Stem Cell Reports, Volume *16*

Supplemental Information

Gene expression dynamics underlying cell fate emergence in 2D micro-patterned human embryonic stem cell gastruloids

Kyaw Thu Minn, Sabine Dietmann, Sarah E. Waye, Samantha A. Morris, and Lilianna Solnica-Krezel

Supplemental Experimental Procedures

BMP4 differentiation in 2D micropatterns

H1 hESC were differentiated with BMP4 in the micropatterned substrates as previously described (Minn et al., 2020). Briefly, H1 hESC of karyotype 46, XY (WiCell, Madison, WI), routinely cultured in Matrigel-coated plates with mTeSR media (Stemcell Technologies, Vancouver, Canada), were differentiated with BMP4 in the micropatterned substrates as previously described (Minn et al., 2020). Briefly, H1 cells were dissociated with Accutase (Sigma Aldrich, St. Louis, MO) incubation at 37°C and 5% CO_2 for 8 min. Equal volume of RPMI medium 1640 (Life Technologies, Carlsbad, CA) was added, and the cell solution was centrifuged at 300 rcf (relative centrifugal force) for 5min. After the supernatant was removed, cells were diluted with fresh mTeSR, and seeded onto micropatterns at 132,000–263,000 cells/cm^2 in mTeSR with 10 μM Rho-associated kinase inhibitor (ROCKi Y-27632, Millipore Sigma, Burlington, MA). After 2h, the medium was replaced with fresh mTeSR, and cultured for additional 3h. Afterwards (6h after initial cell seeding), the medium was replaced with mTeSR containing 50 ng/mL BMP4 (R&D Systems, Minneapolis, MN) for 12h, 24h, and 44h.

scRNA-seq and data analysis

Single cells collection and scRNA-seq and data analyses were performed as described previously (Minn et al., 2020). Briefly, micropatterned cells after 12h, 24h, and 44h BMP4 treatment were washed with PBS, and dissociated into single cells with Accutase incubation at 37°C and 5% CO_2 for 10 min. Equal volume of RPMI medium was added, and the solution was centrifuged 300 rcf for 5 min. After the supernatant was removed, cells were diluted at 20,000 cells per 200 μL of cold DPBS-/-. Afterwards, 800 μL of cold

methanol was added dropwise to the cell solution, and incubated on ice for 15 min. The final solution was kept at -80°C until use.

10x Chromium Single Cell 3’ Library & Gel Bead Kit v2, Chromium Next GEM Single Cell 3’ GEM, Library & Gel Bead Kit v3.1, Chromium Next GEM Single Cell 3' Kit v3.1, Chromium Next GEM Chip G Single Cell Kit, Single Index Kit T Set A, and Dual Index Kit TT Set A and Chromium i7 Multiplex Kit (10X Genomics, Pleasanton, CA) were used according to the manufacturer’s instructions to prepare single-cell library. Agilent TapeStation (Agilent Scientific Instruments, Santa Clara, CA) was used to quantify cDNA libraries. Sequencing was performed on Illumina HiSeq 2500 and NextSeq500 (Illumina, San Diego, CA).

The Cell Ranger v.2.1.0 pipeline was used to align reads to a custom hg19 genome build including transgenes, and to generate a digital gene expression matrix. Seurat package (v.3.2.0) was used for processing and visualization (Butler et al., 2018; Stuart et al., 2019). Default settings were used unless noted otherwise. For each dataset, cells with outlier number of genes and mitochondrial gene expression were excluded. We normalized and scaled the filtered expression matrix to remove unwanted sources of variation driven by mitochondrial gene expression, the number of detected UMIs, and the cell cycle. For 24h and 44h datasets, we combined replicates at each time point using canonical correlation analysis selecting for 30 dimensions and 3,000 anchor features (highly variable genes). Next, we combined all 12h, 24h, and 44h datasets using the same parameters. Nonlinear dimensionality reduction by UMAP was performed on the first 30 principal components using the implementation by Seurat. With previously annotated 44h cells serving as reference (Minn et al., 2020), we

annotated cells at 12h and 24h using FindTransferAnchors and TransferData functions in Seurat. This anchor-based approach calculated prediction score in each 12h and 24h cell for referenced 44h cell types. Each 12h or 24h cell with the highest prediction score for a particular 44h cell type was annotated as that cell type. DEGs across cell types or time course were identified using FindMarkers or FindAllMarkers functions, with threshold settings of 0.25 log fold-change and 25% detection rate.

For comparison with published dataset, CCA was used as described above for combining gastruloid datasets. Average gene expression correlation was calculated using Spearman correlation on the basis of highly variable genes from reference datasets. Prediction score was calculated using the anchor-based approach described above, after the reference and query datasets were combined with CCA. Module score was calculated using AddModuleScore function in Seurat, which calculates the difference in average gene expression between a set of marker genes of interest and that of randomly selected genes (Tirosh et al., 2016).

Diffusion maps for single cells were calculated based on the normalized and scaled gene expression data matrix using the R Bioconductor *destiny* package (Angerer et al., 2016), with a number of k-nearest neighbors, knn = 40, and a Gaussian kernel width, sigma = 8, slightly lower than the optimal value of sigma estimated by *destiny*. A probabilistic breadth-first search of the k-nearest neighbor graph was performed and the results of this search were converted into a pseudotime.

To identify genes differentially expressed between emergent cell populations (Ectoderm versus 0h hESCs, Table 1S), we used the Seurat FindMarkers function to compare against all other cells types and used Seurat threshold of logFC > 0.1 and p

value < 0.05 to consider as significant differences in expression. In addition, we inspected the localized expression in a scatterplot.

