Organ-specific and conserved regulatory logic orchestrates gene expression in the embryonic mesothelium

Quang Minh Dang[1,2], Nicola Smart[1], Andia Nicole Redpath[1,*], and Joaquim Miguel Vieira[1,3,*]

[1]Institute of Developmental and Regenerative Medicine, Department of Physiology, Anatomy & Genetics, University of Oxford, Oxford OX3 7TY, UK.

[2]High-Tech Center, Vinmec International Healthcare System, Hanoi, Vietnam

[3]School of Cardiovascular & Metabolic Medicine & Sciences, King's College London, London SE5 9NU, UK.

*Joint Senior Contribution and Corresponding Author email: andia.redpath@dpag.ox.ac.uk and joaquim.nunes_vieira@kcl.ac.uk

Lead contact: joaquim.nunes_vieira@kcl.ac.uk

Highlights

- **Coelomic mesothelia display distinct transcriptome linked to their morphogenic role**
- **Mesothelial markers associate with organ-specific *trans*- and *cis*-regulatory logic**
- **Pan-mesothelial gene regulatory mechanisms control canonical marker expression**
- **Integrating mouse and human datasets uncovers conserved dynamics of epicardial EMT**

Summary

The embryonic coelomic mesothelium undergoes epithelial-to-mesenchymal transition (EMT) to promote vascular growth and parenchymal development. A prominent example is the epicardium, which plays an essential role during heart development. Little is known about the mechanisms behind gene regulation in the coelomic mesothelium, or the organ-specific enhancer logic that endows specialization. Using gene regulatory network inference via multi-omic analysis, our study reveals *trans*- and *cis*-regulatory elements (CREs) that regulate mesothelial gene expression in three organs: heart, lung, and pancreas. We delineate pivotal transcription factors (TFs) and CREs specific to the epicardium, and show, in contrast, that the TF MAF orchestrates pan-mesothelial gene expression *via* conserved CREs, which are absent in non-mesothelial lineages. MAF may preserve mesothelial identity, evidenced by negative correlation with EMT and mouse-human conservation. Collectively, our work elucidates the regulatory logic behind cell type identity and leverages single-cell integration to gain insights into mammalian organ development.

Introduction

The mesothelium, an epithelium surrounding body cavities and visceral organs, serves as a critical interface between tissues. One of the best-known mesothelial tissues is the epicardium, which encases the heart. From embryonic day (E)11.5, epicardial cells undergo epithelial-to-mesenchymal transition (EMT) to form epicardium-derived cells (EPDCs). These mesenchymal derivatives invade the myocardium, differentiating into cardiac fibroblasts and mural cells. EPDCs thus provide cardiac precursors and, with the epicardium, supply paracrine signals that regulate myocardial growth, maturation, and coronary vessel development [1].

The coelomic mesothelium is also integral to the development of other organs it envelops, such as the lungs, liver, pancreas, and gonads, as reviewed elsewhere [2]. Despite organ diversity, the mesothelium has shared features, including common genetic markers such as Wilms' tumour 1 (*Wt1)*, which distinguish it from other lineages [3-7]. Using these markers, lineage-tracing has revealed another common trait: multipotency. Mesothelial cells undergo EMT, generating diverse mesenchymal fates, including fibroblasts [4, 8, 9] and smooth muscle cells [4, 10-12]. Consequently, *Wt1* knockout impairs EMT, causing significant defects in the heart and lungs [3, 11, 13]. *Wt1* expression diminishes as development progresses and

epicardial cells transition to a mesenchymal state [14], a pattern recapitulated in lung mesothelial cells [4].

Despite commonalities, mesothelial cells exhibit organ-specific differences. A key divergence is their origin. Most coelomic mesothelial cells arise from coelomic lining cells that acquire epithelial features [2], but the epicardium emerges from the proepicardial organ, a transient structure at the heart's venous pole [15]. Another divergence is the terminal fates of their mesenchymal derivatives. Mesothelial cells in chest cavity organs and the liver generate pericytes [5, 10, 16-18]. However, pancreatic pericytes derive from a distinct *Nkx3.2*-expressing mesenchyme [19]. Also, unlike in the liver, mesothelium-derived stellate cells in the pancreas are myofibroblast-like[7, 20]. Additionally, mesothelial cells also have organ-specific functions. In the pancreas, *Wt1*-positive mesothelial cells provide boundary separation from the stomach [21]. Failure of mesothelial layer formation causes adhesion of the dorsal pancreas to the stomach, and systemic *Wt1* deletion leads to abnormal dorsal pancreas localization [7]. Therefore, contrasting the pancreas with the heart and lungs, a prototypical mesothelium-encased organ, can reveal the full spectrum of mesothelial behaviour.

These observations underscore the coelomic mesothelium's importance as a source of organ-specific cell types and regulatory signals that shape visceral morphogenesis. Deciphering the underlying gene regulatory networks (GRNs) is key to understanding mesothelial identity and diversification. GRNs, which map interactions between transcription factors (TFs) and cis-regulatory elements (CREs) including enhancers, control gene expression and cellular identity [22, 23]. A GRN-centric analysis is therefore central to understanding the mesothelium's cellular plasticity.

Here, we performed ATAC-seq of the *in vivo* epicardium at E11.5 and E17.5 and *in vivo* EPDCs at E13.5. We integrated these data with available RNA-seq, ATAC-seq, and CUT&RUN-seq datasets to profile the GRNs governing mesothelial spatiotemporal behaviour in the heart, lung, and pancreas. We identify that specific CREs and TFs likely control spatiotemporal mesothelial expression and highlight their potential relevance to epicardial EMT. We also documented conserved, pan-mesothelial regulatory elements and mechanisms absent from non-mesothelial lineages. We identified several such elements driving the spatiotemporal expression of cytokeratins, linked to MAF. By curating an extensive scRNA-seq dataset of murine and human epicardial cells and an integrated scRNA-seq dataset profiling murine epicardial EMT, we reinforced MAF's human relevance and its importance for mesothelial identity, noting its repression as epicardial cells undergo EMT. By pinpointing pan-mesothelial regulatory mechanisms, our research opens new

avenues for therapeutic strategies leveraging the coelomic mesothelium's regenerative capacity.

Results

Single-cell genomics for benchmarking mesothelial cells across embryonic organs and reappraisal of the role of cadherins in EMT

To understand the GRNs governing mesothelial cell identity, we profiled gene expression and chromatin accessibility in the mouse heart, lung, and pancreas at E13.5, a stage of active EMT in all three organ linings (Figure 1A) [4, 7, 24]. For the epicardium, we used our scRNA-seq [14] and bulk ATAC-seq data from FACS sorted cells [25] (Figures 1B, S1A). For lung and pancreas, we examined public scRNA- and scATAC-seq datasets (Figures 1C, 1D, S1B, S1C) [26]. We confirmed *Wt1* and Uroplakin 3B (*Upk3b*) as robust markers for mesothelial clusters in all three organs and data modalities (Figure 1E-G, S1A–S1C).

A major roadblock to understanding mesothelial cell identity and function is the paucity of knowledge of cadherin activity. Cadherins are critical for EMT by forming adherens junctions and mediating cell-cell adhesion, therefore maintaining epithelial identity [27, 28], but their specific roles in the coelomic mesothelium are contested and unclear. The previously implicated role of CDH1 (E-cadherin) in epicardial EMT has been refuted [3, 13]. We examined cadherin expression and found that only N-cadherin (*Cdh2*), P-cadherin (*Cdh3*), and cadherin-11 (*Cdh11*) were notably expressed across all mouse organ mesothelia (Figure S1D). Notably, E-cadherin (*Cdh1*) is absent from all organ mesothelia at E13.5, consistent with other epicardial studies [3], suggesting it is not a primary target of EMT-inducing transcription factors in the mesothelium.

To further delineate cadherins' roles in EMT, we devised a computational approach leveraging single-cell genomics and lineage tracing to reconstruct the mesothelial EMT trajectory *in silico*, focusing on the heart due to data availability (see Methods, Figure S1E). We integrated our published data with other datasets, creating a 7,332-cell dataset spanning E12.5 to E17.5 (Figure S1F). This integrated dataset includes epicardial cells, their mesenchymal progeny, and differentiated cells (Figure S1F, S1G). Pseudotime scoring predicted branched pathways with distinct fates (epicardial maturation, or differentiation into smooth muscle cell progenitor, cardiac fibroblasts, pericytes, or valvular interstitial cells). The integrated dataset confirms *Cdh1* is absent across the epicardial lineage (Figure S1H). *Cdh2* expression positively correlates with the EMT trajectory (Figure S1H, S1I), supporting its requirement for epicardial migration into the myocardium – a foundational step for EMT [29].

Conversely, the highly expressed cadherins *Cdh3* and *Cdh11* are downregulated along the trajectories (Figure S1H, S1J, S1K), suggesting they may act as guardians of epicardial identity.

Organ mesothelia exhibit transcriptional and epigenomic concordance

To reveal the epicardial GRN, we analyzed gene expression and chromatin accessibility (Figure S1E). We used our E13.5 ATAC-seq data [25] and identified high-quality peaks through an iterative filtering approach [30]. We validated our bulk ATAC-seq dataset (dataset 1) against published E13.5 scATAC-seq data (dataset 2) [26], the only other dataset capturing epicardial cells at this stage, to our knowledge. We subsetted the *Upk3b*-expressing epicardial cluster from dataset 2 (Figure 2A) and found a strong correlation between datasets (Figure 2B, 2C), with dataset 1 capturing 86.6% of peaks in dataset 2. The 81,067 peaks unique to dataset 1 likely reflect the higher number of epicardial cells sampled (~10,000-50,000 vs. 649), capturing more chromatin features. We further leveraged H3K4me3 and H3K27ac CUT&RUN-seq data from cultured murine embryonic epicardial cells (MEC1) to identify active promoters and enhancers [25, 31] (Figure S2A, S2B). Many of these regions were accessible in dataset 2 (18058 enhancers and 16514 promoters; Figure 2D) and other organ mesothelia (23704 enhancers and 14550 promoters; Figure S2C, S2D). To further demonstrate data quality, a weighted gene co-expression analysis identified organ-specific mesothelial gene modules (Figure S2E) and GO term analysis identified relevant terms such as “heart morphogenesis” and “lung development” (Figure S2F). All transcriptional modules were of high quality and conserved in the ATAC-seq datasets as determined by the Z-statistics metric [32] (Figure S2G).

We then compared other organ mesothelia. ATAC-seq peak intersection revealed conserved chromatin accessibility, with 54.9% and 56.0% of peaks in lung and pancreas mesothelium, respectively, being accessible in all three (Figure 2E), which we term consensus mesothelial elements. Epigenomic and transcriptomic concordance between organ mesothelia is strong (Figure 2F, 2G). However, the correlation with the epicardium is weaker, with 47.2% of its ATAC-seq peaks being unique, alluding to organ specificity (Figure 2E). A metacell-based differential gene expression analysis identified enriched markers: Mesothelin (*Msln*), *Tbx20*, and *Twist1* for the epicardium; Hedgehog interacting protein (*Hhip*) and *Gli1* for the lung mesothelium; and Insulin growth factor 1 (*Igf1*) and Hox TF family members *Hoxb6* and *Hoxb8* for the pancreas mesothelium (Figure 2H, Supplementary table 1,2). *Msln*'s epicardium-specificity at E13.5 is intriguing, as it's expressed later in other mesothelia [33, 34]. Among these markers, we identified *Pnoc* (Figure 2I), *Pitx2* (Figure 2J), and *Barx1* (Figure

2K) as distinguishing mesothelial cells within their respective organs. The mesothelium-restricted expression of *Pitx2* and *Barx1* is consistent with previous findings [21, 33], and their chromatin-inferred gene activities reliably distinguish the pancreas mesothelium (Figure S2H, S2I). Conversely, *Pnoc* has not been previously described as a lung mesothelium marker.

Inference of mesothelial GRNs highlights transcriptional regulons as putative encoders of mesothelial identity across organs

To understand the regulatory code underpinning mesothelial identity and function, we inferred GRNs using the SCENIC+ protocol, which predicts TF-enhancer-gene (triplet) relationships as regulons [23]. We created a custom motif database by combining databases from SCENIC+ and Yi et al. [35], with the latter included to enhance mouse regulon discovery. For high-confidence predictions, we selected top importance scores for each triplet (Supplementary table 3,4). We cross-validated putative epicardial enhancers using MEC1 H3K27ac CUT&RUN-seq data and corroborated predicted E-P connections with the Activity-By-Contact (ABC) model [36]. We only considered regulatory relationships predicted by both SCENIC+ and the ABC model in the epicardium. For non-cardiac mesothelia lacking H3K27ac data, which can degrade the ABC model's performance, we relied only on SCENIC+.

Contrasting organ mesothelial GRNs, we identified regulons with enriched activity in the epicardium, including TBX20, which exhibits strong epicardial specificity (Figure 3A). In the lung mesothelium, regulons for Fox family members, particularly FOXF1 and FOXP2, were specific (Figure 3B), consistent with their roles in lung morphogenesis [37, 38]. Genome-wide footprinting confirmed differential DNA binding for FOXF1 and FOXP2 between the epicardium and lung mesothelium (Figure 3D). For the pancreas mesothelium, ISL1 and *PITX2* regulons were central, showing cell-type specificity (Figure 3C) and enhanced TF binding compared to the epicardium (Figure 3E). ISL1 is known for its prominent role in pancreas morphogenesis [39, 40], being required for endocrine islet cells differentiation.

We studied the predicted GRNs in greater detail, focusing on highly specific regulons with many target genes. In the epicardium, TBX20 has the largest number of predicted targets. While known to be indispensable for cardiomyocyte biology [41-43], its role in the epicardium is unstudied. We found epicardium-enriched DNA binding activity for TBX20 (Figure 3D, 3E). GO analysis of its target genes revealed enriched terms relevant to cardiac morphogenesis and, notably, "cell-substrate adhesion", suggesting unique epicardial functions (Figure 3F, Supplementary Table 5). A gene comprising this term is Macrophage Colony-Stimulating Factor (*Csf1*), a cytokine that regulates cardiac macrophages [44]. *Csf1* was enriched in the

epicardium (Figure 3G), consistent with its role in yolk-sac derived macrophage recruitment and cardiac regeneration [45-48]. We identified a CRE (chr3:107793606-107794106) with a TBX20 binding site that has enriched accessibility in the epicardium (Figure 3H). Footprinting showed the TBX20 motif is bound only in the epicardium (Figure 3I, 3J). The region is an active enhancer (H3K27ac mark) and was identified by the ABC model as a *Csf1* regulator (Figure 3H). We thus conclude that TBX20 may be a regulator of *Csf1*.

Using similar strategies, we found other examples of TBX20-driven, epicardium-specific regulation of *Fgf2*, *Sema3d,* and *Rarres2*, genes critical for epicardial function (Figure S3A-I, Supplemental table 3) [15, 25, 49]. TBX20 may also contribute to epicardial EMT through genes regulating cell adhesion. We identified TBX20-mediated regulation of plakophilin-2 (*Pkp2*), a desmosome component expressed mostly only in the epicardium (Figure S3J), potentially reflecting the unique mechanical stress experienced by cardiac tissue. We found two putative epicardium-accessible enhancers (chr16:16219711-16220211 and chr16:256606-16257106) bound by TBX20 linked to *Pkp2* (Figure S3K). Given the role of desmosomes in maintaining cellular adhesion of the mesothelial layer, their disruption may correlate with epicardial EMT. We reason that *Pkp2* expression will change over the course of EMT, and so will the activity of its enhancers. Therefore, we tested the activity of our proposed CREs as an additional strategy to validate their credibility. We generated ATAC-seq data that profiled the *in vivo* EPDCs at E13.5, when EMT peaks. We utilised the inducible *Wt1*^{CreERT2} mouse, crossed with R26R-tdTomato (tdTom) reporter, to trace the epicardial lineage as previously described [14]. Due to their negative surface expression of podoplanin (PDPN) [25, 50], EPDCs were sorted as CD31-tdTom+PDPN- cells for bulk ATAC-seq, and systematically compared against E13.5 epicardial ATAC-seq data [25] . Both enhancers had significantly lower chromatin accessibility in EPDCs (enhancer in S3L: Log2FC -1.15 and FDR 2.23*10-5; enhancer in S3M: Log2FC -2.10 and FDR 1.34*10-6) (Supplementary table 6), aligning with the downregulation of *Pkp2* during EMT (Figure S3N).

TBX20 may also regulate epicardial layer integrity via adherens junctions, which in turn involves cadherins. *Cdh2* (N-cadherin) expression increases during EMT (Figure S1I). We propose this is enabled by TBX20 regulation through a distal enhancer (Figure S3O, Supplemental table 3) that exhibits slightly increased (17% higher) accessibility in EPDCs (Supplementary table 6). Collectively, TBX20 may direct transcriptional programs that provide cardiac-specific context to the epicardial GRN and are linked to EMT.

Master transcriptional regulators may control lung and pancreas mesothelial marker genes via distal enhancers

We then studied prominent regulons in the lung mesothelium. One of the top lung-specific regulons *GLI1*, the Hedgehog signaling effector indispensable for mesothelial cell invasion [4] showed enriched expression in lung mesenchyme (Figure S4A). Our analysis suggests *Hhip* is a target gene of GLI1 (Figure S4B), confirming its known dependence on GLI1 [51].

Another key lung-specific regulon is FOXF1 (Figure 3B), which is essential for lung morphogenesis [37] and not expressed in non-lung mesothelia (Figure S4C). Our findings indicate FOXF1 may regulate *Pnoc*, a lung mesothelium marker (Figure 2I), via a lung-specific CRE (Figure S4D). Another putative target is FOXP2, essential for lung epithelial development [38]. We found *Foxp2* is also highly expressed in the lung mesothelium (Figure S4E), possibly modulated by FOXF1 binding to its promoter (Figure S4F). FOXF1 may also regulate *Gli1* itself via a distal CRE and TBX2, which contributes majorly to lung development and mesothelial lineage [52-54], via a lung-specific CRE (Figure S4G, S4H).

In the pancreas mesothelium, we identified key regulons that may drive marker gene expression (Figure 2H). A prominent example is BARX1, which may regulate *Hoxb8* via a promoter-proximal CRE (Figure S5A). BARX1-deficient mice have spleen defects and abnormal pancreas location, suggesting a role for BARX1 in pancreatic morphogenesis [55]. Another key regulon is *PITX2*, known for left-right patterning function [56] but predominantly expressed in the E13.5 mesothelium (Figure 2J). We found a distal region predicted to regulate *Igf1* via PITX2 (Figure S5B). IGF1 plays a significant role in exocrine pancreas development by acting through its receptor IGF1R [57]. Another PITX2 target is the Cadherin 6 (*Cdh6*) promoter (Figure S5C), uniquely expressed in the pancreas mesothelium (Figure S1D). CDH6, although unstudied in the pancreas, may be critical for maintaining epithelial identity, judging by its involvement in forming the kidney epithelium [58] through formation of adherens junctions [59]. Lastly, the ISL1 regulon is predicted to regulate *Gcg*, a gluconeogenesis hormone, via a CRE with three ISL1 binding sites (Figure S5D, S5E).

Spatiotemporal gene expression of the coelomic mesothelium is conserved and is linked to the activities of *cis*-regulatory elements

We have so far described organ specifications at a single time point. However, expression can be temporal, as highlighted by *Msln*, which is expressed in the epicardium at E13.5 but appears later in other mesothelia [33, 60]. *Msln* is an important marker distinguishing mesothelial cells from other lineages across organs [34, 61], over lifetime [34], and holds significant roles in mesothelioma biology [62-64]. Using *Msln* as a case study, we elucidated principles of temporal gene regulation in the mesothelium.

We scanned the *Msln* promoter for TF motifs and found candidates including RBPJ, SOX proteins, TCF proteins, and AP-1 family members such as NRF2/NFE2l2 (Figure S6A). The promoter is accessible only in the epicardium (Figure 4A, 4B). However, two proximal CREs, marked by H3K27ac, are also linked to the *Msln* promoter (Figure 4B). Both CREs contain WT1 binding sites (Figures 4C, 4D, S6B), suggesting WT1 is a direct regulator. Since *Msln* is majorly downregulated during EMT (Figure S6C), we used our EPDC ATAC-seq to show these CREs are also inactive in mesenchymal derivatives (Figure S6D).

To test if these CREs regulate *Msln* expression over time, we generated bulk ATAC-seq datasets of the *in vivo* epicardium at E11.5 and E17.5. *Msln* expression increased as the epicardium matures (Figure 4E), consistent with protein levels [65]. Mirroring this, chromatin accessibility of CREs increased with development, peaking at E17.5 (Figure 4G, Supplementary table 9, 10). SCENIC+ analysis of our temporal data identified NRF2 as the highest confidence candidate TF (Supplementary table 11). NRF2 has binding sites in the *Msln* promoter and one enhancer (Figure 4A, 4D), and its expression (*Nfe2l2*) increases throughout development, correlating with *Msln* (Figure 4F). We now label these regions as NRF2-linked regions (Figure 4B, Figure S6D).

We tested if these NRF2-linked regions also regulate *Msln*'s temporal expression in other organs. In the pancreas, *Msln* and *Nfe2l2* expression becomes prominent from E17.5 (Figure 4H, 4I). Consistently, ATAC-seq shows NRF2-linked regions have elevated accessibility at E17.5 (Figure 4J). This pattern of accessibility and expression is conserved in the lung mesothelium (Figure 4K, 4L). Here, for the lung, we used postnatal day 3 (P3) ATAC-seq data due to a lack of high-quality data capturing the mesothelium in late embryonic development (between E13.5 and P3). This suggests a conserved mechanism of temporal gene expression, albeit with variation in developmental timing.

