

Supplementary Information

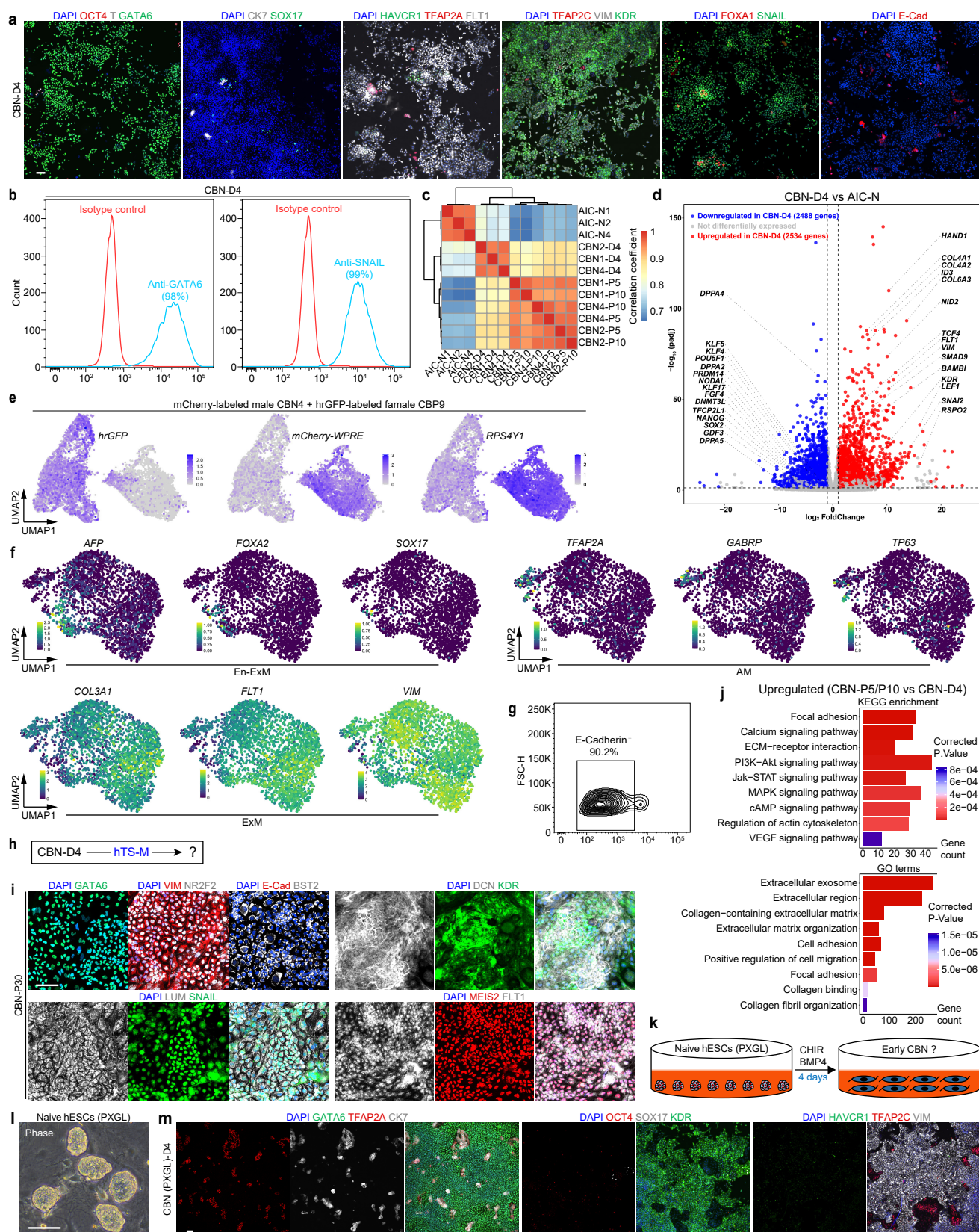
Deciphering signaling mechanisms and developmental dynamics in extraembryonic mesoderm specification from hESCs

Baohua Niu^{1, 2#}, Da Wang^{1, 2#}, Yingjie Hu^{1, 2#}, Yundi Wang^{1, 2#}, Gaohui Shi^{1, 2#},
Zhongying Chen³, Lifeng Xiang⁴, Chi Zhang^{1, 2}, Xuesong Wei^{1, 2}, Ruize Kong⁴,
Hongzhi Cai^{1, 2}, Weizhi Ji^{1, 2, 5*}, Yu Yin^{1, 2, 5*}, Tianqing Li^{1, 2, 5*}, Zongyong Ai^{1, 2, 5*}

