Single-cell analysis of embryoids reveals lineage diversification roadmaps of early human development

Yi Zheng[1 *], Robin Zhexuan Yan[1], Shiyu Sun[1], Mutsumi Kobayashi[2], Lifeng Xiang[3], Ran Yang[4], Alexander Goedel[4], Yu Kang[6], Xuefeng Xue[1], Sajedeh Nasr Esfahani[1], Yue Liu[1], Agnes M. Resto Irizarry[1], Weisheng Wu[5], Yunxiu Li[3], Weizhi Ji[6], Yuyu Niu[6], Kenneth R. Chien[4], Tianqing Li[6], Toshihiro Shioda[2,7], and Jianping Fu[1,8,9 *]

[1]Department of Mechanical Engineering, University of Michigan, Ann Arbor, Michigan 48109, USA; [2]Massachusetts General Hospital Center for Cancer Research, Charlestown, MA 02129, USA; [3]Department of Reproductive Medicine, the First People's Hospital of Yunnan Province, Kunming, China; [4]Department of Cell and Molecular Biology, Karolinska Institutet, Stockholm 171 77, Sweden; [5]BRCF Bioinformatics Core, University of Michigan, Ann Arbor, Michigan 48109, USA; [6]State Key Laboratory of Primate Biomedical Research, Institute of Primate Translational Medicine, Kunming University of Science and Technology, Kunming, China; [7]Department of Medicine, Harvard Medical School, MA 02115, USA; [8]Department of Cell and Developmental Biology, University of Michigan Medical School, Ann Arbor, MI 48109, USA; [9]Department of Biomedical Engineering, University of Michigan, Ann Arbor, Michigan 48109, USA.

*Correspondence and requests for materials should be addressed to Y.Z. and J.F. (email: zhengum@umich.edu, jpfu@umich.edu).

SUMMARY

Despite its clinical and fundamental importance, our understanding of early human development remains limited. Stem cell-derived, embryo-like structures (or embryoids) allowing studies of early development without using natural embryos can potentially help fill the knowledge gap of human development. Herein, transcriptome at the single-cell level of a human embryoid model, which recapitulates aspects of lineage diversification and three-dimensional tissue architecture of the human embryo from the implantation to the onset of gastrulation, was profiled at different time points. Molecular maps of lineage diversifications from the pluripotent human epiblast towards the amniotic ectoderm, primitive streak / mesoderm, and primordial germ cells were constructed and compared with *in vivo* primate data. Similarly, chimpanzee embryoids were generated and profiled to reveal transcriptome dynamics during the early post-implantation chimpanzee development. Our comparative transcriptome analyses reveal a critical role of NODAL signaling in human mesoderm and primordial germ cell specification, which is further functionally validated. Through comparative transcriptome analyses and validations with human blastocysts and *in vitro* cultured *cynomolgus* embryos, we further proposed stringent criteria for distinguishing between human blastocyst trophectoderm and early amniotic ectoderm cells. Altogether, this study provides new knowledge of the lineage diversification roadmap of early human development and will serve as a valuable resource for studying human development.

KEYWORDS

Human embryoid, single-cell transcriptome, NODAL signaling, primate development, amnion, primitive streak, primordial germ cell, mesoderm, trophoblast, microfluidics

INTRODUCTION

Development of multicellular organisms is one of nature's greatest triumphs. Development is a tightly orchestrated process, following stereotypic lineage diversifications and morphogenetic tissue patterning events in a precise spatiotemporal order. Scientists commonly use animal models to study the key transcriptional and signaling activities that underlie pattern formation, morphogenesis, cell differentiation, and tissue growth (Gilbert, 2000; Schoenwolf, 2020; Solnica-Krezel, 2020). However, cross-species genetic and morphological divergence is evident between humans and commonly used animal models (Rossant, 2015; Rossant and Tam, 2017). To address this issue, there is a significant current interest in improving *in vitro* culture protocols of human and non-human primate (NHP) monkey embryos for experimental observations and mechanistic studies (Deglincerti et al., 2016; Ma et al., 2019; Niu et al., 2019; Shahbazi et al., 2016; Xiang et al., 2020; Yang et al., 2021). However, experimentations on human and NHP monkey embryos remain challenging due to limited access to and bioethical constraints on these natural specimens (Clark et al., 2021; Hyun et al., 2016; Lovell-Badge et al., 2021). As such, knowledge of human development remains limited; this is particularly true for early post-implantation human development, when the basic human body plan is laid down and when the human embryo *in vivo* is at its most inaccessible phase for experimentation.

Recently, stem cell-derived embryo-like structures (or embryoids) that could recapitulate certain aspects of mammalian early embryogenesis are emerging as tractable experimental tools for studying human development (Beccari et al., 2018; Haremaki et al., 2019; Harrison et al., 2017; Liu et al., 2021; Moris et al., 2020; Rivron et al., 2018; Shao et al., 2017a; Shao et al., 2017b; Simunovic et al., 2019; Warmflash et al., 2014; Xue et al., 2018; Yanagida et al., 2021; Yu et al., 2021; Zheng et al., 2019a; Zheng et al., 2019b). Particularly, we have recently developed a human pluripotent stem cell (hPSC)-based embryoid, termed post-implantation amniotic sac embryoid (PASE), that appears to recapitulate different developmental events of the early post-implantation human embryo in a three-dimensional (3D), human-relevant tissue architecture (Shao et al., 2017a; Shao et al., 2017b). These developmental events include landmarks of the development of the epiblast (Epi) and amniotic ectoderm (AM) parts of the human embryo, including lumenogenesis of the Epi and the resultant pro-amniotic cavity, symmetry breaking of the Epi sac to form the bipolar amniotic sac, and specification of primordial germ cells (PGCs) and primitive streak cells (PSs). By using controllable microfluidic

85 tools, we have further successfully developed a microfluidic PASE (μPASE) system (Zheng et
86 al., 2019b), allowing the development of PASEs in a highly controllable, reproducible and
87 scalable fashion.

88 To fill the critical knowledge gap of early post-implantation human development, herein
89 we studied transcriptome dynamics during the progressive development of μPASE at the single-
90 cell resolution using single-cell RNA sequencing (scRNA-seq) and provided detailed analyses of
91 cell lineage diversifications, developmental trajectories, regulatory networks and signaling
92 pathways involved in this previously unexplored yet critical stage of human development. We
93 also developed chimpanzee μPASEs using chimpanzee PSCs and profiled their transcriptome
94 using scRNA-seq. The scRNA-seq data from human and chimpanzee μPASEs were compared
95 with recently published data from a Carnegie Stage 7 (CS7) human gastrula (Tyser et al., 2021)
96 and those from *in vivo* and *in vitro* cultured human and NHP monkey embryos (Ma et al., 2019;
97 Nakamura et al., 2016; Sasaki et al., 2016; Yang et al., 2021). Our comparative transcriptome
98 analyses reveal the developmental coordination among different primate species and further
99 highlight a critical role of NODAL signaling in mesoderm (Meso) and PGC specification, which
100 was further functionally validated. To address current confusions in distinguishing between
101 human early AM cells *vs.* blastocyst trophectoderm cells, we established stringent criteria for
102 identifying these two lineages by profiling and comparing related transcriptomes from *in vivo*
103 samples (Blakeley et al., 2015; Petropoulos et al., 2016; Tyser et al., 2021), and further evaluated
104 the authenticity of related cells included in different human embryoids and *in vitro*
105 differentiation protocols. Finally, we proposed a few cell fate markers that could be utilized to
106 reliably distinguish human early AM and blastocyst trophectoderm cells and further validated
107 expression of these markers using *in vitro* fertilization (IVF) human blastocysts and *in vitro*
108 cultured NHP monkey embryos. Altogether, our scRNA-seq datasets and related analyses as well
109 as experimental results from both *in vitro* and *in vivo* models provide new insights into the
110 lineage diversification roadmap of early human development and will serve as a valuable
111 resource for studying human development.
112

113 **RESULTS**

114 **Single-cell transcriptomic profiling of μPASE development**

115 The microfluidic device for generating μPASEs consists of three parallel channels, partitioned by
116 evenly spaced supporting posts (Zheng et al., 2021; Zheng et al., 2019b). Preloaded Geltrex in
117 the central gel channel forms concave gel pockets between adjacent supporting posts during
118 gelation. Singly dissociated hPSCs are loaded into the cell loading channel and are guided to
119 form individual cell clusters in each concave gel pocket. After initial clustering of hPSCs
120 (designated as $t = 0$ h), culture medium in the device is switched to a basal medium comprised of
121 Essential 6 and FGF2, with BMP4 further supplemented only into the cell loading channel
122 (**Figure 1A**). Development of μPASE involves successive cell morphogenetic and lineage
123 specification events that recapitulate early post-implantation human development up to the onset
124 of gastrulation (**Figure 1A**) (Zheng et al., 2021; Zheng et al., 2019b). Specifically, owing to their
125 intrinsic lumenogenic property, each hPSC cluster undergoes lumenogenesis and epithelization
126 to establish apical-basal polarity and form a single central apical lumen by $t = 12$ h (**Figures 1B**
127 **and S1A**). hPSCs exposed directly to exogeneous BMP4 stimulation in each cluster initiate
128 amniogenesis, evidenced by continuous flattening of cell morphology, resolving into a thin layer
129 of squamous amniotic cells (**Figure 1B**). Inductive effects of AMLCs in the μPASEs lead to
130 hPSCs at the opposite pole to undergo epithelial-mesenchymal transition (EMT) and
131 gastrulation-like events (Zheng et al., 2021; Zheng et al., 2019b), with gastrulating cells
132 disseminating away from μPASEs from $t = 36$ h onwards (**Figure 1B**), leading to disintegration
133 of the μPASE structure. By $t = 48$ h, the μPASE contains only AMLCs, MeLCs and PGCLCs,
134 without the presence of EpiLCs (**Figures 1B and 1C**).
135 To investigate dynamics of μPASE development at the transcriptome level, single-cell
136 suspensions of μPASEs at $t = 24$ h, 36 h, and 48 h were prepared before single-cell RNA-
137 sequencing (scRNA-seq) using 10× Genomics. We performed UMAP (uniform manifold
138 approximation and projection) dimension reduction using the Seurat R package (Butler et al.,
139 2018; Satija et al., 2015) for scRNA-seq datasets at each time point (**Figure 1C**), as well as for
140 the integrated scRNA-seq dataset from all three time points (**Figure 1D**). These analyses reveal
141 distinct cell clusters in the μPASE based on expression patterns of key lineage markers (**Figures**
142 **1C-1E, S1B and S1C**). Consistent with our previous findings (Zheng et al., 2021; Zheng et al.,
143 2019b), in the μPASE, hPSCs develop progressively from a pluripotent epiblast-like cell
144 (EpiLC) stage to three distinct cell populations by $t = 48$ h: amniotic ectoderm-like cells

145 (AMLCs), mesoderm-like cells (MeLCs), and primordial germ cell-like cells (PGCLCs)
146 (**Figures 1C-1E**, **S1D and S1E**).
147 To understand how hPSCs transit from the EpiLC stage to AMLCs, MeLCs and PGCLCs
148 during μPASE development, RNA velocity analysis (Bergen et al., 2020; La Manno et al., 2018)
149 was conducted for the integrated scRNA-seq dataset, with RNA velocity vectors overlaid on the
150 integrated UMAP plot (**Figure 1D**). This RNA velocity analysis reveals developmental
151 trajectories of the AMLC lineage (EpiLC → nascent AMLC or NasAMLC → AMLC1 →
152 AMLC2) and MeLC lineage (EpiLC → primitive streak-like cell or PSLC → MeLC1 / MeLC2)
153 (**Figure 1D**). However, developmental trajectory of PGCLCs is not as clear from the RNA
154 velocity analysis (**Figure 1D**). To further examine lineage relations between different μPASE
155 cell clusters, partition-based graph abstraction (PAGA) analysis (Wolf et al., 2018) was
156 conducted, revealing that PGCLCs correlate best with the NasAMLC cluster (**Figure 1F**). To
157 reveal gene regulatory network (GRN) underlying each cell cluster, we performed GRN analysis
158 using SCENIC (Aibar et al., 2017) (single-cell regulatory network inference and clustering;
159 **Figures 1G and 1H**). Notably, AM markers *ISL1* and *GATA3* (Yang et al., 2021), PS / Meso
160 markers *EOMES*, *MIXL1*, *TBX6* and *GATA6*, and PGC markers *SOX17* and *NANOG* are
161 identified by SCENIC as regulons of corresponding cell lineages (**Figures 1G and 1H**).
162
163 **Trajectory inference and gene expression dynamics analysis**
164 To infer developmental trajectories of different μPASE cell lineages, we plotted a three-
165 dimensional (3D) diffusion map based on PCA embeddings of the integrated scRNA-seq dataset
166 (**Figure 2A and Video S1**) (Angerer et al., 2016). The AMLC, MeLC and PGCLC lineages
167 display distinct and well separated trajectories in the 3D diffusion map (**Figure 2A**). To analyze
168 transcriptome dynamics during AMLC lineage development, EpiLC, NasAMLC and AMLC1 / 2
169 clusters were isolated from the integrated scRNA-seq dataset and re-plotted using two-
170 dimensional (2D) diffusion maps (**Figure 2B**). Similarly, to analyze MeLC lineage development,
171 EpiLC, PSLC and MeLC1 / 2 clusters were isolated from the integrated scRNA-seq dataset
172 before re-plotting using 2D diffusion maps (**Figure 2C**). Expression dynamics of selected genes
173 related to AM and PS or Meso was plotted against diffusion pseudotime (dpt) (**Figures 2D, 2E,**
174 **S2A and S2B**). Notably, in the AMLC lineage, expression of *TFAP2A*, *MSX2* and *ID2*, which
175 are commonly used AM markers (Ma et al., 2019; Sasaki et al., 2016; Yang et al., 2021), is

176 quickly upregulated following exogeneous BMP4 stimulation (**Figures 2D and S2A**). Similarly,
177 *GATA3* and *ISL1* become upregulated relatively early during AMLC lineage development
178 (**Figures 2D and S2A**). In contrast, expression of *GABRP*, *IGFBP3* and *WNT6* shows delayed
179 upregulation, whereas *TBXT* is only transiently expressed during early AMLC lineage
180 development (**Figures 2D and S2A**).

181 On the 2D diffusion map, the MeLC lineage branches into two separate paths
182 corresponding to MeLC1 and MeLC2 clusters (**Figure 2C**). Both MeLC1 and MeLC2 lineages
183 show upregulated expression of common Meso markers, yet with some key genes exhibiting
184 distinct expression levels (**Figures 2E and S2B**). For example, compared with MeLC1, MeLC2
185 expresses relatively higher levels of *MIXL1*, *EOMES* and *GATA6* (**Figures 2E and S2B**),
186 presumably corresponding to a lateral plate / intermediate Meso fate (Daoud et al., 2014;
187 Prummel et al., 2019). MeLC1 lineage, in contrast, expresses higher levels of *CDX2* and *HOXB6*
188 (**Figures 2E and S2B**), presumably corresponding to a paraxial Meso fate (Casaca et al., 2016;
189 Chawengsaksophak et al., 2004). We further conducted immunostaining for selected AM and
190 Meso markers in μPASEs at t = 24 h, 36 h, and 48 h to validate lineage fate specification
191 (**Figures 2F and S2C-S2E**). GATA6^{high}TBXT^{high}MIXL1^{high}CDX2^{low} MeLC2 appears as leading
192 cells in the migratory gastrulating cell population, whereas
193 GATA6^{low}TBXT^{low}MIXL1^{low}CDX2^{high} MeLC1 appears as trailing cells in this population
194 (**Figure S2F**).

195 Patterning of AMLCs and PSLCs was evident in the μPASE by t = 24 h, with positive
196 immunostaining for ISL1, GATA3 and TFAP2A in incipient AMLCs and for TBXT and MIXL1
197 in incipient MeLCs (**Figures S2C and S2D, Video S2**). Notably, expression of AM marker
198 GABRP (Yang et al., 2021) is restricted on the apical surface of AMLCs, whereas expression of
199 HEY1, another AM marker (Yang et al., 2021), is evident in both the nucleus and cytoplasm of
200 AMLCs (**Figure 2F**). AMLCs appear to actively proliferate during μPASE development (**Figure**
201 **S2G**). To reveal transcriptome changes during AMLC development, we conducted differentially
202 expressed gene (DEG) and pathway enrichment analysis to compare NasAMLC, AMLC1 and
203 AMLC2 (**Figure S2H and Table S3**).

204 To explore potential mechanisms underlying lineage choices between NasAMLCs *vs.*
205 PSLCs for EpiLCs, we examined DEGs upregulated in NasAMLCs and PSLCs relative to
206 EpiLCs and noticed significant overlaps (144 out of 342 genes for NasAMLCs, 144 out of 184

genes for PSLCs) (**Figure S2I and Table S3**). KEGG pathway analysis suggests that development of both NasAMLCs and PSLCs requires TGF-β signaling (KEGG: 04350; **Figure S2J**), which is not surprising given that exogenous BMP4 is supplemented in the μPASE protocol. However, WNT signaling pathway (KEGG: 04310) is evident only in PSLCs but not in NasAMLCs (**Figure 2G**), suggesting a critical role of WNT signaling in Meso induction but not in amniogenesis. When IWP2, a small molecule inhibitor blocking the transport, secretion or signaling activity of all WNT molecules (Chen et al., 2009; Kadowaki et al., 1996; Zhai et al., 2004), was supplemented into the microfluidic device, development of PSLCs and MeLCs in the μPASE was completely inhibited, as evidenced by negative immunostaining for TBXT or MIXL1 (**Figure 2H**). In contrast, AMLCs still emerge under IWP2 treatment, as evidenced by positive immunostaining for ISL1, TFAP2A and GATA3 in flattened, presumptive AMLCs directly exposed to exogeneous BMP4 stimulation (**Figure 2H**). Consistently, the scRNA-seq data obtained from IWP2-treated μPASEs at t = 48 h show absence of PSLCs or MeLCs; instead, the majority of cells remain pluripotent, retaining the EpiLC identity (**Figures 2I and 2J**), further supporting the requirement of WNT signaling in PSLC/MeLC development. It should be noted that under IWP2 treatment, AMLCs show lower expression of several AM marker genes, such as *ISL1, GABRP* and *GATA3*, and there are much less PGCLCs in IWP2-treated μPASEs (**Figures S2K and 2I, and Table S3**).

PGCLC specification

During μPASE development, incipient PGCLCs emerge together with NasAMLCs and PSLCs by t = 24 h (**Figure 1C**). To explore the origin and specification of PGCLCs in the μPASE, a 2D diffusion map with EpiLC, PSLC, MeLC1 and PGCLC clusters isolated from the integrated scRNA-seq dataset was plotted, which, however, did not show a continuous developmental trajectory connecting EpiLCs, PSLCs, MeLCs with PGCLCs (**Figure S3A**). In addition, in this 2D diffusion map, K-branch algorithm (Chlis et al., 2017) did not identify a branching point or branches with proper confidence (**Figure S3A**). Thus, PGCLCs in the μPASE are unlikely originated from PSLCs or MeLCs.

In contrast, 2D diffusion map analysis with EpiLC, NasAMLC, AMLC1 / 2 and PGCLC clusters clearly shows lineage progression and bifurcation of NasAMLCs into PGCLCs and AMLCs (**Figure 3A**). RNA velocity vectors were also computed and overlaid onto the diffusion

238 map in **Figure 3A**, revealing that NasAMLCs have the closest lineage relation with PGCLCs,
239 consistent with the PAGA analysis in **Figure 1F**. Our observation here is consistent with recent
240 studies of *cynomolgus* monkey embryos reporting that primate PGCs could emerge in the
241 nascent AM prior to the gastrulation (Sasaki et al., 2016). Expression of selected genes relevant
242 to PGC development, including *SOX17*, *NANOG*, *TFAP2C*, *PRDM1*, *NANOS3*, *TFAP2A*, *TBXT*,
243 *SOX15*, *POU5F1*, *ISL1*, *PDPN*, *IGF1*, *PEG10*, *BAMBI*, *WNT5B* and *WNT2*, was also plotted
244 against diffusion pseudotime, revealing their upregulated expression during PGCLC lineage
245 development (**Figures 3B and S3B**). Immunofluorescence analyses of μPASEs at t = 24 h, 36 h,
246 and 48 h further confirmed spatiotemporal expression of some key PGC markers, including
247 TFAP2C, NANOG, SOX17 and BLIMP1, in incipient PGCLCs (**Figures 3C and S3C**).
248 Correlation coefficient analysis based on PGC ontogenic genes identified from the *cynomolgus*
249 embryo transcriptome data suggests that the transcriptome of PGCLCs in the μPASE is similar
250 to that of Day 2 hPGCLCs derived from conventional protocols (Chen et al., 2019; Sasaki et al.,
251 2015) (**Figure S3D**). It is worth noting that the μPASE essentially is a posteriorized embryonic-
252 like structure, mimicking the posterior portion of the amnion and epiblast compartments. As
253 such, the development of μPASE gives rise to a greater number of hPGCLCs but lack the
254 development of ectoderm.
255 Using K-branch algorithm, we identified developmental branches and the branching point
256 for the 2D diffusion map with EpiLC, NasAMLC, AMLC1/2 and PGCLC clusters (**Figure 3A**).
257 The K-branch analysis further allowed us to separate NasAMLCs into three sub-clusters, with
258 each sub-cluster merged with EpiLCs, AMLCs and PGCLCs, respectively, and annotated as
259 EpiLC-branch NasAMLC, AMLC-branch NasAMLC and PGCLC-branch NasAMLC,
260 respectively (**Figure 3A**). To explore mechanisms underlying lineage choices between AMLCs
261 *vs.* PGCLCs for NasAMLCs, DEGs upregulated in AMLC-branch NasAMLCs and PGCLC-
262 branch NasAMLCs, as compared to EpiLC-branch NasAMLCs, were examined (**Figure 3D;**
263 **Table S4**). KEGG pathway analysis of these DEGs reveals that PGCLC-branch NasAMLCs
264 show upregulated WNT signaling, whereas AMLC-branch NasAMLCs exhibit upregulated
265 Hippo activity (**Figure 3E**). We further performed pharmacological inhibition assays to explore
266 the roles of different signaling pathways during μPASE development. When IWP2 was
267 supplemented into the microfluidic device, development of PGCLCs was almost completely
268 abolished in the μPASE (**Figure 3F**). PGCLCs show upregulated HIF-1 signaling activity

compared to PGCLC-branch NasAMLCs (**Figure S3E**). When HIF-1 signaling was repressed by supplementing LY294002, which functions through binding to phosphoinositide 3-kinases (PI3Ks), in the microfluidic device, the percentage of PGCLCs was significantly reduced (**Figures S3F and S3G**).

Our scRNA-seq data analysis suggests that progenitors of PGCLCs likely pass through a transient transcriptome state similar to that of NasAMLCs before their full commitment to the PGCLC fate. Our data, however, do not exclude the possibility that cells at the junction between the AMLC and PSLC / MeLC compartments, or from the PSLC / MeLC pole of the μPASE, with a transient transcriptome state similar to that of NasAMLCs, can also give rise to PGCLCs. After all, NasAMLCs and PSLCs share a similar transcriptome, and cells in the early post-implantation human embryo might remain plastic (Chen et al., 2019; Irie et al., 2015; Kobayashi et al., 2017; Sasaki et al., 2015).

