT nuclei, which covered two stages of OFT development (CS16-17 and 12 pcw) up to the adult derivatives of the OFT, the aortic valves. We detected cardiac (5 clusters), mesenchymal (4 clusters), endothelial (5 clusters) neural (2 clusters) and immune (2 clusters) cell types. We observed that at the earliest stage, cardiac and endothelial nuclei populations already expressed their appropriate cell type markers. In contrast, most mesenchymal nuclei did not exhibit markers of differentiated cell types, implying the embryonic stage captured mesenchymal nuclei before differentiation. Moreover, although embryonic and fetal cells exhibit significant transcriptional diversity, adult cells (with the exception of well separated immune cell types) exhibit shared characteristics that distinguish them from developmental cells. In addition, while embryonic and fetal cells display considerable variations in gene expression, adult cells (with the exception of immune cell types) share distinguishing features that set them apart from developing cells. These feature include high levels of expression of shared markers, such as NEAT1, a type of long non-coding RNA (lncRNA) that acts in the nucleus as a pivotal structural component of paraspeckle sub-organelles [56].

We conducted a spatial analysis of gene expression in the fetal OFT (OFT) to visualize the distribution of different cell types within the organ. Our focus was primarily on mesenchymal cells, which are responsible for building the semilunar valves and the separation of the aorta from the pulmonary artery. Two main cell lineages, the second heart field-derived myocardium and endothelium, and the cardiac neural crest, orchestrate the morphological changes that occur during OFT formation. We were able to identify four subtypes of mesenchymal nuclei at 12 pcw, which largely corresponded to distinct cell populations located in different areas of the OFT. These included two clusters of fibroblast-like cells, one cluster of smooth muscle cells, and one cluster of interstitial valvular cells. 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 [34]. 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 complete expression profiles. Rather than relying on global gene expression patterns, we identified specific and unique transcriptional signatures that distinguish between two distinct embryonic mesenchymal lineages and are inherited by fetal mesenchymal nuclei. This approach enabled us to trace each of the four spatially separated mesenchymal populations in the fetal OFT back to their embryonic precursors. Our analysis revealed that valvular fibroblasts (group 1) and smooth muscle cells (group 2) have separate embryonic origins. Conversely, mesenchymal cells expressing *DCN* but lacking *HAPLN1* were found to derive from both embryonic mesenchymal lineages. Through the use of model systems, significant

advances have been made in understanding the cell lineages that contribute to the OFT. OFT formation is primarily orchestrated by two cell lineages, namely, the myocardium and endothelium derived from the second heart field, as well as the cardiac neural crest. However, our understanding of OFT development in humans remains limited. According to our analysis, mesenchymal cells involved in constructing the semilunar valves are likely to originate from endothelial cells undergoing EMT in the endocardium - the myocardium-adjacent cell layer [32, 33]. Additionally, our findings align with previous observations in mice, indicating that smooth muscle cells in the human aortic root originate from the neural crest and secondary heart field [57].

OFT defects make up one third of all cases of CHD. Anomalies occur when the vessels leaving the heart do not form or remodel correctly, leading to problems with blood circulation and oxygenation, which are fatal and require surgical correction shortly after birth. Bicuspid aortic valve is the most common form of CHD that affects the OFT is often asymptomatic at birth and in childhood but can predispose an individual to more severe heart disease later in life. We utilized our spatial transcriptomic data to visualize expression patterns of genes implicated in OFT defects. We made an interesting observation that genes associated with the same defect can exhibit complementary expression patterns. For example, while mutations in *NR2F2* and *JAG1* result in TOF, *JAG1* transcripts are present in both the aorta and the pulmonary artery, while *NR2F2* transcripts are predominantly located in the cardiac tissues surrounding the vessels and are largely excluded from OFT structures. Our spatial transcriptomic analysis not only allows for the investigation of known genes in human OFT defects, but also has the potential to reveal new candidate genes involved in these conditions. By identifying genes with expression patterns similar to *JAG1*, we were able to uncover new genes whose loss of function leads to OFT abnormalities in mice. Thus, our OFT spatial transcriptomics offers a powerful tool for enhancing our understanding of OFT defects and advancing the discovery of previously unknown genes implicated in these conditions.

We obtained histologically normal aortic valves from hearts intended for transplantation but later rejected. We conducted sequencing on a total of 5,430 individual nuclei extracted from three distinct human aortic valve samples, which were classified into three primary compartments, namely interstitial, endothelial, and immune cells - consistent with the anticipated cell types present in aortic valves. The variability of individual samples from multiple donors can be a confounding factor; however, we observed that patient samples were evenly distributed across the UMAP, with only a few exceptions. Notably, one sample showed a significantly high abundance of nuclei assigned to SMCs. As SMCs are not a primary constituent of aortic valves, we attribute this biased enrichment to the aortic wall tissue

that may have been inadvertently sampled with the aortic valves in this sample. On the other hand, the higher prevalence of clusters 4, 7, and 9, characterized by elevated levels of *CCL2* (in clusters 4 and 7) and *SPARC* (in cluster 9), in AV1 may be attributed to physiological differences between the samples and the onset/presence of inflammatory processes that are more prominent in the AV1 sample. VICs are characterized by the expression of transcripts that encode ECM proteins, and the heterogeneity of VICs can be attributed to distinct subpopulations that produce specific types of ECM during valve development [51]. Overall, our analysis showed moderate enrichment of ECM-encoding transcripts in the nuclei of interstitial cells across all adult samples examined, suggesting a reduced deposition of ECM in adult valves. In conclusion, our findings strongly suggest that adult cells exhibit significantly reduced transcriptional diversity and a higher degree of shared markers compared to their developmental progenitors. These results underscore the critical need for single-cell atlases to identify appropriate markers for accurately defining cell types of interest.

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 and Peyman Zarrineh 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 SB. 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 above are available upon acceptance of the manuscript through the HCA.

References

1. Hoffman, J.I. and S. Kaplan, *The incidence of congenital heart disease.* J Am Coll Cardiol, 2002. **39**(12): p. 1890-900.
2. 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, 2006. **113**(6): p. e85-151.
3. Stefanovic, S., H.C. Etchevers, and S. Zaffran, *Outflow Tract Formation-Embryonic Origins of Conotruncal Congenital Heart Disease.* J Cardiovasc Dev Dis, 2021. **8**(4).

