Single cell transcriptomics reveals a signaling roadmap coordinating endoderm and mesoderm lineage diversification during foregut organogenesis

Lu Han[#1], Hiroyuki Koike[#2], Praneet Chaturvedi[#1], Keishi Kishimoto[1,4,5], Kentaro Iwasawa[2], Kirsten Giesbrecht[2], Phillip C Witcher[1], Alexandra Eicher[1], Talia Nasr[1], Lauren Haines[1], John M Shannon[3] Mitsuru Morimoto[4,5], James M Wells[1], Takanori Takebe[2], Aaron M Zorn[$1]

Center for Stem Cell & Organoid Medicine (CuSTOM), Perinatal Institute, Divisions of [1]Developmental Biology, [2]Gastroenterology and [3]Pulmonary Biology, Cincinnati Children's Hospital, Cincinnati OH, USA 45229. Department of Pediatrics University of Cincinnati College of Medicine. 4 Laboratory for Lung Development, RIKEN Center for Biosystems Dynamics Research (BDR), Kobe, 650-0047, Japan 5 CuSTOM-RIKEN BDR collaborative laboratory, Cincinnati OH, USA 45229.

# equal contribution

$ correspondence: Aaron.zorn@cchmc.org;

Visceral organs, such as the lungs, stomach, liver and pancreas, derive from the fetal foregut through a series of inductive interactions between the definitive endoderm (DE) epithelium and the surrounding splanchnic mesoderm (SM) [1,2]. This foregut patterning, which occurs between embryonic day (E) 8.5 and E9.5 in the mouse embryo, equivalent to 17-23 days of human gestation, defines the landscape of the thoracic cavity and disruptions in this process can lead to severe congenital defects. While patterning of the endoderm lineages has been fairly well studies, the SM which is known to provide many paracrine factors required for organogeneis is virtually unstudied [1,2]. In particular we lack a comprehensive understanding of the molecular nature of SM regional identity, the mechanisms by which SM signaling boundaries are established, the role of the epithelium in SM patterning and how SM and DE lineages are dynamically coordinated during organogenesis. Here we used single cell transcriptomics to generate a high-resolution expression map of the embryonic mouse foregut. This uncovered an unexpected diversity in SM progenitors that developed in close register with the organ-specific epithelium. These data allowed us to infer a spatial and temporal signaling roadmap of the combinatorial endoderm-mesoderm interactions that orchestrate foregut organogenesis. We validated key predictions with mouse genetics, showing importance of epithelial signaling in mesoderm patterning. Finally, we leveraged the signaling road map to generate different SM subtypes from human pluripotent stem cells (hPSCs), which previously have been elusive.

Single cell transcriptomics allows examination of organogenesis at an unprecedented resolution [3-6] , however, to date studies have either taken a broad overview of many organ systems or focused only on the epithelial component of the developing gut [7-9]. To comprehensively characterize the signaling networks orchestrating foregut organogenesis, we performed single cell RNA sequencing (scRNA-seq) of the mouse embryonic foregut at three time points that span early patterning through organ lineage induction: Embryonic day (E) E8.5 (5-10 somites, ‘s’), E9.0 (12-15s) and E9.5 (25-30s) (Fig. 1a, b). We micro-dissected the foregut between the posterior pharynx and the midgut, pooling tissue from 15-20 embryos for each time point. At E9.5, we isolated anterior and posterior regions separately, containing lung/esophagus and liver/pancreas primordia, respectively. A total of 31,268 single-cell transcriptomes passed quality control measures with an average read depth of 3,178 transcripts/cell. Cells were clustered based on the expression of highly variable genes across the population and visualized using t-distributed stochastic neighbor embedding (t-SNE) (Fig. 1c; Supplementary Fig. 1). This identified 9 major

cell lineages: DE, SM, cardiac, other mesoderm (somatic and paraxial), endothelium, blood, ectoderm, neural crest and extraembryonic. DE clusters (4,448 cells) were characterized by co-expression of *Foxa1/2, Cdh1* and/or *Epcam,* whereas SM (10,097 cells) was defined by co-expression of *Foxf1* (Fig. 1d), *Vim* and/or *Pdgfra* as well as being negative for cardiac and other mesoderm specific transcripts.

To comprehensively characterize lineage diversification in the foregut we analyzed the DE and SM transcriptomes separately, defining 26 DE (E8.5, 6 clusters; E9.0, 8 clusters; E9.5, 12 clusters) and 36 SM clusters (E8.5, 7 clusters; E9.0, 12 clusters; E9.5, 17 clusters), which were annotated by comparing their distinguishing genes with published expression patterns of over 160 genes in the Mouse Genome Informatics (MGI) database [10]. This identified all the previously known DE lineages (Fig. 1e) [1, 2, 9, 11] and revealed a surprising number of distinct SM cell types, indicating previously unappreciated mesenchymal diversity in the early foregut (Fig. 1f; Supplementary Fig. S1). Unlike the DE, SM populations were typically defined not just by one or two markers, but rather by the combination of multiple transcripts (Supplementary Table S1). For example, *in situ* hybridization of E9.5 foreguts showed that different combinations of *Nkx6-1, Wnt2*, *Osr1*, *Barx1*, *Hand1* and *Hand2* co-expression defined the esophageal, respiratory, gastric, hepatic and intestinal mesenchyme (Fig. 1k-p).

At E8.5 and E9.0 the level of regional pattern and diversity of SM populations was surprisingly complex and mirrored that of the spatially restricted DE progenitor domains (Supplementary Fig. S1). By E9.5 we identified 17 SM cell populations (Fig. 1f) including five *Tbx1*+ pharyngeal clusters as well as *Nkx6-1/Wnt2*+ respiratory. A *Msc/Wnt4*+ esophageal mesenchyme. *Mtus2*+ cardiac outflow tract cluster had a transcriptional signature resembling both cardiac and respiratory mesenchyme. The liver, the first organ to differentiate in the gut tube, had an unexpected number of SM subtypes. Unlike other GI organs which form by epithelium evagination, the hepatic endoderm delaminates and invades the adjacent septum transversum mesenchyme (stm), an ill-defined mesodermal population that gives rise to the hepatic stellate cells, stromal fibroblasts, perivascular smooth muscle and mesothelium [12]. Consistent with this, the liver bud had five distinct mesenchyme populations; *Wt1/Lhx2/Alcam*-enriched stm, *Vsnl1*+ sinus venosus, a *Msx/Hand1/Col1a1*-enriched fibroblast population and two *Wt1*/*Lhx9/Uroplakin*+ mesothelium populations. Finally, we annotated two *Barx1*+ stomach mesenchyme populations and *Hoxc8*+ duodenum mesenchyme. We were unable to identify pancreas-specific mesenchyme and suspect that these cells were part of the stomach or duodenum clusters.

DE organ lineages have historically been defined by the overlapping expression domains of a few key transcription factors (TFs) [1, 2, 11]. While some regionally expressed TFs have been reported in the SM, the single cell data allowed us to define a comprehensive combinatorial code of differentially expressed TFs that distinguish SM and DE subtypes (Supplementary Fig. 1o & Table S1). This revealed new lineage-restricted markers such as homeodomain TF Nkx6-1. Well known for its expression in the pancreatic endoderm at E10.5 [13, 14], Nkx6-1 was also specifically expressed in the respiratory and esophageal mesoderm at E9.5 (Fig. 1k). This TF code will facilitate lineage tracing experiments and prompt studies testing their role in mesenchymal differentiation.

Different organs form at precise locations along the anterior-posterior (A-P) axis of the gut. To assess whether this was reflected in the single cell transcriptional profiles, we inferred the pseudo-spatial ordering of the cells by plotting the diffusion pseudo-time of each cluster (see Methods). Both the DE and SM cell populations were indeed appropriately ordered along the A-P axis (Fig. 1g-h; Supplementary Fig.1) [6]. E9.5 clusters from the anterior dissections were located in the anterior half of the pseudo-space continuum, compared to posterior tissue, confirming the robustness of the computational ordering. Finally, *Hox* genes are expressed in a co-linear fashion along the A-P axis and accordingly we observed a progressive increase in posterior *Hox* paralogs expression in more posterior clusters (not shown). Combining the pseudo-space analysis, MGI curations and in situ validation, we were able to map each DE and SM population to their approximate location in the gut tube (Fig. 1i, j; Supplementary Fig. 1). This revealed that the SM diversity mirrored DE lineages, indicating that they are closely coordinated from the very beginning of organogenesis.

The transcriptional cell state complexity of the DE and SM doubled in just 24 hours between E8.5 and E9.5, reflecting progenitors forming more specialized cell types. To examine the temporal dynamics of lineage diversification we visualized the single cell data using SPRING (Fig. 2a, b), an algorithm that represents k-nearest neighbors in a force directed graph, facilitating analysis of developmental trajectories[15]. Both the DE and SM trajectories progressed from a continuum of closely related expression states at E8.5 to transcriptionally distinct cell populations at E9.5 (Fig. 2a, b; Supplementary Fig. S2a-d), consistent with the transition from multipotent progenitors to organ specific lineages. Importantly the cell clusters defined by tSNE were well-preserved in SPRING (Supplementary Fig. 2a-d), supporting the robustness of the clustering.

To more clearly visualize the developmental trajectories associated with lineage diversification, we generated a consensus cell state tree of the using a single cell voting method [3], where each cell of a later time point votes for its most likely parent of the previous time point based on gene expression similarity (Fig. 2c, d, Supplementary Fig. S2e, f). While we cannot rule out SM migration bringing distant cell types to a given organ the data support the notion of transcriptionally related cell states arising from the subdivision of common progenitor populations. Given that our time points were generated from pooled embryos of slightly different ages, it was possible that parent-child relationships could exist at a given time. To address this and confirm the single cell voting tree, we assessed each trajectory with a pseudotime analysis (Monocle[4]) that computationally predicts progenitor states in a cell population. In general, the pseudotime analysis agreed with the single cell voting. But in the case of the liver endoderm, Monocle predicted a parent-child relationship within E9.0, where *Hhex*+ posterior foregut endoderm (cluster e_b2) gives rise to both *Prox1/Afp*+ hepatoblasts (e_b5) and *Prox1/Sox17/Pdx1*+ pancreatobiliary progenitors (e_b7) (Supplementary Fig. 2g), consistent with lineage tracing experiments [16, 17]. Paralleling the specialization in the liver epithelium, the trajectory allows mapping of five distinct liver SM-cells, consistent with the liver being the first organ to differentiate.

Overall the DE trajectories are consistent with experimentally determined fate maps [16-18], demonstrating the robustness of our analysis and suggesting that the SM lineages, which have not been well defined, are also correctly. Interestingly the DE and SM trajectories were highly synchronized. For example, at E8.5 the DE lateral foregut population (e_a2) and the spatially neighboring SM population (m_a0) both express the TF *Osr1*, and the trajectories suggests they are multipotent progenitors, giving rise to respiratory, esophageal and gastric epithelium and mesenchyme, respectively (Supplementary Fig. 2h, i). Furthermore, a close examination of the DE trachea cluster revealed a transitional cell population co-expressing *Nkx2-1* and *Sox2*, which was identified to be a rare cell population at the prospective tracheal-esophageal boundary (Fig. 2e-h). Together these analyses revealed novel progenitor populations and the close synchronization of DE and SM development.

We next sought to computationally predict the signaling microenvironment in the foregut that controls these cell fate decisions. We calculated metagene expression profiles for all the ligands, receptors and context-independent response genes in each DE and SM cluster for six major paracrine signaling pathways implicated in organogenesis: BMP, FGF, Hedgehog (HH),

Notch, retinoic acid (RA) and canonical Wnt (Methods & Supplementary Fig. S3, Table S2). Leveraging our spatial map of juxtaposed DE and SM populations in the foregut we then used the metagene expression levels to predict ligand-receptor pairs and the likelihood that a given cell population was responding to local paracrine or autocrine signals (Fig. 3a-c, Supplementary Fig. S4). In addition, we projected the signaling response-metagene expression levels onto the lineage trajectories which revealed spatiotemporally dynamic signaling domains that correlated with cell lineages (Fig. 3d, Supplementary Fig. S5). In general, the transcriptome data predicts local interactions, with the SM being the primary source of BMP, FGF, RA and Wnt ligands signaling to both the adjacent DE and within the SM itself (Fig. 3c). In contrast HH ligands are produced by the DE and signal to the gut tube SM, with no evidence of autocrine activity in the DE (Fig. 3c). The computational predictions identified most known signaling interactions controlling DE lineage specification such as mesoderm derived BMP, FGF and Wnt promoting DE liver and lung fate, and autocrine notch signaling in the DE endocrine pancreas [1, 2, 19, 20]. This suggesting that previously undefined SM signaling predictions are also likely to regulate cell fate. Combining the data for all six signaling pathways onto the cell state trees, we generated a comprehensive roadmap of the combinatorial signals predicted to coordinate the temporal and spatial development of each DE and SM lineage (Fig. 3e,f).

