HEART DEVELOPMENT

Defining the earliest step of cardiovascular lineage segregation by single-cell RNA-seq

Fabienne Lescroart,[1]* Xiaonan Wang,[2,3]* Xionghui Lin,[1]* Benjamin Swedlund,[1] Souhir Gargouri,[1] Adriana Sànchez-Dànes,[1] Victoria Moignard,[2,3] Christine Dubois,[1] Catherine Paulissen,[1] Sarah Kinston,[2,3] Berthold Göttgens,[2,3]†‡ Cédric Blanpain[1,4]†‡

Mouse heart development arises from *Mesp1*-expressing cardiovascular progenitors (CPs) that are specified during gastrulation. The molecular processes that control early regional and lineage segregation of CPs have been unclear. We performed single-cell RNA sequencing of wild-type and *Mesp1*-null CPs in mice. We showed that populations of *Mesp1* CPs are molecularly distinct and span the continuum between epiblast and later mesodermal cells, including hematopoietic progenitors. Single-cell transcriptome analysis of *Mesp1*-deficient CPs showed that Mesp1 is required for the exit from the pluripotent state and the induction of the cardiovascular gene expression program. We identified distinct populations of *Mesp1* CPs that correspond to progenitors committed to different cell lineages and regions of the heart, identifying the molecular features associated with early lineage restriction and regional segregation of the heart at the early stage of mouse gastrulation.

The mammalian heart is composed of different regions (ventricles, atria, and great vessels) and cell types, including cardiomyocytes (CMs), endocardial cells (ECs), smooth muscle cells (SMCs), and epicardial cells (EPs) (*1*). Heart development begins at gastrulation, during which CPs leave the primitive streak (PS) and migrate toward the anterolateral pole of the embryo (*2*). From embryonic day 6.25 (E6.25) to E7.25, *Mesp1* marks the early CPs within the PS, whereas a day later, *Mesp1* is expressed in the somites (*2*, *3*). *Mesp1*+ cells give rise to all heart cells, ECs of the aorta and brain, some muscles of the head and neck, as well as to few somitic derivatives and liver cells from its later expression (*2*, *4–6*). Temporally inducible *Mesp1* lineage tracing shows that at E6.5, *Mesp1*+ cells mark left ventricle (LV) progenitors, whereas the right ventricle, atria, outflow, and inflow tracts and head muscles arise from *Mesp1*+ cells at E7.25, which correspond respectively to the first and second heart fields (FHF and SHF, respectively). No somitic or liver derivatives were labeled at these early time points (*4*, *5*, *7*). In addition, most of the *Mesp1* CPs differentiate into either CMs or ECs, suggesting that lineage segregation occurs early during gastrulation (*4*, *5*). It remains unknown whether molecular heterogeneity between E6.5 and E7.25 *Mesp1*+ CPs reflects stochasticity in gene expression, transcriptional priming, or early lineage and regional segregation.

To investigate the molecular and cellular basis of the earliest stages of CP specification and diversification, we performed single-cell RNA sequencing (scRNA-seq) of *Mesp1* CPs at E6.75 and E7.25. To this end, *Mesp1-rtTA/tetO-H2B–GFP* (green fluorescent protein) mice were treated with doxycycline at different time points after fertilization in order to label only early *Mesp1*-expressing cells and no somitic derivatives (Fig. 1A), embryos were dissociated into single cells, and H2B-GFP–positive CPs were isolated by means of fluorescence-activated cell sorting (FACS) (fig. S1). A total of 172 and 341 *Mesp1* CPs at E6.75 and E7.25, respectively, were sequenced and analyzed further after passing through a stringent quality control pipeline (supplementary materials, materials and methods). We recently reported single-cell transcriptomes for E6.5 epiblast cells, as well as E7.25/7.5 Flk1-expressing progenitors (*8*). Visualization by use of dimensionality reduction techniques allowed us to order the cells along developmental progression and assign a time stamp to each cell, demonstrating that the *Mesp1*+ CPs at E6.75 and E7.25 likely represent a continuum of differentiation (Fig. 1, B and C, and fig. S2) (*8*).

