VGLL1 contributes to both the transcriptome and epigenome of the developing trophoblast compartment

Ruben I. Calderon[1,2,3], Nirvay Sah[1,3], Molly Huang[4], Ryan H. Kittle[1,7], Walee B. Shaik[1,3], Jennifer N. Chousal[1], Sampada Kallol[1,3], Tony Bui[1], Robert Morey[1,3], Alejandra Mitre[1], Norah M.E. Fogarty[5,8], Claudia Gerri[5,9,10], Claire Zheng[1,2], Peter DeHoff [3,4], Pratik Home[6,11], Kathy K. Niakan[5,12,13,14], Heidi Cook-Andersen[2,3,4], Kathleen M. Fisch[4], Soumen Paul[6], Francesca Soncin[1,3,*]

Affiliations:

1. Department of Pathology, School of Medicine, University of California San Diego, La Jolla, CA 92093, USA

2. School of Biological Sciences, University of California San Diego, La Jolla, CA 92093, USA

3. Sanford Consortium for Regenerative Medicine, La Jolla, CA, 92093, USA

4. Department of Obstetrics, Gynecology, and Reproductive Sciences, School of Medicine, University of California San Diego, La Jolla, CA, 92093, USA.

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

6. Department of Pathology & Laboratory Medicine and Institute for Reproductive and Developmental Sciences, University of Kansa Medical Center, Kansas City, KS 66160.

Current addresses:

7. Department of Genetics and Development, Columbia University Irving Medical Center, New York, NY 10032, USA.

8. Centre for Gene Therapy and Regenerative Medicine, King's College London, London, SE1 9RT, UK.

9. Max Planck Institute of Molecular Cell Biology and Genetics, Pfotenhauerstrasse 108, 01307, Dresden, Germany.

10. Center for Systems Biology Dresden, Pfotenhauerstrasse 108, 01307, Dresden, Germany.

11. BioIVT, 1101 W Cambridge Circle Dr. Kansas City, KS 66103.

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

13. Wellcome Trust – Medical Research Council Stem Cell Institute, University of Cambridge, Jeffrey Cheah Biomedical Centre, Puddicombe Way, Cambridge CB2 0AW, UK.

14. Epigenetics Programme, Babraham Institute, Cambridge CB22 3AT, UK.

15. Lead contributor * Corresponding author: fsoncin@health.ucsd.edu

Abstract

The trophectoderm (TE), the first lineage specified during mammalian development, initiates implantation and gives rise to placental trophoblasts. While animal models have elucidated key conserved signaling pathways involved in early TE specification, including BMP, WNT, and HIPPO, species-specific differences during early development emphasize the need for human-specific models. We previously identified VGLL1, a coactivator of TEAD transcription factors, as a human-specific placental marker. In this study, we employed a pluripotent stem cell (PSC)-based model of TE induction by BMP4 to investigate chromatin remodeling and transcriptional dynamics during TE specification. BMP4-induced chromatin accessibility changes promoted a trophoblast gene expression program, while mesoderm lineage markers were only transiently expressed upon canonical WNT activation. We found that VGLL1 was expressed downstream of key TE transcription factors (GATA2/3, TFAP2A/C) but was essential for full trophoblast differentiation by up-regulating EGFR and reinforcing GATA3 expression through positive feedback. Notably, VGLL1 enhanced canonical WNT signaling via direct regulation of WNT receptors and effectors. We also identified KDM6B, a histone demethylase that removes H3K27me3 repressive marks, as a direct VGLL1 target. KDM6B facilitated activation of bivalent promoters associated with TE markers, linking epigenetic regulation to lineage specification. Our findings establish a mechanistic framework positioning VGLL1 as a central regulator that integrates HIPPO, BMP, and WNT signaling pathways to drive human TE and trophoblast development.

Statement of Significance

This research explores human-specific mechanisms for the induction of the trophectoderm (TE) during early development, which is crucial for placental function and pregnancy establishment. Using a pluripotent stem cell system, we showed that BMP4 treatment induced TE specification by a combination of TE-specific gene expression activation and chromatin changes. We showed that the transcription co-factor VGLL1 played a critical role in directing cells toward a TE fate via three mechanisms: (1) regulating the expression of TE-specific genes; (2) enhancing activation of canonical WNT pathway; and (3) activating the expression of KDM6B, which acts on the chromatin to promote TE-specific gene accessibility.

Introduction

The trophectoderm (TE) is the first lineage specified during mammalian development and is important for embryo cavitation, implantation, and pregnancy initiation (1). After implantation, the TE gives rise to the trophoblast compartment of the placenta, contributing to the formation of the maternal/fetal interface (2). Early pregnancy losses and several placental-associated disorders are linked to abnormalities in the establishment and differentiation of the TE compartment during early development (2). However, ethical and technical considerations have limited our understanding of the mechanisms of TE specification and trophoblast development in humans. Studies in animal models have determined that tight spatio-temporal regulation of highly conserved signaling pathways, including BMP, WNT, and HIPPO, are required to establish the TE in the pre-implantation embryos and for early placental development (3–8). However, increasing evidence points to species-specific differences in the composition of regulatory networks that contribute to the development of the extra-embryonic ectoderm compartment (3, 9–11). Therefore, species-specific models are better suited to investigate TE specification and early placental development (3, 9, 12).

The HIPPO signaling pathway is a conserved cascade that initiates the specification of the TE lineage following activation of GATA3 and CDX2 expression by the TEAD4/YAP1 complex (10, 13, 14). In early mouse embryos, differential HIPPO pathway activation between the inner and outer cells leads to the expression of CDX2, which initiates a TE expression program, including induction of GATA3 and ELF5 (13, 15, 16). EOMES is expressed down-stream of CDX2 and, while not required for TE specification, it is fundamental for the maintenance of the TE transcriptional program and maintenance of mouse trophoblast stem cells (TSCs) (17, 18). In humans, four transcription factors, GATA2, GATA3, TFAP2A, and TFAP2C, form a transcriptional network required for TE initiation (TEtra factors) downstream of the HIPPO signaling pathway (19). This network drives the establishment of an extra-embryonic ectoderm signature, which includes mature markers such as VGLL1, TP63, and EGFR, but not EOMES, which is dispensable for human TE and trophoblast development (9, 20). In particular, we have previously identified VGLL1 as a human-specific marker

93 expressed in the trophoblast progenitor compartments of the placenta (9). VGLL1 is a
94 conserved binding partner of the TEAD family of transcription factors and competes with
95 YAP1 to form a transcriptional complex with TEAD4, therefore potentially modulating HIPPO
96 signaling in human placenta (21).

97 Canonical WNT signaling is a well-characterized developmental pathway mediated
98 by nuclear localization of β-catenin, where it binds to TCF/LEF complexes and drives
99 expression of WNT-specific target genes (22). This pathway is required for maintenance and
100 expansion of various human epithelial stem cells, including intestinal and skin stem cells
101 (23, 24). In human pre-implantation embryos, the TE compartment is enriched in WNT
102 components, including ligands and receptors (25–29). Moreover, compared to mouse, both
103 human and non-human primate embryos show higher canonical WNT-related gene
104 expression in the TE (30). WNT genes are also highly expressed in the trophoblast
105 compartment of the placenta. *In vitro*, both human TSCs and trophoblast organoids require
106 activation of canonical WNT signaling for proliferation (31–33) while its transient inhibition is
107 required to initiate differentiation of extravillous trophoblasts (EVT) (32, 34).

108 Bone Morphogenetic Protein (BMP) signaling has well described roles in gastrulation,
109 including specification of both the embryonic mesoderm lineage and primordial germ cells
110 (5). However, recent evidence demonstrated BMP activation in the TE of human, but not
111 murine, blastocysts (5, 12, 35), though autocrine BMP signaling has been reported in both
112 mouse and human TSCs (36). *In vitro*, exogenous BMP4 has been shown to drive
113 differentiation of human pluripotent stem cells (PSCs) into different lineages, namely
114 embryonic mesoderm, primordial germ cells, or extra-embryonic trophoblast lineage
115 progenitors, based on the cellular context and the activation of other signaling pathways (5).
116 In particular, embryonic mesoderm differentiation requires BMP-induced WNT activation,
117 while inhibition of β-catenin-dependent WNT signaling downstream of BMP effectively
118 blocks mesoderm differentiation, instead promoting trophoblast lineage (37). However, the
119 mechanisms underlying WNT activation by BMP4 and the switch from mesoderm to
120 trophoblast lineage in PSC-derived TE-like cells have not been elucidated yet.

121 Regulatory networks drive establishment and maintenance of cellular identity by
122 acting on a lineage-specific epigenetic landscape downstream of signaling pathways (38).
123 DNA methylation and histone marks surrounding lineage-specific genes determine
124 chromatin accessibility for specific transcription factor (TF) binding (38). Recent studies in
125 pre-implantation embryos have described epigenome changes across early development
126 from zygote to blastocyst, including segregation of TE and ICM (39–41). However, there is
127 little mechanistic understanding on how epigenetic marks are laid out in a lineage-specific
128 manner during the specification of the TE and in trophoblast development. Moreover, it is
129 largely unknown how multiple signaling pathways like HIPPO, WNT, and BMP converge
130 towards a lineage-specific outcome and drive trophoblast development.

131 PSCs represent a unique resource to study species-specific genetic and epigenetic
132 mechanisms of TE and trophoblast specification as they are tractable models for genetic
133 and chemical manipulation. Multiple protocols have been published to study TE and
134 trophoblast specification from human PSCs (42–46). We have established a 2-step protocol
135 based on BMP4 treatment of PSCs to first differentiate them into TE-like cells, followed by
136 either further differentiation into mature trophoblasts or derivation of *bona fide* hTSCs (47,
137 48). Based on literature, we further optimized the first step by adding a WNT inhibitor, IWP2,
138 to block concurrent activation of a mesoderm program and obtain a more homogeneous
139 population of TE-like cells (37, 45). We previously reported that VGLL1 is one of the most
140 highly upregulated genes in the BMP4-directed differentiation of human embryonic stem
141 cells (hESCs) into TE-like cells (9). Here, we use this model to first investigate the BMP4-
142 dependent chromatin modifications that occur during the transition from PSC to TE-like cells
143 in relation to gene expression changes and WNT activation status. Then, we present our
144 findings on how VGLL1 directly and indirectly drives a trophoblast-specific gene expression
145 and epigenomic program in this PSC-derived trophoblast model.

146 Results

147 BMP4 mediates chromatin accessibility to stabilize a trophoblast-specific
148 expression program

149 We have previously optimized a protocol for differentiation of PSCs into TE-like cells
150 by treatment with 10ng/mL BMP4 and the WNT inhibitor IWP2 over 4 days, which
151 successfully blocks mesoderm differentiation and promotes a uniform EGFR+ population
152 (45). To start dissecting the transcriptional changes towards TE-like cells during this
153 differentiation protocol, we first performed bulk RNA-seq analysis of H9 hESCs at day 0, day
154 2, day 4 following treatment with BMP4 or BMP4/IWP2 (**Suppl. Fig. S1A**) and identified
155 differentially expressed genes for each timepoint and treatment compared to d0 (**Suppl.**
156 **Table S1**). Genes downregulated in both treatments included pluripotency and embryonic
157 lineage markers, i.e. *NANOG*, *POU5F1*, and *SOX2* (**Suppl. Table S1**). Upregulated genes
158 common across timepoints and treatments included the TEtra factors *GATA2*, *GATA3*,
159 *TFAP2A*, and *TFAP2C*. By day 4, common upregulated genes included *ENPEP*, and
160 trophoblast markers *VGLL1* and *TP63 (***Suppl. Fig. S1B and Suppl. Table S1**). Genes
161 uniquely upregulated by BMP4 treatment on day 2 were associated with various signaling
162 pathways, including WNT signaling (**Fig. 1A, Suppl. Table S1**). As expected, secondary to
163 WNT signaling activation in hESCs, we detected upregulation of mesoderm markers *EOMES*,
164 *TBXT*, and *TBX6* at day 2 of BMP4 treatment (**Fig. 1B**). However, by day 4, the expression of
165 these markers decreased to levels comparable to the treatment with the addition of IWP2
166 (**Fig. 1B**). Accordingly, the WNT signaling pathway was no longer one of the top significantly
167 enriched pathways at day 4 of BMP4 alone treatment (**Suppl. Fig. S1C**). To confirm the
168 activation timeframe of the WNT pathway during BMP4 treatment, we used a TCF/LEF-eGFP
169 reporter hESC line (HUES9) where detection of GFP fluorescence is dependent on canonical
170 WNT activation (49). We detected WNT signaling activation starting at day 2, a peak at day 3,
171 and decreased activity at day 4 of BMP4 treatment (**Suppl. Fig. S1D**). The WNT inhibitor IWP2
172 completely blocked eGFP expression, while another inhibitor, IWR1, showed only partial
173 suppression of WNT activity at the dose used (**Suppl. Fig. S1D**). Accordingly, we detected a
174 dose-dependent inhibition of both WNT targets, *LEF1* and *WNT5B*, as well as mesoderm
175 markers *TBXT* and *EOMES* between the two inhibitors (**Suppl. Fig. S1E**). Overall, these data
176 suggest that the WNT pathway is transiently activated during treatment of hESCs with low

177 dose BMP4, but that both WNT activity and mesoderm gene expression decrease over time
178 as the TE identity is established in this model.

179 To investigate chromatin accessibility changes associated with the differential
180 establishment of the mesoderm and TE lineage during BMP4 treatment of PSCs, we
181 performed Assay for Transposase-Accessible Chromatin (ATAC) followed by sequencing in
182 hESCs treated with BMP4 or BMP4/IWP2 at days 0, 2, and 4. We found that peaks resided at
183 promoter (31%), intergenic (24%), and intronic (23%) regions (**Suppl. Fig. S1F**). Following
184 differential peak analysis, both at day 2 and day 4 of treatment, we found open chromatin
185 regions specific to BMP4 or BMP4/IWP2, as well as regions common to both treatment
186 protocols (**Fig. 1C, Suppl. Table S2**). We performed Hypergeometric Optimization of Motif
187 EnRichment (HOMER) analysis on common and day/treatment-specific regions to identify
188 TF binding motifs. Commons regions at both days 2 and 4 contained binding motifs for well-
189 known TE and trophoblast TFs, including TEAD and AP1-complex members (i.e. JUN and
190 FOS) (**Suppl. Fig S1G**). Binding motifs of both TEtra factors GATA3 and TFAP2C appeared in
191 both common and treatment-specific regions. At day 2 of BMP4-only treatment, we found
192 open chromatin at LEF-1 and TCF3, TFs downstream of the WNT pathway. We then looked at
193 the chromatin regions around the transcription start sites of well-known trophoblast and
194 mesoderm genes (**Fig.1D and Suppl. Fig. S1H**). The open chromatin region across
195 treatments/days showed a sharp peak and remained consistent around trophoblast markers
196 (**Fig. 1D**), including GATA3 and EGFR (**Suppl. Fig. S1H**). Conversely, open chromatin around
197 mesoderm genes showed a broader distribution that decreased between day 2 and day 4 in
198 response to both treatments (**Fig. 1D**), including at EOMES and TBXT (**Suppl. Fig. S1H**) These
199 chromatin accessibility data are largely consistent with the gene expression patterns
200 observed above.

