Title: Direct reprogramming of human embryonic to trophoblast stem cells

Running Title: Human induced trophoblast stem cells

Authors

Norah M.E. Fogarty^{1,2*}, Ahmed Abdelbaki¹, Afshan McCarthy¹, Liani Devito³, Alice E. Chen⁴,
Prabhakaran Munusamy⁵, Paul Blakeley^{1,6}, Kay Elder⁷, Phil Snell⁷, Leila Christie⁷, Paul Serhal⁸
Rabi A. Odia⁸, Mahesh Sangrithi^{2,5,9} and Kathy K. Niakan^{1,10*}

Affiliations

¹Human Embryo and Stem Cell Laboratory, The Francis Crick Institute, 1 Midland Road,
London, NW1 1AT, UK

²Centre for Stem Cells and Regenerative Medicine, King's College London, London, SE1 9RT,
UK

³Human Embryo and Stem Cell Unit, The Francis Crick Institute, 1 Midland Road, London NW1
1AT, UK

⁴Trestle Biotherapeutics, Centre for Novel Therapeutics, 9310 Athena Circle, La Jolla, Ca 92037

⁵KK Women's and Children's Hospital, Division of Obstetrics and Gynecology, 100 Bukit Timah
Road, Singapore 229899, Singapore

⁶A*STAR, Agency for Science, Technology and Research, Singapore

⁷Bourn Hall Clinic, Bourn, Cambridge, CB23 2TN, UK

⁸The Centre for Reproductive & Genetic Health, 230 – 232 Great Portland Street, London,
W1W 5QS, UK

⁹Duke-NUS Graduate Medical School OBGYN Cancer Stem Cell Biology / ACP, USA

¹⁰The Centre for Trophoblast Research, Department of Physiology, Development and
Neuroscience, University of Cambridge, Cambridge, CB2 3EG, UK

*Correspondence: norah.fogarty@kcl.ac.uk or kkn21@cam.ac.uk

Key words

Trophectoderm, Embryonic stem cells, Trophoblast stem cells, Human embryo,
Reprogramming

Summary statement

Transcriptional analysis of human blastocysts reveals transcription factors sufficient to derive induced trophoblast stem cells from primed human embryonic stem cells.

Abstract

During the first week of development, human embryos form a blastocyst comprised of an inner cell mass and trophectoderm (TE) cells, the latter of which are progenitors of placental trophoblast. Here we investigated the expression of transcripts in the human TE from early to late blastocyst stages. We identified enrichment of transcription factors *GATA2*, *GATA3*, *TFAP2C* and *KLF5* and characterised their protein expression dynamics across TE development. By inducible overexpression and mRNA transfection we determined that these factors, together with MYC, are sufficient to establish induced trophoblast stem cells (iTSCs) from human embryonic stem cells. These iTSCs self-renew and recapitulate morphological characteristics, gene expression profiles and directed differentiation potential similar to existing human TSCs. Systematic omission of each or combinations of factors, revealed the critical importance of *GATA2* and *GATA3* for successful iTSC reprogramming. Altogether, these findings provide insights into the transcription factor network that may be operational in the human TE and broaden the methods for establishing cellular models of early human placental progenitor cells, which may be useful in the future to model placental-associated diseases.

Introduction

The correct functioning of the human placenta is crucial to ensure a healthy pregnancy outcome. However, despite this critical role, it is one of the least understood organs. The trophoblast cells of the placenta are derived from the trophectoderm (TE) cells of the human blastocyst. After fertilisation the zygote undergoes a series of cleavage cell divisions to form a tight ball of cells called a morula. The outer cells of the morula become polarised which leads to inactivation of the Hippo signalling pathway and allows the translocation of YAP1 into the nucleus where, together with TEAD4, it drives transcriptional activation of the TE programme (Gerri et al., 2020). The human blastocyst implants into the uterine lining approximately 7-9 days after fertilisation (James et al., 2012). The TE generates the first trophoblast lineages: the mononuclear cytotrophoblast cells (CTB) and the multinucleated

primitive syncytiotrophoblast (STB). The primitive STB is the initial invading interface and enzymatically degrades the uterine lining allowing for implantation of the conceptus (Hertig et al., 1956). CTB cells are the precursor for trophoblast cell types in the placenta. CTB cells sustain the syncytiotrophoblast across gestation by dividing asymmetrically, with one daughter cell fusing with the overlying syncytiotrophoblast and the other daughter remaining in the proliferative pool (Baczyk et al., 2006). The syncytiotrophoblast performs key functions including providing a physical and immunological barrier to pathogens, transporting nutrients and gaseous exchange, as well as producing and secreting hormones required to adapt the maternal physiology to the pregnancy (Bansal et al., 2012; Costa, 2016). CTB cells are also the precursor of extravillous trophoblast cells (EVTs) which invade the decidua to interact with maternal immune cells and remodel spiral arteries to establish blood supply to the placenta (Pijnenborg et al., 1980).

Human trophoblast stem cells (hTSCs) derived from blastocysts and first trimester placental villous tissue provide a paradigm for elucidating molecular mechanisms regulating trophoblast development (Okabe et al., 2018). hTSCs show expression of well-established markers of trophoblast including TP63 (Lee et al., 2007), GATA3 (Chiu and Chen, 2016) and TEAD4 (Soncin et al., 2018) and can be directed to differentiate into both EVT and STB, as identified by the expression of markers HLA-G (Apps et al., 2008) and SDC1 (Jokimaa et al., 1998) respectively. In addition, the establishment of a trophoblast organoid culture system has allowed hTSCs to be cultured in 3D, somewhat resembling the architecture of villous CTB cells and STB of the placenta (Haider et al., 2018; Turco et al., 2018). Most recently, induced trophoblast stem cells (iTSCs) which resemble primary tissue-derived hTSCs have also been captured in the process of reprogramming fibroblasts to human embryonic stem cells (hESCs) in naïve cell culture conditions (Castel et al., 2020; Cinkornpumin et al., 2020; Dong et al., 2020; Guo et al., 2021; Liu et al., 2020) and when established naïve ES cells are cultured in conditions which inhibit ERK and Nodal signalling (Guo et al., 2021). These *in vitro* models can be utilised in studies of trophoblast biology and have the potential to further inform our understanding of molecular mechanisms, transcription factors and signalling pathways regulating human trophoblast biology.

Studies of mouse preimplantation embryogenesis have elucidated mechanisms regulating lineage specification and maintenance of the TE in this species (Hemberger et al., 2020). In turn, this knowledge has led to the development of strategies to capture their *in vitro* counterparts. Blastocyst-derived TSCs can be isolated from the polar TE of blastocyst outgrowths or post-implantation extraembryonic ectoderm cultured in TSC media supplemented with FGF4 and heparin which recapitulates the signalling environment of the *in vivo* TSC compartment at the post-implantation stage (Tanaka et al., 1998). Identification of key transcription factors regulating mouse TE lineage specification and maintenance has also informed reprogramming strategies. Single overexpression of *Cdx2* or *Eomes* in TSC media is sufficient to induce differentiation of mouse ESCs to trophoblast-like cells (Niwa et al., 2005). Transient ectopic expression of the combination of TE-associated transcription factors *Gata3*, *Eomes*, *Tcfap2c*, with either *Ets2* or *Myc*, has been shown to reprogramme fibroblasts into mouse iTSCs (Benchetrit et al., 2015; Kubaczka et al., 2015). These alternatively derived cells recapitulate the morphology, epigenomes and gene expression patterns of blastocyst-derived TSCs and contribute exclusively to the placenta following blastocyst injection.

Here we perform a time-course RNA-sequencing analysis of human TE development and explored gene expression profiles of the human preimplantation TE to identify associated transcription factors that may be employed as reprogramming factors to convert hESCs to iTSCs. We identified four transcription factors (*GATA2*, *GATA3*, *TFAP2C* and *KLF5*) that are enriched in the TE lineage of human blastocysts compared with EPI and PE. We characterised gene expression dynamics across human preimplantation development and confirmed high expression of these factors in TE cells at all stages analysed at both the protein and the transcript level. Similar to strategies used in the mouse, we hypothesized that the transient ectopic expression of these transcription factors, along with the reprogramming enhancing factor *MYC*, would facilitate the establishment of human iTSCs. Here, we demonstrate using both a doxycycline-inducible approach and modified mRNA transfection that the expression of these factors promotes reprogramming of hESCs from primed condition to iTSCs. The iTSCs derived can be maintained stably in culture and recapitulate the expression of markers that have been used previously to identify embryo or primary placental-derived hTSCs (Okoe et al., 2018). RNA-seq analysis reveals that iTSCs are transcriptionally similar to existing hTSCs

and primary cytotrophoblast cells. When subjected to directed differentiation protocols iTSCs can undergo syncytialisation to form multinucleated structures that secrete a syncytiotrophoblast-associated peptide hormone, human chorionic gonadotropin (hCG). Altogether this indicates a novel strategy for the establishment of iTSCs and expands the repertoire of cells to model placenta biology *in vitro*.

Results

Identification and validation of candidate transcription factors

In this study we sought to devise a reprogramming strategy to generate iTSCs informed by the expression of transcription factors that are enriched in the TE. We initially sought to determine which genes overlap in their expression in the TE across preimplantation stages of human development, reasoning that commonly expressed transcripts may be required for the establishment and maintenance of the TE and thereby may facilitate iTSC reprogramming. We cultured human embryos from embryonic day 5 when the human blastocyst is first established and TE cells are discernible, until embryonic day 7 which is the latest stage we can culture preimplantation blastocysts to *in vitro* (Fig. 1A). We microdissected mural TE cells away from EPI and PE cells from human blastocysts and performed RNA-sequencing analysis to determine the gene expression profile for each developmental stage (Table S1). t-Distributed stochastic neighbour embedding (t-SNE) dimensionality reduction analysis indicated that plotting the first two dimensions of the t-SNE separates TE samples into three groups corresponding to the developmental time-points analysed (Fig. 1B). This was confirmed by a principal component analysis (PCA) that showed that when the first five principal components (PC) are plotted against each other, PC1 separates TE samples into the three developmental time-points (Fig. 1C). We compared the global gene expression patterns at these stages of development to identify the transcripts that are commonly expressed and those unique to each stage (Table S2). We found that the majority of transcripts, 4729 genes, were expressed in common at the stages analysed compared with 576, 387 and 250 genes uniquely expressed at day 5, 6 and 7, respectively. This suggests that once initiated, the TE programme remains transcriptionally consistent across preimplantation development (Fig. 1D). Functional enrichment analysis using GeneOntology (Ashburner et al., 2000; Gene Ontology, 2021; Mi et al., 2019) and REACTOME databases (Jassal et al., 2020) revealed enrichment in metabolic processes, translational activities and molecule binding which is

