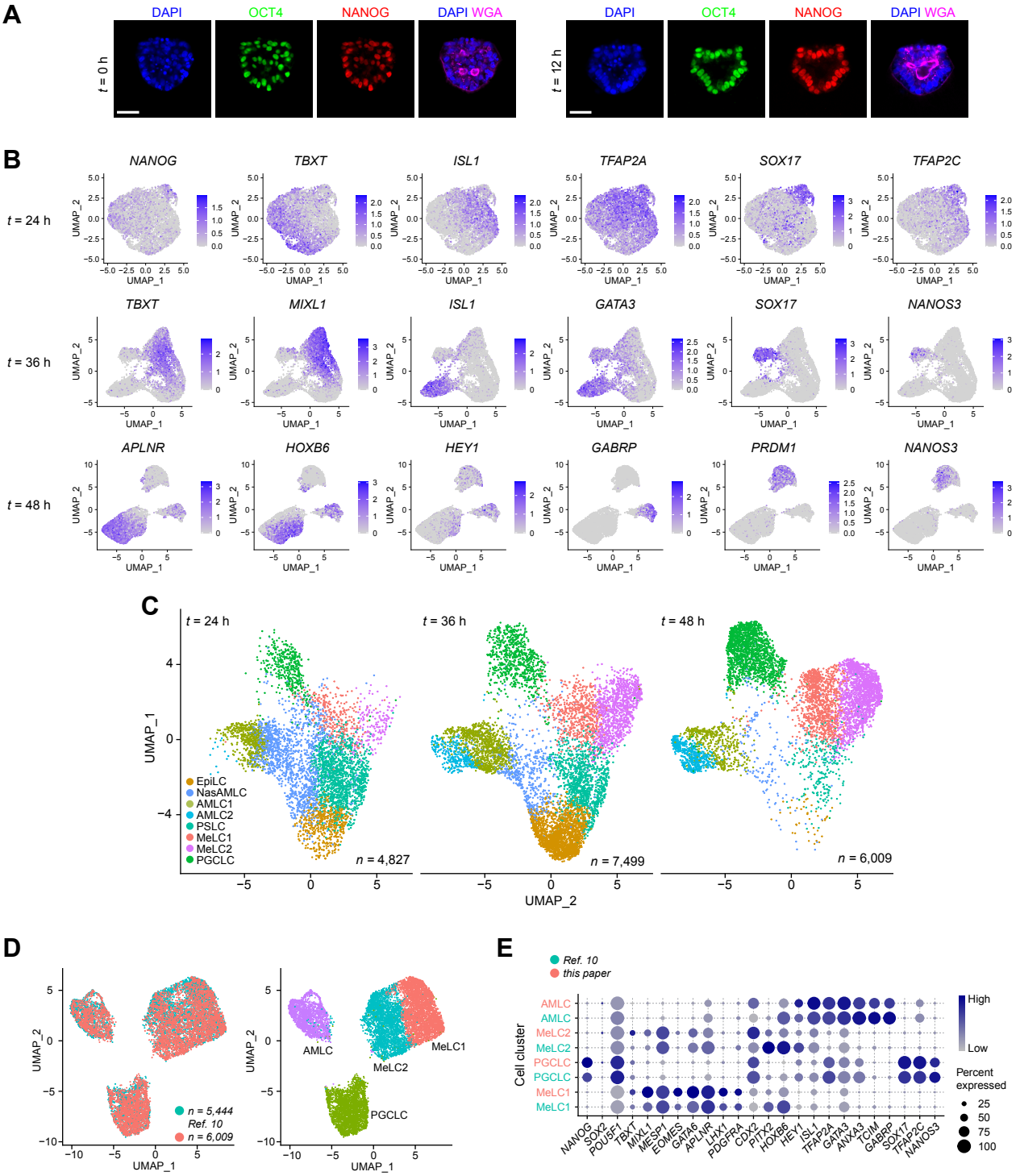
Cell Stem Cell, Volume *29*

Supplemental Information

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

Yi Zheng, Robin Zhexuan Yan, Shiyu Sun, Mutsumi Kobayashi, Lifeng Xiang, Ran Yang, Alexander Goedel, Yu Kang, Xufeng Xue, Sajedeh Nasr Esfahani, Yue Liu, Agnes M. Resto Irizarry, Weisheng Wu, Yunxiu Li, Weizhi Ji, Yuyu Niu, Kenneth R. Chien, Tianqing Li, Toshihiro Shioda, and Jianping Fu

Supplemental Figure 1



Supplemental Figure 1. Single-cell transcriptomic profiling of µPASE development. Related to Figure 1.

(A) Representative confocal micrographs showing µPASEs at *t* = 0 h (*left*) and *t* = 12 h (*right*) stained for OCT4 and NANOG. Plasma membrane is stained with fluorescently labelled wheat germ agglutinin (WGA).