Transcriptomic comparison between μPASEs and Carnegie Stage 7 human gastrula

The μPASE recapitulates certain aspects of human development from early implantation to the onset of gastrulation. Notably, single-cell transcriptome data of a CS7 human gastrula recently became available (Tyser et al., 2021). Thus, we conducted transcriptomic comparison between μPASEs and the CS7 human gastrula. We first downsampled the μPASE dataset by randomly selecting 100 cells from each cluster and integrated this downsampled dataset with the CS7 human gastrula dataset (**Figure S4**). Based on the transcriptome proximity in the UMAP plot, the μPASE contains cells corresponding to "Epiblast", "Primitive Streak", "Nascent Mesoderm", "Emergent Mesoderm", "Amniotic/embryonic ectoderm", and "PGC" cells in the human gastrula. We next selected only these cells from the CS7 human gastrula dataset to integrate with the whole μPASE dataset. A UMAP plot of the integrated scRNA-seq dataset shows cell clustering similar to that from the μPASE scRNA-seq data alone (**Figure 4A**). Cells from the CS7 human gastrula are evident in all cell clusters of the UMAP plot (**Figure 4A**). Furthermore, for each annotated cell cluster, expression patterns of key lineage markers are consistent between cells from the CS7 human gastrula and μPASEs (**Figure 4B**). Most of the cells from the CS7 human gastrula fall into different cell clusters consistent with their lineage annotations in the original publication (Tyser et al., 2021), except for some cells originally classified as "Epiblast" or "Primitive Streak" that fall into NasAM or AM1 clusters (**Figure 4C**). Notably, PGCs in the

300 CS7 human gastrula can be easily distinguished in the PGC cluster of the integrated dataset in
301 the UMAP plot (**Figure 4A**).
302 In the original publication of the CS7 human gastrula (Tyser et al., 2021), a UMAP plot
303 was generated with all the cells in the human gastrula, including those at relatively late
304 developmental stages, such as "Hemogenic Endothelial Progenitors" and "Erythrocytes", which
305 could negatively affect the resolution of cell clustering analysis, especially for those closely
306 related cell lineages. To address this issue, we re-generated a UMAP plot using only the
307 "Epiblast", "Primitive Streak", "Nascent Mesoderm", "Emergent Mesoderm",
308 "Amniotic/embryonic ectoderm" and "PGC" clusters from the human gastrula dataset (**Figure**
309 **4D**). Interestingly, this UMAP plot reveals distinct cell clusters corresponding to Meso and AM,
310 supported by feature plots showing expression patterns of key Meso and AM markers (**Figures**
311 **4D and 4E**). Notably, human PGCs identified in the original publication is clustered together
312 with AM cells in the UMAP plot (**Figure 4D**), supporting their close lineage relation. We further
313 isolated the Epi, AM, and PGCs from the CS7 human gastrula to generate a diffusion map
314 (**Figure 4F**). Surprisingly, AM and PGC seemingly display lineage bifurcation trajectories from
315 the Epi (**Figure 4F**), similar to μPASEs in **Figure 3A**. In addition, as shown in **Figure 4G**,
316 pseudotime gene expression dynamics of AM cells from the CS7 human gastrula is highly
317 consistent with that of AMLCs in μPASEs in **Figure 2D**. However, due to the low number of
318 PGCs, AM and PGC lineage bifurcation is not as evident as that shown in the μPASE diffusion
319 map (**Figure 3A**); and we could not obtain pseudotime gene expression dynamics for PGCs.
320 To characterize pluripotency state transition in EpiLCs during the progressive
321 development of the μPASE, we further performed scRNA-seq for cultured hPSCs and μPASEs
322 at $t = 0$ h and $t = 12$ h. No distinguishable cell clusters or populations were observed when the
323 scRNA-seq data were analyzed using Seurat R package (**Figures S5A-S5C**). Notably, although
324 μPASEs at $t = 12$ h show upregulated *TFAP2A* expression owing to exogeneous BMP4
325 treatment, transcriptomes of cells in the μPASE have not yet changed sufficiently for the cells to
326 emerge as distinct clusters in the UMAP. We then downsampled these datasets and compared
327 them with transcriptome data of the human morula, pre-implantation epiblast from human
328 blastocysts (Petropoulos et al., 2016), early post-implantation epiblast from days post-
329 fertilization (d.p.f) 9 and d.p.f 11 *in-vitro* cultured human embryos (Molè et al., 2021), and late
330 post-implantation epiblast from the CS7 human gastrula (Tyser et al., 2021). As suggested by the

principal component analysis (PCA) plot (**Figures S5D and S5E**), cultured hPSCs and EpiLCs in the μPASE share comparable transcriptome states and are both at a developmental stage between d.p.f 11 and d.p.f 16 epiblast. Thus, there was no compelling evidence showing heterogeneous pluripotency states or transitions at early stages of μPASE development.

Transcriptomic coordination of early development across different primate species

NHP monkey embryos, including *cynomolgus* (*Macaca fascicularis*) embryos, have been used successfully as an *in vivo* model to study primate development (Ma et al., 2019; Nakamura et al., 2016; Sasaki et al., 2016; Yang et al., 2021). Compared with NHP monkeys, chimpanzees share much more of human DNA (99%), making them our closest living relatives (Gibbs et al., 2007; Mikkelsen et al., 2005). Thus, we sought to generate μPASEs from chimpanzee induced pluripotent stem cells (iPSCs) to examine whether chimpanzee μPASEs (or C_μPASEs) would develop in a fashion similar as μPASEs made from hPSCs (or H_μPASEs). To this end, the same microfluidic protocol for H_μPASE development was used for C_μPASE formation. Under exogeneous human BMP4 stimulation, C_μPASEs also display prominent molecular and morphogenetic asymmetry, with the pole exposed to BMP4 differentiating progressively into squamous, flattened AMLCs positive for AM markers ISL1, GATA3 and TFAP2A, and the opposite pole developing into germ layer lineages positive for TBXT and MIXL1 (**Figure 5A**). Similar to H_μPASEs, PGCLCs also emerge in C_μPASEs, as evidenced by positive immunostaining for SOX17, TFAP2C, NANOG and BLIMP1 (**Figure 5A**). However, distinct from H_μPASEs, FOXA2+BLIMP1+ endoderm-like cells (EndoLCs) also emerge in C_μPASEs, and these EndoLCs randomly intermix with MeLCs and PGCLCs (**Figure 5A**). scRNA-seq analysis of C_μPASEs obtained at t = 48 h further confirms the presence of EndoLCs in C_μPASEs (**Figures 5B and 5C**). It is worth noting that C_μPASEs do not develop a prominent central lumen (the pro-amniotic-like cavity) as in H_μPASEs (**Figure 5A**).

Transcriptomic comparisons between related corresponding cell clusters in H_μPASEs, C_μPASEs, and human and cynomolgus embryos (Ma et al., 2019; Nakamura et al., 2016; Sasaki et al., 2016; Tyser et al., 2021; Yang et al., 2021) reveal that such cell clusters from these *in vitro* and *in vivo* systems, regardless in the Meso (**Figure 5D**), AM (**Figure 5E**) or PGC (**Figure 5F**) lineages, show strong correlations based on ontogenic genes identified from *cynomolgus* embryo transcriptome data (Nakamura et al., 2016; Sasaki et al., 2016; Yang et al.,

362 2021). Thus, despite systematic variations that might be resulted from different sequencing
363 platforms, related corresponding cell types of different primate species still demonstrate
364 relatively high transcriptomic correlation and developmental coordination (**Figure 5E-5F**), *e.g.*,
365 human mesoderm cells and MeLCs from the H_μPASE at later developmental stages correlate
366 better with chimpanzee or monkey mesoderm cells at later stages. Additionally, corresponding
367 cell types from the CS7 human gastrula and H_μPASEs correlate better with cells from
368 C_μPASEs, as compared to cells from *cynomolgus* embryos, which may reflect the greater
369 genetic similarity between humans and chimpanzees. Consistently, when IWP2 was
370 supplemented into C_μPASEs, developments of PSLCs and PGCLCs were inhibited, whereas
371 ISL1+GATA3+ AMLCs still emerged, similar with IWP2-treated H_μPASEs (**Figure 5G**).
372

373 **NODAL is essential for mesoderm development**

374 Our DEG and pathway analyses of μPASE scRNA-seq data in **Figures 2G, 2H**, **3D-3F, S2F and**
375 **S2G** suggest complex cell-cell interactions involved in early post-implantation human
376 development. Thus, μPASE scRNA-seq data at t = 24 h were further analyzed using CellChat
377 (Jin et al., 2021) for inference and analysis of ligand-receptor interactions (**Figures S6A-S6C**).
378 Among cell clusters present in μPASEs at t = 24 h, AMLCs and PSLCs were identified as major
379 sources of signaling ligands involved in key developmental pathways, such as BMP and WNT
380 pathways (Ben-Haim et al., 2006; Bernardo et al., 2011; Clevers, 2006; Rivera-Perez and
381 Magnuson, 2005; Wang et al., 2014; Zhao, 2003) (**Figures S6A-S6C**). This is consistent with a
382 recent study of pre-gastrulation *cynomolgus* embryos (Yang et al., 2021), which shows that both
383 AM and PS / Meso cells upregulate *BMP4* and *WNT5B* and extraembryonic mesenchyme cells
384 show high expression of *BMP2* and *BMP4* (**Figures S6D-S6F**). μPASEs lack extraembryonic
385 mesenchyme cells, the effects of which might have been substituted by exogeneous
386 supplementation of BMP4 in the μPASE protocol. For comparison, ligand-receptor interaction
387 analysis was also conducted using CellChat on scRNA-seq data from E6.5 mouse embryos
388 (Pijuan-Sala et al., 2019), revealing the extraembryonic ectoderm as the only source of BMP and
389 WNT signals (**Figures S6G-S6I**). Interestingly, during the development of both μPASEs and
390 pre-gastrulation *cynomolgus* embryos, non-canonical WNT pathways show a greater signaling
391 strength than canonical WNT, whereas the opposite is observed for E6.5 mouse embryos
392 (**Figures S6A, S6D and S6G**).

393 Our ligand-receptor interaction analysis for μPASEs using CellChat further reveals that,
394 compared to BMP and WNT pathways identified as an "incoming signal" for multiple cell
395 clusters, NODAL signals appear to be perceived only by PSLCs (**Figures S6A and S6B**),
396 suggesting a critical role of NODAL signaling in PSLC development. To study the functional
397 role of NODAL, *NODAL*-knockout (KO) hPSC lines were generated and used for μPASE
398 development (**Figures S7A-S7E**). By t = 48 h, majority of cells in *NODAL*-KO μPASEs appear
399 to have developed into flattened ISL1+GATA3+ AMLCs, and there are no cells disseminating
400 away from the μPASE structure, in distinct contrast with wildtype control μPASEs (**Figures 6A-**
401 **6D and S7F**). Consistently, scRNA-seq data obtained from *NODAL*-KO μPASEs at t = 48 h
402 showed a dominant population of AMLCs at the expense of MeLCs (**Figures 6B,6C,6E-6G**),
403 further supporting the critical role of NODAL in MeLC development. PGCLCs were still evident
404 in *NODAL*-KO μPASEs at t = 48 h, albeit with a much less percentage compared with those in
405 wildtype controls (8.1% *vs.* 28.2%; **Figures 6D,6E-6G** and **S7G**). This observation suggests that
406 in μPASEs, NODAL signaling is involved not only in MeLC development, but also in PGCLC
407 specification.
408 We further conducted DEG and Gene Ontology (GO) enrichment analyses for PSLCs
409 from *NODAL*-KO and wildtype control μPASEs (**Figure 6H and 6I; Table S7**). In *NODAL*-KO
410 PSLCs, BMP target genes, such as *ID2*, *TFAP2A* and *ISL1*, are upregulated, whereas PS / Meso-
411 related genes, such as *MESP1* and *MIXL1*, and EMT related genes, including *SNAI1* and *VIM*,
412 are downregulated (**Figure 6H**). PI3K-Akt, WNT and focal adhesion signaling pathways appear
413 to be downstream targets of NODAL signaling in PSLCs (**Figure 6I**). We speculate that lineage
414 bifurcation between AMLCs and PSLCs from EpiLCs might be regulated by a competition
415 between BMP and NODAL signaling. Absence or repression of NODAL signaling in the μPASE
416 could lead to "hyper" BMP activities, which in turn causes excessive AMLC development
417 (marked by an expanded ISL1 domain) and greater BMP activities in PSLCs. We also compared
418 PGCLCs from wildtype and *NODAL*-KO μPASEs (**Figure S7H and Table S7**). Consistently,
419 when SB431542, a pharmacological NODAL signaling inhibitor, was supplemented into the
420 microfluidic device, development of PSLCs/MeLCs in μPASEs was completely inhibited
421 (**Figure S7I**). Impaired development of MeLCs in *NODAL*-KO μPASEs can be efficiently
422 reversed by supplementing ACTIVIN A, a NODAL pathway agonist, into the channel opposite
423 to BMP4 stimulation. In addition to rescuing MeLC development, supplementing ACTIVIN A to

424 *NODAL*-KO μPASEs also leads to the specification of BLIMP1+FOXA2+ EndoLCs (**Figure**
425 **S7J**). Supplementing ACTIVIN A to wildtype μPASEs also results in specification of
426 BLIMP1+FOXA2+ EndoLCs (**Figure S7K**). We also successfully generated μPASEs using a
427 different hPSC line, including its corresponding *NODAL*-KO line, and a different chimpanzee
428 iPSC line, and repeated IWP2 and SB431542 treatment assays with consistent results (**Figure**
429 **S8**).
430
431 **Stringent criteria for distinguishing human trophoblast and amniotic ectoderm**
432 *In vivo*, blastocyst trophoblast and pre-gastrulation AM both appear as flattened, squamous
433 epithelium, and these two cell types share many lineage markers (Blakeley et al., 2015;
434 Petropoulos et al., 2016; Tyser et al., 2021; Yang et al., 2021). Unique markers that can
435 distinguish between these two cell types remain elusive, leading to confusions about true cell
436 lineage identities in different human embryoids (Xu et al., 2002; Zhao et al., 2021). Through
437 comparative transcriptome analysis of AM cells in the CS7 human gastrula (Tyser et al., 2021)
438 and trophoblast cells in human blastocysts (Blakeley et al., 2015; Petropoulos et al., 2016), we
439 identified a Trophoblast_Amnion ontogenic gene list (**Figures S9A and S9B; Table S8**) and
440 applied this list to examine the properties of human trophoblast-like cells or AMLCs reported
441 previously (**Figure 7A**). This ontogenic gene list contains several previously reported AM
442 makers, such as *ISL1, GABRP* and *IGFBP3* (Yang et al., 2021), and trophoblast makers, such as
443 *GCM1, HAVCR1* and *CGA* (Li et al., 2019; Okae et al., 2018; Pillai et al., 2019). Notably,
444 several pan-preimplantation embryo markers, including *DPPA3* and *DNMT3L* (Guo et al., 2021;
445 Io et al., 2021; Yanagida et al., 2021) are found to be quite efficient markers for distinguishing
446 between blastocyst trophoblast and pre-gastrulation AM. Based on quantified correlation
447 coefficients between previously reported human trophoblast-like cells or AMLCs and human
448 trophoblast or AM cells, we conclude that the transcriptome of BMP4-treated primed hPSCs, as
449 in μPASEs, in 2D Transwell membrane-based AMLC differentiation assays (Zheng et al.,
450 2019b) and in 2D patterned gastrulation models (Minn et al., 2020), is similar to that of human
451 AM cells, whereas the transcriptome of trophoblast-like cells derived from naïve hPSCs (Dong et
452 al., 2020; Guo et al., 2021; Io et al., 2021; Yanagida et al., 2021) is consistent with that of human
453 blastocyst trophoblast. However, transcriptome of trophoblast-like cells derived from extended

454 pluripotent stem (EPS) cells as well as that of trophoblast-like cells present in recently reported
455 human blastocyst-like structures fall somewhat in-between (Liu et al., 2021; Yu et al., 2021).
456 Using stringent criteria, we further identified a subset of DEGs within the
457 Trophoblast_Amnion ontogenic gene list useful for distinguishing between human blastocyst
458 trophoblast and pre-gastrulation AM and thus between human trophoblast-like cells and AMLCs
459 (see **Methods**; **Figures 7B**, **S9A and S9B**). Specifically, markers expressed in human blastocyst
460 trophoblast but not in pre-gastrulation AM include *FABP3*, *GCM1*, *S100A14*, *DNMT3L*, *DPPA3*,
461 *HAVCR1*, *CGA*, *GTSF1* and *SNORD99*; conversely, genes expressed in human pre-gastrulation
462 AM but not in blastocyst trophoblast include *ISL1*, *HEY1*, *GABRP*, *MIF*, *PLA2G2A*, *IGFBP7*
463 and *IGFBP3* (**Figure 7B**). Many commonly used markers for human trophoblast or AM,
464 including *GATA3*, *TFAP2A*, *TFAP2C*, *CDX2*, *KRT7* and *KRT19*, are shared between them
465 (**Figure 7B**). We should note that although AM transcriptome is obtained from the CS7 human
466 gastrula (Tyser et al., 2021), which remains as the only *bona fide* human pre-gastrulation AM
467 data currently available, most of pre-gastrulation AM markers identified here, including *ISL1*,
468 *HEY1*, *GABRP* and *IGFBP3*, are upregulated in the pre-gastrulation AM of *cynomolgus* embryos
469 (Yang et al., 2021).
470 We further conducted immunofluorescence analyses of IVF human blastocysts (D6), *in*
471 *vitro* cultured *cynomolgus* embryos (D14), trophoblast stem cells (TSCs) (Okae et al., 2018), and
472 AMLCs derived from BMP4-treated primed hPSCs (Zheng et al., 2019b), to ascertain these
473 newly identified human blastocyst trophoblast and pre-gastrulation AM markers (**Figures 7C-7E**
474 **and S9C**). Consistently, trophoblast cells of D6 human blastocysts show positive
475 immunostaining for GCM1, FABP3, DPPA3 and HAVCR1, but are negative for ISL1 or
476 GABRP (**Figure 7C**). In contrast, AM cells in D14 *cynomolgus* embryos show clear
477 immunostaining for ISL1 and GABRP (**Figure 7D**). GATA3, TFAP2A and TFAP2C show
478 positive immunostaining in both trophoblast cells of D6 human blastocysts and AM cells of D14
479 *cynomolgus* embryos (**Figures 7C and 7E**). Immunofluorescence analyses of TSCs (Okae et al.,
480 2018), AMLCs (Zheng et al., 2019b) and μPASEs at t = 36 h further support that ISL1, GCM1
481 and HAVCR1 can be utilized for distinguishing between trophoblast-like cells and AMLCs,
482 whereas GATA3, TFAP2A or TFAP2C could not (**Figure S9C and S9D**). We should note that
483 ISL1 is a particular useful nuclear marker for distinguishing between human blastocyst
484 trophoblast and pre-gastrulation AM and thus between human trophoblast-like cells and AMLCs

485 (**Figures 7C-7E and S6C**). The function of ISL1 in the AM in inducing Meso development in
486 *cynomolgus* embryos has recently been documented (Yang et al., 2021).
487
488 **DISCUSSION**
489 Due to bioethical restrictions and practical limitations, experimentations with post-implantation
490 human or NHP monkey embryos remain challenging. As an ethically acceptable alternative,
491 human embryoids are emerging as promising tractable experimental tools for advancing
492 fundamental knowledge of human development (Fu et al., 2021; Rossant and Tam, 2021). As
493 demonstrated by μPASEs, with the aid of integrative bioengineering strategies, human
494 embryoids can faithfully recapitulate certain morphogenetic and cell fate patterning events
495 during the early post-implantation human development, with superior controllability and
496 reproducibility. μPASEs are particularly suitable for studying the specification of AM and PGCs,
497 since μPASE development shows spatiotemporal tissue patterning and architecture reminiscent
498 of those in the early post-implantation human embryo. By studying transcriptome kinetics during
499 progressive μPASE development, molecular maps of lineage diversifications from pluripotent
500 Epi towards AM, PS / Meso, and PGCs are constructed and further compared with *in vivo* human
501 and cynomolgus monkey data. Our data reveal that both WNT and NODAL signaling are
502 critically involved in Meso and PGC induction, but not for AM development, which is further
503 confirmed using *NODAL*-KO hPSC lines and drug inhibition assays. Through comparative
504 transcriptome analyses and validations with human and cynomolgus embryos, we further
505 propose some stringent criteria for distinguishing between human trophectoderm and AM cells.
506 Altogether, this study provides new knowledge of the lineage diversification roadmap of early
507 human development and will serve as a valuable resource for studying human development.
508 Experimentation on human embryoids can lead to a better understanding of the
509 mechanisms of human development and offers opportunities for functional genomic studies of
510 disease-causing mechanisms, identification of therapeutic targets, and preclinical modeling of
511 advanced therapeutics for precision medicine (Fu et al., 2021; Rossant and Tam, 2021).
512 Continuous development of human embryoids should lead to more authentic human
513 development models, with *in vivo*-like tissue architectures and spatiotemporal cell patterning and
514 organization reminiscent of those in natural human embryos. This is essential for understanding
515 the autonomy, self-organizing principles, and innate cell-cell interactions involved in human

516 development. Importantly, molecular insights generated from embryoids, if at all possible,
517 should be functionally validated using *in vivo* models (Yang et al., 2021). When developing new
518 embryoids or improving existing embryoids, rigorous scientific scrutiny must be implemented to
519 avoid incorrect interpretations or overstatements (Posfai et al., 2021; Zhao et al., 2021).
520 Currently, there are few lineage markers accepted for distinguishing between human
521 trophectoderm and AM cells, and most existing embryoids are still limited in imitating sequential
522 cell lineage diversifications and 3D tissue organizations exhibited stereotypically in natural
523 embryos. Besides using cell lineage markers, validation and authentication of embryoids are
524 currently commonly conducted through comparative transcriptome studies (*e.g.*, through
525 integration of scRNA-seq datasets from embryoids and *in vivo* models) (Posfai et al., 2021; Zhao
526 et al., 2021). However, caution should be taken when interpreting such integrated datasets, since,
527 although seemingly "unbiased", existing scRNA-seq data analysis tools still have notable
528 artifacts depending on the cell types present, cell numbers and percentages of each cell type, and
529 specific computational algorithms used in the analysis tools. For instance, AM cells only account
530 for a very small cell population in post-implantation primate embryos; such a small AM cell
531 number can cause significant issues when scRNA-seq data of AM cells are integrated with
532 human blastocyst datasets wherein trophoblast cells are abundant. Thus, we propose that in order
533 to validate cell identities in peri-implantation human embryoids, expression of key cell fate
534 markers need to be clearly demonstrated, in addition to transcriptome comparison based on
535 scRNA-seq data. The cell identity markers established in this work for distinguishing between
536 human trophoblast and AM cells represent one step towards this important direction.
537

538 **LIMITATIONS OF STUDY**

539 As demonstrated in this study, μPASEs provide a promising tractable experimental model for
540 exploring previously inaccessible phases of early post-implantation human development.
541 However, μPASEs lack a few key embryonic and extraembryonic lineages in the post-
542 implantation human embryo, including the hypoblast (or extraembryonic endoderm),
543 extraembryonic mesoderm, or trophoblast cells. The hypoblast is known to play a critical role in
544 anterior-posterior patterning of the epiblast prior to gastrulation. Therefore, μPASEs only
545 recapitulate certain aspects of the lineage diversification and development of the posterior end of
546 the embryonic sac during the early post-implantation human development. Additionally,

547 μPASEs disintegrate and lose their embryonic-like structure soon after 48 h in culture, as
548 differentiating MeLCs that are undergoing EMT emigrate from the μPASE structure, which
549 limits the potential of μPASEs for prolonged culture to investigate cell lineage development and
550 embryonic tissue formation at later developmental stages. It should also be noted that cell type
551 annotations and lineage inference analyses of the μPASE in this work are solely based on
552 transcriptomic studies. Thus, caution should be taken when interpreting these results included in
553 this Resource.
554

555 **ACKNOWLEDGEMENTS**

556 Studies of μPASE is supported by the Michigan-Cambridge Research Initiative (J.F.), the
557 National Institutes of Health (R21 NS113518 and R21 HD100931, J.F.), the National Science
558 Foundation (CMMI 1917304 and CBET 1901718, J.F.), and the 21st Century Jobs Trust Fund
559 received through the Michigan Strategic Fund from the State of Michigan (Grant CASE-315037,
560 J.F.). Studies of human blastocysts are supported by the Open Project of Yunnan Provincial
561 Reproductive and Obstetrics and Gynecology Clinical Medicine Center (zx2019-01-01 and
562 2020LCZXKF-SZ04, L.X.). Non-human primate monkey studies are supported by the National
563 Key Research and Development Program of China (2016YFA0101401), the Major Basic
564 Research of Yunnan from China (2019FY002), and a Grant for Swedish-Chinese collaboration
565 from the Swedish Research Council (Dnr 539-2013-7002). MK and TS are partially supported by
566 the RICBAC Foundation. K.C. is partially supported by a Grant for Distinguished Professors
567 from the Swedish Research Council (Dnr 541-2013-8351). A.M.R.I. is partially supported by the
568 National Science Foundation Graduate Research Fellowship under grant no. DGE 1256260. The
569 Lurie Nanofabrication Facility at the University of Michigan is acknowledged for support with
570 microfabrication.
571

572 **AUTHOR CONTRIBUTIONS**

573 Y.Z. and J.F. conceived and initiated the project; Y.Z. designed and performed microfluidic
574 embryoid experiments, and conducted scRNA-seq data analyses and interpretations; R.Z.Y.
575 conducted CellChat analysis and participated in data analysis; S.S. participated in microfluidic
576 embryoid experiments and repeated experiments with additional cell lines; M.K. and T.S.
577 established and characterized *NODAL*-KO hPSC lines; L.X., Y.L. and T.L. performed human

578 embryo experiments; R.Y., A.G., W.J., Y.N., and K.C. performed monkey embryo experiments;
579 X.X., S.N.E., Y.L. and A.R helped with microfluidic embryoid experiments; W.W. helped with
580 scRNA-seq data analyses; Y.Z. and J.F. wrote the manuscript; J.F. supervised the study. All
581 authors edited and approved the manuscript.
582
583 **COMPETING INTERESTS**
584 Two patents related to this work have been filed (US20190321415 / WO2018106997 by J. Fu, Y.
585 Zheng and S.N. Esfahani; US2020049721 / PCT/US20/49721 by J. Fu and Y. Zheng).

