Single cell transcriptome dynamics from pluripotency to FLK1$^+$ mesoderm

Haiyong Zhao and Kyunghee Choi*

ABSTRACT

Hemangiogenic progenitors generating blood and endothelial cells are specified from FLK1-expressing (FLK1$^+$) mesoderm by the transcription factor ETV2. FLK1$^+$ mesoderm also contributes to smooth muscle and cardiomyocytes. However, the developmental process of FLK1$^+$ mesoderm generation and its allocation to various cell fates remain obscure. Recent single cell RNA-sequencing studies of early embryos or *in vitro*-differentiated human embryonic stem (ES) cells have provided unprecedented information on the spatiotemporal resolution of cells in embryogenesis. These snapshots, however, lack information on continuous dynamic developmental processes. Here, we performed single cell RNA sequencing of *in vitro*-differentiated mouse ES cells to capture the continuous developmental process leading to hemangiogenesis. We found that hemangiogenic progenitors from ES cells develop through intermediate gastrulation stages, which are gradually specified by 'relay'-like highly overlapping transcription factor modules. Moreover, the transcriptional program of the *Flk1*$^+$ mesoderm was maintained in the smooth muscle lineage, suggesting that smooth muscle is the default fate of *Flk1*$^+$ mesoderm. We also identified the SRC kinase contributing to ETV2-mediated activation of the hemangiogenic program. This continuous transcriptome map will facilitate both basic and applied studies of mesoderm development.

KEY WORDS: Single cell RNA sequencing, Flk1/Kdr, Etv2, Hemangiogenesis, Vascular smooth muscle, Mouse embryonic stem, Yolk sac

INTRODUCTION

Previous studies have established that blood and endothelial cells differentiate from FLK1 (KDR)-expressing mesoderm (Ema et al., 2006; Kataoka et al., 1997; Kattman et al., 2006; Lugus et al., 2009). *Flk1*$^+$ mesoderm can also contribute to smooth muscle, cardiac lineages and paraxial mesoderm tissues (Ema et al., 2006; Kataoka et al., 1997; Kattman et al., 2006; Motoike et al., 2003; Yamashita et al., 2000). However, high plasticity, heterogeneity and rapid differentiation of early embryonic cells have hampered investigations on the lineage relationships among these cell types, the mechanisms underlying the lineage restriction steps regulating *Flk1*$^+$ mesoderm development, and the following commitment from *Flk1*$^+$ mesoderm to its downstream lineages. Recent progress in single cell RNA profiling technology has advanced our understanding of the complex cell constitution at discrete early embryonic stages (Farrell et al., 2018; Gong et al., 2017; Ibarra-Soria et al., 2018; Lescroart et al.,

Department of Pathology and Immunology, Washington University School of Medicine, Saint Louis, MO 63110, USA.

*Author for correspondence (kchoi@wustl.edu)

H.Z., 0000-0002-9816-901X; K.C., 0000-0002-9634-9375

2018; Mohammed et al., 2017; Moignard et al., 2015; Pijuan-Sala et al., 2019; Scialdone et al., 2016; Wagner et al., 2018) or in *in vitro*-differentiated embryonic stem (ES) cells (Chu et al., 2016; Han et al., 2018; Spangler et al., 2018). However, none of these studies has provided a complete process from pluripotent cells to *Flk1*$^+$ mesoderm owing to missing intermediate states or batch effects. Therefore, there is still a gap in our current knowledge on the continuous sequence of *Flk1*$^+$ mesoderm development and its segregation into the hemangiogenic lineage.

Mouse ES cells in culture differentiate in an asynchronous manner giving rise to embryoid bodies (EBs). Both undifferentiated pluripotent cells and differentiated progenies (Park et al., 2013; Spangler et al., 2018) co-exist in mouse EBs at early stages of differentiation. We reasoned that cells in the intermediate states also exist in the same culture, representing the full spectrum of differentiation, which would provide a unique opportunity to delineate a continuous differentiation process without batch effects. By taking advantage of the asynchronous nature of the differentiation of mouse ES cells and culture conditions that favor hemangiogenic lineage differentiation, we performed single cell RNA sequencing (scRNA-seq) of EB cells to elucidate a continuous developmental path from the pluripotent state to *Flk1*$^+$ mesoderm, and to the hemangiogenic lineage. We found that the cardiac lineage branched out immediately after nascent mesoderm specification. Intermediate states positioned between ES and *Flk1*$^+$ mesoderm were distinguished by a few temporally overlapping transcription factor modules, suggesting a plastic and combinational regulatory mechanism regulating *Flk1*$^+$ mesoderm development. Notably, expression of most of the characteristic genes of *Flk1*$^+$ mesoderm continuously increased in the smooth muscle branch, implying a 'default' route. Our data suggested that ETV2 turns on the hemangiogenic program from *Flk1*$^+$ mesoderm by restraining the smooth muscle fate. We further identified that the SRC kinase potentially regulates ETV2-mediated hemangiogenic lineage specification, possibly through mediating cell adhesion signaling. These studies provide a comprehensive description of mesoderm and hematopoietic/endothelial cell lineage development at the single cell level.

RESULTS

EBs contain undifferentiated pluripotent cells and *Flk1*$^+$ mesoderm tissues

In capturing the dynamics of transcriptome of the hemangiogenic differentiation, we performed scRNA-seq using day 4 EB cells in hemangiogenic differentiation. We reasoned that day 4 would allow access to a range of cells from pluripotent cells to those of the hemangiogenic cell lineage, as hemangiogenic cells extensively emerge and pluripotent stem cells are still present at this time. After filtering out low-quality cells, 1848 cells, which had 4373 genes/27,272 unique mRNAs detected per cell in average (Fig. 1A), were kept for further analysis. Day 4 EB cells were grouped into 18 populations (Fig. 1B, Fig. S1A) and 936
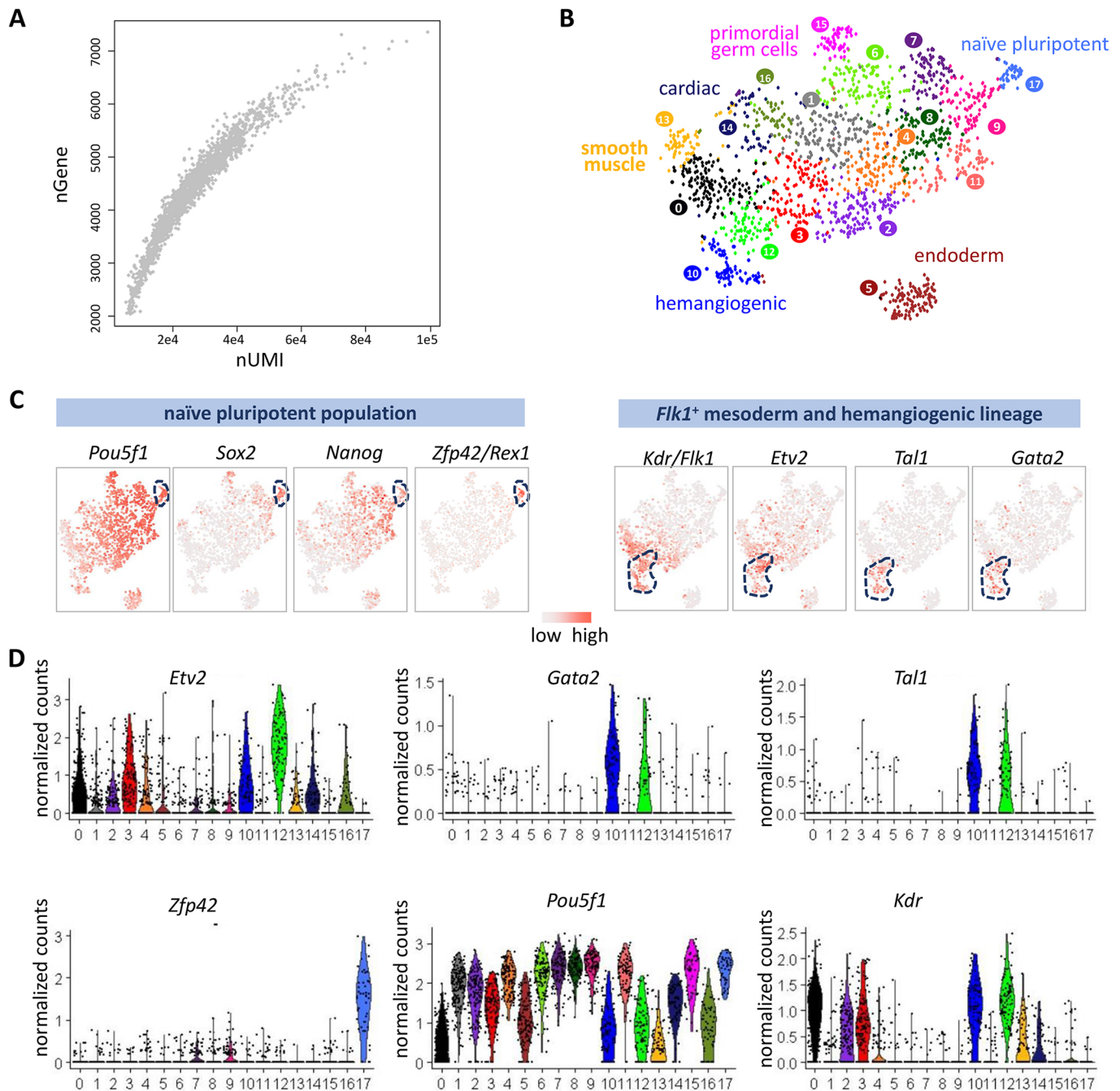
Fig. 1. scRNA-seq of EBs containing undifferentiated pluripotent cells and differentiated cells. (A) Distribution of genes and transcripts detected in individual cells. nGene, number of detected genes in the cell; nUMI, number of detected UMIs in the cell. (B) t-SNE projection of all cells. Cluster numbers correspond to those in Fig. S1A. Potential lineage branches were annotated according to specific marker gene expression patterns. (C,D) Expression of the indicated marker genes for specific cell states/lineage fates in different clusters from Fig. 1B. The colors and numbers of the clusters are consistent with those shown in Fig. 1B. The violin plot shows the normalized counts of given genes in each cluster.

genes were differentially upregulated across the populations (Table S1). t-Stochastic neighbor embedding (t-SNE) analysis indicated that most clusters did not form a distinct, well-segregated population (Fig. 1B). Instead, distinct populations branched out gradually from the trunk, underpinning continuous developmental processes. We assigned different fates to the branches according to the expression of related marker genes (Table S1). SPRING (Weinreb et al., 2018), a different tool for visualization of high-dimensional scRNA-seq data, produced a similar clustering pattern of the cells (Fig. S1B). High expression of *Sox2*, *Pou5f1* (also

known as *Oct4*), *Nanog* and *Zfp42* (also known as *Rex1*) highlights the maintenance of undifferentiated naïve pluripotent cells in the EBs (Fig. 1C,D). *Flk1*+ mesoderm and the hemangiogenic lineage, as marked by high expression of *Flk1* and *Etv2*, and upregulation of the ETV2 target genes *Tal1* and *Gata2*, clustered at the other end of the ES cell population (Fig. 1C,D). Expression of related genes showed a gradual shift in intermediate clusters from the pluripotent ES population to the *Flk1*+ mesoderm branch, revealing continuous intermediate states from ES to *Flk1*+ mesoderm that are accessible in EBs.

scRNA-seq captured multiple intermediate stages from naïve pluripotency to *Flk1*⁺ mesoderm derivatives in EBs

To determine in further detail the developmental relationships among the clusters between the ES and *Flk1*⁺ populations, we first examined the dynamics of the loss of pluripotency and the bifurcation of mesoderm and endoderm fates. Pluripotency can be maintained in two states: 'naïve' and 'primed' (Kalkan and Smith, 2014). The former represents the primitive phase and is characterized by high expression of *Zfp42*, whereas the latter loses *Zfp42* expression and harbors an unstable pluripotency gene network, which marks a nascent stage of gastrulation/mesendoderm (Kalkan and Smith, 2014). We picked out clusters that were located in the endoderm path
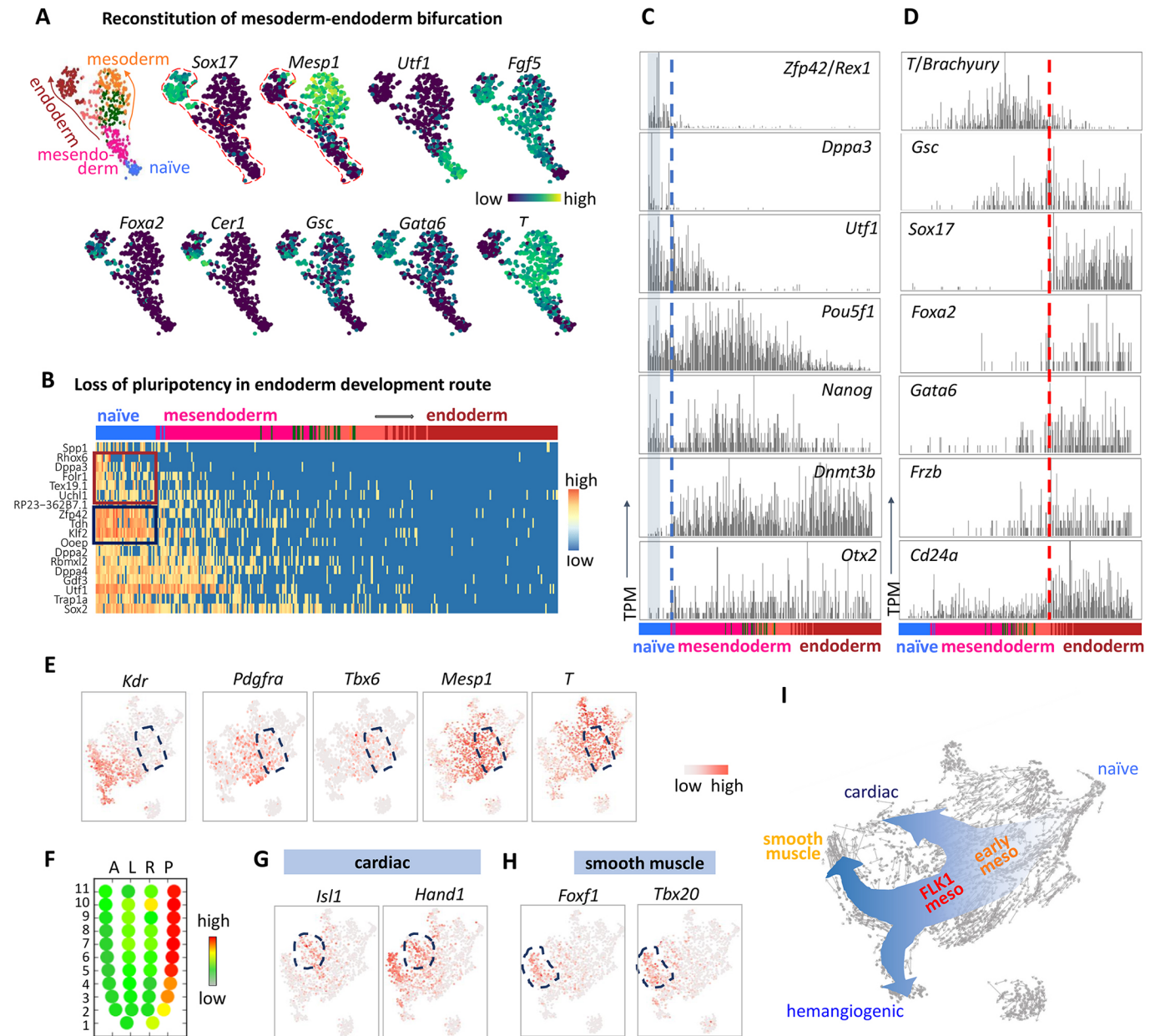

