GATA2/3-TFAP2A/C transcription factor network couples human pluripotent stem cell differentiation to trophectoderm with repression of pluripotency

Christian Krendl[a,1], Dmitry Shaposhnikov[a,1], Valentyna Rishko[a], Chaido Ori[a], Christoph Ziegenhain[b], Steffen Sass[c], Lukas Simon[c], Nikola S. Müller[c], Tobias Straub[d], Kelsey E. Brooks[e], Shawn L. Chavez[e], Wolfgang Enard[b], Fabian J. Theis[c,f], and Micha Drukker[a,2]

[a]Institute of Stem Cell Research, Helmholtz Center Munich, 85764 Neuherberg, Germany; [b]Anthropology and Human Genomics, Department of Biology II, Ludwig-Maximilians University Munich, 82152 Martinsried, Germany; [c]Institute of Computational Biology, Helmholtz Center Munich, 85764 Neuherberg, Germany; [d]Biomedical Center, Core Facility Bioinformatics, Ludwig-Maximilians University Munich, 82152 Martinsried, Germany; [e]Division of Reproductive and Developmental Sciences, Oregon National Primate Research Center, Oregon Health and Science University, Beaverton, OR 97006; and [f]Department of Mathematics, Technical University Munich, 85748 Garching, Germany

Edited by R. Michael Roberts, University of Missouri, Columbia, MO, and approved September 25, 2017 (received for review May 23, 2017)

To elucidate the molecular basis of BMP4-induced differentiation of human pluripotent stem cells (PSCs) toward progeny with trophectoderm characteristics, we produced transcriptome, epigenome H3K4me3, H3K27me3, and CpG methylation maps of trophoblast progenitors, purified using the surface marker APA. We combined them with the temporally resolved transcriptome of the preprogenitor phase and of single APA+ cells. This revealed a circuit of bivalent TFAP2A, TFAP2C, GATA2, and GATA3 transcription factors, coined collectively the "trophectoderm four" (TEtra), which are also present in human trophectoderm in vivo. At the onset of differentiation, the TEtra factors occupy multiple sites in epigenetically inactive placental genes and in *OCT4*. Functional manipulation of *GATA3* and *TFAP2A* indicated that they directly couple trophoblast-specific gene induction with suppression of pluripotency. In accordance, knocking down *GATA3* in primate embryos resulted in a failure to form trophectoderm. The discovery of the TEtra circuit indicates how trophectoderm commitment is regulated in human embryogenesis.

trophectoderm | trophoblast | BMP4 | hESC | differentiation

The earliest cell fate commitment event that takes place during eutherian embryogenesis is the bifurcation of totipotent cells into the inner cell mass that generates the fetus, and trophectoderm (TE) precursors that give rise to the chorion and subsequently the fetal portion of the placenta (1). Studies of TE specification in the mouse revealed the importance of the transcription factors (TFs) *Tead4* (2, 3), *Cdx2* (4, 5), *Eomes* (4, 6), *Gata3* (7), and *AP-2γ (Tfap2c)* (8, 9). Further differentiation of the precursors involves TFs such as the placenta morphogenesis master regulator *Gcm1*, *Elf5*, which promotes self-renewal of mouse trophoblast stem cells (TSCs), and *Hand1* and *Mash2* that regulate giant cell and spongiotrophoblast development, respectively (10–13). The expression of Cdx2 in the outer layer cells of the embryo, which are destined to become trophoblasts, is thought to antagonize pluripotency by interfering with *Oct4* autoregulation (5). In accordance with these key roles, overexpression of *Tead4*, *Cdx2*, *Eomes*, or *Gata3* in mouse embryonic stem cells (ESCs) is sufficient to drive them toward the TE fate (5, 7, 14). Recently, it has also been shown that ectopic expression of *Tfap2c*, *Gata3*, *Eomes*, and either *Myc* or *Ets2* converts mouse fibroblasts to functional trophoblast stem-like cells (15–17).

The molecular mechanism of TE specification in humans has not been elucidated, but expression studies have shown that orthologs of some of the key TFs implicated in mouse TE development, including *CDX2*, *TFAP2C*, *GCM1*, and *GATA3*, are found in human TE progenitors at the blastocyst stage (18–21). The early expression of *TFAP2* TFs could have been inferred from deregulation of their target genes in cases of placental dysfunction (22). Other mouse TFs, however, like *Eomes* and *Elf5*, were not unequivocally assigned to the TE lineage in human embryos (23).

Human ESCs (hESCs) and induced pluripotent stem cells (iPSCs) can differentiate into trophoblast-like cells by treatment with BMP4, BMP5, BMP10, or BMP13 (24–29). Although these morphogens were not initially implicated in trophoblast development in the mouse (30, 31), it was recently found that components of the BMP signaling pathway are indeed differentially expressed in TE-fated cells immediately following the first wave of asymmetric divisions in mouse embryos, and that BMP signaling is required for development of mouse TE in vivo (32). The treatment of human ESCs by BMPs also gives rise to mesoderm lineages, but this process, unlike derivation of trophoblasts, is known to be Wnt dependent (33).

Exposure of human ESCs and iPSCs (collectively PSCs) to BMPs triggers induction of a broad cohort of TFs including *CDX2*, *GATA2*, *GATA3*, *TFAP2A*, *TFAP2C*, *MSX2*, *SSI3*, *HEY1*, *GCM1*, and others, the majority of which by analogy to mouse

Significance

This study provides a mechanistic explanation for the differentiation of trophoblasts from human pluripotent stem cells, a process relying on BMP morphogens. We found that a network of the transcription factors GATA2, GATA3, TFAP2A, and TFAP2C regulates early trophoblast progenitor specification by activating placental genes and inhibiting the pluripotency gene *OCT4*, thus acting to couple trophoblast specification with exit from pluripotency. To demonstrate the relevance of our findings in vivo, we show that down-regulating *GATA3* in primate embryos prevents trophectoderm specification. In addition, we present a genome-wide analysis of active and inactive chromatin during trophoblast progenitor specification. These results provide a basis to guide investigations of human trophectoderm development.

Author contributions: C.K., D.S., S.L.C., and M.D. designed research; C.K., D.S., V.R., C.O., C.Z., and K.E.B. performed research; W.E. and F.J.T. contributed new reagents/analytic tools; C.K., D.S., S.S., L.S., N.S.M., T.S., K.E.B., F.J.T., and M.D. analyzed data; and C.K., D.S., S.L.C., and M.D. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

This open access article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

Data deposition: The raw sequencing data reported in this paper have been deposited in the Gene Expression Omnibus (GEO) database (accession nos. GSE104818, GSE104969, GSE105022, GSE105081, and GSE105258).

[1]C.K. and D.S. contributed equally to this work.

[2]To whom correspondence should be addressed. Email: micha.drukker@helmholtz-muenchen.de.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1708341114/-/DCSupplemental.

knowledge could be considered candidate factors involved in human TE specification (29, 34). However, it has yet to be determined which of these TFs actually participate in the initial specification of human trophoblast progenitors, how the TFs are configured in a circuit that drives further trophoblast development by transcription of placenta-specific genes, and to what extent this network governs primate TE specification in vivo.

The colocalization of the transcription promoting and inhibiting trimethylation modifications on lysine 4 and 27 of the histone H3 tail, namely H3K4me3 and H3K27me3, respectively, contributes to transcriptional poising of developmental genes in ESCs (35, 36). Upon differentiation and depending on the lineage, the H3K4me3 mark dominates the expressed genes that lose the H3K27me3 mark, and conversely, the H3K27me3 mark is enhanced and the H3K4me3 mark is reduced in nonexpressed genes. These bivalent genes harbor cytosine-guanine dinucleotide (CpG)-dense promoter regions [high-CpG promoters (HCPs)] that are silenced by CpG methylation (37). Interestingly, a distinct class of developmental genes does not exhibit bivalency in ESCs and is characterized by low presence of CpGs in promoters [low-CpG promoters (LCPs)] (36). The mode of activation and repression of this class, which is thought to include tissue-specific regulators and structural genes, is considered distinct from that of HCP genes (38).

We have previously identified a panel of cell surface markers, including aminopeptidase A (APA) (or CD249/Ly-51), LIFR, EGFR, and CD117 (c-kit) that are expressed by a trophoblast progenitor population that emerges as early as 48 h after treatment of human PSCs by BMP4. We showed that sorted APA+, but not APA− cells, display cytotrophoblast characteristics and the capacity to further differentiate into multinucleated fused syncytiotrophoblast-like cells, and to express placental hormones in vivo (26). These surface markers therefore could be instrumental for purifying trophoblast progenitors and discovering mechanisms that regulate initial specification and differentiation along this lineage.

Here, we addressed these questions by transcriptome and epigenome analyses of purified BMP4-treated human ESC-derived trophoblast-fated progenitors as bulk population and single cells. We coupled these data to temporally resolved transcriptome changes that occur in the cells before the progenitors emerge. Our results support the identification of the human PSC-derived trophoblast progeny as TE, rather than extraembryonic mesoderm (33). Moreover, we discovered a TF circuit that could explain the coupling of TE specification with suppression of pluripotency, a finding that was supported by mapping the genome-wide binding sites of the TFs, and by the results of functional manipulation of these TFs in human ESCs in vitro and in primate embryos in vivo. Our results also revealed the modes of epigenetic regulation that govern gene induction and suppression along the differentiation axis of human PSCs to trophoblasts.

Results

The Transcriptome of Human Trophoblast Progenitors. To identify cell-intrinsic mechanisms that govern human TE specification in vitro, we used the previously characterized cell surface marker APA (*ENPEP*), which marks trophoblast progenitors that differentiate from human ESCs upon constant exposure to BMP4 (26). The APA+ cell population emerged, peaked, and leveled out when 70–90% of the cells became positive, at differentiation days 2, 3, and 4, respectively (Fig. 1 *A* and *B*). Further culturing using this condition led to production of secreted CG (hCG), a pregnancy hormone expressed by trophoblasts in utero (39) (Fig. 1*C*). This protocol performed equally well in KnockOut serum replacement (KSR)-based and B27-based media (Fig. S1).