4. Anderson, R.H., et al., *Normal and abnormal development of the intrapericardial arterial trunks in humans and mice.* Cardiovasc Res, 2012. **95**(1): p. 108-15.
5. Anderson, R., et al., *The Reappraisal of Normal and Abnormal Cardiac Development*, in *Hemodynamics and Cardiology Neonatology Questions and Controversies*, K. C and S. I, Editors. 2012, Elsevier. p. 391-414.
6. Bamforth, S.D., et al., *Clarification of the identity of the mammalian fifth pharyngeal arch artery.* Clin Anat, 2013. **26**(2): p. 173-82.
7. Kelly, R.G., *The second heart field.* Curr Top Dev Biol, 2012. **100**: p. 33-65.
8. Kirby, M.L. and M.R. Hutson, *Factors controlling cardiac neural crest cell migration.* Cell Adh Migr, 2010. **4**(4): p. 609-21.
9. Waldo, K., et al., *Cardiac neural crest cells provide new insight into septation of the cardiac outflow tract: aortic sac to ventricular septal closure.* Dev Biol, 1998. **196**(2): p. 129-44.
10. Jiang, X., et al., *Fate of the mammalian cardiac neural crest.* Development, 2000. **127**(8): p. 1607-1616.
11. 10xGenomics, *Nuclei Isolation from Complex Tissues for Single Cell Multiome ATAC + Gene Expression Sequencing* 2021.
12. 10xGenomics, *Visium Spatial Protocols – Tissue Preparation Guide*. 2021.
13. 10xGenomics, *Methanol Fixation, H&E Staining & Imaging for Visium Spatial Protocols*. 2021.
14. 10xGenomics, *Visium Spatial Gene Expression Reagent Kits - Tissue Optimization*. 2020.
15. 10xGenomics, *Visium Spatial Gene Expression Reagent Kits*. 2021.
16. 10xGenomics, *Chromium Next GEM Single Cell 3' Reagent Kits v3.1 (Dual Index)* 2020.
17. Wolf, F.A., P. Angerer, and F.J. Theis, *SCANPY: large-scale single-cell gene expression data analysis.* Genome Biology, 2018. **19**(1): p. 15.
18. Lun, A., K. Bach, and J. Marioni, *Pooling across cells to normalize single-cell RNA sequencing data with many zero counts.* Genome Biology, 2016. **17**.
19. Scialdone, A., et al., *Computational assignment of cell-cycle stage from single-cell transcriptome data.* Methods, 2015. **85**: p. 54-61.
20. Langfelder, P., B. Zhang, and S. Horvath, *Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R.* Bioinformatics, 2007. **24**(5): p. 719-720.
21. Lun, A.T., D.J. McCarthy, and J.C. Marioni, *A step-by-step workflow for low-level analysis of single-cell RNA-seq data with Bioconductor.* F1000Res, 2016. **5**: p. 2122.
22. Bergen, V., et al., *Generalizing RNA velocity to transient cell states through dynamical modeling.* Nat Biotechnol, 2020. **38**(12): p. 1408-1414.
23. La Manno, G., et al., *RNA velocity of single cells.* Nature, 2018. **560**(7719): p. 494-498.
24. 10xGenomics. *What is Space Ranger?* 2021; Available from: https://support.10xgenomics.com/spatial-gene-expression/software/pipelines/latest/what-is-space-ranger.
25. Moran, P.A., *Notes on continuous stochastic phenomena.* Biometrika, 1950. **37**(1-2): p. 17-23.
26. Kleshchevnikov, V., et al., *Cell2location maps fine-grained cell types in spatial transcriptomics.* Nature Biotechnology, 2022. **40**(5): p. 661-671.
27. 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, 2022. **20**(1): p. 7.
28. Farahani, R.M. and M. Xaymardan, *Platelet-Derived Growth Factor Receptor Alpha as a Marker of Mesenchymal Stem Cells in Development and Stem Cell Biology.* Stem Cells Int, 2015. **2015**: p. 362753.

29. 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, 2018. **75**(3): p. 547-561.
30. Babij, P., C. Kelly, and M. Periasamy, *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, 1991. **88**(23): p. 10676-80.
31. Fang, M., et al., *Loss of β-catenin promotes chondrogenic differentiation of aortic valve interstitial cells.* Arterioscler Thromb Vasc Biol, 2014. **34**(12): p. 2601-8.
32. Gittenberger-de Groot, A.C., et al., *Epicardium-derived cells contribute a novel population to the myocardial wall and the atrioventricular cushions.* Circulation research, 1998. **82**(10): p. 1043-1052.
33. Kirby, M.L., T.F. Gale, and D.E. Stewart, *Neural crest cells contribute to normal aorticopulmonary septation.* Science, 1983. **220**(4601): p. 1059-1061.
34. Trapnell, C., et al., *The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells.* Nat Biotechnol, 2014. **32**(4): p. 381-386.
35. Bridoux, L., et al., *HOX paralogs selectively convert binding of ubiquitous transcription factors into tissue-specific patterns of enhancer activation.* PLOS Genetics, 2020. **16**(12): p. e1009162.
36. Sawada, H., et al., *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, 2017. **37**(9): p. 1722-1726.
37. Neeb, Z., et al., *Cardiac outflow tract anomalies.* WIREs Developmental Biology, 2013. **2**(4): p. 499-530.
38. Eldadah, Z.A., et al., *Familial Tetralogy of Fallot caused by mutation in the jagged1 gene.* Hum Mol Genet, 2001. **10**(2): p. 163-9.
39. Jiang, J.Q., et al., *Prevalence and spectrum of GATA5 mutations associated with congenital heart disease.* Int J Cardiol, 2013. **165**(3): p. 570-3.
40. Maitra, M., et al., *Identification of GATA6 sequence variants in patients with congenital heart defects.* Pediatr Res, 2010. **68**(4): p. 281-5.
41. Tomita-Mitchell, A., et al., *GATA4 sequence variants in patients with congenital heart disease.* J Med Genet, 2007. **44**(12): p. 779-83.
42. Kodo, K., et al., *GATA6 mutations cause human cardiac outflow tract defects by disrupting semaphorin-plexin signaling.* Proc Natl Acad Sci U S A, 2009. **106**(33): p. 13933-8.
43. Yang, B., et al., *Protein-altering and regulatory genetic variants near GATA4 implicated in bicuspid aortic valve.* Nat Commun, 2017. **8**: p. 15481.
44. Al Turki, S., et al., *Rare variants in NR2F2 cause congenital heart defects in humans.* Am J Hum Genet, 2014. **94**(4): p. 574-85.
45. Sanchez, J., et al., *Conditional inactivation of Foxc1 and Foxc2 in neural crest cells leads to cardiac abnormalities.* genesis, 2020. **58**(7): p. e23364.
46. Wang, Q., et al., *Odd-skipped related 1 (Odd1) is an essential regulator of heart and urogenital development.* Developmental Biology, 2005. **288**(2): p. 582-594.
47. Harris, E.S. and W.J. Nelson, *VE-cadherin: at the front, center, and sides of endothelial cell organization and function.* Current Opinion in Cell Biology, 2010. **22**(5): p. 651-658.
48. Hulin, A., et al., *Macrophage Transitions in Heart Valve Development and Myxomatous Valve Disease.* Arteriosclerosis, Thrombosis, and Vascular Biology, 2018. **38**(3): p. 636-644.
49. Collin, M. and V. Bigley, *Human dendritic cell subsets: an update.* Immunology, 2018. **154**(1): p. 3-20.
50. Kern, C.B., et al., *Reduced versican cleavage due to Adamts9 haploinsufficiency is associated with cardiac and aortic anomalies.* Matrix biology, 2010. **29**(4): p. 304-316.

