Comparative analysis of mouse and human placentae across gestation reveals species-specific regulators of placental development

Francesca Soncin[1,3], Marwa Khater[2,3], Cuong To[2,3], Donald Pizzo[1], Omar Farah[1,3], Anna Wakeland[1,3], Kanaga Arul Nambi Rajan[1,3], Katharine K. Nelson[1,3], Ching-Wen Chang[1,3], Matteo Moretto-Zita[1,3] , David R. Natale[2], Louise C. Laurent[2,3,*], Mana M. Parast[1,3,**]

1. Department of Pathology, University of California San Diego, La Jolla, CA, 92093, USA
2. Department of Reproductive Medicine, University of California San Diego, La Jolla, CA, 92093, USA.
3. Sanford Consortium for Regenerative Medicine, La Jolla, CA, 92037, USA

Corresponding authors: *llaurent@ucsd.edu and **mparast@ucsd.edu

Keywords: Placenta, Trophoblast stem cells, Placental Development, Cytotrophoblast, Placental Progenitors, Comparative Study

Abbreviations:

AP	Affinity propagation
BP	Basal plate
Ch	Chorion
CP	Chorionic Plate
CTB	Cytotrophoblast
DAB	3,3'-diaminobenzidine
DEGs	Differentially expressed genes
EVT	Extravillous Trophoblast
FGF	Fibroblast Growth Factor
GEO	Gene expression omnibus
GO	Gene ontology
hESC	Human embryonic stem cells
ICM	Inner cell mass
IHC	Immunohistochemistry
ISH	*In-situ* Hybridization
mRNA	messenger RNA
PBS	Phosphate-buffered saline
PCA	Principal component analysis
pCC	Proximal cell column
PFA	Paraformaldehyde
STB	Syncytiotrophoblast
TE	Trophectoderm
TF	Transcription factor
TSC	Trophoblast Stem Cell

Abstract

An increasing body of evidence points to significant spatio-temporal differences in early placental development between mouse and human, but a detailed comparison of placentae in these two species is missing. We set out to compare placentae from both species across gestation, with a focus on trophoblast progenitor markers. We found that CDX2 and ELF5, but not EOMES, are expressed in early post-implantation trophoblast subpopulations in both species. Genome-wide expression profiling of mouse and human placentae revealed clusters of genes with distinct co-expression patterns across gestation. Overall, there was a closer fit between patterns observed in the placentae when the inter-species comparison was restricted to human placentae through gestational week 16 (thus excluding term samples), suggesting that the developmental timeline in mouse runs parallel to the first half of human placental development. In addition, we identified VGLL1 as a human-specific marker of proliferative cytotrophoblast, where it is co-expressed with the transcription factor TEAD4. Since TEAD4 is involved in trophectoderm specification in the mouse, we posit a regulatory role for VGLL1 in early events during human placental development.

Introduction

The placenta is a temporary organ dedicated to supporting fetal growth during gestation by providing an interface for nutrient and gas exchange, as well as hormones essential for development. While we have some knowledge of human placental development and function, particularly from later gestational ages, its early development during the peri-implantation period is poorly understood. Abnormal trophoblast differentiation early in gestation is considered the root cause of many placenta-associated pregnancy complications, such as miscarriage, pre-eclampsia, and intra-uterine growth restriction, and clarification of these mechanisms may point to novel interventional approaches for such disorders (Cuffe et al., 2017; Norwitz, 2006; Steegers et al., 2010).

Study of early human placental development is hampered by practical and ethical issues. Therefore, both animal and *in vitro* (cell culture) models are routinely used to probe trophoblast lineage specification and function. Rodents, and in particular mice, have been the primary models used to study placental development. Importantly, both mouse and human placentae are discoid in shape and show a hemochorial gas-nutrient exchange interface. However, fundamental differences exist between rodents and humans, including gestational length, litter size, and the component trophoblast cell types and their organization within the placenta (Soncin et al., 2014). Recently, significant differences have been identified between mouse and human blastocyst-stage embryos (Blakeley et al., 2015; Deglincerti et al., 2016; Niakan and Eggan, 2013; Shahbazi et al., 2016). Specifically, while in mouse blastocysts, the inner cell mass (ICM) and trophectoderm (TE) compartments are clearly marked by mutually exclusive expression of POU5F1/OCT4 and CDX2, there is co-expression of these two markers in human TE (Niakan and Eggan, 2013). In addition, EOMES and ELF5 are absent from pre-implantation human embryos (Blakeley et al., 2015), though ELF5 has been found, along with CDX2, in a subset of cytotrophoblast (trophoblast progenitor cells) in early post-implantation placentae (Hemberger et al., 2010). Instead, GATA3 was identified as a consistent marker of trophectoderm and early cytotrophoblast (Deglincerti et al., 2016; Lee et al., 2016). These data, along with the inability to derive human trophoblast stem cells (TSC) from pre-implantation blastocyst-stage embryo (Kunath et al., 2014), suggest at least a spatio-temporal difference between early events during TE establishment and differentiation in the two species. Nevertheless, the many advantages of transgenic mouse models, including the ability to evaluate contribution of specific genes to both placental and fetal development, make the mouse

system an indispensable tool for identification of pathways involved in trophoblast lineage specification and differentiation, and placental development.

To better understand the strengths and limitations of the mouse system, we set out to compare mouse and human placental development across gestation, from early in the post-implantation period to term, with a focus on studying in detail the cytotrophoblast of the early post-implantation human placenta. We started by evaluating mouse TSC markers, CDX2, ELF5 and EOMES, in the human placenta by immunohistochemistry and/or a highly sensitive *in situ* hybridization method. These three factors regulate a transcriptional program involved in maintenance and expansion of the TSC population in mice (Senner and Hemberger, 2010), but their expression in the human placenta has not been characterized in detail. In addition, we compared the transcriptomes of placentae from these two species across gestation using genome-wide microarray-based gene expression profiling. Since human placental villi contain a continuous layer of cytotrophoblast in early gestation, which becomes discontinuous and eventually sparse in late gestation, we reasoned that differential expression analysis comparing early and late gestation placentae should identify at least some genes that are specific to this trophoblast progenitor cell type, with expression of such genes decreasing across gestation. Also, comparison of trajectories of co-expressed clusters of genes between mouse and human suggests that mouse placental development across gestation corresponds to the first half of gestation in the human placenta. In addition, although there was little overlap in the expression patterns of specific genes during this period of placental development in mouse and human, we found at least a partial overlap in enriched biological process terms associated with up and down-regulated genes over gestation. Finally, we created species-specific networks of co-regulated transcription factors, allowing us to identify common and species-specific "master" regulator genes important for early placental development. Among these, we identified and characterized a member of the vestigial like family of transcriptional co-factors as a novel human-specific cytotrophoblast marker.

Results

Stem cell markers in human placentae across gestation

We determined spatio-temporal distribution of three established mouse trophoblast stem/progenitor markers, CDX2, ELF5 and EOMES, in the human placenta across gestation. We have previously shown by immunohistochemistry (IHC) that CDX2 expression in early gestation human placentae localizes to a subset of cytotrophoblast (CTB) (Horii et al., 2016). On further inspection, we noted that within early gestation placentae (5-8 week gestational age), CDX2+ CTB were most abundant near the chorionic plate (CP), with the percentage of CDX2+ CTB decreasing as villi approach the basal plate (BP) (Fig. 1A,B i-ii); no CDX2 was detected in CTB of anchoring villi (Fig.1B iii). CDX2 expression was also absent from syncytiotrophoblast (STB, Fig. 1C i,ii) and extravillous trophoblast (EVT, Fig. 1C iii) at all gestational ages.

Because IHC using multiple antibodies against ELF5 and EOMES resulted in only non-specific cytoplasmic staining (data not shown), we turned to a novel highly sensitive *in-situ hybridization* (ISH) technique for evaluation of these two transcription factors. Using this method, we detected *ELF5* mRNA primarily in CTB and proximal cell column EVT of early gestation placentae (Fig. 2A), with lower levels detected in STB (Fig. 2A) and more mature EVT (in distal cell column of early, and basal plate of late, gestation placentae, not shown). Expression decreased dramatically in the second trimester and was restricted to the CTB compartment (Fig. 2B); by term, CTB contained only rare copies of the transcripts of this marker (not shown). Using the same method, *EOMES* mRNA expression was absent from all stages of human placenta analyzed. *EOMES*-specific probes were validated in several positive control tissues, including human tonsils and various types of human cancers, namely breast, lung, and cervical tumors (Fig. S1 and data not shown).

Genome-wide RNA profiling

To broaden the scope of our comparative study of mouse and human placentae across gestation beyond a handful of markers, we performed genome-wide microarray-based RNA profiling and compared gene expression both across time and between species, using 54 normal human placenta

samples collected between 4 and 39 weeks gestational age, and 54 mouse placenta samples collected between E7.5 and E18.5 gestational age. For both species, we used principal component analysis (PCA) to identify and remove potential outliers. After filtering the gene lists to retain transcripts with variance cutoff 0.02 in Qlucore, PCA1 showed a correlation with gestational age for both human and mouse placenta (Fig. S2.A and C). Hierarchical clustering using this minimally filtered gene list revealed that mouse placental samples formed distinct clusters according to day of gestation (Figure S2.A). For the mouse dataset, differential expression analysis comparing the samples according to day of gestation was then used to decrease the dimensionality of the dataset ($q \leq 0.05$ and $FC \geq 2$) to 2,947 differentially expressed genes (DEGs) (Fig. 3A, Table S1). Five representative DEGs were randomly selected and their differential expression confirmed using qPCR (Fig. S2.B).

The human placental samples were more variable, but also showed gestational age progression along PCA1 (Fig. S2.C). However, hierarchical clustering did not separate these samples clearly according to the week of gestation (Fig. S2.C). Term samples had a well-defined signature, as expected by the large temporal difference compared to the other samples. The rest of the samples displayed a general correlation with gestational age, with the earliest samples on the left and later samples on the right, but with frequent overlap in the distributions of samples from different timepoints. Therefore, we used hierarchical clustering to separate the samples into groups according to their gene expression patterns, rather than the clinically-determined gestational age. The samples from weeks 8 and 12 showed a broad distribution in the PCA plot, and were therefore removed, along with previously identified outliers, narrowing down the sample set to 31 samples. The hierarchical clustering identified six groups (Fig. 3B): groups A and B included samples from week 4-5 and week 6-7, respectively; groups C and D each contained a combination of samples from week 9 to 11; and finally groups E and F represented second trimester and term samples, respectively. Differential expression analysis between these sample groups identified 1,195 human placenta DEGs (Fig. 3B, Table S1). Five representative DEGs were randomly selected and their differential expression confirmed using qPCR (Fig. S2.D).

Comparing the list of DEGs between the two species, we found 517 genes to be in common, representing only 18% of mouse and 43% of human DEGs (Fig. 3C). For each species, the DEGs displayed two main expression patterns across gestation, up-regulation and down-regulation, comprising 35% and 48% of the common DEGs, respectively, with the remaining genes (approximately 20%) showing other expression patterns (Fig. 3D, center pie chart). We subsequently investigated the common DEGs for direction of expression for the two main expression patterns, down and up-regulation across gestation

(Fig. S3.A-B). In this analysis, 129 genes showed parallel up-regulation, and 157 showed parallel down-regulation across gestation in both species, with the remaining 231 having different patterns of expression (including 48 with completely opposing expression patterns) in mouse and human (Fig. 3D, side pie charts). To confirm that these observations reflected a true difference between mouse and human, and were not technical artifacts, we compared the DEGs from our mouse analysis with a similar but smaller dataset from Knox and Baker (2008). Applying the same differential expression thresholds, about 40% of DEGs were in common between the two mouse studies, of which over 96% showed comparable expression patterns across gestation (Fig. S3.C-D). Moreover, the microarray data from human placentae confirmed CDX2 and ELF5 down-regulation across gestation seen in the IHC and ISH experiments, while the probe for EOMES did not pass the minimum signal threshold to be included in the analysis (Fig. S3.E).

Next, we identified the enriched gene ontology (GO) terms in the Biological Process tree in the commonly up- (red) and down- (green) regulated genesets (Fig. 3E). Genes down-regulated with gestation were enriched for regulation of cell cycle and various DNA and cellular metabolic pathways. In contrast, genes up-regulated with gestation were enriched for functions associated with cell differentiation, maturation and specialisation. When we analysed the GO terms associated with the species-specific differentially expressed genesets, we found modest overlaps in the enriched GO biological process terms (about 30%), despite the absence of common genes in the mouse and human lists (Fig. S4). Down-regulated genes (Fig. S4.A) were associated with cell cycle and protein synthesis, while up-regulated genes (Fig. S4.B) represented cell migration and cell/organ maturation, with a strong focus on blood vessel development and cell signaling activation/regulation.

To start unravelling the complexity of the biology during placental development across gestation we applied the Affinity Propagation (AP) algorithm to each species dataset. For the mouse dataset, AP clustering of the 2,947 DEGs yielded 7 clusters of co-expressed genes (Fig. 4A, Table S1), showing specific patterns of expression across gestation. Three clusters (#2, #3 and #7) showed a trend of down-regulation across gestation, albeit with differences in the precise timing of decrease in gene expression. In contrast, clusters #1, #4 and #5 showed up-regulation with increasing gestational age. One cluster, #6, had a distinct pattern, with the average expression pattern showing up-regulation between E9.5 and E12.5, with peak expression around E11.5, and lower expression at both extremes of gestational age. We hypothesized that genes representing trophoblast stem/progenitor cells would be highly represented in clusters showing an overall trend of down-regulation across gestation. To this end, we

differentiated mTSC for 7 days and subjected undifferentiated (day 0/d0) and day 7/d7 differentiated samples to genome-wide microarray-based RNA profiling (Arul-nambi-rajan et al., 2017). We then selected the DEGs between these two sample sets (v0.02, two-group analysis q0.01 and FC>2.0) and identified the AP clusters from the placental dataset in which they appeared. We found 2,966 DEGs between mTSC d0 and d7 (Table S1), of which 1,369 genes also appeared to be differentially expressed across gestation in the mouse placental dataset. Of the latter, 765 DEGs were up-regulated in d0 TSCs; these genes were enriched (65% vs 39%) in AP clusters #2/3 (genes with high expression in early gestation placentae) and depleted (24% vs 41%) from AP clusters #1/4 (genes with high expression in late gestation placentae) (Fig. 4B). The remaining 604 DEGs were upregulated in d7-differentiated TSCs; these DEGs were enriched (52% vs 41%) in AP clusters #1/4 (genes with high expression in late gestation placentae) and depleted (21% vs 39%) from AP clusters #2/3 (genes with high expression in early gestation placentae) (Fig. 4B). These data provide support for our hypothesis that, in fact, early gestation placentae are enriched in undifferentiated TSC-associated genes.

AP analysis on the 1,195 DEGs of the human dataset yielded 6 clusters of co-expressed genes (Fig. 4C, Table S1). Clusters #5 and #6 both showed a trend of down-regulation across gestation, though with differences in the precise pattern of decrease of gene expression. In contrast, clusters #1 through 4 generally showed up-regulation with increasing gestational age, albeit with some differences, including a subsequent decrease in expression at term in clusters #2 and 4 (Fig. 4C). Clusters #2 and 4 were further distinguished based on the gestational age at which the average expression pattern began to increase: this occurred early in cluster #4 (between wk 4-5 and wk 6-7), and later in cluster #2 (between wk 9-11 and wk 14-16). We next compared these whole placental profiles with gene expression data from cytotrophoblast (CTB) isolated from different gestational age placentae (Fig. S5.A-B). Unsupervised PCA and hierarchical clustering separated the CTB samples into two groups according to gestational age: those from weeks 8 and 10 (which we termed early CTB) and those from weeks 12 and up, including weeks 18, 20 and 39 (which we termed mature CTB) (Fig. S5.A-B). We applied a two-group analysis on Qlucore (q0.01, FC≥2) and identified 1,120 DEGs (Fig. S5.C, Table S1). Of these, 197 were also differentially expressed in human placental samples across gestation (Fig. 4D). Interestingly, we observed an enrichment (56% vs. 34%) of early CTB markers in human AP cluster #5, which showed a pattern of gene expression highest in early gestation placentae, and a relative deficit (20% vs. 36%) of these markers in human AP clusters #1 and 2, which showed a pattern of gene expression highest in late gestation placentae (Fig. 4D). Again, these data suggest that, as with profiling of mouse placentae and differentiating TSC, genes enriched in early human placentae include early CTB-associated genes.

Moreover, it appears that the decrease in expression of early CTB/TSC in placental samples is due not only to the decrease in the fraction of placental cells represented by these cell types, but also because the gene expression profiles of these cells are changing over gestation.

As the datasets from both species spanned the entire gestation, we wanted to see if we could align and correlate specific gestational periods between mouse and human. To this end, we fit non-linear curves to each of the AP clusters and computed the Euclidean distance between the curves for each mouse and each human AP cluster to identify the closest fitting human curve for each mouse curve. The "E.d." values shown are the sum of Euclidean distances measured across 1000 points for each pair of curves. Because of the difference in timeframe between mouse and human gestations and a large gap in our human dataset (from week 16 to term), we performed this analysis both including the term samples ("Human Long (HL)", Fig. 4C, Fig. S6.A, Fig. 5A) and without them ("Human Short (HS)", Fig S5.D, Fig. S6.B, Fig. 5A). Curves for mouse clusters 1, 2, 5, and 6 correlated better (smaller Euclidean distance) with the HS dataset (Fig. 5A). For clusters in which the mouse data showed closer correlation with the HL dataset (mouse clusters 3, 4, and 7), the differences between the Euclidean distances for HL and HS were generally less marked. These results show that the transcriptional patterns observed during mouse placental development more closely resemble those seen during the first half of human placental development, suggesting that human placental development possesses a long maturation phase that is not seen in the mouse. Consistent with the low percentage of overlapping genes in the mouse and human lists of DEGs, the mouse and human clusters with the closest patterns of expression across gestation did not display higher numbers of overlapping genes (Suppl. Fig. S6A-B). When functional enrichment analysis of the genes in each cluster was performed (Table S2), we did observe similarities in the enriched Gene Ontology categories for two out of three pairs of clusters with the closest trajectories (Fig. 5B): mouse cluster 1 and HS cluster 2, which monotonically increase over the course of gestation, are both enriched for genes associated with vascular development and cell migration; and mouse cluster 2 and HS cluster 5, which monotonically decrease over the course of gestation, are both enriched for genes associated with RNA metabolism.

To focus our analysis on the identification of candidate genes with potential regulatory roles in early gestation, we decided to narrow our analysis on transcription factors (TFs). 258 TF were differentially expressed in the mouse dataset (Table S1) and were organized into 6 clusters of co-expressed genes, with the loss of cluster 7 from the AP analysis on all DEGs (Fig. S7A, compared to Fig. 4A). The expression patterns for Clusters #2 and #3 were positively correlated with each other, and

negatively correlated with #1, #4 and #5 (Fig. S7.B). The average expression of the TFs in Cluster #2 was most strongly anti-correlated with Cluster #1, and Cluster #3 was most strongly anti-correlated with Cluster #4 (Fig. S7.B). Cluster #6, which showed up-regulation between E9.5 and E12.5, was not correlated or anti-correlated with any other TF cluster. Many genes known to be involved in establishment and/or maintenance of TSC, including Elf5, Gata2, and Arid3a, were found in clusters #2 and #3, with downregulation of expression with increasing gestational age (Fig S7.A). Surprisingly, however, two TSC-associated genes, Tead4 (Cluster #1) and Cdx2 (Cluster #4), displayed the opposite expression pattern, increasing with gestational age. To validate these latter findings, we performed ISH on mouse placenta samples across gestation, using probes specific to Tead4 (Fig. 6A) or Cdx2 (Fig. 6B). We noted that, consistent with the microarray results, the expression of these transcripts as determined by ISH is significantly elevated later in gestation, with Tead4 expressed through all layers of the mouse placenta (Fig. 6A) and Cdx2 expressed specifically in the junctional zone in later gestation placentae, particularly in PAS-positive glycogen cells (Fig. 6B).

In the human dataset, 115 TFs were differentially expressed (Fig. S7.C, Table S1), and were organized into 5 clusters of co-expressed genes, with the loss of cluster 6 from the AP analysis on all DEGs (Fig. S7.C, compare to Fig. 4C). Human TF cluster #4, similar to the mouse cluster #6, included TFs with highest expression around mid-gestation. Human TF clusters #1, 2, and 3 were anti-correlated to TF cluster #5 (Fig. S7.D); the latter is enriched in genes up-regulated in early CTB and included many known progenitor markers. Since mouse TF clusters #2 and 3 and human TF cluster #5 all were enriched in trophoblast stem/progenitor markers, we compared the gene lists in these clusters. In both species, these clusters included well-known progenitor markers, such as Elf5/ELF5, Arid3a/ARID3A and Gata2/GATA2. However, there were also large numbers of species-specific TFs, which might account for the differences between mouse and human placental development. In particular, we noticed the presence of VGLL1, a homolog of the Drosophila Vestigial factor, in human Cluster #5, but not in any mouse cluster. Recent studies have suggested an interaction between VGLL1 and TEAD4 (also present in human Cluster #5, Fig. S7.D), creating a protein complex potentially able to promote the transcription of specific downstream targets (Pobbati et al., 2012). Given the role of TEAD4 in trophoblast lineage specification in the mouse (Nishioka et al., 2008; Yagi et al., 2007), we considered that VGLL1 may be an interesting regulatory gene candidate as a human-specific early CTB marker. Knowledge regarding the function of VGLL1 during development is extremely limited; therefore, we decided to begin by evaluating the localization and expression patterns of VGLL1/Vgll1 in the placentae of both species across gestation.

VGLL1 is a human-specific CTB marker

In human placentae, IHC revealed high expression of VGLL1 in villous CTB and proximal cell column (pCC) trophoblast in early first trimester samples (week 6, Fig. 7A); there was overlap with TEAD4 expression in villous CTB, but not in pCC trophoblast (Fig. 7A). Moreover, cells expressing VGLL1 co-expressed PCNA, a marker of cell proliferation (Fig. S8.A). The expression of both VGLL1 and TEAD4 was maintained in CTB across gestation (Fig. 7B-C); but unlike TEAD4, VGLL1 was also expressed in mature EVT at the basal plate of term placentae (Fig. 7C). To confirm specificity of VGLL1 antibody, we performed ISH using VGLL1-specific probes, and identified mRNA expression in the same cells in the early human placenta (Fig. S8.B). In contrast, in mouse placentae, we could not detect any Vgll1 mRNA by ISH at any gestational age (Fig. S9.A and data not shown). Human VGLL1 expression in CTB was further confirmed by western blot, with isolated primary CTB showing high levels of VGLL1 protein, with decreasing expression upon differentiation of these cells *in vitro* (Fig. 8A). In contrast, no Vgll1 protein was expressed in either undifferentiated or differentiated mTSC (Fig. S9B), suggesting a species-specific role for VGLL1 in human, but not mouse, trophoblast.

