A single cell trajectory of human archetypal pluripotent stem cell differentiation to trophoblast stem cells reveals induction of endogenous BMP5/7 and GATA3 without transitioning through a naive state

Authors:
Ethan Tietze[a,#], Andre Rocha Barbosa[a,b,c,#], Bruno Henrique Silva Araujo[a,#], Veronica Euclydes[a,b,h,#], Hyeon Jin Cho[a], Yong Kyu Lee[a], Arthur Feltrin[a,d], Bailey Spiegelberg[a,g], Alan Lorenzetti[a], Joyce van de Leemput[a,$], Pasquale Di Carlo[a,i], Tomoyo Sawada[a,b], Gianluca Ursini[a,e], Kynon J. Benjamin[a,b,f], Helena Brentani[c,h], Joel E. Kleinman[a,f], Thomas M. Hyde[a,b,f], Daniel R. Weinberger[a,b,e,f,g], Ronald McKay[a,e], Joo Heon Shin[a,b], Apua C.M. Paquola[a,b], and Jennifer A. Erwin[a,b,e,*]

[a] Lieber Institute for Brain Development, Baltimore, Maryland, USA
[b] Department of Neurology, Johns Hopkins School of Medicine, Baltimore, Maryland, USA
[c] Inter-institutional Graduate Program on Bioinformatics, University of São Paulo, São Paulo, SP, Brazil
[d] Center for Mathematics, Computation and Cognition, Federal University of ABC, Santo André, SP, Brazil
[e] Department of Neuroscience, Johns Hopkins School of Medicine, Baltimore, Maryland, USA
[f] Department of Psychiatry & Behavioral Sciences, Johns Hopkins University School of Medicine, Baltimore, MD, USA
[g] Department of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA
[h] Department of Psychiatry, University of Sao Paulo, Medical School, São Paulo, Brazil
[i] Department of Basic Medical Science, Neuroscience, and Sense Organs – University of Bari Aldo Moro, Bari, Italy

[$] Present address: Center for Precision Disease Modeling and Division of Endocrinology, Diabetes and Nutrition, Department of Medicine, University of Maryland School of Medicine, 670 West Baltimore Street, Baltimore, MD, 21201, USA

[#]Equal contributor

[*]Lead contact and corresponding author: Jennifer.Erwin@libd.org

Author Contributions: JE and ET. conceptualized the study; JE, ET, ARB, BHSA and VE. wrote the original draft; JE, ET, ARB, BHSA, VE, HJC, YKL, AF, JL, PDC, TS, and KJB performed experiments and data analysis; HB, JEK, TMH, BGS, DRW, GU, RM, JHS, ACMP, JE were responsible for writing, reviewing, and editing of the manuscript; JE coordinated the study.

Competing Interest Statement: The authors declare no competing interests.

Classification: Biological Science and Developmental Biology.

Keywords: Placenta, Trophoblast, Trophectoderm, Primed iPSC, BMP.

This PDF file includes:

Main Text
Figures 1 to 6

Abstract

The human placenta is increasingly a focus of research related to early child development and the impact of maternal hyperimmune states. Primary human trophoblast stem cells (hTSC) and human pluripotent stem cells (hPSC) differentiated to hTSC can potentially model placental processes *in vitro*. Yet, it remains controversial how the differentiation of human pluripotent stem cells to trophectoderm relates to *in vivo* development and the factors required for this differentiation. Here, we demonstrate that the primed pluripotent state retains potency to generate trophoblast stem cells by activating EGF and WNT and inhibiting TGFb, HDAC and ROCK signaling without exogenous BMP4 (named TS). We map this specification by temporal single cell RNAseq compared to activating BMP4 or activating BMP4 and inhibiting WNT. TS conditions generate a stable proliferating cell type that is highly similar to six-week placental cytotrophoblasts with activation of endogenous retroviral genes and without amnion expression. Multiple primed iPSC and ES lines differentiate to iPS-derived-TSCs that can be passaged for at least 30 passages and differentiate to pure populations of multinucleated syncytiotrophoblasts and extravillous trophoblast cells. Our findings establish that primed iPS cell specification to hTSC with TS conditions involves induction of *TMSB4X*, *BMP5/7*, GATA3 and *TFAP2A* without transitioning through a naive state. Collectively, our results suggest that the primed state is on a continuum of potency and can differentiate to trophoblast stem cells via multiple paths.

Significance Statement

In the present study, we map the specification of primed induced pluripotent stem cells to trophoblast stem cells (TSC). Primed iPS-derived-TSC share transcriptional, morphological and functional characteristics with human *ex vivo* cytotrophoblasts including capacity of self-renewal and the ability to differentiate to pure extravillous and syncytiotrophoblasts. iPS-derived TSC display a uniquely active transcriptional network of human endogenous retroviruses similar to *in vivo* trophoblast. In addition, the fast conversion of primed iPSC to TSC allows for modeling placental diseases from large pluripotent stem cell cohorts which are traditionally banked at the primed state. Collectively, our results suggest that the primed state is on a continuum of potency which can differentiate to trophoblast stem cells via multiple paths.

Main Text

Introduction

Human models are essential to study human placenta development and function in health and disease. Comparative genome analysis demonstrates that genes involved in placenta and reproduction are amongst the most divergent between human and mouse [1]. For the human placenta, which is increasingly implicated as a risk factor for developmental disorders associated with a maternal hyperimmune state [2], the ability to model human trophectoderm *in vitro* is a powerful platform to identify disease-associated molecular targets, perform phenotypic screens and evaluate toxicity. *In vitro* and *in vivo*, trophoblast stem cells (TSC) are proliferative stem cell precursors that can differentiate into the trophectoderm cell types syncytiotrophoblasts and extravillous cytotrophoblasts [3].

Questions for how to appropriately model disorders in hTSC using hPSC cohorts exist [4]. Cohorts of iPSCs derived from patients can be used to model disease processes and genetically encoded traits. iPSCs cohorts and large iPSC banks are almost universally grown in primed conditions. Thus, the initial stage of most iPSC cohort studies will begin with cells in the primed state.

Conflicting results raise questions about the developmental potential and specification processes of hPSC to TSC [4]. Multiple culture conditions direct human stem cells to various developmental stages, including “naive” conditions which achieve a pre-implantation-like epiblast state [5, 6, 7] and archetypical “primed” conditions which achieve a post implantation-like epiblast state. It is widely accepted that the extraembryonic trophoblast lineage is specified before implantation. Therefore the differentiation of primed and naive pluripotent stem cells to trophectoderm does not normally occur *in vivo*. Questions remain about the heterogenous pluripotent states of archetypical primed PSC and the competency of these states to give rise to trophoblast cells [4]. Some studies suggest that BMP-treated primed hPSC resemble amnion and not trophoblast, while several more recent studies conclusively demonstrate that primed hPSCs can generate *bona fide* TSC with BMP4 treatment [8,9] or TGFb inhibition [10]. These differing reports may be reconciled by the fact that amnion specification in humans may be highly similar to trophectoderm [11] and that non-optimized culture conditions caused conflicting results [10].

Here, we sought to define the developmental trajectory of primed iPSC conversion to trophoblast stem cells and the involvement of exogenous BMP4 and WNT. The TS state can be reached via multiple paths. Four factor reprogramming of human fibroblasts cultured in either primed or naive conditions leads to upregulation of trophectoderm programs, and trophoblast stem cell lines can be derived from both naive and primed reprogramming intermediates [12]. Activn/Nodal, FGF, IGF and WNT signaling regulate pluripotency in human primed PSCs, and manipulating these pathways can induce primed PSC specification to trophectoderm and other embryonic lineages [13]. Bone Morphogenetic Protein 4 (BMP4) induces differentiation of human primed PSCs to trophoblast-like cells [14]. In the presence of BMP4, Inhibition of FGF and activin/nodal signaling [15, 16] or inhibition of WNT signaling generates more uniform trophectoderm [17] and trophoblast stem cells [8]. The differentiation in the presence of BMP4 follows a trajectory through a naive state [8]. Trophoblast can also be generated from human primed iPSC by activation of WNT and EGF while inhibiting Transforming Growth Factor Beta (TGFb), histone deacetylase (HDAC) and Rho-associated protein kinase (ROCK) [3], which is enhanced in the presence of BMP4 or by removing WNT activation [10].

In the present study, we find that activation of WNT and EGF while inhibiting TGFb, HDAC and ROCK (TS condition) in the absence of exogenous BMP4 specifies multiple primed pluripotent stem cell lines to TSC. We demonstrate that primed-derived-TSCs are capable of self-renewal for at least 30 passages and

can differentiate into syncytiotrophoblast and extravillous trophoblasts. We used temporal single-cell sequencing analysis to elucidate the trajectory of this specification. Trophoblast cells generated by TS condition are transcriptionally highly similar to *ex vivo* cytotrophoblasts and uniquely upregulate endogenous retroviral genes. The specification employs established trophectoderm programs including *TMSB4X*, *YAP*, *GATA3, CDX2* and *TFAP2A* and interestingly induces endogenous *BMP5* and *BMP7* expression, indicating a role for endogenous BMP signaling even in the absence of exogenous BMP4 treatment.

Results:

Derivation of proliferative trophoblast-like stem cells (TSC) from primed human embryonic stem cells and primed induced pluripotent stem cells without BMP4

We asked if primed ESC and iPSC can differentiate to TSC without exogenous BMP4 in the conditions used to generate trophoblast stem cells from blastocysts and human villus cytotrophoblasts [3], here referred to as TS which activates EGF and WNT and inhibits TGFb, HDAC and ROCK signaling. While previously described primed iPSC differentiations to trophectoderm have added BMP4, evidence has shown that iPSC can differentiate to trophectoderm in the absence of exogenous BMP4 [18]. 24 hours after clump passage of the human embryonic stem cell line H1 [19] in primed feeder-free stem cell conditions using StemFlex, ES colonies demonstrated tight packing with defined edges characteristic of primed cells (Figure 1a and b). Once self organized colonies were established 24 hours post passage, cells were switched directly to TS condition media. By brightfield imaging over the subsequent 6 days, we observed that cells proliferated rapidly and adopted a flatter appearance with some cells adopting a cobblestone appearance (Figure 1b-d). After subsequent passaging, two morphologically distinct populations emerged. Circular colonies with an epithelial-like appearance were surrounded by phase bright fibroblastic cells (Figure 1e,f). The inner epithelial-like cells continued to proliferate and appeared morphologically similar to TSC derived from blastocysts and human villus cytotrophoblasts (Figure 1g,j,l and n). Immunofluorescence revealed that subpopulations of the inner cells expressed TP63 and/or KRT7 and minimally expressed VIM while surrounding fibroblastic cells strongly expressed VIM (Figure 1h). After several passages, the epithelial-like cells preferentially expand and reach purity between passage 5-7. To confirm the ability to proliferate as TSC-like cells, iPS-derived trophoblast stem cell lTSC lines were passed for up to 32 passages and maintained KRT7 and TP63 expression (Figure 1i). These TSC widely express both TP63 and KRT7 (Figure 1j-o, Figure S1) without expressing the pluripotency genes SOX2 and NANOG (Figure S1k-n). In addition, the majority of cells highly express ITGA6 (89%), a common cell surface marker that is present in cytotrophoblast and human trophoblast stem cells [3] (Figure 1p).