51. Hulin, A., et al., *Maturation of heart valve cell populations during postnatal remodeling.* Development, 2019. **146**(12).
52. Kumar, S.N. and J.M. Boss, *Site A of the MCP-1 distal regulatory region functions as a transcriptional modulator through the transcription factor NF1.* Molecular immunology, 2000. **37**(11): p. 623-632.
53. 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, 1998. **58**(10): p. 2196-9.
54. Monkley, S.J., C.A. Pritchard, and D.R. Critchley, *Analysis of the Mammalian Talin2 Gene TLN2.* Biochemical and Biophysical Research Communications, 2001. **286**(5): p. 880-885.
55. Bradshaw, A.D., et al., *SPARC-null mice exhibit increased adiposity without significant differences in overall body weight.* Proc Natl Acad Sci U S A, 2003. **100**(10): p. 6045-50.
56. Taiana, E., et al., *LncRNA NEAT1 in Paraspeckles: A Structural Scaffold for Cellular DNA Damage Response Systems?* Noncoding RNA, 2020. **6**(3).
57. Sawada, H., et al., *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, 2017. **37**(9): p. 1722-1726.

Figure legends

Figure 1. **The cellular landscape of the developing OFT and its adult derivatives**. A. Experimental schematics. We performed snRNA-seq with nuclei isolated from embryonic (CS 16-17) and fetal (12 pcw) OFTs and the aortic valves (AV) from three adults. Spatial transcriptomics was performed on four sections of the OFT region from a cryo-sectioned fetal (12 pcw) heart. B. Sample correlation visualized by unsupervised clustering and projected on a two-dimensional UMAP. C. Cell clusters visualized by the same UMAP as in B. D. Cluster variation in each sample presented as percentage of cells. E. Dotplot of top differentially expressed genes in clusters identify five main cell types, cardiac, endothelial, mesenchymal (including valve insterstitial), neural and immune. F. Cell types in E visualised by UMAP.

Figure S1. A. Cell types in 1E labelled using established lineage-specific markers. *DCN* identifies fibroblast populations, *MYH11* labels SMCs, *PECAM1* is a marker for endothelial cells and *PDGFRA* and *PDGFRB* mark mesenchymal cells before differentiation. Each nucleus is colored based on the scaled expression of the indicated marker.

Figure 2. Spatial distribution of mesenchymal clusters A. Heart alignment for sectioning, with the OFT region marked in red (left). H&E staining of OFT cryosections used for spatial transcriptomics from the base of the OFT (a) to the pulmonary valves (d). The aorta and pulmonary artery are indicated by blue and green arrowheads, respectively. BC. Mesenchymal cell clusters (B) and sample projection (C) of fetal and embryonic samples following subclustering, projected on a two-dimensional tSNE. D. 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. E. H&E section c annotated to show major structures. F, F^{I}, G, H. Spatial distribution of mesenchymal clusters (purple and yellow). F^{II}, G^{I}, H^{I}. Spatial distribution of lineage specific markers (white and red). F-F^{II}. Fibroblast clusters 3 (F) and 6 (F^{I}) largely overlap with fibroblast lineage marker *DCN* (F^{II}). G-G^{I}. Cluster 9 (G) and smooth muscle lineage specific marker *MYH11* (G^{I}) map to the aortic walls as well as to the pulmonary artery. H-H^{I} Cluster 12 (H) and valve specific marker *HAPLN1* (H^{I}) are mainly found in the valves at the base of the aorta and pulmonary artery.

Figure S2. Mapping of cell types from transcriptomics data to spatial location using Cell2location. A. Cardiac muscle cells in the atria and ventricle. B. Mesenchymal cells consisting of smooth muscle cells inside the lumen of the aorta, fibroblast layer around the vessels and in the aortic and pulmonary semilunar valves. C. Endothelial cells in the aortic semilunar valves. D. Immune cells. E. Neuronal cells. F. Cluster 15 is restricted to the cardiac ventricle and was removed from the analysis of OFT mesenchymal clusters.

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. 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 with genes chosen for gene modules marked by asterisks. D. Heatmap using gene modules in embryonic mesenchymal clusters, obtained using k-means clustering, separates embryonic 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.

Figure S3. A. Separate aggregation of embryonic and fetal samples (without adult samples), re-clustered and presented by tSNE. B. Louvain 0.4 reclustering projected on the same tSNE.

Figure 4. Genes mutated in congenital OFT defects. A. H&E staining of spatial transcriptomics section C, and magnified view of Aortic and Pulmonary valve area (A^{I}). The aorta and pulmonary artery 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 (BI, CI, DI). E. Genes identified as displaying spatially similar expression pattern to *JAG1.*

Figure S4. A. *GATA6* and B. *GATA4*. GATA4 and GATA6 are broadly distributed across the OFT and surrounding cardiac tissue. Mutations in both genes cause BAV; the position of the valves is highlighted by *HAPLN1.*

Figure 5. Cellular constituents of the mature aortic valves A. Aortic valve sample association projected on a two-dimensional UMAP. B. Cell clusters visualized by unsupervised clustering. C. Cell lineages identified by 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 variation in each adult sample presented by percentage of cells, with cluster 1 removed.

Figure S5. A. Cluster variation in each sample presented by percentage of cells. B. Differential gene expression heatmap of adult clusters. C. ECM encoding transcripts in interstitial cells. Each nucleus is colored based on the scaled expression of the indicated marker. D. Proportions of nuclei in each cluster.