Fig. 2. scRNA-seq captured continuous intermediates from naïve pluripotency to *Flk1*⁺ mesoderm derivatives. (A) Re-clustering of populations from naïve pluripotency exit to bifurcation of endoderm and mesoderm. Clusters 4, 5, 8, 9, 11 and 17 in Fig. 1B were picked to re-cluster based on PCA. The first two components were used to do re-clustering. The clusters are in the same colors as they are in Fig. 1B. 'naïve' indicates naïve pluripotency. (B) Cells in the endoderm differentiation route (red dashed line-surrounded region in A) were ordered into pseudo-time line. Arrows indicate the differentiation direction. Heatmap of naïve pluripotent cell-enriched genes are shown. Red box outlines genes heterogeneously expressed in the naïve stage; blue box marks the constantly expressed naïve stage-specific genes. (C) Barplot of expression of selected marker genes along the endoderm pseudo-time line. Shaded region covers cells maintaining high *Dppa3* expression in the naïve population. Blue dashed line marks the exit point from naïve pluripotency. (D) Barplots of expression of marker genes along the pseudo-time line of endoderm differentiation. Red dashed line marks the point at which endoderm fate is committed. (E) Expression of the indicated genes in early mesoderm cells (blue dashed line-surrounded region). (F) Zipcode mapping of early mesoderm cells (cluster 4 in Fig. 1B; mean values of each genes in the cluster were used for the mapping) to E7.0 mouse embryo. The color indicates the similarity between the transcriptome of our population and that of a specific micro-dissected region in E7.0 mouse embryo. A, anterior; L, left; R, right; P, posterior. (G,H) Expression of the indicated genes in cardiac or smooth muscle branches (regions encircled by blue dashed lines). (I) Scheme of developmental routes of the indicated mesoderm lineages (an enlarged and color-coded version is shown as Fig. S2E).

(4, 5, 8, 9, 11 and 17 in Fig. 1B), and re-clustered the cells based on the 163 most differentially expressed genes among these cells (Fig. 2A). A narrow 'path' from the *Utf1*-expressing pluripotent population leads to the *Sox17*-expressing endoderm (the region surrounded by a red dashed line), next to the sprouting *Mesp1*-expressing nascent mesoderm. We next ordered cells in the endoderm path into a pseudo-time developmental line and examined the kinetics of the 163 most differentially expressed genes (Fig. 2B-D, Fig. S2A). At the beginning of the path, we observed an exit from the naïve pluripotency (Fig. 2C, flanking the blue dashed line). Unexpectedly, expression of a small group of genes (represented by *Dppa3*; Fig. 2C and red box in Fig. 2B) already displayed strong heterogeneity in the *Zfp42*-expressing population. Gradual loss of the expression of these genes preceded the exit from naïve pluripotency. This may suggest an extra layer of guarding mechanisms of pluripotency beyond the naïve circuit (as represented by the genes in the blue box in Fig. 2B, which have relatively constant expression in the naïve pluripotency stage). Primitive streak markers such as *T* and *Fgf5* were upregulated in cells exiting from the naïve pluripotent state and entering the mesendoderm stage (Fig. 2A). Shortly after, the endoderm and mesoderm lineages bifurcated. *Fgf5* expression became more enriched in endoderm, whereas *T* expression was exclusively enriched in mesoderm (Fig. 2A). *Gata6* and *Gsc* were expressed in both mesoderm and endoderm, although they were expressed more extensively in the latter (Fig. 2A). The majority of cells in the endoderm population were separated from the trunk, with very few cells expressing *Foxa2* at a high level in the intermediate state. When we lowered the stringency and included more variable genes to cluster the cells, the connection between endoderm and the mesendoderm cluster became more obvious (Fig. S2B,C). These results suggest that the core endoderm genes form a strong self-enforcing circuit, rendering the intermediate specification state relatively unstable.

The immediate upstream population of *Flk1*$^+$ mesoderm expressed *Pdgfra* and *Mesp1* (Fig. 2E, region marked by a blue dashed line). These cells also expressed low levels of *Tbx6*, implying that, to some extent, they are 'primed' for paraxial mesoderm, which generates

tissues including skeletal muscle and cartilage. As such, they might represent an early ancestral state for both paraxial mesoderm and *Flk1*$^+$ lateral plate mesoderm. Consistently, cells in this region had an expression pattern similar to embryonic day (E) 7.0 posterior primitive streak, where nascent mesoderm emerges (Fig. 2F). Furthermore, these cells were transcriptionally similar to *Mesp1*-labeled early mesoderm cell population in mouse E6.75 embryos in recent scRNA-seq data from Lescroart et al., (2018). Similar to EB-derived progenitors, this embryonic population of cells still expressed high levels of the pluripotency gene *Pou5f1*, and already expressed high levels of *Mesp1* and *Pdgfra*. Moreover, although they expressed low levels of *Tbx6*, they have not yet upregulated *Flk1* (cluster 1 in Fig. S2D). The cardiac lineage branched out mainly from the *Pdgfra*$^+$*Flk1*$^-$ early mesoderm population, and a small group of *Flk1*$^+$ mesoderm also expressed the cardiac transcription factor *Isl1* (Fig. 2G). Notably, smooth muscle and hemangiogenic lineages directly bifurcated from an *Flk1*-expressing population, suggesting a common origination (Fig. 2H). RNA velocity estimation of the variable genes, based on the abundance of unspliced and spliced mRNAs (La Manno et al., 2018), revealed persistent overall cell state dynamics from naïve pluripotency to differentiation branches (Fig. S2E). These results suggested a continuous process from pluripotency to *Flk1*$^+$ mesoderm and its derivate tissues (Fig. 2I).

Cells corresponding to different developmental stages proliferate equally

To investigate whether the finding that pluripotent and early gastrulation populations co-exist in day 4 EBs is due to differential cell proliferation, we examined expression of cell cycle-related genes. The expression pattern of typical S-phase marker genes showed no obvious bias in different cell populations (Fig. 3A, Fig. S3A). Using the cell cycle analysis tool in the Seurat package (Butler et al., 2018), we assigned each cell to a specific phase in the cell cycle and found that the majority of cells in the mesoderm developmental route proliferated robustly (Fig. 3B, Fig. S3B). Analysis of cell proliferation using the violet proliferation dye 450 (VPD450) also confirmed that
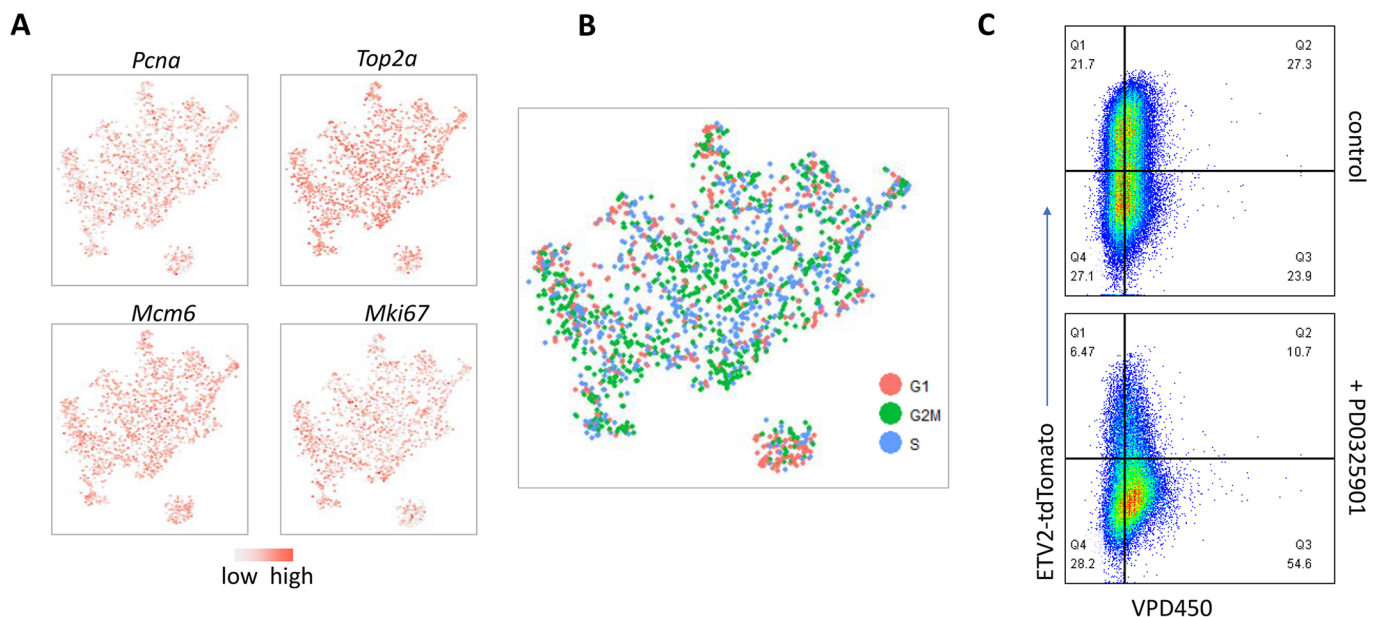

Fig. 3. Cells corresponding to different developmental stages proliferate equally. (A) Expression of the indicated cell cycle genes. (B) Assigned cell cycle phases of each cell. (C) Proliferation of EB cells treated with or without 2 μM of the ERK inhibitor PD0325901 from day 3 was compared using the Dye VPD450. Dye dilution and the *Etv2*-tdTomato reporter gene expression were examined 24 h later using flow cytometry.

EB cells with different levels of differentiation (indicated by the reporter of the hemangiogenic marker gene *Etv2*) proliferated similarly in our system (Fig. 3C, top). As a control, differentiation block and cell cycle delay by the ERK inhibitor PD0325901 was observed in the VPD450 dilution assay (Fig. 3C, bottom).

Transcriptome dynamics in hemangiogenic lineage development

Next, we retrieved a continuous process of the hemangiogenic lineage development by manually choosing the populations corresponding to the shortest path from the naïve pluripotent state to the hemangiogenic branch and reconstructed a fine developmental route (Fig. 4A). A group of *Flk1*-expressing cells achieved the highest level of *Etv2* expression, followed by dramatic upregulation of the ETV2 target genes *Tal1* (also known as *Scl*) and *Gata2*, confirming that a threshold level of *Etv2* expression is necessary for the hemangiogenic lineage specification (Zhao and Choi, 2017) (Fig. 4A). We ordered the cells into a one-dimensional pseudo-time line and assigned the populations to specific stages according to the dynamics of marker gene expression (Fig. 4B). The 68 most variable transcription regulation-related genes were clustered based on their expression along the pseudo-time line (Fig. 4C). They aggregated into a few obvious modules covering different developmental stages (blue dashed line boxed regions in the heatmap). Notably, the naïve pluripotency module (Fig. 4C, box II) and the hemangiogenic module (box III) were relatively independent, whereas each of the intermediate gastrulation modules were shared by adjacent stages (Fig. 4C boxes I, IV and V). The gene network constructed based on the co-expressed genes also highlighted the collaboration among transcription factors of individual modules (Fig. 4D). This stepwise combinatorial usage of different regulatory circuits fits the diverse lineage specifications in gastrulation. Consistent with this, the intermediate populations were less committed, largely interchangeable (Sakurai et al., 2006), and more similar to each other in transcriptome (Fig. 4E, Fig. S4A,B). These results suggested that the intermediate states constitute a 'relay'-like multistep specification process (Fig. 4F). Exit from the naïve pluripotency and activation of the hemangiogenic program might be the two most crucial transitions in hemangiogenic lineage development.

SRC signaling regulates ETV2-mediated hemangiogenic fate specification

Owing to intracellular noises and microenvironmental influences, gene expression within individual cells can fluctuate. scRNA-seq analysis offers an opportunity to utilize this luxuriant gene expression heterogeneity to assess a regulatory network. We previously reported that *Etv2* expression above a threshold level is essential to activate hemangiogenic genes (Zhao and Choi, 2017). We confirmed this at a single cell endogenous mRNA level (Fig. 5A). We noticed that the kinetics of upregulation of ETV2 target genes soon became extremely steep after *Etv2* expression reached a specific level, suggesting an *Etv2*-mediated sensitive 'switch' mechanism in activating its downstream target genes. We explored potential ETV2 cooperating genes for triggering this sensitive switch by examining potential upstream regulators of *Tal1* and *Lmo2*, two ETV2 direct target genes that are crucial for establishing the hemangiogenic lineage (Liu et al., 2015; Porcher et al., 1996). We assumed that if more gene A-expressing cells simultaneously expressed gene B, it is likely that gene A expression depends on gene B (Fig. 5B; see Materials and Methods for more details). From the population in which ETV2 starts to activate hemangiogenic genes (population 12 in Fig. 1B), we assessed the possibility of a given gene being required for expression

of *Tal1* or *Lmo2* (Fig. 5C,D). *Etv2* and *Flk1* were at the top of the list of genes required for *Tal1* or *Lmo2* expression (Fig. 5C,D, Table S2), consistent with the finding that VEGF-FLK1 signaling ensures *Etv2* threshold expression to activate hemangiogenic genes (Zhao and Choi, 2017). The comparison of the top 100 genes required for *Tal1* expression and the top 100 required for *Lmo2* expression generated 66 in common, of which 22 were annotated to cell adhesion-related signaling (Fig. S5A, highlighted in Table S2). The majority of these 22 genes were already highly expressed in populations preceding the hemangiogenic stage (Fig. S5B). To test if the cell adhesion signaling plays a role in ETV2-mediated hemangiogenic gene activation, we utilized a doxycycline (DOX)-inducible *Etv2*-expressing ES cell line (Liu et al., 2012). Because extracellular matrix/cell adhesion signaling converges on the SRC kinase (Cabodi et al., 2010), we added the SRC inhibitor PP2 to day 3 EBs, while simultaneously inducing exogenous *Etv2* expression. After 24 h, as expected, DOX-induced *Etv2* expression greatly enhanced the FLK1$^+$PDGFRα$^-$ population that is enriched in the hemangiogenic lineage cell population (Liu et al., 2012), whereas inhibition of SRC using PP2 was sufficient to reduce *Etv2* overexpression-induced hemangiogenic lineage skewing (Fig. S5C). Given that cell adhesion is necessary for gastrulation as well (Adams and Watt, 1993), it was possible that PP2 affected earlier stages rather than directly affecting the function of ETV2. To minimize this possibility, we induced ETV2-mCherry fusion protein expression ectopically in undifferentiated ES cells, in which the endogenous *Etv2* level is negligible (Zhao and Choi, 2017) and exogenous *Etv2* can induce *Flk1* expression. *Etv2* is also downstream of FLK1 signaling and the reciprocal activation between ETV2 and FLK1 signaling is an important mechanism for hemangiogenic fate determination (Zhao and Choi, 2017). Exogenous ETV2 induced more than 8% of FLK1$^+$ cells in undifferentiated ES cells, and this ratio was significantly reduced by PP2 treatment (Fig. 5E). Unexpectedly, PP2 treatment also dramatically reduced ETV2 protein level, with only a limited number of cells achieving threshold expression of *Etv2* (Fig. 5E). To exclude the possibility that PP2-mediated SRC inhibition affected DOX-induction system or general gene expression processes, we tested DOX-induced expression of another transcription factor, EOMES, and the constitutively active promoter *EF1α*-driven expression of the red fluorescent protein mCherry. We found that DOX-induced EOMES protein level only mildly decreased in the presence of PP2, while *EF1α*-driven mCherry expression was insensitive to this treatment (Fig. S5D). These results suggest that SRC signaling contributes to ETV2-mediated hemangiogenic fate specification possibly through upstream cell adhesion signals.