To further probe the role of VGLL1 in human early CTB, we used a previously-established BMP4-based two-step protocol for differentiation of human embryonic stem cells (hESC) into CTB-like cells (step 1), and then into mature terminally differentiated EVT- and STB-like cells (step 2) (Horii et al., 2016). We evaluated *VGLL1* mRNA expression and found it to be upregulated during the first step of this protocol, along with downregulation of *POU5F1/OCT4* expression and upregulation of other known human early CTB markers, *TP63* and *CDX2* (Fig. 8B); *TEAD4* expression did not change significantly, as previously shown in similar BMP4-treated hESC (Home et al. 2012) (Fig. 8B). VGLL1 protein expression was confirmed in these hESC-derived CTB-like cells by western blot (Fig. 8C). Immunofluorescence staining confirmed lack of VGLL1 in undifferentiated OCT4+ hESC and positive expression in nuclei of KRT7+ hESC-derived CTB-like cells at day 4 (Fig. 8D and Fig. S10.A-C). hESC-derived CTB-like cells also showed nuclear expression of TEAD4, as previously reported (Home et al., 2012), co-localizing with VGLL1 (Fig. S10.C). We also generated hESC lines, using a mix of 5 shRNA constructs targeting VGLL1, and noted that VGLL1 knockdown resulted in blunted expression of the CTB marker TP63 in hESC-derived CTB, without affecting down-regulation of POU5F1/OCT4 (Fig. 8E).

Discussion

Due to both practical and ethical issues, early placental development is often studied in non-human animal models, in particular, the mouse. However, an increasing body of evidence has identified significant species-specific differences in early embryonic development, including in specification and differentiation of the trophectoderm (TE), between mouse and human. Here, we performed a comprehensive comparative study of mouse and human placental development, with a particular focus on defining expression of known mouse trophoblast stem/progenitor markers in the human placenta, and identification of novel human-specific early CTB markers. We began our study by evaluating expression and localization of three key markers of mouse TSC: Cdx2, which is required for TE specification (Strumpf et al., 2005), and Eomes and Elf5, which reinforce and maintain the TSC state (Donnison et al., 2005; Ng et al., 2008; Russ et al., 2000). Consistent with a prior study reporting co-expression of CDX2 and ELF5 proteins in CTB of early gestation human placenta (Hemberger et al., 2010), we also noted that CDX2 was particularly abundant in CTB in early gestation placentae, with a higher percentage of CDX2+ CTB close to the chorionic plate (fetal surface). However, we were not able to detect specific nuclear staining with multiple commercially available antibodies against ELF5, including the one used in a previous publication (Hemberger et al., 2010) (data not shown). Using ISH, we found that *ELF5* mRNA was expressed in both villous CTB and trophoblasts of the proximal column, both of which are proliferative trophoblast (Lee et al., 2007).

Unlike CDX2 and ELF5, however, EOMES expression was conspicuously absent from all human placenta samples tested. EOMES expression has been documented by immunostaining in early human blastocysts (Zdravkovic et al., 2015) and putative trophoblast stem cells (Genbacev et al., 2011); however, cytoplasmic staining patterns noted in these cells raise questions regarding specificity of these antibodies. Our data are consistent with a recent study evaluating gene expression in the pre-implantation human embryo by single cell RNAseq, showing a lack of EOMES expression in human TE (Blakeley et al., 2015). Based on co-expression of ELF5 and CDX2 in a subset of CTB, we concur with Hemberger and colleagues (Hemberger et al., 2010) that it is highly likely that early post-implantation human placenta harbors TSCs. The lack of EOMES expression, however, indicates that establishment and maintenance of TSCs in the human placenta requires other factors.

Examining individual markers in this fashion is time-consuming and labor intensive. We therefore decided to compare mouse and human placental gene expression across gestation using more comprehensive microarray-based gene expression technology. For both species, we started collection at the earliest post-implantation timepoint we could confidently sample: gestational week 4 for human and E7.5 for mouse. While our mouse data spanned most of the post-implantation gestational period in a uniform manner, our human data were bimodally distributed, with dense coverage of the first trimester and the late third trimester. The reasons for this species-specific difference are twofold: one is related to our overall questions, as we were interested in characterizing early gestation human CTB. The second is due to practical reasons, as acquisition of "normal" tissues from late second and early third trimesters is difficult at best.

One challenge encountered in cross-species comparative studies is from integration of species-specific datasets. Although we took particular attention to select results from the most robust and representative probe, when multiple probes for a given transcript were available, this selection process might have introduced bias into the analysis. Moreover, discrepancy between mouse and human gene orthologs further complicated the analysis. However, the design of our analysis, in which we first identified DEGs within each species across gestational age, allowed us to avoid many of the potential artifacts that might arise from the use of different probes for the mouse and human transcripts. The large dataset gathered in this study represents an invaluable resource for both mouse and human placentologists and developmental biologists to evaluate gene expression in this important transient organ across gestation and probe the role of specific genes and pathways during its development. In our overall inter-species comparison of gestational periods, we noted a closer fit between the patterns observed in the mouse and human placentae when the comparison was restricted to mouse placentae from E7.5-18.5 and human placentae from weeks 4-16, rather than including human term placentae, suggesting that the developmental timeline in mouse runs parallel to the first half of human development. In addition, our data support the concept that although general biological processes and regulatory programs associated with different stages of placental development partially overlap between mouse and human, there are distinct species-specific regulatory interactions underlying this process. This is supported by limited overlap between the two species on the individual transcript level, both in the sets of differentially expressed transcripts in early compared to late gestation, and in clusters of co-expressed genes that show similar patterns of expression across gestation in both species. In some cases (e.g. Tead4/TEAD4), the same factor may display different gestational age-specific expression patterns in mouse and human placentae, while in other cases (e.g. Eomes in mouse and VGLL1 in

human), a factor appears to play a role in one species but not the other. Interestingly, our data from human placentae showed significant changes in gene expression even within the first trimester, reflecting rapid development of the placenta in this timeframe. The most rapid changes occurred from week 4 to week 8, with subsequent slowing of the rate of change. In comparison, the mouse placenta showed changes throughout the entire gestation and lack of a long maturational plateau present in human placental development. Although we cannot completely rule out that this observation might be biased due to sparser human samples at later gestational ages, as described above, it is consistent with rapid changes in placental morphology and size in the first trimester. We thus suggest that future studies using first trimester human placental samples specify the week of gestation.

For this study, we focused our analysis on genes that display a progressive decrease in expression according to gestational age in human and mouse placentae. Since trophoblast progenitor cells (CTB in the human placenta and TSC in the mouse) are known to be more abundant in early gestation placentae (Lee et al., 2007; Natale et al., 2017), at least some of the genes enriched in early gestation placentae would thus likely correspond to these cell populations. To identify genes that potentially maintain such cell populations, we further focused our attention on transcription factors (TFs), which are known to orchestrate important cellular functions by regulating gene expression. While some TFs known to be involved in early mouse TSC specification/maintenance were enriched within early gestation placental tissues (including Elf5, Gata2, Arid3a), two TFs, Cdx2 and Tead4, showed a surprisingly different pattern, with increased expression in later gestation placentae; we confirmed these data by ISH. While Tead4 is known to be involved in trophoblast lineage specification in the pre-implantation mouse embryo (Nishioka et al., 2008; Yagi et al., 2007), studies from Jacquemin and colleagues have reported expression of Tead4 in the labyrinth (Jacquemin et al., 1996; Jacquemin et al., 1998). With the sensitive ISH technique used in this study, we observed localization in both the labyrinth and junctional zone of the mouse placenta. Similarly, Cdx2 was found to be expressed not just in extraembryonic ectoderm and primitive streak in the early embryo, as previously shown (Beck et al., 1995; Strumpf et al., 2005), but also in the junctional zone of the later gestation mouse placenta. Report of this expression date back to Beck and colleagues in 1995, but we were able to confirm localization of Cdx2 to PAS-positive glycogen cells. Further studies are required to determine the role of both of these TFs in the late gestation mouse placenta.

Similar analysis of human placenta across gestation identified ELF5, TEAD4, GATA2, and ARID3A to be enriched in early gestation tissues, while neither EOMES nor CDX2 were identified in any specific clusters due to very low/undetectable mRNA levels. While the former correlates to our ISH data, the latter is likely due to random sampling of placental tissues, without regard to proximity to the chorionic plate. This highlights one of the limitations of this study, at least with the human placenta, as we investigated gene expression in a random sampling of villous tissue, containing a mixed cell population, with stromal tissues as well as trophoblasts. Single-cell transcriptome analysis might overcome this limitation in the future, although common issues associated with these techniques, such as sampling and sample size, will need to be addressed. Yet another known early human CTB factor not identified as a DEG in our dataset was GATA3 (Deglincerti et al., 2016; Lee et al., 2016). In fact, in complex with the AP-2 transcription factors, TFAP2A/C, GATA3 was recently identified as an early repressor of pluripotency/inducer of trophectoderm during BMP4-induced trophoblast differentiation of human embryonic stem cells (Krendl et al., 2017). We performed ISH on human placenta samples across gestation with a GATA3-specific probes and found expression of this marker in all trophoblast subtypes, including syncytiotrophoblast and extravillous trophoblast (data not shown); this is consistent with previous immunolocalization studies of this protein in gestational tissues (Banet et al., 2015). This finding highlights yet another limitation of our study, where factors that are not unique to CTB, yet serve an important function in these cells, may have escaped identification.

In the TF cluster enriched in early gestation human placentae, among putative trophoblast progenitor markers including ELF5, ARID3A (Rhee et al., 2017), and TEAD4, we identified VGLL1, a homolog of Drosophila vestigial gene, as uniquely expressed in human (and not the mouse) placenta. Very little is known about VGLL1 function; however, it is known to bind TEAD proteins through its vestigial homology domain (Vaudin et al., 1999). Interestingly, structural analysis of the complex has shown that VGLL1 binds to the same pocket in TEAD4 to which YAP/TAZ proteins bind (Pobbati et al., 2012). Since the combination of Tead4 and Yap occurs in mouse TE, leading to induction of Cdx2 (Home et al., 2012; Nishioka et al., 2009), we hypothesized that VGLL1 may be involved in similar events, but specific to the human placenta. We showed that VGLL1 protein was specifically expressed in CTB and proximal cell column trophoblast, where proliferative trophoblast reside in the early human placenta. However, VGLL1 co-localized with TEAD4 only in CTB. Moreover, we observed VGLL1 to be upregulated during trophoblast differentiation of human embryonic stem cells, induced by BMP4; Krendl et al. (2017)

have recently confirmed this finding, identifying VGLL1 to be induced, following induction of GATA factors, after BMP4 treatment of these cells. Interestingly, in Xenopus, Vgll1 was found to be expressed in the epidermis, and noted to be induced downstream of BMP4 (Faucheux et al., 2010). Downregulation of VGLL1 in hESC resulted in blunted expression of the human CTB-associated marker, TP63, following BMP4 treatment. We have previously shown that TP63, a CTB marker unique to the human placenta, is required for BMP4-induced trophoblast differentiation of hESC (Li et al., 2013). Our new data now indicate that VGLL1 may be a human-specific regulator of trophoblast differentiation, possibly acting in concert with TEAD4, to regulate TP63 expression.

Overall, this study represents the first attempt at a large scale comparison between mouse and human placental development from early post-implantation period to term. Our data have identified species-specific genes and networks of TFs involved in placental development. Specifically, we have begun to characterize VGLL1 as a marker specific to human trophoblast and we posit a potential role for the TEAD4/VGLL1 complex during early human trophoblast differentiation. Further studies will be necessary to investigate the specific role of this complex, its upstream regulation and downstream effectors, in human placental development.

Acknowledgements

This work was supported by the National Institute of Child Health and Human Development (R01-NIH HD07110 to M.M.P.). M.M.P. and L.C.L. were also supported by a grant from the California Institute for Regenerative Medicine (RN3-06396 to M.M.P.). The authors declare no competing interests.

Materials and Methods

Placental samples. Human placental tissues were collected under a UCSD Human Research Protections Program Committee Institutional Review Board-approved protocol; all patients gave informed consent for collection and use of these tissues. A total of 54 normal human placentae were used for the microarray analysis: three to five biological replicates were chosen for each week of gestation from week 4 to week 12; seven second trimester (week 14-16) and five term placental samples were also included. Formalin-fixed paraffin-embedded placental tissues were chosen either from our Perinatal

biobank or from the UC San Diego Pathology Department tissue archives. A total of 22 normal human placental tissues were stained, representing 12 first trimester (one per weeks 4, 5, 7, 8, 9 and 12; two per weeks 7 and 10; and three for week 6), 7 second trimester (one per weeks 13, 14, 15, 17, 18, and two for week 20), and three term samples. Gestational age was determined based on crown-rump length as measured on first trimester ultrasound, and stated in weeks from the first day of the last menstrual period; in the text, it is listed according to the completed week of gestation (i.e. placentae at week 4 day 0, through week 4 day 6 were defined as "week 4"). For placentae from pre-viable gestations, "normal" is defined as a singleton pregnancy without any detectable fetal abnormalities on ultrasound; for term placentae, "normal" is defined by a non-hypertensive, non-diabetic singleton pregnancy, where the placenta is normally grown and shows no gross or histologic abnormalities.

Mouse placental samples were collected according to a UCSD IACUC-approved protocol. CD-1 mice were time-mated and presence of the morning plug represented E0.5 of gestation. Fifty-four placenta samples were used for the microarray analysis: five biological replicates from two different litters were collected at E7.5, 8.5, 9.5, 10.5, 11.5, 12.5, 14.5, 16.8 and 18.5. Two placentas per time point were stained.

Primary cytotrophoblast (CTB) isolation. When not specified, media components in the methods section were purchased from Gibco-Life-Technology (Carlsbad, USA). Human CTB were isolated from first trimester placentae according to the protocol described in Wakeland et al. (Wakeland et al., 2017). Briefly, chorionic villi were minced and subjected to three sequential digestions. Collected cells were separated on a Percoll (Sigma-Aldrich, St. Louis, MO) gradient. About 2 million freshly isolated cells were lysed for western blot and about 2 million cells were plated on fibronectin (20µg/mL; Sigma-Aldrich) and cultured in Dulbecco's modified Eagle's medium/F12 containing 10% fetal bovine serum (Sigma-Aldrich), penicillin-streptomycin (Thermo Fisher, Grand Island, USA), and gentamicin for 4 days in 2% oxygen before lysis. Human cytotrophoblast cells were isolated from second trimester and term placentae, based on the protocol described in Li et al. (2013). CTB from gestational weeks 8, 10, 12, and 18 (two each), one 20 week, and two term placentae were isolated and subjected to microarray-based gene expression profiling as described below.

Stem cell culture. Use of human embryonic stem cells was approved by the UCSD Embryonic Stem Cell Research Oversight Committee. H9/WA09 human embryonic stem cells (hESC) were cultured in Stem-Pro media containing 12ng/ml bFGF (BioPioneer, San Diego, USA) on geltrex-coated plates and passaged as necessary with StemDS (ScienCell, Carlsbad, USA). Trophoblast differentiation was induced according

to the 2-step protocol developed in the lab (Horii et al., 2016) and samples collected at the CTB-like stage. Briefly, 60,000 H9 hESC were seeded onto geltrex-coated 6-well plate wells in EMIM minimal media (Knock-out DMEM/K12 media containing 2mM L-Glutamine, 1mg/ml ITS, 2% BSA, 100ng/ml heparin and MEM non-essential amino acids). After 48h, 10ng/ml BMP4 (R&D Systems, Minneapolis, USA) were added to the EMIM media and cells were cultured for 4 days. Media was changed every day.

Five Mission shRNA Lentiviral constructs targeting the human VGLL1 gene were purchased and packaged into lentiviral particles according to the manufacturer's instructions (Sigma-Aldrich). Lentiviral supernatants containing a mix of all five VGLL1-targeting constructs or a Scramble control sequence were concentrated with PEG-it virus precipitation solution (System Biosciences). H9 hESC were infected with the concentrated viral particles and 8µg/ml polybrene (Sigma). Stable clones were selected with 10µg/ml of puromycin. Packaging and infection efficiency were tested using a GFP-expressing lentivirus. Two independently derived clones, which showed the best knock-down, were used in further experiments.

Mouse trophoblast stem cells derived in the lab were cultured in feeder-free condition in differentiation media containing RPMI 1640 media (Corning, Manassas, USA), 20% fetal bovine serum (Sigma-Aldrich), 1mM sodium pyruvate (Invitrogen, Carlsbad, USA), 2mM L-glutamine, 55nM 2-mercaptoethanol (Invitrogen) and the addition of the growth factors FGF4 (25ng/mL, Sigma-Aldrich), Activin A (10ng/ml, Stemgent) and heparin (1ug/mL, Sigma) to keep them undifferentiated. Cells were differentiated as previously described in Moretto-Zita et al. (Moretto Zita et al., 2015).

Immunohistochemistry and in-situ hybridization. Placental tissue samples were fixed in neutral-buffered formalin and embedded in paraffin. 5µm sections were subjected to either immunohistochemistry (IHC) or in-situ hybridization (ISH), both performed on a Ventana Discovery Ultra automated stainer (Ventana Medical Systems, Tucson, USA). For IHC, standard antigen retrieval was performed for 24-40 minutes at 95°C as per manufacturer's protocol (Ventana Medical Systems). The following primary antibodies were incubated for 1 hour at 37°C: rabbit anti-CDX2 (EPR2764Y, 1:100; Abcam, Cambridge, USA), rabbit anti-VGLL1 (HPA042403, 1:100; Sigma-Aldrich) and rabbit anti-TEAD4 (HPA056896, 1:20 Sigma-Aldrich). Staining was visualized using 3,3'-diaminobenzidine (DAB, Ventana Medical Systems) and slides were counterstained with hematoxylin. For ISH, slides were de-paraffinized, and subjected to antigen retrieval and protease treatment as described by the manufacturer (ACD-Bio, Hayward, USA). ISH was performed using the RNAscope method with probes specific to human ELF5, EOMES, GATA3, and VGLL1, and to mouse Cdx2, Tead4, and Vgll1 as well as a negative control probe

(DapB), all from ACD-Bio. Following amplification steps, the probes were visualized with DAB and slides counterstained with hematoxylin. Both IHC and ISH slides were analyzed by conventional light microscopy on an Olympus BX43 microscope (Olympus, Waltham, USA). For ISH, each dot corresponds to a single RNA message transcript; larger dots may be multiple mRNAs closely localized. Probes and antibodies were tested on known positive control tissues prior to use on the above placental tissues.

Immunofluorescence. hESC were cultured on geltrex-coated coverslips and fixed with 4% PFA/PBS (VWR International, San Diego USA) for 15min. Cells were incubated in 0.1% triton X-100 (Bio-Rad Lab, Hercules, USA) for 15 minutes and in blocking buffer (0.1% triton X-100, 5% goat serum and 1% BSA in PBS) for 1 hour both at room temperature before overnight staining in primary antibodies at 4°C. The following primary antibodies, diluted in blocking buffer, were used: anti-KRT7 (mouse, Invitrogen Clone OV-TL 12/30 1:100 or rabbit, Abcam ab68459 1:100, Cambridge, UK), anti-OCT4 (rabbit polyclonal, Abcam #ab19857 1:200 or mouse monoclonal Santa Cruz sc-5279 1:100, Dallas, USA), anti-TEAD4 (mouse monoclonal, Abcam ab58310 1:100) and anti-VGLL1 (rabbit, Sigma-Aldrich HPA042403 1:100). Immunofluorescence on a paraffin-embedded placenta (week 9) was performed after antigen retrieval according to manufacturer's guidelines. Tissue was incubated overnight with anti-VGLL1 (rabbit, Sigma-Aldrich, 1:500) and anti-PCNA (mouse, Abcam ab29, 1:500). After PBS- washes, cells and tissues were incubated with AlexaFluor-488 or -594-conjugated goat anti-rabbit or anti-mouse secondary antibodies (1:500, Thermo Fisher) for 2 hour at room temperature. Nuclei were stained with DAPI during final washing steps. Coverslips were mounted on glass slides with Hard-set Vectorshield mounting media (Thermo Fisher) and visualised under a Leica STP 6000 fluorescent microscope (Buffalo Grove, USA).

Western blot. Cells were lysed in protein lysis buffer (1% triton X-100 and 0.5% SDS in TBS) containing HALT protease inhibitor cocktail (Thermo Fisher) and 5mM EDTA and sonicated. Protein lysates were quantified using a Pierce BCA assay (Thermo Fisher). Thirty micrograms of protein were loaded onto 12.5% or 14% SDS gels and separated by gel electrophoresis followed by transfer onto PVDF membrane. Membranes were probed with 1ug/ml of anti-human VGLL1 (rabbit, Sigma HPA042403), anti-mouse VGLL1 C-terminal (rabbit, Abcam ab171019) or anti-β-ACTIN (mouse, Sigma-Aldrich #A5441) antibodies overnight at 4°C followed by HRP-conjugated anti-rabbit and anti-mouse antibody (1:2,000, Cell Signaling Technology, Beverly, USA) incubation for one hour at room temperature. Luminol reaction was performed with Pierce ECL Western Blotting Substrate (Thermo Fisher) and exposed film was processed in an automatic medical film processor machine SRK-101A (Konica Minolta, San Diego, USA).

RNA purification, total RNA microarray-based gene expression profiling and q-RT-PCR. Total RNA was purified using the MirVana RNA extraction kit (Ambion, Carlsbad, USA). For microarray analysis, total RNA was quantified using the Ribogreen reagent (Life Technology, Carlsbad, USA), and quality-controlled on a Bioanalyzer (Agilent, La Jolla, USA). Samples with RIN > 8.0 were selected for microarray analysis. Two hundred nanograms of total RNA were amplified and labeled using the TotalPrep kit (Ambion). The labeled product was then hybridized and scanned on a BeadArray Reader (Illumina, San Diego, USA) according to the manufacturer's instructions. Illumina HT12 was used for human placental tissues; MouseRef-8 v2.0 for mouse samples. Data have been uploaded into the Gene Expression Omnibus (GEO) database (Edgar, 2002) under GSE100053 and GSE100279 (Arul-nambi-rajan et al., 2017). For qPCR, total RNA was quantified with a Nanodrop and 500ng were reverse transcribed with iScript RT kit (Bio-Rad). Four microliters of diluted cDNA were used in each qPCR reaction with Power SYBR green PCR Mastermix (Applied Biosystems, Carlsbad, USA) and 1.25µM of target primers. Data were analysed according to the ddCt method using 18S as house-keeping gene. Statistical analysis was performed on the normalized Ct values (dCt) using either the unpaired t-test or ANOVA with Tukey's post-hoc test. Primer sequences are shown in Table S3.

Microarray Data Analysis. Only genes represented in both species were analysed and one single probe per gene was selected when multiple probes were available (the probe with the lowest average detection p-value and highest mean AVG signal was selected). Genes were filtered for an average detection p value <0.01 and normalised in R using Robust Spline Normalisation in the lumi package. Using the PCA and heatmap functions in Qlucore Omics Explorer 3.1 (Qlucore AB, Lund, Sweden), outliers were removed and samples were grouped according to gestational age in a data driven manner, as described in Results. Differentially expressed genes (DEGs) were identified applying a variance cutoff of 0.02, a multi-group comparison statistical analysis (similar to ANOVA) with $q<0.05$, and a fold change of ≥2.0 in at least one of the group pairs. The same analysis was applied to the GEO dataset GSE11224 by Knox and Baker (Knox and Baker, 2008) and the list of DEGs compared to our dataset. Gene Ontology enrichment analysis for Biological Process was performed using Metascape (http://metascape.org) (Tripathi et al., 2015). Affinity Propagation (AP) algorithm (Dueck and Frey, 2007) under an R implementation (Bodenhofer et al., 2011) was used to cluster differentially expressed genes or transcription factors (TF), using Pearson correlations as the similarity measure. Genes were designated to be transcription factors if the Gene Ontology terms assigned to them included the word "transcription". The most highly connected TFs were identified by calculating the sum of all positive and negative Pearson correlations of each TF with the others, called sum score positive and sum score

negative, respectively. TFs with absolute positive and negative sum score values ≥40 for mouse and ≥ 20 for human were retained. For these filtered TF lists, we then applied absolute Pearson correlation thresholds of ≥0.75 for mouse and ≥0.6 for human (i.e. keeping only highly correlated connections), and subsequently TFs with fewer than three highly correlated connections were removed. AP results for transcription factors were visualized in Cytoscape (Shannon et al., 2003), in which the nodes represented TFs and edges represented the Pearson correlation between TFs. For the Mouse data, which contained markedly more TFs, node size was set to a fixed value to enable visualization of each node, while in the human dataset, node size represented the positive or negative sum score (as indicated). Node colors indicate the assigned AP cluster. The edge thickness reflected the Pearson correlation values, where blue indicated positive and red indicated negative correlation. We used the Prefuse Force Directed layout in Cytoscape with Pearson correlation set as the attribute that contains the weights. The default parameter values were used except that the Spring Length was set to 150. In order to compare the gene expression pattern across gestation revealed by the AP analysis between mouse and human, we average the expression values for each cluster at each time point.