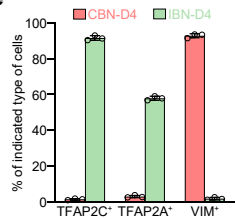
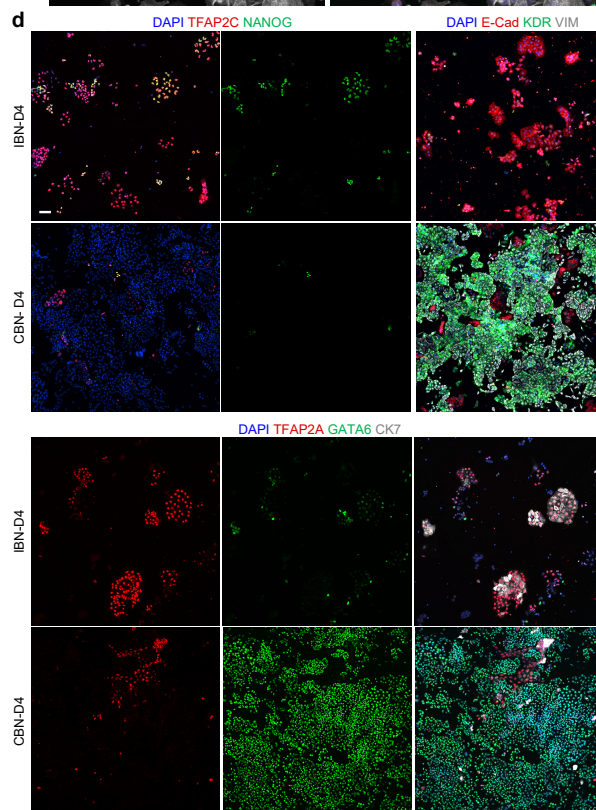
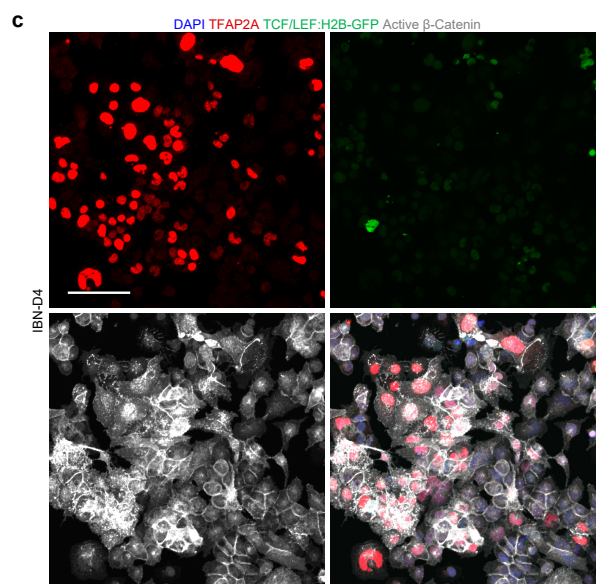
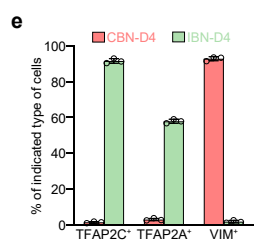
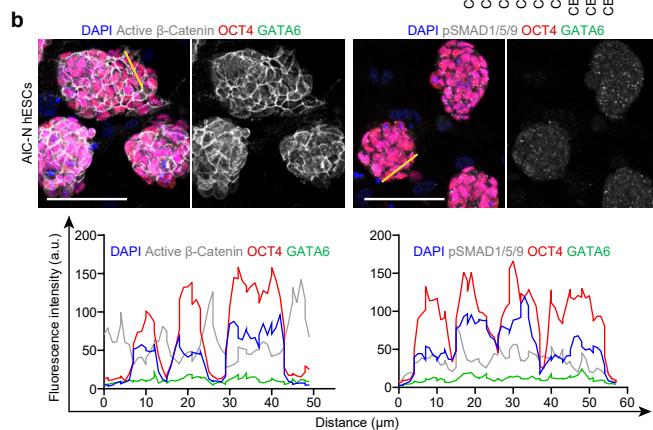
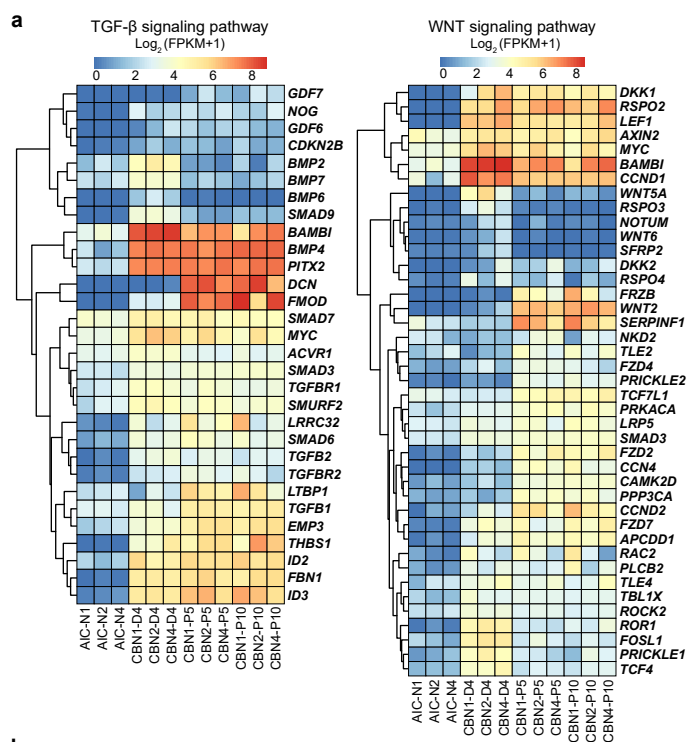
This PDF file includes:

Supplementary Figures 1 to 8

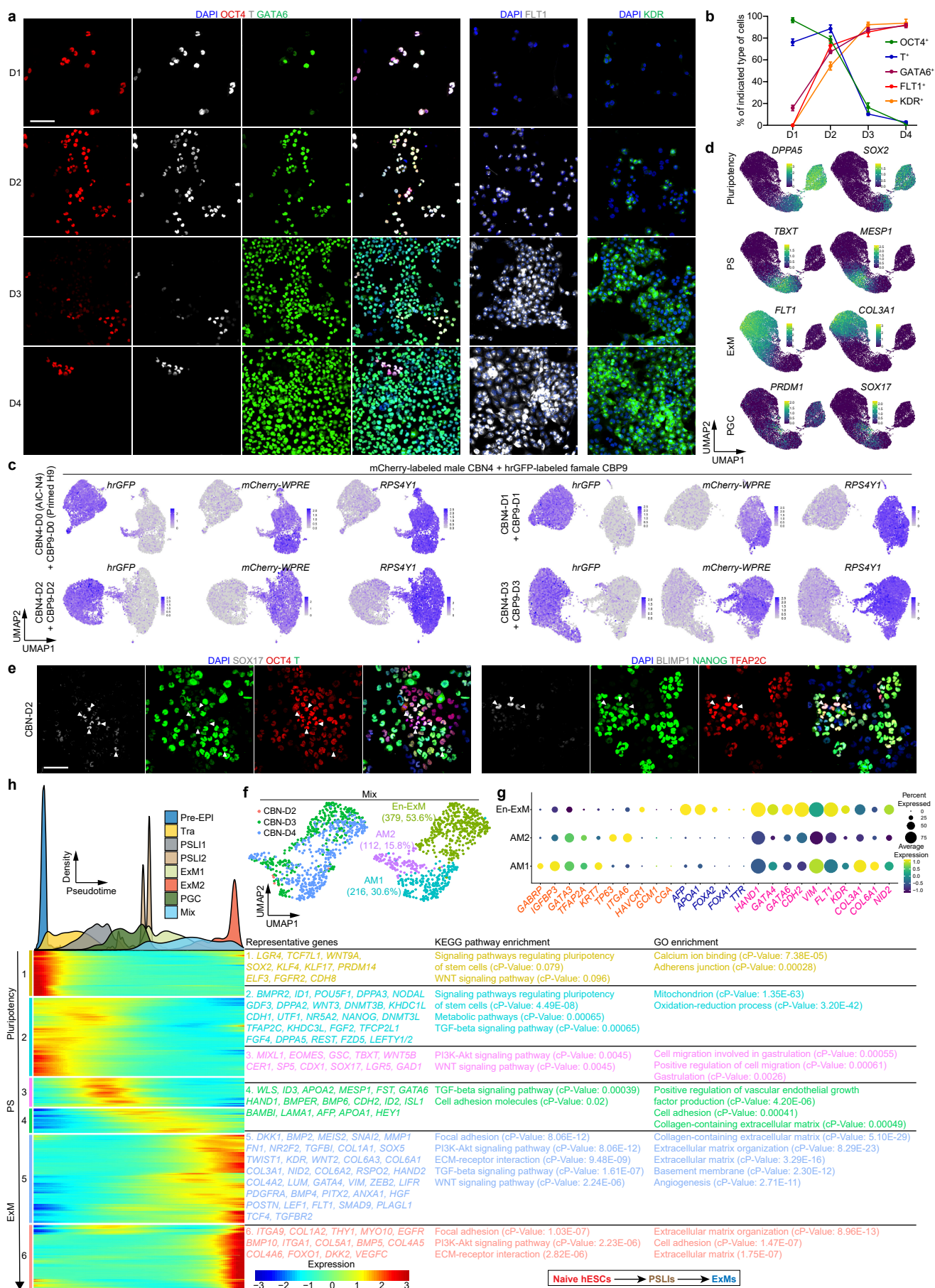
Supplementary Table 1



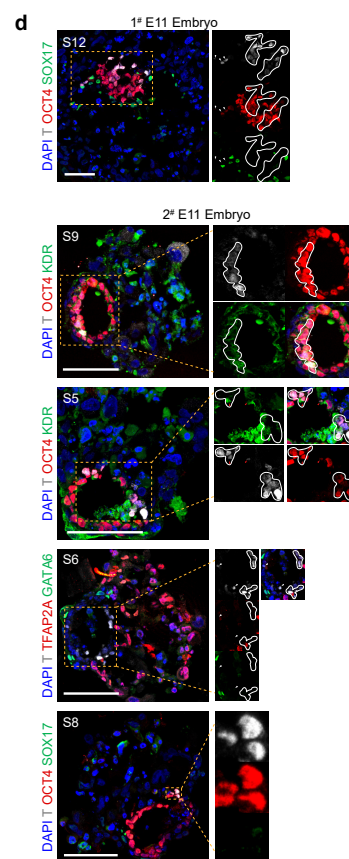
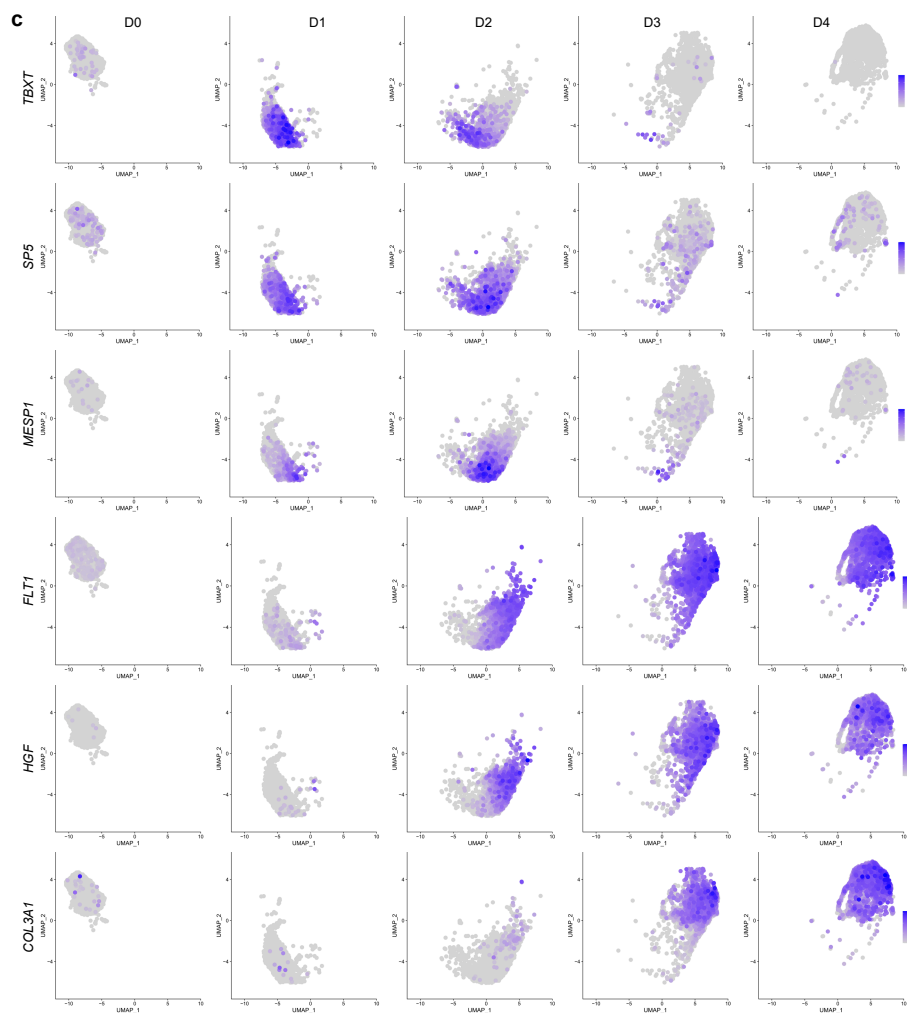
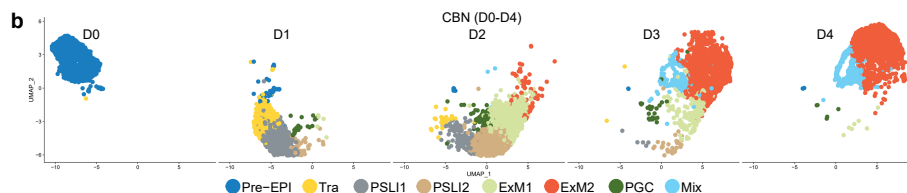
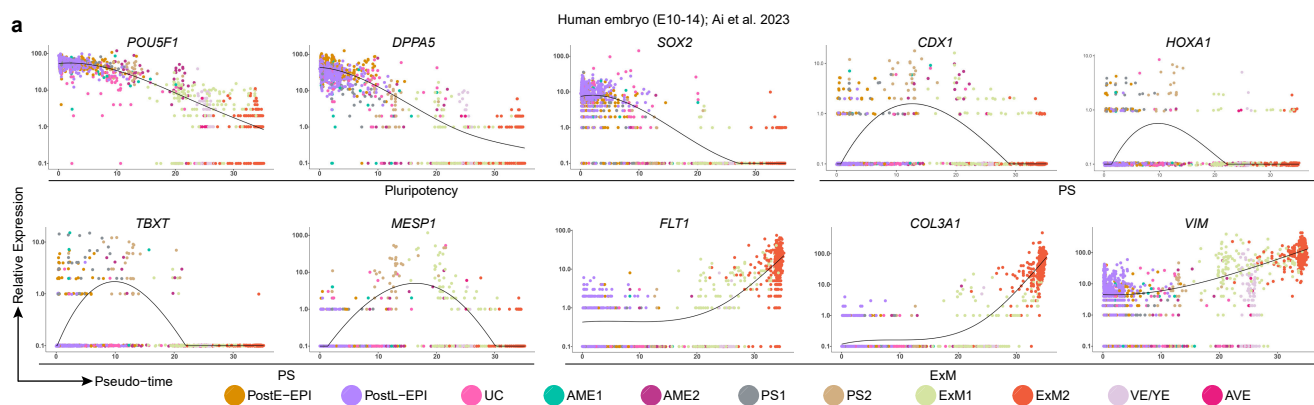
Supplementary Fig. 1 | Identification and extended culture of naive hESC-derived ExMs, related to Fig. 1. **a** Immunofluorescence (IF) staining for various embryonic and extraembryonic lineage markers in CHIR+BMP4-treated naive hESCs on day 4 (CBN-D4). **b** Flow cytometry analysis of GATA6 and SNAIL expression in CBN-D4. **c** Spearman correlation analysis of gene expression patterns in differentiation time courses for CBNs. **d** Volcano plot representing the differentially expressed genes between CBN-D4 and naive (AIC-N) hESCs, $\log_2 \text{FC} \geq 1$ or ≤ -1 , $\text{padj} < 0.05$. Two-sided Wald test, p -values were adjusted using the Benjamini-Hochberg method. **e** UMAP plots showing expressions of *hrGFP*, *mCherry-WPRE*, and Y-chromosome specific gene *RPS4Y1* according to scRNA-seq data from CB-treated AIC-N4 (CBN4-D4, male, mCherry-labeled) and primed H9 (CBP9-D4, female, hrGFP-labeled) hESCs on day 4. The fluorescent protein genes were knocked into AAVS1 site in hESCs according to the published method³. To minimize batch effects and costs in scRNA-seq, we collected the indicated sample pair as pooled sample by mixing equal cell numbers. **f** UMAP plots of the indicated genes expressed in CBN4-D4. ExM, extraembryonic mesoderm; AM, amnion; EnExM, ExM expressing specific endoderm markers. **g** Flow cytometry contour plot of CBN-D4 for E-Cadherin. **h** Schematic of extended culture of D4 CBNs in hTS-M. **i** IF staining for the indicated markers in expandable ExMs at passage 30. **j** Kyoto Encyclopedia of Genes and Genomes (KEGG) and Gene ontology (GO) analysis showing upregulated gene categories in expandable ExMs relative to D4 CBNs. Two-sided hypergeometric test, p -values were adjusted using the Benjamini-Hochberg method. **k–m** Schematic diagram (**k**), brightfield (**l**) and IF staining (**m**) images demonstrating naive (PXGL) hESC derivatives induced by CB conditions (unless otherwise specified, naive hESCs were typically cultured in AIC-N medium). All scRNA-seq data in this study are from 10× Genomics platform. Scale bars, 100 μm . Source data are provided as a Source Data file.