Figure 1

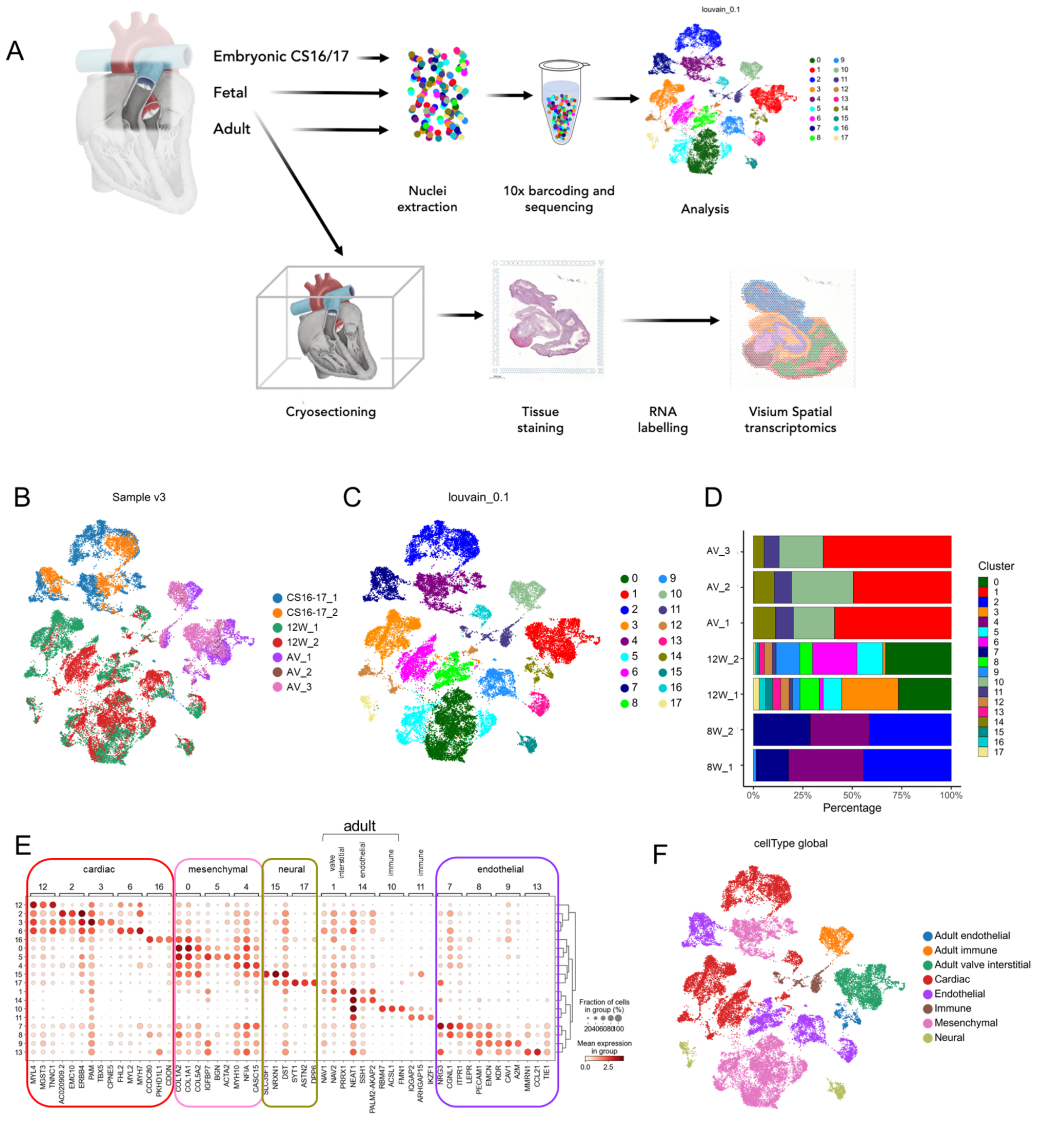

A
Embryonic CS16/17
Fetal
Adult
Nuclei extraction
10x barcoding and sequencing
louvain_0.1
Analysis
Cryosectioning
Tissue staining
RNA labelling
Visium Spatial transcriptomics
B
Sample v3
CS16-17_1
CS16-17_2
12W_1
12W_2
AV_1
AV_2
AV_3
C
louvain_0.1
D
AV_3
AV_2
AV_1
12W_2
12W_1
8W_2
8W_1
0%
25%
50%
75%
100%
Percentage
Cluster
E
adult
cardiac
mesenchymal
neural
endothelial
valve interstitial
endothelial
immune
immune
Fraction of cells in group (%)
Mean expression in group
F
cellType global
Adult endothelial
Adult immune
Adult valve interstitial
Cardiac
Endothelial
Immune
Mesenchymal
Neural

