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 Abdelbaki[1], Afshan McCarthy[1], Liani Devito[3], Alice E. Chen[4], Prabhakaran Munusamy[5], Paul Blakeley[1,6], Kay Elder[7], Phil Snell[7], Leila Christie[7], Paul Serhal[8] Rabi A. Odia[8], Mahesh Sangrithi[2,5,9] and Kathy K. Niakan[1,10*]

Affiliations

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

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

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

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

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

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

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

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

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

[10]The 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

65 primitive syncytiotrophoblast (STB). The primitive STB is the initial invading interface and
66 enzymatically degrades the uterine lining allowing for implantation of the conceptus (Hertig
67 et al., 1956). CTB cells are the precursor for trophoblast cell types in the placenta. CTB cells
68 sustain the syncytiotrophoblast across gestation by dividing asymmetrically, with one
69 daughter cell fusing with the overlying syncytiotrophoblast and the other daughter remaining
70 in the proliferative pool (Baczyk et al., 2006). The syncytiotrophoblast performs key functions
71 including providing a physical and immunological barrier to pathogens, transporting nutrients
72 and gaseous exchange, as well as producing and secreting hormones required to adapt the
73 maternal physiology to the pregnancy (Bansal et al., 2012; Costa, 2016). CTB cells are also the
74 precursor of extravillous trophoblast cells (EVTs) which invade the decidua to interact with
75 maternal immune cells and remodel spiral arteries to establish blood supply to the placenta
76 (Pijnenborg et al., 1980).
77
78 Human trophoblast stem cells (hTSCs) derived from blastocysts and first trimester placental
79 villous tissue provide a paradigm for elucidating molecular mechanisms regulating
80 trophoblast development (Okae et al., 2018). hTSCs show expression of well-established
81 markers of trophoblast including TP63 (Lee et al., 2007), GATA3 (Chiu and Chen, 2016) and
82 TEAD4 (Soncin et al., 2018) and can be directed to differentiate into both EVT and STB, as
83 identified by the expression of markers HLA-G (Apps et al., 2008) and SDC1 (Jokimaa et al.,
84 1998) respectively. In addition, the establishment of a trophoblast organoid culture system
85 has allowed hTSCs to be cultured in 3D, somewhat resembling the architecture of villous CTB
86 cells and STB of the placenta (Haider et al., 2018; Turco et al., 2018). Most recently, induced
87 trophoblast stem cells (iTSCs) which resemble primary tissue-derived hTSCs have also been
88 captured in the process of reprogramming fibroblasts to human embryonic stem cells (hESCs)
89 in naïve cell culture conditions (Castel et al., 2020; Cinkornpumin et al., 2020; Dong et al.,
90 2020; Guo et al., 2021; Liu et al., 2020) and when established naïve ES cells are cultured in
91 conditions which inhibit ERK and Nodal signalling (Guo et al., 2021). These *in vitro* models
92 can be utilised in studies of trophoblast biology and have the potential to further inform our
93 understanding of molecular mechanisms, transcription factors and signalling pathways
94 regulating human trophoblast biology.
95

96 Studies of mouse preimplantation embryogenesis have elucidated mechanisms regulating
97 lineage specification and maintenance of the TE in this species (Hemberger et al., 2020). In
98 turn, this knowledge has led to the development of strategies to capture their *in vitro*
99 counterparts. Blastocyst-derived TSCs can be isolated from the polar TE of blastocyst
100 outgrowths or post-implantation extraembryonic ectoderm cultured in TSC media
101 supplemented with FGF4 and heparin which recapitulates the signalling environment of the
102 *in vivo* TSC compartment at the post-implantation stage (Tanaka et al., 1998). Identification
103 of key transcription factors regulating mouse TE lineage specification and maintenance has
104 also informed reprogramming strategies. Single overexpression of *Cdx2* or *Eomes* in TSC
105 media is sufficient to induce differentiation of mouse ESCs to trophoblast-like cells (Niwa et
106 al., 2005). Transient ectopic expression of the combination of TE-associated transcription
107 factors *Gata3*, *Eomes*, *Tcfap2c*, with either *Ets2* or *Myc*, has been shown to reprogramme
108 fibroblasts into mouse iTSCs (Benchetrit et al., 2015; Kubaczka et al., 2015). These
109 alternatively derived cells recapitulate the morphology, epigenomes and gene expression
110 patterns of blastocyst-derived TSCs and contribute exclusively to the placenta following
111 blastocyst injection.
112
113 Here we perform a time-course RNA-sequencing analysis of human TE development and
114 explored gene expression profiles of the human preimplantation TE to identify associated
115 transcription factors that may be employed as reprogramming factors to convert hESCs to
116 iTSCs. We identified four transcription factors (*GATA2*, *GATA3*, *TFAP2C* and *KLF5*) that are
117 enriched in the TE lineage of human blastocysts compared with EPI and PE. We characterised
118 gene expression dynamics across human preimplantation development and confirmed high
119 expression of these factors in TE cells at all stages analysed at both the protein and the
120 transcript level. Similar to strategies used in the mouse, we hypothesized that the transient
121 ectopic expression of these transcription factors, along with the reprogramming enhancing
122 factor *MYC*, would facilitate the establishment of human iTSCs. Here, we demonstrate using
123 both a doxycycline-inducible approach and modified mRNA transfection that the expression
124 of these factors promotes reprogramming of hESCs from primed condition to iTSCs. The iTSCs
125 derived can be maintained stably in culture and recapitulate the expression of markers that
126 have been used previously to identify embryo or primary placental-derived hTSCs (Okae et
127 al., 2018). RNA-seq analysis reveals that iTSCs are transcriptionally similar to existing hTSCs

128 and primary cytotrophoblast cells. When subjected to directed differentiation protocols iTSCs
129 can undergo syncytialisation to form multinucleated structures that secrete a
130 syncytiotrophoblast-associated peptide hormone, human chorionic gonadotropin (hCG).
131 Altogether this indicates a novel strategy for the establishment of iTSCs and expands the
132 repertoire of cells to model placenta biology *in vitro*.
133
134 **Results**
135 **Identification and validation of candidate transcription factors**
136 In this study we sought to devise a reprogramming strategy to generate iTSCs informed by
137 the expression of transcription factors that are enriched in the TE. We initially sought to
138 determine which genes overlap in their expression in the TE across preimplantation stages of
139 human development, reasoning that commonly expressed transcripts may be required for
140 the establishment and maintenance of the TE and thereby may facilitate iTSC reprogramming.
141 We cultured human embryos from embryonic day 5 when the human blastocyst is first
142 established and TE cells are discernible, until embryonic day 7 which is the latest stage we can
143 culture preimplantation blastocysts to *in vitro* (Fig. 1A). We microdissected mural TE cells
144 away from EPI and PE cells from human blastocysts and performed RNA-sequencing analysis
145 to determine the gene expression profile for each developmental stage (Table S1). t-
146 Distributed stochastic neighbour embedding (t-SNE) dimensionality reduction analysis
147 indicated that plotting the first two dimensions of the t-SNE separates TE samples into three
148 groups corresponding to the developmental time-points analysed (Fig. 1B). This was
149 confirmed by a principal component analysis (PCA) that showed that when the first five
150 principal components (PC) are plotted against each other, PC1 separates TE samples into the
151 three developmental time-points (Fig. 1C). We compared the global gene expression patterns
152 at these stages of development to identify the transcripts that are commonly expressed and
153 those unique to each stage (Table S2). We found that the majority of transcripts, 4729 genes,
154 were expressed in common at the stages analysed compared with 576, 387 and 250 genes
155 uniquely expressed at day 5, 6 and 7, respectively. This suggests that once initiated, the TE
156 programme remains transcriptionally consistent across preimplantation development (Fig.
157 1D). Functional enrichment analysis using GeneOntology (Ashburner et al., 2000; Gene
158 Ontology, 2021; Mi et al., 2019) and REACTOME databases (Jassal et al., 2020) revealed
159 enrichment in metabolic processes, translational activities and molecule binding which is

160 consistent with the rapid expansion of the TE (Fig. 1E). Upon further inspection of the gene
161 lists associated with these terms we observed an enrichment in genes that are required for
162 efficient oxidative phosphorylation, including genes encoding complexes of the electron
163 transport chain (*MT-ND1*, *SDHD*, *CYC1*, *UQCRQ, UQCRC1*, *ATP5E*, *ATP5I* and *ATP5L*). In
164 addition, we observed an enrichment of genes involved in mitochondrial morphology.
165 Interestingly while we observed enrichment of genes regulating mitochondrial fusion (*MFN1*,
166 *MFN1* and *OPA1*), we did not detect expression of genes regulating mitochondrial fission
167 (*DRP1*). In all, this suggests that mitochondrial fusion in the TE may be essential for the
168 electron transport chain favouring oxidative phosphorylation. This is consistent with the
169 observation that OXPHOS activity increases in developing mammalian blastocysts and
170 correlates with the capacity of the embryo to develop to term following implantation
171 (Gardner, 1998; Houghton, 2006; Leese et al., 2008). The genes expressed in common
172 included *PPARG* which is expressed in cytotrophoblast cells and regulates differentiation
173 (Murthi et al., 2013), a marker of cytotrophoblast *SPINT1* (Kohama et al., 2012) and *TCF7L1,*
174 a key transcription factor of the WNT signalling pathway that is implicated in trophoblast
175 proliferation (Meinhardt et al., 2014). *ACE2*, *APOE* and the receptor protein-tyrosine kinase
176 *EFNA4* were detected at the earlier time-points analysed whereas *WLS,* a core component of
177 the WNT secretion pathway (Yu et al., 2014), *HAND1* a transcription factor involved in
178 branching morphogenesis during mouse placentation and *VHL* a regulator of placenta
179 vasculogenesis (Genbacev et al., 2001) were enriched at later timepoints (Fig. 1F).
180
181 We collected a comprehensive transcription factor annotation from Human Transcription
182 Factor Database (Hu et al., 2019) and cross-referenced this against the genes expressed in
183 common at all stages analysed. This identified 232 transcription factors expressed in the TE
184 (Table S3). We similarly mined our previously published single-cell RNA-seq datasets to
185 identify transcription factors that were significantly enriched in the TE compared to the EPI
186 and PE cells (Blakeley et al., 2015; Wamaitha et al., 2020). We further refined these lists of
187 TE-associated transcription factors by screening for transcription factors with a known role in
188 mouse TE development, or a suggested role in the human placenta. In all, from these analyses
189 we identified four transcription factors for further investigation as candidate reprogramming
190 factors: *GATA2, GATA3, TFAP2C* and *KLF5* (Fig. 1G) (Table S4).
191

192 From our analysis of single-cell RNA-seq data using our previously published Shiny App
193 programme (Blakeley et al., 2015; Wamaitha et al., 2020), we found that *GATA2*, *GATA3* and
194 *TFAP2C* share a common pattern of expression, with an onset of embryonic transcription
195 shortly prior to TE initiation at the 8-cell stage (Gerri et al., 2020), and high expression
196 maintained as development proceeds. By contrast, *KLF5* is expressed in human zygotes and
197 its expression increases from the 8-cell stage (Blakeley et al., 2015). Analysis of lineage-
198 specific gene expression patterns showed that *GATA2* and *GATA3* are enriched specifically in
199 the TE (Blakeley et al., 2015). While *TFAP2C* is detected in the TE, it is also detected in the EPI,
200 which we previously confirmed at the protein level (Blakeley et al., 2015). *KLF5* is detected in
201 all three lineages, with the most abundant expression in the TE (Blakeley et al., 2015).
202
203 We next performed immunofluorescence analysis of these transcription factors in human
204 embryos cultured from embryonic day 5 to 7, which corresponds to the key morphological
205 stages of TE development: TE formation, expansion and hatching. At all stages analysed,
206 GATA2 protein expression was detected in the nuclei in TE cells, which were identified by both
207 their position within the blastocyst and the absence of NANOG expression (Fig. 2A). Similarly,
208 GATA3 protein was detected in TE cells across the stages analysed (Fig. 2B). The pattern of
209 GATA2 and GATA3 expression is similar to what has been reported in mouse embryos (Home
210 et al., 2017) and is consistent with our previous analysis of GATA3 expression (Gerri et al.,
211 2020). At embryonic day 5 nuclear KLF5 was detected in the TE, whereas the inner cell mass
212 cells showed some cytoplasmic expression (Fig. 2C). At day 6 KLF5 was abundantly expressed
213 in TE nuclei. Lower levels of nuclear KLF5 expression were also detectable in cells within the
214 inner cell mass, which colocalised with SOX2 expression. This pattern is similar to what has
215 been described in the mouse, where KLF5 is expressed in all cells of preimplantation embryos
216 with lower expression in inner cell mass compared to the TE (Lin et al., 2010). At day 7, KLF5
217 expression was abundant in the TE, and the protein was not detectable in the inner cell mass.
218 By contrast, TFAP2C was detected at similar levels in both NANOG-positive epiblast and TE
219 cells throughout all of the blastocyst stages analysed. As we, and others, reported this is in
220 contrast to the expression pattern in the mouse where the homologue TCFAP2C is exclusively
221 expressed in the TE (Fig. 2D) (Blakeley et al., 2015; Kuckenberg et al., 2010). Altogether, the
222 enriched expression of GATA2, GATA3, TFAP2C and KLF5 in the TE suggests that these
223 transcription factors may have a functional role within the TE transcriptional network.

224 **Generation of iTSCs from hESCs using lentiviral doxycycline inducible system**

225 We next evaluated whether the expression of these transcription factors was sufficient to
226 facilitate the reprogramming of hESCs cultured in primed conditions directly to induced
227 trophoblast stem cells (iTSCs) without a requirement to transit through a naïve hESC state.
228 We engineered doxycycline-inducible hESCs to overexpress *TFAP2C*, *KLF5*, *GATA3* and *GATA2*
229 (Fig. 3A). We also induced the expression of *MYC* because we identified high expression of
230 the transcript across the stages of TE development analysed (Table S4), and it has been shown
231 to enhance the efficiency of cellular reprogramming in other contexts (Nakagawa et al., 2008).
232 Individual lentiviruses expressing each of the five genes were packaged, pooled and used to
233 transduce hESC. These cells are henceforth denoted five factor-hESC (5F-hESCs).
234
235 Transgenes were induced by exposing 5F-hESCs to doxycycline for 20 days (Fig. 3B) in the
236 presence of an hTSC media which has been described previously (Okae at al. 2018). As a
237 control, 5F-hESCs were cultured in hTSC media in the absence of doxycycline (uninduced, Fig.
238 3C). After 20 days exposure to doxycycline epithelial colonies with polygonal hTSC-like
239 morphology were apparent (induced, Fig. 3C). These colonies could be isolated by passaging
240 onto freshly coated collagen plates and continued to maintain their morphology and
241 proliferate as iTSCs in the absence of doxycycline (Fig. 3D). Stable iTSC lines were generated
242 and were maintained in culture up to ten passages. By contrast, uninduced 5F-hESCs did not
243 survive the first passage. Immunofluorescence analysis of the iTSCs after 8 passages in the
244 absence of doxycycline indicated widespread expression of endogenous GATA3 and TFAP2C
245 as well as expression of the TE-associated keratin, KRT18 (Cauffman et al., 2009; Lee et al.,
246 2007), and TP63 (Fig. 3E). By contrast, 5F-hESCs cultured in conventional primed conditions
247 did not exhibit appreciable expression of these factors (Fig. 3F).
248
249 **Generation of iTSCs from hESCs using non-integrating modified mRNAs**

250 Reprogramming via lentiviral transduction is reportedly an inefficient process (from 0.01% to
251 0.1%) (Wernig et al., 2008) and integrated constructs can be spontaneously silenced and
252 reactivated during cell culture and differentiation (Ellis, 2005; Herbst et al., 2012). As an
253 alternative approach to lentiviral transduction, in parallel we employed a strategy of
254 transcription factor overexpression using chemically modified mRNAs, which has been used
255 to generate transgene-free human iPSCs (Mandal and Rossi, 2013; Warren et al., 2010).