consistent with the rapid expansion of the TE (Fig. 1E). Upon further inspection of the gene lists associated with these terms we observed an enrichment in genes that are required for efficient oxidative phosphorylation, including genes encoding complexes of the electron transport chain (*MT-ND1*, *SDHD*, *CYC1*, *UQCRCQ*, *UQCRC1*, *ATP5E*, *ATP5I* and *ATP5L*). In addition, we observed an enrichment of genes involved in mitochondrial morphology. Interestingly while we observed enrichment of genes regulating mitochondrial fusion (*MFN1*, *MFN1* and *OPA1*), we did not detect expression of genes regulating mitochondrial fission (*DRP1*). In all, this suggests that mitochondrial fusion in the TE may be essential for the electron transport chain favouring oxidative phosphorylation. This is consistent with the observation that OXPHOS activity increases in developing mammalian blastocysts and correlates with the capacity of the embryo to develop to term following implantation (Gardner, 1998; Houghton, 2006; Leese et al., 2008). The genes expressed in common included *PPARG* which is expressed in cytotrophoblast cells and regulates differentiation (Murthi et al., 2013), a marker of cytotrophoblast *SPINT1* (Kohama et al., 2012) and *TCF7L1*, a key transcription factor of the WNT signalling pathway that is implicated in trophoblast proliferation (Meinhardt et al., 2014). *ACE2*, *APOE* and the receptor protein-tyrosine kinase *EFNA4* were detected at the earlier time-points analysed whereas *WLS*, a core component of the WNT secretion pathway (Yu et al., 2014), *HAND1* a transcription factor involved in branching morphogenesis during mouse placentation and *VHL* a regulator of placenta vasculogenesis (Genbacev et al., 2001) were enriched at later timepoints (Fig. 1F).

We collected a comprehensive transcription factor annotation from Human Transcription Factor Database (Hu et al., 2019) and cross-referenced this against the genes expressed in common at all stages analysed. This identified 232 transcription factors expressed in the TE (Table S3). We similarly mined our previously published single-cell RNA-seq datasets to identify transcription factors that were significantly enriched in the TE compared to the EPI and PE cells (Blakeley et al., 2015; Wamaitha et al., 2020). We further refined these lists of TE-associated transcription factors by screening for transcription factors with a known role in mouse TE development, or a suggested role in the human placenta. In all, from these analyses we identified four transcription factors for further investigation as candidate reprogramming factors: *GATA2*, *GATA3*, *TFAP2C* and *KLF5* (Fig. 1G) (Table S4).

From our analysis of single-cell RNA-seq data using our previously published Shiny App programme (Blakeley et al., 2015; Wamaitha et al., 2020), we found that *GATA2*, *GATA3* and *TFAP2C* share a common pattern of expression, with an onset of embryonic transcription shortly prior to TE initiation at the 8-cell stage (Gerri et al., 2020), and high expression maintained as development proceeds. By contrast, *KLF5* is expressed in human zygotes and its expression increases from the 8-cell stage (Blakeley et al., 2015). Analysis of lineage-specific gene expression patterns showed that *GATA2* and *GATA3* are enriched specifically in the TE (Blakeley et al., 2015). While *TFAP2C* is detected in the TE, it is also detected in the EPI, which we previously confirmed at the protein level (Blakeley et al., 2015). *KLF5* is detected in all three lineages, with the most abundant expression in the TE (Blakeley et al., 2015).

We next performed immunofluorescence analysis of these transcription factors in human embryos cultured from embryonic day 5 to 7, which corresponds to the key morphological stages of TE development: TE formation, expansion and hatching. At all stages analysed, *GATA2* protein expression was detected in the nuclei in TE cells, which were identified by both their position within the blastocyst and the absence of *NANOG* expression (Fig. 2A). Similarly, *GATA3* protein was detected in TE cells across the stages analysed (Fig. 2B). The pattern of *GATA2* and *GATA3* expression is similar to what has been reported in mouse embryos (Home et al., 2017) and is consistent with our previous analysis of *GATA3* expression (Gerri et al., 2020). At embryonic day 5 nuclear *KLF5* was detected in the TE, whereas the inner cell mass cells showed some cytoplasmic expression (Fig. 2C). At day 6 *KLF5* was abundantly expressed in TE nuclei. Lower levels of nuclear *KLF5* expression were also detectable in cells within the inner cell mass, which colocalised with *SOX2* expression. This pattern is similar to what has been described in the mouse, where *KLF5* is expressed in all cells of preimplantation embryos with lower expression in inner cell mass compared to the TE (Lin et al., 2010). At day 7, *KLF5* expression was abundant in the TE, and the protein was not detectable in the inner cell mass. By contrast, *TFAP2C* was detected at similar levels in both *NANOG*-positive epiblast and TE cells throughout all of the blastocyst stages analysed. As we, and others, reported this is in contrast to the expression pattern in the mouse where the homologue *TCFAP2C* is exclusively expressed in the TE (Fig. 2D) (Blakeley et al., 2015; Kuckenberg et al., 2010). Altogether, the enriched expression of *GATA2*, *GATA3*, *TFAP2C* and *KLF5* in the TE suggests that these transcription factors may have a functional role within the TE transcriptional network.

Generation of iTSCs from hESCs using lentiviral doxycycline inducible system

We next evaluated whether the expression of these transcription factors was sufficient to facilitate the reprogramming of hESCs cultured in primed conditions directly to induced trophoblast stem cells (iTSCs) without a requirement to transit through a naïve hESC state. We engineered doxycycline-inducible hESCs to overexpress *TFAP2C*, *KLF5*, *GATA3* and *GATA2* (Fig. 3A). We also induced the expression of *MYC* because we identified high expression of the transcript across the stages of TE development analysed (Table S4), and it has been shown to enhance the efficiency of cellular reprogramming in other contexts (Nakagawa et al., 2008). Individual lentiviruses expressing each of the five genes were packaged, pooled and used to transduce hESC. These cells are henceforth denoted five factor-hESC (5F-hESCs).

Transgenes were induced by exposing 5F-hESCs to doxycycline for 20 days (Fig. 3B) in the presence of an hTSC media which has been described previously (Okoe et al. 2018). As a control, 5F-hESCs were cultured in hTSC media in the absence of doxycycline (uninduced, Fig. 3C). After 20 days exposure to doxycycline epithelial colonies with polygonal hTSC-like morphology were apparent (induced, Fig. 3C). These colonies could be isolated by passaging onto freshly coated collagen plates and continued to maintain their morphology and proliferate as iTSCs in the absence of doxycycline (Fig. 3D). Stable iTSC lines were generated and were maintained in culture up to ten passages. By contrast, uninduced 5F-hESCs did not survive the first passage. Immunofluorescence analysis of the iTSCs after 8 passages in the absence of doxycycline indicated widespread expression of endogenous *GATA3* and *TFAP2C* as well as expression of the TE-associated keratin, *KRT18* (Cauffman et al., 2009; Lee et al., 2007), and *TP63* (Fig. 3E). By contrast, 5F-hESCs cultured in conventional primed conditions did not exhibit appreciable expression of these factors (Fig. 3F).

Generation of iTSCs from hESCs using non-integrating modified mRNAs

Reprogramming via lentiviral transduction is reportedly an inefficient process (from 0.01% to 0.1%) (Wernig et al., 2008) and integrated constructs can be spontaneously silenced and reactivated during cell culture and differentiation (Ellis, 2005; Herbst et al., 2012). As an alternative approach to lentiviral transduction, in parallel we employed a strategy of transcription factor overexpression using chemically modified mRNAs, which has been used to generate transgene-free human iPSCs (Mandal and Rossi, 2013; Warren et al., 2010).

Individual mRNAs encoding *GATA2*, *GATA3*, *TFAP2C*, *KLF5* and *MYC* were synthesised by *in vitro* transcription. Uridine and cytidine were substituted with the modified nucleotides pseudo-uridine and 5'methylcytidine to prevent cellular immune responses, and mRNA was capped with a modified 5'guanine cap to improve mRNA half-life. A cocktail of mRNA was made by pooling individual mRNAs in equal molar ratios and this was delivered into hESCs by lipofection every day for 20 days (Fig. 3G). Similar to what was observed with the doxycycline-inducible system, transfected cells underwent morphological change towards a polygonal shape (Fig. 3H). Cells were passaged after 20 days of transfections and iTSC-like colonies were observed. Stable iTSC lines that were generated following mRNA transfection were maintained in culture up to fifteen passages (Fig. 3I). Immunofluorescence analysis of the iTSCs after 8 passages following mRNA reprogramming indicated widespread expression of *GATA3*, *TFAP2C*, *KRT18* and *TP63* at levels comparable to established hTSCs (Fig. 3J).

iTSCs are transcriptionally similar to previously established TSCs

We next compared global gene expression of mRNA-generated iTSCs to alternatively derived TSCs by integrating previously published datasets (Dong et al., 2020; Liu et al., 2020; Okae et al., 2018), as well as primary CTB cells (Haider et al., 2018) and primed hESCs (Dong et al., 2020). After adjusting for batch effects, we used the top 500 most variably expressed genes to perform dimensionality reduction analysis. Principal component analysis showed that plotting the first two principal components, which together account for 95% of the variance, separates samples into three groups representing all *in vitro* trophoblast stem cells, primary CTBs, and hESCs (Fig. 4A). To determine which samples were most similar with respect to genes with significant expression changes, we performed a sample-to-sample distances and similarity matrix analysis. Consistent with the principal component analysis, three major clusters were observed: TSCs, CTBs and hESCs. Significantly, this analysis showed that the iTSCs generated in this study were most similar to hTSCs generated by Okae et al., followed by iTSCs generated from reprogrammed fibroblasts (Dong et al., 2020; Liu et al., 2020)(Fig. 4B). Plotting normalised count expression values for selected trophoblast-associated genes *GATA2*, *GATA3*, *TEAD4*, *TFAP2A* (Hubert et al., 2010) and *ENPEP* (Ito et al., 2003) showed shared high expression of these factors between iTSCs from this study with previously derived iTSCs (Dong et al., 2020; Liu et al., 2020) and hTSCs (Okae et al., 2018), as well as primary CTBs (Haider et al., 2018) (Fig. 4C). Conversely, iTSCs had downregulated the pluripotency factor

NANOG. In all, these analyses confirmed that we successfully reprogrammed iTSCs that resemble existing TSC lines based on the global transcriptome.

iTSCs function similarly to blastocyst- and placental-derived TSCs

We next examined whether iTSCs were able to undergo differentiation. We utilised a method to generate terminally differentiated syncytiotrophoblast as has been previously described (Okoe et al., 2018). Accordingly, we observed formation of multinucleated syncytia which exhibited a reduction in filamentous actin (F-actin) expression, indicating reorganisation of the actin cytoskeleton (Fig. 5A). hCG is one of the first peptide hormones that is produced by the syncytiotrophoblast (Kliman et al., 1986). Spent culture media from the iTSC-derived syncytiotrophoblast cells was collected after 6 days of differentiation and subjected to an over-the-counter pregnancy test kit which showed detectable hCG expression, similar to existing hTSCs (Fig. 5B).