To characterize key genes involved in the differentiation of human trophoblast progenitors, we sorted the top 20% brightest APA+ and dimmest APA− cell populations after 60 h (2.5 d) of differentiation, around the time when the size of the APA+ population grows exponentially. To set a baseline for gene expression levels, we sorted the SSEA-5+ cell population from undifferentiated cultures (which includes ∼95% of the cells). This removes spontaneously differentiated cells that can obstruct analysis of cell-intrinsic properties (40). Lineage assessment of these cell populations by qPCR before global transcriptomics analysis indicated a transition from pluripotency to TE fate in the APA+ population evident by *OCT4* down-regulation and a reciprocal *CDX2*, *GCM1*, and *ENPEP (APA)* up-regulation (Fig. S2*A*). Moreover, in the APA− population, we noticed an up-regulation of key mesoderm genes and surface markers, including *T*, *GSC*, *ROR2*, and *CD13*, as well as lower enrichment of trophoblast genes, for example, *GCM1*, indicating that this population consisted of primitive streak-like progenitors and possibly cells in pre-APA phase (Fig. S2*A*).

Next, we globally analyzed differentially expressed (DE) genes in the APA+, APA−, and SSEA-5+ cell populations using Affymetrix oligonucleotide microarrays (Fig. S2*B* and Dataset S1). Comparing APA+, APA−, and SSEA-5+ profiles, we noted ∼700 down- and ∼1,000 up-regulated transcripts (Fig. 1 *D* and *E*, respectively). The cohort of the down-regulated genes included the pluripotency circuitry members *SOX2*, *OCT4*, and *NANOG* (Fig. 1*D*). Importantly, tissue association analysis of the genes that were up-regulated in the APA+ population identified trophoblast and placental tissues as the most overrepresented (Fig. 1*F*; a similar signature was observed when comparing the APA+ and APA− cell populations; Fig. S2*C*). This shows the relevance of the APA+ cell population for identifying key human TE genes. To substantiate this claim, we compared the up-regulated gene set of the APA+ population with the genes that are enriched in human embryonic mural trophoblasts (19), and demonstrated an overlapping set of TFs including *GCM1*, *TP63*, *VGLL1*, *GATA2*, *GATA3*, and *TFAP2C*, as well as the surface marker *ENPEP (APA)*, which was collectively annotated as trophoblast/placenta specific with high confidence (Fig. S2*D*). Finally, the enrichment of *CDX2* and *ELF5* in the APA+ cell population (Fig. S2*A*) is also consistent with commitment to TE fate (5, 13).

To assess the degree of heterogeneity in the APA+ population, we performed global RNA sequencing (RNA-seq) of ∼350 individually sorted APA+ and undifferentiated cells. Analysis of this dataset confirmed that the APA+ population is essentially homogenous, except for fewer than 10% of the cells that clustered with undifferentiated cells (Fig. 1*G* and Fig. S3). Importantly, significantly up-regulated genes in single APA+ cells exhibited associations with trophoblast and placental tissues (Fig. 1*H*), similarly to the bulk APA+ cell population (Fig. 1*F*). Taken together, we concluded that BMP4 treatment of human ESCs leads to the specification of APA+ progenitors that, on the gene expression level, resemble human TE progenitors in vivo and lack pluripotency features, and that the transcriptional network of these early progenitors is enriched for factors which were previously found to be important in mouse TE development.

Histone Modification Redistribution During TE Progenitor Specification. To analyze posttranslational histone modifications that underlie human trophoblast progenitor specification and pluripotency shutdown, we used sorted APA+, APA−, and SSEA-5+ cell populations (matched samples of Fig. 1) to perform chromatin immunoprecipitation and deep sequencing (ChIP-seq) of H3K4me3- and H3K27me3-bound DNA fragments (using validated antibodies; Fig. S4). Comparing APA+ and SSEA-5+ populations, we found that the redistribution of H3K4me3 or H3K27me3 monovalent, bivalent (both marks present), and H3K4me3/K27me3 double-negative genes was markedly different between up- and down-regulated genes (Fig. 2 *A* and *B*, representative maps in Fig. 2*E*). While close to 65% of the genes that were bivalent in undifferentiated SSEA-5+ cells became predominantly H3K4me3 monovalent (by losing the H3K27me3 mark) in the case they were up-regulated, there was a little change in bivalency for genes that are down-regulated. Also, only around 25% of the genes from the H3K4me3 class became bivalent when they are down-regulated. Taken together, this indicates that transcriptional changes often precede changes in histone modifications, both in the direction of gene induction and repression. This conclusion is supported by a group of up-regulated

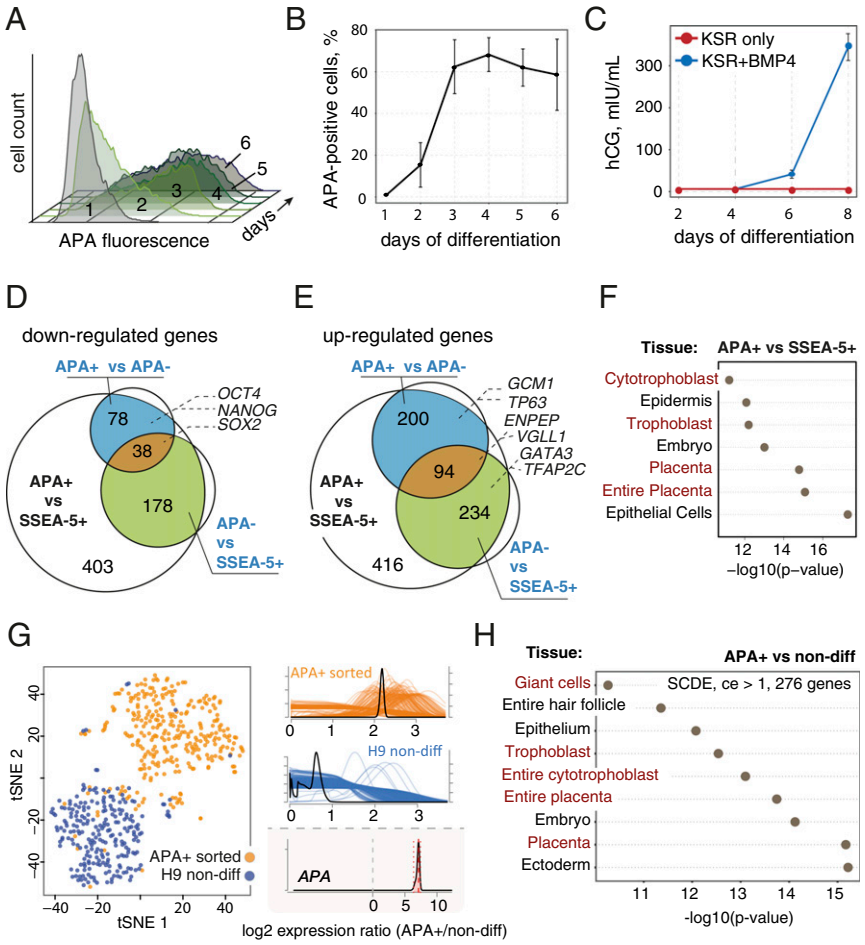

Fig. 1. Purification and characterization of human APA+ progenitors. (*A*) Representative time-course flow cytometry analysis of APA in BMP4-treated hESCs. (*B*) Quantification of the temporal dynamics of the APA+ cells over the course of 6 d (n = 2; mean ± SEM). (*C*) Time-course measurements of the secreted CG (hCG) in hESCs with and without BMP4 in differentiation medium (n = 2; mean ± CI, 95%). (*D* and *E*) Venn diagrams showing overlapping microarray probe sets that were down-regulated (*D*) and up-regulated (*E*) in sorted APA+, APA– cells after 2.5 d of BMP4 treatment [n = 3; false-discovery rate (FDR): adjusted P value < 0.05; fold change ≥ 2]. (*F*) Significant tissue and cell type associations of the up-regulated (FDR: adjusted P value < 0.05; fold change ≥ 2) microarray probe sets in APA+ cells compared with SSEA5+ hESCs, based on the literature mining algorithm of the Genomatix GeneRanker tool. The lowest $\log_{10}$ P values out of three replicates are displayed. (*G*) t-SNE plot of the single-cell RNA-seq data from FACS-sorted APA+ cells (72 h of BMP4 treatment) and undifferentiated hESCs. Gene expression count estimates of the *ENPEP* (APA) gene are shown for APA+ cells in orange, and for undifferentiated hESCs in blue. Estimated logtwo-fold change of *ENPEP* is shown at the *Bottom*. (*H*) Significant tissue and cell type associations of the up-regulated genes [SCDE, conservative fold change estimate ($\log_2$) > 1,276 genes] in the APA+ cells compared with undifferentiated hESCs from single cell RNA-seq data, analyzed as in *F*.

genes that maintained the H3K4me3/K27me3 double-negative phenotype (Fig. 2*A*).

To gain specific insights into the relationship between chromatin states and transcriptional regulation in TE progenitors, we next focused on transcription factors/cofactors (TFs collectively) that were DE between the APA+ and the SSEA-5+ cell population. With respect to the up-regulated TFs, we found that orthologs of key mouse trophoblast-related TFs belong to the bivalent or double-negative class in undifferentiated cells (Fig. 2 *C* and *E* and Dataset S2). This included *CDX2*, *GATA3*, *TFAP2C* and *GCM1*, *VGLL1*, *TP63*, and *ELF5*, respectively. Moreover, a cohort of genes that is specific for trophoblast and placental tissues, including steroid sulfatase (*STS*) and solute carriers *SLC13A4* and *SLC8A1*, was found in the double-negative class (Fig. 2*C*). TFs that were down-regulated in the APA+ cell population, including *SOX2*, *NANOG*, and *OCT4* in most part belonged to the H3K4me3-monovalent class in undifferentiated cells (Fig. 2*D*). Intriguingly, *NANOG* and *OCT4* were among the rare genes in this class that lost H3K4me3 mark without gaining H3K27me3 mark in the APA+ cells (Fig. 2 *B* and *E*).