Pluripotent stem cell lines vary in their propensity to differentiate to different cell types [20]. Therefore, we asked if TS condition could specify a variety of primed human stem cell lines from different sources. We confirmed TSC specification by VIM, KRT7 and TP63 immunofluorescence on iPSC lines reprogrammed from dermal fibroblasts by sendai, named 2014.06 (Supplemental Figure 2 and Figure S1k-r), and two postmortem dura fibroblast-derived iPS cell lines reprogrammed by episome, named LIBD1c8 and LIBD7c6 [21] (Figure 1n and o and S1a-j).

Differentiation of Trophoblast Cells into multinucleated syncytiotrophoblasts and extravillous trophoblast cells

173 To investigate the differentiation potential of primed-derived-TSC, we evaluated the ability of primed-
174 derived-TSC to differentiate to mature trophectoderm cell types. Proliferative trophoblast stem cells are
175 bipotential stem cells with the capacity to differentiate into syncytiotrophoblast (STB) and extravillous
176 trophoblasts (EVT) [3]. EVT differentiation of the primed-derived-TSC H1 line was induced with the TGFP-
177 B inhibitor A83, NRG1 and Matrigel, as previously described for TSC (Figure 2a). After 9 days in EVT
178 differentiation, primed-derived-TSC acquired EVT morphology (Figure 2b) and 88% of cells were positive
179 for the EVT specific marker HLA-G by immunofluorescence staining (Figure 2c,d), thereby resembling
180 previously described derived human TSC [3].
181
182 To confirm differentiation to syncytiotrophoblast (STB), we induced the Primed-derived-TSC differentiation
183 to STB with Forskolin for 3 days as previously described [3]. STB is a terminally differentiated
184 multinucleated epithelial layer that infiltrates the maternal endometrium. The multinucleated cell forms
185 from multiple cytotrophoblast cell fusions. Cells demonstrate STB morphology of multinucleation (Figure
186 2e) and the fusion index revealed that more than 80% of the cells were multinucleated (Figure 2f), similar
187 to STB derived trophoblast cells from human villous cytotrophoblast and from the blastocyst outgrowth [3]
188 and naive TSC [22]. Primed-TSC-derived STB expressed STB specific proteins hCGβ and syndecan-1
189 (SDC1) (Figure 2g-j), which were not detected in iPS cells (Supplemental Figure 1i). In summary, primed-
190 derived-TSC are bipotential stem cells capable of efficient differentiation into both HLA-G positive villus
191 cytotrophoblasts and hCGβ and SDC1 positive syncytiotrophoblasts.
192
193 **Single cell RNA sequencing reveals unique transcriptional programs for primed iPSC TSC**
194 **specification in the absence of BMP4**
195
196 To define the molecular events involved in the specification of primed PSC to trophectoderm, we
197 performed temporal single cell RNAseq analysis of human iPSC specified the TSC using the previously
198 described TS condition which activate EGF and WNT and inhibit TGFb, HDAC and ROCK signaling. Due
199 to previous reports of BMP4 and Wnt involvement, we compared this specification to BMP4 treatment [14]
200 and BMP4 and inhibition of WNT with the small molecule inhibitor IWP2 [17]. 24 hours after clump
201 passaging with EDTA in StemFlex media, the media was changed to differentiation conditions. Single
202 cells were sequenced at iPS cell stage before differentiation (day 0), at day 4 and day 6 after BMP4
203 treatment; on days 2, 4 and 6 for BMP4 + IWP2 treatment, and on days 2, 4, 6 and 8 for TS condition,
204 resulting in 10 single-cell transcriptomes (Figure 3a). In TS condition, the day 8 is a timepoint two days
205 after the first passage annotated as passage 1 (p1). In the BMP4 and BMP4 + IWP2 conditions, no
206 proliferative cells survived passaging, therefore the p1 time point was not sequenced. Importantly we
207 sought to compare the effects of altering BMP4 and WNT but do not aim to compare previously described
208 hPSC specification methods; therefore, all differentiations began with iPS in stem flex media after clump
209 passaging that gave rise to hTSC from multiple iPS lines (see Figure 2) but is a different iPS media from
210 previous BMP4 protocols (see discussion). Using the highly parallel droplet based single cell sequencing
211 method Drop-Seq [23], 9,821 high-quality cells were obtained after removing cells with less than 1000
212 genes detected and more than 20% of mitochondrial mapping. The average ratio of mapping reads for
213 mitochondrial genes in all cells were 3.9%, indicating good viability (Supplemental Figure 3a). The total
214 number of genes detected ranged between 21804 for BMP4 + IWP2 day 4 to 27134 for TS day 4
215 (Supplemental Figure 3b) time point with between 563 to 1572 high quality cells per time point
216 (Supplemental Figure 3c).
217
218 We performed dimensionality reduction for the most variable genes across all cells with Principal
219 Component Analysis (PCA) and Uniform Manifold Approximation and Projection (UMAP) embedding.
220 Generally, single cells are separated by differentiation stage and condition, with UMAP dimension 2
221 generally capturing the differentiation condition (Figure 3b). To identify transcriptionally similar groups of

222 cells (hereafter referred to as clusters), we performed a graph-based clustering analysis of k nearest
223 neighbors using the euclidean distance of the 30 first PCs in Seurat [24]. We identified 19 clusters, and
224 clusters contained mostly a single time-point and media condition (Figure 3b-d and S3d and Table S1).
225 iPSC clusters 1, 2 and 3 mostly contained iPS cells with slight contributions from the first days of all three
226 differentiation conditions. Cells from the BMP4 condition almost exclusively formed four clusters (termed
227 BMP Alone) BA1, BA2, BA3 and BA4. BA1 and BA3 were predominantly composed of cells from day 4 of
228 the differentiation while BA2 and BA4 were predominantly from day 6 cells. Cells from the BMP4+ IWP2
229 condition were found predominantly in a group of five heterogeneous clusters (termed BMP IWP2) BI1,
230 BI2, BI3, BI4 and BI5. BI1 cluster was composed of roughly equal parts iPSCs and BMP4 + IWP2 day 2
231 treated cells. BMP4 + IWP2 cells made up the majority of the cells in the other BIs clusters alongside a
232 smaller population of BMP4 condition. BI2 and BI3 clusters were predominantly derived from BMP4 +
233 IWP2 day 2 cells, while day 4 contributed to BI4 cluster, and day 6 contributed to BI5. In addition, cluster
234 BI5 contained a small contribution of cells originating from the final differentiation day of the TS condition
235 termed T7 cluster, demonstrating capacity to differentiate to TSCs. Cells from the TS condition almost
236 exclusively populated 7 clusters termed T1, T2, T3, T4, T5, T6 and T7 clusters. T1 cluster was exclusively
237 composed of day 2 cells and T 2-5 clusters were composed of a mixture of cells from days 2-6. The final
238 day of TS condition differentiation from cells 24 hours after passaging separated into two clusters named
239 T6 and T7, which confirms the brightfield observations of 2 cell populations. Interestingly, 14 cells from
240 BMP4+IWP2 day6 and 4 cells from the BMP4 day6 contribute to the T7 cluster (Figure 3d, Table S1).
241
242 To understand transcriptional changes during the differentiation, we asked which cells express canonical
243 pluripotent, trophectoderm and mesoderm specific genes. The stem cell transcription factor *SOX2* is
244 highest expressed in the iPSC stage and is absent from the most differentiated clusters (BA4, BI5 and
245 T7) in all conditions (Figure 3e). *SOX2* expression is maintained at a higher level in the initial days of the
246 TS condition compared to BMP4 + IWP2 and BMP4 treatment and absent from the most differentiated
247 trophoblast stem cells. The Paired-like homeodomain transcription factor 2 PITX2 is highly expressed in
248 the later stages of BMP4 differentiation (BA4 clusters), suggesting that BMP4 conditions generate cells
249 with more mesenchyme phenotype (Figure 3f). In single cell RNAseq of the human placenta, *PITX2* was
250 identified as a marker gene for a stromal fetal communicating cell type F1 by Vento-Tormo [25, 26].
251 Keratin 7, *KRT7*, and human chorionic gonadotropin subunit alpha, *CGA*, are expressed in both stem cell
252 derived cytotrophoblasts and *in vivo* placenta [27]. Indicating specification to trophectoderm
253 transcriptional programs, we find both *KRT7* and *CGA* are expressed in the later stages of the BMP4 +
254 IWP2 condition (BI4 and BI5 clusters) and TS condition (T6 and T7 clusters) (Figure 3g-h and Figure S4d
255 and e). Overall, both TS condition and BMP4 + IWP2 treated cells demonstrated an increased
256 expression for most trophoblast markers including *GATA3* and *TFAP2C* (Figure S4a) while the BMP4
257 condition demonstrated a more mesoderm associated expression of *GATA4*, *TBXT* and *PDGFRA* (Figure
258 S4b). In mouse, the Caudal-type homeobox 2 (CDX2) is involved in the segregation of the inner cell mass
259 and trophectoderm lineages at the blastocyst stage by repressing OCT4/POU5F1 and NANOG in the
260 trophectoderm. Overexpression of CDX2 in mouse embryonic stem cells causes differentiation to
261 trophoblast stem cells [28, 29]. In humans, *CDX2* expression is initiated after blastocyst formation and
262 has variable expression patterns in trophectoderm [30]. Similar to human embryo and TSC derived from
263 human blastocyst outgrowth [3], *CDX2* is transiently expressed during the intermediate stages of
264 specification (day2-4) for all conditions and is not expressed in the final TS condition state (Figure S4a).
265
266 Next we asked which genes are preferentially expressed in each cluster by non-parametric Wilcoxon rank
267 sum test (adj.p-value < 0.05; logFC > 0.25), by comparing gene expression of cells in a cluster compared
268 to all other cells (Figure 3i, Table S2). *DUSP6* and *THY1* are primed pluripotent stem cell state genes and
269 are among the top marker genes identified in the iPS cells. The BA cluster genes include WNT signaling
270 pathway *RSPO2* a WNT agonist [31] and *PITX2,* which interacts with WNT signaling and regulates

collagen expression [32]. BI5 cluster genes include *PRSS23* a serine protease involved in Snail-dependent Epithelial to Mesenchymal Transition [33] and Insulin-like growth factor signaling *IGFBP7* [34] and *PAPPA2* [35]. Markers of the T7 cluster include EGFR signaling *IFI6*, and *MUC15* [36] and the YAP downstream effector *ARHGAP18* [37], recently implicated in the specification hiPSC to bipotent trophoblast stem cells [38].

In sum, activation of WNT and EGF and inhibition of TGFb, HDAC and ROCK of primed iPSC induces transcriptional programs that closely related trophoblast-like cells. The differentially expressed genes for all pairwise cluster comparisons are provided in Table S3.

Trophectoderm expression signatures are enriched in iPS-derived trophoblast cells

To compare the iPS-derived cell types to peri-implantation human embryos, we calculated the scaled transcriptional similarity of our scRNAseq data to previously annotated gene signatures for epiblast, hypoblast and trophoblast from cultured human embryos [39] (Figure 4a-d). Two iPSC clusters exhibited the highest similarity to epiblast marker gene expression, with average score of 0.1 (Figure 4b). Interestingly, a population of cells from the BA2 cluster were most similar to hypoblast tissue profiles with a similarity score of 0.3 (Figure 4c). The most mature trophoblast T7 cluster was specific for trophoblast tissue expression with a transcriptional similarity of 0.6 (Figure 4d). The post-passaging TS condition cluster T6 and T7 are enriched in trophoblast gene expression. The T7 cluster contains the highest trophoblast-genes expression, as evaluated by pairwise comparison of the average trophoblast gene expression for all clusters (Wilcoxon Rank Sum Test with Holm-middle combined p value = 3.5e-41). Further confirming the trophoblast transcriptional similarity by gene set enrichment analysis, cluster T7 exhibited the highest enrichment for expression of trophoblast-signature genes (GSEA adjusted p value = 0.001, Figure S3a). In summary, the single cell expression profile of cells treated with TS condition gain a trophoblast identity 8 days after specification and passaging.

