Title

Multiomic analysis reveals developmental dynamics of the human heart in health and disease

James Cranley[1]*, Kazumasa Kanemaru[1]*, Semih Bayraktar[2,3]*, Vincent Knight-Schrijver[2,3], Jan Patrick Pett[1], Krzysztof Polanski[1], Monika Dabrowska[1], Ilaria Mulas[1], Laura Richardson[1], Claudia I. Semprich[1], Rakeshlal Kapuge[1], Shani Perera[1], Xiaoling He[4], Siew Yen Ho[5], Nadav Yayon[1], Liz Tuck[1], Kenny Roberts[1], Jack A. Palmer[1], Hongorzul Davaapil[2,3], Laure Gambardella[1,2,3], Minal Patel[1], Richard Tyser[2,3], Sanjay Sinha[2,3#], Sarah A. Teichmann[1,6#]

Affiliations:

1 Cellular Genetics Programme, Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, UK

2 Wellcome-MRC Cambridge Stem Cell Institute, Jeffrey Cheah Biomedical Centre, Cambridge Biomedical Campus, University of Cambridge, Cambridge, UK

3 Department of Medicine, University of Cambridge, Cambridge, UK

4 John van Geest Centre for Brain Repair, Cambridge University, Cambridge, UK

5 Cardiac Morphology Unit, Royal Brompton Hospital and Imperial College London, UK

6 Department of Physics, Cavendish Laboratory, University of Cambridge, Cambridge, UK

* denotes co-first author

# denotes co-corresponding and co-senior author: ss661@cam.ac.uk, st9@sanger.ac.uk

Abstract

Developmental dynamics encompass both the specification of cell types and their spatial organisation into multicellular niches. Here we harness the power of single-cell and spatial multiomics to unravel embryonic and foetal cardiac tissue niches, which lead to the development of a new tool, TissueTypist. We infer developmental cell trajectories, including evidence for lineage relationships based on shared somatic mutations, within first- and second-trimester human hearts. We reveal that cardiac-resident macrophages likely originate from the yolk sac, forming heterogeneous subsets. CX3CR1^{+} macrophages with a microglia-like profile localise in the sinoatrial node, which may contribute to axon guidance for the innervating autonomic neurons. Foetal pacemaker cells exhibit distinct characteristics compared to their adult counterparts, including the expression of genes that are known to promote parasympathetic innervation. By comparing somatic mutation profiles of cardiomyocytes, we identify an early branching point where pacemaker cells diverge from working cardiomyocytes. We highlight the enhancer-mediated gene regulatory networks governing atrial and ventricular cardiomyocyte specification. The maturation of atrial cardiomyocytes into distinct left and right phenotypes, driven by transcription factors linked to atrial septal defect genes, underscores the significance of this process for healthy heart development. In the ventricle, cellular and transcriptional gradients along both pseudotime and the transmural axis provide a new molecular understanding of myocardial compaction. Finally, generating data from Trisomy 21 hearts and comparing this with the euploid atlas, we reveal a reduced abundance of specific cell types including compact cardiomyocytes. Overall, this extensive dataset and our precomputed models will form a valuable resource for the field.

Introduction

The development of the heart involves the orchestrated activities of multiple cell types, derived from cardiac mesoderm and neural crest, as well as immune lineages from the yolk sac and liver. These cells work together to regulate critical aspects of heart development, including the specification and maturation of cardiomyocytes, formation of coronary vasculature to supply energy to the growing heart and control of heart rate. These complex developmental processes are sensitive to genetic, epigenetic, and environmental perturbation, manifesting as congenital heart disease (CHD), a collection of conditions affecting 1% of live births and accounting for 4% of infant mortality[1].

Highlighting this sensitivity, the altered gene dosage in Trisomy 21 leads to an approximate 50-fold increase in this risk[2]. Although CHD-causal genes have been identified, more than half of CHD patients will receive no genetic diagnosis[3]. Furthermore, mutations in causal genes, which are heavily enriched for transcription factors[4], lead to pleiotropic lesions, pointing to complex networks responsible for the formation of structures across the heart. Together this underlines the need for a better understanding of human cardiogenesis with a greater emphasis on capturing the important, cell type-specific gene regulatory networks and spatiotemporal dynamics at play.

Our knowledge of the cellular and molecular mechanisms underlying cardiac cell origin, specification and maturation in humans is still limited. Many of our insights are either based on *in vitro* systems or animal models[5]. Studies of human pluripotent stem cell-derived cardiomyocytes (hPSC-CM) have uncovered novel regulators and insights into maturation, however, such models can be misleading and caution must be taken when extrapolating findings to *in vivo* development[6]. Recent studies applying 'omic' profiling to human foetal samples have yielded insights into the regulation and spatial patterning of gene expression in the developing heart[7,8]. However, these have been derived from a limited number of heart samples and are generally restricted to a single modality. More comprehensive datasets in terms of developmental span, modality, and spatial resolution are required to shed light on the processes underlying cardiac development.

We therefore constructed a multimodal reference atlas of the developing human heart, which is the largest of its kind. Comprising single-cell (paired RNA and ATAC-seq) and spatial transcriptomic data from euploid hearts spanning the first and second trimesters, as well as trisomy 21. The temporal breadth of this atlas enables us to study cell type maturation and origin. In the myeloid compartment, yolk sac-derived macrophages and later liver-derived monocytes contribute differentially to macrophage subsets, with distinct functional profiles and spatial localisations. Using the spatial data we describe the sinoatrial node niche, which also involves parasympathetic neurons and microglia-like macrophages, and show that it exhibits a burst of intercellular signalling supporting innervation, which wanes by the second trimester.

By leveraging the rich data covering both coding and non-coding regions in our joint snRNA- and ATAC-seq data, we defined the enhancer-mediated gene regulatory mechanisms governing foetal pacemaker cell identity. These data allow us to employ lineage tracing by somatic mutation calling, which clarifies that pacemaker cells diverge from working cardiomyocytes in the developing cardiac conduction

system early in cardiogenesis. We defined the regulatory networks governing the transition of immature atrial cardiomyocytes as they adopt mature left and right phenotypes over the first trimester.

Constructing a spatial common coordinate framework (OrganAxis[9]), a transmural axis, allows us to define the changing distribution of cell types across the ventricular wall during the development and compaction. We integrated single-cell data from trisomy 21 hearts with the euploid atlas, identifying a depletion of specific cardiomyocyte subsets in the trisomic hearts. This depletion may be due to altered activity of key regulons and cell-cell interactions as compared to euploid cells and tissue niches.

Finally, we provide precomputed gene regulatory networks, scArches reference model and a new “TissueTypist” tissue niche model repository, which can be used to map, annotate and analyse newly generated data going forward.

Results

Chromatin dynamics in the developing human heart

Here, and in our accompanying paper, we collected a total of 23 human foetal hearts between the ages of 4 and 20 post-conception weeks that underwent single-cell RNA or single nuclei Multiome (paired RNA (**see Bayraktar *et al.***) and ATAC) sequencing (Fig. 1A, Supp Fig. 1A). The hearts were dissected into several anatomical structures to gain regional insights and the dataset was further supplemented with Visium spatial transcriptomics from 6 hearts (22 capture frames). After quality control of gene expression data, a total of 297,473 cells and nuclei were retained to yield six coarse-grained, 14 mid-grained and 63 fine-grained cell types (**see Bayraktar *et al.***). Of these, 167,022 Multiome nuclei from 11 hearts that span 4 to 20 PCW also passed quality control for ATAC-seq.

Chromatin-accessible regions (peaks) were identified for each fine-grained cell type, and by taking the union, a set of 508,040 peaks was obtained. 66% of the peaks overlapped (more than 100 bp overlap per peak) with the ENCODE candidate cis-regulatory elements (Registry V3)[10]. We found good agreement between ATAC and RNA features within cells, since the transcriptionally identified cell types were also separated in the accessibility profile UMAP embedding (Fig. 1B, Supp. Fig. 1B). When considering the fine-grained cell types within a given coarse-grained compartment, we noted that the peaks shared between them were enriched for promoters, whereas the private peaks were enriched for enhancers (distal and intronic peaks), suggesting that enhancers fine-tune cell identity[11](Supp. Fig. 1C).

The impact of cis-regulatory elements and epigenetic regulation on gene expression during cardiac development and cell type specification is significant[12,13]. Integrating single-nucleus RNA and ATAC-seq data enables us to explore the epigenetic stability, indicative of the chromatin status dynamics. To investigate the epigenetic stability, we initially defined overlapping neighbourhoods[14] in both the RNA and ATAC spaces. For each RNA neighbourhood, the entropy of the corresponding nuclei distribution across the ATAC neighbourhoods was calculated (Supp. Fig. 2A). As a result, in the cardiomyocyte populations, working cardiomyocytes including cycling populations exhibited elevated entropy scores (indicating reduced epigenetic stability) relative to cardiac conduction system (CCS) cells, such as sinoatrial pacemaker cells or distal ventricular conduction system cells (Fig. 1C, D and Supp. Fig. 2B). This suggests that CCS cells have more stable and established epigenetic status, whereas working cardiomyocytes are undergoing chromatin remodelling. It aligns with prior studies that underscore the vital role of epigenetic regulation in cardiomyocyte development[15,16]. The entropy scores of monocyte and macrophage populations were highest among the myeloid cell types (Fig. 1C, D), potentially indicating their enhanced reactivity and dynamic development, influenced by the tissue environment and internal stimuli.

Spatial transcriptomics reveals multicellular niches

To explore the cellular niches in the developing human heart, we first mapped our cell type annotations to spatial transcriptomic data using cell2location[17], inferring the abundance of each cell type per spot (Supp. Fig. 1A). We then integrated the spot-by-cell type abundance data (39,563 spots) from 22 Visium sections (Fig. 1E). These spots were annotated with tissue structures based on clustering, estimated abundance of cell types in each cluster, and histological structure as defined by expert annotations using the associated H&E images (Fig. 1E, F). This analysis showed shared structural gene signatures across chronological age, which highlights the consistency of our spatial dataset.

Each structure showed a unique pattern of cell-type abundance, which can be interpreted as a multicellular niche (Fig. 1G). The SAN showed enrichment of pacemaker cells, as well as CX3CR1^{+} macrophages and parasympathetic neurons. LYVE1^{+} tissue-resident macrophages were highly abundant in vessel-associated structures (adventitia of coronary and great vessels) and also in the epicardium. The ventricular compact myocardium showed a high abundance of coronary vasculature cells, such as capillary endothelial cells and pericytes. The adventitia of the great vessels, unlike coronary vessels, was highly enriched in neuron progenitors. This may reflect the later development of the cardiac sympathetic nervous system, which has enveloped the aortic root but is yet to fully extend along the coronary arterial walls[18].

We found a higher abundance of parasympathetic compared to sympathetic neurons across various tissues, including the node, atrium, and ventricle, particularly at earlier stages (Supp. Fig. 3A). In the single-nucleus RNA-seq data (**see Bayraktar *et al.***), both parasympathetic and sympathetic neurons were predominantly found in atrial tissue samples, and sympathetic neurons were enriched in aorta samples specifically (Supp. Fig. 3B). These findings suggest that ganglionated plexi are primarily localised in the atrial tissues and, notably, parasympathetic axons begin to innervate the entire heart early in development (4-5 PCW). This is in keeping with previous animal studies which suggest parasympathetic innervation begins in the first trimester and precedes sympathetic innervation[18]. It also provides a developmental basis for the observation that nodal innervation is predominantly parasympathetic in the post-natal heart[19].

A total of nineteen tissues were annotated, including seven myocardial and five vascular structures. Having defined this set of tissue structures, we developed a logistic regression model, 'TissueTypist', which is trained on their transcriptomic signatures. This approach enabled us to annotate tissues in new spatial datasets in a reproducible manner (Supp. Fig. 3C). The model showed excellent robustness and reproducibility in classifying tissue structures (Supp. Fig. 3D).

Cardiac macrophage development

Macrophages are critical in the embryonic development of various tissues[20] including the heart[21]. However, in humans, tissue localisations and potential functions of cardiac macrophages in development are not fully clarified. To uncover this, we utilised gene expression and chromatin accessibility data of macrophages and monocytes and created an integrated embedding using multiVI[22] (Fig. 2A). The principal graph was

calculated based on the multiVI representation, and the pseudotime was obtained by setting the youngest milestone as the root (Supp Fig. 4A). Recently, we described monocyte-independent macrophage differentiation in the human yolk sac (YS)[23] that seeds developing organs. To test whether the embryonic heart contains macrophages originating from this route, we trained a CellTypist model with the YS dataset and transferred the labels to our dataset. The population found in the youngest milestone (~5PCW) was predicted to be YS premacrophages (YSpMPs) with high probability (Fig. 2B, Supp Fig. 4B). High activities of the YSpMP regulons (FLI1, MEF2C)[23] at the youngest milestone also suggest that the YSpMPs populate the early-stage heart (Fig. 2C). Based on this finding, we added the "MacrophagesYSderived" label to the cardiac monocyte and macrophage populations identified based on their gene expression profiles (Fig. 2D) (**see Bayraktar *et al.***). Cell abundance testing using Milo[14] across chronological age suggested that YS-derived macrophages were the youngest population, followed by CX3CR1$^+$ macrophages (Fig. 2E). CX3CR1$^+$ macrophages were predicted as a "microglia-like" population using the YS CellTypist model and expressed microglia marker genes (Fig 2F,G), including *CX3CR1*, *TREM2*, *P2RY12*, *ADGRG1* and *C3*.

In contrast, a LYVE1$^+$ macrophage population included cells from a wide range of different chronological ages (Fig. 2E), and was inferred to arise from two separate trajectories: YS-derived macrophages and monocytes differentiating towards mature macrophages (Supp Fig. 4A). We subsetted the cells/nuclei belonging to the trajectory from YS-derived macrophages (Supp Fig. 4A, red arrow) to understand the molecular mechanisms of cardiac macrophage maturation. Amongst these cells, chronological age was correlated with the calculated pseudotime (Fig. 2H), reinforcing the notion that the pseudotime captures the maturation process. The top MSigDB Hallmark[24] pathway enriched in the pseudotime-associated receptor ligand genes was angiogenesis, containing genes such as *VEGFA*, *SPP1*, *APP* and *CXCL8* (Fig. 2H, Supp Fig. 4C). Spatial transcriptomic analysis showed colocalisation of LYVE1+ macrophages and vessels, such as coronary and lymphatic vessels (Fig. 1G, 2I), consistent with a previous study in mice[25]. We validated their localisation in coronary vessels using multiplex smFISH (RNAscope) (Fig. 2J). Finally, cell-cell interaction analysis revealed potential ligand-receptor interactions, suggesting that these LYVE1+ macrophages may facilitate angiogenesis through paracrine communication with smooth muscle and endothelial cells, including VEGFA signalling[26] (Fig. 2K). These results bring new clarity to the emergence and role of cardiac macrophages in the developing heart (Fig. 2L).

Cell circuits of the developing sinoatrial node

The SAN niche showed enrichment of pacemaker cells, parasympathetic neurons, and CX3CR1$^+$ microglia-like macrophages (Fig. 1G, 3A). To investigate the development of pacemaker cells, we initially examined differentially expressed genes (DEGs) of the foetal SAN pacemaker cells by comparing them to the other cardiomyocytes (Fig. 3B). Pathway enrichment analysis of genes upregulated in SAN pacemaker cells highlighted Gene Ontology (GO) processes in SA node and pacemaker cell development pathways, as well as core electrophysiological properties of pacemaker cells (Supp Fig 5A). Additionally, in our ATAC-seq data, we found that SAN pacemaker cells contained the largest number of cell type-specific accessible regions when compared to the other cardiomyocytes (Supp Fig 5B). These regions were linked to genes associated with biological processes involving SA node

development and pacemaker cell activity (Supp Fig 5C). We then constructed an enhancer-mediated gene regulatory network (GRN) targeting the identified pacemaker cell gene signatures using SCENIC+[27] and Multiome data (Supp Table 3). Centrality of the TF interactions highlighted well-described TFs for pacemaker cell development[28], such as ISL1, SHOX2, TBX3, TBX5, TBX18, as hubs within the network (Supp Fig 5D). The network selected for the interactions targeting pacemaker-specific ion channel genes, including *HCN1*, *HCN4*, and *CACNA1D*, also showed the high centrality of the well-described TFs (Fig. 3C).

We then sought to explore foetal-specific properties of pacemaker cells by comparing the expression of the 620 upregulated pacemaker genes between foetal and adult pacemaker cells using our recent reference[12] (Fig 3B). We found 143 foetal-specific genes encompassing those involved in axon guidance processes, such as semaphorins and slit genes (*SEMA3A*, *SLIT2*, *SLIT3*)[13] (Supp Fig 5A). Visium data showed expression of these genes in the SAN (Supp Fig 5E). We noted that these genes were highly expressed in the first-trimester hearts, with downregulation in the second-trimester and adult hearts (Fig. 3D). To understand where guidance cues may be received, we explored their receptors within the cell types of the SA nodal niche (Fig 1G). We found that the parasympathetic neurons highly expressed several cognate receptors, including *CXCR4* (Fig. 3E), suggesting these as neuronal cell recruitment signals from pacemaker cells. Conversely, we observed trans-synaptic interactions and neurotransmitter signalling, such as acetylcholine, from parasympathetic neurons to pacemaker cells (Supp Fig 5F), which may regulate the beating rate of the developing heart. To understand the GRNs governing axon guidance gene expression in pacemaker cells, we applied SCENIC+. The constructed network showed both TFs directly targeting these genes as well as high-level TFs, which may indirectly orchestrate expression (Fig. 3F). We noted that the ligand genes *SEMA3A*, *SLIT3*, *EFNB2*, and *CXCL12* were targeted by multiple TFs, suggesting that their regulation may involve cooperative TF interactions and/or redundancy mechanisms.

In keeping with the CX3CR1+ macrophage localisation in the SA node (Fig 1G), *CX3CL1* expression was specific to the parasympathetic neurons and cells of the vasculature (Supp Fig 5G). This suggests CX3CR1^{+} macrophage enrichment in the SA node is promoted by *CX3CL1* expression in parasympathetic neurons, where the same interaction causes efficient migration of microglia to neurons in the CNS[29]. We also observed the expression of other ligands and receptors, including IGF1-IGF1R, CD200-CD200R1, CD47-SIRPA, P2RY12, and TREM2, which are known to play a role in neuronal homeostasis mediated by microglia[30] (Fig. 2G, Fig 3G). Aligning with the temporal dynamics of axon guidance ligand expression, the abundance of CX3CR1^{+} macrophages was highest in the first-trimester hearts and decayed in the second trimester (Fig. 2E). In the adult heart, macrophage populations do not exhibit expression of *CX3CR1* or any other markers typically associated with microglia (Supp. Fig. 5H).

Recently it has been suggested that para- and autocrine glutamatergic signalling may regulate firing rate in working cardiomyocytes pacemaker cells[31,32]. Amongst the genes upregulated in adult compared to foetal pacemaker cells (Fig. 3B), we found several genes supporting glutamatergic signalling (*SLC1A3*, *SLC17A5*, *GRIA3*, *GRID2*). This supports a role for glutamatergic signalling in adult pacemaker cells and suggests that this capability is obtained after the second trimester (Supp. Fig. 5I).

In summary, foetal pacemaker cells, unlike their adult counterparts, show lower expression of glutamatergic signalling machinery and express a series of genes promoting vagal innervation, particularly in the first trimester. These data outline the foetal SA nodal niche for the first time, showing evidence of crosstalk between colocalised pacemaker cells, parasympathetic neurons as well as microglia-like CX3CR1+ macrophages (Fig. 3H).

Mutational tracing for ordering the separation between working and conduction cardiomyocytes

Lineage tracing studies, most commonly in transgenic mouse lines, have outlined the developmental origins of various cardiomyocyte populations[33]. For instance, the myocytes of the murine SAN derive from a *TBX18*^{pos}/*NKX2-5*^{neg} population at the venous pole of the inflow tract, subsequent to left ventricular myocytes, and with neighbouring cells destined to become atrial cardiomyocytes [28]. However, since these methods are not applicable to human cardiogenesis, it remains uncertain whether the specification of pacemaker cardiomyocytes happens before or after the specification of atrial or ventricular working myocytes (Fig. 4A).

To address this, we hypothesised that somatic mutations could be used as naturally occurring barcodes: accrued during cell division and inherited by progeny in a lineage-restricted manner. To infer somatic mutations we used SComatic[34]. This approach is similar to DEsceRNAMut[35], but enables mutation calling in both RNA and ATAC-seq data. It overcomes the sparsity of single-cell data by aggregating reads from all the cells belonging to each cell type in a given sample. A binary mutation matrix is generated that associates cell types with identified mutations (Methods, Supp. Fig. 6B). This allows the construction of a neighbour-joining tree which describes the relationship between each cell type in terms of somatic mutations. The tree was rooted using a pooled population of neural cell types, since the common ancestor between neural (ectoderm) and cardiomyocytes (mesoderm) is prior to gastrulation. Individual trees were made for each donor (Supp. Fig. 6A), and by concatenating the mutation matrices of multiple donors, we generated a consensus tree (Fig. 4B).

These findings support the view that pacemaker cardiomyocytes separate from working cardiomyocytes prior to atrial/ventricular or compact/trabeculated cardiomyocyte specification. In terms of mutational difference, working atrial and ventricular populations are all similar, in keeping with their more recent development. This is in contrast to their transcriptional differences, which are substantial (Supp. Fig. 6C,D).

Atrial cardiomyocyte maturation

Two transcriptionally distinct populations of working atrial cardiomyocytes (aCMs) were identified, and Visium mapping reinforced that these correspond to left and right atria (AtrialCardiomyocytesRight and Left; RaCMs, LaCMs) (Fig. 4C, Supp. Fig. 7A). Marker genes for the LaCMs (*PITX2* and *PANCR*: PITX2-associated non-coding RNA) have been highlighted in human and mouse left atrial tissue (**see Bayraktar *et al.***) [36,37], and are important in repressing the pacemaker gene programme in the left horn of

the sinus venosus[28]. The PITX2-adjacent 4q25 locus is a well-known risk locus for atrial fibrillation [28,33]. RaCMs exhibit greater expression of genes involved in axon guidance and innervation (Supp. Fig. 7B), including Neurotrimin (*NTM*), *SLIT2* and *3* and Semaphorin 5A, all of which increased in expression in the second trimester (Supp. Fig. 7C). The cognate receptors were expressed by both sympathetic and parasympathetic neurons (Supp. Fig. 7D). This difference in innervation-promoting signalling aligns with the greater density of autonomic innervation in the right atrium post-natally [38].

To explore the maturation of atrial cardiomyocytes we constructed a multiVI embedding using both RNA and ATAC-seq data (Fig. 4C). This revealed LaCM and RaCM profiles in younger hearts are relatively similar, and that the distinction between the two profiles becomes more pronounced during development (Fig. 4C). Mapping the signatures of the immature (< 7PCW) and mature L and RaCMs in Visium from multiple developmental ages (Supp. Fig. 7E) showed a clear trend where the immature state is gradually replaced by its mature counterpart (Fig. 4E), supporting the age-dependent transition seen in the single-cell data. The increased divergence between L and RaCMs over time thus allows us to explore distinct maturation trajectories which promote atrial laterality during human heart development.

A pseudotemporal ordering was created based on the multiVI latent space, which correlated highly with chronological age (Fig. 4D, Supp. Fig. 7F). To understand the mechanisms that govern the initial divergence of the L and RaCM expression profiles during maturation, we first identified 'early genes' that were upregulated at the earliest stage of each trajectory (Supp. Fig. 7G). For the LaCM trajectory, the early genes included *PITX2*, *PANCR*, *SLC8A3* (encoding a 'non-cardiac' Na-Ca exchanger NCX3), *MBP* (myelin basic protein), *KCNMB2* (a calcium-sensitive potassium channel crucial for modulating neuronal excitability). Early genes in the RaCM trajectory included *ANGPT1* (encoding angiopoietin-1, which plays an important role in angiogenesis) and the transcription factor gene *MAF*, which to our knowledge has no ascribed cardiac role. The expression of several early genes (*PITX2*, *PANCR*, *SLC8A3*, *KSR1*, *ANGPT1*, *RFTN1*, *MAF*) is largely restricted to foetal aCMs and is lower or absent in adult cardiomyocytes [32] (Supp. Fig. 7H). We found that the identified early genes were significantly enriched amongst the target genes of several GRNs inferred for aCMs. This includes TCF4, PBX3 and MEIS2 for the RaCM trajectory, and PITX2, TEAD1 and GATA4 for the LaCM (Fig. 4F). Cross-referencing this list of transcription factors with a curated CHD database [4] identified four transcription factors (PITX2, GATA4, MEIS2, SMAD3) implicated in congenital heart disease, with PITX2, GATA4, and MEIS2 specifically linked to atrial septal defects. *MEIS2*, the only CHD gene whose regulon was significantly enriched in RaCM trajectory, is linked to right-sided congenital heart lesions (Ebstein's anomaly and hypoplastic right ventricle) [39]. This suggests that the early phenotypic divergence of LaCMs and RaCMs may go awry in congenital heart disease.