Taken together, this argues that distinct cohorts of regulators which participate in the process of human trophoblast specification are held in divergent chromatin states in undifferentiated human ESCs: one cohort includes bivalent TFs that become predominantly H3K4me3 monovalent in trophoblast progenitors, and a second cohort comprises trophoblast-specific TFs and other genes that are mostly double negative. Interestingly, the initial transcriptional up-regulation of the latter group took place without H3K4me3 histone mark changes.

Temporal Activation of TFs During TE Progenitor Specification. To analyze the order of TF activation of the DE genes in the APA+ progenitors (Fig. 1*E*), we conducted time-series analysis of global transcriptome changes in bulk human ESCs during the first 72 h of BMP4 treatment. The gradual up-regulation and high percentage of overlap between the datasets of DE genes at 48 and 72 h of differentiation, relative to the APA+ progenitor cells, substantiated the use of the time-series information to deduce transcriptional trajectories of genes that are pertinent for TE progenitor specification (Fig. 3 *A* and *B* and Dataset S3). Analysis of the up-regulated TFs according to their temporal trends identified several distinctly clustered groups (Fig. S5*A*), of which three exhibited early (8 h), intermediate (24 h), and late (48 h) induction phases (Fig. 3*C*). Read coverage plots of representative genes from these clusters, including *GATA3*, *CDX2*, and *GCM1*, respectively, as well as of *OCT4* are depicted in Fig. 3*B*. Next, we analyzed the TFs from each group with respect to overrepresented classes of histone modifications in undifferentiated SSEA-5+ cells. This showed that early genes were predominantly bivalent, while the late genes tend to be (although not reaching statistical significance) H3K4me3/K27me3 double negative and bivalent (Fig. 3*D*). The cluster of the intermediate genes did not exhibit a specific overrepresentation of any histone class. While inspecting the TFs from the three groups, we noted that many of the genes that had the highest induction amplitudes (>32-fold at the given time points) are human orthologs of genes that had been previously implicated in mouse TE development (Fig. 3*E*). This included bivalent (in SSEA-5+ cells) early TFs, namely *GATA2*, *GATA3*, *TFAP2A*, and *TFAP2C* (the latter exhibited only 10-fold up-regulation), which have known TE-specific functions (7, 9, 41–44); intermediate-group bivalent TFs, *HAND1* and *CDX2*, which are essential for placental development (4, 11); and late H3K4me3/K27me3 double-negative placental TFs, *GCM1*, *VGLL1*, and *TP63* (45). This further

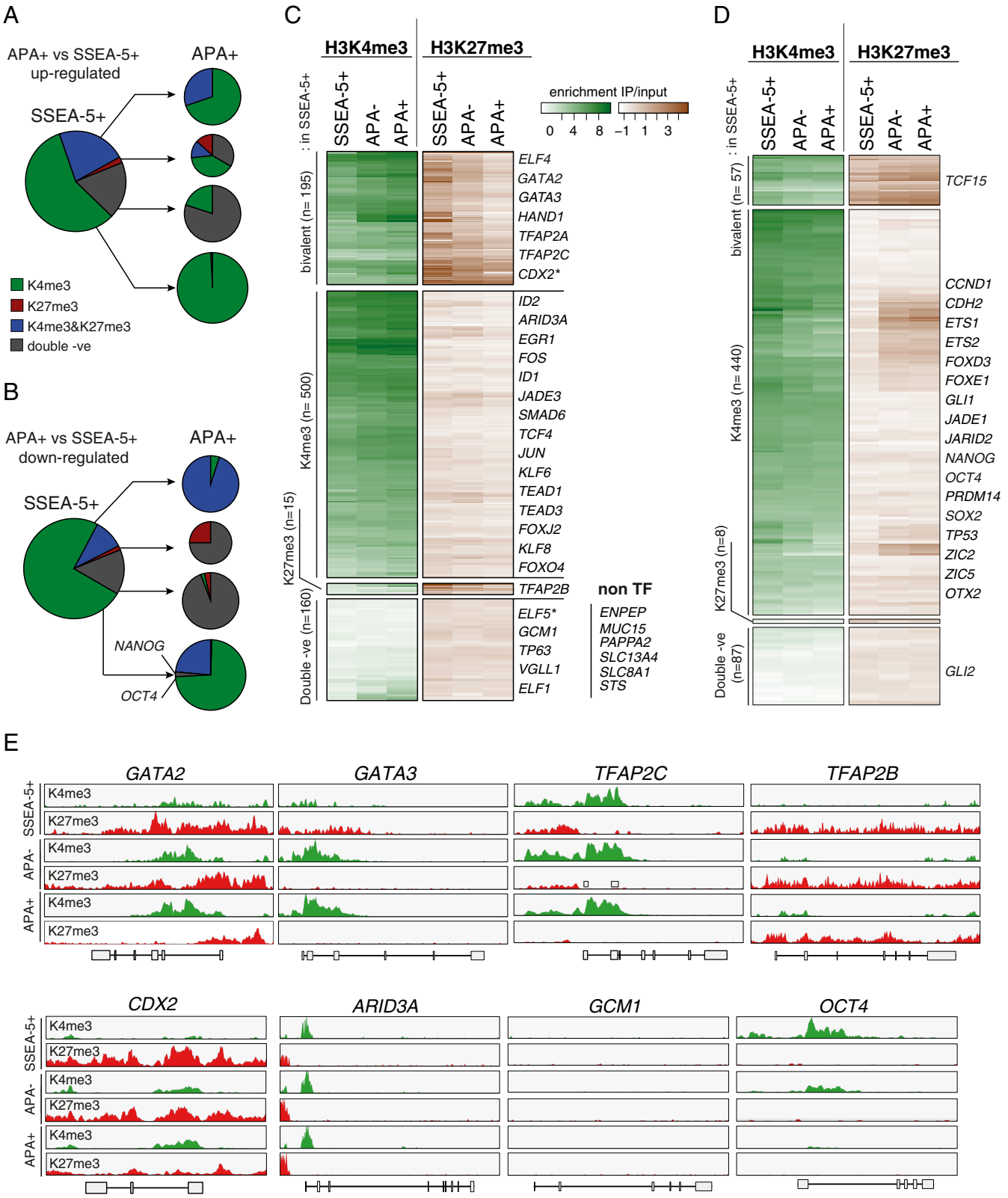

Fig. 2. Changes in the histone modification patterns during APA+ cell differentiation. (*A* and *B*) Histone modification pattern turnover in the sorted APA+ cells compared with SSEA-5+ cells. Plotted are DE genes between the APA+ and the SSEA-5+ cells (870 up- and 592 down-regulated genes, no fold-change cut-off). Histone modification classes are based on the enrichment of H3K4me3 and H3K27me3 in the 4-kb region around known TSS ($n = 3$). (*C* and *D*) Heatmaps of histone mark enrichment in promoters of up-regulated (*C*) and down-regulated (*D*) genes in the APA+ cells compared with SSEA-5+ cells (from Fig. 1 *D* and *E*, respectively). TFs are listed on the side. An *Inset* in *C* shows double-negative non-TF genes that were significantly associated with placental tissues (Genomatix). (*E*) Representative histone mark coverage profiles that correspond to bivalent (*GATA2*, *GATA3*, *TFAP2C*), monovalent H3K4me3 (*ARID3A*, *OCT4*) and H3K27me3 (*TFAP2B*, *CDX2*), and double-negative (*GCM1*) loci in the SSEA-5+ cells. The *y* axis was normalized for each condition/histone modification combination across all genes.

strengthens the notion that the APA+ cell population represents bona fide early TE progenitors. The expression of *MSX2* from the early group was not analyzed further in this context because its early induction is common to other early progenitors (Fig. S5*B*). Analysis of *GATA2*, *GATA3*, *TFAP2A*, and *TFAP2C* in single cells, as well as immunocytochemistry of the respective proteins, indicated that this TE phenotype represents the intrinsic properties of individual cells (Fig. 3*F* and Fig. S5*C*).

High- and low-CpG island promoters (HCPs, LCPs) have been correlated with early embryonic or tissue-specific gene expression, respectively (38). To determine whether the methylation status of the HCP and LCP changes during APA+ progenitor specification, we performed a genome-wide analysis of CpG methylation. Surprisingly, we could not detect significant changes of the methylation states in CpGs of bivalent and H3K4me3/K27me3 double-negative TF genes, which are respectively low and high, comparing the APA+, APA−, and undifferentiated SSEA-5+ cell population (Fig. 3*G*).

Based on these data, we hypothesized that early TE progenitors are specified rather quickly and exit the state of pluripotency by an input from a group of early-response TFs (induced by 8 h) that are bivalent with unmethylated CpG islands in undifferentiated cells, and include *GATA2*, *GATA3*, *TFAP2A*, and *TFAP2C*. Furthermore, these TFs are very likely to govern the induction of late (by 48 h) placenta-specific LCP genes, which are H3K4me3/K27me3 double negative with methylated CpG in undifferentiated state. We postulated that the intermediate group is more heterogeneous with respect to the transcriptional activation mechanisms.

Global Mapping of TFAP2A/C and GATA2/3 TF-Bound Loci. To delineate the binding landscape of the early TFs in TE progenitors, we used human ESCs that were differentiated for 72 h to perform ChIP-seq of GATA2, GATA3, TFAP2A, and TFAP2C. De novo search produced motifs that closely resemble those previously published for these factors (Fig. 4*A*). Next, listing of putative target genes, according to TF binding in −3.5/+5 kb around the transcription start site (TSS), revealed that when more TFs were bound, the correlation to transcriptional up-regulation was higher. Conversely, the transcriptional down-regulation was negatively correlated with the number of bound TFs (Fig. 4*B*). Moreover, GATA3 exhibited the broadest potential of synergy with the other three TFs because it coincided with the other TFs more frequently (Fig. 4*C*). Taken together, this suggests that binding of GATA2/3, TFAP2A/C is predominantly gene activating, and that GATA3 if the chief member of the four TFs, which collectively promote trophoblast differentiation.