(B) Feature plots showing selected lineage markers used for cell identity annotations in the UMAP plots of µPASEs at *t* = 24, 36 and 48 h, respectively.

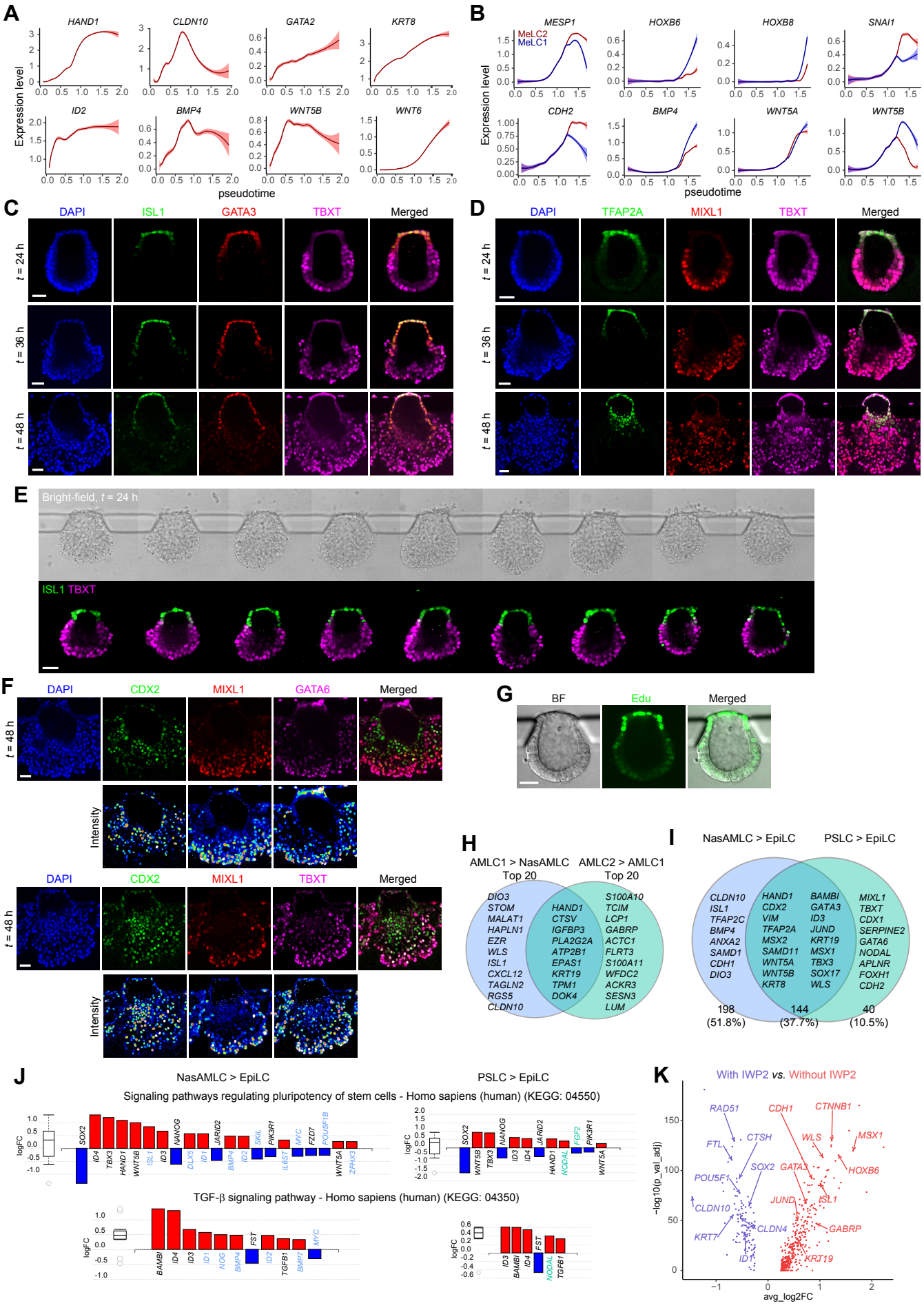
(C) UMAP plots of µPASEs at *t* = 24, 36, 48 h, separated from the integrated UMAP plot in **Figure 1D**. *n* indicates the cell number.

(D) Integrated UMAP plots of datasets from µPASEs at *t* = 48 h and the published dataset in Ref.(Zheng et al., 2019b). *Left:* Datasets are color-coded according to sources, with *n* indicating the cell number; *right:* Cells are color-coded according to cell identity annotations.

(E) Dot plot showing expression of key marker genes across the cell clusters as indicated (µPASEs at *t* = 48 h from this paper (red) and Ref.(Zheng et al., 2019b) (blue)). The sizes and colors of dots indicate the proportion of cells expressing the corresponding genes and their averaged scaled values of log-transformed expression, respectively.

In **A**, experiments were repeated four times with similar results. Nuclei were counterstained with DAPI. Scale bars, 50 µm.