To understand the dynamics during maturation, we identified genes significantly associated (increasing or decreasing) with pseudotime in each trajectory (LaCM trajectory, RaCM trajectory, or both "common"). This showed a common loss of pacemaker-type ion channel expression (*KCNJ3*, *TRPM3*), with an increase in genes conferring a working phenotype (*SCN5A*, *GJA1, CASQ2*) (Supp. Fig. 7I). To compare differences between the trajectories we then used genes2genes, which identifies feature-level alignment between two trajectories [40] (Supp. Fig. 7J). For the processes common to L and RaCM maturation we took the high-similarity genes (>30%), and divided them into those which increase or

decrease along the trajectories. Gene set enrichment analysis revealed a significant increase in 'myogenesis' and a decrease in both 'Myc targets' and 'oxidative phosphorylation' (Fig. 4G). This is consistent with aCMs diminishing proliferative potential in favour of contractility, while suggesting cardiomyocyte metabolic activity is heightened early in development due to active proliferation. Examination of low-similarity genes specific to each trajectory can give insights into trajectory-specific biology. In the RaCM trajectory these genes highlighted cellular migration and patterning with Receptor Tyrosine Kinase-Like Orphan Receptor 1 (*ROR1*) and plexin C1 (*PLXNC1*) (Supp. Fig. 7K). Amongst the cardiomyocyte subtypes, foetal and adult pacemaker cells expressed the genes encoding the ligands for ROR1 (*WNT5A*) and plexin C1 (*SEMA7A*) most highly, suggesting these may enable interactions between pacemaker cells and the surrounding RaCMs (Supp. Fig. 7L). LaCM genes included *VASH2* which encodes a secreted ligand promoting angiogenesis.

Finally, we used genes2genes to compare regulon activities between the trajectories. Looking for regulons with low similarity across the whole trajectory, which potentially exert an early and sustained influence during development, highlighted eleven highly mismatched regulons, nine of which had a higher mean activity in the LaCM trajectory (Fig. 4H, I, Supp. Fig. 7M). Among these, the PITX2 regulon again appeared, underlining its importance in establishing LaCM identity. Interestingly, the activating and repressing activities of two regulons (ETS and MAF) showed reciprocal activity in the two trajectories. This suggests that these TFs may be bifunctional, promoting divergence towards two different cell identities.

In summary, although the morphological atria have formed prior to our earliest sample (around post-conception day 28), we show that atrial cardiomyocytes continue to mature and the L and RaCM populations become more distinct with development.

Compact and trabeculated myocardium development

Consistent with a recent study[41] we identified both compact and trabeculated cardiomyocyte populations in our suspension data and validated their separation through spatial mapping, revealing a consistent overlap with their respective histological structures in corresponding H&E images (Fig. 5A). To analyse the spatiotemporal transition of compact and trabeculated cardiomyocytes across the ventricular wall, we defined a transmural axis between manually annotated epicardial and endocardial layers of the left ventricle using OrganAxis[9] (Fig. 5B, C, Supp. Fig. 8A). This provided a common spatial coordinate framework to compare features, such as genes or cell types, across developmental time points from multiple slides of different sizes of hearts. During development, we found the abundance distribution of compact cardiomyocytes along this axis shifted from a tight distribution at the epicardium, to a more dispersed distribution closer towards the endocardium, while the opposite shift was observed for that of the trabeculated cardiomyocytes (Fig. 5D). Additionally, nuclear density was seen to be highest in the older hearts, particularly towards the epicardial boundary (Fig. 5E, Supp. Fig. 8B). Analysis of gene expression across the transmural axis revealed an enrichment of genes associated with cardiomyocyte

proliferation on the epicardial side. (Supp. Fig. 8C, D). Collectively, these findings suggest that the compact cardiomyocyte population expands in the second-trimester heart.

Of note, the compact cardiomyocytes showed a similar spatiotemporal localisation pattern to coronary pericytes and capillary endothelial cells (Fig. 5C, F) consistent with the cellular niche analysis (Fig. 1G). Cell-cell interaction inference suggested that compact cardiomyocytes promote angiogenesis *via* VEGFs directed towards these coronary capillary endothelial cells (Fig. 5G). On the other hand, trabeculated cardiomyocytes were localised with ventricular conduction system (VCS) cells, endocardial cells, and myofibroblasts at the endocardial boundary (Fig. 5C). Parasympathetic neurons were also enriched at the endocardial side (Fig. 5C), which aligns with the upregulation of genes involved in *neuron differentiation* and *axon guidance* at the trabeculated region of the transmural axis (Supp. Fig. 8C).

To clarify the molecular mechanisms involved in the maturation of compact and trabeculated cardiomyocytes, we utilised gene expression and chromatin accessibility data to create an integrated embedding of left ventricular cardiomyocytes (Fig. 5H). The integrated embedding showed the independent temporal transition for compact and trabeculated populations (Fig. 5H). Pseudotime was calculated by specifying the youngest tip principal point as the root, and the resulting pseudotime assignment correlated with the post-conception age of the samples (Fig. 5H, Supp. Fig. 8E, F). The genes associated with pseudotime in the compact and trabeculated cardiomyocyte trajectories were selected, and the common upregulated gene set was enriched with the genes involved in the pathways related to vCMs maturation, such as *response to calcium ion* or *regulation of heart contraction* (Fig. 5I). The compact-specific signature was enriched for processes involved in lipid transport and metabolism, such as *CD36*, *OSBPL2*, *PITPNC1* and *SREBF2,* suggesting a preparatory mechanism for a metabolic switch (Fig. 5I)[42]. In addition, the gene expression of a major angiogenesis growth factor, *VEGFA*, was upregulated, as included in the *venous blood vessel morphogenesis* pathway (Supp. Fig. 8G), suggesting that compact cardiomyocytes promote increased vascularisation in preparation for greater metabolic demand.

Within the GRNs targeting common and compact cardiomyocyte maturation, the centrality of the TF interactions showed several TFs as hubs within the network, such as *PRDM16, TBX20*, *PPARA*, *GATA4* and *GATA6* (Fig. 5J, Supp Table 5) which have been implicated in cardiomyocyte maturation [43,44]. Our analysis also highlighted TFs that are less well established within cardiomyocytes, including *SREBF2*, which encodes SREBP2. SREBP2 is a key regulator of sterol and fatty acid synthesis[45], a metabolic pathway that plays a crucial role in cardiomyocyte maturation. To predict the effect of TFs in the maturation of compact cardiomyocytes, we performed an *in silico* TF perturbation analysis using CellOracle[46]. This highlighted the implication of several transcription factors, including *HEY2, PRDM16*, *NFE2L1*, and *SREBF2* in the maturation of left compact vCMs (Fig. 5J, K).

Trisomy 21 disrupts compact myocardium formation

Congenital heart disease (CHD) encompasses a wide variety of cardiac structural defects affecting approximately 1% of live births[47]. Its aetiology is complex with both genetic and environmental drivers.

392 Trisomy 21 (T21) is a strong risk factor for CHD, with a frequency of between 40% and 50% amongst
393 T21 live births and confers a particular risk for septal defects (interatrial, interventricular, or combined)[2].
394 Although some specific genes have been implicated[48], the mechanisms of CHD downstream of
395 aneuploidy, and particularly the cellular basis of this risk in humans are poorly understood.

396 We sought to shed light on this using a comparative approach, generating Multiome and Visium data
397 from T21 hearts (aged 13-14 PCW). After quality control, there were 76 thousand nuclei which were
398 integrated into the euploid atlas using scArches[49], and annotations were then provided using a
399 CellTypist[50] model trained on the euploid atlas (Fig. 6A). As a measure of difference between euploid cell
400 types and their T21 counterparts, we compared their inter-centroid distance in both gene expression and
401 ATAC space. This revealed that T21-induced differences were not equal across cell types, with the
402 biggest differences in the cardiomyocytes (compact ventricular and left atrial), natural killer cells, and
403 adventitial fibroblasts (Fig. 6B). To inspect the differential abundance of cell types between euploid and
404 aneuploid datasets, we used Milo[14] on age-matched data from both groups. The most striking difference
405 observed was the depletion of left and right compact vCMs (Fig. 6C).

406 To assess the alteration in spatial data, we utilised our TissueTypist model, a logistic regression
407 model trained on the tissue structure labels in the euploid Visium dataset (Fig. 1E, Supp. Fig. 3C, D). We
408 used this model to add tissue structure labels to the T21 Visium section and assessed the transition of
409 the "Ventricle - Compact" label across the left ventricular transmural axis using OrganAxis (Supp. Fig.
410 9A). This showed lower probability of the compact myocardium layer prediction in the T21 ventricle
411 compared to the age-matched euploid sections (Fig. 6D). To examine the spatial arrangement of
412 specific cell types across the ventricular wall we used their cell2location mappings. In keeping with the
413 snRNA-seq analysis and TissueTypist prediction, we found a significant reduction of the compact
414 cardiomyocytes, as well as reduced coronary endothelial and pericytes (Fig. 6E), suggesting myocardial
415 vascularisation may also be altered.

416 Having identified the depletion of the compact vCMs in T21, we coupled differential expression
417 analysis to CellPhoneDB to identify the most prominent aberrations in cell-cell interactions between
418 compact cardiomyocytes and other cell types in the presence of T21. This highlighted an increase in
419 signalling directed at the adhesion G-protein coupled receptor *ADGRL3* on MesothelialEpicardialCells,
420 coupled with a reduction in *SEMA3A* signalling directed towards compact cardiomyocytes (Fig. 6F). We
421 then evaluated alterations in active GRNs in compact cardiomyocytes by computing regulon activity
422 scores. This analysis showed a marked activity reduction of the several regulons governed by HEY2,
423 PRDM16, and SREBF2 (Fig. 6G), which we previously inferred to be implicated in the network targeting
424 compact-specific signatures of left compact vCMs (Fig. 5J, K).

425 It has been reported that patients with Down's syndrome have an increased likelihood of patent
426 ductus arteriosus (PDA)[51,52], and we compared DASMCs from T21 samples to those from euploid
427 individuals. We observed a reduced expression of *PDE1C* in DASMCs of patients with T21, suggesting
428 potential perturbation in the mechanism underlying post-natal duct closure, while other DASMC-enriched
429 phosphodiesterases *PDE3A*, *PDE4D*, *PDE5A* were comparable (Supp. Fig. 9B). Examining the cell
430 type-specific GRNs between age-matched euploid and T21 samples, we observed a skew toward T21 in
431 the eigenvector centrality score *TFAP2B* (Supp. Fig. 9C). This finding, along with earlier studies that

432 identified mutations in this gene causing PDA[53,54], highlights the importance of *TFA*P2B in defining the
433 identity of DASMCs.

434 Finally, we sourced a list of CHD-associated genes and compared the expression of these genes
435 between euploid and T21 samples. We observed the most significant expression differences in great
436 vessel SMCs and adventitial fibroblasts, and cardiomyocytes (Supp. Fig. 9D).

437 Overall, we showed impairment of ventricular compaction in the T21 heart together with a defect in
438 coronary vascularisation.

Discussion

Here we contribute towards a comprehensive human cardiac developmental atlas through our multimodal dataset spanning part of the first and second trimesters. Combining gene expression, chromatin accessibility and spatial transcriptomics, we shed light on dynamic processes occurring in human cardiogenesis.

Our investigation into macrophage development showed that cardiac tissue-resident macrophages are derived from two sources: foetal liver monocytes and YS-derived macrophages. The majority of YS-derived macrophages were from the 5 PCW heart, the time hematopoiesis in the aorta-gonad-mesonephros initiates. We identified a population of CX3CR1^{+} microglia-like macrophages, localising in the SAN, which express genes encoding ligands and receptors involved in synaptic maintenance[30]. Recently it was shown that CX3CR1^{+} macrophages populate tissues outside the CNS, including skin, testicles, heart and aorta, during human development[55]. However, their precise localisation and function in the developing heart were not clarified. This study reveals they are part of the developing sinoatrial node niche.

We found foetal pacemaker cells exhibit higher expression of axon guidance genes compared to their counterparts in the adult, and colocalise with parasympathetic neurons in our first trimester Visium sections. The elevation in pacemaker cell axon guidance gene expression and increased abundance of these microglia-like macrophages in the first trimester suggest that this time window is critical for parasympathetic innervation of the sinoatrial node. This period also aligns with the first trimester heart rate decrease which may be partly driven by the suppressive influence of this innervation[56]. We observed clear spatial mapping of the parasympathetic, but not sympathetic neurons in the node, this is consistent with previous animal studies which suggests parasympathetic begins in the first trimester and precedes sympathetic innervation[18].

We present the human foetal pacemaker cell enhancer-mediated GRNs for the first time. This underscores the importance of canonical TFs which have been shown to drive the pacemaker phenotype in the mouse heart, such as TBX3, TBX18, SHOX2 and ISL1[57,58]. However, there are also contrasts with both the adult human pacemaker network, such as the absence of FOXP2, and elegant inferences in zebrafish where the *nrf2* family *suppresses* the pacemaker phenotype[59].

Drawing inferences regarding the origins of human cell populations is challenging due to the limited applicability of lineage tracing methods. Through our analysis of somatic mutations detected in both RNA and ATAC-seq reads, we suggest that pacemaker cells diverge from working cardiomyocytes earlier than the atrial/ventricular specification of working cells. This observation is more aligned with development in chick than in mouse. In the former, pacemaker cell identity is specified directly in the lateral plate mesoderm [60], whereas in the latter, pacemaker cells arise from a population of *Nppa*+ atrial myocytes at the venous pole [57,61]. This may suggest that chick models could provide a more accurate representation for studying congenital heart diseases involving the SAN, such as congenital Sick Sinus Syndrome [62]. These findings are suggestive but not conclusive. The sparse coverage in scRNA and ATAC-seq means that each mutation can only be called in a fraction of the cells belonging to each cell type. Future work

with a better ability to resolve mutational signatures at single-cell resolution, such as scDNA or mitochondrial ATAC-seq will be useful to verify these claims.

The process of cardiac looping leads to the physical emergence of the atria by around 4 PCW, prior to the earliest sample in this dataset [63]. Using an embedding reflecting both gene expression and chromatin accessibility, we show that although they are distinguishable, the left and right atrial cardiomyocyte populations are relatively similar prior to approximately 7 PCW, and after that their profiles diverge markedly as they mature. Compared with the left, the right atrial profile is characterised by axon guidance genes, which may provide a developmental and cellular basis for the known denser innervation of the right atrium[38]. It is also notable that, unlike the SAN pacemaker cells, the axon guidance genes continue to rise in the second trimester, suggesting a later window of innervation. Contrasting these diverging trajectories reveals the gene programmes that begin to differ earliest. Overlapping these with the atrial cardiomyocyte gene regulatory network showed enrichment for transcription factors implicated in congenital heart disease, and ASDs in particular, perhaps suggesting that a failure of aCM phenotypic maturation is a key process in normal atrial septation. We identify the core regulons whose activities differ between these two trajectories. This underscored the importance of PITX2 activity in left atrial cardiomyocyte identity, and also implicates ETS2 and MAF as transcription factors whose activating and repressive networks may drive the divergence between these two trajectories.

We showed the spatial transition of the compact and trabeculated cardiomyocyte layers along the left ventricular transmural axis as well as their maturation over time. The transition of the compact cardiomyocyte was associated with that of coronary capillary endothelial cells (Fig. 5F), and both cell types were similarly disrupted in the Trisomy 21 ventricle. This suggests that the two cell types show interdependence during development. Animal studies also suggest that cardiomyocytes play a crucial role in coronary vasculogenesis in the context of both ischemia and development [64,65].

The expression of genes involved in lipid transport and metabolism, such as *CD36*, *OSBPL2*, *PITPNC1*, *SREBF2* (encoding SREBP2) and *SLC27A1* (encoding fatty acid transport protein), were increased along the compact cardiomyocyte maturation. Lipid metabolism is a crucial hallmark of postnatal cardiomyocyte maturation due to the increase of dietary lipid levels after birth, and contributes to more efficient ATP production compared to carbohydrate metabolism[5]. Our result suggests that the groundwork for this postnatal shift in energy substrate may be laid *in utero*. From the gene regulatory network analysis of compact cardiomyocyte maturation genes, we identified TFs, including *PPARA*, which was a target used to promote maturation of iPSC-derived cardiomyocytes[66]. This underscores the advantage of our study for clarifying cardiomyocyte maturation mechanisms in humans, and potential applications in regenerative medicine.

Trisomy 21 represents a high effect size risk factor for congenital heart disease broadly, and defects of the muscular septum in particular [2]. Having obtained these rare samples we generated high quality single-cell and spatial transcriptomic data from them, which will be a valuable reference resource for the study of congenital heart disease mechanisms. We integrated these data with our euploid single-cell atlas and showed the power of our TissueTypist model for spatial data annotation.

Both spatial and single-cell data independently point to a significant depletion of compact cardiomyocytes in the ventricle wall. Impairment of myocardial compaction has been associated with septal defect formation in clinical and preclinical murine studies[67–70], however the molecular detail and cellular profiles our data provide have been lacking. Our findings suggest that in Trisomy 21 compact cardiomyocytes receive reduced signalling from the neighbouring cells of the epicardium and trabeculated cardiomyocytes, in particular the semaphorin 3A-plexinA4 axis is markedly reduced. This same axis has been recently shown to be critical in compact cardiomyocyte differentiation in vitro[41]. Furthermore, the compact cardiomyocytes express reduced *VEGFA*, which may explain the impairment of vascularisation evident in our spatial data. We also find reduced activity in a host of regulons which were central in the euploid compact cardiomyocyte GRN.

Our data do not highlight a specific gene whose increased dosage may be causal to the Trisomy 21 phenotype. None of the genes encoding transcription factors, whose activities were inferred to be suppressed or enhanced in Trisomy 21 data, were located in chromosome 21. In addition, there was no clear enrichment of chromosome 21 genes among the target genes of each transcription factors. This suggests that the impairment is caused by indirect effects. A recent study in a murine model of Trisomy 21 elegantly implicated the gene *Dyrk1a* as being phenotype-causal in that model, with increased expression seen in the disease state [48]. We did not find any difference in *DYRK1A* expression between euploid and trisomy 21 in our human data, however we note the authors did also find that compact ventricular cardiomyocytes were depleted, suggesting this as a common pathway for muscular septal defect formation across species [48].

Overall, here and in our accompanying atlas (**see Bayraktar *et al.***), we present a high-resolution multi-omic map of the developing human heart during the first and second trimesters. This resource and the insights presented here are anticipated to broaden our understanding of human heart development with a range of implications in disease modelling and regenerative medicine.

Methods

Human embryonic and foetal heart tissues

All tissue samples were obtained after Research Ethics Committee approval and written informed consent from mothers. The embryonic and foetal heart samples corresponding to the donor IDs BRC2251, BRC2253, BRC2256, BRC2260, BRC2262, BRC2263, C82, C83, C85, C86, C87, C92, C94, C97, C98, C99, and C104 were provided from terminations of pregnancy from Cambridge University Hospitals NHS Foundation Trust under permission from NHS Research Ethical Committee (96/085). The donor IDs Hst33, Hst36, Hst39, Hst40, Hst41, Hst42, Hst44, and Hst45 were provided from the MRC/Wellcome Trust Human Developmental Biology Resource (University College London (UCL) site REC reference: 18/LO/0822; www.hdbr.org. Registered project title: A multi-omics analysis of fetal cardiac development and the impact of Trisomy 21, Registered project number: 200556). Donor IDs and that information are listed in Supp. Table 1.

Sample collection and processing

The processing of samples with donor IDs of BRC2251,BRC2253, BRC2256, BRC2260, BRC2262 and BRC2263 were described previously[71]. Briefly, these samples were stored at 4 °C overnight in Hibernate-E medium (ThermoFisher Scientific). The next day, the apex and the base of the heart were dissected and separately dissociated using 6.6 mg/mL Bacillus licheniformis protease (Merck), 5 mM CaCl2 (Merck), and 20 U/mL DNase I (NEB). The mixture was triturated on ice for 20 seconds every 5 minutes until clumps of tissue were no longer visible. The digestion was stopped with ice-cold 10% fetal bovine serum (FBS, ThermoFisher Scientific) in phosphate-buffered saline (PBS, ThermoFisher Scientific). Cells were then washed with 10% FBS, resuspended in 1 mL PBS and viability assessed using Trypan blue. Cells were submitted for 10x library preparation with v3.0 chemistry for 3' single-cell sequencing on a NovaSeq 6000 (Illumina) at the Cancer Research UK Cambridge Institute.

For the processing of donors corresponding to the donor IDs C86, C94, C97 and C99, samples were stored at 4 °C overnight in HypoThermosol preservation solution (Merck). Tissue was first minced in a tissue culture dish using scalpel. Minced tissue was digested with type IV collagenase (final concentration of 3 mg/mL; Worthington) in RPMI (Sigma-Aldrich) supplemented with 10% fetal bovine serum (FBS; Gibco), at 37°C for 30 min with intermittent agitation. Digested tissue was then passed through a 100-µm cell strainer, and cells were pelleted by centrifugation at 500g for 5 min at 4°C. Cells were then resuspended in 5 ml of red blood cell lysis buffer (eBioscience) and left for 5 min at room temperature. It was then topped up with a flow buffer (PBS containing 2% (v/v) FBS and 2 mM EDTA) to 45 ml and pelleted by centrifugation at 500g for 5 min at 4°C. The resuspended cell solution was filtered through a 70-µm cell strainer (Corning), and live cells were manually counted by Trypan blue exclusion. Dissociated cells were first incubated with 5uL of FcR blocker for 5 min at room temperature and stained with anti-CD45 antibody (BV785 anti-human CD45 antibody, BioLegend, 304048) and DAPI

575 (Sigma-Aldrich, D9542) prior to sorting. DAPI was used at a final concentration of 2.8 μM, and all
576 antibody solutions were used at a final concentration of 5 μl per 100 μl cell suspensions containing fewer
577 than 5 million cells. DAPI- CD45+ and DAPI- CD45- populations were sorted by FACS using MA900
578 Multi-Application Cell Sorter (Sony) and its proprietary software (Cell Sorter Software v3.1.1). Sorted
579 cells were loaded on the Chromium Controller (10x Genomics) with a targeted cell recovery of
580 5,000–10,000 per reaction. Single-cell cDNA synthesis, amplification, gene expression library was
581 generated according to the manufacturer's instructions of the Chromium Next GEM Single Cell 5' Kit v2
582 (10x Genomics). Libraries were sequenced using NovaSeq 6000 (Illumina) at Wellcome Sanger Institute
583 with a minimum depth of 20,000–30,000 read pairs per cell.
584 Samples used for single nuclei isolation and spatial transcriptomics were flash-frozen (unembedded)
585 or frozen in OCT and stored at −80 °C, or formalin-fixed and subsequently embedded in paraffin blocks.
586 All tissues were stored and transported on ice at all times until freezing or tissue dissociation to minimise
587 any transcriptional degradation. Single nuclei were obtained from flash-frozen tissues using sectioning
588 and mechanical homogenization as previously described. 5-10 mm thickness frozen tissues were first
589 sectioned with cryostat in a 50 μm thickness section. All sections from each sample were homogenised
590 using a 7 ml glass Dounce tissue grinder set (Merck) with 8–10 strokes of a tight pestle (B) in
591 homogenization buffer (250 mM sucrose, 25 mM KCl, 5 mM MgCl2, 10 mM Tris-HCl, 1 mM dithiothreitol
592 (DTT), 1× protease inhibitor, 0.4 U μl−1 RNaseIn, 0.2 U μl−1 SUPERaseIn, 0.05% Triton X-100 in
593 nuclease-free water). Homogenate was filtered through a 40-μm cell strainer (Corning). After
594 centrifugation (500g, 5 min, 4 °C) the supernatant was removed and the pellet was resuspended in
595 storage buffer (1× PBS, 4% bovine serum albumin (BSA), 0.2U μl−1 Protector RNaseIn). Nuclei were
596 stained with 7-AAD Viability Staining Solution (BioLegend) and filtered through 20-μm cell strainer
597 (CellTrics). Positive single nuclei were purified by fluorescent activated cell sorting (FACS) using MA900
598 Multi-Application Cell Sorter (Sony) and its proprietary software (Cell Sorter Software v3.1.1). Nuclei
599 purification and integrity were verified under a microscope, and nuclei were manually counted by Trypan
600 blue exclusion. Nuclei suspension was adjusted to 1000–3,000 nuclei per microlitre and loaded on the
601 Chromium Controller (10x Genomics) with a targeted nuclei recovery of 5,000–10,000 per reaction. 3'
602 gene expression libraries and ATAC libraries were prepared according to the manufacturer's instructions
603 of the Chromium Multiome ATAC+Gene Expression Kits (10x Genomics). Libraries were sequenced
604 using NovaSeq 6000 (Illumina) at Wellcome Sanger Institute with a minimum depth of 20,000–30,000
605 read pairs per nucleus.
606

607 Spatial transcriptomics processing

608 Fresh-frozen (FF) samples: Fresh samples were frozen and embedded in optimal cutting temperature
609 medium (OCT) using a dry ice-cooled bath of isopentane at -45°C. OCT-embedded samples were
610 sectioned using a cryostat (Leica CX3050S) and were cut at 10 μm.
611 Formalin-fixed paraffin-embedded (FFPE) samples: Fresh samples were fixed in >5 times their volume
612 of 4% v/v formalin at ambient temperature for 24 hours before processing to paraffin on a Tissue-Tek

613 Vacuum Infiltration Processor 5 (Sakura Finetek). FFPE blocks were sectioned at 5 µm using a
614 microtome (Leica RM2125RT).
615 RNA integrity number (fresh-frozen samples) or DV200 (formalin-fixed) was obtained using Agilent
616 2100 Bioanalyzer. CytAssist FFPE Visium Spatial Gene Expression (10x genomics) was performed
617 following the manufacturer's protocol. For FF samples, the Tissue Optimization protocol from 10x
618 Genomics was performed to obtain a permeabilisation time of 45 min, and the FF Visium Spatial Gene
619 Expression experiment was performed as per the manufacturer's protocol (10x Genomics). H&E stained
620 Visium Gene Expression slides were imaged at 40× on Hamamatsu NanoZoomer S60. After transcript
621 capture, Visium Library Preparation Protocol from 10x Genomics was performed. Eight cDNA libraries
622 were diluted and pooled to a final concentration of 2.25 nM (200 µl volume) and sequenced on 2× SP
623 flow cells of Illumina NovaSeq 6000.