201 Finally, we looked at the differential expression of known epigenetic modifiers that
202 could account for these changes in chromatin accessibility. We found multiple lineage-
203 specific chromatin modifiers differentially expressed during TE-specification. These
204 included down-regulation of *de novo* DNA methyltransferases *DNMT3A* at day 2, a switch
205 from *TET1* to *TET2* (50) expression, and up-regulation of histone demethylases like *KDM6B*,

206 *KDM4B,* and *KDM1B* (**Fig. 1E**). These data suggest that low levels of BMP4 preferentially
207 enable a trophoblast transcriptional signature via a combination of lineage-specific gene
208 expression activation and finely tuned regulation of the epigenetic landscape.

209 VGLL1 regulates a TE-specific gene expression program

210 In the context of BMP4 treatment of human PSCs, the TEtra factors GATA2, GATA3,
211 TFAP2A, and TFAP2C have been identified as initiators of the TE/trophoblast transcriptional
212 program, upstream of more mature trophoblast markers like TP63, EGFR, ELF5, and
213 VGLL1(9, 19, 20). To further tease apart the timeframe of activation of these TE/trophoblast
214 genes in this BMP4-based model, we investigated the expression of TEtra factors, and
215 mature trophoblast markers by qPCR at short intervals within the first 72 hours of
216 BMP4/IWP2 treatment. As previously reported, TEtra factors *GATA3* and *TFAP2C* were the
217 first to be activated, with upregulation as early as 6h after the start of the treatment and
218 continuous increase over time (**Fig. 2A**). *VGLL1* expression started to rise between 12 hours
219 and 24 hours of BMP4/IWP2 treatment (**Fig. 2A**) while other mature TE/trophoblast markers
220 were observed at 48 hours (*TP63*) and 72 hours (*ELF5* and *EGFR*), respectively (**Fig. 2A**). By
221 immunofluorescence, VGLL1 protein was detected in the nucleus after 24 hours of
222 BMP4/IWP2 treatment, where it co-localized with TEAD4 and YAP1 throughout TE induction
223 (**Suppl. Fig. S2A**). While previous literature considered VGLL1 a mature trophoblast marker
224 (like TP63), these data suggest that VGLL1 might play an earlier role downstream of TEtra
225 factors but before other trophoblast markers.

226 To understand the role of VGLL1 in this context, we established hESC lines that
227 constitutively express VGLL1-specific shRNA for targeted gene knock-down (or shScramble
228 as control, hereafter referred to as Scramble). Compared to Scramble, treatment of shVGLL1
229 hESCs with BMP4/IWP2 led to blunted up-regulation of late trophoblast markers, *TP63*,
230 *ASCL2*, and *GCM1* (**Fig. 2B**). Interestingly, TEtra factor *GATA3* and the early TE marker *CDX2*
231 also showed lower up-regulation. In addition, surface levels of EGFR levels were also lower
232 in the BMP4/IWP2-treated shVGLL1-hESCs, compared to similarly-treated Scramble-hESCs
233 (**Suppl. Fig. S2B**), suggesting an incomplete TE differentiation. Therefore, to further explore

234 the effect of VGLL1 knock-down in TE induction, we performed bulk RNA-seq analysis on two
235 independent shVGLL1 clones and one Scramble control at days 0 and 4 of BMP4/IWP2
236 treatment (**Suppl. Fig. S2C**). Overall, shVGLL1 clones showed a marked delay in
237 differentiation compared to Scramble (**Suppl. Fig. S2C**). We found 7,284 genes that were
238 differentially dysregulated in BMP4/IWP2-treated shVGLL1 hESCs compared to Scramble
239 (adj. p value<0.05), organized in 4 distinct cluster based on pattern of expression (Suppl.
240 **Table S3**). Specifically, 3,056 genes showed blunted down-regulation (cluster #2) compared
241 to Scramble and were associated with transcriptional regulation of pluripotent stem cells
242 and embryo development (**Suppl. Fig. S2D**), including pluripotency markers *POU5F1*, *SOX2*,
243 and *SALL2*. The 4,192 up-regulated genes that showed misregulation in shVGLL1 hESCs
244 compared to Scramble (**Fig. 2C**) organized in 3 clusters of expression. Cluster #3 contained
245 genes that failed to be up-regulated, including TE/trophoblast markers *GATA2*, *NR2F2* (51),
246 and *CCR7* (52), while cluster #4 included genes with blunted expression compared to
247 Scramble, including TEtra factors *GATA3* and *TFAP2C*, TE markers *ENPEP* (53), and
248 trophoblast markers *EGFR*, *ARID3A*, and *ARID5B* (54). Gene ontology analysis confirmed
249 that genes that failed to be fully up-regulated were associated with response to hormones
250 and cell migration and were highly enriched for placental genes (**Suppl. Fig. S2E**).
251 Interestingly, the remaining cluster #1 contained genes that were more highly up-regulated
252 in shVGLL1 cells compared to Scramble and included TFAP2A, a potential compensatory
253 response to lack of VGLL1 and lower expression of other TEtra factors. These data suggest
254 VGLL1 acts through a positive feedback mechanism downstream of the TEtra factors to
255 complete a TE/trophoblast-specific program.

256 We then investigated if VGLL1 was required for TSC derivation. Recently, we
257 demonstrated that hESC-derived TE-like cells (hESCs following 4 days of BMP4/IWP2
258 treatment) can adapt to hTSC culture conditions, resulting in cells that show characteristics
259 of *bona fide* primary hTSCs (48). hTSC culture medium includes a canonical WNT agonist
260 and EGF signaling activation combined with TGFβ and ROCK signaling inhibition. Therefore,
261 we attempted to derive hTSC-like cells from both Scramble and two shVGLL1 hESC clones.
262 Similar to our initial report with wild type PSCs, both Scramble and shVGLL1 hESCs rapidly

263 lost both epithelial morphology and cell surface expression of EGFR during the early stages
264 of adaptation (**Suppl. Fig. S3A-B**). However, while Scramble cells acquired a hTSC-like
265 colony morphology and regained expression of surface EGFR by passage 6, shVGLL1 clones
266 maintained a single cell morphology and failed to regain EGFR (**Fig. 2D-E**). Moreover,
267 shVGLL1 clones expressed lower levels of TEtra factors *GATA2*, *GATA3*, and *TFAP2C*,
268 compared to Scramble hTSCs (**Fig. 2F, Suppl. Fig. S3D**). Interestingly, while the TE marker
269 *ENPEP* also decreased in shVGLL1 cells, *CDX2* increased compared to Scramble (**Suppl.**
270 **Fig. S3C-D**). These data suggest that VGLL1 knock-down PSCs failed to form TSCs and that
271 VGLL1 is required for the maintenance of a trophoblast signature, in particular related to the
272 expression of some TEtra factors and EGFR.

273 We further evaluated VGLL1 activity, uncoupled from BMP4 treatment, to investigate
274 the function of VGLL1 in promoting the TE-specific transcriptional program. We leveraged a
275 system composed of a doxycycline (DOX)-inducible dCas9-VPR and VGLL1-specific guide
276 RNAs for CRISPR-mediated gene activation in undifferentiated hESCs (iVGLL1a) (**Fig. 3A**). In
277 the absence of exogenous BMP4, ectopic expression of VGLL1 in hESCs initiated changes in
278 cell morphology with cells taking on a cobblestone shape, typical of TE/trophoblast cells
279 (**Fig. 3B**), and increased surface expression of EGFR (**Fig. 3C**). Moreover, we observed up-
280 regulation of the TEtra factor *GATA3* and trophoblast marker *EGFR* accompanied by down-
281 regulation of pluripotency marker *NANOG* by qPCR (**Fig. 3D**). We then performed bulk RNA-
282 seq analysis on three independent iVGLL1a clones treated with DOX (or DMSO as carrier
283 control) (**Suppl. Fig. S4A**). We found 6,530 differentially expressed genes (DEGs) in DOX-
284 treated cells compared to DMSO (adj. p-value 0.05), of which 3,053 were up-regulated and
285 3,477 down-regulated (**Fig. 3E, Suppl. Table S4**). Significantly down-regulated genes were
286 enriched in pluripotency genes, including *MYC*, *NANOG*, *POU5F1*, and *SOX2* (**Fig. 3E and**
287 **Suppl. Fig. S4B**). Up-regulated DEGs included TE/trophoblast markers *ARID5B*, *EGFR*, and
288 *GATA3* (**Fig. 3E**). Moreover, gene ontology analysis of pathways and biological processes
289 revealed the activation of signaling pathways important for TE and trophoblast specification
290 and function, including TGFβ, MAPK, and PI3K-Akt pathways (**Suppl. Fig. S4C**). Interestingly,
291 iVGLL1a cells could not be adapted to TSC culture conditions. These data point to a specific

292 role of VGLL1 in controlling a TE/trophoblast transcriptional program but suggest that
293 ectopic expression of VGLL1 alone is insufficient for TSC establishment from hPSCs.

294 VGLL1 regulates canonical WNT pathway effectors

295 Gene ontology analysis of up-regulated genes in response to DOX treatment of
296 iVGLL1a hESCs suggested activation of WNT signaling pathway (**Suppl. Fig. S4C**). Indeed,
297 VGLL1 expression induced up-regulation of WNT pathway components and mediators,
298 including *EPHA2*, *FZD2*, *LEF1*, *LRP4*, *WLS*, *WNT4/5A/5B/6/10A/11*, as well as WNT signaling
299 targets like *FST* (**Fig. 4A and Suppl. Fig S4D**). We also confirmed β-catenin dependent WNT
300 activation using a reporter expressing mCherry under constitutive promoter and GFP under
301 TCF/LEF promoter (55). We observed GFP signal in a subset of mCherry expressing cells
302 following DOX treatment compared to DMSO control (**Fig. 4B**). Interestingly, we observed up-
303 regulation of mesoderm markers on DOX-treated iVGLL1a hESCs by qPCR, including EOMES
304 and TBXT (**Fig. 4C**). During BMP4-induced hESC differentiation, co-treatment with WNT
305 inhibitor IWP2 effectively suppressed mesoderm marker expression. Therefore, we tested
306 the dependence of mesoderm marker expression down-stream of VGLL1 activation in
307 iVGLL1a hESCs on activation of the WNT signaling pathway. Addition of IWP2 during
308 doxycycline induction of VGLL1 effectively reduced the expression of WNT genes *FST*, *LEF1*,
309 and *LGR5* (**Suppl. Fig. S4D**). Moreover, this resulted in reduced expression of mesoderm
310 markers *EOMES* and *TBXT* (**Fig. 4C**) and improved expression levels of *GATA3* and *EGFR* by
311 qPCR and flow cytometry (**Fig. 4D-E**). We compared gene expression profiles of DOX and
312 DOX/IWP2 treated cells (**Fig. 4F, Suppl. Table S4**) and found that DOX/IWP2 treated cells
313 showed significant loss of mesoderm-associated genes, including *EOMES* and *TBXT* (**Suppl.
314 Table S4**). These data support a role for VGLL1 in modulating β-catenin-dependent WNT
315 signaling, an important pathway in trophoblast development and placenta function.

316 VGLL1 directly regulates trophoblast-associated genes at placenta-specific
317 distal regulatory regions

318 We compared the differentially expressed genes in both the constitutive knock-down
319 and gene activation systems to further narrow down VGLL1 contributions to the TE
320 transcriptional program (**Fig. 5A**). We found 528 genes that were concordantly regulated by
321 VGLL1 expression, specifically down-regulated in VGLL1sh hESCs (relative to Scramble)
322 after TE induction, and up-regulated in the iVGLL1a hESCs treated with DOX/IWP2 (**Suppl.**
323 **Table S4**). These genes were involved in cancer pathways, focal adhesion, TGFβ and HIPPO
324 signaling pathways (**Suppl. Fig. S4E**) and were highly associated with placental gene
325 expression (**Suppl. Fig. S5A**). We then set out to investigate whether VGLL1 directly regulates
326 the transcription of a subset of these genes. VGLL1 lacks a DNA binding domain but is a
327 transcriptional co-factor for TEAD family members. TEAD4 is expressed in both hESCs and
328 hESC-derived TE-like cells and plays a role in trophoblast differentiation of hESCs as well as
329 in maintenance of stemness in primary hTSCs (56, 57). Therefore, we predicted that VGLL1
330 binding to regulatory regions might be mediated through its association with TEAD4. To test
331 this, we performed immunoprecipitation with either an anti-TEAD4 or anti-VGLL1 antibody
332 on the choriocarcinoma cell line JEG3 and confirmed co-precipitation of VGLL1 and TEAD4,
333 respectively (**Suppl. Fig. S5B**). We also confirmed binding of VGLL1 and TEAD4 to a validated
334 5'UTR region of *GATA3* by ChIP-qPCR (**Suppl. Fig. S5C**) (58).

335 Having validated the interaction between VGLL1 and TEAD4 in a trophoblast cell line,
336 we next set out to identify binding regions of VGLL1 in complex with TEAD4 in hESC-derived
337 TE-like cells. Therefore, we performed ChIP-seq using both anti-VGLL1 and anti-TEAD4
338 antibodies on BMP4/IWP2-treated hESCs at day 4. VGLL1 ChIP peaks were distributed
339 mainly between intronic (44.4%), intergenic (30.6%), and promoter (20%) regions (**Suppl. Fig**
340 **S5D**), suggesting that VGLL1 might regulate target genes through transcription activation at
341 promoters as well as through interactions at distal enhancer regions. In line with this, we
342 found evidence of VGLL1 occupancy at *EGFR* and *ENPEP* promoter regions (**Fig. 5B**), as well
343 as in distal regulatory regions of *GATA3* and *EGFR* identified as putative active enhancers in
344 the human placenta (59) (**Suppl. Fig. 5E, Suppl. Table S5**). VGLL1 peaks were enriched in
345 TEAD family member motifs, including TEAD4 and TEAD1 (**Fig. 5C**). In our ChIP-seq data,
346 VGLL1 and TEAD4 co-occupied regions at trophoblast-associated genes *EGFR*, *ARID3A*,

347 *CDH1*, *ENPEP*, *KRT7*, *KRT8*, *PTPN14*, and *TEAD1* (**Fig. 5D**, **Suppl. Fig. S5F, Suppl. Table S5**).
348 In addition, we identified VGLL1-specific binding at TE/trophoblast regulators *ASCL2*,
349 *ARID5B*, *CCR7*, CITED2, *NODAL*, *TP63,* and WNT signaling effectors, such as the long non-
350 coding *LEF-AS1*, and components *EPHA2*, *FZD2*, *ITGB1, LGR4*, *LRP5/6*, *WLS*, *WNT5A*, and
351 *WNT5B* (**Fig. 5D and Suppl. Table S5**). The genes bound by VGLL1 were highly associated
352 with the human trophectoderm transcriptional signature and not with mesoderm-derived
353 cell types (**Suppl. Fig. S5G**). Together, these data demonstrate that VGLL1 directly regulates
354 a TE/trophoblast-specific program at promoter regions in complex with TEAD4 and at distal
355 regulatory regions potentially with a novel partner.