Narrowing down transcription factors essential to generate iTSCs

To determine which transcription factors are required for the induction of iTSCs, we examined the effect of withdrawing individual transcription factors from the pool of factors on the formation of hTSCs. MYC was kept constant reasoning that based on its expression pattern in the embryo and reprogramming effect in other contexts that it functions to generally enhance reprogramming rather than being essential for iTSC generation (Nakagawa et al., 2010). To allow us to directly compare the success and efficiency of the iTSC programme induction we maintained an equivalent total amount mRNA in each cocktail by replacing the omitted factors with an equivalent amount of mRNA encoding GFP.

After 10 days of lipofection, cells were immunofluorescently analysed for the expression of KRT18 and TP63 (Fig. 6A; Fig. S1; Fig. S2)(Meistermann et al., 2021). Induction of the iTSC-programme was first determined by the presence of cells co-expressing nuclear TP63 and filamentous KRT18 expression extending from the nucleus to the cell membrane, as we observed in cells treated with the 5-factor cocktail (Fig. 6B:A). Immunofluorescence analysis for the detection of the additional markers, TEAD4 and KRT7, further confirmed the identity of iTSCs (Fig. S4:A). We observed that only the combination of *GATA2*, *GATA3* and *MYC* produced appreciable colonies of cells showing co-expression of TP63 and KRT18 (Fig. 6B:K),

albeit at lower proportions compared to the 5-factor cocktail. This 3-factor cocktail also produced cells expressing nuclear TEAD4 and KRT7 (Fig. S3K; yellow arrowheads). By contrast, while the other conditions showed some upregulation of KRT7 and KRT18, the expression pattern of the latter was patchy and disorganised, and the cells did not co-express TP63 nor exhibit nuclear TEAD4 (Fig. 6B)(Fig. S3). This suggests a functional requirement for GATA2 and GATA3 for the induction of iTSCs.

Discussion

In this study we investigate the sufficiency of four TE-associated transcription factors in reprogramming hESCs to iTSCs. In the mouse CDX2, GATA3, TFAP2C and EOMES were identified as key regulators of the TE lineage (Kuckenberget al., 2010; Ralston et al., 2010; Russ et al., 2000; Strumpf et al., 2005). These factors are highly enriched in mouse TE and mTSCs and have been shown to be capable of reprogramming fibroblasts towards mouse iTSC (Benchetrit et al., 2015; Kubaczka et al., 2015). In this study we sought to identify human-specific TE transcription factors capable of reprogramming alternative cell types to iTSCs.

We identified GATA2, GATA3, TFAP2C and KLF5 as factors that are enriched in the human TE. Together with MYC, these transcription factors were capable of reprogramming hESCs to iTSCs that were transcriptionally equivalent to existing hTSCs and can be directed to differentiate to alternative trophoblast populations. Future experiments to verify the *in vivo* bipotent differential potential of these hTSCs may include engraftment studies into mouse tissues and histological assessment to confirm their differentiation to both EVT and STB. In support of GATA3 as a candidate hTSC reprogramming factor is the recent work implicating GATA3 in the initiation of the TE lineage in human embryogenesis (Gerri et al., 2020). GATA2 is exclusively expressed in the TE of mouse, cow and human embryos, and regulates some trophoblast-related genes in mouse and cow (Bai et al., 2011; Gerri et al., 2020; Ma et al., 1997). The mouse homologue of *TFAP2C* (*Tcfap2c*) is required post-implantation with knockout mice dying at embryonic day 7.5 due to proliferation and differentiation defects in the TE compartment (Auman et al., 2002; Werling and Schorle, 2002). In mouse TSCs, TCFAP2C is detected at the promoters of key genes including *Elf5*, *Gata3*, *Hand1*, *Id2* and *Tead4* (Kidder and Palmer, 2010). By contrast, the role of *TFAP2C* in human trophoblast biology is not fully understood. It is known, however, that TFAP2C is a defining marker of CTBs

across gestation (Lee et al., 2016) and it stimulates human placenta lactogen and hCG production (Richardson et al., 2001). Functional analysis of KLF5 in human TE and trophoblast has not yet been performed, but in the mouse this transcription factor is required for the establishment of the inner cell mass and TE. *Klf5*^{-/-} embryos arrest at the blastocyst stage due to defects in TE development resulting in a failure to hatch (Lin et al., 2010). The expression patterns of these factors during human embryogenesis, together with their associated roles in TE and trophoblast biology, implicate *GATA2*, *GATA3*, *TFAP2C* and *KLF5* as candidate regulators of a human TE programme.

Recent studies have shown that reprogramming primed hESCs to naïve hESCs generates a small side-population of hTSCs that can be propagated in culture and recapitulate established hTSCs in terms of molecular markers, transcriptome, and directed differentiation potential (Castel et al., 2020; Cinkornpumin et al., 2020; Dong et al., 2020; Guo et al., 2021; Liu et al., 2020). Additionally, naïve hESCs treated with either small molecule inhibitors of ERK/mitogen-activated protein kinase (MAPK) and Nodal signalling for 3-5 days (Guo et al., 2020; Guo et al., 2021), or for 3 days with transient addition of BMP4 and JAK1 inhibitor (Io et al., 2021), differentiate to hTSCs when placed in conventional hTSC media. Human naïve hESCs represent an earlier, less-fixed developmental state that reflects gene expression and epigenetic profiles similar to that of the early epiblast or late morula (Gafni et al., 2013; Takashima et al., 2015; Theunissen et al., 2014). A feature of naïve pluripotency is the expression of *TFAP2C*. *TFAP2C* plays a critical role during primed to naïve reversion by facilitating the opening of naïve-specific enhancers, as well as regulating the expression of the pluripotency factors *OCT4* (Pastor et al., 2018) and *KLF4* (Chen et al., 2018). Indeed, CRISPR/Cas9-mediated knockout of *TFAP2C* in naïve hESCs significantly affected the efficiency of differentiation to hTSCs (Guo et al., 2021). This suggests a dual functionality of *TFAP2C* in human embryogenesis and stem cell counterparts in terms of maintaining naïvety and regulating a TE transcriptional network. Indeed, *TFAP2C* is suggested to both activate and repress target genes (Eloranta and Hurst, 2002; Pastor et al., 2018). Further molecular analysis of hTSCs may reveal the TE-specific repertoire of *TFAP2C* gene targets and provide further insight into the molecular mechanisms underpinning the TE transcriptional network.

Among the five transcription factors, we found that *GATA2* and *GATA3* are essential for upregulation of the hTSC programme. Surprisingly, and in contrast to the mouse, *TFAP2C* appears dispensable. While our reprogramming strategies worked, in the future, the efficiency of hTSC reprogramming may be further enhanced by either equivalently increasing the amount of each factor or by altering the stoichiometry of transcription factors, as has been demonstrated for induced pluripotent stem cell and cardiomyocyte reprogramming (Carey et al., 2011; Muraoka and Ieda, 2015). In addition, functional examination of these factors in knock-out hTSCs may further refine the combination of essential factors required for the maintenance of these cell types and these will be important studies in the future. The mRNA-based strategy is advantageous towards these goals as it easily allows for the generation of bespoke combinations of transcription factor as well as the alteration of the stoichiometry of factors in the cocktail.

Defective TE specification, trophoblast differentiation and maturation leading to abnormal placentation underlies miscarriage and preeclampsia (Burton and Jauniaux, 2017; Fisher, 2015). hTSCs and trophoblast organoids present a paradigm for modelling placenta development and disease *in vitro*. However, a caveat of existing cytotrophoblast- and blastocyst-derived hTSCs and trophoblast organoids is that it cannot be ascertained whether the starting population of cells would have given rise to a normal or a disease-affected placenta. Attempts to derive hTSCs and trophoblast organoids after 12 weeks' gestation, or at even later stages coincident with placental-related disease manifestation, have not been successful (Haider et al., 2018; Okae et al., 2018; Turco et al., 2018). In addition, as we do not have an understanding of the genetic basis of placental-related diseases we cannot currently apply CRISPR/Cas9-mediated genome editing to create mutations for disease modelling in existing hTSCs. Instead, we propose that the mRNA reprogramming strategy presented here could be applied and further refined to generate iTSCs from patient fibroblasts, or fibroblasts or mesenchymal stromal cells isolated from disease-affected placentas (Pelekanos et al., 2016). This strategy could open up the possibility of generating patient-specific human TSCs. Alternatively, if fibroblasts are not amenable to this conversion, reprogramming could be employed to generate patient-specific primed iPSCs as the starting population before applying our five-factor strategy. This may have an advantage over transiting through naïve iPSCs if it is found that epigenetic imprints are lost during iTSC generation depending on the

starting cell type or if there are persistent issues of karyotypic instability in the starting cell type. Additionally, mRNA transfection has the benefit of being readily transferable to future clinical application as it avoids the random integration of transgenes that can result in genomic modification and tumorigenicity (Warren et al., 2010). In all, this strategy could allow for the generation of a catalogue of clinically normal and disease-associated hTSCs that would provide a tool for basic research into trophoblast biology as well as a powerful tool for elucidating placental defects including recurrent miscarriage, preeclampsia, intrauterine growth restriction and stillbirth as well as a future drug screening platform.

Methods

Human embryo thaw and culture conditions

Human embryos that were surplus to family building requirements were donated to the Francis Crick Institute for use in research projects under the UK Human Fertilisation and Embryology Authority License number R0162 and the Health Research Authority's Research Ethics Committee (Cambridge Central reference number 16/EE/0067). Slow-frozen blastocysts (day 5 and day 6) were thawed using the BlastThaw (Origio; Cat. No. 10542010A) kit using the manufacturer's instructions. Vitrified blastocysts (day 5 and day 6) were thawed using the vitrification thaw kit (Irvine Scientific; Cat. No. 90137-SO) following the manufacturer's instructions. Human embryos were cultured in pre-equilibrated Global Media supplemented with 5 mg/ml Life Global HSA (both LifeGlobal; Cat. No. LGG-020 and LGPS-605) and overlaid with mineral oil (Origio; Cat. No. ART-4008-5P) and incubated in Embryoscope+ time lapse incubator (Vitrolife). For the collection of day 5 samples, embryos were fixed approximately 2 h after thawing to allow them to recover. To collect day 6 and 7 samples, day 5 or 6 embryos were cultured to the appropriate time before fixation.