References:

Angerer P, Haghverdi L, Büttner M, Theis FJ, Marr C, Buettner F. 2016. destiny: diffusion maps for large-scale single-cell data in R. *Bioinfor* **32**:1241–1243. doi:10.1093/bioinformatics/btv715

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

Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, Hao Y, Stoeckius M, Smibert P, Satija R. 2019. Comprehensive Integration of Single-Cell Data. *Cell* **177**:1888-1902.e21. doi:10.1016/j.cell.2019.05.031

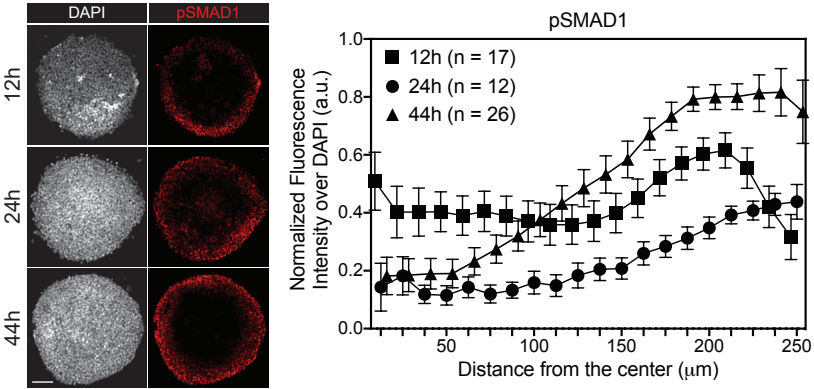
Tirosh I, Izar B, Prakadan SM, Wadsworth MH, Treacy D, Trombetta JJ, Rotem A, Rodman C, Lian C, Murphy G, Fallahi-Sichani M, Dutton-Regester K, Lin JR, Cohen O, Shah P, Lu D, Genshaft AS, Hughes TK, Ziegler CGK, Kazer SW, Gaillard A, Kolb KE, Villani AC, Johannessen CM, Andreev AY, Van Allen EM, Bertagnolli M, Sorger PK, Sullivan RJ, Flaherty KT, Frederick DT, Jané-Valbuena J, Yoon CH, Rozenblatt-Rosen O, Shalek AK, Regev A, Garraway LA. 2016. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science (80-)* **352**:189–196. doi:10.1126/science.aad0501

Figure 1S

A

Experiment	# colonies per experiment	# of cells after QC	Average genes per cell	Average UMI per cell
0h Replicate 1	36	1,422	3,121	11,549
0h Replicate 2	36	1,397	3,825	17,352
12h Replicate 1	36	766	4,194	23,756
12h Replicate 2	36	378	4,274	19,899
24h Replicate 1	36	947	4,525	30,756
24h Replicate 2	36	1,588	4,011	25,130
44h Replicate 1	36	1,858	4,981	29,609
44h Replicate 2	36	811	3,801	14,612

B



C

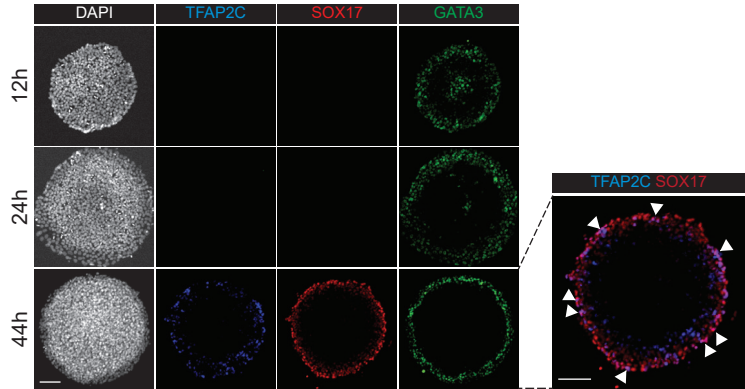


Figure 1S. Supplemental information related to Figure 1

(A) Selected parameters for scRNA-seq analysis.

(B) Immunofluorescence images (left) and quantification of fluorescence intensity (right) for BMP4 downstream effector pSMAD1. Data are represented as mean ± SEM.

(C) Immunofluorescence analysis of SOX17, TFAP2C, and GATA3. Arrow heads indicate selected hPGCLCs co-expressing SOX17 and TFAP2C.

Scale bar is 100μm.