Next, we compared the specification map with *in vivo* amnion. Recent studies suggest that primed pluripotent stem cells patterned with BMP4 give rise to amnion-like cells [22, 40, 41]. Importantly, genes expressed predominantly in human amnion were not enriched in cluster T7 containing mostly differentiated TS treated cells (GSEA adjusted p value = 1, Figure S5b,c), instead cluster T7 is most enriched for trophoblast genes (Figure 4e). We evaluated the expression of 198 genes predominantly expressed in human amnion [42] and previously reported to be upregulated in primed BMP4 condition [22]. The T7 cluster have the lowest average expression for the 198 amnion genes with 0.068 log normalized average expression compared to all other cell clusters (Wilcoxon Rank Sum Test with Holm-middle combined p value = 4.8e-06), compared to 0.47 for the BA cluster BMP4 treated cells, 0.46 for the BMP4 +IWP2 treated cells, and 0.16, 0.13 and 0.16 for the iPSC clusters (Figure 4e, S6, Table S4, S5 and S6). Instead, amnion genes are expressed in clusters from the earlier stages of TS condition specification and BMP4 treatment. For example, *POSTN* gene is significantly upregulated in cluster BA4 (adjusted p value < 2.225074e-308), which predominantly contains d6 BMP4 alone treated cells (Figure 4f). Moreover, the amnion gene IGFBP5 is significantly upregulated four and six days after differentiation in both TS and BPM4+IWP2 conditions (BI 4 adjusted p value = 4.389994e-95 and BI5 adjusted p value = 1.909090e-147, mostly containing day 4 and 6 BMP4 + IWP2 cells; and T3 cluster adjusted p value = 4.301357e-148; mostly containing day 4 TS treatment) (Figure 4g). Cinkornpumin et al report high expression of the amnion gene *ITGB6* in trophoblast cells derived from primed hPSC [40]. *ITGB6* is nearly absent from all TS condition single cells while the BMP4 + IWP2 D6 (Wilcoxon Rank Sum test adjusted p value = 9.332e-264) sample and BI5 cluster (Wilcoxon Rank Sum test adjusted p value = 8.890e-286) show significant higher upregulation (Figure 4h). In summary, 8 days after specification and passaging in TS condition, primed iPS cells are specified to trophoblast without amnion expression.

iPS-derived trophoblast cells resemble first trimester placental cytotrophoblasts

We next asked how similar are the iPS-derived placental cells to *ex vivo* cells found in human first trimester placenta. To characterize the expression profiles of the 19 clusters, we analyzed the overlap with cell type-specific genes identified by two published single-cell studies of early human placenta [25, 26]. We performed gene set enrichment analyses using Cell-Specific Expression Analysis (CSEA) [43]. We identified cell type specific genes that were preferentially expressed in each cell type/cluster of the 19 clusters from iPS-derived cells by the specificity index probability (pSI) statistic at thresholds of $p< 0.05$. Next, we tested whether cell-type specific genes previously identified by single cell studies of human placenta are over-represented in the cell-type specific genes from our iPSC specification by hypergeometric test and applied the Bonferroni correction for multiple comparisons, considering all the tested gene lists [$\alpha = 0.05/ (19\times (38+14)) = 5.1\times10^{-5}$]. Vento-Tormo identified a set of 38 cell types from maternal (decidua and blood) and fetal tissues (placenta) of first trimester placenta and reported 30 cell-type specific genes per cell type [26]. We found that the transcriptional profile from all iPSC clusters and clusters from d2-d4 timepoints did not have significant enrichment for any of the placental cell type clusters (Figure S7a,b). In contrast, the most mature iPS-derived TS cluster T7 was highly enriched for the syncytiotrophoblast (STB) and villous cytotrophoblast (VCT) cluster ($p=\log_{10}^{-32}$ and $\log_{10}^{-17}$) (Figure 5a). The expression profile of the iPS-derived TS T7 cluster was highly specific for STB and VCT with no significant enrichment for any other cell-type in the maternal fetal interface. The cluster BA4 (predominantly day 6 BMP4) was most enriched in the two fibroblast clusters (F1 and F2) ($p=\log_{10}^{-26}$ and $p=\log_{10}^{-13}$). Vento-Tormo described these placental fibroblasts as mesenchymal stromal cells of fetal origin that derive from the primitive endoderm expressing *GATA4*, *GATA6*, *PDGFRa*, and *SOX17* (Figure S7 and S8). The BI clusters, which contain mostly BMP4 + IWP2 treated cells, were not strongly enriched for specific placental cell type and instead had weaker enrichment for several clusters including syncytiotrophoblast (STB), and maternal decidua derived F1 endodermal cell type and dS1. In summary, iPS-derived TS cells are highly enriched for expression of genes specific to syncytiotrophoblast and villous cytotrophoblast of *ex vivo* human placenta.

Next, we compared the expression profiles to an early placental dataset from the 8 week and 24 week placenta [25] (Figure 5b). Again, we found strong enrichment for the iPS-derived TS cells in the eight week fusion competent CTB (cytotrophoblast, $\log_{10}^{-9}$). BA and BI clusters that contain mostly BMP4 and BMP4 + IWP2 conditions were enriched for two different mesoderm cells identified by Liu et al [25] ($\log_{10}^{-23}$ and $\log_{10}^{-4}$). To visualize placental cell-type specific expression patterns among the iPS-derived cell types, we plotted a heatmap of the top unique markers of the fetal cell types identified in Vento-Tormo and polar trophectoderm markers identified in Petropoulos [44] (Figure 5c). Consistent with the Gene Set Enrichment Analysis, we observed high expression for a majority of unique marker genes previously identified in polar trophectoderm villus cytotrophoblasts (VCT) and syncytiotrophoblasts (STB). We also compared the expression profiles of all iPS-derived cellular clusters to previously identified lineage and marker genes from pre-implantation datasets [45] (Figure S8a). Again, we found the highest expression of trophoblast genes *KRT18*, *TEAD3*, *GATA3* and *GATA2* in iPS-derived TS clusters, with moderate expression in BI clusters. Primitive endoderm genes *SOX17*, *GATA4*, *GATA6* were highest expressed in BA clusters. When compared to primary trophoblast cells (Figure S8b), we confirmed that the T7 cluster has the highest level of cytotrophoblast, trophectoderm and EVT enriched genes.

In summary, iPS cells specified via TS condition have a transcriptional profile highly similar to cytotrophoblasts taken directly from human placenta while the BMP4 differentiation conditions used here generates more heterogeneous placental cells types with evidence of amnion expression.

Primed iPSC specify to trophoblast by activation of TFAP2A without activation of naive stem cell programs

It has been reported that primed stem cells are restricted in their potency and unable to differentiate to TSC while naive stem cells readily differentiate to TSC [41]. Therefore, we asked if the primed iPSC or TS condition cells adopt a naive stem cell program during differentiation. For 12 previously identified naive marker genes [46], naive transcripts were nearly absent. Five genes *HORMAD1*, *ALPPL2*, *KHD3L*, *TRIM60*, and *HYAL4* had no observed expression in any cell (Figure 6a, Table S7). Less than 1% of cells for any condition had detectable expression for six other marker genes *ALPP*, *OLAH*, *LYZ*, *MEG8*, *KDC1L*, and *FAM151A*. *ZNF729* was the only gene expressed in 0.51% of TSp1 cells to 14.69% of IWP2d2. In contrast, primed marker genes were consistently expressed in iPSC clusters and often throughout the differentiation state. For example, *DUSP6* was detected in 74.87% of the iPS cells, indicating a primed state. In sum, primed iPSC transitions to cytotrophoblasts without transitioning through a naive state.

We asked how BMP signaling contributes to the differentiation by evaluating BMP family signaling in all samples. Low percentages of cells expressed BMP4 in TS condition, while BMP4 was significantly upregulated in several cell clusters of the BMP4 and BMP4 + IWP2 treated cells compared to iPSC and TS condition (Figure 6b, Table S7). BMP5 and BMP7 were expressed in a larger portion of TS condition cells compared to iPSC (BMP5: 0-0.67% iPSC vs 24.72% T1 cells and BMP7: 1.01-4.71% iPSC vs 29.06% T1 cells). The BMP effector proteins SMAD1,4, and 5 are expressed in the first days of all conditions, which can indicate that the endogenous BMP signaling may be important during the initial differentiation process.

To identify the most significant gene expression changes associated with the initial stages of TS condition specification, we applied a single-cell trajectory inference and pseudotime estimation (STREAM) [47] to the cells from the iPSC and TS condition day 2 samples. Briefly, single cells were ordered along probabilistic trajectories and a numeric value referred to as pseudotime was assigned to each cell to indicate how far it progresses along the dynamic differentiation. STREAM identifies an initial branch composed of both iPSC and TS condition cells at day 2 (S3-S0) which transition to three branches: TS condition day 2 (S0-S1) branch reflecting the exit from pluripotency, and iPS cell branches (S0-S2, S0-S4), reflecting the pluripotency continuum (Figure 6c and Table S8, S9, S10, S11, S12). We next identified the genes involved in the specification that are correlated with the transition along the branches (see methods). Pluripotency to TS (branch S0-S1) have upregulation of cytokeratins (*KRT18*, *KRT19*, *KRT8*) and calcium binding proteins (*S100A11*, *S100A10*) known to be expressed in trophectoderm [48, 49] (Figure S9d). Top positively correlated transition genes along the exit from pluripotency branch (S0-S1) include *LRRN3* which amplifies MAPK signaling through EGFR [50], *MEIS2* which codes for a TALE homeobox protein that promotes tumor progression through the combined action of two isoforms increasing YAP nuclear translocation and activating the WNT/b-catenin pathway [51], *BMP5*, *BMPR1B* and *TFAP2A* a transcription factor key to the suppression of pluripotency and expression of trophoblast associated genes [52] (Figure 6c, S9c). Also found as a top transition gene, *TMSB4X* is implicated in stemness of progenitor trophoblast cells of first trimester human placentas by increasing NOTCH1 activity [53, 54, 55]. In summary, trophectoderm transcriptional programming occurs early in the specification and utilizes previously identified trophectoderm associated regulatory factors including *TMSB4X*, *YAP*, *BMP5*, and *TFAP2A*.

Human endogenous retrovirus derived genes participate in a regulatory subnetwork within primed-derived trophoblast stem cells

Endogenous retroviruses play a fundamental role in the evolutionary diversification of the mammalian placenta with multiple instances of human endogenous retroviruses (hERVs) co-option [56, 57, 58, 59]. We reasoned that hERV expression may be altered because primed PSC, naive PSC and TSC differ in the regulation of human endogenous retroviruses [60], and therefore, we investigated how hERVs are expressed during specification. While there are potentially 1500 ERV-derived genes capable of encoding proteins [57], we restrict our analysis to the 20 ERV-derived genes that are currently annotated as human genes including *ERV3-1, Syppressyn (ERVH48-1)*, *syncytin-1 (ERVW-1)* and *syncytin-2 (ERVFRD-1),* which function in antiviral responses and syncytial fusion. We found specific upregulation of five ERV-derived genes (*ERVH48-1*, *ERV3-1*, *ERVMER34-1*, *ERVW-1* and *ERVFRD-1*) in the most differentiated cells in the TS condition, cluster T7 (Figure S10a).