Microdissection of TE from human blastocysts

Embryos were placed in drops of G-MOPS solution (Vitrolife; Cat. No. 10129) on a petri dish overlaid with mineral oil. The plate was placed on a microscope stage (Olympus IX70) and the embryos were held with an opposing holding pipette and blastomere biopsy pipette (Research Instruments) using micromanipulators (Narishige, Japan). The biopsy mode of a Saturn 5 laser (Research Instruments) was used to separate the mural TE from the ICM and polar TE. The separated mural TE was transferred to individual low bind RNase-free tube

containing 0.25 µl RNase inhibitor, 4.75 µl Dilution buffer (SMARTer Ultra Low Input RNA kit, Clontech; Cat. No. 634820) and 5 µl nuclease-free water on a pre-chilled CoolRack (Biocision, CA). Samples were stored at - 80°C until ready to be processed.

cDNA synthesis and library preparation of TE samples

cDNA was synthesized using SMARTer Ultra Low Input RNA for Illumina Sequencing-HV kit (Clontech Laboratories; Cat. No. 634820) according to the manufacturer's instructions and as previously published (Blakeley et al., 2015; Hyslop et al., 2016). cDNA was sheared using Covaris S2 with the modified settings 10% duty, intensity 5, burst cycle 200 for 2 min. Libraries were prepared using Low Input Library Prep Kit (Clontech Laboratories; Cat. No. 634900) according to the manufacturer's instructions. Library quality was assessed with an Agilent 2100 BioAnalyser and concentration measured by QuBit broad range assay (Thermo Fisher; Cat. No. Q32850). Prepared libraries were submitted for 50-bp paired-end sequencing on Illumina HiSeq 2000.

cDNA synthesis and library preparation of bulk cell lines

For bulk RNA-seq of cell lines, RNA was isolated using TRI reagent (Sigma-Aldrich; Cat. No. T9424) and DNase I-treated (Ambion; Cat. No. AM2222). Libraries were prepared using the polyA KAPA mRNA HyperPrep Kit (Roche; Cat. No. 8098115702). Quality of submitted RNA samples and the resulting cDNA libraries was determined by ScreenTape Assay on a 4200 TapeStation (Agilent). Prepared libraries were submitted for single-end 75 bp sequencing on an Illumina HiSeq 4000 (Illumina).

RNA-seq analysis of TE samples

RNA-seq data from human TE was analysed as previously described (Blakeley et al., 2015; Hyslop et al., 2016). Briefly, the reference human genome sequence was obtained from Ensembl, along with the gene annotation (GTF) file. The reference sequence was indexed using the bowtie2-build command. Reads were aligned to the reference human genome sequence using Tophat2 (Kim et al., 2013), with gene annotations to obtain BAM files for each sample. BAM files were then sorted by read coordinates and converted into SAM files using SAMtools. The process of mapping and processing BAM files was automated using a custom Perl script. The number of reads mapping to each gene were counted using the program

HTSeq-count (Version 0.6.1; Anders et al., 2015). The resulting count files for each sample were used as input for differential expression analysis using DESeq2 (Anders and Huber, 2010). Firstly, the function 'estimateSizeFactors' and 'estimateDispersions' were used to estimate biological variability and calculate normalised relative expression values across the different blastocyst samples. Initially, this was performed without sample labels (option: method='blind') to allow unsupervised clustering of the blastocyst samples using principal components analysis and hierarchical clustering. The dispersion estimates were recalculated with the sample labels included and with the option: method='pooled'. The function 'nbinomTest' was then used to calculate p-values to identify genes that show significant differences in expression between different developmental stages.

An RPKM >5 threshold was applied to generate the lists of genes expressed at each stage analysed. Overlap between the gene lists was determined using the online web application GeneVenn (Pirooznia et al., 2007). Gene lists were used to perform a Gene Ontology and REACTOME functional enrichment analysis (Ashburner et al., 2000; Fabregat et al., 2018) to identify overrepresented categories using a significance threshold of $p \leq 0.05$.

Cell culture

H9 hESCs cells were routinely cultured in mTeSR1 (Stem Cell Technologies; Cat. No. 85850) and growth factor reduced Matrigel (BD Biosciences; Cat. No. 356321). For hTSC reprogramming experiments, H9 cells were plated in mTeSR1 media on Matrigel coated cell culture dishes. Media was changed for hTSC media on the first day of reprogramming. hTSC media was prepared as described previously (Okoe et al., 2018) (Chir99021, BioTechne, Cat. No. 4423/10; EGF, Tebu Bio Ltd, Cat. No. 167AF-100-15-a; ITS-X supplement, Fisher Sci, Cat. No. 10524233; L-ascorbic acid, Tocris, Cat. No. 4055/50; A83-01, Sigma-Aldrich Cat. No. SML0788/5MG; SB431542, Cambridge Bioscience, Cat. No. SM33-2; Valproic acid, Sigma-Aldrich Cat. No. V-006-1ML; Y27632, Stem Cell Technologies, Cat. No. 72302). Media was changed every second day. For transgene induction, doxycycline was added daily at a concentration of 1 µg/ml over 20 days for transgene induction. hTSCs were passaged by dissociation with TrypLE Express (Fisher Thermo Scientific, Cat. No. 12604013) for 15 min at 37 C and passaged onto collagen IV-coated (5 µg/ml; Corning, Cat. No. 354233) 10 cm plates and maintained in hTSC media.

Immunostaining

Embryos were fixed with 4% paraformaldehyde in PBS for 1 h at 4°C. Immunofluorescence staining was performed as described previously (Fogarty et al. 2018). The primary antibodies used are listed in Table S5. Embryos were placed on μ -Slide 8 Well coverslip dishes (Ibidi; Cat. No. 80826) for confocal imaging. Imaging was performed using a Leica SP5 confocal microscope and 3 μ m thick optical sections were collected. hESC, iTSC and hTSC lines were fixed with PBS 4% PFA for 1 h at 4°C, then washed three times with PBS (Life Technologies; Cat. No. 14190-094). Blocking was achieved by incubation with PBS with 10% donkey serum (Sigma-Aldrich; Cat. No. D9663) and 0.1% Triton X-100 (Sigma-Aldrich; Cat. No. T8787) for 30 mins at room temperature. Permeabilisation was performed by incubation in PBS with 0.5% Triton X-100. Primary antibodies were diluted in blocking solution and each incubated overnight at 4 C. Following each incubation, the cells underwent 3 x 5 min washes with PBS with 0.1% Triton X-100. Secondary antibodies were diluted 1:300 in PBS with 0.1% Triton X-100 for 1 h at room temperature. After 3 x 5 min washes in PBS with 0.1% Triton X-100, 5 μ g/ml DAPI (Sigma-Aldrich; D9542) was added for 2 mins during the final wash in order to perform nuclear staining. Epifluorescence images were performed on an Olympus IX73 using CellSens software (Olympus).

Generation of inducible system

Doxycycline-inducible overexpression of transcription factors was achieved using the Lenti-X Tet-On 3G Inducible Expression System (Clontech; Cat. No. 631363) following the manufacturer's protocol, and as previously described (Wamaitha et al., 2015). Coding sequences were sub-cloned from the template plasmids into the pLVX-TRE3G vector to generate individual pLVX-TRE3G-Gene of Interest (GOI) vectors. Lentiviral packaging was achieved using 7 μ g of pLVX-TRE3G-GOI and the Lenti-X Packaging Single Shot reagents in HEK-293T cells. Transfection media was replaced after 6 h with fresh MEF media. Lentiviral supernatants were subsequently collected after transfection of HEK 293T cells with either the pLVX-TRE3G-GOI or the pLVX-Tet3G vector using the X-fect reagent (Clontech; Cat. No. 631317). Supernatants were concentrated by ultracentrifugation. Equal volumes of lentivirus supernatant encoding *GATA2*, *GATA3*, *TFAP2C*, *MYC* and *KLF5* were pooled to generate a 5-factor cocktail which was aliquoted into single use 10 μ l volumes and stored at -80 C until needed. H9 hESCs were sequentially transduced with the pLVX-Tet3G and the 5-factor

cocktail. hESCs were plated on growth factor reduced Matrigel (BD Biosciences; Cat. No. 356230) in mTeSR1 medium (Stem Cell Technologies; Cat. No. 85850) and grown to 70% confluency. Cells were first transduced with the pLVX-tet3G lentivirus followed by selection with G418 (250 µg/ml) for one week. Resistant cells were selected for at least 2 passages and then transduced with the pLVX-TRE3G-GOI lentivirus pool, followed by selection with puromycin (0.5 µg/ml) for four days. For the induction of GOI expression, doxycycline was added to the mTeSR1 media at a concentration of 1 µg/ml. Clonal lines were generated and screened for transgene integration. Lines were cultured in the presence of doxycycline for 48 h before cDNA was harvested and RT-qPCR was performed to determine the levels of transgene induction.

Generation of templates for lentivirus production and in vitro transcription (IVT)

The vector design strategy was informed from previously published reports (Mandal and Rossi, 2013). Briefly, nucleotide sequences of the open reading frames for the canonical isoforms of reprogramming factors were identified from the Ensembl genome browser, and the sequences were verified against the corresponding amino acid sequences in Uniprot. Individual *in vitro* transcription (IVT) template constructs for *GATA2*, *GATA3*, *TFAP2C* and *KLF5* consisting of a T7 promoter-5'UTR-Open Reading Frame-3'UTR-T7 terminator cassette cloned into a pUC57 backbone were custom made (Genewiz UK Ltd). 5'UTR and 3'UTRs were added to maximize stability of mRNA transcripts and to increase protein translation. The ORFs of *MYC* and *GFP* were templated from plasmids bearing human MYC and GFP and ligated into the pUC57 backbone flanked by the 5'UTR and 3'UTR sequences. Annotated sequence files of all constructs are provided in Table S6.