Figure 2S

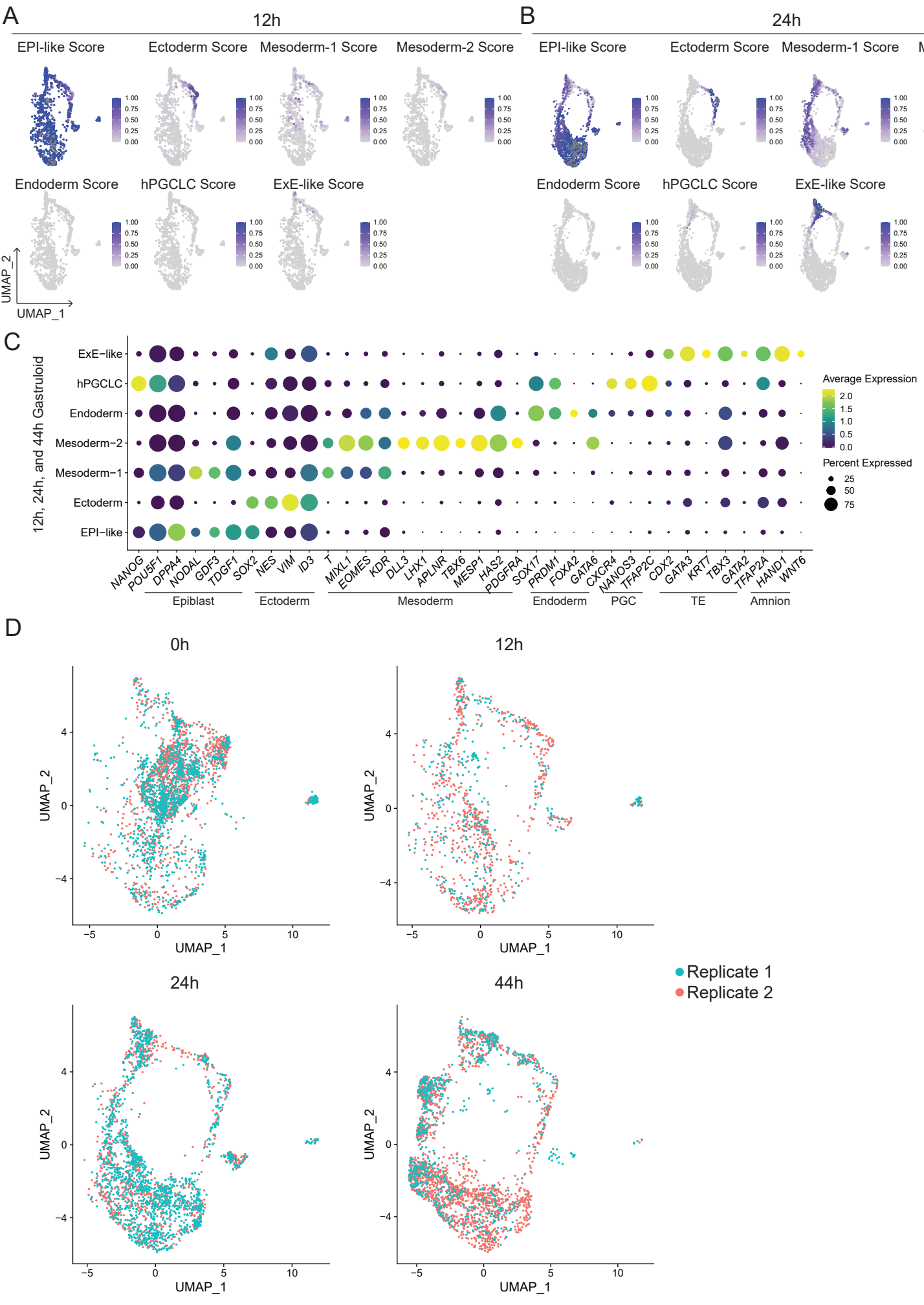

A
12h
B
24h
EPI-like Score
Ectoderm Score
Mesoderm-1 Score
Mesoderm-2 Score
Endoderm Score
hPGCLC Score
ExE-like Score
1.00
0.75
0.50
0.25
0.00
UMAP_2
UMAP_1
C
12h, 24h, and 44h Gastruloid
ExE−like
hPGCLC
Endoderm
Mesoderm−2
Mesoderm−1
Ectoderm
EPI−like
NANOG
POU5F1
DPPA4
NODAL
GDF3
TDGF1
SOX2
NES
VIM
ID3
T
MIXL1
EOMES
KDR
DLL3
LHX1
APLNR
TBX6
MESP1
HAS2
PDGFRA
SOX17
PRDM1
FOXA2
GATA6
CXCR4
NANOS3
TFAP2C
CDX2
GATA3
KRT7
TBX3
GATA2
TFAP2A
HAND1
WNT6
Epiblast
Ectoderm
Mesoderm
Endoderm
PGC
TE
Amnion
Average Expression
2.0
1.5
1.0
0.5
0.0
Percent Expressed
25
50
75
D
0h
12h
24h
44h
UMAP_2
UMAP_1
Replicate 1
Replicate 2

Figure 2S. Supplemental information related to Figure 2

(A and B) Cell type prediction scores for cells treated with BMP4 for 12h (A), and 24h (B), using 44h gastruloid cells as reference.

(C) Dot plots displaying expression of canonical markers of indicated cell types in gastruloid clusters at indicated time points.

(D) UMAP display of overlap between replicates at each indicated time point.