356 VGLL1 contributes to BMP4-dependent epigenetic changes by regulating the 357 histone demethylation enzyme KDM6B

358 To investigate if VGLL1 might contribute to the epigenetic changes for TE induction of
359 hESC downstream of BMP4-treatment, we first looked at chromatin modifiers whose
360 expression patterns followed VGLL1 manipulation (**Fig. 5A, Suppl. Table S4**). Among the 528
361 genes concordantly regulated, we identified *TET2*, involved in DNA demethylation, and
362 *KDM6B*, an enzyme responsible for demethylation of histone 3 lysine residue 27
363 (H3K27me3/2/1). Furthermore, we detected VGLL1 occupancy at *KDM4B* and *KDM6B* loci in
364 BMP4/IWP2-treated hESCs (**Fig. 5D**) as well as in hTSCs (data not shown), indicating these
365 lysine demethylases are direct targets of VGLL1. KDM6B has been reported to regulate the
366 demethylation of H3K27me3 at promoter regions of *Gata3* and *Cdx2*, thus promoting their
367 expression in the TE of pre-implantation murine embryos (60). Therefore, we treated hESCs
368 with a KDM6B demethylase activity inhibitor, GSKJ4, during BMP4/IWP2-mediated
369 differentiation, and, similar to VGLL1 knock-down, we found decreased levels of TE markers
370 *GATA3* and *CDX2*, compared to vehicle control (**Fig. 5F**). Collectively, these data suggest that
371 VGLL1 further contributes to the establishment of a TE/trophoblast transcriptional signature
372 by regulating the expression of chromatin remodeler enzymes.

373 VGLL1 and KDM6B are co-expressed in the TE compartment of pre-374 implantation embryos

375 To extend our findings beyond PSC-based models, we examined VGLL1 expression in
376 human embryos. Using published datasets, we investigated its temporal and spatial
377 expression during early embryonic development. Yan *et al*. previously performed RNA
378 sequencing (RNA-seq) of individual early human embryos from the oocyte to pre-
379 implantation blastocyst (27). Analysis of this data revealed that *VGLL1* was expressed at very
380 low levels until the blastocyst stage, at which point its expression began to increase (**Suppl.**
381 **Fig. S6A**). Blakeley *et al*. investigated the transcriptional signature of the three main cell
382 types found in the pre-implantation blastocyst: TE, epiblast (EPI), and primitive endoderm
383 (PE) with single-cell RNA-seq (3). Examination of datasets from these cell types revealed that
384 *VGLL1* was highly expressed in TE, with some expression in the PE but not in EPI (**Suppl. Fig.**
385 **S6B**). To further distinguish between polar TE (adjacent to the inner cell mass) and mural TE
386 (adjacent to the blastocoele) clusters, we re-analyzed single-cell RNA-seq data of
387 dissociated human embryo cells from another study that performed extended embryo
388 culture (61). We found that *VGLL1* expression was increased in the TE relative to the EPI and
389 PE, with slightly higher expression in the polar TE compared to the mural TE. By contrast,
390 *CDX2*, another TE marker central to the TE program in humans and mice, was similarly
391 increased in TE compared to EPI and PE cells but was higher in the mural compared to the
392 polar compartment (**Fig 6A**). To assess if protein expression correlated with our
393 transcriptomic results, we co-stained pre-implantation embryos with VGLL1 and both extra-
394 embryonic (GATA3, TP63) and embryonic (NANOG) markers. We observed VGLL1 positive
395 cells at the morula stage in both inside and outside cells and co-staining with GATA3 in a
396 subset of outside cells (**Fig 6B**). At the blastocyst stage, VGLL1 expression was localized in
397 the nuclei of TE cells (**Fig 6B, Suppl. Fig. S6C**). These data confirm VGLL1 activity in the TE
398 compartment of human embryos, further suggesting a role for this transcriptional co-factor
399 in establishment/maintenance of this lineage.

400 Finally, we examined KDM6B expression in both human embryos and first trimester
401 placentas and found that its expression correlated with VGLL1 in the TE/trophoblast
402 compartment. In human pre-implantation embryos, *KDM6B* was highly expressed in the TE,
403 with increased expression in the polar compared to mural TE (**Fig. 6C**). In the first trimester

404 placenta, KDM6B was expressed in the cytotrophoblast (CTB) compartment and proximal
405 cell column as well as in the syncytiotrophoblast compartment (STB) (**Fig. 6D**). Taken
406 together with our ChIPseq, these data point to a direct role for VGLL1 in regulating KDM6B
407 expression in the progenitor compartment of first trimester placenta. Overall, this study
408 demonstrates that VGLL1 regulates both genetic and epigenetic features to support a
409 trophoblast lineage-specific transcriptional program.

410 Discussion

411 Correct establishment of the TE lineage is fundamental for placental development
412 and a successful pregnancy outcome. Because of the ethical and technical limitations of
413 working with human embryos, various protocols have been reported to differentiate human
414 pluripotent stem cells into trophoblast in order to study species-specific mechanisms of
415 development. In this work, we used a BMP4-model for TE induction from primed PSCs
416 previously established in our lab, which includes addition of the WNT inhibitor IWP2 to block
417 mesoderm differentiation, followed by TSC culture adaptation (48). First, we further explore
418 the transcriptional programs and lineages elicited by BMP4 in the presence and absence of
419 IWP2. We demonstrated that activation of the canonical WNT pathway and consequent up-
420 regulation of mesoderm markers was limited to the first 2 days of treatment, with
421 transcription directed towards a TE-like program by day 4. Universal TE markers including
422 *GATA3*, *ENPEP*, and *TACSTD2* were all induced by BMP4 treatment. Our ATAC-seq data
423 comparing BMP4-driven differentiation with and without WNT signaling inhibition indicates
424 that, in primed hPSCs, the chromatin surrounding trophoblast lineage genes was accessible
425 for transcription throughout differentiation. The changes in chromatin reported here
426 accompany the temporal increase of TE/trophoblast markers that we observed in our RNA-
427 seq data. On the other hand, we showed that, at these concentrations of BMP4, chromatin
428 near mesoderm lineage genes closed over time even in the absence of WNT inhibition.

429 Temporal activation of TE/trophoblast lineage in our model was similar to previously
430 published studies of BMP4-dependent induction of the trophoblast lineage, with TEtra
431 factors GATA2/3 and TFAP2A/C expressed early during differentiation followed by more

432 mature markers (19). However, in their study Krendl and colleagues defined VGLL1 as a late
433 marker. In this study, we determined that *VGLL1* induction is downstream of TEtra factors
434 but upstream of other mature markers like *EGFR* and *TP63*. Knock-down of VGLL1 prevented
435 a complete exit from pluripotency with maintenance of factors like *POU5F1* and *SOX2*, and
436 blunted expression TE markers including *EGFR*, *ENPEP*, and *NR2F2*. Interestingly, we also
437 observed reduced expression of TEtra factors *GATA3*, *GATA2*, and *TFAP2C*. These data
438 suggest that while VGLL1 is not required for the initiation of the TE/trophoblast program, it
439 belongs to a second wave of factors necessary for the completion of this program in part via
440 a positive feedback loop onto the TEtra factors.

441 Such a role is not unique to VGLL1. In the mouse, Eomesodermin (Eomes),
442 downstream of Cdx2, is not required to initiate the TE program (18, 62) but instead plays a
443 key role as regulator of TE differentiation and maintenance of TSCs in the extra-embryonic
444 compartment (17). Embryos lacking Eomes are arrested in the early post-implantation
445 period, initiate formation of TE (e.g. by cavitation), but fail to form a proper chorion (17). While
446 EOMES is not involved in human TE/trophoblast development (3, 9), our data suggest that
447 VGLL1 might play a similar function in reinforcing the TE/TSC identity, downstream of early
448 key TE lineage regulators and allowing hTSC expansion. In fact, VGLL1 knock-down TE-like
449 cells adapted to hTSC culture conditions showed decreased expression of GATA2 and
450 GATA3, and failed to regain cell surface expression of EGFR, demonstrating lack of *bona fide*
451 hTSC derivation.

452 Canonical WNT signaling is a crucial pathway to establish and maintain various
453 epithelial stem cells (23). As mentioned, canonical WNT signaling supports mesoderm
454 differentiation of BMP4-treated hPSCs (37). VGLL1 is uniquely expressed in the
455 TE/trophoblast lineage during early human development, thus the up-regulation of
456 mesoderm genes *EOMES*, *MIXL1*, and *TBXT* in hPSCs as a consequence of ectopic
457 expression of VGLL1 was unexpected. As in our BMP4-based differentiation protocol,
458 addition of IWP2 significantly decreased the expression of mesoderm genes in iVGLL1a
459 hESCs, suggesting that canonical WNT activation, but not mesoderm genes, may be directly
460 regulated by VGLL1. In fact, temporal activation of the canonical WNT reporter at 48-72h of

461 BMP4-treatment correlated with the timeline of VGLL1 induction. We also detected active
462 WNT signaling in DOX-treated iVGLL1a hESCs. However, our data also suggest that VGLL1
463 alone cannot sustain long-term autocrine activation of the canonical WNT pathway. We
464 detected VGLL1 occupancy at WNT coreceptors *LRP5/6* as well as the WNT agonist R-
465 Spondin receptor *LGR4*. Canonical WNT pathway activation is necessary for hTSC and
466 trophoblast organoid derivation from placental tissue. Our data therefore suggest that
467 VGLL1 is key in supporting and maintaining the trophoblast stem cell state, potentially via
468 up-regulation of canonical WNT signaling pathway receptors and coreceptors.

469 Recently, Yang and colleagues explored the role of VGLL1 and TEAD4 in TE
470 specification, starting from naïve PSCs, which represent cells of the pre-implantation
471 embryonic compartment (63). Similar to our results, they demonstrated that loss of VGLL1
472 led to maintenance of pluripotency genes, blunted expression of TE genes, including *GATA3*,
473 *ENPEP* and *NR2F2*, and inability to derive hTSCs. They demonstrated that the VGLL1/TEAD4
474 complex directly regulates expression of markers such as *GATA3* and *ENPEP* at promoter and
475 intergenic regions. In contrast, our model utilizes primed PSCs, which represent a post-
476 implantation epiblast state. Despite the different initial states, our data largely corroborate
477 their findings with VGLL1 directly regulating markers such as *GATA3* and *ENPEP*.
478 Interestingly, we showed VGLL1 occupancy at distal enhancer regions of *GATA3* and *EGFR*.
479 These regions were recently mapped to predicted enhancers active in the human placenta
480 across gestation (59), suggesting a role for VGLL1 in trophoblast-specific distal regulatory
481 regions. In fact, when we looked at the presence of these distal enhancer peaks in the Yang
482 *et al* data, we found similar peaks, corroborating that they might represent key regulatory
483 regions for TE and trophoblast identity.

484 The TE and ICM lineages segregate early during mammalian development and are
485 distinguished by unique transcriptomes and epigenomes. Both naïve and primed hPSCs
486 contain bivalent promoters at developmental genes, which maintain them in a
487 transcriptionally poised state (19, 64, 65). Such bivalent regions are characterized by
488 H3K4me3, a post-translational histone modification mark associated with active promoters,
489 and H3K27me3, a marker correlated with transcriptional repression. Global distribution of

490 H3K27me3 differs between primed and naive hPSCs. However, many of the genes marked
491 by H3K27me3 and shared between these two pluripotency states are trophoblast regulators,
492 including GATA3 (64, 65). In their study, Yang and colleagues found a correlation between
493 loss of VGLL1 and lower levels of H3K27ac, a histone mark associated with active promoters
494 and enhancers, at TE-specific genes (63). However, the authors did not show a direct link
495 between VGLL1 expression and H3K27ac levels at TE genes. Moreover, before H3K27 can be
496 modified by acetylation, the methyl groups need to be removed from the bivalent promoter
497 containing H3K27me3. We have also observed decreased levels of global H3K27ac in our
498 primed model (data not shown) and identified KDM6B, one of two H3K27me3 specific
499 demethylases in mammals, as a potential mediator of VGLL1-dependent chromatin
500 remodeling. The transcript levels of *KDM6B* were largely dependent on VGLL1 expression
501 and we observed direct VGLL1 binding on *KDM6B* in our ChIP-seq data. In the murine
502 blastocyst, KDM6B removes the repressive H3K27me3 at the promoter regions of *Gata3* and
503 *Cdx2* specifically in the TE (60). Loss of *Kdm6b* results in maintenance of H3K27me3 at TE-
504 specific gene regulatory regions, lower expression of Gata3 and Cdx2 in the TE, and an
505 impaired ability for the embryos to implant into the uterus. In humans, *VGLL1* and *KDM6B*
506 are highly expressed in the polar TE of the human blastocyst, which attaches to the uterine
507 wall during implantation. It remains unclear whether KDM6B is involved in human
508 implantation, but our results show that chemical inhibition of KDM6B demethylase activity
509 phenocopies loss of VGLL1 and results in the blunted expression of *GATA3* and *CDX2*, *in*
510 *vitro*. Thus, we posit a model whereby VGLL1 directly up-regulates the expression of KDM6B,
511 which contributes to demethylation of H3K27me3 at bivalent/poised promoters prior to
512 deposition of the acetyl group by acetyltransferases at trophoblast-specific regulatory
513 regions to mark a transcriptionally permissive chromatin state (**Fig.7**). Evidence suggests
514 that, while mouse TE is fully specified in the pre-implantation blastocyst, human TE
515 commitment might not be completed until post-implantation (51). VGLL1-dependent
516 regulation of chromatin remodelers like KDM6B might provide a mechanism explaining the
517 higher plasticity of human post-implantation pluripotent cells.