624

625 # Read mapping

626 After sequencing, samples were demultiplexed and stored as CRAM files. Each sample of single-cell
627 data was mapped to the human reference genome (GRCh38-2020-A) provided by 10x Genomics using
628 the CellRanger software (CellRanger v.6.0.2 or v.6.1.0, or CellRanger ARC v.2.0.0) with default
629 parameters. Each sample of Visium was mapped to the human reference genome (OCT: GRCh38-3.0.0,
630 FFPE: GRCh38-2020-A) the SpaceRanger software (OCT: v.1.1.0, FFPE: v.2.0.0) with default
631 parameters. For Visium samples, SpaceRanger was also used to align paired histology images with
632 mRNA capture spot positions in the Visium slides. Part of the single-cell samples were mixed with
633 different donors after the nuclei isolation for cost-efficient experimental design (Supp Table 2) and
634 computationally demultiplexed (Soupercell, v.2.0)[72] based on genetic variation between the donors.

635

636 # Preprocessing and quality control

637 For the single-cell transcriptome data, the CellBender algorithm (remove-background)(v.0.2)[73] was applied
638 to remove ambient and background RNA from each count matrix produced using the CellRanger pipeline.
639 Downstream analysis was performed using the Scanpy package (v.1.7.1). Doublets with a score of >0.15
640 were removed using Scrublet[74]. We performed quality control and filtering based on the following settings:
641 number of genes>500, total counts>1000, % mitochondrial<5 (nuclei), % mitochondrial<20 (cell), %
642 ribosomal<5 (nuclei), % ribosomal<20 (cell), red blood cell score<1 (nuclei), red blood cell score<2 (cell).
643 For the multiome ATAC data, the data processed using CellRanger ARC were further analysed using
644 ArchR (v.1.0.2)[75]. Quality control was performed, considering, among other factors, transcription start site
645 enrichment, nucleosomal banding patterns, the number and fraction of fragments in peaks, reads falling into
646 ENCODE blacklist regions as well as doublet scores computed by ArchR. For high-quality cells, reads were
647 mapped to 500-bp bins across the reference genome (hg38)(TileMatrix). Gene scores based on chromatin
648 accessibility around genes were computed from TileMatrix using the createArrowFiles function to check

their consistency with measured expression values. Before peak calling, pseudo-bulk replicates were generated (addGroupCoverages) for each fine-grained cell state annotated using the paired gene expression data. Peak calling (501 bp fixed-width peaks) was performed for each cell state, and the peak sets were merged to obtain a unified peak set (addReproduciblePeakSet). A cell-by-peak count matrix was obtained using the addPeakMatrix function.

For the visium data, the histological tissue annotation was performed based on the H&E image, and the visium spots on the irrelevant regions ('Out of tissue' and 'Cavity') were removed. Spots of each sample were further filtered for more than 500 UMI (OCT) or 5000 UMI (FFPE) counts and 300 (OCT) or 1000 (FFPE) genes.

Spatial mapping of the fine-grained cell types

We used cell2location (v.0.1) method[17] to spatially map the developing heart fine-grained cell types defined using single-cell transcriptomics data (**see Bayraktar *et al.***) in the Visium data. In brief, we first estimated reference signatures of cell types using sc/snRNA-seq data and a negative binomial regression model provided in the cell2ocation package. The inferred reference cell type signatures were used for cell2location cell-type mapping that estimates the abundance of each cell type in each Visium spot by decomposing spot mRNA counts. The mapping was performed per group of sections of each time point (4, 5, 7, 13, 16, and 20 PCW). The H&E images of the Visium slides were used to determine the average number of nuclei per Visium spot (n = 20) in the tissue and used as a hyperparameter in the cell2location pipeline.
For the immature versus mature atrial cardiomyocyte analysis, a temporary (i.e. used only for this analysis) annotation was made where the scFates milestones prior to the divergence of the principal graph were labelled as AtrialCardiomyocytesImmature (all other fine_grain annotations remained) (Supp. Fig. 7C). These signatures were mapped in the Visium data using cell2location. The per spot cell type abundances were extracted for spots with the structural annotation 'atria'. For each of the 4 labels (immature/mature for each of L/RaCM) a proportion was calculated: the abundance of that label divided by the combined total of the immature and mature abundances for the corresponding cell type. For instance $proportion_{LaCMimmature} = abundance_{LaCMimmature} / (abundance_{LaCMimmature} + abundance_{LaCMmature})$. This enables a comparison between the mature and immature versions of the same cell type.

Single-cell data integration

After pre-processing and removal of low-quality cells and nuclei, single-cell transcriptomics data were integrated using scVI(Gayoso et al. 2022; Lopez et al. 2018) accounting for categorical covariates of donor, cell_or_nuclei, and kit_10x, as well as continuous covariates of total_counts, %mito, and %ribo. Neighbourhood identification, high-resolution leiden clustering, and further dimensional reduction using UMAP were performed based on the scVI latent space. Cells and nuclei of extra-cardiac origin were removed based on marker gene expressions: lung epithelial cells (IGFBP2, EPCAM, SOX2, and NKX2-1); hepatocytes (ALB, APOA1, TTR); hepatic stellate cells (CRHBP, FCN3, OIT3); parathyroid cells (PTH,

686 MAFB, GATA3); and further erythrocytes (FTH1, HBA1, HBA2, HBB). The remaining data were integrated
687 using scVI with the same variables, followed by the downstream processing. To annotate cell types, cell
688 barcodes comprising major clusters were subsampled from the initial raw count matrix for subclustering
689 analysis and processed separately through the same downstream analysis as described. Subclusters were
690 annotated based on immunohistochemical validation and literature evidence.
691 For ATAC data, post-QC data were integrated using cell-by-peak count matrix peakVI(Ashuach et al.
692 2022) accounting for batch variation (donor). Neighbourhood identification and dimensional reduction using
693 UMAP were performed based on the peakVI latent space.
694

695 Epigenetic stability

696 Multiome data of each mid-grained cell type was separately integrated for the gene expression (GEX) or
697 chromatin accessibility (ATAC) data using scVI or peakVI, accounting for batch variation (donor and
698 sampled region). Overlapping Milo neighbourhoods were identified using the calculated neighbourhood
699 graph based on the scVI (n_neighbors=15) or peakVI (n_neighbors=100) latent spaces
700 (*milo.make_nhoods*). For each gene expression neighbourhood, the entropy of the distribution of
701 corresponding nuclei across ATAC neighbourhoods was calculated (*scipy.stats.entropy*). The effects of GEX
702 neighbourhood size (number of nuclei per neighbourhood) and variance of nucleus neighbours in the scVI
703 latent space on the entropy score were regressed out using a multiple linear regression model
704 (*sklearn.linear_model.LinearRegression*). The corrected entropy score indicates the extent of diffusion of a
705 group of transcriptionally similar nuclei in the ATAC space, which can be interpreted as the epigenetic
706 stability of the group of nuclei.
707

708 Visium data integration

709 Both visium OCT and FFPE data (4, 5, 7, 13, and 16 PCW) were integrated using spot-by-cell type
710 estimated abundance (cell2location) matrix and BBKNN[76] accounting for batch variation (section).
711 Batch-balanced neighbour graph was used for leiden clustering and UMAP visualisation. To annotate tissue
712 niches, major tissues were first annotated based on the clustering result and the estimated cell type
713 abundances. Each major cluster was subsampled and subjected to the downstream workflow as described
714 to annotate finer-grained tissues. Some annotations were fully (AVN, lymph node, and epicardium) or partly
715 (AV ring, AV valves, aortic valve, valve apparatus, coronary artery, and great vessels) defined based on the
716 histological tissue structures manually annotated using the paired H&E images.
717

718 TissueTypist model training

719 A CellTypist model was trained using the spot-by-gene expression matrix of euploid OCT data and the
720 tissue structure labels (TissueTypist model of the developing heart). To balance the cell number of each
721 structure label, for labels which comprised > 100 spots, index spots of transcriptomically similar

neighbourhoods were selected using Milo and used for the model training. We also generated a model for the ventricle region (ventricle model) using all spots from relevant tissue regions (Ventricle - Compact, Ventricle - Trabeculated, Endocardium - Ventricular, Epicardium, VCS - Distal, VCS - Proximal, Coronary vessel). The performance of the model was assessed by cross-validation of the spots used for training. For mapping to Trisomy 21 data, the healthy 13 PCW section was removed from the euploid dataset, and the remaining data was used for training similarly to the ventricle model. The trained model was used for the prediction of tissue structures in both the euploid and trisomy 21 sections.

Trajectory analysis

We first obtained a latent space based on both the RNA-seq and ATAC-seq data using \texttt{MultiVI}. The latent space was used for obtaining two-dimensional embeddings and a series of trajectory analyses using scFates[77].

For monocyte and macrophages, both the cells and nuclei samples with region=='whole heart' were integrated with accounting for categorical covariates of donor, cell_or_nuclei, and kit_10x, as well as continuous covariates of total_counts, %mito, %ribo, and cell cycle scores (S and G2/M phase). Multiscale diffusion space was obtained using the MultiVI latent space and the Palantir algorithm [78]. The neighbourhood graph was obtained using the Palantir diffusion space and was used for force-directed graph embedding. The Palantir diffusion space was also used for obtaining a principal graph with SimplePPT algorithm. Pseudotime calculation and the assignment of milestones and segments were performed with setting two milestones (29: YS-derived macrophages, 2: monocytes) as roots (*scf.tl.roots*, *scf.tl.pseudotime*). To explore genes involved in the LYVE1+ macrophage maturation, the segments belonging to the trajectory (i.e. which formed a contiguous path) were subsampled, and the genes whose expressions were associated and upregulated along the pseudotime were selected (scf.tl.test_association, sc.tl.rank_genes_groups). The pathways enrichment analysis was performed using the selected genes and GSEApy[79].

For left ventricular cardiomyocytes, multiome data of compact and trabeculated cardiomyocytes were integrated with accounting for categorical covariates of donor and sampled region, as well as continuous covariates of total_counts, %mito, %ribo, and cell cycle scores (S and G2/M phase). The downstream steps were performed similarly to the analysis above. Pseudotime calculation and the assignment of milestones and segments were performed by setting the youngest milestones (17) as the root. The genes whose expressions were associated and upregulated along the pseudotime were selected either by subsampling compact or trabeculated populations or by using all nuclei.

For atrial cardiomyocytes, multiome data of left and right cardiomyocytes were integrated with multiVI accounting for the same covariates as before. To avoid an over-representation of either cell type, data where the whole heart had been processed was used. The downstream steps were performed similarly to the analysis above. Pseudotime calculation and the assignment of milestones and segments were performed by setting milestone 49 as the root. Two trajectories, consisting of milestones leading from the root to terminal milestones in the left and right atrial cardiomyocyte clusters, respectively, were defined.

Association of gene expression with pseudotime was tested along each of these two trajectories independently, as well as across the whole embedding.

Fitted gene expression

Using the scikit-learn Python library, we employed a polynomial regression analysis pipeline to fit the expression levels of genes as a function of pseudotime. Specifically, the make_pipeline function was used to streamline the construction of a modelling pipeline, which included PolynomialFeatures for generating polynomial and interaction features, followed by LinearRegression to perform the linear regression fitting. This process allowed us to assess and select the optimal polynomial degree for each gene based on the highest cross-validated score, as determined by the cross_val_score function with a 5-fold cross-validation strategy.

For each gene, we initially normalised the pseudotime to a range between 0 and 1 to ensure uniformity in model fitting and prediction. We then explored polynomial degrees ranging from 1 to 5 to identify the model that best captured the relationship between gene expression and pseudotime for each gene. The selected model was subsequently used to predict gene expression levels across a uniformly spaced pseudotime grid for visualization purposes.

Cardiomyocyte gene signatures

The gene expression profile of SAN pacemaker cells ("SinoatrialNodeCardiomyocytes") were compared to that of other cardiomyocytes and 675 DEGs were obtained (scanpy.tl.rank_genes_groups, logFC>1, mean expression>0.1). For the foetal-specific gene signatures, the foetal SAN pacemaker cell population was compared to the adult counterpart (SAN_P_cell in [32]). For each cell type, donors were treated as replicates by making pseudo-bulk data (mean). Groups which have less than 10 cells were removed from the analysis. Wilcoxon rank-sum test was performed to identify DEGs (620 genes)(logFC>1, mean expression>0.1). Intersection of the two DEGs were used as foetal-specific pacemaker cell signatures (143 genes) in the downstream analysis.

The gene signatures of the left ventricular cardiomyocyte populations were selected as follows. For compact or trabeculated cardiomyocyte-specific signatures, genes associated with and upregulated along the pseudofime (calculated per population) and expressed higher in one population compared to the other (either compact-vs-trabeculated or trabeculated-vs-compact) were selected (compact: 130 genes, trabeculated: 107 genes). For the common signature, genes associated with and upregulated along the pseudotime of both populations were selected (336 genes).

Pathway enrichment analysis was performed for each gene signature using GSEApy[79].

Gene regulatory network analysis

The SCENIC+[27] pipeline was used to predict TFs and putative target genes as well as regulatory genomic regions harbouring TF binding sites. For each analysis, gene expression and chromatin accessible region (peak) counts of multiome data was used as input.

For monocyte and macrophage populations, multiome data were subsetted from the data used for trajectory analysis, and metacells were not created because of the small nucleus number. For the analysis of pacemaker cells, all the cardiomyocytes were used as input and metacells were created by clustering cells into groups of approximately 10 cells based on their RNA profiles and subsequent aggregation of counts and fragments. For the maturation analysis of atrial and left ventricle cardiomyocytes, the same multiome data used for the trajectory analysis were used as input. Metacells were created using SEACells[80] (10 nuclei per metacell) based on the MultiVI latent space and subsequent aggregation of counts and fragments.

CisTopic was applied to identify region topics and differentially accessible regions from the fragment counts as candidate regions for TF binding. CisTarget was then run to scan the regions for transcription factor binding sites and GRNBoost2 was used to link TFs and regions to target genes based on co-expression/accessibility. Enriched TF motifs in the regions linked to target genes were used to construct TF-region and TF-gene regulons. Finally, regulon activity scores (AUC) were computed with AUCell based on target-gene expression and target region accessibility. Networks of TFs, regions and target genes (eGRNs) were constructed by linking individual regulons. TF-region-gene links for all analyses can be found in Supplementary Table 3-5.

Constructed network was further pruned for downstream analysis and visualisation. For the SAN pacemaker cells, GRNs which target SAN pacemaker cell gene signatures, ion channel genes, or axon guidance genes were first selected. For each TF-region-gene interaction, either TF (logFC>1, mean expression>0.1, proportion of expression>0.1) or region (logFC>0, proportion of accessibility>0.01) being SAN pacemaker cell-specific (compared to other cardiomyocytes) were further selected. For the left ventricular cardiomyocytes, GRNs which target “compact” and “common” gene signatures were selected similarly to the GRNs of SAN pacemaker cells (compact cardiomyocyte GRNs).

In-silico TF perturbation analysis was performed using CellOracle for the left ventricular cardiomyocyte analysis. CellOracle object was made using the genes included in the left ventricle cardiomyocyte SCENIC+ regulons. Input baseGRN was also prepared using the same regulon (only activator regulons). The cluster labels were generated by combining fine-grained cell types and binned chronological age (n_bin = 3). TFs which govern the pruned GRNs for left compact cardiomyocytes were tested for in-silico perturbation.

Mutational lineage tracing

The SComatic [34] pipeline was used to infer somatic mutations using RNA and ATAC seq reads. A merged BAM file containing all the mapped data for each donor-modality (RNA or ATAC) combination was generated, this was the input to the pipeline. A mappings file (linking barcode to cell type annotation) was also created. SComatic first splits the input BAM file into one BAM file for each cell type (using the

mappings file). It then generates a base count matrix, which records the counts for each base at each genomic position in each cell type. Candidate somatic mutations are distinguished from germline polymorphisms and sequencing artefacts using a set of hard filters and a beta-binomial test parameterized using non-neoplastic samples for each of RNA and ATAC data. Candidate mutations overlapping known RNA-editing sites or single-nucleotide polymorphisms (SNPs) with population frequencies greater than 1% in the gnomAD database are also filtered out (as likely germline). In addition, SComatic uses a panel of normals (PON) generated using non-neoplastic samples to exclude recurrent sequencing and mapping artefacts. To make a mutation call, SComatic requires a sequencing depth of at least five reads in the cell type in which the mutation is detected, and that the mutation is detected in at least three reads from at least two different individual cells of the same cell type. By default it will then remove mutations which are present in >1 cell types from the final output (since this increases the probability of the mutation being non-germline). To construct a phylogenetic relationship, features (e.g. mutations) must be present in >1 taxon to enable a distance to be calculated. Therefore, we parameterised the pipeline to report all mutations.

We parsed the SComatic output by first extracting a cell type by mutation matrix where values were the absolute number of cells with sufficient depth at that genomic position for the mutation to be callable if it were present ('NC' output). We then extracted a matrix of the same dimensions where values were the fraction of cells of that cell type which were called as having the mutation ('CCF' output).

We then binarised the matrix, counting a mutation as being present in a cell type if it was called in at least 10% of the cells of that cell type (CCF > 0.1). The values in this matrix were then 1 (if the mutation site was callable and the mutation was present), 0 (if the mutation site was callable and the mutation was absent), or NA (if the mutation site was not callable, due to insufficient depth). Less populous cell types are more likely to have insufficient coverage at a given genomic site. To mitigate this we retained mutations in the analysis if they were callable (i.e. 1 or 0) in >80% of cell types. We also removed mutations which were callable and present (i.e. 1) in all cell types, since these are likely to be germline. A Hamming distance matrix was then generated, which computed distance at non-NA mutations for all pairwise combinations of cell types. The R package ape was used to construct neighbor-joining trees and bootstrap them. To create a consensus tree we concatenated the mutation matrices of multiple donors, then constructed the distance matrix and tree as before.

Transmural axis analysis using OrganAxis

The spots which belong to left ventricle regions were manually selected (Supp. Fig. 8A). The inner (endocardial side) and outer (epicardial side) most layer were manually annotated. Using the OrganAxis algorithm[9], we calculated the relative distances of each spot from the two manually annotated layers. we calculated the L2 distance of all spots to the closest corresponding points in each annotation by number of KNN points (dist2cluster, kNN parameter = 4). To construct an axis anchored to the interface between two structures ("transmural axis"), we fitted a normalised signoidal curve to the relative distances of each spot and placed splots in relative positions in respect of the two structures (axis_2p_norm).

The gene expressions and cell type abundances were aggregated per bin along the transmural axis and clustered along the transmural axis and visualised (seaborn.clustermap). For gene selection, transmural axis was binned into three part (outer, middle, and inner) and differentially expressed genes for each part (compared to rest) were selected (logFC>0.5, mean expression>0.01). Pathway enrichment analysis was performed for the genes upregulated in outer or inner part.

Nuclei segmentation of the H&E images of Visium sections was performed using Cellpose[81] implemented in Squidpy[82] (squidpy.im.segment, method=cellpose_he, flow_threshold=0.8, channel_cellpose=1).

CellTypist model training

To identify Yolk sac (YS) derived macrophage population, macrophage populations (MACROPHAGE_IRON_RECYCLING, MACROPHAGE_KUPFFER_LIKE, MACROPHAGE_LYVE1_HIGH, MACROPHAGE_MICROGLIA, MACROPHAGE_PRE_MAC) were selected from the YS single-cell RNA-seq dataset[23]. MACROPHAGE_IRON_RECYCLING, MACROPHAGE_KUPFFER_LIKE, and MACROPHAGE_LYVE1_HIGH were merged as “MACROPHAGE” based on the transcriptional profiles shown in Goh et al, *Science*, 2023. Highly variable genes were selected (sc.pp.highly_variable_genes) and cell number of each cell type was balanced by subsampling cells using geosketch, a geometric sketching algorithm[83]. The selected genes and cells were used for training a logistic regression model (celltypist.train).

Cell-cell interaction inference

We used CellPhoneDB to infer cell-cell interactions between annotated clusters. Expression matrices and the metadata relating cell-barcode to cluster were submitted to CellPhoneDB with default parameters. Interactions with a p-value > 0.05 were removed, a selection of the interactions were reported with p-value and mean expressions. Interactions were selected based on ranking after removing overrepresented protein complex interactions, in order to display a diverse set of interactions.

smFISH

FFPE foetal heart tissue sections (5 µm) were placed onto SuperFrost Plus slides. Staining was performed using the RNAscope Multiplex Fluorescent Reagent kit v2 assay (Advanced Cell Diagnostics, Bio-Techne), automated with a Leica BOND RX, according to the manufacturers’ instructions. Automated pretreatment included baking at 60°C for 30 minutes, dewaxing, heat-induced epitope retrieval at 95°C for 15 min in buffer ER2, and digestion with Protease III for 15 min. Following probe hybridisation, tyramide signal amplification with Opal 520 and Opal 650 (Akoya Biosciences) and TSA-biotin (TSA Plus Biotin Kit, Akoya) with streptavidin-conjugated Atto 425 (Sigma Aldrich) was used to develop RNAscope channels. All nuclei

were DAPI stained. Stained sections were imaged using a Perkin Elmer Opera Phenix High-Content Screening System with a ×20 water-immersion objective (NA of 0.16, 0.299 µm per pixel). The following channels were used: DAPI (excitation, 375 nm; emission, 435–480 nm); Opal 520 (excitation, 488 nm; emission, 500–550 nm); Opal 650 (excitation, 640 nm; emission, 650–760 nm); and Atto 425 (excitation, 425 nm; emission, 463–501 nm).

Trisomy 21 data analysis

To integrate the Trisomy 21 (T21) data with the euploid atlas a transfer learning approach was taken using scArches [49]. The GEX portion of the T21 data underwent the same preprocessing and QC as had been performed for the atlas data (described above). The atlas was subset to match the age range of the T21 data (PCW13-14), yielding ~74k cells across all cell types. Using this as the reference data a 'mid-grain' cell-type-label-aware SCANVI model was created correcting the same categorical ('donor', 'cell_or_nuclei', 'kit_10x') and continuous ('total_counts', 'pct_counts_mt', 'pct_counts_ribo') covariates as when using scVI. This base model was then updated with the query (T21) data (*scarches.models.SCANVI.load_query_data*()). This updated model was then used to predict cell type labels in the T21 data, and a harmonised latent space incorporating both atlas and T21 data was extracted. To transfer 'fine-grain' labels, a series of CellTypist models (one per mid-grain compartment) were trained in the euploid data and applied on the Trisomy 21 data.

The scArches latent space was then used to test for differential abundance associated with T21 status using Milo [14]. Each library was considered a unique sample and to minimise differences explained by age as a confounder, a GLM design of `~ age + T21_status` was used. A 'spatial FDR' of <0.05 was deemed significant.

To assess the impact of T21 on cellular profiles, we used a distance-based approach. In each of the peakVI (ATAC) and scArches (GEX) latent spaces we computed the euclidean distance between the centroid of the euploid and T21 cells from each cell type. To adjust for variations in cell type representation within the latent spaces, we normalized these distances. The normalization factor for each cell type was determined by calculating the median Euclidean distance of all cells of that type to their respective centroid. Subsequently, the distance between each pair of cell types was divided by the average of the median distances for those cell types. To facilitate a comparative analysis between the ATAC and GEX data, min-max normalisation was then applied to the distances in each dataset. This scaling, which adjusted values to a range between 0 and 1, was used to generate the bidirectional bar plot.

To analyse the activity of GRNs, compact cardiomyocyte GRNs constructed using euploid data was utilised (see "Gene regulatory network analysis" section). The activity of the GRNs in euploid and T21 cardiomyocytes was computed with AUCell based on target-gene expression. Differential activity score analysis was performed to identify GRNs which were altered in T21 cells (scanpy.tl.rank_genes_groups, adjusted *p*-value<0.05).

CHD gene expression

A list of CHD causal genes was downloaded from https://chdgene.victorchang.edu.au/ [4]. A simplified categorisation of five congenital heart lesions was constructed: Septal Defects, Outflow Tract defects, Heterotaxy, Obstructive Lesions and Functional Single Ventricle. Since many CHD genes are associated with more than one CHD lesion, each gene was given a True or False association with each of the five categories. To compare expression of the CHD genes between the euploid and trisomy 21 data, we identified which of the 63 fine_grain cell states were represented by at least 100 cells in both conditions (40 cell states). For each of those cell states differential expression testing was performed (using sc.tl.rank_gene_groups()). For each of the CHD genes and cell types, the log fold change values was extracted and plotted.

Data availability

Open-access datasets will be available from ArrayExpress (www.ebi.ac.uk/arrayexpress) with accession numbers (Multiome snATAC-seq and Visium). All of the processed data are available for browsing gene expression and download from the Heart Cell Atlas (https://www.heartcellatlas.org/foetal.html, currently password protected and will be publicly available at the time of publication). TissueTypist models trained on this euploid atlas are available for download from the Heart Cell Atlas web portal for automated tissue annotation of other cardiac datasets. scArches models trained on the euploid atlas is available at the Heart Cell Atlas (https://www.heartcellatlas.org) for cell mapping of a query dataset. Inferred gene regulatory networks are available in Supplementary Table 3-5. The human reference genome (GRCh38) used for read mapping is available from 10x Genomics (https://support.10xgenomics.com/single-cell-gene-expression/software/release-notes/build).

Code availability

The code used for all analysis performed in this study will be available on GitHub at https://github.com/Teichlab/Developing_heart_atlas.

Acknowledgements

We thank the donors for granting access to the tissue samples. We also thank staff at the Wellcome Sanger Cytometry Core Facility, Cellular Genetics Informatics team, Spatial Genomics Platform team, and Core DNA Pipelines team for their support; F. Muyas and R. Li for their help on SComatic analysis; A. Oszlanczi for her help on sample management; B. Cakir and J. Cox for their help on the Heart Cell Atlas web portal; and A. Wilk for administrative assistance; R. Barker for tissue procurement.