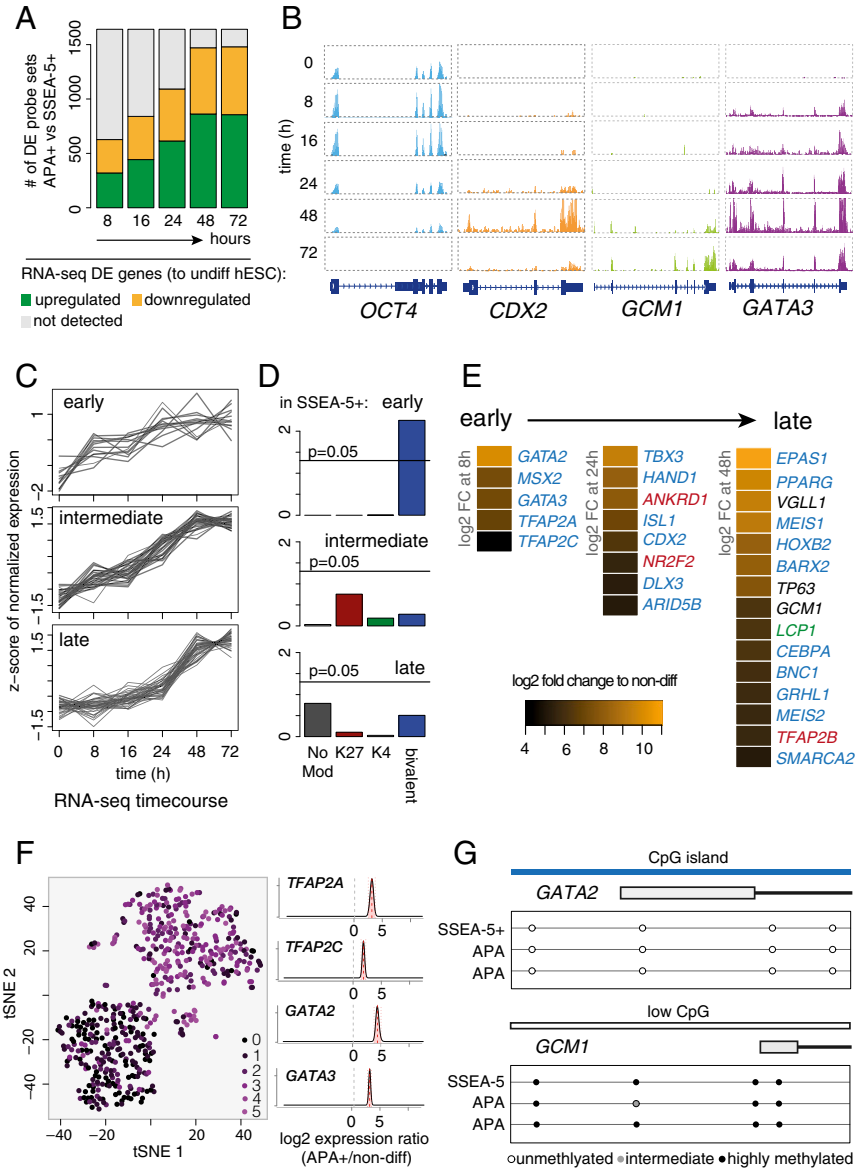

Fig. 3. Expression kinetics of TE regulators and their epigenetic characteristics. (*A*) Correspondence of the microarray dataset to the time-course RNA-seq analysis of bulk BMP4-treated hESCs. Bars represent the total number of APA+ versus SSEA-5+ DE genes (n = 3; FDR: adjusted P value < 0.05). Green and yellow represent the number of DE genes that are respectively up- or down-regulated in the indicated time point (8, 16, 24, 48, and 72 h) of BMP4 treatment (every time point compared with undifferentiated hESCs; n = 2; FDR: adjusted P value < 0.05). Gray represent genes not detected as DE by RNA-seq. (*B*) Representative RNA-seq read coverage plots (bulk BMP4-treated hESCs) of the pluripotency gene *OCT4*, early-response cluster gene *GATA3*, intermediate cluster gene *CDX2*, and late cluster gene *GCM1*. Per gene, the y-axis scale is equal for all time points. (*C*) Clusters of TFs produced by k-means analysis that exhibit early (8 h), intermediate (24 h), and late (48 h) up-regulation during 72-h treatment of hESCs with BMP4. Plotted are significantly up-regulated TFs in APA+ cells compared with SSEA-5+ cells (Fig. 1*E*). Representative clusters shown here were selected by visual inspection (all clusters in Fig. S5*A*). (*D*) Over-representation analysis of the histone modification classes in genes from *C* (P values, Fisher's exact test for the observed number of overlaps between the genes with the respective histone modifications and the genes in all three clusters). (*E*) TFs from the early, intermediate, and late clusters [$\log_2$ fold change (FC) of ≥5 at 8, 24, and 48 h, with the exception of *TFAP2C* with $\log_2$ FC = 3.42 at 8 h]. A $\log_2$ FC = 4.1 of *ELF5* was detected at 72 h. Font color corresponds to the histone modification classes from *D*. (*F*) t-SNE plot of single-cell RNA-seq dataset (as in Fig. 1*G*) colored according to trophoblast gene expression score (*Materials and Methods*). Estimated logtwofold changes between APA+ cells and undifferentiated hESCs for *TFAP2A*, *TFAP2C*, *GATA2*, and *GATA3* are shown on the *Right*. (*G*) DNA methylation status of the CpG sites around TSS of *GATA2* and *GCM1* genes in SSEA-5+, APA+, and APA− cells (n = 3; unmethylated: 0–20%; intermediate: 20–60%; highly methylated: 61–100%).

Indeed, of the 204 genes that were bound by the four TFs, 122 (60%) were up-regulated in the APA+ cell population (Dataset S4). This includes 11 TFs, among them *CDX2* and *ANKRD1*, which have been implicated in TE specification, and *GCM1* that is bound by three TFs (Fig. 4 *D* and *E*). Importantly, *CDX2* displayed potential binding of all four TFs in the first intron (Fig. 4*E*), which is in line with the previous reports showing that the binding of GATA3 and TFAP2C to the first intron of *Cdx2* in mouse TSCs activates its transcription (46, 47). Interestingly, this intronic site is occupied by OCT4 and NANOG in undifferentiated human ESCs (Fig. S6*B*). The up-regulated non-TF genes that exhibited binding of the four TFs, included *ENPEP*, *STS*, *VTCN1*, and other placenta-specific genes (Fig. 4*D*). In addition, we noted in ∼50% of the possible cases, autofactor or cross-factor interactions between the TFs GATA2/3 and TFAP2A/C (Fig. S6*A*). Strikingly, in the very few genes where promoter CpG demethylation did take place during the differentiation from SSEA-5+ to APA− and APA+ populations, these genes were bound by the four TFs (asterisk-labeled genes in Fig. 4*D*).

Finally, although only two down-regulated TFs were bound by GATA2/3 and TFAP2A/C, this binding is likely important for down-regulation of pluripotency, as it took place in the first intron of *OCT4*, and in *JADE1*, which promotes histone acetylation and was implicated in embryogenesis (48) (Fig. 4 *D* and *E*). Taken together, these data indicate that GATA2/3 and TFAP2A/C TFs form a feedforward network that regulates pluripotency and TE genes.

Functional Validation of GATA3 and TFAP2A in TE Specification. To functionally analyze the roles of *GATA2/3* and *TFAP2A/C* in human TE differentiation, we manipulated the expression of the one factor from either pair that is more likely to have a broader regulation. We chose *GATA3* because it exhibited the highest co-occupancy potential among the four TFs (Fig. 4*C*), and *TFAP2A* since it was induced to a higher extent than *TFAP2C* (Fig. 3*E*). We used the HUES9 human ESC line that carries an inducible