Microarray dataset comparison between species. Cytoscape (Shannon et al., 2003) was used to visualize the species-specific GO enrichment analysis and to connect terms with overlapping functions. To fit a curve to the pattern of expression across gestation for each AP cluster, we calculated the average expression level for all genes in each cluster, scaled the average values for each cluster across samples (mean=0, variance=2), and then identified the optimal function f(x), where x is the gestational age. To identify the best functions, we tested several methods, including linear regression (e.g. Gaussian processes, kernel quartile, principal components, partial least squares and least squares) and nonlinear regression (e.g. support vector with polynomial kernel and genetic algorithms, in which the elementary functions used are power and exponential functions). For all methods, the objective function used was the error function. For all clusters, the functions obtained using genetic algorithms were always the best fit with the experimental data (i.e. had the smallest errors, Table S4). Therefore, the curves included in the manuscript are those derived from the genetic algorithm approach. Gene Ontology enrichment analysis for Biological Process was performed using Metascape (http://metascape.org) (Tripathi et al., 2015) and visualized on REVIGO with TreeView (Supek et al., 2011).

Bibliography

Arul-nambi-rajan, K., Khater, M., Soncin, F., Pizzo, D., Moretto-Zita, M., Pham, J., Stus, O., Iyer, P., Tache, V., Laurent, L. C., et al. (2017). Sirtuin1 is required for proper trophoblast differentiation and placental development in mice. *Placenta* In Press.

Banet, N., Gown, A. M., Shih, I.-M., Li, Q. K., Roden, R. B. S., Nucci, M. R., Cheng, L., Przybycin, C. G., Nasseri-Nik, N., Wu, L.-S.-F., et al. (2015). GATA-3 expression in trophoblastic tissues: an immunohistochemical study of 445 cases, including diagnostic utility. *Am. J. Surg. Pathol.* **39**, 101–108.

Beck, F., Erler, T., Russell, a and James, R. (1995). Expression of Cdx-2 in the mouse embryo and placenta: possible role in patterning of the extra-embryonic membranes. *Dev. Dyn.* **204**, 219–227.

Blakeley, P., Fogarty, N. M. E., 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**, 3613–3613.

Bodenhofer, U., Kothmeier, A. and Hochreiter, S. (2011). Apcluster: An R package for affinity propagation clustering. *Bioinformatics* **27**, 2463–2464.

Cuffe, J. S. M., Holland, O., Salomon, C., Rice, G. E. and Perkins, A. V. (2017). Review: Placental derived biomarkers of pregnancy disorders. *Placenta* 104–110.

Deglincerti, A., Croft, G. F., Pietila, L. N., Zernicka-Goetz, M., Siggia, E. D. and Brivanlou, A. H. (2016). Self-organization of the in vitro attached human embryo. *Nature* **533**, 251–254.

Donnison, M., Beaton, A., Davey, H. W., Broadhurst, R., L'Huillier, P. and Pfeffer, P. L. (2005). Loss of the extraembryonic ectoderm in Elf5 mutants leads to defects in embryonic patterning. *Development* **132**, 2299–2308.

Dueck, D. and Frey, B. J. (2007). Clustering by Passing Messages Between Data Points. *Science* **315**, 972–976.

Edgar, R. (2002). Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* **30**, 207–210.

Faucheux, C., Naye, F., Tréguer, K., Fédou, S., Thiébaud, P. and Thézé, N. (2010). Vestigial like gene family expression in Xenopus: Common and divergent features with other vertebrates. *Int. J. Dev. Biol.* **54**, 1375–1382.

Genbacev, O., Donne, M., Kapidzic, M., Gormley, M., Lamb, J., Gilmore, J., Larocque, N., Goldfien, G., Zdravkovic, T., McMaster, M. T., et al. (2011). Establishment of human trophoblast progenitor cell lines from the chorion. *Stem Cells* **29**, 1427–36.

Hemberger, M., Udayashankar, R., Tesar, P., Moore, H. and Burton, G. J. (2010). ELF5-enforced transcriptional networks define an epigenetically regulated trophoblast stem cell compartment in the human placenta. *Hum. Mol. Genet.* **19**, 2456–2467.

Home, P., Saha, B., Ray, S., Dutta, D., Gunewardena, S., Yoo, B., Pal, A., Vivian, J. L., Larson, M., Petroff, M., et al. (2012). Altered subcellular localization of transcription factor TEAD4 regulates fi

rst mammalian cell lineage commitment. *Proc. Natl. Acad. Sci.* **109**, 7362–7367.

Horii, M., Li, Y., Wakeland, A. K., Pizzo, D. P., Nelson, K. K., Sabatini, K., Laurent, L. C., Liu, Y. and Parast, M. M. (2016). Human pluripotent stem cells as a model of trophoblast differentiation in both normal development and disease. *Proc. Natl. Acad. Sci.* **113**, E3882–E3891.

Jacquemin, P., Hwang, J. J., Martial, J. A., Dollé, P. and Davidson, I. (1996). A novel family of developmentally regulated mammalian transcription factors containing the TEA/ATTS DNA binding domain. *J. Biol. Chem.* **271**, 21775–21785.

Jacquemin, P., Sapin, V., Alsat, E., Evain-Brion, D., Dollé, P. and Davidson, I. (1998). Differential expression of the TEF family of transcription factors in the murine placenta and during differentiation of primary human trophoblasts in vitro. *Dev. Dyn.* **212**, 423–436.

Knox, K. and Baker, J. C. (2008). Genomic evolution of the placenta using co-option and duplication and divergence. *Genome Res.* **18**, 695–705.

Krendl, C., Shaposhnikov, D., Rishko, V., Ori, C., Ziegenhain, C., Sass, S., Simon, L., Müller, N. S., Straub, T., Brooks, K. E., et al. (2017). GATA2/3-TFAP2A/C transcription factor network couples human pluripotent stem cell differentiation to trophectoderm with repression of pluripotency. *Proc. Natl. Acad. Sci.* **114**, E9579–E9588.

Kunath, T., Yamanaka, Y., Detmar, J., MacPhee, D., Caniggia, I., Rossant, J. and Jurisicova, a. (2014). Developmental differences in the expression of FGF receptors between human and mouse embryos. *Placenta* **35**, 1079–1088.

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

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

Li, Y., Moretto-Zita, M., Soncin, F., Wakeland, A., Wolfe, L., Leon-Garcia, S., Pandian, R., Pizzo, D., Cui, L., Nazor, K., et al. (2013). BMP4-directed trophoblast differentiation of human embryonic stem cells is mediated through a ΔNp63+ cytotrophoblast stem cell state. *Development* **140**, 3965–76.

Moretto Zita, M., Soncin, F., Natale, D., Pizzo, D. and Parast, M. (2015). Gene Expression Profiling Reveals a Novel Regulatory Role for Sox21 in Mouse Trophoblast Stem Cell Differentiation. *J. Biol. Chem.* 1–22.

Natale, B. V., Schweitzer, C., Hughes, M., Globisch, M. A., Kotadia, R., Tremblay, E., Vu, P., Cross, J. C. and Natale, D. R. C. (2017). Sca-1 identifies a trophoblast population with multipotent potential in the mid-gestation mouse placenta. *Sci. Rep.* **7**, 5575.

Ng, R. K., Dean, W., Dawson, C., Lucifero, D., Madeja, Z., Reik, W. and Hemberger, M. (2008). Epigenetic restriction of embryonic cell lineage fate by methylation of Elf5. *Nat. Cell Biol.* **10**, 1280–1290.

Niakan, K. K. and Eggan, K. (2013). Analysis of human embryos from zygote to blastocyst reveals distinct

gene expression patterns relative to the mouse. *Dev. Biol.* **375**, 54–64.

Nishioka, N., Yamamoto, S., Kiyonari, H., Sato, H., Sawada, A., Ota, M., Nakao, K. and Sasaki, H. (2008). Tead4 is required for specification of trophectoderm in pre-implantation mouse embryos. *Mech. Dev.* **125**, 270–283.

Nishioka, N., Inoue, K. I., Adachi, K., Kiyonari, H., Ota, M., Ralston, A., Yabuta, N., Hirahara, S., Stephenson, R. O., Ogonuki, N., et al. (2009). The Hippo Signaling Pathway Components Lats and Yap Pattern Tead4 Activity to Distinguish Mouse Trophectoderm from Inner Cell Mass. *Dev. Cell* **16**, 398–410.

Norwitz, E. R. (2006). Defective implantation and placentation: laying the blueprint for pregnancy complications. *Reprod. Biomed. Online* **13**, 591–599.

Pobbati, A. V., Chan, S. W., Lee, I., Song, H. and Hong, W. (2012). Structural and functional similarity between the Vgll1-TEAD and the YAP-TEAD complexes. *Structure* **20**, 1135–1140.

Rhee, C., Edwards, M., Dang, C., Harris, J., Brown, M., Kim, J. and Tucker, H. O. (2017). ARID3A is required for mammalian placenta development. *Dev. Biol.* **422**, 83–91.

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

Senner, C. E. and Hemberger, M. (2010). Regulation of early trophoblast differentiation - Lessons from the mouse. *Placenta* **31**, 944–950.

Shahbazi, M. N., Jedrusik, A., Vuoristo, S., Recher, G., Hupalowska, A., Bolton, V., Fogarty, N. M. E., Campbell, A., Devito, L. G., Ilic, D., et al. (2016). Self-organization of the human embryo in the absence of maternal tissues. *Nat. Cell Biol.* **18**, 700–708.

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B. and Ideker, T. (2003). Cytoscape: A software Environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504.

Soncin, F., Natale, D. and Parast, M. M. (2014). Signaling pathways in mouse and human trophoblast differentiation: a comparative review. *Cell. Mol. Life Sci.* **72**, 1291–1302.

Steegers, E. A. P., Von Dadelszen, P., Duvekot, J. J. and Pijnenborg, R. (2010). Pre-eclampsia. In *The Lancet*, pp. 631–644.

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

Supek, F., Bošnjak, M., Škunca, N. and Šmuc, T. (2011). Revigo summarizes and visualizes long lists of gene ontology terms. *PLoS One* **6**, e21800.

Tripathi, S., Pohl, M. O., Zhou, Y., Rodriguez-Frandsen, A., Wang, G., Stein, D. A., Moulton, H. M., Dejesus, P., Che, J., Mulder, L. C. F., et al. (2015). Meta- and Orthogonal Integration of Influenza "oMICs" Data Defines a Role for UBR4 in Virus Budding. *Cell Host Microbe* **18**, 723–735.

Vaudin, P., Delanoue, R., Davidson, I., Silber, J. and Zider, A. (1999). TONDU (TDU), a novel human protein related to the product of vestigial (vg) gene of Drosophila melanogaster interacts with vertebrate TEF factors and substitutes for Vg function in wing formation. *Development* **126**, 4807 LP-4816.

Yagi, R., Kohn, M. J., Karavanova, I., Kaneko, K. J., Vullhorst, D., DePamphilis, M. L. and Buonanno, A. (2007). Transcription factor TEAD4 specifies the trophectoderm lineage at the beginning of mammalian development. *Development* **134**, 3827–3836.

Zdravkovic, T., Nazor, K. L., Larocque, N., Gormley, M., Donne, M., Hunkapillar, N., Giritharan, G., Bernstein, H. S., Wei, G., Hebrok, M., et al. (2015). Human stem cells from single blastomeres reveal pathways of embryonic or trophoblast fate specification. *Development* **142**, 4010–25.

Figures

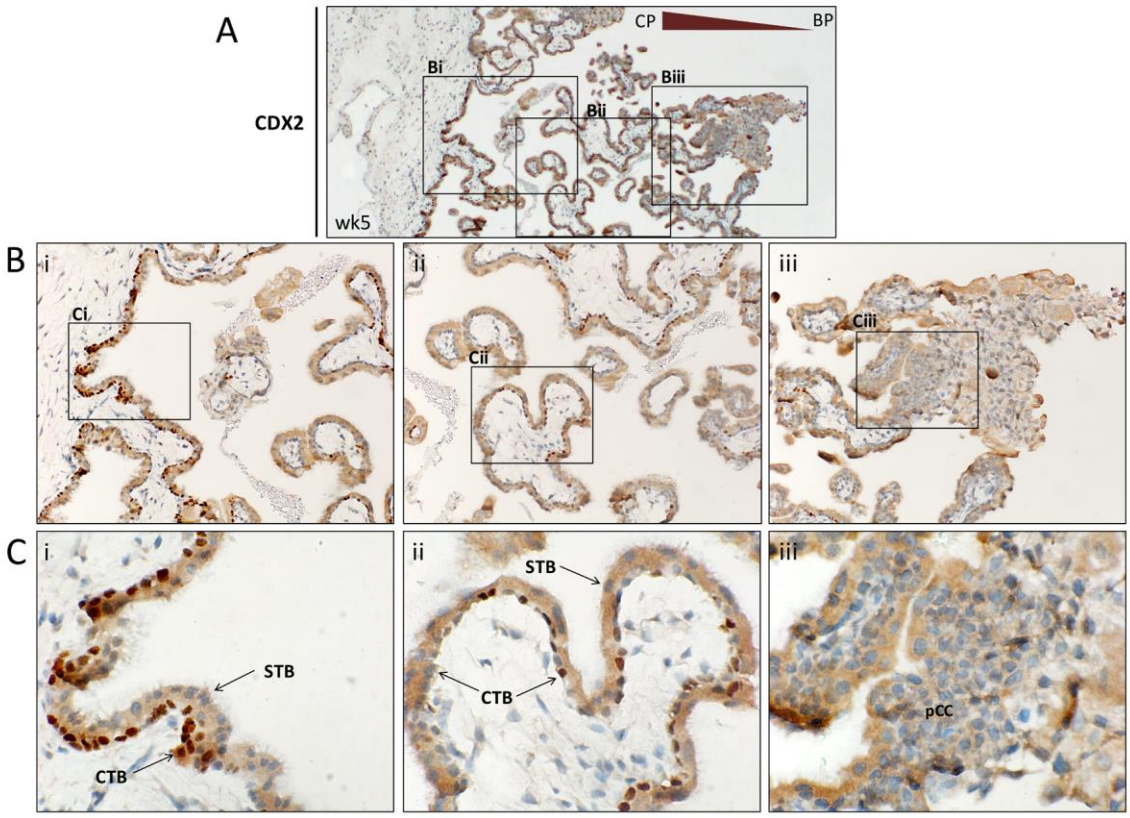


Figure 1. CDX2 is highly expressed in a sub-population of human villous cytotrophoblast near the chorionic plate. Immunohistochemistry for CDX2 in early post-implantation human placenta (week 5). **A-C.** Nuclear expression was noted in a subpopulation of villous CTB, concentrated near the chorionic plate (CP), and became less frequent toward the basal plate (BP). No nuclear staining was noted in syncytiotrophoblast (STB; Ci-ii) and proximal cell column trophoblasts (pCC; Ciii) **A.** Magnification 20x. Squares define areas magnified in B. **B.** Magnification 100x. Squares define areas magnified in C. **C.** Magnification 300x. CTB=cytotrophoblast; pCC=proximal cell column trophoblast; STB=syncytiotrophoblast; wk=week.

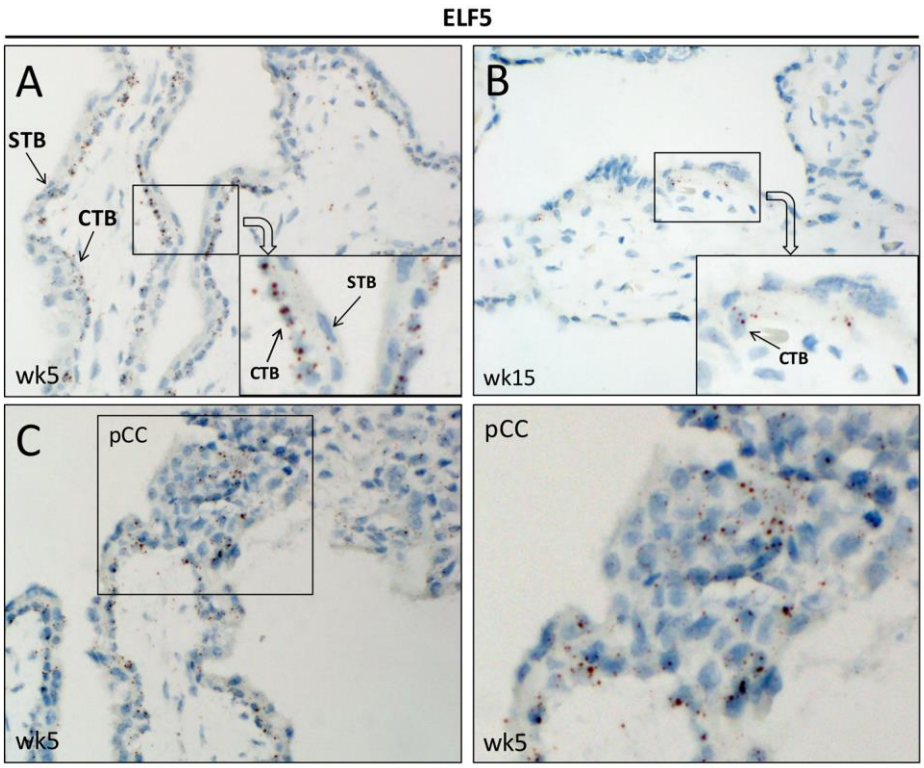


Figure 2. ELF5 is expressed in both villous cytotrophoblast and proximal cell column trophoblast in early human placenta. In-situ hybridization was performed using an ELF5-specific probe, showing a wider expression pattern in first trimester human placentae than previously reported. **A.** ELF5 expression in first trimester villous samples was highest in CTB. Magnification 300x **B.** By the second trimester, ELF5 expression was decreased, and became restricted to CTB. Magnification 300x. **C-D.** Expression of ELF5 in week 5 extravillous trophoblast located in the proximal cell columns (pCC). C: Magnification 300x; inset: magnification 600x. CTB=cytotrophoblast; pCC=proximal cell column; STB=syncytiotrophoblast; wk=week.

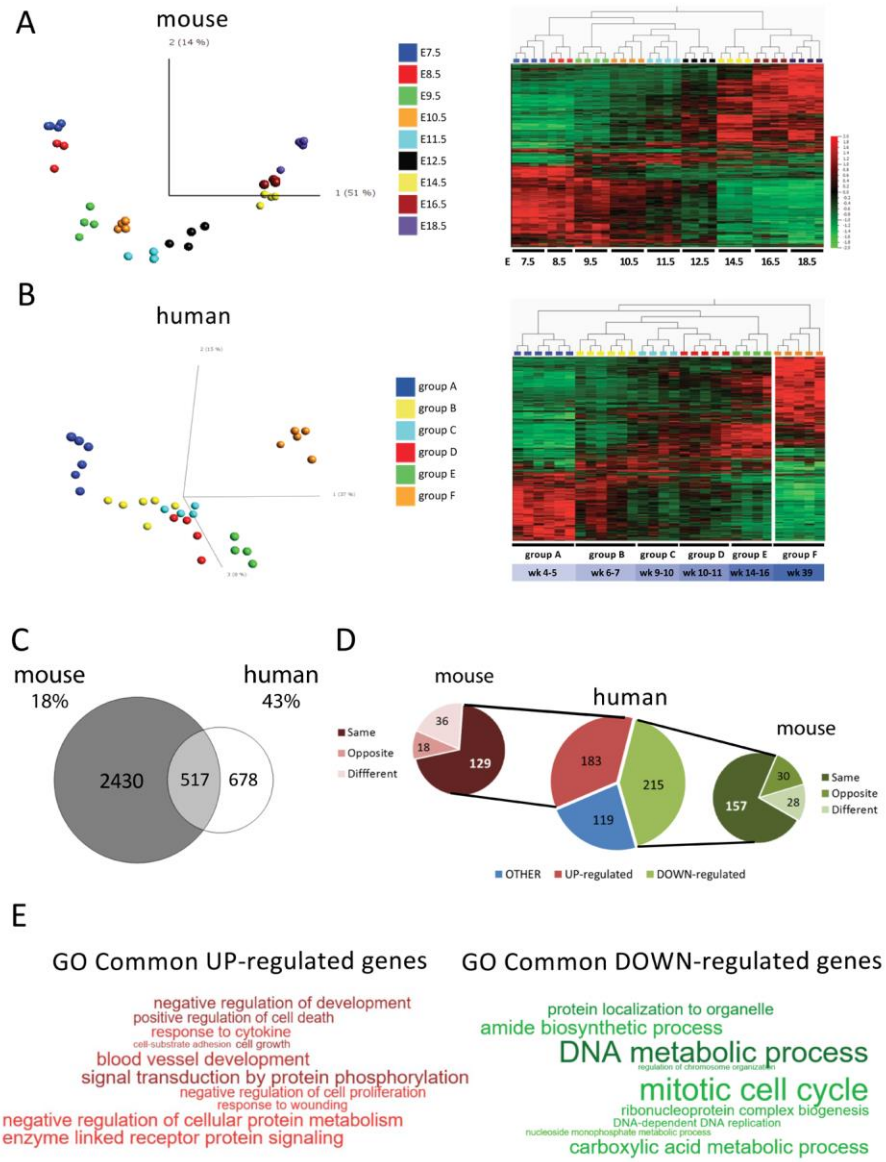
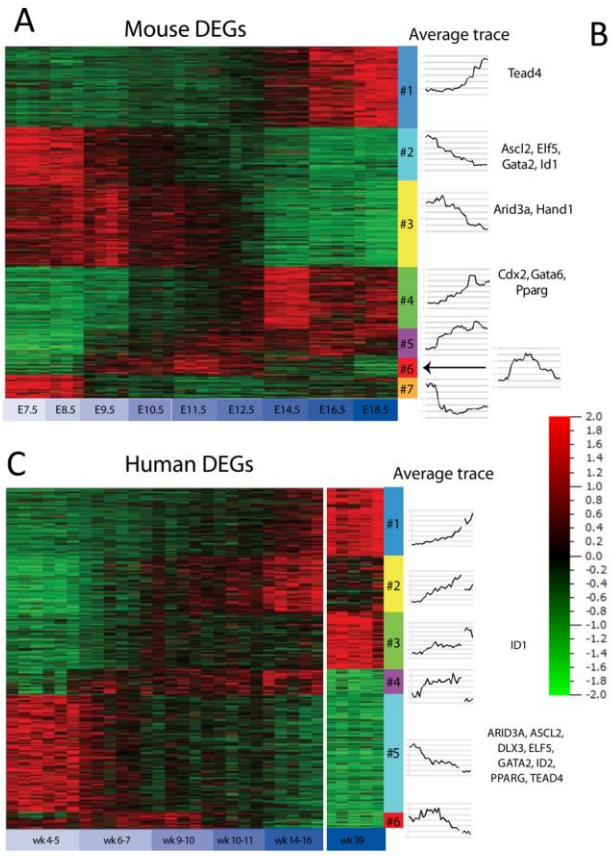


Figure 3. Comparison of mouse and human placenta gene expression signatures across gestation. A. Genome-wide expression profiles of mouse placentae across gestation. Samples are colour-coded according to gestational age. Heatmap shows expression profiles of 2,947 differentially expressed genes (DEGs, v=0.02, a multi-group analysis q<0.05, FC≥2.0). **B.** Genome-wide expression profiles of human placentae across gestation. Samples are colour-coded according to data-driven groups used for statistical analysis. Heatmap shows expression profiles of 1,195 DEGs (v=0.02, a multi-group analysis q<0.05, FC≥2.0). **C.** Venn diagram representing overlap of differentially expressed genes (DEGs) between

mouse and human datasets. **D.** Pie charts showing comparison in expression pattern direction between human and mouse common DEGs (517) (heatmaps in Fig. S3.A). Central pie chart shows expression pattern in human placentae. Genes with expression pattern other than UP and DOWN-regulation across gestation were combined and labelled as “Other”. The side charts show expression pattern comparison between the human sub-classes UP and DOWN-regulated genes and their expression patterns in mouse. **E.** Biological Function gene ontology (GO) terms enriched in commonly up- (red) and down- (green) regulated genes in mouse and human placentae. Font size is proportional to the enrichment -logP values with bigger font size representing higher enrichment.