Illustrations in Figs. 1A, 2L, 3H, and 4A were created using BioRender (https://biorender.com). This work was made possible by a partnership between the Wellcome Sanger Institute and Wellcome-MRC Cambridge Stem Cell Institute at University of Cambridge.

This project was made possible in part by the Wellcome Trust (WT206194 to S.A.T); the Chan Zuckerberg Foundation (2021-237882 to S.A.T.); the Wellcome Trust Clinical PhD Fellowship to J.C.; the Overseas Research Fellowship of the Takeda Science Foundation to K.K.; the British Heart Foundation (BHF) Senior Fellowship (FS/18/46/33663) (S.S. and L.G.); the Oxbridge BHF Centre for Regenerative Medicine (RM/17/2/33380) (V.K.S.); and BHF grants PG/17/24/32886 (L.G.) and RG/17/5/32936 (H.D.); This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie-Skłodowska-Curie grant agreement No. 101026233 to J.P.P.; We also acknowledge core support from the Wellcome Trust, the Medical Research Council and the Wellcome Trust–Medical Research Council Cambridge Stem Cell Institute. This research was funded, in whole or in part, by the Wellcome Trust (grant no. 203151/Z/16/Z).

Author contributions

Project design, Data generation, Analysis, Manuscript: J.C., K.K.
Project design, Analysis, Manuscript: S.B., V.K.S.
Analysis (GRN): J.P.P., J.P.
Analysis (SComatic, Visium mapping): K.P.
Data generation (Visium): M.D., I.M., M.P.
Data generation (Multiome): L.R., C.I.S., R.K., S.P.
Data generation (tissue procurement): X.H.
Analysis (Tissue structure annotation): S.Y.H.
Analysis (OrganAxis): N.Y.
Data generation (smFISH): L.T., K.R.
Heart collection, processing and sequencing: H.D., L.G.
Guide on the analysis, reviewing/editing manuscript: R.T.
Supervision: S.S., S.A.T.

Competing interests

S.A.T. is a scientific advisory board member of ForeSite Labs, Qiagen and Element Biosciences, and a co-founder and equity holder of TransitionBio and EnsoCell Therapeutics, and a part-time employee of GlaxoSmithKline. S.S. is a co-founder and equity holder of ABS Biotechnologies. The remaining authors declare no competing interests.

References

1. Zimmerman, M. S. *et al.* Global, regional, and national burden of congenital heart disease, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *The Lancet Child & Adolescent Health* **4**, 185–200 (2020).
2. Dimopoulos, K. *et al.* Cardiovascular Complications of Down Syndrome: Scoping Review and Expert Consensus. *Circulation* **147**, 425–441 (2023).
3. Sifrim, A. *et al.* Distinct genetic architectures for syndromic and nonsyndromic congenital heart defects identified by exome sequencing. *Nat. Genet.* **48**, 1060–1065 (2016).
4. Yang, A. *et al.* CHDgene: A Curated Database for Congenital Heart Disease Genes. *Circ Genom Precis Med* **15**, e003539 (2022).
5. Karbassi, E. *et al.* Cardiomyocyte maturation: advances in knowledge and implications for regenerative medicine. *Nat. Rev. Cardiol.* **17**, 341–359 (2020).
6. Ahmed, R. E., Anzai, T., Chanthra, N. & Uosaki, H. A Brief Review of Current Maturation Methods for Human Induced Pluripotent Stem Cells-Derived Cardiomyocytes. *Front Cell Dev Biol* **8**, 178 (2020).
7. Asp, M. *et al.* A Spatiotemporal Organ-Wide Gene Expression and Cell Atlas of the Developing Human Heart. *Cell* **179**, 1647–1660.e19 (2019).
8. Ameen, M. *et al.* Integrative single-cell analysis of cardiogenesis identifies developmental trajectories and non-coding mutations in congenital heart disease. *Cell* **185**, 4937–4953.e23 (2022).
9. Yayon, N. *et al.* A spatial human thymus cell atlas mapped to a continuous tissue axis. *bioRxiv* (2023) doi:10.1101/2023.10.25.562925.
10. ENCODE Project Consortium *et al.* Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature* **583**, 699–710 (2020).
11. Corces, M. R. *et al.* Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. *Nat. Genet.* **48**, 1193–1203 (2016).
12. Anene-Nzelu, C. G., Lee, M. C. J., Tan, W. L. W., Dashi, A. & Foo, R. S. Y. Genomic enhancers in cardiac development and disease. *Nat. Rev. Cardiol.* **19**, 7–25 (2022).
13. George, R. M. & Firulli, A. B. Epigenetics and Heart Development. *Front Cell Dev Biol* **9**, 637996 (2021).
14. Dann, E., Henderson, N. C., Teichmann, S. A., Morgan, M. D. & Marioni, J. C. Differential abundance testing on single-cell data using k-nearest neighbor graphs. *Nat. Biotechnol.* **40**, 245–253 (2022).

1038 15. Gilsbach, R. *et al.* Distinct epigenetic programs regulate cardiac myocyte development and disease in the
1039 human heart in vivo. *Nat. Commun.* **9**, 391 (2018).
1040 16. Gomez-Velazquez, M. *et al.* CTCF counter-regulates cardiomyocyte development and maturation
1041 programs in the embryonic heart. *PLoS Genet.* **13**, e1006985 (2017).
1042 17. Kleshchevnikov, V. *et al.* Cell2location maps fine-grained cell types in spatial transcriptomics. *Nat.*
1043 *Biotechnol.* (2022) doi:10.1038/s41587-021-01139-4.
1044 18. Hasan, W. Autonomic cardiac innervation: development and adult plasticity. *Organogenesis* **9**, 176–193
1045 (2013).
1046 19. Chow, L. T., Chow, S. S., Anderson, R. H. & Gosling, J. A. Autonomic innervation of the human cardiac
1047 conduction system: changes from infancy to senility--an immunohistochemical and histochemical analysis.
1048 *Anat. Rec.* **264**, 169–182 (2001).
1049 20. Wu, Y. & Hirschi, K. K. Tissue-Resident Macrophage Development and Function. *Front Cell Dev Biol* **8**,
1050 617879 (2020).
1051 21. Zaman, R. & Epelman, S. Resident cardiac macrophages: Heterogeneity and function in health and
1052 disease. *Immunity* **55**, 1549–1563 (2022).
1053 22. Ashuach, T. *et al.* MultiVI: deep generative model for the integration of multimodal data. *Nat. Methods* **20**,
1054 1222–1231 (2023).
1055 23. Goh, I. *et al.* Yolk sac cell atlas reveals multiorgan functions during human early development. *Science*
1056 **381**, eadd7564 (2023).
1057 24. Liberzon, A. *et al.* The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst* **1**,
1058 417–425 (2015).
1059 25. Cahill, T. J. *et al.* Tissue-resident macrophages regulate lymphatic vessel growth and patterning in the
1060 developing heart. *Development* **148**, (2021).
1061 26. Fantin, A. *et al.* Tissue macrophages act as cellular chaperones for vascular anastomosis downstream of
1062 VEGF-mediated endothelial tip cell induction. *Blood* **116**, 829–840 (2010).
1063 27. Bravo González-Blas, C. *et al.* SCENIC+: single-cell multiomic inference of enhancers and gene
1064 regulatory networks. *Nat. Methods* 1–13 (2023).
1065 28. van Eif, V. W. W., Devalla, H. D., Boink, G. J. J. & Christoffels, V. M. Transcriptional regulation of the
1066 cardiac conduction system. *Nat. Rev. Cardiol.* **15**, 617–630 (2018).
1067 29. Lauro, C. *et al.* The chemokine CX3CL1 reduces migration and increases adhesion of neurons with

1068 mechanisms dependent on the beta1 integrin subunit. *J. Immunol.* **177**, 7599–7606 (2006).
1069 30. De Schepper, S., Crowley, G. & Hong, S. Understanding microglial diversity and implications for neuronal
1070 function in health and disease. *Dev. Neurobiol.* **81**, 507–523 (2021).
1071 31. Liang, D. *et al.* Sinoatrial node pacemaker cells share dominant biological properties with glutamatergic
1072 neurons. *Protein Cell* **12**, 545–556 (2021).
1073 32. Kanemaru, K. *et al.* Spatially resolved multiomics of human cardiac niches. *Nature* **619**, 801–810 (2023).
1074 33. van Weerd, J. H. & Christoffels, V. M. The formation and function of the cardiac conduction system.
1075 *Development* **143**, 197–210 (2016).
1076 34. Muyas, F. *et al.* De novo detection of somatic mutations in high-throughput single-cell profiling data sets.
1077 *Nat. Biotechnol.* (2023) doi:10.1038/s41587-023-01863-z.
1078 35. Li, R. *et al.* Mapping single-cell transcriptomes in the intra-tumoral and associated territories of kidney
1079 cancer. *Cancer Cell* **40**, 1583–1599.e10 (2022).
1080 36. van Ouwerkerk, A. F. *et al.* Epigenetic and Transcriptional Networks Underlying Atrial Fibrillation. *Circ.*
1081 *Res.* **127**, 34–50 (2020).
1082 37. Steimle, J. D. *et al.* Decoding the PITX2-controlled genetic network in atrial fibrillation. *JCI Insight* **7**,
1083 (2022).
1084 38. Zandstra, T. E. *et al.* Asymmetry and Heterogeneity: Part and Parcel in Cardiac Autonomic Innervation
1085 and Function. *Front. Physiol.* **12**, 665298 (2021).
1086 39. Verheije, R. *et al.* Heterozygous loss-of-function variants of MEIS2 cause a triad of palatal defects,
1087 congenital heart defects, and intellectual disability. *Eur. J. Hum. Genet.* **27**, 278–290 (2019).
1088 40. Sumanaweera, D. *et al.* Gene-level alignment of single cell trajectories informs the progression of in vitro
1089 T cell differentiation. *bioRxiv* 2023.03.08.531713 (2023) doi:10.1101/2023.03.08.531713.
1090 41. Farah, E. N. *et al.* Spatially organized cellular communities form the developing human heart. *Nature* 1–11
1091 (2024).
1092 42. Kim, T. T. & Dyck, J. R. B. The role of CD36 in the regulation of myocardial lipid metabolism. *Biochim.*
1093 *Biophys. Acta* **1861**, 1450–1460 (2016).
1094 43. Wu, T. *et al.* PRDM16 Is a Compact Myocardium-Enriched Transcription Factor Required to Maintain
1095 Compact Myocardial Cardiomyocyte Identity in Left Ventricle. *Circulation* **145**, 586–602 (2022).
1096 44. Guo, Y. & Pu, W. T. Cardiomyocyte Maturation: New Phase in Development. *Circ. Res.* **126**, 1086–1106
1097 (2020).

1098 45. Madison, B. B. Srebp2: A master regulator of sterol and fatty acid synthesis. *Journal of lipid research* vol.
1099 57 333–335 (2016).
1100 46. Kamimoto, K. *et al.* Dissecting cell identity via network inference and in silico gene perturbation. *Nature*
1101 1–10 (2023).
1102 47. Nees, S. N. & Chung, W. K. Genetic Basis of Human Congenital Heart Disease. *Cold Spring Harb.*
1103 *Perspect. Biol.* **12**, (2020).
1104 48. Lana-Elola, E. *et al.* Increased dosage of DYRK1A leads to congenital heart defects in a mouse model of
1105 Down syndrome. *Sci. Transl. Med.* **16**, (2024).
1106 49. Lotfollahi, M. *et al.* Mapping single-cell data to reference atlases by transfer learning. *Nat. Biotechnol.* **40**,
1107 121–130 (2022).
1108 50. Domínguez Conde, C. *et al.* Cross-tissue immune cell analysis reveals tissue-specific features in humans.
1109 *Science* **376**, eabl5197 (2022).
1110 51. Frid, C., Drott, P., Lundell, B., Rasmussen, F. & Annerén, G. Mortality in Down's syndrome in relation to
1111 congenital malformations. *J. Intellect. Disabil. Res.* **43 (Pt 3)**, 234–241 (1999).
1112 52. Benhaourech, S., Drighil, A. & Hammiri, A. E. Congenital heart disease and Down syndrome: various
1113 aspects of a confirmed association. *Cardiovasc. J. Afr.* **27**, 287–290 (2016).
1114 53. Satoda, M. *et al.* Mutations in TFAP2B cause Char syndrome, a familial form of patent ductus arteriosus.
1115 *Nat. Genet.* **25**, 42–46 (2000).
1116 54. Wang, J. *et al.* Tfap2b mutation in mice results in patent ductus arteriosus and renal malformation. *J.*
1117 *Surg. Res.* **227**, 178–185 (2018).
1118 55. Wang, Z. *et al.* An immune cell atlas reveals the dynamics of human macrophage specification during
1119 prenatal development. *Cell* **0**, (2023).
1120 56. Männer, J. When Does the Human Embryonic Heart Start Beating? A Review of Contemporary and
1121 Historical Sources of Knowledge about the Onset of Blood Circulation in Man. *J Cardiovasc Dev Dis* **9**,
1122 (2022).
1123 57. Wiese, C. *et al.* Formation of the sinus node head and differentiation of sinus node myocardium are
1124 independently regulated by Tbx18 and Tbx3. *Circ. Res.* **104**, 388–397 (2009).
1125 58. Espinoza-Lewis, R. A. *et al.* Shox2 is essential for the differentiation of cardiac pacemaker cells by
1126 repressing Nkx2-5. *Dev. Biol.* **327**, 376–385 (2009).
1127 59. Martin, K. E., Ravisankar, P., Beerens, M., MacRae, C. A. & Waxman, J. S. Nr2f1a maintains atrial

1128 expression to repress pacemaker identity within venous atrial cardiomyocytes of zebrafish. *Elife* **12**,
1129 (2023).
1130 60. Bressan, M., Liu, G. & Mikawa, T. Early mesodermal cues assign avian cardiac pacemaker fate potential
1131 in a tertiary heart field. *Science* **340**, 744–748 (2013).
1132 61. Hoogaars, W. M. H. *et al.* Tbx3 controls the sinoatrial node gene program and imposes pacemaker
1133 function on the atria. *Genes Dev.* **21**, 1098–1112 (2007).
1134 62. Benson, D. W. *et al.* Congenital sick sinus syndrome caused by recessive mutations in the cardiac sodium
1135 channel gene (SCN5A). *J. Clin. Invest.* **112**, 1019–1028 (2003).
1136 63. Sizarov, A. *et al.* Formation of the building plan of the human heart: morphogenesis, growth, and
1137 differentiation. *Circulation* **123**, 1125–1135 (2011).
1138 64. Arita, Y. *et al.* Myocardium-derived angiopoietin-1 is essential for coronary vein formation in the
1139 developing heart. *Nat. Commun.* **5**, 4552 (2014).
1140 65. Gladka, M. M. *et al.* Cardiomyocytes stimulate angiogenesis after ischemic injury in a ZEB2-dependent
1141 manner. *Nat. Commun.* **12**, 84 (2021).
1142 66. Funakoshi, S. *et al.* Generation of mature compact ventricular cardiomyocytes from human pluripotent
1143 stem cells. *Nat. Commun.* **12**, 3155 (2021).
1144 67. Monserrat, L. *et al.* Mutation in the alpha-cardiac actin gene associated with apical hypertrophic
1145 cardiomyopathy, left ventricular non-compaction, and septal defects. *Eur. Heart J.* **28**, 1953–1961 (2007).
1146 68. Wessels, M. W. *et al.* Compound heterozygous or homozygous truncating MYBPC3 mutations cause
1147 lethal cardiomyopathy with features of noncompaction and septal defects. *Eur. J. Hum. Genet.* **23**,
1148 922–928 (2015).
1149 69. Hirono, K., Hata, Y., Ibuki, K. & Yoshimura, N. Familial Ebstein's anomaly, left ventricular noncompaction,
1150 and ventricular septal defect associated with an MYH7 mutation. *J. Thorac. Cardiovasc. Surg.* **148**,
1151 e223–6 (2014).
1152 70. Yang, J. *et al.* Inhibition of Notch2 by Numb/Numblike controls myocardial compaction in the heart.
1153 *Cardiovasc. Res.* **96**, 276–285 (2012).
1154 71. Knight-Schrijver, V. R. *et al.* A single-cell comparison of adult and fetal human epicardium defines the
1155 age-associated changes in epicardial activity. *Nat Cardiovasc Res* **1**, 1215–1229 (2022).
1156 72. Heaton, H. *et al.* Souporcell: robust clustering of single-cell RNA-seq data by genotype without reference
1157 genotypes. *Nat. Methods* **17**, 615–620 (2020).

1158 73. Fleming, S. J. *et al.* Unsupervised removal of systematic background noise from droplet-based single-cell
1159 experiments using CellBender. *Nat. Methods* **20**, 1323–1335 (2023).
1160 74. Wolock, S. L., Lopez, R. & Klein, A. M. Scrublet: Computational Identification of Cell Doublets in
1161 Single-Cell Transcriptomic Data. *Cell Syst* **8**, 281–291.e9 (2019).
1162 75. Granja, J. M. *et al.* ArchR is a scalable software package for integrative single-cell chromatin accessibility
1163 analysis. *Nat. Genet.* **53**, 403–411 (2021).
1164 76. Polański, K. *et al.* BBKNN: fast batch alignment of single cell transcriptomes. *Bioinformatics* **36**, 964–965
1165 (2020).
1166 77. Faure, L., Soldatov, R., Kharchenko, P. V. & Adameyko, I. scFates: a scalable python package for
1167 advanced pseudotime and bifurcation analysis from single-cell data. *Bioinformatics* **39**, (2023).
1168 78. Setty, M. *et al.* Characterization of cell fate probabilities in single-cell data with Palantir. *Nat. Biotechnol.*
1169 **37**, 451–460 (2019).
1170 79. Fang, Z., Liu, X. & Peltz, G. GSEApy: a comprehensive package for performing gene set enrichment
1171 analysis in Python. *Bioinformatics* **39**, (2023).
1172 80. Persad, S. *et al.* SEACells infers transcriptional and epigenomic cellular states from single-cell genomics
1173 data. *Nat. Biotechnol.* (2023) doi:10.1038/s41587-023-01716-9.
1174 81. Stringer, C., Wang, T., Michaelos, M. & Pachitariu, M. Cellpose: a generalist algorithm for cellular
1175 segmentation. *Nat. Methods* **18**, 100–106 (2021).
1176 82. Palla, G. *et al.* Squidpy: a scalable framework for spatial omics analysis. *Nat. Methods* **19**, 171–178
1177 (2022).
1178 83. Hie, B., Cho, H., DeMeo, B., Bryson, B. & Berger, B. Geometric Sketching Compactly Summarizes the
1179 Single-Cell Transcriptomic Landscape. *Cell Syst* **8**, 483–493.e7 (2019).

Figures

Figure 1

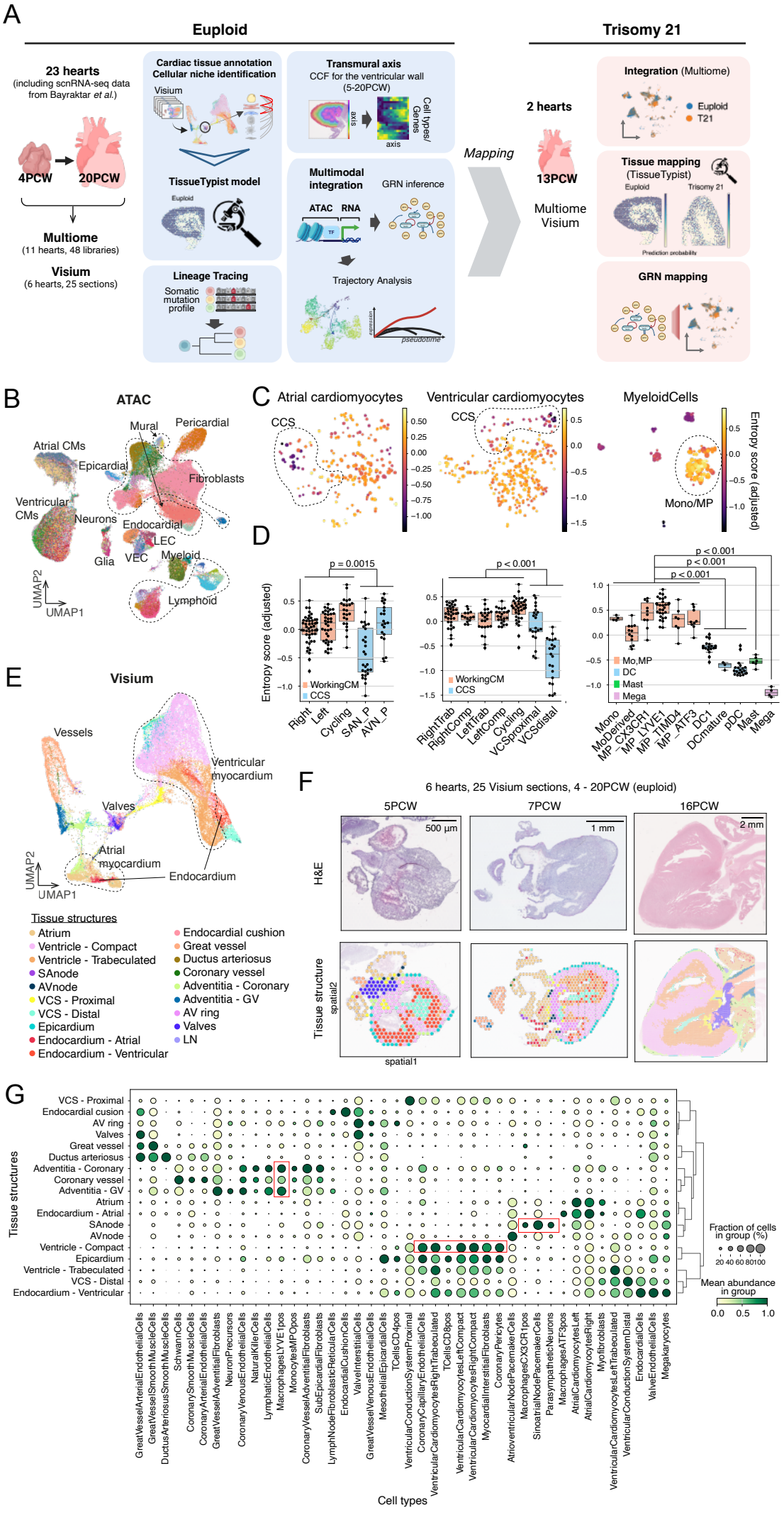

A
Euploid
Trisomy 21
23 hearts
(including scnRNA-seq data from Bayraktar et al.)
4PCW
20PCW
Multiome
(11 hearts, 48 libraries)
Visium
(6 hearts, 25 sections)
Cardiac tissue annotation
Cellular niche identification
Visium
TissueTypist model
Euploid
Lineage Tracing
Somatic mutation profile
Transmural axis
CCF for the ventricular wall
(5-20PCW)
axis
Cell types/ Genes
axis
Multimodal integration
GRN inference
ATAC
RNA
TF
Trajectory Analysis
expression
pseudotime
Mapping
2 hearts
13PCW
Multiome
Visium
Integration (Multiome)
Euploid
T21
Tissue mapping
(TissueTypist)
Euploid
Trisomy 21
Prediction probability
GRN mapping
B
ATAC
Mural
Pericardial
Atrial CMs
Epicardial
Fibroblasts
Ventricular CMs
Neurons
Endocardial
LEC
Glia
VEC
Myeloid
Lymphoid
UMAP2
UMAP1
C
Atrial cardiomyocytes
Ventricular cardiomyocytes
MyeloidCells
CCS
CCS
Mono/MP
Entropy score (adjusted)
D
p = 0.0015
p < 0.001
p < 0.001
p < 0.001
p < 0.001
Entropy score (adjusted)
WorkingCM
CCS
Right
Left
Cycling
SAN_P
AVN_P
RightTrab
RightComp
LeftTrab
LeftComp
Cycling
VCSproximal
VCSdistal
Mo,MP
DC
Mast
Mega
Mono
MoDerived
MP_CX3CR1
MP_LYVE1
MP_TIMD4
MP_ATF3
DC1
DCmature
pDC
Mast
Mega
E
Visium
Vessels
Ventricular myocardium
Valves
Atrial myocardium
Endocardium
UMAP2
UMAP1
Tissue structures
Atrium
Ventricle - Compact
Ventricle - Trabeculated
SAnode
AVnode
VCS - Proximal
VCS - Distal
Epicardium
Endocardium - Atrial
Endocardium - Ventricular
Endocardial cushion
Great vessel
Ductus arteriosus
Coronary vessel
Adventitia - Coronary
Adventitia - GV
AV ring
Valves
LN
F
6 hearts, 25 Visium sections, 4 - 20PCW (euploid)
5PCW
7PCW
16PCW
500 µm
1 mm
2 mm
H&E
Tissue structure
spatial2
spatial1
G
Tissue structures
VCS - Proximal
Endocardial cusion
AV ring
Valves
Great vessel
Ductus arteriosus
Adventitia - Coronary
Coronary vessel
Adventitia - GV
Atrium
Endocardium - Atrial
SAnode
AVnode
Ventricle - Compact
Epicardium
Ventricle - Trabeculated
VCS - Distal
Endocardium - Ventricular
Fraction of cells in group (%)
20 40 60 80 100
Mean abundance in group
0.0 0.5 1.0
GreatVesselArterialEndothelialCells
GreatVesselSmoothMuscleCells
DuctusArteriosusSmoothMuscleCells
SchwannCells
CoronarySmoothMuscleCells
CoronaryArterialEndothelialCells
GreatVesselAdventitialFibroblasts
NeuronPrecursors
CoronaryVenousEndothelialCells
NaturalKillerCells
LymphaticEndothelialCells
MacrophagesLYVE1pos
MonocytesMPopos
CoronaryVesselAdventitialFibroblasts
SubEpicardialFibroblasts
LymphNodeFibroblasticReticularCells
EndocardialCushionCells
ValveInterstitialCells
GreatVesselVenousEndothelialCells
MesothelialEpicardialCells
TCellsCD4pos
VentricularConductionSystemProximal
CoronaryCapillaryEndothelialCells
VentricularCardiomyocytesRightTrabeculated
TCellsCD8pos
VentricularCardiomyocytesLeftCompact
VentricularCardiomyocytesRightCompact
MyocardialInterstitialFibroblasts
CoronaryPericytes
AtrioventricularNodePacemakerCells
MacrophagesCX3CR1pos
SinoatrialNodePacemakerCells
ParasympatheticNeurons
MacrophagesATF3pos
AtrialCardiomyocytesLeft
AtrialCardiomyocytesRight
Myofibroblasts
VentricularCardiomyocytesLeftTrabeculated
VentricularConductionSystemDistal
EndocardialCells
ValveEndothelialCells
Megakaryocytes
Cell types

Figure 1 Multiomic atlas of the developing human heart

A: Overview of study design, data modalities, and analysis strategies. Single-cell and nucleus RNA-seq (Bayraktar S. *et al.*), Multiome, and Visium spatial transcriptomics data were generated from a total of 23 human fetal hearts between the ages of 4 and 20 PCW. Euploid datasets and models were integrated with and mapped to the newly generated Trisomy 21 datasets (13 and 14 PCW).
B: UMAP embedding of chromatin accessibility data of 63 cell types of the developing human heart and the great vessels, spread across 14 mid- and 6 coarse-grains. UMAP embedding of the cells based on the chromatin accessibility profile shares similar segregation patterns with the gene expression profile.
C-D: Entropy score (Methods) of neighbourhoods of atrial or ventricular cardiomyocytes or myeloid cells. The UMAP embeddings (C) show neighbourhoods (Milo) identified using gene expression data and scVI latent space. The dot size shows the neighbourhood size and the colour shows the entropy score. The box plots (D) indicate the entropy scores of the neighbourhoods in each assigned cell type label. Data points represent neighbourhoods. *p*-values are provided for the comparison among groups of cell types as indicated (two-sided Kruskal-Wallis test, for multiple comparisons, post hoc analysis with Dunn's test was performed with adjusting p-values using the Bonferroni method).
E: UMAP embedding of estimated cell type abundances (cell2location) of the Visium spots across 22 spatial sections. Spots were annotated based on clustering, estimated abundance of cell types in each cluster, and histological structure as defined by expert annotations using H&E images (Methods).
F: Representative Visium sections, which capture the whole heart, with structural annotations.
G: Dotplot displaying the estimated fine-grained cell (columns) abundance per tissue structure annotation (rows). Summarised abundances were normalised for each cell type.
AV: atrioventricular, AVN: AV node, GV: great vessel, H&E: hematoxylin and eosin, VCS: ventricular conduction system, VCSp: VCS proximal, CCF: common coordinate framework, GRN: gene regulatory network, PCW: post-conception weeks.