Supplemental Figure 2



Supplemental Figure 2. Trajectory inference and gene expression dynamics during µPASE development. Related to Figure 2.

(A) Expression dynamics (pseudotime) of selected genes during AMLC lineage development. Level of confidence (0.95) is indicated by band width.

(B) Expression dynamics (pseudotime) of selected genes during PSLC / MeLC lineage development. Level of confidence (0.95) is indicated by band width.

(C) Representative confocal micrographs showing µPASEs at indicated time points stained for ISL1, GATA3 and TBXT.

(D) Representative confocal micrographs showing µPASEs at indicated time points stained for TFAP2A, MIXL1 and TBXT.

(E) Bright-field and immunofluorescence images showing an array of µPASEs at t = 24 h, stained for ISL1 and TBXT.

(F) Representative confocal micrographs showing µPASEs at t = 48 h stained for CDX2, MIXL1 and GATA6 (*top*); CDX2, MIXL1 and TBXT (*bottom*). Intensity maps show relative intensities of corresponding markers as indicated.

(G) Representative confocal micrographs showing proliferating cells in µPASEs. Images were taken at t = 27 h. Cell nuclei with newly synthesized DNA within the last 3 hours were labeled using Click-iT EdU Imaging Kit (Invitrogen). Note: The low intensity of the Edu signal from cells embedded in the gel is caused by insufficient dye diffusion.

(H) Venn diagram showing top 20 upregulated genes in AMLC1 *vs.* NasAMLC and AMLC2 *vs.* AMLC1. The full DEG lists can be found in **Mendeley Data Table 3**.

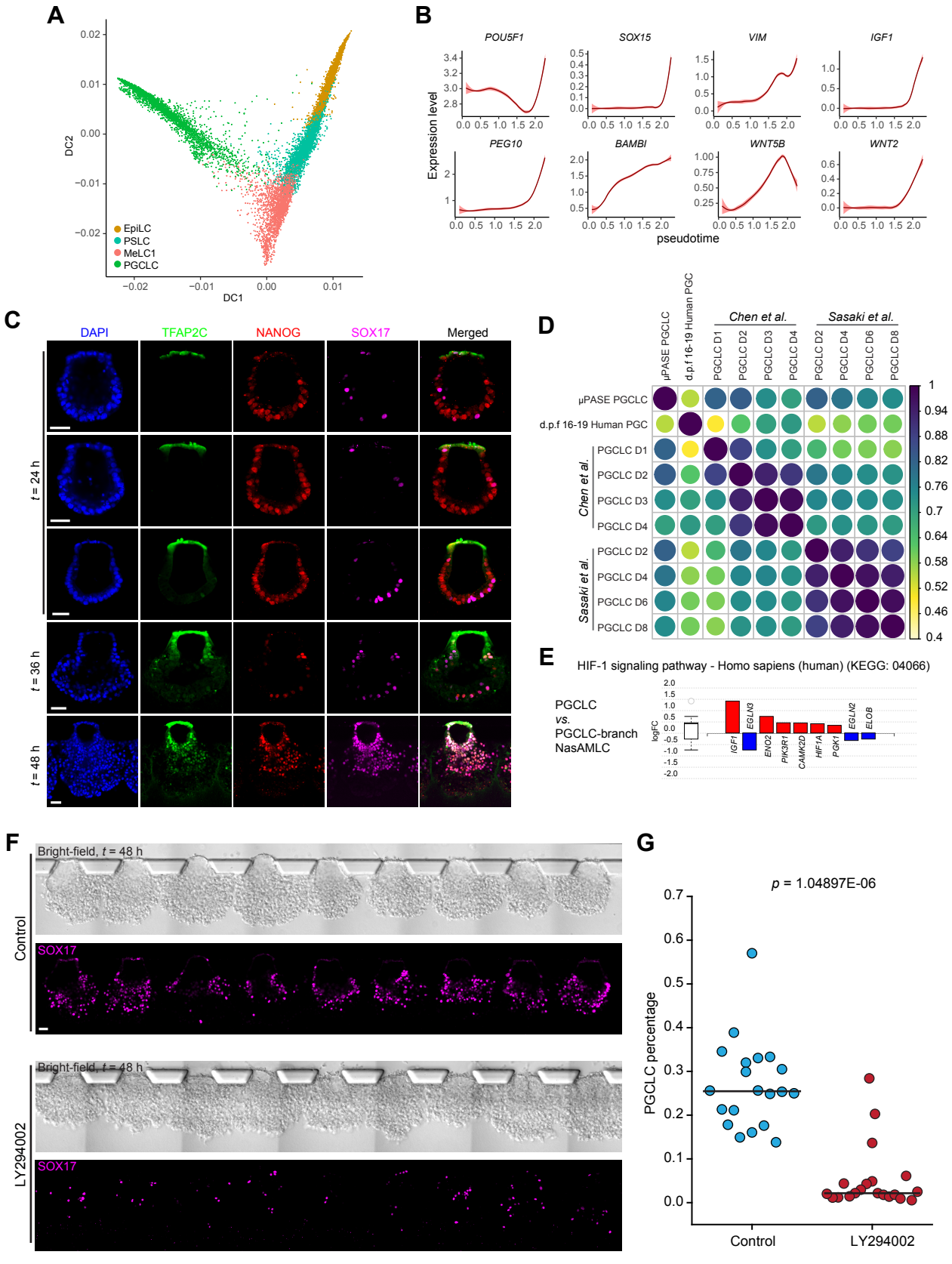
(I) Venn diagram showing selected upregulated genes in NasAMLC and PSLC, as compared to EpiLC. The full DEG lists can be found in **Mendeley Data Table 3**.