586 **METHODS**

587 **Ethics statement**

588 μPASE lacks both the primitive endoderm and trophoblast and thus cannot form the yolk sac or
589 placenta. Therefore, μPASE does not have human organismal form or potential. Furthermore,
590 μPASE disassembles and lose structural integrity after t = 48 h, and all experiments were
591 terminated by no later than t = 72 h. All protocols with hPSCs were approved by the Human
592 Pluripotent Stem Cell Research Oversight Committee at the University of Michigan, Ann Arbor.
593 Protocols of human embryo experiments were approved by the Medicine Ethics Committee of
594 the First People's Hospital of Yunnan Province (KHLL2020-KY064). This Medicine Ethics
595 Committee has 13 members, including lawyers, scientists and clinicians with relevant expertise.
596 This committee evaluated the scientific merits and ethical justification of human embryo
597 experiments and conducted a full review of the donation and use of human embryo samples. All
598 human embryos donated to this study were surplus frozen embryos from couples who had at
599 least one healthy baby after *in vitro* fertilization. All donor couples signed informed consents for
600 voluntary donations at the Department of Reproductive Medicine in the First People's Hospital
601 of Yunnan Province. No economic benefits were offered during the process. Couples were
602 informed that their embryos would be used for experimental studies of human development and
603 that their donation would not affect their *in vitro* fertilization processes.

604

605 **Cell culture**

606 H9 (WA09, WiCell; NIH registration number: 0062, female) and ESI-017 (BioTime, Inc.; NIH
607 registration number: 0093, male) hPSCs and chimpanzee iPSC lines (C3651, male and C4955,
608 female) (Gallego Romero et al., 2015; Pavlovic et al., 2018) were maintained in a feeder-free
609 culture system using mTeSR medium (STEMCELL Technologies). For hPSCs, culture plate was
610 coated with 1% lactate dehydrogenase-elevating virus (LDEV)-free, hESC cell-qualified reduced
611 growth factor basement membrane matrix Geltrex (Thermo Fisher Scientific; derived from
612 Engelbreth-Holm-Swarm mouse tumors) before cell seeding. For chimpanzee iPSCs, culture
613 plate was coated with 1% Matrigel (Thermo Fisher Scientific; extracted from Engelbreth-Holm-
614 Swarm mouse sarcoma). Cells were visually examined during each passage to ensure absence of
615 spontaneously differentiated, mesenchymal-like cells in culture. Cells were used before P70.
616 Both human and chimpanzee cells have been authenticated by original sources as well as in-

house by immunostaining for pluripotency markers and successful differentiation into the three germ layers. Cells were maintained for at least ten passages and authenticated as karyotypically normal. Karyotype analysis was performed by Cell Line Genetics. Both human and chimpanzee cell lines were tested negative for mycoplasma contamination (LookOut Mycoplasma PCR Detection Kit, Sigma-Aldrich).

Generation of *NODAL*-knockout hPSCs

To generate *NODAL*-knockout (KO) hPSCs, a 58-bp portion of genomic DNA within NODAL exon 1 was deleted by CRISPR/Cas9 using two crRNA purchased from Thermo Fisher Scientific [NODAL_crRNA_1: 5'-AGGCUCAGCAUGUACGCCAG-3'; NODAL_crRNA_2: 5'-AGACAUCAUCCGCAGCCUAC-3'] (**Figures S7A-S7C**). Deplexes of crRNA:tracrRNA were prepared using a standard protocol and introduced into H9 hPSCs with the Cas9 enzyme and the pCXLE-EGFP expression plasmid (a gift from Shinya Yamanaka; Addgene plasmid # 27082; RRID: Addgene_27082) for constitutional expression of EGFP using the NEON electroporation system (Thermo Fisher Scientific). EGFP-expressing single cells were collected and seeded onto Matrigel-coated 96-well plates by fluorescence-activated cell sorting (FACSAria Fusion, BD Biosciences) with CloneR single-cell culture supplement diluted with mTeSR Plus medium (STEMCELL Technologies). To detect the anticipated deletion, genomic DNA was isolated from single-cell derived clones and subjected to PCR using the following primers designed for amplification of NODAL exon 1 [Forward Primer: 5'-CTTCCTTCTGCACGCCTGGTGG-3'; Reverse Primer: 5'-CCAACCCACAGCACTTCCCGAG-3']. Resulting amplicons were subjected to Sanger sequencing using a primer 5'-CTTCCTTCTGCACGCCTGGTGG-3'. ESI-017 *NODAL*-KO hPSC line is a generous gift from Aryeh Warmflash at Rice University (Chhabra et al., 2019).

Western blotting

Wildtype and *NODAL*-KO hPSCs were exposed to GSK3 inhibitor CHIR99021 (CHIR, 10 μM; Cayman Chemical) for 24 h to augment expression of NODAL protein before cells were lysed with RIPA buffer containing the cOmplete protease inhibitor cocktail (Roche). Protein concentration was determined by the Bradford assay using Protein Assay Dye Reagent Concentrate (Bio-Rad). An equal amount of protein (80 μg) from cell lysates of widetype and

648 *NODAL*-KO hPSCs was resolved on 10% SDS-PAGE and transferred onto PVDF membranes
649 (Thermo Fisher Scientific). Immunostaining was performed by blocking PVDF membranes with
650 5% skim milk for 1 h at room temperature followed by incubation overnight at 4 ℃ with mouse
651 monoclonal antibodies to human NODAL (Abcam ab55676; 1:500 dilution) or human GAPDH
652 (Sigma-Aldrich G8795; 1:20000 dilution) diluted in 5% skim milk. Membranes were washed
653 with phosphate buffered saline (PBS) containing 0.1% Tween-20 (PBS-T) and stained with
654 HRP-conjugated anti-mouse IgG antibody (Santa Cruz sc-516102; 1:3000 dilution) diluted in 5%
655 skim milk for 1 h at room temperature. Protein bands were detected by a chemiluminescence
656 assay using SuperSignal West Pico Plus and SuperSignal West Femto reagents (Thermo Fisher
657 Scientific).
658
659 **Cynomolgus macaque**
660 Healthy cynomolgus monkeys (*Macaca fascicularis*), aged from 5 to 8 years old, were used in
661 this study. All animals were housed either at the facility of the Yunnan Key Laboratory of
662 Primate Biomedical Research (LPBR) in China, or at the Astrid Fagræus laboratory of the
663 Karolinska Institutet in Sweden. Both facilities are accredited by AAALAC international.
664 Experimental protocols for using cynomolgus macaque embryos were approved by the
665 Institutional Animal Care and Use Committee of LPBR in China (KBI K001115033/01,01) and
666 by the Jordbruksverket in Sweden (Ethical Permit Number N277/14). Animals involved in this
667 study were never used for other treatments.
668
669 ***In vitro* fertilization and culture of cynomolgus macaque embryo**
670 *In vitro* fertilized cynomolgus macaque embryos were generated as described previously (Niu et
671 al., 2014). Briefly, healthy female cynomolgus monkeys aged 5 to 8 years old with regular
672 menstrual cycles were selected as oocyte donors. Cynomolgus monkeys were treated with
673 recombinant human follicle stimulation hormone (Merck, Gonal-f) for 8 days, followed by
674 administration of recombinant human chorionic gonadotropin (Merck, Ovidrel) on day 9. After
675 32-35 h, oocytes were collected by laparoscopic follicular aspiration. Metaphase II (MII) oocytes
676 were used for intracytoplasmic sperm injection to generate zygotes, and fertilization was
677 confirmed by the presence of two pronuclei. Zygotes were cultured in embryo culture medium-9
678 (polyvinyl alcohol (0.1 mg/mL), calcium chloride (1.9 mM), magnesium chloride (0.46 mM),

679 potassium chloride (3.0 mM), sodium chloride (113.8 mM), sodium bicarbonate (25.0 mM),
680 sodium lactate (4.5 mM), Minimum Essential Medium (MEM) amino acid, MEM nonessential
681 amino acid, and gentamicin (10 mg/mL)) containing 10% fetal calf serum in 37 °C incubator
682 supplied with 5% CO_2 until the blastocyst stage. *In vitro* culture of NHP monkey blastocysts
683 beyond the implantation stage has been described previously (Yang et al., 2021). In brief, frozen
684 NHP monkey blastocysts were thawed using Thawing Media (Kizatato) and cultured in
685 blastocyst culture medium (Origio) for at least 4 h to recover. Blastocysts were then treated with
686 Acidic Tyrode's solution (Sigma) to remove zona pellucida before being transferred onto an
687 ibiTreat 8-well μ-plate (Ibidi) containing 300 μL of pre-equilibrated *in vitro* culture medium 1
688 (advanced DMEM/F12, 20% FBS, l-Glutamine, l-cysteine, 1x Penicillin/Streptomycin, 1x ITS-
689 X, supplemented with beta-estradiol, progesterone). On the second day, 150 μL of culture
690 medium was aspirated, before 200 μL of pre-equilibrated *in vitro* culture medium 2 (advanced
691 DMEM/F12, 30% KSR, l-Glutamine, l-cysteine, 1x Penicillin/Streptomycin, 1x ITS-X,
692 supplemented with beta-estradiol, progesterone) was added into the ibiTreat 8-well μ-plate.
693 Embryo growth was recorded daily, and culture medium was replenished every two days till Day
694 14.
695
696 **NHP monkey embryo cryosection and immunocytochemistry**
697 Day 14 NHP monkey embryos were fixed using 2% paraformaldehyde (PFA; buffered in PBS)
698 overnight at 4 °C before being washed with PBS. Fixed embryos were dehydrated by 30%
699 sucrose overnight at 4 ℃ before embedded in Tissue-Tek O.C.T. compound (Sakura) and frozen
700 in liquid nitrogen. Frozen blocks were used for cryosection using CryoStar NX70 Cryostat
701 (Thermo Fisher Scientific) according to manufacturer's protocol. Immunofluorescence staining
702 was then performed according to a standard procedure. Briefly, sections were thawed and air-
703 dried at room temperature. After washing with PBS, sections were incubated in blocking buffer
704 (3% FBS) diluted in PBS with 0.1% Triton X-100) for 1 h at room temperature and then
705 incubated overnight at 4 °C with primary antibodies diluted in blocking buffer. The sections
706 were washed with PBS-T and incubated with secondary antibodies diluted in blocking buffer for
707 2 h at room temperature. After washing thoroughly with PBS-T, sections were mounted and
708 imaged. Secondary antibodies were used in dilution of 1:500. Images were acquired by a Zeiss
709 700 LSM Confocal Microscope and analyzed by iMaris.

710

711 **Immunocytochemistry for D6 human embryos**

712 D6 human embryos were fixed with 4% PFA for 20 min at room temperature in a 96-well plate,
713 washed with PBS, and then permeabilized and blocked by 0.2% Triton X-100 supplemented with
714 3% bovine serum albumin (BSA) overnight at 4 ℃. Embryos were then transferred onto a new
715 well with primary antibodies for 16 - 18 h at 4 ℃. Embryos were washed 3 times with PBS
716 containing 0.05% Tween-20, 15 min every time, before being transferred to secondary antibody
717 solutions for 4 h at room temperature. Embryos were then washed 3 times in PBS containing
718 0.05% Tween-20, 15 min every time, before being transferred onto an ibiTreat 8-well μ-plate.
719 All antibodies were diluted by 1% BSA solutions.

720

721 **Generation of μPASEs**

722 The microfluidic device was fabricated by bonding a polydimethylsiloxane (PDMS) structure
723 layer to a glass coverslip. Singly dissociated hPSCs or chimpanzee iPSCs were loaded into the
724 cell loading channel and guided to settle into preformed, concave Geltrex pockets by tilting the
725 device by 90° for 10 min. hPSCs were then maintained in mTeSR containing 10 μM Y27632
726 (Tocris) for 18 h to allow for cell clustering. At t = 0 h, culture medium in all medium reservoirs
727 of the device was switched to a fresh basal medium comprising Essential 6 medium (E6; Thermo
728 Fisher Scientific) and FGF2 (20 ng mL^{-1}; GlobalStem), with additional 50 ng mL^{-1} BMP4 (R&D
729 Systems) supplemented only in the cell loading channel. A detailed protocol for microfluidic
730 device fabrication and generation of μPASEs can be found elsewhere(Zheng et al., 2021; Zheng
731 et al., 2019b). To examine possible involvements of different signaling pathways involved in the
732 development of μPASEs, IWP2 (10 μM; Tocris), SB 431542 (10 μM; Cayman Chemical) or
733 LY294002 (20 μM; BioVision) was supplemented into basal medium from t = 0 h. For assays
734 with ACTVIN A, 100 ng mL^{-1} ACTIVIN A (R&D Systems) was supplemented into the channel
735 opposite to BMP4 stimulation from t = 0 h. For *in situ* proliferation measurements, Click-iT EdU
736 Kit (Invitrogen) was used as per manufacturer's protocol. Diluted EdU solution (10 μM) was
737 introduced into all reservoirs of the device at t = 24 h. After 3 h of incubation, the μPASEs was
738 fixed and imaged.

739

740 **Immunocytochemistry for μPASEs**

741 μPASEs were fixed with 4% PFA for 12 h before being permeabilized in 0.1% SDS solution
742 (sodium dodecyl sulphate, dissolved in PBS) for 3 h. μPASEs were then blocked in 4% donkey
743 serum (Sigma-Aldrich) at 4 °C for 24 h, followed by incubation with primary antibody solutions
744 at 4 °C for another 24 h. Samples were then labelled with donkey-raised secondary antibodies
745 (1:500 dilution) at 4 °C for 24 h. 4′,6-diamidino-2-phenylindole (DAPI; Thermo Fisher
746 Scientific) was used for counterstaining cell nuclei. Both primary and secondary antibodies were
747 prepared in 4% donkey serum supplemented with 0.1% NaN_3. 70 μL antibody solutions were
748 added to each medium reservoir of the microfluidic device for immunostaining. All primary and
749 secondary antibodies used in this study are listed in **Table S9**. All confocal micrographs of
750 μPASEs were acquired by a NIKON A1SI Confocal Microscope equipped with a
751 photomultiplier tube (PMT) detector and processed using ImageJ 1.53c.
752
753 **Quantification of SOX17+ cells**
754 μPASEs at t = 48 h with or without LY294002 were stained for SOX17 using the protocol
755 described above. All confocal micrographs captured at the central focal plane of structure (50 μm
756 above the microfluidic device bottom surface) were used for quantification. The percentage of
757 SOX17+ cells was calculated as the ratio between the area of SOX17 channel and DAPI channel
758 for each μPASE. The binarization and measurement of the images was conducted using ImageJ
759 1.53c. Quantitative results were analyzed using independent, two-tailed Student's t-test in Excel
760 (Microsoft). $P < 0.05$ was considered statistically significant.
761
762 **Trophoblast stem cells and derivation of AMLCs using BMP4**
763 Human trophoblast stem cells (hTSCs) derived from blastocysts were generously provided by
764 Dr. H. Okae and Dr. T. Arima (Okae et al., 2018). hTSCs were maintained in 1% Geltrex coated
765 6-well plates in DMEM/F12 supplemented with 0.05 mM 2-mercaptoethanol, 0.2% fetal bovine
766 serum (FBS), 0.5% knockout serum replacement (KSR), 0.5% penicillin-streptomycin, 0.3%
767 BSA, 1% ITS-X supplement, 1.5 μg mL^{-1} L-ascorbic acid, 50 ng mL^{-1} epidermal growth factor
768 (EGF), 2 μM CHIR99021, 1 μM A83-01, 1 μM SB431542, 0.8 mM valproic acid (VPA) and 5
769 μM Y27632, and passaged using TrypLE.
770 To obtain AMLCs by treating hPSCs with BMP4, singly dissociated hPSCs were
771 suspended in mTesR1 containing 10 μM Y27632 and seeded in a 1% Geltrex coated 6-well plate

772 at a density of 2.5×10^3 cells cm^{-2}. Note that this cell seeding density is optimized to avoid
773 extensive cell death (when cell density is too low) or emergence of PSLCs in culture through an
774 unspecified inductive effect of incipient AMLCs (Zheng et al., 2019b). 18 h after cell seeding,
775 culture medium was switched to Essential 6 medium containing FGF2 (20 ng mL^{-1}) and BMP4
776 (50 ng mL^{-1}). Resulting AMLCs were fixed and stained after 48 h of BMP4 treatment.
777
778 **Single cell dissociation and RNA-sequencing**
779 μPASEs (H9) at different time points were washed twice with DMEM/F12 for 10 min and
780 incubated with Accutase for 1 h. After incubation, μPASEs in the microfluidic device were
781 dissociated into single cells by gentle agitating. Single cells from six microfluidic devices were
782 collected and pooled into PBS containing 0.5% BSA before being centrifuged at 300 g for 5 min.
783 The resultant cell pellet was re-suspended in PBS containing 0.5% BSA. Within 1 h after cell
784 dissociation, cells were loaded into the 10X Genomics Chromium system (10X Genomics). 10X
785 Genomics v.3 libraries were prepared according to the manufacturer's instructions. Libraries
786 were then sequenced using paired-end sequencing with a minimum coverage of 20,000 raw reads
787 per cell using an Illumina NovaSeq-6000. scRNA-seq data were aligned and quantified using
788 Cell Ranger Single-Cell Software Suite (v.3.1.0, 10X Genomics) against the Homo sapiens
789 (human) genome assembly GRCh38.p13 from ENSEMBL. Chimpanzee μPASEs (C3651) were
790 dissociated and sequenced following the same protocol. scRNA-seq data from chimpanzee
791 μPASEs were aligned against Pan_tro_3.0 from ENSEMBL.
792
793 **Data integration, dimensionality reduction and clustering**
794 Analysis of scRNA-seq data and integration of scRNA-seq datasets were performed using Seurat
795 R package (v.4.0.0.0, https://satijalab.org/seurat/) (Butler et al., 2018; Satija et al., 2015). Default
796 setups were used unless noted otherwise. Briefly, a single batch of scRNA-seq dataset was
797 filtered based on total number of genes detected and percentage of total mitochondrial genes.
798 Gene expression was then calculated by normalizing the raw count with the total count before
799 multiplying by 10,000 and log transformed. Cell cycle was regressed out based on cell cycle
800 scores (CellCycleScoring) during data scaling process (ScaleData). PCA analysis (RunPCA) was
801 then performed on filtered data followed by embedding into low dimensional space with
802 Uniform Manifold Approximation and Projection (UMAP; RunUMAP). Identification of cell

803 clusters by a shared nearest neighbor (SNN) modularity optimization-based clustering algorithm
804 was achieved using the FindClusters function in the Seurat R package. For integration of
805 different scRNA-seq datasets, count matrices of different datasets were filtered and normalized
806 separately before being integrated using the IntegrateData function based on 2,000 anchor
807 features. After integration, the integrated scRNA-seq dataset was analyzed following the
808 standard Seurat pipeline. Annotation of cell clusters was based on expression of canonical
809 lineage marker genes.
810
811 **RNA velocity analysis and partition-based graph abstraction analysis (PAGA)**
812 Bam files generated by the Cell Ranger pipeline was used for RNA velocity analysis. Genome
813 annotations GRCh38.p13 were used for counting spliced and unspliced mRNA of individual
814 cells. Python package scVelo (v.0.2.2, https://scvelo.readthedocs.io) was employed to perform
815 RNA velocity analysis using dynamical modeling (scv.tl.velocity) (Bergen et al., 2020; La
816 Manno et al., 2018). Function 'scv.pl.velocity_embedding_stream' was used to project RNA
817 velocities onto UMAP plots or diffusion maps. All default parameters were used unless noted
818 otherwise. Python package Scanpy (v.1.8.0, https://scanpy.readthedocs.io/en/stable/) was used
819 for evaluating the relationship between different cell clusters by the partition-based graph
820 abstraction (PAGA) analysis (Wolf et al., 2018). Briefly, Seurat object was converted to "Loom"
821 file and passed to Scanpy. Neighborhood graph of observations was then computed using 20 PCs
822 (sc.pp.neighbors). Finally, PAGA graph was plotted with a "eq_tree" layout.
823
824 **Gene regulatory network analysis**
825 Regulatory activity of transcription factors associated with specific cell types was assessed using
826 the R-package SCENIC (Single Cell rEgulatory Network Inference and Clustering, v.1.1.2-2,
827 https://github.com/aertslab/SCENIC) (Aibar et al., 2017). Briefly, regulatory modules were first
828 identified by inferring co-expression with transcription factors using GENIE3. Each co-
829 expression module was then analyzed using cis-regulatory motif analyses (RcisTarget). Only
830 modules with significant motif enrichment of the correct upstream regulator were retained. The
831 human motif collection v9 and the cisTarget databases for hg38 were used in the pipeline
832 (https://resources.aertslab.org/cistarget/). Filtered counts of the integrated Seurat object were
833 used as input of SCENIC. All default parameters were used in SCENIC unless noted otherwise.

834

835 **Trajectory interference using diffusion map and pseudotime**

836 Diffusion maps were obtained by the R-package Density
837 (https://bioconductor.org/packages/release/bioc/html/destiny.html), which computes kernel
838 density estimates with parametric starts and asymmetric kernels (Angerer et al., 2016). To
839 generate three-dimensional diffusion maps, PCA embeddings of the integrated Seurat object
840 were used as input of 'diffmap' function. Cells in the three-dimensional diffusion map were
841 color-coded consistently with the UMAP plot of the integrated Seurat object. Three-dimensional
842 visualization was realized using R-package rgl (3D Visualization Using OpenGL). For trajectory
843 inference and pseudotime analysis of specific cell lineages, relevant cell clusters of the integrated
844 Seurat object were extracted using Subset function (Seurat). PCA embeddings of selected cell
845 clusters were then used as the input of 'diffmap' function. To visualize gene expression
846 dynamics, diffusion pseudotime (dpt) was utilized. The roots of the diffusion map were
847 automatically chosen by the algorithm; and the EpiLC cluster was arbitrarily made as the initial
848 cell type. Expression levels of selected genes were fitted as a function of the pseudotime with
849 "loess" method by using 'geom_smooth' function in ggplot2 R package (v.3.3.3). Cells with
850 extreme dpt values were counted as outliner and excluded from gene expression dynamic
851 analysis. RNA velocity vectors were computed as described above and superimposed onto the
852 diffusion maps. In the case of lineage bifurcation, branching point and branches of different
853 lineages were identified by locally fitting half-lines to single-cell data in the diffusion map using
854 K-Branches R package (https://github.com/theislab/kbranches) (Chlis et al., 2017).

855

856 **Differential gene expression, gene ontology enrichment and pathway analyses**

857 Gene expression data depicted in feature plots and dot plots in this paper were calculated from
858 raw counts after NormalizeData function followed by ScaleData function of the Seurat package,
859 unless noted otherwise. Differentially expressed genes (DEGs) between different cell types were
860 identified using FindMarkers function (Seurat), with a minimal fold difference of 0.25 in the
861 logarithmic scale and > 10% detection rate in either of the two cell types under comparison.
862 Gene ontology (GO) enrichment and pathway analyses were performed using online tool
863 iPathwayGuide (Advaita Bioinformatics) referencing AmiGO Gene Ontology database and
864 KEGG PATHWAY Database, respectively.

865

866 **Integration, re-analysis and PCA of published human embryo data**

867 The scRNA-seq dataset of the CS7 human gastrula was integrated with μPASE scRNA-seq
868 datasets. Since μPASEs contain AMLCs, PGCLCs and PSLCs / MeLCs, only relevant cell types
869 from the human gastrula (Amniotic/embryonic ectoderm, Epiblast, Primitive Streak, Nascent
870 Mesoderm, Emergent Mesoderm) were extracted according to the annotations in the original
871 publication. Human gastrula scRNA-seq dataset was generated using Smart-seq2. To compare
872 with μPASE scRNA-seq datasets generated using the 10X Genomics Chromium system, raw
873 counts of each cell in the human gastrula were normalized to exon sizes before being utilized to
874 create the Seurat object (*input count* = *raw count* / *exon size* × 1,000). Exon size information was
875 obtained from GRCh38.p13, ENSEMBL. To integrate scRNA-seq data from the human gastrula
876 with those of μPASEs, μPASE scRNA-seq datasets from the three time points (t = 24 h, 36 h and
877 48 h) were first filtered. Normalization, Scaling (including cell cycle regression) and PCA were
878 then performed separately on each of the datasets after which they were combined using the
879 reciprocal PCA approach (IntegrateData function, Seurat) based on 30 dimensions and 2,000
880 anchor features. Dimensionality reduction and clustering were then performed as described
881 above. Downsampling of large scRNA-seq datasets was performed using Seurat R package
882 (Subset), which randomly selects 100 cells from every cluster in the original datasets to form a
883 new Seurat object. This step helps prevent larger datasets from dominating the downstream
884 analysis.

885 For re-analysis of human gastrula scRNA-seq dataset, relevant cell types were processed
886 with the default Seurat pipeline. Trajectory inference using diffusion map and pseudotime
887 analysis was achieved using the same pipeline as for μPASEs, as descried earlier. For PCA,
888 large datasets were first downsampled; then the normalized counts of epiblast ontogenetic genes
889 (**Table S5**) from different datasets were merged into a matrix. The matrix was scaled to ensure
890 each gene has zero average and unit variance. PCAs was calculated and visualized using Seurat
891 R package.