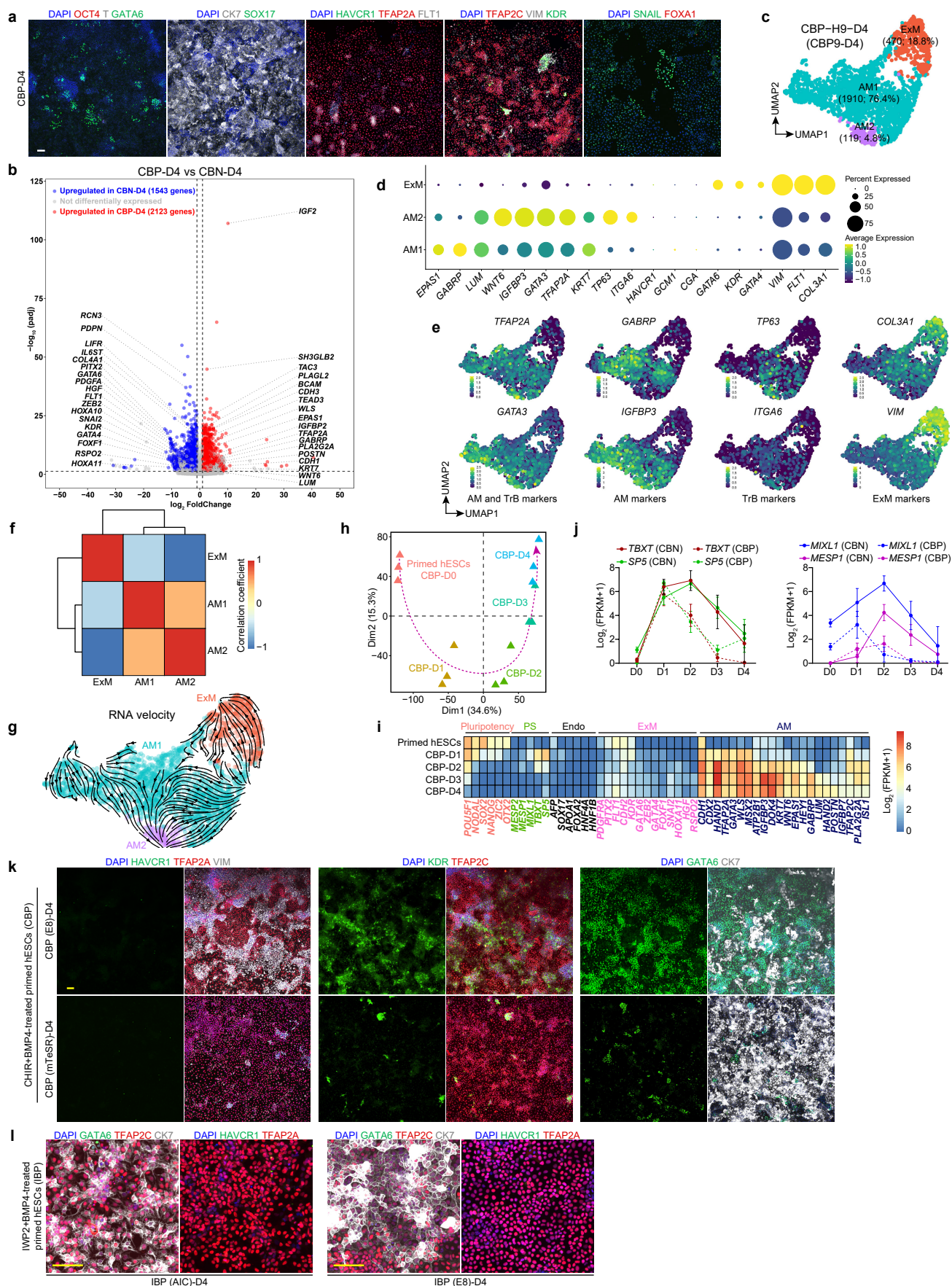
Supplementary Fig. 2 | WNT signaling governs ExM and AM fate specification in naive hESCs, related to Fig. 2. **a** Heatmaps showing the expression of genes in TGF- β and WNT signalling pathways in AIC-N hESCs and CBNs. Values represent $\log_2$ (FPKM+1). **b** IF staining (top) and fluorescence intensity profiles (bottom) for signal transducers and lineage markers in AIC-N hESCs. Yellow lines indicate planes used to plot intensity profiles of indicated markers. **c** IF staining of AIC-N2 TCF/LEF:H2B-GFP hESCs grown under IWP2+BMP4 conditions for 4 days with indicated markers. IBN-D4, IWP2+BMP4-treated AIC-N hESCs on D4. **d** IF staining for the indicated markers in IBN-D4 and CBN-D4. **e** Proportion of the indicated cell types in CBN-D4 and IBN-D4. $n = 3$ independent experiments; data are presented as mean $\pm$ SD. Scale bars, 100 μ m. Source data are provided as a Source Data file.