(J) DEGs related to pluripotency signaling pathway (KEGG: 04550) and TGF-β signaling pathway (KEGG: 04350) in NasAMLC and PSLC, as compared to EpiLC. Blue and green colors highlight genes identified only for NasAMLC and PSLC, respectively.

(K) Volcano plots showing DEGs between AMLCs from µPASEs with or without IWP2 at t = 48 h, with selected genes labelled. The full DEG list can be found in **Mendeley Data Table 3.**

In **C-G**, experiments were repeated three times with similar results. Nuclei were counterstained with DAPI. Scale bars, 50 µm.

Supplemental Figure 3



Supplemental Figure 3. PGCLC specification during µPASE development. Related to Figure 3.

(A) Diffusion map of EpiLC, PSLC, MeLC1 and PGCLC clusters from the UMAP plot in **Figure 1D**. PGCLC cluster is discontinuous from other clusters. K-Branches algorithm failed to identify branches or branching points with sufficient confidence.

(B) Expression dynamics (pseudotime) of selected genes during PGCLC lineage development. Level of confidence (0.95) is indicated by band width.

(C) Representative confocal micrographs showing µPASEs at indicated time points stained for TFAP2C, NANOG and SOX17.

(D) Heat map of correlation matrix for PGCLCs in µPASEs, human PGCs (Ref. (Tyser et al., 2021)) and PGCLCs derived using other protocols (Ref. (Chen et al., 2019; Sasaki et al., 2015)). Correlation coefficients between indicated cell types are calculated based on PGC ontogenic genes identified for cynomolgus embryo transcriptome data.

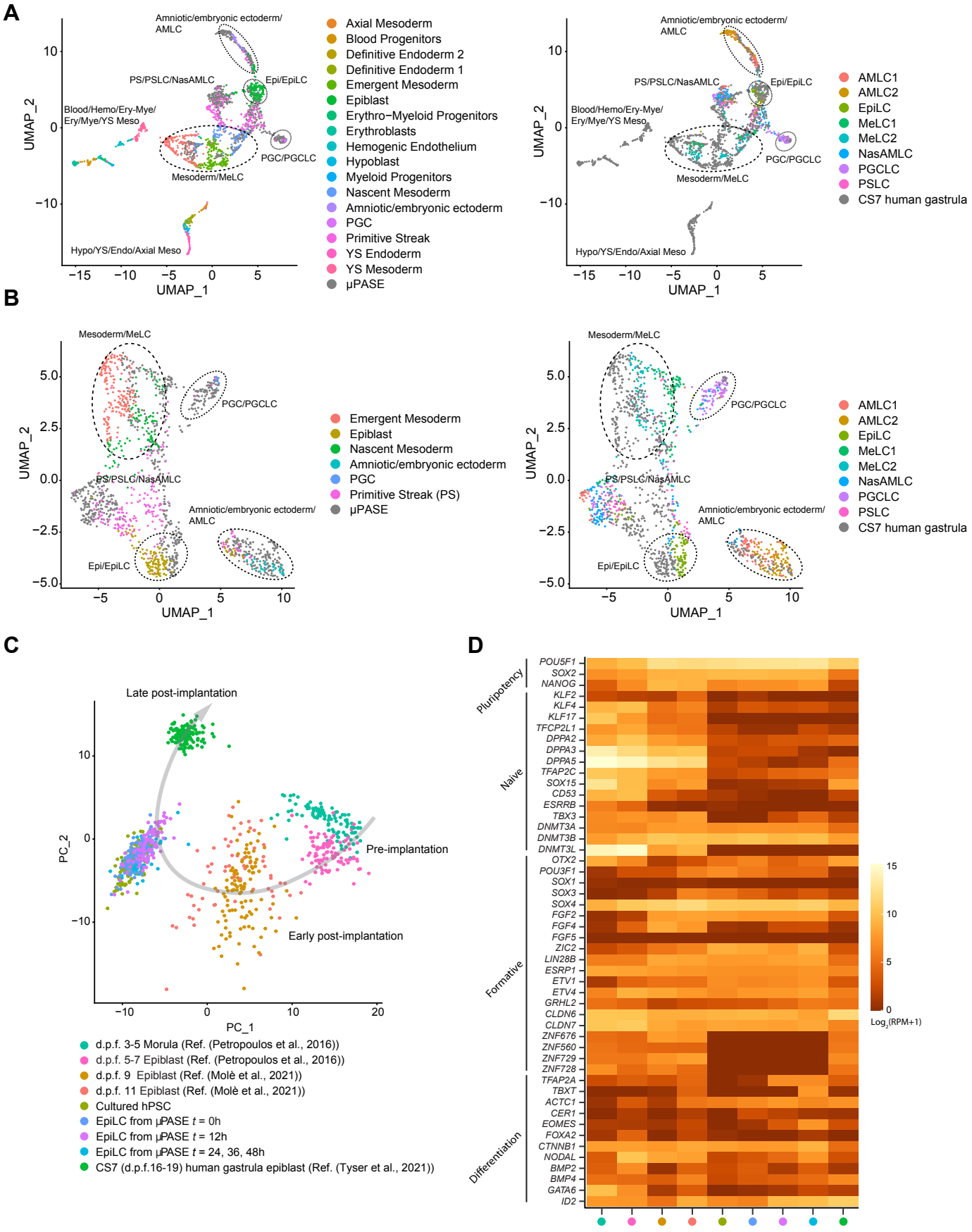
(E) DEGs related to HIF-1 signaling pathway (KEGG: 04066) in PGCLC as compared to PGCLC-branch NasAMLC.