To determine the role of Mesp1 in regulating the cardiovascular differentiation program and the heterogeneity of early CPs, we performed scRNA-seq of FACS-isolated *Mesp1*-expressing cells in *Mesp1* knockout (KO) context (fig. S3) (*2*). We sequenced transcriptomes of 85 single *Mesp1*-null cells isolated at E6.75, before the appearance of the developmental defect associated with *Mesp1* deficiency (Fig. 2A). Pseudotime analysis revealed that *Mesp1* KO cells presented a developmental block, being stuck in the gene expression program of epiblast cells (Fig. 2B). Principal components analysis showed that principal component 2 captured expression differences between wild-type (WT) and *Mesp1* KO cells (fig. S4), with 206 down-regulated and 136 up-regulated genes (table S1). We found a highly significant overlap for genes differentially expressed between WT and *Mesp1* KO cells in vivo and genes that are down- or up-regulated after

[1]Université Libre de Bruxelles, Laboratory of Stem Cells and Cancer, Brussels B-1070, Belgium. [2]Department of Haematology, Cambridge Institute for Medical Research, University of Cambridge, Cambridge CB2 0XY, UK. [3]Wellcome and Medical Research Council Cambridge Stem Cell Institute, University of Cambridge, Cambridge, UK. [4]WELBIO, Université Libre de Bruxelles, Brussels B-1070, Belgium.
*These authors contributed equally to this work. †These authors contributed equally to this work. ‡Corresponding author. Email: bg200@cam.ac.uk (B.G.); cedric.blanplain@ulb.ac.be (C.B.)
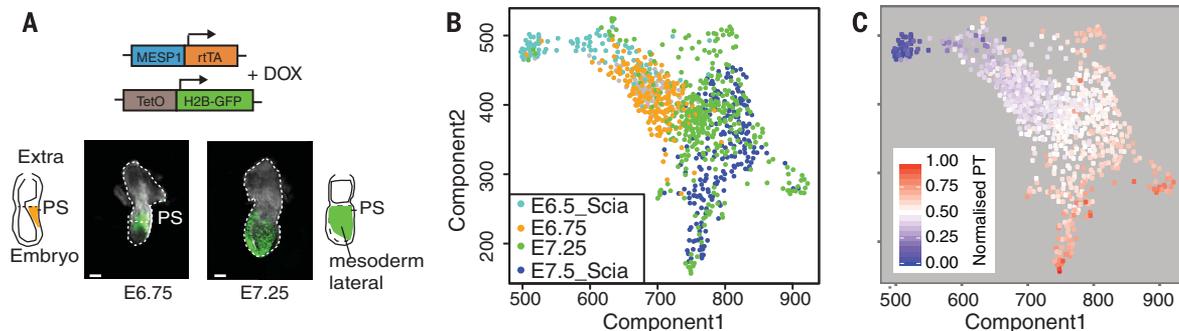
Fig. 1. scRNA-seq of *Mesp1*+ CPs fills the gap between E6.5 epiblast cells and E7.5 mesodermal cells. (**A**) Scheme of the experimental strategy used for isolating *Mesp1*-expressing CPs in vivo. Scale bar, 200 μm. (**B**) SPRING plot of 892 cells showing *Mesp1* CPs at E6.75 and E7.25 and the published epiblast cells (E6.5_Scia) and E7.5 Flk1+ progenitors (E7.5_Scia) with read count of Mesp1 > 0. (**C**) SPRING plot colored by the inferred pseudotime time for all 892 cells.

Mesp1-induced gain of function in embryonic stem cells (ESCs) in vitro (9), many of which are direct Mesp1 target genes (Fig. 2C, fig. S5, and table S1). Several well-known regulators of pluripotency—including *Nanog*, *Eras*, *Pou5f1/Oct4* (10), and markers of the epiblast, including *E-Cadherin/Cdh1*, *Epcam*, *Cldn6*, and *Cldn7*—were up-regulated in single *Mesp1* KO cells (Fig. 2, D to F, and table S1), which is consistent with the defect of exiting the pluripotent epiblast stage. By contrast, the genes down-regulated in *Mesp1* KO cells were greatly enriched for Mesp1 target genes controlling epithelial-mesenchymal transition (EMT) (*Snai1* and *Zeb2*), migration (*Rasgrp3*), and cardiovascular commitment (*Etv2*, *Hand1*, *Myl7*, *Gata4*, *Flk1*, and *Pdgfra*) (Fig. 2, F and G, and fig. S5) (9, 11). Pdgfra/Flk1–expressing cells that

mark *Mesp1* CPs in human and mouse ESC differentiation in vitro and during mouse gastrulation in vivo (5, 12, 13) were much reduced in *Mesp1* KO cells, supporting the absence of CP specification (fig. S6).