892

893 **Cross-species comparison**

894 Cross-species comparisons utilize scRNA-seq datasets from the CS7 human gastrula, human
895 μPASEs (t = 24 h, 36 h and 48 h), chimpanzee μPASEs (t = 48 h), *in vivo* cynomolgus embryos

896 (D13 - 17), and *in vitro* cultured cynomolgus embryos (D11 - 17) (Ma et al., 2019; Nakamura et
897 al., 2016; Sasaki et al., 2016; Tyser et al., 2021; Yang et al., 2021). To mitigate artifacts resulted
898 from different sequencing platforms, for datasets generated using Smart-seq2, raw counts were
899 first normalized to exon length, as described above. Exon size information of cynomolgus
900 monkey was obtained from Macaca_fascicularis_6.0, ENSEMBLE. Average raw counts
901 normalized by exon length for specific cell types in different datasets were then transformed into
902 log2(counts per million reads + 1) and used as gene expression levels for calculating the
903 correlation coefficient matrix and heat map plotting.
904 Ontogenic genes of mesoderm and PGC lineages were identified through DEG analysis
905 of *in vivo* cynomolgus embryo datasets. Briefly, for mesoderm lineage ontogenic genes, DEGs of
906 "Gast1", "Gast2a" and "Gast2b" as compared to "PostL-EPI" were identified using FindMarkers
907 function (Seurat)(Nakamura et al., 2016). All DEGs, with a minimal fold difference of 2 in the
908 logarithmic scale (logfc), > 10% detection rate in either of the two cell types under comparison,
909 and adjusted *P*-value < 0.05, were used as mesoderm lineage ontogenic genes (**Table S6**). To
910 identify PGC lineage ontogenic genes, DEGs of "ePGC" as compared to "PostL-EPI"(Sasaki et
911 al., 2016), with a minimal fold difference of 2 in the logarithmic scale (logfc), > 10% detection
912 rate in either of the two cell types under comparison, and adjusted *P*-value < 0.05, were used as
913 PGC lineage ontogenic genes (**Table S6**). Amniotic ectoderm ontogenic genes were obtained
914 through DEG analysis of Day 14 *in vitro* cultured cynomolgus embryos (Yang et al., 2021).
915 Specifically, DEGs of "early-Amnion" and "late-Amnion" as compared to "EPI" were identified
916 using FindMarkers function (Seurat). All DEGs, with a minimal fold difference of 0.75 in the
917 logarithmic scale (logfc), > 10% detection rate in either of the two cell types under comparison,
918 and adjusted *P*-value < 0.01, were used as amniotic lineage ontogenic genes (**Table S6**).
919
920 **Analysis of cell-cell interactions**
921 R package CellChat was used to perform cell-cell communication analysis
922 (http://www.cellchat.org/) (Jin et al., 2021). Briefly, based on manually curated databases that
923 consider known structural compositions of ligand-receptor interactions, CellChat infers and
924 analyzes intercellular communication networks from scRNA-seq data using network analysis and
925 pattern recognition. Seurat object including count matrix and clustering results from each dataset
926 is imported to CellChat. The default human database was used for μPASE and cynomolgus

927 embryo dataset analyses, whereas the default mouse database was used for mouse embryo
928 dataset analysis. For both databases, only secreted signaling pathways from Kyoto Encyclopedia
929 of Genes and Genomes (KEGG) were used. Default values were used for all parameters, except
930 that the truncated mean was lowered to 5% to increase algorithm sensitivity.
931
932 **Comparisons between human trophoblast and amniotic ectoderm**
933 Several published RNA-seq datasets related to human blastocyst trophoblast and amniotic
934 ectoderm were surveyed in this work. Datasets generated from bulk RNA-sequencing or Smart-
935 seq2 were first normalized by exon sizes as described earlier. Specifically, blastocyst trophoblast
936 cells annotated in (Blakeley et al., 2015; Petropoulos et al., 2016) were utilized as the reference
937 of human blastocyst trophoblast. Both two datasets were analyzed in this work to mitigate batch
938 effects and systematic variations. Amniotic/embryonic ectoderm cells from the CS7 human
939 gastrula were utilized as the reference of human amniotic ectoderm (Tyser et al., 2021). DEGs
940 between trophoblast and amniotic ectoderm were first identified using trophoblast cells from
941 (Blakeley et al., 2015; Petropoulos et al., 2016), respectively, as compared to
942 amniotic/embryonic ectoderm cells from the CS7 human gastrula, with a minimal fold difference
943 of 0.75 in the logarithmic scale (logfc) and adjusted *P*-value < 0.05. Overlapped DEGs identified
944 from both comparative transcriptome analyses were used as Trophoblast_Amnion ontogenic
945 genes (n = 299) to calculate correlation coefficients of published datasets as well as AMLCs
946 from this work with human blastocyst trophoblast and amniotic/embryonic ectoderm (**Table**
947 **S8**). Correlation coefficients and gene expression heatmap were calculated based on averages of
948 experimental repeats, if any. The average gene expression level of the two trophoblast datasets
949 was used to calculate correlation coefficients of surveyed cells with the human blastocyst
950 trophoblast. To identify the most reliable markers for distinguishing trophoblast cells from
951 amniotic ectoderm cells, more stringent criteria were applied than those for identifying
952 Trophoblast_Amnion ontogenic genes. Among Trophoblast_Amnion ontogenic genes with
953 logfc > 2, only the genes that are negative in the other cell type (normalized count < 0.05) were
954 identified as reliable markers. Overlapped markers revealed for both trophoblast datasets were
955 proposed as markers for distinguishing human trophoblast cells from amniotic ectoderm cells.
956
957 **Code availability**

RStudio and Python scripts used in this work are available from the corresponding authors upon request.

Data availability

All processed datasets are available at https://umichibbl.shinyapps.io/shinyapp/. Data supporting findings of this study are available within the article and its Supplemental Information files and from the corresponding authors upon request. scRNA-seq data are also available at the Gene Expression Omnibus under accession no. GSE185643. All source data for graphs included in the paper are available in the online version of the paper.

967 **REFERENCES**

968 Aibar, S., Gonzalez-Blas, C.B., Moerman, T., Huynh-Thu, V.A., Imrichova, H., Hulselmans, G., Rambow, F.,
969 Marine, J.C., Geurts, P., Aerts, J.*, et al.* (2017). SCENIC: single-cell regulatory network inference and
970 clustering. Nat Methods *14*, 1083-1086.

971 Angerer, P., Haghverdi, L., Buttner, M., Theis, F.J., Marr, C., and Buettner, F. (2016). destiny: diffusion
972 maps for large-scale single-cell data in R. Bioinformatics *32*, 1241-1243.

973 Beccari, L., Moris, N., Girgin, M., Turner, D.A., Baillie-Johnson, P., Cossy, A.C., Lutolf, M.P., Duboule, D.,
974 and Arias, A.M. (2018). Multi-axial self-organization properties of mouse embryonic stem cells into
975 gastruloids. Nature *562*, 272-276.

976 Ben-Haim, N., Lu, C., Guzman-Ayala, M., Pescatore, L., Mesnard, D., Bischofberger, M., Naef, F.,
977 Robertson, E.J., and Constam, D.B. (2006). The nodal precursor acting via activin receptors induces
978 mesoderm by maintaining a source of its convertases and BMP4. Dev Cell *11*, 313-323.

979 Bergen, V., Lange, M., Peidli, S., Wolf, F.A., and Theis, F.J. (2020). Generalizing RNA velocity to transient
980 cell states through dynamical modeling. Nat Biotechnol *38*.

981 Bernardo, A.S., Faial, T., Gardner, L., Niakan, K.K., Ortmann, D., Senner, C.E., Callery, E.M., Trotter, M.W.,
982 Hemberger, M., Smith, J.C.*, et al.* (2011). BRACHYURY and CDX2 Mediate BMP-Induced Differentiation of
983 Human and Mouse Pluripotent Stem Cells into Embryonic and Extraembryonic Lineages. Cell Stem Cell *9*,
984 144-155.

985 Blakeley, P., Fogarty, N.M.E., del Valle, I., Wamaitha, S.E., Hu, T.X., Elder, K., Snell, P., Christie, L., Robson,
986 P., and Niakan, K.K. (2015). Defining the three cell lineages of the human blastocyst by single-cell RNA-
987 seq (vol 142, pg 3151, 2015). Development *142*, 3613-3613.

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

990 Casaca, A., Nóvoa, A., and Mallo, M. (2016). Hoxb6 can interfere with somitogenesis in the posterior
991 embryo through a mechanism independent of its rib-promoting activity. Development *143*, 437-448.

992 Chawengsaksophak, K., de Graaff, W., Rossant, J., Deschamps, J., and Beck, F. (2004). Cdx2 is essential
993 for axial elongation in mouse development. Proc Natl Acad Sci U S A *101*, 7641-7645.

994 Chen, B., Dodge, M.E., Tang, W., Lu, J., Ma, Z., Fan, C.-W., Wei, S., Hao, W., Kilgore, J., Williams, N.S.*, et*
995 *al.* (2009). Small molecule–mediated disruption of Wnt-dependent signaling in tissue regeneration and
996 cancer. Nature Chemical Biology *5*, 100-107.

997 Chen, D., Sun, N., Hou, L., Kim, R., Faith, J., Aslanyan, M., Tao, Y., Zheng, Y., Fu, J., Liu, W.*, et al.* (2019).
998 Human Primordial Germ Cells Are Specified from Lineage-Primed Progenitors. Cell Rep *29*, 4568-4582
999 e4565.

1000 Chhabra, S., Liu, L., Goh, R., Kong, X., and Warmflash, A. (2019). Dissecting the dynamics of signaling
1001 events in the BMP, WNT, and NODAL cascade during self-organized fate patterning in human
1002 gastruloids. PLOS Biology *17*, e3000498.

1003 Chlis, N.K., Wolf, F.A., and Theis, F.J. (2017). Model-based branching point detection in single-cell data
1004 by K-branches clustering. Bioinformatics *33*, 3211-3219.

1005 Clark, A.T., Brivanlou, A., Fu, J.P., Kato, K., Mathews, D., Niakan, K.K., Rivron, N., Saitou, M., Surani, A.,
1006 Tang, F.C.*, et al.* (2021). Human embryo research, stem cell-derived embryo models and in vitro
1007 gametogenesis: Considerations leading to the revised ISSCR guidelines. Stem Cell Rep *16*, 1416-1424.

1008 Clevers, H. (2006). Wnt/beta-catenin signaling in development and disease. Cell *127*, 469-480.

1009 Daoud, G., Kempf, H., Kumar, D., Kozhemyakina, E., Holowacz, T., Kim, D.W., Ionescu, A., and Lassar, A.B.
1010 (2014). BMP-mediated induction of GATA4/5/6 blocks somitic responsiveness to SHH. Development *141*,
1011 3978-3987.

1012 Deglincerti, A., Croft, G.F., Pietila, L.N., Zernicka-Goetz, M., Siggia, E.D., and Brivanlou, A.H. (2016). Self-
1013 organization of the in vitro attached human embryo. Nature *533*, 251-254.

1014 Dong, C., Beltcheva, M., Gontarz, P., Zhang, B., Popli, P., Fischer, L.A., Khan, S.A., Park, K.-m., Yoon, E.-J.,
1015 Xing, X.*, et al.* (2020). Derivation of trophoblast stem cells from naïve human pluripotent stem cells. Elife
1016 *9*, e52504.

1017 Fu, J., Warmflash, A., and Lutolf, M.P. (2021). Stem-cell-based embryo models for fundamental research
1018 and translation. Nat Mater *20*, 132-144.

1019 Gallego Romero, I., Pavlovic, B.J., Hernando-Herraez, I., Zhou, X., Ward, M.C., Banovich, N.E., Kagan, C.L.,
1020 Burnett, J.E., Huang, C.H., Mitrano, A.*, et al.* (2015). A panel of induced pluripotent stem cells from
1021 chimpanzees: a resource for comparative functional genomics. Elife *4*, e07103.

1022 Gao, X.F., Nowak-Imialek, M., Chen, X., Chen, D.S., Herrmann, D., Ruan, D.G., Chen, A.C.H., Eckersley-
1023 Maslin, M.A., Ahmad, S., Lee, Y.L.*, et al.* (2019). Establishment of porcine and human expanded potential
1024 stem cells. Nat Cell Biol *21*, 687-+.

1025 Gibbs, R.A., Rogers, J., Katze, M.G., Bumgarner, R., Weinstock, G.M., Mardis, E.R., Remington, K.A.,
1026 Strausberg, R.L., Venter, J.C., Wilson, R.K.*, et al.* (2007). Evolutionary and biomedical insights from the
1027 rhesus macaque genome. Science *316*, 222-234.

1028 Gilbert, S. (2000). Developmental Biology, 6th edition (Sinauer Associates).

1029 Guo, G., Stirparo, G.G., Strawbridge, S.E., Spindlow, D., Yang, J., Clarke, J., Dattani, A., Yanagida, A., Li,
1030 M.A., Myers, S.*, et al.* (2021). Human naive epiblast cells possess unrestricted lineage potential. Cell
1031 Stem Cell *28*, 1040-1056 e1046.

1032 Haremaki, T., Metzger, J.J., Rito, T., Ozair, M.Z., Etoc, F., and Brivanlou, A.H. (2019). Self-organizing
1033 neuruloids model developmental aspects of Huntington's disease in the ectodermal compartment. Nat
1034 Biotechnol *37*, 1198-1208.

1035 Harrison, S.E., Sozen, B., Christodoulou, N., Kyprianou, C., and Zernicka-Goetz, M. (2017). Assembly of
1036 embryonic and extraembryonic stem cells to mimic embryogenesis in vitro. Science *356*, eaal1810.

1037 Hyun, I., Wilkerson, A., and Johnston, J. (2016). Revisit the 14-day rule. Nature *533*, 169-171.

1038 Io, S., Kabata, M., Iemura, Y., Semi, K., Morone, N., Minagawa, A., Wang, B., Okamoto, I., Nakamura, T.,
1039 Kojima, Y.*, et al.* (2021). Capturing human trophoblast development with naive pluripotent stem cells in
1040 vitro. Cell Stem Cell *28*, 1023-1039 e1013.

1041 Irie, N., Weinberger, L., Tang, W.W.C., Kobayashi, T., Viukov, S., Manor, Y.S., Dietmann, S., Hanna, J.H.,
1042 and Surani, M.A. (2015). SOX17 Is a Critical Specifier of Human Primordial Germ Cell Fate. Cell *160*, 253-
1043 268.

1044 Jin, S., Guerrero-Juarez, C.F., Zhang, L., Chang, I., Ramos, R., Kuan, C.H., Myung, P., Plikus, M.V., and Nie,
1045 Q. (2021). Inference and analysis of cell-cell communication using CellChat. Nat Commun *12*, 1088.

1046 Kadowaki, T., Wilder, E., Klingensmith, J., Zachary, K., and Perrimon, N. (1996). The segment polarity
1047 gene porcupine encodes a putative multitransmembrane protein involved in Wingless processing. Genes
1048 & Development *10*, 3116-3128.

1049 Kinoshita, M., Barber, M., Mansfield, W., Cui, Y., Spindlow, D., Stirparo, G.G., Dietmann, S., Nichols, J.,
1050 and Smith, A. (2021). Capture of Mouse and Human Stem Cells with Features of Formative Pluripotency.
1051 Cell Stem Cell *28*, 453-471.e458.

1052 Kobayashi, T., Zhang, H., Tang, W.W.C., Irie, N., Withey, S., Klisch, D., Sybirna, A., Dietmann, S.,
1053 Contreras, D.A., Webb, R.*, et al.* (2017). Principles of early human development and germ cell program
1054 from conserved model systems. Nature *546*, 416-420.

1055 La Manno, G., Soldatov, R., Zeisel, A., Braun, E., Hochgerner, H., Petukhov, V., Lidschreiber, K., Kastriti,
1056 M.E., Lonnerberg, P., Furlan, A.*, et al.* (2018). RNA velocity of single cells. Nature *560*, 494-498.

1057 Li, Z., Kurosawa, O., and Iwata, H. (2019). Establishment of human trophoblast stem cells from human
1058 induced pluripotent stem cell-derived cystic cells under micromesh culture. Stem Cell Research &
1059 Therapy *10*, 245.

1060 Liu, X., Tan, J.P., Schroder, J., Aberkane, A., Ouyang, J.F., Mohenska, M., Lim, S.M., Sun, Y.B.Y., Chen, J.,
1061 Sun, G.*, et al.* (2021). Modelling human blastocysts by reprogramming fibroblasts into iBlastoids. Nature
1062 *591*, 627-632.

1063 Lovell-Badge, R., Anthony, E., Barker, R.A., Bubela, T., Brivanlou, A.H., Carpenter, M., Charo, R.A., Clark,
1064 A., Clayton, E., Cong, Y.L.*, et al.* (2021). ISSCR Guidelines for Stem Cell Research and Clinical Translation:
1065 The 2021 update. Stem Cell Rep *16*, 1398-1408.

1066 Ma, H., Zhai, J., Wan, H., Jiang, X., Wang, X., Wang, L., Xiang, Y., He, X., Zhao, Z.A., Zhao, B.*, et al.* (2019).
1067 In vitro culture of cynomolgus monkey embryos beyond early gastrulation. Science *366*, eaax7890.

1068 Mikkelsen, T.S., Hillier, L.W., Eichler, E.E., Zody, M.C., Jaffe, D.B., Yang, S.P., Enard, W., Hellmann, I.,
1069 Lindblad-Toh, K., Altheide, T.K.*, et al.* (2005). Initial sequence of the chimpanzee genome and
1070 comparison with the human genome. Nature *437*, 69-87.

1071 Minn, K.T., Fu, Y.C., He, S., Dietmann, S., George, S.C., Anastasio, M.A., Morris, S.A., and Solnica-Krezel, L.
1072 (2020). High-resolution transcriptional and morphogenetic profiling of cells from micropatterned human
1073 ESC gastruloid cultures. Elife *9*.

1074 Molè, M.A., Coorens, T.H.H., Shahbazi, M.N., Weberling, A., Weatherbee, B.A.T., Gantner, C.W., Sancho-
1075 Serra, C., Richardson, L., Drinkwater, A., Syed, N.*, et al.* (2021). A single cell characterisation of human
1076 embryogenesis identifies pluripotency transitions and putative anterior hypoblast centre. Nat Commun
1077 *12*, 3679.

1078 Moris, N., Anlas, K., van den Brink, S.C., Alemany, A., Schroder, J., Ghimire, S., Balayo, T., van
1079 Oudenaarden, A., and Martinez Arias, A. (2020). An in vitro model of early anteroposterior organization
1080 during human development. Nature *582*, 410-415.

1081 Nakamura, T., Okamoto, I., Sasaki, K., Yabuta, Y., Iwatani, C., Tsuchiya, H., Seita, Y., Nakamura, S.,
1082 Yamamoto, T., and Saitou, M. (2016). A developmental coordinate of pluripotency among mice,
1083 monkeys and humans. Nature *537*, 57-62.

1084 Niu, Y., Shen, B., Cui, Y., Chen, Y., Wang, J., Wang, L., Kang, Y., Zhao, X., Si, W., Li, W.*, et al.* (2014).
1085 Generation of gene-modified cynomolgus monkey via Cas9/RNA-mediated gene targeting in one-cell
1086 embryos. Cell *156*, 836-843.

1087 Niu, Y., Sun, N., Li, C., Lei, Y., Huang, Z., Wu, J., Si, C., Dai, X., Liu, C., Wei, J.*, et al.* (2019). Dissecting
1088 primate early post-implantation development using long-term in vitro embryo culture. Science *366*,
1089 eaaw5754.

1090 Okae, H., Toh, H., Sato, T., Hiura, H., Takahashi, S., Shirane, K., Kabayama, Y., Suyama, M., Sasaki, H., and
1091 Arima, T. (2018). Derivation of Human Trophoblast Stem Cells. Cell Stem Cell *22*, 50-63.e56.

1092 Pavlovic, B.J., Blake, L.E., Roux, J., Chavarria, C., and Gilad, Y. (2018). A Comparative Assessment of
1093 Human and Chimpanzee iPSC-derived Cardiomyocytes with Primary Heart Tissues. Sci Rep *8*, 15312.

1094 Petropoulos, S., Edsgard, D., Reinius, B., Deng, Q.L., Panula, S.P., Codeluppi, S., Reyes, A.P., Linnarsson,
1095 S., Sandberg, R., and Lanner, F. (2016). Single-Cell RNA-Seq Reveals Lineage and X Chromosome
1096 Dynamics in Human Preimplantation Embryos. Cell *165*, 1012-1026.

1097 Pijuan-Sala, B., Griffiths, J.A., Guibentif, C., Hiscock, T.W., Jawaid, W., Calero-Nieto, F.J., Mulas, C., Ibarra-
1098 Soria, X., Tyser, R.C.V., Ho, D.L.L.*, et al.* (2019). A single-cell molecular map of mouse gastrulation and
1099 early organogenesis. Nature *566*, 490-+.

1100 Pillai, V.V., Siqueira, L.G.B., Das, M., Kei, T.G., Tu, L.N., Herren, A.W., Phinney, B.S., Cheong, S.H., Hansen,
1101 P.J., and Selvaraj, V. (2019). Physiological profile of undifferentiated bovine blastocyst-derived
1102 trophoblasts. Biology Open *8*.

1103 Posfai, E., Schell, J.P., Janiszewski, A., Rovic, I., Murray, A., Bradshaw, B., Yamakawa, T., Pardon, T., El
1104 Bakkali, M., Talon, I.*, et al.* (2021). Evaluating totipotency using criteria of increasing stringency. Nat Cell
1105 Biol *23*.

1106 Prummel, K.D., Hess, C., Nieuwenhuize, S., Parker, H.J., Rogers, K.W., Kozmikova, I., Racioppi, C.,
1107 Brombacher, E.C., Czarkwiani, A., Knapp, D.*, et al.* (2019). A conserved regulatory program initiates
1108 lateral plate mesoderm emergence across chordates. Nat Commun *10*, 3857.

1109 Rivera-Perez, J.A., and Magnuson, T. (2005). Primitive streak formation in mice is preceded by localized
1110 activation of Brachyury and Wnt3. Dev Biol *288*, 363-371.

1111 Rivron, N.C., Frias-Aldeguer, J., Vrij, E.J., Boisset, J.C., Korving, J., Vivie, J., Truckenmuller, R.K., van
1112 Oudenaarden, A., van Blitterswijk, C.A., and Geijsen, N. (2018). Blastocyst-like structures generated
1113 solely from stem cells. Nature *557*, 106-111.

1114 Rossant, J. (2015). Mouse and human blastocyst-derived stem cells: vive les differences. Development
1115 *142*, 9-12.

1116 Rossant, J., and Tam, P.P.L. (2017). New Insights into Early Human Development: Lessons for Stem Cell
1117 Derivation and Differentiation. Cell Stem Cell *20*, 18-28.

1118 Rossant, J., and Tam, P.P.L. (2021). Opportunities and challenges with stem cell-based embryo models.
1119 Stem Cell Rep *16*, 1031-1038.

1120 Rostovskaya, M., Andrews, S., Reik, W., and Rugg-Gunn, P.J. (2022). Amniogenesis occurs in two
1121 independent waves in primates. Cell Stem Cell *29*, 744-759.e746.

1122 Sasaki, K., Nakamura, T., Okamoto, K., Yabuta, Y., Iwatani, C., Tsuchiya, H., Seita, Y., Nakamura, S.,
1123 Shiraki, N., Takakuwa, T.*, et al.* (2016). The Germ Cell Fate of Cynomolgus Monkeys Is Specified in the
1124 Nascent Amnion. Dev Cell *39*, 169-185.

1125 Sasaki, K., Yokobayashi, S., Nakamura, T., Okamoto, I., Yabuta, Y., Kurimoto, K., Ohta, H., Moritoki, Y.,
1126 Iwatani, C., Tsuchiya, H.*, et al.* (2015). Robust In Vitro Induction of Human Germ Cell Fate from
1127 Pluripotent Stem Cells. Cell Stem Cell *17*, 178-194.

1128 Satija, R., Farrell, J.A., Gennert, D., Schier, A.F., and Regev, A. (2015). Spatial reconstruction of single-cell
1129 gene expression data. Nat Biotechnol *33*, 495-502.

1130 Schoenwolf, G., Bleyl, S., Brauer, P., Francis-West, P. (2020). Larsen's Human Embryology 6th Edition
1131 (Elsevier).

1132 Shahbazi, M.N., Jedrusik, A., Vuoristo, S., Recher, G., Hupalowska, A., Bolton, V., Fogarty, N.M.E.,
1133 Campbell, A., Devito, L.G., Ilic, D.*, et al.* (2016). Self-organization of the human embryo in the absence of
1134 maternal tissues (vol 18, pg 700, 2016). Nat Cell Biol *18*.

1135 Shao, Y., Taniguchi, K., Gurdziel, K., Townshend, R.F., Xue, X., Yong, K.M.A., Sang, J., Spence, J.R.,
1136 Gumucio, D.L., and Fu, J. (2017a). Self-organized amniogenesis by human pluripotent stem cells in a
1137 biomimetic implantation-like niche. Nat Mater *16*, 419-425.