(F) Representative confocal micrographs showing arrays of µPASEs at t = 48 h, stained for SOX17; control (*top*); with LY294002 supplemented into the basal medium from t = 0 h (*bottom*).

(G) Percentage of SOX17+ PGCLCs in µPASEs at t = 48 h under indicated conditions. n = 20 µPASEs for each condition. Data were pooled from n = 2 independent experiments. Red lines represent the median.

In **C** and **F**, experiments were repeated three times with similar results. Nuclei were 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 and single-cell transcriptomic profiling of cultured hPSCs, µPASEs at *t* = 0 h and *t* = 12 h. Related to Figure 4.

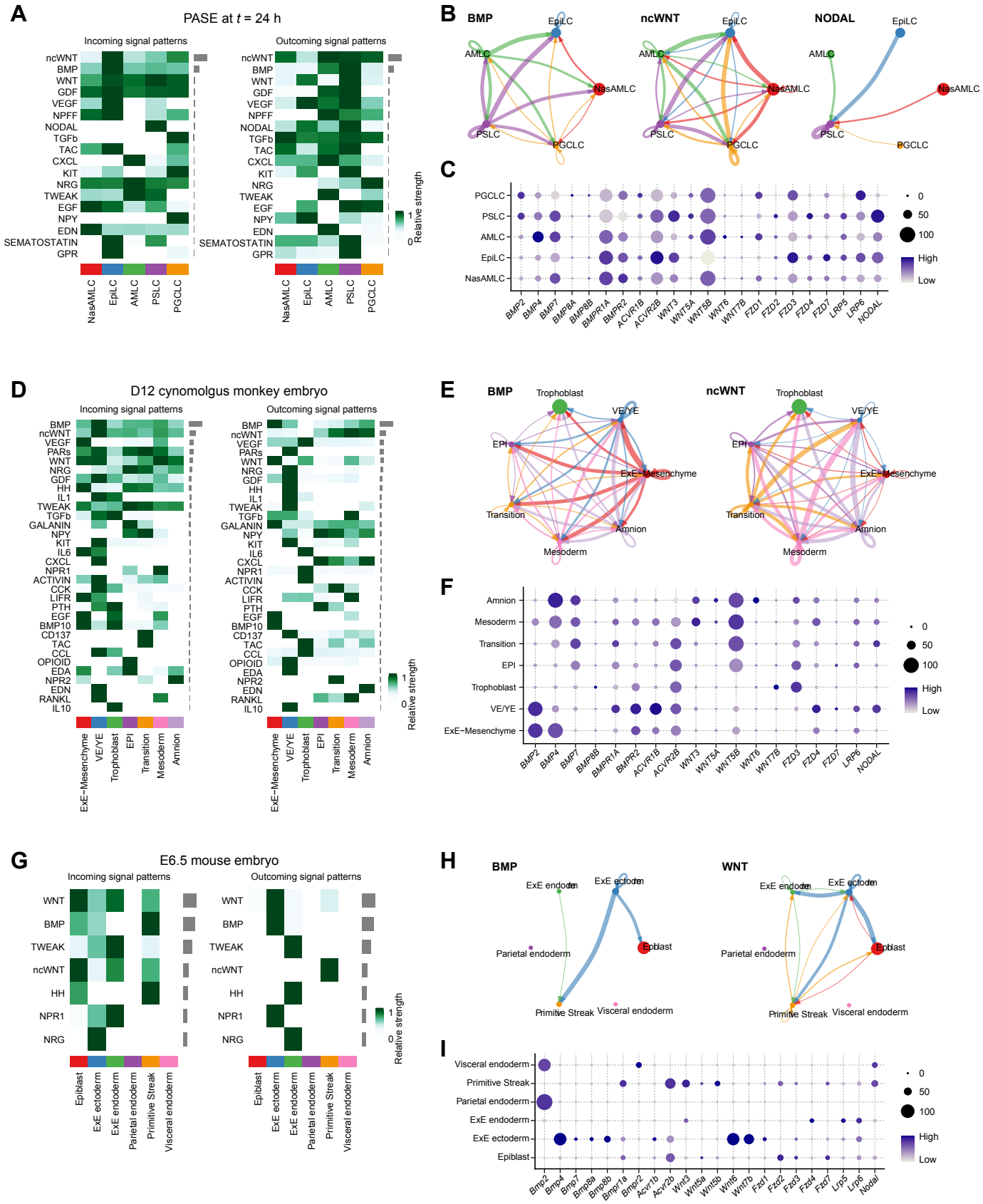
(A) UMAP plots of integrated dataset of CS7 human gastrula (all 1,195 cells) and downsampled µPASEs (from **Figure 1D**, 100 cells per cluster). *Left*: color-coded according to original cell identity annotations of CS7 human gastrula; grey color indicates cells from µPASEs. *Right:* color-coded according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