SPRING analysis, which allows visualizing high-dimensional single-cell expression data (14), of WT *Mesp1*-expressing cells at E6.75 and E7.25 identified five distinct destination cell types (DCTs) protruding from a core of intermingled cells (Fig. 3, A and B). All cells present within the DCTs came from E7.25 embryos, which is consistent with cell fate diversification of *Mesp1*-derived lineages during the late stages of gastrulation. To further define the nature of the five DCTs, we identified genes with specifically elevated expression in each of the five groups (Fig. 3,

B and C; fig. S7; and table S2). This analysis identified both known and previously unrecognized genes associated with cardiovascular development. DCT1 was enriched among others in *Sox7*, *Etv2*, and *Tal1* transcripts, which are markers of the endothelial or endocardial lineage (15, 16). DCT2 was marked by the expression of *Hand1*, *Bmp4*, *Tnnc1*, *Tbx3*, *Hand2*, *Tbx20*, *Gata4*, *Myl4*, and *Mef2c*, which are well-known CM markers (Fig. 3C, fig. S7, and table S2) (7, 17). Bmp4 promotes CM differentiation (18). Moreover, Hand1 lineage tracing showed that *Hand1*-expressing cells contribute to the LV and to the myocardial and epicardial lineages, with no contribution to the endocardium (17). These data suggest that DCT2 corresponds to CPs committed to the CM lineage. In situ hybridization of *Sox7* (EC
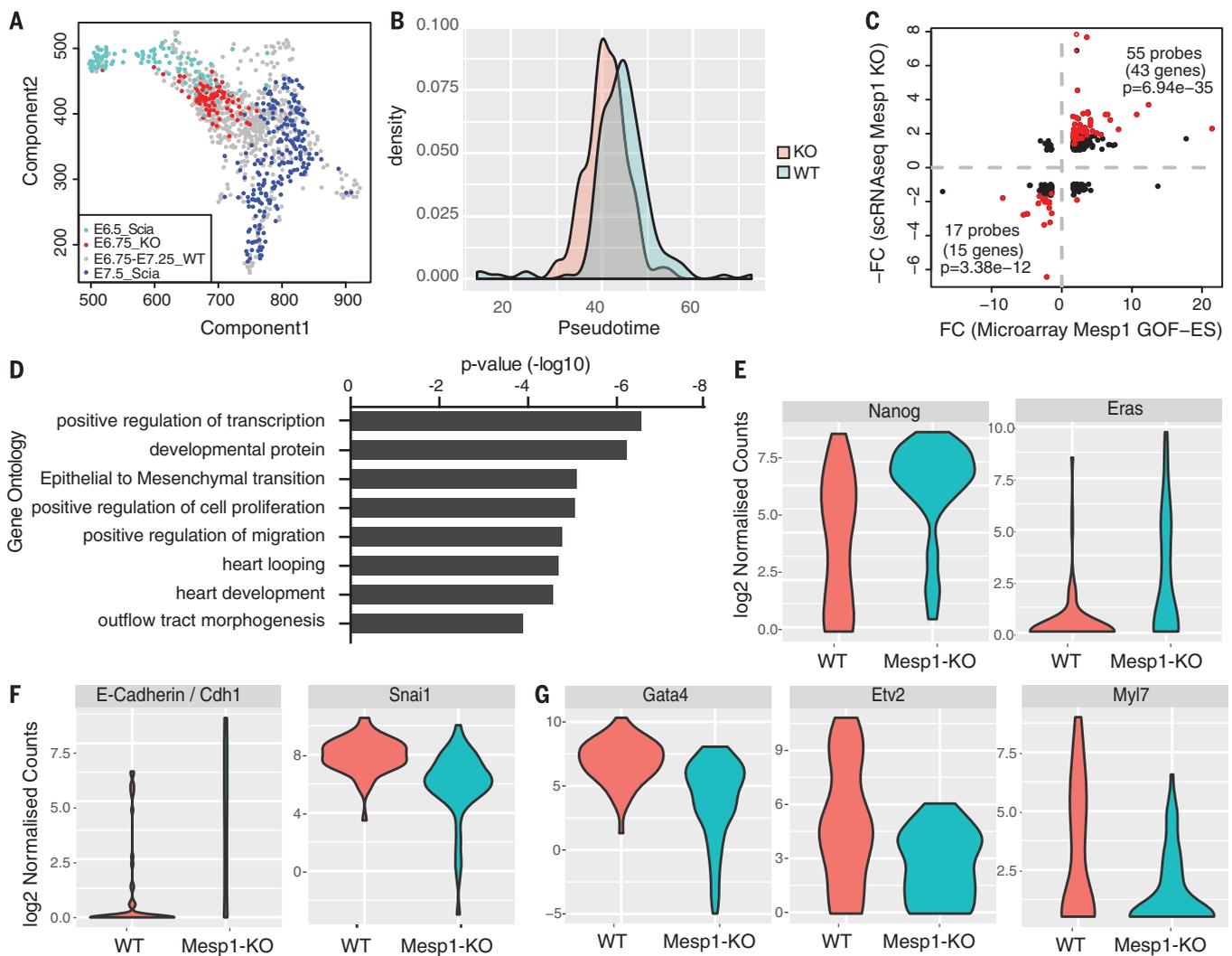
Fig. 2. Mesp1 controls the exit from pluripotency, EMT, and cardiovascular specification. (**A**) SPRING plot of all 892 cells, including *Mesp1* KO cells colored by cell types. (**B**) Pseudotime time distribution for WT and *Mesp1* KO cells at E6.75. (**C**) Comparison of the genes differentially expressed in scRNA-seq experiments between control and *Mesp1* KO cells and the genes regulated by Mesp1 gain of function (GOF) in ESCs. The 58 genes in agreement with the scRNA-seq experiment with false