Fig S1

A

Figure 2

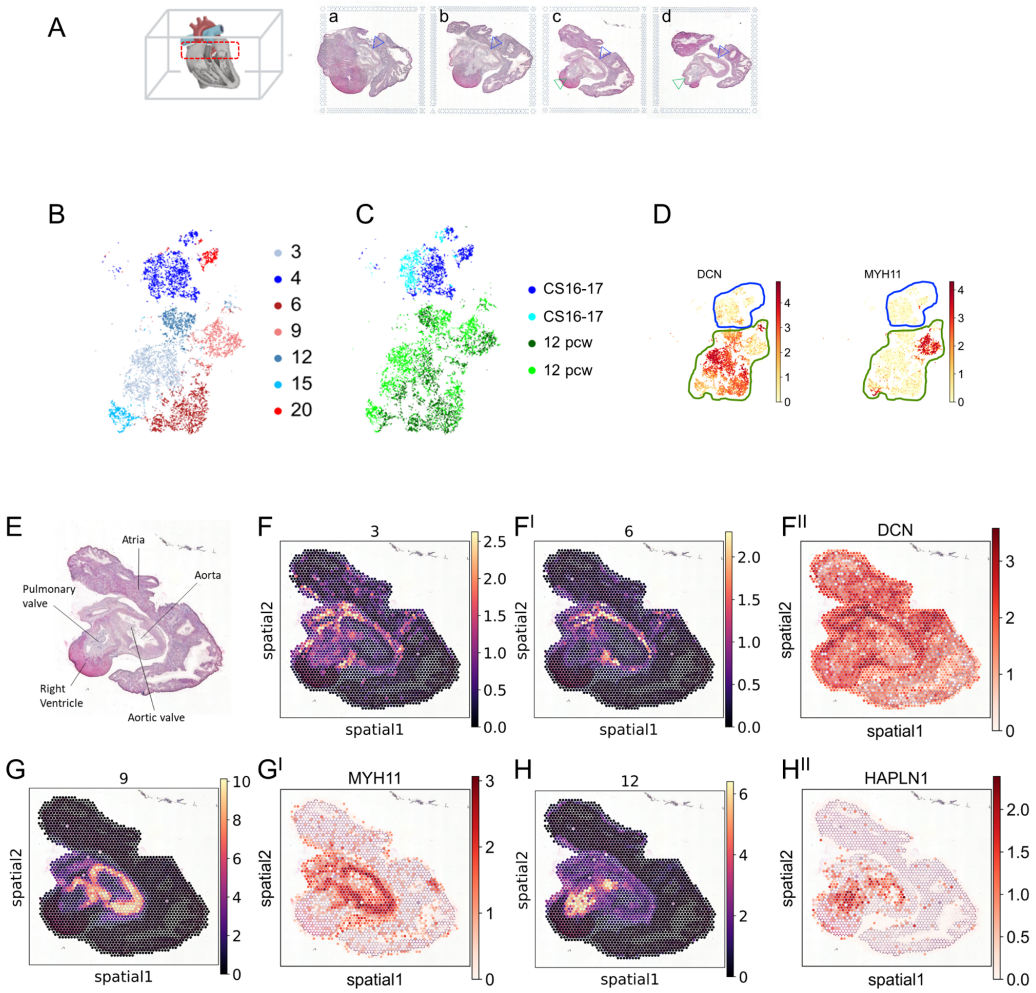

A
a
b
c
d
B
3
4
6
9
12
15
20
C
CS16-17
CS16-17
12 pcw
12 pcw
D
DCN
MYH11
E
Atria
Aorta
Pulmonary valve
Right Ventricle
Aortic valve
F
3
F^I
6
F^II
DCN
G
9
G^I
MYH11
H
12
H^II
HAPLN1
spatial1
spatial2

Fig S2

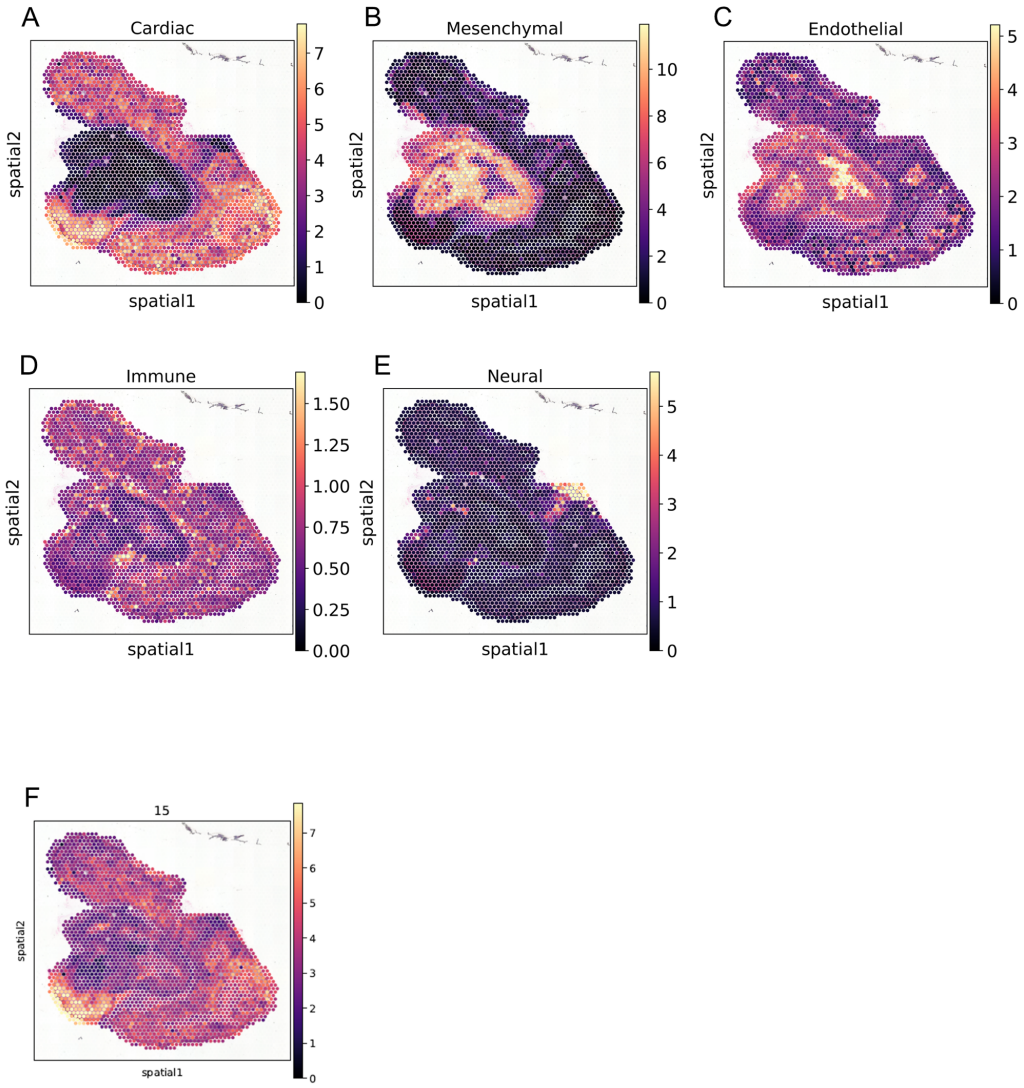

A
Cardiac
spatial2
spatial1
7
6
5
4
3
2
1
0
B
Mesenchymal
spatial2
spatial1
10
8
6
4
2
0
C
Endothelial
spatial2
spatial1
5
4
3
2
1
0
D
Immune
spatial2
spatial1
1.50
1.25
1.00
0.75
0.50
0.25
0.00
E
Neural
spatial2
spatial1
5
4
3
2
1
0
F
15
spatial2
spatial1
7
6
5
4
3
2
1
0