To genetically test the predictive value of the signaling roadmap, we focused on HH activity, which the scRNA-seq suggests is high in gut tube SM (esophagus, respiratory, stomach and duodenum) but low in the pharyngeal and liver SM (Fig. 4a-c). HH ligands stimulate the activation of Gli2 and Gli3 TFs, which in turn promote the transcription of HH-target genes (e.g. *Gli1)*[21]. As expected mouse embryo sections confirmed that *Shh* ligand was expressed in the gut tube DE with high levels of *Gli1-LacZ* expression in the adjacent SM. By contrast hepatoblasts did not express *Shh*[22] and the liver SM had very few if any *Gli1-LacZ* positive cells (Fig. 4d). To comprehensively define the impact of HH on SM development, we performed bulk RNA-seq on foreguts from *Gli2-/-;Gli3-/-* double mutant embryos, which lack all HH activity and fail to specify respiratory fate [20]. Comparing homozygous mutants to heterozygous littermates, we identified 156 HH/Gli-regulated transcripts (Fig. 4e; Supplementary Table. S3). Examining the expression of these transcripts in the scRNA-seq data revealed that most were localized to the SM. Transcripts downregulated in Gli2/3-mutants (n=80) were normally enriched in the gut tube SM, whereas upregulated transcript (n=76) were normally enriched in the liver or pharyngeal SM (Fig. 4e-g). Interestingly HH/Gli-regulated transcripts encoded, BMP, FGF and Wnt signaling components, as well as 28 TFs, including downregulated TFs (Osr1, Tbx4/5, Foxf1/2) and upregulated TFs

(Tbx18, Lhx2 and Wt1) that have been implicated in respiratory and liver development respectively (Fig. 3e; supplementary Table S3) [23]. This genetic analysis confirmed the predictive value of the signaling roadmap where differential HH activity promotes gut tube versus liver and pharyngeal SM, in part by regulating other lineage specifying TFs and signaling proteins.

Our data, together with previous work, suggest a model where the reciprocal epithelial-mesenchymal signaling network coordinates DE and SM lineages during organogenesis. In this model SM-derived RA induces a local region of Shh in the DE by E9.0 [20], which then signals back to the SM establishing broad pharynx, gut tube and liver domains. Other SM ligands (BMP, FGF, Notch, RA and Wnt), with regionally restricted expression in these 3 broad domains then progressively subdivide DE and SM progenitor populations in a coordinated manner. In the future it will be important to test this model by cell-specific genetic manipulations.

We next tested whether our new SM markers and signaling roadmap could be used to direct the differentiation of distinct SM subtypes from human pluripotent stem cells (hPSC), which to date have been elusive. Previous studies have established protocols to differentiate hPSC into lateral plate mesoderm (lpm), the precursors of SM, and into cardiac tissue [24]. The single cell data suggested that, in mouse the early SM experiences more RA signaling than the early cardiac mesoderm, which was confirmed by RA-responsive *RARE:lacZ* transgene expression in embryos (Supplementary Fig. S6). Addition of RA to the lpm cultures on days (d) 2-4 down-regulated the cardiac markers *NKX2-5/ISL1/TBX20* and promoted the SM markers *FOXF1/HOXA1*and *HOXA5* (Fig. 5b, Supplemental Fig S6), as predicted. We next treated SM cultures with different combinations of HH, RA, Wnt and BMP agonists or antagonist from d4-d7 (Fig. 5a), to drive distinct SM-like lineages based on the roadmap. As predicted the HH-agonist promoted respiratory and gut tube identity and efficiently blocked liver fates. In the HH treated cultures addition of retinoic acid plus BMP4 (RA/BMP4) RA/BMP, followed by WNT on d6-7 promoted expression of respiratory mesenchymal markers (*NKX6-1/TBX5/WNT2*), whereas RA and BMP4-antagoist on d6-7 promoted esophageal/gastric identity (*MSC/BARX1/WNT4/NKX3-2*) (Fig. 5b, c; Supplementary Fig. S6). In the absence of HH agonist, cells treated with RA/BMP expressed liver-stm markers (*WT1/TBX18/LHX2/UPK1B/3B*), whereas RA/BMP4/WNT treated cells expressed liver-fibroblast markers (HAND1/*MSX1/2*). These data provide a proof of principle that manipulating combinatorial signaling in PSC-derived lpm cultures to mimic the microenvironment of the foregut can direct the differentiation of organ specific SM subtypes based on the prediction of the signaling roadmap.

Most previous efforts to direct differentiation of human PSCs into specific cell and tissue types have depended on combining data from many published papers describing how individual pathways control each developmental step. Here, in one study, using single cell transcriptomics we inferred a roadmap for the entire foregut revealing an unexpected diversity of early mesenchymal lineages that are synchronized with the epithelial identity. Moreover, this roadmap predicts the combinatorial signaling microenvironment orchestrating foregut organogenesis which we leveraged to direct the formation of organ-specific mesenchyme from hPSCs; an important step towards engineering complex foregut tissue for regenerative medicine. A similar approach could be used to identify cell-cell interactions in other organ systems or those that drive pathological states such as the cancer niche.

Acknowledgements

This work was supported by grant NICHD P01HD093363 to AMZ and JWM. TT is a New York Stem Cell Foundation – Robertson Investigator. MM research is supported by a Japanese Grant-in-Aid for Scientific Research (B)(17H04185). KK is supported by a Memorial Foundation postdoctoral fellowship and a RIKEN-CuSTOM collaborative grant for Promotion of Joint International Research (A) (18KK0423). This project was supported in part by NIH P30 DK078392 (Integrative Morphology, Sequencing and Pluripotent Stem Cell and Organoid Cores) of the Digestive Disease Research Core Center in Cincinnati. We are grateful to James Briggs and Caleb Weinreb from the Alon Klein lab for advice on SPRING and members of the Rahul Satija lab for help with Seurat. Drs. Rafi Kopan, Yan Xu, Minzhe Guo, Emily Miraldi and Nathan Salomonis provided critical discussions and advice.

Author Contributions

AMZ, TT, JMW and HK conceived the project. JMS and HK prepared the mouse foregut single cell samples. PC performed the computational analysis with input from AMZ, LH, HK and KG. AMZ, LH and HK performed manual data curation. LH, TN, KI and PW and performed mouse embryo validation. LH, JMW and AMZ devised the differentiation protocol. LH, KK, KI, AE and LH performed the hPSC differentiation experiments. MM and KK shared unpublished data on mouse ESC differentiations. AMZ, LH and PC wrote the manuscript with input from the authors.

Methods

Embryo collection and single cell dissociation

All mouse experiments were performed in accordance with protocols approved by the Cincinnati Children's Hospital Medical Center Institutional Animal Care and Use Committee (IACUC). No statistical sample size estimates were performed prior to the experiment, sufficient embryos to generate the material need for the experiments were used. No randomization was utilized as no particular treatment was performed in different groups. Timed matings were set up between C57BL/6J mice and the day where a plug was detected was considered embryonic day 0.5. Staging was validated by counting somite numbers E8.5 (5-10 somites; "s"), E9.0 (12-15s) and E9.5 (25-30s) (Fig. 1a, b). We micro-dissected the foregut between the posterior pharynx and the midgut, removing most of the heart and paraxial tissue and excluded the thyroid. At E9.5, we isolated anterior and posterior regions separately, containing lung/esophagus and liver/pancreas primordia, respectively. We pooled dissected foregut tissue from 16 , 20, 18 and 15 embryos for E8.5, E9.0 and E9.5 anterior and E9.5 posterior, respectively isolated from 2-3 litters.

Single cell dissociation by cold active protease protocol was performed as described previously[25]. Rapidly dissected C57BL/6J mouse embryo tissues were transferred to ice-cold PBS with 5 mM $CaCl_2$, 10 mg/ml of *Bacillus Licheniformis* protease (Sigma) and 125 U/ml DNAse (Qiagen) and incubated on ice with mixing by pipet. After 7 min, single cell dissociation was confirmed with microscope. Cells were then transferred to a 15 ml conical tube, and 3ml ice cold PBS with 10 % FBS (FBS/PBS) was added. Cells were pelleted (1200 G for 5 min), and resuspended in 2 ml PBS/FBS. Cells were washed three times in 5 ml PBS/0.01%BSA (PBS/BSA) and resuspended in a final cell concentration of 100,000 cells/ml for scRNAseq. Single cell suspensions of each stage were loaded onto the Chromium Single Cell Controller instrument (10x Genomics) to generate single-cell gel beads in emulsion. Single cell RNA-Seq libraries for high-throughput sequencing were prepared using the Chromium Single Cell 5' Library and Gel Bead Kit (10x Genomics). All samples were multiplexed together and sequenced in an Illumina HiSeq 2500. The technician was blinded during the RNA extraction, library preparation and sequencing.

Immunofluorescence staining and whole mount *in situ* for dissected foreguts

Mouse embryos were harvested at indicated stages and fixed in 4% paraformaldehyde (PFA) at 4°C for overnight. The fixed samples were washed with PBS 10 min 3 times and the foreguts were micro-dissected when indicated. Embryos or dissected foreguts were then processed as described previously [26] by antibody staining (supplementary table.S4) or processed for in situ hybridization

Pre-processing 10x Genomics raw scRNA-seq data

Raw scRNA-Seq data was processed using cellranger (v2.0.0 10xGenomics) (https://github.com/10XGenomics/cellranger). Reads were aligned to mouse genome [mm10] to produce counts of genes across barcodes. Barcodes with less than ~5k UMI counts were not included in downstream analysis. Percentage of reads mapped to transcriptome was ~70% each sample. The resulting data comprised 9748 cells in E8.5, 9265 cells in E9.0 7208 cells in E9.5 anterior samples and 5085 cells in E9.5 posterior samples.

Quality control, dimensionality reduction, clustering and marker prediction

Subsequent QC, and clustering was performed using Seurat [v2.3.4] package in R[27, 28]. Basic filtering was carried out where all genes expressed $\geq$ 3 cells and all cells with at least 100 detected genes were included. QC was based on nGene and percent.mito parameters to remove the multiplets and cells with high mitochondrial gene expression using the following parameters for each sample [E8.5: [ngene(-Inf,7500), percent.mito(-Inf,0.15)], E9.0: [ngene(-Inf,7000), percent.mito(-Inf,0.15)] and E9.5: [ngene(-Inf,7500), percent.mito(-Inf,0.15)]]. After filtering 9748, 9265 and 12255 cells were retained in E8.5, E9.0 and E9.5 samples respectively. Global scaling was used to normalize counts across all cells in each sample [scale factor: 10000] and cell cycle effect was removed by regressing out difference between S phase and G2M phase from normalized data using default parameters. We first clustered each developmental stage separately to identify major cell lineages. Approximately 1500 Highly variable genes (HVG) across each population were selected by marking outliers from dispersion vs. avgExp plot. PCA was performed using HVG, and the first 20 Principal Components were used for cells clustering, which then was visualized using t-distributed stochastic neighbor embedding (tSNE). Marker genes defining each cluster were identified using 'FindAllMarkers' function (Wilcoxon Rank Sum Test) in Seurat and these were used to annotate clusters based on well-known cell type specific genes. Canonical correlation analysis (CCA) was used to visualize all of the cells from all stages together. Briefly, the HVGs from

latest stage [E9.5] and 30 CCs (canonical correlation components) were used for multi-CCA (integration analysis) and clustering was performed with default parameters.