Figure 3S

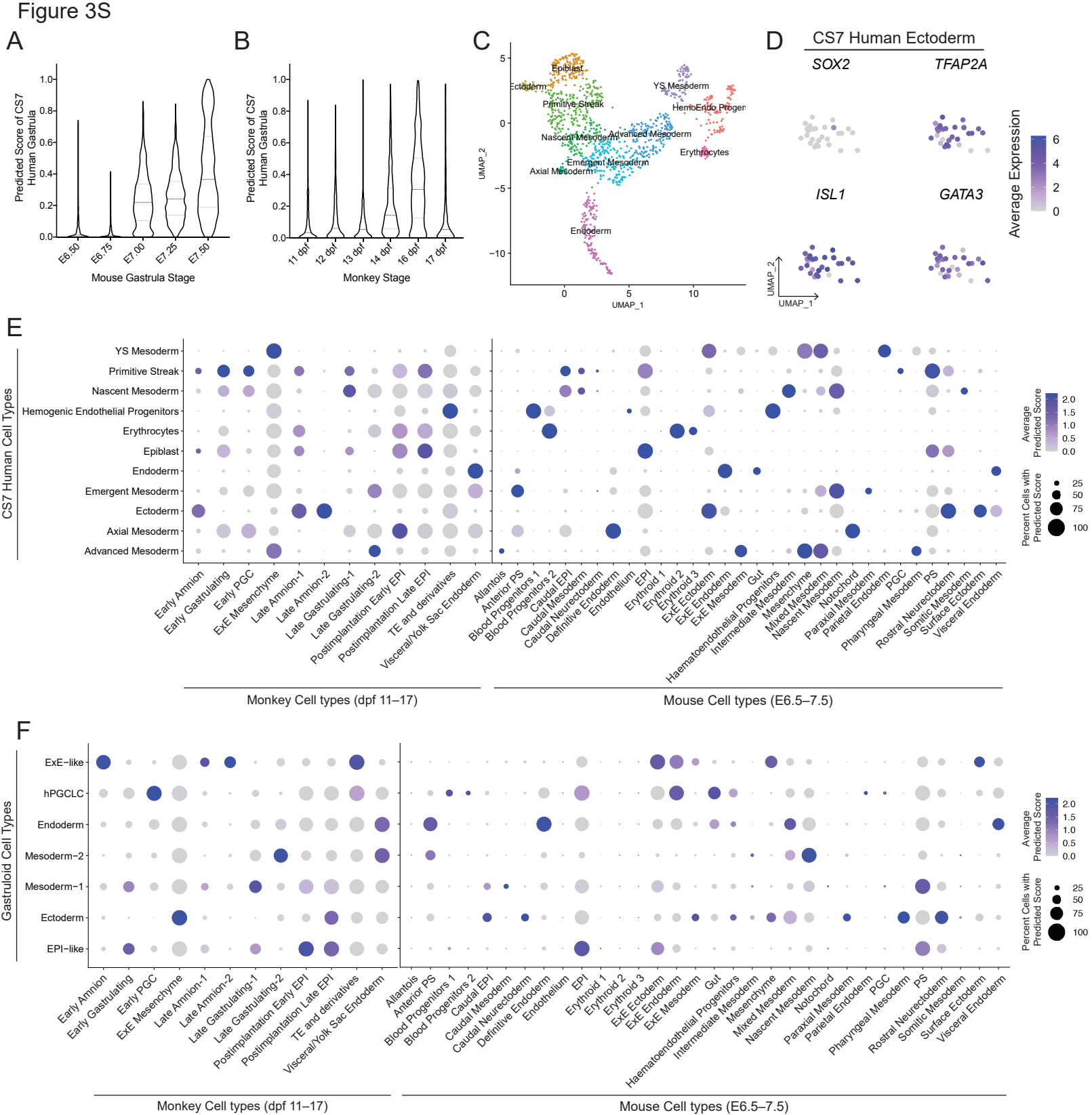

A
Predicted Score of CS7 Human Gastrula
Mouse Gastrula Stage
E6.50
E6.75
E7.00
E7.25
E7.50
B
Predicted Score of CS7 Human Gastrula
Monkey Stage
11 dpf
12 dpf
13 dpf
14 dpf
16 dpf
17 dpf
C
UMAP_1
UMAP_2
Epiblast
Primitive Streak
Nascent Mesoderm
Axial Mesoderm
Emergent Mesoderm
Advanced Mesoderm
YS Mesoderm
HemoEndo Progen
Erythrocytes
Endoderm
D
CS7 Human Ectoderm
SOX2
TFAP2A
ISL1
GATA3
Average Expression
E
CS7 Human Cell Types
YS Mesoderm
Primitive Streak
Nascent Mesoderm
Hemogenic Endothelial Progenitors
Erythrocytes
Epiblast
Endoderm
Emergent Mesoderm
Ectoderm
Axial Mesoderm
Advanced Mesoderm
Early Amnion
Early Gastrulating
Early PGC
ExE Mesenchyme
Late Amnion-1
Late Amnion-2
Late Gastrulating-1
Late Gastrulating-2
Postimplantation Early EPI
Postimplantation Late EPI
TE and derivatives
Visceral/Yolk Sac Endoderm
Allantois
Anterior PS
Blood Progenitors 1
Blood Progenitors 2
Caudal EPI
Caudal Mesoderm
Caudal Neurectoderm
Definitive Endoderm
Endothelium
EPI
Erythroid 1
Erythroid 2
Erythroid 3
ExE Ectoderm
ExE Endoderm
ExE Mesoderm
Gut
Haematoendothelial Progenitors
Intermediate Mesoderm
Mesenchyme
Mixed Mesoderm
Nascent Mesoderm
Notochord
Paraxial Mesoderm
Parietal Endoderm
PGC
Pharyngeal Mesoderm
PS
Rostral Neurectoderm
Somitic Mesoderm
Surface Ectoderm
Visceral Endoderm
Monkey Cell types (dpf 11–17)
Mouse Cell types (E6.5–7.5)
Average Predicted Score
Percent Cells with Predicted Score
F
Gastruloid Cell Types
ExE−like
hPGCLC
Endoderm
Mesoderm−2
Mesoderm−1
Ectoderm
EPI−like
Monkey Cell types (dpf 11–17)
Mouse Cell types (E6.5–7.5)
Average Predicted Score
Percent Cells with Predicted Score

Figure 3S. Supplemental information related to Figure 3

(A and B) Prediction score of mouse (A) and monkey (B) gastrula stages in CS7 human gastrula.

(C) UMAP projection showing 11 major cell types reported in CS7 human (Tyser et al., 2020).

(D) UMAP display of indicated markers expression in CS7 Ectoderm.

(E and F) Prediction score of E6.5–7.5 mouse and 11–17 dpf monkey gastrulating cell types in CS7 human gastrula (E) and gastruloids (F).