Our temporal data also reinforced other CRE predictions we have made. In the epicardium, the expression of TBX20 and its targets *Fgf2* and *Csf1* decreases from E13.5 to E17.5 (Figure S6E), as does the accessibility of their nominated CREs (comparing E17.5 against E13.5: *Csf1* CRE: Log2FC -0.64, FDR $5.39*10^{-14}$; *Fgf2* CRE: Log2FC -0.81, FDR $1.15*10^{-10}$; Figure S6F, S6G). In the pancreas, expression of BARX1 and its target gene correlates over time (Figure S6H), mirrored by the accessibility of the *Hoxb8*-linked CRE (Figure S6I). A similar correlation was seen for ISL1/*Gcg* (Figure S6J, S6K). Collectively, these examples illustrate GRN-dependent temporal gene regulation in the mesothelium.

Comparison against other cardiac cell types reveals mesothelial-specific *cis*-regulatory grammar and regulons

We next explored the regulatory dynamics that distinguish the coelomic mesothelium from other cell lineages, which may reveal master regulators of mesothelial identity. Using the heart as an example, we selected ventricular cardiomyocytes (CM) and cardiac vascular endothelial cells (EC) for comparison (Figure 5A). Using markers *Myl2* and *Fabp4*, we obtained transcriptomic and chromatin profiling of E13.5 CM and vascular EC (Figures 5B, S7A, S7B). As expected, epicardial cells are more similar to other organ mesothelia than to other cardiac cell types, based on chromatin accessibility and transcriptomics (Figure S7C-E). Combined analysis with cardiac lineages reveals enriched expression of previous markers that distinguish organ mesothelia (Figure S7F).

GRN inference highlighted cell-type-specific regulons, including MAF and TCF21 for the epicardium (Figure 5C), ESRRG and TBX5 for CMs (Figure 5D), and SOX17 and SOX18 for vascular ECs (Figure 5E). These TFs are known regulators of their respective cell types, with ESRRG and TBX5 being required for CM maturation, while SOX17 and SOX18 are regulators of the endothelial/hematopoietic lineage commitment [66-69]. Differential footprinting analysis using curated TF motifs (Figure S7G) confirmed their cell-type-specific binding and the epicardium-enriched binding of MAF and TCF21 (Figure S7H, S7I). Examining these regulons, we identified validated TF-gene relationships, such as ESRRG regulating *Mdh1* in CMs [70] (Figure 5F) and SOX18 regulating *Cldn5* in vascular ECs [71] (Figure 5G). The strength of our inference, demonstrated in other cell types, reinforces our subsequent findings on candidate TFs that distinguish mesothelial cells from other cardiac lineages.

Evolutionarily conserved *cis*-regulatory elements drive mesothelial gene expression patterns shared across organs

Based on the cardiac lineage comparison, we studied the regulatory targets of the epicardial-specific regulons MAF and TCF21. TCF21 is highly expressed in the epicardium and required for EPDC commitment to a fibroblast fate [72, 73]. MAF, on the other hand, has not been studied in coelomic mesothelial cells. Both TFs have limited expression in CMs and CECs (Figure S8A). MAF, in particular, is expressed in both cardiac and non-cardiac mesothelia.

MAF's regulon revealed its control of cytokeratins (Krt), which form intermediate filaments that facilitate epithelial cell adhesion by interacting with desmosomes [74]. Several cytokeratins, including *Krt7*, *Krt8*, and *Krt18,* are epicardial markers [61, 75]. We provide evidence that MAF

regulates *Krt8* and *Krt18* via putative enhancers conserved across multiple organ mesothelia. *Krt18* showed pan-mesothelial expression (Figure 6A), and we identified a shared accessible enhancer (chr5:102016032-102016532) linked to its promoter (Figure 6B). As cytokeratins are critical for intercellular adhesion and thus may be inactive after EMT, we confirmed this CRE has significantly lower accessibility (Log2FC: -2.04, FDR: 4.67*10-6) in EPDCs (Figure 6C). A similar logic was applied to *Krt8*, another pan-mesothelial cytokeratin (Figure 6D), which has a conserved CRE (chr15:102005234-102005734) (Figure 6E) with reduced accessibility in EPDCs (Log2FC: -1.79, FDR: 5.88*10-7; Figure 6F). The reduced CRE activity correlates with the downregulation of *Krt8*, *Krt18*, and *Maf* during epicardial EMT (Figure 6G), reinforcing MAF as a regulator of cytokeratins and mesothelial identity.

We also documented conserved regulation of *Krt7*, potentially via a TCF21-linked enhancer (Figure 6H, 6I) that is less accessible in EPDCs (Log2FC: -2.30, FDR: 7.68*10-7; Figure 6J), mirroring *Krt7*'s downregulation during EMT (Figure 6K). While TCF21 could regulate this CRE in the heart mesothelium, its limited expression elsewhere (Figure S8A) suggests that other TFs and CREs are involved. The regulation of *Upk3b*, a robust mesothelial marker (Figure 6L), exemplifies this tissue-specific control. In the epicardium, *Upk3b* is likely regulated by an enhancer linked to the TF TCF21 (Figure 6M, 6N). In other tissues, however, its expression may be controlled by different mechanisms, such as other TFs acting on the same enhancer, or by the pan-mesothelial TF WT1 binding directly to the *Upk3b* promoter (Figure 6O). Both *Upk3b*-linked elements showed significantly decreased accessibility in EPDCs (Figure 6N, 6O), correlating with *Upk3b*'s downregulation during EMT (Figure 6P).

These CREs also appear to drive temporal expression. *Maf* and its target genes *Krt8* and *Krt18* all have increased expression as the epicardium matures (Figure S8B), as does the accessibility of their respective CREs (Figure S8C, S8D). On the other hand, *Upk3b* expression shows limited temporal variation (Figure S8B). This finding is consistent with the accessibility of *Upk3b*-linked elements (Figure S8E), closely mirroring expression and reinforcing its regulatory role. In the pancreas, the *Krt18*-linked enhancer's accessibility mirrors the gene's complex expression pattern (downregulation by E14.5 followed by upregulation by E17.5), and *Maf* expression generally correlates with *Krt8* and *Krt18* (Figure S8F, S8G, S8H). For *Krt8*, a contribution from other regulatory elements, such as the *Krt8* promoter, is substantial, as *Krt8* promoter accessibility temporally correlated more strongly with gene expression (Figure S8H). Unlike in the epicardium, *Upk3b* expression in the pancreas mesothelium increases over time (Figure S8F), which correlates with the accessibility values of both *Upk3b*-linked elements (Figure S8I). Similar temporal correlations between CRE accessibility and gene expression for *Krt8*, *Krt18*, *Maf*, and *Upk3b*

were also observed in the lung (Figure S8L-O). Overall, our data suggest MAF governs cytokeratin expression and mesothelial identity in multiple mesothelia.

Cross-species comparison supports the role of MAF and enables a re-evaluation of epicardial markers

To test if these regulators, which may modulate gene expression in both epicardium-specific and pan-mesothelial manners in mouse, play similar roles in humans, through multiple rounds of single-cell integration (Figure 7A), we generated a cross-species integrated dataset of mouse (E10.5-P6) and human (PCW 6-22) epicardial cells. After benchmarking integration methods (Supplemental Table 12), we used scVI [76] to harmonize the datasets, yielding an integrated scRNA-seq dataset of 5,784 epicardial cells (Figures 7B, S9A, S9B).

Clustering, gene marker identification, and GO pathway enrichment identified states for non-proliferative (State 1), proliferative (State 2; marked by MKI67), lubrication (State 3, marked by PRG4 and MUC16), and epicardial adipose tissue (State 4, marked by ITLN1) (Figure S9C, S9D). We also captured a migratory cell state (State 6), similar to one described in a human fetal heart atlas [77], which highly expresses POSTN and SEMA3D, indicative of migration and paracrine control of cardiac innervation, roles associated with EPDCs [78-80] (Figure 7C). This state is likely resolvable due to the multi-layered nature of the human ventricular epicardium [49, 50]. We hypothesise this structural difference enables the resolution of very early EMT stages in human tissue, as cells retain some epicardial characteristics during their prolonged migration. In mouse, the rapid transition from epicardium to sub-epicardium likely obscures these initial phenotypic changes.

Using CytoTrace [81] to infer developmental potential (Figure 7D), we uncovered a trajectory from proliferating epicardium to an intermediate state (State 5) and then to the migratory subpopulation (Figure 7E). Along this trajectory, MAF, TCF21, and their targets (KRT8, KRT18, *KRT7*, and UPK3B) were downregulated as cells lost their mesothelial identity, evidenced by the loss of MSLN expression, and supporting their roles in maintaining mesothelial identity (Figure 7F).

The cross-species dataset confirmed that our inferences on cadherins are conserved. Analyzing mouse epicardial cells from E11.5 to E13.5, and human cells at approximately equivalent developmental stages (post-conception week 6 to 7), CDH2, CDH3, and CDH11 are expressed in both human and mouse epicardium (Figure 7G). Importantly, E-cadherin/CDH1 shows species-specific expression: it is expressed in human epicardial cells (albeit at a low level compared to other cadherins) but is absent in the mouse epicardium

from E10.5 to P6 (Figure 7G, S9E). Using our integrated murine heart atlas (Figure S9F), we confirmed *Cdh1* expression is absent from the entire mouse heart (Figure S9G), consistent with Figure S1H and in line with another mouse heart atlas [82] (Figure S9H). CDH1 is likely silenced by H3K27me3 modification near its promoter (Figure S9I), a process potentially driven by PRC2 binding. PRC2 binding to the *Cdh1* promoter has been demonstrated in mouse E14.5 ventricular apex [83]. PRC2-mediated heterochromatin deposition may contribute to cross-species variation, offering an avenue to improve the translatability of non-human models. Altogether, our cross-species dataset verified the human relevance of our findings and represents a valuable resource for cardiac biology.

Discussion

The coelomic mesothelium's role in visceral morphogenesis underscores its potential for organ regeneration. Mesothelial dysfunction contributes to congenital diseases [77, 84-87], fibrosis [88-90] , cardiomyopathies [91], and mesothelioma [92, 93]. Despite its importance, the gene regulatory logic governing mesothelial identity remains elusive. Decoding these GRNs would improve understanding of this layer's therapeutic importance and provide a basis for reactivating the quiescent adult mesothelium for regeneration [94]. We present a cross-organ comparison to examine conserved and organ-specific mesothelial roles. We find *Upk3b* robustly marks the mesothelial compartment, suggesting it could be used for lineage tracing to overcome the non-specific expression of *Wt1* in coronary endothelial cells [95]. We further inferred mechanisms behind *Upk3b* regulation mediated by a series of CREs and TFs, including WT1.

Our analysis shows each organ's mesothelium has a distinct gene expression and regulation landscape. We identified enriched markers that distinguish mesothelia from each other and from other cell types in their organ (e.g., *Pnoc* for lung, *Pitx2* and *Barx1* for pancreas). Our initial finding of *Msln* being exclusive to the E13.5 epicardium reflects the earlier morphogenesis of the heart; the gene is expressed later in lung and pancreas mesothelia. We propose *Msln*'s spatiotemporal expression is regulated by a conserved set of CREs, potentially driven by common TFs such as WT1 and NRF2.

Our investigation of the lung mesothelium revealed new regulatory complexity. We identified FOXF1 as a potential upstream regulator of the Hedgehog signal transducer *GLI1*, linking it to lung development. This is supported by similar phenotypes (lung hypoplasia) from *Foxf1* haploinsufficiency, mesenchyme-specific *Foxf1* inactivation, and disruptions to Hedgehog signaling [37, 96, 97]. These phenotypes may result from disrupted mesothelial EMT. As FOXF1

is also a known Hedgehog target, our findings suggest a potential bidirectional interplay between FOXF1 and the Hedgehog signaling pathway.

Our analysis also unveiled new insights into the developing pancreas's genomic landscape. We propose TFs such as PITX2, ISL1, and BARX1 mediate mesothelial function via CREs whose activity correlates with temporal gene expression. In particular, PITX2 may be critical for cell layer integrity via *Cdh6* and may influence exocrine pancreas development through its proposed regulation of *Igf1* [57]. BARX1's potential regulation of *Hoxb8* may have implications for pancreas organogenesis and warrants further exploration, as other HOX family members such as HOXB6 are involved in endocrine cell differentiation [98].

A comparative analysis also identified key factors contributing to epicardial function and identity. We identified the TBX20 regulon, which may control genes with critical roles in the embryonic epicardium, including *Csf1, Sema3d, Fgf2*, and *Rarres2*. Through these genes, TBX20 may influence epicardial behaviours such as establishing immune niches, axon pathfinding, and EMT, which should be investigated further. We also proposed that TBX20 regulates plakophilin-2 (*Pkp2*), a key desmosomal component that protects tissues under mechanical stress and may safeguard epicardial identity. *Pkp2* expression may therefore negatively correlate with EMT, as its loss can increase cell migration [99].

Besides desmosomes, epicardial cells rely on adherens junctions formed by cadherins like *Cdh2, Cdh3*, and *Cdh11* to maintain intercellular junctions and layer integrity. Through an integrated, cross-species analysis, we proposed a negative role for *Cdh3* and *Cdh11* in EMT and confirmed their expression in the human epicardium. Our *in silico* approach also clarified the contested role of *Cdh1* (E-cadherin). While present in the human epicardium at low levels, its transcription in the murine heart is negligible, likely due to epigenetic silencing. Using CDH1 as the case study, we suggest that divergent gene expression patterns across species may arise due to adaptive evolutionary genomics. Nevertheless, our analysis highlights the cross-species similarity between the mouse and human embryonic epicardium, evident by transcriptomic state conservation and similar cadherin expression.

Mesothelial biology may be conserved not only across species but also organs. By investigating the conserved regulatory architecture of the coelomic mesothelium, we identified MAF as a potential gatekeeper of mesothelial identity across organs. We showed a negative association between MAF and EMT in both the mouse and human epicardium. We then confirmed the conserved nature of mesothelial enhancers that may be regulated by MAF. Through these enhancers, MAF may modulate cytokeratin genes such as *Krt8* and *Krt18*, which form part of the keratin cytoskeleton connected to desmosomes [100]. This finding

is consistent with previous observations that identify MAF as a driver of epidermal progenitor differentiation into keratinocytes by regulating keratin gene expression [101].

Overall, our work provides the first comprehensive overview of the gene regulatory landscape of the coelomic mesothelium. By leveraging cell-type-specific and cross-species genomics, we provide initial insights into organ-specific and conserved mesothelial regulatory programs. We identify high-confidence enhancers that can be manipulated to gain a deeper understanding of mesothelial biology, laying the groundwork for future studies on the mesothelium's role in disease and regeneration.

Limitations of the study

Our study relies on computational predictions of E-P connections from transcriptomic and epigenomic data, which require experimental validation. Data availability was a limitation, particularly the lack of chromatin conformation data (e.g., Hi-C) for mesothelial cells to confirm E-P contacts. Our histone CUT&RUN-seq data were from cultured epicardial cells, which may differ from the *in vivo* epigenome. While we identified plausible enhancers for each gene, other regulatory elements may contribute to gene expression. Our cross-species inferences were limited by the current lack of human epicardial ATAC-seq data and limited RNA-seq data from early, EMT-relevant developmental stages. A more comprehensive cross-species comparison will require techniques such as Hi-C and cell-type-specific mapping of PRC2 binding.

Acknowledgements

The Dunn School and Ludwig Institute for Cancer Research Flow Cytometry facility for advice and assistance, and Biomedical Services staff for animal husbandry. This work was funded by the British Heart Foundation (BHF) Intermediate Basic Science Research Fellowship (JMV: FS/19/31/34158), BHF Ian Fleming Senior Basic Science Research Fellowship (NS: FS/19/32/34376) and Oxford BHF Centre of Research Excellence awards to JMV and ANR, and to NS and ANR (both RE/18/3/34214). QD is funded by a doctoral training scholarship from Vinmec International Healthcare System.

Author Contributions

Conceptualisation, QD, ANR, JMV; Methodology, QD, ANR; Investigation, QD, ANR; Formal Analysis, QD, ANR; Visualisation, QD; Writing – Original Draft, QD; Writing – Review and Editing, QD, ANR, NS, JNV; Funding Acquisition, ANR, NS, JNV; Supervision, ANR, JNV, NS.

Declaration of interests

The authors declare no competing interests.

Figure titles and legends

Figure 1 Multi-omic profiling of the embryonic mesothelia across organs

(A) Schematic illustrating cellular contributions of the coelomic mesothelium to the embryonic heart (1), lung (2), and pancreas (3) via epithelial-to-mesenchymal transition (EMT). Created with Biorender.com

(B) Uniform manifold approximation and projection (UMAP) of the E13.5 mouse heart (7,867 cells) based on gene expression. Black arrow indicates the epicardium cluster.

(C) UMAP of the E13.5 mouse lung (5,925 cells).

(D) UMAP of the E13.5 mouse pancreas (7,113 cells).

(E) Dot-plot of marker gene expression in the E13.5 heart. The mesothelium was identified by *Upk3b* and *Wt1* expression. Dot size represents the fraction of cells expressing each gene.

(F) Dot-plot of marker gene expression in the E13.5 lung.

(G) Dot-plot of marker gene expression in the E13.5 pancreas.

AT2: Alveolar-type 2; CM: Cardiomyocyte; EC: Endothelial cell.

See also Figure S1

Figure 2 Transcriptomic and epigenomic concordance of the coelomic mesothelium across organs

(A) UMAP of the E13.5 mouse heart scATAC-seq dataset (7,292 cells). Epicardial cells (649 cells) are highlighted, identified by enriched *Upk3b* gene activity.

(B) Venn diagram of peak intersection from the E13.5 epicardium across two studies. Dataset 1 is from [25]; dataset 2 is from [26].

(C) Correlation matrix for epicardial ATAC-seq from two studies based on accessible peak signatures. Pearson's correlation coefficient shown and.strong correlation is indicated in blue.

(D) Heatmap of normalized peak scores of epicardial ATAC-seq profiles [98] in active epicardial promoter and enhancer regions. Active enhancers and promoters were identified from H3K27ac and H3K4me3 CUT&RUN-seq of the MEC1 epicardial cell line [25].

(E) Venn diagram of ATAC-seq peak intersection from the E13.5 mesothelium across the heart, lung, and pancreas.

(F) Correlation matrix for E13.5 organ mesothelia based on chromatin accessibility signatures.

(G) Correlation matrix for E13.5 organ mesothelia based on transcriptomic signatures.

(H) Matrix dot-plot showing mean expression of enriched marker genes for each E13.5 mesothelium.

(I) Violin plot of normalised *Pnoc* expression in the E13.5 mouse lung.

(J) Violin plot of normalised *Pitx2* expression in the E13.5 mouse pancreas.

(K) Violin plot of normalised *Barx1* expression in the E13.5 mouse pancreas.

See also Figure S2

Figure 3 Mesothelial gene expression is powered by organ-specific gene regulatory networks

(A) Top activator regulons in the heart mesothelium ranked by cell-type specificity. Top 10 specific regulons are highlighted in red.

(B) Top activator regulons in the lung mesothelium ranked by cell-type specificity.

(C) Top activator regulons in the pancreas mesothelium ranked by cell-type specificity.

(D) Volcano plot of differential TF binding activity comparing the heart and lung mesothelium. Differential binding activity against the −log10(p value) of all investigated TF motifs is shown; each dot represents one motif. Only TFs expressed in at least one organ mesothelium are retained. Significantly differential motifs are coloured in red (FC>0.08, $-\log_{10}P > 100$)

(E) Volcano plot of differential TF binding activity comparing the heart and pancreas mesothelium.

(F) GO term over-representation of TBX20 regulon target genes.

(G) Stacked violin plot showing median expression of *Tbx20* and *Csf1* in E13.5 organ mesothelial cells.

(H) Genome tracks showing a distal CRE, a putative enhancer (highlighted), regulated by TBX20, and predicted to regulate the *Csf1* promoter. Merged normalised ATAC-seq tracks displayed. Merged histone tracks show MEC1 epicardial H3K27ac (N = 3) and H3K4me3 (N = 4) CPM-normalised scores. The link plot shows predicted enhancer-promoter (E-P) connections in the epicardium; link colour denotes the ABC model score (the confidence of E-P prediction).

(I) Footprints of the *Csf1* enhancer highlighted in (H) across organ mesothelia. The position of TOBIAS predicted TBX20 binding site is highlighted.

(J) The TBX20 mouse motif logo.

See also Figure S3, S4, S5

Figure 4. Spatiotemporal regulation of *Msln* is conserved across organ coelomic mesothelia.

(A) Footprints of the *Msln* promoter. The NRF2/NFE2L2 motif is highlighted.

(B) Genome tracks of the *Msln* locus. The promoter and two distal CREs are highlighted.

(C) Footprints of the CRE (chr17:25755800-25756200) linked to the *Msln* promoter. WT1 motifs are shown.

(D) Footprints of the CRE (chr17:25756700-25757300). WT1 and NRF2/NFE2L2 motifs are shown.

(E) Violin plot of normalised *Msln* expression in the mouse epicardium at E11.5, E13.5, and E17.5.

(F) Violin plot of normalised *Nfe2l2* expression in the mouse epicardium at E11.5, E13.5, and E17.5.