Generation of modified mRNAs by in vitro synthesis

The mRNA synthesis protocol has been described previously (Mandal and Rossi, 2013). Briefly, dsDNA templates were linearised from cDNA clones in pLVX vectors for *GATA2*, *GATA3*, *TFAP2C*, *KLF5* and *MYC*. A small amount of digestion mix was run on a gel to confirm complete digestion. Linearised plasmid was purified using PCR purification kit (Qiagen; Cat. No. 28104). A nanodrop was used to confirm the purity of the eluted product according to the 260/280 ratios. Poly(A) tail was added using KAPA PCR ready mix (2X), Xu-F1 and Xu-T120 primers (Integrate DNA Technologies) and digested plasmid adjusted to 10 ng/ul. Tail PCR was run for

32 cycles and purified using PCR purification kit. *In vitro* transcription was performed using MEGAscript T7 kit (Thermo Fisher; Cat. No. AMB13345): custom NTP mix was prepared with 3'-O-Me-m7G cap analogue (60mM, NEB), GTP (75mM, MEGAscript T7 kit), ATP (75mM, MEGAscript T7 kit), Me-CTP (100 mM, TriLink; Cat. No. N-1014-1) and pseudo-UTP (100mM, Tri-link; Cat. No. O-0263). Reaction was heated at 37 C for 2 h. 2 ul of Turbo DNase (Thermo Fisher; Cat. No. AM2238) was added and incubated at 37 C for 15 min. DNase treated reaction mix was purified using RNAeasy kit (Qiagen; Cat. No. 74104) according to the manufacturer's instructions. RNA was phosphatase-treated using Antarctic phosphatase (New England BioLabs; Cat. No. M0289S) and purified using MEGAclean kit (Thermo Fisher; Cat. No. AM1909). Concentration and quality were measured using a nanodrop and adjusted to 100 ng/ul. For reprogramming experiments a modified mRNA cocktail was prepared by mixing the five factors in an equal molar ratio. The cocktail was aliquoted into single use aliquots containing a total mRNA amount of 1 µg and stored at -80 C until needed. For transcription factor combinatorial experiments cocktails were made up to a total amount of 500 ng of mRNA with the relevant quantity of mRNA encoding GFP replacing the omitted factors.

Lipofection of hESCs

For one well of a 6-well plate, 1 µg (1x) of mRNA was mixed with 86 µl of OptiMEM reduced serum media in an Eppendorf tube. In a second tube 0.5 x volume of lipofectamine RNAi max (Thermofisher; Cat. No. 13778100) was mixed with 93 µl OptiMEM. Tubes were incubated at room temperature for 5 min. The tubes were then mixed together and incubated at room temperature for 20 min. The lipofection mix was added to the well in a dropwise manner and mixed well by rocking the plate from side to side. The cells were incubated at 37 C in 20% O₂ for 4 h. The media was then replaced with hTSC media containing 200 ng/ml recombinant B18R to prevent gamma-interferon response (Stem Cell Technologies; Cat. No. 78075). For reprogramming experiments lipofection was performed at the same time every day for 20 days. For transcription factor combinatorial experiments H1 hESCs were lipofected daily for 10 days.

Directed differentiation

For the induction of syncytiotrophoblast, iTSCs were grown to 70% confluency in 6-well plates. Cells were passaged by incubating in TrypLE (Thermo Fisher Scientific; Cat. No.

12604013) for 10 min at 37 C. Cells were seeded in a 6-well plate pre-coated with 2.5 mg/ml Col IV and cultured in 2 mL of syncytiotrophoblast medium comprised of DMEM/F12 supplemented with 0.1 mM 2-mercaptoethanol, 0.5% penicillin-streptomycin, 0.3% BSA, 1% ITS-X supplement, 2.5 mM Y27632, 2 mM forskolin, and 4% KSR]. The medium was replaced at day 3, and the cells were fixed at day 6 for analysis.

RNA-seq data analysis of induced trophoblast stem cells (iTSCs) and related lines

Data from this study

Sequencing data from sample replicates comprising iTSCs (n = 4), H9 ES cells (n = 3) and primary CT27 cells (n = 3) were first checked using the *FastQC* package (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Adapter removal was performed using *Trimalore v0.6.6* (<https://github.com/FelixKrueger/TrimGalore>), and the trimmed read data were re-checked for conformity and quality with *FastQC*.

Accessory datasets

Primary cytotrophoblast gene expression data was downloaded from GEO with accession number GSE109976 (Haider et al., 2018). Additional primed hESCs gene expression data was downloaded from GEO with accession numbers GSM4116153 and GSM4116151 (Dong et al., 2020). Trophoblast stem cells derived from human blastocysts gene expression data was downloaded from GSE138762 (Dong et al., 2020). Trophoblast stem cells derived from fibroblast reprogramming gene expression data was downloaded from GSE150616 (Liu et al., 2020) and GSE138762 (Dong et al., 2020).

Next, sequences were aligned to the reference genome (Homo_sapiens.GRCh38) using *HISAT2 v2.2.1* (Kim et al., 2019), with trimming of 5' and 3' bases performed based on the QC of each of each of these samples. Counts were generated using *FeatureCounts v2.0.1*, and the count matrix was analysed further using *DESeq2* (Liao et al., 2019; Love et al., 2014). Genes with fewer than 10 counts detected across all 22 samples were excluded, leaving 17498 genes for further processing. A variance stabilizing transformation (VST) was applied, and subsequently batch related effects in combining these different datasets were corrected for using the *limma* package (i.e. *limma::removeBatchEffect*) (Law et al., 2016). Sample to sample

distances were computed and plotted as a heatmap. Principal Component Analysis (PCA) was performed using the top 500 most highly variable genes.

Acknowledgements

We are very grateful to the donors of human embryos whose contributions enable this research. We thank all members of the Niakan lab for help and comments on the paper. We are grateful to the Francis Crick Institute's Science Technology Platforms: Lyn Healey from Human Embryonic Stem Cell Facility, Robert Goldstone and Deb Jackson from the Advanced Sequencing Facility; and the Advanced Light Microscopy Facility.

Funding

Work in the K.K.N. laboratory was supported by the Francis Crick Institute, which receives its core funding from Cancer Research UK (FC001120), the UK Medical Research Council (FC001120), and the Wellcome Trust (FC001120). The M.S. lab at Centre for Stem Cells and Regenerative Medicine, King's College London is supported by a Wellcome Trust Clinical Career Development Fellowship (222052/Z/20/Z), and the M.S. group in Singapore (including P.M.) was supported by a National Medical Research Council, Singapore Clinician Scientist Investigator Award (CSAINV17may001). N.M.E.F is supported by a King's Prize Fellowship. For the purpose of Open Access, the authors have applied a CC-BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

Competing Interests

No competing interests declared.

Author Contributions

N.M.E.F and K.K.N. conceived the study; K.K.N supervised the project; N.M.E.F and K.K.N designed the experiments. N.M.E.F. performed the majority of experiments with assistance from A.C. A.A., L.D. and A.M. K.K.N. performed embryo dissections. K.E., P.S., L.C. and R.A.O. coordinated the donation of embryos to the research project. P.B. and P.M. performed bioinformatic analysis of RNA-seq data. N.M.E.F. and K.K.N. wrote the manuscript with help from all the authors.

Data availability

RNA-seq FastQ files have been deposited at ArrayExpress with the accession numbers E-MTAB-10749 for data from microdissected human embryos and E-MTAB-10748 for data arising from cell lines.

Figure Legends

Figure 1- Transcriptome analysis of TE in human blastocysts.

(A) Brightfield images of human blastocysts microdissected for TE isolation ahead of RNA-sequencing analysis. (B) t-Distributed Stochastic Neighbour Embedding (t-SNE) plot based on VST-normalisation of the top 6000 most variable expressed genes. A colour coded sample key is provided. (C) Principal component analysis matrix using the first five principal components for the top 12,000 most variable expressed genes. (D) Venn diagram indicating the number of transcripts with overlap or unique expression at day 5, day 6 and day 7 in the mural TE (RPKM>5). (E) The most enriched GO terms for genes associated with biological process (BP), cellular component (CC), molecular function (MF), and REACTOME pathways associated with genes expressed in common are shown. (F) Scatterplots showing selected genes that are expressed in common at all stages analysed, expressed in common at days 5 and 6, or expressed in common at days 6 and 7. Horizontal line denotes the mean RPKM value of each timepoint, dotted line shows RPKM value = 5. (G) Scatterplots showing expression of *GATA2*, *GATA3*, *TFAP2C*, *KLF5* and *MYC* at days 6 and 7. Horizontal line denotes the mean RPKM value of each timepoint, dotted line shows RPKM value = 5.

Figure 2 – Identification of TE-associated transcription factors in the human blastocyst.

Immunofluorescence analysis of (A) *GATA2*, (B) *GATA3*, (C) *KLF5* and (D) *TFAP2C* (green) and DAPI nuclear staining (blue) in human blastocysts cultured to day 5, 6 and 7 post-fertilisation. Inner cell mass cells are detected by the expression of *NANOG* or *SOX2* (red). Scale bars: 50 μ m.

Figure 3 – Expression of 5 transcription factors reprograms hESCs to hTSC-like cell.

(A) Schematic diagram of the generation of *GATA2*, *GATA3*, *TFAP2C*, *KLF5* and *MYC* inducible hESCs by lentiviral transduction. (B) Schematic representation of the strategy for reprogramming 5F-hESCs to hTSCs. 5F-hESCs were plated in mTeSR1 media for 24 h after

which media was replaced with hTSC media. Doxycycline was administered daily for 20 days in hTSC media for transgene induction. (C) Brightfield images of 5F-hESCs cultured in hTSC media alone for 20 days (uninduced) and hESCs cultured in hTSCs in the presence of dox for 20 days (induced). (D) Brightfield images of stable iTSC lines derived from transgene overexpression. Stable iTSCs were maintained up to 10 passages in culture. Scale bar: 200 μ m. (E) Immunofluorescence analysis for the detection of KRT18, GATA3, TFAP2C and TP63 (green) and DAPI nuclear expression (blue) in stable iTSCs and in previously established control hTSCs (Okoe et al., 2018). Scale bars: 50 μ m. (F) Immunofluorescence analysis for the detection of GATA3, TFAP2C, KRT18 and TP63 (green) and DAPI nuclear expression (blue) in 5F-hESCs cultured in mTeSR1 media. Scale bars: 50 μ m. (G) Schematic representation of the strategy for reprogramming hESCs to hTSCs using modified mRNAs. hESCs were plated in mTeSR1 media for 24 h after which media was replaced with hTSC media. mRNAs encoding *GATA2*, *GATA3*, *TFAP2C*, *KLF5* and *MYC* administered daily for 20 days via lipofection-mediated transfection. (H) Brightfield images of mock transfected cells and 5-factor mRNA-cocktail transfected cells on day 20. (I) Brightfield image of stable iTSCs derived from mRNA reprogramming. Scale bar: 200 μ m. (J) Immunofluorescence analysis for the detection of KRT18, GATA3, TFAP2C and TP63 (green) and DAPI nuclear expression (blue) in stable iTSCs and previously established control hTSCs (Okoe et al., 2018). Scale bars: 50 μ m.