discovery rates of <0.1 were highlighted in red. The significance of the overlap was calculated by means of hypergeometric test using the phyper function in R. (**D**) Gene ontology enrichment for genes down-regulated in *Mesp1* KO cells. (**E** to **G**) Violin plots showing the mean and variance difference between WT and *Mesp1* KO cells of (E) genes regulating pluripotency (*Nanog* and *Eras*), (F) EMT (*Cdh1* and *Snai1*), and (G) cardiovascular fate (*Gata4*, *Etv2*, and *Myl7*).
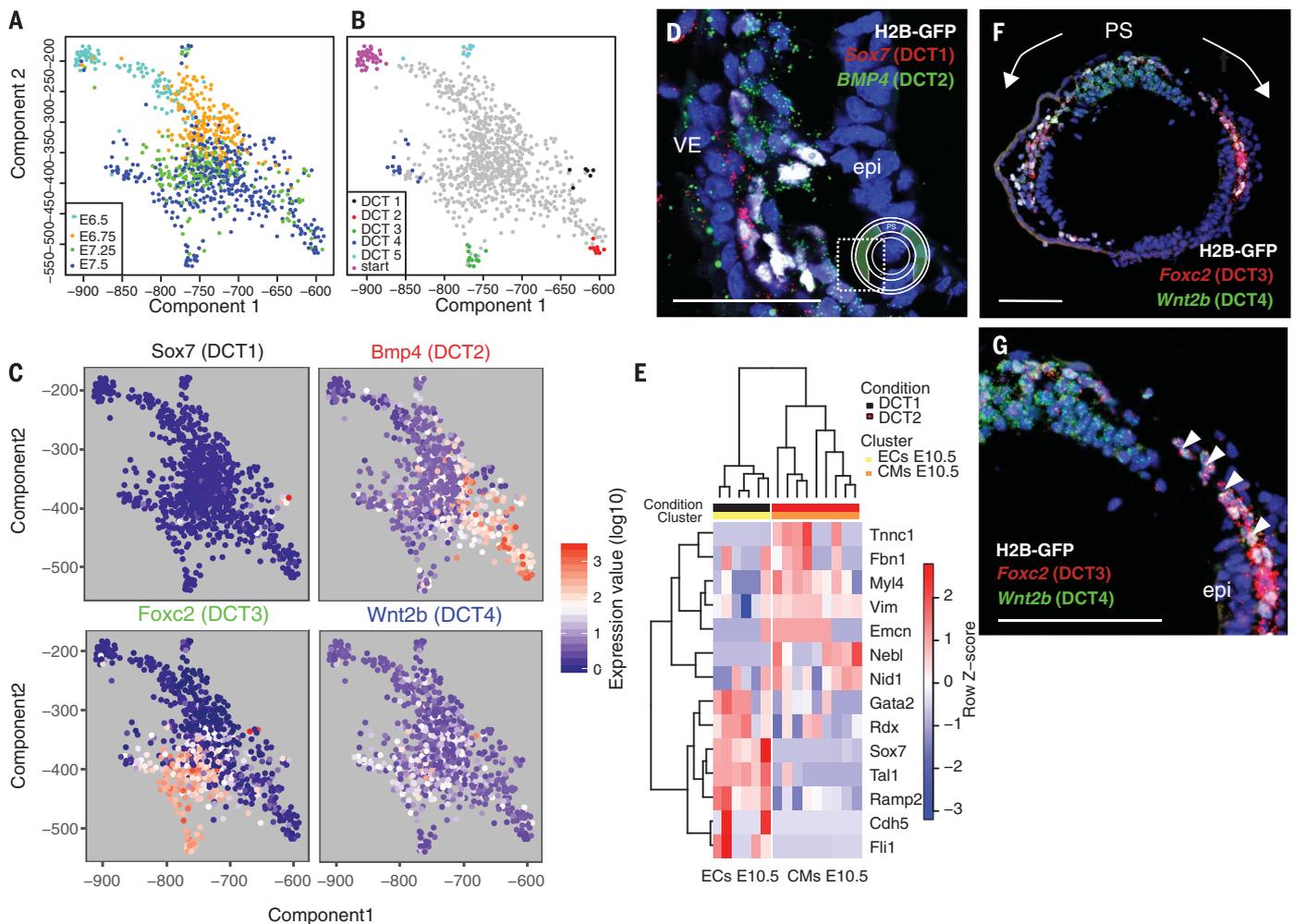
Fig. 3. *Mesp1* single-cell analysis identifies different progenitors committed to different fates and heart regions. (**A**) SPRING analysis of the 807 WT Mesp1-H2B-GFP+ cells at E6.75 and E7.25 and *Mesp1*+ Scia cells. (**B**) The five end points revealed by means of SPRING analysis were considered as five distinct cell types (DCT1-5). (**C**) Expression of key genes specific for DCT1-4. (**D**) RNA-FISH of *Sox7* and *Bmp4* on sections of *Mesp1-rtTA/tetO-H2B-GFP* embryos at E7.25. (**E**) Heatmap of DCT1 and DCT2 end point cells based on unsupervised clustering of the expression of CM and EC marker genes identified at E10.5 (*24*) combined with newly identified genes enriched in DCT1 or DCT2. (**F** and **G**). RNA-FISH of *Foxc2* and *Wnt2b* on sections of *Mesp1-rtTA/tetO-H2B-GFP* embryos at E7.25. Higher magnification is found in (G). Scale bar, 50 μm.