(G) Genome tracks of the *Msln* locus in epicardial cells at E11.5, E13.5, and E17.5. The promoter is highlighted. Merged normalised ATAC-seq tracks displayed (N = 3). The link plot shows ABC model predicted enhancer-promoter (E-P) connections from all three stages.

(H) Violin plot of normalised *Msln* expression in the mouse pancreas mesothelium at E12.5, E13.5, E14.5 and E17.5.

(I) Violin plot of normalised *Nfe2l2* expression in the mouse pancreas mesothelium at E12.5, E13.5, E14.5 andE17.5.

(J) Genome tracks of the *Msln* locus in pancreas mesothelial cells at E13.5, E14.5, and E17.5. NRF2-linked regions are highlighted. All tracks were pseudo-bulked and TF-IDF-normalised.

(K) Violin plot of normalised *Msln* expression in the mouse lung mesothelium at E12.5, E13.5, E17.5, and P3.

(L) Violin plot of normalised *Nfe2l2* expression in the mouse lung mesothelium at E12.5, E13.5, E17.5, and P3.

(M) Genome tracks of the *Msln* locus in lung mesothelial cells at E13.5 and P3. NRF2-linked regions are highlighted.

See also Figure S6

Figure 5 Comparison with other cardiac cell types reveal mesothelial-specific gene regulation landscape

(A) UMAP of E13.5 mouse cardiac cell types (3,700 cells).

(B) UMAP feature plots of cardiac cell type markers: *Myl2* (CM), *Upk3b* (heart mesothelium), Fabp4 (vascular EC).

(C) Top activator regulons in the epicardium ranked by cell-type specificity. Top 10 specific regulons are highlighted in red.

(D) Top activator regulons in CM ranked by cell-type specificity.

(E) Top activator regulons in vascular EC ranked by cell-type specificity.

(F) Genome tracks showing an ESRRG-regulated CRE predicted to regulate *Mdh1*. The violin plot shows normalised *Mdh1* expression.

(G) Genome tracks showing a SOX18-regulated CRE predicted to regulate *Cldn5*. The violin plot shows normalised *Cldn5* expression.

CM: Ventricular cardiomyocytes; EC: Endothelial cells

See also Figure S7

Figure 6 Conserved CREs regulate mesothelial gene expression

(A) Violin plot of normalised *Krt18* expression in mesothelia, CM, and Vascular EC.

(B) Genome track of the *Krt18* locus showing a putative enhancer (highlighted).

(C) Normalised ATAC-seq signals of the CRE from (B) in FACS sorted epicardial cells and EPDCs. ATAC-seq read counts were normalised using the trimmed mean of M values (TMM) method. The MAF motif is shown.

(D) Violin plot of normalised *Krt8* expression in mesothelia, CM, and CECs.

(E) Genome track of the *Krt8* locus showing a putative enhancer (highlighted).

(F) Normalised ATAC-seq signals of the CRE from (E) in FACS sorted epicardial cells and EPDCs. Position of MAF binding site in the enhancer is shown.

(G) Imputed expression of *Maf*, *Krt8*, and *Krt18* along the epicardial EMT pseudotime trajectory. Three trajectories are shown, associated with terminal fates: Pericytes, Cardiac Fibroblasts (CFb), and valvular interstitial cells (VICs). Plot sections are coloured to represent distinct phases along the EMT and differentiation trajectory, progressing from Heart Mesothelium/Epicardium to EPDC, and subsequently to terminal fates. Vertical lines indicate the boundary pseudotime values demarcating these trajectory phases. MAGIC imputed expression values plotted.

(H) Violin plot of normalised *Krt7* expression in mesothelia, CM, and CECs.

(I) Genome tracks of the *Krt7* locus showing a putative enhancer (highlighted).

(J) Normalised ATAC-seq signals of the CRE from (I) in FACS sorted epicardial cells and EPDCs. Positions of TCF21 binding sites in the enhancer are shown.

(K) Imputed expression of *Krt7* along the epicardial EMT and differentiation pseudotime trajectory.

(L) Violin plot of normalised *Upk3b* expression in mesothelia, CM, and CECs.

(M) Genome tracks of the *Upk3b* locus showing a putative enhancer and promoter (highlighted).

(N) Normalised ATAC-seq signals of the CRE from (M) in FACS sorted epicardial cells and EPDCs. Positions of TCF21 binding sites in the enhancer are shown.

(O) Normalised ATAC-seq signals of the promoter from (M) in FACS sorted epicardial cells and EPDCs. The WT1 motif is shown.

(P) Imputed expression of *Upk3b* along the epicardial EMT and differentiation pseudotime trajectory.

CM: ventricular cardiomyocytes, EC: endothelial cells, CFb: Cardiac fibroblast, EPDC: epicardium-derived cell, Heart meso: Heart mesothelium, VIC: Valvular interstitial cell

See also Figure S8

Figure 7 Human and mouse cross-species comparison of transcriptional regulators and gene markers in the epicardium

(A) Schematic of the cross-species comparison methodology.

(B) UMAP of the integrated human-mouse embryonic/fetal epicardial cells, showing transcriptional cell states.

(C) Stacked violin plot of median SEMA3D and POSTN expression across cell states.

(D) UMAP showing CytoTrace pseudotime of epicardial cells.

(E) CellRank transition matrix UMAP showing predicted trajectories between cell states.

(F) Imputed expression of MAF, TCF21, and their target genes along the inferred epicardial EMT pseudotime trajectory. A trajectory associated with epicardial cells initiating EMT and losing mesothelial features is shown. Plot sections are coloured to represent distinct phases along the trajectory, progressing from State 2 to State 5 and State 6. Vertical lines indicate the boundary pseudotime values demarcating these trajectory phases. MAGIC imputed expression values plotted.

(G) Stacked violin plots of median cadherin expression in mouse (E11.5-E13.5) and human (PCW 6-7) epicardial cells.

See also Figure S9

STAR Methods

Key resources table

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
APC/Cy7 anti-mouse Podoplanin (1:100, FC)	Biolegend	Cat# 127418 RRID:AB_2629804
Brilliant Violet 421™ anti-mouse CD31 (1:50, FC)	Biolegend	Cat# 102424 RRID:AB_2650892
SYTOX™ Green Dead Cell Stain (1:50000)	ThermoFisher Scientific/ Invitrogen	Cat# S34860
Critical commercial assays		
Neonatal Heart Dissociation Kit, mouse and rat	Miltenyi Biotec	Cat# 130-098-373
Deposited data - mouse		
scRNA-seq E10.5 heart	[102]	GSE198905
scRNA-seq E11.5 heart	[103]	GSE193746
scRNA-seq E12.5 heart	[104]	CNP0002316
scRNA-seq E12.5 heart (sorted epicardial lineage)	[105]	GSE154715
scRNA-seq E13.5 heart	[106]	GSE205797
scRNA-seq E16.5 heart (sorted epicardial lineage)	[105]	GSE154715
scRNA-seq E17.5 heart	[25]	GSE299996
scRNA-seq P1, P2, P4, P6 hearts	[107]	GSE232466
scRNA-seq heart (multiple embryonic and postnatal timepoints)	[82]	GSE193346
scRNA-seq E13.5 lung, E13.5 pancreas	[26]	CRA003910
scRNA-seq E12.5 pancreas, E14.5 pancreas, E17.5 pancreas	[33]	GSE101099
scRNA-seq E12.5 lung, E15.5 lung, E17.5 lung, and P3 lung	[108]	GSE149563
ATAC-seq E11.5 epicardial cells	This paper	GSE300631
ATAC-seq E13.5 epicardial cells	[25]	GSE299995
ATAC-seq E17.5 epicardial cells	This paper	GSE300631
ATAC-seq E13.5 EPDC	This paper	GSE300631
scATAC-seq E13.5 heart, E13.5 lung, E13.5 pancreas	[26]	CRA003910
scATAC-seq E14.5 pancreas, E17.5 pancreas	[21, 108]	GSE210569
scATAC-seq P3 lung	[108]	GSE149563
Cut&Run-seq MEC1 epicardial	[25]	GSE299998

cell line		
ChIP-seq E13.5 heart	[109]	GSE82469
ChromHMM model E13.5 heart	[110]	
Deposited data - human		
scRNA-seq PC17W, 22W, 23W and 25W hearts	[111]	GSE106118
scRNA-seq 40 to 80-day hearts	[49]	GSE216019
scRNA-seq 42- and 56-day hearts	[112]	GSE181346
scRNA-seq 14W and 15W sorted epicardial cells	[113]	GSE213669
scRNA-seq 10W and 11W hearts	[114]	GSE241128
Experimental models: Organisms/strains		
Mouse: Gt(ROSA)26Sor$^{tm9(CAG-tdTomato)Hze}$/J	The Jackson Laboratory	MGI ID: 3809523 JAX Stock No: 007909 RRID:IMSR_JAX: 007909
Mouse: *Wt1*$^{tm2(cre/ERT2)Wtp}$/J	The Jackson Laboratory	MGI ID: 3801682 JAX Stock No: 010912 RRID:IMSR_JAX: 010912
Mouse: C57BL/6J Wild-types	Charles River	-
Software and algorithms		
Pycistopic v1.0a0	[23]	https://github.com/aertslab/pycisTopic
Pysam v0.22.0	[115]	https://github.com/pysam-developers/pysam
Activity-by-Contact v0.2.2	[36]	https://github.com/broadinstitute/ABC-Enhancer-Gene-Prediction
Cellranger v7.2.0	10x Genomics	https://support.10xgenomics.com/single-cell-gene-expression/software/ RRID:SCR_023221
Cellranger ATAC v2.1.0	10x Genomics	https://github.com/10XGenomics/cellranger-atac
ChIPseeker v1.40.0	[116]	https://bioconductor.org/packages/release/bioc/html/ChIPseeker.html
edgeR v4.2.2	[117]	https://bioconductor.org/packages/devel/bioc/html/edgeR.html
MACS2	[118]	https://github.com/macs3-project/MACS
Samtools v1.18.0	[119]	https://github.com/samtools/samtools
Bedtools v2.31.0	[120]	https://github.com/arq5x/bedtools2
SEACells v0.3.3	[121]	https://github.com/dpeerlab/SEACells
CellRank v2.0.0	[122]	https://github.com/theislab/cellrank
Omicverse v1.5.8	[123]	https://github.com/Starlitnightly/omicverse

SnapATAC2 v2.5.3	[124]	https://github.com/kaizhang/SnapATAC2
DeepTools v3.5.4	[125]	https://github.com/deeptools/deepTools
FeatureCounts v2.0.6	[126]	https://subread.sourceforge.net/featureCounts.html
TOBIAS v0.16.0	[127]	https://github.com/loosolab/TOBIAS
Seqkit	[128]	https://bioinf.shenwei.me/seqkit/
Cluster-Buster	[129]	https://github.com/weng-lab/cluster-buster
R v 4.4.1	The R Foundation	https://www.r-project.org RRID:SCR_001905
RStudio	Posit Software	https://posit.co RRID:SCR_000432
Signac v1.14.0	[130]	https://github.com/stuart-lab/signac
EnhancedVolcano v1.20.0	[131]	https://igv.org/ RRID:SCR_011793
IRanges v2.38.1	[132]	https://github.com/Bioconductor/IRanges
Integrative Genome Viewer v2.19.5	[133]	
Scanpy v1.9.5	[134]	https://github.com/scverse/scanpy
SoupX v1.3.0	[135]	https://github.com/constantAmateur/SoupX
ScDblFinder v3.12	[136]	https://github.com/plger/scDblFinder
Scran v1.30.0	[137]	https://bioconductor.org/packages/release/bioc/html/scran.html
Scry v1.14.0	[138]	https://bioconductor.org/packages/scry.html
Palantir v.1.40	[121, 139]	https://github.com/dpeerlab/Palantir
DESeq2 v1.42.0	[140]	https://bioconductor.org/packages/release/bioc/html/DESeq2.html
Mousipy	[141]	https://github.com/stefanpeidli/mousipy
hdWGCNA v0.4.0	[142]	https://github.com/smorabit/hdWGCNA
ClusterProfiler v4.10.0	[143]	https://bioconductor.org/packages/release/bioc/html/clusterProfiler.html
CellTypist v1.6.1	[144]	https://github.com/Teichlab/celltypist
nf-core/cutandrun v2.0.0	10.5281/zenodo.6624266	https://github.com/nf-core/cutandrun/tree/2.0
nf-core/atacseq v2.0.0	10.5281/zenodo.2634132	https://github.com/nf-core/atacseq
Softwares and algorithms – scRNAseq data integration		
BBKNN	[145]	https://github.com/Teichlab/bbknn
Harmony v0.18	[146]	https://github.com/lilab-bcb/harmony-pytorch

LIGER v0.2.0	[147]	https://github.com/welch-lab/liger
Scanorama v1.7.4	[148]	https://github.com/brianhie/scanorama
scVI v1.0.4	[149]	https://github.com/scverse/scvi-tools
scANVI v1.6.1	[150]	https://github.com/scverse/scvi-tools
scGen v2.1.0	[151]	https://github.com/theislab/scgen

Resource availability

Lead contact

Further information and requests for resources should be directed to the lead contact, Joaquim M. Vieira (joaquim.nunes_vieira@kcl.ac.uk).

Data and code availability

ATAC-seq data have been deposited at GEO as GSE300631 and are publicly available as of the date of publication.

This paper analyzes existing, publicly available data, as detailed in STAR Methods.

Original code and custom pipelines have been deposited at https://github.com/loganminhdang/Mesothelium_paper_2025.

Any additional information required to reanalyse the data reported in this work paper is available from the lead contact upon request.

Method details

Animals

Mouse Strains

Males homozygous for *Gt(ROSA)26Sortm14(CAG-tdTomato)Hze/J* or *Gt(ROSA)26Sortm9(CAG-tdTomato)Hze/J* (*Rosa26tdTomato*), and heterozygous for *Wt1tm2(cre/ERT2)Wtp/J* (*Wt1*$^{CreERT2/+}$) were crossed with *Rosa26tdTomato* or C57BL/6J wild-type females for genetic-lineage tracing studies. Pregnant females were administered 100 mg/kg tamoxifen by oral gavage at E9.5. All procedures were approved by the University of Oxford Animal Welfare and Ethical Review Board, in accordance with Animals (Scientific Procedures) Act 1986 (Home Office, UK).

Fluorescence-activated cell sorting (FACS)

E11.5, E12.5 or E13.5, and E17.5 hearts (*Wt1*$^{CreERT2/+}$*;Rosa26tdTomato*) from littermate embryos were pooled prior to dissociation (E11.5: 6 hearts/sample; E13.5: 9-13 hearts/sample; E17.5: 4 hearts/sample) to maximise cell numbers for FACS and corresponding FMO controls, and achieve a target sort of 50,000 epicardial cells/sample for the downstream ATAC-seq protocol. Single-cell suspensions in cell staining buffer (Biolegend) were pre-treated for 5 minutes on ice with TruStain FcX antibody, followed by surface staining with BV421 anti-CD31 and APC/Cy7 anti-Podoplanin (PDPN) antibody for 30 minutes on ice, then washed with PBS before being incubated with SYTOX Green Dead Cell Stain for 5 minutes on ice. After PBS washes, cells were resuspended in sorting buffer containing 2% FBS in PBS. Primary antibodies and SYTOX stain used are detailed in the Key Resources Table. BD FACSAria III was used to sort 10,000-50,000 live cells negative for SYTOX Green and CD31, and positive for tdTomato and PDPN, into 0.5% BSA in PBS.

ATAC-seq data generation

Using FACS, we isolated live epicardial cells (CD31-tdTomato+PDPN+) from E11.5 and E17.5 hearts, and EPDCs (CD31-tdTomato+PDPN-) from E12.5 and E13.5 hearts (E12.5 and E13.5 hearts were both used to represent the E13.5 stage due to very strong concordance between these two stages). Cells were subjected to the ATAC-seq protocol as previously described [152]. In brief, 10,000-50,000 sorted cells were centrifuged at 500 ×g. for 5 minutes, the supernatant was removed, and the pellet was resuspended in 50µL of ATAC Resuspension Buffer (RSB: 10mM Tris-HCl pH 7.5, 10mM NaCl, 3mM MgCl2) supplemented with 0.1% NP-40, 0.1% Tween-20, and 0.01% Digitonin to extract nuclei. After a 3-minute incubation, 1mL of ATAC RSB with 0.1% Tween-20 was added, mixed by inversion, and centrifuged at 500 ×g for 10 minutes. The supernatant was carefully removed, and nuclei were resuspended in 50µL of transposition mix (100mM tagmentation enzyme, 0.01% Digitonin, 0.1% Tween-20, 0.33X PBS, in TD buffer). The tagmentation enzyme amount was adjusted according to cell count (for 40,000 cells, the enzyme was diluted to a

0.8× concentration in TD buffer before being added to the transposition mix). Incubation and centrifugation steps were performed on ice or at 4°C, respectively, up to this point. Transposition reactions were carried out at 37°C for 30 minutes with shaking at 1,000 RPM using an Eppendorf ThermoMixer.

ATAC-seq DNA libraries were generated as previously described [152], with minor modifications. Reactions were purified using the MinElute Reaction Cleanup Kit (Qiagen) following the manufacturer's instructions. Pre-amplification of transposed fragments (5 cycles) was performed using NEBNext Ultra II Q5 Master Mix (NEB) and Nextera i7 index adapters (IDT). The number of additional amplification cycles (typically 4) was determined using the qPCR method [153]. Final PCR products were purified using a double-sided bead purification method with SPRIselect beads (0.5X volume followed by 1.3X volume of SPRI beads). Final cleanup was performed with the MinElute Reaction Cleanup Kit (Qiagen). Libraries were assessed for quality using an Agilent TapeStation with a D1000 DNA ScreenTape assay, and quantified with the KAPA Library Quantification Kit. Next-generation sequencing (NGS) was conducted on equimolar pooled libraries using the NextSeq 500/550 High Output v2.5 kit (75 cycles) (Illumina), with paired-end reads (40 bp x 2) and 8 bp single index reads.

Bulk ATAC-seq data processing

Briefly, the ATAC-seq nf-core Nextflow pipeline (version 2.0) [154] was used to pre-process ATAC-seq sequencing data, consisting of the following steps: quality control, adapter trimming, alignment to the UCSC mm10 genome, filtering unwanted reads, and merging across biological replicates. Filtering removed: reads aligning to mitochondrial DNA and blacklisted regions [155]: PCR duplicates, multi-mapped reads, and reads that have an insert size > 2kb and contain > 4 mismatches.

Alignment files were converted to bed files using the randsample function of MACS2 (version 2.2.9.1) [118] with 100% of tags kept (-p 100). Peak-calling was performed with MACS2 using the following parameters: -f BEDPE -g mm -q 0.01 --nomodel --shift -73 --extsize 146 --call-summits --cutoff-analysis --keep-dup all.

The resulting peak summits were then extended by 250 bp to both sides to a fixed width of 501 bp. Peaks that extend beyond chromosome ends were filtered. The remaining peaks went through an iterative removal process to filter out less significant peaks that overlap with more significant ones, as described in [30]. For joint analysis of all developmental stages, we merged biological replicates for each stage into a single bam file represent each stage, then

sort and index using Samtools (version 1.18.0) [119]. We then merged the sorted bam files of E11.5, E13.5, and E17.5 samples into a single bam file using the samtools merge function. MACS2 peak calling and processing were performed exactly as the procedure described above. Intersection of peaks between samples was performed using either the bedtools intersect function of Bedtools (version 2.31.0) [120] or the findOverlaps function of the IRanges library (version 2.38.1) [132].

Single-cell RNA-seq data pre-processing

Single-cell RNA-seq datasets were processed via CellRanger (version 7.2.0) using the mm10 reference and underwent quality control and normalisation according to published best practices [156]. Briefly, we filtered out low-quality cells, which are defined as barcodes that: (1) are outside the range of 5 median absolute deviations (MADs) for count depth and genes per barcode; (2) outside of 3 MADs for the fraction of counts from mitochondrial genes per barcode; (3) cells with a percentage of mitochondrial counts exceeding a manual selected threshold. We used a dynamic threshold for percentage of mitochondrial counts due to the observed large proportion of mitochondrial DNA in cardiac cell types [157], ranging from 0 to 8%. For non-cardiac datasets, we used a threshold of 15%. Next, ambient RNA correction and doublet detection were done via soupX (version 1.3.0) [135] and scDblFinder (version 3.12) [136], respectively. Single-cell counts were normalised using the shifted logarithm approach implemented in Scanpy (version 1.9.5) [134]. To select highly variable genes (HVG), for individual datasets, corrected counts were normalized using the Scran's pooling-based size factor estimation method implemented in the scran (version 1.30.0) R library [137]. Scran-normalized counts were then used to select top 4000 HVGs via the R scry (version 1.14.0) library [138]. Dimensionality reduction and clustering were performed using the UMAP [158] and Leiden [159] algorithms, respectively. We calculated the UMAP embedding by finding nearest neighbors in the principal component analysis space.

Single-cell ATAC-seq data processing

Raw sequencing files were processed via the CellRanger-ATAC pipeline (version 2.1.0) using the mm10 reference. Pre-processing was performed using SnapATAC2 (version 2.5.3) [124]. Briefly, a minimum TSS enrichment of 10 and a minimum number of fragments of 4000 was chosen to filter for high-quality cells. After removing features that map to blacklisted regions, we count fragments instead of reads to better preserve regulatory information, as demonstrated in Martens et al. (2024) [160]. Doublet removal was performed prior to dimensionality reduction via spectral embedding and clustering via the Leiden algorithm.