Figure 4 – Global transcriptional analysis of iTSCs and similarity to previously established TSCs

Analysis of mRNA-derived iTSCs (this study), previously established hTSCs (CT27, BT5) (Okoe et al., 2018), iTSCs derived from naïve reprogramming experiments (Dong et al., 2020; Liu et al., 2020), primary placental CTBs (Haider et al., 2018) and primed hESCs (this study) (Dong et al., 2020). (A) Principal component analysis (PCA) using the top 500 most variable expressed genes. (B) Sample-to-sample distances and similarity matrix analysis. iTSCs generated in this study are labelled in red. (C) Dot plots of log₂normalised count data for selected trophoblast-associated genes and *NANOG*. iTSCs generated in this study are labelled in red.

Figure 5 – iTSCs can be directed to differentiate to syncytiotrophoblast-like cells.

(A) Immunostaining analysis of F-actin (green) and DAPI nuclear expression (blue) in undifferentiated and differentiated iTSCs and existing hTSCs on the sixth day of

differentiation. Scale bars: 50 μ m. (B) Detection of hCG in spent culture media from syncytiotrophoblast generated from iTSCs and previously established control hTSCs using an over-the-counter pregnancy test. 1. hTSC media, 2. iTSC-derived syncytiotrophoblast replicate 1, 3. iTSC-derived syncytiotrophoblast replicate 2, 4. iTSC-derived syncytiotrophoblast replicate 3, 5. hTSC-derived syncytiotrophoblast, 6. unconditioned syncytiotrophoblast media. Key shows the interpretation of the test results.

Figure 6- GATA2 and GATA3 are required for induction of hTSC programme

(A) Schematic representation of the strategy for testing factors required for iTSC reprogramming. hESCs were plated in primed hESC culture conditions for 24 h after which media was replaced with hTSC media. hESCs were transfected with cocktails of modified mRNAs where omitted factors were replaced with equivalent amount of mRNA encoding GFP to keep the molar ratio consistent. Cells were transfected daily for 10 days at which point they were fixed for immunofluorescence analysis (B) Immunofluorescence analysis for the detection of TP63 (red) and KRT18 (yellow) and DAPI nuclear expression (blue) in transfected cells. Merged images are shown. Key for mRNA cocktail combinations: B, GATA2, TFAP2C, KLF5 and MYC (GATA3 omitted); C, GATA3, TFAP2C, KLF5 and MYC (GATA2 omitted); D, GATA2, GATA3, TFAP2C and MYC (KLF5 omitted); E, GATA2, GATA3, KLF5 and MYC (TFAP2C omitted); F, TFAP2C, KLF5 and MYC (GATA2 and GATA3 omitted); G, GATA2, TFAP2C and MYC (GATA3 and KLF5 omitted); H, GATA2, KLF5 and MYC (GATA3 and TFAP2C omitted); I, GATA3, TFAP2C and MYC (GATA2 and KLF5 omitted); J, GATA3, KLF5 and MYC (GATA2 and TFAP2C omitted); K, GATA2, GATA3 and MYC (TFAP2C and KLF5 omitted); L, TFAP2C, KLF5 and MYC (KLF5, GATA2 and GATA3 omitted); M, GATA2 and MYC (GATA3, TFAP2C and KLF5); N, KLF5 and MYC (GATA2, GATA3 and TFAP2C omitted). Scale bars: 50 μ m.

Figure S1 – Short-term exogenous expression of 5 TE-associated transgenes upregulates

markers of trophoblast. (A) Schematic representation of the strategy for testing duration of transgene induction required for upregulation of trophoblast markers. 5F-hESCs were plated in mTeSR1 media for 24 h after which media was replaced with hTSC media. Transgenes were induced by addition of doxycycline for 10 days. Cells were cultured in the absence of doxycycline for a further 10 days at which point they were fixed for immunofluorescence analysis. (B) Immunofluorescence analysis for the detection of KRT18, GATA3 and TP63

(green) and DAPI nuclear expression (blue) in induced and uninduced 5F-hESCs on day 20.
Scale bars: 50 µm.

Figure S2 - GATA2 and GATA3 are required for induction of hTSC programme

Extended immunofluorescence analysis for the detection of TP63, KRT18 and DAPI nuclear expression in transfected cells. Key for mRNA cocktail combinations: A, Five factors positive controls; B, *GATA2*, *TFAP2C*, *KLF5* and *MYC* (*GATA3* omitted); C, *GATA3*, *TFAP2C*, *KLF5* and *MYC* (*GATA2* omitted); D, *GATA2*, *GATA3*, *TFAP2C* and *MYC* (*KLF5* omitted); E, *GATA2*, *GATA3*, *KLF5* and *MYC* (*TFAP2C* omitted); F, *TFAP2C*, *KLF5* and *MYC* (*GATA2* and *GATA3* omitted); G, *GATA2*, *TFAP2C* and *MYC* (*GATA3* and *KLF5* omitted); H, *GATA2*, *KLF5* and *MYC* (*GATA3* and *TFAP2C* omitted); I, *GATA3*, *TFAP2C* and *MYC* (*GATA2* and *KLF5* omitted); J, *GATA3*, *KLF5* and *MYC* (*GATA2* and *TFAP2C* omitted); K, *GATA2*, *GATA3* and *MYC* (*TFAP2C* and *KLF5* omitted); L, *TFAP2C*, *KLF5* and *MYC* (*KLF5*, *GATA2* and *GATA3* omitted); M, *GATA2* and *MYC* (*GATA3*, *TFAP2C* and *KLF5*); N, *KLF5* and *MYC* (*GATA2*, *GATA3* and *TFAP2C* omitted). hESCs cultured in mTeSR1 media are included as a negative controls. Scale bars: 50 µm.

Figure S3. Immunofluorescence analysis of additional markers in combinatorial experiments

Immunofluorescence analysis for the detection of TEAD4, KRT7 and DAPI nuclear expression in transfected cells. Key for mRNA cocktail combinations: A, Five factors positive controls; B, *GATA2*, *TFAP2C*, *KLF5* and *MYC* (*GATA3* omitted); C, *GATA3*, *TFAP2C*, *KLF5* and *MYC* (*GATA2* omitted); D, *GATA2*, *GATA3*, *TFAP2C* and *MYC* (*KLF5* omitted); E, *GATA2*, *GATA3*, *KLF5* and *MYC* (*TFAP2C* omitted); F, *TFAP2C*, *KLF5* and *MYC* (*GATA2* and *GATA3* omitted); G, *GATA2*, *TFAP2C* and *MYC* (*GATA3* and *KLF5* omitted); H, *GATA2*, *KLF5* and *MYC* (*GATA3* and *TFAP2C* omitted); I, *GATA3*, *TFAP2C* and *MYC* (*GATA2* and *KLF5* omitted); J, *GATA3*, *KLF5* and *MYC* (*GATA2* and *TFAP2C* omitted); K, *GATA2*, *GATA3* and *MYC* (*TFAP2C* and *KLF5* omitted); L, *TFAP2C*, *KLF5* and *MYC* (*KLF5*, *GATA2* and *GATA3* omitted); M, *GATA2* and *MYC* (*GATA3*, *TFAP2C* and *KLF5*); N, *KLF5* and *MYC* (*GATA2*, *GATA3* and *TFAP2C* omitted). hESCs culture in mTeSR1 media are included as a negative controls. Scale bars: 50 µm.

References

- Apps, R., Gardner, L., Hiby, S.E., Sharkey, A.M. and Moffett, A.** (2008). Conformation of human leucocyte antigen-C molecules at the surface of human trophoblast cells. *Immunology*. **124**, 322-328.
- Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., et al.** (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet*. **25**, 25-29.
- Auman, H.J., Nottoli, T., Lakiza, O., Winger, Q., Donaldson, S. and Williams, T.** (2002). Transcription factor AP-2gamma is essential in the extra-embryonic lineages for early postimplantation development. *Development*. **129**, 2733-2747.
- Baczyk, D., Dunk, C., Huppertz, B., Maxwell, C., Reister, F., Giannoulas, D. and Kingdom, J.C.** (2006). Bi-potential behaviour of cytotrophoblasts in first trimester chorionic villi. *Placenta*. **27**, 367-374.
- Bai, H., Sakurai, T., Someya, Y., Konno, T., Ideta, A., Aoyagi, Y. and Imakawa, K.** (2011). Regulation of trophoblast-specific factors by GATA2 and GATA3 in bovine trophoblast CT-1 cells. *J Reprod Dev*. **57**, 518-525.
- Bansal, A.S., Bora, S.A., Saso, S., Smith, J.R., Johnson, M.R. and Thum, M.Y.** (2012). Mechanism of human chorionic gonadotrophin-mediated immunomodulation in pregnancy. *Expert Rev Clin Immunol*. **8**, 747-753.
- Benchetrit, H., Herman, S., van Wietmarschen, N., Wu, T., Makedonski, K., Maoz, N., Yom Tov, N., Stave, D., Lasry, R., Zayat, V., et al.** (2015). Extensive Nuclear Reprogramming Underlies Lineage Conversion into Functional Trophoblast Stem-like Cells. *Cell Stem Cell*. **17**, 543-556.
- Blakeley, P., Fogarty, N.M., del Valle, I., Wamaitha, S.E., Hu, T.X., Elder, K., Snell, P., Christie, L., Robson, P. and Niakan, K.K.** (2015). Defining the three cell lineages of the human blastocyst by single-cell RNA-seq. *Development*. **142**, 3151-3165.
- Burton, G.J. and Jauniaux, E.** (2017). The cytotrophoblastic shell and complications of pregnancy. *Placenta*. **60**, 134-139.
- Carey, B.W., Markoulaki, S., Hanna, J.H., Faddah, D.A., Buganim, Y., Kim, J., Ganz, K., Steine, E.J., Cassady, J.P., Creighton, M.P., et al.** (2011). Reprogramming factor stoichiometry influences the epigenetic state and biological properties of induced pluripotent stem cells. *Cell Stem Cell*. **9**, 588-598.
- Castel, G., Meistermann, D., Bretin, B., Firmin, J., Blin, J., Loubersac, S., Bruneau, A., Chevolleau, S., Kilens, S., Chariau, C., et al.** (2020). Induction of Human Trophoblast Stem Cells from Somatic Cells and Pluripotent Stem Cells. *Cell Rep*. **33**, 108419.
- Cauffman, G., De Rycke, M., Sermon, K., Liebaers, I. and Van de Velde, H.** (2009). Markers that define stemness in ESC are unable to identify the totipotent cells in human preimplantation embryos. *Hum Reprod*. **24**, 63-70.
- Chen, D., Liu, W., Zimmerman, J., Pastor, W.A., Kim, R., Hosohama, L., Ho, J., Aslanyan, M., Gell, J.J., Jacobsen, S.E., et al.** (2018). The TFAP2C-Regulated OCT4 Naive Enhancer Is Involved in Human Germline Formation. *Cell Rep*. **25**, 3591-3602 e3595.
- Chiu, Y.H. and Chen, H.** (2016). GATA3 inhibits GCM1 activity and trophoblast cell invasion. *Sci Rep*. **6**, 21630.
- Cinkornpumin, J.K., Kwon, S.Y., Guo, Y., Hossain, I., Sirois, J., Russett, C.S., Tseng, H.W., Okae, H., Arima, T., Duchaine, T.F., et al.** (2020). Naive Human Embryonic Stem Cells Can Give Rise to Cells with a Trophoblast-like Transcriptome and Methylome. *Stem Cell Reports*. **15**, 198-213.