marker) and *Bmp4* (CM marker) on *Mesp1-rtTA/tetO-H2B-GFP* embryos showed that these two markers did not colocalize in Mesp1-H2B-GFP–expressing cells at E7.25, which is consistent with the notion that DCT1 and DCT2 mark two distinct *Mesp1* populations committed to EC and CM differentiation (Fig. 3D). Unsupervised hierarchical clustering of DCT1 and DCT2 markers with CM and EC markers identified by means of scRNA-seq of more mature mouse hearts at E10.5 (*19*) further showed that DCT1 and DCT2 cells clustered respectively with EC and CM lineage (Fig. 3E).

DCT3 and DCT4 are enriched in genes expressed and regulating SHF development (*Tbx1, Foxc2, Hoxb1*, and *Hoxa1*) (Fig. 3C, fig. S7, and table S2) (*20–22*). Lineage tracing experiments have previously shown that *Tbx1* (DCT3) marks anterior SHF progenitors, whereas *Hoxb1* and *Hoxa1* (DCT4) mark posterior SHF (*20*, *22*), suggesting that DCT3 and DCT4 correspond to *Mesp1*

CPs committed to the anterior and posterior SHFs. RNA–fluorescence in situ hybridization (FISH) experiments further showed that *Foxc2* (DCT3) and *Wnt2b* (DCT4) were largely nonoverlapping, with *Wnt2b* localized closer to the PS, whereas *Foxc2*-expressing cells were found more anterolaterally at E7.25 (Fig. 3, F and G). A subset of SHF progenitors, called cranio-pharyngeal progenitors, express *Tcf21* and contribute to the formation of some head muscles (*23*). RNA-FISH showed that *Tcf21* is preferentially expressed in a subset of DCT3 *Mesp1* H2B-GFP expressing *Foxc2*, whereas little overlap was observed between DCT4 progenitors expressing *Wnt2b* and *Tcf21* (fig. S8). Consistent with the notion that DCT3 marks cranio-pharyngeal progenitors, the "pharyngeal" cluster found in our previous Flk1+ scRNA-seq (*8*) are closest to DCT3 (fig. S7). DCT5 expressed endoderm markers such as *Sox17* and *Foxa2* and may have no relation with cardiac development (fig. S7 and table S2).