(C) Principal component analysis (PCA) plot of cultured hPSCs, µPASEs at *t* = 0 h and µPASEs at *t* = 12 h, EpiLCs from µPASEs in **Figure 1D** (at *t* = 24, 36 and 48 h), d.p.f. 9 (*n* = 108) and d.p.f. 11 (*n* = 62) epiblasts from *in vitro* cultured human embryos (Ref.(Molè et al., 2021)), epiblasts from CS7 human gastrula (*n* = 133) (Ref. (Tyser et al., 2021)), d.p.f. 5-7 epiblasts from human blastocysts and d.p.f. 3-5 human morula cells (Ref.(Petropoulos et al., 2016)). PCAs were calculated using epiblast ontogenic genes identified from cynomolgus embryos (Nakamura et al., 2016, see **Mendeley Data Table 5**). To prevent datasets with high number of cells dominating PCA calculation, cultured hPSCs, µPASEs at *t* = 0 h, µPASEs at *t* = 12 h, EpiLCs from µPASEs (at *t* = 24, 36 and 48 h), d.p.f. 5-7 epiblasts from human blastocysts and d.p.f. 3-5 human morula cells were downsampled to 100 cells.

(D) Heatmap showing expression levels of selected genes reportedly related to human epiblast pluripotency states (Kinoshita et al., 2021; Takashima et al., 2014; Wang et al., 2021). Color codes are consistent with **C**.