Figure 2

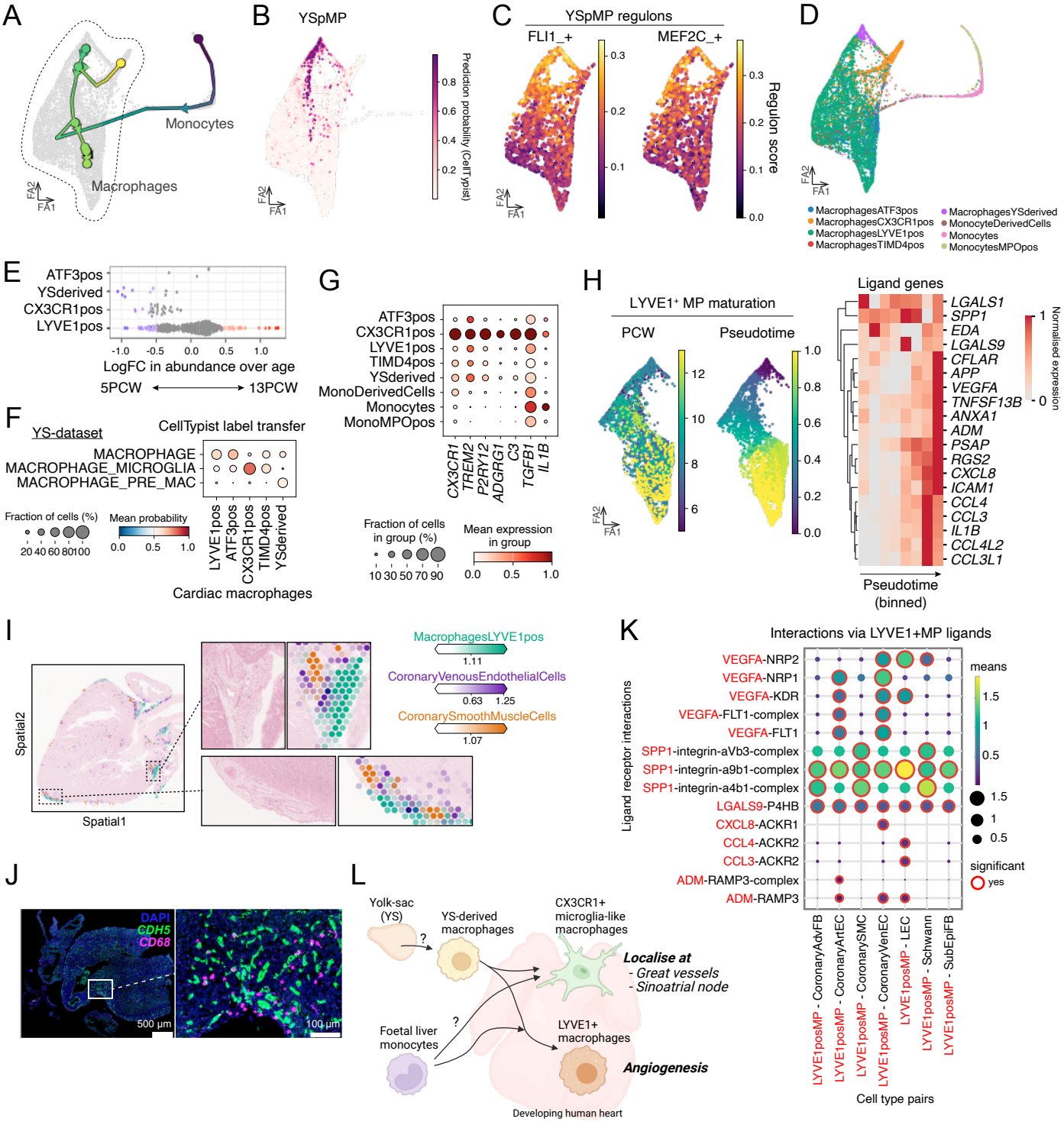

A
Monocytes
Macrophages
FA2
FA1
B
YSpMP
Prediction probability (CellTypist)
0.8
0.6
0.4
0.2
C
YSpMP regulons
FLI1_+
MEF2C_+
Regulon score
0.3
0.2
0.1
0.0
D
MacrophagesATF3pos
MacrophagesCX3CR1pos
MacrophagesLYVE1pos
MacrophagesTIMD4pos
MacrophagesYSderived
MonocyteDerivedCells
Monocytes
MonocytesMPOpos
E
ATF3pos
YSderived
CX3CR1pos
LYVE1pos
-1.0 -0.5 0.0 0.5 1.0
LogFC in abundance over age
5PCW
13PCW
F
YS-dataset
CellTypist label transfer
MACROPHAGE
MACROPHAGE_MICROGLIA
MACROPHAGE_PRE_MAC
LYVE1pos
ATF3pos
CX3CR1pos
TIMD4pos
YSderived
Fraction of cells (%)
20 40 60 80 100
Mean probability
0.0 0.5 1.0
Cardiac macrophages
G
ATF3pos
CX3CR1pos
LYVE1pos
TIMD4pos
YSderived
MonoDerivedCells
Monocytes
MonoMPOpos
CX3CR1
TREM2
P2RY12
ADGRG1
C3
TGFB1
IL1B
Fraction of cells in group (%)
10 30 50 70 90
Mean expression in group
0.0 0.5 1.0
H
LYVE1⁺ MP maturation
PCW
Pseudotime
12
10
8
6
1.0
0.8
0.6
0.4
0.2
0.0
Ligand genes
LGALS1
SPP1
EDA
LGALS9
CFLAR
APP
VEGFA
TNFSF13B
ANXA1
ADM
PSAP
RGS2
CXCL8
ICAM1
CCL4
CCL3
IL1B
CCL4L2
CCL3L1
Normalised expression
1
0
Pseudotime (binned)
I
Spatial2
Spatial1
MacrophagesLYVE1pos
1.11
CoronaryVenousEndothelialCells
0.63 1.25
CoronarySmoothMuscleCells
1.07
J
DAPI
CDH5
CD68
500 µm
100 µm
K
Interactions via LYVE1+MP ligands
Ligand receptor interactions
VEGFA-NRP2
VEGFA-NRP1
VEGFA-KDR
VEGFA-FLT1-complex
VEGFA-FLT1
SPP1-integrin-aVb3-complex
SPP1-integrin-a9b1-complex
SPP1-integrin-a4b1-complex
LGALS9-P4HB
CXCL8-ACKR1
CCL4-ACKR2
CCL3-ACKR2
ADM-RAMP3-complex
ADM-RAMP3
means
1.5
1
0.5
significant
yes
LYVE1posMP - CoronaryAdvFB
LYVE1posMP - CoronaryArtEC
LYVE1posMP - CoronarySMC
LYVE1posMP - CoronaryVenEC
LYVE1posMP - LEC
LYVE1posMP - Schwann
LYVE1posMP - SubEpiFB
Cell type pairs
L
Yolk-sac (YS)
YS-derived macrophages
CX3CR1+ microglia-like macrophages
Localise at
- Great vessels
- Sinoatrial node
Foetal liver monocytes
LYVE1+ macrophages
Angiogenesis
Developing human heart

Figure 2 Cardiac macrophage development

A: Force-directed graph of macrophages and monocytes using MultiVI latent space (based on both gene expression and chromatin accessibility). The principal graph coloured with pseudotime was calculated based on the MultiVI latent space and projected on the embedding.
B: Label transferring of yolk-sac (YS) macrophages to the foetal heart monocytes and macrophages was performed using CellTypist. The prediction probability of YS premacrophage (YSpMP) is projected on the UMAP.
C: Calculated activity scores of the YSpMP regulons (FLI1 and MEF2C)[23] are projected on the embedding.
D: Monocyte and macrophage fine-grained cell types including YS-derived macrophages projected on the embedding.
E: Bee-swarm plot of log-fold change (x-axis) in cell abundance across chronological age (Milo[14] based on the MultiVI latent space). Neighbourhoods overlapping the same cell type are grouped (y-axis). Neighbourhoods with significant differential abundance are coloured (Spatial FDR<0.1). The range of the age in the analysis is 5 to 13 PCW.
F: Dot plot showing mean probability and proportion of CellTypist prediction results for each macrophage. YS macrophage dataset[23] was used for the model training.
G: Dot plot showing the expression of maker genes for CX3CR1-positive microglia-like macrophages.
H: Cells and nuclei belong to the segments associated with the LYVE1+ macrophage maturation (Supp Fig 4A, red arrow). The colour scales show chronological age (PCW, left embedding) or normalised pseudotime (right embedding). Heatmap shows the normalised gene expression of the pseudotime-associated ligand genes (HGNC Group ID: 542). The pseudotime was binned and the mean of the values were calculated for each bin.
I: Estimated cell abundance of LYVE1+ macrophages and coronary vessel cells (cell2location). The colocation of the cell types is highlighted.
J: Multiplex smFISH (RNAscope) of the 7 PCW foetal heart region for *CD68* (yellow) and *CDH5* (red). DAPI was used to stain nuclei.
K: Inferred cell-cell interactions between LYVE1+ macrophages and the other cell types in the coronary vessel niches. Colour scale and dot size represent the mean expression levels of the interacting ligand-receptor partners. Cell-type specific interactions are delineated with a red border.
L: Schematic representation of the macrophage development in the developing heart.
MP: macrophages, PCW: post-conception weeks.

Figure 3

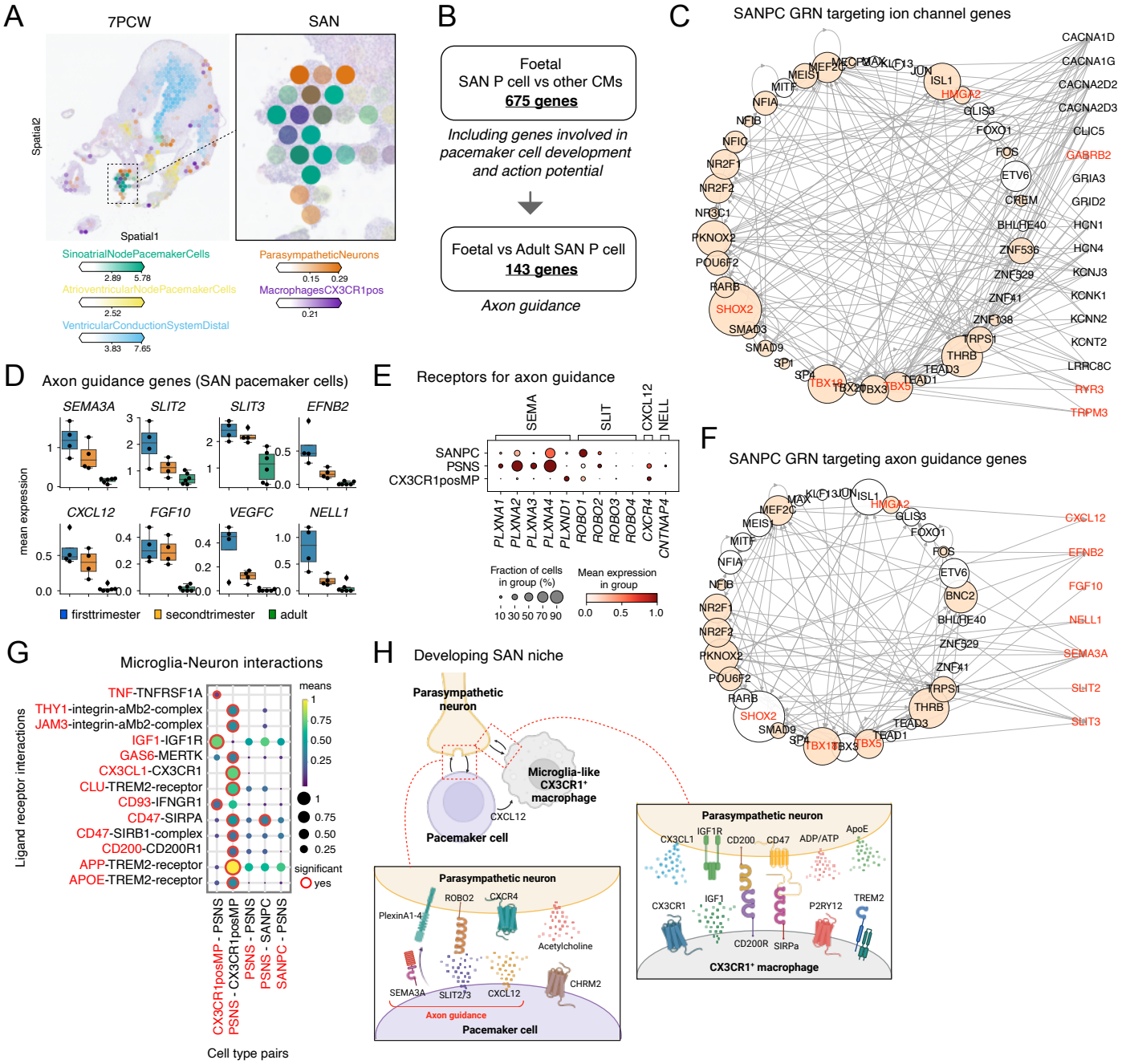

A
7PCW
SAN
Spatial2
Spatial1
SinoatrialNodePacemakerCells
2.89 5.78
AtrioventricularNodePacemakerCells
2.52
VentricularConductionSystemDistal
3.83 7.65
ParasympatheticNeurons
0.15 0.29
MacrophagesCX3CR1pos
0.21
B
Foetal
SAN P cell vs other CMs
675 genes
Including genes involved in pacemaker cell development and action potential
Foetal vs Adult SAN P cell
143 genes
Axon guidance
C
SANPC GRN targeting ion channel genes
CACNA1D
CACNA1G
CACNA2D2
CACNA2D3
CLIC5
GABRB2
GRIA3
GRID2
HCN1
HCN4
KCNJ3
KCNK1
KCNN2
KCNT2
LRRC8C
RYR3
TRPM3
D
Axon guidance genes (SAN pacemaker cells)
SEMA3A
SLIT2
SLIT3
EFNB2
CXCL12
FGF10
VEGFC
NELL1
mean expression
firsttrimester
secondtrimester
adult
E
Receptors for axon guidance
SEMA
SLIT
CXCL12
NELL
SANPC
PSNS
CX3CR1posMP
PLXNA1
PLXNA2
PLXNA3
PLXNA4
PLXND1
ROBO1
ROBO2
ROBO3
ROBO4
CXCR4
CNTNAP4
Fraction of cells in group (%)
10 30 50 70 90
Mean expression in group
0.0 0.5 1.0
F
SANPC GRN targeting axon guidance genes
CXCL12
EFNB2
FGF10
NELL1
SEMA3A
SLIT2
SLIT3
G
Microglia-Neuron interactions
Ligand receptor interactions
TNF-TNFRSF1A
THY1-integrin-aMb2-complex
JAM3-integrin-aMb2-complex
IGF1-IGF1R
GAS6-MERTK
CX3CL1-CX3CR1
CLU-TREM2-receptor
CD93-IFNGR1
CD47-SIRPA
CD47-SIRB1-complex
CD200-CD200R1
APP-TREM2-receptor
APOE-TREM2-receptor
means
significant
yes
CX3CR1posMP - PSNS
PSNS - CX3CR1posMP
PSNS - PSNS
PSNS - SANPC
SANPC - PSNS
Cell type pairs
H
Developing SAN niche
Parasympathetic neuron
Microglia-like CX3CR1+ macrophage
CXCL12
Pacemaker cell
Parasympathetic neuron
PlexinA1-4
ROBO2
CXCR4
Acetylcholine
SEMA3A
SLIT2/3
CXCL12
CHRM2
Axon guidance
Pacemaker cell
Parasympathetic neuron
CX3CL1
IGF1R
CD200
CD47
ADP/ATP
ApoE
CX3CR1
IGF1
CD200R
SIRPa
P2RY12
TREM2
CX3CR1+ macrophage

Figure 3 Cellular niche of the developing sinoatrial node

A: Estimated cell abundance of the indicated cell types is projected on the 7 PCW heart. Colocation of pacemaker cells, parasympathetic neurons, and CX3CR1+ macrophages at the SA node is highlighted.
B: Workflow of differentially expressed gene analysis for SAN pacemaker cells. Pathways identified by gene set enrichment analysis (Supp Fig 5A) are italicised.
C, F: Gene regulatory network (GRN) targeting ion channel (C) or axon guidance (F) genes specifically expressed in SAN pacemaker cells. Transcription factors (TFs) are shown in a circular arrangement with ion channel target genes (TGs) on the right. The edges represent TF-TG activation. The size of the TF nodes represents the centrality score (eigenvector) in the TF network. TFs, which directly interact with the target genes, are coloured by salmon. Genes in red are expressed higher in foetal pacemaker cells compared to their adult counterpart.
D: Expression of the genes involved in axon guidance in SAN pacemaker cells across foetal and adult[32] life stages are shown. The dots show the mean expression of each donor. For the box plots, the centre line shows the median, the box limits represent the 25th and 75th percentiles, and the whiskers show the minimum and maximum values.
E: Gene expression of the receptors of axon guidance genes in the colocating cell types at the SA node.
G: Inferred cell-cell interactions of the SAN niche. The interactions are selected based on the ligands and receptors known for neuron-microglia crosstalk. Colour scale and dot size represent the mean expression levels of the interacting ligand-receptor partners. Cell-type specific interactions are delineated with a red border.
H: Schematic representation of the developing sinoatrial nodal niche.
SAN: sinoatrial node, SANPC: SAN pacemaker cell, GRN: gene regulatory network.

Figure 4

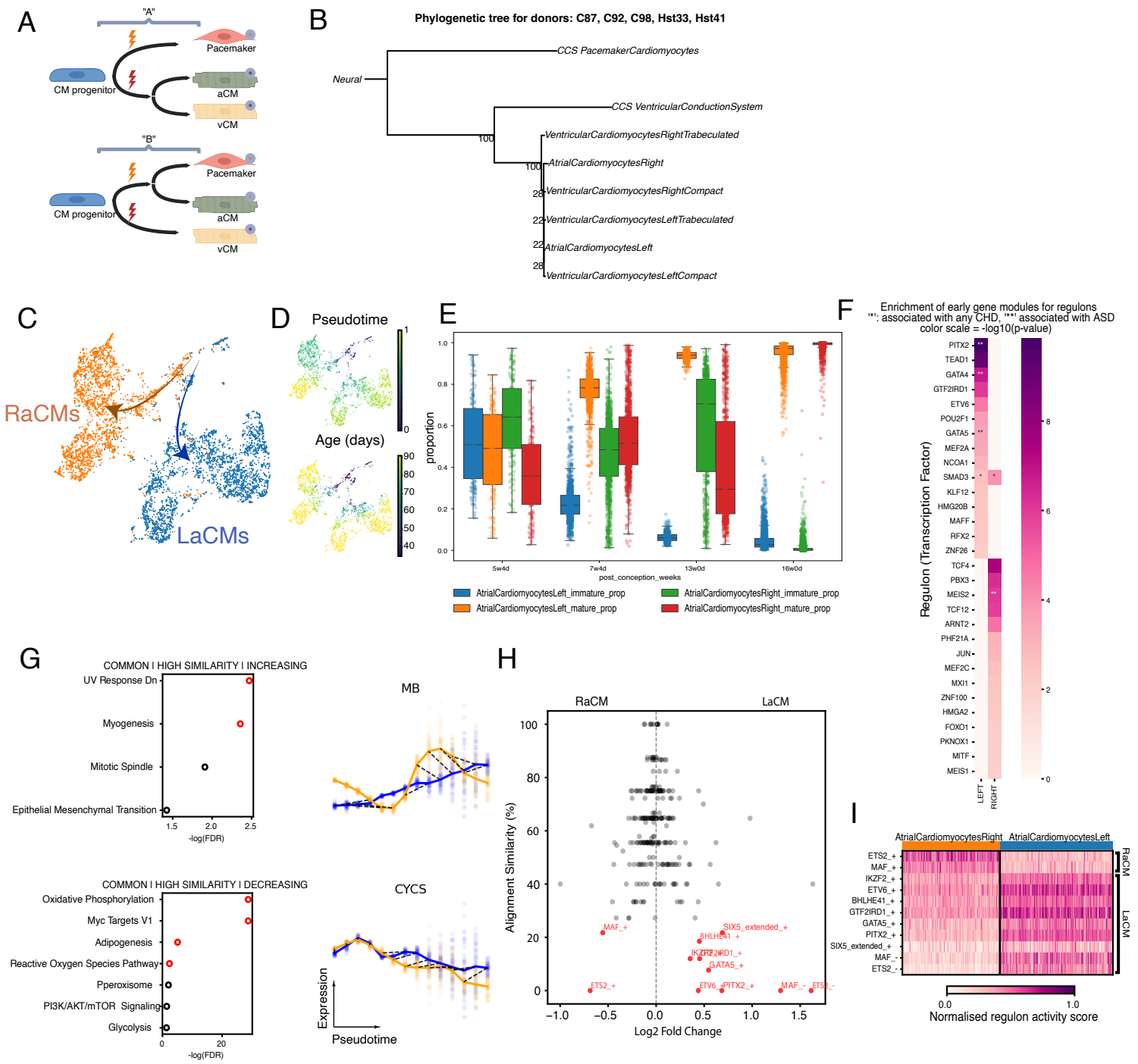

A
"A"
Pacemaker
CM progenitor
aCM
vCM
"B"
Pacemaker
CM progenitor
aCM
vCM
B
Phylogenetic tree for donors: C87, C92, C98, Hst33, Hst41
Neural
CCS PacemakerCardiomyocytes
CCS VentricularConductionSystem
VentricularCardiomyocytesRightTrabeculated
AtrialCardiomyocytesRight
VentricularCardiomyocytesRightCompact
VentricularCardiomyocytesLeftTrabeculated
AtrialCardiomyocytesLeft
VentricularCardiomyocytesLeftCompact
C
RaCMs
LaCMs
D
Pseudotime
Age (days)
E
proportion
5w4d
7w4d
13w0d
16w0d
post_conception_weeks
AtrialCardiomyocytesLeft_immature_prop
AtrialCardiomyocytesLeft_mature_prop
AtrialCardiomyocytesRight_immature_prop
AtrialCardiomyocytesRight_mature_prop
F
Enrichment of early gene modules for regulons
'*': associated with any CHD, '**' associated with ASD
color scale = -log10(p-value)
Regulon (Transcription Factor)
PITX2
TEAD1
GATA4
GTF2IRD1
ETV6
POU2F1
GATA5
MEF2A
NCOA1
SMAD3
KLF12
HMG20B
MAFF
RFX2
ZNF26
TCF4
PBX3
MEIS2
TCF12
ARNT2
PHF21A
JUN
MEF2C
MXI1
ZNF100
HMGA2
FOXO1
PKNOX1
MITF
MEIS1
LEFT
RIGHT
G
COMMON I HIGH SIMILARITY I INCREASING
UV Response Dn
Myogenesis
Mitotic Spindle
Epithelial Mesenchymal Transition
-log(FDR)
COMMON I HIGH SIMILARITY I DECREASING
Oxidative Phosphorylation
Myc Targets V1
Adipogenesis
Reactive Oxygen Species Pathway
Pperoxisome
PI3K/AKT/mTOR Signaling
Glycolysis
-log(FDR)
MB
CYCS
Expression
Pseudotime
H
RaCM
LaCM
Alignment Similarity (%)
Log2 Fold Change
MAF_+
SIX5_extended_+
BHLHE41_+
GATA5_+
ETS2_+
ETV6_+
PITX2_+
MAF_-
ETS2_-
I
AtrialCardiomyocytesRight
AtrialCardiomyocytesLeft
RaCM
LaCM
Normalised regulon activity score

Figure 4 Atrial cardiomyocyte maturation

A: Schematic illustrating alternative possible origins of cardiomyocyte subtypes. In scenario “A”, a progenitor diverges into working and pacemaker cells, with subsequent divergence of atrial and ventricular lineages. In scenario “B” the atrial vs. ventricular divergence occurs first. Lightning bolts/circles represent mutations and their accrual to decipher this split.
B: A phylogenetic tree linking cardiomyocyte subtype based on somatic mutational profiles. Branch length represents mutational distance. Numbers represent bootstrapping values for each bifurcation.
C: UMAP embedding of left and right atrial cardiomyocytes (L, RaCMs). Based on multiVI space. Arrows represent divergence over development.
D: UMAP embedding from (C) coloured by pseudotime (scFates) and chronological age (days).
E: Boxplot showing proportions of immature and mature aCMs over time using Visium data. Cell2location was used to map cell labels to Visium data using a modified fine_grain annotation which included four aCM labels: immature and mature for each of LaCMs and RaCMs (see Supp. Fig. 7E). Estimated cell type abundance data from atrial (manually annotation ‘atrial’) spots was used. For each spot the proportion of the mature or immature cell type was calculated using the total of immature and mature as the denominator (Methods).
F: Heatmap showing enrichment of SCENIC-defined regulons for L and RaCM early gene modules. Single asterisk denotes TF is a known congenital heart causal gene, double asterisk if associated with atrial septal defect.
G: Pathway enrichment results for 'high similarity' genes (>30% similarity). Term is shown if FDR < 0.1. Term is red if FDR <0.01. Scatterplot shows expression of a gene from the myogenesis term (myoglobin, *MB*) and oxidative phsophrylation (cytochrome c, *CYCS*).
H: Scatterplot of genes2genes alignment results comparing regulon activity scores as features between L and RaCM trajectories. Low similarity (<25%) regulons are highlighted in red.
I: Heatmap showing the per-cell regulon activity scores.