Figure 4S

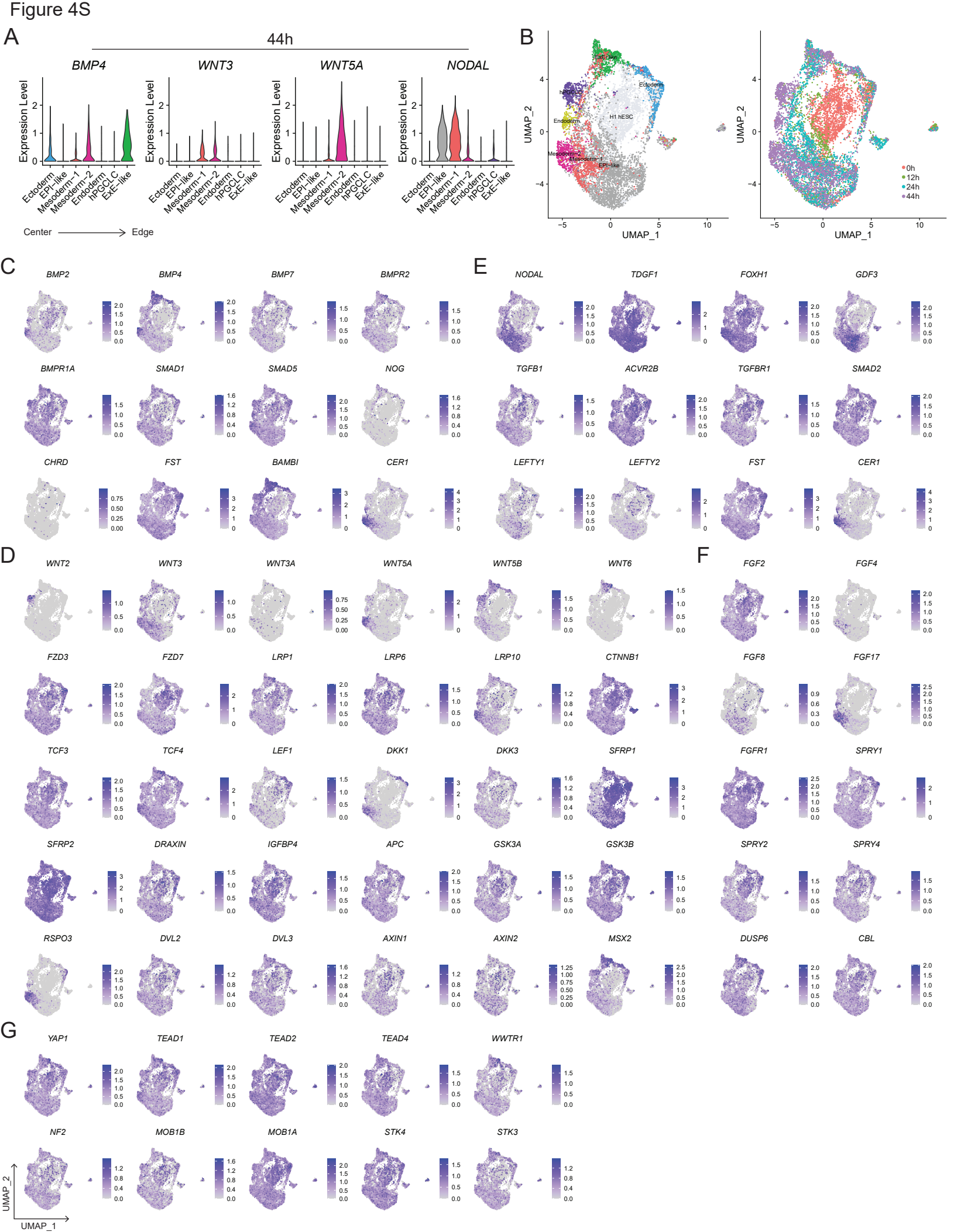

A
44h
BMP4
WNT3
WNT5A
NODAL
Expression Level
Ectoderm
EPI−like
Mesoderm−1
Mesoderm−2
Endoderm
hPGCLC
EXE−like
Center
Edge
B
UMAP_1
UMAP_2
0h
12h
24h
44h
C
BMP2
BMP4
BMP7
BMPR2
BMPR1A
SMAD1
SMAD5
NOG
CHRD
FST
BAMBI
CER1
D
WNT2
WNT3
WNT3A
WNT5A
WNT5B
WNT6
FZD3
FZD7
LRP1
LRP6
LRP10
CTNNB1
TCF3
TCF4
LEF1
DKK1
DKK3
SFRP1
SFRP2
DRAXIN
IGFBP4
APC
GSK3A
GSK3B
RSPO3
DVL2
DVL3
AXIN1
AXIN2
MSX2
E
NODAL
TDGF1
FOXH1
GDF3
TGFB1
ACVR2B
TGFBR1
SMAD2
LEFTY1
LEFTY2
FST
CER1
F
FGF2
FGF4
FGF8
FGF17
FGFR1
SPRY1
SPRY2
SPRY4
DUSP6
CBL
G
YAP1
TEAD1
TEAD2
TEAD4
WWTR1
NF2
MOB1B
MOB1A
STK4
STK3
UMAP_2
UMAP_1

Figure 4S. Supplement information related to Figure 4

(A) Expression of genes encoding indicated signaling molecules in 44h gastruloid.

(B) UMAP display of annotated gastruloid cell types at indicated time points.

(C–G) Gene expression of indicated components of BMP (B), NODAL (C), WNT (D), FGF (E), and HIPPO (F) signaling pathway in all cell types across all time points.