In silico selection and clustering for definitive endoderm and splanchnic mesenchyme

Definitive Endoderm (DE) clusters (4,448 cells) were defined by the co-expression of *Foxa1/2, Cdh1* and/or *Epcam,* whereas the splanchnic SM (10,097 cells) were defined by co-expression of *Foxf1*, *Vim* and/or *Pdgfra* as well as being negative for cardiac, somatic and paraxial mesoderm specific transcripts. Cells from DE and SM clusters were extracted from each time point and re-clustered using Seurat [v2.3.4] to define lineage subtypes. Prior to re-clustering blood, mitochondrial, ribosomal and strain-dependent noncoding RNA genes were regressed from the data. QC parameters for sub-setting DE cells were [E8.5: [ngene(-Inf,5000), percent.mito(-Inf,1)], E9.0: [ngene(-Inf,5500), percent.mito(-Inf,1)] and E9.5: [ngene(-Inf,6000), percent.mito(-Inf,1)]] and SM cells were [E8.5: [ngene(-Inf,5000), percent.mito(-Inf,1)], E9.0: [ngene(-Inf,5000), percent.mito(-Inf,1)] and E9.5: [ngene(-Inf,6000), percent.mito(-Inf,1)]]. Dimensionality reduction, clustering and marker prediction steps were performed as described above. DE and SM cell subtypes were annotated by manual curation comparing the cluster marker genes with overpublished expression profiles of over 160 different genes in the MGI database[10] and our own gene expression validations.

Transcription factor code for mesoderm lineages

To identify TFs with enriched expression specific to different DE and SM cell types 'FindAllMarkers' function in Seurat [v3.0] was utilized on set of 1623 TFs expressed in the mouse genome [AnimalTFDB] [29]. Raw counts of TFs were normalized and scaled in Seurat [v3.0].Cells in clusters served as replicates in finding differential TFs for each lineage. Wilcoxon rank sum test was used in identifying differential TFs.

Pseudo-space ordering of cells

To computationally predict the ordering of cells along the anterior-poster axis, the density of marker gene expression in each cluster was plotted against pseudotime using URD [v1.0] [30] . The most anterior pharyngeal cluster was set as the root of X-axis of the pseudotime continuum. In order to calculate pseudotime, URD first obtains transition probabilities between cells to construct a diffusion map. Diffusion maps for several pair of dimensions help understanding the structure of differentiation. Next, the most anterior pharyngeal cluster was set as the root [cells in this cluster served as start point] and then a probabilistic breadth-first

graph search using transition probabilities was carried out. Breadth-first graph moves step-wise outward from root cells, until all the cells in the graph have been traversed through. Lastly, pseudotime equals average iteration that visited each cell, after running several simulations.

Analysis of cell trajectories

To examine cell trajectories across time, we implemented SPRING [v1.0] [15]. SPRING uses k-Nearest Neighbors (KNN) graph (5 nearest neighbors) to obtain force-directed layout of cells and their neighbors. To understand transcriptional change across cell states (lineages), first 40 principle components (PC) were learnt from the latest time point E9.5, and this PC space was used to transform the entire data set (E8.5, E9.0 and E9.5). This transformed data was used to generate a distance matrix to obtain the KNN graph using the default parameters.

To visualize the trajectories in a simple transcriptional cell state tree, we used a parent-child single cell voting approach based on the KNN classification algorithm, similar as previously described [3]. First, a normalized counts matrix was generated using the distinguishing marker genes from all DE or SM clusters as features. Marker genes were used as features to train KNN, during which the KNN learns the distance among cells in the training set based on the feature expression. Each cell was classified based on the clusters it was assigned. Cells of a later time point vote for their most likely parent cells in the earlier time point as follows: train KNN using E8.5 cells and test by E9.0 cells voting for E8.5 cells. KNN resulted in vote probability for each cell in E9.0 against each cluster in E8.5, which was subsequently averaged for each cluster in E9.0 against each cluster in E8.5. This approach was repeated with E9.5 cells voting for E9.0 parents. All of the average vote probability for a given cluster are tabulated, normalized for cluster size and represented as a % of total votes. The top winning votes linking later time points back to the preceding time point were displayed as a solid line on the tree. Prominent second choices with >60% of winning votes were reported on the tree as dashed lines. We also compared this vote probability with the confusion matrix resulting from the SPRING KNN to assess our transcriptional cell-state tree. In more than 99% cases, these two methods resulted in the same first and second choices, thereby validating deduced parent-child relationships.

To validate the SPRING trajectory and cell state tree using pseudotime analysis, Monocle [v3.0.0] [31-33] was deployed on individual lineages/cell states. tSNE was used for Dimensionality Reduction and principle graph was learnt using SimplePPT. All the other parameters were set to default. In general Monocle pseudotime replicated real time and we

found one instance of a strong parent-child relationship between cells of a given stage (Supplementary Fig. S2g).

Calculation of metagene profiles and

For six of the major paracrine signaling pathways implicated in foregut organogenesis (BMP, FGF, HH, Notch RA and canonical Wnt), we curated a list of all the well-established ligands, receptors and context-independent pathway response genes that were encoded in the mouse genome (supplementary table S2a -excel tab1). We then calculated a "*ligand-metagene"*, "*receptor-metagene*" and "*response-metagene*" profiles by summing the normalized expression of each individual gene for each pathway (e.g.: Wnt-ligand metagene =ΣWnt1+Wnt2+Wnt2b+Wnt3…Wnt10b expression) in each cell and cluster as follows:

Assuming that there are x genes in the gene set and n cells. Gene1 has (a1,a2…an) counts, Gene2 has (b1,b2…bn) counts and so on.

Step1: Each gene's counts were normalized using the gene's max count across all DE and SM cells (n=14,545 cells) : Gene1_norm = (a1,a2,..an)/max(a1,a2…an)

Step2: Normalized counts of genes were summed up, for each cell, to generate a metagene_v1 with counts across cells: metagene_v1 = Gene1_norm+ Gene2_norm+..+Genex_norm

Presuming summed up counts are: m1,m2,..mn

Step3: Summed counts of metagene_v1 were normalized by max counts of the metagene_v1, to create a meta gene profile for each cell : MetaGene = (m1,m2,..mn)/max(m1,m2…mn)

The average Metagene expression profiles for ligands, receptors and response genes in each DE and SM cluster were then calculated in Seurat [v3.0] using 'AverageExpression' function. The average expression profile of metagene across all DE and SM clusters were visualized as a Dotplot using Seurat (v3.0).

Prediction of receptor-ligand interactions

A given cell type was scored to be expressing enough ligand to send a signal or enough receptor to respond to ligand if the average ligand-metagene or receptor-metagene expression level was $\geq$ -1 and expressed in $\geq$ 25% of cells. (Except for the Notch ligand-metagene where expression threshold of $\geq$ -1.5 was used due to low overall expression in all cells). These thresholds empirically set to be conservative and benchmarked against experimentally validate signaling interactions in DE liver, lung and pancreas from the literature. Furthermore, we determined the likely hood that a given cell population was responding based on the context-independent

pathway response-metagene expression level being ≥ -1 and expressed in ≥ 25% of cells. Context-independent response genes are those genes that are known from the literature to be directly transcribed in most cell types that are responding to a ligand-receptor activation.

DE and SM cell clusters of each stage are ordered along the A-P axis consistent with the location of organ primordia *in vivo* with spatially adjacent DE and SM cell types across from one another in the diagram. To assign receptor-ligand interactions for each cell cluster we first determined if a given cluster was responding based on having response-metagene and receptor-metagene levels ≥ -1 threshold. If the responding cluster also expressed the ligand-metagene level ≥ -1, an autocrine signaling was established. For paracrine signaling, we then identified adjacent cell populations, within the same tissue layer and from the adjacent layer that expressed the ligand-metagene above the threshold and then established a receptor-ligand interaction (arrow). The signal strength was calculated as the sum of the ligand-metagene and the response-metagene values. If this value was ≥ 1, the signal was considered "strong".

Comparison of bulk RNA-Seq vs scRNA-Seq:

Foregut tissue was dissected from E9.5 double mutant *Gli2-/-;Gli3-/-* (n=3) and *Gli2+/-;Gli3+/-* heterozygous litter mate controls (n=3). Each dissected foregut was separately used for RNA extraction, library preparation and bulk RNA-Seq. These mice were of mixed strains, and the sex of the embryos were unknown. The CSBB [v3.0] (https://github.com/csbbcompbio/CSBB-v3.0) pipeline was used to aligned to mouse genome [mm10] and determine differentially expressed transcripts between the two gene types were obtained using RUVSeq (LogFC ≥ |1| and FDR ≤ 0.1). Differentially expressed (DE) genes were clustered using hierarchical clustering and visualized in Morpheus (https://software.broadinstitute.org/morpheus/) across samples.

To compare our bulk analysis to scRNA-Seq, we visualized the expression of DE genes across cells in scClusters. We utilized the 'DoHeatmap' function in Seurat [v3.0.0]. Cells were arranged according to Anterior/Posterior axis position of their respective clusters and genes were ordered as returned from the clustering order obtained above. We also performed Gene Set Enrichment Analysis (GSEA) [v3.0] [34] to examine statistical enrichment of the DE genes in the gut tube SM (respiratory, esophagus, gastric, duodenum), pharynx and liver SM clusters. Normalized counts of genes across cells and up/down-regulated genes from bulk RNA sequencing were used as custom gene sets to perform the GSEA analysis.

Maintenance of PSCs

Two hPSC lines were used in this study; 1) WA01-H1 human embryonic stem cells purchased from WiCell (NIH approval number NIHhESC-10-0043 and NIHhESC-10-0062) and 2) human iPSC72_3 generated by the CCHMC Pluripotent Stem Cell Facility. Both cell lines have been authenticated as follows: i) Cell identity; via STR profiling by Genetica DNA Laboratory (a LabCorp brand; Burlington, NC), ii) Genetic stability; by standard metaphase spread and G-banded karyotype analysis in CCHMC Cytogenetics Laboratory and iii) Functional pluripotency; cells were subjected to analysis of functional pluripotency by teratoma assay demonstrating ability to differentiate into each of the three germ layer. Both cell lines routinely tested negative for mycoplasma contamination. hPSC lines were maintained on feeder-free conditions in mTeSR1 medium (StemCell technologies, Vancouver, Canada) on six-well Nunclon surface plates (Nunc) coated with Geltrex (ThermoFisher Scientific) and maintained in mTESR1 media (Stem Cell Technologies) at 37°C with 5% CO_2. Cells were checked daily and differentiated cells were manually removed. Cells were passaged every 4 days using Dispase solution (ThermoFisher Scientific).

Differentiation of PSCs into mesenchyme

Differentiation of hPSCs into lateral plate mesoderm was induced using previously described methods [24] with modifications. In brief, partially confluent hPSCs were dissociated into very fine clumps in Accutase (Invitrogen) and passaged 1:18 onto new Geltrex-coated 24-well plates for immunocytochemistry and 12-well plates for RNA preparation in mTeSR1 with 1uM thiazovivin (Tocris) (Day -1). Next day, a brief wash with DMEM/F12 is followed with Day0 medium 30ng/ml Activin A (Cell Guidance Systems) + 40ng/ml BMP4 (R&D Systems) + 6uM CHIR99021 (Tocris) +20ng/ml FGF2 (Thermo Fisher Scientific) +100nM PIK90 (EMD Millipore) for 24 hours. A basal media composed of Advanced DMEM/F-12, N2, B27, 15 mM HEPES, 2 mM L-glutamine, penicillin-streptomycin is used for this and all subsequent differentiation. On Day 1, a brief wash with with DMEM/F12 is followed with Day 1 medium 1uM A8301 (Tocris) +30ng/ml BMP4+1uM C59 (Cellagen Technology) for 24 hours. For cardiac mesoderm generation, cells are cultured in 1uM A8301+30ng/ml BMP4+1uM C59+20ng/ml FGF2 from Day2 to Day4 (medium changed every day). From Day 4, cells are cultured in 200ug/ml 2-phospho-Ascorbic acid (Sigma) + 1uM XAV939 (Sigma) + 30ng/ml Bmp4 for 3 days. For splanchnic mesoderm generation, cells are cultured in 1uM A8301+30ng/ml BMP4+1uM C59+20ng/ml FGF2 + 2uM RA (Sigma) from Day2 to Day4 (medium changed every day). To further direct regional splanchnic mesoderm, either 2uM RA+40ng/ml BMP4 is used to promote STM fate for 3 days; 2uM RA+ 2uM PMA (Tocris) is

used for 2 days, and then 2uM RA+ 2uM PMA + 100ng/ml Noggin (R&D Systems) is used at the last 1 day to promote esophageal/gastric mesenchyme fate; 2uM RA+ 40ng/ml BMP4 + 2uM PMA is used for 2 days, and then 2uM RA+ 40ng/ml BMP4 + 2uM PMA + 1uM CHIR99021 is used at the last 1 day to promote respiratory mesenchyme fate. Medium was changed every day. Similar results were obtained with WA-01 hES cells and human iPSC 72_3.