518 Overall, our data extend previous findings and add a substantial mechanistic hypothesis on
519 the role of VGLL1 in TE and trophoblast lineage specification. We also demonstrated that
520 primed hPSCs and the BMP4-dependent differentiation protocol remain a functional and
521 practical *in vitro* model to study fundamental aspects of genetic and epigenetic regulation
522 of TE and trophoblast cell identity in humans (66).

523 Material and Methods

524 **Human embryo and Placental Tissue Collection.** Human embryos were donated to the
525 research project by informed consent under the UK Human Fertilisation and Authority
526 Licence number R0162 and independently reviewed by the Health Research Authority's
527 Research Ethics Committee IRAS projects 308099. Human placental tissues were collected
528 under a protocol approved by the Human Research Protections Program Committee of the
529 University of California, San Diego Institutional Review Board; all patients gave informed
530 consent for the collection and use of these tissues.

531 **Culture of human embryonic stem cells.** Human embryonic stem cells H9/WA09 (H9) and
532 HUES9TOPGFP were cultured and passaged in mTeSR Plus (StemCell Technologies cat. no.
533 100-1130) on Geltrex (Gibco cat. no. A14133-02)-coated plates as previously described (67).
534 shRNA H9 hESC lines were previously derived and maintained in 10 µg/ml puromycin (Sigma
535 Aldrich cat. no. P8833-25MG)(9).

536 **Derivation of dCas9-VPR VGLL1 gRNAs (iVGLL1a) H9 hESCs.** Five RNA guides for VGLL1
537 activation were designed using the sgRNA Designer tool by the Broad Institute (68) and
538 cloned into the Golden Gate pMA Cas9 Guide vector system (69), modified to incorporate an
539 Eif1a-Bsd antibiotic resistance cassette for selectable incorporation into a mammalian
540 cells. Each of the guides were cloned into the U6 promoter driven cloning site on the
541 plasmid, and sequence verified. Plasmids containing DOX-inducible dCas9-VPR (gift from
542 Dr. Church) (70), VGLL1-specific guide RNAs, and TOPGFP.mC (Addgene cat. no. 35491) (55)
543 were introduced in H9 ESCs using the Amaxa Nucleofector II electroporation system
544 following manufacturer instructions (setting B-016). Briefly, 1 million cells were collected,
545 resuspended and mixed with 2 µg of plasmid DNA before nucleofection using the Lonza

546 Human Stem Cell Nucleofector Kit 1 (Lonza cat. no. VPH-5012). Cells were immediately re-
547 plated, and selection started 48 hours later. iVGLL1a cell lines were maintained with 2 µg/mL
548 puromycin and 8 µg/mL blasticidin S (Gibco cat. no. A1113903). Ectopic expression of VGLL1
549 was induced with 2 µg/mL doxycycline (Millipore Sigma cat. no. D3447-500MG) for 4 days.

550 **Trophoblast differentiation of hESC.** hESCs were first differentiated into TE-like cells
551 followed by adaptation to hTSC culture conditions according to a previously published
552 protocol (67). Briefly, hESCs were dissociated with TrypLE Express (Gibco cat. no. 12604021)
553 and plated at a density of 10,000 cells per cm^2 on Geltrex-coated plates in mTeSR Plus with
554 5µM Y27632 (Selleck Chemicals cat. no. S1049). After 24 hours, media was changed to
555 differentiation media composed of DMEM/F12 (Corning, cat. no. 10-092-CV), 1x ITS
556 (MilliporeSigma, cat. no. I1884), L-ascorbic acid 64 µg/mL (MilliporeSigma, cat. no. A8960),
557 sodium bicarbonate 543 µg/mL (ThermoFisher, cat. no. 25080094), 2% BSA (Gemini Bio-
558 Products, cat. no. 700-110), 10 ng/mL recombinant human BMP4 (R&D Systems cat. no. 314-
559 BP-010), and 2 µM IWP2 (Selleck Chemicals cat. no. S7085). Media was changed daily. At
560 day 4, cells were either collected for analysis or re-plated on collagen IV (5µg/mL)-coated
561 plates (MilliporeSigma cat. no. C0543-1VL) and media changed to iCTB media, composed of
562 Advanced DMEM/F12 (ThermoFisher, cat. no. 12634028), 1X N2 (ThermoFisher, cat. no.
563 17502001), 1X B27 (ThermoFisher, cat. no. 17504001), 1X GlutMax (ThermoFisher, cat. no.
564 35050061), 150 µM 1-thioglycerol (MilliporeSigma, cat. no. M6145), 0.05% BSA (Gemini Bio,
565 cat. no. 700-110-100), 1% knockout serum replacement (ThermoFisher, cat. no. 10828028),
566 2 µM CHIR99021 (MilliporeSigma SML1046-25MG), 0.5 µM A83-01 (Tocris, cat. no. 2939), 1
567 µM SB431542 (Millipore Sigma cat. no. 616464-5MG), 5 µM Y27632 (Selleck Chemicals, cat.
568 no. S1049), 130 µg/mL VPA sodium salt (MilliporeSigma, cat. no. 676380), 100 ng/mL
569 recombinant human FGF2 (BioPioneer, cat. no. HRP-0011), 50 ng/mL recombinant human
570 EGF (R&D Systems, cat. no. 236-EG), 50 ng/mL recombinant human HGF (STEMCELL
571 Technologies, cat. no. 78019.2), and 20 ng/mL Noggin (R&D Systems cat. no. 6057-NG).
572 IWP2 was withheld in the BMP4 only treatment. 2 µM IWR1 (MilliporeSigma, cat. no. I0161)
573 was used as an alternative WNT inhibitor for HUES9TOPGFP. 1 µM GSK-J4 (MedChem
574 Express cat. no. HY-15648B) was used for KDM6B activity inhibition.

575 **Immunohistochemistry.** Formalin-fixed, paraffin-embedded placental tissue sections
576 were stained with anti-KDM6B (Abcam cat. no. ab38113) and Goat anti-Rabbit IgG
577 secondary antibody conjugated with AlexaFluor 488 (Invitrogen cat. no. A11008) following
578 previously published protocol (20). Embryos were thawed according to recommendations
579 from Bourn Hall Clinic, the *in vitro* fertilization clinic coordinating donations. hESCs and
580 embryos were fixed in 4% paraformaldehyde (Electron Microscopy Sciences cat. no. 15710)
581 and stained as previously described(3). The following antibodies were used: anti-VGLL1
582 (Sigma Aldrich cat. no. HPA042403-100UL), anti-TEAD4 (Abcam cat. no. ab58310), anti-YAP1
583 (Abcam cat. no. ab52771), anti-NANOG (Company), anti-GATA3, mouse anti-p40 (ΔNp63-
584 specific antibody, clone BC28; Biocare Medical). Embryos were imaged on a Leica SP5
585 inverted confocal microscope (Leica Microsystems). Cells and placentas were imaged on an
586 ECHO Rebel microscope (ECHO).

587 **Flow cytometry**. Flow cytometry was conducted using live cells as previously
588 described(67). Briefly, single cell suspensions were stained with EGFR-APC (Biolegend cat.
589 no. 352906), or APC-conjugated mouse IgG (BioLegend cat. no. 400120) antibodies, in flow
590 cytometry buffer (1% FBS, 2% BSA, and 0.03% sodium azide) for 1 hour at 4°C. Cells were
591 then washed three times in PBS and fixed with 2% PFA (Electron Microscopy Sciences cat.
592 no. 15710). Flow cytometry analysis was performed on a Fortessa X20 flow cytometer (BD
593 Biosciences). Data were visualized using FlowJo Software (v10).

594 **RNA isolation, cDNA preparation, and quantitative real-time PCR**. Total RNA was purified
595 using NucleoSpin isolation kit (Macherey-Nagel cat. no. 740955.250) and RNA concentration
596 measured using NanoDrop 2000C spectrophotometer (ThermoScientific). cDNA was
597 prepared from total RNA using the PrimeScript RT-Kit (Takara Bio cat. no. RR037A).
598 Quantitative real-time PCR (qPCR) was performed using TB Green Premix Ex Taq (Takara bio
599 cat. no. RR420A) on a QuantStudio5 thermocycler (AppliedBiosystems). The primer
600 sequences used are listed in **Supplemental Table 2**. Relative expression of each transcript
601 was calculated using ΔΔCT method, normalized to 18S rRNA. Statistical analysis and graphs
602 were generated using GraphPad Prism 10. Bar chart data display mean fold change and
603 standard deviation. Student's t tests were performed to determine significance between two

604 groups. Two-way ANOVA was performed when analyzing significant differences between
605 three or more groups. All qRT-PCR experiments were done using three replicates and
606 statistical tests were performed using log transformed fold change. Level of significance is
607 represented with asterisks.

608 **RNA sequencing and analysis**. Total RNA extraction with mirVana (Ambion cat. no. AM1561)
609 and library preparation for RNA-sequencing were performed as previously described (71).
610 Quality control was performed using FastQC (v.0.12.1) and multiQC (v.1.19). Two samples
611 with a lower signal quality than their replicates were eliminated. Reads were mapped to
612 GRCh38.p14 (GENCODE release 45) using STAR (v.2.7.9a) (72) and annotated using
613 featureCounts (subread v.2.0.6). Ensembl genes without at least three samples with 10 or
614 more reads were removed from analysis. Normalization and differential expression analysis
615 was performed using the R (v.4.4.0) package DESeq2 (v.1.46.0) (73). Genes with an adjusted
616 p-value < 0.05 and $\log_2$ fold change > 1 were considered differentially expressed. In iVGLL1a
617 models, adjusted p-value < 0.05 and $\log_2$ fold change > 0.5 were used to include subtle but
618 biologically relevant changes. Heatmaps and PCA were generated with Qlucore. Statistical
619 analysis was performed using a two-group comparison or multi-group comparison with q-
620 value <0.05. Gene list enrichment analysis was performed with Enrichr on multiple database
621 as specified in the text (74).

622 **Assay for Transposase-Accessible Chromatin (ATAC)-sequencing.** Nuclei extraction and
623 ATAC-seq library prep ware performed according to previously published protocol (75). The
624 bcbio-nextgen (v1.2.8) ATAC-seq pipeline was used to process the ATAC-seq reads. The
625 nucleosome free fraction was normalized using deeptools (v 3.5.1) (76) bamCoverage by
626 CPM. One sample with a lower signal quality than its replicates, as determined by limma-
627 voom multi-dimensional scaling plots, was eliminated. Bigwigs for replicates were
628 combined using wiggletools (v 1.2.11) (77) mean to obtain one average track from biological
629 replicates. Genomic tracks were visualized using IGV Viewer (v2.11.7). Heatmaps were
630 generated using deeptools computeMatrix and plotHeatmap. Binding profiles were
631 generated deeptools computeMatrix and plotProfile. Consensus peaks were annotated
632 using ChIPpeakAnno (v 3.32.0)(78) with EnsDb.Hsapiens.v86 (v 2.99.0). Chromosome

633 regions were annotated using ChIPpeakAnno assignChromosomeRegion with
634 TxDb.Hsapiens.UCSC.hg38.knownGene (v 3.16.0). Differential expression analysis was
635 carried out using edgeR (v 3.40.2)(79) with limma-voom normalization (80). HOMER (v
636 5.1)(81) findMotifsGenome.pl was then used to identify transcription factor binding motifs.

637 **Chromatin Immunoprecipitation (ChIP)-sequencing.** Chromatin immunoprecipitation
638 was performed following a previously published protocol(58). Briefly, cells were crosslinked
639 with 1% formaldehyde (Pierce cat. no. 28906) for 10 minutes at room temperature on a
640 horizontal rocker. Crosslinking was stopped with 2.5 M glycine. Cells were collected into
641 15mL conical tubes and centrifuged 1200rpm for 5 minutes at 4°C. Cell pellets were washed
642 in PBS two times and cell pellets were stored in -80°C until ChIP. Frozen cell pellets were
643 resuspended in cell lysis buffer with protease inhibitors and incubated at 4°C for 10 minutes.
644 Following centrifugation, nuclear pellets were resuspended in nuclei lysis buffer with
645 protease inhibitors and incubated at 4°C for 10 minutes. Covaris shearing buffer was added
646 and chromatin was sheared in microtube-500 AFA (Covaris cat. no. 520185) using Covaris
647 E220 Focused-Ultrasonicator with settings 5% duty, PiP 140, 15 minutes in 30 second bursts,
648 cycles 200. Sodium chloride concentration was adjusted to 150mM and Triton-X to 1%
649 following shearing. Pre-clearing was done with agarose beads (Cell Signaling Technology
650 cat. no. 9863S) for 30 minutes at 4°C on a rocker. Input was collected and stored in -20°C.
651 Cleared lysates were incubated with 6 µg of either anti-TEAD4 (Abcam cat. no. ab58310) or
652 anti-VGLL1 antibody (Sigma Aldrich cat. no. HPA042403-100UL) overnight at 4°C on a rocker.
653 Agarose beads (50 µL) were added and incubated at 4°C for 1 hour on a rocker. Lysates were
654 centrifuged at 14000rpm for 5 minutes at 4°C, and washed five times. DNA-protein-antibody
655 complexes were eluted in two steps with 0.1 M sodium bicarbonate, 1% SDS, vortexed
656 briefly, and centrifuged at 7500rpm for 2 minutes at room temperature. Reverse crosslinking
657 for eluted and input DNA was done with 3 µg/mL RNAse A (ThermoScientific cat. no.
658 EN0531) and 0.3 M sodium chloride (Invitrogen cat. no. AM9759) on a heat block at 67°C for
659 5 hours. Proteinase K (Invitrogen cat. no. 100005393) was added (0.2 mg/mL) and incubated
660 at 45°C overnight. DNA purification was performed with Monarch PCR and DNA Cleanup kit

(NEB #T1030) and submitted for library preparation and sequencing conducted at the IGM Genomics Center, University of California, San Diego, La Jolla, CA.

Adapters were trimmed from reads using Trim Galore (v0.6.10) -q 20 --fastqc --paired. Trimmed reads were aligned to GRCh38 using bowtie2 (v2.4.1) with setting -p 24. samtools (v1.10) were used to sort and index bam files. Peaks were called using MACS2 (v2.2.6) callpeak with settings -f BAMPE -g 2.7e+9 -q 0.05 --keep-dup all --call-summits. DiffBind (v3.16.0) was used to find consensus peaks among replicates using minOverlapmembers = 2. BigWig files were generated with deeptools (v3.5.5) bamCoverage with settings --binSize 20 --smoothLength 60 --effectiveGenomeSize 2913022398 –normalizeUsing CPM --extendReads 150 --centerReads --scaleFactorsMethod None --minMappingQuality 30 --ignoreForNormalization chrM, then replicates were averaged using bigwigAverage with settings -bs 20. Heatmaps were generated with deeptools computeMatrix reference-point --referencePoint center -b 1000 -a 1000 -missingDataAsZero -refPointLabel 0 and visualized using plotHeatmap --kmeans 2. ChIPpeakAnno (v3.14.2) was used to annotate peaksets. Motif analysis was performed using HOMER (v5.1) findMotifsGenome.pl with -size 200. Pie charts of genomic annotations were generated with ChIPpeakAnno pie1 function and genomic element distribution gED. Genomic tracks were visualized using IGV Viewer (v2.11.7). Computationally intensive sequencing analyses were done on the San Diego Supercomputer Center (2025): Expanse through the University of California San Diego(82).