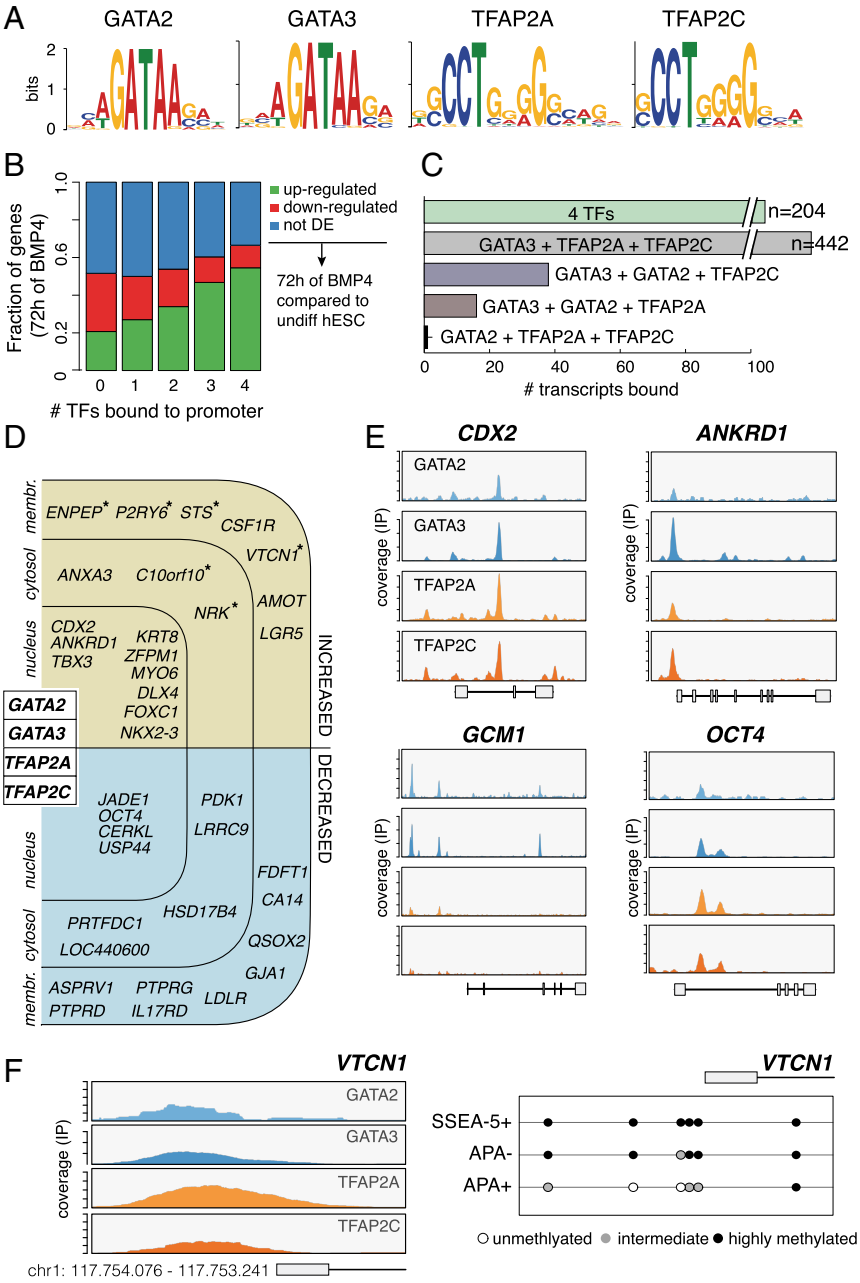

Fig. 4. Identification of GATA2/3 and TFAP2A/C genomic binding sites. (*A*) De novo binding motif analysis of GATA2, GATA3, TFAP2A, and TFAP2C from ChIP-seq of hESCs treated by BMP4 for 72 h. (*B*) Correlation plot of all expressed genes in 72 h of BMP4 treatment (RNA-seq) and the number of TFs (GATA2, GATA3, TFAP2A, and TFAP2C) bound in their promoter regions (−5. . .+3.5 kb around TSS). Color denotes DE status compared with undifferentiated hESCs: up-regulated (green), down-regulated (red), and not DE (blue) ($n = 2$; FDR: adjusted P value < 0.05). (*C*) The total number of gene promoters (−5. . .+3.5 kb around known TSS) that are bound by either all four (GATA2, GATA3, TFAP2A, and TFAP2C) or different combinations of three TFs as determined by ChIP-seq ($n = 3$ for GATA2/3 and TFAP2A; $n = 1$ for TFAP2C). (*D*) An overview of the genes that were bound by all four TFs. Only TFs and genes significantly associated with placental tissues are displayed. Asterisks denote genes with promoter CpG demethylation in the APA+ cells. Binding within the four TEtra TFs is depicted in Fig. S6*A*. (*E*) Representative ChIP-seq coverage profiles of TEtra TFs in the *CDX2*, *ANKRD1*, *GCM1*, and *OCT4* genes. (*F*) An example of ChIP-seq coverage profiles of TEtra TFs in the promoter of the LCP trophoblast-associated gene *VTCN1*. DNA methylation in CpG sites is shown below.

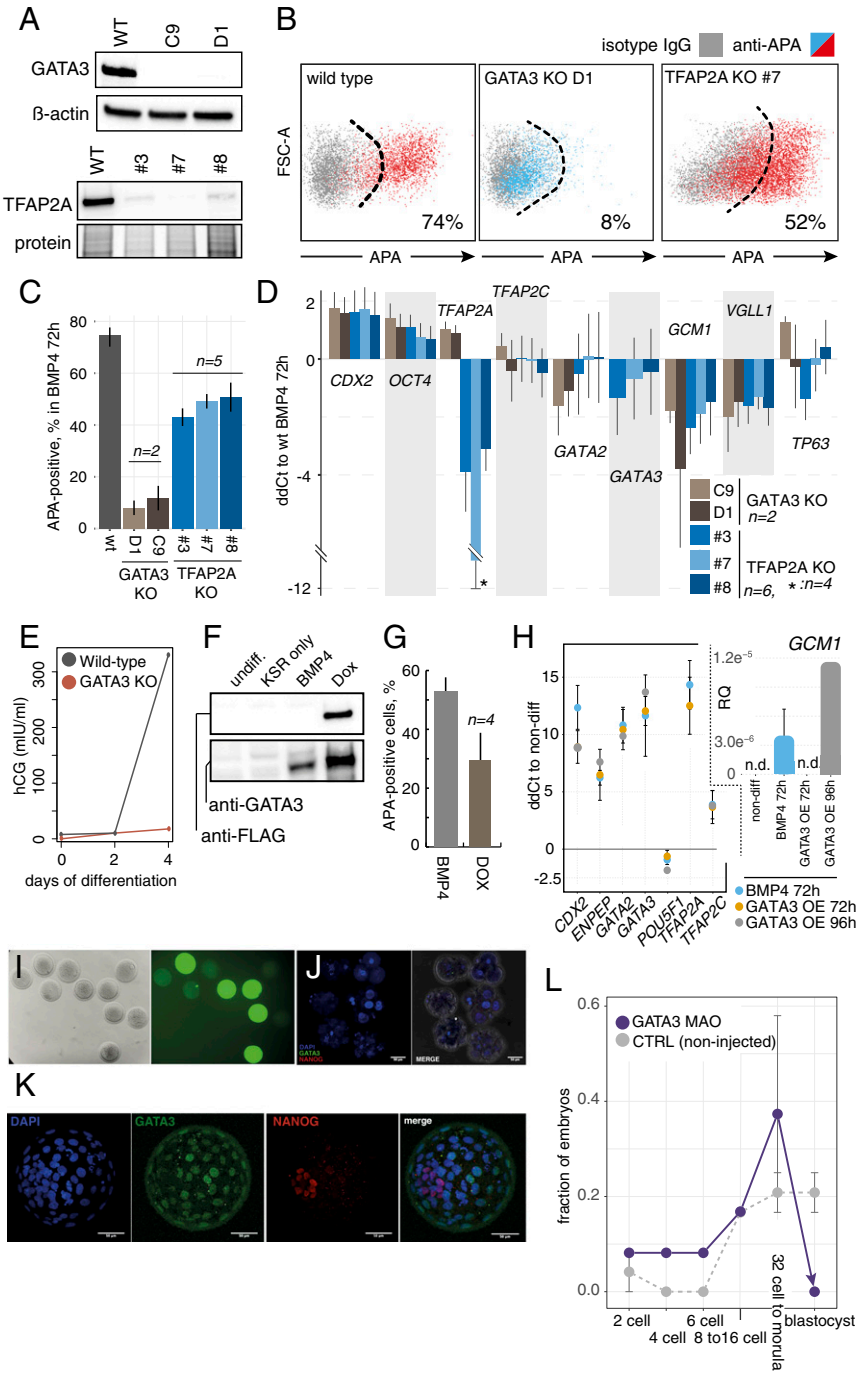

Fig. 5. Functional validation of the human TE specification circuit. (*A*) Western blot analysis of *GATA3* and *TFAP2A* KO clones treated with BMP4 for 72 h. *TFAP2A* clones 3 and 8 demonstrate residual expression (likely due to clonal impurities). (*B*) Representative flow cytometry analysis of APA-expressing cells in *GATA3* and *TFAP2A* KO clones treated with BMP4 for 72 h. The parental iCRISPR cell line was used as the control. (*C*) Quantification of *B*. For *GATA3* KO clones: $n = 2$, mean ± SEM; for wild type (wt) iCRISPR and *TFAP2A* KO clones: $n = 5$, mean ± SEM. (*D*) qPCR analysis of a set of TE genes and *OCT4* in *GATA3* and *TFAP2A* KO clones treated with BMP4 for 72 h. Results are plotted as logtwofold changes (ddCt) compared with identically treated parental cells. [For *GATA3* KO, $n = 2$, mean ± SEM; for *TFAP2A* KO, $n = 6$ (except where indicated), mean ± SEM.] (*E*) Time-course analysis of secreted hCG in *GATA3* KO clones and wild-type iCRISPR cells treated with BMP4 ($n = 2$; mean ± SEM). (*F*) Western blot analysis of an inducible *GATA3* hESC line. Cells were either left in differentiation medium alone (KSR only) or treated with BMP4 or doxycycline (Dox) for 72 h. (*G*) Quantification of the APA+ cells following inducible overexpression of *GATA3* (as in *B*). BMP4-treated hESCs served as the positive control ($n = 4$; mean ± SEM). (*H*) qPCR analysis of a set of TE genes and *OCT4* following overexpression of *GATA3* in differentiation medium for 72 and 96 h, and in 72-h BMP4-treated hESCs. *Inset* shows *GCM1* gene that is first detected after 96 h of Dox treatment ($n = 3$, except 96 h of Dox, where $n = 1$, mean ± SEM; n.d., not detected). (*I*) Representative microinjected rhesus macaque embryos with control 3′-COF–labeled morpholino antisense oligonucleotides (MAO). The intensity of the green signal corresponds to the amount of MAO delivered to each embryo (brightfield on *Left*). (*J*) Rhesus macaque embryos microinjected with *GATA3* MAO and immunostained with GATA3 (green), NANOG (red), and DAPI (blue), shown at various arrested stages. (*K*) Noninjected controls of *J* reached the blastocyst stage and expressed GATA3 in the TE layer, and NANOG in the inner cell mass. (*L*) A summary of the results of MAO microinjection into rhesus macaque embryos according to the developmental stages. Noninjected embryos ($n = 15$), and GATA3 MAO ($n = 19$) from two independent in vitro fertilization experiments (error bars indicate SEM; $n = 2$).

Cas9 cassette (iCRISPR) (49) to delete the boundary of the first intron and the second exon in both genes (325 and 149 bp, respectively). Analysis of differentiated clones (72-h BMP4), harboring homozygous deletions, demonstrated complete absence of GATA3 protein in two clones and of TFAP2A in one clone, and faint bands of TFAP2A in additional two clones (likely due to residual heterozygous or wild-type cells) (Fig. 5*A*). We observed that, while the knockout (KO) of *GATA3* led to a drastic reduction in the number of APA+ progenitors (Fig. 5 *B* and *C*) and hCG levels (Fig. 5*E*), which were comparable to undifferentiated cells, *TFAP2A* KO exhibited a much milder effect, leading to ∼30% reduction in the number of APA+ progenitors (Fig. 5 *B* and *C*).