To understand the place of ERV-derived genes in the regulatory TS regulatory network, we analyzed the participation of ERV-derived genes in a gene regulatory network. We built a transcription factor and target gene network model using the Passing Attributes between Networks for Data Assimilation (PANDA) algorithm to all clusters among the different conditions used here. PANDA integrates information from TF-sequence-motif data, gene expression and protein-protein interaction (PPI) in a message-passing approach [61]. We found that ERV interactions (Transcription Factors + ERV-derived genes) are prominent in the most differentiated TS cluster TS 7 and in no other cell clusters. We found regulatory interactions for five ERV-derived genes (*ERVH48-1*, *ERV3-1*, *ERVMER34-1*, *ERVW-1* and *ERVFRD-1*) are highly unique to the TS 7 cluster network, while only 0 or 1 ERV-derived gene was found in the other cell clusters (Figure S10b, S11a). Interestingly, genes encompassed in ERV regulatory interactions of primed-derived TSC were enriched in biological pathways related to hormone metabolism, cell differentiation and the immune system (adj. p value < 0.01) (Figure S10c, Table S13). In addition, ERV regulatory interactions are also enriched for categories such as placenta development, trophectodermal cell differentiation and syncytium formation (Table S13). To understand the significance of the ERV-subnetwork interactions, we next asked what is the strength of ERV edges in the network context of the TS cluster. Weights of ERV regulatory interactions were highly ranked among all network connections (Wilcoxon Rank Sum p-value=$1.6778e^{-31}$, permuted p-value=0.001), indicating a significant contribution to the gene regulatory network of the TSC state. In sum, we find that ERV-derived genes participate in a regulatory subnetwork within primed-derived TSC.

Discussion

Here we map the specification of primed iPSC to trophoblast by exogenous induction of EGF, WNT and Valproic acid and inhibition of TGF-β and HDAC and demonstrate the involvement of GATA3, BMP5/7, TFAP2A without transitioning through a naive state. Primed-derived-TSC are highly transcriptionally similar to cytotrophoblasts found in human placenta. Primed-derived-TSCs proliferate rapidly as a stem cell and differentiate into multinucleated extravillous and syncytiotrophoblast. After eight days of specification, primed-derived-TSC displays a uniquely active transcriptional network of human endogenous retroviruses similar to *in vivo* trophoblast.

The ability of primed human stem cells to differentiate into true trophectoderm lineages has been contested. Here we confirm that primed PSC retains a broad potency to generate extra embryonic lineages which are transcriptionally highly similar to placenta. While recent studies conclusively prove that primed PSC can be specified to TSC, studies report absent [62] or inefficient conversion to TSC with the direct application of the TSC-media, as described here. Optimized culturing conditions and differences in the primed PSC state at the initiation of TSC-media exposure likely explain the more robust differentiation described here. We induce differentiation after self-organized multicellular colonies are established by

clump-passaging primed cells from StemFlex conditions, plating cells in StemFlex media and inducing differentiation 24hrs after passaging all in 20% O2. Instead of StemFlex, others have initiated differentiation with primed PSC cultured on MEFs (Viukov et al., 2022), TSER or E8, a xeno-free media, and used either 20% or 5% O2. StemFlex is proprietary media containing bovine serum albumin and not all components are known. While PSCs grown in StemFlex and E8 have similar transcriptional profiles, PSC in StemFlex demonstrate less karyotypic abnormalities, better expression of Nanog, more efficient single cell cloning functional differences such as enhanced single-cell cloning and robustness to automation (Daniszewski et al., 2018). While the conditions optimized here improve the specification of TS from primed iPS, removing WNT for the initial days would likely improve the efficiency even further, as recently published for primed PSC grown on MEFS [10].

We also confirm that the exogenous BMP4 protocols can generate cells with key features of trophoblast, but the majority of cells specified by BMP4 are less similar to *in vivo* cytotrophoblasts. Importantly, the BMP4 methods used here differ from the more established BAP method, which are optimized to produce more efficient trophectoderm differentiation. In those optimized conditions, transient BMP4 application induces a naive state. For the direct TSCM method here, we find that endogenous BMP5 and BMP7 are induced, without inducing BMP4 or naive state. Similar to induction with BMP4, direct TSCM method induces transient initial CDX2 expression, and CDX2 is also transiently upregulated in an enhanced primed-to TSC conversion without WNT stimulation and inhibiting TGF-B.

Our data suggest that while PSC can differentiate via multiple paths, the paths share many established trophectoderm specification factors and the final TSCs are similar. Single cell sequencing allows deconvolution of the cellular specification by complete transcriptional profiles. The specification generates two separate cell clusters both with trophectoderm expression, and the primed-derived TSCs preferentially expand to purity with passaging (Figure 1p). Further supporting similarity to primary TSC derived from human villous cytotrophoblast and from the blastocyst outgrowth, the primed-derived-TSC also begin as mixed populations of cells which are passaged to generate pure TSC lines [3]. Our trajectory analysis indicates that *CDX2* and *TFAP2C/A* are important in the initial exit from pluripotency in primed cells and facilitates trophectoderm lineage specification, similar to human trophectoderm specification *in vivo.* TFAP2A is a transcription factor involved in both trophoblast specification and a trophoblast-like amnion speciation. In humans, the first wave of amnion specification may follow a trophectoderm-like pathway [11]. We use multiple statistical tests examining single genes, gene sets and complete transcriptional profile similarities to demonstrate that the primed-derived-TSC are most transcriptionally similar to trophectoderm and not amnion.

We find that five hERV-derived ORFs annotated as human genes are expressed in primed-derived TSC, similar to in vivo [57]. hERV are abundant sequences within the human genome that are derived from ancient retroviral infections in human ancestral germline. hERV-derived Long Terminal Repeats function as human-specific enhancers during placentation and inflammation. hERV-derived envelope proteins create coding ORFs. A specific hERV-envelope protein named syncytian-1 (hERVW-1) is best characterized for cell-cell fusion during embryo attachment and syncytialization where it is upregulated [58]. Like many of the other 668 hERV-envelope proteins expressed in the early embryo and placenta, Syncytian-1 is widely expressed in ICM, trophectoderm and CTB [57]. Other hERV-derived ORFs inhibit cell fusion (syncytin-2), provide immunosuppressive activities (syncytin-2, ERV3-1 and suppressyn) [59], and restrict retroviral infection in morula, ICM, TE and ESC (supressyn) [57].

The differentiation of primed iPSC to trophectoderm allows for modeling of placental diseases from patient-specific iPS cells without first trimester primary tissue and without transitioning through a naïve state. While the final TS product of these different models appears similar, the naive vs primed states

differ in molecular features which could later impact hTSC models. In naive state, global DNA demethylation and genomic instability occur in both t2i/L/Go or 5i/L/A conditions [64, 60], and imprinted loci may be aberrantly demethylated [40]. Mammalian imprinting is most pronounced in the placenta [66] and cellular conditions that maintain imprinting may be important in a cellular model of human placenta. X chromosome inactivation (XCI) is reset in naive conditions compared to inactivated or eroded in primed conditions, and it remains to be seen if there are differences in XCI.

Placental adaptive responses to insults can alter fetal development and program the fetus for later disease including diabetes and neurodevelopmental disorders [67, 68]. As both the protective barrier and a component of the innate immune system, cytotrophoblasts and their progeny play important roles in the outcome of insults [69]. Recent evidence that Covid19 leads to a hyperimmune state has implications for placental biology and fetal development, potentially analogous to the effects of other hyperimmune states during pregnancy [2]. Genetic variation among individuals influences how a pathogen affects the fetus. iPSC models of this genetic variation could be harnessed for therapeutic intervention. We anticipate that primed-derived TS cells will be a powerful model to study the placental contribution to disease and the genetic regulation of human placental evolution.

Material and Methods

iPSC culture: Human ES and iPSC were maintained on Collagen IV (1mg/12 ml DMEM/F12) coated plates in StemFlex media. Media was changed every 48hrs in accordance with manufacturer recommendations. Cells were passaged in small clusters using Versene solution. See supplemental methods for details.

Differentiations and Cell Maintenance: 24 hours post passaging with Versene, cells were washed with DPBS and media was changed to differentiation media. TS medium: [DMEM/F12 with Glutamax supplemented with 0.1 mM 2-mercaptoethanol, 0.2% FBS, 0.3% BSA, 1% ITS-X supplement, 1.5 µg/ml L-ascorbic acid, 50 ng/ml EGF, 2 µM CHIR99021, 0.5 µM A83-01, 1 µM SB431542, 0.8 mM VPA and 5 µM Y27632]. BMP4: 10 ng/ml BMP4 with/without 2uM IWP2 was added to basal differentiation medium (DMEM/F12 with Glutamax supplemented with ITS and L-ascorbic acid 2-phosphate magnesium). Cells were collected on day 2, 4, and 6 for RNA sequencing. For the TS condition, after 6 days of differentiation cells were passaged using TrypLE express at a split ratio of 1:3 and plated on new Collagen IV (5 ug/mL) coated plates. After the first passage, media was changed every 48 hours and split at a ratio of 1:3-1:6 every 3-4 days. See supplemental methods for more details.

EVT and STB differentiation and characterization*:* Differentiation of TSCs were performed as previously described [3], with minor modifications. For EVT differentiation, 24-well plates ibidi were coated with 1 µg/mL Collagen IV overnight. 2×10^5 TSC were seeded per well in 500µL EVT basal medium [DMEM/F12 supplemented with 0.1 mM β-mercaptoethanol, 0.5% penicillin-streptomycin, 0.3% BSA, 1% ITS-X, 7.5 µM A83-01, 2.5 µM Y27632] supplemented with 4% KSR and 100 ng/mL NRG1 alpha. Matrigel was added to a 2% final concentration shortly after resuspending TSC in the medium. On day 3, the media were replaced with 500µL EVT basal medium supplemented with 4% KSR, and Matrigel was added to a 0.5% final concentration. At day 6, the media were replaced with 500µL EVT basal medium, and Matrigel was added to a 0.5% final concentration. At day 9, the cells were ready for analysis

For STB differentiation, 24-well plates ibidi were coated with 2.5 µg/mL Collagen IV overnight. 2×10^5 TSCs were seeded per well in 500µl STB medium [DMEM/F12 supplemented with 0.1 mM β-

mercaptoethanol, 0.5% penicillin-streptomycin, 0.3% BSA, 1% ITS-X, 2.5 μM Y-27632, 2 μM Forskolin and 4% KSR]. The media was changed at day 3, and at day 4 the cells were ready for analysis.

Data Availability: The Single cell RNA sequencing datasets generated and/or analyzed during the current study have been deposited in ArrayExpress with the accession code E-MTAB-9526. Analysis pipelines are publicly deposited on Github:https://github.com/paquolalab/placenta_ips

Acknowledgements

Supported by funding from the Lieber Institute for Brain Development and a NARSAD Young Investigator Grant from the Brain & Behavior Research Foundation to J.A.E, the National Institute of Health (NIH) T32 Fellowship KJB (T32MH015330). The authors are grateful for the financial support of the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) - Financing code 001 (Andre Barbosa, Veronica Euclydes and Arthur Feltrin). This work was supported in part by FAPESP - Sao Paulo Research Foundation (http://www.fapesp.br/en/): Grant Number 2018/18560-6. We thank Bill Ulrich for the single cell RNAseq browser. We thank Tricia Nilles and Worod Allak from the Becton Dickinson Immune Function Laboratory at the Johns Hopkins Bloomberg School of Public Health, for flow cytometry technical assistance. The facility was supported in part by CFAR: P30AI094189-04 (Chaisson).