Flk1$^+$ mesoderm is fated to smooth muscle lineage, but ETV2 skews it into the hemangiogenic fate

Based on the observation of the direct bifurcation of the smooth muscle and hemangiogenic lineages from *Flk1*$^+$ mesoderm (Fig. 2I), we analyzed the cell populations corresponding to the developmental route of the smooth muscle lineage (Fig. 6A, Fig. S6A), and examined expression of the 427 most variable genes along the pseudo-time of differentiation (Fig. 6B). Many of the variable genes aggregated as modules. The modules corresponding to mesendoderm, nascent mesoderm, and *Flk1*-expressing mesoderm stages displayed largely overlapping 'relay'-like patterns. Three groups of genes were most extensively expressed in the smooth muscle branch (Fig. 6B, gene clusters a, b and c in the black dashed line boxed region). Of the three groups of genes, groups b and c were already upregulated prominently in *Flk1*$^+$ mesoderm, with cluster c being upregulated even earlier from the mesendoderm stage. The
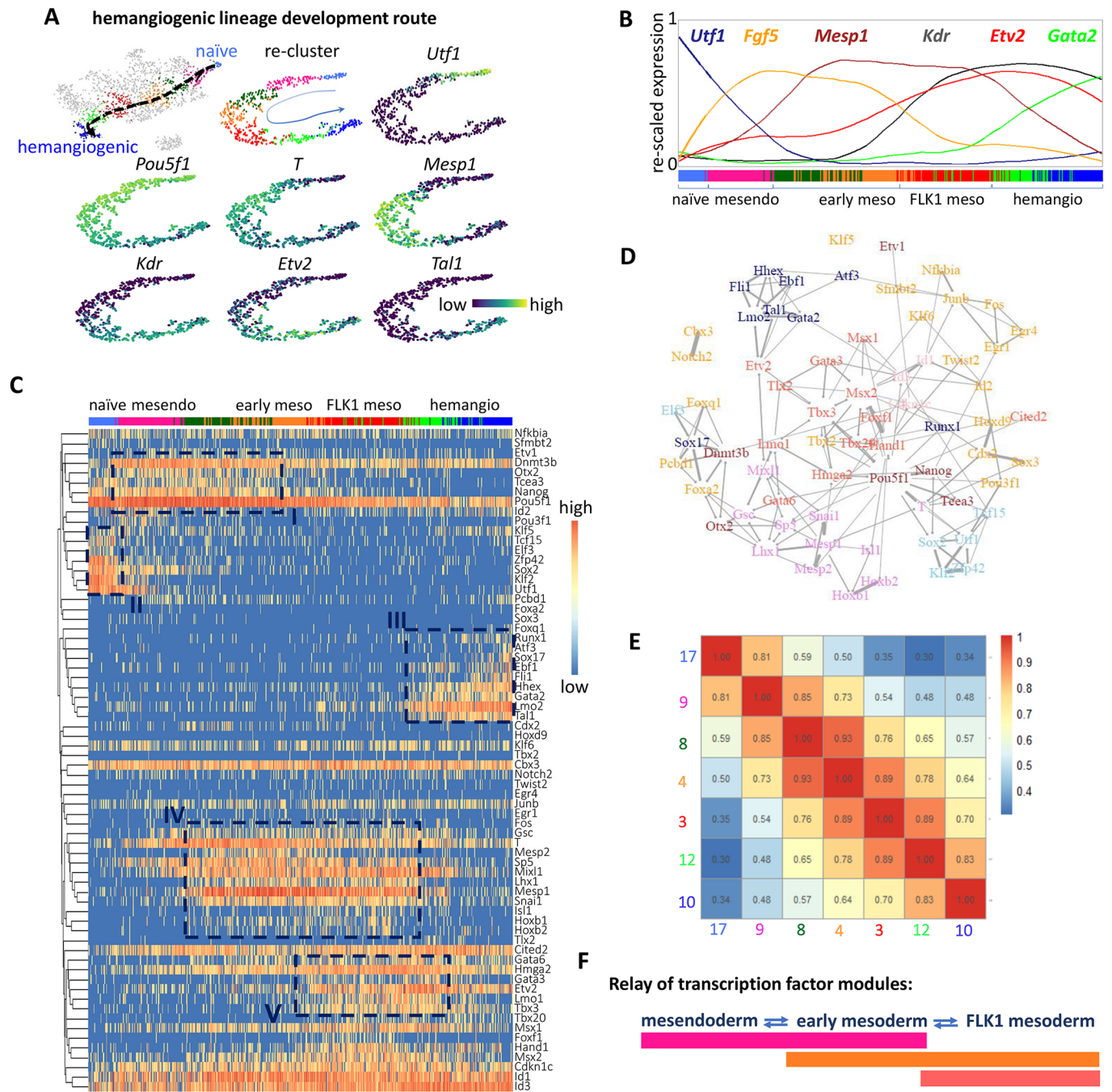
Fig. 4. Transcription factor modules underlying hemangiogenic lineage development. (A) Re-clustering of populations in the route of hemangiogenic lineage development. Clusters 3, 4, 8, 9, 10, 12 and 17 in Fig. 1B were picked to re-cluster. The first two components of PCA are shown for re-clustered cells. The clusters are in the same colors as they are in Fig. 1B. Arrows indicate the differentiation direction. Expression of the representative marker genes in the re-clustered populations is shown. (B) The re-clustered cells were ordered into a pseudo-time line of differentiation. Expression dynamics of representative marker genes along the pseudo-time line is shown. Gene expression data were smoothened. The expression curves and the gene names are in same colors. At the bottom, color bars indicate the cells' original identities as in Fig. 1B. Cell states were annotated as follows: naïve, naïve pluripotent state; mesendo, mesendoderm; early meso, early mesoderm; FLK1 meso, *Flk1*$^+$ mesoderm; hemangio, hemangiogenic lineage. (C) Expression dynamics of transcription regulation-related factors from the 427 most variable genes are shown along the pseudo-time line of hemangiogenic lineage development. Dashed line boxes indicate the obvious stage-specific gene modules. (D) Gene regulatory network of the transcription regulation-related factors shown in Fig. 4C. Only the top 250 gene-gene links are shown. The width of the edges indicate the weight of the links. (E) Gene expression similarities among individual clusters along the hemangiogenic development route. The mean values of gene expression of cells in individual clusters were used to calculate the Pearson's correlation scores. (F) Scheme of the 'relay'-like transcription factor modules underlying *Flk1*$^+$ mesoderm development.

transcription factors *Hand1*, *Tbx20* and *Foxf1* were found in gene cluster b (Fig. 6B,C). *Hand1* and *Foxf1* have been reported to be important for smooth muscle development (Mahlapuu et al., 2001;

Morikawa and Cserjesi, 2004). Expression of the genes in clusters b and c remained similar in the hemangiogenic branch. Only a small cluster of genes was exclusively upregulated after cells entered the
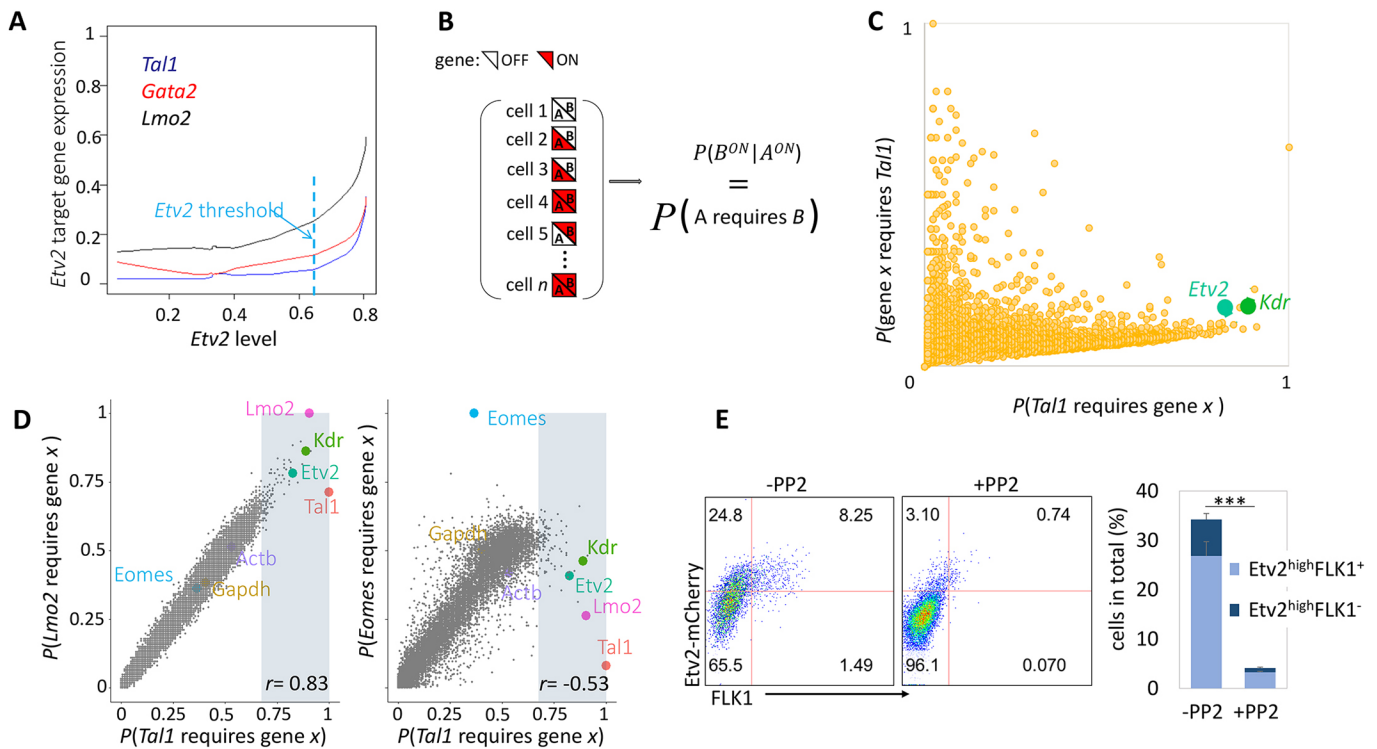
Fig. 5. Initiation of the hemangiogenic program requires cell adhesion signaling. (A) ETV2 target genes sensitively respond to *Etv2* dosage. The input data were from smoothed and re-scaled expression of related genes along the pseudo-time line of hemangiogenic lineage development. Only cells before *Etv2* achieves its highest level in the pseudo-time array were calculated, because after that *Etv2* starts to be downregulated. (B) Scheme for identification of potential regulators of ETV2-mediated hemangiogenic fate specification. If gene B is required for gene A expression, then in a cell in which gene A is expressed gene B should also be expressed. In a population in which more gene A-expressing cells have gene B expression, it is more likely that gene B is required for gene A expression. (C) All detected genes' relationship with *Tal1*. The *x*-axis is the possibility that *Tal1* expression requires a given gene *x* in group 12 of Fig. 1C, where *Etv2* achieves threshold expression and starts to activate hemangiogenic genes. The *y*-axis is the possibility that *Tal1* is required for a specific gene's expression. *Etv2* and *Flk1*, two known upstream regulators of *Tal1*, are labeled in the plot. (D) Comparison of genes required for *Tal1* expression to that required for *Lmo2* or *Eomes*. *Eomes* is a transcription factor expressed in *Flk1*+ mesoderm that has not been reported to regulate hemangiogenic genes. For the top 100 genes that are most tightly required for *Tal1* expression in cluster 12 (dots in shaded regions), the Pearson's correlation coefficient (*r*) between the *x*-axis and *y*-axis are shown, respectively. Genes of interest are shown in color. Housekeeping genes *Actb* (beta-actin) and *Gapdh* are shown as controls. (E) A2lox ES cells with DOX-inducible expression of ETV2-mCherry fusion protein were transferred to feeder-free condition and cultured in the presence of 0.5 µg/ml of DOX, simultaneously treated with or without PP2 (5 µM) for 24 h. Cells were then analyzed for FLK1 and mCherry levels using flow cytometry. The horizontal red lines mark presumptive ETV2 threshold. The statistics summary is shown on the right. ***$P<0.001$ (Student's *t*-test). Mean±s.d.

smooth muscle branch (gene cluster a). Importantly, all the smooth muscle transcription factor genes, except *Tbx2*, were already extensively expressed in the *Flk1*+ mesoderm and constituted a major characteristic of the *Flk1*+ mesoderm (Fig. S6B, in the black dashed box). These results suggest that smooth muscle might be the default fate of the *Flk1*+ mesoderm.

We investigated further populations 0, 3, 10, 12 and 13 (as labeled in Fig. 1B), corresponding to the bifurcation of hemangiogenic and smooth muscle fates (Fig. 6D). Expression of *Bmp4*, which can enhance vascular smooth muscle (VSM) development (Cheung et al., 2012), was dramatically upregulated in the smooth muscle branch [from population 0 (black) to population 13 (yellow)], but not in population 12 (green), where *Etv2* initiates specification of the hemangiogenic fate (Fig. 6D). *Hand1* and *Foxf1* were dramatically downregulated in population 12 compared with the upstream population 3, implying an active repression of them by *Etv2* (Fig. 6E). Consistent with this, by analyzing the work of Wareing and colleagues (Wareing et al., 2012) we found that re-expression of *Etv2* in *Etv2*-knockout EBs inhibited expression of *Hand1* and *Foxf1* (Fig. 6F). These results suggest that *Etv2* represses the default smooth muscle program when specifying the hemangiogenic lineage from *Flk1*+ mesoderm.