Figure 3

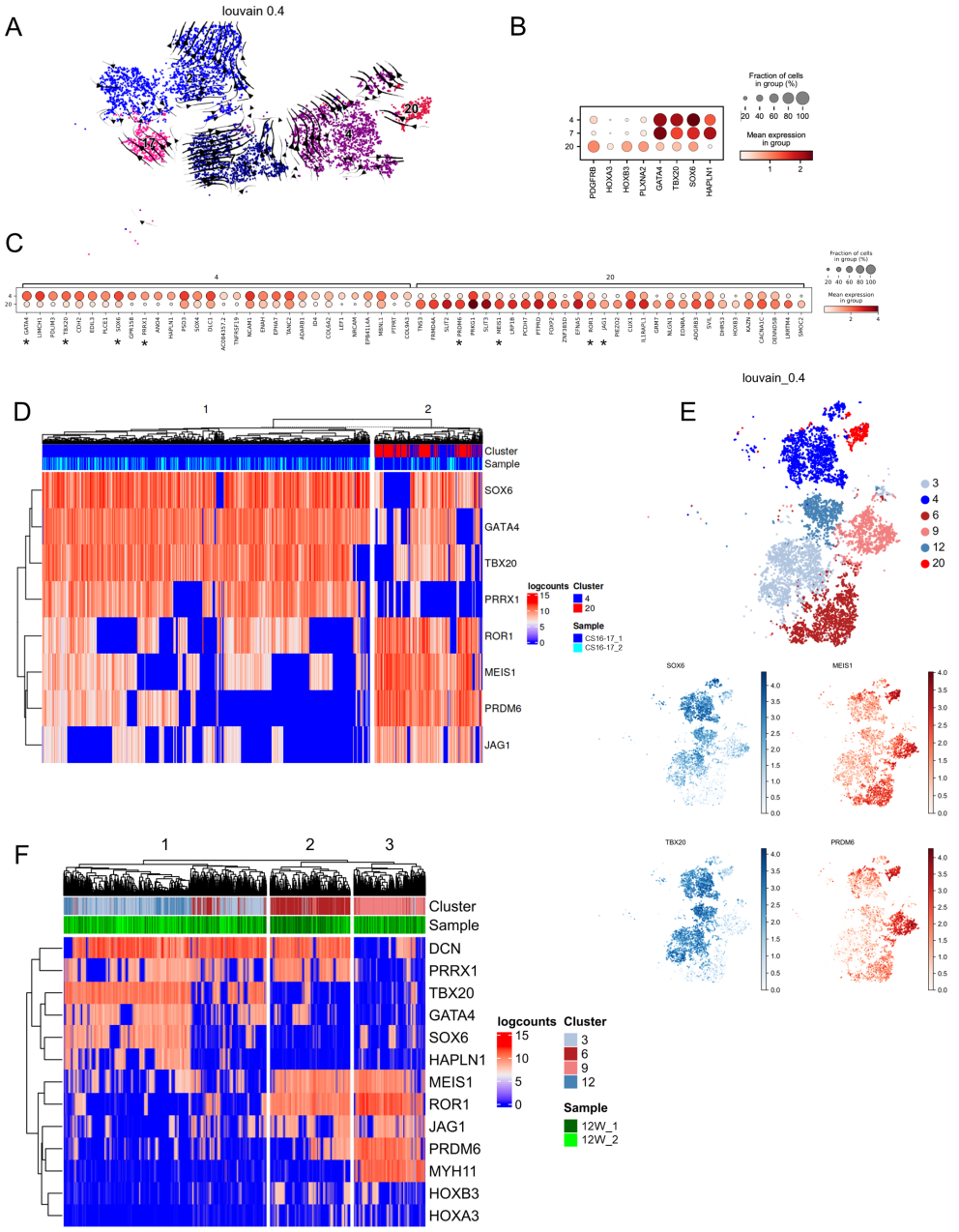

A
louvain 0.4
B
Fraction of cells in group (%)
Mean expression in group
PDGFRB
HOXA3
HOXB3
PLXNA2
GATA4
TBX20
SOX6
HAPLN1
C
D
Cluster
Sample
SOX6
GATA4
TBX20
PRRX1
ROR1
MEIS1
PRDM6
JAG1
logcounts
Cluster
Sample
CS16-17_1
CS16-17_2
E
louvain_0.4
SOX6
MEIS1
TBX20
PRDM6
F
Cluster
Sample
DCN
PRRX1
TBX20
GATA4
SOX6
HAPLN1
MEIS1
ROR1
JAG1
PRDM6
MYH11
HOXB3
HOXA3
logcounts
Cluster
Sample
12W_1
12W_2

Figure S3

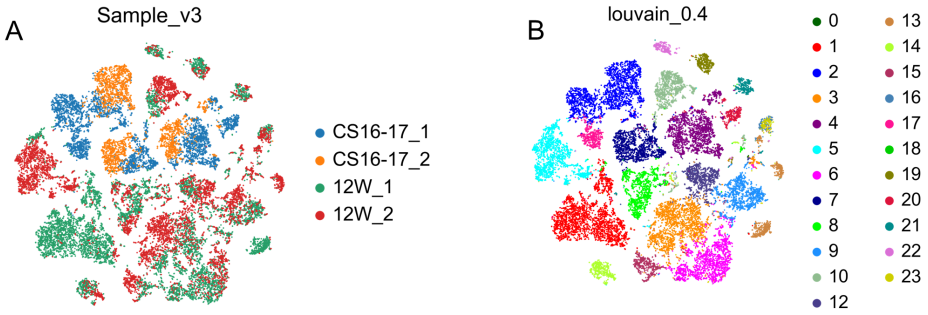

A
Sample_v3
CS16-17_1
CS16-17_2
12W_1
12W_2
B
louvain_0.4
0
1
2
3
4
5
6
7
8
9
10
12
13
14
15
16
17
18
19
20
21
22
23