597 **Figures**

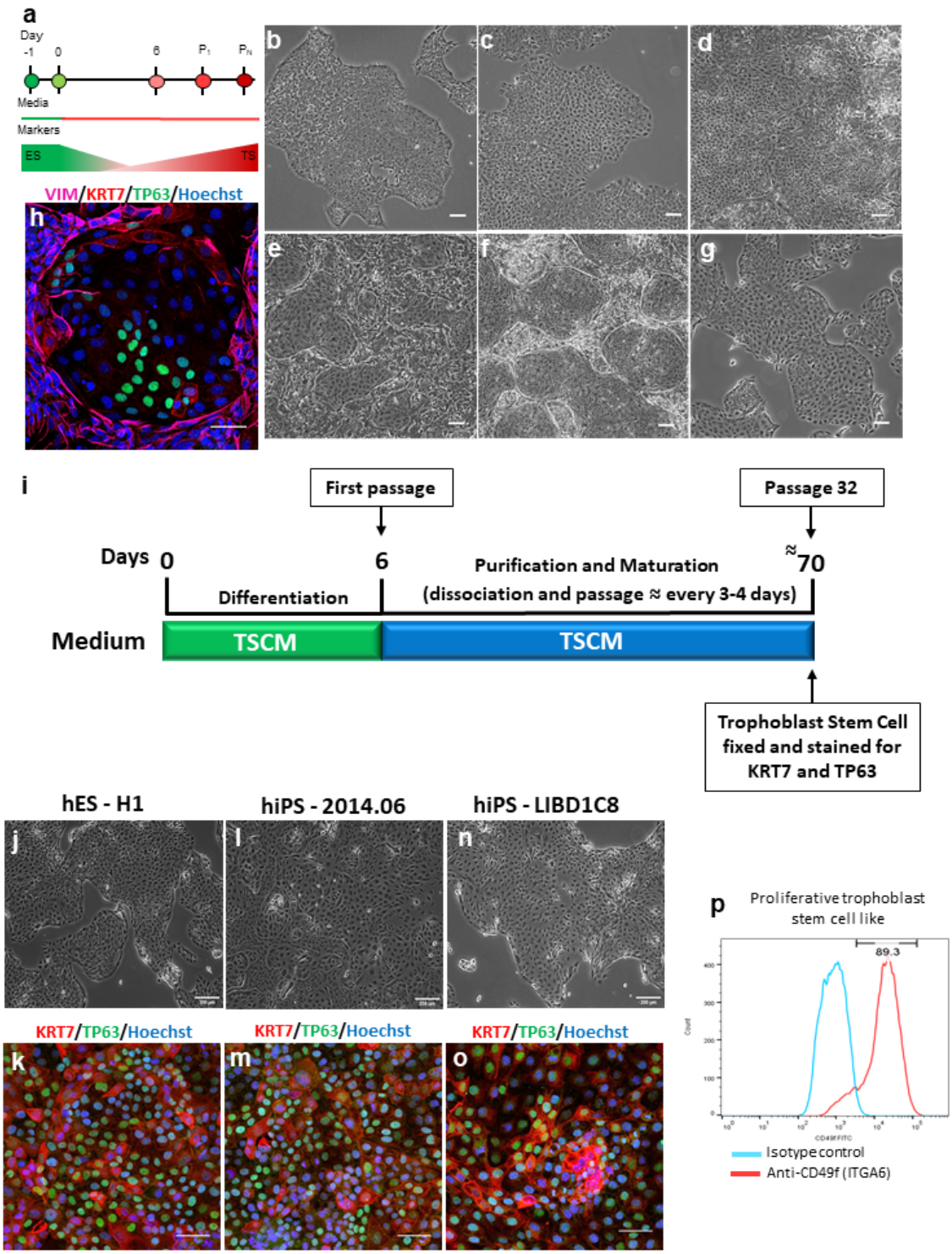

Figure 1: Derivation of trophoblast stem cell like cells from human pluripotent stem cells.

598

599 **Figure 1: Derivation of Trophoblast Stem Cell Like (TS) Cells From Human Pluripotent Stem Cells.**

600 **(a).** Schematic of protocol used to derive TS cells from human pluripotent stem cells by activating
601 wingless/Integrated (WNT) and epidermal growth factor (EGF) and inhibiting Transforming Growth Factor
602 Beta (TGFb) histone deacetylase (HDAC) and Rho-associated protein kinase (ROCK).

603
604 **(b-d).** Representative phase contrast images of differentiating TS cells (ES line H1). Images of
605 differentiating cells on days 2 (b), 4 (c), and 6 (d) after addition of differentiation (TS) media.

606 **(e,f).** Images of TS cell morphology 24 (e) and 72 (f) hours following passage (P1) on day 6.

607 **(g).** Mature TS lines generated from ES line H1 after 10 passages.

608 **(h).** Immunofluorescence for VIM (magenta), KRT7 (red), TP63 (green), and Hoechst 33342 (blue) in TS
609 cells after 1 passage (P_0)(iPSC line LIBD7c6).

610 **(i).** Schematic representation of the differentiation protocol of iPSC to TSC. TSCM = Trophoblast Stem
611 Cell Medium.

612 **(j,l,n)**. Representative phase contrast images of TS cells H1 (j), 2014.06 (l) and LIBD1C8 (n).

613 **(k,m,o).** Confocal images of KRT7 (red), TP63 (green), and Hoechst 33342 (blue) for TS cells generated
614 from passage 11 H1 (k) and 2014.06 (m) and passage 32 LIBD1C8 (o).

615 **(p).** Flow cytometry histograms of ITGA6 expression in mature TSC derived from H1 line.

616 All phase contrast images scale bars indicate 100 μm. All immunofluorescence images scale bars
617 indicate 50 μm.

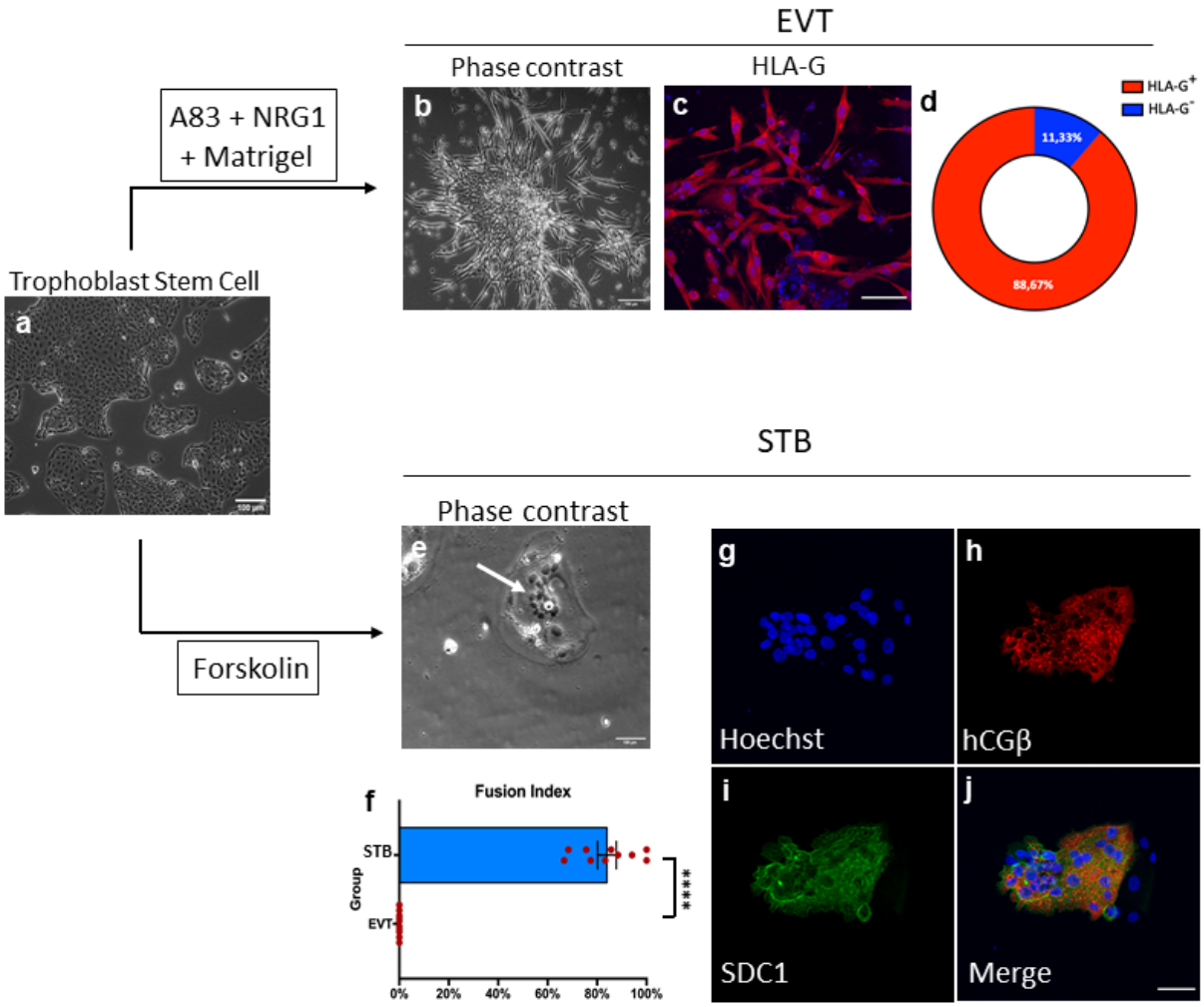

Figure 2: Primed Pluripotent derived TS cells differentiate into extra villous cytotrophoblasts (EVT) and syncytial cytotrophoblast (STB).

Figure 2: iPS-derived TS Cells Differentiate Into Syncytiotrophoblast, Extravillous Cytotrophoblasts and Villi-like Structures.

621 **(a).** Representative phase contrast images of differentiating TS cells. Scale bars indicate 200 μm

622 **(b)**. Representative image of EVT derived cells. Scale bars indicate 100 μm.

623 **(c)**. Immunofluorescence image of EVT derived cell from mature TSC line H1 stained for Hoechst 33342
624 (blue) and HLA-G (red). Scale bars indicate 100 μm.

625 **(d)**. Percentage of HLA-G positive cells after 6 days of differentiation. Scale bars indicate 100 μm.

626 **(e) .** Representative image of derived STB. Arrow indicates the multinucleated ST characteristic. Scale
627 bars indicate 100 μm.

628 **(f)**. Fusion efficiency of STB. H1 ESC - STs were analyzed. Mean + SEM. $****p \leq 0.0001$ by Wilcoxon-
629 Mann-Whitney test.