256 Individual mRNAs encoding *GATA2, GATA3, TFAP2C, KLF5* and *MYC* were synthesised by *in*
257 *vitro* transcription. Uridine and cytidine were substituted with the modified nucleotides
258 pseudo-uridine and 5'methylcytidine to prevent cellular immune responses, and mRNA was
259 capped with a modified 5'guanine cap to improve mRNA half-life. A cocktail of mRNA was
260 made by pooling individual mRNAs in equal molar ratios and this was delivered into hESCs by
261 lipofection every day for 20 days (Fig. 3G). Similar to what was observed with the doxycycline-
262 inducible system, transfected cells underwent morphological change towards a polygonal
263 shape (Fig. 3H). Cells were passaged after 20 days of transfections and iTSC-like colonies were
264 observed. Stable iTSC lines that were generated following mRNA transfection were
265 maintained in culture up to fifteen passages (Fig. 3I). Immunofluorescence analysis of the
266 iTSCs after 8 passages following mRNA reprogramming indicated widespread expression of
267 GATA3, TFAP2C, KRT18 and TP63 at levels comparable to established hTSCs (Fig. 3J).
268
269 **iTSCs are transcriptionally similar to previously established TSCs**
270 We next compared global gene expression of mRNA-generated iTSCs to alternatively derived
271 TSCs by integrating previously published datasets (Dong et al., 2020; Liu et al., 2020; Okae et
272 al., 2018), as well as primary CTB cells (Haider et al., 2018) and primed hESCs (Dong et al.,
273 2020). After adjusting for batch effects, we used the top 500 most variably expressed genes
274 to perform dimensionality reduction analysis. Principal component analysis showed that
275 plotting the first two principal components, which together account for 95% of the variance,
276 separates samples into three groups representing all *in vitro* trophoblast stem cells, primary
277 CTBs, and hESCs (Fig. 4A). To determine which samples were most similar with respect to
278 genes with significant expression changes, we performed a sample-to-sample distances and
279 similarity matrix analysis. Consistent with the principal component analysis, three major
280 clusters were observed: TSCs, CTBs and hESCs. Significantly, this analysis showed that the
281 iTSCs generated in this study were most similar to hTSCs generated by Okae et al., followed
282 by iTSCs generated from reprogrammed fibroblasts (Dong et al., 2020; Liu et al., 2020)(Fig.
283 4B). Plotting normalised count expression values for selected trophoblast-associated genes
284 *GATA2*, *GATA3*, *TEAD4*, *TFAP2A* (Hubert et al., 2010) and *ENPEP* (Ito et al., 2003) showed
285 shared high expression of these factors between iTSCs from this study with previously derived
286 iTSCs (Dong et al., 2020; Liu et al., 2020) and hTSCs (Okae et al., 2018), as well as primary CTBs
287 (Haider et al., 2018) (Fig. 4C). Conversely, iTSCs had downregulated the pluripotency factor

288 *NANOG.* In all, these analyses confirmed that we successfully reprogrammed iTSCs that
289 resemble existing TSC lines based on the global transcriptome.
290
291 **iTSCs function similarly to blastocyst- and placental-derived TSCs**
292 We next examined whether iTSCs were able to undergo differentiation. We utilised a method
293 to generate terminally differentiated syncytiotrophoblast as has been previously described
294 (Okae et al., 2018). Accordingly, we observed formation of multinucleated syncytia which
295 exhibited a reduction in filamentous actin (F-actin) expression, indicating reorganisation of
296 the actin cytoskeleton (Fig. 5A). hCG is one of the first peptide hormones that is produced by
297 the syncytiotrophoblast (Kliman et al., 1986). Spent culture media from the iTSC-derived
298 syncytiotrophoblast cells was collected after 6 days of differentiation and subjected to an
299 over-the-counter pregnancy test kit which showed detectable hCG expression, similar to
300 exisiting hTSCs (Fig. 5B).
301
302 **Narrowing down transcription factors essential to generate iTSCs**
303 To determine which transcription factors are required for the induction of iTSCs, we examined
304 the effect of withdrawing individual transcription factors from the pool of factors on the
305 formation of hTSCs. MYC was kept constant reasoning that based on its expression pattern in
306 the embryo and reprogramming effect in other contexts that it functions to generally enhance
307 reprogramming rather than being essential for iTSC generation (Nakagawa et al., 2010). To
308 allow us to directly compare the success and efficiency of the iTSC programme induction we
309 maintained an equivalent total amount mRNA in each cocktail by replacing the omitted
310 factors with an equivalent amount of mRNA encoding GFP.
311
312 After 10 days of lipofection, cells were immunofluorescently analysed for the expression of
313 KRT18 and TP63 (Fig. 6A; Fig. S1; Fig. S2)(Meistermann et al., 2021). Induction of the iTSC-
314 programme was first determined by the presence of cells co-expressing nuclear TP63 and
315 filamentous KRT18 expression extending from the nucleus to the cell membrane, as we
316 observed in cells treated with the 5-factor cocktail (Fig. 6B:A). Immunofluorescence analysis
317 for the detection of the additional markers, TEAD4 and KRT7, further confirmed the identity
318 of iTSCs (Fig. S4:A). We observed that only the combination of *GATA2*, *GATA3* and *MYC*
319 produced appreciable colonies of cells showing co-expression of TP63 and KRT18 (Fig. 6B:K),

320 albeit at lower proportions compared to the 5-factor cocktail. This 3-factor cocktail also
321 produced cells expressing nuclear TEAD4 and KRT7 (Fig. S3K; yellow arrowheads). By contrast,
322 while the other conditions showed some upregulation of KRT7 and KRT18, the expression
323 pattern of the latter was patchy and disorganised, and the cells did not co-express TP63 nor
324 exhibit nuclear TEAD4 (Fig. 6B)(Fig. S3). This suggests a functional requirement for GATA2 and
325 GATA3 for the induction of iTSCs.
326
327 **Discussion**
328 In this study we investigate the sufficiency of four TE-associated transcription factors in
329 reprogramming hESCs to iTSCs. In the mouse CDX2, GATA3, TFAP2C and EOMES were
330 identified as key regulators of the TE lineage (Kuckenberg et al., 2010; Ralston et al., 2010;
331 Russ et al., 2000; Strumpf et al., 2005). These factors are highly enriched in mouse TE and
332 mTSCs and have been shown to be capable of reprogramming fibroblasts towards mouse iTSC
333 (Benchetrit et al., 2015; Kubaczka et al., 2015). In this study we sought to identify human-
334 specific TE transcription factors capable of reprogramming alternative cell types to iTSCs.
335
336 We identified GATA2, GATA3, TFAP2C and KLF5 as factors that are enriched in the human TE.
337 Together with MYC, these transcription factors were capable of reprogramming hESCs to
338 iTSCs that were transcriptionally equivalent to existing hTSCs and can be directed to
339 differentiate to alternative trophoblast populations. Future experiments to verify the *in vivo*
340 bipotent differential potential of these hTSCs may include engraftment studies into mouse
341 tissues and histological assessment to confirm their differentiation to both EVT and STB. In
342 support of GATA3 as a candidate hTSC reprogramming factor is the recent work implicating
343 GATA3 in the initiation of the TE lineage in human embryogenesis (Gerri et al., 2020). GATA2
344 is exclusively expressed in the TE of mouse, cow and human embryos, and regulates some
345 trophoblast-related genes in mouse and cow (Bai et al., 2011; Gerri et al., 2020; Ma et al.,
346 1997). The mouse homologue of *TFAP2C* (*Tcfap2c*) is required post-implantation with
347 knockout mice dying at embryonic day 7.5 due to proliferation and differentiation defects in
348 the TE compartment (Auman et al., 2002; Werling and Schorle, 2002). In mouse TSCs,
349 TCFAP2C is detected at the promoters of key genes including *Elf5, Gata3, Hand1, Id2* and
350 *Tead4* (Kidder and Palmer, 2010). By contrast, the role of *TFAP2C* in human trophoblast
351 biology is not fully understood. It is known, however, that TFAP2C is a defining marker of CTBs

352 across gestation (Lee et al., 2016) and it stimulates human placenta lactogen and hCG
353 production (Richardson et al., 2001). Functional analysis of KLF5 in human TE and trophoblast
354 has not yet been performed, but in the mouse this transcription factor is required for the
355 establishment of the inner cell mass and TE. *Klf5*$^{-/-}$ embryos arrest at the blastocyst stage due
356 to defects in TE development resulting in a failure to hatch (Lin et al., 2010). The expression
357 patterns of these factors during human embryogenesis, together with their associated roles
358 in TE and trophoblast biology, implicate *GATA2*, *GATA3*, *TFAP2C* and *KLF5* as candidate
359 regulators of a human TE programme.
360
361 Recent studies have shown that reprogramming primed hESCs to naïve hESCs generates a
362 small side-population of hTSCs that can be propagated in culture and recapitulate established
363 hTSCs in terms of molecular markers, transcriptome, and directed differentiation potential
364 (Castel et al., 2020; Cinkornpumin et al., 2020; Dong et al., 2020; Guo et al., 2021; Liu et al.,
365 2020). Additionally, naïve hESCs treated with either small molecule inhibitors of
366 ERK/mitogen-activated protein kinase (MAPK) and Nodal signalling for 3-5 days (Guo et al.,
367 2020; Guo et al., 2021), or for 3 days with transient addition of BMP4 and JAK1 inhibitor (Io
368 et al., 2021), differentiate to hTSCs when placed in conventional hTSC media. Human naïve
369 hESCs represent an earlier, less-fixed developmental state that reflects gene expression and
370 epigenetic profiles similar to that of the early epiblast or late morula (Gafni et al., 2013;
371 Takashima et al., 2015; Theunissen et al., 2014). A feature of naïve pluripotency is the
372 expression of TFAP2C. TFAP2C plays a critical role during primed to naïve reversion by
373 facilitating the opening of naïve-specific enhancers, as well as regulating the expression of the
374 pluripotency factors OCT4 (Pastor et al., 2018) and KLF4 (Chen et al., 2018). Indeed,
375 CRISPR/Cas9-mediated knockout of *TFAP2C* in naïve hESCs significantly affected the efficiency
376 of differentiation to hTSCs (Guo et al., 2021). This suggests a dual functionality of TFAP2C in
377 human embryogenesis and stem cell counterparts in terms of maintaining naïvety and
378 regulating a TE transcriptional network. Indeed, TFAP2C is suggested to both activate and
379 repress target genes (Eloranta and Hurst, 2002; Pastor et al., 2018). Further molecular analysis
380 of hTSCs may reveal the TE-specific repertoire of TFAP2C gene targets and provide further
381 insight into the molecular mechanisms underpinning the TE transcriptional network.
382

383 Among the five transcription factors, we found that *GATA2* and *GATA3* are essential for
384 upregulation of the hTSC programme. Surprisingly, and in contrast to the mouse, *TFAP2C*
385 appears dispensable. While our reprogramming strategies worked, in the future, the
386 efficiency of hTSC reprogramming may be further enhanced by either equivalently increasing
387 the amount of each factor or by altering the stoichiometry of transcription factors, as has
388 been demonstrated for induced pluripotent stem cell and cardiomyocyte reprogramming
389 (Carey et al., 2011; Muraoka and Ieda, 2015). In addition, functional examination of these
390 factors in knock-out hTSCs may further refine the combination of essential factors required
391 for the maintenance of these cell types and these will be important studies in the future. The
392 mRNA-based strategy is advantageous towards these goals as it easily allows for the
393 generation of bespoke combinations of transcription factor as well as the alteration of the
394 stoichiometry of factors in the cocktail.
395
396 Defective TE specification, trophoblast differentiation and maturation leading to abnormal
397 placentation underlies miscarriage and preeclampsia (Burton and Jauniaux, 2017; Fisher,
398 2015). hTSCs and trophoblast organoids present a paradigm for modelling placenta
399 development and disease *in vitro*. However, a caveat of existing cytotrophoblast- and
400 blastocyst-derived hTSCs and trophoblast organoids is that it cannot be ascertained whether
401 the starting population of cells would have given rise to a normal or a disease-affected
402 placenta. Attempts to derive hTSCs and trophoblast organoids after 12 weeks' gestation, or
403 at even later stages coincident with placental-related disease manifestation, have not been
404 successful (Haider et al., 2018; Okae et al., 2018; Turco et al., 2018). In addition, as we do not
405 have an understanding of the genetic basis of placental-related diseases we cannot currently
406 apply CRISPR/Cas9-mediated genome editing to create mutations for disease modelling in
407 existing hTSCs. Instead, we propose that the mRNA reprogramming strategy presented here
408 could be applied and further refined to generate iTSCs from patient fibroblasts, or fibroblasts
409 or mesenchymal stromal cells isolated from disease-affected placentas (Pelekanos et al.,
410 2016). This strategy could open up the possibility of generating patient-specific human TSCs.
411 Alternatively, if fibroblasts are not amenable to this conversion, reprogramming could be
412 employed to generate patient-specific primed iPSCs as the starting population before
413 applying our five-factor strategy. This may have an advantage over transiting through naïve
414 iPSCs if it is found that epigenetic imprints are lost during iTSC generation depending on the