Flk1+ mesoderm characteristics are enriched in smooth muscle tissues *in vivo*

To understand better the relationship between *Flk1*+ mesoderm and the smooth muscle cell program, we further analyzed the overall expression score of the group b genes in Fig. 6B (listed in Table S3). The group b gene score reflected the dynamics of *Flk1*+ mesoderm identity in day 4 EB data (Fig. S7A). The cardiac lineage progenitors also had a high group b gene score; as such, we termed it the lateral plate mesoderm score, 'LPM-score' (both *Flk1*+ mesoderm and *Flk1*− cardiac tissues arise from lateral plate mesoderm). If the *Flk1*+ mesoderm's default fate is indeed smooth muscle lineage *in vivo*, it was expected that *Flk1*+ mesoderm identity would persist in the smooth muscle lineage. To determine whether this is the case, we performed scRNA-seq of a mixed cell pool of E9.5 and E10.5 mouse yolk sac, an extra-embryonic tissue that generates the first-arising hemangiogenic lineages and vasculature structure (Fig. 7A). We chose E9.5 and E10.5 yolk sac to ensure sufficient representation of fully committed smooth muscle lineage, because published single cell analysis of early mouse embryos from implantation to E8.5 (Ibarra-Soria et al., 2018; Pijuan-Sala et al., 2019) did not have good coverage of extra-embryonic smooth muscle tissues. We found that the LPM-score was particularly high in smooth muscle cells of the mouse yolk sac [VSM; Fig. 7B], exemplified by the co-expression of
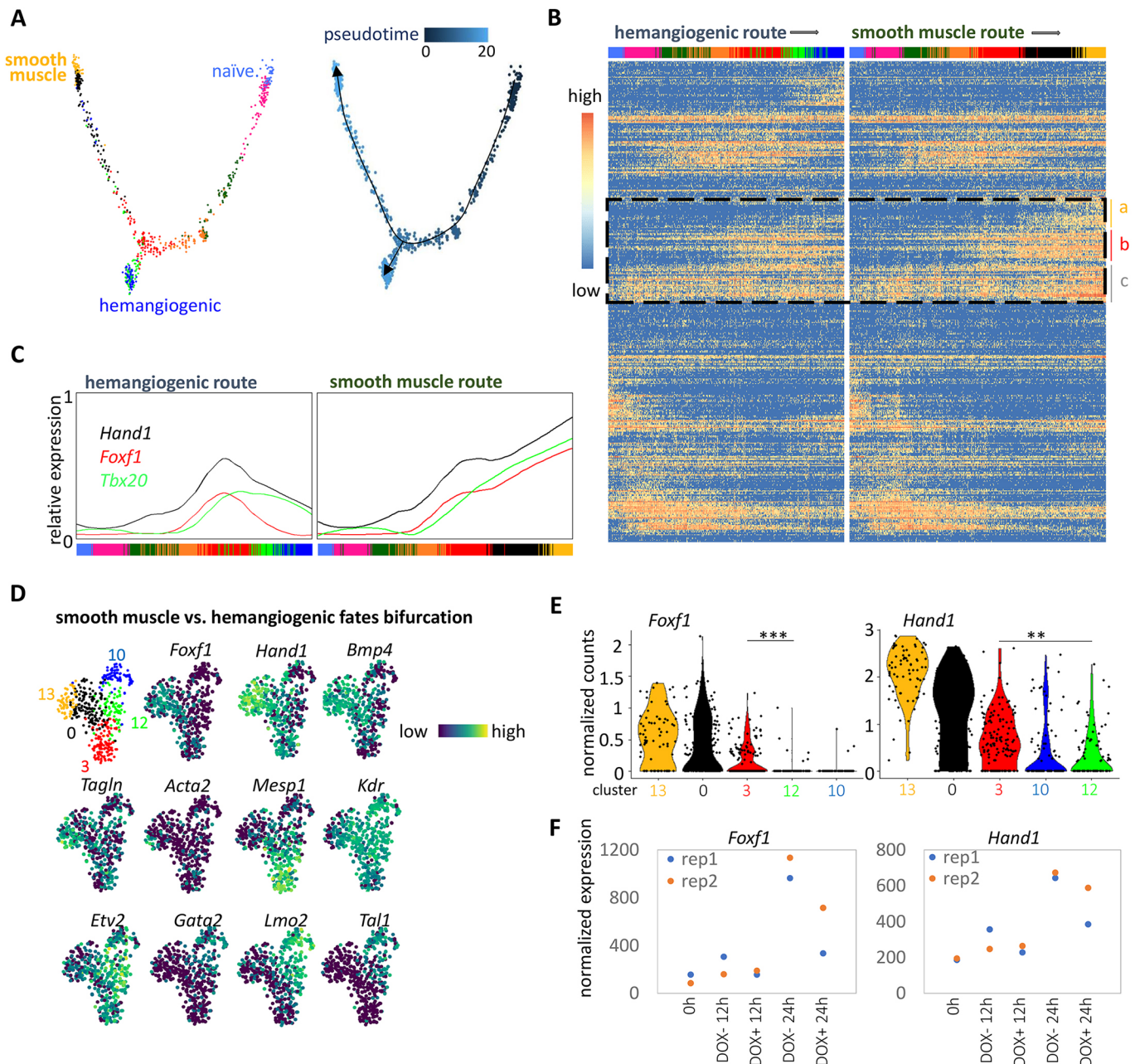
Fig. 6. 'Default' smooth muscle fate of *Flk1*⁺ mesoderm. (A) Left: Re-clustering of populations in the routes of smooth muscle and hemangiogenic lineage development. Clusters 0, 3, 4, 8, 9, 10, 12, 13 and 17 in Fig. 1B were picked, and the first two components of PCA are shown. Right: Pseudo-time analysis of the re-clustered cells. (B) The comparison of the most variable 427 genes' dynamics along the smooth muscle or hemangiogenic routes is shown. The genes were clustered based on Pearson's correlation. Color bars indicate cells' original identities. a, b and c, shown in a dashed line box, mark three groups of genes that are mostly enriched in the smooth muscle branch. (C) Expression dynamics of indicated genes along the pseudo-time lines of hemangiogenic and smooth muscle lineage development are shown. The expression values were smoothened. The colors of curves and gene names are consistent. (D) Clusters 0, 3, 10, 12 and 13, corresponding to the smooth muscle/hemangiogenic lineage bifurcation fork in Fig. 1B were picked and re-clustered. Cells are colored according to the expression levels of the indicated genes. (E) Violin plots comparing indicated genes' expression in different populations. **P<0.01; ***P<0.001 (Student's *t*-test). (F) Important smooth muscle transcription factors respond to *Etv2* overexpression. Day 2.5 *Etv2*-knockout EBs were induced for exogenous *Etv2* expression using DOX (Wareing et al., 2012). Two replicates of samples corresponding to DOX induction for 0 h, 12 h or 24 h were collected for microarray analysis.

the smooth muscle actin gene *Acta2* and the transcription factors *Foxf1* and *Hand1*, and the absence of the expression of the hematopoietic and endothelial transcription factor *Tal1* (Fig. 7C, Fig. S7B). It has been reported that the extra-embryonic amnion expresses smooth muscle genes (Fleury et al., 2015). In recently reported scRNA-seq of E8.25 mouse embryos (Ibarra-Soria et al., 2018), the amnion population had a high LPM-score and expressed smooth muscle markers (Fig. 7D-F, Fig. S7C,D). Consistent with

this, the 72 marker genes shared by day 4 EB smooth muscle, yolk sac VSM, and amnion were enriched with smooth muscle function annotation (Fig. 7G,H). Collectively, smooth muscle lineage might be the default fate of *Flk1*⁺ mesoderm.

DISCUSSION

It has been reported that single *Flk1*⁺ mesoderm cells can form both VSM and endothelial cells in culture and *in vivo* (Yamashita et al.,
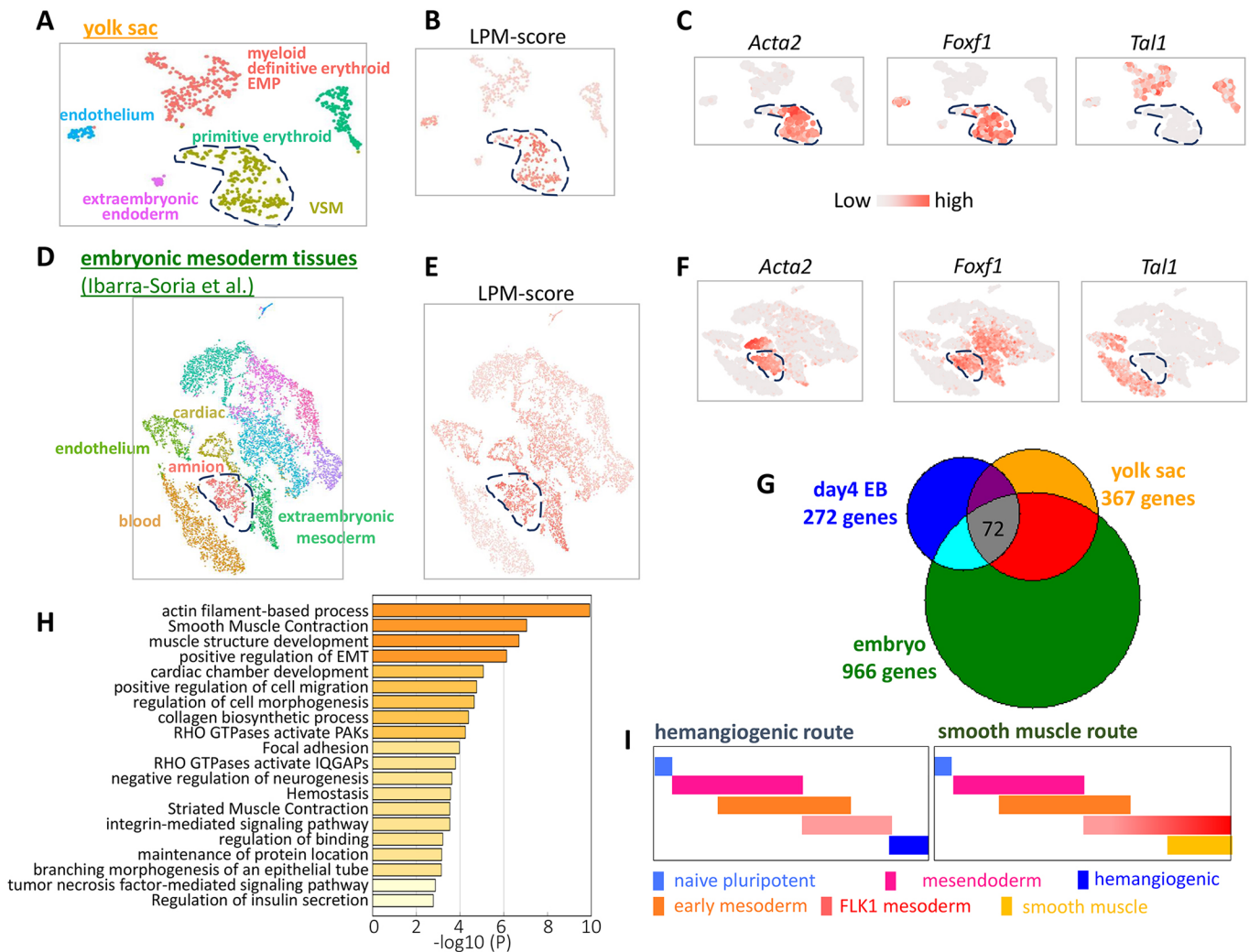
Fig. 7. Characteristics of *Flk1*⁺ mesoderm are enriched in smooth muscle tissue *in vivo*. (A) t-SNE plot of scRNA-seq analysis of yolk sac. (B) LPM-score gene expression is apparent in smooth muscle cells of the yolk sac. (C) Expression of the indicated genes in smooth muscle cells of the yolk sac. (D) Re-analysis of E8.25 embryo scRNA-seq data published by Ibarra-Soria et al. (2018). Only mesoderm tissues are shown. (E,F) Expression of the indicated genes in Ibarra-Soria et al.'s mesoderm population. Blue dashed line encircles the smooth muscle population of the yolk sac in A-C and the embryonic population defined as amnion by Ibarra-Soria et al. (a population expressing smooth muscle lineage genes) in D-F. (G) Comparison of smooth muscle marker genes from the indicated datasets. (H) Annotation of the shared 72 smooth muscle genes by the three different datasets. (I) Model of gene module dynamics along smooth muscle and hemangiogenic lineage development. Hemangiogenic lineage develops through a gradual shift of gene modules. Note that the FLK1 mesoderm gene module is further enforced in smooth muscle lineage, implying a default smooth muscle fate of FLK1 mesoderm.

2000). Moreover, sorted FLK1⁺ cells are able to form cardiomyocytes and even skeletal muscle (Liu et al., 2012; Sakurai et al., 2006). However, the lineage relationships among cardiac, VSM, skeletal muscle and hemangiogenic tissues concerning *Flk1*⁺ mesoderm allocation remain elusive. Our single cell transcriptome profiling of EB cells revealed a continuous differentiation trajectory. We showed that the cardiac branch mainly segregates out from *Pdgfra*⁺*Flk1*⁻ nascent mesoderm. Unexpectedly, smooth muscle and hemangiogenic lineages have the closest lineage relationship and directly bifurcate from a common group of *Flk1*-expressing cells. Notably, most *Flk1*⁺ mesoderm stage markers were further upregulated in the smooth muscle branch, suggesting that smooth muscle might be the default fate of the *Flk1*-expressing mesoderm. We propose that ETV2 drives the branching of the hemangiogenic lineage from the smooth muscle fate. Expression of cell adhesion signaling coincides with ETV2 target gene activation and hemangiogenic lineage specification. Consistent with this,

inhibition of SRC, a kinase important for cell adhesion signaling, was sufficient to repress hemangiogenic lineage skewing induced by *Etv2* overexpression. Meanwhile, crucial smooth muscle transcription factors, *Hand1* and *Foxf1*, become actively downregulated. Notably, neither *Hand1* nor *Foxf1* are direct ETV2 target genes (Liu et al., 2015). It will be important in the future to explore how this repression is achieved.