Costa, M.A. (2016). The endocrine function of human placenta: an overview. *Reprod Biomed Online*. **32**, 14-43.

Dong, C., Beltcheva, M., Gontarz, P., Zhang, B., Popli, P., Fischer, L.A., Khan, S.A., Park, K.M., Yoon, E.J., Xing, X., et al. (2020). Derivation of trophoblast stem cells from naive human pluripotent stem cells. *Elife*. **9**.

Ellis, J. (2005). Silencing and variegation of gammaretrovirus and lentivirus vectors. *Hum Gene Ther*. **16**, 1241-1246.

Eloranta, J.J. and Hurst, H.C. (2002). Transcription factor AP-2 interacts with the SUMO-conjugating enzyme UBC9 and is sumoylated in vivo. *J Biol Chem*. **277**, 30798-30804.

Fabregat, A., Jupe, S., Matthews, L., Sidiropoulos, K., Gillespie, M., Garapati, P., Haw, R., Jassal, B., Korninger, F., May, B., et al. (2018). The Reactome Pathway Knowledgebase. *Nucleic Acids Res*. **46**, D649-D655.

Fisher, S.J. (2015). Why is placentation abnormal in preeclampsia? *Am J Obstet Gynecol*. **213**, S115-122.

Gafni, O., Weinberger, L., Mansour, A.A., Manor, Y.S., Chomsky, E., Ben-Yosef, D., Kalma, Y., Viukov, S., Maza, I., Zviran, A., et al. (2013). Derivation of novel human ground state naive pluripotent stem cells. *Nature*. **504**, 282-286.

Gardner, D.K. (1998). Changes in requirements and utilization of nutrients during mammalian preimplantation embryo development and their significance in embryo culture. *Theriogenology*. **49**, 83-102.

Genbacev, O., Krtolica, A., Kaelin, W. and Fisher, S.J. (2001). Human cytotrophoblast expression of the von Hippel-Lindau protein is downregulated during uterine invasion in situ and upregulated by hypoxia in vitro. *Dev Biol*. **233**, 526-536.

Gene Ontology, C. (2021). The Gene Ontology resource: enriching a GOLD mine. *Nucleic Acids Res*. **49**, D325-D334.

Gerri, C., McCarthy, A., Alanis-Lobato, G., Demtschenko, A., Bruneau, A., Loubersac, S., Fogarty, N.M.E., Hampshire, D., Elder, K., Snell, P., et al. (2020). Initiation of a conserved trophectoderm program in human, cow and mouse embryos. *Nature*. **587**, 443-447.

Guo, G., Stirparo, G.G., Strawbridge, S., Yang, J., Clarke, J., Li, M.A., Ozel, B.N., Nichols, J. and Smith, A. (2020). Human Naïve Epiblast Cells Possess Unrestricted Lineage Potential. *BioRxiv*.

Guo, G., Stirparo, G.G., Strawbridge, S.E., Spindlow, D., Yang, J., Clarke, J., Dattani, A., Yanagida, A., Li, M.A., Myers, S., et al. (2021). Human naive epiblast cells possess unrestricted lineage potential. *Cell Stem Cell*.

Haider, S., Meinhardt, G., Saleh, L., Kunihs, V., Gamperl, M., Kaendl, U., Ellinger, A., Burkard, T.R., Fiala, C., Pollheimer, J., et al. (2018). Self-Renewing Trophoblast Organoids Recapitulate the Developmental Program of the Early Human Placenta. *Stem Cell Reports*. **11**, 537-551.

Hemberger, M., Hanna, C.W. and Dean, W. (2020). Mechanisms of early placental development in mouse and humans. *Nat Rev Genet*. **21**, 27-43.

Herbst, F., Ball, C.R., Tuorto, F., Nowrouzi, A., Wang, W., Zavidij, O., Dieter, S.M., Fessler, S., van der Hoeven, F., Kloz, U., et al. (2012). Extensive methylation of promoter sequences silences lentiviral transgene expression during stem cell differentiation in vivo. *Mol Ther*. **20**, 1014-1021.

Hertig, A.T., Rock, J. and Adams, E.C. (1956). A description of 34 human ova within the first 17 days of development. *Am J Anat*. **98**, 435-493.

Home, P., Kumar, R.P., Ganguly, A., Saha, B., Milano-Foster, J., Bhattacharya, B., Ray, S., Gunewardena, S., Paul, A., Camper, S.A., et al. (2017). Genetic redundancy of GATA factors

in the extraembryonic trophoblast lineage ensures the progression of preimplantation and postimplantation mammalian development. *Development*. **144**, 876-888.

Houghton, F.D. (2006). Energy metabolism of the inner cell mass and trophectoderm of the mouse blastocyst. *Differentiation*. **74**, 11-18.

Hu, H., Miao, Y.R., Jia, L.H., Yu, Q.Y., Zhang, Q. and Guo, A.Y. (2019). AnimalTFDB 3.0: a comprehensive resource for annotation and prediction of animal transcription factors. *Nucleic Acids Res.* **47**, D33-D38.

Hubert, M.A., Sherritt, S.L., Bachurski, C.J. and Handwerger, S. (2010). Involvement of transcription factor NR2F2 in human trophoblast differentiation. *PLoS One*. **5**, e9417.

Hyslop, L.A., Blakeley, P., Craven, L., Richardson, J., Fogarty, N.M., Fragouli, E., Lamb, M., Wamaitha, S.E., Prathalingam, N., Zhang, Q., et al. (2016). Towards clinical application of pronuclear transfer to prevent mitochondrial DNA disease. *Nature*. **534**, 383-386.

Io, S., Kabata, M., Iemura, Y., Semi, K., Morone, N., Minagawa, A., Wang, B., Okamoto, I., Nakamura, T., Kojima, Y., et al. (2021). Capturing human trophoblast development with naive pluripotent stem cells in vitro. *Cell Stem Cell*. **28**, 1023-1039 e1013.

Ito, N., Nomura, S., Iwase, A., Ito, T., Ino, K., Nagasaka, T., Tsujimoto, M., Kobayashi, M. and Mizutani, S. (2003). Ultrastructural localization of aminopeptidase A/angiotensinase and placental leucine aminopeptidase/oxytocinase in chorionic villi of human placenta. *Early Hum Dev.* **71**, 29-37.

James, J.L., Carter, A.M. and Chamley, L.W. (2012). Human placentation from nidation to 5 weeks of gestation. Part I: What do we know about formative placental development following implantation? *Placenta*. **33**, 327-334.

Jassal, B., Matthews, L., Viteri, G., Gong, C., Lorente, P., Fabregat, A., Sidiropoulos, K., Cook, J., Gillespie, M., Haw, R., et al. (2020). The reactome pathway knowledgebase. *Nucleic Acids Res.* **48**, D498-D503.

Jokimaa, V., Inki, P., Kujari, H., Hirvonen, O., Ekholm, E. and Anttila, L. (1998). Expression of syndecan-1 in human placenta and decidua. *Placenta*. **19**, 157-163.

Kidder, B.L. and Palmer, S. (2010). Examination of transcriptional networks reveals an important role for TCFAP2C, SMARCA4, and EOMES in trophoblast stem cell maintenance. *Genome Res.* **20**, 458-472.

Kim, D., Paggi, J.M., Park, C., Bennett, C. and Salzberg, S.L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol.* **37**, 907-915.

Kliman, H.J., Nestler, J.E., Sermasi, E., Sanger, J.M. and Strauss, J.F., 3rd (1986). Purification, characterization, and in vitro differentiation of cytotrophoblasts from human term placentae. *Endocrinology*. **118**, 1567-1582.

Kohama, K., Kawaguchi, M., Fukushima, T., Lin, C.Y. and Kataoka, H. (2012). Regulation of pericellular proteolysis by hepatocyte growth factor activator inhibitor type 1 (HAI-1) in trophoblast cells. *Hum Cell*. **25**, 100-110.

Kubaczka, C., Senner, C.E., Cierlitz, M., Arauzo-Bravo, M.J., Kuckenberg, P., Peitz, M., Hemberger, M. and Schorle, H. (2015). Direct Induction of Trophoblast Stem Cells from Murine Fibroblasts. *Cell Stem Cell*. **17**, 557-568.

Kuckenberg, P., Buhl, S., Woynecki, T., van Furden, B., Tolkunova, E., Seiffe, F., Moser, M., Tomilin, A., Winterhager, E. and Schorle, H. (2010). The transcription factor TCFAP2C/AP-2gamma cooperates with CDX2 to maintain trophectoderm formation. *Mol Cell Biol*. **30**, 3310-3320.

Law, C.W., Alhamdoosh, M., Su, S., Smyth, G.K. and Ritchie, M.E. (2016). RNA-seq analysis is easy as 1-2-3 with limma, Glimma and edgeR. *F1000Res*. **5**, 1408.

938 **Lee, C.Q., Gardner, L., Turco, M., Zhao, N., Murray, M.J., Coleman, N., Rossant, J.,**
939 **Hemberger, M. and Moffett, A.** (2016). What Is Trophoblast? A Combination of Criteria Define
940 Human First-Trimester Trophoblast. *Stem Cell Reports*. **6**, 257-272.