415 starting cell type or if there are persistent issues of karyotypic instability in the starting cell
416 type. Additionally, mRNA transfection has the benefit of being readily transferable to future
417 clinical application as it avoids the random integration of transgenes that can result in
418 genomic modification and tumorigenicity (Warren et al., 2010). In all, this strategy could allow
419 for the generation of a catalogue of clinically normal and disease-associated hTSCs that would
420 provide a tool for basic research into trophoblast biology as well as a powerful tool for
421 elucidating placental defects including recurrent miscarriage, preeclampsia, intrauterine
422 growth restriction and stillbirth as well as a future drug screening platform.
423
424 **Methods**
425 *Human embryo thaw and culture conditions*
426 Human embryos that were surplus to family building requirements were donated to the
427 Francis Crick Institute for use in research projects under the UK Human Fertilisation and
428 Embryology Authority License number R0162 and the Health Research Authority's Research
429 Ethics Committee (Cambridge Central reference number 16/EE/0067). Slow-frozen
430 blastocysts (day 5 and day 6) were thawed using the BlastThaw (Origio; Cat. No. 10542010A)
431 kit using the manufacturer's instructions. Vitrified blastocysts (day 5 and day 6) were thawed
432 using the vitrification thaw kit (Irvine Scientific; Cat. No. 90137-SO) following the
433 manufacturer's instructions. Human embryos were cultured in pre-equilibrated Global Media
434 supplemented with 5 mg/ml Life Global HSA (both LifeGlobal; Cat. No. LGG-020 and LGPS-
435 605) and overlaid with mineral oil (Origio; Cat. No. ART-4008-5P) and incubated in
436 Embryoscope+ time lapse incubator (Vitrolife). For the collection of day 5 samples, embryos
437 were fixed approximately 2 h after thawing to allow them to recover. To collect day 6 and 7
438 samples, day 5 or 6 embryos were cultured to the appropriate time before fixation.
439
440 *Microdissection of TE from human blastocysts*
441 Embryos were placed in drops of G-MOPS solution (Vitrolife; Cat. No. 10129) on a petri dish
442 overlaid with mineral oil. The plate was placed on a microscope stage (Olympus IX70) and the
443 embryos were held with an opposing holding pipette and blastomere biopsy pipette
444 (Research Instruments) using micromanipulators (Narishige, Japan). The biopsy mode of a
445 Saturn 5 laser (Research Instruments) was used to separate the mural TE from the ICM and
446 polar TE. The separated mural TE was transferred to individual low bind RNAse-free tube

447 containing 0.25 µl RNase inhibitor, 4.75 µl Dilution buffer (SMARTer Ultra Low Input RNA kit,
448 Clontech; Cat. No. 634820) and 5 µl nuclease-free water on a pre-chilled CoolRack (Biocision,
449 CA). Samples were stored at - 80°C until ready to be processed.
450
451 *cDNA synthesis and library preparation of TE samples*
452 cDNA was synthesized using SMARTer Ultra Low Input RNA for Illumina Sequencing-HV kit
453 (Clontech Laboratories; Cat. No. 634820) according to the manufacturer's instructions and as
454 previously published (Blakeley et al., 2015; Hyslop et al., 2016). cDNA was sheared using
455 Covaris S2 with the modified settings 10% duty, intensity 5, burst cycle 200 for 2 min. Libraries
456 were prepared using Low Input Library Prep Kit (Clontech Laboratories; Cat. No. 634900)
457 according to the manufacturer's instructions. Library quality was assessed with an Agilent
458 2100 BioAnalyser and concentration measured by QuBit broad range assay (Thermo Fisher;
459 Cat. No. Q32850). Prepared libraries were submitted for 50-bp paired-end sequencing on
460 Illumina HiSeq 2000.
461
462 *cDNA synthesis and library preparation of bulk cell lines*
463 For bulk RNA-seq of cell lines, RNA was isolated using TRI reagent (Sigma-Aldrich; Cat. No.
464 T9424) and DNase I-treated (Ambion; Cat. No. AM2222). Libraries were prepared using the
465 polyA KAPA mRNA HyperPrep Kit (Roche; Cat. No. 8098115702). Quality of submitted RNA
466 samples and the resulting cDNA libraries was determined by ScreenTape Assay on a 4200
467 TapeStation (Agilent). Prepared libraries were submitted for single-end 75 bp sequencing on
468 an Illumina HiSeq 4000 (Illumina).
469
470 *RNA-seq analysis of TE samples*
471 RNA-seq data from human TE was analysed as previously described (Blakeley et al., 2015;
472 Hyslop et al., 2016). Briefly, the reference human genome sequence was obtained from
473 Ensembl, along with the gene annotation (GTF) file. The reference sequence was indexed
474 using the bowtie2-build command. Reads were aligned to the reference human genome
475 sequence using Tophat2 (Kim et al., 2013), with gene annotations to obtain BAM files for each
476 sample. BAM files were then sorted by read coordinates and converted into SAM files using
477 SAMtools. The process of mapping and processing BAM files was automated using a custom
478 Perl script. The number of reads mapping to each gene were counted using the program

479 HTSeq-count (Version 0.6.1; Anders et al., 2015). The resulting count files for each sample
480 were used as input for differential expression analysis using DESeq2 (Anders and Huber,
481 2010). Firstly, the function 'estimateSizeFactors' and 'estimateDispersions' were used to
482 estimate biological variability and calculate normalised relative expression values across the
483 different blastocyst samples. Initially, this was performed without sample labels (option:
484 method='blind') to allow unsupervised clustering of the blastocyst samples using principal
485 components analysis and hierarchical clustering. The dispersion estimates were recalculated
486 with the sample labels included and with the option: method='pooled'. The function
487 'nbinomTest' was then used to calculate p-values to identify genes that show significant
488 differences in expression between different developmental stages.
489
490 An RPKM >5 threshold was applied to generate the lists of genes expressed at each stage
491 analysed. Overlap between the gene lists was determined using the online web application
492 GeneVenn (Pirooznia et al., 2007). Gene lists were used to perform a Gene Ontology and
493 REACTOME functional enrichment analysis (Ashburner et al., 2000; Fabregat et al., 2018) to
494 identify overrepresented categories using a significance threshold of $p \leq 0.05$.
495
496 *Cell culture*
497 H9 hESCs cells were routinely cultured in mTeSR1 (Stem Cell Technologies; Cat. No. 85850)
498 and growth factor reduced Matrigel (BD Biosciences; Cat. No. 356321). For hTSC
499 reprogramming experiments, H9 cells were plated in mTeSR1 media on Matrigel coated cell
500 culture dishes. Media was changed for hTSC media on the first day of reprogramming. hTSC
501 media was prepared as described previously (Okae et al., 2018) (Chir99021, BioTechne, Cat.
502 No. 4423/10; EGF, Tebu Bio Ltd, Cat. No. 167AF-100-15-a; ITS-X supplement, Fisher Sci, Cat.
503 No. 10524233; L-ascorbic acid, Tocris, Cat. No. 4055/50; A83-01, Sigma-Aldrich Cat. No.
504 SML0788/5MG; SB431542, Cambridge Bioscience, Cat. No. SM33-2; Valproic acid, Sigma-
505 Aldrich Cat. No. V-006-1ML; Y27632, Stem Cell Technologies, Cat. No. 72302). Media was
506 changed every second day. For transgene induction, doxycycline was added daily at a
507 concentration of 1 µg/ml over 20 days for transgene induction. hTSCs were passaged by
508 dissociation with TrypLE Express (Fisher Thermo Scientific, Cat. No. 12604013) for 15 min at
509 37 C and passaged onto collagen IV-coated (5 µg/ml; Corning, Cat. No. 354233) 10 cm plates
510 and maintained in hTSC media.

511 *Immunostaining*

512 Embryos were fixed with 4% paraformaldehyde in PBS for 1 h at 4°C. Immunofluorescence
513 staining was performed as described previously (Fogarty et al. 2018). The primary antibodies
514 used are listed in Table S5. Embryos were placed on µ-Slide 8 Well coverslip dishes (Ibidi; Cat.
515 No. 80826) for confocal imaging. Imaging was performed using a Leica SP5 confocal
516 microscope and 3 µm thick optical sections were collected. hESC, iTSC and hTSC lines were
517 fixed with PBS 4% PFA for 1 h at 4°C, then washed three times with PBS (Life Technologies;
518 Cat. No. 14190-094). Blocking was achieved by incubation with PBS with 10% donkey serum
519 (Sigma-Aldrich; Cat. No. D9663) and 0.1% Triton X-100 (Sigma-Aldrich; Cat. No. T8787) for 30
520 mins at room temperature. Permeabilisation was performed by incubation in PBS with 0.5%
521 Triton X-100. Primary antibodies were diluted in blocking solution and each incubated
522 overnight at 4 C. Following each incubation, the cells underwent 3 x 5 min washes with PBS
523 with 0.1% Triton X-100. Secondary antibodies were diluted 1:300 in PBS with 0.1% Triton X-
524 100 for 1 h at room temperature. After 3 x 5 min washes in PBS with 0.1% Triton X-100, 5
525 µg/ml DAPI (Sigma-Aldrich; D9542) was added for 2 mins during the final wash in order to
526 perform nuclear staining. Epifluorescence images were performed on an Olympus IX73 using
527 CellSens software (Olympus).

528

529 *Generation of inducible system*

530 Doxycycline-inducible overexpression of transcription factors was achieved using the Lenti-X
531 Tet-On 3G Inducible Expression System (Clontech; Cat. No. 631363) following the
532 manufacturer's protocol, and as previously described (Wamaitha et al., 2015). Coding
533 sequences were sub-cloned from the template plasmids into the pLVX-TRE3G vector to
534 generate individual pLVX-TRE3G-Gene of Interest (GOI) vectors. Lentiviral packaging was
535 achieved using 7 µg of pLVX-TRE3G-GOI and the Lenti-X Packaging Single Shot reagents in
536 HEK-293T cells. Transfection media was replaced after 6 h with fresh MEF media. Lentiviral
537 supernatants were subsequently collected after transfection of HEK 293T cells with either the
538 pLVX-TRE3G-GOI or the pLVX-Tet3G vector using the X-fect reagent (Clontech; Cat. No.
539 631317). Supernatants were concentrated by ultracentrifugation. Equal volumes of lentivirus
540 supernatant encoding *GATA2*, *GATA3*, *TFAP2C*, *MYC* and *KLF5* were pooled to generate a 5-
541 factor cocktail which was aliquoted into single use 10 µl volumes and stored at -80 C until
542 needed. H9 hESCs were sequentially transduced with the pLVX-Tet3G and the 5-factor

543 cocktail. hESCs were plated on growth factor reduced Matrigel (BD Biosciences; Cat. No.
544 356230) in mTeSR1 medium (Stem Cell Technologies; Cat. No. 85850) and grown to 70%
545 confluency. Cells were first transduced with the pLVX-tet3G lentivirus followed by selection
546 with G418 (250 µg/ml) for one week. Resistant cells were selected for at least 2 passages and
547 then transduced with the pLVX-TRE3G-GOI lentivirus pool, followed by selection with
548 puromycin (0.5 µg/ml) for four days. For the induction of GOI expression, doxycycline was
549 added to the mTeSR1 media at a concentration of 1 µg/ml. Clonal lines were generated and
550 screened for transgene integration. Lines were cultured in the presence of doxycycline for 48
551 h before cDNA was harvested and RT-qPCR was performed to determine the levels of
552 transgene induction.
553
554 *Generation of templates for lentivirus production and in vitro transcription (IVT)*
555 The vector design strategy was informed from previously published reports (Mandal and
556 Rossi, 2013). Briefly, nucleotide sequences of the open reading frames for the canonical
557 isoforms of reprogramming factors were identified from the Ensembl genome browser, and
558 the sequences were verified against the corresponding amino acid sequences in Uniprot.
559 Individual *in vitro* transcription (IVT) template constructs for *GATA2, GATA3, TFAP2C* and *KLF5*
560 consisting of a T7 promoter-5'UTR-Open Reading Frame-3'UTR-T7 terminator cassette cloned
561 into a pUC57 backbone were custom made (Genewiz UK Ltd). 5'UTR and 3'UTRs were added
562 to maximize stability of mRNA transcripts and to increase protein translation. The ORFs of
563 *MYC* and *GFP* were templated from plasmids bearing human MYC and GFP and ligated into
564 the pUC57 backbone flanked by the 5'UTR and 3'UTR sequences. Annotated sequence files of
565 all constructs are provided in Table S6.
566
567 *Generation of modified mRNAs by in vitro synthesis*
568 The mRNA synthesis protocol has been described previously (Mandal and Rossi, 2013). Briefly,
569 dsDNA templates were linearised from cDNA clones in pLVX vectors for *GATA2, GATA3,*
570 *TFAP2C, KLF5* and *MYC*. A small amount of digestion mix was run on a gel to confirm complete
571 digestion. Linearised plasmid was purified using PCR purification kit (Qiagen; Cat. No. 28104).
572 A nanodrop was used to confirm the purity of the eluted product according to the 260/280
573 ratios. Poly(A) tail was added using KAPA PCR ready mix (2X), Xu-F1 and Xu-T120 primers
574 (Integrate DNA Technologies) and digested plasmid adjusted to 10 ng/ul. Tail PCR was run for

575 32 cycles and purified using PCR purification kit. *In vitro* transcription was performed using
576 MEGAscript T7 kit (Thermo Fisher; Cat. No. AMB13345): custom NTP mix was prepared with
577 3'-O-Me-m7G cap analogue (60mM, NEB), GTP (75mM, MEGAscript T7 kit), ATP (75mM,
578 MEGAscript T7 kit), Me-CTP (100 mM, TriLink; Cat. No. N-1014-1) and pseudo-UTP (100mM,
579 Tri-link; Cat. No. O-0263). Reaction was heated at 37 C for 2 h. 2 ul of Turbo DNase (Thermo
580 Fisher; Cat. No. AM2238) was added and incubated at 37 C for 15 min. DNAse treated reaction
581 mix was purified using RNAeasy kit (Qiagen; Cat. No. 74104) according to the manufacturer's
582 instructions. RNA was phosphatase-treated using Antarctic phosphatase (New England
583 BioLabs; Cat. No. M0289S) and purified using MEGAclear kit (Thermo Fisher; Cat. No.
584 AM1909). Concentration and quality were measured using a nanodrop and adjusted to 100
585 ng/ul. For reprogramming experiments a modified mRNA cocktail was prepared by mixing the
586 five factors in an equal molar ratio. The cocktail was aliquoted into single use aliquots
587 containing a total mRNA amount of 1 µg and stored at -80 C until needed. For transcription
588 factor combinatorial experiments cocktails were made up to a total amount of 500 ng of
589 mRNA with the relevant quantity of mRNA encoding GFP replacing the omitted factors.
590
591 *Lipofection of hESCs*
592 For one well of a 6-well plate, 1 µg (1x) of mRNA was mixed with 86 µl of OptiMEM reduced
593 serum media in an Eppendorf tube. In a second tube 0.5 x volume of lipofectamine RNAi max
594 (Thermofisher; Cat. No. 13778100) was mixed with 93 µl OptiMEM. Tubes were incubated at
595 room temperature for 5 min. The tubes were then mixed together and incubated at room
596 temperature for 20 min. The lipofection mix was added to the well in a dropwise manner and
597 mixed well by rocking the plate from side to side. The cells were incubated at 37 C in 20% O_2
598 for 4 h. The media was then replaced with hTSC media containing 200 ng/ml recombinant
599 B18R to prevent gamma-interferon response (Stem Cell Technologies; Cat. No. 78075). For
600 reprogramming experiments lipofection was performed at the same time every day for 20
601 days. For transcription factor combinatorial experiments H1 hESCs were lipofected daily for
602 10 days.
603
604 *Directed differentiation*
605 For the induction of syncytiotrophoblast, iTSCs were grown to 70% confluency in 6-well
606 plates. Cells were passaged by incubating in TrypLE (Thermo Fisher Scientific; Cat. No.