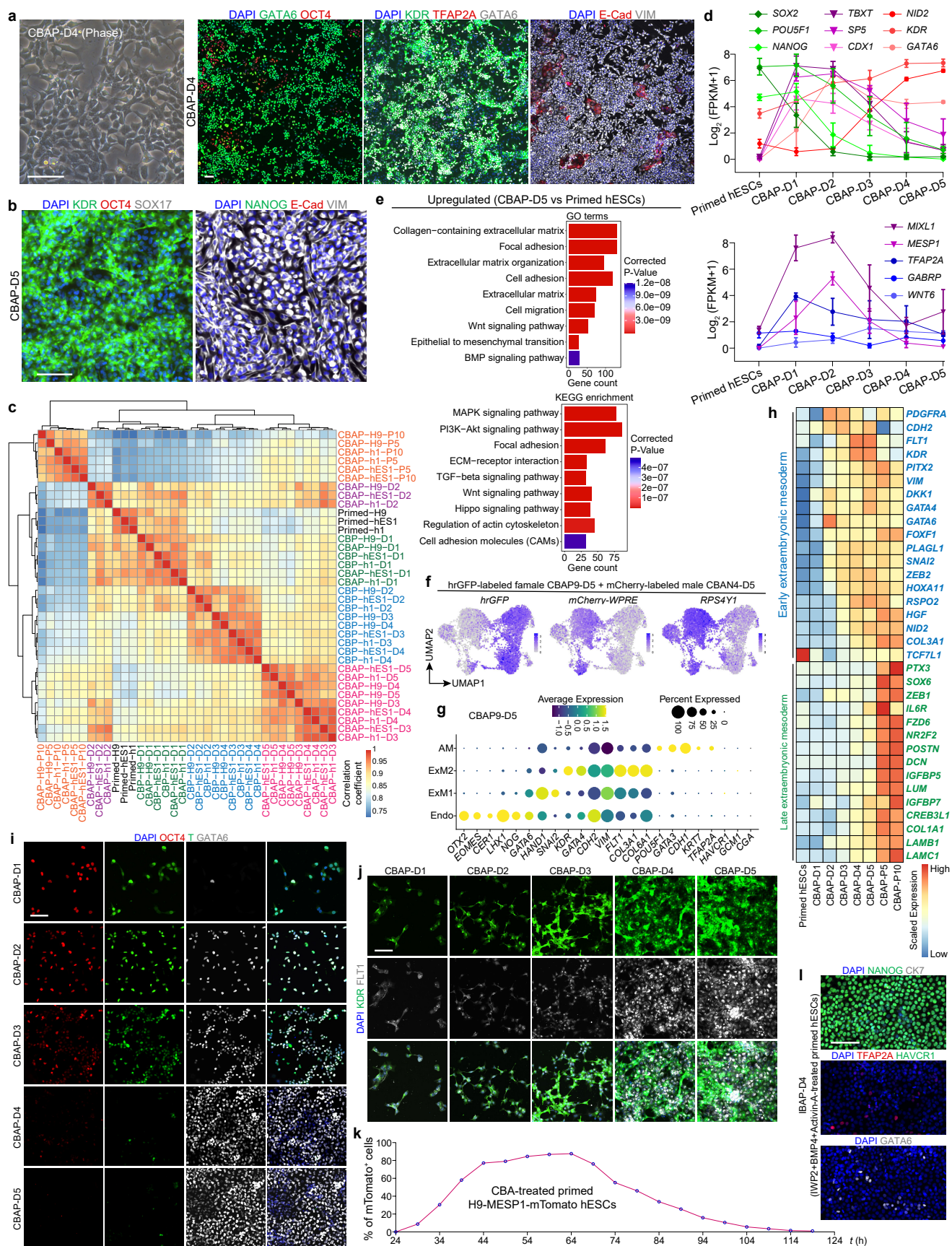
Supplementary Fig. 3 | ExM specification recapitulates the progression from naive pluripotency to PSLI to ExM, related to Fig. 3. **a** IF staining to demonstrate the dynamic expression of the indicated markers in differentiation time courses for CBNs. **b** Proportion of the indicated cell types in **a**. $n = 3$ independent experiments; data are presented as mean $\pm$ SD. **c** UMAP plots showing expressions of *hrGFP*, *mCherry-WPRE*, and *RPS4Y1* according to scRNA-seq data from differentiation time courses for mCherry-labeled male CBN4 and hrGFP-labeled female CBP9. To minimize batch effects and costs in scRNA-seq, we collected the indicated sample pair as pooled sample by mixing equal cell numbers. **d** UMAP plots of the indicated genes expressed in differentiation time courses for CBN4. PS, primitive streak; PGC, primordial germ cell. **e** IF staining for PGC markers in CBNs on day 2. Arrowheads indicate putative PGC-like cells. **f** Subclustering analysis of scRNA-seq data of Mix cluster in Fig. 3e. UMAP plot is color-coded according to cell subcluster identity annotations. **g** Dot plots of candidate genes specific for the indicated cell subtypes. **h** Pseudotime heatmap showing the top 1500 most variably expressed genes in differentiation time courses for CBNs (top). Temporal sequence schematic of appearance of primitive streak-like intermediates (PSLI) and ExMs (bottom). Color bar denotes relative expression level. Representative genes (left), KEGG (middle) and GO (right) enrichment analysis are shown. Two-sided hypergeometric test, p -values were adjusted using the Benjamini-Hochberg method. cP-Value, corrected p -value. Scale bars, 100 μ m. Source data are provided as a Source Data file.