For peak calling, ATAC-seq fragments for each cell type were pseudo-bulked using the function export_pseudobulk of the Pycistopic package (version 1.0a0), based on the UCSC mm10 genome. Peak-calling was performed on the pseudo-bulk profiles, using the following parameters: --format BEDPE --gsize mm --qvalue 0.01 --nomodel --shift 73 --extsize 146 --keep-dup all --call-summits –nolambda. Consensus peaks were inferred using the summit extension and iterative filtering approach used in bulk ATAC-seq data processing.

ATAC-seq data analysis

To generate a list of consensus non-overlapping peaks across tissue mesothelia, the procedure from Corces et al. (2018) [30] was implemented. Briefly, to normalize the differences between samples in terms of read depth and quality, the individual MACS2 501-bp peak score ("-log10(pvalue)") was divided by the sum of all peak scores in each sample and multiplied by 1 million. The peak sets from tissue mesothelia were then merged using Bedtools and iteratively filtered again to retain significant peaks from all samples to generate a master peak set representing producible mesothelial peaks.

BigWig files used to calculate the correlation between organ mesothelia and cardiac cell types were generated using Pycistopic. Average scores for consensus mesothelial peaks between bigWig files were calculated using the multiBigwigSummary function of DeepTools (version 3.5.4). Overall similarity between samples was determined by computing Pearson correlation coefficients based on read coverage in consensus peaks. Sample signals over the set of consensus peaks were calculated using the computeMatrix function of Deeptools [125] using the following parameters: --referencePoint center -b 500 -a 500 –missingDataAsZero and visualised using the plotHeatmap function.

The bigWig files for the epicardium bulk ATAC-seq, including the E13.5 stage, were normalised using the trimmed mean of M values normalisation method to account for differences in the peak landscape (signal-to-noise ratio) between samples. We used FeatureCounts to extract read counts for peaks and create a region-by-sample matrix. We then used this matrix to calculate a scale factor for each sample, which was then used to scale bigWig files accordingly using the –scaleFactor parameter of the bamCoverage tool.

Differential chromatin accessibility analysis

We adapted the procedure from Reske et al. (2020) [161] and conducted differential analysis using three biological replicates for each sample (epicardium: E11.5, E13.5, E17.5; EPDC: E13.5). To select for peaks and normalisation method, we tested four peaking calling and normalisation strategies as recommended by the original report (MACS2 + TMM

normalisation, MACS2 + loess normalisation, csaw *de novo* peak identification by local enrichment + TMM normalisation, and csaw *de novo* peaks + loess normalisation). For MACS2, we used peak-calling parameters and iterative peak filtering strategy as described in the "ATAC-seq data analysis" method section. For identifying *de* novo peaks using csaw, we performed local enrichment by counting BAM reads in 300 bp window and filtering features by local enrichment (threshold of 3-fold increase in enrichment over 2kb window). We found that our MACS2-dependent peak-calling strategy provides comparable performance to the csaw peak identification. TMM normalisation, when used in conjunction with our MACS2 strategy, does not create global normalisation bias.

For differential accessibility analysis, we used edgeR (version 4.2.2) [117]. Using the ATAC-seq count matrices with TMM-normalised factors, dispersions were stabilised by empirical Bayes, followed by quasi-likelihood GLM fitting for hypothesis testing for differentially accessible windows (500bp). Windows were merged, with the most significant windows retained as the statistical representation. FDR threshold for detecting significantly differentially accessible peaks was set at 0.05. Differential peaks were annotated with gene information using the ChIPseeker R library (1.40.0) [116], using the UCSC mm10 database.

Single-cell RNA-seq batch correction and integration

For batch correction and data integration of multiple datasets (mouse embryonic epicardium, mouse embryonic and postnatal lung mesothelium, mouse embryonic pancreas mesothelium), we downsampled large datasets to a maximum of 10,000 cells using the "subsample" function in Scanpy. HVGs were selected in a batch-aware manner using Scanpy by calculate HVGs for each batch separately and select HVGs that are highly variable in the highest number of batches. Using only HVGs, we ran seven selected data integration methods according to default parameters obtained from paper methods and tutorials. For further details on the methods, please see Supplementary Table 12. As some methods, such as scANVI or scGen, required prior cell type annotation, we annotated cells in the concatenated dataset prior to integration. For cardiac samples, we used automated annotation via CellTypist (version 1.6.1) [144] based on cell-type annotation from existing datasets [82] for mouse and [112] for human). Briefly, we normalised raw counts from these datasets to 10000 counts per cell, then transformed via shifted logarithm. Via Celltypist, a logistic regression model was trained using filtered genes as features. We then used this model to annotate cell types in single-cell datasets, followed by careful manual inspection of marker genes to ensure correct annotation. For other datasets, cell type annotation was performed manually by matching cluster-specific genes to established transcriptomic markers.

We performed benchmarking of integration based on the procedure described in Luecken et al. (2022) [162], using the metrics grouped into two broad categories: (1) removal of batch effects and (2) conservation of biological variance. For cross-species integration, as we focused exclusively on the epicardium, we ranked methods based on their ability to correct batch effects.

Metacell-based differential expression analysis

To circumvent issues associated with single-cell pseudo-bulking approaches, we created metacells – artificial aggregates of single cells. The use of metacells to increase statistical rigor has been applied previously [163]. In our case, we generated metacells to create statistical dispersion that enhances the performance of differential analysis frameworks such as DESeq2 [140].

We used SEACells (version 0.3.3) [121] to build transcriptomic metacells for all organ mesothelia. The input is an AnnData object comprising of the entire pre-filtered gene, cell type assignment, and a low-dimensional UMAP representation. To build metacells – coarse aggregate of phenotypically similar cells, we first sampled 10 cells to uniformly cover the phenotypic landscape. We define one metacell per 75 single cells and used a convergence percentage of 1e-5. Following the construction of metacells, we sum the unnormalized soupX-corrected counts of individual cell in each metacell to generate counts of all genes in each metacell. Each metacell was then assigned a cell type label based on which cell type was most prominent amongst individual cell used to construct that metacell. Following the generation of a counts x gene matrix, we treat each metacell as a biological “replicate” and performed differential gene expression analysis using DESeq2 (version 1.42.0). We filtered genes with low count (sum of expression count across all metacells less than 50). Count data was normalized using DESeq2’s built-in median-of-ratios method. Differentially expressed genes between organ mesothelia are considered genes with Bonferroni-Hochberg-adjusted p-value less than 0.05 (to account for multiple comparisons) and Log2FC > 1.

Mouse epicardial lineage integration

To mitigate the confounding influence of inter-platform technical variation on biological inference, we selected only datasets generated with UMI-based technologies (10X Genomics). Datasets from other technologies, such as the full-length Smart-seq2 method, which is known to exhibit higher gene detection sensitivity, are deliberately excluded. This intrinsic difference in transcript recovery between technologies can introduce non-biological

variance upon single-cell integration, thereby confounding downstream analyses such as pseudotime inference.

We performed quality control on individual datasets, which included ambient RNA correction and doublet detection, before concatenating the datasets and following the steps outlined in the section on batch correction and integration methods. We selected scANVI due to its performance in integration metrics. We extracted the latent representation of each cell generated by scANVI and calculated a new UMAP embedding using this corrected representation. Following a round of clustering using Leiden (resolution 0.5), we identified cell-type marker genes using the Wilcoxon rank-sum test implemented in Scanpy. Cell types associated with epicardial EMT, based on marker annotation and UMAP distance, are then clustered out to form a new subsetted dataset. We identified cell types based on established markers: *Upk3b* (epicardium), *Postn* (EPDC), *Rgs5* (Pericytes), *Dpt* and *Col15a1* (CFb/Fb progenitors), *Adamts19*, *Col9a3*, and *Lef1* (VIC), and *Myl9* (SMC progenitor) [164-169].

Further data cleaning was performed to ensure epicardial lineage by only retaining cells with tdTomato reporter expression. Gene expression counts were then re-normalised using the shifted logarithm method. A further round of marker gene identification was then performed to identify cell states.

Cross-species data integration

For each species, we pre-processed individual datasets as outlined earlier and integrated them using the best-performing method according to our benchmarking (scGen for mouse embryonic/postnatal heart and scANVI for human fetal heart). Following integration, we clustered out epicardial cells based on their expression of WT1 and UPK3B. For the human fetal heart, as the epicardial cluster is confounded by the presence of cells expressing high levels of haemoglobin genes, we performed multiple rounds of clustering to remove these cells and minimise noise in the dataset. We then concatenated the clustered-out mouse and human epicardial cells into a single dataset. To harmonise human and mouse gene symbols, we mapped mouse (GRCm39) genes to their corresponding human (GRCh38.p13) orthologs using mousipy [141]. We then normalised raw counts in the combined dataset using the shifted log approach, followed by dimensionality reduction and clustering. 5000 HVGs were selected on the basis of being highly variable in at least one species. Data integration was then performed using HVGs. We only selected methods that are agnostic to cell-type label and found scVI outperformed at correcting species difference (using raw counts, a 30-dimensional latent space and two hidden layers for encoder and decoder neural networks).

We used the cell latent representation generated by the scVI model to compute a new UMAP embedding.

Gene co-expression network inference

We applied hdWGCNA [142] to scRNA-seq dataset consisting of three organ mesothelia to detect network modules of genes whose expression profiles are tightly intertwined. Briefly, we construct metacells based on the low-dimensional embedding of the dataset (aggregating 75 cells for each metacell). We then constructed a gene-gene correlation adjacency matrix to infer co-expression relationships between genes, selecting a soft power threshold of 5 to reduce noise and retain only strong correlations. Modules were projected from the murine mesothelial scRNA-seq reference to murine mesothelial ATAC-seq query using the hdWGCNA function ProjectModules. We assessed module preservation using the Z statistic [32]. Over-representation analyses based on regulon target genes were done via the enrichGO function of R library ClusterProfiler (version 4.10.0) [143], using the GO Biological Process database and with Benjamini-Hochberg (BH)-adjusted p-value and q-value cutoff of 0.01 and 0.05 respectively.

Differentiation potential analysis

For the integrated mouse epicardial lineage dataset, we used Palantir (version 1.4.0) [139]. 20 diffusion components were computed to determine the diffusion map of the data, followed by low-dimensional embedding and imputation. We pre-selected cells at the extreme for diffusion components as initial and terminal cells before computing pseudotime.

For inferring cellular progression across the integrated cross-species dataset, we utilized CytoTRACE [81]. This choice was predicated on CytoTRACE's core mechanism, which primarily leverages transcriptional diversity (i.e., the number of expressed genes per cell) to estimate differentiation potential. We posited that this fundamental cellular property is likely more robustly conserved across species compared to the specific, detailed gene expression programs and precise manifold topologies upon which algorithms like Palantir depend. Indeed, embryonic epicardium is active and proliferative [24], thus may express more genes and satisfy the assumption of CytoTRACE. Furthermore, Palantir's reliance on constructing a continuous low-dimensional manifold from transcriptional similarity can be confounded by inter-species transcriptional divergence. This divergence is prominent, as specific transcriptional states of epicardial cells are enriched in one species. CytoTRACE's approach, being less directly dependent on the integrity of a unified manifold, is anticipated to offer more reliable pseudotime inference. Based on CytoTRACE pseudotime, we used CellRank

(version 2.0.0) to compute a transition matrix and predict the initial cell state [122]. This initial state was used to predict terminal states. Visualisation of the transition matrix is done by projecting it into the cross-species integrated UMAP embedding.

Biological pathway analysis

We used the AUCell algorithm [170] to score the activity of gene sets associated with GO terms (Biological Process 2021 database). Visualisation of epicardial cell state-specific pathways was done using Omicverse (version 1.5.8) [123].

Enhancer-driven gene regulatory network inference

Cis-regulatory topics were identified using Pycistopic (version 1.0a0). Briefly, topic modelling was performed to infer cis-regulatory topics, which we then used to infer differentially accessible regions and gene activity. In the case of bulk ATAC-seq samples, we simulated single cells to generate sparsity to enhance the performance of topic modelling. Using Pysam (version 0.22.0), single cells were simulated by randomly sampling 20000 reads from the pool of bulk ATAC-seq aligned reads and using 2000 reads as an alignment profile for a single cell. Following topic discovery and annotation (topic threshold 0.2), accessibility of regions was imputed (scale factor 10^6), which generated a binarised cell-peak matrix as input. From here, differentially accessible regions were identified between organ mesothelia and cardiac cell types using the Wilcoxon rank-sum test.

For motif enrichment to identify TF cistromes, we first generated two cisTarget databases based on two consensus peak sets, one from mesothelial across organs and another for cardiac cell types. We supplemented this database by including 1636 consensus mouse TR motifs based on Yi et al. (2021) [35]. This expansion creates a more comprehensive set of high-quality murine TF motifs, improving the statistical power for the subsequent motif enrichment and GRN analysis. Briefly, we used the UCSC mm10 reference genome and chromosome size to generate fasta files from consensus regions, adding 1kb of background padding, and subsequently generated cisTarget databases using the “create_cistarget_motif_databases” function.

Combining motif database and processed regions, inference of enhancer-driven Regulons was done using the SCENIC+ method (version 1.0a1) in unpaired multiome mode and “mmusculus” as the species. Pseudogenes and ribosomal genes were excluded from the analysis. We focused exclusively on regulons of the activator archetype, which include TFs that bind to and open the chromatin of target genomic regions to promote the expression of target genes. We did not include repressive elements. Negative associations between TF-

CRE-genes do not allow reliable inference of transcriptional repressors, as they can be confounded by the data sparsity inherent to scATAC-seq. High-quality regulons are filtered according to their triplet ranking and metrics comprising this ranking (high importance_R2G and importance_TF2G).

Regulon specificity scores were computed as described previously [23]. In each regulon, only the top 50 target genes (ranked by triplet score) were considered. Over-representation analyses based on regulon target genes were done via the enrichGO function of R library ClusterProfiler (version 4.10.0) .

Activity-by-Contact analysis

Epicardial enhancer-promoter connections were predicted using the Activity-by-Contact (ABC) model (version 0.2.2). A custom gene and promoter annotation files were compiled by combining NCBI RefSeq annotation for mm10 genome (build 38.1) with the ENSEMBL mouse database (BiomaRt version 2.58.2). Using NCBI RefSeq gene body annotation, we computed gene length and normalised pseudo-bulked epicardial expression profile to create TPM-normalised epicardial expression. Pseudo-bulked expression for each gene was computed as the sum of expression of each epicardial cell in a sample (i.e., E13.5).

Briefly, to infer E-P connections in the E13.5 epicardium, 501-bp candidate enhancer elements were predicted based on ATAC-seq peaks. Enhancer activities within the 150000 strongest elements were quantified using a combination of ATAC-seq, H3K27ac signal, and TPM-normalized gene expression. ABC scores were then computed by combining region activity and Hi-C contact frequency estimated by a power-law function of genomic distance. The powerlaw exponent was set to be 0.87. To flag genes that are impervious to the effect of distal enhancers, a list of ubiquitously expressed genes across mouse tissues was collected from [171]. To infer E-P links across all developmental stages, we used the consensus epicardial alignment file (merged from all three time points) and filtered peaks called from this file to run the ABC model. The threshold for ABC score is 0.02, which corresponds to roughly 70% recall in human benchmarks.

Transcription factor footprint pre-processing and analysis

Calculation of TF footprint scores was done using the TOBIAS pipeline v0.17.0 [127]. Briefly, the pipeline performs Tn5 Transposase insertion bias correction and footprinting score calculation from the bias-corrected cut sites. Footprint scores across tissue mesothelia were compared to identify TF differential binding. Only motifs annotated to TFs that are expressed in at least one tissue mesothelium were considered.

To identify genes that may be subjected to promoter-mediated regulation, we selected genes that are differentially enriched in the epicardium, in addition to having bound footprints of relevant TFs within the 500 bp region surrounding their TSS.

To scan for all TF binding sites and associated footprints in the *Msln* promoter, we used the *submerge* function of TOBIAS.

Epigenomic data analysis

CUT&RUN-seq data was pre-processed as previously described [25]. Briefly, H3K27ac and H3K4me3 Cut&Run-seq data of the MEC1 mouse epicardial cell line were pre-processed using the nf-core NextFlow pipeline (version 2.0). The following steps were performed: quality control, and alignment to both target and spike-in genomes with a minimum q-score of 10. Peaks were normalised against IgG controls, scaling background control by a factor of 0.8. Read counts were normalised using Counts Per Million (CPM), and consensus peaks, representing active promoters and enhancers, were identified and merged from at least three biological replicates.

Individual replicates were merged, sorted, and indexed using Samtools. Merged bam files were converted into BigWig files using the bamCoverage tool of Deeptools, normalised using the CPM method, with bin size 10. The signal distribution of histone modification was calculated using the computeMatrix function of Deeptools with the following parameters: --beforeRegionStartLength 1000 --regionBodyLength 5000 --afterRegionStartLength 1000 –skipZeros (H3K4me3, active promoters) and --beforeRegionStartLength 3000 --regionBodyLength 20000 --afterRegionStartLength 3000 (H3K27ac, active enhancers). Signal is then visualised using the plotProfile function.

We determine consensus enhancer and promoter elements by comparing H3K27ac/ H3K4me3-marked regions and consensus mesothelial ATAC-seq peaks. To visualise the accessibility of consensus mesothelial enhancers and promoters, we compute ATAC-seq peak score for each element using the computeMatrix function with following parameters: reference-point --referencePoint center -b 500 -a 500 --missingDataAsZero and visualise using the plotHeatmap function of Deeptools.

Murine H3K27ac, H3K4me3 and H3K27me3 ChIP-seq datasets of E13.5 whole heart were retrieved from the ENCODE project [109]. Signal p-values were visualised using IGV or Signac where relevant.

We also downloaded the 15-chromatin-state, 8-histone-mark ChromHMM assignments for the mouse E12.5 heart from van de Velde et al. (2021) [110].

Quantification and statistical analysis

Statistics were performed in R (https://cran.r-project.org/; version 4.4.1) and Python. Differential gene expression analysis was performed using DESeq2, and genes that had BH-adjusted p value < 0.05 was considered to be differential. Statistical significance for differential accessible peaks was calculated using Wilcoxon rank test. To generate chromatin tracks, ATAC-seq data was normalized using the term frequency-inverse document frequency (TF-IDF) method and then visualised either using Signac (version 1.14.0) [130] or Integrative Genome Viewer v2.19.5 [133]. Motif logos were visualised via the ggseqlogo R package (version 0.2) [172]. Motif locations within open chromatin regions were identified using seqkit [128] or Cluster Buster [129] and visualised using svist4get [173]. Heatmap of differential TF binding was made using the EnhancedVolcano R library (version 1.20.0) [131].

Supplementary information titles and legends

Figure S1. Cadherins profile in the coelomic mesothelium and dynamics of EMT and differentiation in the epicardial lineage

(A) Genome tracks showing the chromatin accessibility profiles of epicardial cells at E13.5. Track signal is in frequency-inverse document frequency (TF-IDF) normalized counts and merged from N=3 bulk ATAC-seq independent replicates. High-quality peaks were identified by merging replicates and called using MACS2, followed by iterative filtering.

(B) UMAP dimensionality reduction of the mouse lung ATAC-seq dataset (11,638 cells) at E13.5. Mesothelial cells (663 cells) are highlighted by their enriched gene activity of mesothelial markers (*Upk3b* and *Wt1*). Gene activities were imputed based on chromatin accessibility at the relevant gene locus.

(C) UMAP dimensionality reduction of the mouse pancreas ATAC-seq dataset (8,086 cells) at E13.5. Mesothelial cells (750 cells) are highlighted by their enriched gene activity of mesothelial markers (*Upk3b* and *Wt1*).

(D) Matrix plot showing average expression of cadherins across organ mesothelia.

(E) An overview of the analytical workflow used in this manuscript. Created with Biorender.com.

(F) UMAP dimensionality reduction of the integrated epicardial lineage dataset (7,332 cells), combining four datasets profiling four developmental stages from E12.5 to E17.5. Cell type annotations are shown. Arrows indicate direction of differentiation.

(G) Dot-plot showing the mean expression of marker genes and their corresponding cell states in the integrated epicardial lineage scRNA-seq dataset. Dot size represents the fraction of cells in the cell type that express each gene.

(H) Matrix plot showing average expression of cadherins across cell types in our integrated epicardial lineage scRNA-seq dataset.

(I) Plot showing the imputed expression of *Cdh2* over the pseudotime associated with epicardial lineage EMT and differentiation trajectories. An integrated scRNA-seq dataset of lineage-traced epicardial cells, EPDCs, and derivatives in differentiated cell fates was constructed from independent datasets. Branched trajectories associated with EPDC differentiation into specific terminal cell fates are then inferred via pseudotime calculation

using Palantir [139]. Three trajectories are shown, associated with differentiated fates: Pericytes, Cardiac Fibroblasts, and VICs. Plot sections are coloured to represent distinct phases along the EMT and differentiation trajectory, progressing from heart mesothelium/epicardium to EPDC, and subsequently to terminal fates. Vertical lines indicate the boundary pseudotime values demarcating these trajectory phases. MAGIC imputed expression values plotted.