Hemangiogenic lineage development can be clearly annotated by a series of transcription factor modules, therefore providing strong molecular support for the current definition of early embryo developmental stages (Fig. 4C). Cells in the gastrulation stages, after exiting pluripotency and before entering the hemangiogenic state, are overall similar to each other in transcriptome, and the underlying transcription factor modules show a 'relay'-like highly overlapping pattern. Consistent with this, FLK1⁻PDGFRα⁻, FLK1⁻PDGFRα⁺ and FLK1⁺PDGFRα⁺ populations in early differentiation of ES cells are reversible/interchangeable (Sakurai et al., 2006). These results

suggest that exit from the naïve pluripotent state and activation of the hemangiogenic program might be the two rate-limiting steps in hemangiogenic lineage development, and that the intermediate gastrulation stages are plastic and adaptable to multiple lineage fates through combinational usage of limited regulatory circuits (Fig. 7I).

In summary, our work describes a continuous process of a mesoderm lineage tree leading to hemangiogenic and smooth muscle fates and the underlying molecular networks, thereby providing comprehensive information on early embryo and blood/blood vessel development. These findings are fundamental for the study of basic cell fate determination and gene network structures, and for designing more effective strategies to generate hematopoietic and endothelial cells for regenerative medicine.

MATERIALS AND METHODS
Mouse ES cell culture and differentiation
Mouse ES cells were maintained and differentiated as previously reported (Liu et al., 2012; Zhao and Choi, 2017). Briefly, ES cells were maintained on mouse embryo fibroblast (MEF) feeder cells in Dulbecco-modified Eagle medium containing 15% fetal bovine serum, 100 units/ml LIF, 1× MEM Non-Essential Amino Acids Solution (Gibco), 1× Glutamax Supplement (Gibco) and $4.5×10^{-4}$ M 1-thioglycerol (MTG, Sigma-Aldrich). For feeder-free culture, ES cells were maintained in a gelatin-coated dish in Iscove's modified Dulbecco medium (IMDM) with the same supplements used for maintenance. For EB differentiation, ES cells were first cultured in a feeder-free condition for 2 days. After culture, single cell suspensions were prepared and 8000 cells were added per ml of differentiation medium containing IMDM, 15% pre-screened fetal calf serum, 1× Glutamax, 50 μg/ml ascorbic acid and $4.5×10^{-4}$ M MTG. Differentiation was carried out on a bacterial Petri dish. DOX-inducible ETV2 (Liu et al., 2012) or ETV2-mCherry (Zhao and Choi, 2017) ES cells have been previously described in detail. Briefly, coding sequences of ETV2 or ETV2-mCherry were targeted into the *Rosa26* locus by homologous recombination using A2Lox ES cells carrying the TRE-loxP-Δneo inducible target locus (Iacovino et al., 2009). The SRC inhibitor PP2 (Sigma-Aldrich) and DOX (Sigma-Aldrich) treatment were performed as indicated in the text.

scRNA-seq
Yolk sacs were dissected from equal numbers of mouse E9.5 and E10.5 embryos, mixed together and dissociated with 0.25% collagenase at 37°C for 30 min. The cells were stored at −80°C in fetal bovine serum with 10% DMSO. When the 10x Genomics instrument was ready, the cells were thawed and washed, and stained using antibody against Ter119 (Ly76) (1:200, 12-5921-82, eBioscience). Dead cells and Ter119+ cells were excluded by sorting to enrich live non-erythroid cells.

Day 4 EBs were dissociated with Accutase solution (Sigma-Aldrich) for 5 min into single cells. The single cell suspension was washed and re-suspended in PBS.

Single cell suspension at 300 cells/μl in PBS were subjected to Chromium 10x Genomics library construction and HiSeq2500 sequencing (The Genome Technology Access Center, Washington University in St. Louis). The sequenced reads were mapped to the GRCm38 assembly using Cell Range 2.0.1 (10x Genomics).

scRNA-seq data quality control
The scRNA-seq output was imported into Seurat 1.4 (Butler et al., 2018), and genes expressed in at least three cells were kept for analysis. Cells with more than 5% mitochondria reads or fewer than 2000 unique genes detected were filtered out; 1848 cells, which had 4373 genes/27,272 unique mRNAs detected per cell in average, were kept for further analysis. Gene expression counts were normalized and log transformed.

scRNA-seq data visualization
According to the average expression and dispersion gene expression of each gene, we selected the 427 most variable genes for further analysis of the remaining cells. nUMIs (number of detected UMIs in each cell) were re-scaled to regress the effects of detection depth of each cell and mitochondria reads. The re-scaled data were run for principal components analysis (PCA) to reduce dimensions. According to the PCA results, the cells were clustered into 18 populations, and also used for t-SNE presentation. Cells in selected populations were imported into Monocle 2 (Trapnell et al., 2014) for re-clustering, PCA plotting and pseudo-time ordering.

For hierarchical clustering and heatmap plotting of selected genes or cells, the R package 'pheatmap' was used. All scRNA-seq data analyses were finished in R environment.

For cell cycle score and phase assignment with Seurat, the 'CellCycleScoring' function was invoked.

For LPM-score gene analysis, group b genes in Fig. 6B were processed by invoking the 'AddModuleScore' function in Seurat.

For SPRING visualization, a matrix of expression counts of the 427 most variable genes versus the filtered cells was uploaded (https://kleintools.hms.harvard.edu/tools/spring.html) per the guidance (Weinreb et al., 2018). Cell cluster information from Seurat analysis was also loaded for viewing.

For RNA velocity estimation, the bam file in the 10x Genomics output data was first re-counted with the velocyto counting pipeline in Python according to the manual (La Manno et al., 2018). The generated loom file was loaded to velocyto.R. Only the counts of the 427 most variable genes were used for RNA velocity estimation. Finally, coordinates from Seurat's t-SNE analysis were used to embed the velocity results.

For gene regulatory network analysis, the R package GENIE3 (Huynh-Thu et al., 2010) was used. To simplify the results, only transcription-related genes in the 427 most variable genes were used to calculate the gene-gene link matrix, and only the top 250 weighted link were adopted for visualization. To view the network, the R package igraph was used.

For Zipcode mapping, the mean expression of all the genes in chosen cells was uploaded to www.itranscriptome.org/ according to the developer's guidance (Peng et al., 2016).

Identification of genes required for ETV2-target gene activation
Cells within cluster 12 were divided into two groups, one with *Tal1* expression above the mean *Tal1* level across the whole population, and the other with *Tal1* expression below the mean level. Cells in the first group (with gene *x* expression above its mean level) were counted separately from cells in the second group (with gene *x* expression below its mean level), and the sum of the two counts was divided by the total cell number of cluster 12. This ratio was considered as the possibility of gene *x* following *Tal1*. In fact, close gene-gene correlation can be attributed to three possibilities (with *Tal1* used as an example): (1) gene *x* lies upstream of *Tal1* and is required for its expression, (2) gene *x* is TAL1's target, (3) both gene *x* and *Tal1* are ETV2's target and thus respond to ETV2 in a similar way. However, in the population in which expression of *Tal1* has just been initiated, it is not likely that TAL1 activates its own target genes yet, reducing the likelihood of possibility 2. We can refer to a gene's expression pattern in the differentiation trajectory to test possibility 3. In fact, most of the genes tightly following *Tal1* and *Lmo2* were already extensively activated before ETV2 started to function.

Dye dilution assays
Etv2-tdTomato ES cells (Zhao and Choi, 2017) were dissociated into single cell suspension and incubated with 1 μM of VPD450 (BD Horizon) in PBS for 10 min at room temperature. Cells were then aggregated for normal differentiation.

Microarray data analysis
Wareing and colleagues' microarray data (Wareing et al., 2012) were re-analyzed using the Expression Console software (Affymetrix).

Gene functional annotation
Gene functional annotation was performed using Metascape (http://metascape.org/gp/index.html) (Tripathi et al., 2015).

Flow cytometry

Single cells were incubated with anti-mouse FLK1 (1:400, 136406, BioLegend) and anti-mouse PDGFRα (1:500, 135908, BioLegend). Data were acquired on LSR-Fortessa flow cytometer (BD Biosciences) and analyzed using the FlowJo (Treestar) software.

Statistical analysis

The results of flow cytometry were analyzed using Student's *t*-test. $P<0.05$ was considered significant.

Data availability

The data generated in this study can be downloaded from the NCBI Gene Expression Omnibus under accession number GSE130146.

Supplementary information

Supplementary information available online at
http://dev.biologists.org/lookup/doi/10.1242/dev.182097.supplemental

References

Adams, J. C. and Watt, F. M. (1993). Regulation of development and differentiation by the extracellular matrix. *Development* **117**, 1183-1198.

Butler, A., Hoffman, P., Smibert, P., Papalexi, E. and Satija, R. (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411-420. doi:10.1038/nbt.4096

Cabodi, S., del Pilar Camacho-Leal, M., Di Stefano, P. and Defilippi, P. (2010). Integrin signalling adaptors: not only figurants in the cancer story. *Nat. Rev. Cancer* **10**, 858-870. doi:10.1038/nrc2967

Cheung, C., Bernardo, A. S., Trotter, M. W. B., Pedersen, R. A. and Sinha, S. (2012). Generation of human vascular smooth muscle subtypes provides insight into embryological origin-dependent disease susceptibility. *Nat. Biotechnol.* **30**, 165-173. doi:10.1038/nbt.2107

Chu, L.-F., Leng, N., Zhang, J., Hou, Z., Mamott, D., Vereide, D. T., Choi, J., Kendziorski, C., Stewart, R. and Thomson, J. A. (2016). Single-cell RNA-seq reveals novel regulators of human embryonic stem cell differentiation to definitive endoderm. *Genome Biol.* **17**, 173. doi:10.1186/s13059-016-1033-x

Ema, M., Takahashi, S. and Rossant, J. (2006). Deletion of the selection cassette, but not cis-acting elements, in targeted Flk1-lacZ allele reveals Flk1 expression in multipotent mesodermal progenitors. *Blood* **107**, 111-117. doi:10.1182/blood-2005-05-1970

Farrell, J. A., Wang, Y., Riesenfeld, S. J., Shekhar, K., Regev, A. and Schier, A. F. (2018). Single-cell reconstruction of developmental trajectories during zebrafish embryogenesis. *Science* **360**, eaar3131. doi:10.1126/science.aar3131

Fleury, M., Eliades, A., Carlsson, P., Lacaud, G. and Kouskoff, V. (2015). FOXF1 inhibits hematopoietic lineage commitment during early mesoderm specification. *Development* **142**, 3307-3320. doi:10.1242/dev.124685

Gong, W., Rasmussen, T. L., Singh, B. N., Koyano-Nakagawa, N., Pan, W. and Garry, D. J. (2017). Dpath software reveals hierarchical haemato-endothelial lineages of Etv2 progenitors based on single-cell transcriptome analysis. *Nat. Commun.* **8**, 14362. doi:10.1038/ncomms14362

Han, X., Chen, H., Huang, D., Chen, H., Fei, L., Cheng, C., Huang, H., Yuan, G.-C. and Guo, G. (2018). Mapping human pluripotent stem cell differentiation pathways using high throughput single-cell RNA-sequencing. *Genome Biol.* **19**, 47. doi:10.1186/s13059-018-1426-0

Huynh-Thu, V. A., Irrthum, A., Wehenkel, L. and Geurts, P. (2010). Inferring regulatory networks from expression data using tree-based methods. *PLoS ONE* **5**, e12776. 10.1371/journal.pone.0012776

Iacovino, M., Hernandez, C., Xu, Z., Bajwa, G., Prather, M. and Kyba, M. (2009). A conserved role for Hox paralog group 4 in regulation of hematopoietic progenitors. *Stem Cells Dev.* **18**, 783-792. doi:10.1089/scd.2008.0227

Ibarra-Soria, X., Jawaid, W., Pijuan-Sala, B., Ladopoulos, V., Scialdone, A., Jörg, D. J., Tyser, R. C. V., Calero-Nieto, F. J., Mulas, C., Nichols, J. et al. (2018). 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. doi:10.1038/s41556-017-0013-z

Kalkan, T. and Smith, A. (2014). Mapping the route from naive pluripotency to lineage specification. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **369**, 20130540. doi:10.1098/rstb.2013.0540

Kataoka, H., Takakura, N., Nishikawa, S., Tsuchida, K., Kodama, H., Kunisada, T., Risau, W., Kita, T. and Nishikawa, S.-I. (1997). Expressions of PDGF receptor alpha, c-Kit and Flk1 genes clustering in mouse chromosome 5 define distinct subsets of nascent mesodermal cells. *Dev. Growth Differ.* **39**, 729-740. doi:10.1046/j.1440-169X.1997.t01-5-00009.x

Kattman, S. J., Huber, T. L. and Keller, G. M. (2006). Multipotent flk-1+ cardiovascular progenitor cells give rise to the cardiomyocyte, endothelial, and vascular smooth muscle lineages. *Dev. Cell* **11**, 723-732. doi:10.1016/j.devcel.2006.10.002

La Manno, G., Soldatov, R., Zeisel, A., Braun, E., Hochgerner, H., Petukhov, V., Lidschreiber, K., Kastriti, M. E., Lonnerberg, P., Furlan, A. et al. (2018). RNA velocity of single cells. *Nature* **560**, 494-498. doi:10.1038/s41586-018-0414-6

Lescroart, F., Wang, X., Lin, X., Swedlund, B., Gargouri, S., Sànchez-Dànes, A., Moignard, V., Dubois, C., Paulissen, C., Kinston, S. et al. (2018). Defining the earliest step of cardiovascular lineage segregation by single-cell RNA-seq. *Science* **359**, 1177-1181. doi:10.1126/science.aao4174

Liu, F., Kang, I., Park, C., Chang, L.-W., Wang, W., Lee, D., Lim, D.-S., Vittet, D., Nerbonne, J. M. and Choi, K. (2012). ER71 specifies Flk-1+ hemangiogenic mesoderm by inhibiting cardiac mesoderm and Wnt signaling. *Blood* **119**, 3295-3305. doi:10.1182/blood-2012-01-403766

Liu, F., Li, D., Yu, Y. Y. L., Kang, I., Cha, M. J., Kim, J. Y., Park, C., Watson, D. K., Wang, T. and Choi, K. (2015). Induction of hematopoietic and endothelial cell program orchestrated by ETS transcription factor ER71/ETV2. *EMBO Rep.* **16**, 654-669. doi:10.15252/embr.201439939

Lugus, J. J., Park, C., Ma, Y. D. and Choi, K. (2009). Both primitive and definitive blood cells are derived from Flk-1+ mesoderm. *Blood* **113**, 563-566. doi:10.1182/blood-2008-06-162750

Mahlapuu, M., Ormestad, M., Enerback, S. and Carlsson, P. (2001). The forkhead transcription factor Foxf1 is required for differentiation of extra-embryonic and lateral plate mesoderm. *Development* **128**, 155-166.