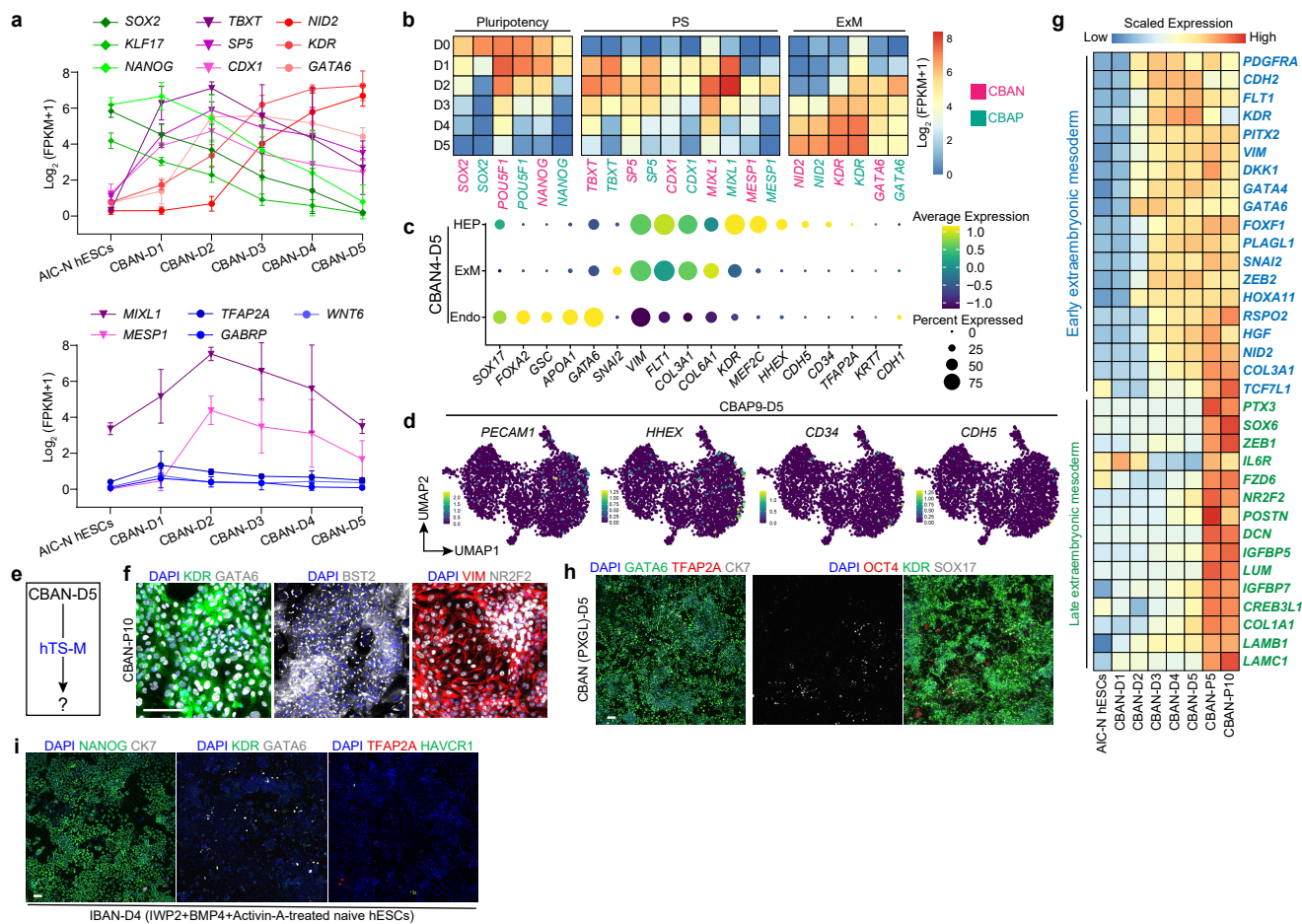
Supplementary Fig. 4 | PSLI specification precedes ExM formation in both cultured human embryos and CBNs, related to Fig. 4. **a** Dynamic expression of pluripotency, PS, and ExM marker genes in cultured human embryos from embryonic day (E10–14) over pseudotime. Each black line indicates the fitted expression trend of a gene over pseudotime. **b** Split UMAP visualization of scRNA-seq data from differentiation time courses for CBN4. **c** UMAP plots showing the dynamics of the indicated marker genes expressed in differentiation time courses for CBN4. **d** IF staining of the indicated markers in the 1[#] (top) and 2[#] (bottom) extended cultured human embryos at E11 (see also Fig. 4g). White numbers indicate section (S) numbers, white arrowheads indicate T⁺OCT4^{+/weak}SOX17⁻ (1[#] embryo, S12), T⁺OCT4⁻KDR⁻ (2[#] embryo, S5), and T⁺GATA6⁻TFAP2A⁻ (2[#] embryo, S6) PSLIs. White solid circles indicate T⁺OCT4^{+/weak}SOX17⁻ (1[#] embryo, S12), T^{weak}OCT4⁺KDR^{weak} (2[#] embryo, S5), T⁺OCT4⁺ (2[#] embryo, S5), and T⁺GATA6⁻TFAP2A⁻ (2[#] embryo, S6) PSLIs. Red arrowheads indicate T⁺OCT4⁺KDR⁺ potential ExM precursors. Scale bars, 100 μ m. Source data are provided as a Source Data file.