607 12604013) for 10 min at 37 C. Cells were seeded in a 6-well plate pre-coated with 2.5 mg/ml
608 Col IV and cultured in 2 mL of syncytiotrophoblast medium comprised of DMEM/F12
609 supplemented with 0.1 mM 2-mercaptoethanol, 0.5% penicillin-streptomycin, 0.3% BSA, 1%
610 ITS-X supplement, 2.5 mM Y27632, 2 mM forskolin, and 4% KSR]. The medium was replaced
611 at day 3, and the cells were fixed at day 6 for analysis.
612
613 *RNA-seq data analysis of induced trophoblast stem cells (iTSCs) and related lines*
614 *Data from this study*
615 Sequencing data from sample replicates comprising iTSCs (n = 4), H9 ES cells (n = 3) and
616 primary CT27 cells (n = 3) were first checked using the *FastQC* package
617 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Adapter removal was
618 performed using *Trimgalore v0.6.6* (https://github.com/FelixKrueger/TrimGalore), and the
619 trimmed read data were re-checked for conformity and quality with *FastQC*.
620
621 *Accessory datasets*
622 Primary cytotrophoblast gene expression data was downloaded from GEO with accession
623 number GSE109976 (Haider et al., 2018). Additional primed hESCs gene expression data was
624 downloaded from GEO with accession numbers GSM4116153 and GSM4116151 (Dong et al.,
625 2020). Trophoblast stem cells derived from human blastocysts gene expression data was
626 downloaded from GSE138762 (Dong et al., 2020). Trophoblast stem cells derived from
627 fibroblast reprogramming gene expression data was downloaded from GSE150616 (Liu et al.,
628 2020) and GSE138762 (Dong et al., 2020).
629
630 Next, sequences were aligned to the reference genome (Homo_sapiens.GRCh38) using
631 *HISAT2 v2.2.1* (Kim et al., 2019), with trimming of 5' and 3' bases performed based on the QC
632 of each of each of these samples. Counts were generated using *FeatureCounts v2.0.1*, and the
633 count matrix was analysed further using *DESeq2* (Liao et al., 2019; Love et al., 2014). Genes
634 with fewer than 10 counts detected across all 22 samples were excluded, leaving 17498 genes
635 for further processing. A variance stabilizing transformation (VST) was applied, and
636 subsequently batch related effects in combining these different datasets were corrected for
637 using the *limma* package (i.e. *limma::removeBatchEffect)* (Law et al., 2016). Sample to sample

638 distances were computed and plotted as a heatmap. Principal Component Analysis (PCA) was
639 performed using the top 500 most highly variable genes.
640

641 **Acknowledgements**
642 We are very grateful to the donors of human embryos whose contributions enable this
643 research. We thank all members of the Niakan lab for help and comments on the paper. We
644 are grateful to the Francis Crick Institute's Science Technology Platforms: Lyn Healey from
645 Human Embryonic Stem Cell Facility, Robert Goldstone and Deb Jackson from the Advanced
646 Sequencing Facility; and the Advanced Light Microscopy Facility.
647

648 **Funding**
649 Work in the K.K.N. laboratory was supported by the Francis Crick Institute, which receives its
650 core funding from Cancer Research UK (FC001120), the UK Medical Research Council
651 (FC001120), and the Wellcome Trust (FC001120). The M.S. lab at Centre for Stem Cells and
652 Regenerative Medicine, King's College London is supported by a Wellcome Trust Clinical
653 Career Development Fellowship (222052/Z/20/Z), and the M.S. group in Singapore (including
654 P.M.) was supported by a National Medical Research Council, Singapore Clinician Scientist
655 Investigator Award (CSAINV17may001). N.M.E.F is supported by a King's Prize Fellowship. For
656 the purpose of Open Access, the authors have applied a CC-BY public copyright licence to any
657 Author Accepted Manuscript version arising from this submission.
658

659 **Competing Interests**
660 No competing interests declared.
661

662 **Author Contributions**
663 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
664 designed the experiments. N.M.E.F. performed the majority of experiments with assistance
665 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.
666 coordinated the donation of embryos to the research project. P.B. and P.M. performed
667 bioinformatic analysis of RNA-seq data. N.M.E.F. and K.K.N. wrote the manuscript with help
668 from all the authors.
669

670 **Data availability**

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

674

675 **Figure Legends**

676 **Figure 1- Transcriptome analysis of TE in human blastocysts.**

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

691

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

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

697

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

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

702 which media was replaced with hTSC media. Doxycycline was administered daily for 20 days
703 in hTSC media for transgene induction. (C) Brightfield images of 5F-hESCs cultured in hTSC
704 media alone for 20 days (uninduced) and hESCs cultured in hTSCs in the presence of dox for
705 20 days (induced). (D) Brightfield images of stable iTSC lines derived from transgene
706 overexpression. Stable iTSCs were maintained up to 10 passages in culture. Scale bar: 200 µm.
707 (E) Immunofluorescence analysis for the detection of KRT18, GATA3, TFAP2C and TP63
708 (green) and DAPI nuclear expression (blue) in stable iTSCs and in previously established
709 control hTSCs (Okae et al., 2018). Scale bars: 50 µm). (F) Immunofluorescence analysis for the
710 detection of GATA3, TFAP2C, KRT18 and TP63 (green) and DAPI nuclear expression (blue) in
711 5F-hESCs cultured in mTeSR1 media. Scale bars: 50 µm. (G) Schematic representation of the
712 strategy for reprogramming hESCs to hTSCs using modified mRNAs. hESCs were plated in
713 mTeSR1 media for 24 h after which media was replaced with hTSC media. mRNAs encoding
714 *GATA2*, *GATA3*, *TFAP2C*, *KLF5* and *MYC* administered daily for 20 days via lipofection-
715 mediated transfection. (H) Brightfield images of mock transfected cells and 5-factor mRNA-
716 cocktail transfected cells on day 20. (I) Brightfield image of stable iTSCs derived from mRNA
717 reprogramming. Scale bar: 200 µm. (J) Immunofluorescence analysis for the detection of
718 KRT18, GATA3, TFAP2C and TP63 (green) and DAPI nuclear expression (blue) in stable iTSCs
719 and previously established control hTSCs (Okae et al., 2018). Scale bars: 50 µm.
720
721 **Figure 4 – Global transcriptional analysis of iTSCs and similarity to previously established**
722 **TSCs**
723 Analysis of mRNA-derived iTSCs (this study), previously established hTSCs (CT27, BT5) (Okae
724 et al., 2018), iTSCs derived from naïve reprogramming experiments (Dong et al., 2020; Liu et
725 al., 2020), primary placental CTBs (Haider et al., 2018) and primed hESCs (this study)(Dong et
726 al., 2020). (A) Principal component analysis (PCA) using the top 500 most variable expressed
727 genes. (B) Sample-to-sample distances and similarity matrix analysis. iTSCs generated in this
728 study are labelled in red. (C) Dot plots of log_2normalised count data for selected trophoblast-
729 associated genes and *NANOG*. iTSCs generated in this study are labelled in red.
730
731 **Figure 5 – iTSCs can be directed to differentiate to syncytiotrophoblast-like cells.**
732 (A) Immunostaining analysis of F-actin (green) and DAPI nuclear expression (blue) in
733 undifferentiated and differentiated iTSCs and existing hTSCs on the sixth day of

734 differentiation. Scale bars: 50 µm. (B) Detection of hCG in spent culture media from
735 syncytiotrophoblast generated from iTSCs and previously established control hTSCs using an
736 over-the-counter pregnancy test. 1. hTSC media, 2. iTSC-derived syncytiotrophoblast
737 replicate 1, 3. iTSC-derived syncytiotrophoblast replicate 2, 4. iTSC-derived
738 syncytiotrophoblast replicate 3, 5. hTSC-derived syncytiotrophoblast, 6. unconditioned
739 syncytiotrophoblast media. Key shows the interpretation of the test results.
740
741 **Figure 6- GATA2 and GATA3 are required for induction of hTSC programme**
742 (A) Schematic representation of the strategy for testing factors required for iTSC
743 reprogramming. hESCs were plated in primed hESC culture conditions for 24 h after which
744 media was replaced with hTSC media. hESCs were transfected with cocktails of modified
745 mRNAs where omitted factors were replaced with equivalent amount of mRNA encoding GFP
746 to keep the molar ratio consistent. Cells were transfected daily for 10 days at which point
747 they were fixed for immunofluorescence analysis (B) Immunofluorescence analysis for the
748 detection of TP63 (red) and KRT18 (yellow) and DAPI nuclear expression (blue) in transfected
749 cells. Merged images are shown. Key for mRNA cocktail combinations: B, GATA2, TFAP2C,
750 KLF5 and MYC (GATA3 omitted); C, GATA3, TFAP2C, KLF5 and MYC (GATA2 omitted); D,
751 GATA2, GATA3, TFAP2C and MYC (KLF5 omitted); E, GATA2, GATA3, KLF5 and MYC (TFAP2C
752 omitted); F, TFAP2C, KLF5 and MYC (GATA2 and GATA3 omitted); G, GATA2, TFAP2C and MYC
753 (GATA3 and KLF5 omitted); H, GATA2, KLF5 and MYC (GATA3 and TFAP2C omitted); I, GATA3,
754 TFAP2C and MYC (GATA2 and KLF5 omitted); J, GATA3, KLF5 and MYC (GATA2 and TFAP2C
755 omitted); K, GATA2, GATA3 and MYC (TFAP2C and KLF5 omitted); L, TFAP2C, KLF5 and MYC
756 (KLF5, GATA2 and GATA3 omitted); M, GATA2 and MYC (GATA3, TFAP2C and KLF5); N, KLF5
757 and MYC (GATA2, GATA3 and TFAP2C omitted). Scale bars: 50 µm.
758
759 **Figure S1 – Short-term exogenous expression of 5 TE-associated transgenes upregulates**
760 **markers of trophoblast.** (A) Schematic representation of the strategy for testing duration of
761 transgene induction required for upregulation of trophoblast markers. 5F-hESCs were plated
762 in mTeSR1 media for 24 h after which media was replaced with hTSC media. Transgenes were
763 induced by addition of doxycycline for 10 days. Cells were cultured in the absence of
764 doxycycline for a further 10 days at which point they were fixed for immunofluorescence
765 analysis. (B) Immunofluorescence analysis for the detection of KRT18, GATA3 and TP63

766 (green) and DAPI nuclear expression (blue) in induced and uninduced 5F-hESCs on day 20.
767 Scale bars: 50 µm.
768
769 **Figure S2 - GATA2 and GATA3 are required for induction of hTSC programme**
770 Extended immunofluorescence analysis for the detection of TP63, KRT18 and DAPI nuclear
771 expression in transfected cells. Key for mRNA cocktail combinations: A, Five factors positive
772 controls; B, *GATA2*, *TFAP2C*, *KLF5* and *MYC* (*GATA3* omitted); C, *GATA3*, *TFAP2C*, *KLF5* and
773 *MYC* (*GATA2* omitted); D, *GATA2*, *GATA3*, *TFAP2C* and *MYC* (*KLF5* omitted); E, *GATA2*, *GATA3*,
774 *KLF5* and *MYC* (*TFAP2C* omitted); F, *TFAP2C*, *KLF5* and *MYC* (*GATA2* and *GATA3* omitted); G,
775 *GATA2*, *TFAP2C* and *MYC* (*GATA3* and *KLF5* omitted); H, *GATA2*, *KLF5* and *MYC* (*GATA3* and
776 *TFAP2C* omitted); I, *GATA3*, *TFAP2C* and *MYC* (*GATA2* and *KLF5* omitted); J, *GATA3*, *KLF5* and
777 *MYC* (*GATA*2 and *TFAP2C* omitted); K, *GATA2*, *GATA3* and *MYC* (*TFAP2C* and *KLF5* omitted); L,
778 *TFAP2C*, *KLF5* and *MYC* (*KLF5*, *GATA2* and *GATA3* omitted); M, *GATA2* and *MYC* (*GATA3*,
779 *TFAP2C* and *KLF5*); N, *KLF5* and *MYC* (*GATA2*, *GATA3* and *TFAP2C* omitted). hESCs cultured in
780 mTeSR1 media are included as a negative controls. Scale bars: 50 µm.
781
782 **Figure S3. Immunofluorescence analysis of additional markers in combinatorial**
783 **experiments**
784 Immunofluorescence analysis for the detection of TEAD4, KRT7 and DAPI nuclear expression
785 in transfected cells. Key for mRNA cocktail combinations: A, Five factors positive controls; B,
786 *GATA2*, *TFAP2C*, *KLF5* and *MYC* (*GATA3* omitted); C, *GATA3*, *TFAP2C*, *KLF5* and *MYC* (*GATA2*
787 omitted); D, *GATA2*, *GATA3*, *TFAP2C* and *MYC* (*KLF5* omitted); E, *GATA2*, *GATA3*, *KLF5* and
788 *MYC* (*TFAP2C* omitted); F, *TFAP2C*, *KLF5* and *MYC* (*GATA2* and *GATA3* omitted); G, *GATA2*,
789 *TFAP2C* and *MYC* (*GATA3* and *KLF5* omitted); H, *GATA2*, *KLF5* and *MYC* (*GATA3* and *TFAP2C*
790 omitted); I, *GATA3*, *TFAP2C* and *MYC* (*GATA2* and *KLF5* omitted); J, *GATA3*, *KLF5* and *MYC*
791 (*GATA2* and *TFAP2C* omitted); K, *GATA2*, *GATA3* and *MYC* (*TFAP2C* and *KLF5* omitted); L,
792 *TFAP2C*, *KLF5* and *MYC* (*KLF5*, *GATA2* and *GATA3* omitted); M, *GATA2* and *MYC* (*GA*TA3,
793 *TFAP2C* and *KLF5*); N, *KLF5* and *MYC* (*GATA2*, *GATA3* and *TFAP2C* omitted). hESCs culture in
794 in mTeSR1 media are included as a negative controls. Scale bars: 50 µm.
795
796
797

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., Giannoulias, 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., Creyghton, 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.