Next, we analyzed the expression of *OCT4* and early, intermediate, and late group up-regulated genes that overlapped between the time course induction and enrichment in the APA+ population (Fig. 3*D*). In accordance with the direction of regulation during trophoblast differentiation, the up-regulation of *GCM1* and *VGLL1* significantly diminished and *OCT4* increased in the *GATA3*$^{-/-}$ and *TFAP2A*$^{-/-}$ clones that were treated by BMP4 compared with isogenic cells, but *TP63* did not show a clear pattern of deregulation, at least not at the 72-h time point (Fig. 5*D*). Unexpectedly, *CDX2* expression increased more in the KO clones compared with isogenic control cells. Interestingly, the effects of the KOs on the other *GATA/TFAP2* TFs did not follow an obvious pattern; *GATA2* up-regulation decreased, *TFAP2A* increased, and *TFAP2C* did not show significant change in the *GATA3*$^{-/-}$ clones, and in the *TFAP2A*$^{-/-}$ clones the expression of the other three TFs was not significantly altered (Fig.

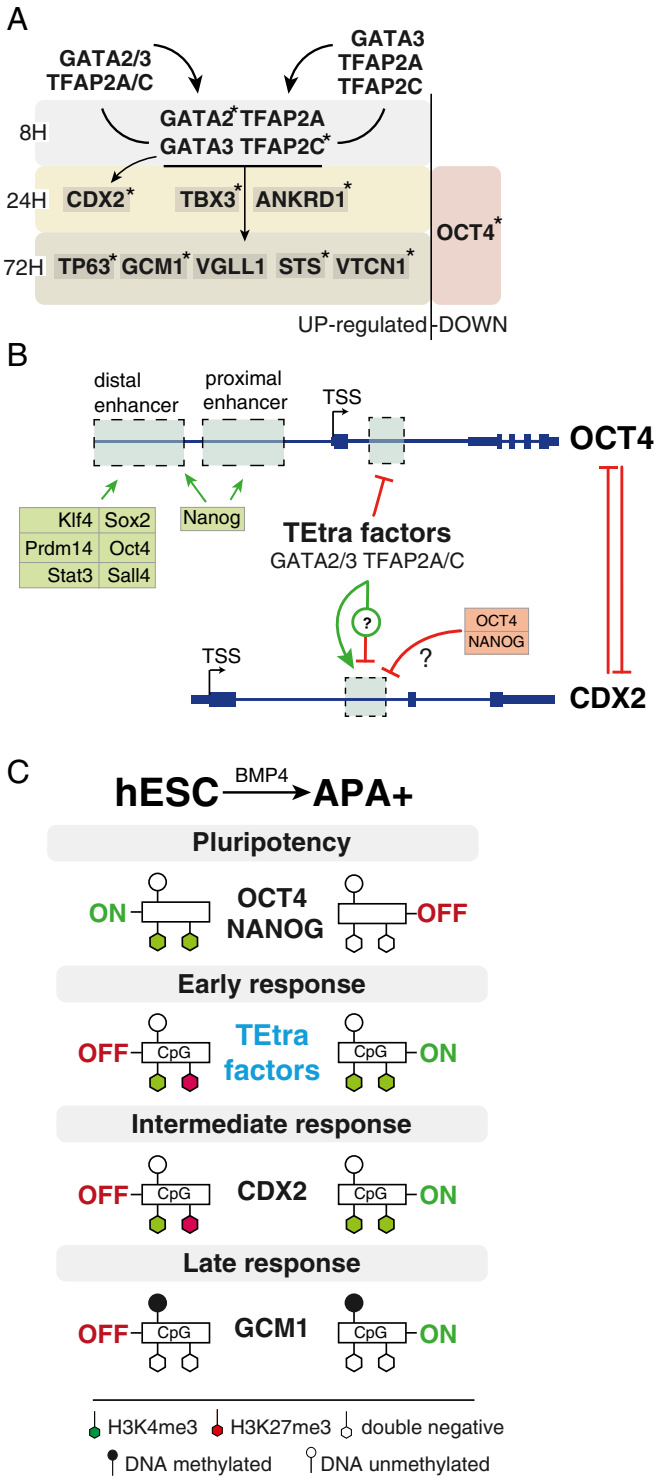

Fig. 6. Proposed models of human TE specification. (*A*) A summary depicting up- and down-regulated TE-specific and pluripotency genes during the time course of BMP4 treatment of hESCs. Reciprocal interactions (binding in the promoter regions) between the TEtra TFs are indicated by arrows, and genes bound by three or more TEtra TFs are labeled with asterisks. (*B*) A model outlining the coupling of *OCT4* and *CDX2* regulation by the TEtra factors during the transition from pluripotency to TE progenitor fate. The unique binding site of GATA2, GATA3, TFAP2A, and TFAP2C is highlighted in the first intron of *OCT4*. *CDX2* is bound by *NANOG* and *OCT4* in hESCs (Fig. S6*A*) in the same region regulated by GATA2, GATA3, TFAP2A, and TFAP2C binding in hESCs that are treated by BMP4. The model includes a hypothetical position of a suppressor of *CDX2* that is induced by the TEtra. (*C*) A model outlining gene expression, histone modifications, and DNA-methylation turnover during the transition from the pluripotency state of undifferentiated hESCs toward APA+ TE progenitor fate. The key pluripotency genes, *OCT4* and *NANOG*, lose the active transcription mark H3K4me3, while the bivalency of early- and intermediate-group genes, which harbor unmethylated HCP, is resolved to the H3K4me3 state. Conversely, late TE regulators, structural and hormone genes, are generally LCP and do not exhibit H3K4me3 and H3K27me3 in undifferentiated cells or following specification to TE progenitors.

5*D*). We postulate that this is due to the complex web of interactions between the GATA and TFAP2 TFs (Fig. S6*A*).

To substantiate the central function of GATA3 in TE specification, we analyzed the outcome of *GATA3* overexpression in human ESCs cells (Fig. 5*F*). This led to a phenotype that closely mimicked 72 h of BMP4 treatment, including the appearance of APA+ progenitors (albeit to a lesser extent), and the up-regulation of *GATA2/3*, *TFAP2A/C*, *CDX2*, and *GCM1*, as well as the down-regulation of *OCT4* (Fig. 5 *G* and *H*). Finally, to validate the role of *GATA3* in primate development in vivo, we knocked down its expression in rhesus macaque embryos. We injected zygotes with morpholino antisense oligonucleotides and monitored the development of the embryos (Fig. 5 *I*–*L*). While control noninjected embryos reached the blastocyst stage, with a typical frequency of macaque embryos in vitro of ∼30%, *GATA3* knockdown led to a failure of blastocyst formation and an embryonic arrest at 32-cell/morula stage (Fig. 5 *J*–*L*). These results suggest that GATA3 is vital for TE specification and embryonic development in primates.

Discussion

In this study, we provide multiple lines of evidence that BMP4 treatment of human ESCs promotes differentiation of trophoblast-like progeny that highly resembles human TE progenitors in vivo. We find that, when the cells exit from the state of pluripotency under this condition, mesendodermal commitment can be detected in the fraction of cells that do not express the cell surface marker APA (Fig. S1). APA had been previously characterized as a trophoblast-specific marker, and sorted APA+ cells were shown to give rise to placental-like tissues upon engraftment in vivo (26). More recently, *APA (ENPEP)* was classified as TE-specific based on its expression in preimplantation human embryos (23, 50). We establish here that gene expression analysis of the APA+ cells reliably classifies them as rather homogenous based on single-cell RNA-seq data, and combined -omics datasets that we produced show that the APA+ population has a pronounced trophoblast gene signature. For example, recently published criteria for defining human mononuclear trophoblasts (39) are fully met by the APA+ cells with respect to expression of marker genes (e.g., *TFAP2C*, *GATA3*, *KRT7*, *ELF5*). We do not, however, detect hypomethylated CpGs in the *ELF5* promoter at day 3 APA+ progenitor state, arguing that the epigenetic remodeling at this locus begins at a later stage. Indeed, Lee et al. have assessed methylation changes at day 6 of BMP4 treatment. Additionally, we find that APA+ cells express the majority of the trophoblast-specific TFs from human mural trophoblasts (19), as well as TE-enriched genes characterized in human preimplantation embryos, including *GATA3*, *CDX2*, *KRT18*, *DAB2*, *EFNA1*, *PPARG*, and *TEAD3* (23, 50). Interestingly, three putative TE-specific markers proposed by Blakeley et al.—*PLAC8*, *CLDN10*, and *TRIML1*—were not detected as up-regulated in the APA+ cells. *PLAC8* and *CLDN10* were transiently up-regulated in bulk BMP4-treated cells from 8 to 24 h and from 24 to 48 h, respectively, arguing that they are not relevant for classifying trophoblasts in vitro, at least using the BMP4 protocol. Overall, based on the gene expression similarities, we are confident to claim that APA marks progenitors closely resembling human embryonic TE, making these cells a suitable model for studying TE progenitor development.

Our combined analysis highlighted three groups of genes with early, intermediate, and late up-regulation phases. We hypothesized that TFs displaying strong induction at the earliest time point of 8 h

and persistent expression at least until the progenitors emerge on day 2.5 of differentiation, could be crucial for governing the transition from pluripotency to the trophoblast fate. Here, we propose that the identified set of four genes—*GATA2*, *GATA3*, *TFAP2A*, and *TFAP2C*—are likely to be the earliest drivers of the trophoblast specification in human ESCs. We named them the trophectoderm four ("TEtra"). Although all four factors have been implicated in the trophoblast development in mouse, and the GATA2–GATA3 redundancy was demonstrated (17, 51), it is important to point out that the earliest known TFs induced during murine TE development—*Tead4* and *Klf5*—were not found to be induced in our analysis (for review of mouse TE commitment, see ref. 52).