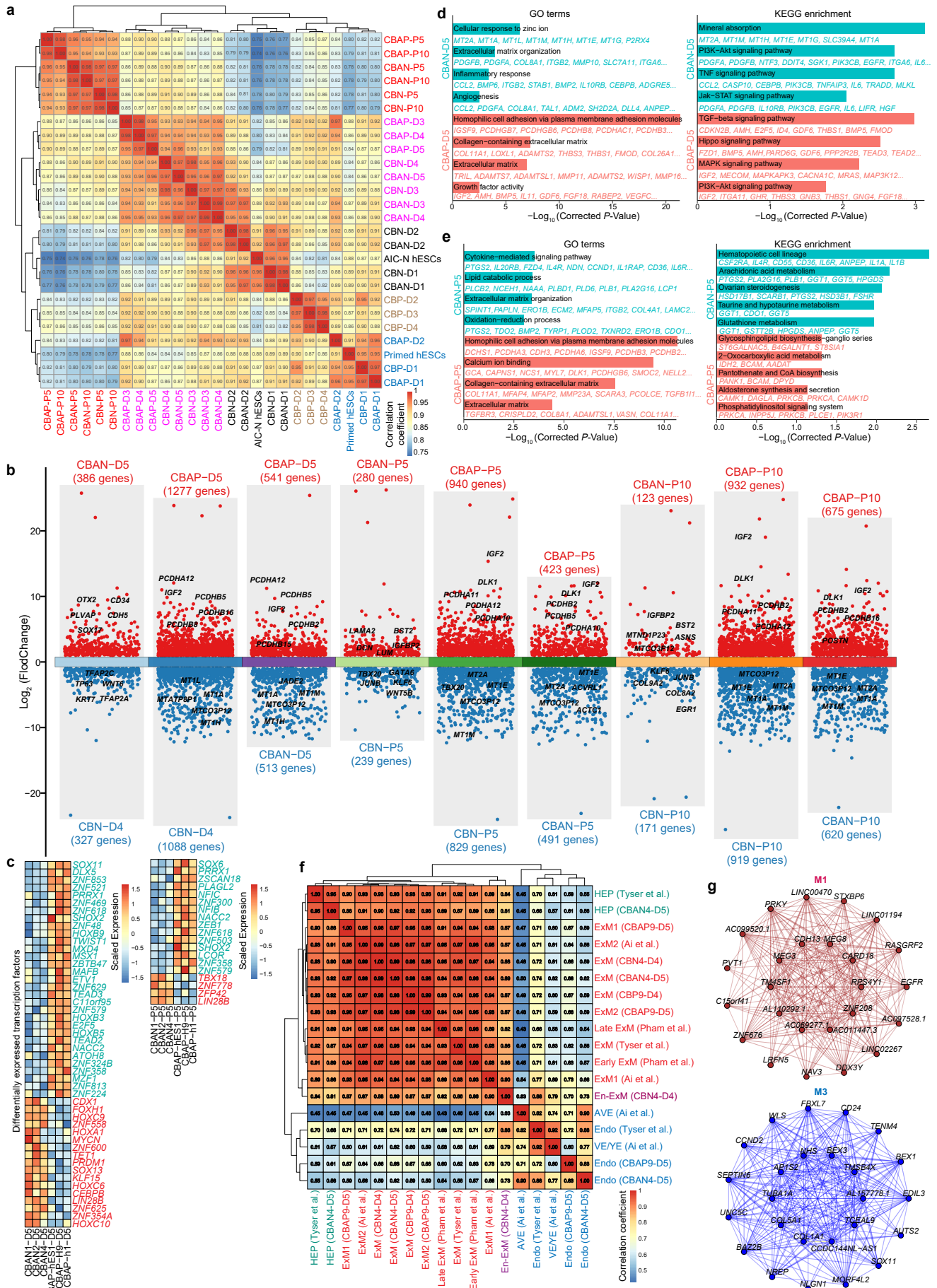
Supplementary Fig. 5 | Primed hESCs respond differently to CB conditions compared to naive hESCs, related to Fig. 5. **a** IF staining for the indicated embryonic and extraembryonic lineage markers in CB-treated primed hESCs on day 4 (CBP-D4). **b** Volcano plot representing the differentially expressed genes between CBP-D4 and CBN-D4, $\log_2 \text{FC} \geq 1$ or ≤ -1 , $p_{\text{adj}} < 0.05$. Two-sided Wald test, p -values were adjusted using the Benjamini-Hochberg method. **c** UMAP visualization of single-cell transcriptomics from CBP9-D4. UMAP plot is color-coded according to cell cluster identity annotations. **d** Dot plots of candidate genes specific for cell subtypes. **e** UMAP plots of the indicated genes expressed in CBP9-D4. **f** Spearman correlation heatmap of distinct cell subpopulations, the colors indicate the Spearman correlation coefficient. **g** RNA velocity vectors projected onto the UMAP-based embeddings of the scRNA-seq dataset shown in **c**. **h** PCA of bulk RNA-seq data from differentiation time courses for CBPs, computed using the genes with $\text{FPKM} \geq 1$ in at least one sample. **i** Heatmap of representative pluripotency and lineage marker genes in differentiation time courses for CBPs. **j** Expression dynamics of PS marker genes in differentiation time courses for CBNs and CBPs. $n = 3$ hESC lines; data are presented as mean $\pm$ SD. **k** IF staining for the indicated markers in D4 CBPs, primed hESCs were cultured in E8 (top) or mTeSR1 (bottom) medium (unless otherwise specified, primed hESCs were typically cultured in AIC medium). **l** IF staining for the indicated markers in IWP2+BMP4-treated primed hESCs on day 4 (IBP-D4), primed hESCs were cultured in AIC (left) or E8 (right) medium. Scale bars, 100 μm . Source data are provided as a Source Data file.



Supplementary Fig. 6 | Primed hESCs efficiently differentiate into ExMs via PSLIs under CBA conditions, related to Fig. 6. **a** Representative brightfield and IF staining images for the indicated markers in D4 CBAPs. **b** Representative IF staining images for the indicated markers in D5 CBAPs. **c** Spearman correlation analysis of gene expression patterns in differentiation time courses for primed hESCs under CBP and CBAP conditions. **d** Expression dynamics of the indicated marker genes in differentiation time courses for CBAPs. $n = 3$ hESC lines; data are presented as mean $\pm$ SD. **e** GO and KEGG analyses showing upregulated gene categories in D5 CBAPs relative to primed hESCs. Two-sided hypergeometric test, p -values were adjusted using the Benjamini-Hochberg method. **f** UMAP plots showing expressions of *hrGFP*, *mCherry-WPRE*, and Y-chromosome specific gene *RPS4Y1* according to scRNA-seq data from CBA-treated primed H9 (CBAP9-D5, hrGFP-labeled) and AIC-N4 (CBAN4-D5, mCherry-labeled) hESCs on day 5. To minimize batch effects and costs in scRNA-seq, we collected the indicated sample pair as pooled sample by mixing equal cell numbers. **g** Dot plots of candidate genes specific for cell subtypes in CBAP9-D5. **h** Heatmap of representative early and late ExM marker genes in differentiation time courses for CBAPs. Values represent $\log_2$ (FPKM+1) scaled by gene expression across samples. **i, j** IF staining to demonstrate the dynamic expression of the indicated markers in differentiation time courses for CBAPs. **k** The dynamics of the proportion of mTomato⁺ cells in the differentiation time courses for MESP1-mTomato knock-in reporter CBAPs, constructed using data from Supplementary Movies 1, 2. **l** IF staining for the indicated markers in IWP2+BMP4+Activin-A-treated primed hESCs on day 4 (IBAP-D4). Scale bars, 100 μ m. Source data are provided as a Source Data file.



Supplementary Fig. 7 | Naive hESCs efficiently differentiate into ExMs via PSLIs under CBA conditions, related to Fig. 7. a Expression dynamics of the indicated marker genes during differentiation time courses for CBANs. $n = 3$ hESC lines; data are presented as mean $\pm$ SD. **b** Heatmap of representative pluripotency, PS, and ExM marker genes during differentiation time courses for CBANs and CBAPs. **c** Dot plots of candidate genes specific for cell subtypes in CBAN4-D5. Endo, endoderm. **d** UMAP plots of the hematopoietic marker genes expressed in CBAP9-D5. **e** Schematic of CBAN-D5 inoculated into hTS-M for extended culture. **f** IF staining images for the indicated markers in expandable CBANs at passage 10. **g** Heatmap depicting representative early and late ExM marker gene expression across differentiation time courses for CBANs. Values represent $\log_2$ (FPKM+1) scaled by gene expression across samples. **h** IF staining images for the indicated markers in CBAN-D5, naive hESCs were cultured in PXGL medium. **i** IF staining for the indicated markers in IWP2+BMP4+Activin-A-treated naive hESCs on day 4 (IBAN-D4). Scale bars, 100 μ m. Source data are provided as a Source Data file.