845 **Costa, M.A.** (2016). The endocrine function of human placenta: an overview. *Reprod Biomed*
846 *Online*. **32**, 14-43.
847 **Dong, C., Beltcheva, M., Gontarz, P., Zhang, B., Popli, P., Fischer, L.A., Khan, S.A., Park, K.M.,**
848 **Yoon, E.J., Xing, X., et al.** (2020). Derivation of trophoblast stem cells from naive human
849 pluripotent stem cells. *Elife*. **9**.
850 **Ellis, J.** (2005). Silencing and variegation of gammaretrovirus and lentivirus vectors. *Hum Gene*
851 *Ther*. **16**, 1241-1246.
852 **Eloranta, J.J.and Hurst, H.C.** (2002). Transcription factor AP-2 interacts with the SUMO-
853 conjugating enzyme UBC9 and is sumolated in vivo. *J Biol Chem*. **277**, 30798-30804.
854 **Fabregat, A., Jupe, S., Matthews, L., Sidiropoulos, K., Gillespie, M., Garapati, P., Haw, R.,**
855 **Jassal, B., Korninger, F., May, B., et al.** (2018). The Reactome Pathway Knowledgebase.
856 *Nucleic Acids Res*. **46**, D649-D655.
857 **Fisher, S.J.** (2015). Why is placentation abnormal in preeclampsia? *Am J Obstet Gynecol*. **213**,
858 S115-122.
859 **Gafni, O., Weinberger, L., Mansour, A.A., Manor, Y.S., Chomsky, E., Ben-Yosef, D., Kalma,**
860 **Y., Viukov, S., Maza, I., Zviran, A., et al.** (2013). Derivation of novel human ground state naive
861 pluripotent stem cells. *Nature*. **504**, 282-286.
862 **Gardner, D.K.** (1998). Changes in requirements and utilization of nutrients during mammalian
863 preimplantation embryo development and their significance in embryo culture.
864 *Theriogenology*. **49**, 83-102.
865 **Genbacev, O., Krtolica, A., Kaelin, W.and Fisher, S.J.** (2001). Human cytotrophoblast
866 expression of the von Hippel-Lindau protein is downregulated during uterine invasion in situ
867 and upregulated by hypoxia in vitro. *Dev Biol*. **233**, 526-536.
868 **Gene Ontology, C.** (2021). The Gene Ontology resource: enriching a GOld mine. *Nucleic Acids*
869 *Res*. **49**, D325-D334.
870 **Gerri, C., McCarthy, A., Alanis-Lobato, G., Demtschenko, A., Bruneau, A., Loubersac, S.,**
871 **Fogarty, N.M.E., Hampshire, D., Elder, K., Snell, P., et al.** (2020). Initiation of a conserved
872 trophectoderm program in human, cow and mouse embryos. *Nature*. **587**, 443-447.
873 **Guo, G., Stirparo, G.G., Strawbridge, S., Yang, J., Clarke, J., Li, M.A., Ozel, B.N., Nichols, J.and**
874 **Smith, A.** (2020). Human Naïve Epiblast Cells Possess Unrestricted Lineage Potential. *BioRXIV*.
875 **Guo, G., Stirparo, G.G., Strawbridge, S.E., Spindlow, D., Yang, J., Clarke, J., Dattani, A.,**
876 **Yanagida, A., Li, M.A., Myers, S., et al.** (2021). Human naive epiblast cells possess unrestricted
877 lineage potential. *Cell Stem Cell*.
878 **Haider, S., Meinhardt, G., Saleh, L., Kunihs, V., Gamperl, M., Kaindl, U., Ellinger, A., Burkard,**
879 **T.R., Fiala, C., Pollheimer, J., et al.** (2018). Self-Renewing Trophoblast Organoids Recapitulate
880 the Developmental Program of the Early Human Placenta. *Stem Cell Reports*. **11**, 537-551.
881 **Hemberger, M., Hanna, C.W.and Dean, W.** (2020). Mechanisms of early placental
882 development in mouse and humans. *Nat Rev Genet*. **21**, 27-43.
883 **Herbst, F., Ball, C.R., Tuorto, F., Nowrouzi, A., Wang, W., Zavidij, O., Dieter, S.M., Fessler, S.,**
884 **van der Hoeven, F., Kloz, U., et al.** (2012). Extensive methylation of promoter sequences
885 silences lentiviral transgene expression during stem cell differentiation in vivo. *Mol Ther*. **20**,
886 1014-1021.
887 **Hertig, A.T., Rock, J.and Adams, E.C.** (1956). A description of 34 human ova within the first
888 17 days of development. *Am J Anat*. **98**, 435-493.
889 **Home, P., Kumar, R.P., Ganguly, A., Saha, B., Milano-Foster, J., Bhattacharya, B., Ray, S.,**
890 **Gunewardena, S., Paul, A., Camper, S.A., et al.** (2017). Genetic redundancy of GATA factors

891 in the extraembryonic trophoblast lineage ensures the progression of preimplantation and
892 postimplantation mammalian development. *Development*. **144**, 876-888.
893 **Houghton, F.D.** (2006). Energy metabolism of the inner cell mass and trophectoderm of the
894 mouse blastocyst. *Differentiation*. **74**, 11-18.
895 **Hu, H., Miao, Y.R., Jia, L.H., Yu, Q.Y., Zhang, Q.and Guo, A.Y.** (2019). AnimalTFDB 3.0: a
896 comprehensive resource for annotation and prediction of animal transcription factors.
897 *Nucleic Acids Res*. **47**, D33-D38.
898 **Hubert, M.A., Sherritt, S.L., Bachurski, C.J.and Handwerger, S.** (2010). Involvement of
899 transcription factor NR2F2 in human trophoblast differentiation. *PLoS One*. **5**, e9417.
900 **Hyslop, L.A., Blakeley, P., Craven, L., Richardson, J., Fogarty, N.M., Fragouli, E., Lamb, M.,**
901 **Wamaitha, S.E., Prathalingam, N., Zhang, Q., et al.** (2016). Towards clinical application of
902 pronuclear transfer to prevent mitochondrial DNA disease. *Nature*. **534**, 383-386.
903 **Io, S., Kabata, M., Iemura, Y., Semi, K., Morone, N., Minagawa, A., Wang, B., Okamoto, I.,**
904 **Nakamura, T., Kojima, Y., et al.** (2021). Capturing human trophoblast development with naive
905 pluripotent stem cells in vitro. *Cell Stem Cell*. **28**, 1023-1039 e1013.
906 **Ito, N., Nomura, S., Iwase, A., Ito, T., Ino, K., Nagasaka, T., Tsujimoto, M., Kobayashi, M.and**
907 **Mizutani, S.** (2003). Ultrastructural localization of aminopeptidase A/angiotensinase and
908 placental leucine aminopeptidase/oxytocinase in chorionic villi of human placenta. *Early Hum*
909 *Dev*. **71**, 29-37.
910 **James, J.L., Carter, A.M.and Chamley, L.W.** (2012). Human placentation from nidation to 5
911 weeks of gestation. Part I: What do we know about formative placental development
912 following implantation? *Placenta*. **33**, 327-334.
913 **Jassal, B., Matthews, L., Viteri, G., Gong, C., Lorente, P., Fabregat, A., Sidiropoulos, K., Cook,**
914 **J., Gillespie, M., Haw, R., et al.** (2020). The reactome pathway knowledgebase. *Nucleic Acids*
915 *Res*. **48**, D498-D503.
916 **Jokimaa, V., Inki, P., Kujari, H., Hirvonen, O., Ekholm, E.and Anttila, L.** (1998). Expression of
917 syndecan-1 in human placenta and decidua. *Placenta*. **19**, 157-163.
918 **Kidder, B.L.and Palmer, S.** (2010). Examination of transcriptional networks reveals an
919 important role for TCFAP2C, SMARCA4, and EOMES in trophoblast stem cell maintenance.
920 *Genome Res*. **20**, 458-472.
921 **Kim, D., Paggi, J.M., Park, C., Bennett, C.and Salzberg, S.L.** (2019). Graph-based genome
922 alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol*. **37**, 907-915.
923 **Kliman, H.J., Nestler, J.E., Sermasi, E., Sanger, J.M.and Strauss, J.F., 3rd** (1986). Purification,
924 characterization, and in vitro differentiation of cytotrophoblasts from human term placentae.
925 *Endocrinology*. **118**, 1567-1582.
926 **Kohama, K., Kawaguchi, M., Fukushima, T., Lin, C.Y.and Kataoka, H.** (2012). Regulation of
927 pericellular proteolysis by hepatocyte growth factor activator inhibitor type 1 (HAI-1) in
928 trophoblast cells. *Hum Cell*. **25**, 100-110.
929 **Kubaczka, C., Senner, C.E., Cierlitza, M., Arauzo-Bravo, M.J., Kuckenberg, P., Peitz, M.,**
930 **Hemberger, M.and Schorle, H.** (2015). Direct Induction of Trophoblast Stem Cells from
931 Murine Fibroblasts. *Cell Stem Cell*. **17**, 557-568.
932 **Kuckenberg, P., Buhl, S., Woynecki, T., van Furden, B., Tolkunova, E., Seiffe, F., Moser, M.,**
933 **Tomilin, A., Winterhager, E.and Schorle, H.** (2010). The transcription factor TCFAP2C/AP-
934 2gamma cooperates with CDX2 to maintain trophectoderm formation. *Mol Cell Biol*. **30**, 3310-
935 3320.
936 **Law, C.W., Alhamdoosh, M., Su, S., Smyth, G.K.and Ritchie, M.E.** (2016). RNA-seq analysis is
937 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 Sturmey, 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., Grosveld, 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 Knofler,**
961 **M.** (2014). Wnt-dependent T-cell factor-4 controls human etravillous trophoblast motility.
962 *Endocrinology*. **155**, 1908-1920.
963 **Meistermann, D., Bruneau, A., Loubersac, S., Reignier, 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 **Mochiduki, 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.

985 **Okae, H., Toh, H., Sato, T., Hiura, H., Takahashi, S., Shirane, K., Kabayama, Y., Suyama, M.,**
986 **Sasaki, H.and Arima, T.** (2018). Derivation of Human Trophoblast Stem Cells. *Cell Stem Cell*.
987 **22**, 50-63 e56.
988 **Pastor, W.A., Liu, W., Chen, D., Ho, J., Kim, R., Hunt, T.J., Lukianchikov, A., Liu, X., Polo, J.M.,**
989 **Jacobsen, S.E., et al.** (2018). TFAP2C regulates transcription in human naive pluripotency by
990 opening enhancers. *Nat Cell Biol*. **20**, 553-564.
991 **Pelekanos, R.A., Sardesai, V.S., Futrega, K., Lott, W.B., Kuhn, M.and Doran, M.R.** (2016).
992 Isolation and Expansion of Mesenchymal Stem/Stromal Cells Derived from Human Placenta
993 Tissue. *J Vis Exp*.
994 **Pijnenborg, R., Dixon, G., Robertson, W.B.and Brosens, I.** (1980). Trophoblastic invasion of
995 human decidua from 8 to 18 weeks of pregnancy. *Placenta*. **1**, 3-19.
996 **Pirooznia, M., Nagarajan, V.and Deng, Y.** (2007). GeneVenn - A web application for
997 comparing gene lists using Venn diagrams. *Bioinformation*. **1**, 420-422.
998 **Ralston, A., Cox, B.J., Nishioka, N., Sasaki, H., Chea, E., Rugg-Gunn, P., Guo, G., Robson, P.,**
999 **Draper, J.S.and Rossant, J.** (2010). Gata3 regulates trophoblast development downstream of
1000 Tead4 and in parallel to Cdx2. *Development*. **137**, 395-403.
1001 **Richardson, B.D., Cheng, Y.H., Langland, R.A.and Handwerger, S.** (2001). Differential
1002 expression of AP-2gamma and AP-2alpha during human trophoblast differentiation. *Life Sci*.
1003 **69**, 2157-2165.
1004 **Russ, A.P., Wattler, S., Colledge, W.H., Aparicio, S.A., Carlton, M.B., Pearce, J.J., Barton, S.C.,**
1005 **Surani, M.A., Ryan, K., Nehls, M.C., et al.** (2000). Eomesodermin is required for mouse
1006 trophoblast development and mesoderm formation. *Nature*. **404**, 95-99.
1007 **Soncin, F., Khater, M., To, C., Pizzo, D., Farah, O., Wakeland, A., Arul Nambi Rajan, K.,**
1008 **Nelson, K.K., Chang, C.W., Moretto-Zita, M., et al.** (2018). Comparative analysis of mouse
1009 and human placentae across gestation reveals species-specific regulators of placental
1010 development. *Development*. **145**.
1011 **Strumpf, D., Mao, C.A., Yamanaka, Y., Ralston, A., Chawengsaksophak, K., Beck, F.and**
1012 **Rossant, J.** (2005). Cdx2 is required for correct cell fate specification and differentiation of
1013 trophectoderm in the mouse blastocyst. *Development*. **132**, 2093-2102.
1014 **Takashima, Y., Guo, G., Loos, R., Nichols, J., Ficz, G., Krueger, F., Oxley, D., Santos, F., Clarke,**
1015 **J., Mansfield, W., et al.** (2015). Resetting Transcription Factor Control Circuitry toward
1016 Ground-State Pluripotency in Human. *Cell*. **162**, 452-453.
1017 **Tanaka, S., Kunath, T., Hadjantonakis, A.K., Nagy, A.and Rossant, J.** (1998). Promotion of
1018 trophoblast stem cell proliferation by FGF4. *Science*. **282**, 2072-2075.
1019 **Theunissen, T.W., Powell, B.E., Wang, H., Mitalipova, M., Faddah, D.A., Reddy, J., Fan, Z.P.,**
1020 **Maetzel, D., Ganz, K., Shi, L., et al.** (2014). Systematic Identification of Culture Conditions for
1021 Induction and Maintenance of Naive Human Pluripotency. *Cell Stem Cell*. **15**, 524-526.
1022 **Turco, M.Y., Gardner, L., Kay, R.G., Hamilton, R.S., Prater, M., Hollinshead, M.S.,**
1023 **McWhinnie, A., Esposito, L., Fernando, R., Skelton, H., et al.** (2018). Trophoblast organoids
1024 as a model for maternal-fetal interactions during human placentation. *Nature*. **564**, 263-267.
1025 **Wamaitha, S.E., del Valle, I., Cho, L.T., Wei, Y., Fogarty, N.M., Blakeley, P., Sherwood, R.I.,**
1026 **Ji, H.and Niakan, K.K.** (2015). Gata6 potently initiates reprograming of pluripotent and
1027 differentiated cells to extraembryonic endoderm stem cells. *Genes Dev.* **29**, 1239-1255.
1028 **Wamaitha, S.E., Grybel, K.J., Alanis-Lobato, G., Gerri, C., Ogushi, S., McCarthy, A.,**
1029 **Mahadevaiah, S.K., Healy, L., Lea, R.A., Molina-Arcas, M., et al.** (2020). IGF1-mediated
1030 human embryonic stem cell self-renewal recapitulates the embryonic niche. *Nat Commun*.
1031 **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 1