1138 Shao, Y., Taniguchi, K., Townshend, R.F., Miki, T., Gumucio, D.L., and Fu, J. (2017b). A pluripotent stem
1139 cell-based model for post-implantation human amniotic sac development. Nat Commun *8*, 208.

1140 Simunovic, M., Metzger, J.J., Etoc, F., Yoney, A., Ruzo, A., Martyn, I., Croft, G., You, D.S., Brivanlou, A.H.,
1141 and Siggia, E.D. (2019). A 3D model of a human epiblast reveals BMP4-driven symmetry breaking. Nat
1142 Cell Biol *21*, 900-910.

1143 Solnica-Krezel, L. (2020). Gastrulation: From Embryonic Pattern to Form (Academic Press).

1144 Takashima, Y., Guo, G., Loos, R., Nichols, J., Ficz, G., Krueger, F., Oxley, D., Santos, F., Clarke, J.,
1145 Mansfield, W.*, et al.* (2014). Resetting Transcription Factor Control Circuitry toward Ground-State
1146 Pluripotency in Human. Cell *158*, 1254-1269.

1147 Tyser, R.C.V., Mahammadov, E., Nakanoh, S., Vallier, L., Scialdone, A., and Srinivas, S. (2021). Single-cell
1148 transcriptomic characterization of a gastrulating human embryo. Nature *600*, 285-289.

1149 Wang, R.N., Green, J., Wang, Z., Deng, Y., Qiao, M., Peabody, M., Zhang, Q., Ye, J., Yan, Z., Denduluri, S.*,*
1150 *et al.* (2014). Bone Morphogenetic Protein (BMP) signaling in development and human diseases. Genes
1151 Dis *1*, 87-105.

1152 Wang, X., Xiang, Y., Yu, Y., Wang, R., Zhang, Y., Xu, Q., Sun, H., Zhao, Z.-A., Jiang, X., Wang, X.*, et al.*
1153 (2021). Formative pluripotent stem cells show features of epiblast cells poised for gastrulation. Cell
1154 Research *31*, 526-541.

1155 Warmflash, A., Sorre, B., Etoc, F., Siggia, E.D., and Brivanlou, A.H. (2014). A method to recapitulate early
1156 embryonic spatial patterning in human embryonic stem cells. Nat Methods *11*, 847-854.

1157 Wolf, F.A., Angerer, P., and Theis, F.J. (2018). SCANPY: large-scale single-cell gene expression data
1158 analysis. Genome Biol *19*, 15.

1159 Xiang, L., Yin, Y., Zheng, Y., Ma, Y., Li, Y., Zhao, Z., Guo, J., Ai, Z., Niu, Y., Duan, K.*, et al.* (2020). A
1160 developmental landscape of 3D-cultured human pre-gastrulation embryos. Nature *577*, 537-542.

1161 Xu, R.H., Chen, X., Li, D.S., Li, R., Addicks, G.C., Glennon, C., Zwaka, T.P., and Thomson, J.A. (2002). BMP4
1162 initiates human embryonic stem cell differentiation to trophoblast. Nat Biotechnol *20*, 1261-1264.

1163 Xue, X., Sun, Y., Resto-Irizarry, A.M., Yuan, Y., Aw Yong, K.M., Zheng, Y., Weng, S., Shao, Y., Chai, Y.,
1164 Studer, L.*, et al.* (2018). Mechanics-guided embryonic patterning of neuroectoderm tissue from human
1165 pluripotent stem cells. Nat Mater *17*, 633-641.

1166 Yanagida, A., Spindlow, D., Nichols, J., Dattani, A., Smith, A., and Guo, G. (2021). Naive stem cell
1167 blastocyst model captures human embryo lineage segregation. Cell Stem Cell *28*, 1016-1022 e1014.

1168 Yang, R., Goedel, A., Kang, Y., Si, C., Chu, C., Zheng, Y., Chen, Z., Gruber, P.J., Xiao, Y., Zhou, C.*, et al.*
1169 (2021). Amnion signals are essential for mesoderm formation in primates. Nat Commun *12*, 5126.

1170 Yu, L., Wei, Y., Duan, J., Schmitz, D.A., Sakurai, M., Wang, L., Wang, K., Zhao, S., Hon, G.C., and Wu, J.
1171 (2021). Blastocyst-like structures generated from human pluripotent stem cells. Nature *591*, 620-626.

1172 Zhai, L., Chaturvedi, D., and Cumberledge, S. (2004). Drosophila Wnt-1 Undergoes a Hydrophobic
1173 Modification and Is Targeted to Lipid Rafts, a Process That Requires Porcupine. Journal of Biological
1174 Chemistry *279*, 33220-33227.

1175 Zhao, C., Reyes, A.P., Schell, J.P., Weltner, J., Ortega, N., Zheng, Y., Björklund, Å.K., Rossant, J., Fu, J.,
1176 Petropoulos, S.*, et al.* (2021). Reprogrammed iBlastoids contain amnion-like cells but not
1177 trophectoderm. bioRxiv, 2021.2005.2007.442980.

1178 Zhao, G.Q. (2003). Consequences of knocking out BMP signaling in the mouse. Genesis *35*, 43-56.

1179 Zheng, Y., Shao, Y., and Fu, J. (2021). A microfluidics-based stem cell model of early post-implantation
1180 human development. Nat Protoc *16*, 309-326.

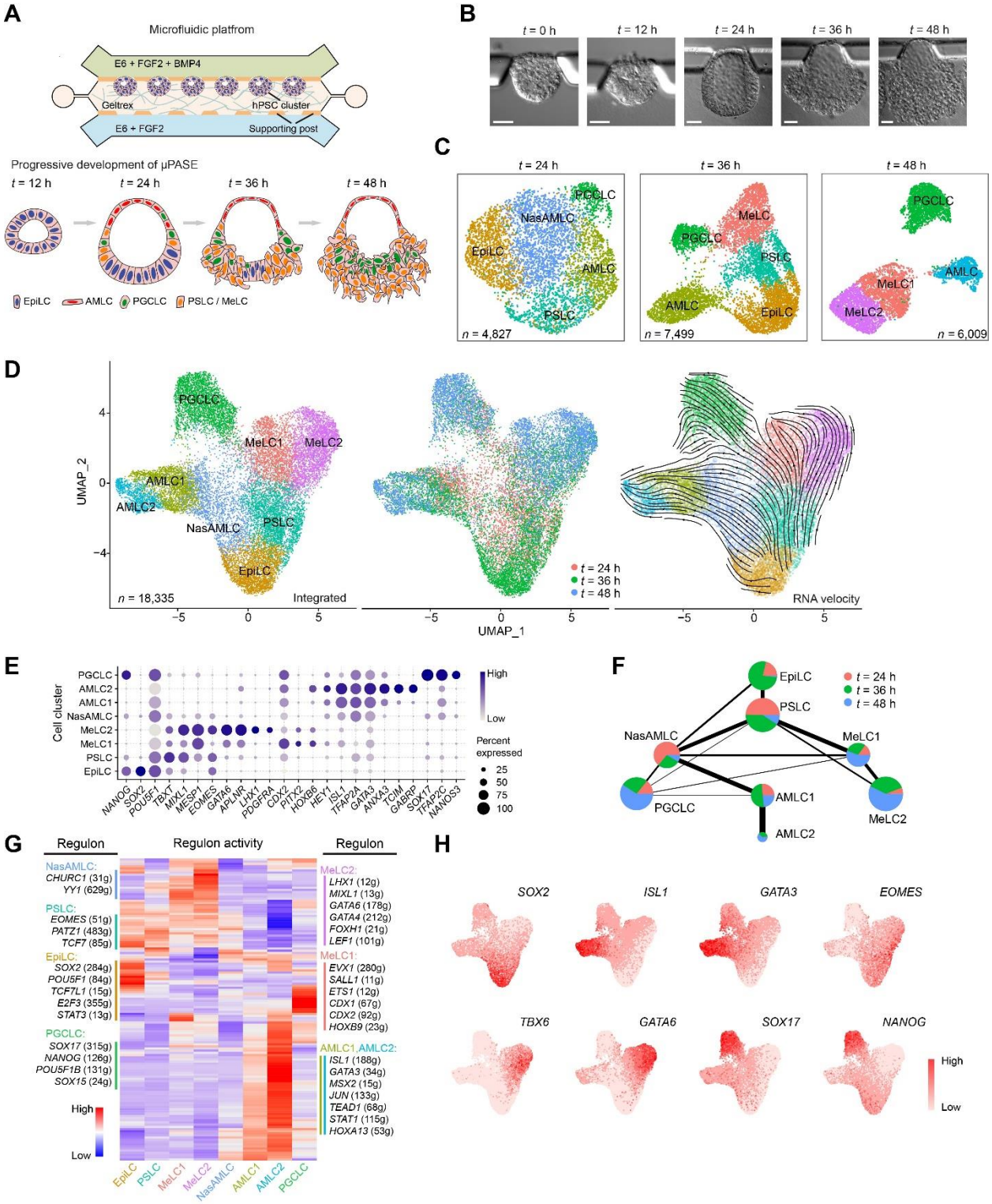
1181 Zheng, Y., Xue, X., Resto-Irizarry, A.M., Li, Z., Shao, Y., Zheng, Y., Zhao, G., and Fu, J. (2019a). Dorsal-
1182 ventral patterned neural cyst from human pluripotent stem cells in a neurogenic niche. Sci Adv *5*,
1183 eaax5933.

1184 Zheng, Y., Xue, X., Shao, Y., Wang, S., Esfahani, S.N., Li, Z., Muncie, J.M., Lakins, J.N., Weaver, V.M.,
1185 Gumucio, D.L.*, et al.* (2019b). Controlled modelling of human epiblast and amnion development using
1186 stem cells. Nature *573*, 421-425.

1187

FIGURES

Figure 1



1191 **Figure 1. Single-cell transcriptomic profiling of μPASE development.**

1192 **(A)** Development of μPASEs. Single hPSCs were guided to form uniform-sized clusters at
1193 prescribed locations in the microfluidic device. Asymmetric stimulation with exogeneous BMP4
1194 from t = 0 – 48 h led to progressive development of μPASEs. E6: Essential 6. See **Methods** for
1195 the μPASE protocol.

1196 **(B)** Bright-field images showing progressive development of μPASEs over time, including
1197 thinning and flattening of the incipient amniotic ectoderm and thickening of the incipient
1198 mesoderm cells before their dissemination from the μPASE structure.

1199 **(C)** Dimension reduction presentation *via* UMAP and cell identity annotations of single-cell
1200 transcriptome datasets obtained for μPASEs at indicated time points. n indicates cell numbers
1201 analyzed for each time point.

1202 **(D)** UMAP of integrated single-cell transcriptome datasets of μPASEs from t = 24, 36 and 48 h
1203 (shown in **(C)**), color-coded according to cell identity annotations (*left*) or time points (*middle*).
1204 RNA velocity vectors were projected onto the UMAP-based embeddings (*right*). n indicates the
1205 total number of cells combined from the three time points.

1206 **(E)** Dot plot showing expression of key marker genes across the cell clusters as indicated. The
1207 sizes and colors of dots indicate the proportion of cells expressing the corresponding genes and
1208 their averaged scaled values of log-transformed expression, respectively.

1209 **(F)** Partition-based graph abstraction (PAGA) analysis of the integrated single-cell transcriptome
1210 dataset shown in **(D)**. The thickness of lines connecting pairs of cell clusters indicates the degree
1211 of correlation between the cell cluster pairs. Lines with a correlation weight less than 0.05 are not
1212 shown. Pie charts for each cell cluster show percentages of indicated cell types from the three
1213 time points. Pie chart size is proportional to the total number of indicated cell types. See **Table**
1214 **S1**.

1215 **(G)** Heatmap of regulon activities calculated from gene regulatory network interference. Selected
1216 master regulators of different cell clusters are depicted as indicated. See **Table S2 and**
1217 **Supplemental HTML Document 1**.

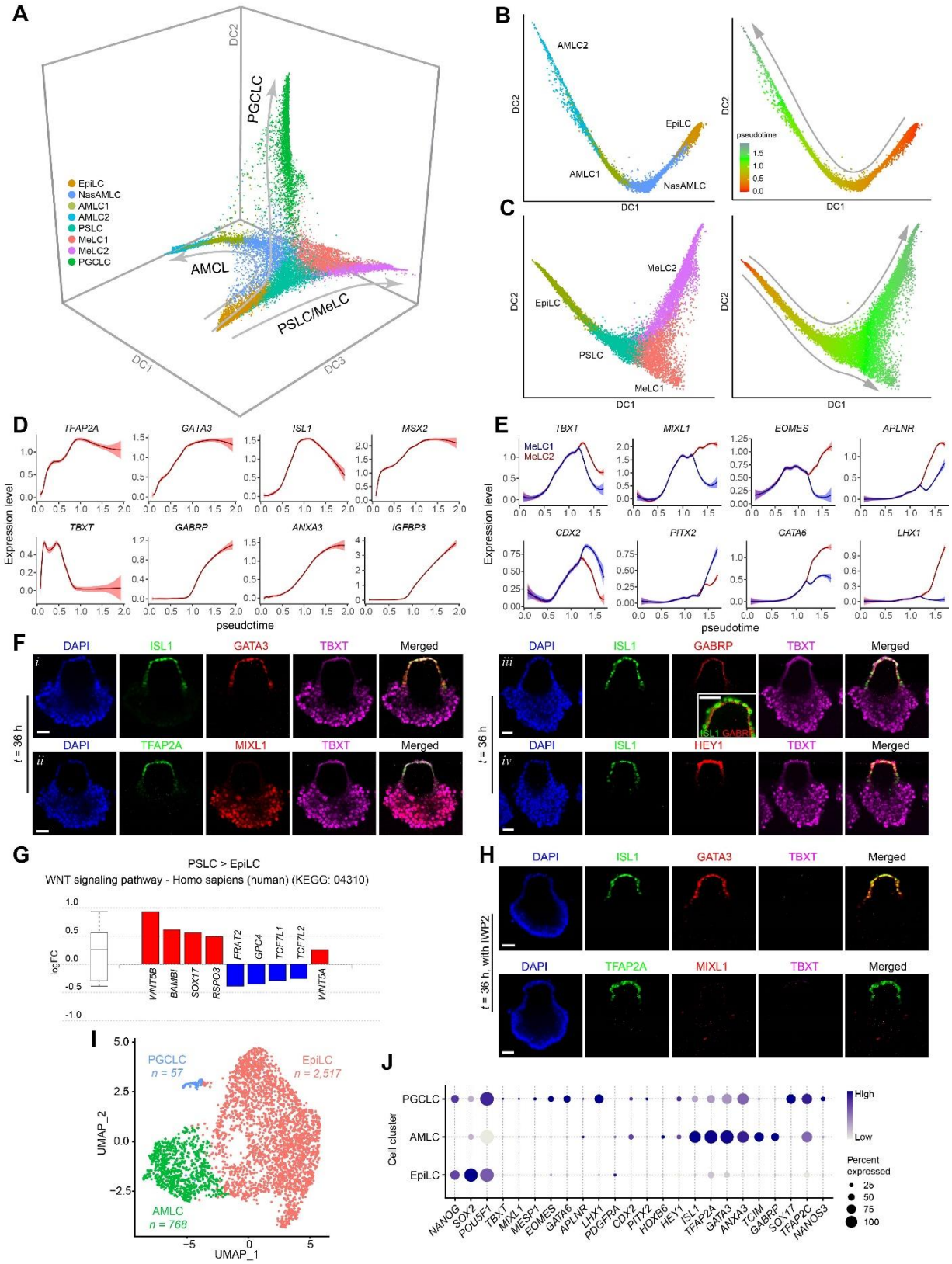
1218 **(H)** Gene set activity of selected regulons overlaid on the integrated UMAP plot from **(D)**.

1219

1220 EpiLC: epiblast-like cell; PSLC: primitive streak-like cell; MeLC1/2: mesoderm-like cell 1/2;
1221 AMLC1/2: amniotic ectoderm-like cell 1/2; NasAMLC: nascent amniotic ectoderm-like cell;

1222 PGCLC: primordial germ cell-like cell. In **(B)**, experiments were repeated more than twenty
1223 times with similar results. Scale bars, 50 μm.
1224

Figure 2



1225

1226 **Figure 2. Trajectory inference and gene expression dynamics during μPASE development.**
1227 **(A)** Three-dimensional diffusion map based on embeddings of the UMAP plot in **Figure 1D**,
1228 showing developmental trajectories of AMLC, PSLC / MeLC and PGCLC lineages. The UMAP
1229 plot is color-coded according to cell identity annotations. See **Video S1**
1230 **(B)** *left:* Trajectory inference (diffusion map) of AMLC lineage (EpiLC, NasAMLC, AMLC1
1231 and AMLC2). *right:* Pseudotime analysis (color-coded) based on the AMLC lineage diffusion
1232 map.
1233 **(C)** *left:* Trajectory inferences (diffusion map) of PSLC / MeLC lineage (EpiLC, PSLC and
1234 MeLC1 or MeLC2). *right:* Pseudotime analysis (color-coded) based on the PSLC / MeLC
1235 lineage diffusion map.
1236 **(D)** Expression dynamics (pseudotime) of selected genes during AMLC lineage development.
1237 Level of confidence (0.95) is indicated by band width.
1238 **(E)** Expression dynamics (pseudotime) of selected genes during PSLC / MeLC lineage
1239 development. Level of confidence (0.95) is indicated by band width.
1240 **(F)** Representative confocal micrographs showing μPASEs at t = 36 h stained for ISL1, GATA3
1241 and TBXT (*i*); TFAP2A, MIXL1 and TBXT (*ii*); ISL1, GABRP and TBXT (*iii*, with zoom-in
1242 view showing apical expression of GABRP); ISL1, HEY1, and TBXT (*iv*).
1243 **(G)** Differentially expressed genes (DEGs) related to WNT signaling pathway (KEGG: 04310)
1244 in PSLC compared to EpiLC.
1245 **(H)** Representative confocal micrographs showing μPASEs at t = 36 h stained for ISL1, GATA3
1246 and TBXT (*top*); TFAP2A, MIXL1 and TBXT (*bottom*) with IWP2 supplemented into the basal
1247 medium from t = 0 h.
1248 **(I)** UMAP and cell identity annotations of single-cell transcriptome data obtained for μPASEs at
1249 t = 48 h with IWP2 supplemented into the basal medium from t = 0 h. n indicates cell numbers of
1250 indicated cell types.
1251 **(J)** Dot plot showing expression of key marker genes across the cell clusters as indicated. The
1252 sizes and colors of dots indicate the proportion of cells expressing the corresponding genes and
1253 their averaged scaled values of log-transformed expression, respectively.
1254
1255 In **F** and **H**, experiments were repeated four times with similar results. Nuclei were
1256 counterstained with DAPI. Scale bars, 50 μm.

Figure 3

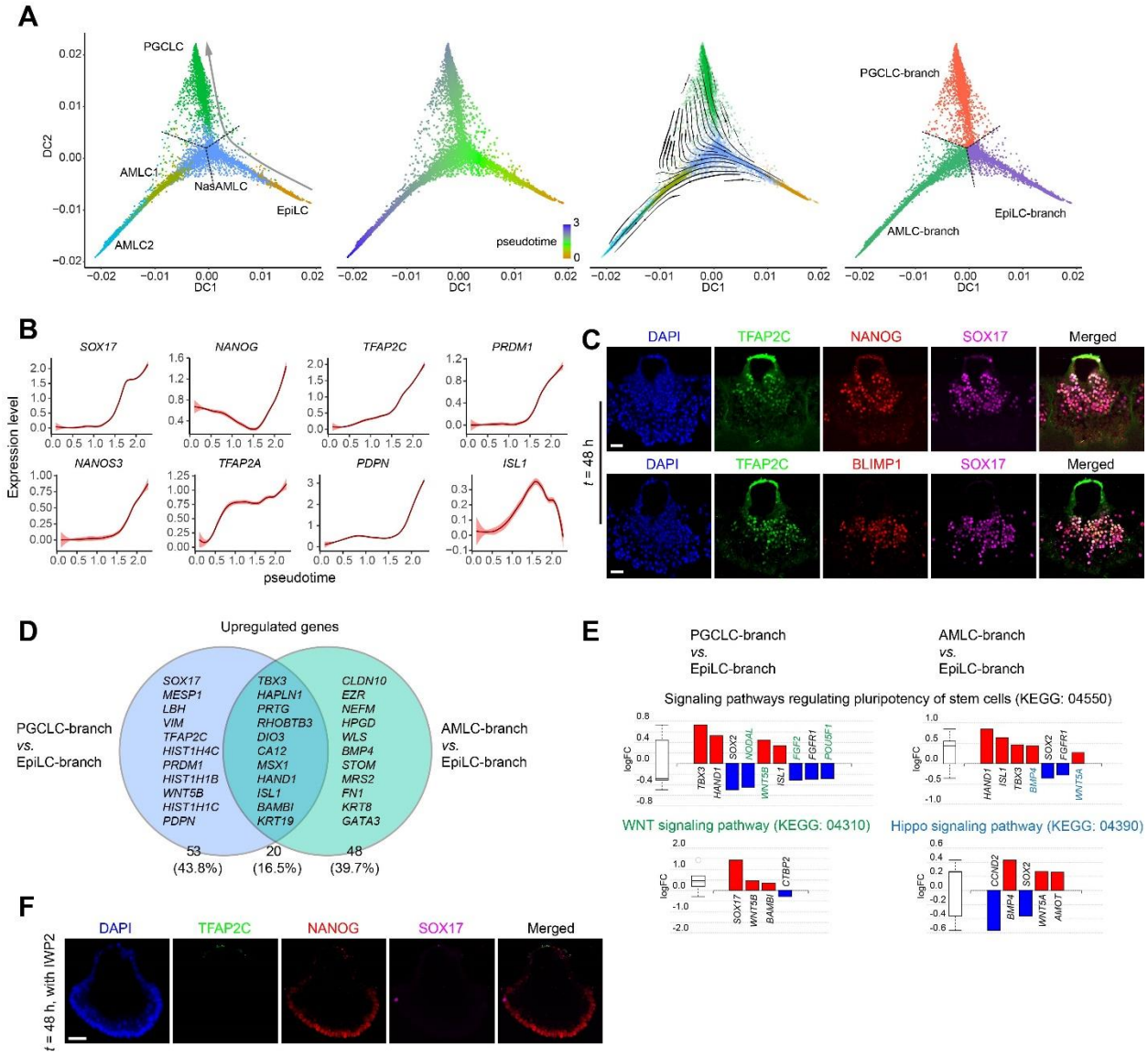
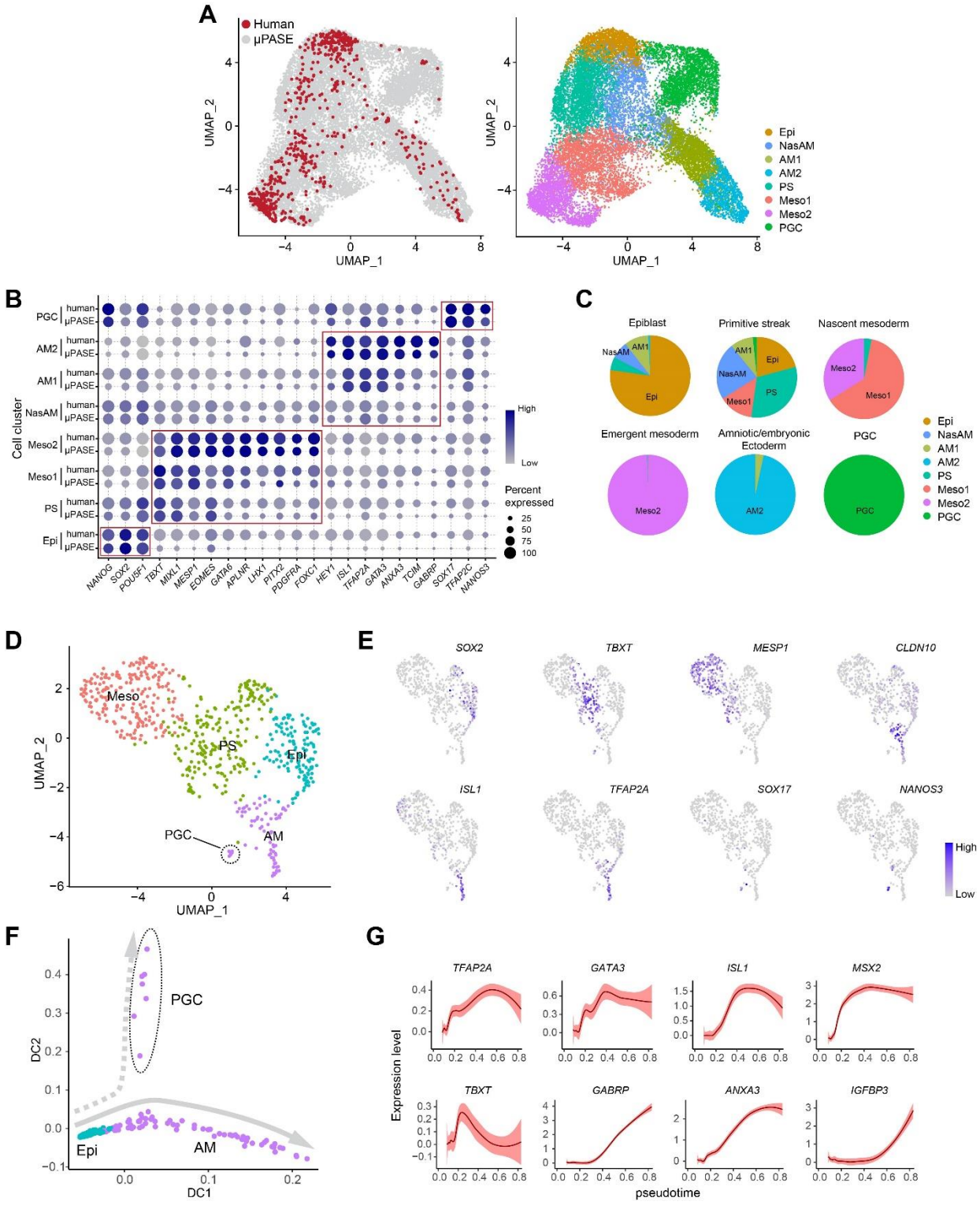


Figure 3. PGCLC specification during µPASE development.