Mohammed, H., Hernando-Herraez, I., Savino, A., Scialdone, A., Macaulay, I., Mulas, C., Chandra, T., Voet, T., Dean, W., Nichols, J. et al. (2017). Single-cell landscape of transcriptional heterogeneity and cell fate decisions during mouse early gastrulation. *Cell Rep.* **20**, 1215-1228. doi:10.1016/j.celrep.2017.07.009

Moignard, V., Woodhouse, S., Haghverdi, L., Lilly, A. J., Tanaka, Y., Wilkinson, A. C., Buettner, F., Macaulay, I. C., Jawaid, W., Diamanti, E. et al. (2015). Decoding the regulatory network of early blood development from single-cell gene expression measurements. *Nat. Biotechnol.* **33**, 269-276. doi:10.1038/nbt.3154

Morikawa, Y. and Cserjesi, P. (2004). Extra-embryonic vasculature development is regulated by the transcription factor HAND1. *Development* **131**, 2195-2204. doi:10.1242/dev.01091

Motoike, T., Markham, D. W., Rossant, J. and Sato, T. N. (2003). Evidence for novel fate of Flk1+ progenitor: contribution to muscle lineage. *Genesis* **35**, 153-159. doi:10.1002/gene.10175

Park, K.-S., Cha, Y., Kim, C.-H., Ahn, H.-J., Kim, D., Ko, S., Kim, K.-H., Chang, M.-Y., Ko, J.-H., Noh, Y.-S. et al. (2013). Transcription elongation factor Tcea3 regulates the pluripotent differentiation potential of mouse embryonic stem cells via the Lefty1-Nodal-Smad2 pathway. *Stem Cells* **31**, 282-292. doi:10.1002/stem.1284

Peng, G., Suo, S., Chen, J., Chen, W., Liu, C., Yu, F., Wang, R., Chen, S., Sun, N., Cui, G. et al. (2016). Spatial transcriptome for the molecular annotation of lineage fates and cell identity in mid-gastrula mouse embryo. *Dev. Cell* **36**, 681-697. doi:10.1016/j.devcel.2016.02.020

Pijuan-Sala, B., Griffiths, J. A., Guibentif, C., Hiscock, T. W., Jawaid, W., Calero-Nieto, F. J., Mulas, C., Ibarra-Soria, X., Tyser, R. C. V., Ho, D. L. L. et al. (2019). A single-cell molecular map of mouse gastrulation and early organogenesis. *Nature* **566**, 490-495. doi:10.1038/s41586-019-0933-9

Porcher, C., Swat, W., Rockwell, K., Fujiwara, Y., Alt, F. W. and Orkin, S. H. (1996). The T cell leukemia oncoprotein SCL/tal-1 is essential for development of all hematopoietic lineages. *Cell* **86**, 47-57. doi:10.1016/S0092-8674(00)80076-8

Sakurai, H., Era, T., Jakt, L. M., Okada, M., Nakai, S., Nishikawa, S. and Nishikawa, S.-I. (2006). In vitro modeling of paraxial and lateral mesoderm differentiation reveals early reversibility. *Stem Cells* **24**, 575-586. doi:10.1634/stemcells.2005-0256

Scialdone, A., Tanaka, Y., Jawaid, W., Moignard, V., Wilson, N. K., Macaulay, I. C., Marioni, J. C. and Göttgens, B. (2016). Resolving early mesoderm diversification through single-cell expression profiling. *Nature* **535**, 289-293. doi:10.1038/nature18633

Spangler, A., Su, E. Y., Craft, A. M. and Cahan, P. (2018). A single cell transcriptional portrait of embryoid body differentiation and comparison to progenitors of the developing embryo. *Stem Cell Res.* **31**, 201-215. doi:10.1016/j.scr.2018.07.022

Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., Lennon, N. J., Livak, K. J., Mikkelsen, T. S. and Rinn, J. L. (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* **32**, 381-386. doi:10.1038/nbt.2859

Tripathi, S., Pohl, M. O., Zhou, Y., Rodriguez-Frandsen, A., Wang, G., Stein, D. A., Moulton, H. M., DeJesus, P., Che, J., Mulder, L. C. F. et al. (2015). Meta- and orthogonal integration of influenza "OMICs" data defines a role for UBR4 in virus budding. *Cell Host Microbe* **18**, 723-735. doi:10.1016/j.chom.2015.11.002

Wagner, D. E., Weinreb, C., Collins, Z. M., Briggs, J. A., Megason, S. G. and Klein, A. M. (2018). Single-cell mapping of gene expression landscapes and lineage in the zebrafish embryo. *Science* **360**, 981-987. doi:10.1126/science.aar4362

Wareing, S., Mazan, A., Pearson, S., Göttgens, B., Lacaud, G. and Kouskoff, V. (2012). The Flk1-Cre-mediated deletion of ETV2 defines its narrow temporal requirement during embryonic hematopoietic development. *Stem Cells* **30**, 1521-1531. doi:10.1002/stem.1115

Weinreb, C., Wolock, S. and Klein, A. M. (2018). SPRING: a kinetic interface for visualizing high dimensional single-cell expression data. *Bioinformatics* **34**, 1246-1248. doi:10.1093/bioinformatics/btx792

Yamashita, J., Itoh, H., Hirashima, M., Ogawa, M., Nishikawa, S., Yurugi, T., Naito, M., Nakao, K. and Nishikawa, S.-I. (2000). Flk1-positive cells derived from embryonic stem cells serve as vascular progenitors. *Nature* **408**, 92-96. doi:10.1038/35040568

Zhao, H. and Choi, K. (2017). A CRISPR screen identifies genes controlling Etv2 threshold expression in murine hemangiogenic fate commitment. *Nat. Commun.* **8**, 541. doi:10.1038/s41467-017-00667-5

DEVELOPMENT

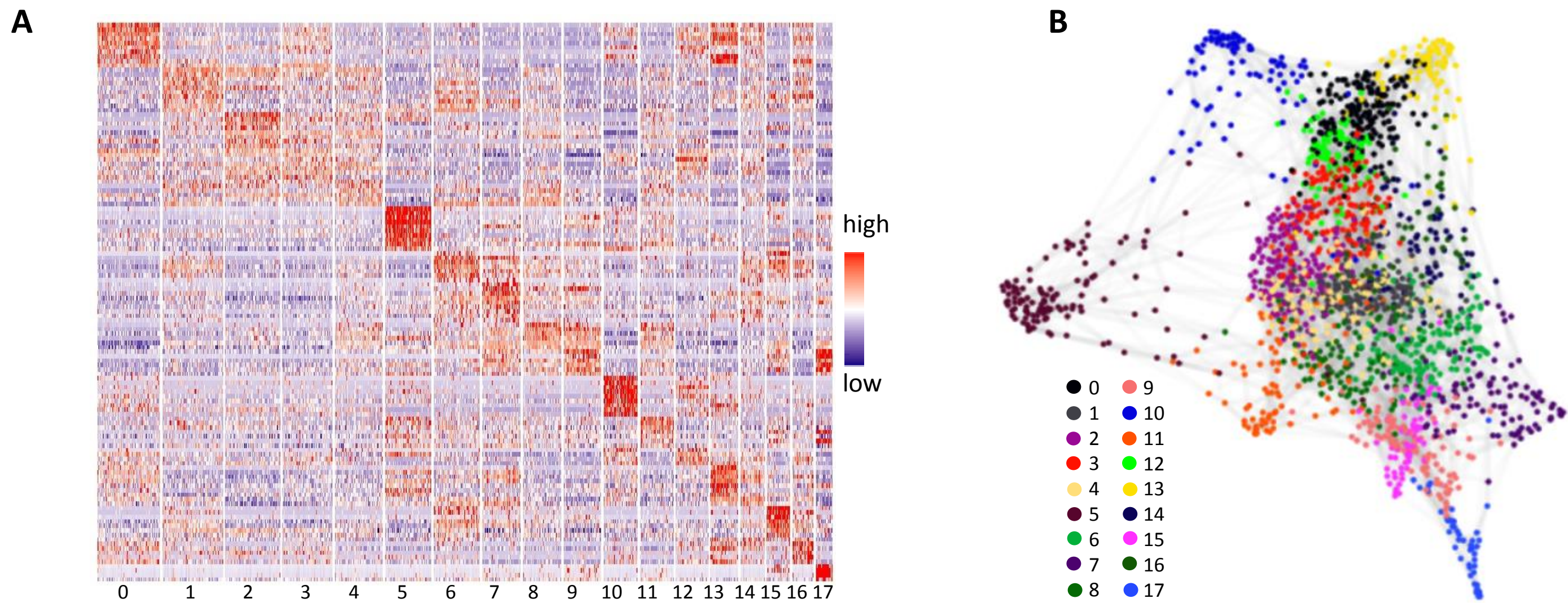# SUPPLEMENTARY FIGURES

A

B

Figure S1.

(A) Representative marker gene expression in each cluster. Only 10 representative marker genes at most are displayed for each cluster. (B) SPRING plot of the single cell transcriptome. Each cell is in the same color as it is in Figure 1B.
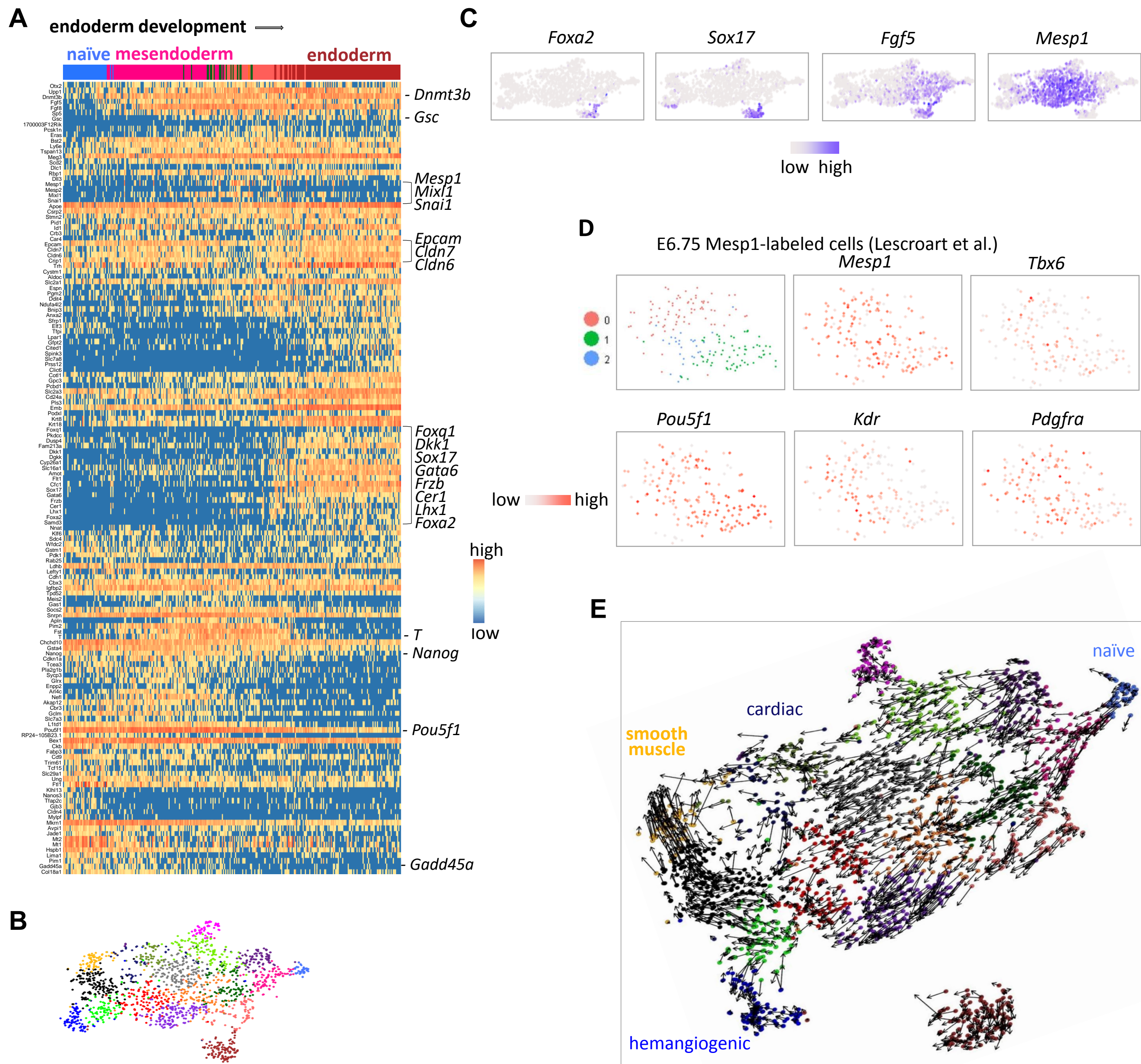
A endoderm development →

naïve mesendoderm endoderm

B

C *Foxa2* *Sox17* *Fgf5* *Mesp1*

low high

D E6.75 Mesp1-labeled cells (Lescroart et al.)

Mesp1 *Tbx6*

Pou5f1 *Kdr* *Pdgfra*

low high

E naïve

cardiac

smooth muscle

hemangiogenic

Figure S2.