Our investigation of the TEtra's chromatin occupancy revealed with high confidence gene-activating binding to the promoters of trophoblast-specific genes from all temporal groups—early (Fig. 6*A*), intermediate (*CDX2*), and late (*GCM1*, *VGLL1*, *TP63*, *STS*, and *ENPEP*) (3, 12, 45, 53–55). In line with the reports of refs. 56 and 57, we found that one member of TEtra—*GATA3*—exhibited certain characteristics of a "pioneer factor." Its promoter was the only one not bound by any other TEtra member, while it occupied the promoters of all of the other TEtra members (except its own, Fig. S6*A*). In addition, it showed the highest degree of chromatin co-occupancy the with other TEtra members (Fig. 4*C*). On the functional level, ablation of GATA3 in human ESCs and in primate embryos in vivo completely abolished the formation of TE progenitors, while overexpression of *GATA3* in human ESCs produced a phenotype that closely mimics that of BMP4 treatment (Fig. 5 *G* and *H*). Overall, this leads us to postulate that GATA3 occupies the very top of the TEtra hierarchy.

The chromatin occupancy analysis showed that *GATA2*, is bound by the three other TEtra factors and itself, pointing to its place at a step lower in the TEtra hierarchy compared with *GATA3* (Fig. 6*A*). Interestingly, while we observed a significant decrease of *GATA2* up-regulation in the *GATA3* KO BMP4-treated cells, others have reported an opposite effect of *Gata3* knockdown in mouse and rat TSCs (58). This highlights possible interspecies differences in GATA factor connectivity, which are expected according to studies showing the evolutionary plasticity of developmental gene regulatory network architectures (59). *TFAP2C* was also bound by three factors, except GATA2, while *TFAP2A*, only by GATA3 and TFAP2C. Although *TFAP2A* KO did not lead to a drastic down-regulation of the APA+ progenitor amount and expression of the other TEtra TFs, it led to similar perturbations of *GCM1*, *VGLL1*, and *OCT4*, as noted for the KO of *GATA3*. Further combinatorial TEtra gene KO and overexpression studies should reveal the finer details of the architecture of this gene-regulatory network and the connectivity to placental genes.

Rather unexpectedly, we noted up-regulation of *CDX2* in BMP4-treated *GATA3* and *TFAP2A* KO cells, even though we observed binding of all TEtra members in the same intronic region of the gene that was previously described to be activated by GATA3 and TFAP2C in the mouse (46, 48). It appears thus that varying the stoichiometric ratios of the TEtra effect the output of *CDX2*, or, alternatively, the TEtra regulate an unknown factor that suppresses *CDX2*, forming so-called "incoherent feedforward loop" (Fig. 6*B*), which is often seen in developmental gene networks (60). Importantly, because up-regulation of *CDX2* in *GATA3* KO did not rescue the ablation of the APA+ progenitors, this argues against its critical role in human TE specification.

Importantly, we provide evidence that the down-regulation of *OCT4* and pluripotency is mechanistically linked to the up-regulation of the TEtra (Fig. 6*B*). This is based on the observations that these processes take place during the relatively short phase of APA+ progenitor specification, that the TEtra bind in the first intron of *OCT4*, and on the phenotype of *GATA3* and *TFAP2A* KO that leads to weaker *OCT4* down-regulation during differentiation (Figs. 4*E* and 5*D*). We speculate that TEtra promote the formation of a repressive complex on this unique site in *OCT4*, which has not reported in the mouse, and that the inhibition of *OCT4* could be connected to the activation of the placental hormone hGC during BMP4-induced differentiation (61). Additional experiments are warranted to reveal the precise mechanism.

In line with previous findings that early developmental and later tissue-specific genes exhibit distinct chromatin configurations (38), we further observed that the majority of the earliest BMP4-induced genes (including TEtra) are bivalent with HCPs and hypomethylated CpGs in undifferentiated cells, while the genes from the late group (e.g., hCG) are respectively H3K27me3/H3K4me3 double negative with LCPs and hypermethylated CpGs. Somewhat unexpectedly, we did not detect significant changes in DNA methylation and acquisition of H3K4me3 mark in late up-regulated genes with LCP, arguing that the transcriptional induction can precede epigenetic changes during human ESC differentiation (Fig. 6*C*).

Taken together, we predict that our validation of the APA+ progenitor cells as a model for studying the biology of human TE in vitro and our discoveries of the TEtra gene-regulatory network lay a foundation for understanding the mechanisms of human placental development and pathologies related to placental dysfunction. Furthermore, we point out important features of regulation of pluripotency dissolution and TE specification that could be unique to human and primates, thus opening a path to understand the evolution of the placenta from a developmental perspective.

Materials and Methods

The H9 (WA09) hESC line was cultured on irradiated mouse embryonic fibroblasts (MEFs) in hESC medium (DMEM/F12; 11320074; Life Technologies), supplemented with 20% KSR (10828028; Thermo Fisher Scientific), Glutamax (35050038; Life Technologies), nonessential amino acids (1140050; Thermo Fisher Scientific), β-mercaptoethanol (31350-010; Thermo Fisher Scientific), 10 ng/mL FGF2 (100-18B; Peprotech), and 1% penicillin–streptomycin (15140122; Thermo Fisher Scientific), or on Geltrex-coated (A1413202; Life Technologies) plates with mTeSR1 medium (05850; STEMCELL Technologies). Colonies were passaged by treatment with 2 mg/mL collagenase IV (17104019; Thermo Fisher Scientific) in DMEM/F12 medium for 30–45 min. HUES9 iCRISPR and H9-GATA3 lines were cultured in mTeSR1 medium in plates coated with Matrigel (356234; BD Biosciences).

Multiple preovulatory follicles from rhesus macaque females of average maternal age (~8 y old) were obtained via controlled ovarian stimulation from the Oregon National Primate Research Center (ONPRC) Assisted Reproductive Technologies (ART) Core according to the Institutional Animal Care and Use Committee (IACUC) approved protocol #0095 entitled "Assisted Reproduction in Macaques." The IACUC is fully accredited by the Association for Assessment and Accreditation of Laboratory Animal Care (AAALAC) and Oregon Health & Science University (OHSU)/ONPRC has an Animal Welfare Assurance on file with the NIH Office for Laboratory Animal Welfare (OLAW; #A3304-01).

ACKNOWLEDGMENTS. For prolific scientific discussions and critical reading of the manuscript, we thank Tal Raveh. We also acknowledge Deutsche Forschungsgemeinschaft Grant DR1008/1-1 (to M.D.).

1. Rossant J, Cross JC (2001) Placental development: Lessons from mouse mutants. *Nat Rev Genet* 2:538–548.
2. Yagi R, et al. (2007) Transcription factor TEAD4 specifies the trophectoderm lineage at the beginning of mammalian development. *Development* 134:3827–3836.
3. Nishioka N, et al. (2008) Tead4 is required for specification of trophectoderm in pre-implantation mouse embryos. *Mech Dev* 125:270–283.
4. Strumpf D, et al. (2005) Cdx2 is required for correct cell fate specification and differentiation of trophectoderm in the mouse blastocyst. *Development* 132:2093–2102.
5. Niwa H, et al. (2005) Interaction between Oct3/4 and Cdx2 determines trophectoderm differentiation. *Cell* 123:917–929.
6. Russ AP, et al. (2000) Eomesodermin is required for mouse trophoblast development and mesoderm formation. *Nature* 404:95–99.
7. Ralston A, et al. (2010) Gata3 regulates trophoblast development downstream of Tead4 and in parallel to Cdx2. *Development* 137:395–403.
8. Werling U, Schorle H (2002) Transcription factor gene AP-2 gamma essential for early murine development. *Mol Cell Biol* 22:3149–3156.
9. Auman HJ, et al. (2002) Transcription factor AP-2gamma is essential in the extra-embryonic lineages for early postimplantation development. *Development* 129:2733–2747.
10. Guillemot F, Nagy A, Auerbach A, Rossant J, Joyner AL (1994) Essential role of Mash-2 in extraembryonic development. *Nature* 371:333–336.
11. Riley P, Anson-Cartwright L, Cross JC (1998) The Hand1 bHLH transcription factor is essential for placentation and cardiac morphogenesis. *Nat Genet* 18:271–275.
12. Anson-Cartwright L, et al. (2000) The glial cells missing-1 protein is essential for branching morphogenesis in the chorioallantoic placenta. *Nat Genet* 25:311–314.