(A) Trajectory inference of PGCLC lineage. *left:* Diffusion map using EpiLC, NasAMLC, AMLC1, AMLC2, and PGCLC clusters from the UMAP plot in **Figure 1D**. Dotted lines show the branching point and branches identified by K-Branches algorithm. *middle left*: Pseudotime analysis based on the diffusion map. *middle right:* RNA velocity vectors overlaid on the diffusion map. *right:* Branches and the branching point identified by K-Branches algorithm. Note that NasAMLC cluster is separated into three branches, which after merging with EpiLC,

1265 AMLC1/2 and PGCLC, respectively, are annotated as EpiLC-branch NasAMLC, AMLC-branch
1266 NasAMLC and PGCLC-branch NasAMLC, respectively.
1267 **(B)** Expression dynamics (pseudotime) of selected genes during PGCLC lineage development.
1268 Level of confidence (0.95) is indicated by band width.
1269 **(C)** Representative confocal micrographs showing μPASEs at t = 48 h stained for TFAP2C,
1270 NANOG and SOX17 (*top*) and TFAP2C, BLIMP1 and SOX17 (*bottom*).
1271 **(D)** Venn diagram showing upregulate genes of AMLC-branch and PGCLC-branch NasAMLCs
1272 when compared to EpiLC-branch NasAMLCs. Note that only a subset of these genes is specified
1273 in the diagram.
1274 **(E)** Pathway analysis of DEGs in PGCLC-branch NasAMLCs and AMLC-branch NasAMLCs,
1275 as compared to EpiLC-branch NasAMLCs. Green and blue colors indicate genes or pathways
1276 identified only for PGCLC-branch NasAMLCs and AMLC-branch NasAMLCs, respectively.
1277 **(F)** Representative confocal micrographs showing μPASEs at t = 48 h stained for TFAP2C,
1278 NANOG and SOX17, with IWP2 supplemented into the basal medium from t = 0 h.
1279
1280 In **C** and **F**, experiments were repeated four times with similar results. Nuclei were
1281 counterstained with DAPI. Scale bars, 50 μm.
1282

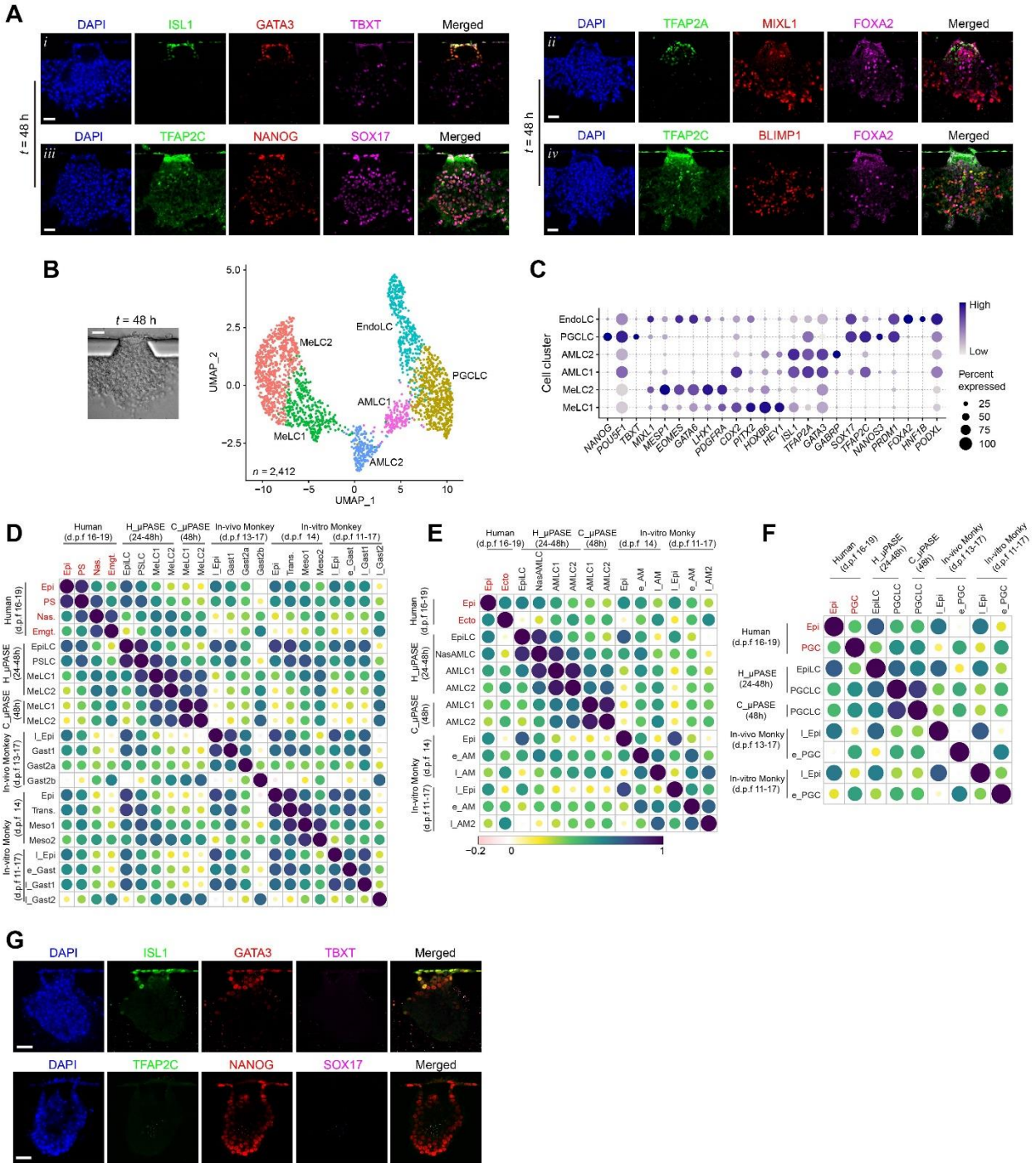
Figure 4



1283

1284 **Figure 4. Transcriptomic comparison between μPASEs and Carnegie Stage 7 human**
1285 **gastrula.**
1286 **(A)** *left*: UMAP of integrated dataset of μPASEs from t = 24, 36 and 48 h (18,335 cells; grey)
1287 and CS7 human gastrula (647 cells, excluding irrelevant cells; red). *right*: UMAP project of
1288 integrated dataset with cell identity annotations.
1289 **(B)** Dot plot comparing expression of key marker genes across different cell clusters from
1290 μPASEs and CS7 human gastrula as indicated. The sizes and colors of dots indicate the
1291 proportion of cells expressing the corresponding genes and their averaged scaled values of log-
1292 transformed expression, respectively.
1293 **(C)** Comparisons between human gastrula cell annotations in the original publication and
1294 annotations in the integrated dataset. The original annotations are indicated above the pie charts.
1295 See **Table S5**
1296 **(D)** Re-analysis of related cells from CS7 human gastrula (647 cells). Cell identity annotations
1297 are color coded as indicated. Note that primordial germ cells (PGCs) appear in the
1298 amniotic/embryonic ectoderm (AM) cluster.
1299 **(E)** Feature plots showing expression of selected lineage markers used for cell cluster
1300 annotations in **(D)**.
1301 **(F)** Trajectory inference (diffusion map) of the AM cluster (including PGCs) in **(D)**.
1302 **(G)** Expression dynamics (pseudotime) of selected genes during AM lineage development. Level
1303 of confidence (0.95) is indicated by band width.
1304
1305 Epi: Epiblast; PS: primitive steak; Meso: Mesoderm; AM: amniotic ectoderm; PGC: primordial
1306 germ cell.
1307

Figure 5


A
DAPI
ISL1
GATA3
TBXT
Merged
TFAP2A
MIXL1
FOXA2
TFAP2C
NANOG
SOX17
BLIMP1
t = 48 h
B
t = 48 h
UMAP_2
UMAP_1
EndoLC
MeLC2
PGCLC
AMLC1
MeLC1
AMLC2
n = 2,412
C
Cell cluster
EndoLC
PGCLC
AMLC2
AMLC1
MeLC2
MeLC1
High
Low
Percent expressed
25
50
75
100
D
Human (d.p.f 16-19)
H_μPASE (24-48h)
C_μPASE (48h)
In-vivo Monkey (d.p.f 13-17)
In-vitro Monkey (d.p.f 14)
In-vitro Monkey (d.p.f 11-17)
E
F
G

1309 **Figure 5. Transcriptomic coordination of primate early post-implantation development.**
1310 **(A)** Representative confocal micrographs showing chimpanzee μPASEs at *t* = 48 h stained for
1311 ISL1, GATA3 and TBXT (*i*); TFAP2A, MIXL1 and FOXA2 (*ii*); TFAP2C, NANOG, SOX17
1312 (*iii*); TFAP2C, BLIMP1 and FOXA2 (*iv*).
1313 **(B)** *Left*: Bright-field image showing a chimpanzee μPASE at *t* = 48 h. *Right*: UMAP plot of
1314 single-cell transcriptome dataset from chimpanzee μPASEs at *t* = 48 h. Cell identity annotations
1315 are color coded as indicated. *n* indicates the total cell number. EndoLC: Endoderm-like cell.
1316 **(C)** Dot plot showing expression of key marker genes across the cell clusters as indicated. The
1317 sizes and colors of dots indicate the proportion of cells expressing the corresponding genes and
1318 their averaged scaled values of log-transformed expression, respectively.
1319 **(D)** Heat map of correlation matrix for primitive streak / mesoderm-related lineages including
1320 those reported by others. Correlation coefficients between indicated cell types are calculated
1321 based on mesoderm ontogenic genes identified from cynomolgus embryo transcriptome data
1322 (131 in common). See **Table S6**.
1323 **(E)** Heat map of correlation matrix for amnion-related lineages including those reported by
1324 others. Correlation coefficients between indicated cell types are calculated based on amnion
1325 ontogenic genes identified from cynomolgus embryo transcriptome data (142 in common). See
1326 **Table S6**.
1327 **(F)** Heat map of correlation matrix for PGC-related lineages including those reported by others.
1328 Correlation coefficients between indicated cell types are calculated based on PGC ontogenic
1329 genes identified for cynomolgus embryo transcriptome data (194 in common). See **Table S6**.
1330 **(G)** Representative confocal micrographs showing chimpanzee μPASEs at *t* = 48 h with IWP2
1331 supplemented into the basal medium from *t* = 0 h, stained for ISL1, GATA3 and TBXT (*top*);
1332 TFAP2C, NANOG and SOX17 (*bottom*),
1333
1334
1335 In **D-F**, correlation coefficients are calculated using average ontogenic gene expression of single
1336 cells. Original cell annotations in published datasets are used. Log_2(reads per million + 1) and
1337 Log_2 (transcripts per million + 1) are used for transcriptome datasets generated using 10×
1338 Chrome and Smart-seq2, respectively.
1339

1340 Human: human gastrula from Ref.(Tyser et al., 2021). Epi: epiblast; PS: primitive streak; Nas.:
1341 Nascent mesoderm; Emgt.: emergent mesoderm; Ecto.: amniotic/embryonic ectoderm; PGC:
1342 primordial germ cell. H_μPASE and C_μPASE: μPASEs generated from human and
1343 chimpanzee cells, respectively. *In vivo* monkey: *in vivo* cynomolgus embryo from
1344 Ref.(Nakamura et al., 2016; Sasaki et al., 2016). l_EPI: late epiblast; Gast: gastrulation; e_PGC:
1345 early primordial germ cell. *In vitro* monkey: *in vitro* cultured cynomolgus embryo from Ref.(Ma
1346 et al., 2019; Yang et al., 2021). Trans: Transition; Meso: mesoderm; l_EPI: late epiblast; e_Gast:
1347 early gastrulation; l_Gast: late gastrulation; e_AM: early amniotic ectoderm; l_AM: late
1348 amniotic ectoderm.
1349
1350 Expression of ontogenic genes is included in **Table S6**. In **A** and **G**, experiments were repeated
1351 four times with similar results. Nuclei were counterstained with DAPI. Scale bars, 50 μm.

Figure 6

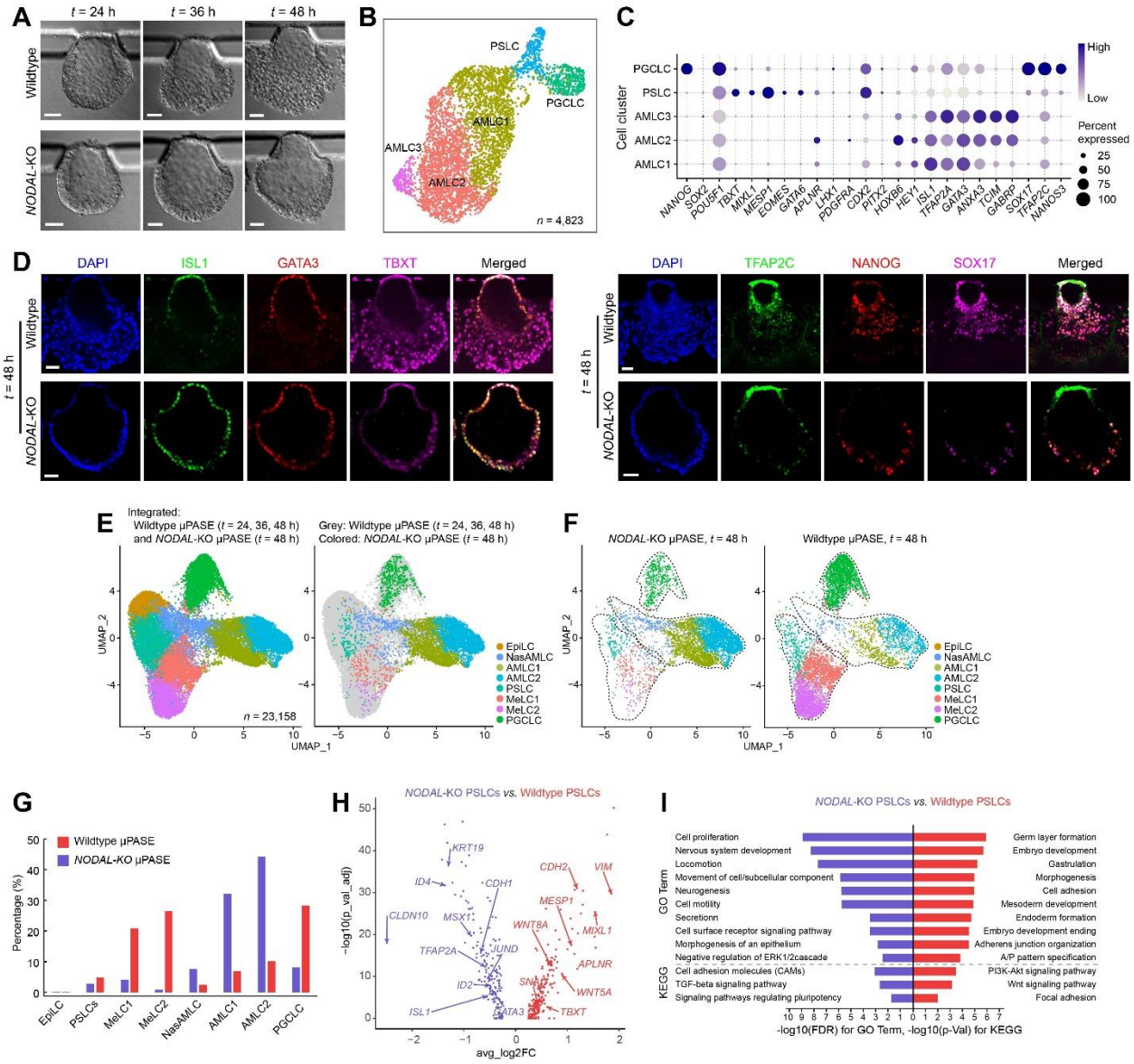


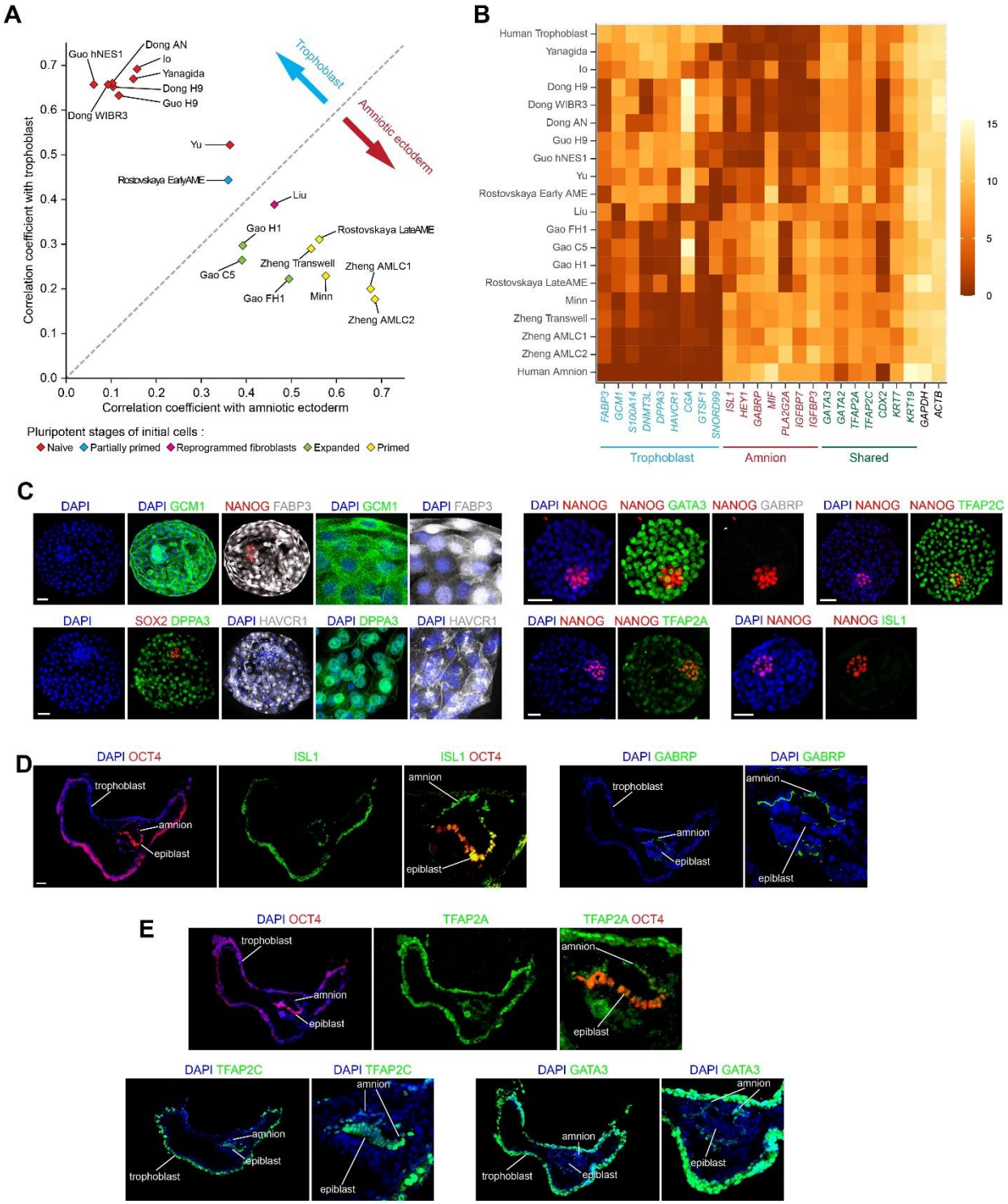
Figure 6. NODAL is essential for mesoderm development in μPASEs.

(A) Representative bight-field images showing progressive development of μPASEs generated from wildtype (*top*) and *NODAL*-KO (*bottom*) hPSCs.

(B) UMAP plot of single-cell transcriptome data from *NODAL*-KO μPASEs at $t = 48$ h, with cell identity annotations color coded. *n* indicates the total cell number.

1358 **(C)** Dot plot showing expression of key marker genes across the cell clusters in *NODAL*-KO
1359 μPASEs as indicated. The sizes and colors of dots indicate the proportion of cells expressing the
1360 corresponding genes and their averaged scaled values of log-transformed expression,
1361 respectively.
1362 **(D)** Representative confocal micrographs showing μPASEs generated from wildtype and
1363 *NODAL*-KO hPSCs at t = 48 h, stained for ISL1, GATA3 and TBXT (*left*) and TFAP2C,
1364 NANOG and SOX17 (*right*).
1365 **(E)** *left:* Integrated UMAP plot of wildtype μPASEs at t = 24, 36, 48 h and NODAL-KO
1366 μPASEs at t = 48 h, color-coded according to cell identity annotations. *right:* data from wildtype
1367 and *NODAL*-KO μPASEs are shown in different colors as indicated.
1368 **(F)** UMAP plots of *NODAL*-KO (*left*) and wildtype (*right*) μPASEs at t = 48 h, with data
1369 isolated from **E**. Doted lines contour cell clusters corresponding to AMLC, PSLC / MeLC and
1370 PGCLC lineages.
1371 **(G)** Percentages of indicted cell types in wildtype (red) and *NODAL*-KO (blue) μPASEs.
1372 **(H)** Volcano plots showing DEGs between PSLCs from wildtype and *NODAL*-KO μPASEs,
1373 with selected genes labelled.
1374 **(I)** Enriched GO categories and KEGG pathways among DEGs between wildtype and *NODAL*-
1375 KO PSLCs.
1376
1377 In **D**, experiments were repeated three times with similar results. Nuclei were counterstained
1378 with DAPI. Scale bars, 50 μm.
1379