Data availability

All datasets are available in Geo: GSE295121, GSE295122, GSE295124, GSE295126.

Acknowledgments

All sequencing was conducted at the IGM Genomics Center, University of California, San Diego, La Jolla, CA. We thank you Karl Willert for the generous donation of the TOPGFP HUES9 cell line, and Marisa Korody for the IWR1. We are grateful to Dr. Mana Parast for the continuous encouragement and support. The contents of this publication are solely the

687 responsibility of the authors and do not necessarily represent the official views of CIRM or
688 any other agency of the State of California.

689 Fundings

690 FS and RC were supported by the Eunice Kennedy Shriver National Institute of Child Health
691 and Human Development (NICHD) grant R01HD096260 to FS and Diversity supplement to
692 RC. This research was also supported by the California Institute for Regenerative Medicine
693 (CIRM) awards DISC0-13757 to FS and DISC0-13816 to HCA. NS was funded by CIRM EDUC4
694 – 12804 Interdisciplinary Stem Cell Training Grant. KF was partially supported by NIH Grant
695 UL1TR001442 and MH was partially supported by T32 GM145427. AM was funded by the
696 CIRM Bridges program. PDH was supported by the UC San Diego OB/GYN and Reproductive
697 Sciences. Work in the laboratory of K.K.N. was supported by the Wellcome Trust
698 (221856/Z/20/Z). Work in the laboratory of KKN was also supported by the Francis Crick
699 Institute, which receives its core funding from Cancer Research UK CC2074, the Medical
700 Research Council CC2074, and Wellcome Trust CC2074. This research was supported by
701 NIH grants HD062546, HD0098880 and HD079363 to SP. This publication includes data
702 generated at the UC San Diego IGM Genomics Center utilizing an Illumina NovaSeq 6000
703 that was purchased with funding from a National Institutes of Health SIG grant (#S10
704 OD026929).

705

Bibliography

1. V. B. Alarcon, Y. Marikawa, Trophectoderm formation: regulation of morphogenesis and gene expressions by RHO, ROCK, cell polarity, and HIPPO signaling HHS Public Access. *Reproduction* **164**, 75–86 (2022).

2. M. Y. Turco, A. Moffett, Development of the human placenta. (2019). https://doi.org/10.1242/dev.163428.

3. P. Blakeley, *et al.*, Defining the three cell lineages of the human blastocyst by single-cell RNA-seq (Development, (2015) 142, 3151-3165). *Dev.* **142**, 3613 (2015).

4. C. Gerri, *et al.*, Initiation of a conserved trophectoderm program in human, cow and mouse embryos. *Nature* **587**, 443–447 (2020).

5. R. M. Roberts, *et al.*, The role of BMP4 signaling in trophoblast emergence from pluripotency. *Cell. Mol. Life Sci.* **79** (2022).

6. N. Nishioka, *et al.*, The Hippo Signaling Pathway Components Lats and Yap Pattern Tead4 Activity to Distinguish Mouse Trophectoderm from Inner Cell Mass. *Dev. Cell* **16**, 398–410 (2009).

7. A. C. Denicol, *et al.*, Canonical WNT signaling regulates development of bovine embryos to the blastocyst stage. *Sci. Reports 2013 31* **3**, 1–7 (2013).

8. K. B. Lee, J. K. Folger, S. K. Rajput, G. W. Smith, Temporal regulation of mRNAs for select bone morphogenetic proteins (BMP), BMP receptors and their associated SMAD proteins during bovine early embryonic development: Effects of exogenous BMP2 on embryo developmental progression. *Reprod. Biol. Endocrinol.* **12**, 1–10 (2014).

9. F. Soncin, *et al.*, Comparative analysis of mouse and human placentae across gestation reveals species-specific regulators of placental development. *Dev.* **145** (2018).

10. C. Gerri, *et al.*, A conserved role of the Hippo signalling pathway in initiation of the first lineage specification event across mammals. *Dev.* **150** (2023).

11. M. Regin, *et al.*, Lineage segregation in human pre-implantation embryos is specified by YAP1 and TEAD1. *Hum. Reprod.* **38**, 1484–1498 (2023).

12. B. A. T. Weatherbee, *et al.*, Distinct pathways drive anterior hypoblast specification in the implanting human embryo. *Nat. Cell Biol.* **26**, 353–365 (2024).

13. H. Sasaki, Roles and regulations of Hippo signaling during preimplantation mouse development. *Dev. Growth Differ.* **59**, 12–20 (2017).

14. M. Azagury, Y. Buganim, Unlocking trophectoderm mysteries: In vivo and in vitro perspectives on human and mouse trophoblast fate induction. *Dev. Cell* **59**, 941–960 (2024).

742 15. A. Ralston, *et al.*, Gata3 regulates trophoblast development downstream of Tead4
743 and in parallel to Cdx2. *Development* **137**, 395–403 (2010).

744 16. H. Sasaki, Position- and polarity-dependent Hippo signaling regulates cell fates in
745 preimplantation mouse embryos. *Semin. Cell Dev. Biol.* **47–48**, 80–87 (2015).

746 17. A. P. Russ, *et al.*, Eomesodermin is required for mouse trophoblast development and
747 mesoderm formation. *Nature* **404**, 95–99 (2000).

748 18. D. Strumpf, *et al.*, Cdx2 is required for correct cell fate specification and
749 differentiation of trophectoderm in the mouse blastocyst. *Development* **132**, 2093–
750 2102 (2005).

751 19. C. Krendl, *et al.*, GATA2/3-TFAP2A/C transcription factor network couples human
752 pluripotent stem cell differentiation to trophectoderm with repression of
753 pluripotency. *Proc. Natl. Acad. Sci. U. S. A.* **114**, E9579–E9588 (2017).

754 20. Y. Li, *et al.*, BMP4-directed trophoblast differentiation of human embryonic stem
755 cells is mediated through a ΔNp63+ cytotrophoblast stem cell state. *Dev.* **140**, 3965–
756 3976 (2013).

757 21. H. M. Sonnemann, B. Pazdrak, D. A. Antunes, J. Roszik, G. Lizée, Vestigial-like 1
758 (VGLL1): An ancient co-transcriptional activator linking wing, placenta, and tumor
759 development. *Biochim. Biophys. Acta - Rev. Cancer* **1878**, 188892 (2023).

760 22. K. Qin, *et al.*, Canonical and noncanonical Wnt signaling: Multilayered mediators,
761 signaling mechanisms and major signaling crosstalk. *Genes Dis.* **11**, 103–134 (2024).

762 23. A. Haegebarth, H. Clevers, Wnt signaling, lgr5, and stem cells in the intestine and
763 skin. *Am. J. Pathol.* **174**, 715–721 (2009).

764 24. D. Pinto, H. Clevers, Wnt control of stem cells and differentiation in the intestinal
765 epithelium. *Exp. Cell Res.* **306**, 357–363 (2005).

766 25. J. Adjaye, *et al.*, Primary Differentiation in the Human Blastocyst: Comparative
767 Molecular Portraits of Inner Cell Mass and Trophectoderm Cells. *Stem Cells* **23**,
768 1514–1525 (2005).

769 26. L. Aghajanova, *et al.*, Comparative transcriptome analysis of human trophectoderm
770 and embryonic stem cell-derived trophoblasts reveal key participants in early
771 implantation. *Biol. Reprod.* **86**, 11–12 (2012).

772 27. L. Yan, *et al.*, Single-cell RNA-Seq profiling of human preimplantation embryos and
773 embryonic stem cells. *Nat. Publ. Gr.* (2013). https://doi.org/10.1038/nsmb.2660.

774 28. M. Krivega, W. Essahib, H. Van de Velde, WNT3 and membrane-associated β-catenin
775 regulate trophectoderm lineage differentiation in human blastocysts. *Mol. Hum.*
776 *Reprod.* **21**, 711–722 (2015).

777 29. S. Sonderegger, J. Pollheimer, M. Knöfler, Wnt signalling in implantation,

decidualisation and placental differentiation--review. *Placenta* **31**, 839–847 (2010).

30. Y. Hu, *et al.*, Single-cell analysis of nonhuman primate preimplantation development in comparison to humans and mice. *Dev. Dyn.* **250**, 974–985 (2021).

31. H. Okae, *et al.*, Derivation of Human Trophoblast Stem Cells. *Cell Stem Cell* **22**, 50-63.e6 (2018).

32. S. Haider, *et al.*, Self-Renewing Trophoblast Organoids Recapitulate the Developmental Program of the Early Human Placenta. *Stem Cell Reports* **11**, 537–551 (2018).

33. M. Y. Turco, *et al.*, Trophoblast organoids as a model for maternal–fetal interactions during human placentation. *Nature* **564**, 263–281 (2018).

34. B. Dietrich, S. Haider, G. Meinhardt, J. Pollheimer, · Martin Knöfler, WNT and NOTCH signaling in human trophoblast development and differentiation. **79**, 292 (2022).

35. A. S. Brumm, *et al.*, Initiation and maintenance of the pluripotent epiblast in pre-implantation human development is independent of NODAL signaling. *Dev. Cell* **60** (2025).

36. J. Au, *et al.*, Role of autocrine bone morphogenetic protein signaling in trophoblast stem cells. *Biol. Reprod.* (2021). https://doi.org/10.1093/biolre/ioab213.

37. D. Kurek, *et al.*, Endogenous WNT signals mediate BMP-induced and spontaneous differentiation of epiblast stem cells and human embryonic stem cells. *Stem Cell Reports* **4**, 114–128 (2015).

38. N. Sah, F. Soncin, Conserved and divergent features of trophoblast stem cells. *J. Mol. Endocrinol.* **72** (2024).

39. L. Li, *et al.*, Single-cell multi-omics sequencing of human early embryos. *Nat. Cell Biol. 2018 207* **20**, 847–858 (2018).

40. L. Liu, *et al.*, An integrated chromatin accessibility and transcriptome landscape of human pre-implantation embryos. *Nat. Commun. 2019 101* **10**, 1–11 (2019).

41. A. L. Wilkinson, I. Zorzan, P. J. Rugg-Gunn, Epigenetic regulation of early human embryo development. *Cell Stem Cell* **30**, 1569–1584 (2023).

42. S. Io, *et al.*, Capturing human trophoblast development with naive pluripotent stem cells in vitro. *Cell Stem Cell* **28**, 1023-1039.e13 (2021).

43. G. Guo, *et al.*, Human naive epiblast cells possess unrestricted lineage potential. *Cell Stem Cell* **28**, 1040-1056.e6 (2021).

44. J. K. Cinkornpumin, *et al.*, Naive Human Embryonic Stem Cells Can Give Rise to Cells with a Trophoblast-like Transcriptome and Methylome. *Stem Cell Reports* **15**, 198–213 (2020).

813 45. M. Horii, T. Bui, O. Touma, H. Y. Cho, M. M. Parast, An Improved Two-Step Protocol
814 for Trophoblast Differentiation of Human Pluripotent Stem Cells. *Curr. Protoc. Stem*
815 *Cell Biol.* **50**, 1–21 (2019).

816 46. M. Amita, *et al.*, Complete and unidirectional conversion of human embryonic stem
817 cells to trophoblast by BMP4. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 1–10 (2013).

818 47. M. Horii, *et al.*, Human pluripotent stem cells as a model of trophoblast
819 differentiation in both normal development and disease. *Proc. Natl. Acad. Sci. U. S.*
820 *A.* **113**, E3882–E3891 (2016).

821 48. F. Soncin, *et al.*, Derivation of functional trophoblast stem cells from primed human
822 pluripotent stem cells. *Stem Cell Reports* **17**, 1303–1317 (2022).

823 49. N. Moya, J. Cutts, T. Gaasterland, K. Willert, D. A. Brafman, Endogenous WNT
824 signaling regulates hPSC-derived neural progenitor cell heterogeneity and specifies
825 their regional identity. *Stem Cell Reports* **3**, 1015–1028 (2014).

826 50. W. Huang, *et al.*, Uncovering the role of TET2-mediated ENPEP activation in
827 trophoblast cell fate determination. *Cell. Mol. Life Sci.* **81** (2024).

828 51. D. Meistermann, *et al.*, Integrated pseudotime analysis of human pre-implantation
829 embryo single-cell transcriptomes reveals the dynamics of lineage specification.
830 *Cell Stem Cell* **28**, 1625-1640.e6 (2021).

831 52. X. Luan, *et al.*, Down-regulation of CCR7 via AKT pathway and GATA2 inactivation
832 suppressed trophoblast migration and invasion in recurrent spontaneous abortion†.
833 *Biol. Reprod.* **102**, 424–433 (2020).

834 53. M. Drukker, *et al.*, Isolation of primitive endoderm, mesoderm, vascular endothelial
835 and trophoblast progenitors from human pluripotent stem cells. *Nat. Biotechnol.* **30**,
836 531–542 (2012).

837 54. C. Dong, *et al.*, A genome-wide CRISPR-Cas9 knockout screen identifies essential
838 and growth-restricting genes in human trophoblast stem cells. *Nat. Commun. 2022*
839 *131* **13**, 1–16 (2022).

840 55. D. Horst, *et al.*, Differential WNT activity in colorectal cancer confers limited
841 tumorigenic potential and is regulated by MAPK signaling. *Cancer Res.* **72**, 1547–
842 1556 (2012).

843 56. B. Saha, *et al.*, TEAD4 ensures postimplantation development by promoting
844 trophoblast self-renewal: An implication in early human pregnancy loss. *Proc. Natl.*
845 *Acad. Sci. U. S. A.* **117**, 17864–17875 (2020).

846 57. G. Meinhardt, *et al.*, Pivotal role of the transcriptional co-activator YAP in trophoblast
847 stemness of the developing human placenta. *Proc. Natl. Acad. Sci. U. S. A.* **117**,
848 13562–13570 (2020).

849 58. P. Home, *et al.*, Altered subcellular localization of transcription factor TEAD4

850 regulates first mammalian cell lineage commitment. *Proc. Natl. Acad. Sci. U. S. A.*
851 **109**, 7362–7367 (2012).