630

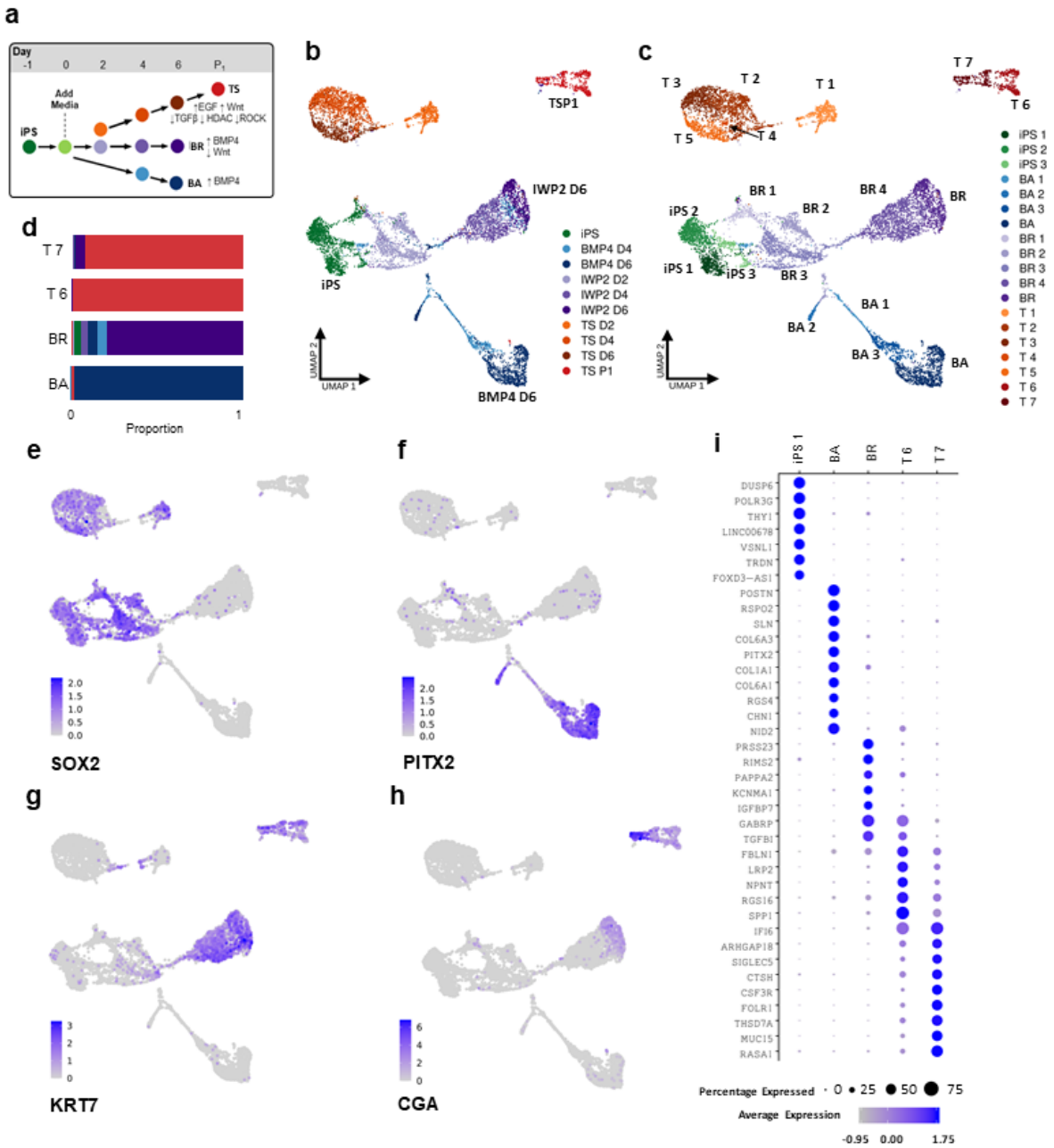

Figure 3: Temporal single Cell RNA Sequencing of iPS to trophoblast differentiation

631
632 **Figure 3: Temporal Single Cell RNA Sequencing Of iPSC To Trophoblast Differentiation**

633 **(a).** Experimental schematic of trophoblast differentiation experiment. hiPS cells were plated and allowed
634 to adhere to culture dishes for 24 hours before addition of differentiation media. Cells were collected for
635 single-cell RNA sequencing at indicated time points.

636 **(b,c).** Dimensionality reduction for the most variable genes across all cells with Principal Component
637 Analysis (PCA) and Uniform Manifold Approximation and Projection (UMAP) embedding of 9,821 single
638 cell transcriptomes from three distinct media treatments with groupings based on (b) sample identity or (c)
639 gene expression clusters calculated by k nearest neighbors using the Euclidean distance of the 30 first
640 PCs which identifies 19 clusters. Cells from the BMP4 condition almost exclusively formed four clusters
641 termed BA (BMP4 alone cell indicated in blue). Cells from the BMP4+ IWP-2 condition were found
642 predominantly in a group of five heterogeneous clusters termed BI (BMP responsive indicated in purple).
643 iPSC indicated in green and TS clusters are indicated in orange.

644 **(d).** Plot of the proportion of cells from each differentiation condition which are present in the final k-
645 means clusters TS, T6, BI5 and BA4. Color represents the differentiation condition scale in 3B
646 BMP4+IWP2: purple, BMP4: blue, iPSC: green, TS: orange. Cell proportions for all clusters can be found
647 in supplemental table S1.

648 **(e-h).** Single cell visualization of gene expression of selected markers of hiPS cells (e), placental stromal
649 cells (f), and trophoblasts (g,h) on UMAP projections. Grey dots indicate cells with no expression and blue
650 scale indicates normalized expression.

651 **(i).** Dot plot displaying differentially upregulated marker genes for the most mature cell clusters from each
652 treatment BA (BMP4 Alone), BI (BMP4+IWP2), and TS 6, TS (TS media) compared to the average
653 expression of all other clusters. Size of dot indicates the percent of cells with expression and intensity of
654 the blue indicates the average expression level.

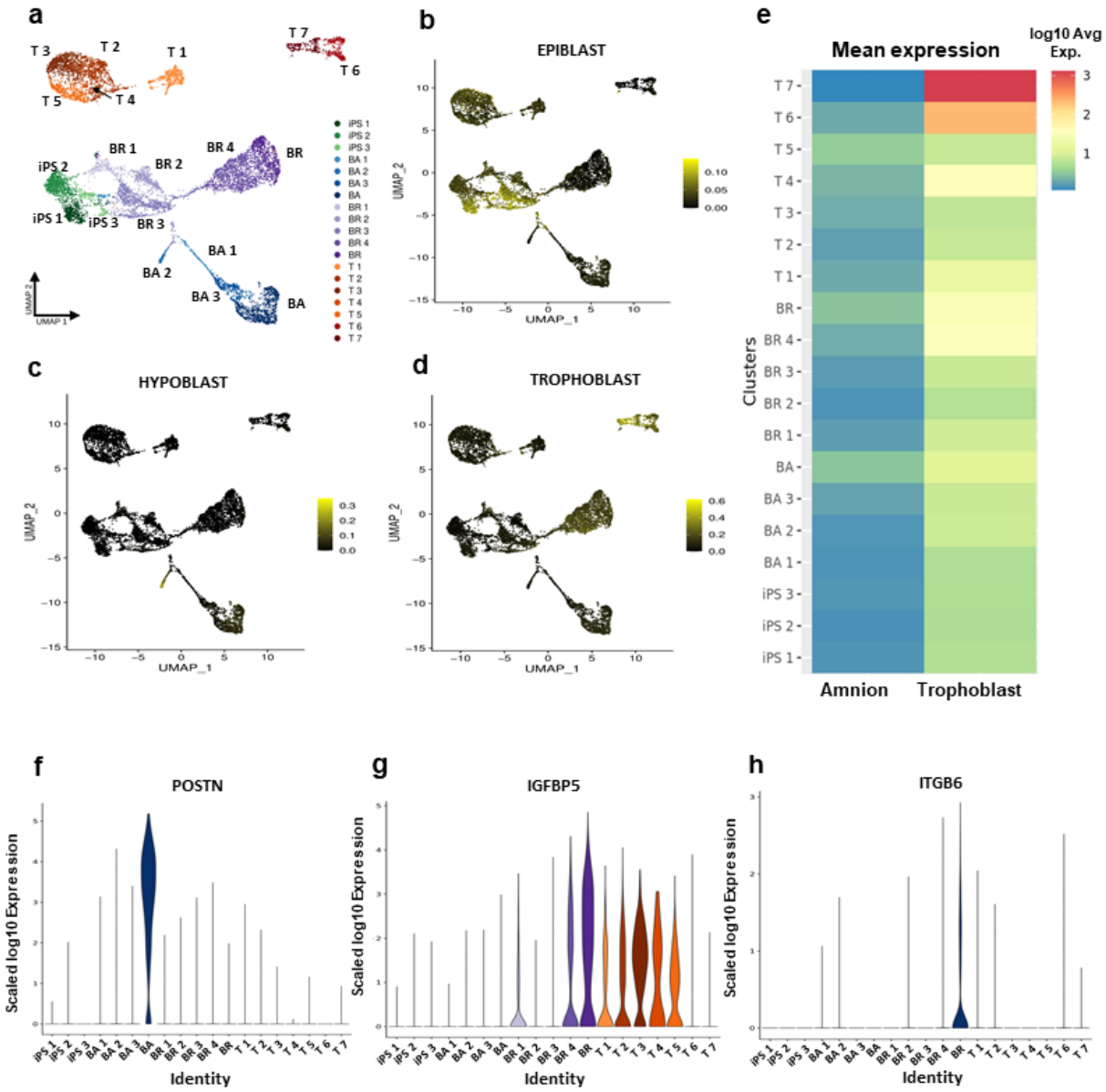

Figure 4: iPS-derived trophoblast cells specified without BMP4 are enriched for trophectoderm expression and are depleted for amnion expression.

Figure 4: iPS-derived trophoblast cells specified without BMP4 are enriched for trophectoderm expression and are depleted for amnion expression

(a). UMAP of single cell transcriptomes from three distinct conditions with cluster labels.

(b,c,d). Expression similarity score for single cells (a) based on gene signatures from Epiblast, Hypoblast and Trophoblast of human embryo [39]. (b) Two iPSC day 0 clusters exhibited the highest similarity to epiblast (average score of 0.1). (c) SC day 4 treatment was the most similar to hypoblast (similarity score of 0.3). (d) The T7 cluster was the most specific for trophoblast tissue (similarity score of 0.6).

(e). Mean expression of previously identified amnion [42] and Trophoblast gene lists [39] in each cluster. The T7 cluster contains the higher average trophoblast-gene expression (Wilcoxon Rank Sum Test with Holm-middle combined p value = 3.5e-41) and also the lowest average expression for the amnion genes (Wilcoxon Rank Sum Test with Holm-middle combined p value = 4.8e-06), compared to all other cell clusters.

(f,g,h). Violin plot indicating the expression distribution of *POSTN, IGFBP5 and ITGB6,* amnion markers, in all clusters. (f) *The POSTN* expression is significantly upregulated in the BA cluster (adjusted p value < 2.225074e-308). (g) The *IGFBP5* is significantly upregulated in BI5 (adjusted p value = 1.909090e-147) and T3 (adjusted p value = 4.301357e-148) clusters. (h) *ITGB6* expression is significantly upregulated in the BI5 cluster (adjusted p value = 8.890e-286).