(J) Plot showing the imputed expression of *Cdh3* over the pseudotime associated with epicardial lineage trajectories

(K) Plot showing the imputed expression of *Cdh11* over the pseudotime associated with epicardial lineage trajectories

EPDC: Epicardium-derived cells; VIC: Valvular interstitial cells; CFb/Fb: Cardiac fibroblast/Fibroblast; SMC: Smooth muscle cell

Figure S2. Epigenomic profiling of the developing epicardium

(A) Profile plots for H3K4me3 histone enrichment scores over gene body regions (±1kb from the transcription start and end sites). Scores were computed using counts per million (CPM)-normalized score of H3K4me3 CUT&RUN-seq generated from the MEC1 epicardial cells.

(B) Profile plots for H3K27ac histone enrichment scores over gene body regions (±1kb from the transcription start and end sites).

(C) Heatmap of normalized peak scores of organ mesothelia in active epicardial enhancer regions, identified using SEACR-called consensus peaks called from H3K27ac CUT&RUN-seq of MEC1 epicardial cells.

(D) Heatmap of normalized peak scores of organ mesothelia in active epicardial promoter regions, identified using SEACR-called consensus peaks called from H3K4me3 CUT&RUN-seq of MEC1 epicardial cells.

(E) Dot-plot showing the percentage of cells and the average scaled expression of co-expression modules across organ mesothelia.

(F) Gene Ontology (GO) term over-representation analysis of genes comprising organ-specific mesothelial modules (heart mesothelium: Meso-M2; lung mesothelium: Meso-M6;

pancreas mesothelium: Meso-M3). GO terms were acquired from the GO “Biological Process” database. P- and q-value cutoffs were set at 0.01 and 0.05, respectively.

(G) Module preservation statistics of organ-specific modules Meso-M2, Meso-M3, and Meso-M6 in the corresponding scATAC-seq datasets of the organ mesothelia. Projection of modules to the scATAC-seq datasets were performed by scoring these modules using gene activity computed by chromatin accessibility counts over gene body and promoter regions. The quality of modules and their degree of preservation across modalities were determined through computing Z-summary statistics.

(H) UMAP plot showing the gene activity of *Pitx2* in the E13.5 mouse pancreas scATAC-seq dataset.

(I) UMAP plot showing the gene activity of *Barx1* in the E13.5 mouse pancreas scATAC-seq dataset.

Figure S3. Epicardial gene regulation by TBX20

(A) Violin plot showing normalised expression of *Fgf2* in organ mesothelial cells

(B) Genome track showing the proposed gene regulation of the *Sema3d* promoter by a distal putative enhancer (chr3:37398983-37399483; highlighted) linked to TBX20. TF-IDF normalised ATAC-seq tracks show bulk (epicardium) and pseudo-bulk (lung and pancreas mesothelium). Histone tracks show MEC1 epicardial H3K27ac and H3K4me3 CPM-normalized enrichment scores. The link plot shows predicted enhancer-promoter (E-P) connections in the epicardium; link colour denotes the ABC score (the strength of E-P prediction).

(C) Footprint signals across organ mesothelia of the proposed *Fgf2* enhancer. CPM-normalised H3K27ac and H3K4me3 enrichment signal is shown. TOBIAS predicted TBX20 binding site is highlighted.

(D) Violin plot showing the normalised expression of *Sema3d* in organ mesothelial cells

(E) Genome tracks showing the proposed gene regulation of the *Sema3d* promoter by an enhancer (chr5:12507516-12508016; highlighted) linked to TBX20.

(F) Footprint (DNA-binding) signals across organ mesothelia of the proposed *Sema3d* enhancer.

(G) Violin plot showing the normalised expression of *Rarres2* in organ mesothelial cells.

(H) Genome tracks showing the proposed gene regulation of the *Rarres2* promoter by a proximal region linked to TBX20. The CRE is highlighted.

(I) Footprint signals across organ mesothelia of the proposed *Rarres2* CRE.

(J) Violin plot showing the normalised expression of *Pkp2* in organ mesothelial cells.

(K) Genome tracks showing the proposed gene regulation of the *Sema3d* promoter by two enhancers (chr16:16219711-16220211 and chr16:16256606-16257106; highlighted) linked to TBX20.

(L) Normalised ATAC-seq signals of the CRE (chr16:16219711-16220211) from (K) in FACS sorted epicardial cells and EPDCs. Positions of TBX20 binding sites in the enhancer are shown.

(M) Normalised ATAC-seq signals of the CRE (chr16:16256606-16257106) from (K) in FACS sorted epicardial cells and EPDCs. Positions of the TBX20 binding site in the enhancer are shown.

(N) Plot showing the imputed expression of *Pkp2* over the pseudotime associated with epicardial lineage trajectories.

(O) Genome tracks showing the proposed gene regulation of the *Cdh2* promoter by an enhancer (chr18:16740742-16741242; highlighted) linked to TBX20.

(P) Normalised ATAC-seq signals of the CRE (chr18:16740742-16741242) from (O) in FACS sorted epicardial cells and EPDCs. Positions of TBX20 binding sites in the enhancer are shown.

Epi: Epicardium; EPDC: Epicardium-derived cells; VIC: Valvular interstitial cells; CFb/Fb: Cardiac fibroblast/Fibroblast; SMC: Smooth muscle cells

Figure S4. Gene markers for the lung mesothelium are putatively regulated by lung-specific regulons

(A) Violin plot showing the normalised expression of *Gli1* in E13.5 embryonic lung

(B) Footprint signals across organ mesothelia of the *Hhip* promoter. The exact position of the GLI1 motif is shown. Spearman correlations (rho) between *Gli1*, the predicted bound region, and *Hhip* expression are shown.

(C) Violin plot showing the normalised expression of *Foxf1* in organ mesothelial cells

(D) Genome track showing the proposed gene regulation of the *Pnoc* promoter by a CRE (chr14:65388142-65388642; highlighted) bound by FOXF1. The E-P link predicted by SCENIC+ was illustrated. The FOXF1 binding positions within the CRE are shown.

(E) Violin plot showing the normalised expression of *Foxp2* in organ mesothelial cells

(F) Footprint signals across organ mesothelia of the *Foxp2* promoter. The FOXF1 binding sites in the CRE are shown.

(G) Genome track showing the proposed gene regulation of the *Gli1* promoter by a distal CRE (chr10:127426586-127427086) linked to FOXF1. The CRE is highlighted. The FOXF1 binding positions within the CRE are shown.

(H) Genome track showing the proposed gene regulation of the *Tbx2* promoter by a proximal region linked to FOXF1. The FOXF1-bound region is highlighted. A violin plot showing the normalised expression of *Tbx2* in organ mesothelial cells is shown.

Figure S5. Gene markers for the pancreas mesothelium are putatively regulated by pancreas-specific regulons

(A) Genome tracks showing the proposed gene regulation of the *Hoxb8* promoter by a proximal region (highlighted) linked to *BARX1*. The *BARX1* binding site is shown.

(B) Genome track showing the proposed gene regulation of the *Igf1* promoter by a distal region (chr10:87850833-87851333; highlighted) linked to *PITX2*.

(C) Footprint signals across organ mesothelia of the *Cdh6* promoter. The location of PITX2 binding site is shown. The curated PITX2 motif is illustrated.

(D) Violin plot showing the normalised expression of *Isl1* and *Gcg* in organ mesothelial cells

(E) Genome track showing the proposed gene regulation of the *Gcg* promoter by a distal CRE (highlighted) linked to ISL1. The curated ISL1 motif is illustrated and ISL1 binding sites within the CRE are shown.

Figure S6. Mesothelial CREs drive temporal gene expression

(A) Stacked violin plot showing the mesothelial median expression of TFs that have motif binding sites in the *Msln* promoter. TFs were identified using the “submerge” function of TOBIAS, which merges all TF binding sites and collects associated footprint information in a region of interest.

(B) Motif logos for WT1 and NRF2/NFE2L2 used in this study.

(C) Plot showing the imputed expression of *Msln* over the pseudotime associated with epicardial lineage trajectories.

(D) Genome tracks showing the regulation of the *Msln* promoter in epicardial EMT and differentiation. The promoter region and two CREs are highlighted. Normalised ATAC-seq tracks of E13.5 epicardial cells and EPDCs are shown. Normalised histone tracks of MEC1 epicardial H3K27ac and H3K4me3.

(E) Violin plot showing the normalised expression of *Tbx20, Fgf2,* and *Csf1* in epicardial cells at E11.5, E13.5, and E17.5.

(F) Genome track showing the regulation of the *Csf1* promoter in epicardial cells over embryonic development. Merged normalised ATAC-seq tracks of E11.5, E13.5, and E17.5epicardial cells (N = 3 per stage). Normalised histone tracks show MEC1 epicardial H3K27ac and H3K4me3. The enhancer predicted to regulate *Csf1* is highlighted (Refer to Figure 3H).

(G) Genome tracks showing the regulation of the *Fgf2* promoter in epicardial cells over embryonic development. The enhancer predicted to regulate *Fgf2* is highlighted (Refer to Figure S3B).

(H) Violin plot showing the normalised expression of *Barx1 and Hoxb8* in pancreas mesothelial cells at E12.5, E13.5, E14.5, and E17.5.

(I) Genome tracks showing the regulation of the *Hoxb8* promoter in the pancreas mesothelium over embryonic development. Normalised ATAC-seq signal tracks of E13.5,

E14.5, and E17.5 pancreas mesothelial cells are shown. The enhancer predicted to regulate *Hoxb8* is highlighted (Refer to Figure S5A).

(J) Violin plot showing the normalised expression of *Isl1 and Gcg* in pancreas mesothelial cells at E12.5, E13.5, E14.5, and E17.5.

(K) Genome tracks showing the regulation of the *Gcg* promoter in the pancreas mesothelium over embryonic development. The enhancer predicted to regulate *Gcg* is highlighted (Refer to Figure S5E).

Epi: Epicardium

Figure S7. Inference of cardiac cell-type specific gene regulatory networks

(A) UMAP dimensionality reduction of the single-cell E13.5 mouse heart ATAC-seq dataset (7,292 cells). CM (2,262 cells) are highlighted by the black arrow, inferred by *Myl2* gene activity. Gene activities were imputed based on chromatin accessibility at the relevant gene locus.

(B) UMAP dimensionality reduction of the single-cell E13.5 mouse heart ATAC-seq dataset. Cardiac vascular ECs (216 cells) are highlighted by the black arrow, inferred by *Fabp4* gene activity.

(C) Correlation matrix of E13.5 cardiac cell types and organ mesothelia based on chromatin-accessible peak signatures.

(D) Upset plot showing the intersection of chromatin peaks from the E13.5 mesothelium across the heart, lung, and pancreas and E13.5 cardiac cell types (CM and CEC).

(E) Pearson correlation matrix of E13.5 cardiac cell types and three organ mesothelia based on transcriptomic signatures. Colour is based on correlation coefficient values (higher value indicates stronger positive correlation)

(F) Dot-plot showing the expression of marker genes and their corresponding cell types. Cells are grouped based on their similarities in a principal component analysis (PCA) representation using a dendrogram.

(G) Motif logos of MAF and TCF21.

(H) Pairwise comparison of TF activity between the heart mesothelium and vascular ECs. The volcano plots show the differential binding activity against the −log10(p value) of all investigated TF motifs; each dot represents one motif. Only motifs for TFs expressed in at least one organ mesothelium are retained. Organ-mesothelium-specific TFs are labelled in red.

(I) Pairwise comparison of TF activity between the heart mesothelium and ventricular CM.

(C)EC: (cardiac) endothelial cells; CM: cardiomyocytes

Figure S8. Conserved CREs regulate mesothelial gene marker expression

(A) Violin plot showing the normalised expression of *Tcf21* and *Maf* in mesothelial cells and cardiac cell types.

(B) Stacked violin plot showing the median expression of *Maf, Krt8, Krt18,* and *Upk3b* in epicardial cells at E11.5, E13.5, and E17.5. Higher expression is shown in yellow.

(C) Genome tracks showing the regulation of the *Krt18* promoter in epicardial cells over embryonic development. The enhancer predicted to regulate *Krt18* is highlighted (Refer to Figure 6B).

(D) Genome tracks showing the regulation of the *Krt8* promoter in epicardial cells over embryonic development. The enhancer predicted to regulate *Krt8* is highlighted (Refer to Figure 6E).

(E) Genome tracks showing the regulation of the *Upk3b* promoter in the epicardium over embryonic development. The enhancer predicted to regulate *Upk3b* is highlighted (refer to Figure 6M).

(F) Stacked violin plot showing the median expression of *Maf, Krt8, Krt18,* and *Upk3b* in pancreas mesothelial cells at E12.5, E13.5, E14.5, and E17.5. Higher expression is shown in yellow.

(G) Genome tracks showing the regulation of the *Krt18* promoter in the pancreas mesothelium over embryonic development. The *Krt18*-linked enhancer and *Krt18* promoter are highlighted.

(H) Genome tracks showing the regulation of the *Krt8* promoter in the pancreas mesothelium over embryonic development. The *Krt8*-linked enhancer and *Krt8* promoter are highlighted.

(I) Genome tracks showing the regulation of the *Upk3b* promoter in the pancreas mesothelium over embryonic development. The *Upk3b*-linked enhancer and the *Upk3b* promoter are highlighted.

(L) Stacked violin plot showing the median expression of *Maf, Krt8, Krt18,* and *Upk3b* in lung mesothelial cells at E12.5, E13.5, E15.5, E17.5, and P3. Higher expression is shown in yellow.

(G) Genome track showing the regulation of the *Krt18* promoter in the lung mesothelium over embryonic development.

(H) Genome track showing the regulation of the *Krt8* promoter in the lung mesothelium over embryonic development.

(I) Genome track showing the regulation of the *Upk3b* promoter in the lung mesothelium over embryonic development.

Figure S9. Mouse-Human comparison of the embryonic epicardium

(A) UMAP dimensionality reduction of the cross-species integrated epicardial cells labelled by species.

(B) UMAP plot of the cross-species integrated epicardial cells labelled by the developmental stage

(C) Stacked violin plot showing the median expression of marker genes that demarcate epicardial cell states in the cross-species dataset. Higher expression is shown in yellow.

(D) Heatmap showing the enrichment of GO terms in epicardial cell states. Depth of colours indicates AUCell enrichment score (darker colour indicates higher score).

(E) Violin plot showing the normalised expression of *Cdh1* between species in the cross-species integrated epicardial dataset.

(F) UMAP plot showing the cell types in mouse embryonic and postnatal heart integrated dataset (72,120 cells), spanning early development (E10.5 to P6).

(G) Violin plot showing the expression of *Cdh1* in mouse embryonic and postnatal heart integrated dataset.

(H) Violin plot showing the expression of *Cdh1* in a published integrated mouse embryonic heart dataset [82].

(I) Genome tracks showing the *Cdh1* promoter and proposing evidence for PRC2-mediated H3K27me3 deposition as a mechanism that silences the expression of this gene in the mouse heart. Track signals include normalised ATAC-seq profiles of organ mesothelia and cardiac cell types; Normalized MEC1 epicardial H3K4me3 enrichment score; Normalized H3K4me3 enrichment score of the mouse whole heart at E13.5 from the ENCODE project; Normalized H3K27me3 enrichment score of the mouse whole heart at E13.5 from the ENCODE project; Mouse embryonic heart ChromHMM-based model with 9 histone marks and 15 states. The ChromHMM model was used to annotate the *Cdh1* promoter.

EC: Endothelial cell; CM: Cardiomyocyte

Supplementary Table 12. Analysis steps and parameterization used for single-cell integration algorithms used in this study

INTEGRATION METHOD	ANALYSIS STEPS
Batch-balanced k-nearest neighbor (BBKNN)	We ran BBKNN using the bbknn function implemented in Scanpy with default parametrization of k = 3 neighbors within each batch.
Harmony	We performed HVG selection in a batch-aware manner and used these HVGs to calculate a PCA embedding for the unintegrated dataset. We then ran the Pytorch implementation of Harmony (version 0.1.8), using HVGs as input.
LIGER	Similar to Harmony, we used HVGs as input. We provided raw counts as input and used LIGER (version 0.2.0) to normalize by library size with a scale factor of 1, scaling without zero-centering, run integrative non-negative matrix factorization with k=30, and perform quantile normalization.
Scanorama	Scanorama (version 1.7.4) was ran using default parameters, using only HVGs as input.
scVI	We ran scVI (version 1.0.4) using raw counts as input and parameterization includes a neural network with negative binomial reconstruction loss, a 30-dimensional latent space, 128 nodes in the hidden layer, and n_layers = 2.
scANVI	We pre-trained a neural network with scVI and provided cell-type label as input. For running scANVI, we used 20 epochs and 100 subsamples for each cell type label class to sample per epoch.
scGen	scGen (version 2.1.0) was ran using default parameters according to the authors' online tutorial.

References

1. Trembley, M.A., et al., *Myocardin-related transcription factors control the motility of epicardium-derived cells and the maturation of coronary vessels.* Development, 2015. **142**(1): p. 21-30.
2. Koopmans, T. and Y. Rinkevich, *Mesothelial to mesenchyme transition as a major developmental and pathological player in trunk organs and their cavities.* Commun Biol, 2018. **1**: p. 170.
3. von Gise, A., et al., *WT1 regulates epicardial epithelial to mesenchymal transition through β-catenin and retinoic acid signaling pathways.* Dev Biol, 2011. **356**(2): p. 421-31.
4. Dixit, R., X. Ai, and A. Fine, *Derivation of lung mesenchymal lineages from the fetal mesothelium requires hedgehog signaling for mesothelial cell entry.* Development, 2013. **140**(21): p. 4398-406.
5. Asahina, K., et al., *Septum transversum-derived mesothelium gives rise to hepatic stellate cells and perivascular mesenchymal cells in developing mouse liver.* Hepatology, 2011. **53**(3): p. 983-95.
6. Liu, C., et al., *Lineage specification of ovarian theca cells requires multicellular interactions via oocyte and granulosa cells.* Nat Commun, 2015. **6**: p. 6934.
7. Ariza, L., et al., *Role of the Wilms' tumor suppressor gene Wt1 in pancreatic development.* Dev Dyn, 2018. **247**(7): p. 924-933.
8. Wessels, A., et al., *Epicardially derived fibroblasts preferentially contribute to the parietal leaflets of the atrioventricular valves in the murine heart.* Dev Biol, 2012. **366**(2): p. 111-24.
9. Li, Y., J. Wang, and K. Asahina, *Mesothelial cells give rise to hepatic stellate cells and myofibroblasts via mesothelial-mesenchymal transition in liver injury.* Proc Natl Acad Sci U S A, 2013. **110**(6): p. 2324-9.
10. Que, J., et al., *Mesothelium contributes to vascular smooth muscle and mesenchyme during lung development.* Proc Natl Acad Sci U S A, 2008. **105**(43): p. 16626-30.
11. Cano, E., R. Carmona, and R. Muñoz-Chápuli, *Wt1-expressing progenitors contribute to multiple tissues in the developing lung.* Am J Physiol Lung Cell Mol Physiol, 2013. **305**(4): p. L322-32.
12. Wilm, B., et al., *The serosal mesothelium is a major source of smooth muscle cells of the gut vasculature.* Development, 2005. **132**(23): p. 5317-28.
13. Martínez-Estrada, O.M., et al., *Wt1 is required for cardiovascular progenitor cell formation through transcriptional control of Snail and E-cadherin.* Nat Genet, 2010. **42**(1): p. 89-93.
14. Lupu, I.E., A.N. Redpath, and N. Smart, *Spatiotemporal Analysis Reveals Overlap of Key Proepicardial Markers in the Developing Murine Heart.* Stem Cell Reports, 2020. **14**(5): p. 770-787.
15. Katz, T.C., et al., *Distinct compartments of the proepicardial organ give rise to coronary vascular endothelial cells.* Dev Cell, 2012. **22**(3): p. 639-50.
16. Cai, C.L., et al., *A myocardial lineage derives from Tbx18 epicardial cells.* Nature, 2008. **454**(7200): p. 104-8.
17. Volz, K.S., et al., *Pericytes are progenitors for coronary artery smooth muscle.* Elife, 2015. **4**.