Supplemental Figure 5



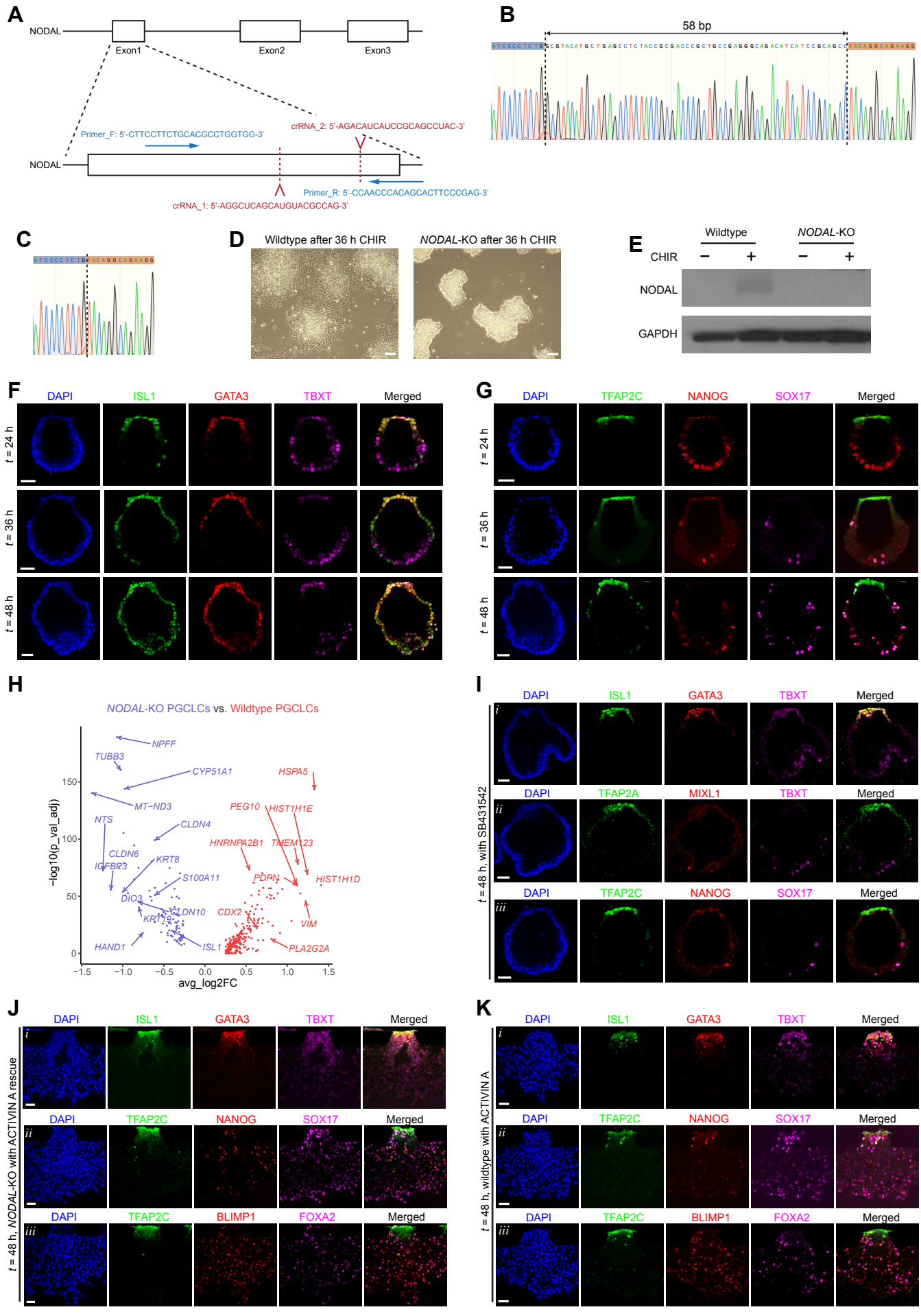
Supplemental Figure 5. Intercellular communication network analysis of μPASEs, cynomolgus embryos and mouse embryos at the peri-gastrulation stage. Related to Figure 6.

(A), **(D)**, **(G)**, Heatmaps showing contributions of individual signaling pathways as incoming and outgoing signals for each cell type in μPASEs at *t* = 24 h (**A**), *in vitro* cultured cynomolgus embryos at Day 12 (**D**), and mouse embryos at E6.5 (**G**). Grey bars on the right indicate relative signal strengths of each pathway across all cell types within local tissue environments.

(B), **(E)**, **(H)**, Circle plots showing inferred signal networks of selected pathways in μPASEs at *t* = 24 h (**B**), *in vitro* cultured cynomolgus embryos at Day 12 (**E**), and mouse embryos at E6.5 (**H**). The dot size is proportional to the number of cells for each cell type, and the line thickness corresponds to the communication probability.

(C), **(F)**, **(I)**, Dot plots showing expression levels of selected ligands and receptors in μPASEs at *t* = 24 h (**C**), *in vitro* cultured cynomolgus embryos at Day 12 (**F**), and mouse embryos at E6.5 (**I**). The sizes and colors of dots indicate the proportion of cells expressing the corresponding genes and their averaged scaled values of log-transformed expression, respectively.

Supplemental Figure 6



Supplemental Figure 6. NODAL is essential for MeLC lineage development in µPASEs. Related to Figure 6.

(A) Generation of *NODAL*-KO hPSCs. Two crRNA:tracrRNA duplexes were used simultaneously to delete a 58-bp portion of genomic DNA within the exon1 of *NODAL*.

(B), (C) Genomic DNA sequences of *NODAL* exon1 before and after CRISPR/Cas9-mediated gene deletion.

(D) Phase-contrast microscopy images of wildtype and *NODAL*-KO hPSC clones after exposure to CHIR99021 (10 µM) for 36 h.

(E) Western blot showing NODAL protein expression in wildtype and *NODAL*-KO hPSCs after exposure to CHIR 99021 (10 µM) for 24 h.

(F) Representative confocal micrographs showing *NODAL*-KO µPASEs at indicated time points stained for ISL1, GATA3 and TBXT.

(G) Representative confocal micrographs showing *NODAL*-KO µPASEs at indicated time points stained for TFAP2C, NANOG and SOX17.

(H) Volcano plot showing DEGs between PGCLCs from wildtype and *NODAL*-KO µPASEs, with selected genes labelled. The full DEG lists can be found in **Mendeley Data Table 7**.