Figure 5S

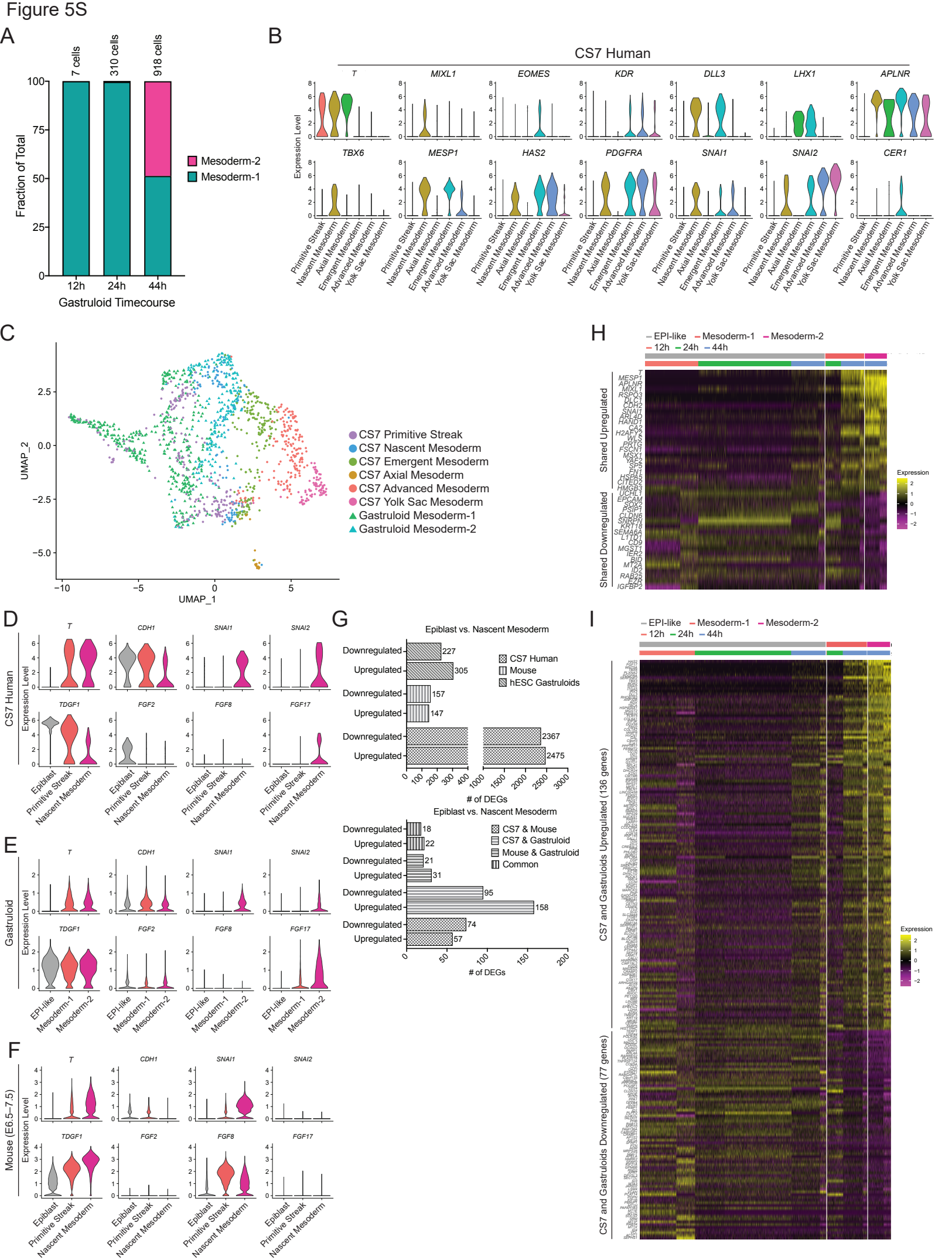

A
7 cells
310 cells
918 cells
Fraction of Total
Mesoderm-2
Mesoderm-1
12h
24h
44h
Gastruloid Timecourse
B
CS7 Human
T
MIXL1
EOMES
KDR
DLL3
LHX1
APLNR
TBX6
MESP1
HAS2
PDGFRA
SNAI1
SNAI2
CER1
Expression Level
Primitive Streak
Nascent Mesoderm
Axial Mesoderm
Emergent Mesoderm
Advanced Mesoderm
Yolk Sac Mesoderm
C
UMAP_1
UMAP_2
CS7 Primitive Streak
CS7 Nascent Mesoderm
CS7 Emergent Mesoderm
CS7 Axial Mesoderm
CS7 Advanced Mesoderm
CS7 Yolk Sac Mesoderm
Gastruloid Mesoderm-1
Gastruloid Mesoderm-2
D
CS7 Human
Expression Level
T
CDH1
SNAI1
SNAI2
TDGF1
FGF2
FGF8
FGF17
Epiblast
Primitive Streak
Nascent Mesoderm
E
Gastruloid
Expression Level
EPI-like
Mesoderm-1
Mesoderm-2
F
Mouse (E6.5–7.5)
Expression Level
G
Epiblast vs. Nascent Mesoderm
Downregulated
Upregulated
227
305
157
147
2367
2475
CS7 Human
Mouse
hESC Gastruloids
# of DEGs
Epiblast vs. Nascent Mesoderm
18
22
21
31
95
158
74
57
CS7 & Mouse
CS7 & Gastruloid
Mouse & Gastruloid
Common
# of DEGs
H
EPI-like
Mesoderm-1
Mesoderm-2
12h
24h
44h
Shared Upregulated
Shared Downregulated
Expression
I
EPI-like
Mesoderm-1
Mesoderm-2
12h
24h
44h
CS7 and Gastruloids Upregulated (136 genes)
CS7 and Gastruloids Downregulated (77 genes)
Expression

Figure 5S. Supplement information related to Figure 5

(A) Fraction of gastruloid Mesoderm-1 and -2 cells at indicated time points.

(B) Expression of indicated PS and mesoderm markers in CS7 PS and mesoderm derivatives.

(C) UMAP projection of CS7 human primitive streak and mesoderm derivatives, and gastruloid Mesoderm-1 and -2 cells.