B

AP clusters	Placenta dataset %	mTSC Up at d0 %	mTSC Up at d7 %
cluster 1	23	**15**	21
cluster 2	16	**27**	**9**
cluster 3	23	**38**	**12**
cluster 4	18	**9**	**31**
cluster 5	8	5	11
cluster 6	5	4	7
cluster 7	7	4	9
clusters 1+4	**41**	**24**	**52**
clusters 2+3	**39**	**65**	**21**

D

AP clusters	Placenta dataset %	Early CTB %
cluster 1	20	**12**
cluster 2	16	**8**
cluster 3	17	14
cluster 4	8	5
cluster 5	34	**56**
cluster 6	5	5
clusters 1+2	**36**	**20**

Figure 4. Affinity Propagation (AP) analysis of mouse and human DEGs. A. AP analysis of the 2,947 DEGs in the mouse dataset distinguished 7 clusters of co-regulated genes. Heatmap represents expression values of genes; line graphs (right) represent the average trace for each cluster. **B.** Table showing percent distribution of mouse placental DEGs (2,947 genes) across the 7 AP clusters ("Placenta dataset"), and of the DEGs up-regulated in mTSC at day 0 ("Up at d0," 765 genes) or in day 7-differentiated mTSC ("Up at d7," 604 genes). Clusters where genes are enriched in either mTSC d0 or d7 over the placenta dataset are highlighted in red, while those where there is a relative depletion are highlighted in green. **C.** AP analysis of the 1,195 DEGs in the human dataset distinguished 6 clusters of co-regulated genes. Heatmap represents expression values of genes; line graphs (right) represent the average trace for each cluster. **D.** Table showing percent distribution of human placental DEGs (1,195 genes) across the 6 AP clusters ("Placental dataset") and of the DEGs up-regulated in early (first trimester) CTB (week 8-10) compared to mature CTB (week 12-39) (197 genes) (see analysis in Fig. S5.A-

C). Clusters where genes are enriched in early CTB over the placenta dataset are highlighted in red, while those where there is a relative depletion are highlighted in green.

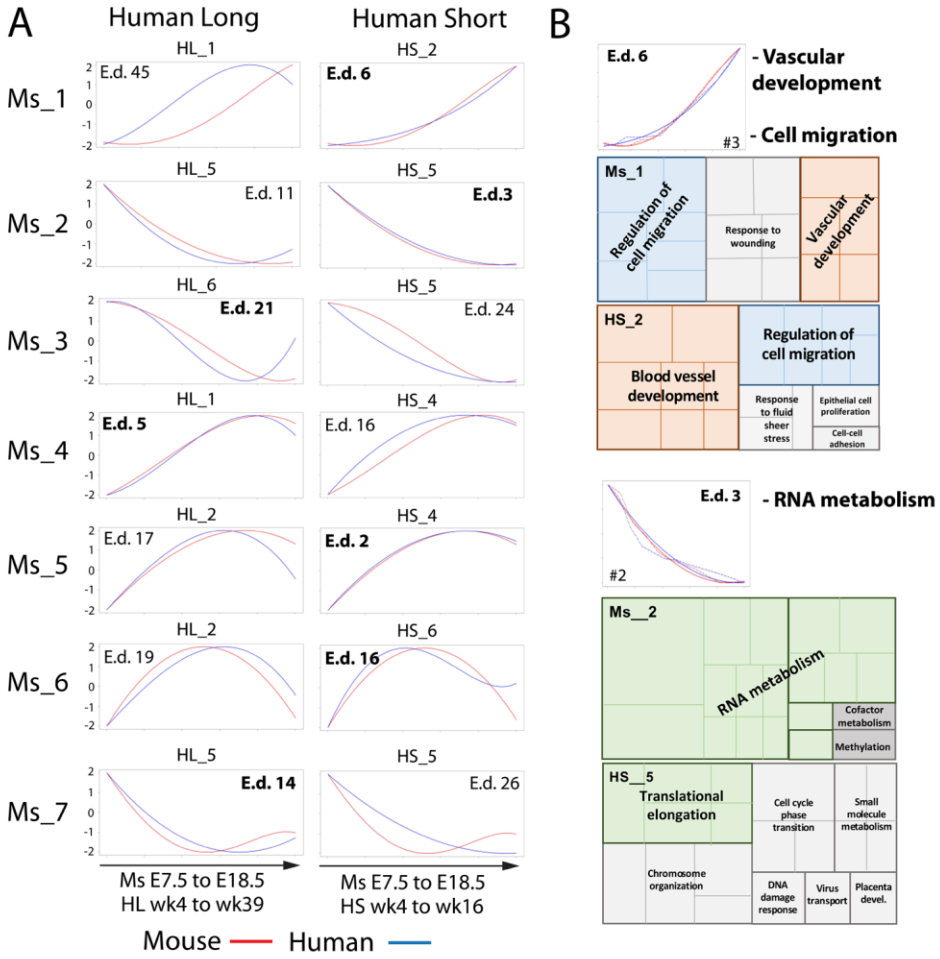
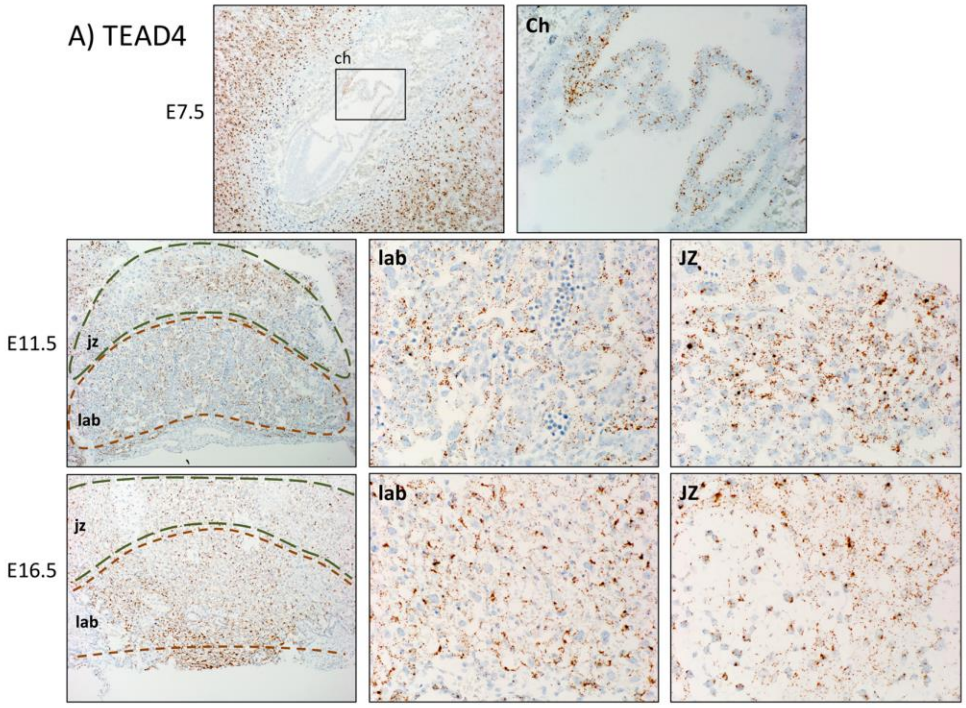


Figure 5. Comparison of expression profiles of best correlated mouse and human co-expression clusters. A. Graphs showing the AP clusters from the HL dataset (which included data from term human placentae) and HS dataset (which did not include data from term human placentae) that showed the highest similarity to each mouse cluster. Data from mouse clusters are shown in red and data from human clusters are shown in blue. The Euclidean distance (E.d) value for each pair of curves is shown. **B.** Gene Ontology enrichment analysis shows overlapping function between the genes included in the cluster pairs with closely fitting curves, Ms_1/HS_2 and Ms_2/HS_5. # = number of genes in common between the mouse and human clusters. Each rectangle is a representative GO term from the enrichment analysis. The representatives are joined into "super-groups" of loosely related terms, visualized with different colors. Size of the rectangles is proportional to the p-value. Overlapping functions between mouse and human clusters are marked with matching colors.



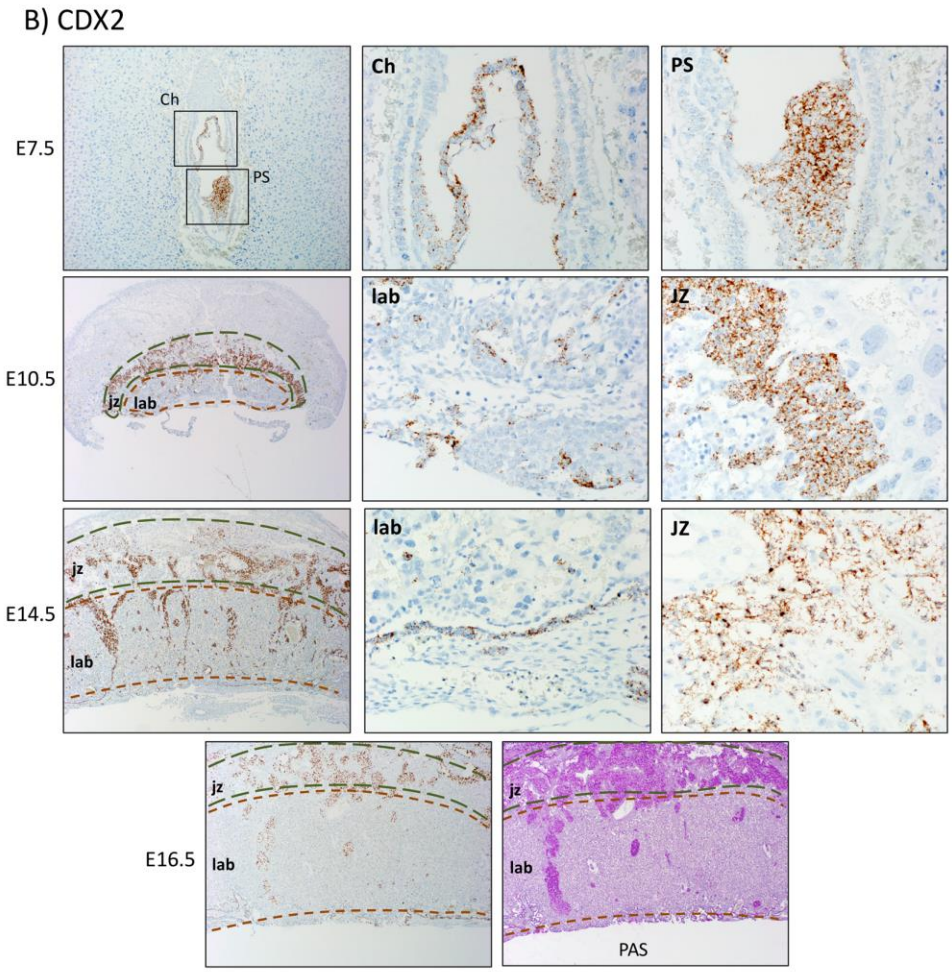


Figure 6. In-situ hybridization for Cdx2 and Tead4 in mouse placentae confirms up-regulation with increasing gestational age. A. ISH for Tead4. Tead4 expression was observed in the chorion (ch) at E7.5, and in both the labyrinth (lab) and the junctional zone (JZ) at E11.5 and E16.5. Left pictures: magnification 50x; middle and right pictures: magnification 200x. **B.** ISH for Cdx2. Cdx2 expression was observed in the chorion (ch) and primitive streak (PS) at E7.5, and in the labyrinth (near the chorionic plate, lab) and junctional zone (jz) later in gestation. PAS staining in a consecutive section at E16.5 suggests Cdx2 expression in glycogen cells. Magnification: E7.5 - Left 50x; middle and right 200x; E11.5 and E14.5 – Left 20x; middle and right 200x; E16.5 – 20x. Ch = chorion; jz = junctional zone; lab = labyrinth; PS = primitive streak.

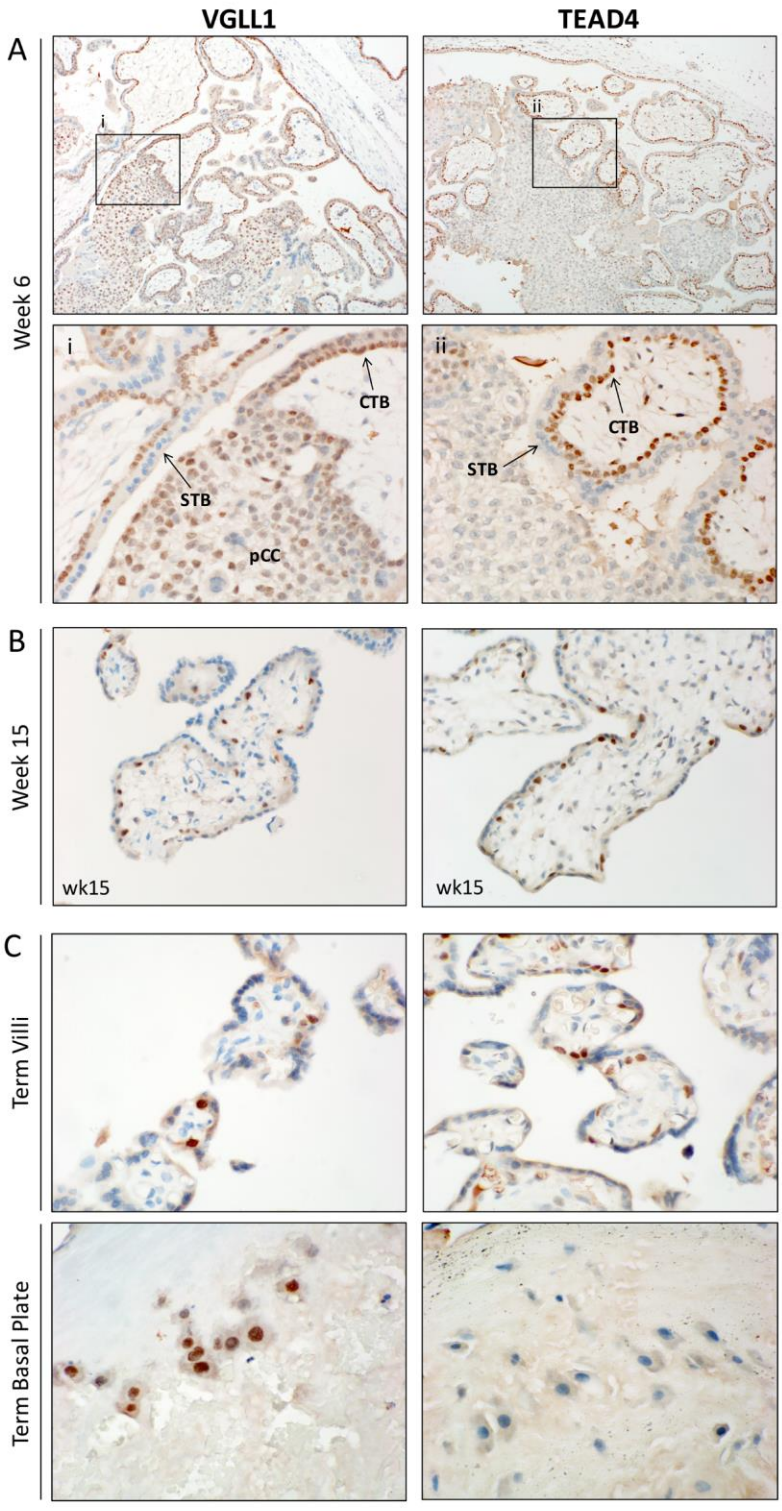


Figure 7. VGLL1 and TEAD4 are co-expressed in human villous cytotrophoblast (CTB) across gestation. **A.** In early post-implantation human placenta, VGLL1 and TEAD4 are co-expressed in villous CTB. Extravillous trophoblast in proximal cell column (pCC) express VGLL1 but not TEAD4. Magnification: Top panels – 50x; bottom panels -- 200x. **B.** In second trimester placentae, villous CTB (now discontinuous) maintain expression of both VGLL1 and TEAD4. Magnification: 200x. **C.** At term, VGLL1 expression is visible both in persistent villous CTB (term villi) but also in mature extravillous trophoblast (term basal plate). TEAD4 expression is restricted to villous CTB. Magnification: 300x. CTB = cytotrophoblast; pCC = proximal cell column; STB=syncytiotrophoblast; wk=week.