941 **Lee, Y., Kim, K.R., McKeon, F., Yang, A., Boyd, T.K., Crum, C.P. and Parast, M.M.** (2007). A
942 unifying concept of trophoblastic differentiation and malignancy defined by biomarker
943 expression. *Hum Pathol*. **38**, 1003-1013.

944 **Leese, H.J., Baumann, C.G., Brison, D.R., McEvoy, T.G. and Sturme, R.G.** (2008). Metabolism
945 of the viable mammalian embryo: quietness revisited. *Mol Hum Reprod*. **14**, 667-672.

946 **Liao, Y., Smyth, G.K. and Shi, W.** (2019). The R package Rsubread is easier, faster, cheaper and
947 better for alignment and quantification of RNA sequencing reads. *Nucleic Acids Res*. **47**, e47.

948 **Lin, S.C., Wani, M.A., Whitsett, J.A. and Wells, J.M.** (2010). Klf5 regulates lineage formation
949 in the pre-implantation mouse embryo. *Development*. **137**, 3953-3963.

950 **Liu, X., Ouyang, J.F., Rossello, F.J., Tan, J.P., Davidson, K.C., Valdes, D.S., Schroder, J., Sun,**
951 **Y.B.Y., Chen, J., Knaupp, A.S., et al.** (2020). Reprogramming roadmap reveals route to human
952 induced trophoblast stem cells. *Nature*. **586**, 101-107.

953 **Love, M.I., Huber, W. and Anders, S.** (2014). Moderated estimation of fold change and
954 dispersion for RNA-seq data with DESeq2. *Genome Biol*. **15**, 550.

955 **Ma, G.T., Roth, M.E., Groskopf, J.C., Tsai, F.Y., Orkin, S.H., Grosfeld, F., Engel, J.D. and Linzer,**
956 **D.I.** (1997). GATA-2 and GATA-3 regulate trophoblast-specific gene expression in vivo.
957 *Development*. **124**, 907-914.

958 **Mandal, P.K. and Rossi, D.J.** (2013). Reprogramming human fibroblasts to pluripotency using
959 modified mRNA. *Nat Protoc*. **8**, 568-582.

960 **Meinhardt, G., Haider, S., Haslinger, P., Proestling, K., Fiala, C., Pollheimer, J. and Knöfler,**
961 **M.** (2014). Wnt-dependent T-cell factor-4 controls human extravillous trophoblast motility.
962 *Endocrinology*. **155**, 1908-1920.

963 **Meistermann, D., Bruneau, A., Loubesac, S., Reigner, A., Firmin, J., Francois-Campion, V.,**
964 **Kilens, S., Lelievre, Y., Lammers, J., Feyeux, M., et al.** (2021). Integrated pseudotime analysis
965 of human pre-implantation embryo single-cell transcriptomes reveals the dynamics of lineage
966 specification. *Cell Stem Cell*.

967 **Mi, H., Muruganujan, A., Ebert, D., Huang, X. and Thomas, P.D.** (2019). PANTHER version 14:
968 more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools.
969 *Nucleic Acids Res*. **47**, D419-D426.

970 **Muraoka, N. and Ieda, M.** (2015). Stoichiometry of transcription factors is critical for cardiac
971 reprogramming. *Circ Res*. **116**, 216-218.

972 **Murthi, P., Kalionis, B., Cocquebert, M., Rajaraman, G., Chui, A., Keogh, R.J., Evain-Brion,**
973 **D. and Fournier, T.** (2013). Homeobox genes and down-stream transcription factor
974 PPARgamma in normal and pathological human placental development. *Placenta*. **34**, 299-
975 309.

976 **Nakagawa, M., Koyanagi, M., Tanabe, K., Takahashi, K., Ichisaka, T., Aoi, T., Okita, K.,**
977 **Mochizuki, Y., Takizawa, N. and Yamanaka, S.** (2008). Generation of induced pluripotent
978 stem cells without Myc from mouse and human fibroblasts. *Nat Biotechnol*. **26**, 101-106.

979 **Nakagawa, M., Takizawa, N., Narita, M., Ichisaka, T. and Yamanaka, S.** (2010). Promotion of
980 direct reprogramming by transformation-deficient Myc. *Proc Natl Acad Sci U S A*. **107**, 14152-
981 14156.

982 **Niwa, H., Toyooka, Y., Shimosato, D., Strumpf, D., Takahashi, K., Yagi, R. and Rossant, J.**
983 (2005). Interaction between Oct3/4 and Cdx2 determines trophectoderm differentiation. *Cell*.
984 **123**, 917-929.

Okae, H., Toh, H., Sato, T., Hiura, H., Takahashi, S., Shirane, K., Kabayama, Y., Suyama, M., Sasaki, H. and Arima, T. (2018). Derivation of Human Trophoblast Stem Cells. *Cell Stem Cell*. **22**, 50-63 e56.

Pastor, W.A., Liu, W., Chen, D., Ho, J., Kim, R., Hunt, T.J., Lukianchikov, A., Liu, X., Polo, J.M., Jacobsen, S.E., et al. (2018). TFAP2C regulates transcription in human naive pluripotency by opening enhancers. *Nat Cell Biol*. **20**, 553-564.

Pelekanos, R.A., Sardesai, V.S., Futrega, K., Lott, W.B., Kuhn, M. and Doran, M.R. (2016). Isolation and Expansion of Mesenchymal Stem/Stromal Cells Derived from Human Placenta Tissue. *J Vis Exp*.

Pijnenborg, R., Dixon, G., Robertson, W.B. and Brosens, I. (1980). Trophoblastic invasion of human decidua from 8 to 18 weeks of pregnancy. *Placenta*. **1**, 3-19.

Pirooznia, M., Nagarajan, V. and Deng, Y. (2007). GeneVenn - A web application for comparing gene lists using Venn diagrams. *Bioinformatics*. **1**, 420-422.

Ralston, A., Cox, B.J., Nishioka, N., Sasaki, H., Chea, E., Rugg-Gunn, P., Guo, G., Robson, P., Draper, J.S. and Rossant, J. (2010). Gata3 regulates trophoblast development downstream of Tead4 and in parallel to Cdx2. *Development*. **137**, 395-403.

Richardson, B.D., Cheng, Y.H., Langland, R.A. and Handwerger, S. (2001). Differential expression of AP-2gamma and AP-2alpha during human trophoblast differentiation. *Life Sci*. **69**, 2157-2165.

Russ, A.P., Wattler, S., Colledge, W.H., Aparicio, S.A., Carlton, M.B., Pearce, J.J., Barton, S.C., Surani, M.A., Ryan, K., Nehls, M.C., et al. (2000). Eomesodermin is required for mouse trophoblast development and mesoderm formation. *Nature*. **404**, 95-99.

Soncin, F., Khater, M., To, C., Pizzo, D., Farah, O., Wakeland, A., Arul Nambi Rajan, K., Nelson, K.K., Chang, C.W., Moretto-Zita, M., et al. (2018). Comparative analysis of mouse and human placentae across gestation reveals species-specific regulators of placental development. *Development*. **145**.

Strumpf, D., Mao, C.A., Yamanaka, Y., Ralston, A., Chawengsaksophak, K., Beck, F. and Rossant, J. (2005). Cdx2 is required for correct cell fate specification and differentiation of trophectoderm in the mouse blastocyst. *Development*. **132**, 2093-2102.

Takashima, Y., Guo, G., Loos, R., Nichols, J., Ficz, G., Krueger, F., Oxley, D., Santos, F., Clarke, J., Mansfield, W., et al. (2015). Resetting Transcription Factor Control Circuitry toward Ground-State Pluripotency in Human. *Cell*. **162**, 452-453.

Tanaka, S., Kunath, T., Hadjantonakis, A.K., Nagy, A. and Rossant, J. (1998). Promotion of trophoblast stem cell proliferation by FGF4. *Science*. **282**, 2072-2075.

Theunissen, T.W., Powell, B.E., Wang, H., Mitalipova, M., Faddah, D.A., Reddy, J., Fan, Z.P., Maetzel, D., Ganz, K., Shi, L., et al. (2014). Systematic Identification of Culture Conditions for Induction and Maintenance of Naive Human Pluripotency. *Cell Stem Cell*. **15**, 524-526.

Turco, M.Y., Gardner, L., Kay, R.G., Hamilton, R.S., Prater, M., Hollinshead, M.S., McWhinnie, A., Esposito, L., Fernando, R., Skelton, H., et al. (2018). Trophoblast organoids as a model for maternal-fetal interactions during human placentation. *Nature*. **564**, 263-267.

Wamaitha, S.E., del Valle, I., Cho, L.T., Wei, Y., Fogarty, N.M., Blakeley, P., Sherwood, R.I., Ji, H. and Niakan, K.K. (2015). Gata6 potently initiates reprogramming of pluripotent and differentiated cells to extraembryonic endoderm stem cells. *Genes Dev*. **29**, 1239-1255.

Wamaitha, S.E., Grybel, K.J., Alanis-Lobato, G., Gerri, C., Ogushi, S., McCarthy, A., Mahadevaiah, S.K., Healy, L., Lea, R.A., Molina-Arcas, M., et al. (2020). IGF1-mediated human embryonic stem cell self-renewal recapitulates the embryonic niche. *Nat Commun*. **11**, 764.

1032 **Warren, L., Manos, P.D., Ahfeldt, T., Loh, Y.H., Li, H., Lau, F., Ebina, W., Mandal, P.K., Smith,**
1033 **Z.D., Meissner, A., et al.** (2010). Highly efficient reprogramming to pluripotency and directed
1034 differentiation of human cells with synthetic modified mRNA. *Cell Stem Cell*. **7**, 618-630.
1035 **Werling, U. and Schorle, H.** (2002). Transcription factor gene AP-2 gamma essential for early
1036 murine development. *Mol Cell Biol*. **22**, 3149-3156.
1037 **Wernig, M., Lengner, C.J., Hanna, J., Lodato, M.A., Steine, E., Foreman, R., Staerk, J.,**
1038 **Markoulaki, S. and Jaenisch, R.** (2008). A drug-inducible transgenic system for direct
1039 reprogramming of multiple somatic cell types. *Nat Biotechnol*. **26**, 916-924.
1040 **Yu, J., Chia, J., Canning, C.A., Jones, C.M., Bard, F.A. and Virshup, D.M.** (2014). WLS
1041 retrograde transport to the endoplasmic reticulum during Wnt secretion. *Dev Cell*. **29**, 277-
1042 291.
1043

Figure 3

Figure 5

Figure 6
A

B

Supplementary Figure 1

Supplementary Figure 2

Supplementary Figure 3