(D–F) Violin plots showing expression of indicated mesoderm markers in transition from EPI to nascent mesoderm cells in CS7 human gastrula (D), gastruloids (E), and mouse gastrula (F).

(G) Bar plots indicating number of downregulated and upregulated genes (top) and shared downregulated and upregulated genes (bottom) in Nascent Mesoderm (gastruloid Mesoderm-2 equivalent) compared to EPI (gastruloid EPI-like equivalent) in human, mouse, and hESC gastruloids.

(H) Heatmap depicting downregulated and upregulated genes of Nascent Mesoderm shared in human, mouse, and hESC gastruloids.

(I) Heatmap depicting downregulated and upregulated genes of Nascent Mesoderm shared in human, and hESC gastruloids but not in mouse.

Figure 6S

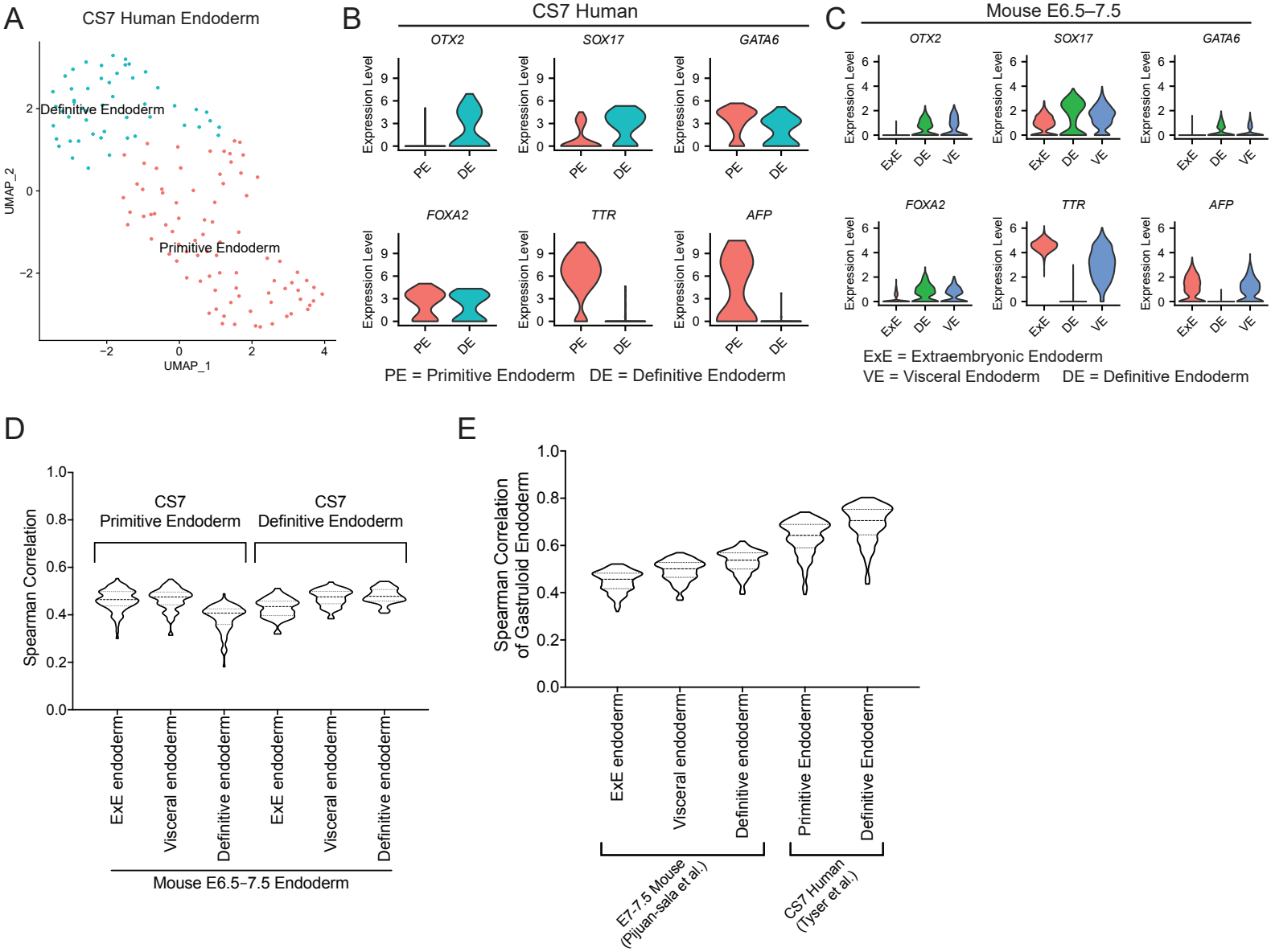

A
CS7 Human Endoderm
Definitive Endoderm
Primitive Endoderm
UMAP_1
UMAP_2
B
CS7 Human
OTX2
SOX17
GATA6
FOXA2
TTR
AFP
Expression Level
PE
DE
PE = Primitive Endoderm DE = Definitive Endoderm
C
Mouse E6.5–7.5
ExE
DE
VE
ExE = Extraembryonic Endoderm
VE = Visceral Endoderm DE = Definitive Endoderm
D
Spearman Correlation
CS7
Primitive Endoderm
CS7
Definitive Endoderm
ExE endoderm
Visceral endoderm
Definitive endoderm
Mouse E6.5–7.5 Endoderm
E
Spearman Correlation
of Gastruloid Endoderm
ExE endoderm
Visceral endoderm
Definitive endoderm
Primitive Endoderm
Definitive Endoderm
E7.5 Mouse
(Pijuan-sala et al.)
CS7 Human
(Tyser et al.)

Figure 6S. Supplemental information related to Figure 6

(A) UMAP projection of CS7 human Endoderm identifying PE and DE.