13. Donnison M, et al. (2005) Loss of the extraembryonic ectoderm in Elf5 mutants leads to defects in embryonic patterning. *Development* 132:2299–2308.
14. Nishioka N, et al. (2009) The Hippo signaling pathway components Lats and Yap pattern Tead4 activity to distinguish mouse trophectoderm from inner cell mass. *Dev Cell* 16:398–410.
15. Kubaczka C, et al. (2015) Direct induction of trophoblast stem cells from murine fibroblasts. *Cell Stem Cell* 17:557–568.
16. Benchetrit H, et al. (2015) Extensive nuclear reprogramming underlies lineage conversion into functional trophoblast stem-like cells. *Cell Stem Cell* 17:543–556.
17. Kuckenberg P, et al. (2010) The transcription factor TCFAP2C/AP-2gamma cooperates with CDX2 to maintain trophectoderm formation. *Mol Cell Biol* 30:3310–3320.
18. Chen AE, et al. (2009) Optimal timing of inner cell mass isolation increases the efficiency of human embryonic stem cell derivation and allows generation of sibling cell lines. *Cell Stem Cell* 4:103–106.
19. Bai Q, et al. (2012) Dissecting the first transcriptional divergence during human embryonic development. *Stem Cell Rev* 8:150–162.
20. Niakan KK, Eggan K (2013) Analysis of human embryos from zygote to blastocyst reveals distinct gene expression patterns relative to the mouse. *Dev Biol* 375:54–64.
21. Deglincerti A, et al. (2016) Self-organization of the in vitro attached human embryo. *Nature* 533:251–254.
22. Sõber S, et al. (2015) Extensive shift in placental transcriptome profile in preeclampsia and placental origin of adverse pregnancy outcomes. *Sci Rep* 5:13336.
23. Blakeley P, et al. (2015) Defining the three cell lineages of the human blastocyst by single-cell RNA-seq. *Development* 142:3613.
24. Amita M, et al. (2013) Complete and unidirectional conversion of human embryonic stem cells to trophoblast by BMP4. *Proc Natl Acad Sci USA* 110:E1212–E1221.
25. Xu RH (2006) In vitro induction of trophoblast from human embryonic stem cells. *Methods Mol Med* 121:189–202.
26. Drukker M, et al. (2012) Isolation of primitive endoderm, mesoderm, vascular endothelial and trophoblast progenitors from human pluripotent stem cells. *Nat Biotechnol* 30:531–542.
27. Horii M, et al. (2016) Human pluripotent stem cells as a model of trophoblast differentiation in both normal development and disease. *Proc Natl Acad Sci USA* 113: E3882–E3891.
28. Telugu BP, et al. (2013) Comparison of extravillous trophoblast cells derived from human embryonic stem cells and from first trimester human placentas. *Placenta* 34:536–543.
29. Xu RH, et al. (2002) BMP4 initiates human embryonic stem cell differentiation to trophoblast. *Nat Biotechnol* 20:1261–1264.
30. Winnier G, Blessing M, Labosky PA, Hogan BL (1995) Bone morphogenetic protein-4 is required for mesoderm formation and patterning in the mouse. *Genes Dev* 9: 2105–2116.
31. Fujiwara T, Dehart DB, Sulik KK, Hogan BL (2002) Distinct requirements for extraembryonic and embryonic bone morphogenetic protein 4 in the formation of the node and primitive streak and coordination of left-right asymmetry in the mouse. *Development* 129:4685–4696.
32. Graham SJ, et al. (2014) BMP signalling regulates the pre-implantation development of extra-embryonic cell lineages in the mouse embryo. *Nat Commun* 5:5667.
33. Kurek D, et al. (2015) Endogenous WNT signals mediate BMP-induced and spontaneous differentiation of epiblast stem cells and human embryonic stem cells. *Stem Cell Rep* 4:114–128.
34. Lichtner B, Knaus P, Lehrach H, Adjaye J (2013) BMP10 as a potent inducer of trophoblast differentiation in human embryonic and induced pluripotent stem cells. *Biomaterials* 34:9789–9802.
35. Bernstein BE, et al. (2006) A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* 125:315–326.
36. Mikkelsen TS, et al. (2007) Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* 448:553–560.
37. Gifford CA, et al. (2013) Transcriptional and epigenetic dynamics during specification of human embryonic stem cells. *Cell* 153:1149–1163.
38. Xie W, et al. (2013) Epigenomic analysis of multilineage differentiation of human embryonic stem cells. *Cell* 153:1134–1148.
39. Lee CQ, et al. (2016) What is trophoblast? A combination of criteria define human first-trimester trophoblast. *Stem Cell Rep* 6:257–272.
40. Tang C, et al. (2011) An antibody against SSEA-5 glycan on human pluripotent stem cells enables removal of teratoma-forming cells. *Nat Biotechnol* 29:829–834.
41. Ma GT, et al. (1997) GATA-2 and GATA-3 regulate trophoblast-specific gene expression in vivo. *Development* 124:907–914.
42. Johnson W, et al. (1997) Regulation of the human chorionic gonadotropin alpha- and beta-subunit promoters by AP-2. *J Biol Chem* 272:15405–15412.
43. Biadasiewicz K, et al. (2011) Transcription factor AP-2α promotes EGF-dependent invasion of human trophoblast. *Endocrinology* 152:1458–1469.
44. Knöfler M, et al. (2004) Transcriptional regulation of the human chorionic gonadotropin beta gene during villous trophoblast differentiation. *Endocrinology* 145: 1685–1694.
45. Li Y, et al. (2013) BMP4-directed trophoblast differentiation of human embryonic stem cells is mediated through a ΔNp63+ cytotrophoblast stem cell state. *Development* 140:3965–3976.
46. Home P, et al. (2009) GATA3 is selectively expressed in the trophectoderm of periimplantation embryo and directly regulates Cdx2 gene expression. *J Biol Chem* 284: 28729–28737.
47. Cao Z, et al. (2015) Transcription factor AP-2γ induces early Cdx2 expression and represses HIPPO signaling to specify the trophectoderm lineage. *Development* 142:1606–1615.
48. Tzouanacou E, Tweedie S, Wilson V (2003) Identification of Jade1, a gene encoding a PHD zinc finger protein, in a gene trap mutagenesis screen for genes involved in anteroposterior axis development. *Mol Cell Biol* 23:8553–8562.
49. González F, et al. (2014) An iCRISPR platform for rapid, multiplexable, and inducible genome editing in human pluripotent stem cells. *Cell Stem Cell* 15:215–226.
50. Petropoulos S, et al. (2016) Single-cell RNA-seq reveals lineage and X chromosome dynamics in human preimplantation embryos. *Cell* 167:285.
51. Home P, et al. (2017) Genetic redundancy of GATA factors in the extraembryonic trophoblast lineage ensures the progression of preimplantation and post-implantation mammalian development. *Development* 144:876–888.
52. Pfeffer PL, Pearton DJ (2012) Trophoblast development. *Reproduction* 143:231–246.
53. Baczyk D, et al. (2009) Glial cell missing-1 transcription factor is required for the differentiation of the human trophoblast. *Cell Death Differ* 16:719–727.
54. Lee Y, et al. (2007) A unifying concept of trophoblastic differentiation and malignancy defined by biomarker expression. *Hum Pathol* 38:1003–1013.
55. Ugele B, Regemann K (2000) Differential increase of steroid sulfatase activity in XX and XY trophoblast cells from human term placenta with syncytia formation in vitro. *Cytogenet Cell Genet* 90:40–46.
56. Takaku M, et al. (2016) GATA3-dependent cellular reprogramming requires activation-domain dependent recruitment of a chromatin remodeler. *Genome Biol* 17:36.
57. Wei G, et al. (2011) Genome-wide analyses of transcription factor GATA3-mediated gene regulation in distinct T cell types. *Immunity* 35:299–311.
58. Ray S, et al. (2009) Context-dependent function of regulatory elements and a switch in chromatin occupancy between GATA3 and GATA2 regulate Gata2 transcription during trophoblast differentiation. *J Biol Chem* 284:4978–4988.
59. Hinman VF, Davidson EH (2007) Evolutionary plasticity of developmental gene regulatory network architecture. *Proc Natl Acad Sci USA* 104:19404–19409.
60. Goentoro L, Shoval O, Kirschner MW, Alon U (2009) The incoherent feedforward loop can provide fold-change detection in gene regulation. *Mol Cell* 36:894–899.
61. Gupta R, Ezashi T, Roberts RM (2012) Squelching of ETS2 transactivation by POU5F1 silences the human chorionic gonadotropin CGA subunit gene in human choriocarcinoma and embryonic stem cells. *Mol Endocrinol* 26:859–872.
62. Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) method. *Methods* 25:402–408.
63. Wickham H (2009) *ggplot2: Elegant Graphics for Data Analysis* (Springer, New York).
64. Brandl C, et al. (2014) Creation of targeted genomic deletions using TALEN or CRISPR/Cas nuclease pairs in one-cell mouse embryos. *FEBS Open Bio* 5:26–35.
65. Soumillon M, Cacchiarelli D, Semrau S, van Oudenaarden A, Mikkelsen TS (March 5, 2014) Characterization of directed differentiation by high-throughput single-cell RNA-Seq. *bioRxiv*, 10.1101/003236.
66. Ziegenhain C, et al. (2017) Comparative analysis of single-cell RNA sequencing methods. *Mol Cell* 65:631–643.e4.
67. Chavez SL, et al. (2012) Dynamic blastomere behaviour reflects human embryo ploidy by the four-cell stage. *Nat Commun* 3:1251.
68. Chavez SL, et al. (2014) Comparison of epigenetic mediator expression and function in mouse and human embryonic blastomeres. *Hum Mol Genet* 23:4970–4984.
69. Schindelin J, et al. (2012) Fiji: An open-source platform for biological-image analysis. *Nat Methods* 9:676–682.
70. Gautier L, Cope L, Bolstad BM, Irizarry RA (2004) affy–Analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* 20:307–315.
71. Bourgon R, Gentleman R, Huber W (2010) Independent filtering increases detection power for high-throughput experiments. *Proc Natl Acad Sci USA* 107:9546–9551.
72. Smyth GK (2004) Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol* 3:1–25.
73. Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10:R25.
74. Heinz S, et al. (2010) Simple combinations of lineage-determining transcription factors prime *cis*-regulatory elements required for macrophage and B cell identities. *Mol Cell* 38:576–589.
75. Bailey TL, et al. (2009) MEME SUITE: Tools for motif discovery and searching. *Nucleic Acids Res* 37:W202–W208.
76. Trapnell C, Pachter L, Salzberg SL (2009) TopHat: Discovering splice junctions with RNA-seq. *Bioinformatics* 25:1105–1111.
77. Kent WJ, et al. (2002) The human genome browser at UCSC. *Genome Res* 12: 996–1006.
78. Law CW, Chen Y, Shi W, Smyth GK (2014) voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol* 15:R29.
79. Thorvaldsdóttir H, Robinson JT, Mesirov JP (2013) Integrative genomics viewer (IGV): High-performance genomics data visualization and exploration. *Brief Bioinform* 14: 178–192.
80. Renaud G, Stenzel U, Maricic T, Wiebe V, Kelso J (2015) deML: Robust demultiplexing of Illumina sequences using a likelihood-based approach. *Bioinformatics* 31:770–772.
81. Dobin A, et al. (2013) STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* 29: 15–21.
82. Macosko EZ, et al. (2015) Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* 161:1202–1214.
83. Kharchenko P, Fan J (2016) scde: Single Cell Differential Expression, Version 2.0.1. Available at pklab.med.harvard.edu/scde. Accessed January 20, 2017.
84. Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: A bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26: 139–140.
85. Lun AT, Bach K, Marioni JC (2016) Pooling across cells to normalize single-cell RNA sequencing data with many zero counts. *Genome Biol* 17:75.
86. Donaldson J (2016) tsne: T-Distributed Stochastic Neighbor Embedding for R (t-SNE), Version 0.1-3. Available at https://github.com/jdonaldson/rtsne/. Accessed January 20, 2017.