Quantitative RT-PCR

Total RNA was prepared from differentiating human ES cells by using Nucleospin kit according to manufacturer's protocol. Reverse transcription PCR was performed by Superscript VILO cDNA synthesis kit. QuantStudio 5 and 6 were used for qPCR analyses. Primers for qPCR were listed in Supplementary Table 4. Statistics were performed with PRISM8 (GraphPad Software). Significance was determined by one-way ANOVA, followed by Tukey's test.

Immunocytochemistry

Cells were fixed with 4% PFA/PBS for 30min at room temperature. After perforation with 0.5% Triton X-100/PBS for 10 min, cells were incubated with 5% normal donkey serum for 2 hours. Cells were incubated with primary antibodies (listed in Supplementary Table 4) overnight at 4°C. Next day, cells were washed with PBS, and then incubated with secondary antibodies for 1 hour at room temperature.

Data and Code Availability

The scRNA-seq and bulk RNA-seq data are available at Gene Expression Omnibus (GEO): GSE136689 and GSE136687. All the code (scripts, R-packages and software) and their documentation has been uploaded to GitHub [https://github.com/ZornLab/Single-cell-transcriptomics-reveals-a-signaling-roadmap-coordinating-endoderm-and-mesoderm-lineage]. All the deposited code is available to use with GPLv3.0.

Literature

1. Zorn, A.M. & Wells, J.M. Vertebrate endoderm development and organ formation. *Annu Rev Cell Dev Biol* **25**, 221-251 (2009).
2. Kraus, M.R. & Grapin-Botton, A. Patterning and shaping the endoderm in vivo and in culture. *Curr Opin Genet Dev* **22**, 347-353 (2012).
3. Briggs, J.A. *et al.* The dynamics of gene expression in vertebrate embryogenesis at single-cell resolution. *Science* **360** (2018).
4. Cao, J. *et al.* The single-cell transcriptional landscape of mammalian organogenesis. *Nature* **566**, 496-502 (2019).

5. Ibarra-Soria, X. *et al.* Defining murine organogenesis at single-cell resolution reveals a role for the leukotriene pathway in regulating blood progenitor formation. *Nat Cell Biol* **20**, 127-134 (2018).
6. Pijuan-Sala, B. *et al.* A single-cell molecular map of mouse gastrulation and early organogenesis. *Nature* **566**, 490-495 (2019).
7. Baron, M. *et al.* A Single-Cell Transcriptomic Map of the Human and Mouse Pancreas Reveals Inter- and Intra-cell Population Structure. *Cell Syst* **3**, 346-360 e344 (2016).
8. Li, L.C. *et al.* Single-cell transcriptomic analyses reveal distinct dorsal/ventral pancreatic programs. *EMBO Rep* **19** (2018).
9. Nowotschin, S. *et al.* The emergent landscape of the mouse gut endoderm at single-cell resolution. *Nature* **569**, 361-367 (2019).
10. Bult, C.J. *et al.* Mouse Genome Database (MGD) 2019. *Nucleic Acids Res* **47**, D801-D806 (2019).
11. Sherwood, R.I., Chen, T.Y. & Melton, D.A. Transcriptional dynamics of endodermal organ formation. *Dev Dyn* **238**, 29-42 (2009).
12. Asahina, K., Zhou, B., Pu, W.T. & Tsukamoto, H. Septum transversum-derived mesothelium gives rise to hepatic stellate cells and perivascular mesenchymal cells in developing mouse liver. *Hepatology* **53**, 983-995 (2011).
13. Sander, M. *et al.* Homeobox gene Nkx6.1 lies downstream of Nkx2.2 in the major pathway of beta-cell formation in the pancreas. *Development* **127**, 5533-5540 (2000).
14. Pedersen, J.K. *et al.* Endodermal expression of Nkx6 genes depends differentially on Pdx1. *Dev Biol* **288**, 487-501 (2005).
15. Weinreb, C., Wolock, S. & Klein, A.M. SPRING: a kinetic interface for visualizing high dimensional single-cell expression data. *Bioinformatics* **34**, 1246-1248 (2018).
16. El Sebae, G.K. *et al.* Single-cell murine genetic fate mapping reveals bipotential hepatoblasts and novel multi-organ endoderm progenitors. *Development* **145** (2018).
17. Spence, J.R. *et al.* Sox17 regulates organ lineage segregation of ventral foregut progenitor cells. *Dev Cell* **17**, 62-74 (2009).
18. Franklin, V. *et al.* Regionalisation of the endoderm progenitors and morphogenesis of the gut portals of the mouse embryo. *Mech Dev* **125**, 587-600 (2008).
19. Zaret, K.S. From Endoderm to Liver Bud: Paradigms of Cell Type Specification and Tissue Morphogenesis. *Curr Top Dev Biol* **117**, 647-669 (2016).
20. Rankin, S.A. *et al.* A Retinoic Acid-Hedgehog Cascade Coordinates Mesoderm-Inducing Signals and Endoderm Competence during Lung Specification. *Cell Reports* **16**, 66-78 (2016).
21. Rubin, L.L. & de Sauvage, F.J. Targeting the Hedgehog pathway in cancer. *Nat Rev Drug Discov* **5**, 1026-1033 (2006).
22. Bort, R., Martinez-Barbera, J.P., Beddington, R.S. & Zaret, K.S. Hex homeobox gene-dependent tissue positioning is required for organogenesis of the ventral pancreas. *Development* **131**, 797-806 (2004).
23. Abe, T. & Fujimori, T. Reporter mouse lines for fluorescence imaging. *Dev Growth Differ* **55**, 390-405 (2013).
24. Loh, K.M. *et al.* Mapping the Pairwise Choices Leading from Pluripotency to Human Bone, Heart, and Other Mesoderm Cell Types. *Cell* **166**, 451-467 (2016).
25. Adam, M., Potter, A.S. & Potter, S.S. Psychrophilic proteases dramatically reduce single-cell RNA-seq artifacts: a molecular atlas of kidney development. *Development* **144**, 3625-3632 (2017).
26. Han, L. *et al.* Osr1 functions downstream of Hedgehog pathway to regulate foregut development. *Developmental biology* (2017).

27. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol* **36**, 411-420 (2018).
28. Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902 e1821 (2019).
29. Hu, H. *et al.* AnimalTFDB 3.0: a comprehensive resource for annotation and prediction of animal transcription factors. *Nucleic Acids Res* **47**, D33-D38 (2019).
30. Farrell, J.A. *et al.* Single-cell reconstruction of developmental trajectories during zebrafish embryogenesis. *Science* **360** (2018).
31. Qiu, X. *et al.* Single-cell mRNA quantification and differential analysis with Census. *Nat Methods* **14**, 309-315 (2017).
32. Qiu, X. *et al.* Reversed graph embedding resolves complex single-cell trajectories. *Nat Methods* **14**, 979-982 (2017).
33. Trapnell, C. *et al.* The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat Biotechnol* **32**, 381-386 (2014).
34. Mootha, V.K. *et al.* PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet* **34**, 267-273 (2003).

Fig1 single cell analysis foregut endoderm and splanchinc mesoderm lineages

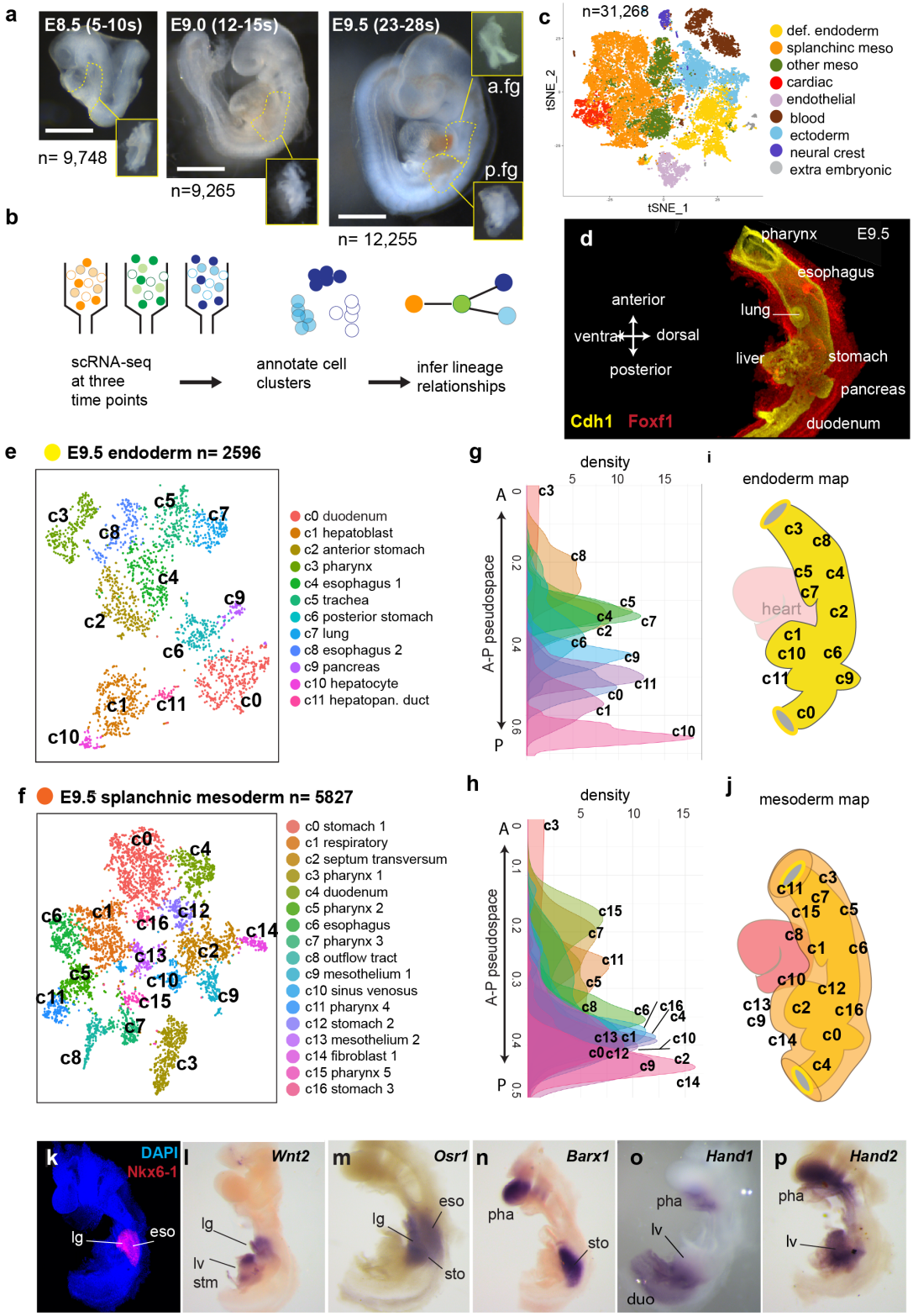


Fig. 1. Single cell analysis of the mouse foregut endoderm and mesoderm lineages. a, Representative mouse embryo images at three developmental stages showing the foregut region (dashed) that was microdissected (insets) to generate single cells. At E9.5 anterior foregut (a.fg) and posterior foregut (p.fg) were isolated separately. E, embryonic day; s, somite number; n, number of cells. Scale bar 1 mm. **b,** Schematic of the RNA-seq work flow. **c,** t-SNE visualization of 31,268 cells isolated from pooled samples of all three stages. Cells are colored based on major cell lineages. **d.** Whole-mount immunostaining of a dissected E9.5 mouse foregut, showing the juxtaposition of Cdh1+ endoderm and Foxf1+ splanchnic mesoderm. **e** and **f,** t-SNE plot of *in-silico* isolated E9.5 endoderm (**e**) and splanchnic mesoderm (**f**) cells. **g** and **h,** Pseudo-spatial ordering of E9.5 endoderm (**g**) and mesoderm (**h**) cells along the anterior-posterior (A-P) axis. **i** and **j,** Schematic showing the predicted location of E9.5 cell types mapped onto **i**) the embryonic mouse foregut endoderm (yellow) and **j**) mesoderm (orange)**. k-p,** Whole-mount immunostaining (**k**) or *in situ* hybridization (**l-p**) of dissected E9.5 foregut tissue revealed distinct expression of mesoderm markers. def. definitive; meso, mesoderm; lg, lung; eso, esophagus; lv, liver; stm, septum transversum mesenchyme; sto, stomach; pha, pharynx.