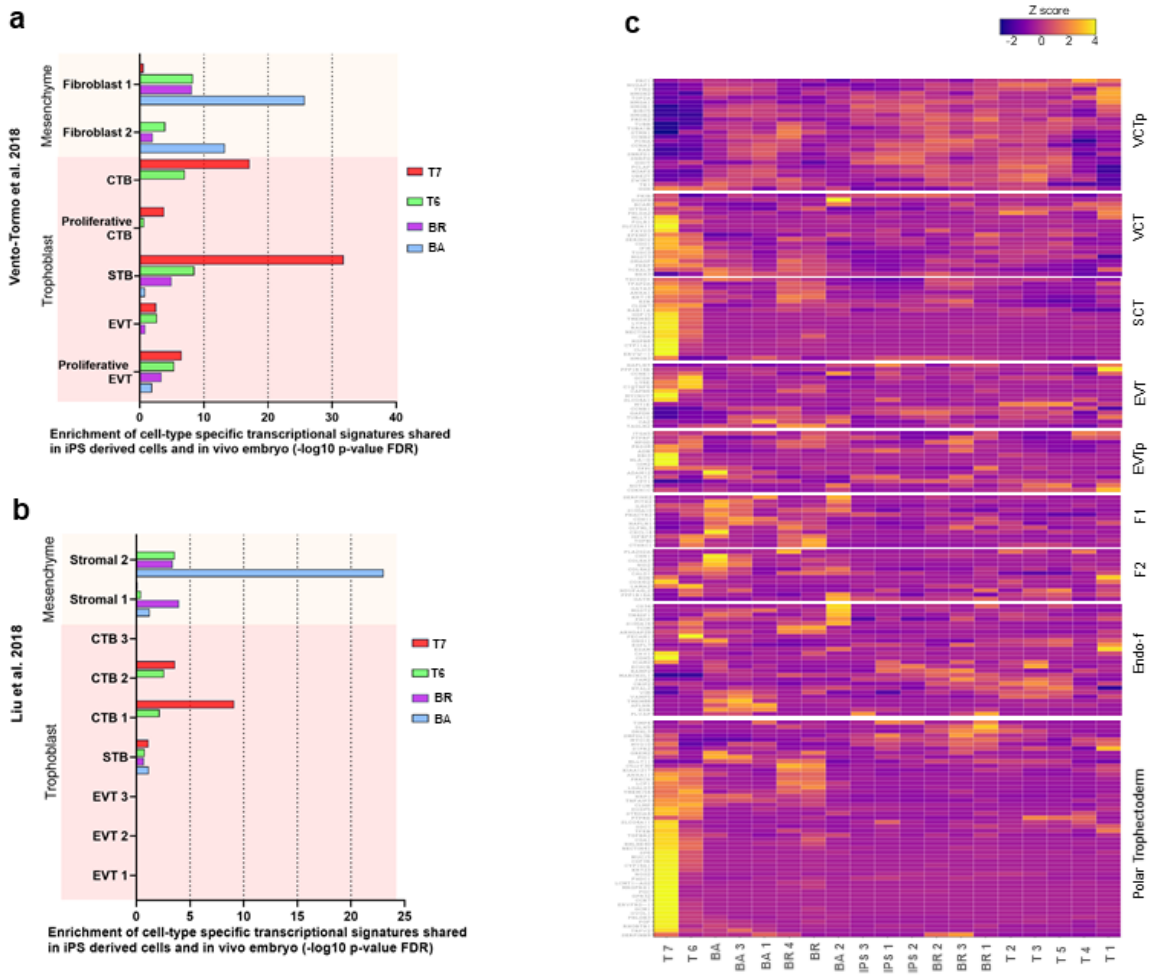

Figure 5: Trophoblast differentiations are enriched for first trimester placental cell types.

673
674 **Figure 5: Trophoblast Differentiations Are Enriched For First Trimester Placental Cell Types**

675 **(a,b).** Gene set enrichment analysis of most mature iPS-derived cell clusters as shown in figure 1c (BA4,
676 BI5, T 6 and T 7n) to cell types identified in previous single cell sequencing of first trimester placenta
677 identified in Vento-Tormo et al. 2018 [26] and Liu et al. 2018 [25].

678

679 **(c).** Composite heatmap of normalized gene expression of the top unique markers of placental cell types
680 of fetal origin identified in Vento-Tormo et al. 2018 and polar trophectoderm markers (top 50) identified in
681 Petropoulos et al. 2016 [44].

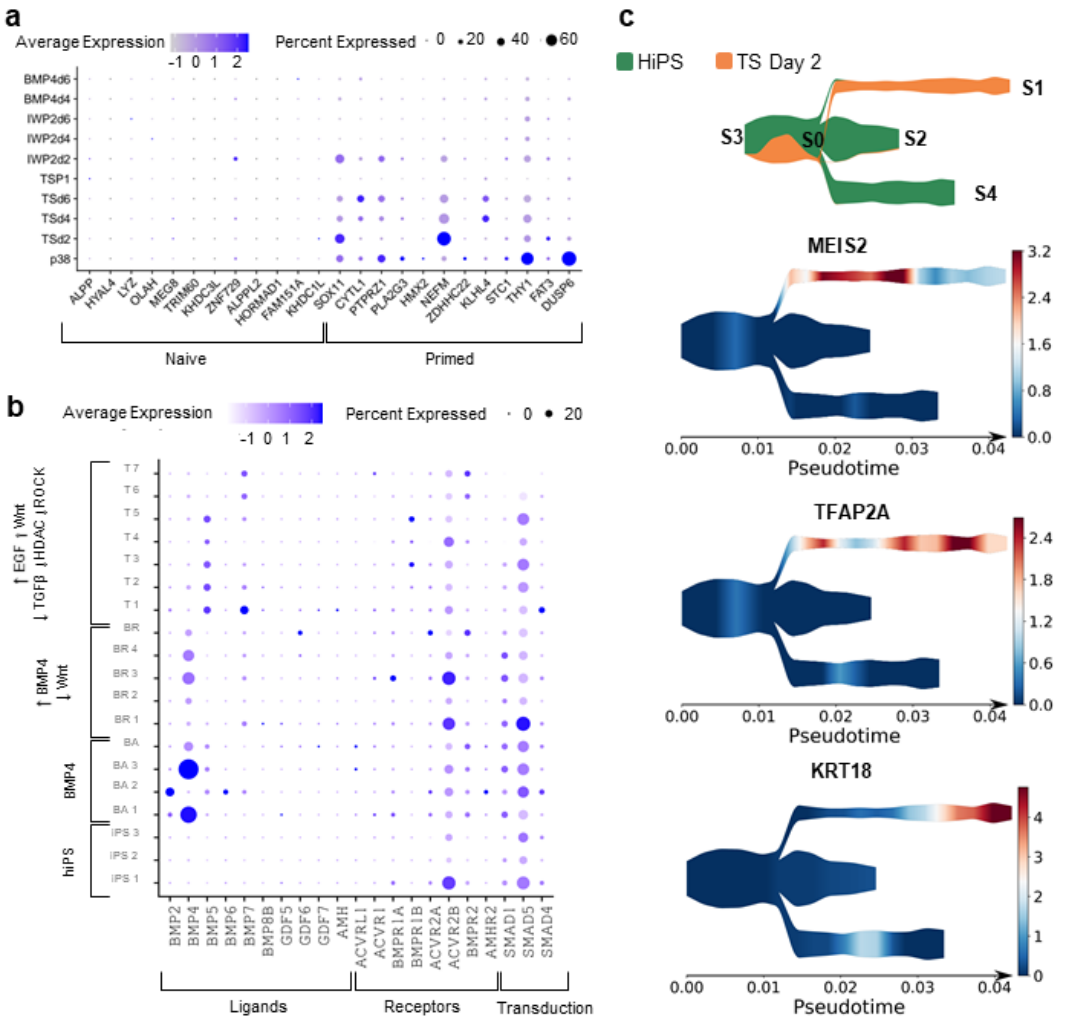

Figure 6: Primed iPS specification to trophoblast stem cells initiates with TFAP2A rewiring without activation of naïve stem cell programs

682

683 **Figure 6: Primed iPSC Specification To Trophoblast Stem Cells Initiates With TFAP2A Rewiring**
684 **Without Activation Of Naïve Stem Cell Programs**

685 **(a).** Dotplot displaying expression of naïve and primed hPSC marker genes in the three differentiation
686 treatments and days. See table S3 for source data containing log fold change and False discovery rate
687 for each pairwise comparison by Wilcoxon Rank Sum Test with Holm-middle combined p value.

688 **(b).** Dotplot displaying BMP signaling in each cellular cluster organized by the treatment identity for the
689 majority of cells in the cluster. See table S3 for source data containing log fold change and False
690 discovery rate for each pairwise comparison by Wilcoxon Rank Sum Test with Holm-middle combined p
691 value.

692 **(c).** STREAM pseudotime trajectory of hiPS differentiation for hiPS and TS D2 cells visualizing cells
693 proportion along the smoothed pseudotime axis (orange represents TS, green represents hiPS). Nodes
694 are labeled with numbers 0-4. Branches are defined as the cells between 2 nodes (Branch 1,4). Bottom:
695 Top ranked (by STAT) genes with expression significantly correlated with the pseudotime transition of the
696 branches. MEIS2, TFAP2A, and less differentiated TS d2 to more differentiated TSd2 (S0-S1) KRT18
697 (STAT 0.692). Ranks 2-4 for these branches is found in figure S6 c-d. Additional graphs showing the
698 branch assignment with cells and a cell level representation of the pseudotime with the clusters figure S7
699 and table S7.

700

701

702

703 Bibliography

704 1. Rossant, J. Genetic Control of Early Cell Lineages in the Mammalian Embryo. *Annu. Rev. Genet.*
705 **52**, 185–201 (2018).
706 2. Al-Haddad, B.J.S. *et al.* (2019) The fetal origins of mental illness. *Am. J. Obstet. Gynecol.* 221,
707 549–562
708 3. Okae, H. *et al.* Derivation of Human Trophoblast Stem Cells. *Cell Stem Cell* **22**, 50-63.e6 (2018).
709 4. Pera, M. F. Pluripotent cell states and unexpected fates. Stem Cell Reports 17, 1235–1236
710 (2022).
711 5. Gafni, O., Weinberger, L., Mansour, A. *et al.* Derivation of novel human ground state naive
712 pluripotent stem cells. *Nature* 504, 282–286 (2013).
713 6. Hanna, J. *et al.* Human embryonic stem cells with biological and epigenetic characteristics similar
714 to those of mouse ESCs. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 9222–9227 (2010).
715 7. Nichols, J. & Smith, A. Naive and Primed Pluripotent States. *Cell Stem Cell* **4**, 487–492 (2009).
716 8. Soncin, F. et al. Derivation of functional trophoblast stem cells from primed human pluripotent
717 stem cells. Stem Cell Reports 17, 1303–1317 (2022).
718 9. Seetharam, A. S. *et al.* The product of BMP-directed differentiation protocols for human primed
719 pluripotent stem cells is placental trophoblast and not amnion. *Stem Cell Reports* **17**, 1289–1302
720 (2022).
721 10. Viukov, S. *et al.* Human primed and naïve PSCs are both able to differentiate into trophoblast
722 stem cells. *Stem Cell Reports* **17**, 2484–2500 (2022).
723 11. Rostovskaya M, Andrews S, Reik W, Rugg-Gunn PJ (2022) Amniogenesis occurs in two
724 independent waves in primates. Cell Stem Cell 29:744-759.e6.
725 12. Liu, X., Ouyang, J.F., Rossello, F.J. *et al.* Reprogramming roadmap reveals route to human
726 induced trophoblast stem cells. *Nature* 586, 101–107 (2020).
727 13. Roberts, R. M. *et al.* The role of BMP4 signaling in trophoblast emergence from pluripotency.
728 *Cell. Mol. Life Sci.* **79**, 447 (2022).
729 14. Xu, R.-H. *et al.* BMP4 initiates human embryonic stem cell differentiation to trophoblast. *Nat*
730 *Biotechnol* **20**, 1261–1264 (2002).
731 15. Sudheer, S., Bhushan, R., Fauler, B., Lehrach, H. & Adjaye, J. FGF inhibition directs BMP4-
732 mediated differentiation of human embryonic stem cells to syncytiotrophoblast. *Stem Cells Dev.*
733 **21**, 2987–3000 (2012).
734 16. Amita, M. *et al.* Complete and unidirectional conversion of human embryonic stem cells to
735 trophoblast by BMP4. *Proc Natl Acad Sci USA* **110**, E1212-21 (2013).
736 17. Horii, M., Bui, T., Touma, O., Cho, H. Y. & Parast, M. M. An Improved Two-Step Protocol for
737 Trophoblast Differentiation of Human Pluripotent Stem Cells. *Current Protocols in Stem Cell*
738 *Biology* **50**, (2019).
739 18. Wei, Y. *et al.* Efficient derivation of human trophoblast stem cells from primed pluripotent stem
740 cells. *Sci. Adv.* **7**, eabf4416 (2021).
741 19. Thomson, J. A. *et al.* Embryonic Stem Cell Lines Derived from Human Blastocysts. *Science* **282**,
742 1145–1147 (1998).
743 20. Osafune, K. *et al.* Marked differences in differentiation propensity among human embryonic stem
744 cell lines. *Nat Biotechnol* **26**, 313–315 (2008).
745 21. Sawada, T. *et al.* Generation of four postmortem dura-derived iPS cell lines from four control
746 individuals with genotypic and brain-region-specific transcriptomic data available through the
747 BrainSEQ consortium. *Stem Cell Res.* **46**, 101806 (2020).
748 22. Io, S. *et al.* Capturing human trophoblast development with naive pluripotent stem cells in vitro.
749 *Cell Stem Cell* **28**, 1023-1039.e13 (2021).
750 23. Macosko, E. Z. et al. Highly Parallel Genome-wide Expression Profiling of Individual Cells Using