Figure 5

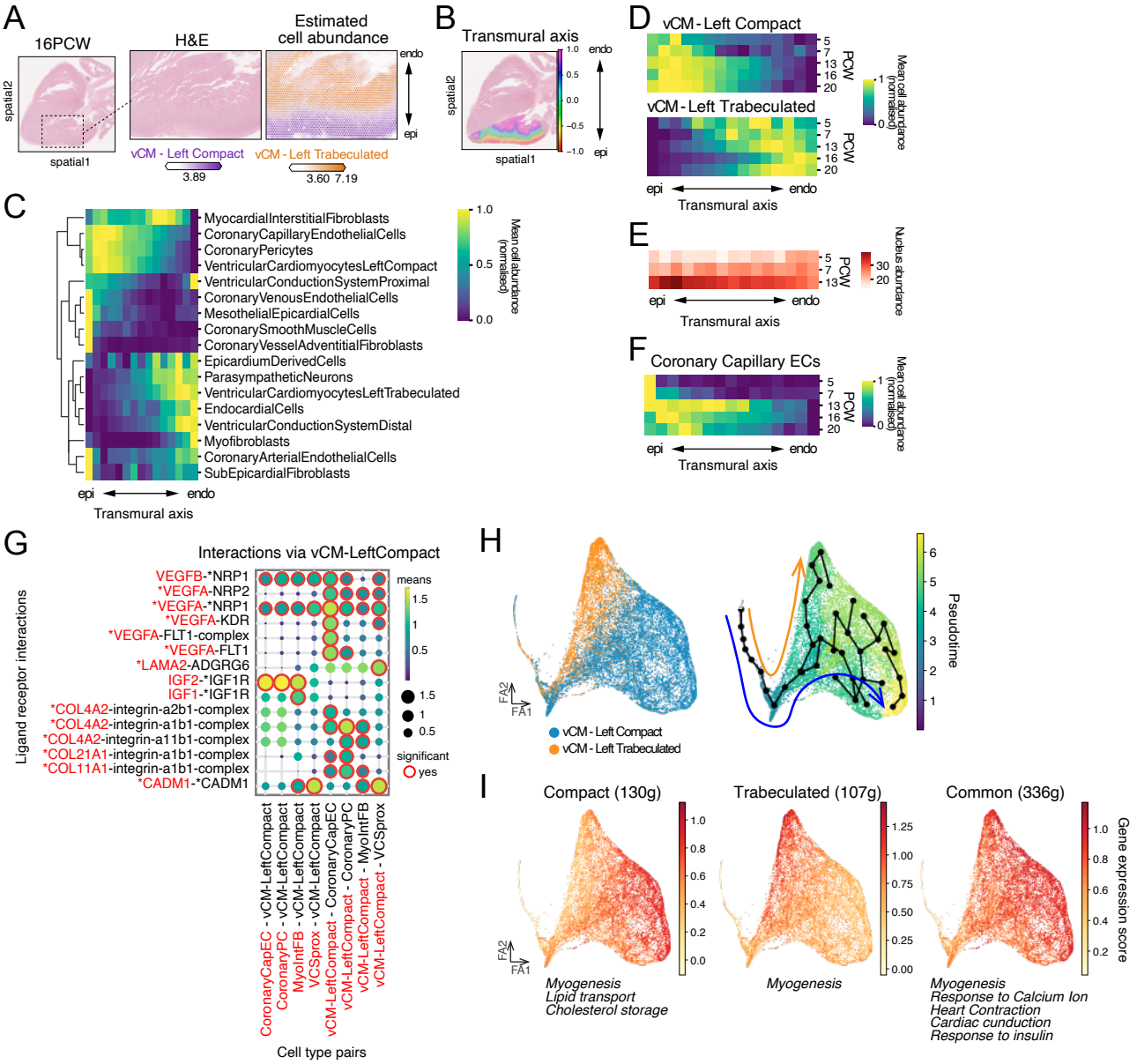

A
16PCW
H&E
Estimated cell abundance
spatial2
spatial1
endo
epi
vCM - Left Compact
vCM - Left Trabeculated
3.89
3.60 7.19
B
Transmural axis
C
MyocardialInterstitialFibroblasts
CoronaryCapillaryEndothelialCells
CoronaryPericytes
VentricularCardiomyocytesLeftCompact
VentricularConductionSystemProximal
CoronaryVenousEndothelialCells
MesothelialEpicardialCells
CoronarySmoothMuscleCells
CoronaryVesselAdventitialFibroblasts
EpicardiumDerivedCells
ParasympatheticNeurons
VentricularCardiomyocytesLeftTrabeculated
EndocardialCells
VentricularConductionSystemDistal
Myofibroblasts
CoronaryArterialEndothelialCells
SubEpicardialFibroblasts
Mean cell abundance (normalised)
Transmural axis
D
vCM - Left Compact
vCM - Left Trabeculated
PCW
Mean cell abundance (normalised)
E
Nucleus abundance
F
Coronary Capillary ECs
G
Interactions via vCM-LeftCompact
Ligand receptor interactions
VEGFB-*NRP1
*VEGFA-NRP2
*VEGFA-*NRP1
*VEGFA-KDR
*VEGFA-FLT1-complex
*VEGFA-FLT1
*LAMA2-ADGRG6
IGF2-*IGF1R
IGF1-*IGF1R
*COL4A2-integrin-a2b1-complex
*COL4A2-integrin-a1b1-complex
*COL4A2-integrin-a11b1-complex
*COL21A1-integrin-a1b1-complex
*COL11A1-integrin-a1b1-complex
*CADM1-*CADM1
means
significant
yes
CoronaryCapEC - vCM-LeftCompact
CoronaryPC - vCM-LeftCompact
MyoIntFB - vCM-LeftCompact
VCSprox - vCM-LeftCompact
vCM-LeftCompact - CoronaryCapEC
vCM-LeftCompact - CoronaryPC
vCM-LeftCompact - MyoIntFB
vCM-LeftCompact - VCSprox
Cell type pairs
H
FA2
FA1
vCM - Left Compact
vCM - Left Trabeculated
Pseudotime
I
Compact (130g)
Trabeculated (107g)
Common (336g)
Gene expression score
Myogenesis
Lipid transport
Cholesterol storage
Myogenesis
Myogenesis
Response to Calcium Ion
Heart Contraction
Cardiac cunduction
Response to insulin
J
TF network for compact cardiomyocyte maturation
Perturbation score (sum)
K
HEY2 KO
PRDM16 KO
NFE2L1 KO
SREBF2 KO

Figure 5 Compact and trabeculated myocardium development

A: Abundance of compact and trabeculated cardiomyocytes at the left ventricle region of 16 PCW Visium section as estimated by cell2location.
B: Assignment of transmural axis for the Visium spots on the left ventricle region.
C: Cell type abundance transition across the transmural axis of Visium data. The colour scale shows the scaled value (min-max normalisation per cell type) of the mean cell abundance per transmural axis bin.
D-F: Normalised cell abundance (D, F) or nuclei number per spot (E) along the left ventricular transmural axis and the chronological age. The transmural axis was binned and the mean of the values were calculated for each bin and age.
G: Inferred cell-cell interactions between the left ventricular compact cardiomyocytes (vCM-LeftCompact) and the other cell types in the ventricular compact niche. The colour scale and dot size represent the mean expression levels of the interacting ligand-receptor partners. Cell type-specific interactions are delineated with a red border. Compact cardiomyocyte signature genes are highlighted with asterisks.
H: Force-directed graph of both gene expression and chromatin accessibility data for left ventricular compact cardiomyocytes. Pseudotime was calculated based on the MultiVI latent space (plot on the right).
I: Compact-enriched (130 genes), trabeculated-enriched (107 genes) and commonly shared (336 genes) signatures associated with the pseudotime in (H). Pathways identified by gene set enrichment analysis (Supp Fig 8G) are italicised.
J: TF network targeting compact-specific and commonly upregulated signatures. GRN was constructed using SCENIC+, and the interactions for the TF network were selected as described in Methods. The size of the nodes represents the centrality score (eigenvector) in the TF network. The colour scale indicates the *in-silico* perturbation score obtained from CellOracle analysis (Methods). A negative score means that the TF perturbation will impede the trajectory (compact cardiomyocyte maturation).
K: Representative plots of TF perturbation simulation results. Showing perturbation scores for impact on the maturation of compact cardiomyocytes overlaid on the compact cardiomyocyte regions of the UMAP embedding (trabeculated cardiomyocyte regions are intentionally blank).
vCM: ventricular cardiomyocytes, EC: endothelial cell, epi: epicardium, endo: endocardium, TF: transcription factor, KO: knock out.

Figure 6

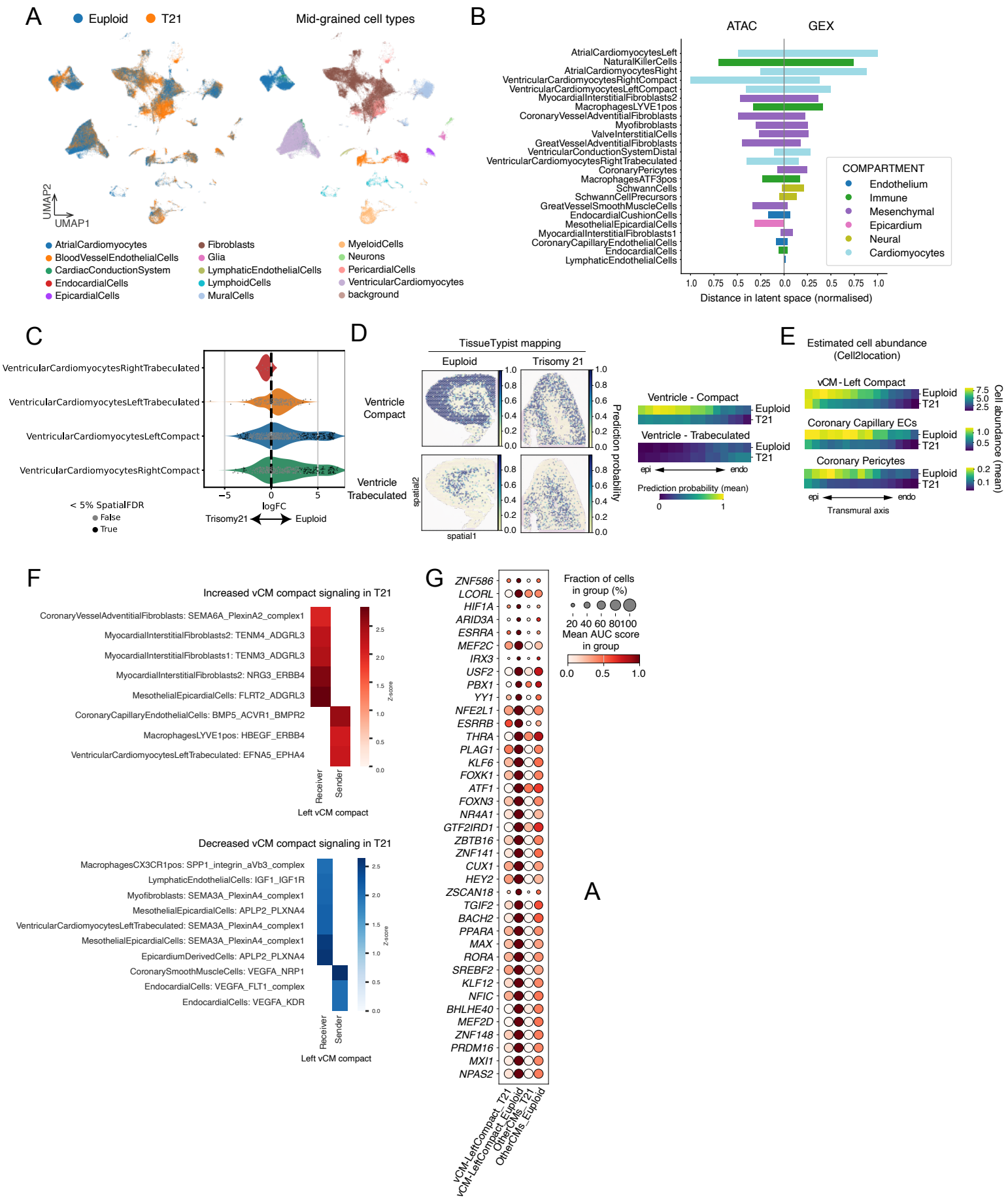
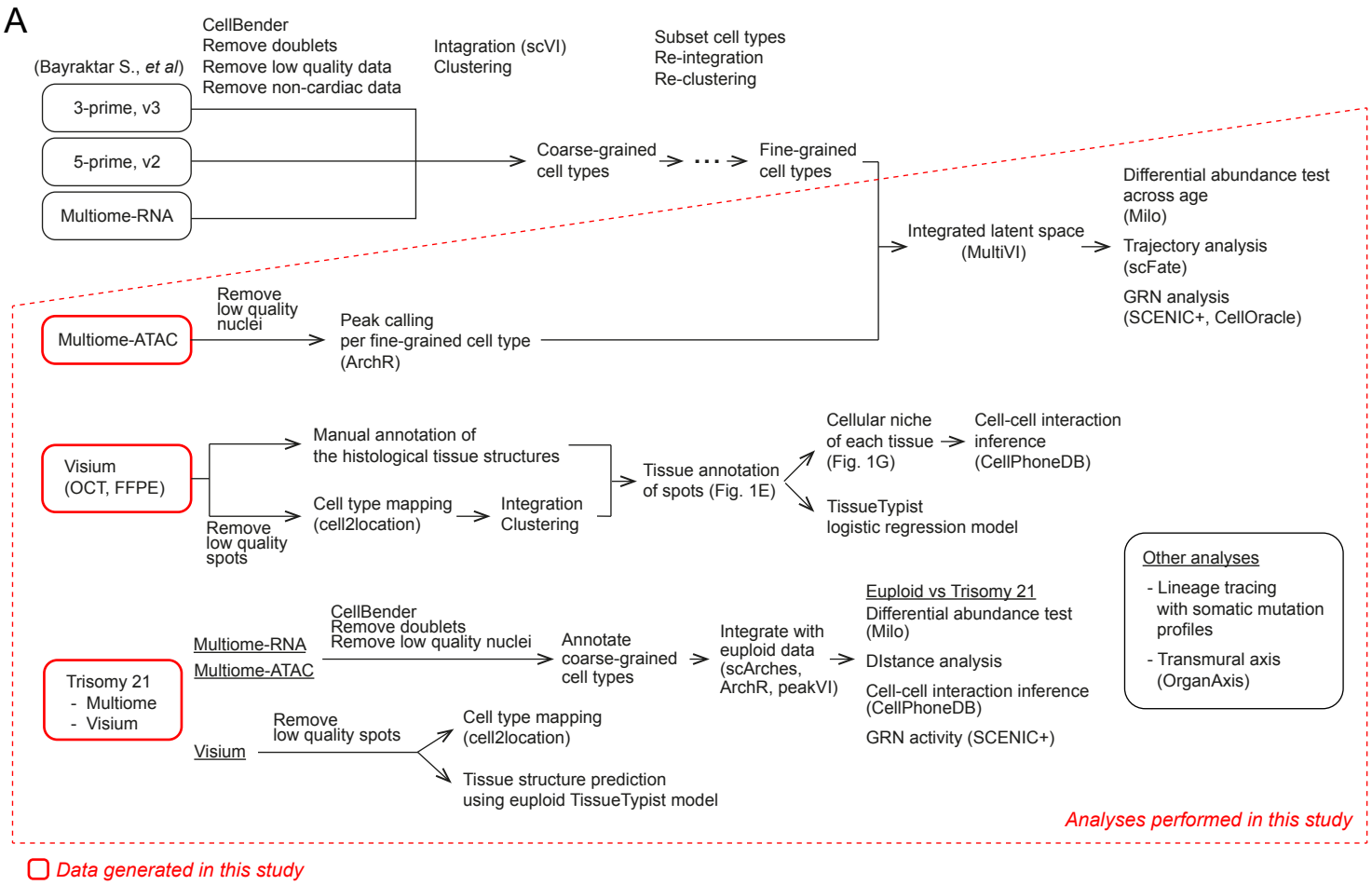
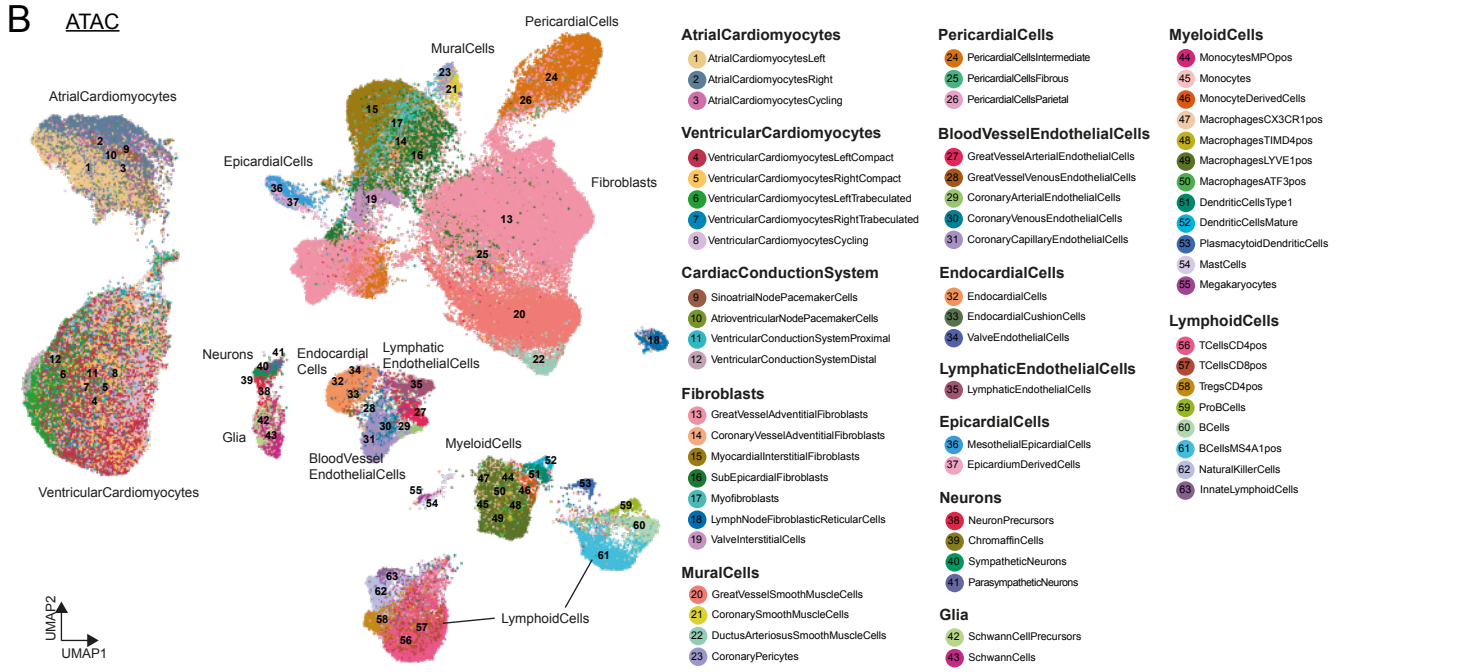

A
Euploid
T21
Mid-grained cell types
UMAP2
UMAP1
AtrialCardiomyocytes
BloodVesselEndothelialCells
CardiacConductionSystem
EndocardialCells
EpicardialCells
Fibroblasts
Glia
LymphaticEndothelialCells
LymphoidCells
MuralCells
MyeloidCells
Neurons
PericardialCells
VentricularCardiomyocytes
background
B
ATAC
GEX
AtrialCardiomyocytesLeft
NaturalKillerCells
AtrialCardiomyocytesRight
VentricularCardiomyocytesRightCompact
VentricularCardiomyocytesLeftCompact
MyocardialInterstitialFibroblasts2
MacrophagesLYVE1pos
CoronaryVesselAdventitialFibroblasts
Myofibroblasts
ValveInterstitialCells
GreatVesselAdventitialFibroblasts
VentricularConductionSystemDistal
VentricularCardiomyocytesRightTrabeculated
CoronaryPericytes
MacrophagesATF3pos
SchwannCells
SchwannCellPrecursors
GreatVesselSmoothMuscleCells
EndocardialCushionCells
MesothelialEpicardialCells
MyocardialInterstitialFibroblasts1
CoronaryCapillaryEndothelialCells
EndocardialCells
LymphaticEndothelialCells
COMPARTMENT
Endothelium
Immune
Mesenchymal
Epicardium
Neural
Cardiomyocytes
Distance in latent space (normalised)
C
VentricularCardiomyocytesRightTrabeculated
VentricularCardiomyocytesLeftTrabeculated
VentricularCardiomyocytesLeftCompact
VentricularCardiomyocytesRightCompact
< 5% SpatialFDR
False
True
logFC
Trisomy21
Euploid
D
TissueTypist mapping
Euploid
Trisomy 21
Ventricle Compact
Ventricle Trabeculated
spatial1
spatial2
Prediction probability
Ventricle - Compact
Ventricle - Trabeculated
Euploid
T21
epi
endo
Prediction probability (mean)
E
Estimated cell abundance (Cell2location)
vCM - Left Compact
Coronary Capillary ECs
Coronary Pericytes
Cell abundance (mean)
Transmural axis
F
Increased vCM compact signaling in T21
CoronaryVesselAdventitialFibroblasts: SEMA6A_PlexinA2_complex1
MyocardialInterstitialFibroblasts2: TENM4_ADGRL3
MyocardialInterstitialFibroblasts1: TENM3_ADGRL3
MyocardialInterstitialFibroblasts2: NRG3_ERBB4
MesothelialEpicardialCells: FLRT2_ADGRL3
CoronaryCapillaryEndothelialCells: BMP5_ACVR1_BMPR2
MacrophagesLYVE1pos: HBEGF_ERBB4
VentricularCardiomyocytesLeftTrabeculated: EFNA5_EPHA4
Receiver
Sender
Left vCM compact
Z-score
Decreased vCM compact signaling in T21
MacrophagesCX3CR1pos: SPP1_integrin_aVb3_complex
LymphaticEndothelialCells: IGF1_IGF1R
Myofibroblasts: SEMA3A_PlexinA4_complex1
MesothelialEpicardialCells: APLP2_PLXNA4
VentricularCardiomyocytesLeftTrabeculated: SEMA3A_PlexinA4_complex1
MesothelialEpicardialCells: SEMA3A_PlexinA4_complex1
EpicardiumDerivedCells: APLP2_PLXNA4
CoronarySmoothMuscleCells: VEGFA_NRP1
EndocardialCells: VEGFA_FLT1_complex
EndocardialCells: VEGFA_KDR
G
ZNF586
LCORL
HIF1A
ARID3A
ESRRA
MEF2C
IRX3
USF2
PBX1
YY1
NFE2L1
ESRRB
THRA
PLAG1
KLF6
FOXK1
ATF1
FOXN3
NR4A1
GTF2IRD1
ZBTB16
ZNF141
CUX1
HEY2
ZSCAN18
TGIF2
BACH2
PPARA
MAX
RORA
SREBF2
KLF12
NFIC
BHLHE40
MEF2D
ZNF148
PRDM16
MXI1
NPAS2
Fraction of cells in group (%)
20 40 60 80100
Mean AUC score in group
0.0 0.5 1.0
vCM-LeftCompact_T21
vCM-LeftCompact_Euploid
OtherCMs_T21
OtherCMs_Euploid
A

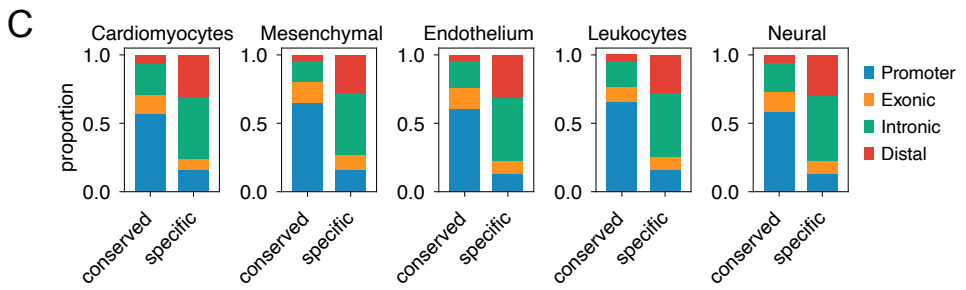
Figure 6 Trisomy 21 disrupts compact myocardium formation

A: UMAP embedding showing snRNA-seq data generated from Trisomy 21 (T21) integrated with the euploid dataset using scArches (left) with corresponding mid-grained cell type labels (right).
B: T21 impacts the expression and chromatin accessibility profiles of different cell types unevenly. Bar chart showing normalised inter-centroid distances calculated from the scArches latent space for each fine-grained label between the euploid and T21 dataset, with the largest shifts observed in cardiomyocytes, natural killer cells, and mesenchymal cells. Analysis was restricted to fine-grained cell labels with more than 50 cells in the T21 dataset.
C: Bee-swarm plot of log-fold change (x-axis) in cell abundance between euploid and T21 (Milo[14] based on the scArches latent space). Neighbourhoods overlapping the same cell type are grouped (y-axis). Neighbourhoods with significant differential abundance are coloured black (Spatial FDR<0.1). The analysis shows that a large number of neighbourhoods, in both left and right ventricular compact cardiomyocytes, are significantly enriched for euploid cells. This indicates compact cardiomyocytes are less abundant in T21 hearts.
D: TissueTypist model prediction of 13PCW euploid and trisomy 21 visium sections using the model trained on the euploid visium dataset and the spot annotations (ventricle model). 13 PCW euploid data was excluded from the training dataset for this analysis. The probability of the “Ventricle - Compact” label is shown on the Visium section and across the transmural axis.
E: Cell distribution across the transmural axis is abnormal in T21 heats. Heatmaps show scaled cell abundances for left ventricular compact cardiomyocytes, coronary capillary endothelial cells and coronary pericytes across the transmural axis, calculated using cell2location.
F: DEG-informed differential cell-cell signalling in T21. Indicates compact cardiomyocyte-centric interactions which are increased (red) or decreased (blue) in T21 compared to euploid.
G: Dot plot showing the regulon activity scores (AUC score) of compact cardiomyocyte GRNs, whose scores were altered in the T21 left compact cardiomyocytes compared to the euploid counterpart (logFC>0.3).
T21: trisomy 21, vCM: ventricular cardiomyocytes, GEX: gene expression.