Fig. 2 lineage trajectory

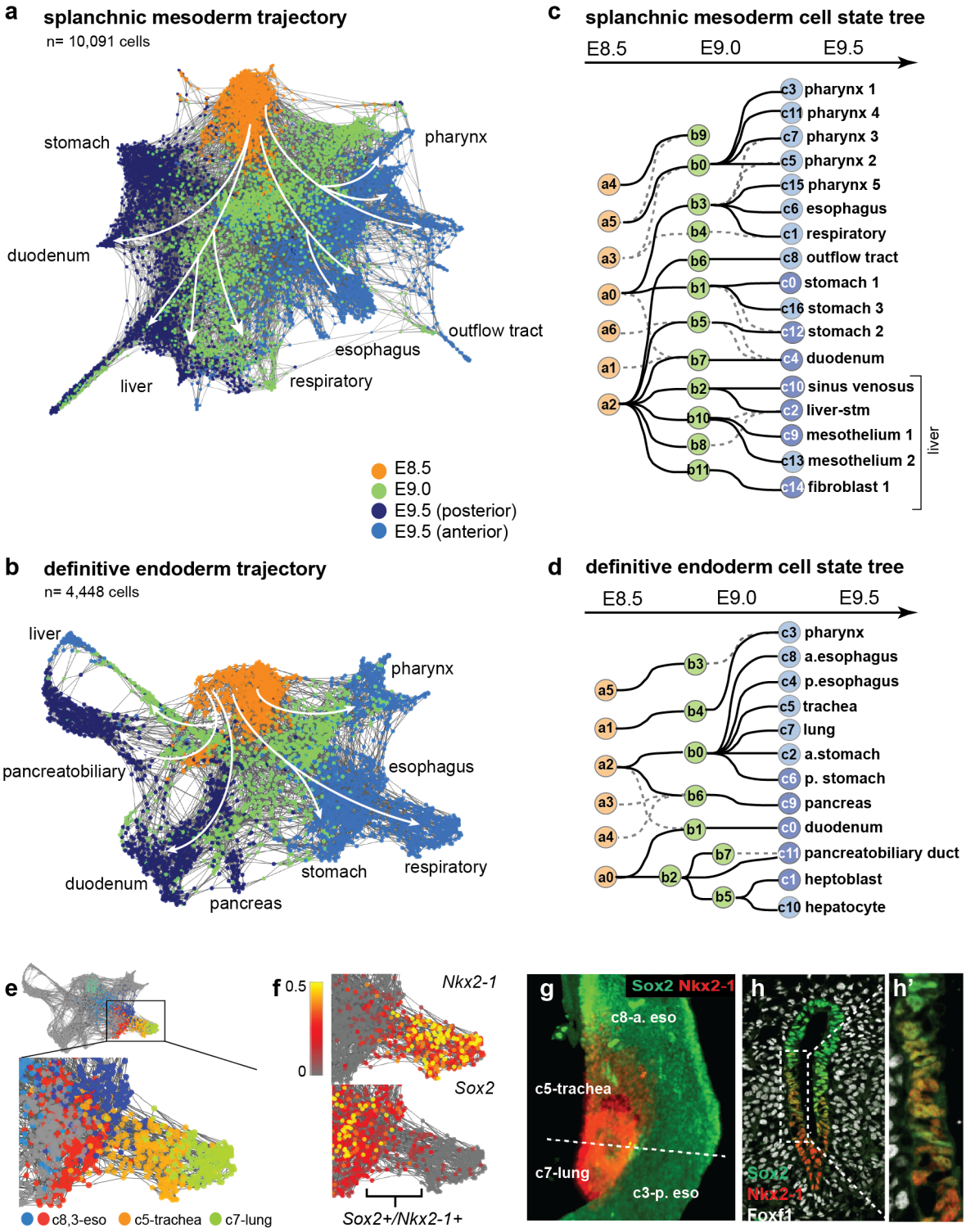


Fig. 2. Coordinated diversification of endoderm and mesoderm lineages. a and **b,** Force directed SPRING visualization of the **a**) splanchnic mesoderm (n=10,097 and **b**) definitive endoderm (**n**=4,448) cell trajectories. Cells are colored by developmental stage. White arrows indicate cell lineage progression. **c** and **d,** Transcriptional cell state trees of **c**) SM and **d**) DE lineages predicted by single cell voting (see Methods). The top choice linking cell states of sequential time points are indicated by solid lines, and prominent second choices are shown by dashed lines. Nodes are colored by stages and annotated with the cluster numbers. E8.5,clusters are designated “a”, E9.0; “b”, and E9.5; “c”. **e,** SPRING plot of the DE esophageal-respiratory lineages **f.** *Nkx2-1* and *Sox2* expressions are projected onto the SPRING plot, showing co-expression across the esophageal-tracheal boundary. **g,** Sox2 and Nkx2-1 whole mount immunostaining of a E9.5 mouse foregut. **h**, Sox2, Nkx2-1and Foxf1 immunostaining of a transverse E9.5 foregut section, confirming a rare population of Sox2/Nkx2-1 co-expressing cells. **h',** Higher magnification of box in h .

Fig 3 Signaling road map

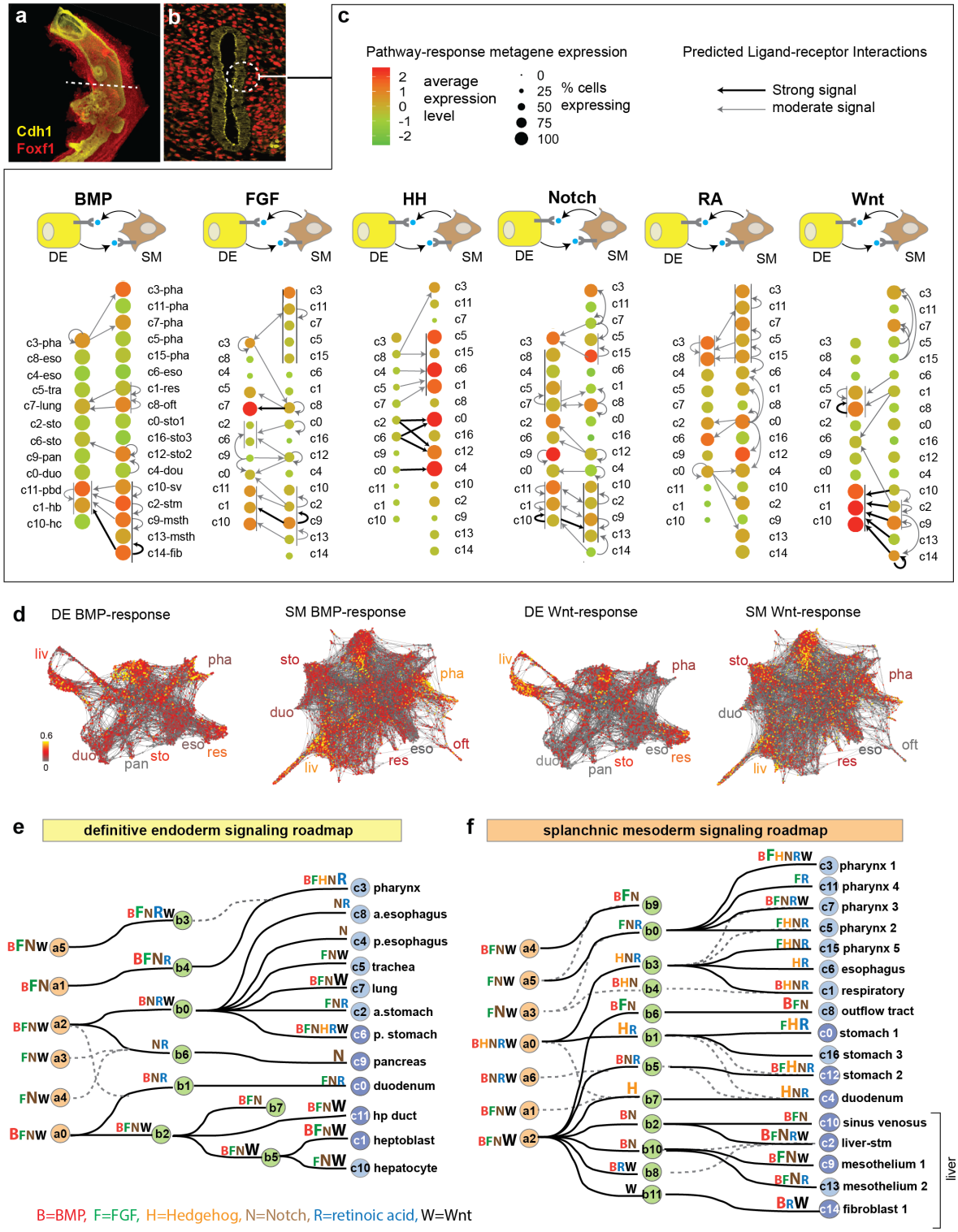


Fig. 3. Computationally inferred receptor-ligand interactions predict a signaling roadmap of foregut organogenesis. a and **b,** Cdh1 (epithelium) and Foxf1 (mesenchyme) immunostaining in **a**) whole mount and **b**) section of a mouse E9.5 foregut, showing the epithelial-mesenchymal tissue microenvironment (dashed circle). **c,** Predicted receptor-ligand interactions between different foregut cell populations. The schematics show paracrine signaling between the DE (yellow cells) and the SM (brown cells) for six major pathways. Below the schematics, E9.5 DE and SM cell clusters are ordered along the anterior to posterior axis as the organ primordia locations *in vivo*, while spatially adjacent DE and SM cell types are across from one another. Colored circles indicate the relative pathway response-metagene expression levels, predicting the likelihood that a given cell population is responding to the growth factor signal (metagene expression > -1 and > 25% of cells). Thin vertical lines next to clusters indicate different cell populations in spatial proximity that are all responding. Arrows represent the predicted source of the ligands showing paracrine and autocrine receptor-ligand pairs inferred from metagene expression profiles (see Methods). **d**, Examples of BMP and Wnt response-metagene expression levels for individual DE or SM cell projected onto the SPRING plot. **e** and **f,** Projecting the integrated signaling activity for all 6 pathways onto the **e**) DE and **f**) SM cell state trees resulted in a roadmap of the combinatorial signals at each developmental stage predicted to control lineage diversification. The letters on each branch of the tree indicate the indicated the signals experienced by the cell population at each step, with larger font indicating a strong signaling response. a, anterior; p, posterior; hp, hepatopancreatic; stm, septum transversum mesenchyme.