852 59. D. M. Owen, *et al.*, Genome-wide identification of transcriptional enhancers during
853 human placental development and association with function, differentiation, and
854 disease†. *Biol. Reprod.* **109**, 965–981 (2023).

855 60. B. Saha, *et al.*, EED and KDM6B coordinate the first mammalian cell lineage
856 commitment to ensure embryo implantation. *Mol. Cell. Biol.* **33**, 2691–2705 (2013).

857 61. F. Zhou, *et al.*, Reconstituting the transcriptome and DNA methylome landscapes of
858 human implantation. *Nature* **572**, 660–664 (2019).

859 62. H. Niwa, *et al.*, Interaction between Oct3/4 and Cdx2 determines trophectoderm
860 differentiation. *Cell* **123**, 917–929 (2005).

861 63. Y. Yang, *et al.*, VGLL1 cooperates with TEAD4 to control human trophectoderm
862 lineage specification. *Nat. Commun.* **15** (2024).

863 64. B. Kumar, *et al.*, Polycomb repressive complex 2 shields naïve human pluripotent
864 cells from trophectoderm differentiation. *Nat. Cell Biol. 2022 246* **24**, 845–857
865 (2022).

866 65. D. W. Zijlmans, *et al.*, Integrated multi-omics reveal polycomb repressive complex 2
867 restricts human trophoblast induction. **24**, 858–871 (2022).

868 66. A. S. Seetharam, *et al.*, The product of BMP-directed differentiation protocols for
869 human primed pluripotent stem cells is placental trophoblast and not amnion. *Stem*
870 *Cell Reports* **17. 10.**, /j.stemcr.2022.04.014 (2022).

871 67. V. C. Cheung, *et al.*, Current Strategies of Modeling Human Trophoblast Using
872 Human Pluripotent Stem Cells in vitro. (2023). https://doi.org/10.1002/cpz1.875.

873 68. H. K. Kim, *et al.*, Deep learning improves prediction of CRISPR–Cpf1 guide RNA
874 activity. *Nat. Biotechnol. 2018 363* **36**, 239–241 (2018).

875 69. J. Vad-Nielsen, L. Lin, L. Bolund, A. L. Nielsen, Y. Luo, Golden Gate Assembly of
876 CRISPR gRNA expression array for simultaneously targeting multiple genes. *Cell.*
877 *Mol. Life Sci.* **73**, 4315–4325 (2016).

878 70. A. Bertero, *et al.*, Optimized inducible shRNA and CRISPR/Cas9 platforms for in vitro
879 studies of human development using hPSCs. *Development* **143**, 4405–4418 (2016).

880 71. R. Morey, *et al.*, Transcriptomic Drivers of Differentiation, Maturation, and Polyploidy
881 in Human Extravillous Trophoblast. *Front. Cell Dev. Biol.* **9** (2021).

882 72. A. Dobin, *et al.*, STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21
883 (2013).

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

74. E. Y. Chen, *et al.*, Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* **14** (2013).

75. F. C. Grandi, H. Modi, L. Kampman, M. R. Corces, Chromatin accessibility profiling by ATAC-seq. https://doi.org/10.1038/s41596-022-00692-9.

76. F. Ramírez, *et al.*, deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160–W165 (2016).

77. D. R. Zerbino, N. Johnson, T. Juettemann, S. P. Wilder, P. Flicek, WiggleTools: parallel processing of large collections of genome-wide datasets for visualization and statistical analysis. *Bioinformatics* **30**, 1008–1009 (2014).

78. L. J. Zhu, *et al.*, ChIPpeakAnno: a Bioconductor package to annotate ChIP-seq and ChIP-chip data. *BMC Bioinformatics* **11** (2010).

79. Y. Chen, L. Chen, A. T. L. Lun, P. L. Baldoni, G. K. Smyth, edgeR v4: powerful differential analysis of sequencing data with expanded functionality and improved support for small counts and larger datasets. *Nucleic Acids Res*. **53**, 13–14 (2025).

80. C. W. Law, Y. Chen, W. Shi, G. K. Smyth, Voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol.* **15**, 1–17 (2014).

81. S. Heinz, *et al.*, Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).

82. S. Strande, *et al.*, Expanse: Computing without Boundaries: Architecture, Deployment, and Early Operations Experiences of a Supercomputer Designed for the Rapid Evolution in Science and Engineering. *ACM Int. Conf. Proceeding Ser.* (2021). https://doi.org/10.1145/3437359.3465588.

910 ## Figure Legends

911 **Figure 1. BMP4 treatment of hESCs induces chromatin remodeling. A.** Venn diagram of
912 up-regulated genes following treatment of hESCs with BMP4 with and without IWP2 for 2
913 days. Gene list enrichment analysis on the KEGG pathway database (2021) for genes
914 uniquely up-regulated with BMP4 alone treatment is presented, a=WNT-mediated b=FGF-
915 mediated c=TGFβ-mediated. **B.** Gene expression levels of markers for the mesoderm lineage
916 across treatment/days by bulk RNA-seq. **C.** Heatmaps illustrating open chromatin loci by
917 ATAC-seq. **D.** Signal intensity of open chromatin at mesoderm- and TE-specific genes. **E.**
918 Heatmap of differentially expressed chromatin modifiers genes across treatment/days.

919 **Figure 2. Loss of VGLL1 prevents completion of the TE/trophoblast program. A.** Gene
920 expression levels of early and late trophoblast markers in the first 72h of BMP4/IWP2
921 treatment of hESCs by qPCR. Statistical analysis relative to 0h **B.** Gene expression levels of
922 early and late trophoblast markers in shVGLL1 compared to Scramble hESCs during
923 BMP4/IWP2 treatment by qPCR. Asterisks represent statistical differences within cell lines
924 compared to d0. Letters represent statistical significance of shVGLL1 compared to same
925 timepoint on Scramble. **C.** Heatmaps of differentially expressed genes between shVGLL1
926 and Scramble hESCs at d0 and d4 of BMP4/IWP2 treatment by RNA-seq organized by gene
927 cluster analysis. **D.** Bright field images of Scramble and shVGLL1 hESC-derived TSCs. **E.** Cell
928 surface expression of EGFR in hESC-derived TSCs by flow cytometry. **F.** Gene expression
929 levels of TEtra factors and EGFR in hESC-derived TSCs by qPCR. Statistical analysis
930 compared to Scramble. * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$, **** $p \leq 0.0001$.

931 **Figure 3. VGLL1 regulates a TE/trophoblast program independently of BMP4 treatment.**
932 **A.** Schematic representation of the inducible CRISPR-mediated gene activation system,
933 which includes a doxycycline (DOX)-inducible dCas9-VPR and VGLL1-specific guide RNAs.
934 **B.** Bright field images of DMSO (control) and DOX-treated hESCs at day 4. **C.** Cell surface
935 expression of EGFR in DOX-treated iVGLL1a hESCs at day 4 by flow cytometry. **D.** Gene
936 expression levels of lineage markers in DOX-treated iVGLL1a hESCs at day 4 by qPCR.
937 Statistical analysis compared to DMSO. **E.** Heatmap of differentially expressed genes in
938 iVGLL1a hESCs at day 4 of DMSO or DOX treatment. * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$, ****
939 $p \leq 0.0001$

940 **Figure 4. VGLL1 regulates expression of canonical WNT effectors. A.** Heatmap of WNT
941 target genes differentially expressed in iVGLL1a hESCs at day 4 of DMSO or DOX treatment.
942 **B.** β-catenin-dependent WNT activation in DOX-treated hESC using a TCF/LEF-GFP reporter
943 plasmid. **C.** Gene expression levels of mesoderm markers in DOX-treated iVGLL1a hESCs
944 with and without IWP2 by qPCR. **D.** Gene expression levels of early and late trophoblast
945 markers in DOX-treated iVGLL1a hESCs with and without IWP2 by qPCR. **E.** Cell surface

946 expression of EGFR in DOX-treated iVGLL1a hESCs with and without IWP2 at day 4 by flow
947 cytometry. **F.** Heatmap of differentially expressed genes in iVGLL1a hESCs at day 4 of DMSO
948 or DOX treatment, with and without IWP2. * p≤0.05, ** p≤0.01, *** p≤0.001, **** p≤0.0001.

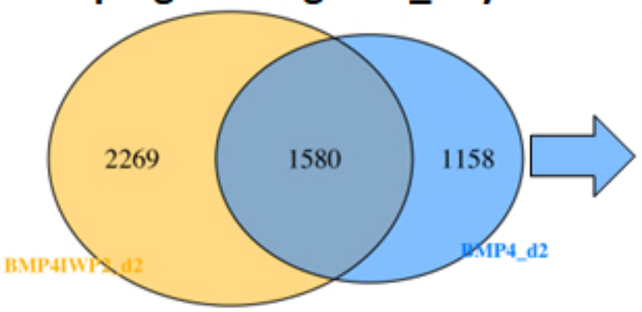
949 **Figure 5. VGLL1 directly regulates chromatin modifier gene KDM6B. A.** Venn diagram of
950 concordantly regulated genes downregulated in shVGLL1 clones and up-regulated in DOX-
951 treated iVGLL1a hESCs. **B.** Genome browser tracks showing VGLL1 and TEAD4 genomic
952 enrichment peaks (ChIP-seq) and chromatin accessibility peaks (ATAC-seq) for
953 representative TE gene loci of BMP4/IWP2-treated hESCs at day 4. **C.** Motif analysis using
954 HOMER showing the significantly enriched DNA-binding motifs at VGLL1-binding sites in
955 BMP4/IWP2-treated hESCs at day 4. **D.** Co-occupancy analysis by signal density pileups of
956 VGLL1 and TEAD4 peaks in ChIP-seq of BMP4/IWP2-treated hESCs at day 4. **E.** Genome
957 browser tracks showing VGLL1 genomic enrichment peaks (ChIP-seq) and chromatin
958 accessibility peaks (ATAC-seq) for chromatin modifiers, KDM6B and KDM4B loci of
959 BMP4/IWP2-treated hESCs at day 4. **F.** Gene expression levels of early trophoblast markers
960 after 4 days of treatment of hESCs with BMP4/IWP2 in the presence of KDM6B inhibitor
961 GSKJ4 (or DMSO as control). * p≤0.05, ns = non-significant

962 **Figure 6. VGLL1 and KDM6B are highly expressed in the polar TE of pre-implantation**
963 **blastocysts. A.** Violin plots showing expression levels of VGLL1 and CDX2 in the mural TE,
964 polar TE, epiblast, and primitive endoderm (PE) cells of d6 blastocysts (data from Zhou *et al*.
965 2019). **B.** Immunofluorescence staining of human morula and day 5 blastocyst with VGLL1,
966 TP63, and GATA3 or NANOG. **C.** Violin plots showing expression levels of KDM6B in the mural
967 TE, polar TE, epiblast, and primitive endoderm (PE) cells of d6 blastocysts (data from Zhou
968 *et al*. 2019). **D.** Immunofluorescence staining of KDM6B in week 6 placenta. CTB=
969 cytotrophoblast, STB=syncytiotrophoblast, CC=cell column

970 **Figure 7. Graphical abstract.**

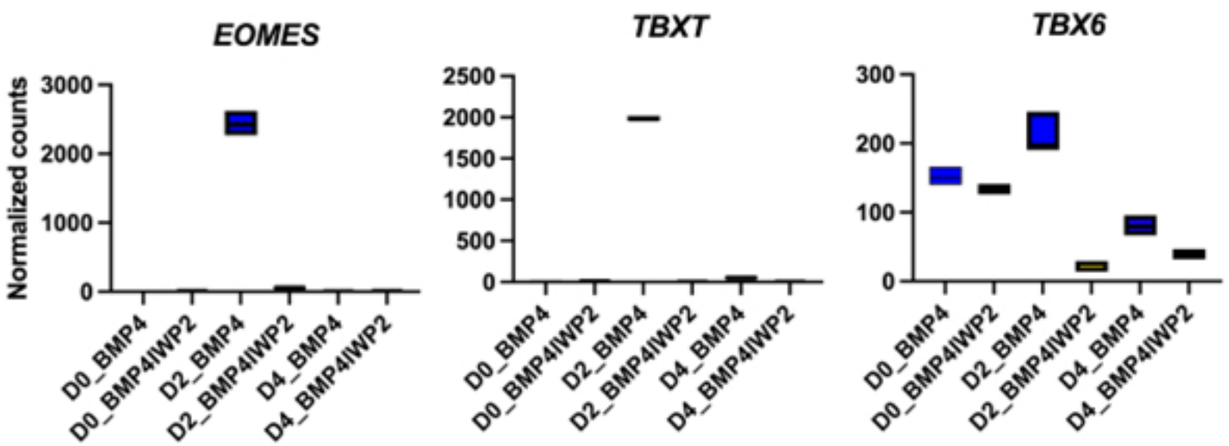
971

A

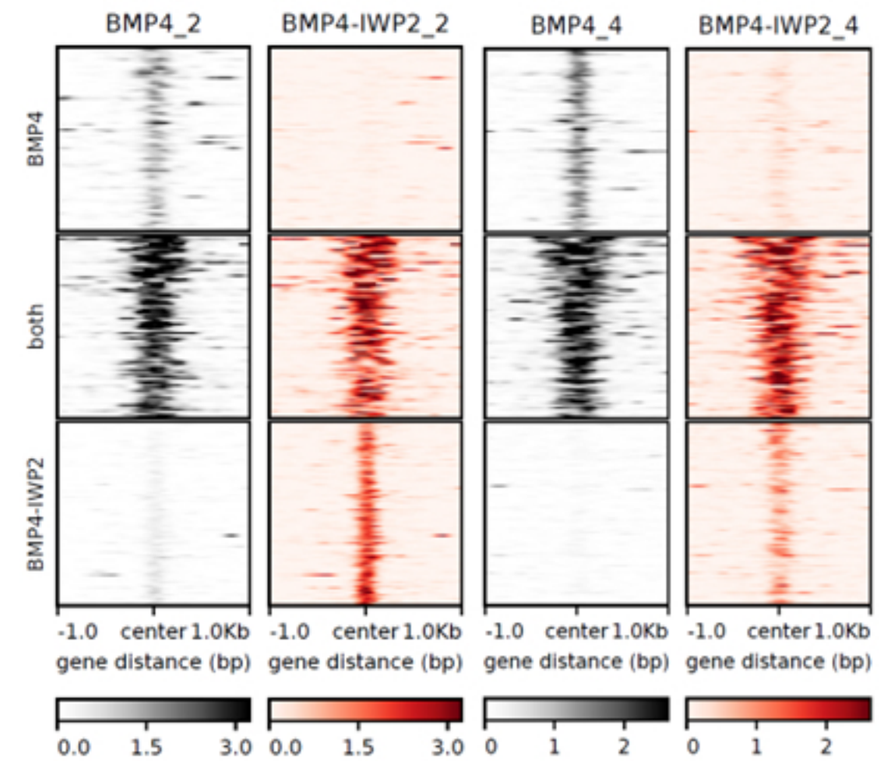


Term	Adj. P-value
Gastric cancer[a,b,c]	2.96E-04
Hippo signaling pathway	6.77E-04
Wnt signaling pathway	1.56E-03
Basal cell carcinoma[a]	1.56E-03
Breast cancer[a,b]	6.95E-03
Regulation of actin cytoskeleton	7.32E-03