Supplementary Figure 1





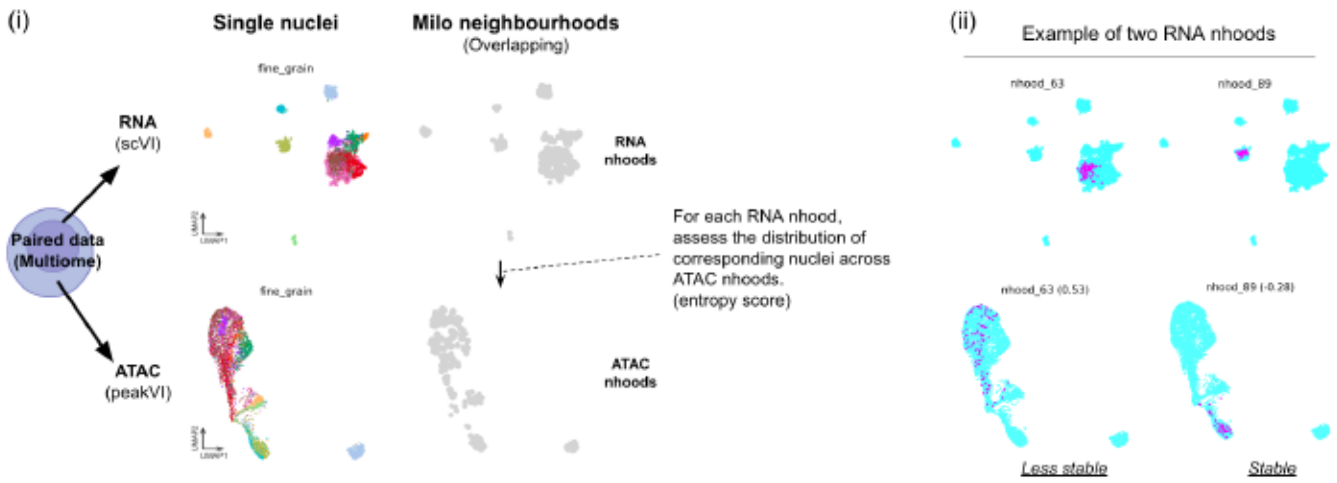


Supplementary Figure 1

A: Workflow of the data processing and analyses performed in this study.
B: UMAP embedding of the chromatin accessibility data labelled with the fine-grained cell types. 63 fine-grained cell type annotations are provided adjacent to the embeddings, grouped under their respective mid-grained cell types.
C: The proportion of the annotated regions (ArchR) in conserved or fine-grained-cell-type specific accessible regions for each coarse-grained cell type.

Supplementary Figure 2

A



B

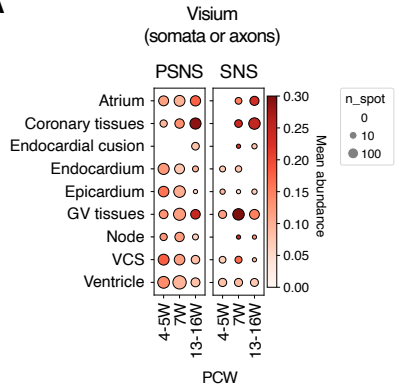
Supplementary Figure 2

A: (i) The workflow of the epigenetic stability assessment. Using Multiome data, for each mid-grained cell type, the overlapping neighbourhoods (Milo) were found using gene expression data and scVI latent space (RNA neighbourhoods) or chromatin accessibility data and peakVI latet space (ATAC neighbourhoods). For each RNA neighbourhood, the entropy of the distribution of corresponding nuclei across the ATAC neighbourhoods was calculated. Data points represent neighbourhoods. (ii) Example distribution of two RNA neighbourhoods in the ATAC space (UMAP embedding based on peakVI latent space) are shown. Data points represent cells. RNA neighbourhood-63 shows greater scattered distribution in the ATAC space and a higher entropy score (0.53), suggesting that the neighbourhood cells are epigenetically unstable and potentially in the process of transitioning to another state.

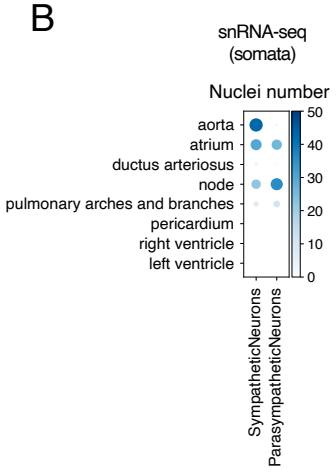
B: Entropy score (plot on the left of each mid-grained cell type) or assigned fine-grained cell type labels (plot on the right) projected on the RNA space (UMAP embedding based on scVI latent space). Data points represent RNA neighbourhoods, and the dot size shows the neighbourhood size.

Supplementary Figure 3

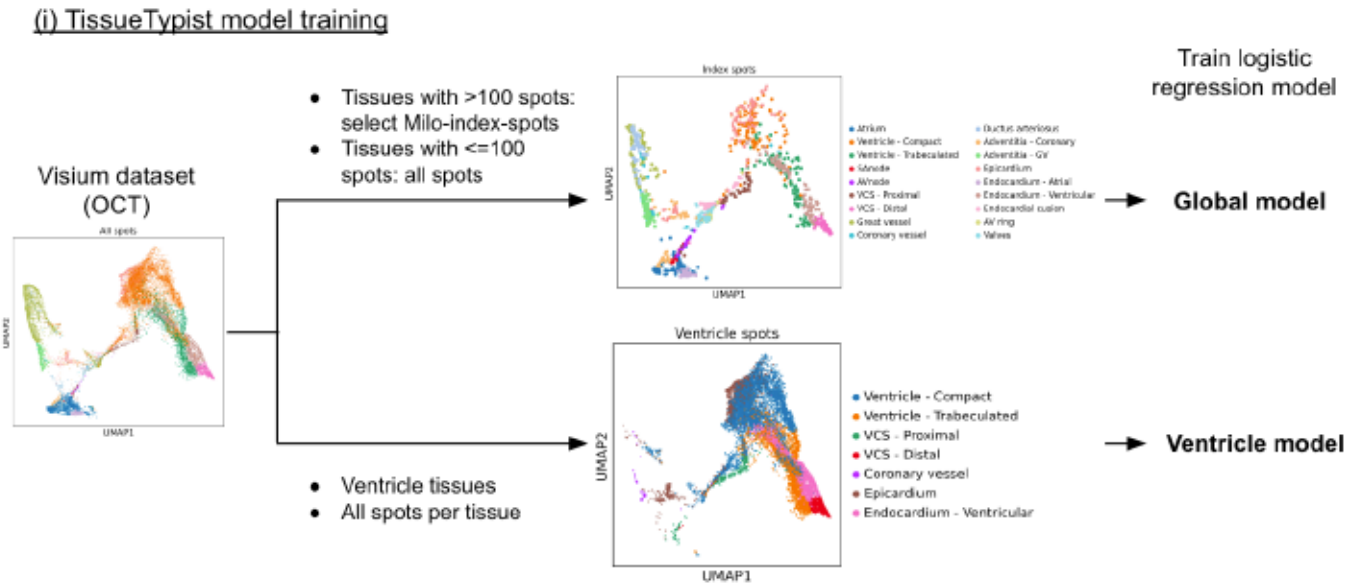
A



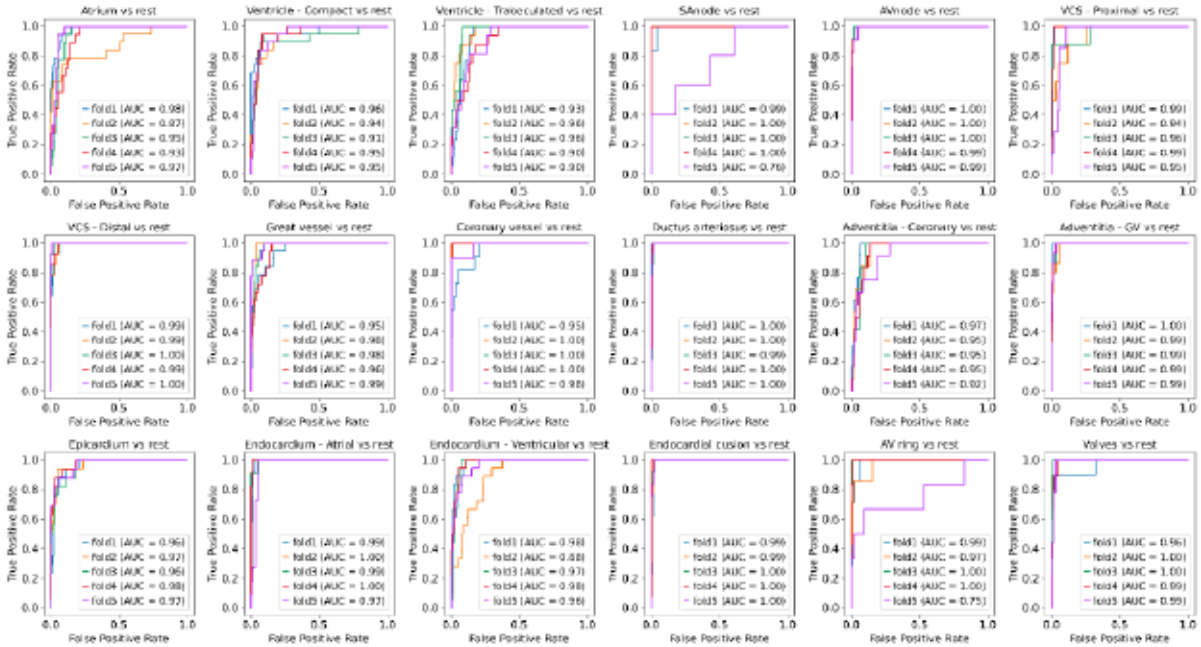
B



C



D



Supplementary Figure 3

A: Dot plot showing estimated cell abundance (cell2location) of parasympathetic neurons (PSNS) or sympathetic neurons (SNS) along age in the Visium data. The colour scale represents the mean cell abundance, and the dot size represents the number of spots which include PSNS or SNS (abundance threshold=0.05) in each annotated tissue and age bin as indicated. This result includes localisations of both somata and axons of the neuronal populations.

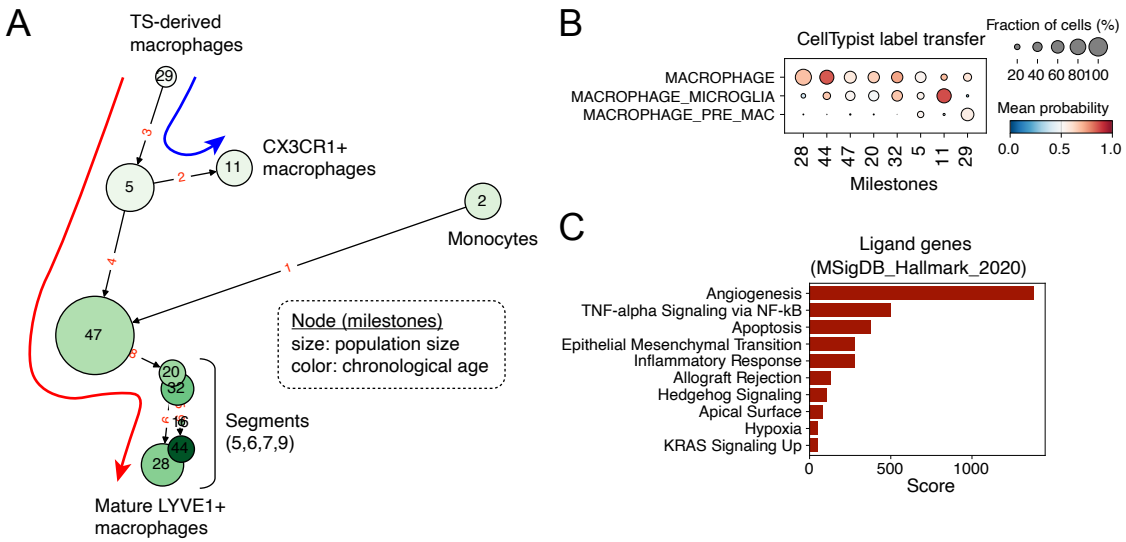
B: Dot plot showing nuclei number of sympathetic or parasympathetic neurons in each anatomical region. The snRNA-seq data with the anatomical regions dissected was used for this analysis (14, 15, and 20 PCW hearts). The colour scale and the dot size represent the nuclei number. This result indicates localisations of somata.

C: The workflow of the TissueTypist model training. The Visium OCT datasets were integrated, and tissue annotation of each spot was performed based on clustering, estimated abundance of cell types in each cluster, and histological structure as defined by expert annotations using H&E images (Methods). The global model was trained by using the spots from all the annotated tissues. To balance the data point number of each tissue, spot neighbourhoods (Milo) were found for the tissues with more than 100 spots, and the index spot of each neighbourhood was used for the training. For the ventricle model, spots from the ventricular tissues (Ventricle - Compact, Ventricle - Trabeculated, Endocardium - Ventricular, Epicardium, VCS - Distal, VCS - Proximal, Coronary vessel) were used for the training.

D: The performance of the global model was assessed through cross-validation using the same spots that were employed for training, and the model's efficacy was quantified by calculating the AUC score derived from the ROC curve.

PSNS: parasympathetic neurons, SNS: sympathetic neurons.

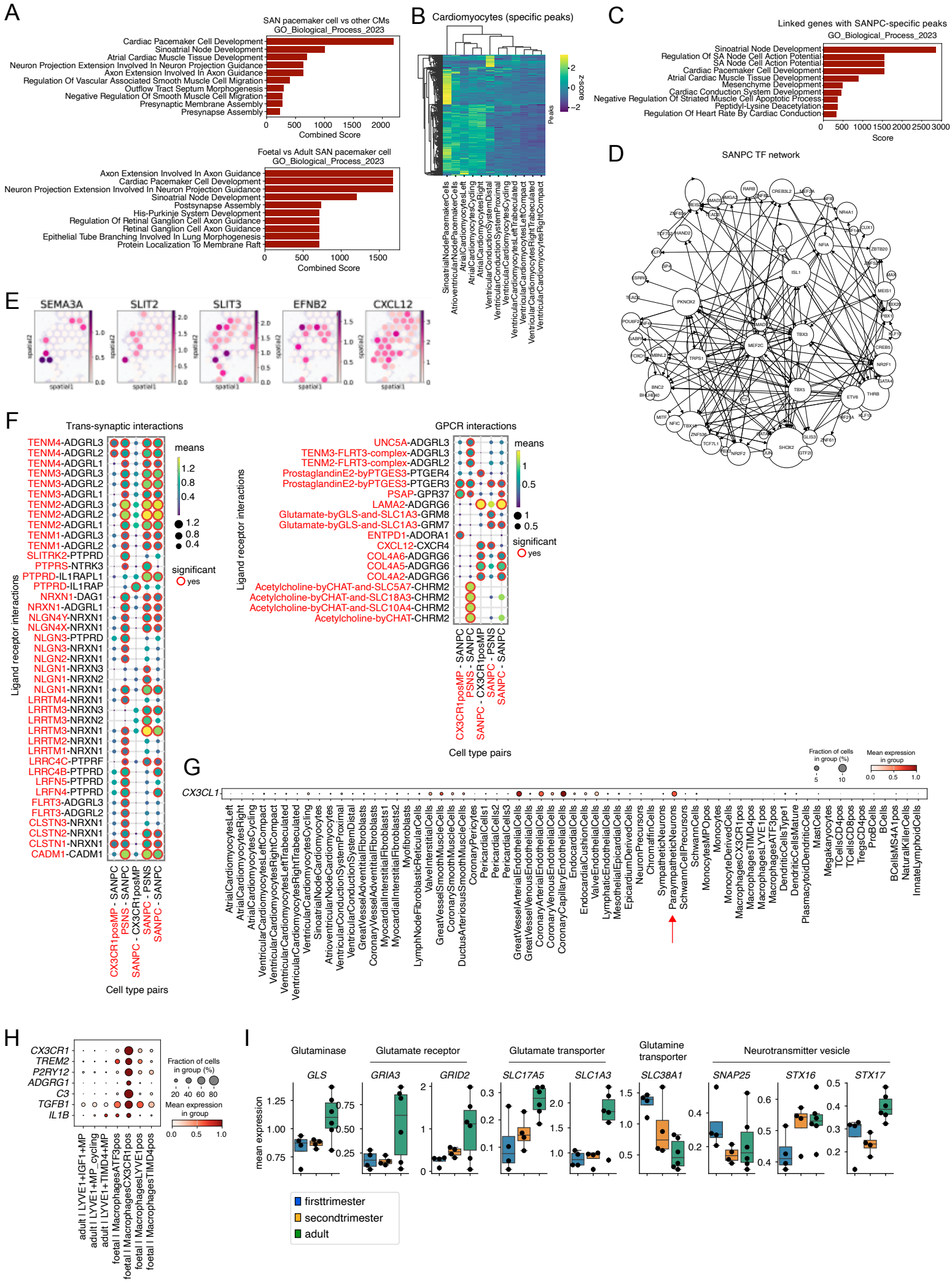
Supplementary Figure 4



Supplementary Figure 4

A: Graph showing macrophage and monocyte trajectory milestones and segments defined using MultiVI latent space and scFate workflow. The trajectories of LYVE1+ (red) and CX3CR1+ (blue) macrophage maturation are highlighted with arrows. The youngest milestone of macrophage (milestone 29) and monocyte (milestone 2) were set as the roots for the pseudotime calculation.
B: Dot plot showing mean probability and proportion of CellTypist prediction results for each milestone. YS macrophage dataset was used for the model training.
C: Pathway enrichment analysis of the genes associated with the LYVE1+ macrophage maturation trajectory.

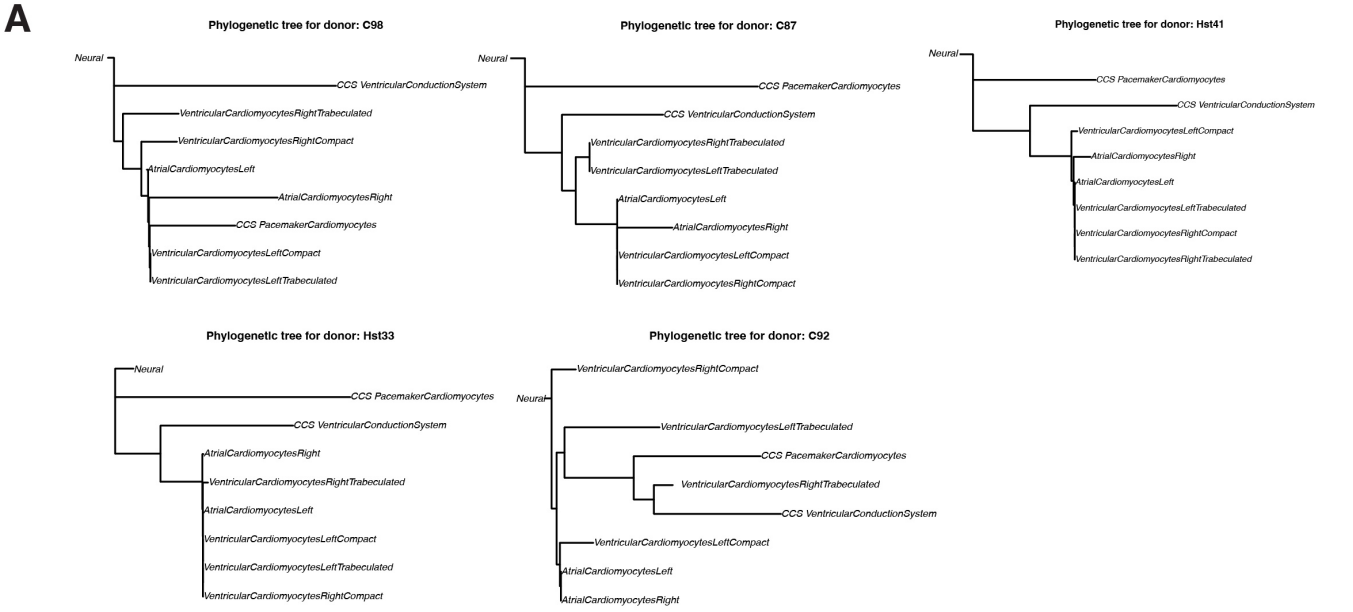
Supplementary Figure 5

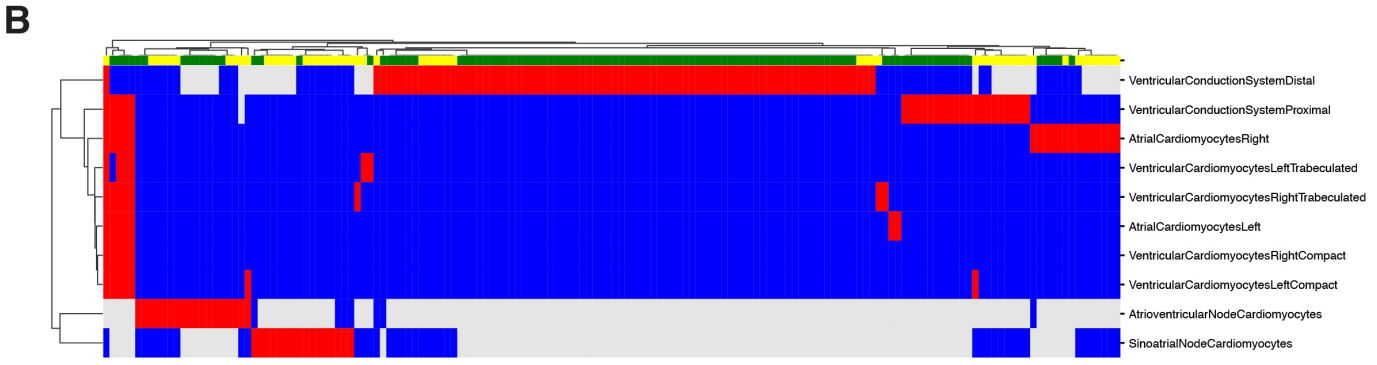


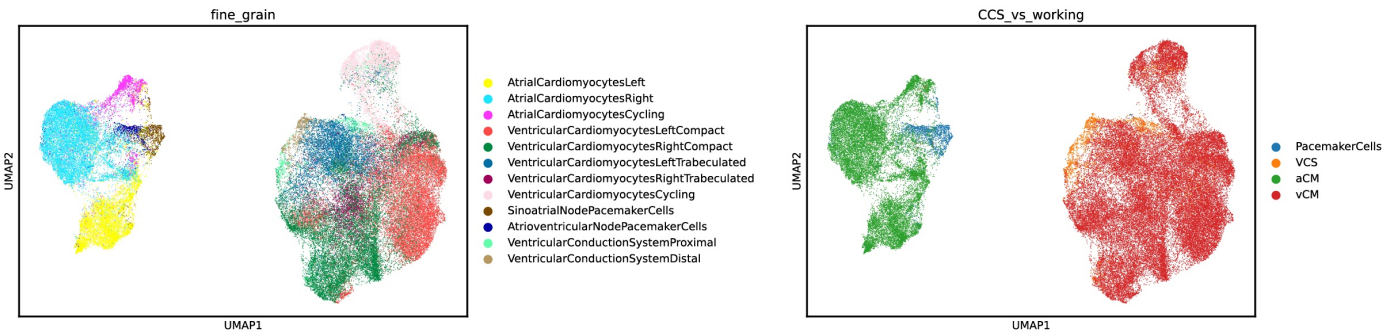
Supplementary Figure 5

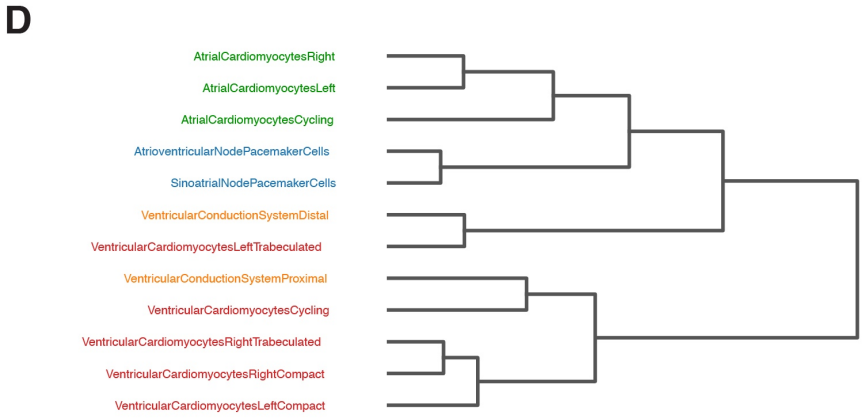
A: Pathway enrichment analysis (GO, biological process) of the genes highly expressed in the foetal SAN pacemaker cells compared to the other foetal atrial cardiomyocytes (top) or the genes narrowed down by comparing to the adult SAN pacemaker cells (bottom).
B: Heatmap showing fine-grained-cell-type specific accessible regions in cardiomyocytes. Colour scale shows scaled peak counts (z-score).
C: Pathway enrichment analysis (GO, biological process) of the genes linked with the SAN pacemaker cell-specific accessible regions. Peak-to-gene linking was obtained by performing SCENIC+ analysis using cardiomyocyte data.
D: TF-TF interactions of the gene regulatory network targeting SAN pacemaker cell gene signatures (675 genes, Fig.3B). The node size represents the eigenvector centrality in the TF network.
E: Spatial expression of the genes involved in axon guidance in the SA node of a Visium section (7 PCW, same section as Fig 3A).
F: Inferred cell-cell interactions (trans-synaptic, GPCR) between the SAN niche cell types. Colour scale and dot size represent the mean expression levels of the interacting ligand-receptor partners. Cell-type specific interactions are delineated with a red border.
G: Dotplot showing the *CX3CL1* expression across the fine-grained cell types from the all time points.
H: Dot plot showing expression of microglial genes involved in neuronal development and homeostasis, comparing adult and foetal cardiac macrophage populations.
I: Expression of the genes involved in glutamate signalling in the SAN pacemaker cells are shown. The dots show the mean expression of each donor. For the box plots, the centre line shows the median, the box limits represent the 25th and 75th percentiles, and the whiskers show the minimum and maximum values.