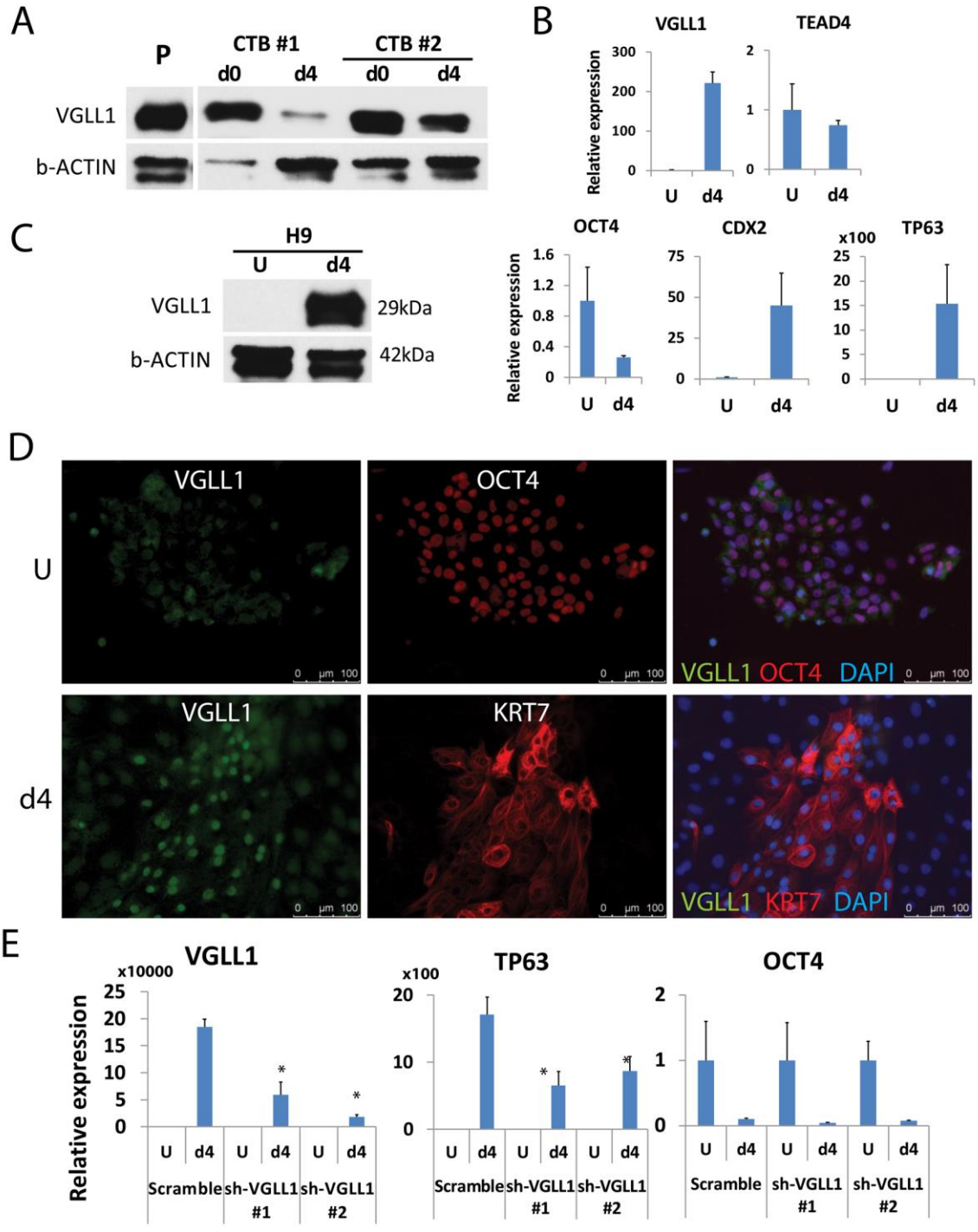
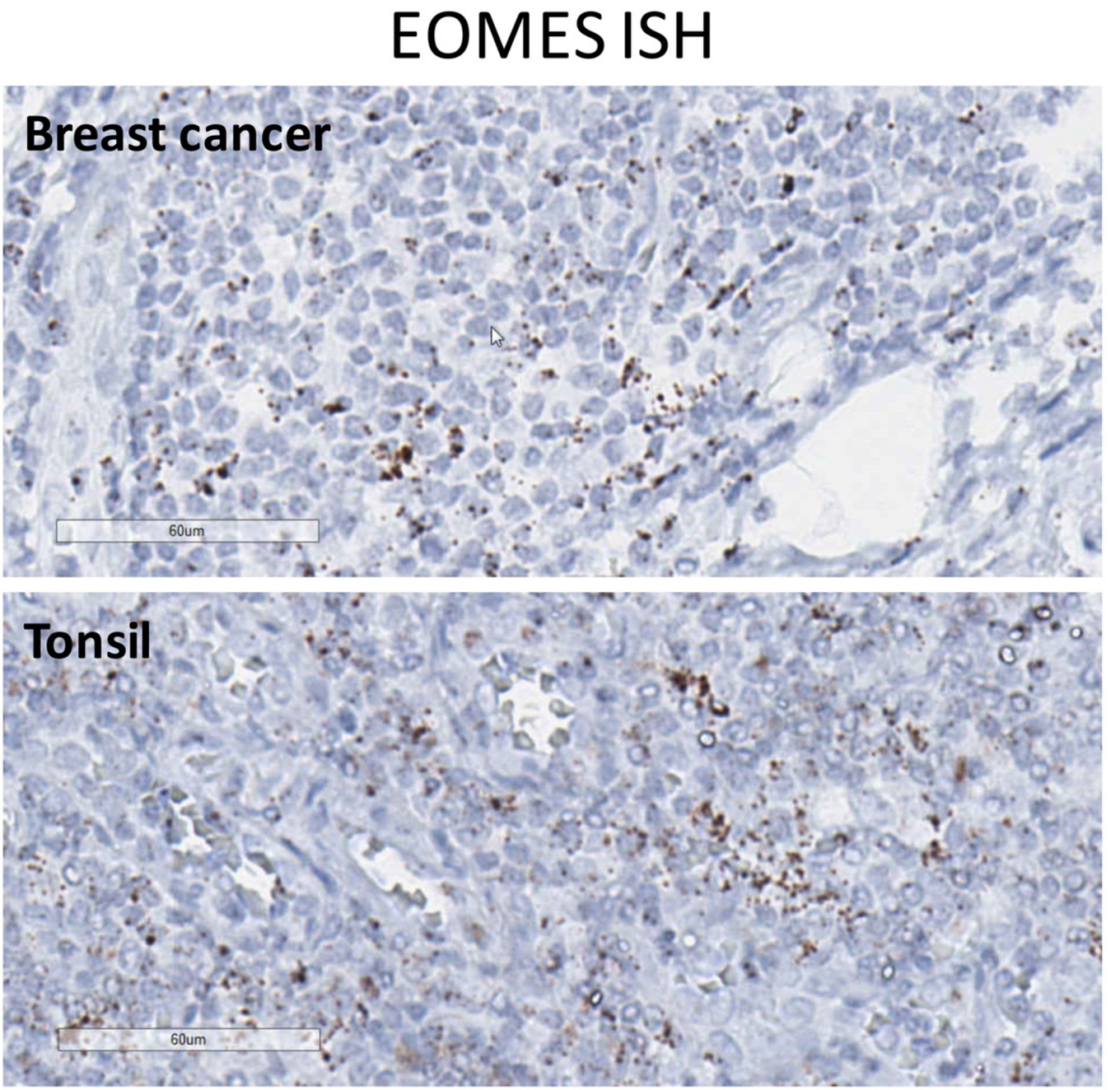
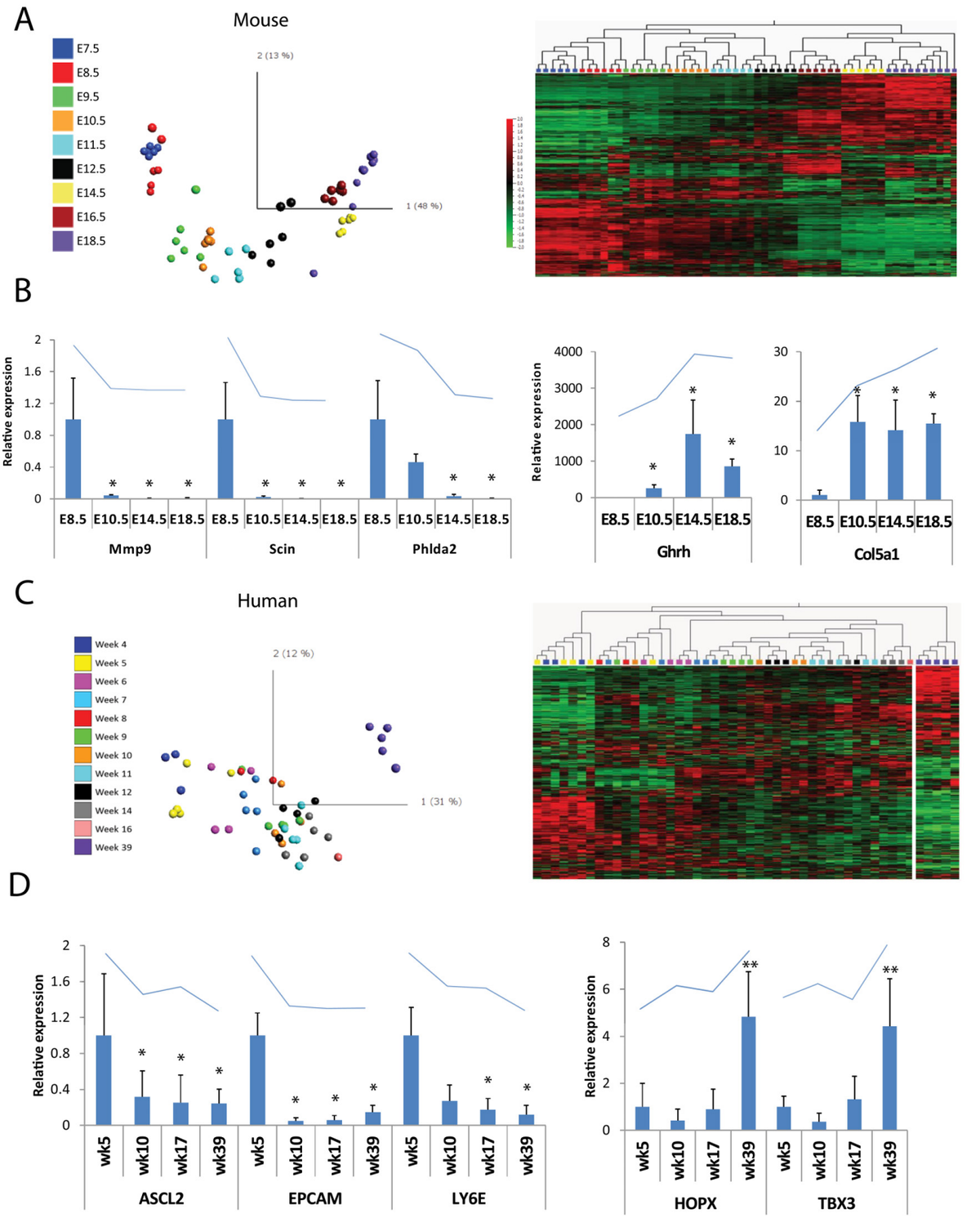


Figure 8. VGLL1 is expressed in primary and hESC-derived CTB *in vitro*. A. Western blot analysis of VGLL1 and β-actin (loading control) in week 6 whole placenta (P), and primary first trimester CTB (#1 week 13; #2 week 11). Expression was analysed both in freshly isolated CTB (d0) and differentiated CTB (d4). **B.** qRT-PCR analysis of VGLL1, TEAD4, CTB markers (TP63 and CDX2), and pluripotency marker

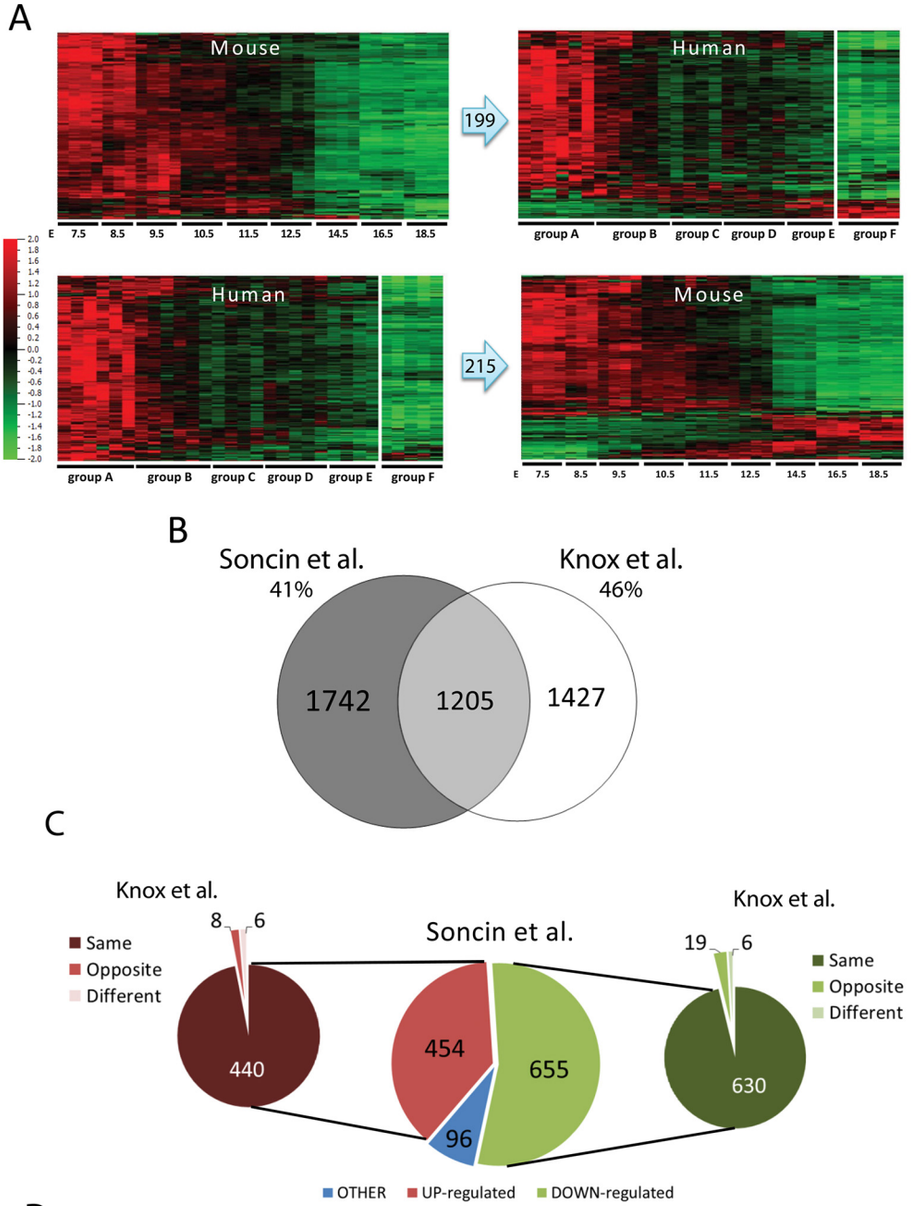
(OCT4) in undifferentiated hESC (U) and in hESC-derived CTB-like cells (d4) following step 1 of our two-step BMP4-based trophoblast differentiation protocol (Horii et al. 2016) (n=3). **C.** Western blot analysis of VGLL1 and β-actin (loading control) in undifferentiated H9 hESCs (U) and in hESC-derived CTB-like cells (d4). **D.** Double immunostaining of undifferentiated H9 hESC (U) and hESC-derived CTB-like cells (d4) with VGLL1 and either OCT4 (top) or pan-trophoblast marker KRT7 (bottom). **E.** qRT-PCR analysis of VGLL1, TP63, and OCT4 in hESC clones, stably expressing either Scramble or VGLL1-targeting shRNA. U =undifferentiated cells, d4= hESC-derived CTB-like cells. n=3 *$p<0.05$ compared to Scramble d4.



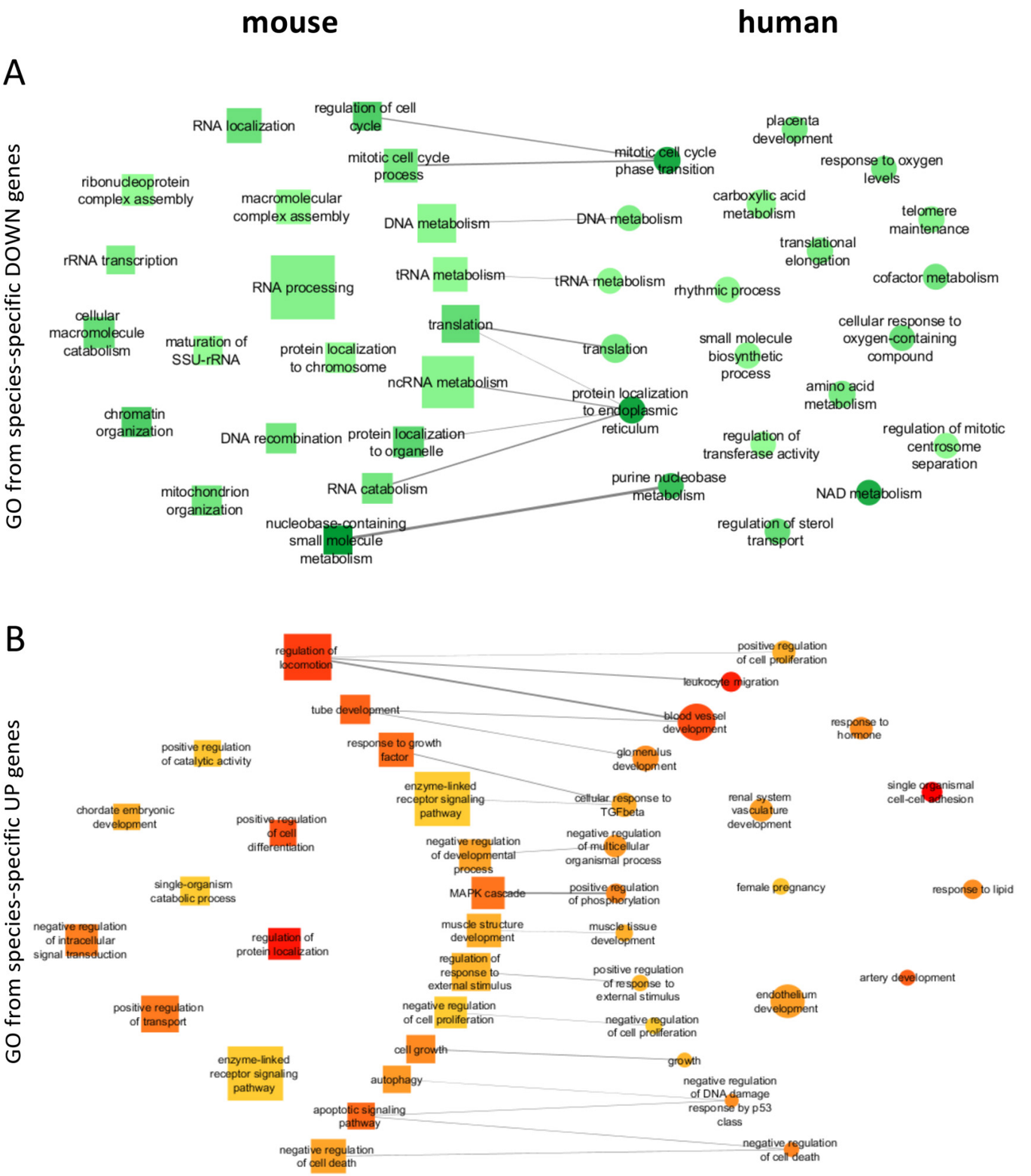
Supplementary Figure S1. In-situ hybridization using EOMES-specific probes on positive control human tissues.



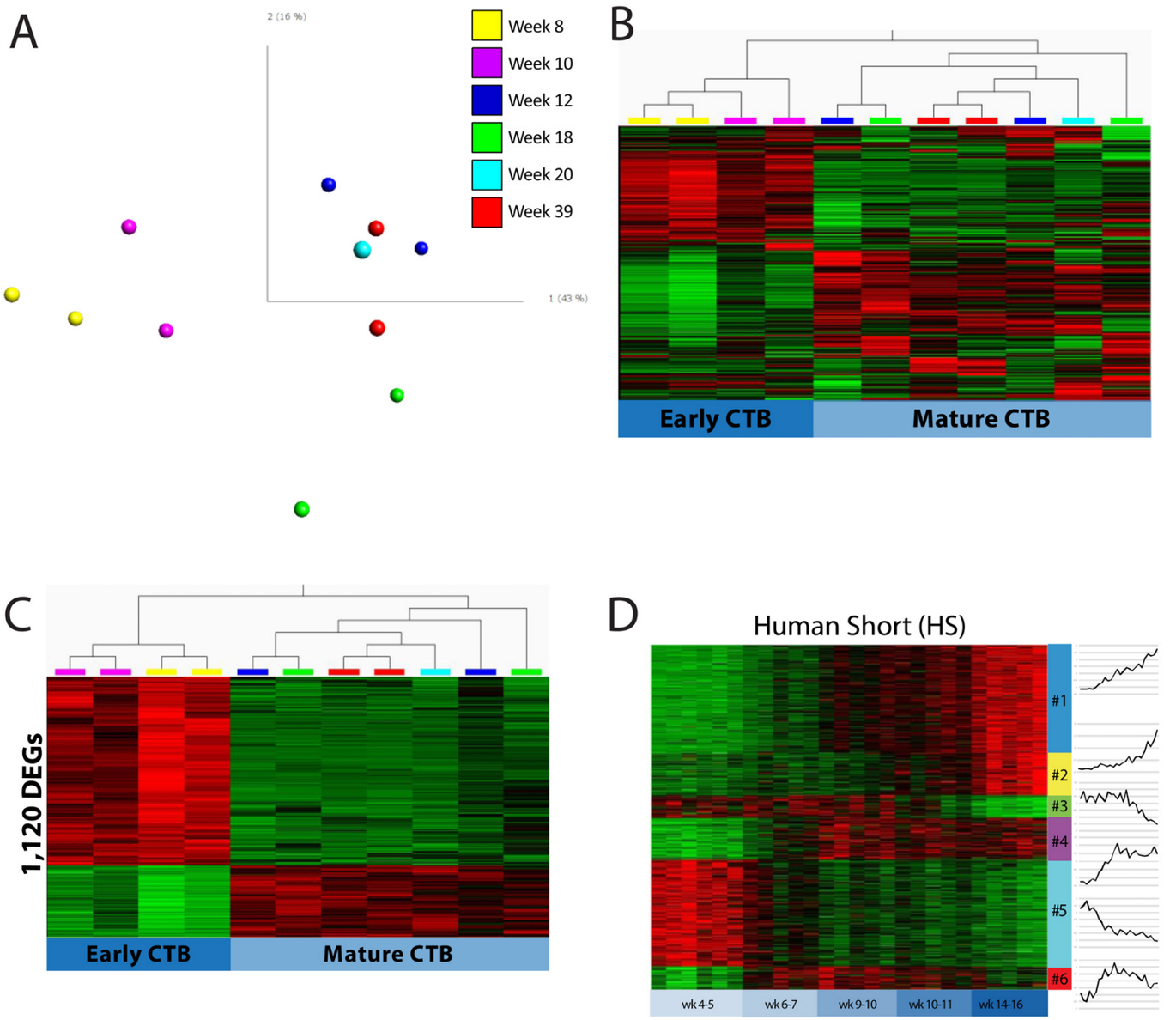
Supplementary Figure S2. Microarray data analysis of mouse and human placenta samples. A. Principle component analysis (PCA) and heatmap of mouse dataset using Qlucore (variance 0.02). Samples were colour-coded according to gestational age. **B.** qRT-PCR analysis (bar graph) for Mmp9, Scin, Phlda2, Ghrh, and Col5a1 in mouse placentas at E8.5, E10.5, E14.5 and E18.5 confirmed expression pattern observed in the microarray (line above bar graph). **C.** PCA and heatmap of human dataset using Qlucore (variance 0.02). Samples were colour-coded according to gestational age. **D.** qRT-PCR analysis (bar graph) for ASCL2, EPCAM, LY6E, HOPX, and TBX3 in human placenta from gestational week 5, 10, 17, and 39 confirmed expression pattern observed in the microarray (line above bar graph). n=3 * $p<0.05$ compared to E8.5 for mouse and Week 5 for human; ** $p<0.05$ compared to week 10.



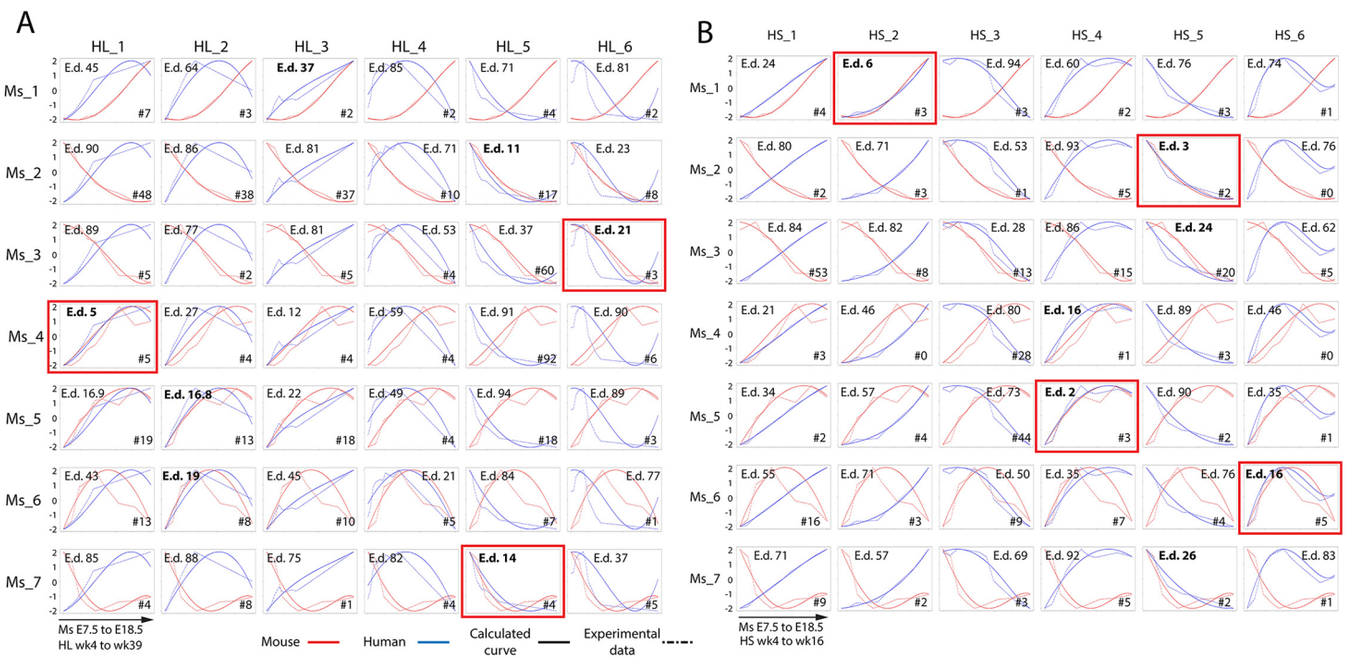
Supplementary Figure S3. Comparison between datasets. A. Top: Heatmap of mouse differentially expressed genes (DEG) downregulated with gestational age (199 genes, left) and their expression pattern in human (heatmap on the right). Bottom: Heatmap of human DEGs (215 genes, left) downregulated with gestational age and their expression pattern in mouse (heatmap on the right). These data were used to create the pie chart in Fig.3D. **B.** Venn diagram representing the overlapping of DEGs between our mouse data and the Knox et al. dataset. We analysed the mouse dataset published by Knox and Baker (2008) using similar parameters applied to our inter-species analysis (v 0.02, $q<0.05$, FC>2.0). We identified 2,632 DEGs in the Knox et al. dataset, of which 1,205 were in common with our analysis, representing over 40%. **C.** Pie charts showing the comparison in expression pattern direction of the common DEGs (1,205) between our data and that of Knox et al. Central pie chart shows the expression pattern in the current study. Genes with expression pattern other than UP and DOWN-regulation across gestation were combined and labelled as "Other." The side charts show the expression pattern comparison between the sub-classes UP- (red) and DOWN- (green) regulated genes in our analysis and their expression in the Knox et al. dataset. Over 96% of the common DEGs showed the same pattern of expression (either UP or DOWN) in both datasets. **D.** Expression levels of CDX2 and ELF5 in human placentas across gestation from our microarray dataset. Note the high variability of CDX2 expression between samples.