On the basis of the cumulative evidence that suggested our scRNA-seq captures the developmental progression from epiblast to early cardiovascular lineage segregation, we next investigated the expression of genes with known biological function, which revealed specific expression of cardiovascular, mesodermal, and other genes regulating signaling pathways across the various DCT populations (fig. S9). Analysis of genes dynamically expressed during the trajectory toward DCT1 and DCT2 revealed the existence of gene clusters peaking in expression at sequential points along the pseudotemporal ordering. These gene clusters showed enrichment for gene ontology categories associated with developmental progression in which the predicted trajectory provided consistency between the pseudotime and the real developmental time (fig. S10 and table S3). Moreover, cells coexpressing genes enriched in both DCT3 and DCT4 presented early pseudotime values than either DCT3 or DCT4 cells,
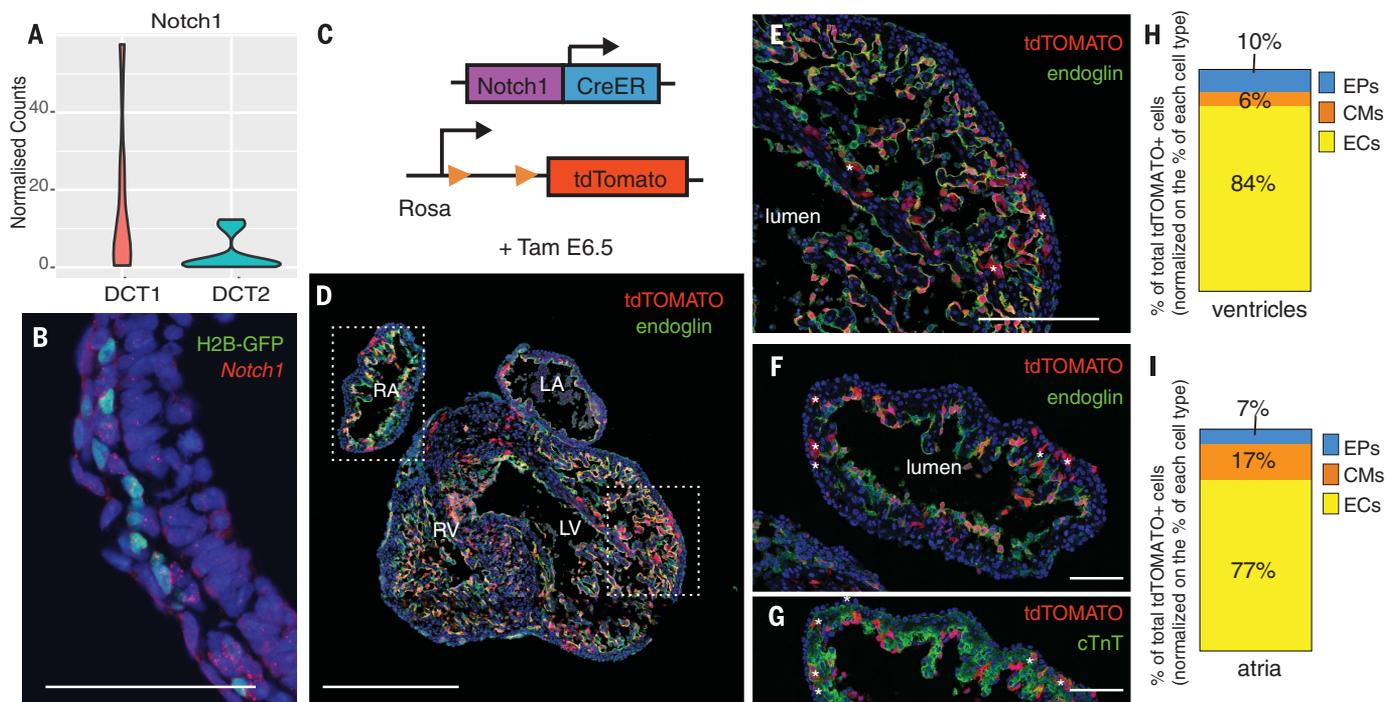
Fig. 4. Notch1 marks *Mesp1* progenitors committed to the endocardial fate. (**A**) Violin plot of *Notch1* expression in DCT1 and DCT2 cells. (**B**) *Notch1* RNA-FISH on a section of an E7.25 *Mesp1-rtTA/tetO-H2B-GFP* embryo. Scale bar, 50 μm. (**C**) Experimental strategy used for tracing *Notch1*-expressing cells at E6.5. (**D** to **G**) Confocal analysis of immunostaining for [(D) to (F)] endoglin (EC marker) and (G) cTNT (CM marker) of *Notch1-CreERT2/Rosa-tdTomato* heart sections at E12.5. (D) Lower magnification. Scale bar, 500 μm. Higher magnifications of the (E) LV and (G) RA showed that most tdTomato⁺ cells are ECs, although rare CMs are also marked (asterisks). Scale bars, 200 μm. (**H** and **I**) Percentage of tdTomato⁺ cells in ECs, CMs, and EPs in (H) the ventricles (7804 cells counted; *n* = 6 embryos from three different litters) and (I) the atria (4819 cells counted; *n* = 7 embryos from three different litters).