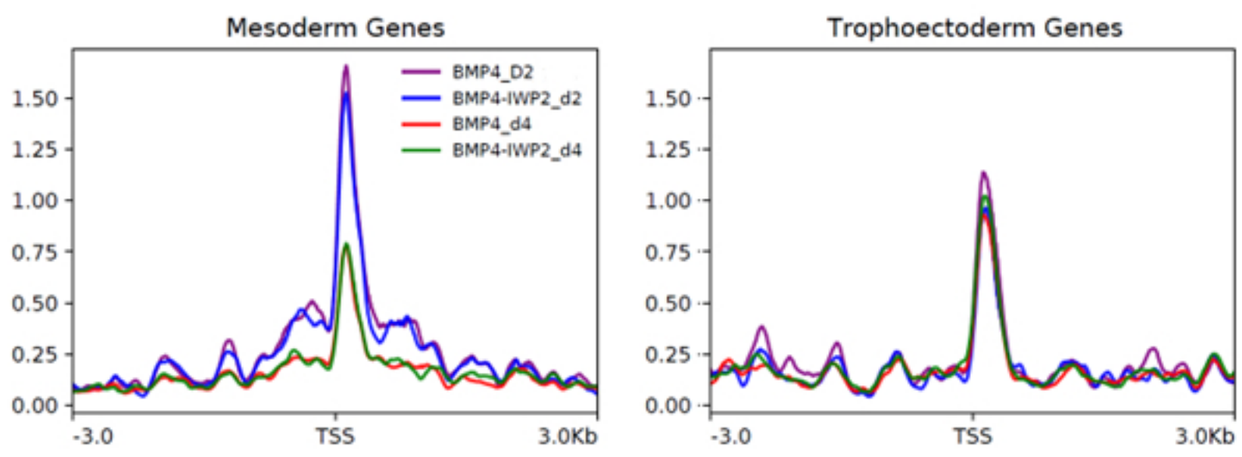
B



C



D



E

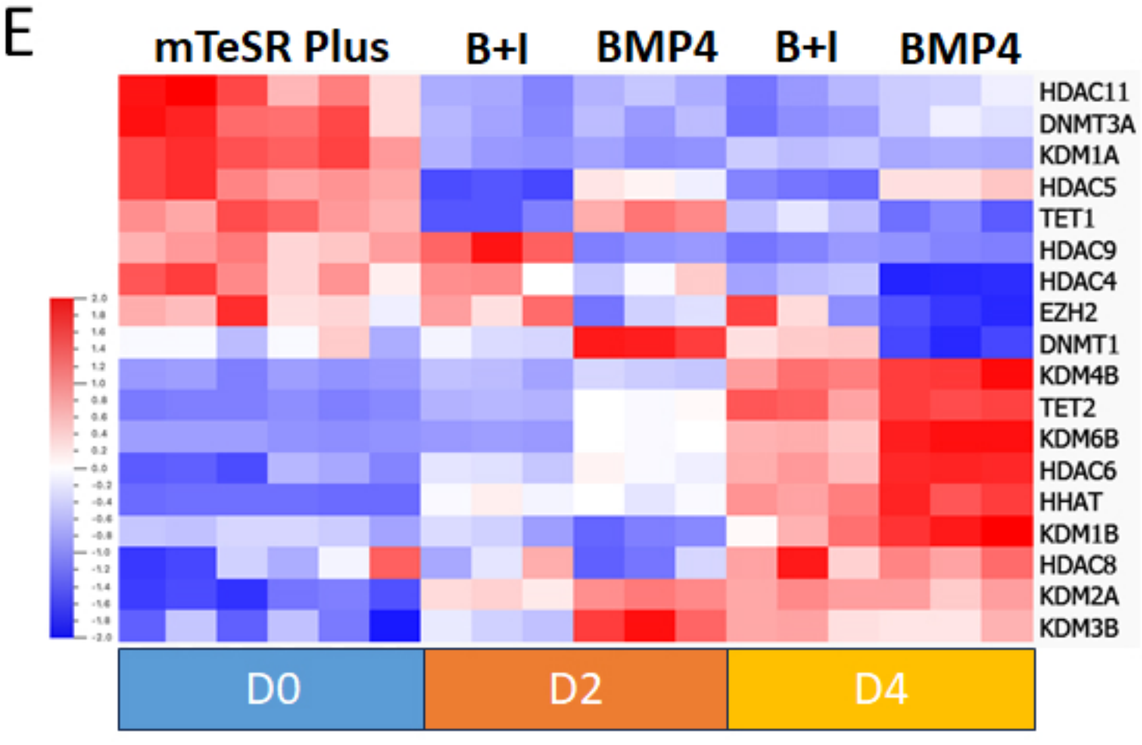


Figure 1.

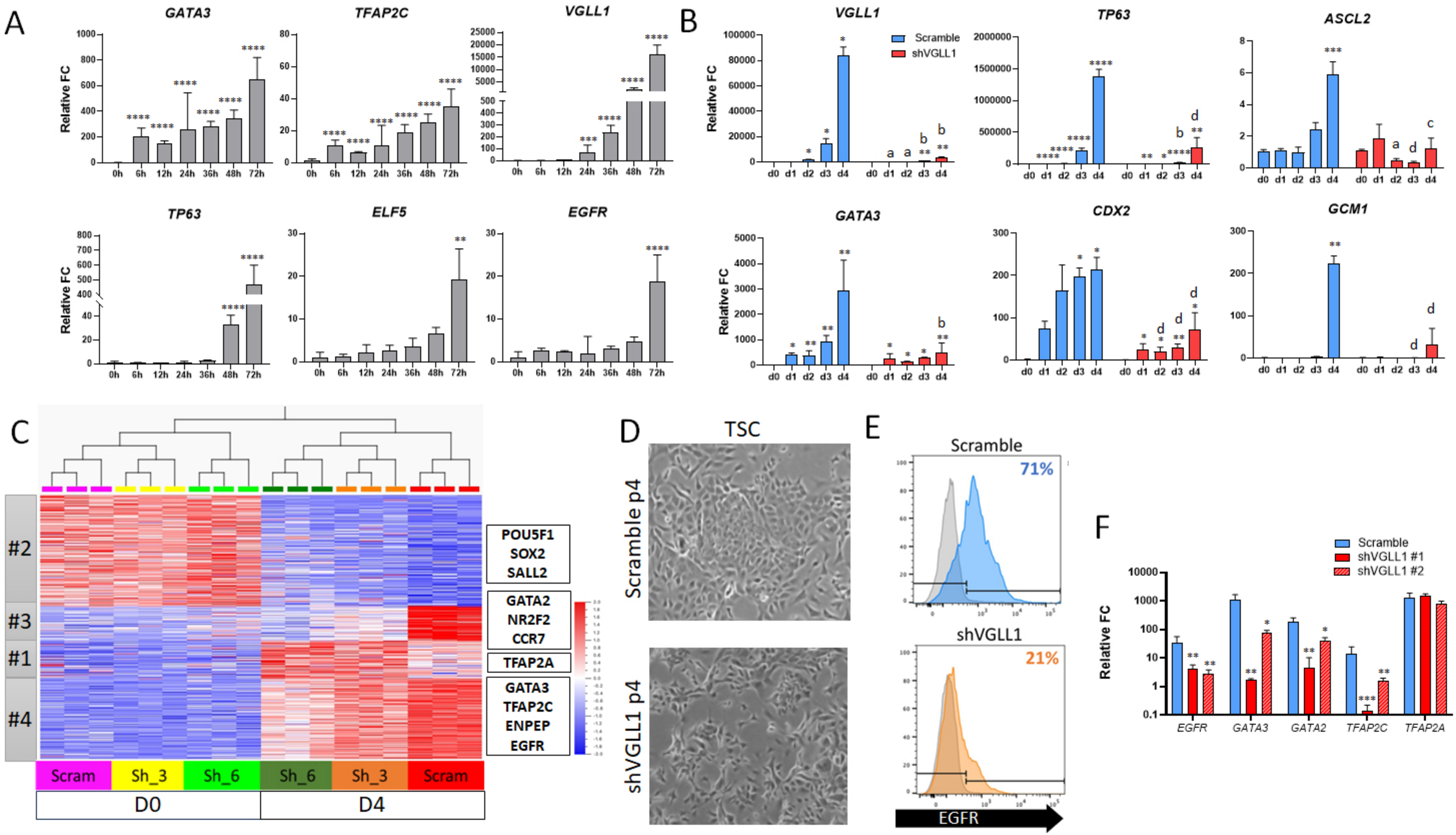

A
GATA3
TFAP2C
VGLL1
TP63
ELF5
EGFR
Relative FC
0h 6h 12h 24h 36h 48h 72h
B
VGLL1
TP63
ASCL2
GATA3
CDX2
GCM1
Scramble
shVGLL1
d0 d1 d2 d3 d4
C
#2
#3
#1
#4
POU5F1
SOX2
SALL2
GATA2
NR2F2
CCR7
TFAP2A
GATA3
TFAP2C
ENPEP
EGFR
Scram
Sh_3
Sh_6
Sh_6
Sh_3
Scram
D0
D4
D
TSC
Scramble p4
shVGLL1 p4
E
Scramble
71%
shVGLL1
21%
EGFR
F
Scramble
shVGLL1 #1
shVGLL1 #2
Relative FC
EGFR
GATA3
GATA2
TFAP2C
TFAP2A

Figure 2.

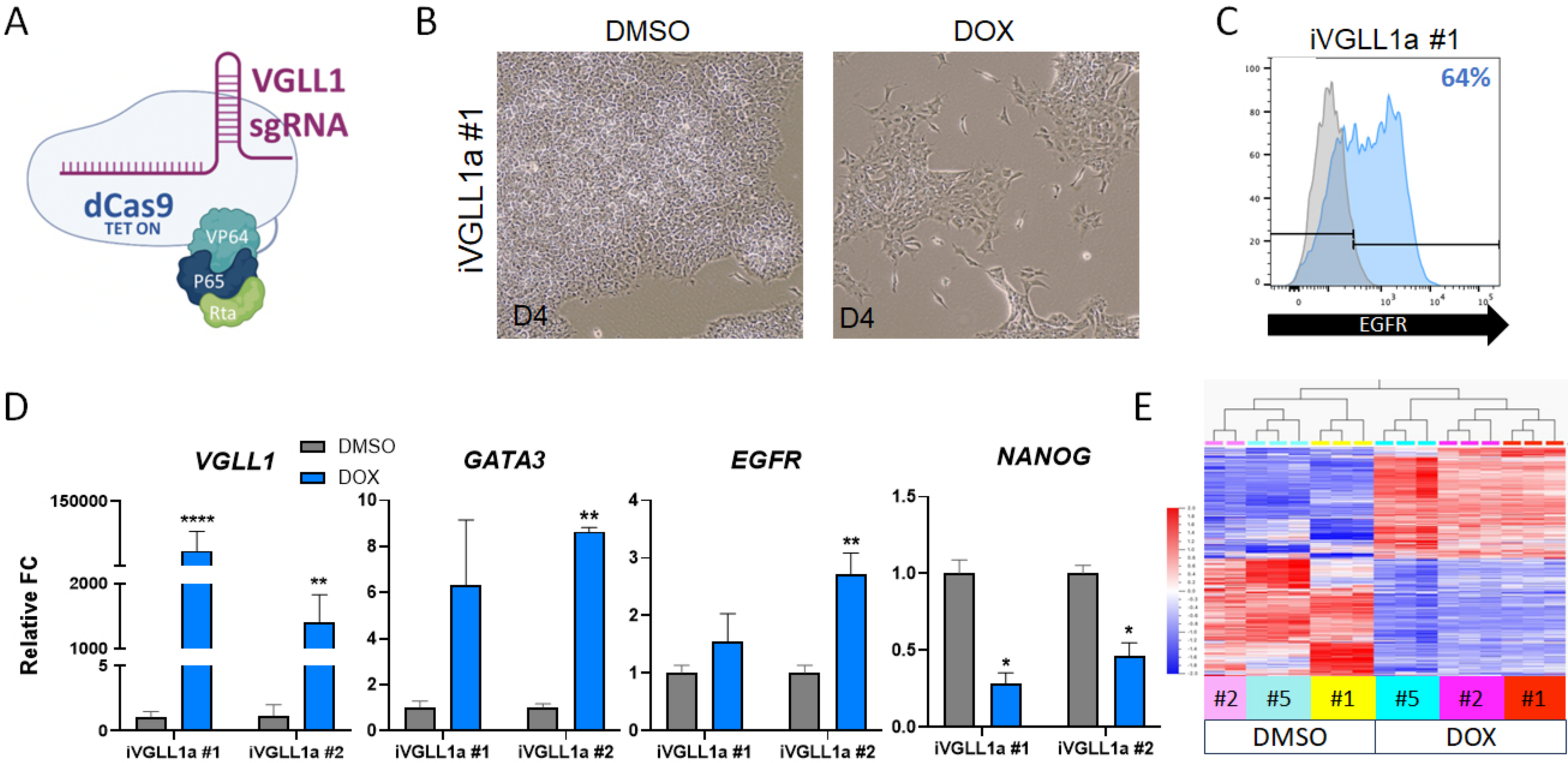

A
VGLL1
sgRNA
dCas9
TET ON
VP64
P65
Rta
B
DMSO
DOX
iVGLL1a #1
D4
D4
C
iVGLL1a #1
64%
EGFR
D
VGLL1
GATA3
EGFR
NANOG
DMSO
DOX
Relative FC
iVGLL1a #1
iVGLL1a #2
E
#2
#5
#1
#5
#2
#1
DMSO
DOX

Figure 3.