Figure 4

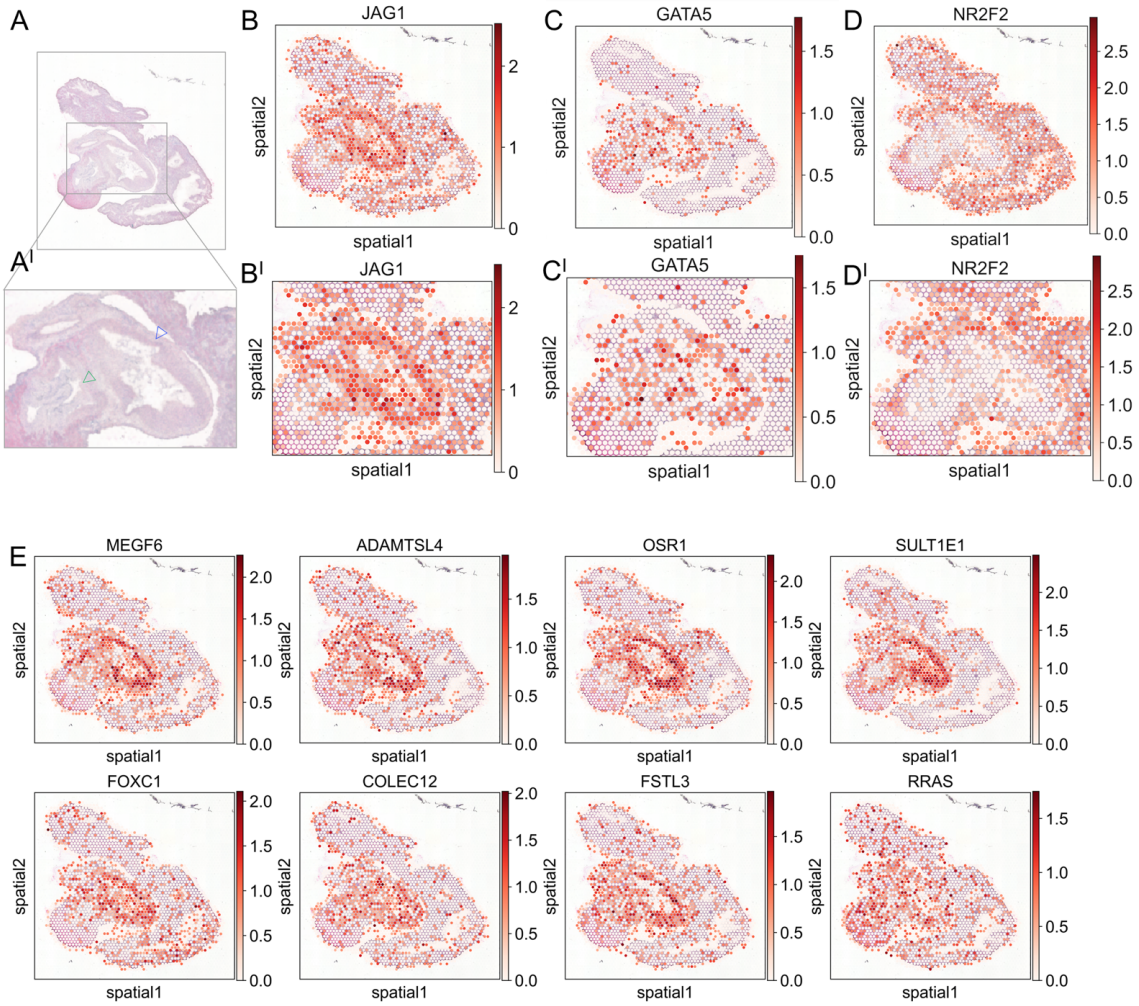

A
A'
B
JAG1
B'
JAG1
C
GATA5
C'
GATA5
D
NR2F2
D'
NR2F2
E
MEGF6
ADAMTSL4
OSR1
SULT1E1
FOXC1
COLEC12
FSTL3
RRAS
spatial1
spatial2

Figure S4

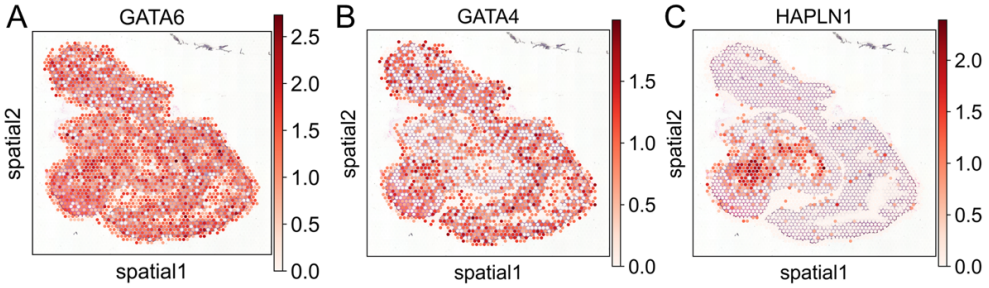

A
GATA6
spatial2
spatial1
2.5
2.0
1.5
1.0
0.5
0.0
B
GATA4
spatial2
spatial1
1.5
1.0
0.5
0.0
C
HAPLN1
spatial2
spatial1
2.0
1.5
1.0
0.5
0.0