(A) Expression dynamics of the 163 most variable genes along the endoderm developmental route. The genes were clustered based on Pearson's correlations. (B) t-SNE plot of the cells embedded with marker genes using less stringent selection. The cells' colors show their original identities as in Figure 1B. (C) Gene expression in the new t-SNE plot shown in B. (D) Expression of indicated genes in E6.75 *Mesp1*-labeled cells (re-analysis of Lescroart et al.'s data (E) RNA velocity estimation of the cells. The arrows indicate estimated differentiation trajectory of given cells. The colors of the cells are same as those in Figure 1B.
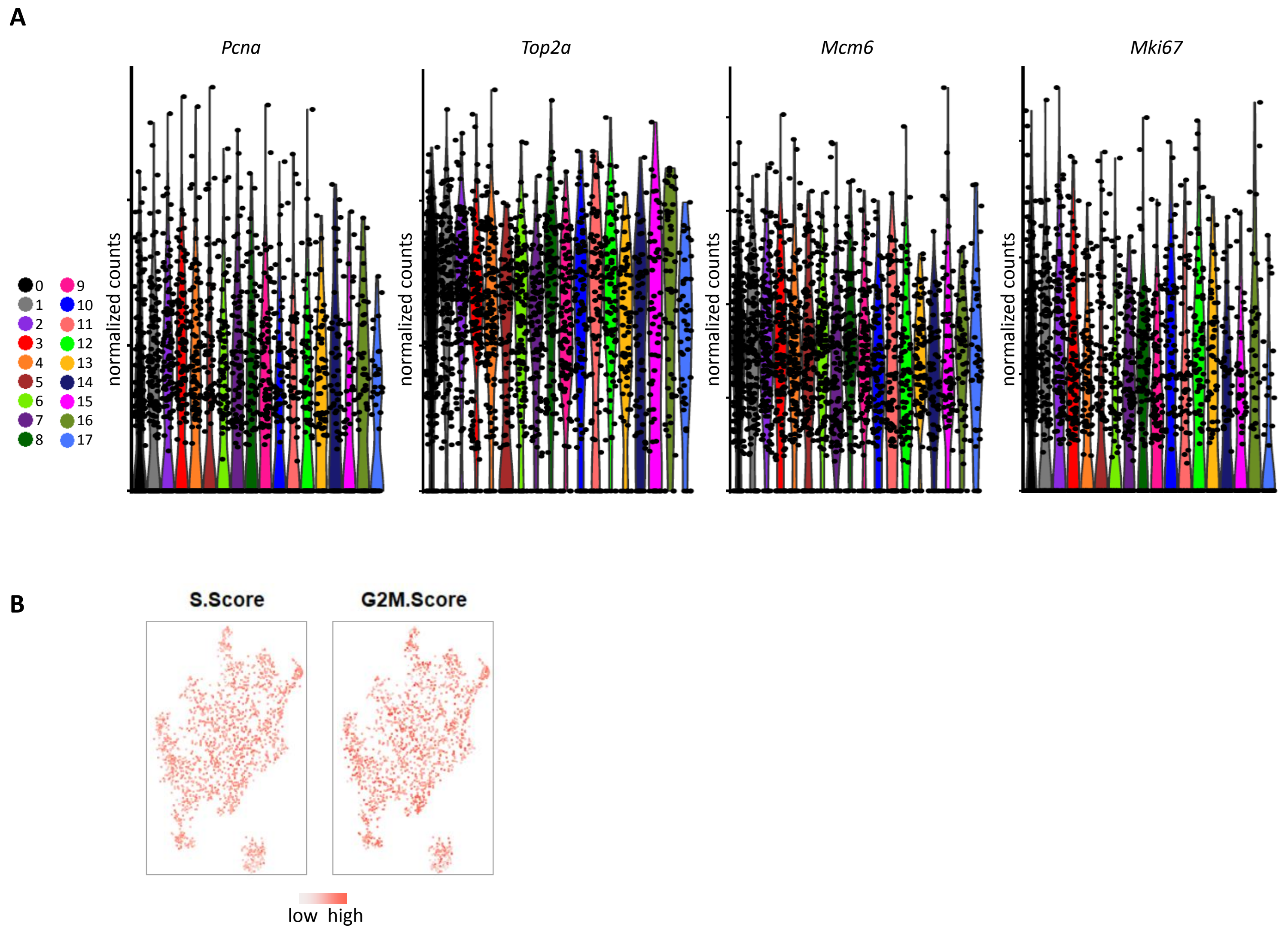
A

B

Figure S3.

(A) Violin plot of cell cycle genes in different clusters. (B) Cell cycle score of each cell.
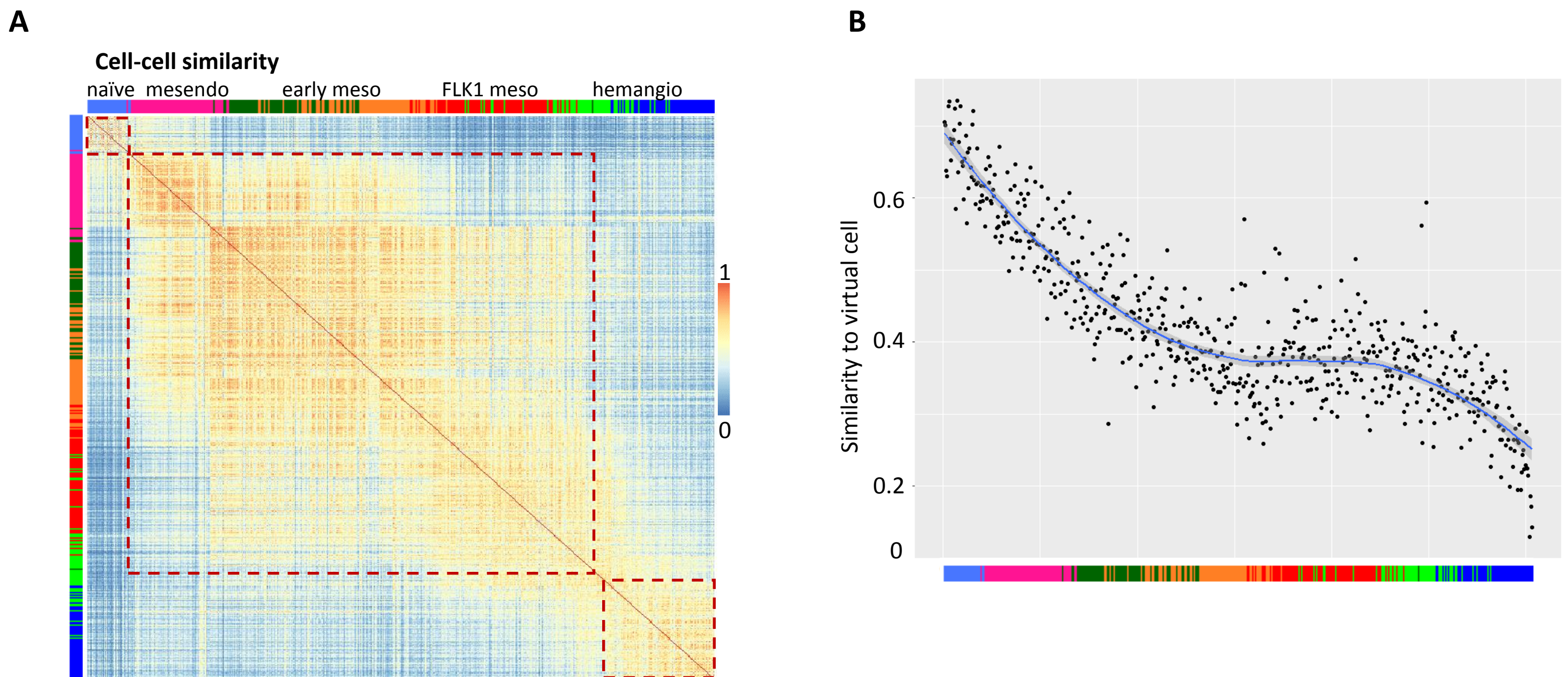
Figure S4.

(A)   ll-cell Pearson's correlations along the pseudo-time line based on the expression of 427 most variable genes.

The brown dash boxes mark the relatively isolated naïve pluripotent population, 'gastrulation' stage, and

hemangiogenic lineage. (B) The first 20 cells in the hemangiogenic lineage route were picked and their mean

expression was treated as a virtual cell. Each cell's similarity in the developmental route to this virtual cell was then

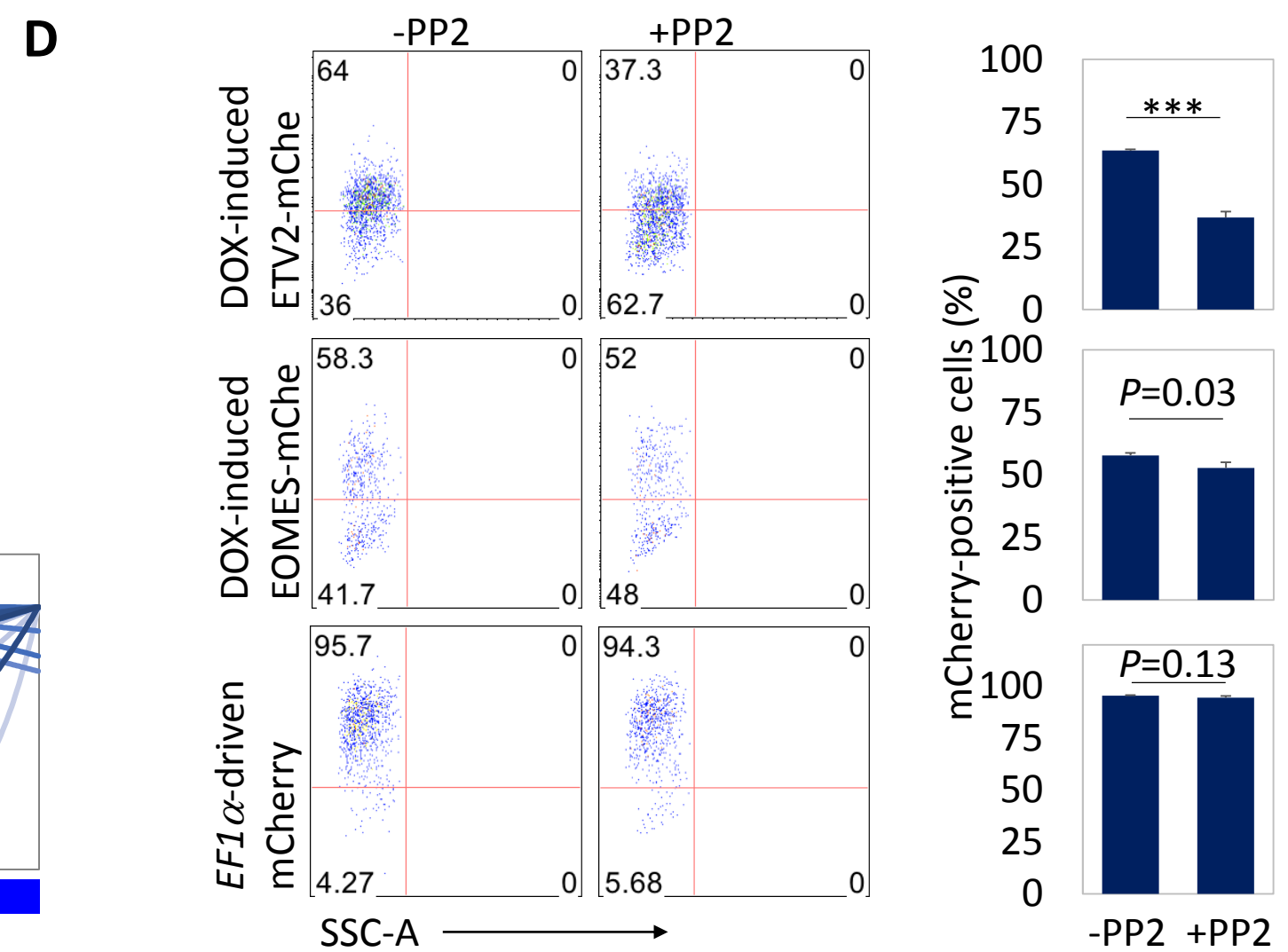indicated in the dot plot. Bars at the bottom show each cell's original identity.
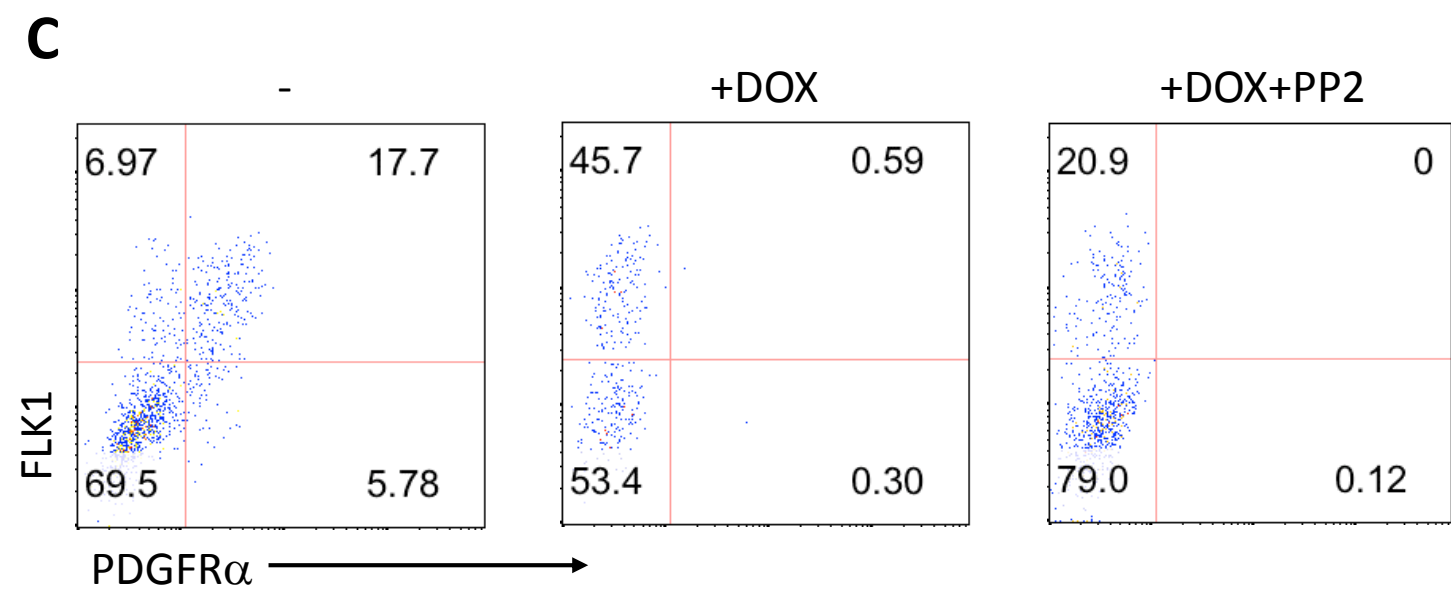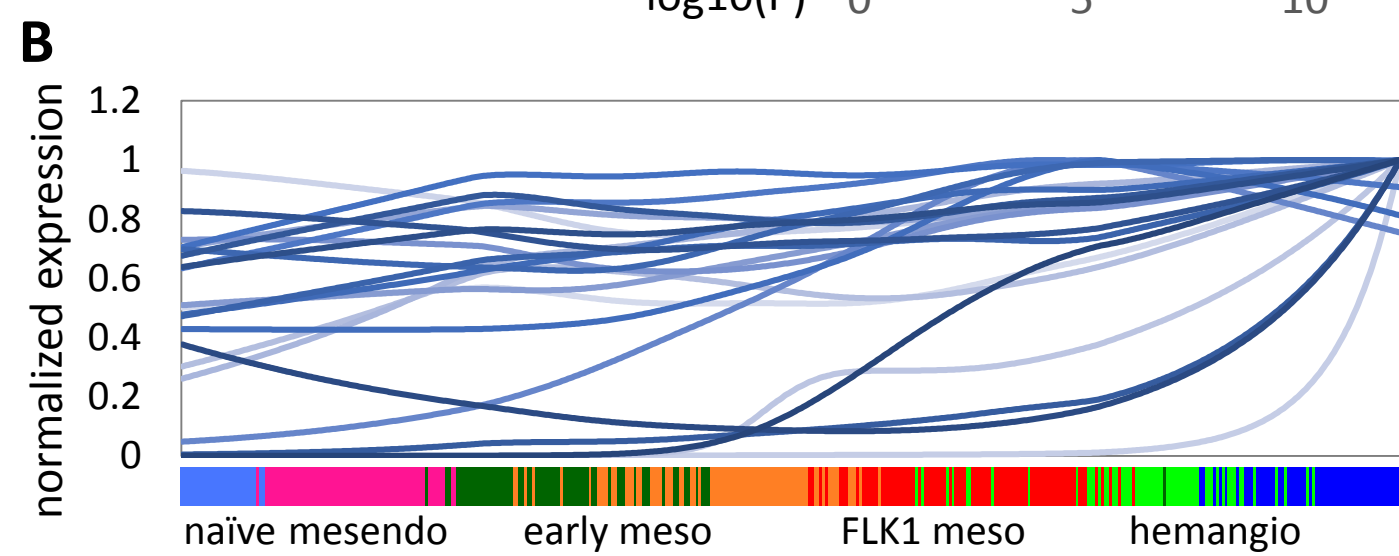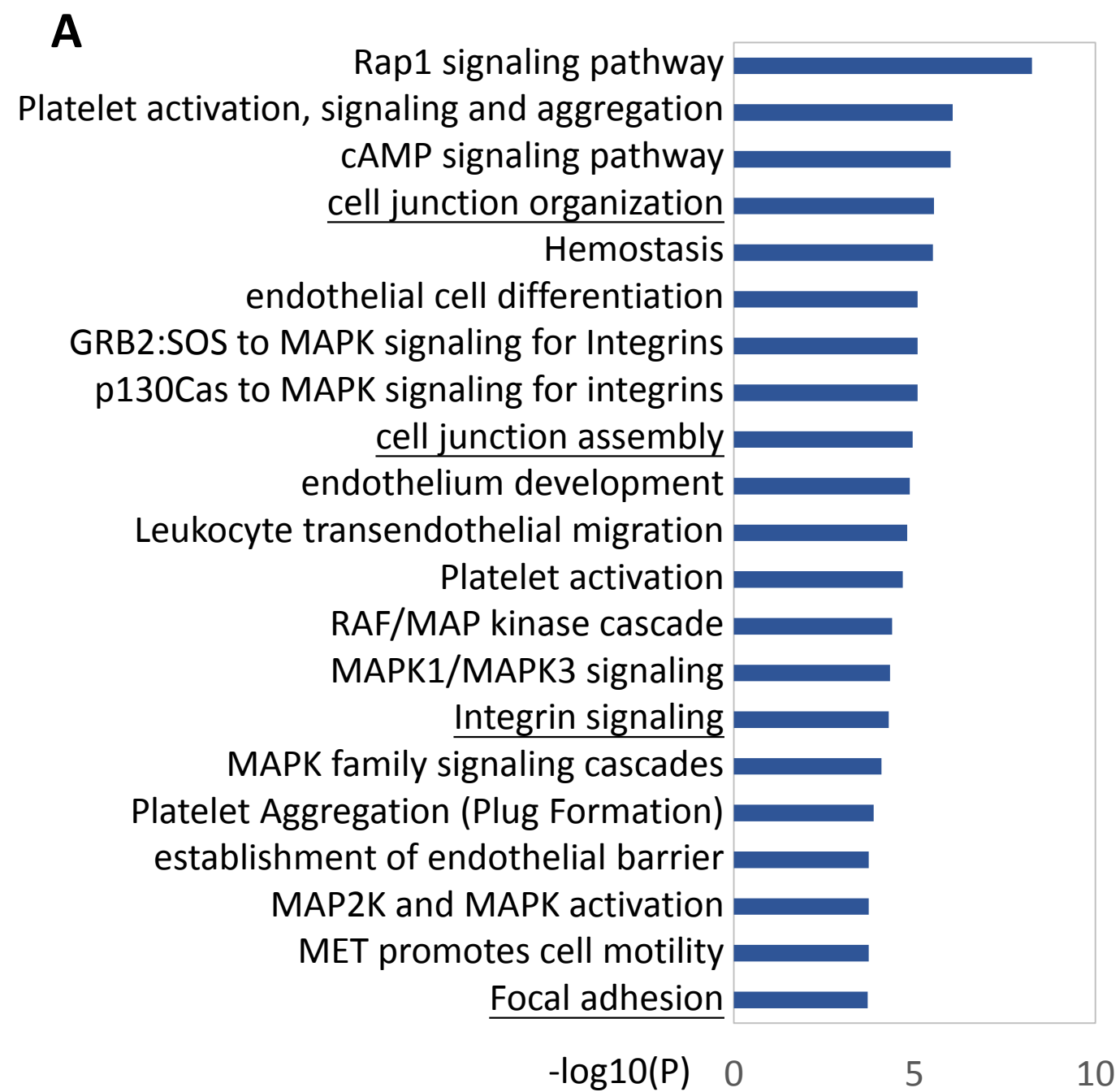
A