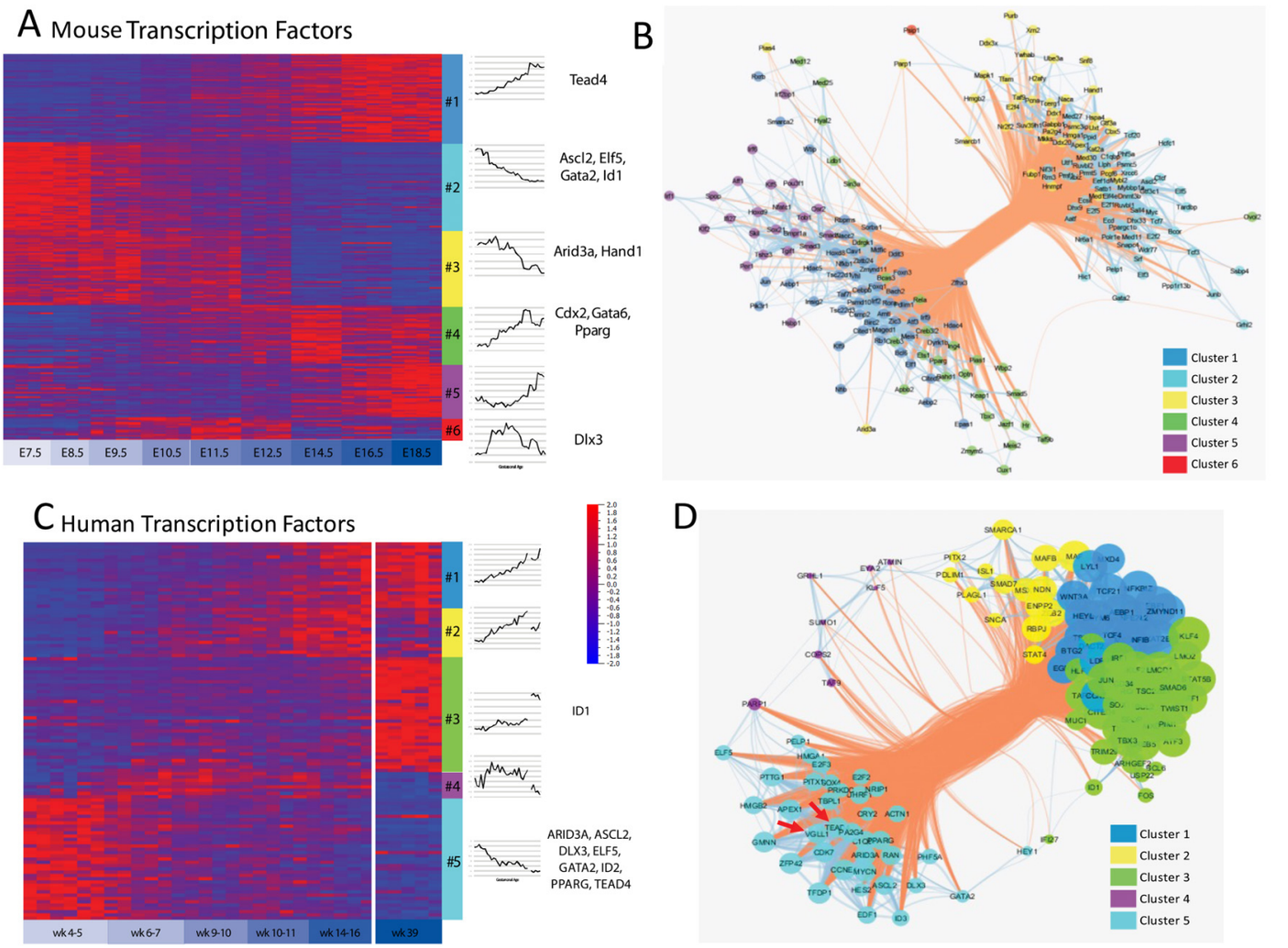
Supplementary Figure S4. Enrichment analysis of Gene Ontology Biological Functions of species-specific up and down-regulated genes in mouse and human placentas. A. Gene ontology of genes downregulated across gestation. **B.** Gene ontology of genes upregulated across gestation. Node shape represents the species (mouse – square; human – circle). The size of the nodes is correlated to the log(p) of the enriched term. Lines (edges) connect mouse and human terms that include overlapping functions. The thickness of the edges is proportional to the number of terms in common between the species.



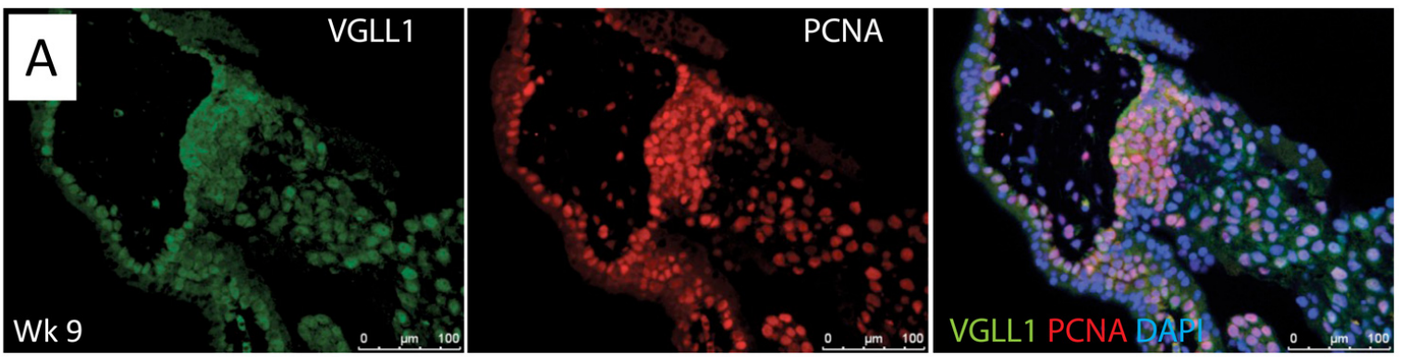
Supplementary Figure S5. Details of microarray data analysis of human primary CTB and placental tissues. A. Principal component analysis (PCA) of primary CTB expression profiles. Note separation into two main groups along PCA1: weeks 8 and 10 on one side and weeks 12, 18, 20, and 39 on the other. **B.** Heatmap of human primary CTB genes filtered for variance (0.05). Note hierarchical clustering of samples into two groups, labelled "early" CTB and "mature" CTB, which were used for subsequent differential expression analysis. **C.** Heatmap of 1,120 differentially expressed genes (DEGs) between the two identified groups of CTB (two-group analysis $q<0.01$). Of these DEGs, 197 genes were also differentially expressed in human placentas across gestation and included in the AP analysis (see Figure 6B). **D.** Affinity Propagation analysis of the 665 DEGs in the "short" human dataset, which did not include term samples. AP distinguished 6 clusters of co-regulated genes. Heatmap represents expression values of genes; line graphs (right) represent the average trace for each cluster.

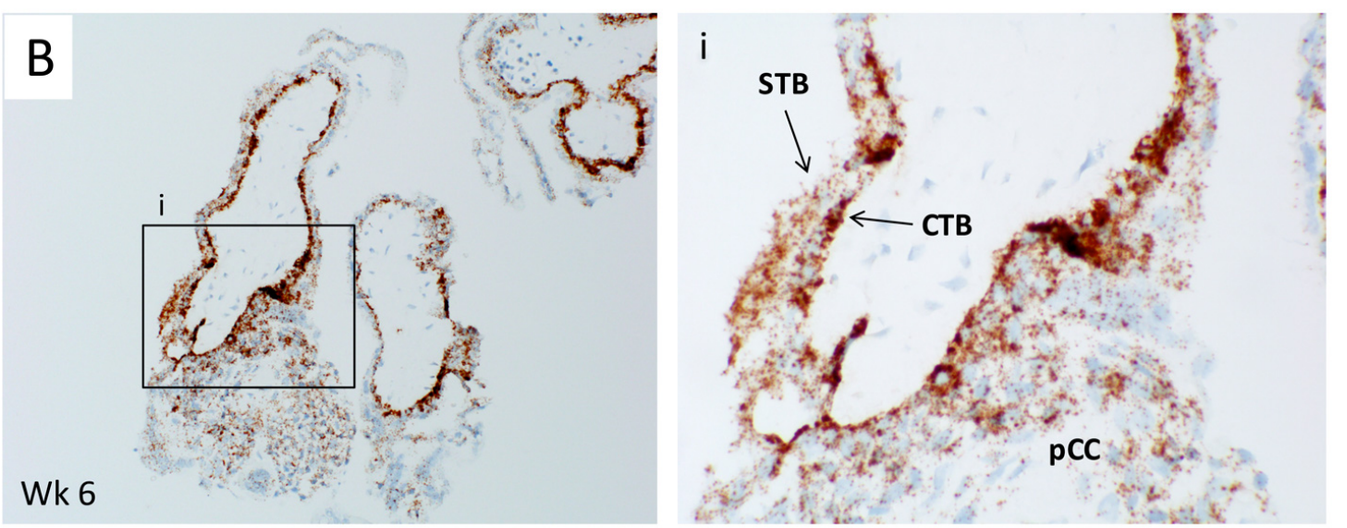


Supplementary Figure S6. Comparison of expression profiles of all mouse and human co-expression clusters. A. Matrix of graphs showing the relationship between each mouse and human AP co-expression cluster (from the HL dataset, which included data from term human placentae). **B.** Matrix of graphs showing the relationship between each mouse and human AP co-expression cluster (from the HS dataset, which did not include data from term human placentae). For both matrices, both the experimental data (dashed lines) and corresponding best-fit curves (solid lines) for each AP cluster are shown. Data from mouse clusters are shown in red and data from human clusters are shown in blue. The Euclidean distance (E.d) values for each pair of curves is shown. # = number of genes in common between the mouse and human clusters. For each mouse cluster, the HL or HS cluster with the most similar pattern of expression of gestation, as measured by the Euclidean distance, is boxed in red.

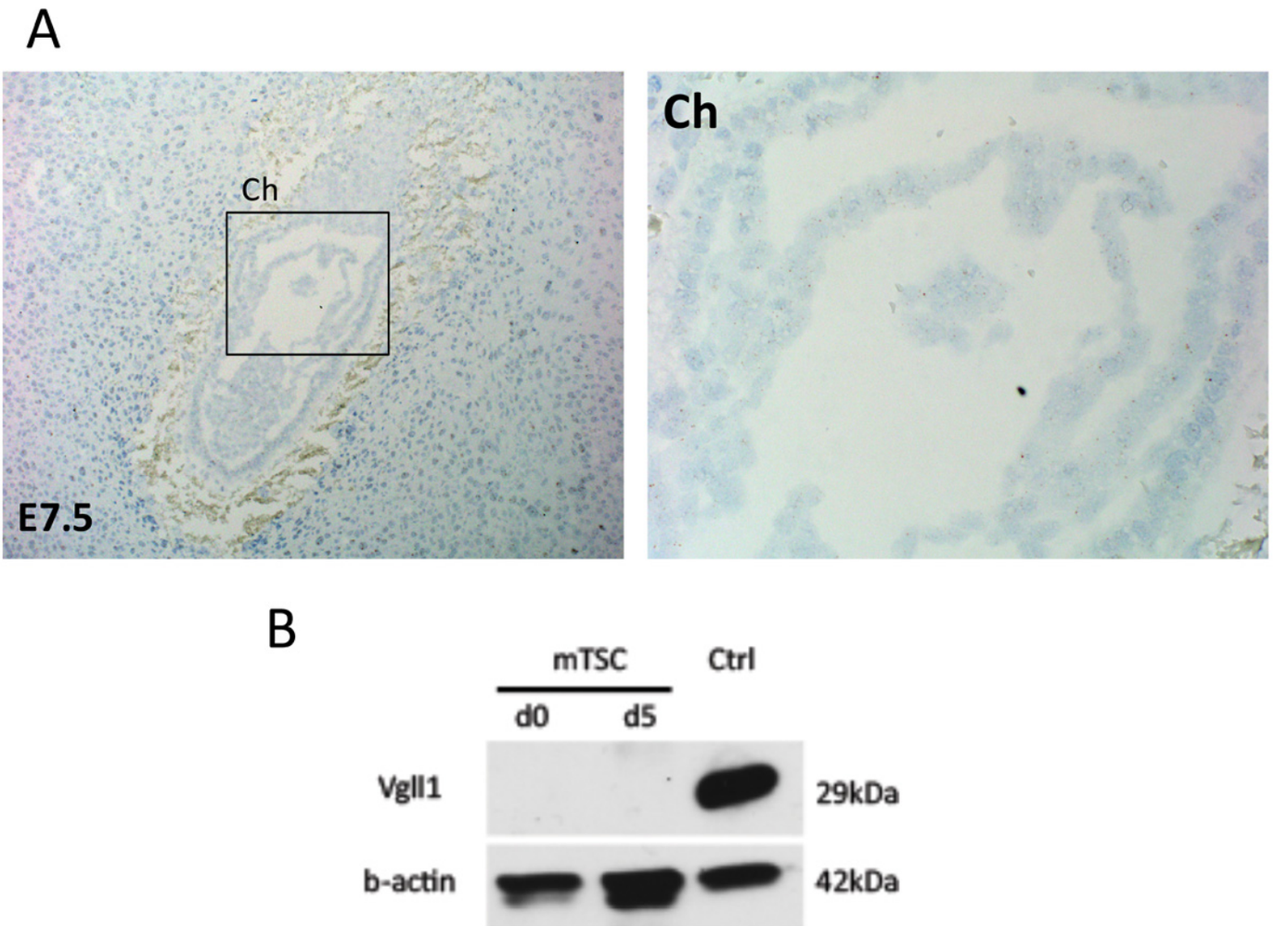


Supplementary Figure S7. A. Affinity Propagation (AP) analysis of transcription factors (TFs) in the mouse dataset distinguished 6 clusters of co-regulated genes. Heatmap represents expression values of TFs; line graphs (right) represent the average trace for each cluster. Representative gene(s) in each cluster are shown. **B.** Mouse TF cluster network. For clarity, node size was set at a fixed value. Nodes are colour-coded according to the cluster they belong to; red edges indicate negative correlation, while blue edges show positive correlation between nodes. **C.** Affinity Propagation (AP) analysis of transcription factors in the human dataset distinguished 5 clusters of co-regulated genes. Heatmap representing expression values of TFs; line graphs (right) represent the average trace for each cluster. Representative gene(s) in each cluster are shown. **D.** Human TF cluster network. Nodes are colour-coded according to the AP cluster number, while their size represents the positive sum scores. Red edges indicate negative correlation, while blue edges show positive correlation between nodes. Edge thickness identifies Pearson correlation absolute value. Red arrows show the position of TEAD4 and VGLL1.

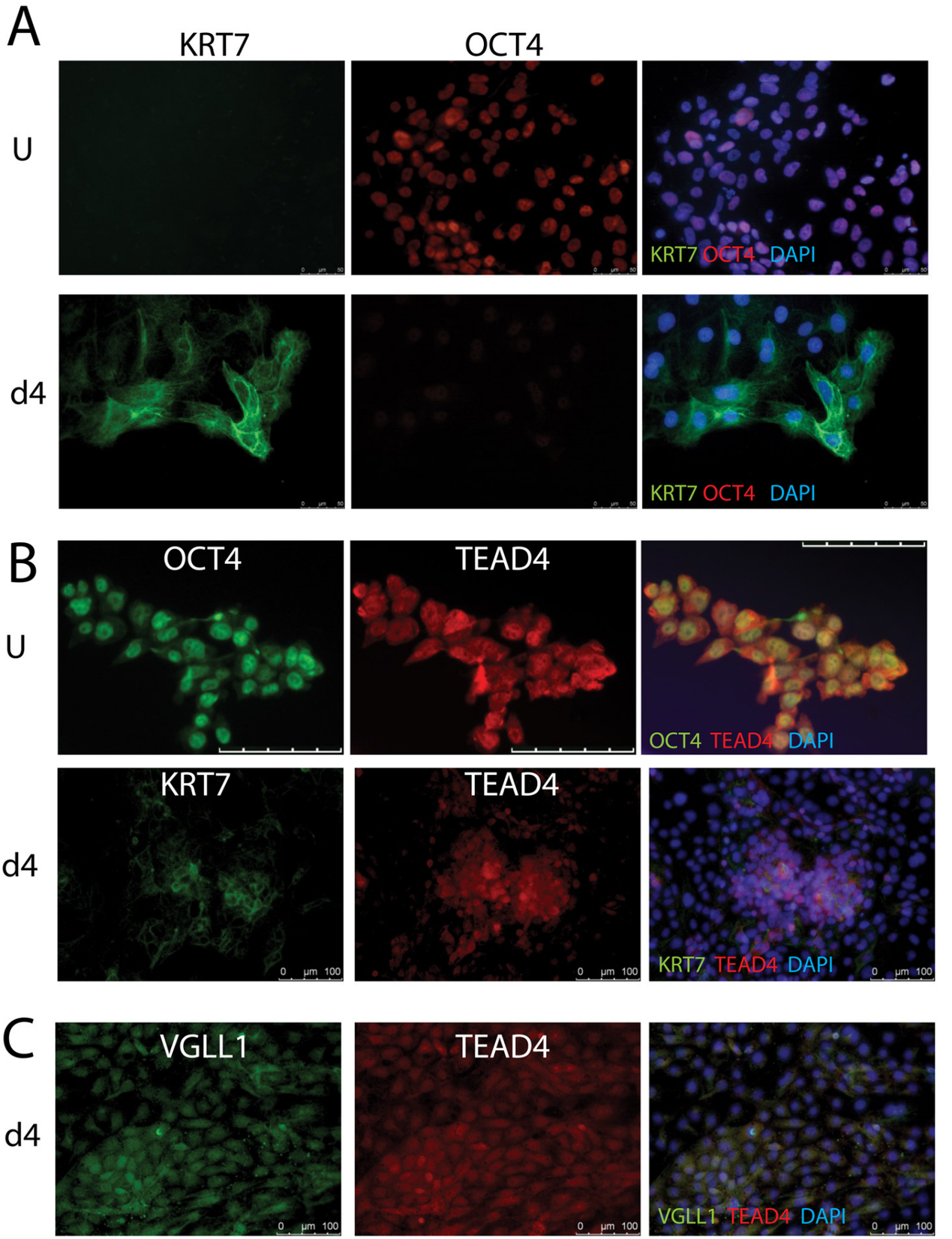




Supplementary Figure S8. IHC and ISH of first trimester human placentas. A. Co-staining of week 9 placenta with VGLL1 (green), the proliferation marker PCNA (red), and Dapi (blue). **B.** In situ hybridization with VGLL1-specific probes in week 6 placenta. Expression of VGLL1 mRNA is noted in cytotrophoblast (CTB), syncytiotrophoblast (STB), as well as in trophoblast of the proximal cell column (pCC), although it is most enriched in CTB. Magnification: Left – 100x; Right: 300x.



Supplementary Figure S9. Vgll1 is not expressed in mouse placenta and trophoblast. A. No Vgll1 RNA was noted in E7.5 placentas by in-situ hybridization. Ch= chorion. Magnification: 50x (left); 200x (right). **B.** Western blot analysis of undifferentiated (d0) and differentiated (day 5/d5) mouse trophoblast stem cells (mTSC) for Vgll1 and β-actin (loading control). Commercially available mouse whole stomach lysate (Abcam) was used as positive control.



Supplementary Figure S10. Expression of OCT4, KRT7 and TEAD4 in the hESC-based *in vitro* model of human trophoblast differentiation. A. Immunostaining for OCT4 (red) and KRT7 (green) in pluripotent (undifferentiated) H9 hESC (U) and in hESC-derived CTB-like cells (d4). OCT4 and KRT7 in this model are mutually exclusive. **B.** Immunostaining for TEAD4, along with either OCT4 in undifferentiated hESC (U) or with KRT7 in hESC-derived CTB-like cells (d4). **C.** Double immunostaining of hESC-derived CTB-like cells (d4) with VGLL1 and TEAD4, showing co-localization in nuclei of these cells.