Figure 5

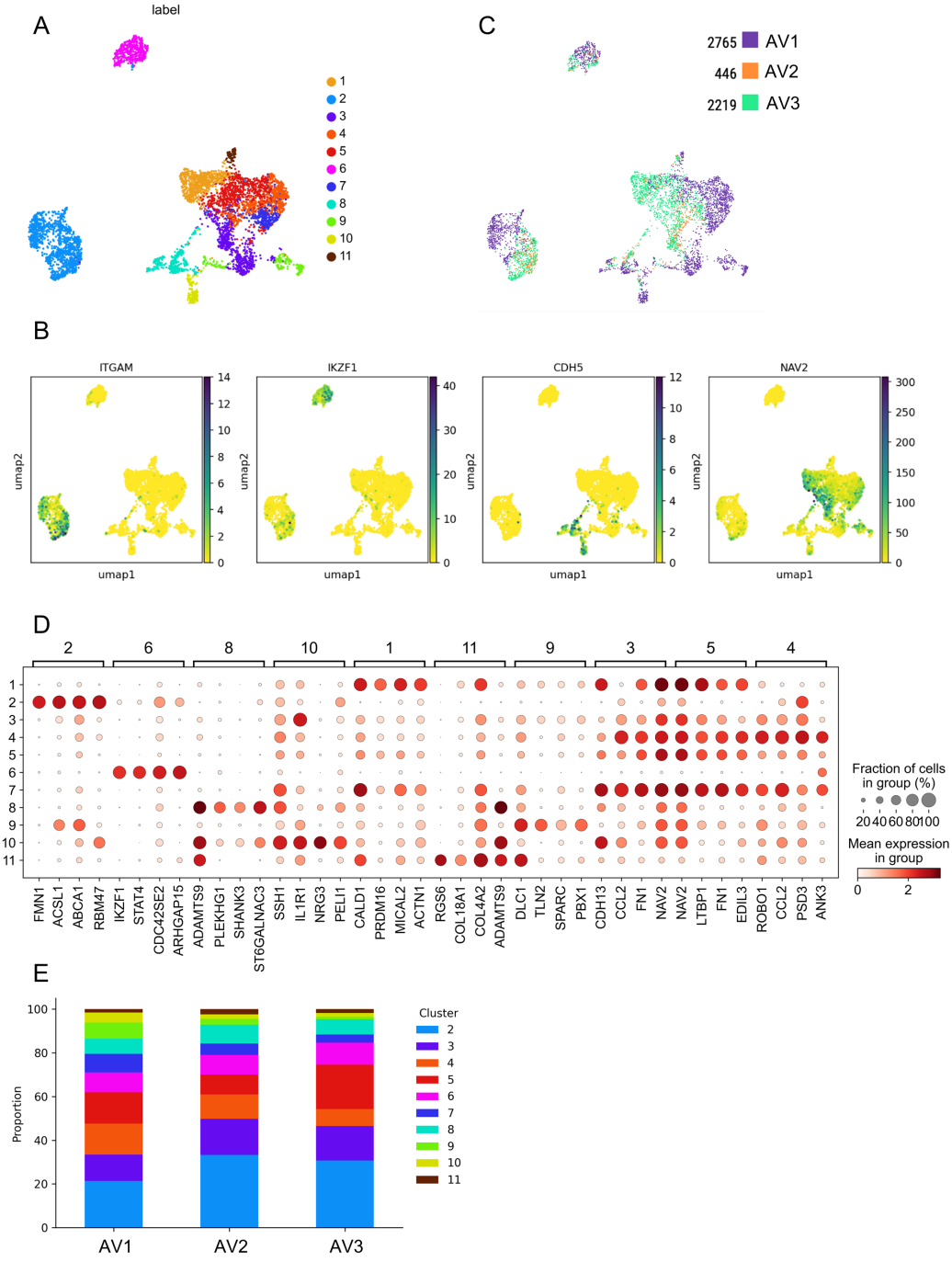
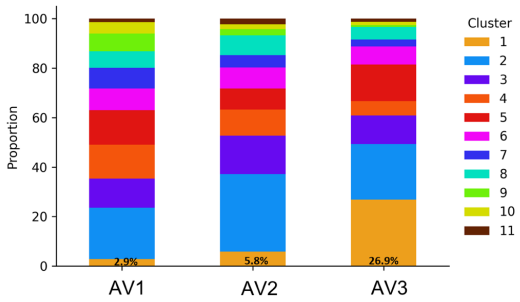
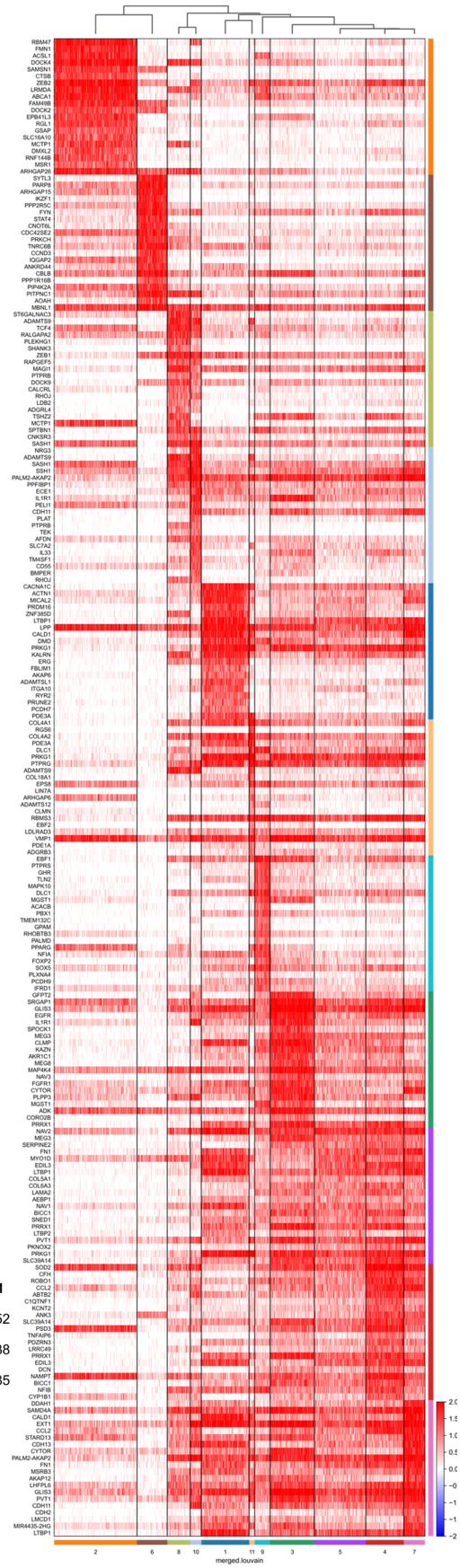

A
label
1
2
3
4
5
6
7
8
9
10
11
C
2765 AV1
446 AV2
2219 AV3
B
ITGAM
IKZF1
CDH5
NAV2
umap1
umap2
D
2 6 8 10 1 11 9 3 5 4
FMN1 ACSL1 ABCA1 RBM47 IKZF1 STAT4 CDC42SE2 ARHGAP15 ADAMTS9 PLEKHG1 SHANK3 ST6GALNAC3 SSH1 IL1R1 NRG3 PELI1 CALD1 PRDM16 MICAL2 ACTN1 RGS6 COL18A1 COL4A2 ADAMTS9 DLC1 TLN2 SPARC PBX1 CDH13 CCL2 FN1 NAV2 NAV2 LTBP1 FN1 EDIL3 ROBO1 CCL2 PSD3 ANK3
Fraction of cells in group (%)
20 40 60 80 100
Mean expression in group
0 2
E
Proportion
Cluster
AV1
AV2
AV3

Figure S5

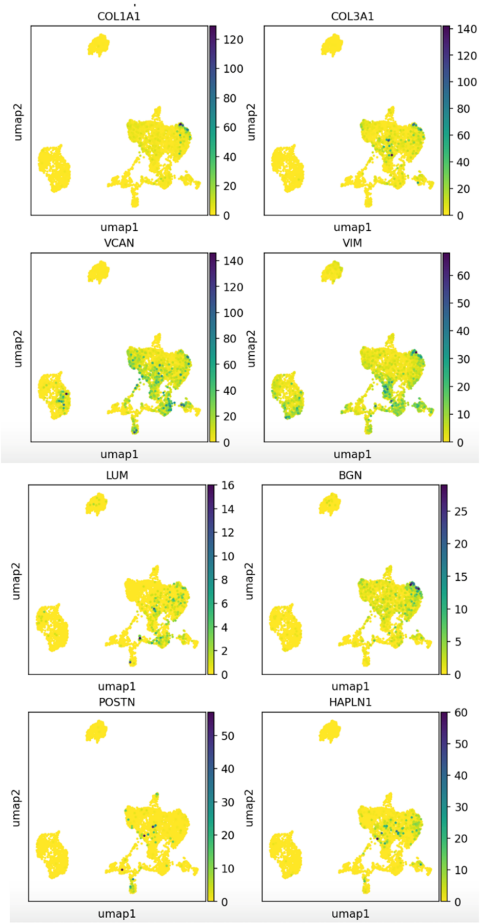
A



B



C



D

label	2	3	4	5	6	7	8	9	10	11
AV_1	21.42	12.01	14.17	14.36	9.00	8.51	6.95	7.32	4.68	1.52
AV_2	33.33	16.43	11.19	9.04	9.04	5.23	8.67	2.61	2.14	2.38
AV_3	30.70	15.78	7.82	20.28	10.04	3.69	6.96	1.10	1.72	1.85