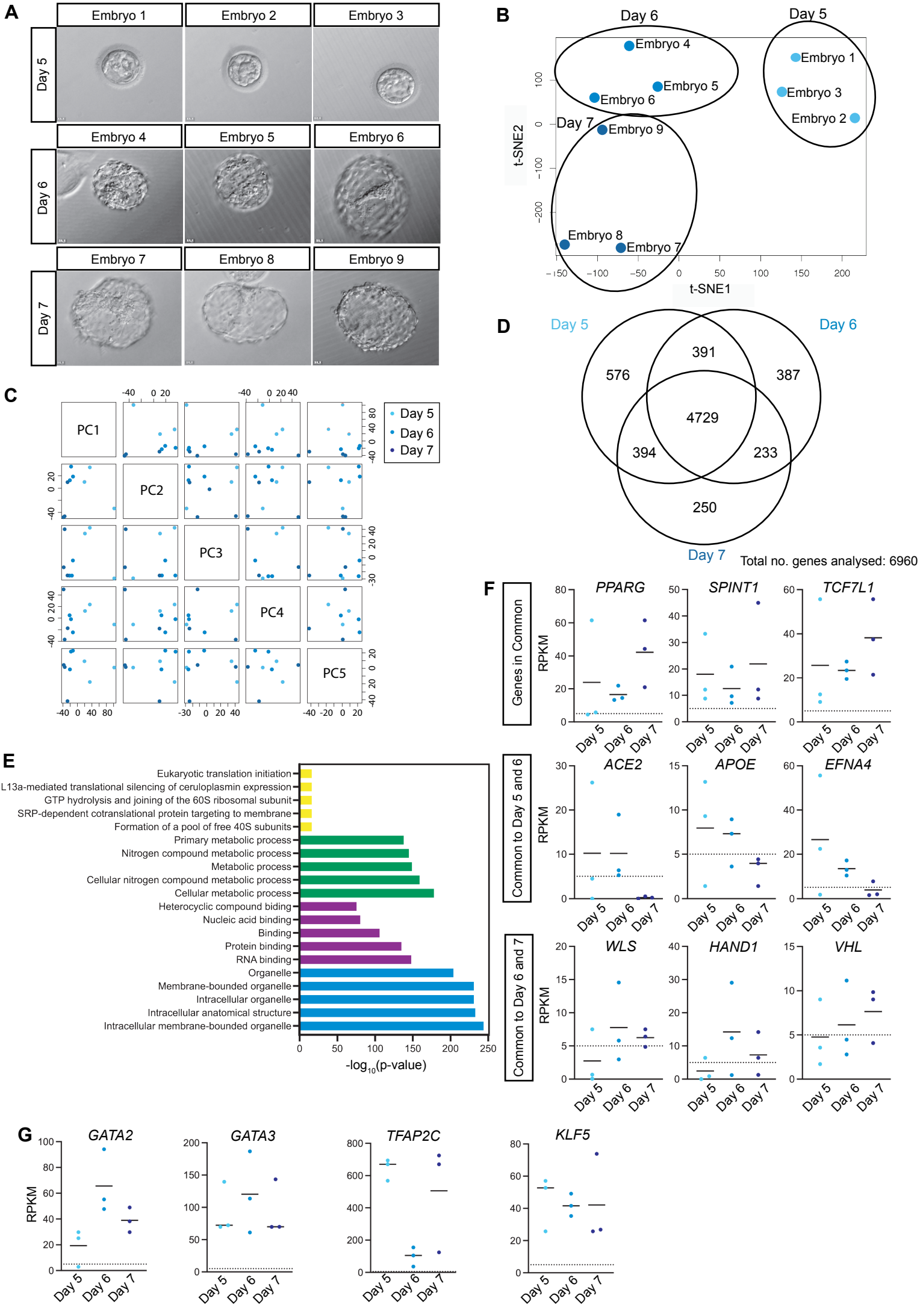

A
Embryo 1
Embryo 2
Embryo 3
Day 5
Embryo 4
Embryo 5
Embryo 6
Day 6
Embryo 7
Embryo 8
Embryo 9
Day 7
B
Day 6
Day 5
Embryo 4
Embryo 5
Embryo 6
Embryo 1
Embryo 3
Embryo 2
Day 7
Embryo 9
Embryo 8
Embryo 7
t-SNE2
t-SNE1
C
PC1
PC2
PC3
PC4
PC5
Day 5
Day 6
Day 7
D
Day 5
Day 6
Day 7
576
391
387
4729
394
233
250
Total no. genes analysed: 6960
E
Eukaryotic translation initiation
L13a-mediated translational silencing of ceruloplasmin expression
GTP hydrolysis and joining of the 60S ribosomal subunit
SRP-dependent cotranslational protein targeting to membrane
Formation of a pool of free 40S subunits
Primary metabolic process
Nitrogen compound metabolic process
Metabolic process
Cellular nitrogen compound metabolic process
Cellular metabolic process
Heterocyclic compound biding
Nucleic acid binding
Binding
Protein binding
RNA binding
Organelle
Membrane-bounded organelle
Intracellular organelle
Intracellular anatomical structure
Intracellular membrane-bounded organelle
-log10(p-value)
F
Genes in Common
PPARG
SPINT1
TCF7L1
Common to Day 5 and 6
ACE2
APOE
EFNA4
Common to Day 6 and 7
WLS
HAND1
VHL
RPKM
Day 5
Day 6
Day 7
G
GATA2
GATA3
TFAP2C
KLF5
RPKM

Figure 2

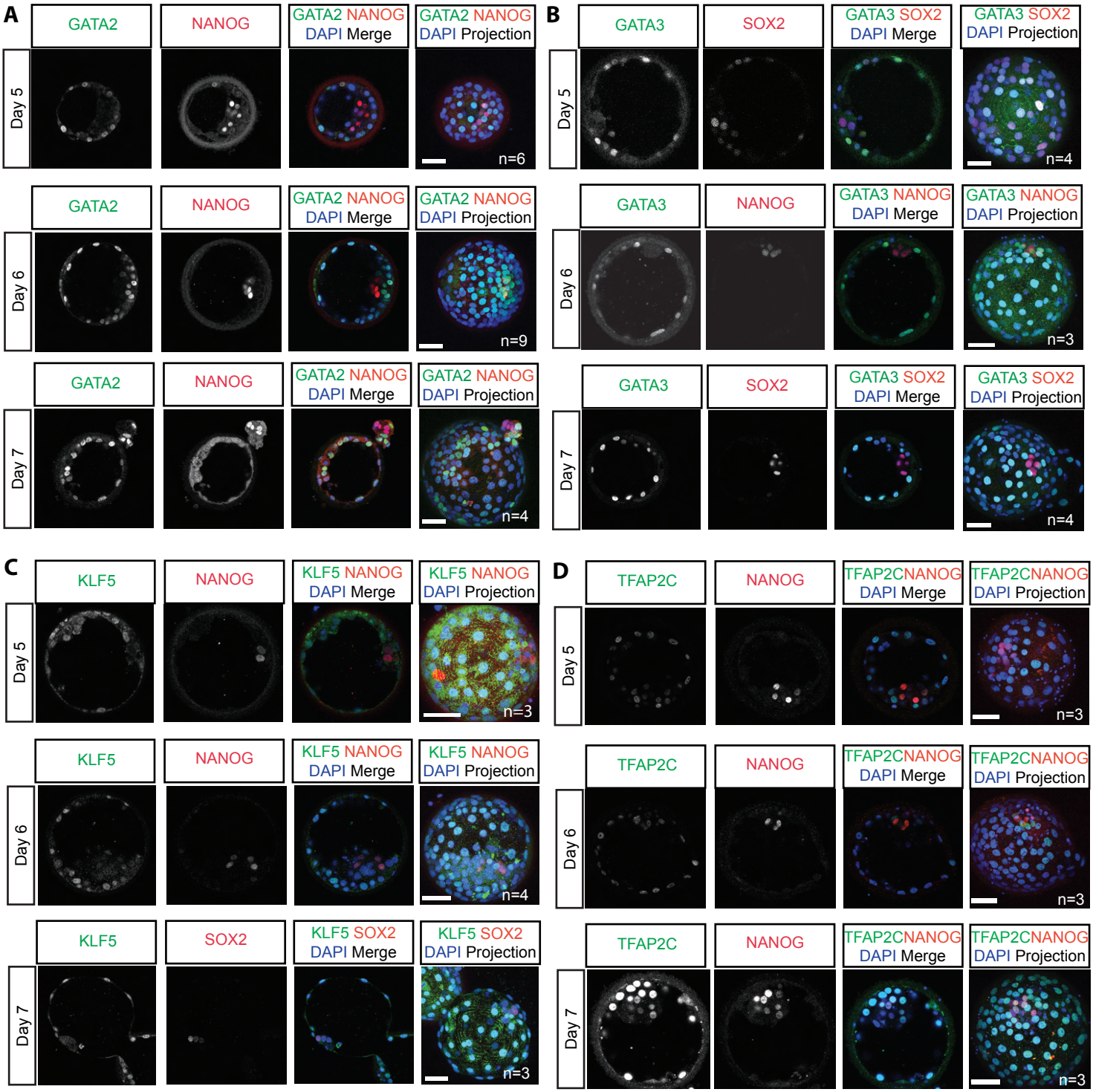

A
GATA2
NANOG
GATA2 NANOG DAPI Merge
GATA2 NANOG DAPI Projection
Day 5
n=6
Day 6
n=9
Day 7
n=4
B
GATA3
SOX2
GATA3 SOX2 DAPI Merge
GATA3 SOX2 DAPI Projection
Day 5
n=4
GATA3 NANOG DAPI Merge
GATA3 NANOG DAPI Projection
Day 6
n=3
Day 7
n=4
C
KLF5
KLF5 NANOG DAPI Merge
KLF5 NANOG DAPI Projection
Day 5
n=3
Day 6
n=4
KLF5 SOX2 DAPI Merge
KLF5 SOX2 DAPI Projection
Day 7
n=3
D
TFAP2C
TFAP2CNANOG DAPI Merge
TFAP2CNANOG DAPI Projection
Day 5
n=3
Day 6
n=3
Day 7
n=3

Figure 3

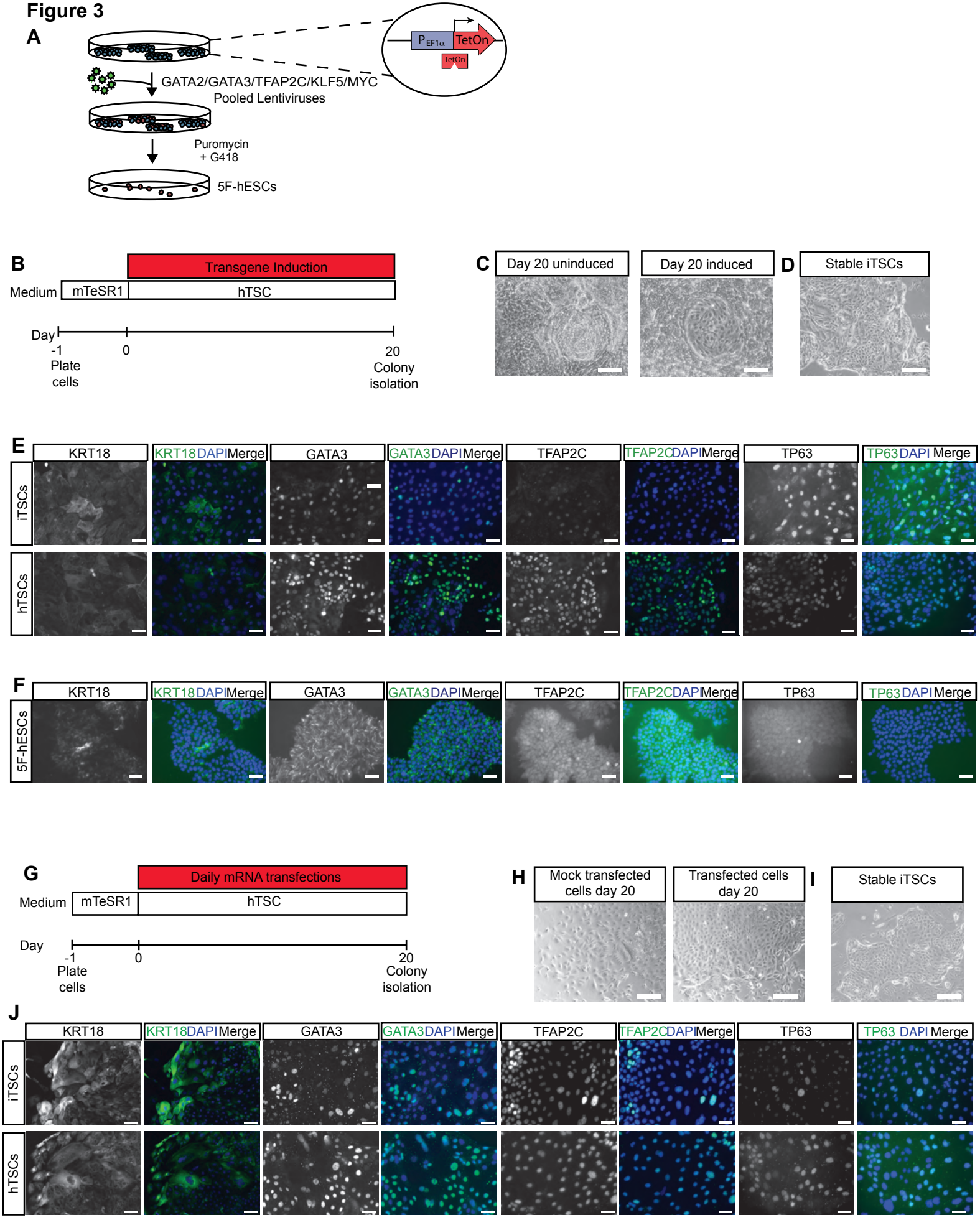
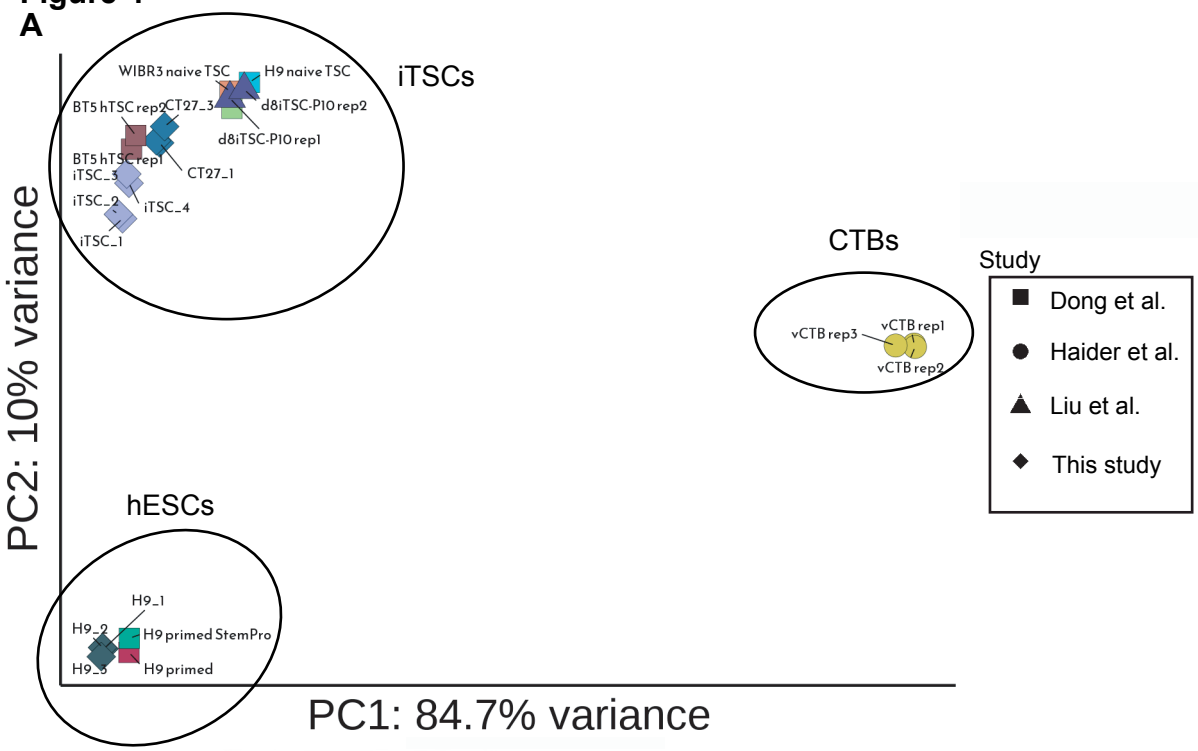
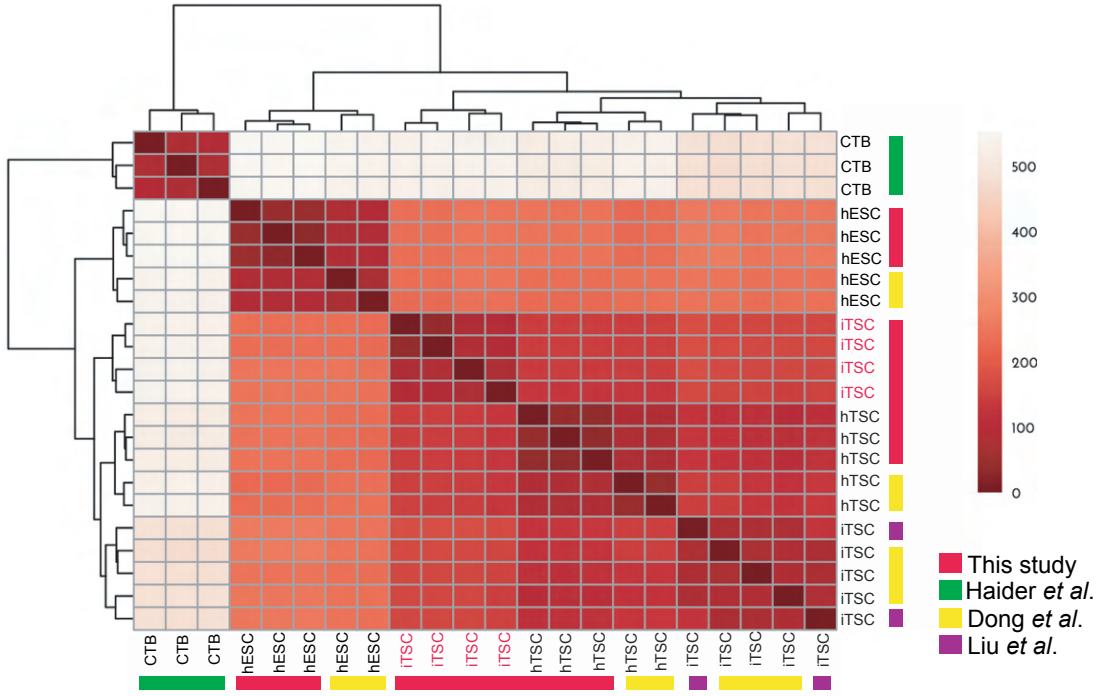