Table S1. Genelists

Click here to Download Table S1

Table S2. GO enrichment

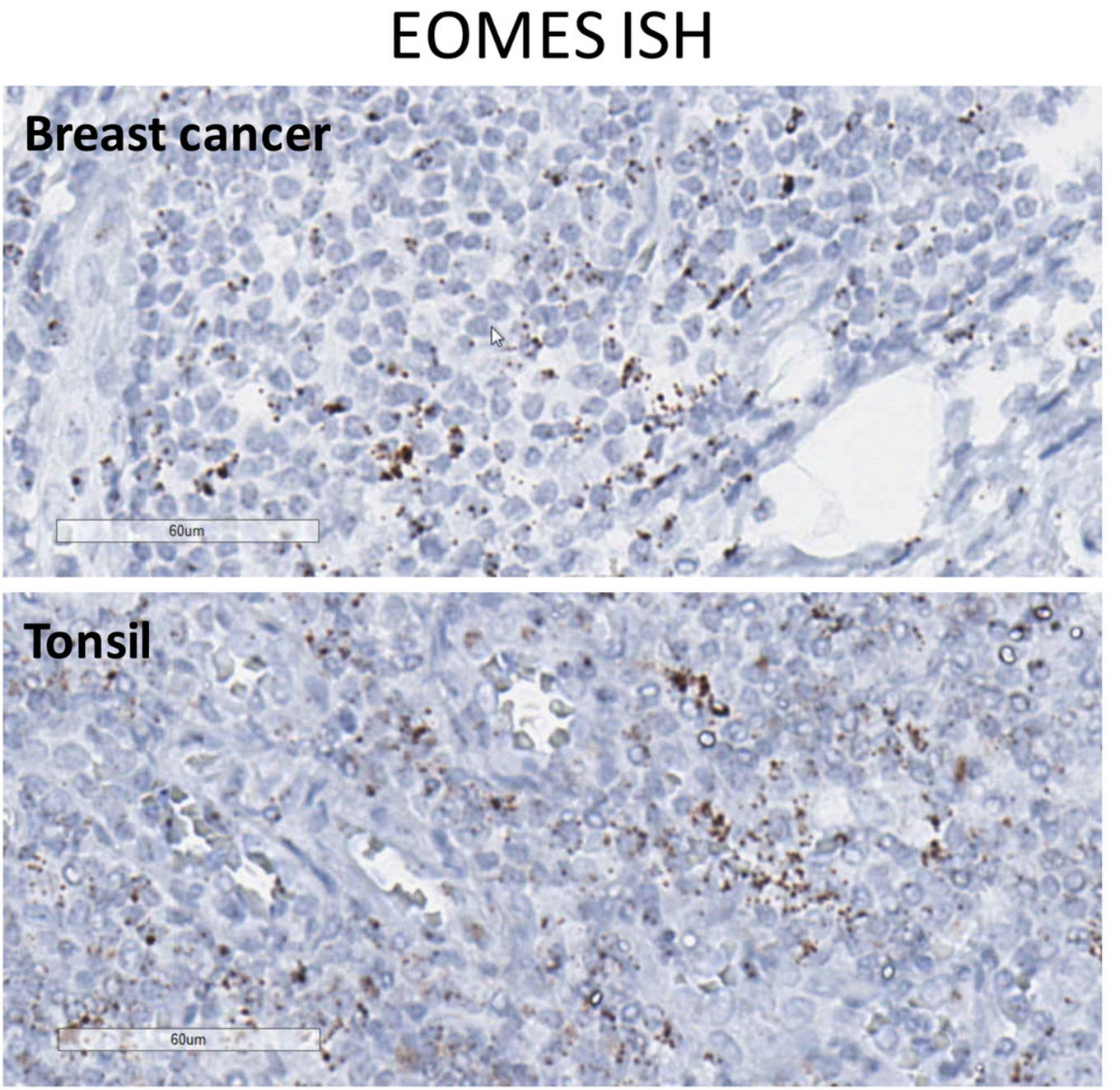
Click here to Download Table S2

Table S3. q-RT-PCR primers

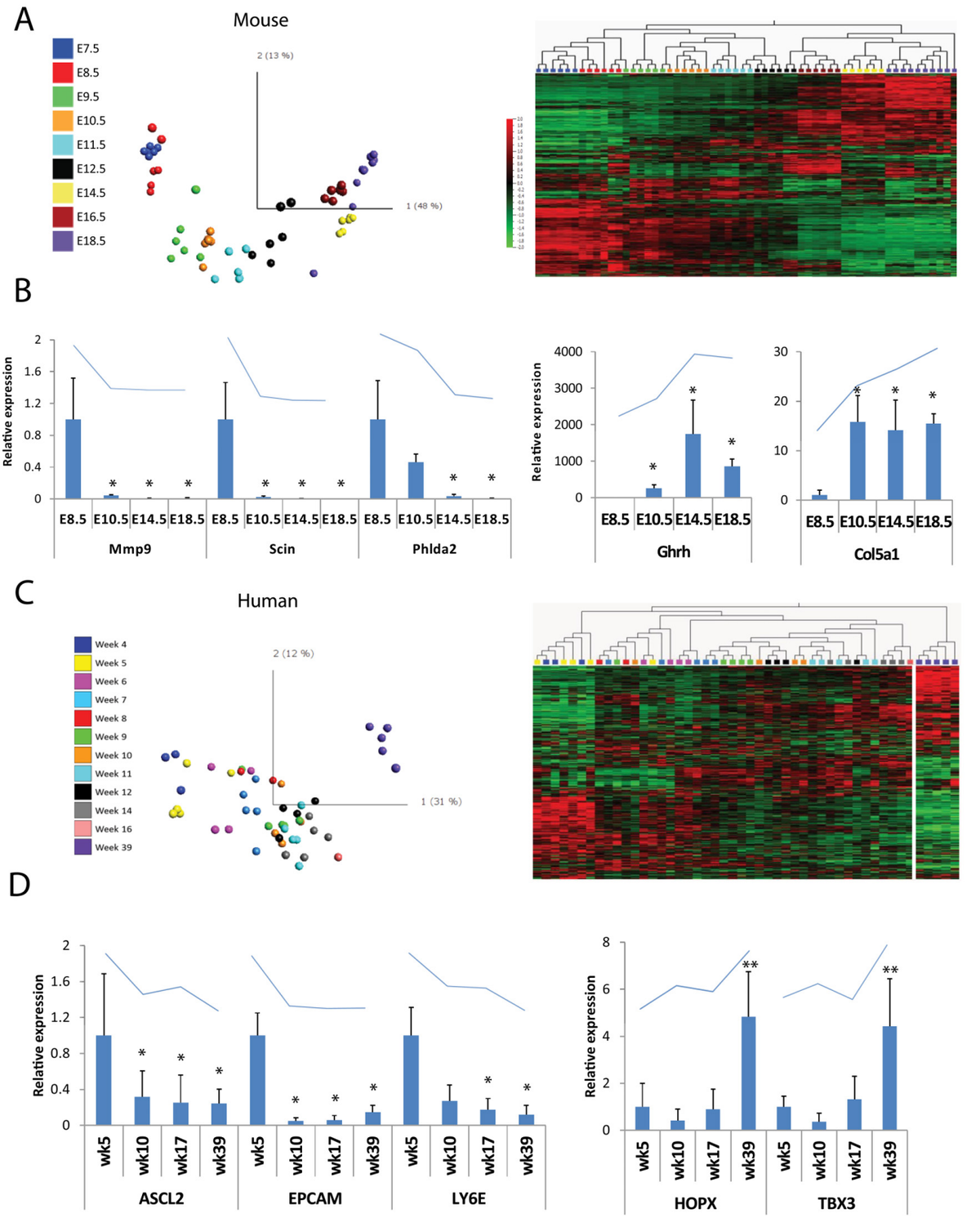
Gene_ID	Sequence
h18S F	CGCCGCTAGAGGTGAAATTCT
h18S R	CGAACCTCCGACTTTCGTTCT
hASCL2 F	CACTGCTGGCAAACGGAGAC
hASCL2 R	AAAACTCCAGATAGTGGGGGC
hCDX2 F	TTCACTACAGTCGCTACATCACC
hCDX2 R	TTGATTTTCCTCTCCTTTGCTC
hEPCAM F	AATCGTCAATGCCAGTGTACTT
hEPCAM R	TCTCATCGCAGTCAGGATCATAA
hHOPX F	GACAAGCACCCGGATTCCA
hHOPX R	GTCTGTGACGGATCTGCACTC
hLY6E F	GGGAATCTCGTGACATTTGGC
hLY6E R	ACACCAACATTGACGCCTTCT
hOCT4 F	TGGGCTCGAGAAGGATGTG
hOCT4 R	GCATAGTCGCTGCTTGATCG
hTBX3 F	CCCGGTTCCACATTGTAAGAG
hTBX3 R	GTATGCAGTCACAGCGATGAAT
hTEAD4 F	CAGTATGAGAGCCCCGAGAA
hTEAD4 R	TGCTTGAGCTTGTGGATGAA
hTP63 F	CTGGAAAACAATGCCCAGA
hTP63 R	AGAGAGCATCGAAGGTGGAG
hVGLL1 F	CTCCCGGCTCAGTTCACTATAA
hVGLL1 R	CCCAGTGGTTTGGTGGTGTA
m18S F	CGCGGTTCTATTTTGTTGGT
m18S R	AACCTCCGACTTTCGTTCTTG
mCol5a1 F	AGGGAGCCAGAATCACTTCTT
mCol5a1 R	GCATCCACATAGGAGAGCAGT
mGhrh F	CACAACATCACAGAGTCCCACC
mGhrh R	CTGGTGAGGATGAGGATCACAA
mMmp9 F	GCCGACTTTTGTGGTCTTCC
mMmp9 R	TACAAGTATGCCTCTGCCAGC
mPhlda2 F	TGCAAGACTTTCCCCGCTAC
mPhlda2 R	GTGCGTTTCACGGACCCA
mScin F	CCAGAATTGTGGAGGTTGACG
mScin R	GCCATTGTTTCGTGGCAGTT

Table S4. Model errors

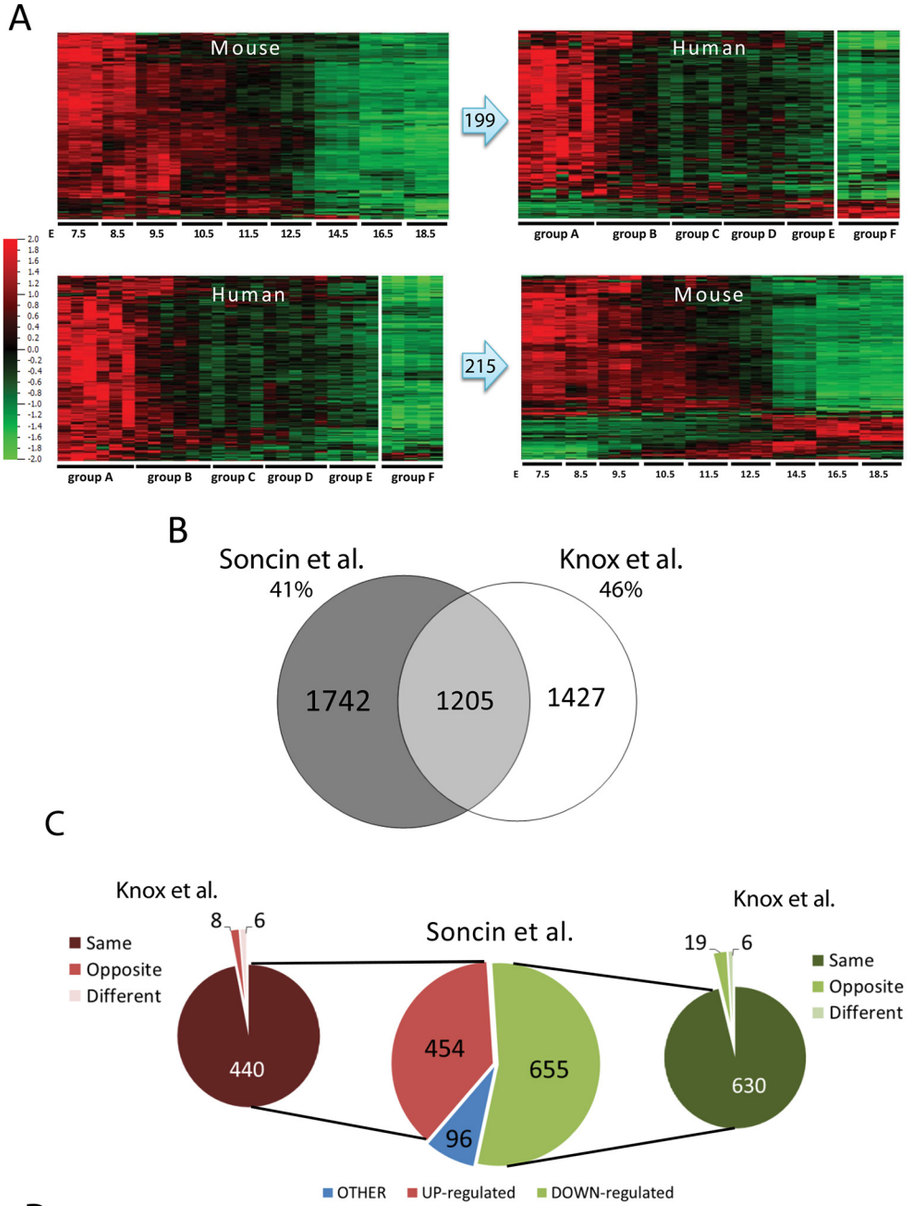
Cluster #	Error_value	Cluster #	Error_value	Cluster #	Error_value
Ms_1	1.7058	**HL_1**	0.2514	**HS_1**	0.2867
Ms_2	0.2037	**HL_2**	0.4030	**HS_2**	0.5153
Ms_3	0.6422	**HL_3**	1.5269	**HS_3**	1.1920
Ms_4	1.2510	**HL_4**	1.8015	**HS_4**	1.2262
Ms_5	1.6053	**HL_5**	1.7795	**HS_5**	1.3352
Ms_6	1.4172	**HL_6**	2.2254	**HS_6**	1.3348
Ms_7	2.6252				



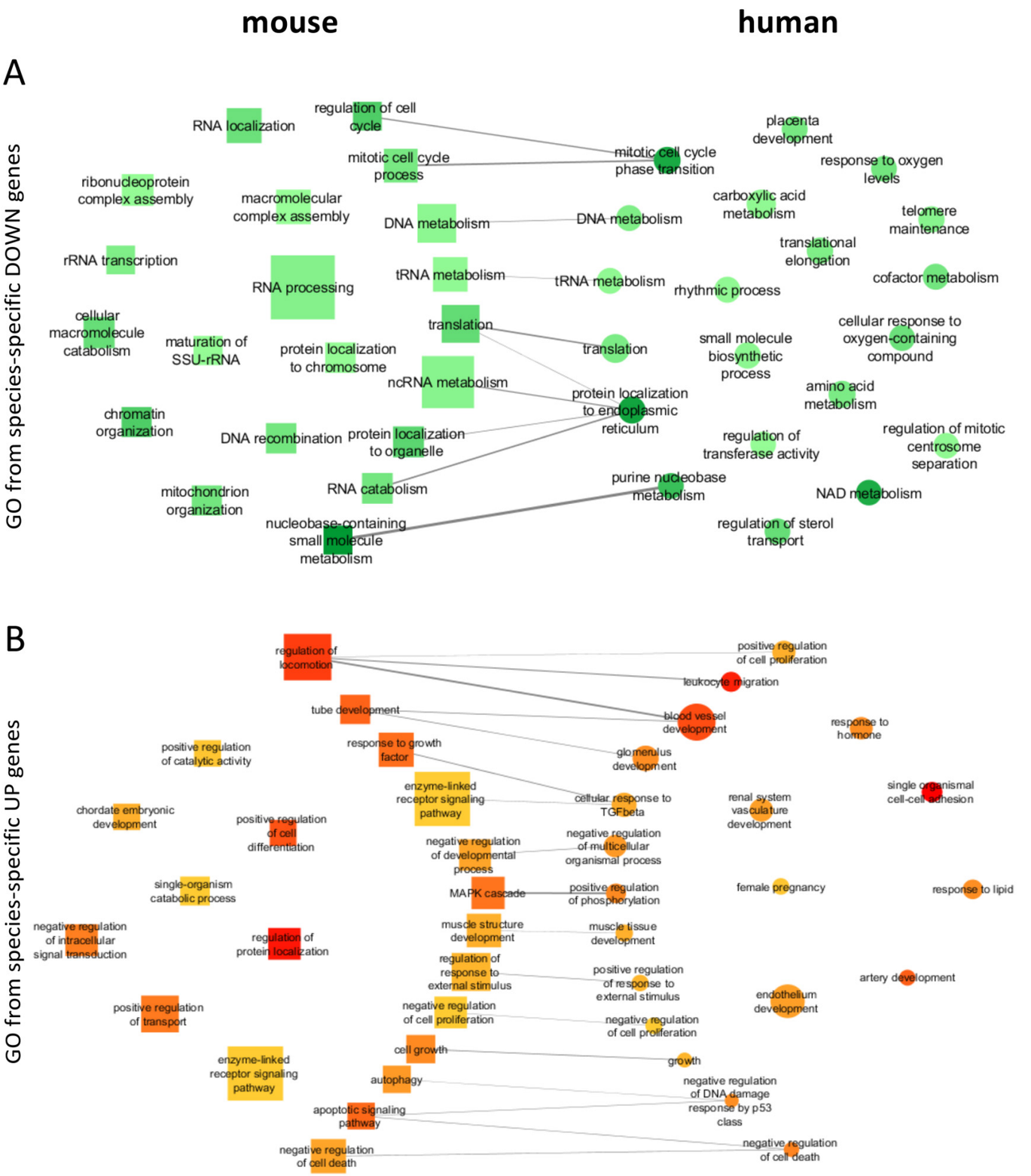
Supplementary Figure S1. In-situ hybridization using EOMES-specific probes on positive control human tissues.



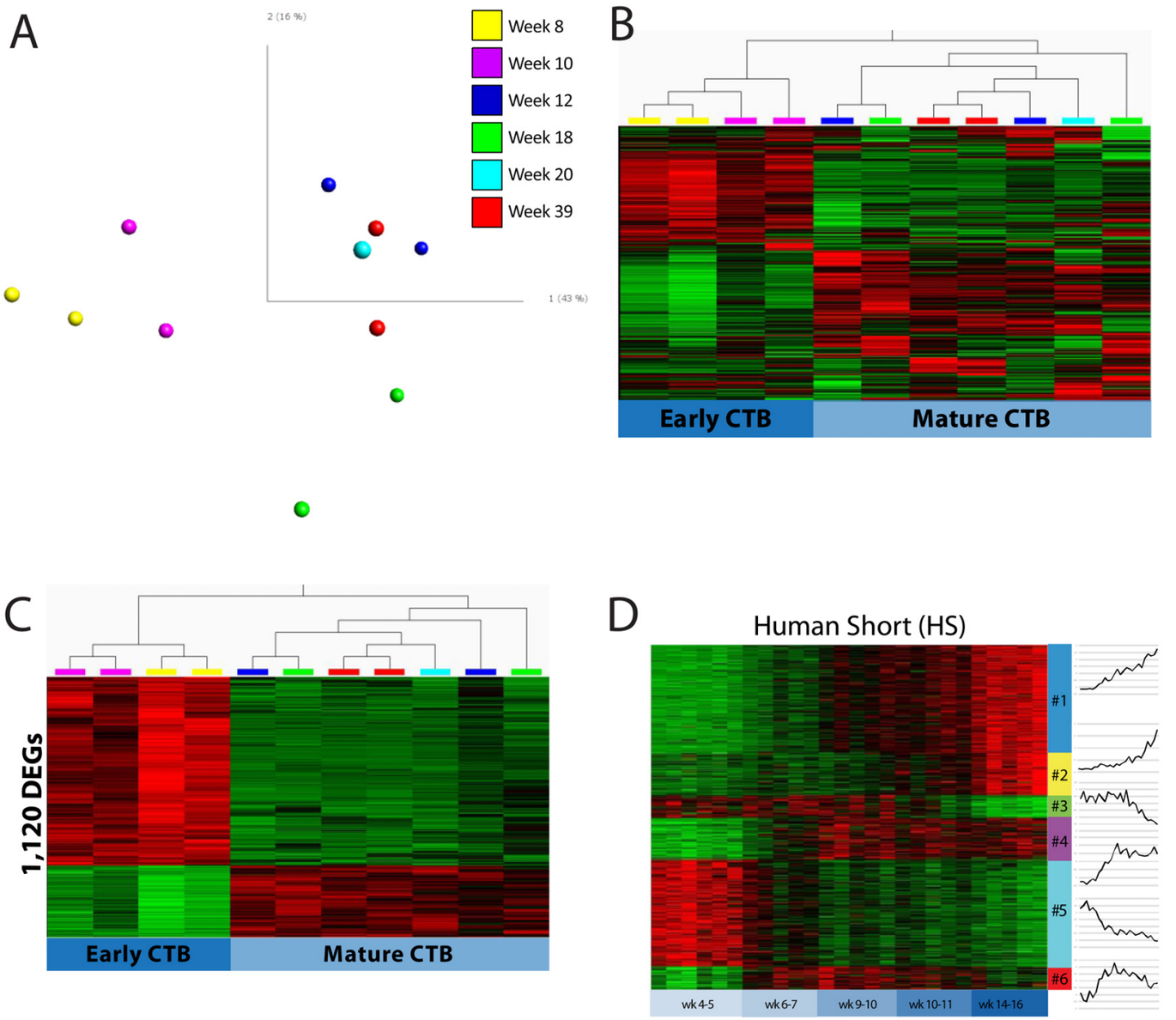
Supplementary Figure S2. Microarray data analysis of mouse and human placenta samples. A. Principle component analysis (PCA) and heatmap of mouse dataset using Qlucore (variance 0.02). Samples were colour-coded according to gestational age. **B.** qRT-PCR analysis (bar graph) for Mmp9, Scin, Phlda2, Ghrh, and Col5a1 in mouse placentas at E8.5, E10.5, E14.5 and E18.5 confirmed expression pattern observed in the microarray (line above bar graph). **C.** PCA and heatmap of human dataset using Qlucore (variance 0.02). Samples were colour-coded according to gestational age. **D.** qRT-PCR analysis (bar graph) for ASCL2, EPCAM, LY6E, HOPX, and TBX3 in human placenta from gestational week 5, 10, 17, and 39 confirmed expression pattern observed in the microarray (line above bar graph). n=3 * $p<0.05$ compared to E8.5 for mouse and Week 5 for human; ** $p<0.05$ compared to week 10.



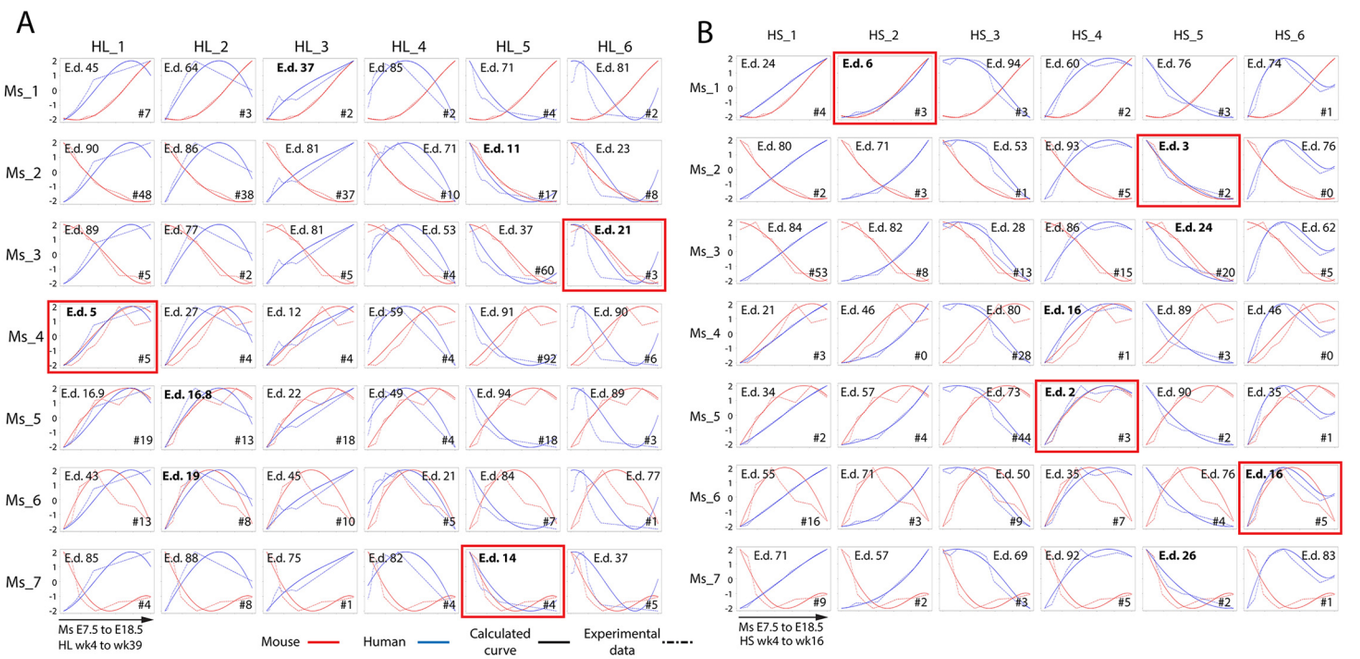
Supplementary Figure S3. Comparison between datasets. A. Top: Heatmap of mouse differentially expressed genes (DEG) downregulated with gestational age (199 genes, left) and their expression pattern in human (heatmap on the right). Bottom: Heatmap of human DEGs (215 genes, left) downregulated with gestational age and their expression pattern in mouse (heatmap on the right). These data were used to create the pie chart in Fig.3D. **B.** Venn diagram representing the overlapping of DEGs between our mouse data and the Knox et al. dataset. We analysed the mouse dataset published by Knox and Baker (2008) using similar parameters applied to our inter-species analysis (v 0.02, $q<0.05$, FC>2.0). We identified 2,632 DEGs in the Knox et al. dataset, of which 1,205 were in common with our analysis, representing over 40%. **C.** Pie charts showing the comparison in expression pattern direction of the common DEGs (1,205) between our data and that of Knox et al. Central pie chart shows the expression pattern in the current study. Genes with expression pattern other than UP and DOWN-regulation across gestation were combined and labelled as "Other." The side charts show the expression pattern comparison between the sub-classes UP- (red) and DOWN- (green) regulated genes in our analysis and their expression in the Knox et al. dataset. Over 96% of the common DEGs showed the same pattern of expression (either UP or DOWN) in both datasets. **D.** Expression levels of CDX2 and ELF5 in human placentas across gestation from our microarray dataset. Note the high variability of CDX2 expression between samples.