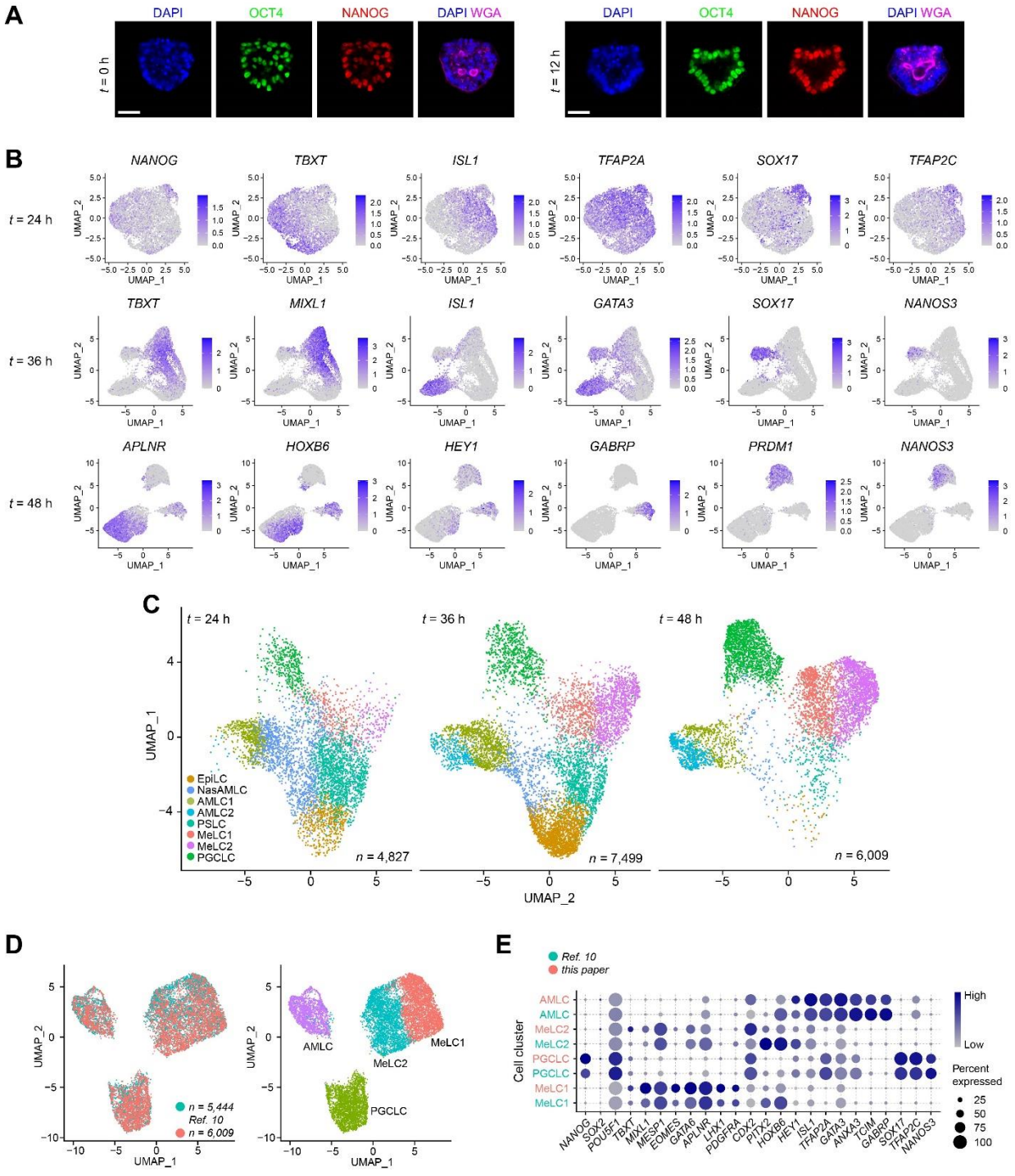
Figure 7



1380

1381 **Figure 7. Stringent criteria for identifying human trophoblast and amniotic ectoderm.**
1382 **(A)** Correlation coefficients of *in vitro* derived cells with human trophoblast and amniotic
1383 ectoderm, computed using Trophoblast_Amnion ontogenic genes. See **Table S8**.
1384 **(B)** Heatmap showing expression levels of selected genes related to human trophoblast and
1385 amniotic ectoderm. Identified human trophoblast and amniotic ectoderm markers are highlighted
1386 in blue and red, respectively. Markers shared between human trophoblast and amniotic ectoderm
1387 are highlighted in green.
1388 **(C)** Representative confocal micrographs showing E6 human blastocysts, stained for GCM1,
1389 NANOG and FABP3; SOX2, DPPA3 and HAVCR1 (Images on the right show magnified views
1390 of the trophectoderm); NANOG, GATA3 and GABRP; NANOG and TFAP2C; NANOG and
1391 TFAP2A; and NANOG and ISL1.
1392 **(D)** Representative confocal micrographs showing *in vitro* cultured D14 cynomolgus embryos,
1393 stained for OCT4 and ISL1; and GABRP. Images on the right show magnified views of the
1394 epiblast and amniotic ectoderm.
1395 **(E)** Representative confocal micrographs showing *in vitro* cultured D14 cynomolgus embryos,
1396 stained for OCT4 and TFAP2A, TFAP2C and GATA3. Images on the right show magnified
1397 views of the epiblast and amniotic ectoderm.
1398
1399 Human trophoblast (Ref.(Blakeley et al., 2015; Petropoulos et al., 2016)), Human amnion
1400 (Ref.(Tyser et al., 2021)), Zheng Transwell (Ref.(Zheng et al., 2019b)), Gao C5, Gao H1, Gao
1401 FH1(Ref.(Gao et al., 2019)), Minn (Ref.(Minn et al., 2020)), Liu (Ref.(Liu et al., 2021)), Yu
1402 (Ref.(Yu et al., 2021)), Guo H9, Guo hNES1(Ref.(Guo et al., 2021)), Io (Ref.(Io et al., 2021)),
1403 Yanagida (Ref.(Yanagida et al., 2021)), Dong H9, Dong WIBR3, Dong AN (Ref.(Dong et al.,
1404 2020)), Rostovskaya EarlyAME, Rostovskaya LateAME (Ref.(Rostovskaya et al., 2022))
1405
1406 Correlation coefficients and gene expression heatmap were calculated based on averages of
1407 experimental repeats, if any. When the published work contains transcriptome datasets from
1408 multiple time points, datasets from the experimental endpoint are utilized.
1409
1410 In **C-E**, experiments were repeated twice with similar results. Nuclei were counterstained with
1411 DAPI. Scale bars, 50 μm.

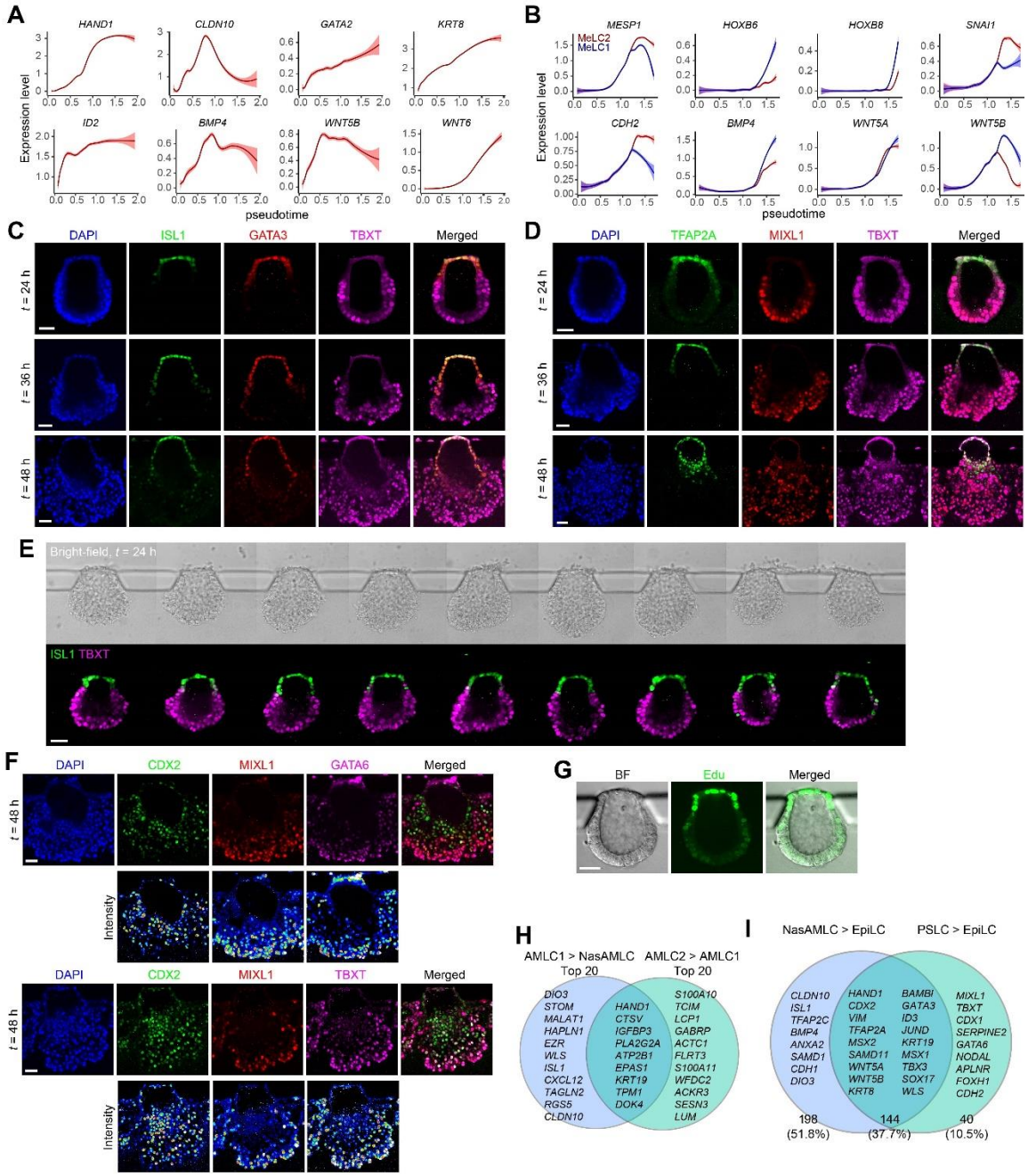
Supplemental Figure 1



1412

1413 **Supplemental Figure 1. Single-cell transcriptomic profiling of μPASE development.**
1414 **Related to Figure 1.**
1415 **(A)** Representative confocal micrographs showing μPASEs at *t* = 0 h (*left*) and *t* = 12 h (*right*)
1416 stained for OCT4 and NANOG. Plasma membrane is stained with fluorescently labelled wheat
1417 germ agglutinin (WGA).
1418 **(B)** Feature plots showing selected lineage markers used for cell identity annotations in the
1419 UMAP plots of μPASEs at *t* = 24, 36 and 48 h, respectively.
1420 **(C)** UMAP plots of μPASEs at *t* = 24, 36, 48 h, separated from the integrated UMAP plot in
1421 **Figure 1D**. *n* indicates the cell number.
1422 **(D)** Integrated UMAP plots of datasets from μPASEs at *t* = 48 h and the published dataset in
1423 Ref.(Zheng et al., 2019b). *Left:* Datasets are color-coded according to sources, with *n* indicating
1424 the cell number; *right:* Cells are color-coded according to cell identity annotations.
1425 **(E)** Dot plot showing expression of key marker genes across the cell clusters as indicated
1426 (μPASEs at *t* = 48 h from this paper (red) and Ref.(Zheng et al., 2019b) (blue)). The sizes and
1427 colors of dots indicate the proportion of cells expressing the corresponding genes and their
1428 averaged scaled values of log-transformed expression, respectively.
1429
1430 In **A**, experiments were repeated four times with similar results. Nuclei were counterstained with
1431 DAPI. Scale bars, 50 μm.

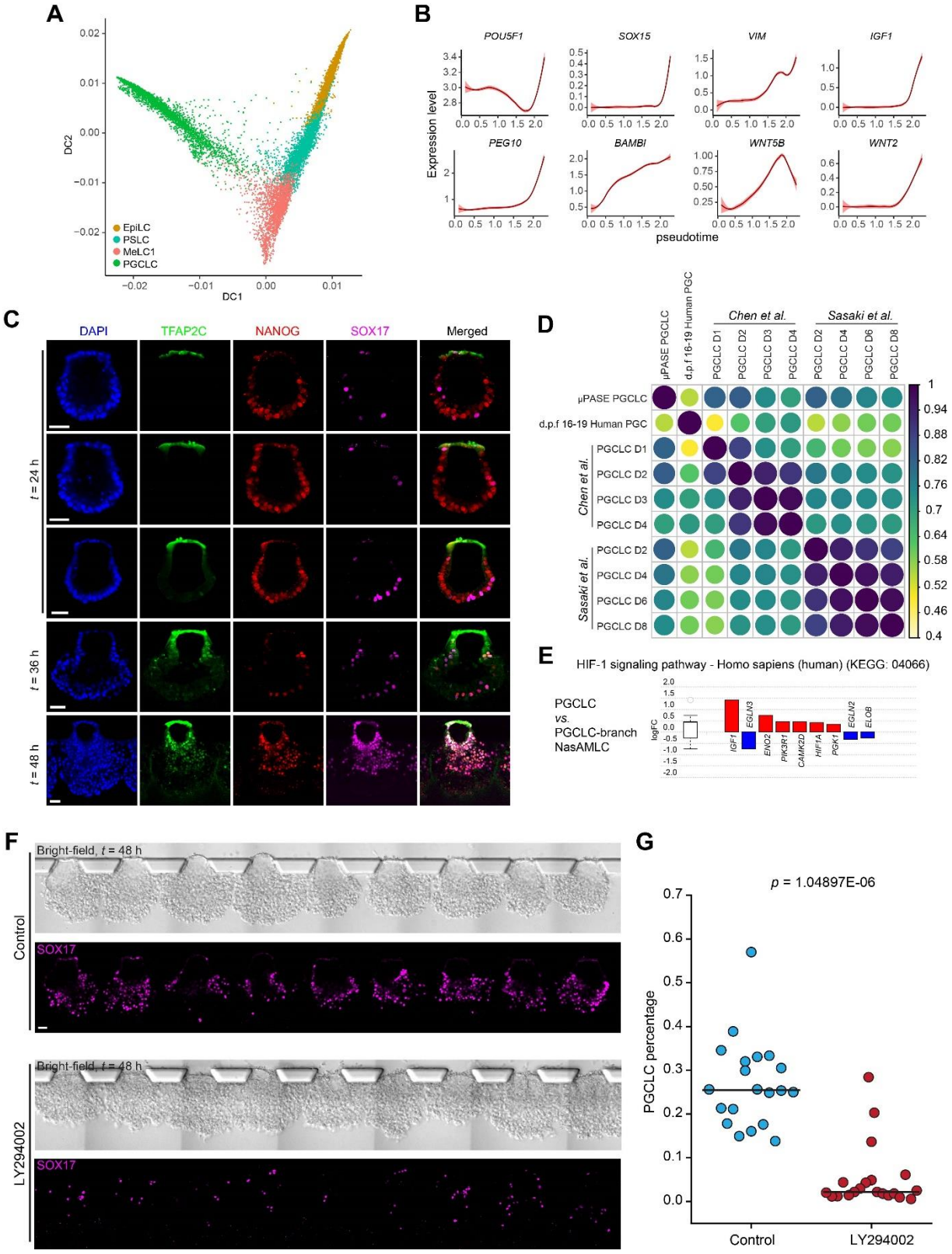
Supplemental Figure 2



1432

1433 **Supplemental Figure 2. Trajectory inference and gene expression dynamics during μPASE**
1434 **development. Related to Figure 2.**
1435 **(A)** Expression dynamics (pseudotime) of selected genes during AMLC lineage development.
1436 Level of confidence (0.95) is indicated by band width.
1437 **(B)** Expression dynamics (pseudotime) of selected genes during PSLC / MeLC lineage
1438 development. Level of confidence (0.95) is indicated by band width.
1439 **(C)** Representative confocal micrographs showing μPASEs at indicated time points stained for
1440 ISL1, GATA3 and TBXT.
1441 **(D)** Representative confocal micrographs showing μPASEs at indicated time points stained for
1442 TFAP2A, MIXL1 and TBXT.
1443 **(E)** Bright-field and immunofluorescence images showing an array of μPASEs at t = 24 h,
1444 stained for ISL1 and TBXT.
1445 **(F)** Representative confocal micrographs showing μPASEs at t = 48 h stained for CDX2, MIXL1
1446 and GATA6 (*top*); CDX2, MIXL1 and TBXT (*bottom*). Intensity maps show relative intensities
1447 of corresponding markers as indicated.
1448 **(G)** Representative confocal micrographs showing proliferating cells in μPASEs. Images were
1449 taken at t = 27 h. Cell nuclei with newly synthesized DNA within the last 3 hours were labeled
1450 using Click-iT EdU Imaging Kit (Invitrogen). Note: The low intensity of the Edu signal from
1451 cells embedded in the gel is caused by insufficient dye diffusion.
1452 **(H)** Venn diagram showing top 20 upregulated genes in AMLC1 *vs.* NasAMLC and AMLC2 *vs.*
1453 AMLC1. The full DEG lists can be found in **Table S3**.
1454 **(I)** Venn diagram showing selected upregulated genes in NasAMLC and PSLC, as compared to
1455 EpiLC. The full DEG lists can be found in **Table S3**.
1456 **(J)** DEGs related to pluripotency signaling pathway (KEGG: 04550) and TGF-β signaling
1457 pathway (KEGG: 04350) in NasAMLC and PSLC, as compared to EpiLC. Blue and green colors
1458 highlight genes identified only for NasAMLC and PSLC, respectively.
1459 **(K)** Volcano plots showing DEGs between AMLCs from μPASEs with or without IWP2 at t =
1460 48 h, with selected genes labelled. The full DEG list can be found in **Table S3.**
1461
1462 In **C-G**, experiments were repeated three times with similar results. Nuclei were counterstained
1463 with DAPI. Scale bars, 50 μm.

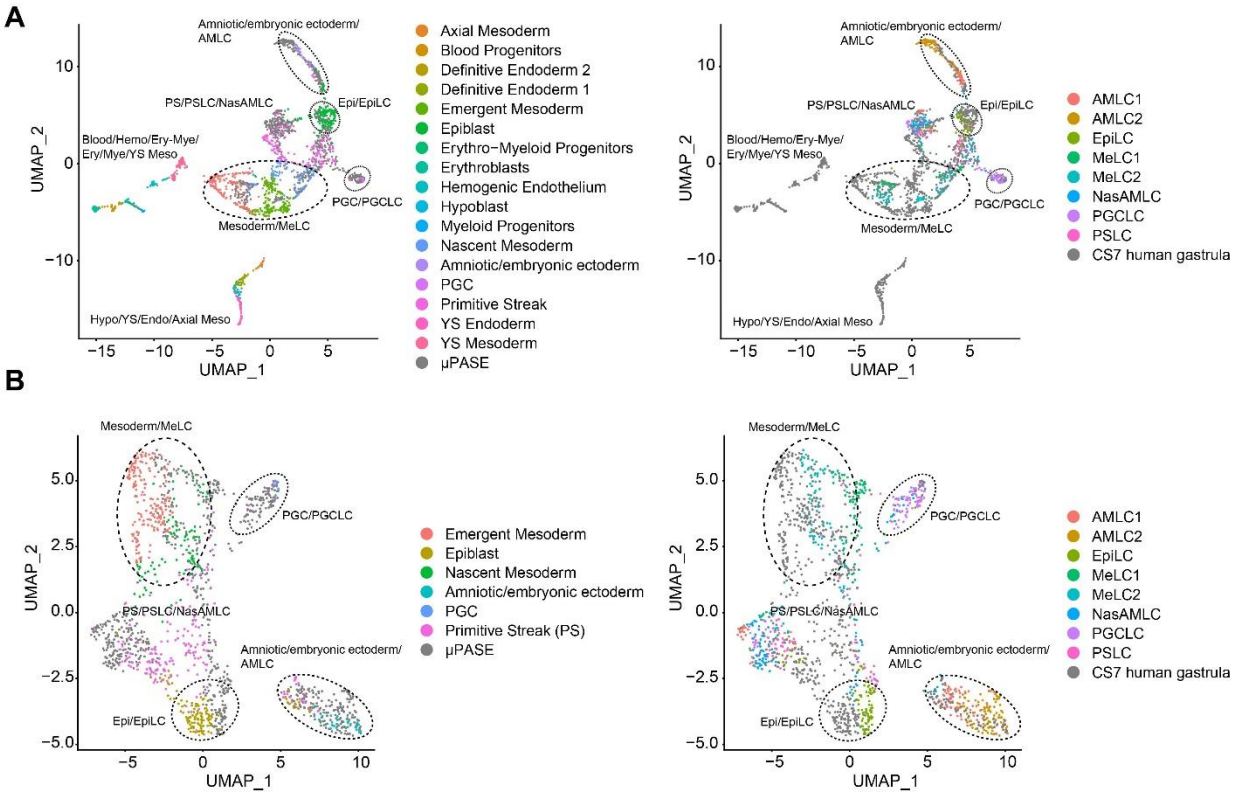
Supplemental Figure 3



1464

1465 **Supplemental Figure 3. PGCLC specification during μPASE development. Related to**
1466 **Figure 3.**
1467 **(A)** Diffusion map of EpiLC, PSLC, MeLC1 and PGCLC clusters from the UMAP plot in
1468 **Figure 1D**. PGCLC cluster is discontinuous from other clusters. K-Branches algorithm failed to
1469 identify branches or branching points with sufficient confidence.
1470 **(B)** Expression dynamics (pseudotime) of selected genes during PGCLC lineage development.
1471 Level of confidence (0.95) is indicated by band width.
1472 **(C)** Representative confocal micrographs showing μPASEs at indicated time points stained for
1473 TFAP2C, NANOG and SOX17.
1474 **(D)** Heat map of correlation matrix for PGCLCs in μPASEs, human PGCs (Ref. (Tyser et al.,
1475 2021)) and PGCLCs derived using other protocols (Ref. (Chen et al., 2019; Sasaki et al., 2015)).
1476 Correlation coefficients between indicated cell types are calculated based on PGC ontogenic
1477 genes identified for cynomolgus embryo transcriptome data.
1478 **(E)** DEGs related to HIF-1 signaling pathway (KEGG: 04066) in PGCLC as compared to
1479 PGCLC-branch NasAMLC.
1480 **(F)** Representative confocal micrographs showing arrays of μPASEs at t = 48 h, stained for
1481 SOX17; control (*top*); with LY294002 supplemented into the basal medium from t = 0 h
1482 (*bottom*).
1483 **(G)** Percentage of SOX17+ PGCLCs in μPASEs at t = 48 h under indicated conditions. n = 20
1484 μPASEs for each condition. Data were pooled from n = 2 independent experiments. Red lines
1485 represent the median.
1486
1487 In **C** and **F**, experiments were repeated three times with similar results. Nuclei were
1488 counterstained with DAPI. Scale bars, 50 μm.

Supplemental Figure 4

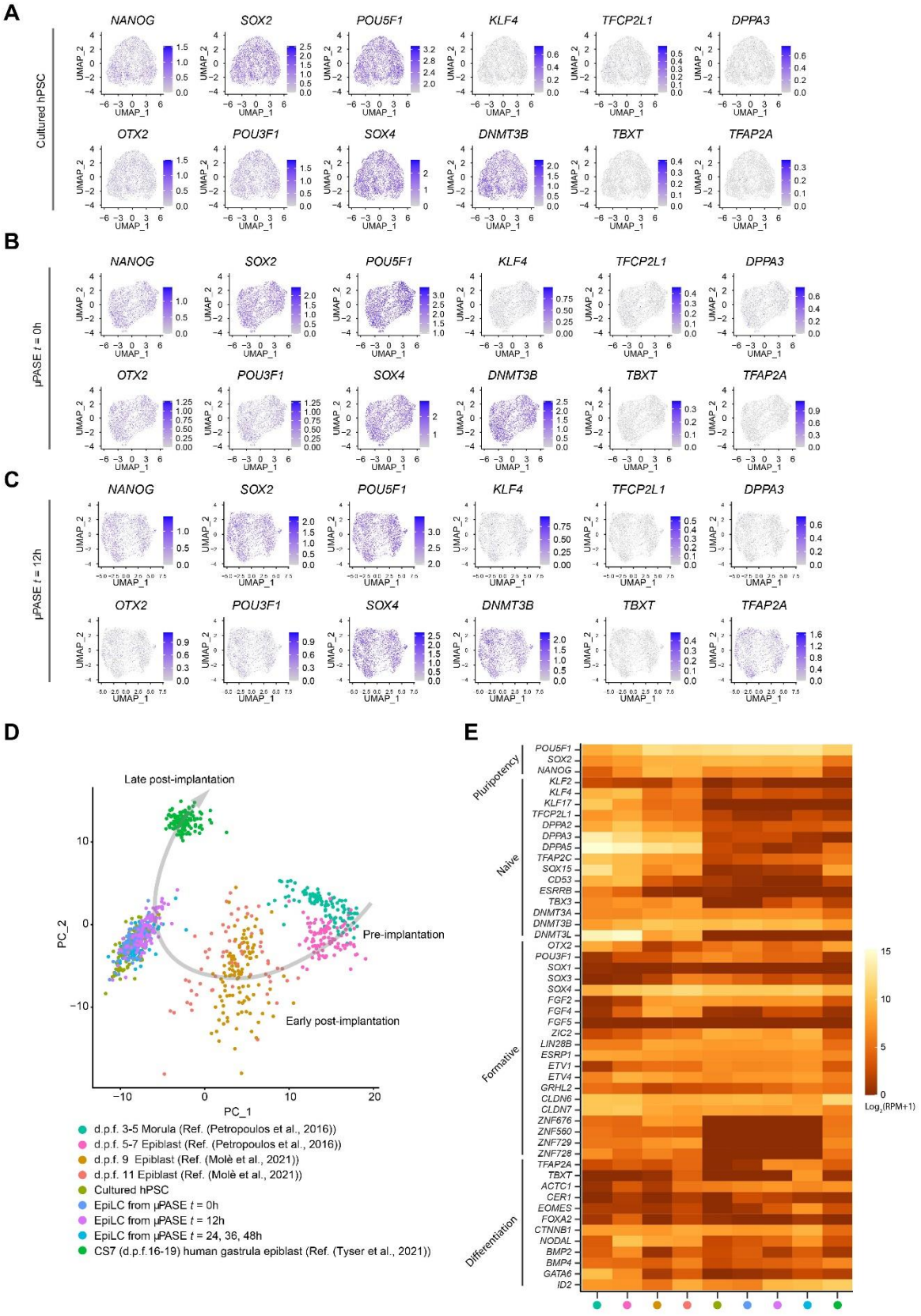


Supplemental Figure 4. scRNA-seq data integration of Carnegie Stage 7 human gastrula and downsampled μPASEs. Related to Figure 4.

(A) UMAP plots of integrated dataset of CS7 human gastrula (all 1,195 cells) and downsampled μPASEs (from **Figure 1D**, 100 cells per cluster). *Left*: color-coded according to original cell identity annotations of CS7 human gastrula; grey color indicates cells from μPASEs. *Right:* color-coded according to cell identity annotations of μPASEs as indicated in **Figure 1D**; grey color indicates cells from CS7 human gastrula.