A
GATA2/GATA3/TFAP2C/KLF5/MYC
Pooled Lentiviruses
Puromycin
+ G418
5F-hESCs
P EF1α
TetOn
TetOn
B
Transgene Induction
Medium
mTeSR1
hTSC
Day
-1
Plate
cells
0
20
Colony
isolation
C
Day 20 uninduced
Day 20 induced
D
Stable iTSCs
E
KRT18
KRT18 DAPI Merge
GATA3
GATA3 DAPI Merge
TFAP2C
TFAP2C DAPI Merge
TP63
TP63 DAPI Merge
iTSCs
hTSCs
F
KRT18
KRT18 DAPI Merge
GATA3
GATA3 DAPI Merge
TFAP2C
TFAP2C DAPI Merge
TP63
TP63 DAPI Merge
5F-hESCs
G
Daily mRNA transfections
Medium
mTeSR1
hTSC
Day
-1
Plate
cells
0
20
Colony
isolation
H
Mock transfected
cells day 20
Transfected cells
day 20
I
Stable iTSCs
J
KRT18
KRT18 DAPI Merge
GATA3
GATA3 DAPI Merge
TFAP2C
TFAP2C DAPI Merge
TP63
TP63 DAPI Merge
iTSCs
hTSCs

Figure 4

A


iTSCs
WIBR3 naive TSC
H9 naive TSC
BT5 hTSC rep2
CT27_3
d8iTSC-P10 rep2
d8iTSC-P10 rep1
BT5 hTSC rep1
CT27_1
iTSC_3
iTSC_2
iTSC_4
iTSC_1
CTBs
vCTB rep3
vCTB rep1
vCTB rep2
Study
Dong et al.
Haider et al.
Liu et al.
This study
hESCs
H9_1
H9_2
H9 primed StemPro
H9_3
H9 primed
PC2: 10% variance
PC1: 84.7% variance

B


CTB
hESC
iTSC
hTSC
500
400
300
200
100
0
This study
Haider et al.
Dong et al.
Liu et al.

C

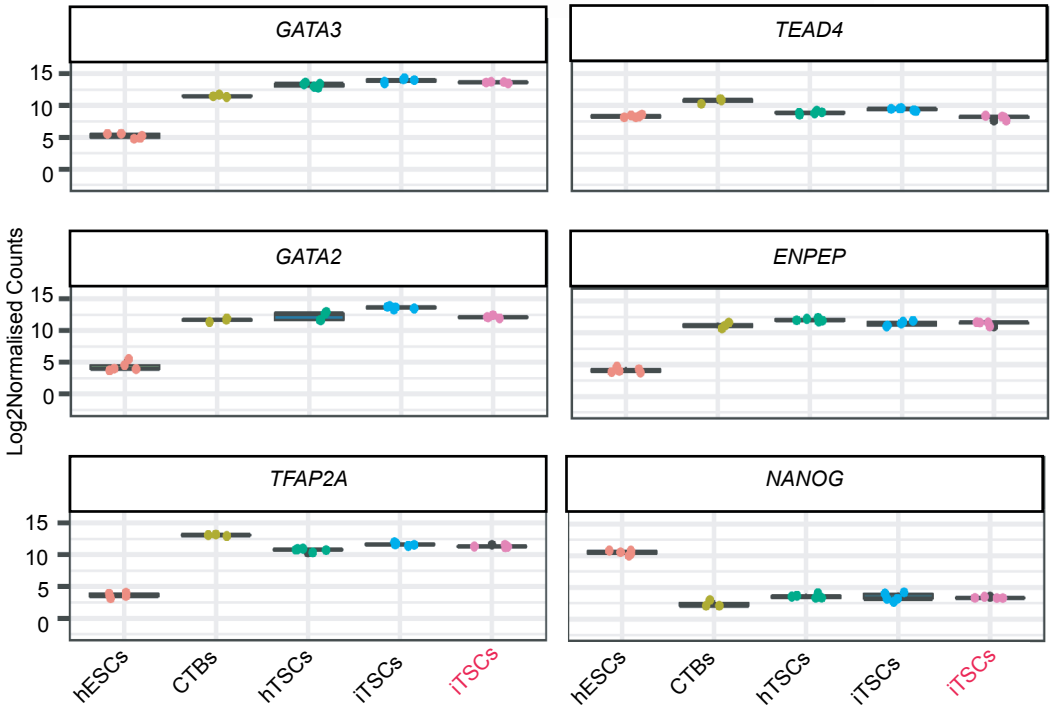

GATA3
TEAD4
GATA2
ENPEP
TFAP2A
NANOG
Log2Normalised Counts
15
10
5
0
hESCs
CTBs
hTSCs
iTSCs

Figure 5

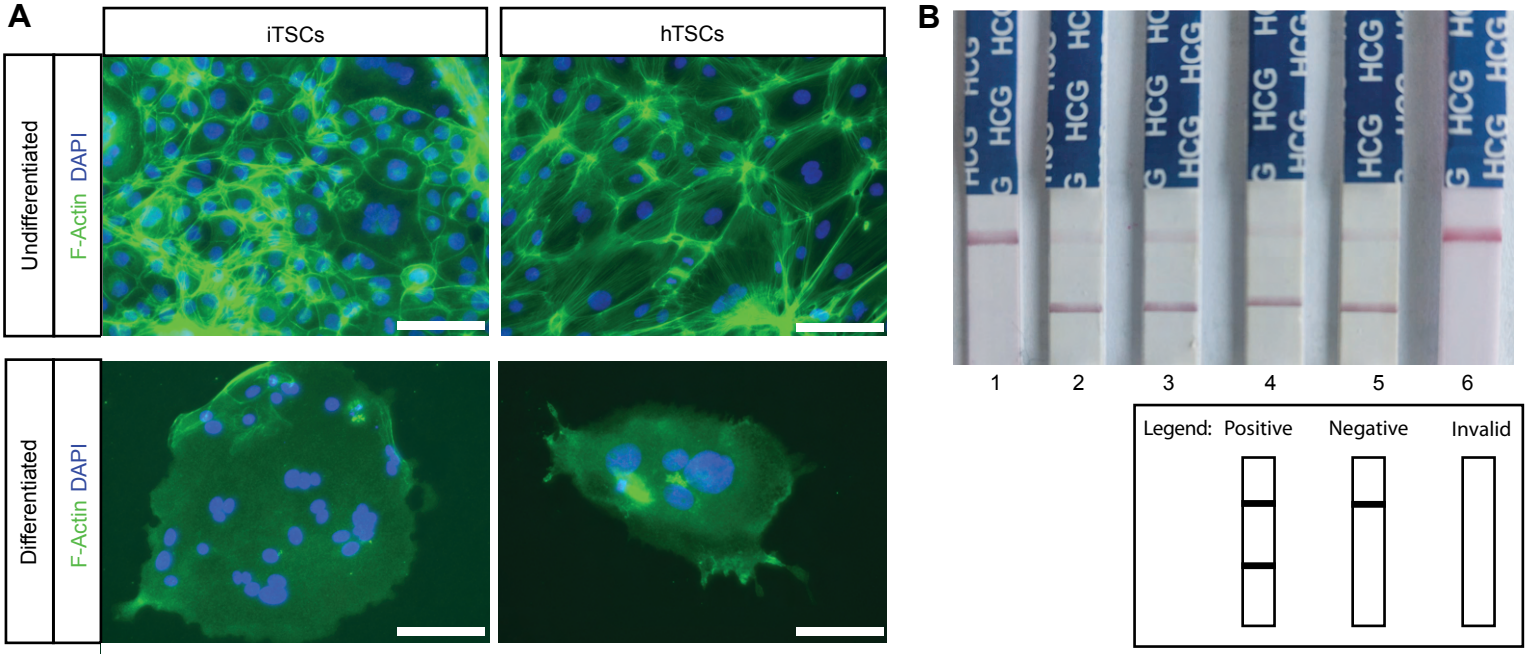
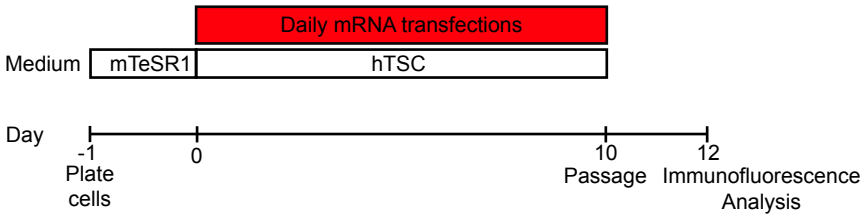
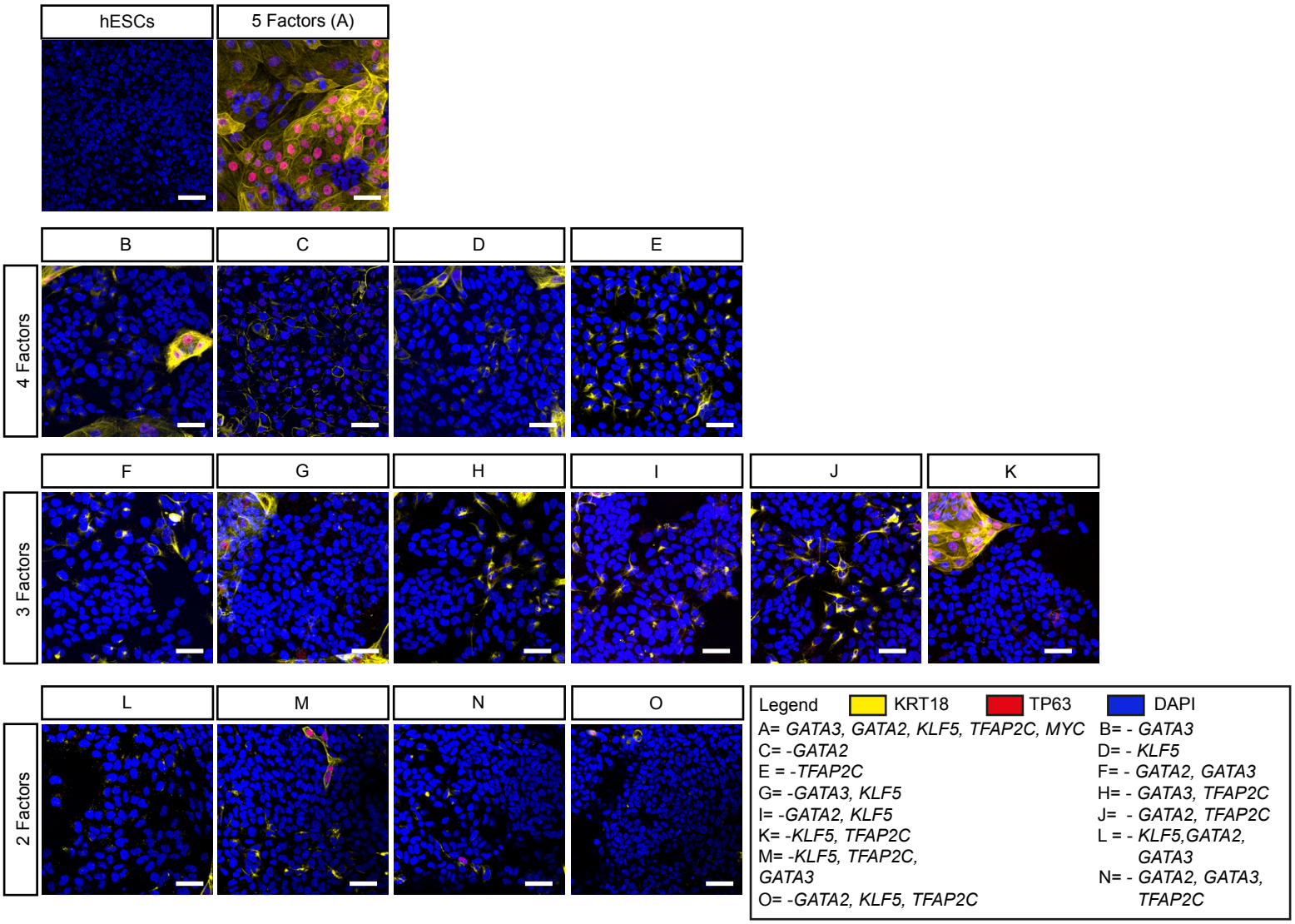