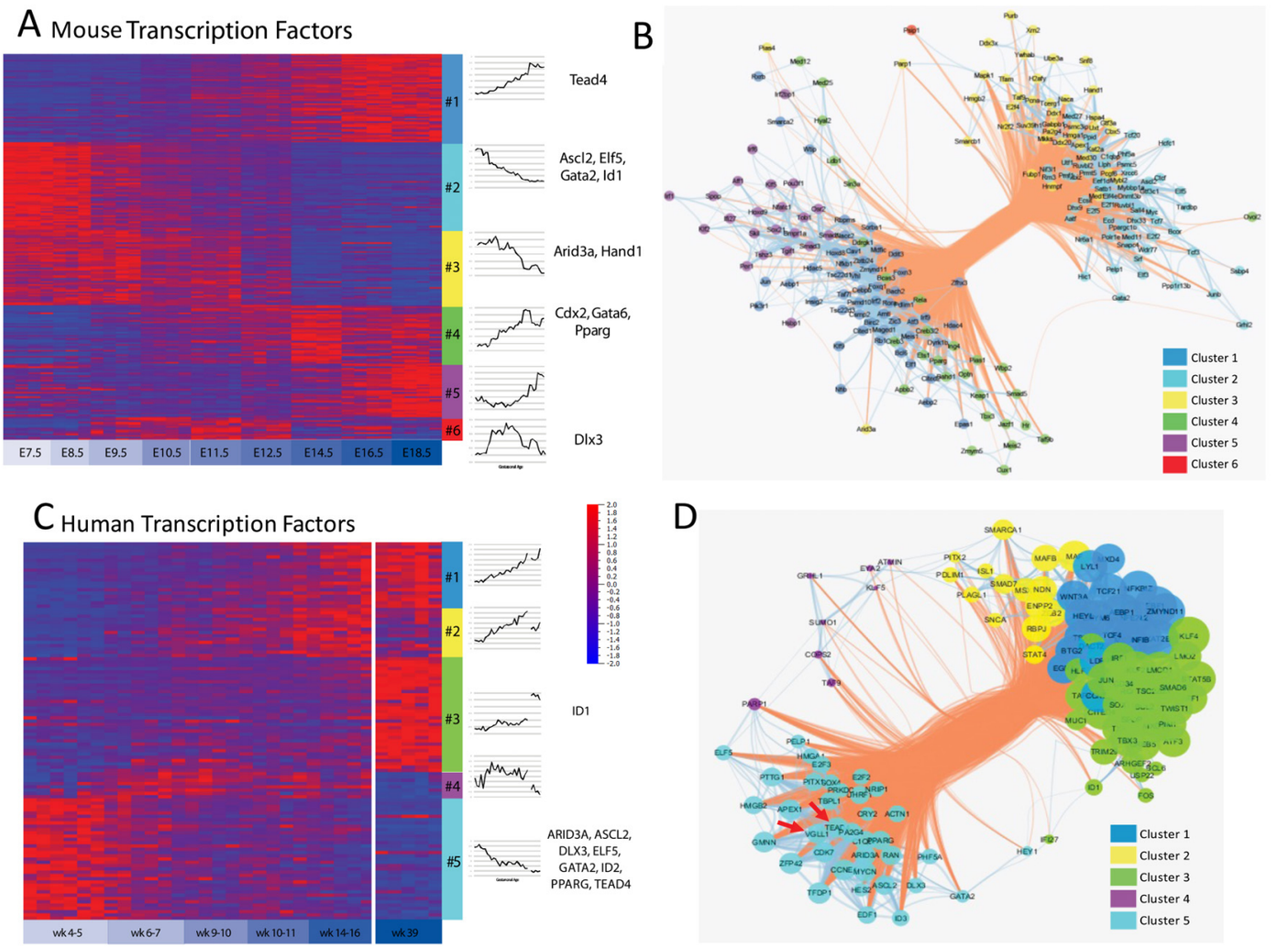
Supplementary Figure S4. Enrichment analysis of Gene Ontology Biological Functions of species-specific up and down-regulated genes in mouse and human placentas. A. Gene ontology of genes downregulated across gestation. **B.** Gene ontology of genes upregulated across gestation. Node shape represents the species (mouse – square; human – circle). The size of the nodes is correlated to the log(p) of the enriched term. Lines (edges) connect mouse and human terms that include overlapping functions. The thickness of the edges is proportional to the number of terms in common between the species.



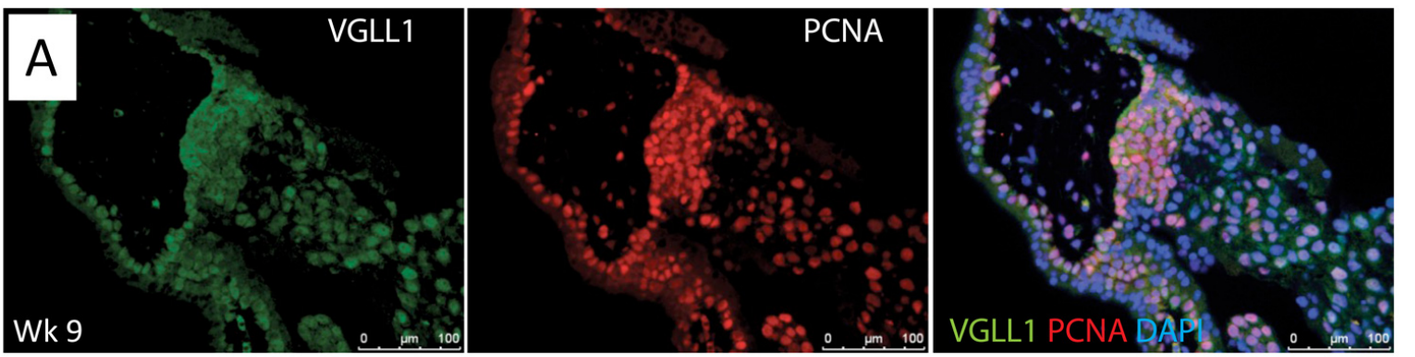
Supplementary Figure S5. Details of microarray data analysis of human primary CTB and placental tissues. A. Principal component analysis (PCA) of primary CTB expression profiles. Note separation into two main groups along PCA1: weeks 8 and 10 on one side and weeks 12, 18, 20, and 39 on the other. **B.** Heatmap of human primary CTB genes filtered for variance (0.05). Note hierarchical clustering of samples into two groups, labelled "early" CTB and "mature" CTB, which were used for subsequent differential expression analysis. **C.** Heatmap of 1,120 differentially expressed genes (DEGs) between the two identified groups of CTB (two-group analysis $q<0.01$). Of these DEGs, 197 genes were also differentially expressed in human placentas across gestation and included in the AP analysis (see Figure 6B). **D.** Affinity Propagation analysis of the 665 DEGs in the "short" human dataset, which did not include term samples. AP distinguished 6 clusters of co-regulated genes. Heatmap represents expression values of genes; line graphs (right) represent the average trace for each cluster.

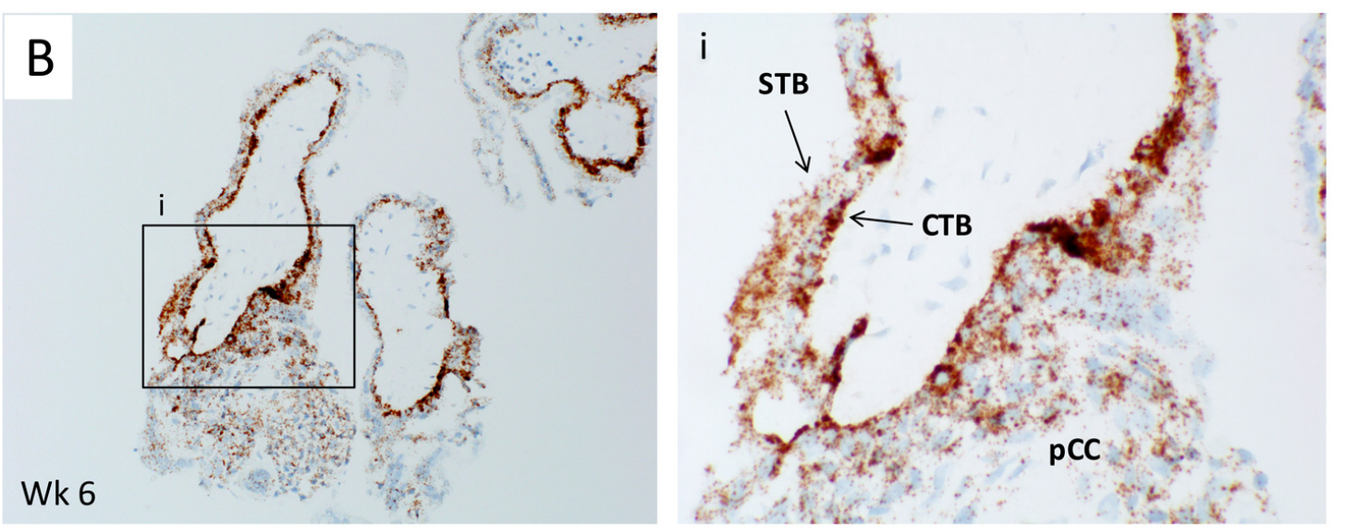


Supplementary Figure S6. Comparison of expression profiles of all mouse and human co-expression clusters. A. Matrix of graphs showing the relationship between each mouse and human AP co-expression cluster (from the HL dataset, which included data from term human placentae). **B.** Matrix of graphs showing the relationship between each mouse and human AP co-expression cluster (from the HS dataset, which did not include data from term human placentae). For both matrices, both the experimental data (dashed lines) and corresponding best-fit curves (solid lines) for each AP cluster are shown. Data from mouse clusters are shown in red and data from human clusters are shown in blue. The Euclidean distance (E.d) values for each pair of curves is shown. # = number of genes in common between the mouse and human clusters. For each mouse cluster, the HL or HS cluster with the most similar pattern of expression of gestation, as measured by the Euclidean distance, is boxed in red.

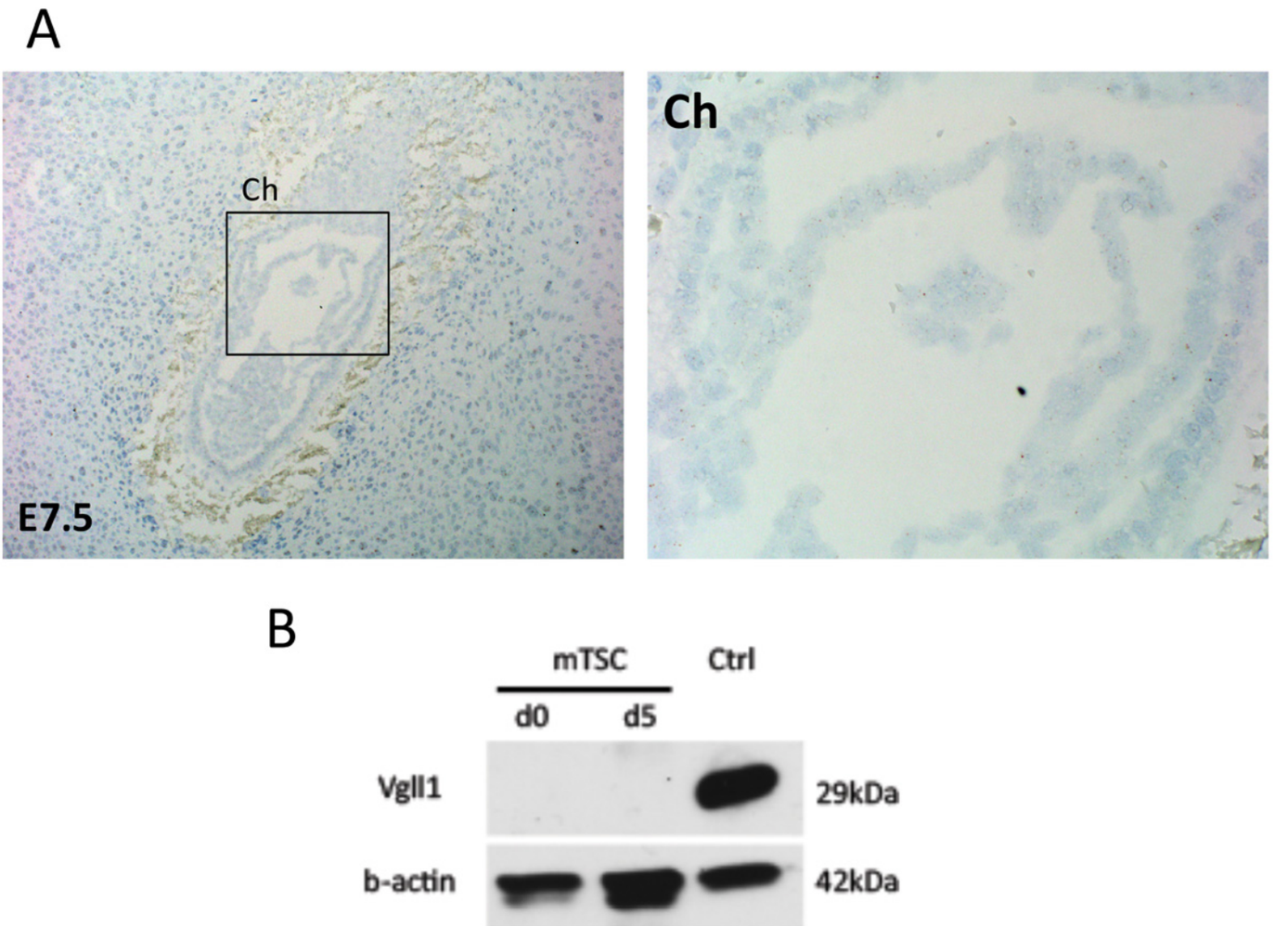


Supplementary Figure S7. A. Affinity Propagation (AP) analysis of transcription factors (TFs) in the mouse dataset distinguished 6 clusters of co-regulated genes. Heatmap represents expression values of TFs; line graphs (right) represent the average trace for each cluster. Representative gene(s) in each cluster are shown. **B.** Mouse TF cluster network. For clarity, node size was set at a fixed value. Nodes are colour-coded according to the cluster they belong to; red edges indicate negative correlation, while blue edges show positive correlation between nodes. **C.** Affinity Propagation (AP) analysis of transcription factors in the human dataset distinguished 5 clusters of co-regulated genes. Heatmap representing expression values of TFs; line graphs (right) represent the average trace for each cluster. Representative gene(s) in each cluster are shown. **D.** Human TF cluster network. Nodes are colour-coded according to the AP cluster number, while their size represents the positive sum scores. Red edges indicate negative correlation, while blue edges show positive correlation between nodes. Edge thickness identifies Pearson correlation absolute value. Red arrows show the position of TEAD4 and VGLL1.

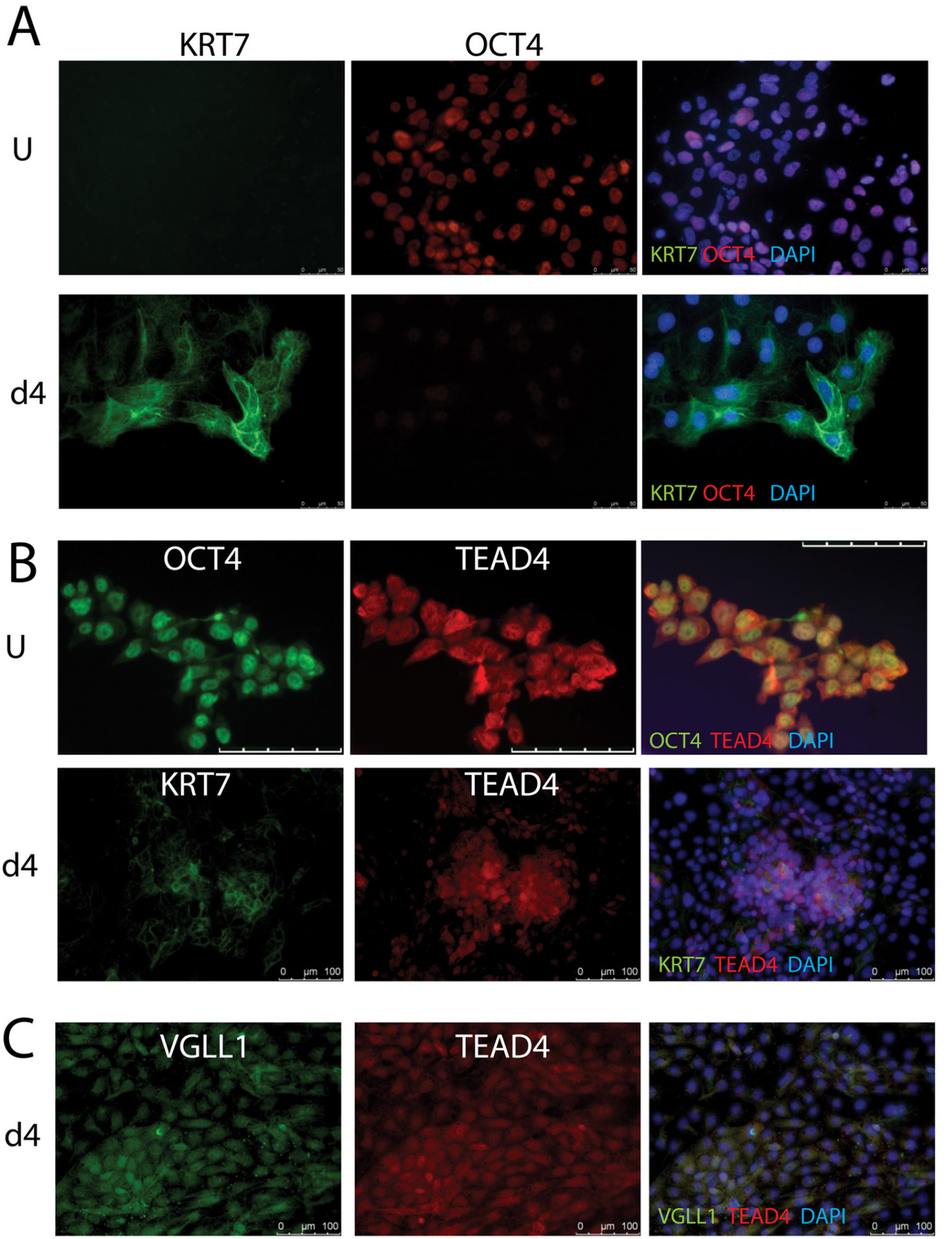




Supplementary Figure S8. IHC and ISH of first trimester human placentas. A. Co-staining of week 9 placenta with VGLL1 (green), the proliferation marker PCNA (red), and Dapi (blue). **B.** In situ hybridization with VGLL1-specific probes in week 6 placenta. Expression of VGLL1 mRNA is noted in cytotrophoblast (CTB), syncytiotrophoblast (STB), as well as in trophoblast of the proximal cell column (pCC), although it is most enriched in CTB. Magnification: Left – 100x; Right: 300x.



Supplementary Figure S9. Vgll1 is not expressed in mouse placenta and trophoblast. A. No Vgll1 RNA was noted in E7.5 placentas by in-situ hybridization. Ch= chorion. Magnification: 50x (left); 200x (right). **B.** Western blot analysis of undifferentiated (d0) and differentiated (day 5/d5) mouse trophoblast stem cells (mTSC) for Vgll1 and β-actin (loading control). Commercially available mouse whole stomach lysate (Abcam) was used as positive control.



Supplementary Figure S10. Expression of OCT4, KRT7 and TEAD4 in the hESC-based *in vitro* model of human trophoblast differentiation. A. Immunostaining for OCT4 (red) and KRT7 (green) in pluripotent (undifferentiated) H9 hESC (U) and in hESC-derived CTB-like cells (d4). OCT4 and KRT7 in this model are mutually exclusive. **B.** Immunostaining for TEAD4, along with either OCT4 in undifferentiated hESC (U) or with KRT7 in hESC-derived CTB-like cells (d4). **C.** Double immunostaining of hESC-derived CTB-like cells (d4) with VGLL1 and TEAD4, showing co-localization in nuclei of these cells.

Table S1. Genelists

Click here to Download Table S1

Table S2. GO enrichment

Click here to Download Table S2

Table S3. q-RT-PCR primers

Gene_ID	Sequence
h18S F	CGCCGCTAGAGGTGAAATTCT
h18S R	CGAACCTCCGACTTTCGTTCT
hASCL2 F	CACTGCTGGCAAACGGAGAC
hASCL2 R	AAAACTCCAGATAGTGGGGGC
hCDX2 F	TTCACTACAGTCGCTACATCACC
hCDX2 R	TTGATTTTCCTCTCCTTTGCTC
hEPCAM F	AATCGTCAATGCCAGTGTACTT
hEPCAM R	TCTCATCGCAGTCAGGATCATAA
hHOPX F	GACAAGCACCCGGATTCCA
hHOPX R	GTCTGTGACGGATCTGCACTC
hLY6E F	GGGAATCTCGTGACATTTGGC
hLY6E R	ACACCAACATTGACGCCTTCT
hOCT4 F	TGGGCTCGAGAAGGATGTG
hOCT4 R	GCATAGTCGCTGCTTGATCG
hTBX3 F	CCCGGTTCCACATTGTAAGAG
hTBX3 R	GTATGCAGTCACAGCGATGAAT
hTEAD4 F	CAGTATGAGAGCCCCGAGAA
hTEAD4 R	TGCTTGAGCTTGTGGATGAA
hTP63 F	CTGGAAAACAATGCCCAGA
hTP63 R	AGAGAGCATCGAAGGTGGAG
hVGLL1 F	CTCCCGGCTCAGTTCACTATAA
hVGLL1 R	CCCAGTGGTTTGGTGGTGTA
m18S F	CGCGGTTCTATTTTGTTGGT
m18S R	AACCTCCGACTTTCGTTCTTG
mCol5a1 F	AGGGAGCCAGAATCACTTCTT
mCol5a1 R	GCATCCACATAGGAGAGCAGT
mGhrh F	CACAACATCACAGAGTCCCACC
mGhrh R	CTGGTGAGGATGAGGATCACAA
mMmp9 F	GCCGACTTTTGTGGTCTTCC
mMmp9 R	TACAAGTATGCCTCTGCCAGC
mPhlda2 F	TGCAAGACTTTCCCCGCTAC
mPhlda2 R	GTGCGTTTCACGGACCCA
mScin F	CCAGAATTGTGGAGGTTGACG
mScin R	GCCATTGTTTCGTGGCAGTT

Table S4. Model errors

Cluster #	Error_value	Cluster #	Error_value	Cluster #	Error_value
Ms_1	1.7058	**HL_1**	0.2514	**HS_1**	0.2867
Ms_2	0.2037	**HL_2**	0.4030	**HS_2**	0.5153
Ms_3	0.6422	**HL_3**	1.5269	**HS_3**	1.1920
Ms_4	1.2510	**HL_4**	1.8015	**HS_4**	1.2262
Ms_5	1.6053	**HL_5**	1.7795	**HS_5**	1.3352
Ms_6	1.4172	**HL_6**	2.2254	**HS_6**	1.3348
Ms_7	2.6252				