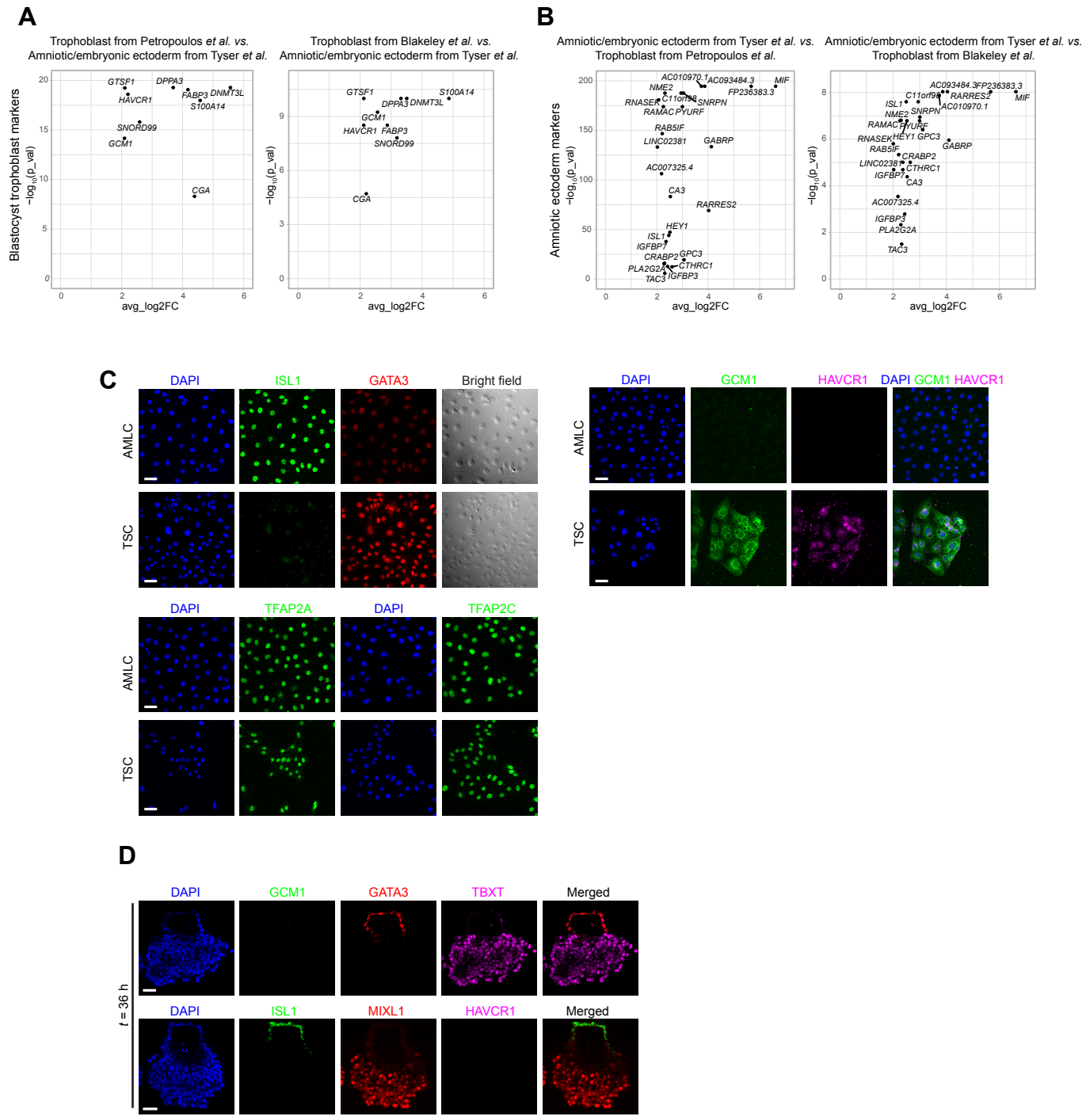
(I) Representative confocal micrographs showing wildtype µPASEs with SB431542 supplemented into the basal medium from t = 0 h, stained for ISL1, GATA3 and TBXT (*i*); TFAP2A, MIXL1 and TBXT (*ii*); TFAP2C, NANOG and SOX17 (*iii*).

(J) Representative confocal micrographs showing *NODAL*-KO µPASEs with ACTIVIN A supplemented into the channel opposite to BMP4 stimulation from t = 0 h, stained for ISL1, GATA3 and TBXT (*i*); TFAP2C, NANOG and SOX17 (*ii*); TFAP2C, BLIMP1 and FOXA2 (*iii*).

(K) Representative confocal micrographs showing wildtype µPASEs with ACTIVIN A supplemented into the channel opposite to BMP4 stimulation from t = 0 h, stained for ISL1, GATA3 and TBXT (*i*); TFAP2C, NANOG and SOX17 (*ii*); TFAP2C, BLIMP1 and FOXA2 (*iii*).

In **D**, **E and I-K**, experiments were repeated twice with similar results. In **F** and **G**, experiments were repeated three times with similar results, and nuclei were counterstained with DAPI. Scale bars, 100 µm (**D**) and 50 µm (**F**, **G, I-K**).

Supplemental Figure 7



Supplemental Figure 7. 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.

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

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

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

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