18. Morimoto, M., et al., *Canonical Notch signaling in the developing lung is required for determination of arterial smooth muscle cells and selection of Clara versus ciliated cell fate.* J Cell Sci, 2010. **123**(Pt 2): p. 213-24.
19. Harari, N., et al., *Pancreatic pericytes originate from the embryonic pancreatic mesenchyme.* Dev Biol, 2019. **449**(1): p. 14-20.
20. Omary, M.B., et al., *The pancreatic stellate cell: a star on the rise in pancreatic diseases.* J Clin Invest, 2007. **117**(1): p. 50-9.
21. Patzek, S., et al., *Loss of Fgf9 in mice leads to pancreatic hypoplasia and asplenia.* iScience, 2023. **26**(4): p. 106500.
22. Yang, J.H. and A.S. Hansen, *Enhancer selectivity in space and time: from enhancer-promoter interactions to promoter activation.* Nat Rev Mol Cell Biol, 2024. **25**(7): p. 574-591.
23. Bravo González-Blas, C., et al., *SCENIC+: single-cell multiomic inference of enhancers and gene regulatory networks.* Nat Methods, 2023. **20**(9): p. 1355-1367.
24. Redpath, A.N. and N. Smart, *Recapturing embryonic potential in the adult epicardium: Prospects for cardiac repair.* Stem Cells Transl Med, 2021. **10**(4): p. 511-521.
25. Redpath, A.N., et al., *Essential regulation of heparan sulfate proteoglycan signalling controls cell behaviour to support cardiac development.* bioRxiv, 2025: p. 2025.07.23.666364.
26. Jiang, S., et al., *Single-cell chromatin accessibility and transcriptome atlas of mouse embryos.* Cell Rep, 2023. **42**(3): p. 112210.
27. Gheldof, A. and G. Berx, *Cadherins and epithelial-to-mesenchymal transition.* Prog Mol Biol Transl Sci, 2013. **116**: p. 317-36.
28. Na, T.Y., et al., *The functional activity of E-cadherin controls tumor cell metastasis at multiple steps.* Proc Natl Acad Sci U S A, 2020. **117**(11): p. 5931-5937.
29. Wu, M., et al., *Epicardial spindle orientation controls cell entry into the myocardium.* Dev Cell, 2010. **19**(1): p. 114-25.
30. Corces, M.R., et al., *The chromatin accessibility landscape of primary human cancers.* Science, 2018. **362**(6413).
31. Li, P., et al., *IGF signaling directs ventricular cardiomyocyte proliferation during embryonic heart development.* Development, 2011. **138**(9): p. 1795-805.
32. Langfelder, P., et al., *Is my network module preserved and reproducible?* PLoS Comput Biol, 2011. **7**(1): p. e1001057.
33. Byrnes, L.E., et al., *Lineage dynamics of murine pancreatic development at single-cell resolution.* Nat Commun, 2018. **9**(1): p. 3922.
34. Rinkevich, Y., et al., *Identification and prospective isolation of a mesothelial precursor lineage giving rise to smooth muscle cells and fibroblasts for mammalian internal organs, and their vasculature.* Nat Cell Biol, 2012. **14**(12): p. 1251-60.
35. Yi, X., et al., *Interrogating cell type-specific cooperation of transcriptional regulators in 3D chromatin.* iScience, 2021. **24**(12): p. 103468.
36. Nasser, J., et al., *Genome-wide enhancer maps link risk variants to disease genes.* Nature, 2021. **593**(7858): p. 238-243.
37. Ustiyan, V., et al., *FOXF1 transcription factor promotes lung morphogenesis by inducing cellular proliferation in fetal lung mesenchyme.* Dev Biol, 2018. **443**(1): p. 50-63.

38. Shu, W., et al., *Foxp2 and Foxp1 cooperatively regulate lung and esophagus development.* Development, 2007. **134**(10): p. 1991-2000.
39. Ahlgren, U., et al., *Independent requirement for ISL1 in formation of pancreatic mesenchyme and islet cells.* Nature, 1997. **385**(6613): p. 257-60.
40. Habener, J.F. and D.A. Stoffers, *A newly discovered role of transcription factors involved in pancreas development and the pathogenesis of diabetes mellitus.* Proc Assoc Am Physicians, 1998. **110**(1): p. 12-21.
41. Boogerd, C.J., et al., *Tbx20 Is Required in Mid-Gestation Cardiomyocytes and Plays a Central Role in Atrial Development.* Circ Res, 2018. **123**(4): p. 428-442.
42. Lu, F., A. Langenbacher, and J.N. Chen, *Tbx20 drives cardiac progenitor formation and cardiomyocyte proliferation in zebrafish.* Dev Biol, 2017. **421**(2): p. 139-148.
43. Tang, Y., et al., *TBX20 Improves Contractility and Mitochondrial Function During Direct Human Cardiac Reprogramming.* Circulation, 2022. **146**(20): p. 1518-1536.
44. Weinberger, T., et al., *Resident and recruited macrophages differentially contribute to cardiac healing after myocardial ischemia.* Elife, 2024. **12**.
45. Cahill, T.J., et al., *Tissue-resident macrophages regulate lymphatic vessel growth and patterning in the developing heart.* Development, 2021. **148**(3).
46. Stevens, S.M., et al., *Epicardium is required for cardiac seeding by yolk sac macrophages, precursors of resident macrophages of the adult heart.* Dev Biol, 2016. **413**(2): p. 153-159.
47. Leid, J., et al., *Primitive Embryonic Macrophages are Required for Coronary Development and Maturation.* Circ Res, 2016. **118**(10): p. 1498-511.
48. Sun, J., et al., *ptx3a(+) fibroblast/epicardial cells provide a transient macrophage niche to promote heart regeneration.* Cell Rep, 2024. **43**(4): p. 114092.
49. Knight-Schrijver, V.R., et al., *A single-cell comparison of adult and fetal human epicardium defines the age-associated changes in epicardial activity.* Nat Cardiovasc Res, 2022. **1**(12): p. 1215-1229.
50. Risebro, C.A., et al., *Characterisation of the human embryonic and foetal epicardium during heart development.* Development, 2015. **142**(21): p. 3630-6.
51. Diao, Y., et al., *Identification of novel GLI1 target genes and regulatory circuits in human cancer cells.* Mol Oncol, 2018. **12**(10): p. 1718-1734.
52. Lüdtke, T.H., et al., *Tbx2 controls lung growth by direct repression of the cell cycle inhibitor genes Cdkn1a and Cdkn1b.* PLoS Genet, 2013. **9**(1): p. e1003189.
53. Lüdtke, T.H., et al., *Combined genomic and proteomic approaches reveal DNA binding sites and interaction partners of TBX2 in the developing lung.* Respir Res, 2021. **22**(1): p. 85.
54. Wojahn, I., et al., *TBX2-positive cells represent a multi-potent mesenchymal progenitor pool in the developing lung.* Respir Res, 2019. **20**(1): p. 292.
55. Kim, B.M., et al., *Independent functions and mechanisms for homeobox gene Barx1 in patterning mouse stomach and spleen.* Development, 2007. **134**(20): p. 3603-13.
56. Welsh, I.C., et al., *Integration of left-right Pitx2 transcription and Wnt signaling drives asymmetric gut morphogenesis via Daam2.* Dev Cell, 2013. **26**(6): p. 629-44.

57. Kido, Y., et al., *Effects of mutations in the insulin-like growth factor signaling system on embryonic pancreas development and beta-cell compensation to insulin resistance.* J Biol Chem, 2002. **277**(39): p. 36740-7.
58. Mah, S.P., et al., *Kidney development in cadherin-6 mutants: delayed mesenchyme-to-epithelial conversion and loss of nephrons.* Dev Biol, 2000. **223**(1): p. 38-53.
59. Jia, L., et al., *Distinct roles of cadherin-6 and E-cadherin in tubulogenesis and lumen formation.* Mol Biol Cell, 2011. **22**(12): p. 2031-41.
60. Heidenreich, A.C., et al., *Identification of mouse and human embryonic pancreatic cells with adult Procr(+) progenitor transcriptomic and epigenomic characteristics.* Front Endocrinol (Lausanne), 2025. **16**: p. 1543960.
61. Du, J., et al., *Single-cell and spatial heterogeneity landscapes of mature epicardial cells.* J Pharm Anal, 2023. **13**(8): p. 894-907.
62. Chu, G.J., et al., *High mesothelin expression by immunohistochemistry predicts improved survival in pleural mesothelioma.* Histopathology, 2023. **83**(2): p. 202-210.
63. Servais, E.L., et al., *Mesothelin overexpression promotes mesothelioma cell invasion and MMP-9 secretion in an orthotopic mouse model and in epithelioid pleural mesothelioma patients.* Clin Cancer Res, 2012. **18**(9): p. 2478-89.
64. Melaiu, O., et al., *MSLN gene silencing has an anti-malignant effect on cell lines overexpressing mesothelin deriving from malignant pleural mesothelioma.* PLoS One, 2014. **9**(1): p. e85935.
65. Tian, Y., et al., *Generation of mature epicardium derived from human-induced pluripotent stem cells via inhibition of mTOR signaling.* Nat Commun, 2025. **16**(1): p. 5902.
66. Sakamoto, T., et al., *A Critical Role for Estrogen-Related Receptor Signaling in Cardiac Maturation.* Circ Res, 2020. **126**(12): p. 1685-1702.
67. Sweat, M.E., et al., *Tbx5 maintains atrial identity in post-natal cardiomyocytes by regulating an atrial-specific enhancer network.* Nat Cardiovasc Res, 2023. **2**(10): p. 881-898.
68. Lee, S.H., et al., *Notch pathway targets proangiogenic regulator Sox17 to restrict angiogenesis.* Circ Res, 2014. **115**(2): p. 215-26.
69. Yao, Y., J. Yao, and K.I. Boström, *SOX Transcription Factors in Endothelial Differentiation and Endothelial-Mesenchymal Transitions.* Front Cardiovasc Med, 2019. **6**: p. 30.
70. Sakamoto, T., et al., *The nuclear receptor ERR cooperates with the cardiogenic factor GATA4 to orchestrate cardiomyocyte maturation.* Nat Commun, 2022. **13**(1): p. 1991.
71. Fontijn, R.D., et al., *SOX-18 controls endothelial-specific claudin-5 gene expression and barrier function.* Am J Physiol Heart Circ Physiol, 2008. **294**(2): p. H891-900.
72. Tandon, P., et al., *Tcf21 regulates the specification and maturation of proepicardial cells.* Development, 2013. **140**(11): p. 2409-21.
73. Acharya, A., et al., *The bHLH transcription factor Tcf21 is required for lineage-specific EMT of cardiac fibroblast progenitors.* Development, 2012. **139**(12): p. 2139-49.
74. Kuburich, N.A., et al., *Vimentin and cytokeratin: Good alone, bad together.* Semin Cancer Biol, 2022. **86**(Pt 3): p. 816-826.

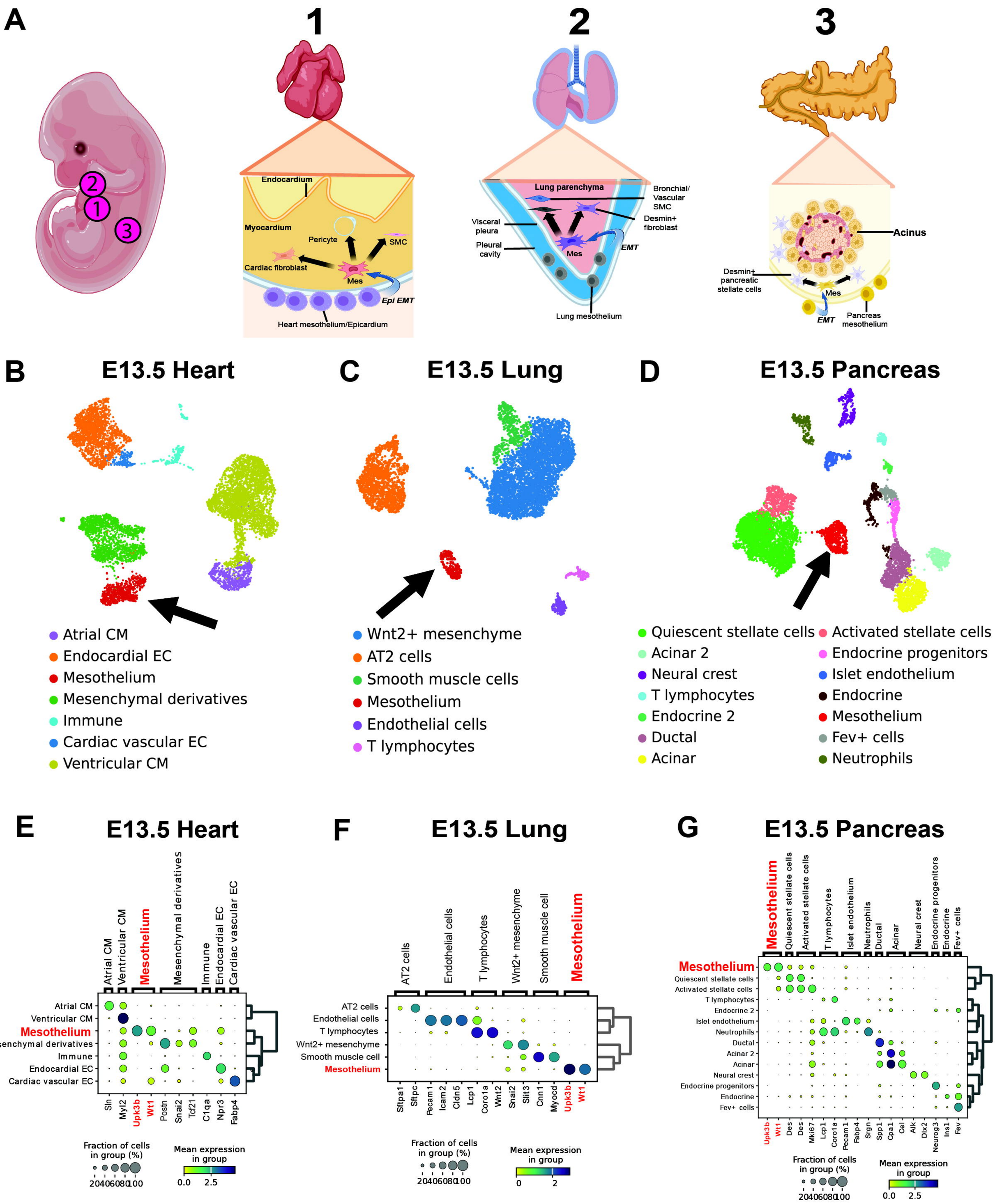
75. Meier, A.B., et al., *Epicardioid single-cell genomics uncovers principles of human epicardium biology in heart development and disease.* Nat Biotechnol, 2023. **41**(12): p. 1787-1800.
76. Gayoso, A., et al., *A Python library for probabilistic analysis of single-cell omics data.* Nat Biotechnol, 2022. **40**(2): p. 163-166.
77. Ma, X.R., et al., *Molecular convergence of risk variants for congenital heart defects leveraging a regulatory map of the human fetal heart.* medRxiv, 2024.
78. Winter, E.M. and A.C. Gittenberger-de Groot, *Epicardium-derived cells in cardiogenesis and cardiac regeneration.* Cell Mol Life Sci, 2007. **64**(6): p. 692-703.
79. Wolman, M.A., et al., *Repulsion and attraction of axons by semaphorin3D are mediated by different neuropilins in vivo.* J Neurosci, 2004. **24**(39): p. 8428-35.
80. Ge, Y., et al., *Human epicardium-derived cells reinforce cardiac sympathetic innervation.* J Mol Cell Cardiol, 2020. **143**: p. 26-37.
81. Gulati, G.S., et al., *Single-cell transcriptional diversity is a hallmark of developmental potential.* Science, 2020. **367**(6476): p. 405-411.
82. Feng, W., et al., *Single-cell transcriptomic analysis identifies murine heart molecular features at embryonic and neonatal stages.* Nat Commun, 2022. **13**(1): p. 7960.
83. He, A., et al., *Polycomb repressive complex 2 regulates normal development of the mouse heart.* Circ Res, 2012. **110**(3): p. 406-15.
84. Kohela, A., et al., *Epicardial differentiation drives fibro-fatty remodeling in arrhythmogenic cardiomyopathy.* Sci Transl Med, 2021. **13**(612): p. eabf2750.
85. Yuan, P., et al., *Single-Cell RNA Sequencing Uncovers Paracrine Functions of the Epicardial-Derived Cells in Arrhythmogenic Cardiomyopathy.* Circulation, 2021. **143**(22): p. 2169-2187.
86. Gilbert, R.M., L.E. Schappell, and J.P. Gleghorn, *Defective mesothelium and limited physical space are drivers of dysregulated lung development in a genetic model of congenital diaphragmatic hernia.* Development, 2021. **148**(10).
87. Suri, M., et al., *WT1 mutations in Meacham syndrome suggest a coelomic mesothelial origin of the cardiac and diaphragmatic malformations.* Am J Med Genet A, 2007. **143a**(19): p. 2312-20.
88. Ali, S.R., et al., *Developmental heterogeneity of cardiac fibroblasts does not predict pathological proliferation and activation.* Circ Res, 2014. **115**(7): p. 625-35.
89. Liu, F., et al., *Crosstalk between pleural mesothelial cell and lung fibroblast contributes to pulmonary fibrosis.* Biochim Biophys Acta Mol Cell Res, 2020. **1867**(11): p. 118806.
90. Karki, S., et al., *Wilms' tumor 1 (Wt1) regulates pleural mesothelial cell plasticity and transition into myofibroblasts in idiopathic pulmonary fibrosis.* Faseb j, 2014. **28**(3): p. 1122-31.
91. Suffee, N., et al., *Reactivation of the Epicardium at the Origin of Myocardial Fibro-Fatty Infiltration During the Atrial Cardiomyopathy.* Circ Res, 2020. **126**(10): p. 1330-1342.
92. Hiriart, E., R. Deepe, and A. Wessels, *Mesothelium and Malignant Mesothelioma.* J Dev Biol, 2019. **7**(2).
93. Greenbaum, A. and H.R. Alexander, *Peritoneal mesothelioma.* Transl Lung Cancer Res, 2020. **9**(Suppl 1): p. S120-s132.

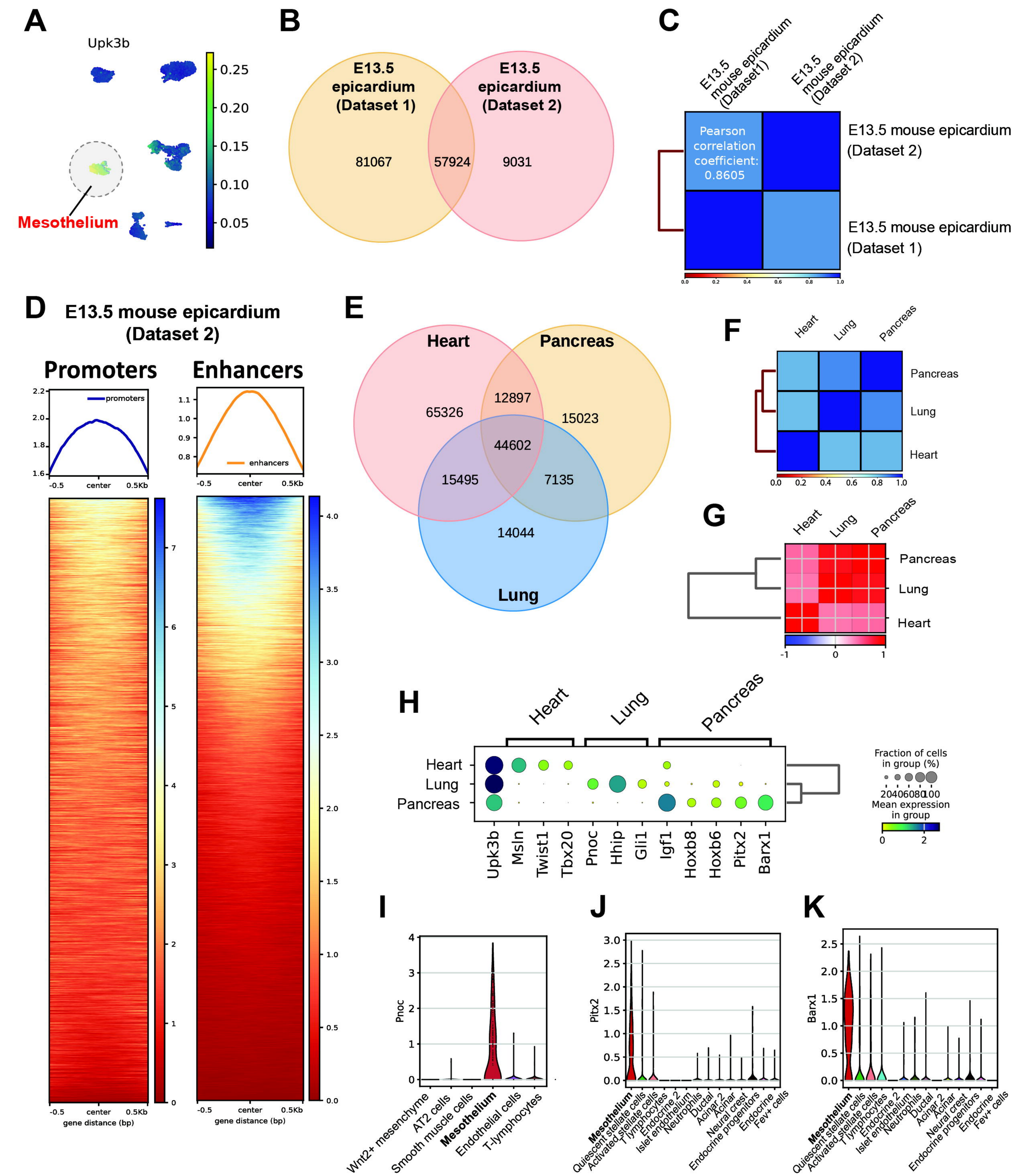
94. Mutsaers, S.E. and S. Wilkosz, *Structure and function of mesothelial cells.* Cancer Treat Res, 2007. **134**: p. 1-19.
95. Zhou, B. and W.T. Pu, *Genetic Cre-loxP assessment of epicardial cell fate using Wt1-driven Cre alleles.* Circ Res, 2012. **111**(11): p. e276-80.
96. Mahlapuu, M., S. Enerbäck, and P. Carlsson, *Haploinsufficiency of the forkhead gene Foxf1, a target for sonic hedgehog signaling, causes lung and foregut malformations.* Development, 2001. **128**(12): p. 2397-406.
97. Grindley, J.C., et al., *Evidence for the involvement of the Gli gene family in embryonic mouse lung development.* Dev Biol, 1997. **188**(2): p. 337-48.
98. Larsen, B.M., et al., *Mesenchymal Hox6 function is required for mouse pancreatic endocrine cell differentiation.* Development, 2015. **142**(22): p. 3859-68.
99. Matthes, S.A., S. Taffet, and M. Delmar, *Plakophilin-2 and the migration, differentiation and transformation of cells derived from the epicardium of neonatal rat hearts.* Cell Commun Adhes, 2011. **18**(4): p. 73-84.
100. Hatzfeld, M., R. Keil, and T.M. Magin, *Desmosomes and Intermediate Filaments: Their Consequences for Tissue Mechanics.* Cold Spring Harb Perspect Biol, 2017. **9**(6).
101. Lopez-Pajares, V., et al., *A LncRNA-MAF:MAFB transcription factor network regulates epidermal differentiation.* Dev Cell, 2015. **32**(6): p. 693-706.
102. Nishino, T., et al., *Single Cell Multimodal Analyses Reveal Epigenomic and Transcriptomic Basis for Birth Defects in Maternal Diabetes.* Nat Cardiovasc Res, 2023. **2**(12): p. 1190-1203.
103. Manivannan, S., et al., *Single-cell transcriptomic profiling unveils dysregulation of cardiac progenitor cells and cardiomyocytes in a mouse model of maternal hyperglycemia.* Commun Biol, 2022. **5**(1): p. 820.
104. Sun, T., et al., *Dbh(+) catecholaminergic cardiomyocytes contribute to the structure and function of the cardiac conduction system in murine heart.* Nat Commun, 2023. **14**(1): p. 7801.
105. Quijada, P., et al., *Coordination of endothelial cell positioning and fate specification by the epicardium.* Nat Commun, 2021. **12**(1): p. 4155.
106. Lupu, I.-E., et al., *SRSF3 is a key regulator of epicardial formation.* bioRxiv, 2025: p. 2021.11.25.470003.
107. Shen, J., et al., *Single-Cell Atlas of Neonatal Mouse Hearts Reveals an Unexpected Cardiomyocyte.* J Am Heart Assoc, 2023. **12**(23): p. e028287.
108. Zepp, J.A., et al., *Genomic, epigenomic, and biophysical cues controlling the emergence of the lung alveolus.* Science, 2021. **371**(6534).
109. He, Y., et al., *Spatiotemporal DNA methylome dynamics of the developing mouse fetus.* Nature, 2020. **583**(7818): p. 752-759.
110. van der Velde, A., et al., *Annotation of chromatin states in 66 complete mouse epigenomes during development.* Commun Biol, 2021. **4**(1): p. 239.
111. Cui, Y., et al., *Single-Cell Transcriptome Analysis Maps the Developmental Track of the Human Heart.* Cell Rep, 2019. **26**(7): p. 1934-1950.e5.
112. Ameen, M., et al., *Integrative single-cell analysis of cardiogenesis identifies developmental trajectories and non-coding mutations in congenital heart disease.* Cell, 2022. **185**(26): p. 4937-4953.e23.
113. Streef, T.J., et al., *Single-cell analysis of human fetal epicardium reveals its cellular composition and identifies CRIP1 as a modulator of EMT.* Stem Cell Reports, 2023. **18**(7): p. 1421-1435.