(B) UMAP plots of integrated dataset of CS7 human gastrula (647 cells, excluding irrelevant cells) and downsampled μPASEs (from **Figure 1D**,100 cells per cluster). *Left*: color-coded according to original cell identity annotations of CS7 human gastrula; grey color indicates cells from μPASEs. *Right:* color-coded according to cell identity annotations of μPASEs as indicated in **Figure 1D**; grey color indicates cells from CS7 human gastrula.

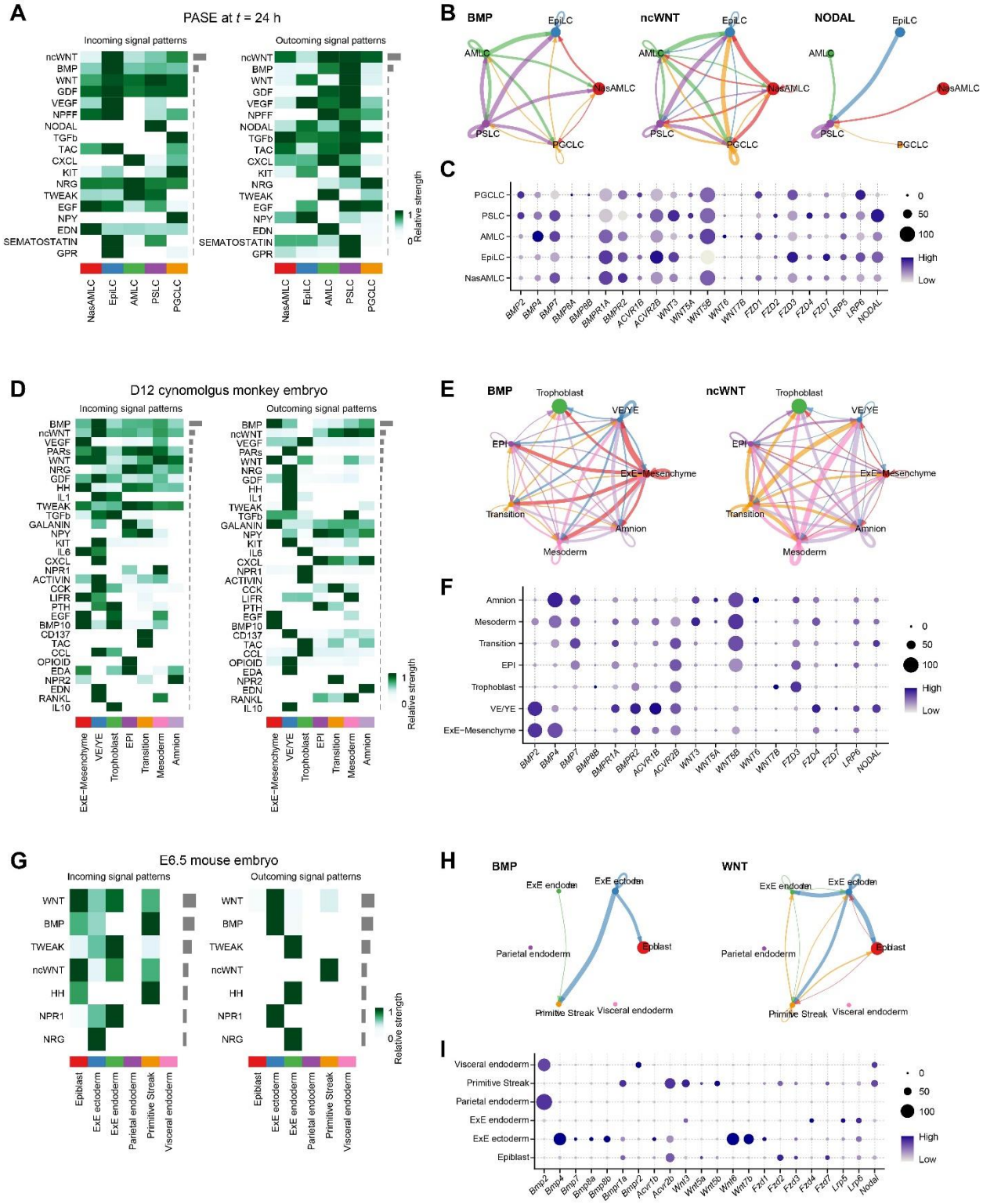
Supplemental Figure 5



1502

1503 **Supplemental Figure 5. Single-cell transcriptomic profiling of cultured hPSCs, µPASEs at *t***
1504 **= 0 h and *t* = 12 h. Related to Figure 4.**
1505 **(A)** Feature plots showing expression of selected markers in the UMAP plot of cultured hPSCs. *n*
1506 = 3782
1507 **(B)** Feature plots showing selected expression of markers in the UMAP plot of µPASEs at *t* = 0
1508 h. *n* = 2948
1509 **(C)** Feature plots showing selected expression of markers in the UMAP plot of µPASEs at *t* = 12
1510 h. *n* = 2732
1511 **(D)** Principal component analysis (PCA) plot of cultured hPSCs, µPASEs at *t* = 0 h and µPASEs
1512 at *t* = 12 h, EpiLCs from µPASEs in **Figure 1D** (at *t* = 24, 36 and 48 h), d.p.f. 9 (*n* = 108) and
1513 d.p.f. 11 (*n* = 62) epiblasts from *in vitro* cultured human embryos (Ref.(Molè et al., 2021)),
1514 epiblasts from CS7 human gastrula (*n* = 133) (Ref. (Tyser et al., 2021)), d.p.f. 5-7 epiblasts from
1515 human blastocysts and d.p.f. 3-5 human morula cells (Ref.(Petropoulos et al., 2016)). PCAs were
1516 calculated using epiblast ontogenic genes identified from cynomolgus embryos (Nakamura et al.,
1517 2016, see **Table S5**). To prevent datasets with high number of cells dominating PCA
1518 calculation, cultured hPSCs, µPASEs at *t* = 0 h, µPASEs at *t* = 12 h, EpiLCs from µPASEs (at *t*
1519 = 24, 36 and 48 h), d.p.f. 5-7 epiblasts from human blastocysts and d.p.f. 3-5 human morula cells
1520 were downsampled to 100 cells.
1521 **(E)** Heatmap showing expression levels of selected genes reportedly related to human epiblast
1522 pluripotency states (Kinoshita et al., 2021; Takashima et al., 2014; Wang et al., 2021). Color
1523 codes are consistent with **D**.
1524
1525 In **A-C**, we could not identify reasonable distinct populations using unsupervised clustering
1526 algorithm (Seurat R package).

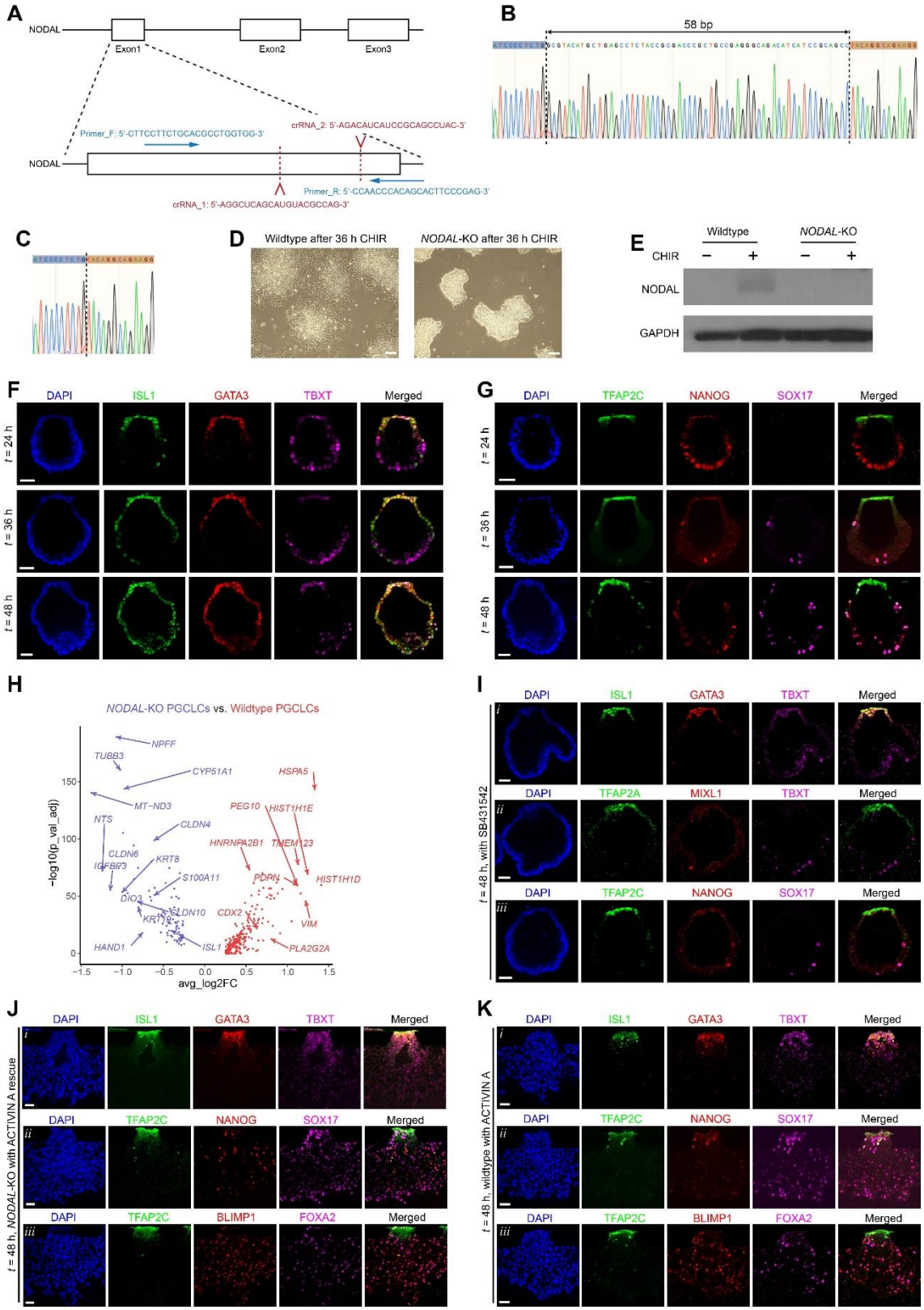
Supplemental Figure 6



1527

1528 **Supplemental Figure 6. Intercellular communication network analysis of μPASEs,**
1529 **cynomolgus embryos and mouse embryos at the peri-gastrulation stage. Related to Figure**
1530 **6.**
1531 **(A)**, **(D)**, **(G)**, Heatmaps showing contributions of individual signaling pathways as incoming
1532 and outgoing signals for each cell type in μPASEs at *t* = 24 h (**A**), *in vitro* cultured cynomolgus
1533 embryos at Day 12 (**D**), and mouse embryos at E6.5 (**G**). Grey bars on the right indicate relative
1534 signal strengths of each pathway across all cell types within local tissue environments.
1535 **(B)**, **(E)**, **(H)**, Circle plots showing inferred signal networks of selected pathways in μPASEs at *t*
1536 = 24 h (**B**), *in vitro* cultured cynomolgus embryos at Day 12 (**E**), and mouse embryos at E6.5
1537 (**H**). The dot size is proportional to the number of cells for each cell type, and the line thickness
1538 corresponds to the communication probability.
1539 **(C)**, **(F)**, **(I)**, Dot plots showing expression levels of selected ligands and receptors in μPASEs at
1540 *t* = 24 h (**C**), *in vitro* cultured cynomolgus embryos at Day 12 (**F**), and mouse embryos at E6.5
1541 (**I**). The sizes and colors of dots indicate the proportion of cells expressing the corresponding
1542 genes and their averaged scaled values of log-transformed expression, respectively.

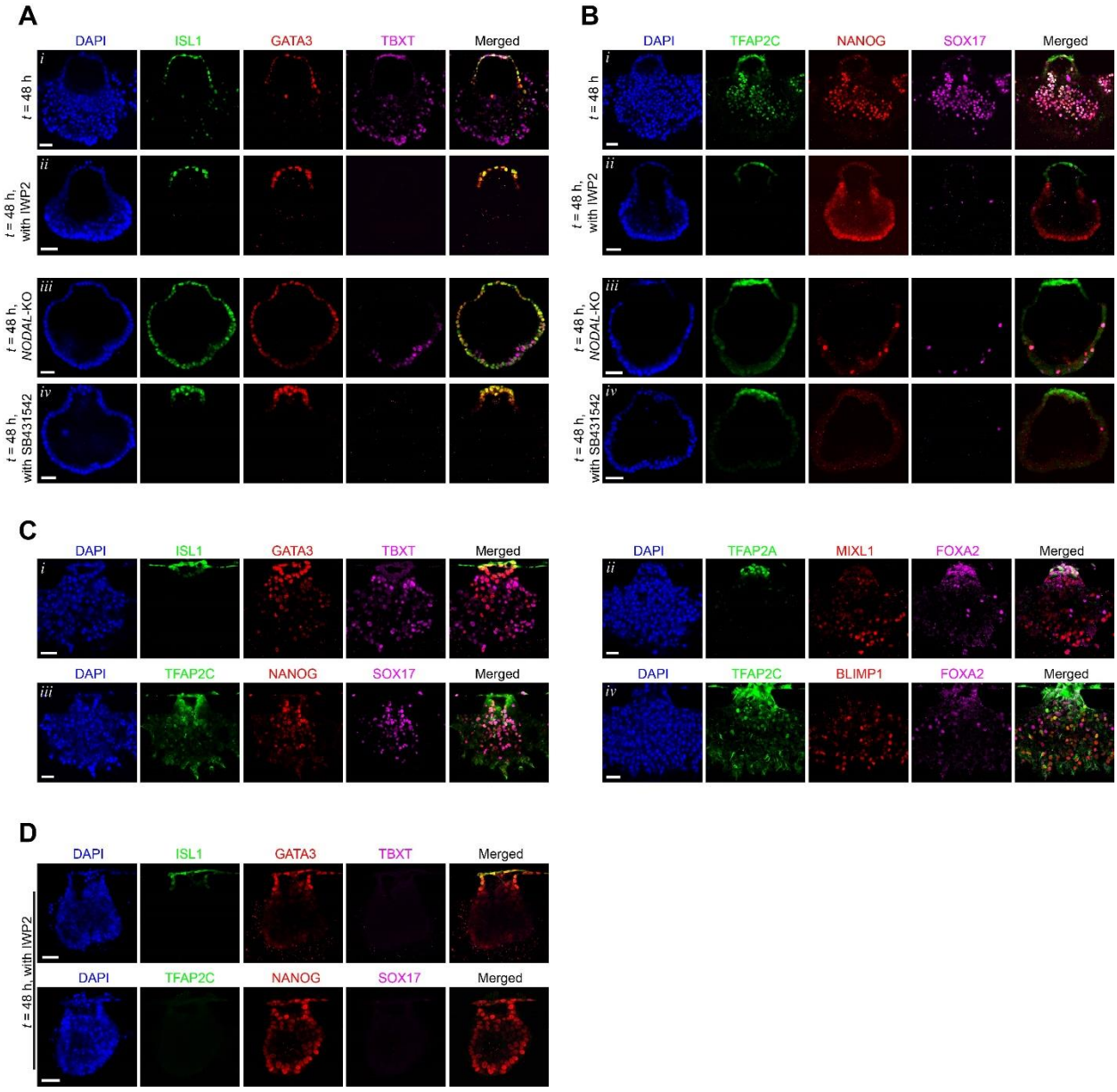
Supplemental Figure 7



1543

1544 **Supplemental Figure 7. NODAL is essential for MeLC lineage development in μPASEs.**
1545 **Related to Figure 6.**
1546 **(A)** Generation of *NODAL*-KO hPSCs. Two crRNA:tracrRNA duplexes were used
1547 simultaneously to delete a 58-bp portion of genomic DNA within the exon1 of *NODAL*.
1548 **(B), (C)** Genomic DNA sequences of *NODAL* exon1 before and after CRISPR/Cas9-mediated
1549 gene deletion.
1550 **(D)** Phase-contrast microscopy images of wildtype and *NODAL*-KO hPSC clones after exposure
1551 to CHIR99021 (10 μM) for 36 h.
1552 **(E)** Western blot showing NODAL protein expression in wildtype and *NODAL*-KO hPSCs after
1553 exposure to CHIR 99021 (10 μM) for 24 h.
1554 **(F)** Representative confocal micrographs showing *NODAL*-KO μPASEs at indicated time points
1555 stained for ISL1, GATA3 and TBXT.
1556 **(G)** Representative confocal micrographs showing *NODAL*-KO μPASEs at indicated time points
1557 stained for TFAP2C, NANOG and SOX17.
1558 **(H)** Volcano plot showing DEGs between PGCLCs from wildtype and *NODAL*-KO μPASEs,
1559 with selected genes labelled. The full DEG lists can be found in **Table S7**.
1560 **(I)** Representative confocal micrographs showing wildtype μPASEs with SB431542
1561 supplemented into the basal medium from $t = 0$ h, stained for ISL1, GATA3 and TBXT (*i*);
1562 TFAP2A, MIXL1 and TBXT (*ii*); TFAP2C, NANOG and SOX17 (*iii*).
1563 **(J)** Representative confocal micrographs showing *NODAL*-KO μPASEs with ACTIVIN A
1564 supplemented into the channel opposite to BMP4 stimulation from $t = 0$ h, stained for ISL1,
1565 GATA3 and TBXT (*i*); TFAP2C, NANOG and SOX17 (*ii*); TFAP2C, BLIMP1 and FOXA2
1566 (*iii*).
1567 **(K)** Representative confocal micrographs showing wildtype μPASEs with ACTIVIN A
1568 supplemented into the channel opposite to BMP4 stimulation from $t = 0$ h, stained for ISL1,
1569 GATA3 and TBXT (*i*); TFAP2C, NANOG and SOX17 (*ii*); TFAP2C, BLIMP1 and FOXA2
1570 (*iii*).
1571
1572 In **D**, **E and I-K**, experiments were repeated twice with similar results. In **F** and **G**, experiments
1573 were repeated three times with similar results, and nuclei were counterstained with DAPI. Scale
1574 bars, 100 μm (**D**) and 50 μm (**F**, **G, I-K**).

Supplemental Figure 8

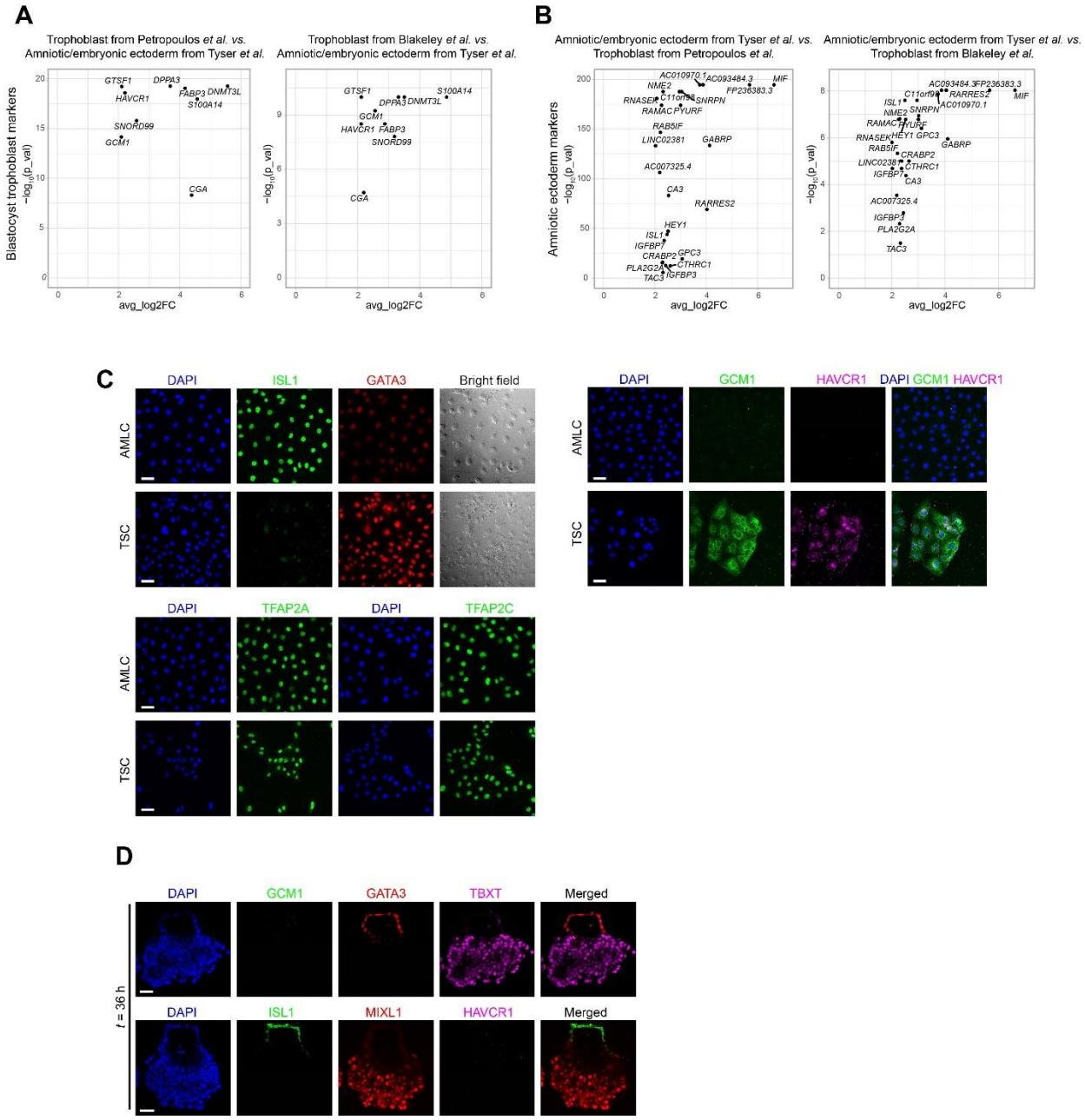


Supplemental Figure 8. µPASEs generated with ESI-017 hPSC line and C4955 chimpanzee iPSC line.

(A) Representative confocal micrographs showing µPASEs generated with ESI-017 hPSCs at *t* = 48 h stained for ISL1, GATA3 and TBXT: wildtype (*i*), wildtype supplemented with IWP2 (*ii*), *NODAL*-KO (*iii*), wildtype supplemented with SB431542 (*iv*).

1581 **(B)** Representative confocal micrographs showing µPASEs generated with ESI-017 hPSCs at t =
1582 48 h stained for TFAP2C, NANOG and SOX17: wildtype (*i*), wildtype supplemented with IWP2
1583 (*ii*), *NODAL*-KO (*iii*), wildtype supplemented with SB431542 (*iv*).
1584 **(C)** Representative confocal micrographs showing chimpanzee µPASEs generated with C4955
1585 chimpanzee iPSCs at *t* = 48 h stained for ISL1, GATA3 and TBXT (*i*); TFAP2A, MIXL1 and
1586 FOXA2 (*ii*); TFAP2C, NANOG and SOX17 (*iii*); TFAP2C, BLIMP1 and FOXA2 (*iv*).
1587 **(D)** Representative confocal micrographs showing chimpanzee µPASEs generated with C4955
1588 chimpanzee iPSCs, supplemented with IWP2, at *t* = 48 h stained for ISL1, GATA3 and TBXT
1589 (*top*); TFAP2C, NANOG and SOX17 (*bottom*).
1590
1591 Experiments were repeated twice with similar results. Nuclei were counterstained with DAPI.
1592 Scale bars, 50 µm.

Supplemental Figure 9



Supplemental Figure 9. Stringent criteria for identifying human trophoblast and amniotic ectoderm. Related to Figure 7.

(A) Highly expressed genes identified using stringent criteria in human trophoblast, as compared with the amniotic/embryonic ectoderm from Ref. (Tyser et al., 2021) shared between the trophoblast data from Ref. (Petropoulos et al., 2016) and Ref. (Blakeley et al., 2015) as indicated.

1599 **(B)** Highly expressed genes identified using stringent criteria in human amniotic/embryonic
1600 ectoderm from Ref. (Tyser et al., 2021), as compared with the trophoblast from Ref.
1601 (Petropoulos et al., 2016) and Ref. (Blakeley et al., 2015) as indicated.

1602 **(C)** Representative micrographs showing human trophoblast stem cells (TSCs) and amniotic
1603 ectoderm-like cells derived by treating hPSCs with BMP4, stained for ISL1 and GATA3; GCM1
1604 and HAVCR1; TFAP2A or TFAP2C, as indicated.

1605 **(D)** Representative micrographs showing μPASEs at t = 36 h stained for GCM1, GATA3 and
1606 TBXT (*top*); ISL1, MIXL1 and HAVCR1 (*bottom*).

1607 In **C** and **D**, experiments were repeated twice with similar results. Nuclei were counterstained
1608 with DAPI. Scale bars, 50 μm.