751 Nanoliter Droplets. Cell 161, 1202–1214 (2015).
752 24. Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902.e21 (2019).
753 25. Liu, Y. *et al.* Single-cell RNA-seq reveals the diversity of trophoblast subtypes and patterns of
754 differentiation in the human placenta. *Cell Res* **28**, 819–832 (2018).
755 26. Vento-Tormo, R. *et al.* Single-cell reconstruction of the early maternal–fetal interface in humans.
756 *Nature* **563**, 347–353 (2018).
757 27. Yabe, S. *et al.* Comparison of syncytiotrophoblast generated from human embryonic stem cells
758 and from term placentas. *Proc. Natl. Acad. Sci. U.S.A.* **113**, (2016).
759 28. Niwa, H. *et al.* Interaction between Oct3/4 and Cdx2 Determines Trophectoderm Differentiation.
760 *Cell* **123**, 917–929 (2005).
761 29. Erwin, J. A., del Rosario, B., Payer, B. & Lee, J. T. An *ex Vivo* Model for Imprinting: Mutually
762 Exclusive Binding of Cdx2 and Oct4 as a Switch for Imprinted and Random X-Inactivation.
763 *Genetics* **192**, 857–868 (2012).
764 30. Niakan, K. K. & Eggan, K. Analysis of human embryos from zygote to blastocyst reveals distinct
765 gene expression patterns relative to the mouse. *Developmental Biology* **375**, 54–64 (2013).
766 31. Kazanskaya, O. *et al.* R-Spondin2 Is a Secreted Activator of Wnt/β-Catenin Signaling and Is
767 Required for Xenopus Myogenesis. *Developmental Cell* **7**, 525–534 (2004).
768 32. Hendee, K. E. *et al.* PITX2 deficiency and associated human disease: insights from the zebrafish
769 model. *Human Molecular Genetics* **27**, 1675–1695 (2018).
770 33. Chen, I.-H. *et al.* PRSS23 is essential for the Snail-dependent endothelial-to-mesenchymal
771 transition during valvulogenesis in zebrafish. *Cardiovascular Research* **97**, 443–453 (2013).
772 34. Oh, Y. *et al.* Synthesis and Characterization of Insulin-like Growth Factor-binding Protein
773 (IGFBP)-7. *Journal of Biological Chemistry* **271**, 30322–30325 (1996).
774 35. Overgaard, M. T. *et al.* Pregnancy-associated Plasma Protein-A2 (PAPP-A2), a Novel Insulin-like
775 Growth Factor-binding Protein-5 Proteinase. *Journal of Biological Chemistry* **276**, 21849–21853
776 (2001).
777 36. Wang, R. *et al.* MUC15 Inhibits Dimerization of EGFR and PI3K–AKT Signaling and Is Associated
778 With Aggressive Hepatocellular Carcinomas in Patients. *Gastroenterology* **145**, 1436-1448.e12
779 (2013).
780 37. Porazinski, S. *et al.* YAP is essential for tissue tension to ensure vertebrate 3D body shape.
781 *Nature* **521**, 217–221 (2015).
782 38. Mischler, A. *et al.* Two distinct trophectoderm lineage stem cells from human pluripotent stem
783 cells. *Journal of Biological Chemistry* **296**, 100386 (2021).
784 39. Xiang, L. *et al.* A developmental landscape of 3D-cultured human pre-gastrulation embryos.
785 *Nature* **577**, 537–542 (2020).
786 40. Cinkornpumin, J. K. *et al.* Naive Human Embryonic Stem Cells Can Give Rise to Cells with a
787 Trophoblast-like Transcriptome and Methylome. *Stem Cell Reports* **15**, 198–213 (2020).
788 41. Dong, C. et al. Derivation of trophoblast stem cells from naïve human pluripotent stem cells. eLife
789 9, e52504 (2020).
790 42. Roost, M. S. *et al.* KeyGenes, a Tool to Probe Tissue Differentiation Using a Human Fetal
791 Transcriptional Atlas. *Stem Cell Reports* **4**, 1112–1124 (2015).
792 43. Xu, X., Wells, A. B., O'Brien, D. R., Nehorai, A. & Dougherty, J. D. Cell Type-Specific Expression
793 Analysis to Identify Putative Cellular Mechanisms for Neurogenetic Disorders. *Journal of*
794 *Neuroscience* **34**, 1420–1431 (2014).
795 44. Petropoulos, S. *et al.* Single-Cell RNA-Seq Reveals Lineage and X Chromosome Dynamics in
796 Human Preimplantation Embryos. *Cell* **165**, 1012–1026 (2016).
797 45. Stirparo, G. G. et al. Integrated analysis of single-cell embryo data yields a unified transcriptome
798 signature for the human preimplantation epiblast. Development dev.158501 (2018)
799 doi:10.1242/dev.158501.

46. Messmer, T. *et al.* Transcriptional Heterogeneity in Naive and Primed Human Pluripotent Stem Cells at Single-Cell Resolution. *Cell Reports* **26**, 815-824.e4 (2019).
47. Chen, H. *et al.* Single-cell trajectories reconstruction, exploration and mapping of omics data with STREAM. *Nat Commun* **10**, 1903 (2019).
48. Abd El-Aleem, S. A. & Dekker, L. V. Assessment of the cellular localisation of the annexin A2/S100A10 complex in human placenta. *J Mol Hist* **49**, 531–543 (2018).
49. Liu, A.-X. *et al.* Proteomic Analysis on the Alteration of Protein Expression in the Placental Villous Tissue of Early Pregnancy Loss. *Biology of Reproduction* **75**, 414–420 (2006).
50. Fukamachi, K. et al. Evaluation of a biomarker for the diagnosis of pancreas cancer using an animal model. J Toxicol Pathol 32, 135–141 (2019).
51. Guan, L. *et al.* MEIS2C and MEIS2D promote tumor progression via Wnt/β-catenin and hippo/YAP signaling in hepatocellular carcinoma. *J Exp Clin Cancer Res* **38**, 417 (2019).
52. Krendl, C. et al. GATA2/3-TFAP2A/C transcription factor network couples human pluripotent stem cell differentiation to trophectoderm with repression of pluripotency. Proc. Natl. Acad. Sci. U.S.A. 114, (2017).
53. Huang, D., Wang, S., Wang, A., Chen, X. & Zhang, H. Thymosin beta 4 silencing suppresses proliferation and invasion of non-small cell lung cancer cells by repressing Notch1 activation. *ABBS* **48**, 788–794 (2016).
54. Lee, C. Q. E. *et al.* A niche of trophoblast progenitor cells identified by integrin α2 is present in first trimester human placentas. *Development* dev.162305 (2018) doi:10.1242/dev.162305.
55. Lv, S., Cheng, G., Zhou, Y. & Xu, G. Thymosin beta4 induces angiogenesis through Notch signaling in endothelial cells. *Mol Cell Biochem* **381**, 283–290 (2013).
56. Johnson, W. E. Origins and evolutionary consequences of ancient endogenous retroviruses. *Nat Rev Microbiol* **17**, 355–370 (2019).
57. Frank, J. A. *et al.* Evolution and antiviral activity of a human protein of retroviral origin. *Science* **378**, 422–428 (2022).
58. Lavialle, C. *et al.* Paleovirology of ' *syncytins* ', retroviral *env* genes exapted for a role in placentation. *Phil. Trans. R. Soc. B* **368**, 20120507 (2013).
59. Mangeney, M. *et al.* Placental syncytins: Genetic disjunction between the fusogenic and immunosuppressive activity of retroviral envelope proteins. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 20534–20539 (2007).
60. Theunissen, T. W. *et al.* Molecular Criteria for Defining the Naive Human Pluripotent State. *Cell Stem Cell* **19**, 502–515 (2016).
61. Glass, K., Huttenhower, C., Quackenbush, J. & Yuan, G.-C. Passing Messages between Biological Networks to Refine Predicted Interactions. *PLoS ONE* **8**, e64832 (2013).
62. Weinberger, L., Ayyash, M., Novershtern, N. & Hanna, J. H. Dynamic stem cell states: naive to primed pluripotency in rodents and humans. *Nat Rev Mol Cell Biol* **17**, 155–169 (2016).
63. Daniszewski, M. *et al.* Single-Cell Profiling Identifies Key Pathways Expressed by iPSCs Cultured in Different Commercial Media. *iScience* **7**, 30–39 (2018).
64. Pastor, W.A. *et al.* (2016) Naive human pluripotent cells feature a methylation landscape devoid of blastocyst or germline memory. *Cell Stem Cell* 18, 323–329
65. Theunissen, T.W. *et al.* (2016) Molecular criteria for defining the naive human pluripotent state. *Cell Stem Cell* 19, 502–515
66. Tunster, S. J., Jensen, A. B. & John, R. M. Imprinted genes in mouse placental development and the regulation of fetal energy stores. *Reproduction* **145**, R117-37 (2013).
67. Myatt, L. Placental adaptive responses and fetal programming: Placenta and fetal programming. *The Journal of Physiology* **572**, 25–30 (2006).
68. Ursini, G. *et al.* Placental genomic risk scores and early neurodevelopmental outcomes. *Proc Natl Acad Sci USA* **118**, (2021).

849 69. Guleria, I. and Pollard, J.W. (2000) The trophoblast is a component of the innate immune system
850 during pregnancy. *Nat. Med.* 6, 589–593.