Rap1 signaling pathway
Platelet activation, signaling and aggregation
cAMP signaling pathway
cell junction organization
Hemostasis
endothelial cell differentiation
GRB2:SOS to MAPK signaling for Integrins
p130Cas to MAPK signaling for integrins
cell junction assembly
endothelium development
Leukocyte transendothelial migration
Platelet activation
RAF/MAP kinase cascade
MAPK1/MAPK3 signaling
Integrin signaling
MAPK family signaling cascades
Platelet Aggregation (Plug Formation)
establishment of endothelial barrier
MAP2K and MAPK activation
MET promotes cell motility
Focal adhesion

-log10(P)

B

normalized expression

naïve mesendo   early meso   FLK1 meso   hemangio

C

-        +DOX   +DOX+PP2

FLK1

PDGFRα

D

-PP2   +PP2

DOX-induced ETV2-mChe
DOX-induced EOMES-mChe
EF1α-driven mCherry

SSC-A

mCherry-positive cells (%)

-PP2   +PP2

Figure S5.

(A) The comparison of the top 100 genes that are most required for *Tal1* expression and the top 100 most required for *Lmo2* expression generated 66 in common. These 66 common genes were analyzed for functional enrichment. (B) Expression of 22 cell adhesion-related genes along the pseudo-time line in the hemangiogenic lineage development. The expression values were smoothened for visualization. (C) A2lox ES cells with DOX-inducible *Etv2* expression were differentiated in serum medium for 3 days, EBs were then treated with or without the SRC kinase inhibitor PP2 (5mM) for another 24 hours in the presence of 2mg/mL of DOX. EBs without PP2 or DOX treatment were used as control. EBs were collected and analyzed for surface marker expression as shown. (D) A2lox ES cells with DOX-inducible ETV2-mCherry or EOMES-mCherry expression, or *EF1α*-driven constitutive mCherry expression were cultured in ES cell medium, treated with or without the SRC kinase inhibitor PP2 (5mM) in the presence of 0.5mg/mL of DOX for 1 day. Cells were collected and analyzed using flow cytometry for mCherry and FLK1 expression. The statistics is shown on the right. ***, P<0.001.
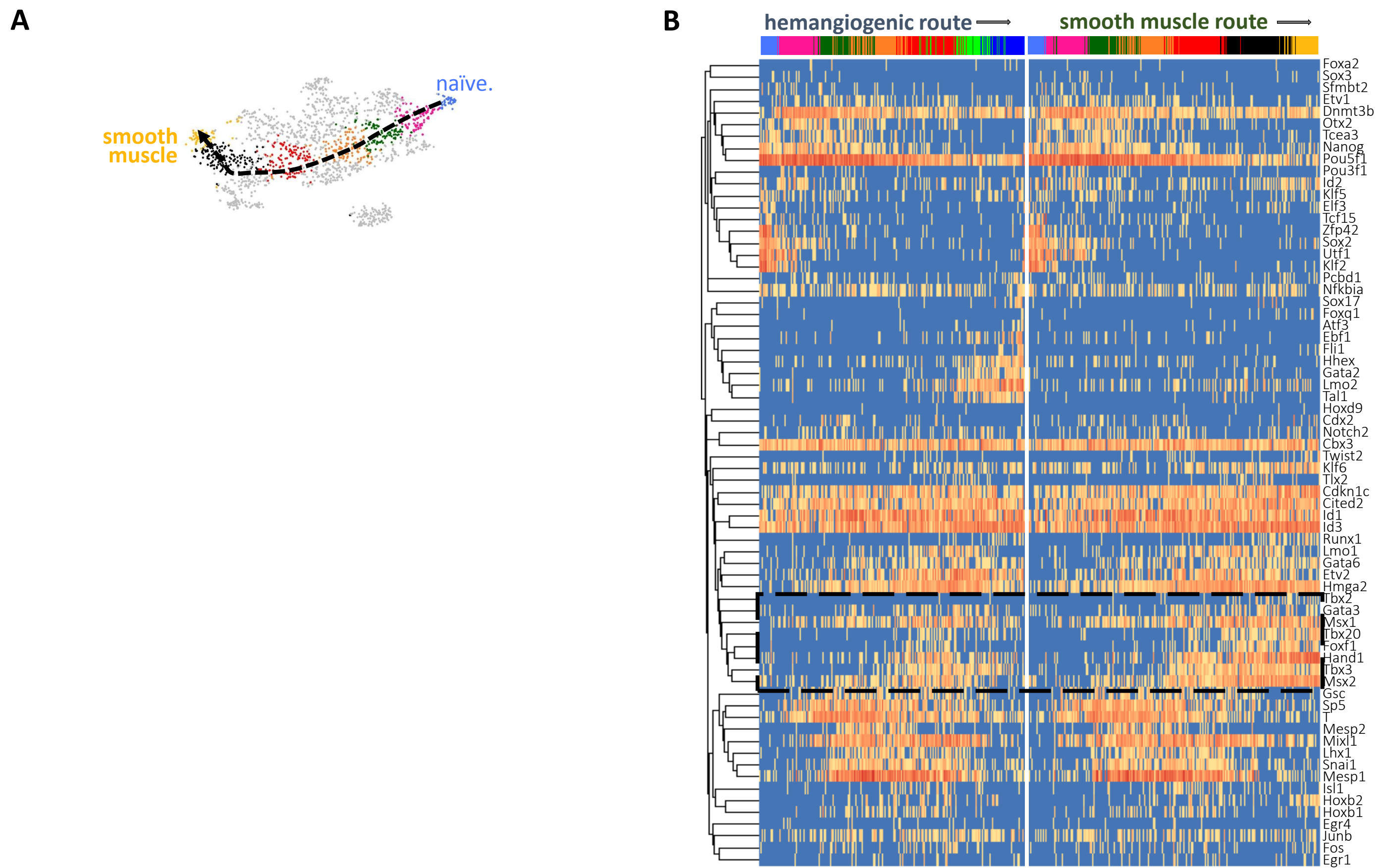
Figure S6.

(A) Reconstitution of the populations selected for smooth muscle developmental route. (B) Comparison of expression

dynamics of transcription regulation-related factors along smooth muscle vs. hemangiogenic lineage development.

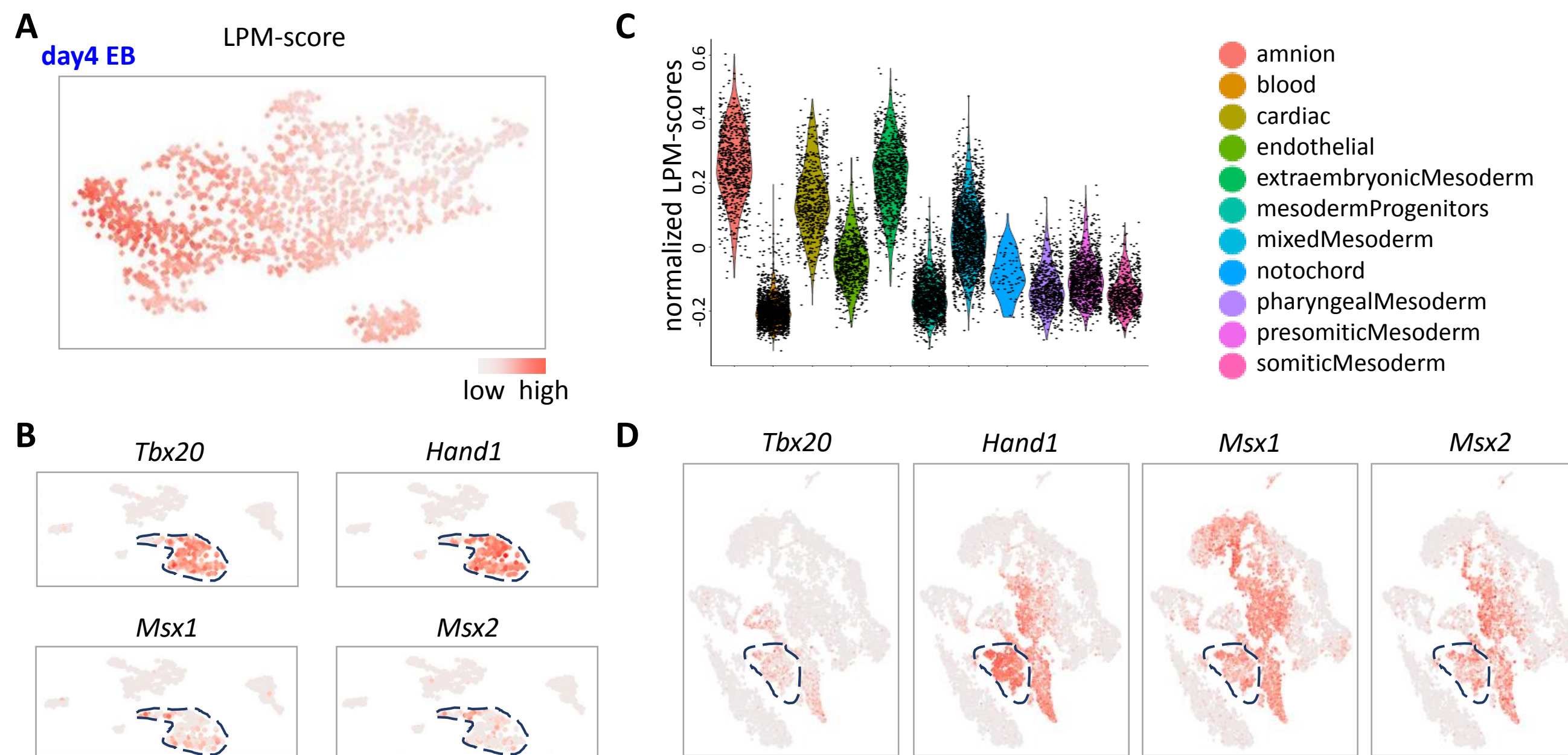The genes were clustered based on the Pearson's correlations.

Figure S7.

(A) LPM-scores of day 4 EB cells. (B) Expression of indicated genes in yolk sac cells. (C) Violin plot of LPM-scores in cells of different mesoderm tissues from E8.25 embryo. (D) Expression of indicated genes in E8.25 embryonic mesoderm cells.

Table S1. Marker genes detected in each population.

Click here to Download Table S1

Table S2. Genes approving Tal1, Lmo2 or Eomes in population12. (Highlighted genes are annotated as cell adhesion-related)

Click here to Download Table S2

Table S3. Genes assembled as LPM-meta gene

Click here to Download Table S3