Fig. 4 testing roadmap prediction with mouse genetics

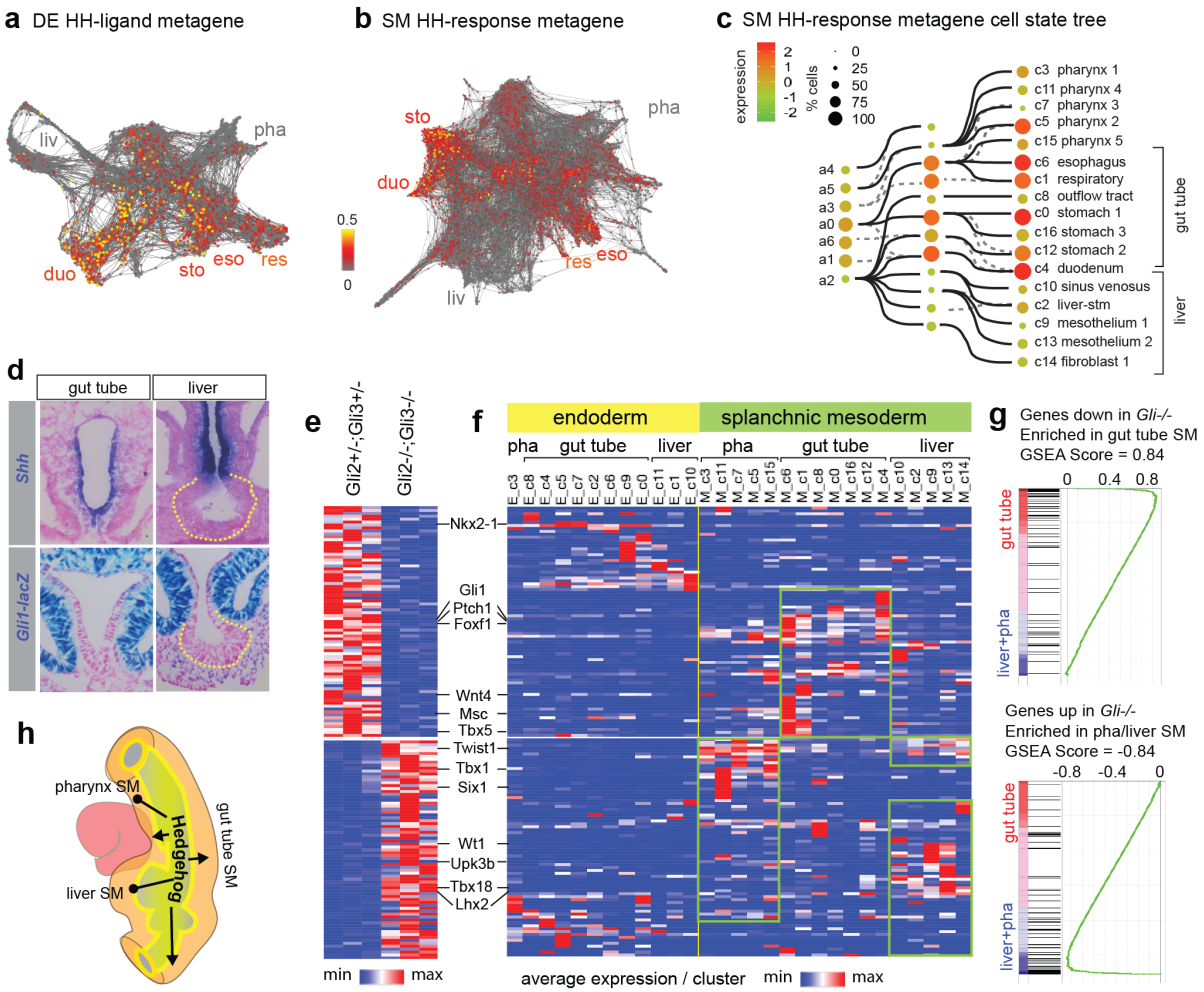


Fig 4. Genetic test of the signaling roadmap revealed that HH promotes gut tube versus liver mesenchyme. a and **b,** SPRING visualization of **a**) the HH ligand-metagene expression in DE cells and **b**) HH response-metagene expression in SM cells. **c**, The HH response-metagene expression projected onto the SM cell state tree showing low HH activity in the liver and pharynx SM but high activity in the gut tube. **d,** *Shh* is expressed in the gut tube epithelium but not in the hepatic epithelium (outlined). *Gli1-lacZ,* a HH-response transgene, is active in the gut tube mesenchyme but not in the liver stm. **e.** Differentially expressed genes between *Gli2-/-, Gli3-/-* and *Gli2+/-, Gi3+/-* mouse E9.5 foreguts through bulk RNA-sequencing (log2 FC >1, FDR < 5%). **f,** Heatmap showing average expression of HH/Gli-regulated genes (from e) in E9.5 DE and SM single cell clusters. **g,** Gene set enrichment analysis (GSEA) reveals specific cell type enrichment of HH/Gli-regulated genes. **h**, Schematic of HH activity in the foregut.

Fig. 5

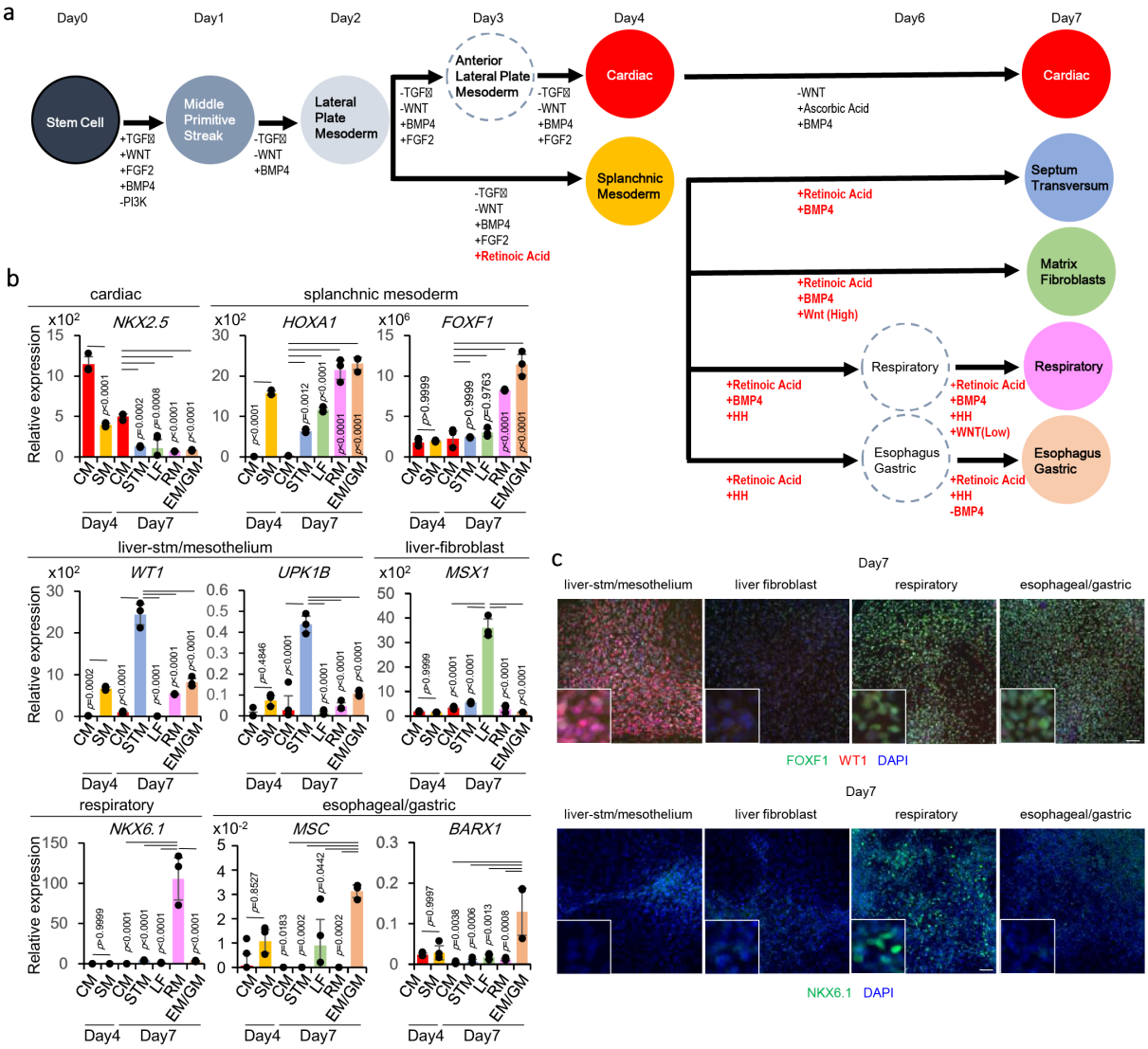
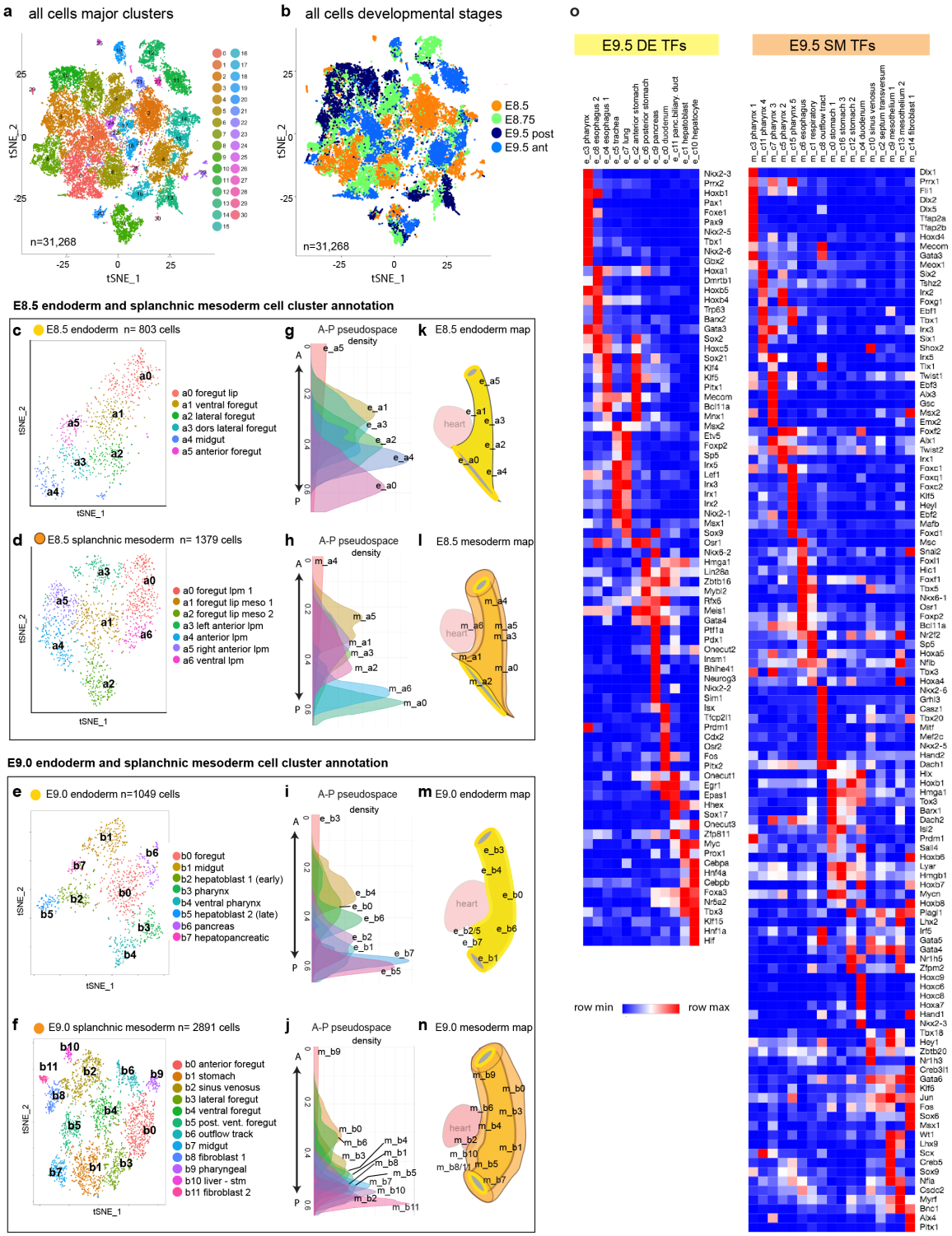


Fig. 5 Generation of splanchnic mesoderm progenitors from human PSCs. a, Schematic of the protocol to differentiate hPSCs into SM subtypes. Factors in red indicate signals predicted from the mouse single cell signaling roadmap. **b,** RT-PCR of markers with enriched expression in specific SM subtypes based on the mouse single cell data; cardiac (*NKX2-5*), early SM (*FOXF1, HOXA1*); liver-stm/mesothelium (*WT1, UKP1B*), liver-fibroblast (*MSX1*), respiratory SM (*NKX6-1+/MSC-*), esophageal/gastric (*MSC, BARX1*). Columns show the means ± S.D. Tukey's test was used for statistical analyses. **c,** Immunostaining of day 7 cultures.