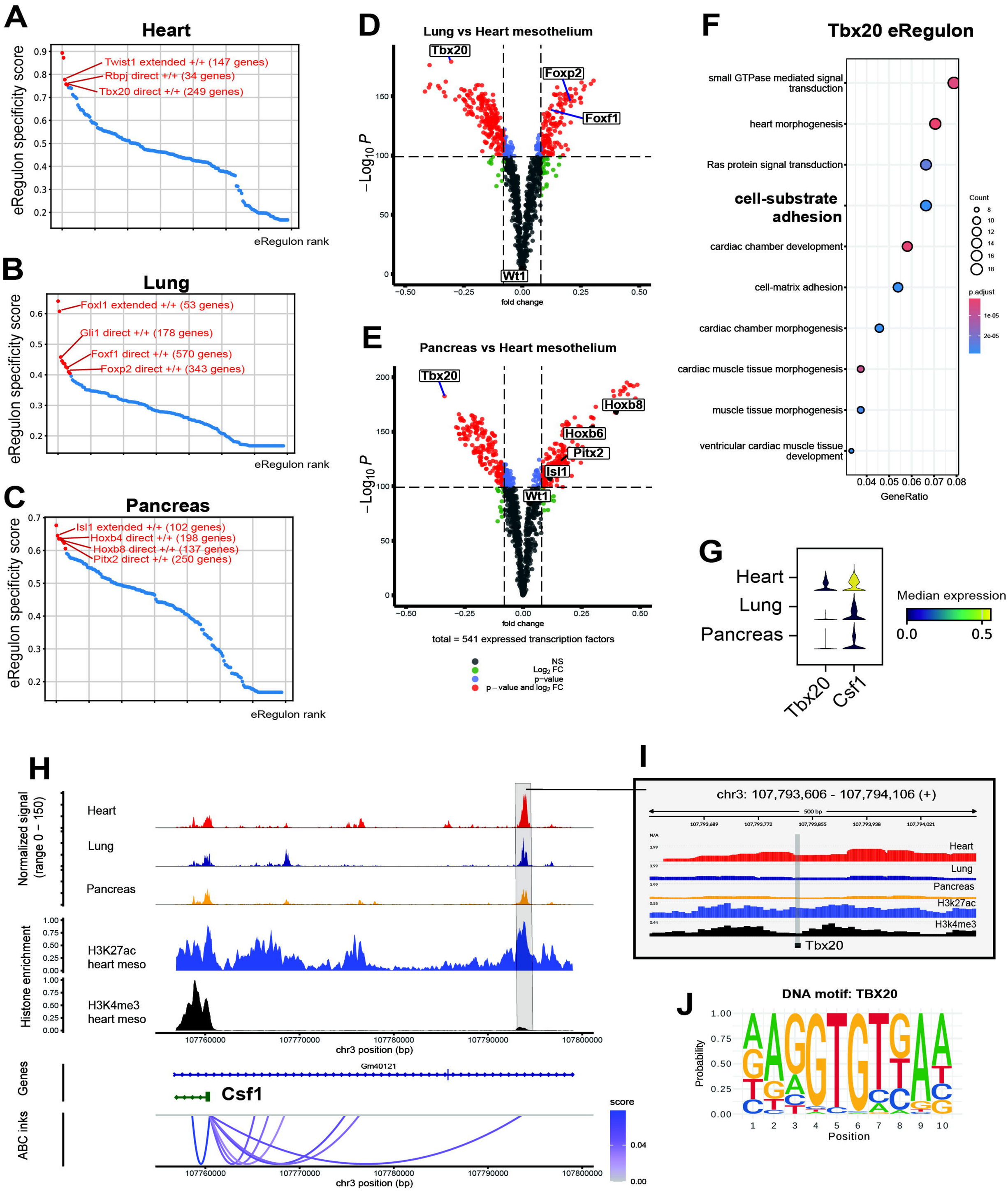
114. Travisano, S.I., et al., *Single-nuclei multiomic analyses identify human cardiac lymphatic endothelial cells associated with coronary arteries in the epicardium.* Cell Rep, 2023. **42**(9): p. 113106.
115. Danecek, P., et al., *Twelve years of SAMtools and BCFtools.* Gigascience, 2021. **10**(2).
116. Yu, G., L.G. Wang, and Q.Y. He, *ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization.* Bioinformatics, 2015. **31**(14): p. 2382-3.
117. Robinson, M.D., D.J. McCarthy, and G.K. Smyth, *edgeR: a Bioconductor package for differential expression analysis of digital gene expression data.* Bioinformatics, 2010. **26**(1): p. 139-40.
118. Zhang, Y., et al., *Model-based analysis of ChIP-Seq (MACS).* Genome Biol, 2008. **9**(9): p. R137.
119. Li, H., et al., *The Sequence Alignment/Map format and SAMtools.* Bioinformatics, 2009. **25**(16): p. 2078-9.
120. Quinlan, A.R. and I.M. Hall, *BEDTools: a flexible suite of utilities for comparing genomic features.* Bioinformatics, 2010. **26**(6): p. 841-2.
121. Persad, S., et al., *SEACells infers transcriptional and epigenomic cellular states from single-cell genomics data.* Nat Biotechnol, 2023. **41**(12): p. 1746-1757.
122. Weiler, P., et al., *CellRank 2: unified fate mapping in multiview single-cell data.* Nat Methods, 2024. **21**(7): p. 1196-1205.
123. Zeng, Z., et al., *OmicVerse: a framework for bridging and deepening insights across bulk and single-cell sequencing.* Nat Commun, 2024. **15**(1): p. 5983.
124. Zhang, K., et al., *A fast, scalable and versatile tool for analysis of single-cell omics data.* Nat Methods, 2024. **21**(2): p. 217-227.
125. Ramírez, F., et al., *deepTools: a flexible platform for exploring deep-sequencing data.* Nucleic Acids Res, 2014. **42**(Web Server issue): p. W187-91.
126. Liao, Y., G.K. Smyth, and W. Shi, *featureCounts: an efficient general purpose program for assigning sequence reads to genomic features.* Bioinformatics, 2014. **30**(7): p. 923-30.
127. Bentsen, M., et al., *ATAC-seq footprinting unravels kinetics of transcription factor binding during zygotic genome activation.* Nat Commun, 2020. **11**(1): p. 4267.
128. Shen, W., et al., *SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation.* PLoS One, 2016. **11**(10): p. e0163962.
129. Frith, M.C., M.C. Li, and Z. Weng, *Cluster-Buster: Finding dense clusters of motifs in DNA sequences.* Nucleic Acids Res, 2003. **31**(13): p. 3666-8.
130. Stuart, T., et al., *Single-cell chromatin state analysis with Signac.* Nat Methods, 2021. **18**(11): p. 1333-1341.
131. Kevin Blighe, S.R.a.M.L., *EnhancedVolcano: Publication-ready volcano plots with enhanced colouring and labeling*. 2020, Github: https://github.com/kevinblighe/EnhancedVolcano.
132. Lawrence, M., et al., *Software for computing and annotating genomic ranges.* PLoS Comput Biol, 2013. **9**(8): p. e1003118.
133. Thorvaldsdóttir, H., J.T. Robinson, and J.P. Mesirov, *Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration.* Brief Bioinform, 2013. **14**(2): p. 178-92.

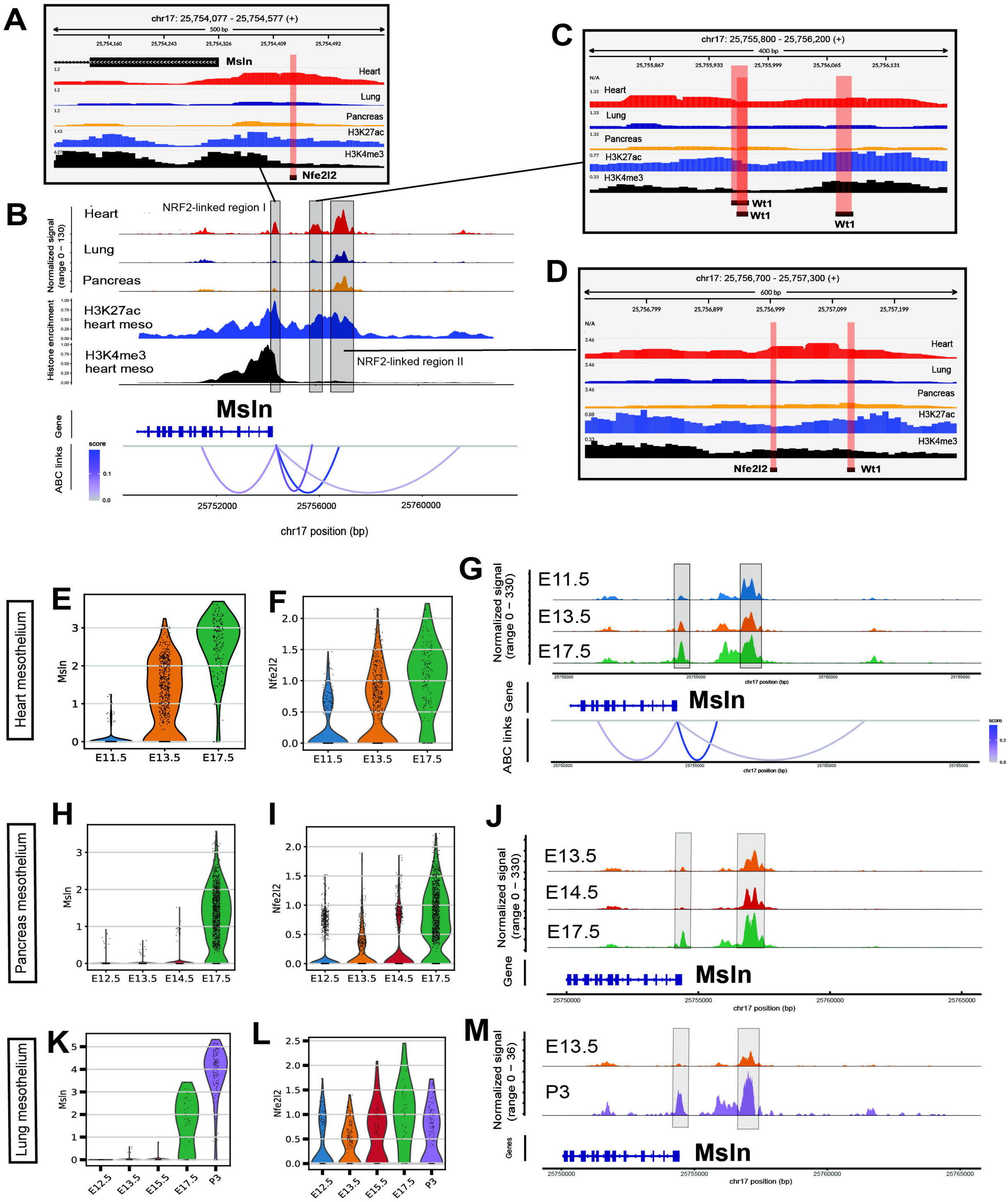
134. Wolf, F.A., P. Angerer, and F.J. Theis, *SCANPY: large-scale single-cell gene expression data analysis.* Genome Biol, 2018. **19**(1): p. 15.
135. Young, M.D. and S. Behjati, *SoupX removes ambient RNA contamination from droplet-based single-cell RNA sequencing data.* Gigascience, 2020. **9**(12).
136. Germain, P.L., et al., *Doublet identification in single-cell sequencing data using scDblFinder.* F1000Res, 2021. **10**: p. 979.
137. Lun, A.T., K. Bach, and J.C. Marioni, *Pooling across cells to normalize single-cell RNA sequencing data with many zero counts.* Genome Biol, 2016. **17**: p. 75.
138. Kelly Street, F.W.T., Davide Risso, Stephanie Hicks, *scry: Small-Count Analysis Methods for High-Dimensional Data*. 2025: https://bioconductor.org/packages/scry.
139. Setty, M., et al., *Characterization of cell fate probabilities in single-cell data with Palantir.* Nat Biotechnol, 2019. **37**(4): p. 451-460.
140. Love, M.I., W. Huber, and S. Anders, *Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2.* Genome Biol, 2014. **15**(12): p. 550.
141. Stefan Peidli, i., & pakiessling, *stefanpeidli/mousipy: v0.1.7 (v0.1.7)*. 2025: Zenodo.
142. Morabito, S., et al., *hdWGCNA identifies co-expression networks in high-dimensional transcriptomics data.* Cell Rep Methods, 2023. **3**(6): p. 100498.
143. Wu, T., et al., *clusterProfiler 4.0: A universal enrichment tool for interpreting omics data.* Innovation (Camb), 2021. **2**(3): p. 100141.
144. Domínguez Conde, C., et al., *Cross-tissue immune cell analysis reveals tissue-specific features in humans.* Science, 2022. **376**(6594): p. eabl5197.
145. Polański, K., et al., *BBKNN: fast batch alignment of single cell transcriptomes.* Bioinformatics, 2020. **36**(3): p. 964-965.
146. Korsunsky, I., et al., *Fast, sensitive and accurate integration of single-cell data with Harmony.* Nat Methods, 2019. **16**(12): p. 1289-1296.
147. Welch, J.D., et al., *Single-Cell Multi-omic Integration Compares and Contrasts Features of Brain Cell Identity.* Cell, 2019. **177**(7): p. 1873-1887.e17.
148. Hie, B., B. Bryson, and B. Berger, *Efficient integration of heterogeneous single-cell transcriptomes using Scanorama.* Nat Biotechnol, 2019. **37**(6): p. 685-691.
149. Lopez, R., et al., *Deep generative modeling for single-cell transcriptomics.* Nat Methods, 2018. **15**(12): p. 1053-1058.
150. Xu, C., et al., *Probabilistic harmonization and annotation of single-cell transcriptomics data with deep generative models.* Mol Syst Biol, 2021. **17**(1): p. e9620.
151. Lotfollahi, M., F.A. Wolf, and F.J. Theis, *scGen predicts single-cell perturbation responses.* Nat Methods, 2019. **16**(8): p. 715-721.
152. Shapira, S., *ATAC Sequencing Protocol*. 2021: Protocols.io.
153. Buenrostro, J.D., et al., *ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide.* Curr Protoc Mol Biol, 2015. **109**: p. 21.29.1-21.29.9.
154. Ewels, P.A., et al., *The nf-core framework for community-curated bioinformatics pipelines.* Nat Biotechnol, 2020. **38**(3): p. 276-278.

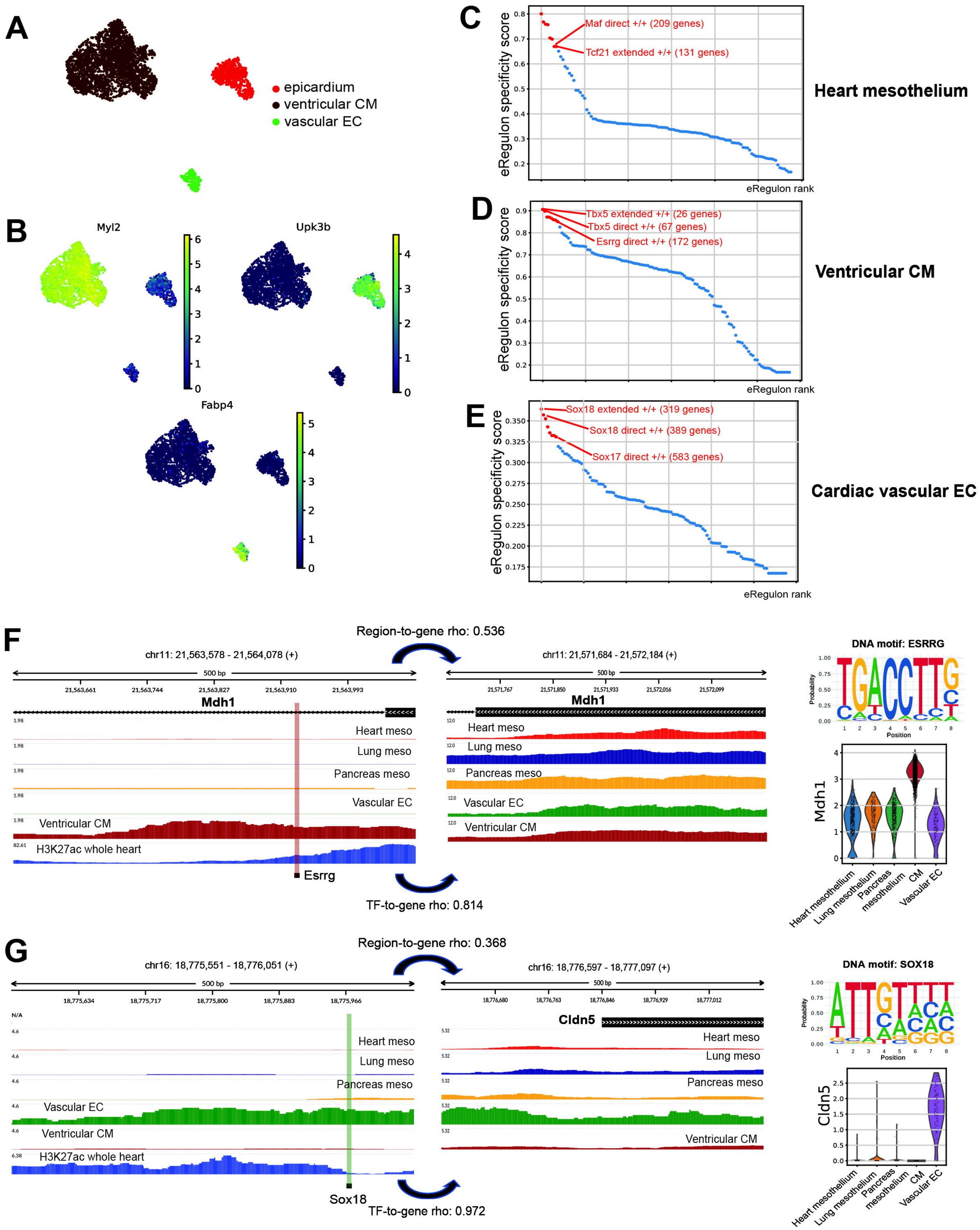
155. Amemiya, H.M., A. Kundaje, and A.P. Boyle, *The ENCODE Blacklist: Identification of Problematic Regions of the Genome.* Sci Rep, 2019. **9**(1): p. 9354.
156. Luecken, M.D. and F.J. Theis, *Current best practices in single-cell RNA-seq analysis: a tutorial.* Mol Syst Biol, 2019. **15**(6): p. e8746.
157. Galow, A.M., et al., *Quality control in scRNA-Seq can discriminate pacemaker cells: the mtRNA bias.* Cell Mol Life Sci, 2021. **78**(19-20): p. 6585-6592.
158. McInnes, L., et al., *UMAP: Uniform Manifold Approximation and Projection.* Journal of Open Source Software, 2018. **3**(29).
159. Traag, V.A., L. Waltman, and N.J. van Eck, *From Louvain to Leiden: guaranteeing well-connected communities.* Sci Rep, 2019. **9**(1): p. 5233.
160. Martens, L.D., et al., *Modeling fragment counts improves single-cell ATAC-seq analysis.* Nat Methods, 2024. **21**(1): p. 28-31.
161. Reske, J.J., M.R. Wilson, and R.L. Chandler, *ATAC-seq normalization method can significantly affect differential accessibility analysis and interpretation.* Epigenetics Chromatin, 2020. **13**(1): p. 22.
162. Luecken, M.D., et al., *Benchmarking atlas-level data integration in single-cell genomics.* Nat Methods, 2022. **19**(1): p. 41-50.
163. O'Leary, K. and D. Zheng, *Metacell-based differential expression analysis identifies cell type specific temporal gene response programs in COVID-19 patient PBMCs.* NPJ Syst Biol Appl, 2024. **10**(1): p. 36.
164. Rudat, C., et al., *Upk3b is dispensable for development and integrity of urothelium and mesothelium.* PLoS One, 2014. **9**(11): p. e112112.
165. Mitchell, T.S., et al., *RGS5 expression is a quantitative measure of pericyte coverage of blood vessels.* Angiogenesis, 2008. **11**(2): p. 141-51.
166. Buechler, M.B., et al., *Cross-tissue organization of the fibroblast lineage.* Nature, 2021. **593**(7860): p. 575-579.
167. Gao, Y., et al., *Cross-tissue human fibroblast atlas reveals myofibroblast subtypes with distinct roles in immune modulation.* Cancer Cell, 2024. **42**(10): p. 1764-1783.e10.
168. Wünnemann, F., et al., *Loss of ADAMTS19 causes progressive non-syndromic heart valve disease.* Nat Genet, 2020. **52**(1): p. 40-47.
169. Kwong, G., et al., *Generation of a Purified iPSC-Derived Smooth Muscle-like Population for Cell Sheet Engineering.* Stem Cell Reports, 2019. **13**(3): p. 499-514.
170. Aibar, S., et al., *SCENIC: single-cell regulatory network inference and clustering.* Nat Methods, 2017. **14**(11): p. 1083-1086.
171. Li, B., et al., *A Comprehensive Mouse Transcriptomic BodyMap across 17 Tissues by RNA-seq.* Sci Rep, 2017. **7**(1): p. 4200.
172. Wagih, O., *ggseqlogo: a versatile R package for drawing sequence logos.* Bioinformatics, 2017. **33**(22): p. 3645-3647.
173. Egorov, A.A., et al., *svist4get: a simple visualization tool for genomic tracks from sequencing experiments.* BMC Bioinformatics, 2019. **20**(1): p. 113.