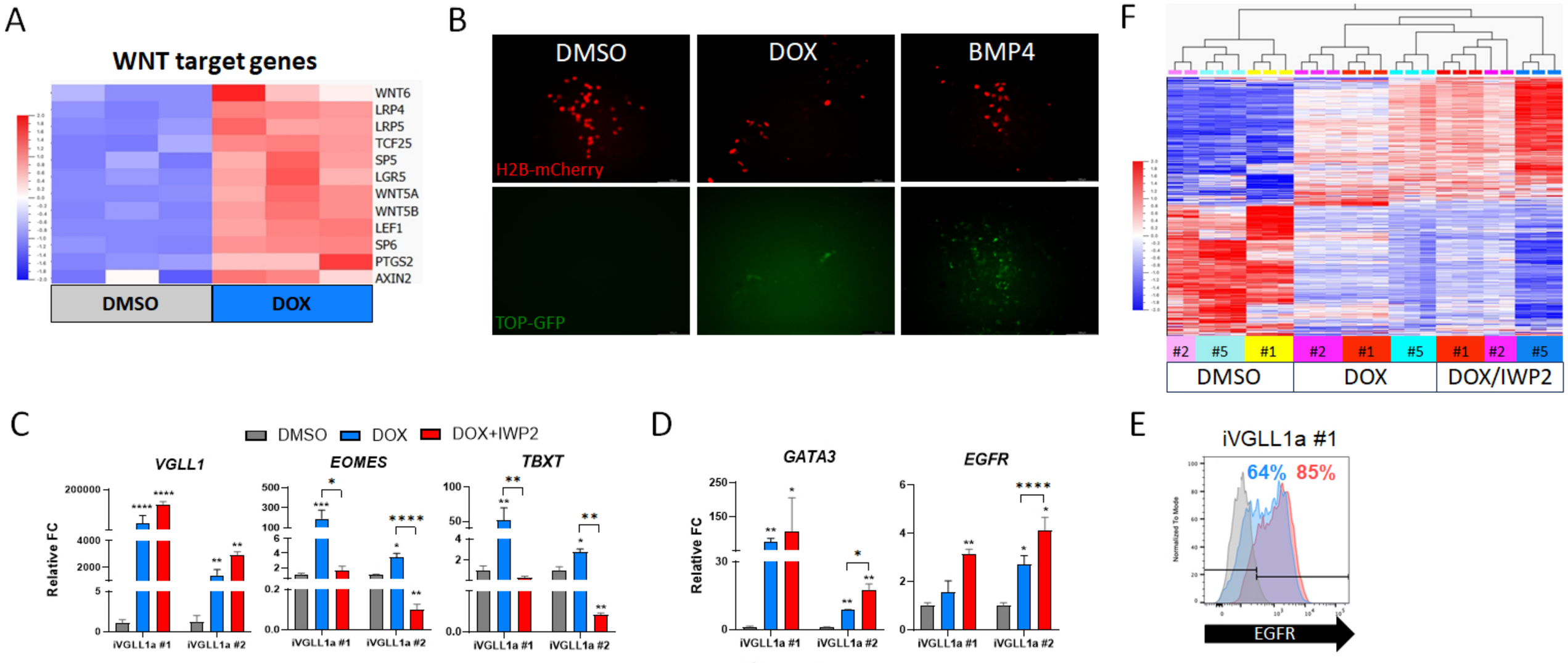

A
WNT target genes
WNT6
LRP4
LRP5
TCF25
SP5
LGR5
WNT5A
WNT5B
LEF1
SP6
PTGS2
AXIN2
DMSO
DOX
B
DMSO
DOX
BMP4
H2B-mCherry
TOP-GFP
F
#2
#5
#1
#2
#1
#5
#1
#2
#5
DMSO
DOX
DOX/IWP2
C
DMSO
DOX
DOX+IWP2
VGLL1
EOMES
TBXT
Relative FC
iVGLL1a #1
iVGLL1a #2
D
GATA3
EGFR
Relative FC
iVGLL1a #1
iVGLL1a #2
E
iVGLL1a #1
64%
85%
EGFR

Figure 4.

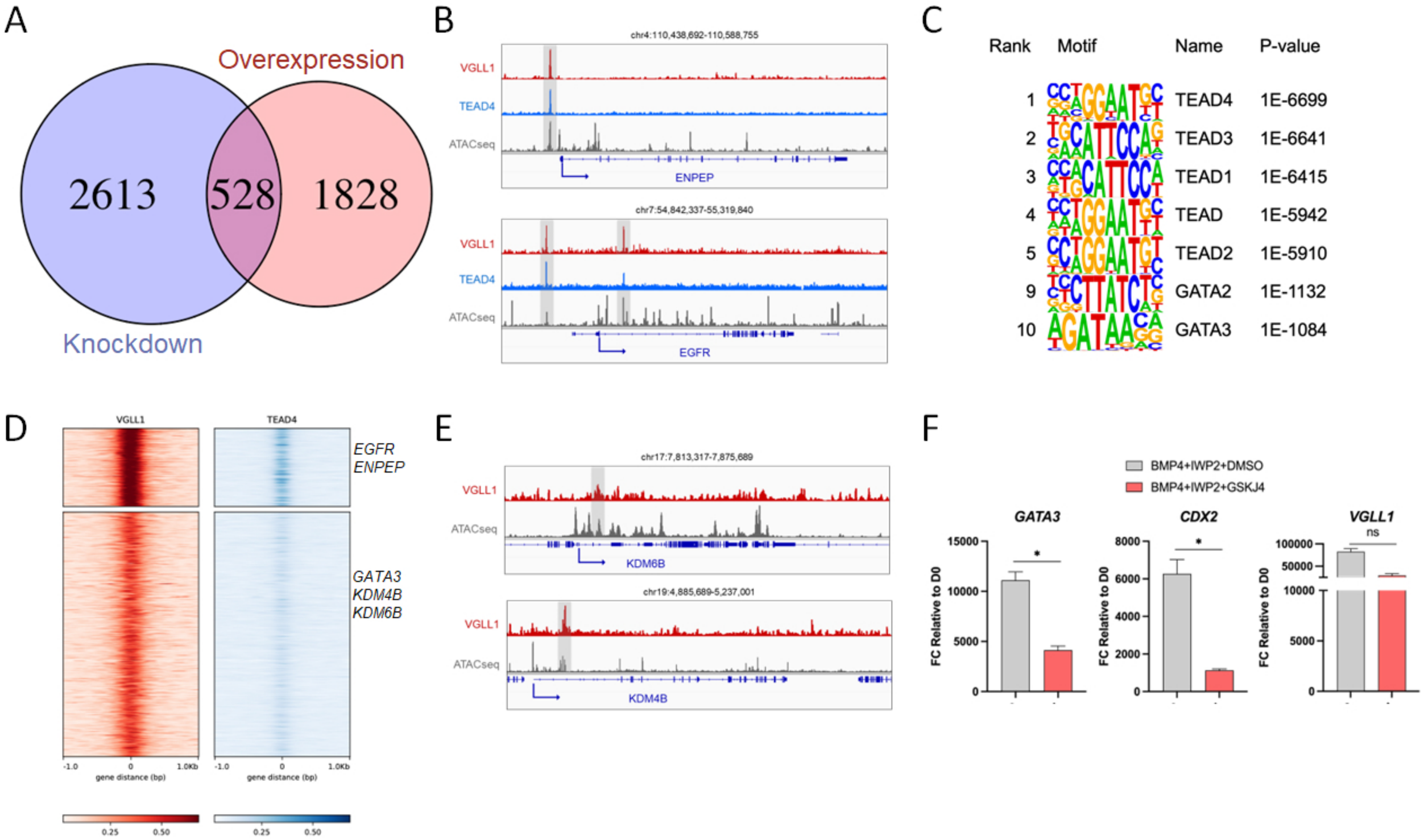

A
Overexpression
2613
528
1828
Knockdown
B
chr4:110,438,692-110,588,755
VGLL1
TEAD4
ATACseq
ENPEP
chr7:54,842,337-55,319,840
VGLL1
TEAD4
ATACseq
EGFR
C
Rank
Motif
Name
P-value
1 TEAD4 1E-6699
2 TEAD3 1E-6641
3 TEAD1 1E-6415
4 TEAD 1E-5942
5 TEAD2 1E-5910
9 GATA2 1E-1132
10 GATA3 1E-1084
D
VGLL1
TEAD4
EGFR
ENPEP
GATA3
KDM4B
KDM6B
-1.0
0
1.0Kb
gene distance (bp)
0.25
0.50
E
chr17:7,813,317-7,875,689
VGLL1
ATACseq
KDM6B
chr19:4,885,689-5,237,001
VGLL1
ATACseq
KDM4B
F
BMP4+IWP2+DMSO
BMP4+IWP2+GSKJ4
GATA3
CDX2
VGLL1
ns
FC Relative to D0
15000
10000
5000
0
8000
6000
4000
2000
100000
50000
*

Figure 5.

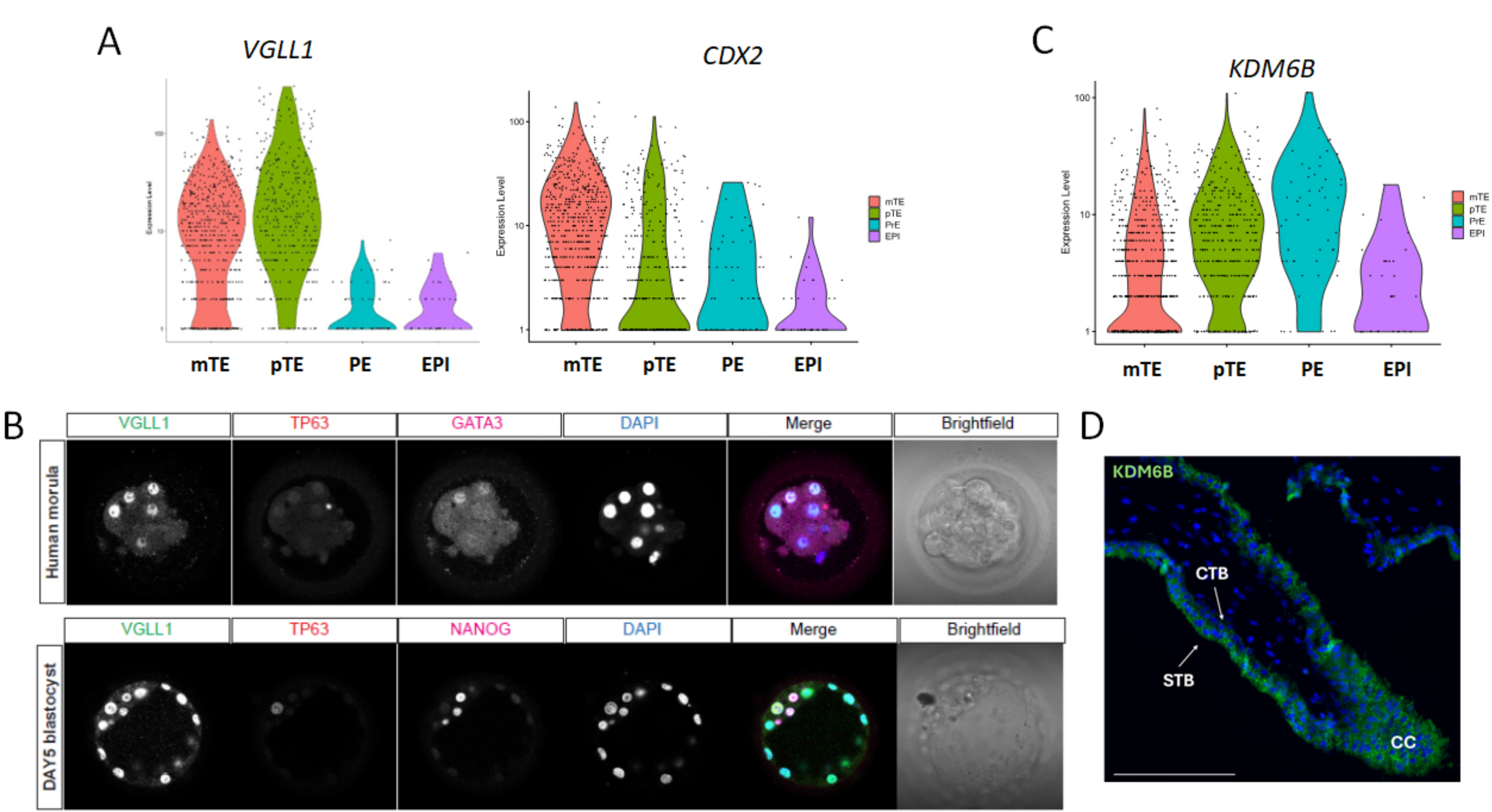

A
VGLL1
CDX2
Expression Level
100
10
1
mTE
pTE
PE
EPI
mTE
pTE
PrE
EPI
C
KDM6B
B
VGLL1
TP63
GATA3
DAPI
Merge
Brightfield
Human morula
NANOG
DAY5 blastocyst
D
KDM6B
CTB
STB
CC

Figure 6.

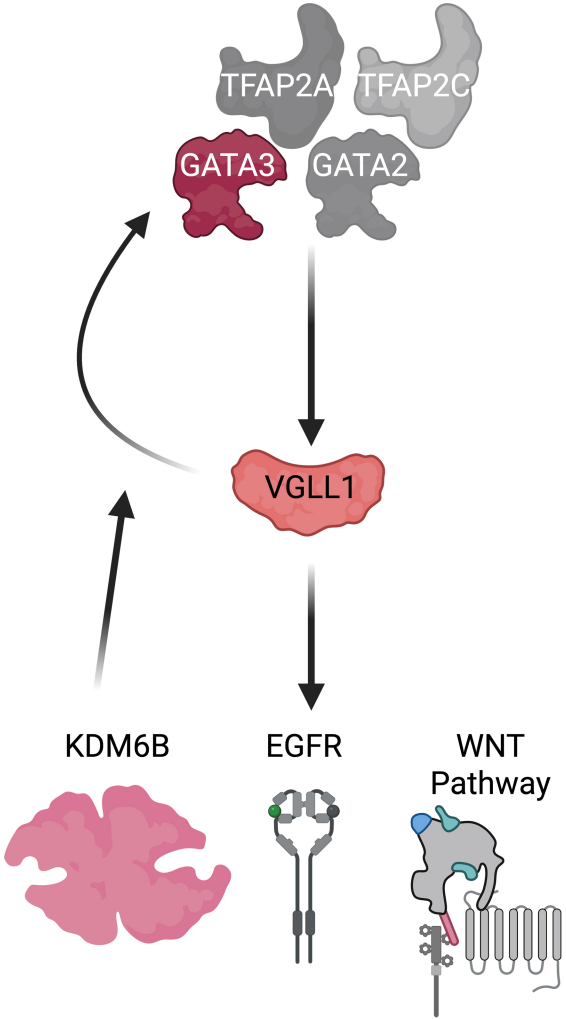


Figure 7