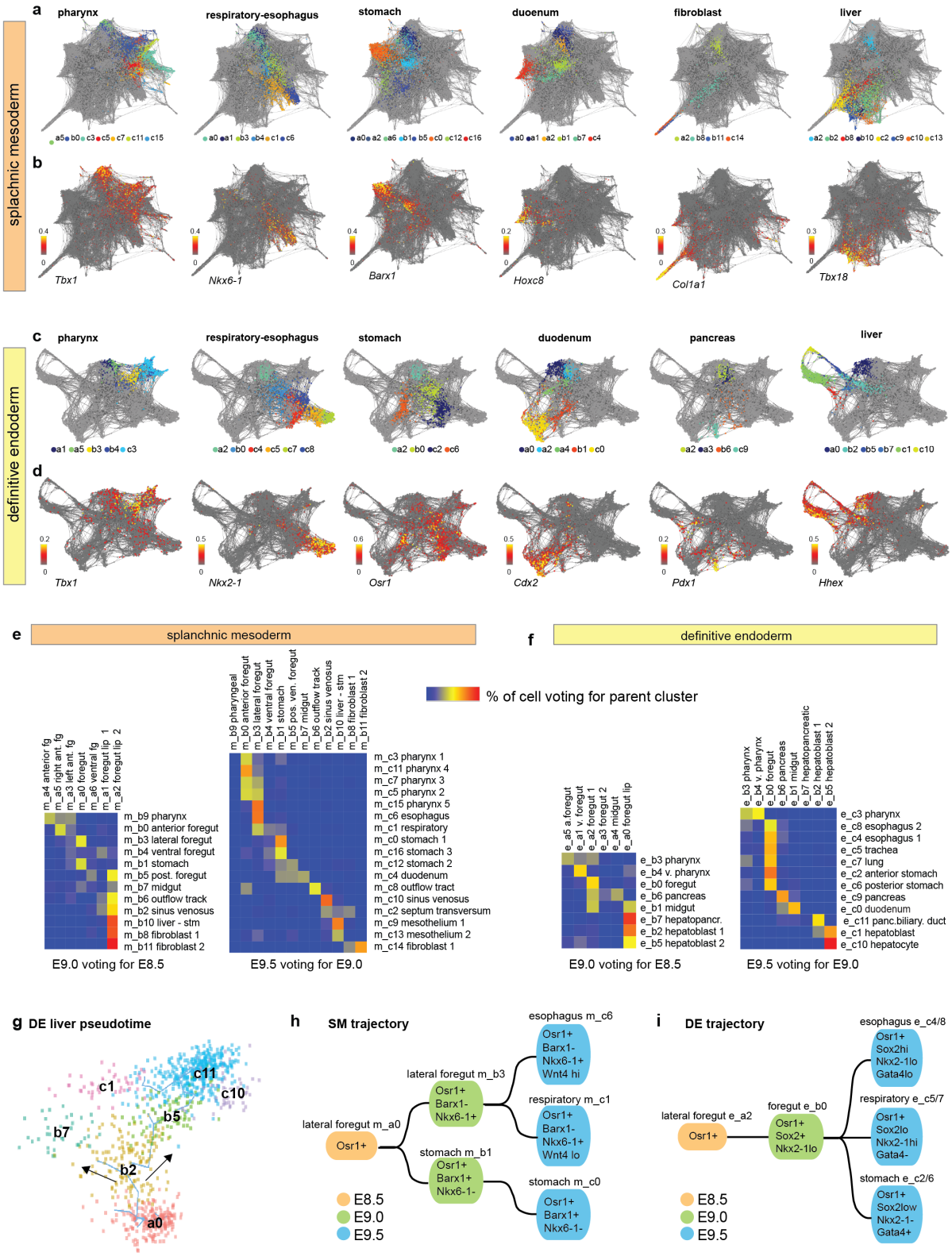
Supplementary Material

Supplementary Fig. S1. annotation of DE and SM cell clusters



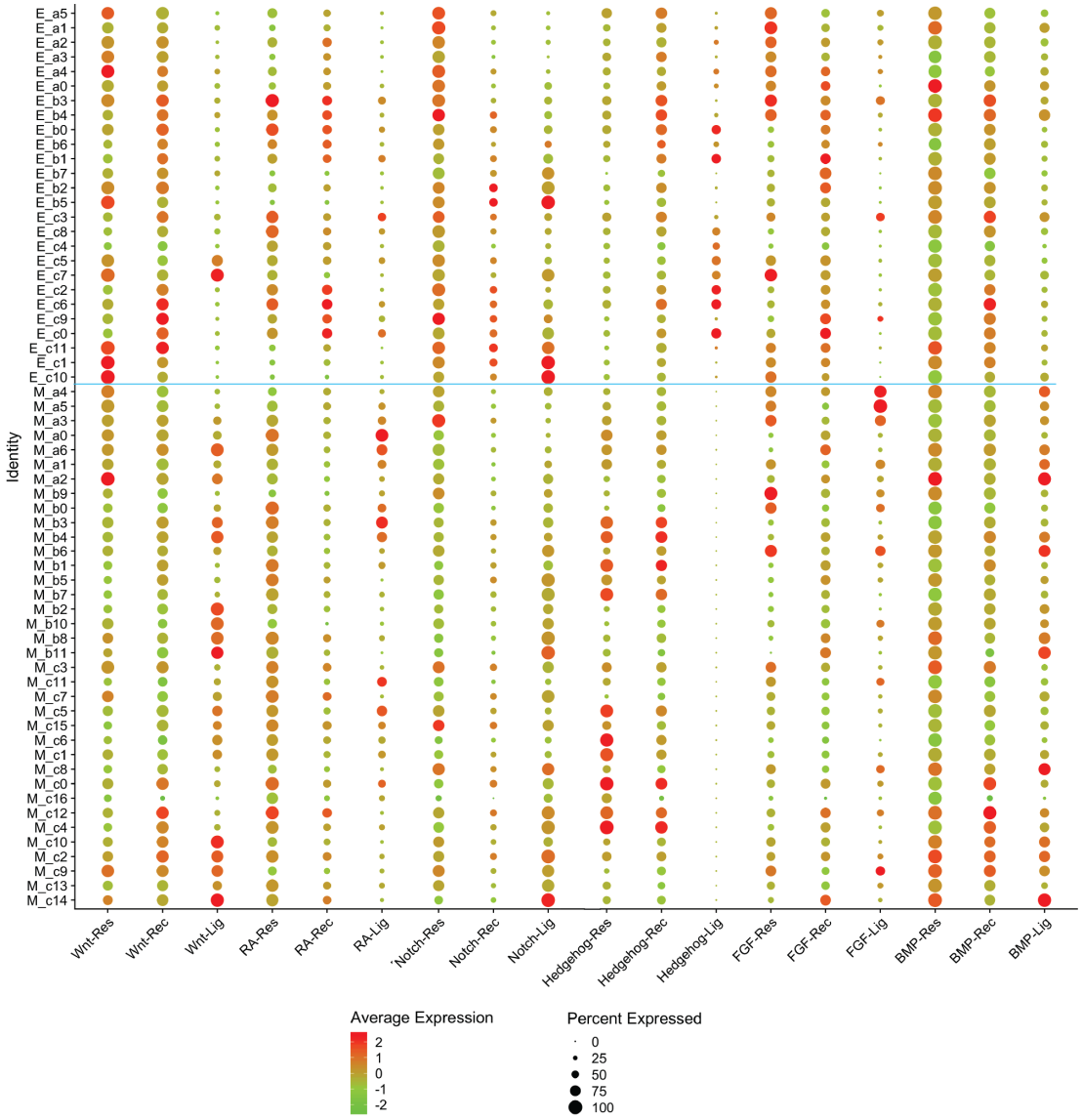
Supplementary Fig. S1 Single cell analysis of the mouse foregut endoderm and mesoderm lineages. a and **b,** tSNE visualization of all cells (n=31,268) from all stages colored by **a**) predicted clusters and **b**) developmental stages. **c-f,** t-SNE plot of *in-silico* isolated, E8.5 DE (**c**), E8.5 SM (**d**), E9.0 DE (**e**) and E9.0 SM cells (**f**) with cell type annotations. E8.5,clusters are designated "a", E9.0; "b", and E9.5; "c". **g-j,** Pseudo-spatial ordering of E8.5 DE (**g**), E8.5 SM (**h**), E9.0 DE (**i**) and E9.0 SM cells (**j**) along the anterior-posterior (A-P) axis of the gut tube. **k-n,** Schematics of the mouse embryonic foregut showing the predicted location of E8.5 DE (**k**), E8.5 SM (**l**), E9.0 DE (**m**) and E9.0 SM (**n**) cell types mapped onto **k, m**) the endoderm (yellow) and **l, n**) mesoderm (orange)**. o,** Heatmap of transcription factor (TFs) with enriched expression in E9.5 DE and SM populations defines a combinatorial TF code for different cell types. ant, anterior; post, posterior; vent, ventral; lpm; lateral plate mesoderm; meso, mesoderm; stm, septum transversum mesenchyme.

Supplementary Fig. S2. lineage trajectory



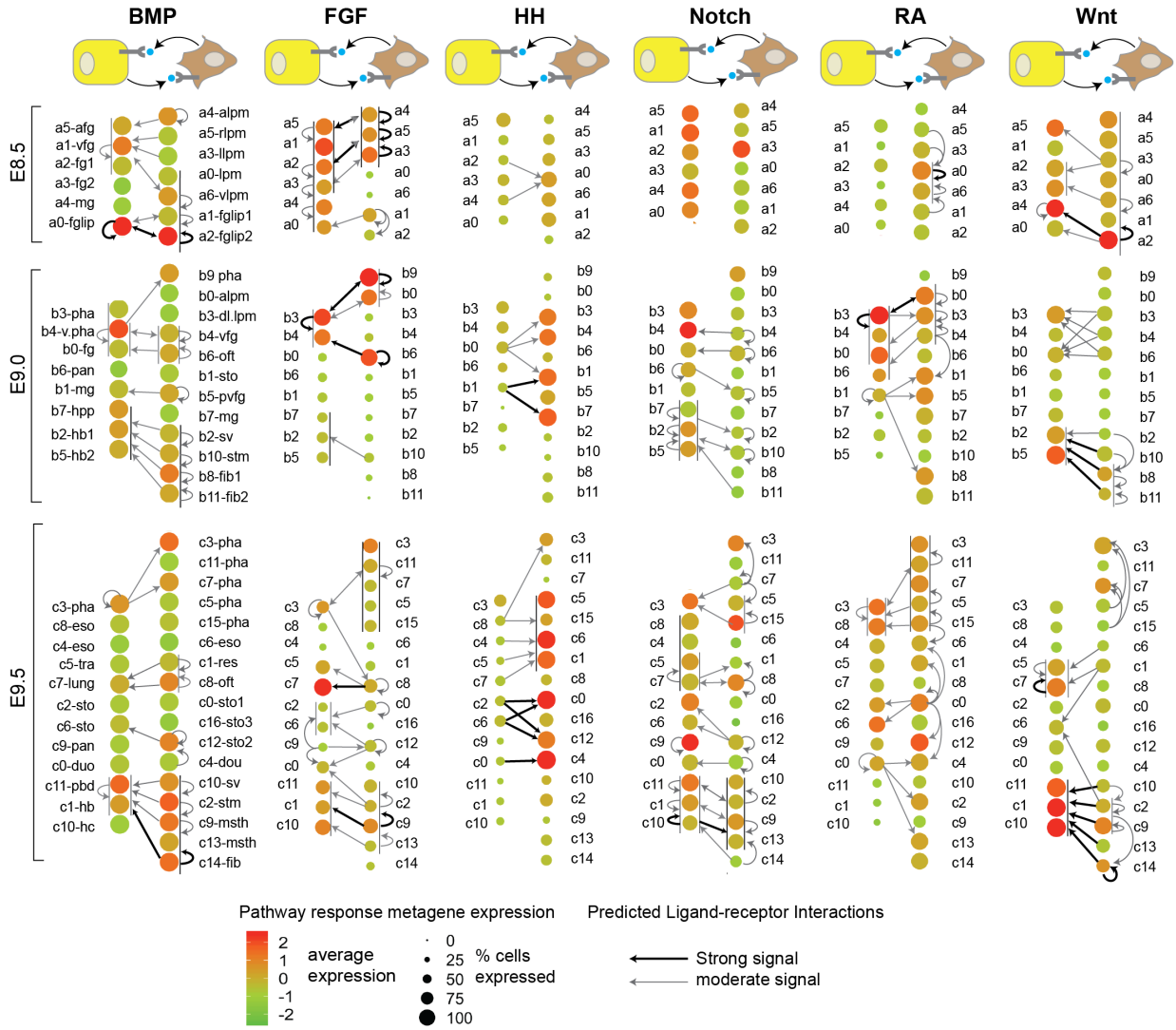
Supplementary Fig. S2. Coordinated diversification of endoderm and mesoderm trajectories. a and **b,** Spring plots of all SM cells (n=10,097) showing **a**) SM lineage colored by cell clusters and **b**) expression of key marker genes in SM cells. **c** and **d,** Spring plots of all DE cells (n=4,448) showing **c**) DE lineage colored by cell clusters and **b**) expression of key marker genes in DE cells. **e** and **f,** Confusion matrix heatmap showing result of "parent-child" single cell voting for **e**) DE and **f**) SM cells, which was used to construct the transcriptional cell stage tree of trajectories in Fig. 2c,d. Each cell in a cluster at the later time point (y-axis) votes for its most similar cell at the preceding time point (x-axis) based on transcriptome similarity (k-nearest neighbors). All of the single cell votes for a give cluster are tabulated, normalized for cluster size (see Methods for details) and represented as a % of votes in the heatmap. E8.5,clusters are designated "a", E9.0; "b", and E9.5; "c". **g,** Pseudotime analysis of the hepatic DE lineage using Molocle_v3 suggests that at E9.0, the e_b2 cluster (early hepatoblasts) is a common progenitor of e_b5 (later hepatoblasts) and e_b7 (hepatopancreatic duct progenitors). **h** and **i**, Graphical illustration of the esophageal-respiratory-gastric cell state trajectories for **h**) SM and **i**) DE with key marker genes. This suggests the coordinated development of *Osr1+* multi-lineage progenitors in endodermal and mesodermal layers, respectively. Nodes are colored by developmental stage.