A
iTSCs
hTSCs
Undifferentiated
F-Actin DAPI
Differentiated
F-Actin DAPI
B
1
2
3
4
5
6
Legend: Positive
Negative
Invalid

Figure 6

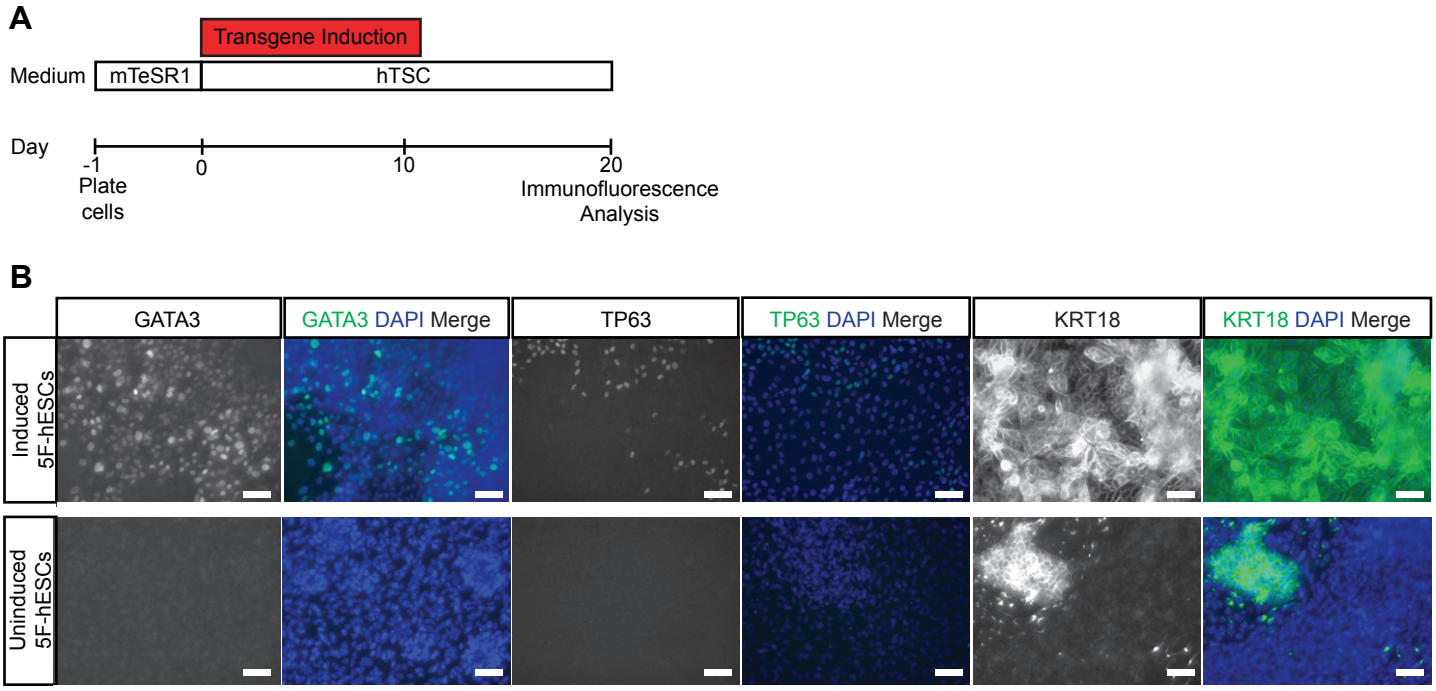
A


Daily mRNA transfections
Medium
mTeSR1
hTSC
Day
-1
Plate cells
0
10
Passage
12
Immunofluorescence Analysis

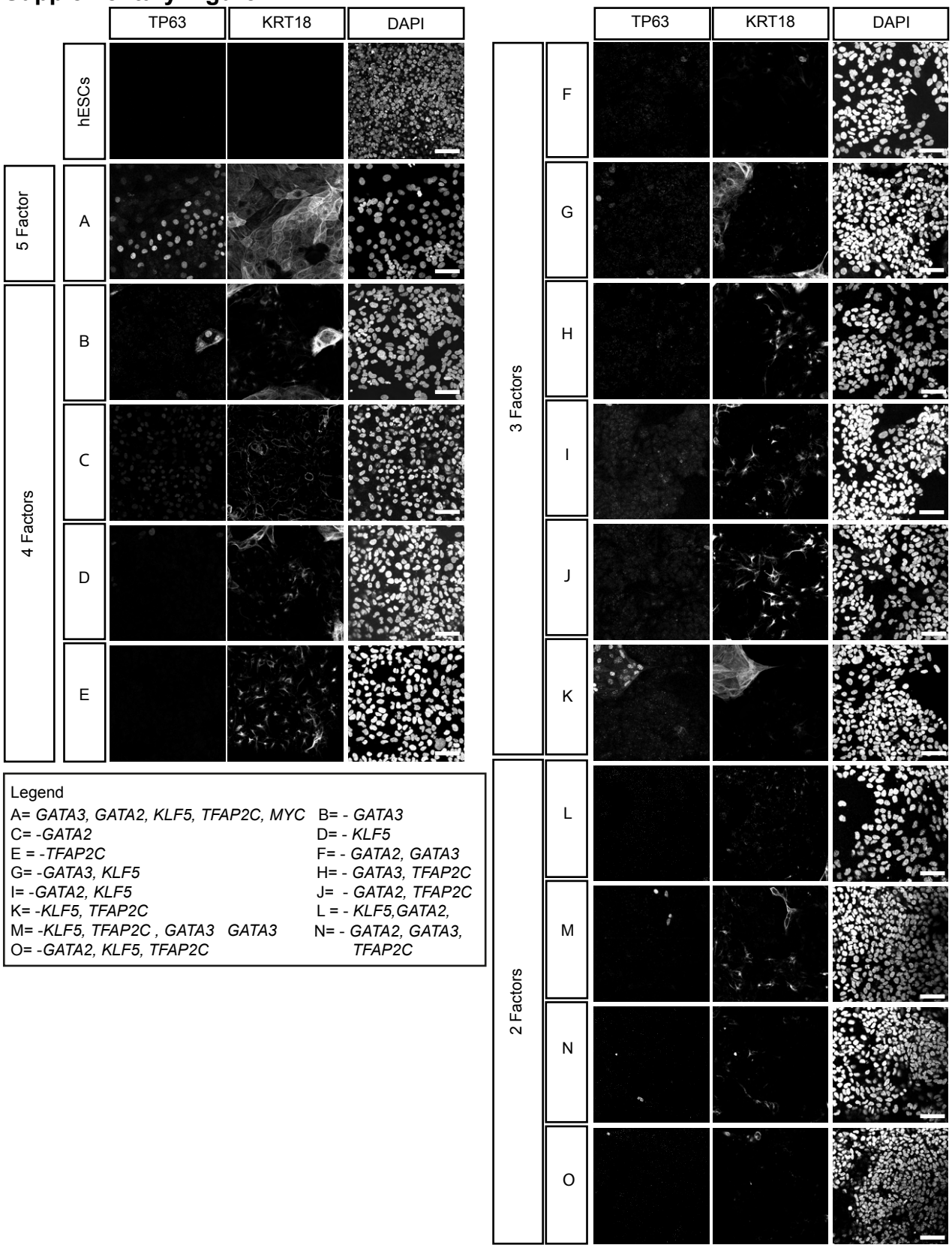
B


hESCs
5 Factors (A)
B
C
D
E
4 Factors
F
G
H
I
J
K
3 Factors
L
M
N
O
2 Factors
Legend
KRT18
TP63
DAPI
A= GATA3, GATA2, KLF5, TFAP2C, MYC
B= - GATA3
C= -GATA2
D= - KLF5
E = -TFAP2C
F= - GATA2, GATA3
G= -GATA3, KLF5
H= - GATA3, TFAP2C
I= -GATA2, KLF5
J= - GATA2, TFAP2C
K= -KLF5, TFAP2C
L = - KLF5,GATA2, GATA3
M= -KLF5, TFAP2C, GATA3
N= - GATA2, GATA3, TFAP2C
O= -GATA2, KLF5, TFAP2C

Supplementary Figure 1



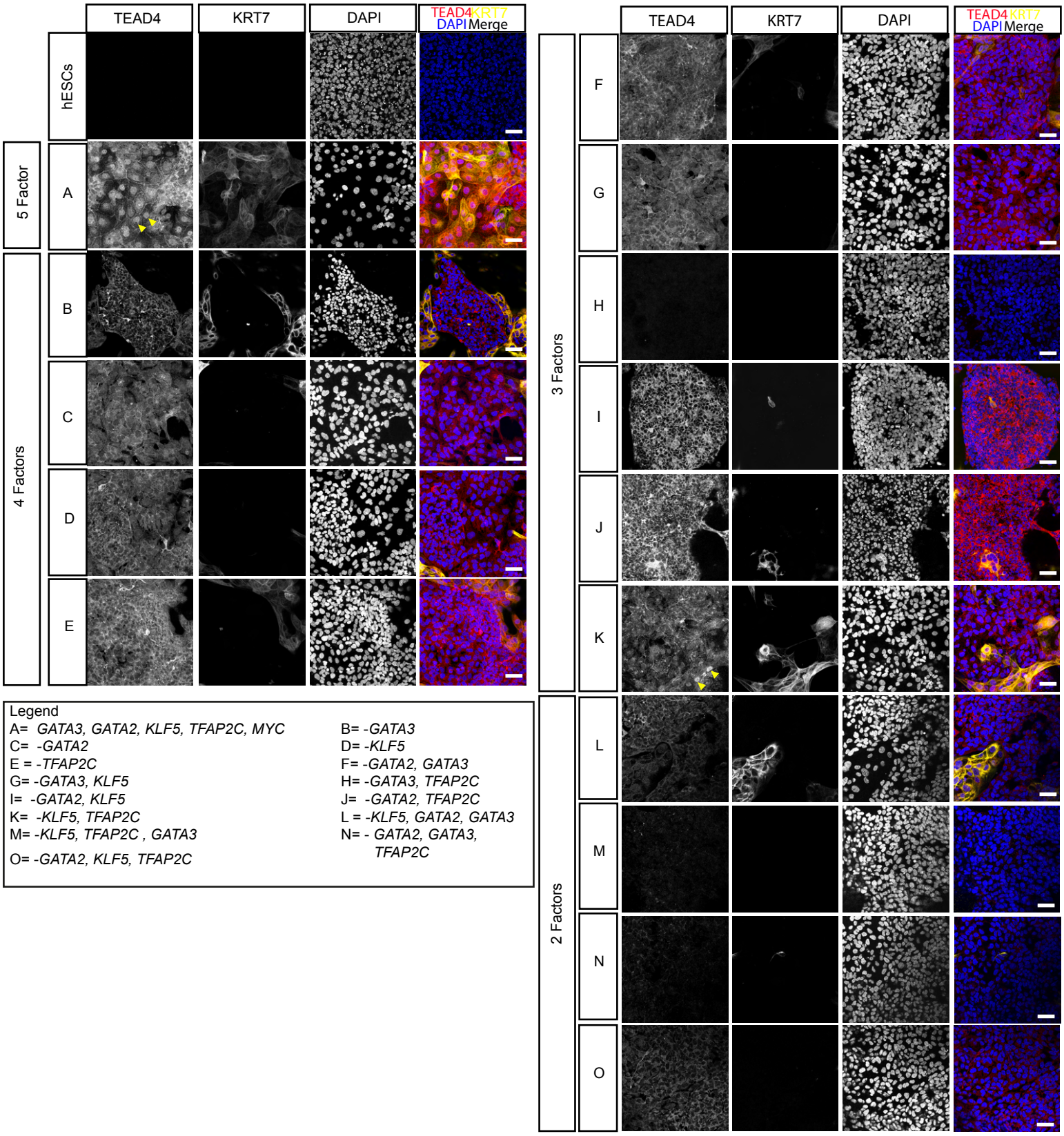
Supplementary Figure 2



Legend

A= *GATA3, GATA2, KLF5, TFAP2C, MYC*
B= - *GATA3*
C= -*GATA2*
D= - *KLF5*
E = -*TFAP2C*
F= - *GATA2, GATA3*
G= -*GATA3, KLF5*
H= - *GATA3, TFAP2C*
I= -*GATA2, KLF5*
J= - *GATA2, TFAP2C*
K= -*KLF5, TFAP2C*
L = - *KLF5,GATA2,*
M= -*KLF5, TFAP2C , GATA3 GATA3*
N= - *GATA2, GATA3, TFAP2C*
O= -*GATA2, KLF5, TFAP2C*

Supplementary Figure 3



Legend

A= *GATA3, GATA2, KLF5, TFAP2C, MYC*
B= *-GATA3*
C= *-GATA2*
D= *-KLF5*
E = *-TFAP2C*
F= *-GATA2, GATA3*
G= *-GATA3, KLF5*
H= *-GATA3, TFAP2C*
I= *-GATA2, KLF5*
J= *-GATA2, TFAP2C*
K= *-KLF5, TFAP2C*
L = *-KLF5, GATA2, GATA3*
M= *-KLF5, TFAP2C , GATA3*
N= *- GATA2, GATA3, TFAP2C*
O= *-GATA2, KLF5, TFAP2C*