Supplementary Figure 6





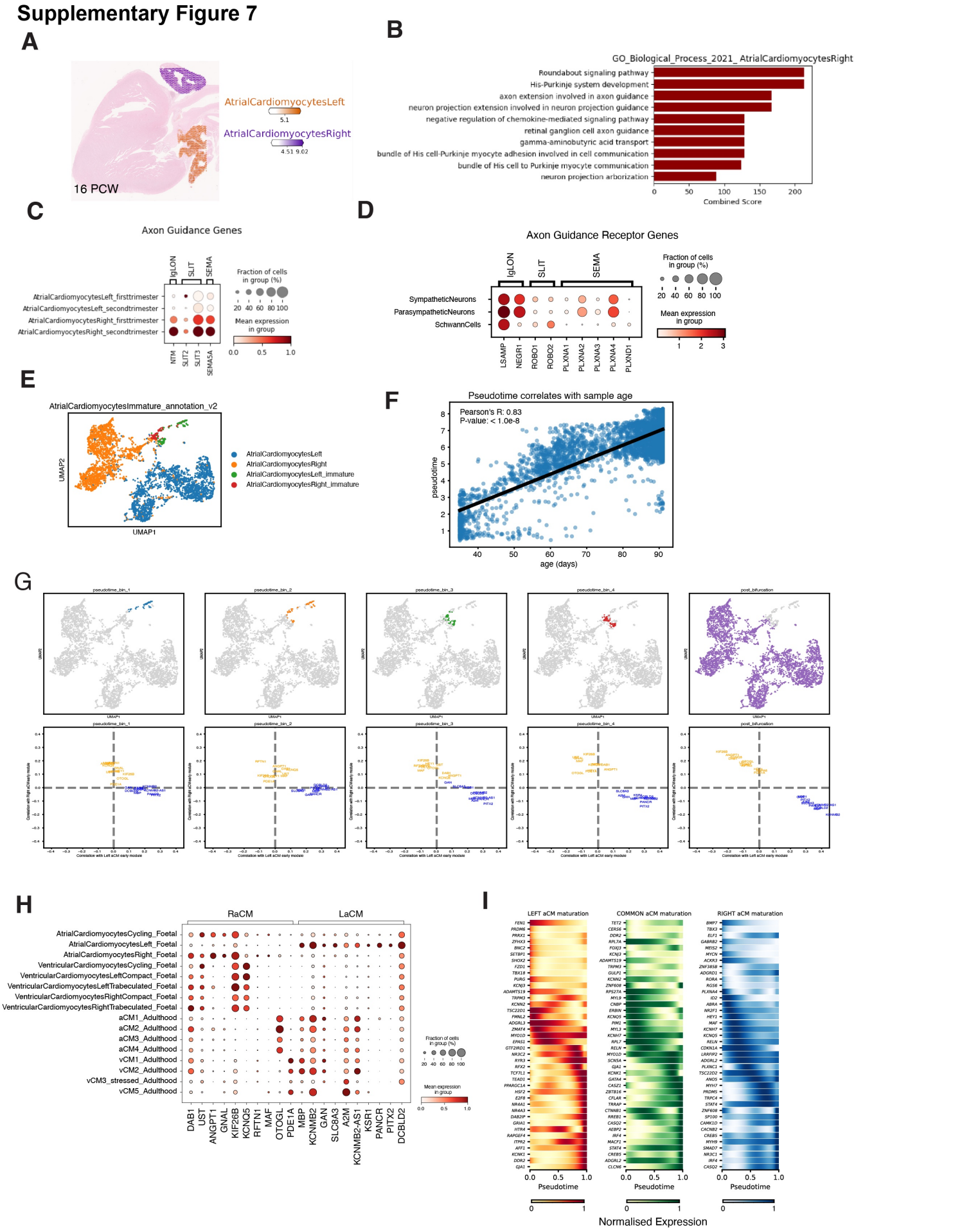




Supplementary Figure 6

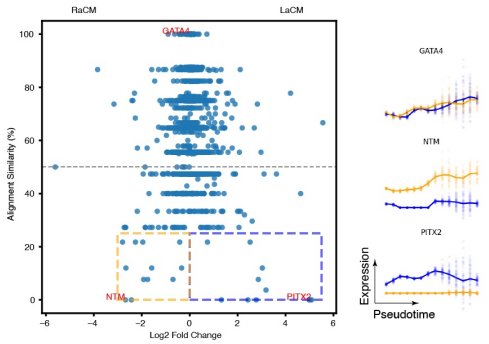
A: Phylogenetic trees for five individual donors.
B: Binarised cell type by mutation matrix for donor Hst41. Red: mutation callable and present. Blue: mutation callable and absent. Grey: mutation not callable (insufficient depth). Column colours represent mutation source (RNA in yellow, ATAC in green). Row and column hierarchical clustering based on euclidean distance.
C: UMAP embeddings (scVI space) of cardiomyocyte subtypes, coloured by fine-grained annotation and showing working vs conduction system cell types.
D: Dendrogram based on euclidean distance in scVI space linking cardiomyocyte subtypes. Label coloured according to working vs conduction system annotation (as per C).

Supplementary Figure 7

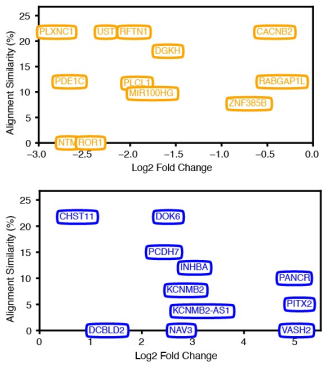


Supplementary Figure 7

J

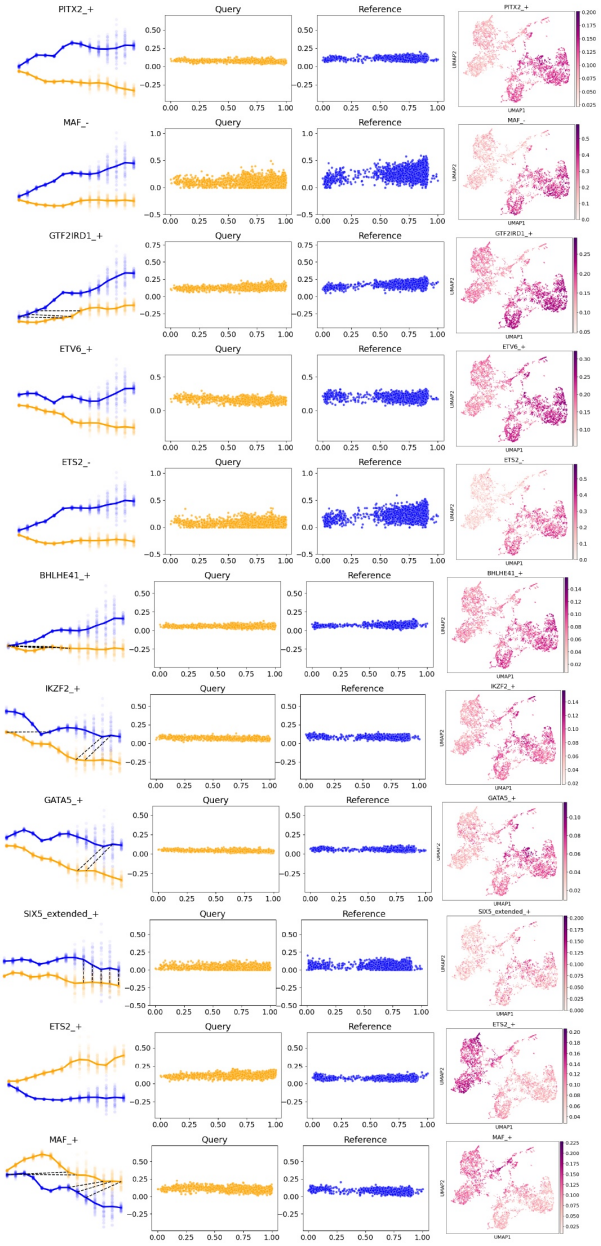


K



L

M



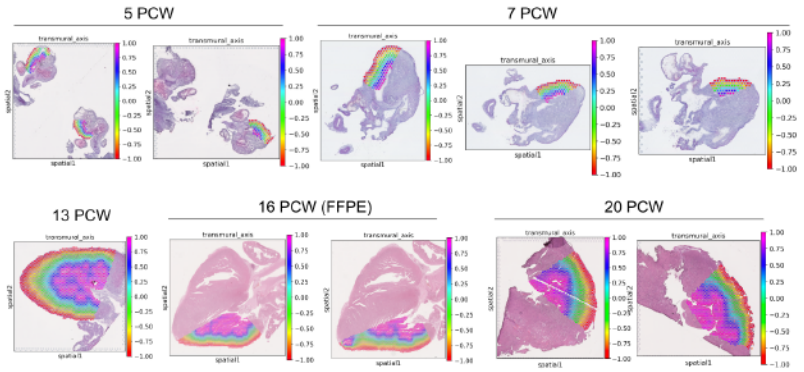
Supplementary Figure 7

A: Cell2location mapping of L and R atrial cardiomyocyte signatures in a 16PCW heart.
B: Bar plot showing pathway enrichment analysis result for the genes upregulated in RaCMs compared to LaCMs. Threshold set at log fold change > 2, adjusted p-values <0.05, method: t-test.
C: Dot plot showing expression of four axon guidance genes which are upregulated in RaCMs vs LaCMs. Data is grouped by a concatenation of cell type and trimester.
D: Dot plot showing mean expression of receptors for the ligands in C in neural cell types.
E: UMAP embedding (multiVI space) showing annotation of immature atrial cardiomyocytes (these signatures were used for cell2location analysis in Fig. 4D,E.
F: Scatterplot showing correlation between chronological age in day (smoothed across neighbourhood graph) and pseudotime.
G: Divergence analysis. Upper row: UMAP plots showing pseudotumoral bins pre and post the divergence of the principal graph. Lower row: Scatterplots based on mean expression in the corresponding bin of cells. Plots show 'early genes' for LaCM (blue) and RaCM (orange) trajectories. Axis represent correlation of each gene with expression of other genes in its own early module, vs correlation with the expression of the genes in the other module. Genes become increasingly anti-correlated towards the divergence.
H: Dot plot showing expression of the L and RaCM early genes in foetal and adult[32] cardiomyocyte subtypes.
I: Heatmap showing gene expression fitted to pseudotime for each of L and RaCM genes as well as genes associated with pseudotime in both trajectories (see Methods). Genes plotted are transcription factors (using http://humantfs.ccbr.utoronto.ca/download/v_1.01/DatabaseExtract_v_1.01.csv), ion channel genes (HGNC group 177), GPCR genes (HGNC group 139) and myosin genes (HGNC groups 656, 657). The forty genes with the most significant association with pseudotime are plotted. Each gene's expression values were min-max normalised for plotting.
J: Scatterplot showing genes2genes results. A genes2genes aligner was constructed for the pseudotime-associated genes (from L and RaCM trajectories). Each gene is plotted with its % similarity and log2 Fold Change. Smaller scatterplots showing expression over pseudotime are shown for the genes in red.
K: Scatterplot (with gene names as points) showing low similarity (<25% similarity) genes. Plots refer to dashed boxes in J.
L: Matrix plot showing mean expression of *WNT5A* and *SEMA7A* in foetal and adult[32] cardiomyocyte subtypes.
M: Scatterplots and UMAP plots showing activity of low similarity (<25%) regulons from Fig. 4I. LaCM data is in blue, RaCM data is in orange.

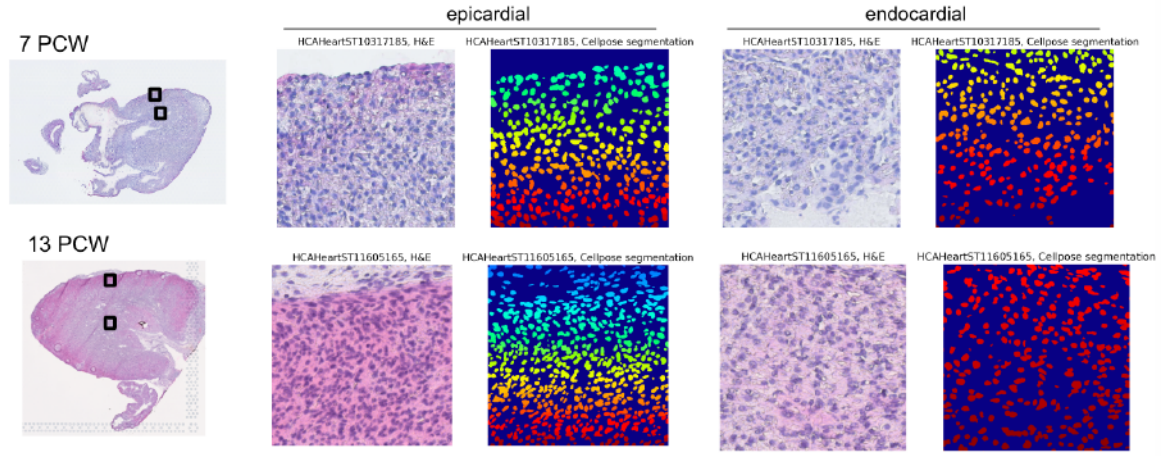
Supplementary Figure 8

A

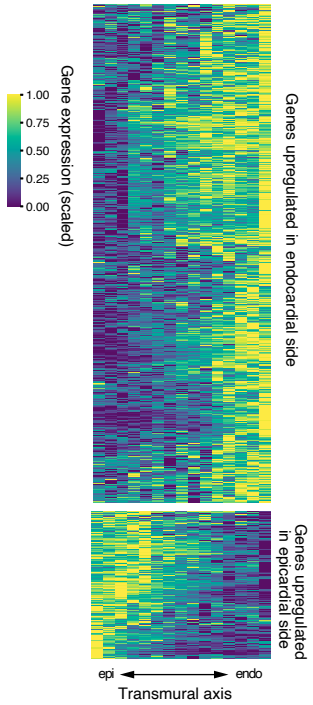
Transmural axis calculated at the LV free wall

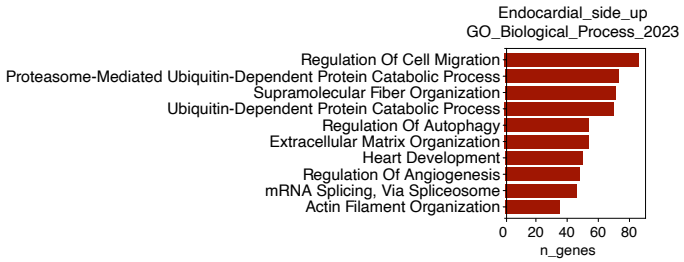


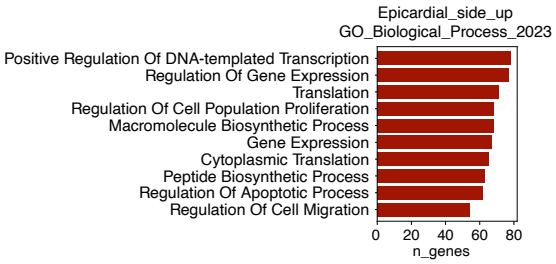
B



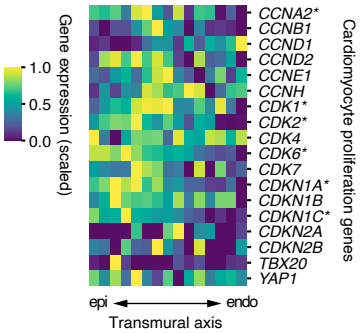
C



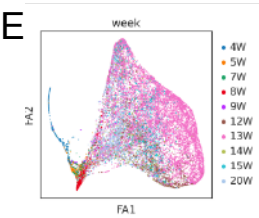




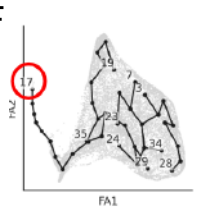
D



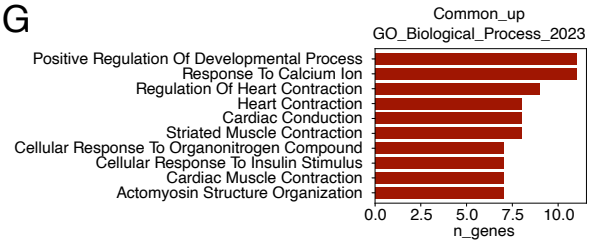
E

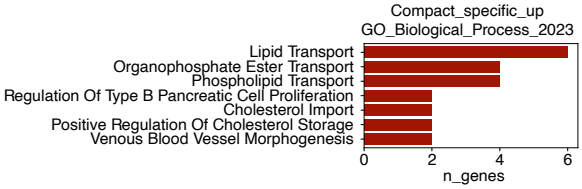


F



G

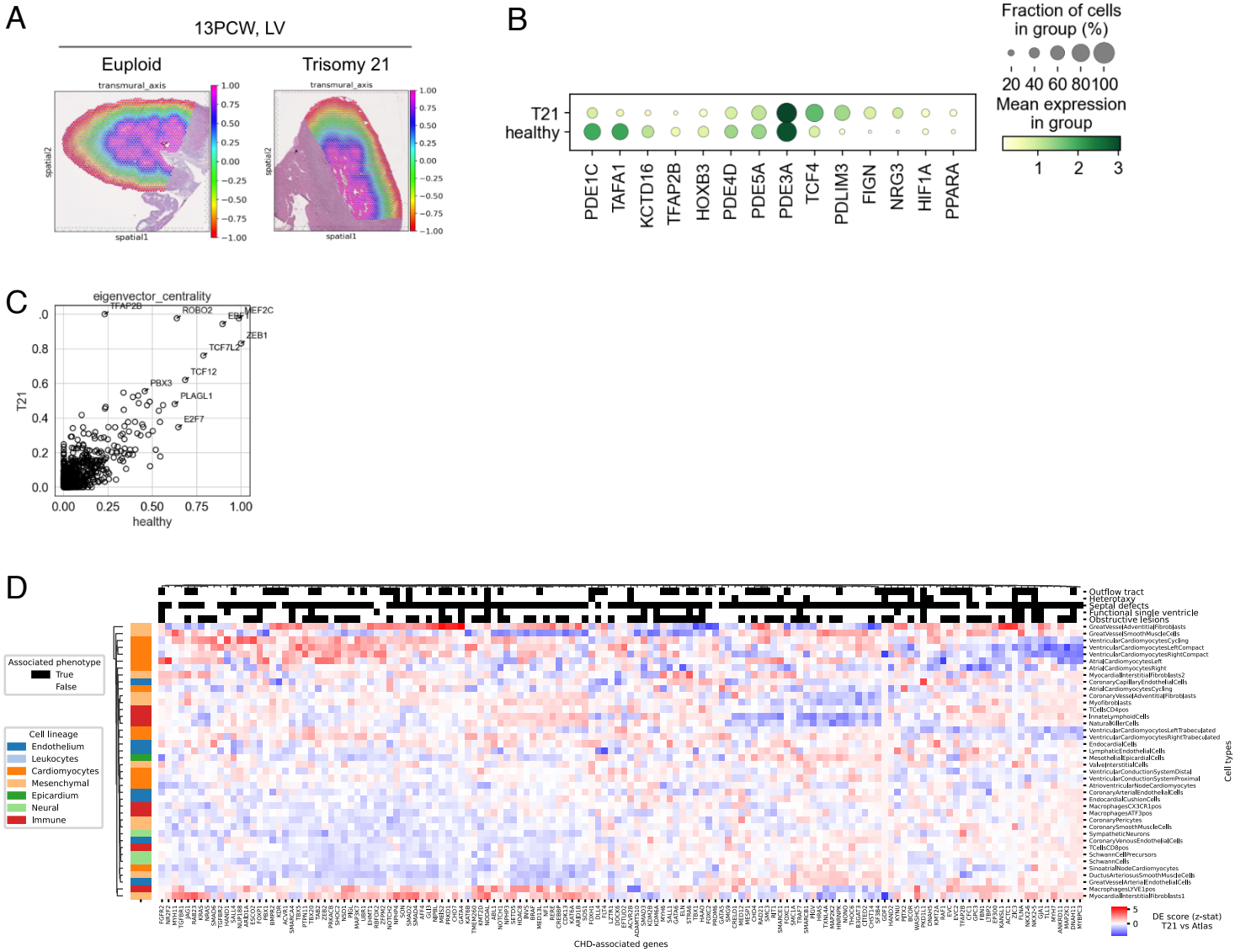




Supplementary Figure 8

A: Common coordinate framework assignment on the left ventricle free wall using OrganAxis[9] (transmural axis) of the Visium sessions across multiple ages.
B: Representative nucleus segmentation result (CellPose) of left ventricle free wall.
C: Transition of scaled gene expression (min-max normalisation per gene) across the transmural axis. Pathways enrichment analysis (GO, biological process) of the genes either downregulated or upregulated toward endocardial side are shown.
D: Transmural transition of genes previously shown to be involved in cardiomyocyte proliferation. Genes which were significantly higher in the epicardial side or lower in the endocardial side were highlighted with asterisks.
E, F: Force-directed graph of both gene expression and chromatin accessibility data for left vCMs. The color scales show chronological age (week)(D). The principal graph (E) was calculated based on the multiVI latent space and projected on the embedding. The youngest milestone (milestone 17) was set as the root for the pseudotime calculation.
G: Pathway enrichment analysis of the commonly upregulated signatures (left) and the compact-specific signatures (right).

Supplementary Figure 9



Supplementary Figure 9

A: Common coordinate framework assignment on the euploid and trisomy 21 left ventricle free wall (13PCW) using OrganAxis (transmural axis).
B: Dot plot showing gene expressions of DASMCs from euploid and trisomy 21 hearts.
C: Eigenvector centrality score (CellOracle) of cell type-specific GRNs in age-matched euploid and T21 DASMCs.
D: Analysis of expression differences for a panel of CHD genes between the fine-grain cell labels of euploid and T21 hearts. Differential testing was performed comparing the euploid and trisomy 21 cells of each fine_grain cell state. The values plotted are the log fold change values for each CHD gene for each pairwise DE test.