Supplementary Fig. S3 Receptor-, Ligand- and Response-metagene expression



Supplementary Fig. S3. Metagene expression for all ligands, receptors and context-independent response genes. Dot plot showing the average scaled expression (2 to -2) of metagenes (X-axis) in each DE and SM cluster (Y-axis). For each cell signaling pathway (BMP, FGF, HH, Notch RA and canonical Wnt), we calculated a "*ligand-metagene*", "*receptor-metagene*" and "*response-metagene*" by averaging the normalizing expression of each individual gene for each pathway (e.g.: Wnt-ligand metagene =ΣWnt1+Wnt2+Wnt2b+Wnt3…Wnt10b expression/n) in each cell and cluster (see Methods for details). Color and size of the dot represents the metagene expression level in each cluster. (See table S2 for list of genes that make up each metagene and the numeric metagene expression data used to generate the plot).

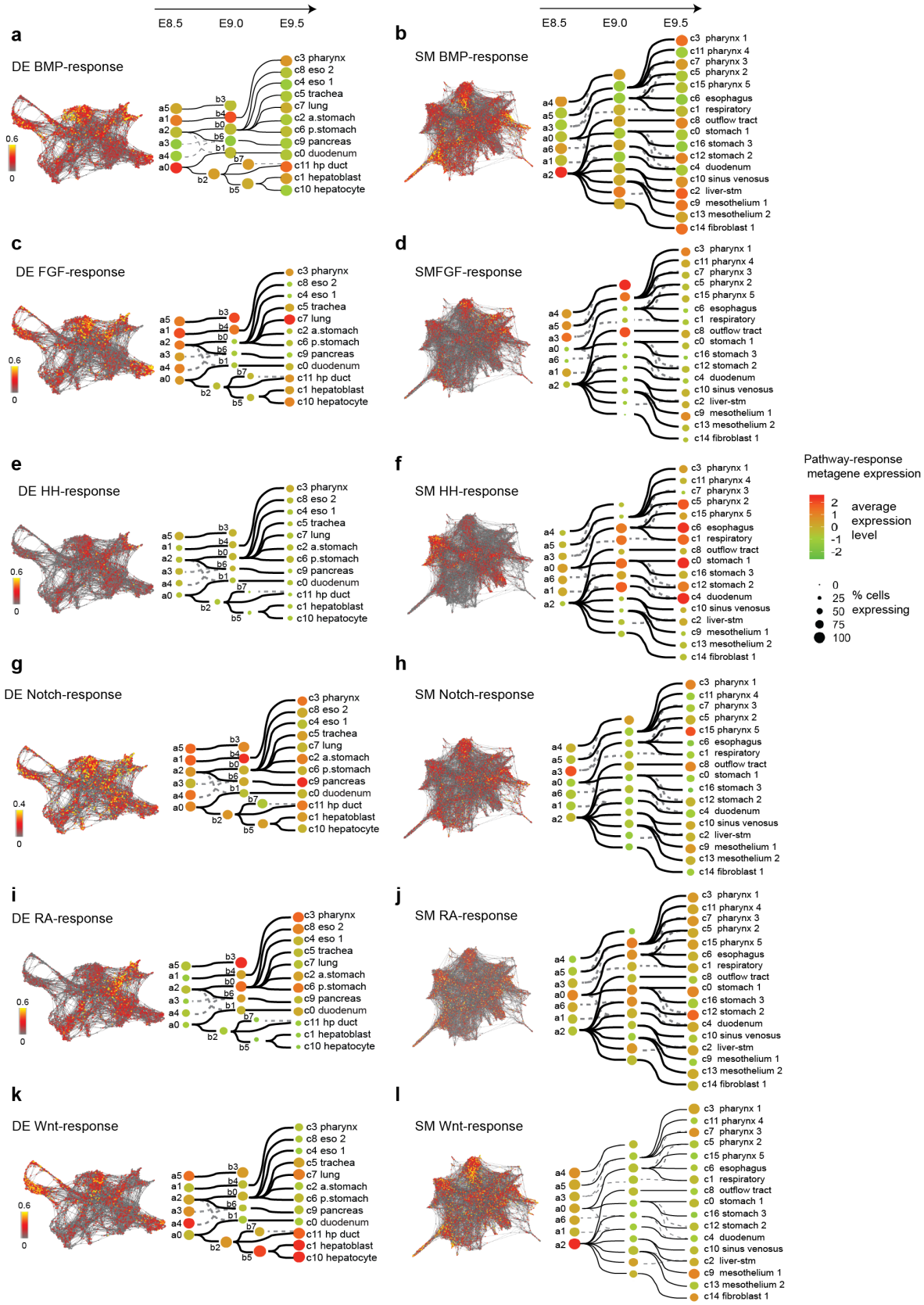
Supplementary Fig. S4 Receptor-Ligand interactions



Supplementary Fig. S4 Computationally predicted receptor-ligand interactions. Predicted receptor-ligand interactions between different foregut cell populations. The schematics show paracrine signaling between the DE (yellow cells) and the SM (brown cells) for six major pathways. Below the schematics DE and SM cell clusters of each stage are ordered along the A-P axis consistent with the location of organ primordia *in vivo*; spatially adjacent DE and SM cell types are across from one another. Colored circles for each cluster indicate the likelihood that a given cell population is responding to the growth factor signal based on the pathway response metagene expression levels (from Fig. S3). A cluster was considered to be responding if the

metagene expression level was > -1 and expressed in > 25% of cells. Similarly, a cluster was judged to express sufficient levels of ligand or receptor in order to signal or respond respectively if the metagene expression > -1 and expressed in > 25% of cells (details in Methods). Arrows represent the predicted source of the ligands showing paracrine and autocrine receptor-ligand pairs inferred from metagene expression profiles. Receptor-ligand pairing (arrows) were restricted to cell populations in close spatial proximity (see Methods for details). Thin vertical lines next to a group of clusters indicate that different cell populations in spatial proximity that are all responding similarly.

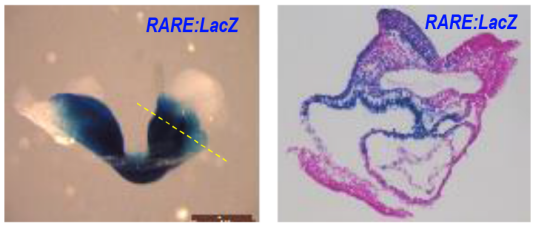
Supplementary Fig. S5 Temporal and spatial dynamics of signaling response



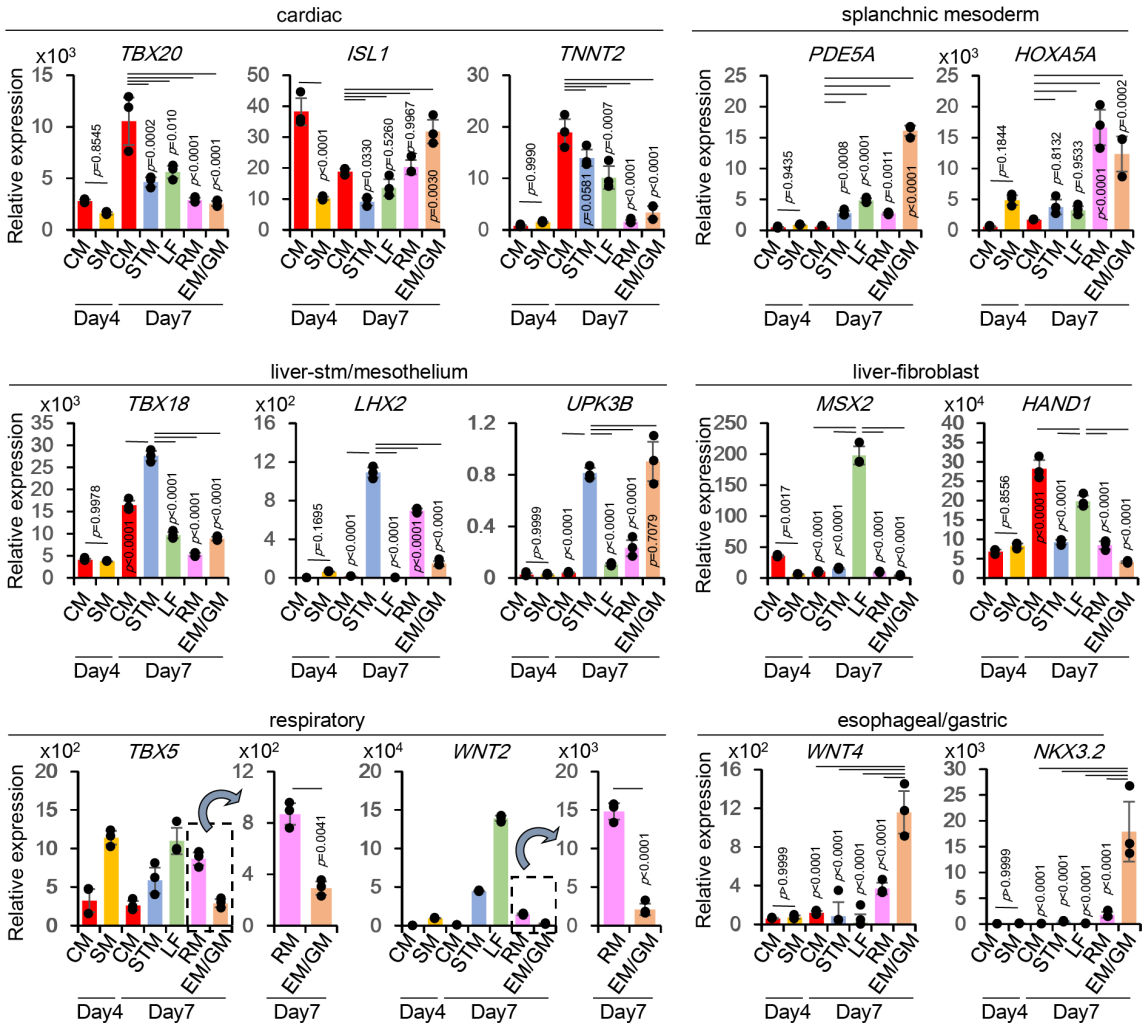
Supplementary Fig. S5 Temporal and spatial dynamics of signaling responses. a-i, Expression levels of the pathway response-metagene (From S3) projected onto the DE and SM SPRING plots and cell state trees for the BMP (a-b), FGF (c-d), HH (e-f), Notch (g-h), RA (i-j) and canonical Wnt (k-l) pathways. This shows how coordinated spatial domains of signaling activity that correspond to cell lineages, change over 24 hours from E8.5 – E9.5.

Supplemental Fig. S6

a



b



Supplementary Fig. S6. Leveraging the single cell signaling roadmap to generate splanchnic mesoderm linages from hPSCs a. Staining of *RARE-lacZ* transgenic mouse embryos confirms the single cell RA response metagene prediction that RA activity is higher in the splanchnic mesenchyme, than the cardiac mesenchyme at E8.5. **b.** RT-PCR of markers with enriched expression in specific mesoderm subtypes based on the mouse single cell data; cardiac (*TBX20, ISL1, TNNT2*), early SM (*PDE5A, HOXA5*); liver-stm/mesothelium (*TBX18, LHX2, UPK3B*), liver-fibroblast (*MSX2, HAND1*), esophageal/gastric (*WNT4, NKX3-2*). Columns show the means ± S.D. Tukey's test was used for statistical analyses

Supplementary Table S1. a, Top distinguishing marker genes for endoderm clusters (excel file tab-1). **b,** Top distinguishing marker genes for splanchnic mesoderm clusters (tab-2). **c**, Transcription factors with enriched expression in endoderm clusters (tab-3). **d**, Transcription factors with enriched expression in splanchnic mesoderm clusters (tab-4).

Supplementary Table S2. a. List of BMP, FGF, HH, Notch and canonical Wnt pathway genes used to calculate metagene profiles. **b.** Normalized and scaled average expression of each metagene in each DE and SM cluster.

Supplementary Table S3. Differentially expressed transcripts in the bulk RNA-sequencing of E9.5 mouse foregut comparing *Gli2-/-;Gli3-/-* to *Gli2+/-;Gli3+/-* littermates.

Supplementary Table S4. Antibody and primer information.