which is consistent with the presence of immature progenitors that undergo multilineage priming before making cell fate decision into either DCT3 or DCT4 lineages (fig. S11).

To more clearly differentiate between the putative differentiation paths to the DCT1/EC and DCT2/CM progenitors, we determined the genes involved in embryonic development that are specifically up- or down-regulated in a given DCT (fig. S12). Of particular interest, DCT2 cells showed reduced *Notch1* expression (Fig. 4A). Different studies have shown the importance of Notch1 in the latter stages of cardiovascular development, in regulating endocardium, valve formation, trabeculation, and myocardium compaction (*24*). However, a role for Notch in the early steps of cardiovascular lineage commitment during gastrulation has not been previously described. RNA-FISH and immunostaining showed that *Notch1* was indeed expressed and active in a subset of Mesp1-H2B-GFP⁺ cells (Fig. 4B and fig. S13). To determine whether *Notch1* expression in DCT1 cluster marks *Mesp1*⁺ CPs committed to the EC fate, we induced lineage tracing by administrating tamoxifen to *Notch1-CreERT2/Rosa-tdTomato* mice at E6.5 (*25*) and assessed the fate of marked cells at E12.5 (Fig. 4, C to G). tdTomato⁺ cells in the ventricles were almost exclusively found in the EC (83.9 ± 3.0%), with minor contribution to the CM (6.0 ± 1.4%) and to the EPs (10.2 ± 2.1%) (Fig. 4H), which is consistent with the notion that DCT1 marks early *Mesp1/Notch1*

double-positive CPs committed to the EC fate. In the atria, *Notch1-CreERT2* marked preferentially the ECs (76.9% ± 2.9), although a smaller contribution to the CM lineage was also observed (16.6% ± 2.7) (Fig. 4I), which is consistent with a lower expression of *Notch1* in DCT2 cluster. In addition to marking ECs of the heart, *Notch1-CreERT2*, similar to *Mesp1-Cre*, also marked, at the early stage of gastrulation, ECs of the aorta, intersomitic, and brain vessels (fig. S14).

Altogether, our single-cell profiling of early CPs shows that *Mesp1* CPs segregate rapidly from the epiblast into distinct cardiovascular lineages. The analysis of *Mesp1* KO cells showed that Mesp1 is required to exit the pluripotent state of the epiblast and promotes EMT, migration, and cardiovascular specification in vivo. Consistent with the early regional and lineage segregation found with clonal analysis (*4, 5*), our scRNA-seq demonstrates that *Mesp1* CPs are also molecularly heterogeneous, as previously suggested by scRNA-seq during in vitro ESC differentiation (*26*), and identifies temporally and spatially distinct *Mesp1* subpopulations that likely correspond to CPs committed to the different cardiovascular lineages and regions of the heart at the early stages of gastrulation (fig. S15). Future studies will be required to determine whether the early cardiovascular lineage segregation uncovered here is also occurring for the other mesodermal and endodermal cells and whether defects in the early commitment steps are associated with congen-

ital diseases and heart malformations. Last, our results will be important to design new strategies to direct the differentiation of ESC into a specific cardiovascular lineage.