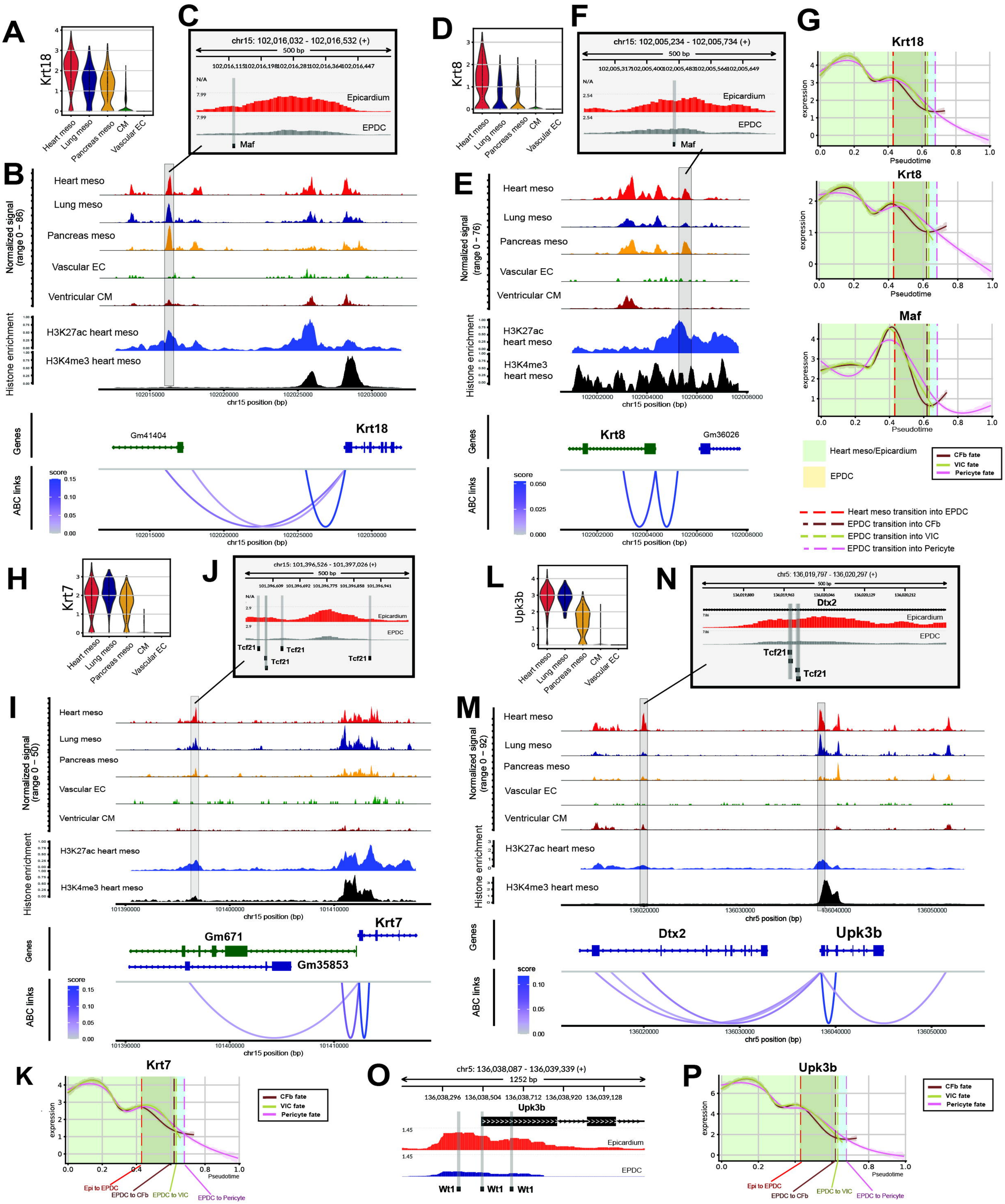

A
1
2
3
Endocardium
Myocardium
Pericyte
SMC
Cardiac fibroblast
Mes
Epi EMT
Heart mesothelium/Epicardium
Lung parenchyma
Bronchial/ Vascular SMC
Desmin+ fibroblast
Visceral pleura
Pleural cavity
Mes
EMT
Lung mesothelium
Acinus
Desmin+ pancreatic stellate cells
Mes
EMT
Pancreas mesothelium
B
E13.5 Heart
Atrial CM
Endocardial EC
Mesothelium
Mesenchymal derivatives
Immune
Cardiac vascular EC
Ventricular CM
C
E13.5 Lung
Wnt2+ mesenchyme
AT2 cells
Smooth muscle cells
Mesothelium
Endothelial cells
T lymphocytes
D
E13.5 Pancreas
Quiescent stellate cells
Activated stellate cells
Acinar 2
Endocrine progenitors
Neural crest
Islet endothelium
T lymphocytes
Endocrine
Endocrine 2
Mesothelium
Ductal
Fev+ cells
Acinar
Neutrophils
E
E13.5 Heart
Atrial CM
Ventricular CM
Mesothelium
Mesenchymal derivatives
Immune
Endocardial EC
Cardiac vascular EC
Sln
Myl2
Upk3b
Wt1
Postn
Snai2
Tcf21
C1qa
Npr3
Fabp4
Fraction of cells in group (%)
20 40 60 80 100
Mean expression in group
0.0 2.5
F
E13.5 Lung
AT2 cells
Endothelial cells
T lymphocytes
Wnt2+ mesenchyme
Smooth muscle cell
Mesothelium
Sftpa1
Sftpc
Pecam1
Icam2
Cldn5
Lcp1
Coro1a
Wnt2
Snai2
Slit3
Cnn1
Myocd
Upk3b
Wt1
Fraction of cells in group (%)
20 40 60 80 100
Mean expression in group
0 2
G
E13.5 Pancreas
Mesothelium
Quiescent stellate cells
Activated stellate cells
T lymphocytes
Islet endothelium
Neutrophils
Ductal
Acinar
Neural crest
Endocrine progenitors
Endocrine
Fev+ cells
Upk3b
Wt1
Des
Des
Mki67
Lcp1
Coro1a
Pecam1
Fabp4
Srgn
Spp1
Cpa1
Cel
Alk
Dlx2
Neurog3
Ins1
Fev
Fraction of cells in group (%)
20 40 60 80 100
Mean expression in group
0.0 2.5

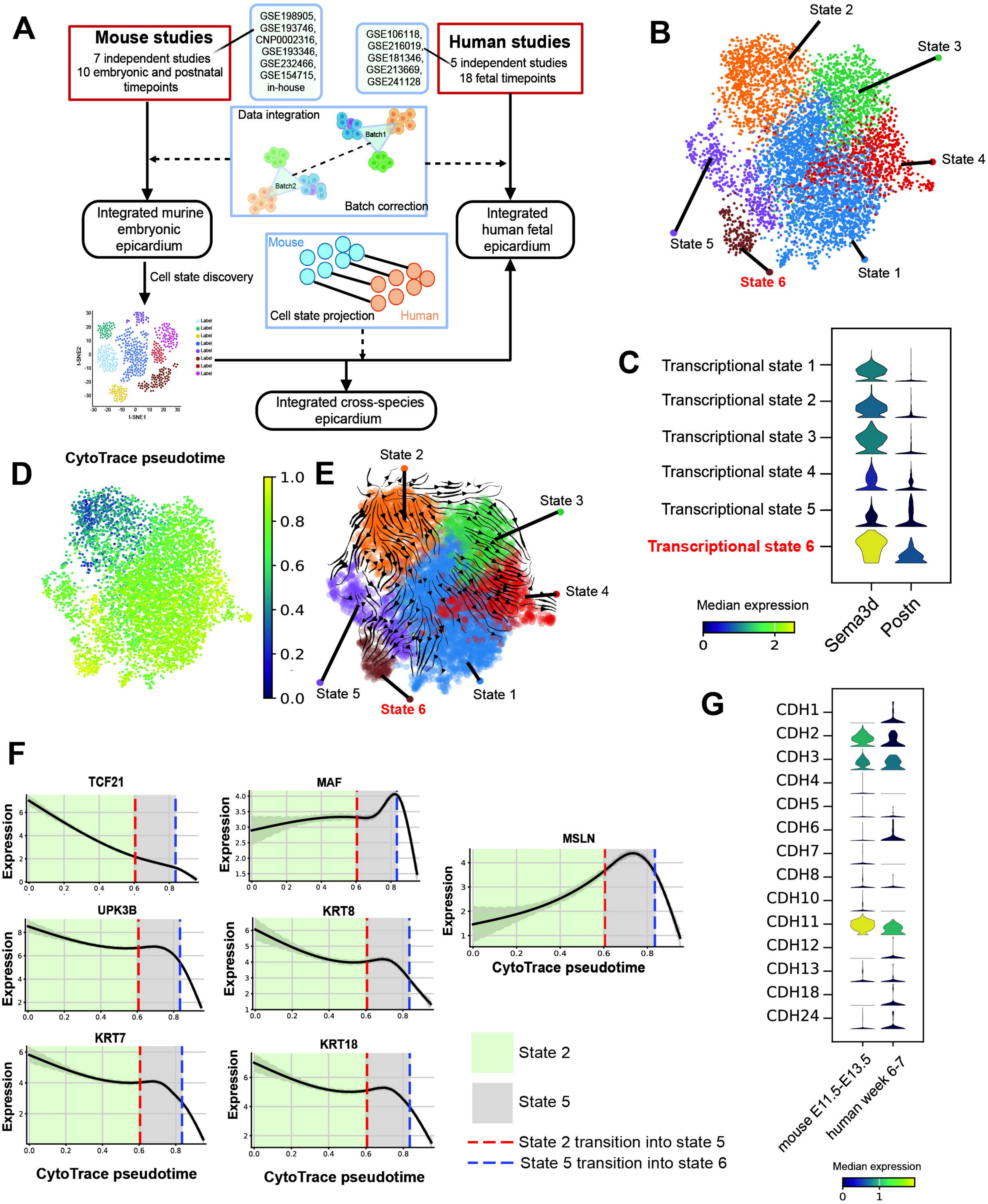

A
Upk3b
Mesothelium
0.25
0.20
0.15
0.10
0.05
B
E13.5 epicardium (Dataset 1)
E13.5 epicardium (Dataset 2)
81067
57924
9031
C
E13.5 mouse epicardium (Dataset1)
E13.5 mouse epicardium (Dataset 2)
Pearson correlation coefficient: 0.8605
E13.5 mouse epicardium (Dataset 2)
E13.5 mouse epicardium (Dataset 1)
D
E13.5 mouse epicardium (Dataset 2)
Promoters
Enhancers
promoters
enhancers
-0.5
center
0.5Kb
gene distance (bp)
E
Heart
Pancreas
Lung
65326
12897
15023
44602
15495
7135
14044
F
Heart
Lung
Pancreas
G
Heart
Lung
Pancreas
H
Heart
Lung
Pancreas
Upk3b
Msln
Twist1
Tbx20
Pnoc
Hhip
Gli1
Igf1
Hoxb8
Hoxb6
Pitx2
Barx1
Fraction of cells in group (%)
Mean expression in group
I
Pnoc
Wnt2+ mesenchyme
AT2 cells
Smooth muscle cells
Mesothelium
Endothelial cells
T-lymphocytes
J
Pitx2
Mesothelium
Quiescent stellate cells
Activated stellate cells
T lymphocytes
Endocrine 2
Islet endothelium
Neutrophils
Ductal
Acinar 2
Acinar
Neural crest
Endocrine progenitors
Endocrine
Fev+ cells
K
Barx1


A
Heart
eRegulon specificity score
Twist1 extended +/+ (147 genes)
Rbpj direct +/+ (34 genes)
Tbx20 direct +/+ (249 genes)
eRegulon rank
B
Lung
Foxl1 extended +/+ (53 genes)
Gli1 direct +/+ (178 genes)
Foxf1 direct +/+ (570 genes)
Foxp2 direct +/+ (343 genes)
C
Pancreas
Isl1 extended +/+ (102 genes)
Hoxb4 direct +/+ (198 genes)
Hoxb8 direct +/+ (137 genes)
Pitx2 direct +/+ (250 genes)
D
Lung vs Heart mesothelium
Tbx20
Foxp2
Foxf1
Wt1
−Log10 P
fold change
E
Pancreas vs Heart mesothelium
Tbx20
Hoxb8
Hoxb6
Pitx2
Isl1
Wt1
total = 541 expressed transcription factors
NS
Log2 FC
p−value
p − value and log2 FC
F
Tbx20 eRegulon
small GTPase mediated signal transduction
heart morphogenesis
Ras protein signal transduction
cell-substrate adhesion
cardiac chamber development
cell-matrix adhesion
cardiac chamber morphogenesis
cardiac muscle tissue morphogenesis
muscle tissue morphogenesis
ventricular cardiac muscle tissue development
GeneRatio
Count
p.adjust
G
Heart
Lung
Pancreas
Tbx20
Csf1
Median expression
0.0
0.5
H
Normalized signal (range 0 – 150)
Heart
Lung
Pancreas
Histone enrichment
H3K27ac heart meso
H3K4me3 heart meso
chr3 position (bp)
Genes
Gm40121
Csf1
ABC inks
score
I
chr3: 107,793,606 - 107,794,106 (+)
500 bp
Heart
Lung
Pancreas
H3k27ac
H3k4me3
Tbx20
J
DNA motif: TBX20
Probability
Position


A
chr17: 25,754,077 - 25,754,577 (+)
500 bp
Msln
Heart
Lung
Pancreas
H3K27ac
H3K4me3
Nfe2l2
B
NRF2-linked region I
NRF2-linked region II
Heart
Lung
Pancreas
H3K27ac heart meso
H3K4me3 heart meso
Normalized signal (range 0 – 130)
Histone enrichment
Msln
Gene
ABC links
score
chr17 position (bp)
C
chr17: 25,755,800 - 25,756,200 (+)
400 bp
Heart
Lung
Pancreas
H3K27ac
H3K4me3
Wt1
D
chr17: 25,756,700 - 25,757,300 (+)
600 bp
Heart
Lung
Pancreas
H3K27ac
H3K4me3
Nfe2l2
Wt1
Heart mesothelium
E
Msln
E11.5
E13.5
E17.5
F
Nfe2l2
G
E11.5
E13.5
E17.5
Normalized signal (range 0 – 330)
Gene
ABC links
Pancreas mesothelium
H
E12.5
E13.5
E14.5
E17.5
I
J
E13.5
E14.5
E17.5
Lung mesothelium
K
E12.5
E13.5
E15.5
E17.5
P3
L
M
E13.5
P3
Normalized signal (range 0 – 36)
Genes


A
epicardium
ventricular CM
vascular EC
B
Myl2
Upk3b
Fabp4
C
Maf direct +/+ (209 genes)
Tcf21 extended +/+ (131 genes)
eRegulon specificity score
eRegulon rank
Heart mesothelium
D
Tbx5 extended +/+ (26 genes)
Tbx5 direct +/+ (67 genes)
Esrrg direct +/+ (172 genes)
Ventricular CM
E
Sox18 extended +/+ (319 genes)
Sox18 direct +/+ (389 genes)
Sox17 direct +/+ (583 genes)
Cardiac vascular EC
F
Region-to-gene rho: 0.536
chr11: 21,563,578 - 21,564,078 (+)
chr11: 21,571,684 - 21,572,184 (+)
500 bp
Mdh1
Heart meso
Lung meso
Pancreas meso
Vascular EC
Ventricular CM
H3K27ac whole heart
Esrrg
TF-to-gene rho: 0.814
DNA motif: ESRRG
Probability
Position
Heart mesothelium
Lung mesothelium
Pancreas mesothelium
CM
Vascular EC
G
Region-to-gene rho: 0.368
chr16: 18,775,551 - 18,776,051 (+)
chr16: 18,776,597 - 18,777,097 (+)
Cldn5
Sox18
TF-to-gene rho: 0.972
DNA motif: SOX18


A
Krt18
Heart meso
Lung meso
Pancreas meso
CM
Vascular EC
C
chr15: 102,016,032 - 102,016,532 (+)
500 bp
Epicardium
EPDC
Maf
D
Krt8
F
chr15: 102,005,234 - 102,005,734 (+)
G
Krt18
Krt8
Maf
Pseudotime
expression
B
Normalized signal (range 0 – 86)
Heart meso
Lung meso
Pancreas meso
Vascular EC
Ventricular CM
Histone enrichment
H3K27ac heart meso
H3K4me3 heart meso
chr15 position (bp)
Genes
Gm41404
Krt18
ABC links
score
E
Normalized signal (range 0 – 76)
Gm36026
Heart meso/Epicardium
EPDC
CFb fate
VIC fate
Pericyte fate
Heart meso transition into EPDC
EPDC transition into CFb
EPDC transition into VIC
EPDC transition into Pericyte
H
Krt7
J
chr15: 101,396,526 - 101,397,026 (+)
Tcf21
L
Upk3b
N
chr5: 136,019,797 - 136,020,297 (+)
Dtx2
I
M
chr5 position (bp)
Gm671
Gm35853
K
O
chr5: 136,038,087 - 136,039,339 (+)
1252 bp
Wt1
P
Epi to EPDC
EPDC to CFb
EPDC to VIC
EPDC to Pericyte


A
Mouse studies
7 independent studies
10 embryonic and postnatal timepoints
GSE198905, GSE193746, CNP0002316, GSE193346, GSE232466, GSE154715, in-house
GSE106118, GSE216019, GSE181346, GSE213669, GSE241128
Human studies
5 independent studies
18 fetal timepoints
Data integration
Batch1
Batch2
Batch correction
Integrated murine embryonic epicardium
Integrated human fetal epicardium
Cell state discovery
Mouse
Human
Cell state projection
Integrated cross-species epicardium
t-SNE1
t-SNE2
B
State 1
State 2
State 3
State 4
State 5
State 6
C
Transcriptional state 1
Transcriptional state 2
Transcriptional state 3
Transcriptional state 4
Transcriptional state 5
Transcriptional state 6
Median expression
0
2
Sema3d
Postn
D
CytoTrace pseudotime
1.0
0.8
0.6
0.4
0.2
0.0
E
State 1
State 2
State 3
State 4
State 5
State 6
F
TCF21
MAF
MSLN
UPK3B
KRT8
KRT7
KRT18
Expression
CytoTrace pseudotime
State 2
State 5
State 2 transition into state 5
State 5 transition into state 6
G
CDH1
CDH2
CDH3
CDH4
CDH5
CDH6
CDH7
CDH8
CDH10
CDH11
CDH12
CDH13
CDH18
CDH24
mouse E11.5-E13.5
human week 6-7
Median expression
0
1