(B and C) Violin plots showing expression of indicated endoderm markers in human (B) and mouse (C).

(D) Average gene expression correlation of indicated human and mouse endoderm derivatives.

(E) Average gene expression correlation of gastruloid endoderm to indicated human and mouse endoderm derivatives.

Figure 7S

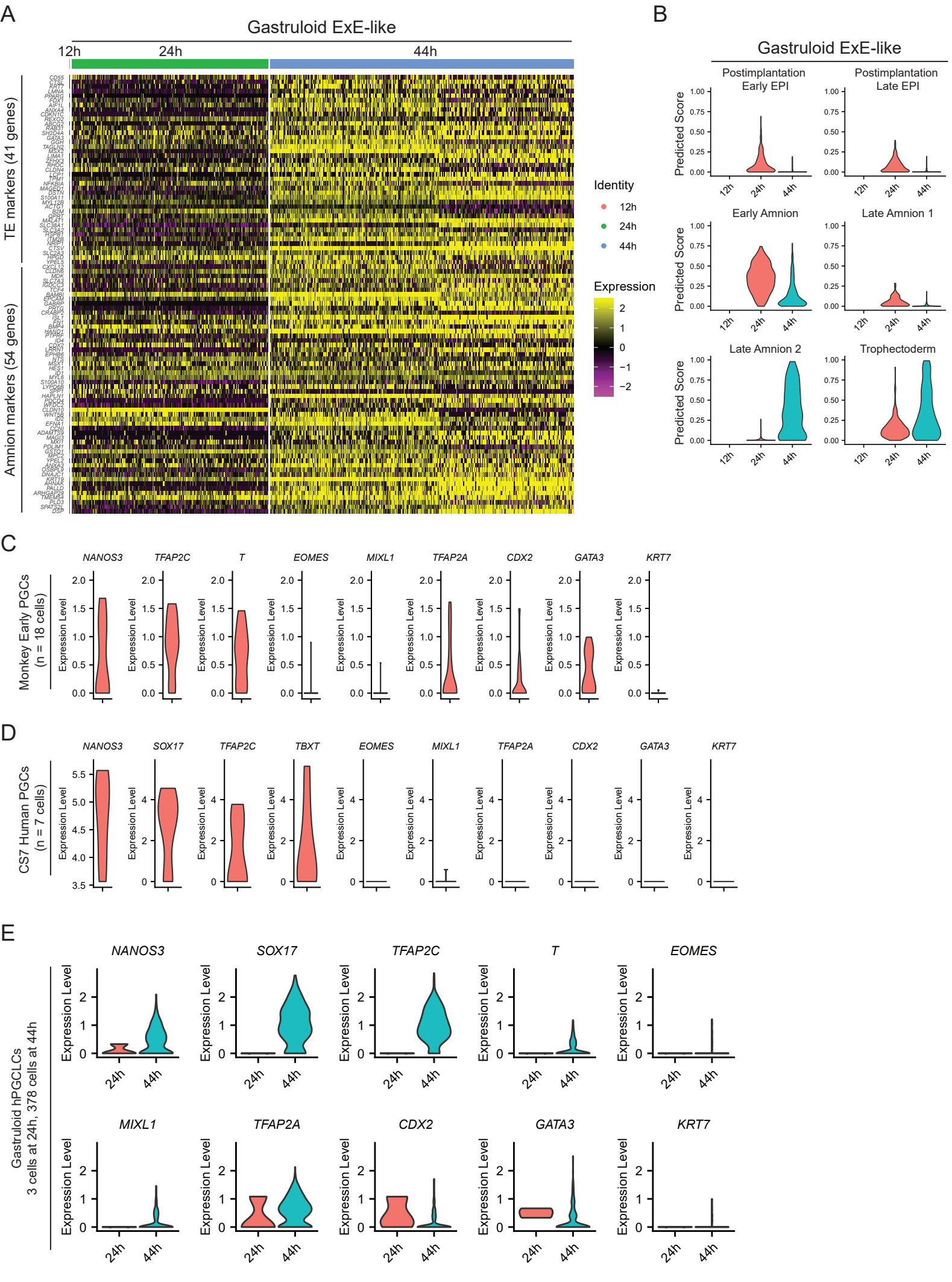

A
Gastruloid ExE-like
12h
24h
44h
TE markers (41 genes)
Amnion markers (54 genes)
Identity
12h
24h
44h
Expression
2
1
0
−1
−2
B
Gastruloid ExE-like
Postimplantation Early EPI
Postimplantation Late EPI
Early Amnion
Late Amnion 1
Late Amnion 2
Trophectoderm
Predicted Score
C
Monkey Early PGCs (n = 18 cells)
NANOS3
TFAP2C
T
EOMES
MIXL1
TFAP2A
CDX2
GATA3
KRT7
Expression Level
D
CS7 Human PGCs (n = 7 cells)
NANOS3
SOX17
TFAP2C
TBXT
EOMES
MIXL1
TFAP2A
CDX2
GATA3
KRT7
Expression Level
E
Gastruloid hPGCLCs
3 cells at 24h, 378 cells at 44h
NANOS3
SOX17
TFAP2C
T
EOMES
MIXL1
TFAP2A
CDX2
GATA3
KRT7
Expression Level
24h
44h

Figure 7S. Supplemental information related to Figure 7

(A) Heatmap showing expression of human and monkey TE markers in gastruloid ExE-like cells along the time course.

(B) Prediction score of monkey EPI, TE, and amnion in gastruloid ExE-like cells over the time course.

(C–E) Violin plot showing expression of indicated markers in monkey (C) and CS7 human (D) PGCs, and gastruloid hPGCLCs (E).