REFERENCES AND NOTES

1. S. Martin-Puig, Z. Wang, K. R. Chien, *Cell Stem Cell* **2**, 320–331 (2008).
2. Y. Saga *et al.*, *Development* **126**, 3437–3447 (1999).
3. A. Bondue, C. Blanpain, *Circ. Res.* **107**, 1414–1427 (2010).
4. W. P. Devine, J. D. Wythe, M. George, K. Koshiba-Takeuchi, B. G. Bruneau, *eLife* **3**, e03848 (2014).
5. F. Lescroart *et al.*, *Nat. Cell Biol.* **16**, 829–840 (2014).
6. I. Harel *et al.*, *Dev. Cell* **16**, 822–832 (2009).
7. M. Buckingham, S. Meilhac, S. Zaffran, *Nat. Rev. Genet.* **6**, 826–835 (2005).
8. A. Scialdone *et al.*, *Nature* **535**, 289–293 (2016).
9. A. Bondue *et al.*, *Cell Stem Cell* **3**, 69–84 (2008).
10. G. Martello, A. Smith, *Annu. Rev. Cell Dev. Biol.* **30**, 647–675 (2014).
11. R. C. Lindsley *et al.*, *Cell Stem Cell* **3**, 55–68 (2008).
12. A. Bondue *et al.*, *J. Cell Biol.* **192**, 751–765 (2011).
13. S. J. Kattman *et al.*, *Cell Stem Cell* **8**, 228–240 (2011).
14. C. Weinreb, S. Wolock, A. Klein, *Bioinformatics* 10.1093/bioinformatics/btx792 (2017).
15. T. Behrens *et al.*, *BMC Cancer* **16**, 395 (2016).
16. S. Palencia-Desai *et al.*, *Development* **138**, 4721–4732 (2011).
17. R. M. Barnes, B. A. Firulli, S. J. Conway, J. W. Vincentz, A. B. Firulli, *Dev. Dyn.* **239**, 3086–3097 (2010).
18. T. M. Schultheiss, J. B. Burch, A. B. Lassar, *Genes Dev.* **11**, 451–462 (1997).
19. D. M. DeLaughter *et al.*, *Dev. Cell* **39**, 480–490 (2016).
20. T. Huynh, L. Chen, P. Terrell, A. Baldini, *Genesis* **45**, 470–475 (2007).
21. T. Kume, H. Jiang, J. M. Topczewska, B. L. Hogan, *Genes Dev.* **15**, 2470–2482 (2001).
22. N. Bertrand *et al.*, *Dev. Biol.* **353**, 266–274 (2011).
23. I. Harel *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **109**, 18839–18844 (2012).

24. D. MacGrogan, M. Nus, J. L. de la Pompa, *Curr. Top. Dev. Biol.* **92**, 333–365 (2010).
25. S. Fre *et al.*, *PLOS ONE* **6**, e25785 (2011).
26. S. S. Chan, H. H. W. Chan, M. Kyba, *Biochem. Biophys. Res. Commun.* **474**, 469–475 (2016).

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/359/6380/1177/suppl/DC1
Materials and Methods
Figs. S1 to S15
Tables S1 to S4
References (27–50)

Science

Defining the earliest step of cardiovascular lineage segregation by single-cell RNA-seq

Fabienne Lescroart, Xiaonan Wang, Xionghui Lin, Benjamin Swedlund, Souhir Gargouri, Adriana Sànchez-Dànes, Victoria Moignard, Christine Dubois, Catherine Paulissen, Sarah Kinston, Berthold Göttgens and Cédric Blanpain

Committing the heart
The heart is a complex organ composed of multiple cell types such as cardiomyocytes and endothelial cells. Cardiovascular cells arise from *Mesp1*-expressing progenitor cells. Lescroart *et al.* performed single-cell RNA-sequencing analysis of mouse wild-type and *Mesp1*-deficient cardiovascular progenitor cells at early gastrulation (see the Perspective by Kelly and Sperling). When *Mesp1* was eliminated, embryonic cells remained pluripotent and could not differentiate into cardiovascular progenitors. During gastrulation, the different *Mesp1* progenitors rapidly became committed to a particular cell fate and heart region. *Notch1* expression marked the earliest step of cardiovascular lineage segregation.
Science, this issue p. 1177; see also p. 1098

ARTICLE TOOLS	http://science.sciencemag.org/content/359/6380/1177
SUPPLEMENTARY MATERIALS	http://science.sciencemag.org/content/suppl/2018/01/24/science.aao4174.DC1
RELATED CONTENT	http://science.sciencemag.org/content/sci/359/6380/1098.full
REFERENCES	This article cites 46 articles, 12 of which you can access for free http://science.sciencemag.org/content/359/6380/1177#BIBL
PERMISSIONS	http://www.sciencemag.org/help/reprints-and-permissions