Supplementary Fig. 8 | Naive and primed hESC-derived ExMs exhibit both shared and distinct transcriptional characteristics, related to Fig. 8. **a** Spearman correlation analysis of gene expression patterns during differentiation time courses for naive and primed hESCs under CBP and CBAP conditions. **b** Differential gene expression analysis across nine different comparison groups. $p\text{-adj} < 0.05$ and the y-axis represents the $\log_2$ fold change. Two-sided Wald test, p -values were adjusted using the Benjamini-Hochberg method. **c** Heatmap of the differentially expressed transcription factors between CBANs and CBAPs at different time points (D5/P5). $\log_2(\text{FC}) \geq 1$, $p\text{-adj} < 0.05$, and $\text{FPKM} \geq 3$ in at least two samples. Values represent $\log_2(\text{FPKM}+1)$ scaled by gene expression across samples. Two-sided Wald test, p -values were adjusted using the Benjamini-Hochberg method. **d, e** Bar plots displaying the GO and KEGG analyses of differentially expressed genes between CBANs and CBAPs on D5 (**d**) and at P5 (**e**). Two-sided hypergeometric test, p -values were adjusted using the Benjamini-Hochberg method. **f** Spearman correlation heatmap of different cell clusters, the colors indicate the Spearman correlation coefficient. **g** Co-expression network plots for M1 and M3 in Fig. 8h. M, module. Source data are provided as a Source Data file.

Supplementary Table 1: A list of antibodies used in this study

Antibody information for flow cytometry					
Antibody (species/isotype)	Source	Cat#	Lot	Dilution	
GATA6 (D61E4) XP® Rabbit mAb (PE Conjugate)	Cell Signaling Technology	26452	4	1:600	
Rabbit (DA1E) mAb IgG XP® Isotype Control (PE Conjugate)	Cell Signaling Technology	5742	17	1:600	
Goat Anti-Human SNAIL	R&D Systems	AF-3639	XRS021707 1	1:400	
Goat IgG Isotype Control	Thermo Fisher Scientific	31245	0960621	1:400	
APC anti-human CD324 (E-Cadherin) Antibody	Biolegend	324108	B379257	1:100	
APC Mouse IgG1, κ Isotype Ctrl (FC) Antibody	Biolegend	400122	B379803	1:100	
Recombinant Anti-Transcription factor AP-2-alpha antibody [EPR2688(2)]	Abcam	ab108311	1014077-1	1:200	
Recombinant Rabbit IgG, monoclonal [EPR25A] - Isotype Control	Abcam	ab172730	1011223-6	1:200	
Alexa Fluor 555 donkey anti-rabbit IgG	Thermo Fisher Scientific	A-31572	1866859	1:1000	
Alexa Fluor 488 donkey anti-goat IgG	Thermo Fisher Scientific	A-11055	1827671	1:1000	

Antibody information for immunofluorescence staining					
Antibody (species/isotype)	Source	Cat#	Lot#	Dilution	
NANOG (goat IgG)	R&D Systems	AF1997	KKJ0922031	1:400	
OCT4 (mouse IgG)	Santa Cruz	sc-5279	C3121	1:600	
BLIMP1 (rabbit IgG)	Cell Signaling Technology	9115S	6	1:100	
MEIS2 (mouse IgM)	Novus Biologicals	H00004212-M01	M3011-EH4	1:200	
E-Cadherin (mouse IgG)	Abcam	ab76055	GR3360021-7	1:200	
BST2 (rabbit IgG)	Abcam	ab243230	1022489-8	1:200	
Active β -Catenin (rabbit IgG)	Cell Signaling Technology	19807S	4	1:400	
Phospho-Smad1/Smad5/Smad9 (rabbit IgG)	Cell Signaling Technology	13820S	3	1:400	
GATA6 (goat IgG)	R&D Systems	AF1700	KWT0523031	1:600	
GATA6 (rabbit IgG)	Cell Signaling Technology	5851S	5	1:600	
SOX17 (goat IgG)	R&D Systems	AF1924	KGA0815042 2	1:600	
SOX17 (rabbit IgG)	Cell Signaling Technology	81778S	1	1:500	
CK7/KRT7 (rabbit IgG)	Abcam	ab181598	GR3321316-1 7	1:500	
NR2F2 (rabbit IgG)	Abcam	ab211776	1006695-11	1:200	
TFAP2C (mouse IgG)	Santa Cruz	sc-12762	D2717	1:400	
Brachyury/T (goat IgG)	R&D Systems	AF2085	KQP0719121	1:400	
Brachyury/T (rabbit IgG)	Cell Signaling Technology	81694S	1	1:800	
HAVCR1 (goat IgG)	R&D Systems	AF1750	GTB0618111	1:200	
TFAP2A (mouse IgG)	Santa Cruz	sc-12726	L2619	1:400	

SNAIL (goat IgG)	R&D Systems	AF-3639	XRS0217071	1:400
VIMENTIN (mouse IgG)	eBioscience	14-9897	2297463	1:2000
VIMENTIN (rabbit IgG)	Abcam	ab137321	GR3425580-7	1:1000
CD34 (rabbit IgG)	Abcam	ab81289	1071545-11	1:100
CD31 (mouse IgG)	Abcam	ab9498	1035700-12	1:100
LUM (rabbit IgG)	Invitrogen	MA5-29402	35FC7A02	1:100
DCN (rabbit IgG)	Abcam	ab151988	GR122269-56	1:100
FLT1 (rabbit IgG)	Abcam	ab32152	GR3321125-1 7	1:200
FOXA1 (mouse IgG)	Abcam	ab55178	GR3448710-4	1:100
KDR (goat IgG)	R&D Systems	AF357	CUE0620061	1:400
Alexa Fluor 488 donkey anti-rabbit IgG	Jackson ImmunoResearch	711-545-152	139808	1:600
Alexa Fluor 568 donkey anti-mouse IgG	Thermo Fisher Scientific	A-10037	1827879	1:600
Alexa Fluor 488 donkey anti-mouse IgG	Thermo Fisher Scientific	A-21202	1820538	1:600
Alexa Fluor 647 donkey anti-rabbit IgG	Thermo Fisher Scientific	A-31573	1874788	1:600
Alexa Fluor 555 donkey anti-rabbit IgG	Thermo Fisher Scientific	A-31572	1866859	1:600
Alexa Fluor 647 donkey anti-goat IgG	Thermo Fisher Scientific	A-21447	1841382	1:600
Alexa Fluor 488 donkey anti-goat IgG	Thermo Fisher Scientific	A-11055	1827671	1:600