1 Comparison analysis on transcriptomic of different human trophoblast development model

2

3 Yajun Liu ^{1,2*}, Yilin Guo ¹, Ya Gao ¹, Guiming Hu ³, Jingli Ren ³, Jun Ma ⁴, Jinquan Cui ^{1,2}

4

5 1. Department of Obstetrics and Gynecology, the Second Affiliated Hospital of Zhengzhou
6 University, Zhengzhou, China;

7 2. Academy of Medical Sciences of Zhengzhou University Translational Medicine platform,
8 Zhengzhou University, No.100 Science Avenue, Zhengzhou, China;

9 3. Department of Clinical Laboratory, the Second Affiliated Hospital of Zhengzhou University,
10 Zhengzhou, China;

11 4. Department of Pathology, the Second Affiliated Hospital of Zhengzhou University, Zhengzhou,
12 China; Postcode: 450001

13

14 *Corresponding authors

15 Yajun Liu

16 Contact information: Second Affiliated Hospital of Zhengzhou University, Henan No. 2, Jingba

17 Road, Zhengzhou, China. e-mail : liuyajun_biology@126.com

18

19 ORCID for corresponding author:

20 Yajun Liu 0000-0002-8203-5762

21

22

23 Abstract

24 **Aims:** Multiple models of trophoblastic cell development were developed. However, systematic
25 comparisons of these cell models are lacking.

26 **Methods and Results:** In this study, first-trimester chorionic villus and decidua tissues were
27 collected. Transcriptome data was acquired by RNA-seq and the expression levels of trophoblast
28 specific transcription factors were identified by immunofluorescence and RNA-seq data analysis.

1 Differentially expressed genes between chorionic villus and decidua tissues and its related
 2 biological functions were identified. We identified genes that were relatively highly expressed and
 3 enriched transcription factors in trophoblast cells of different trophoblast cell models.

4 **Conclusions:** This analysis is of certain significance for further exploration of the development of
 5 placenta and the occurrence of pregnancy-related diseases in the future. The datasets and
 6 analysis provide a useful source for the researchers in the field of the maternal-fetal interface and
 7 the establishment of pregnancy.

8 **Keywords :** placenta, RNA-seq, transcription factor

1

2 **Introduction**

3

4 Implantation failure and insufficient placental development are important causes of female
5 infertility, recurrent miscarriage, and other pregnancy-related problems [1]. Based on different
6 trophoblastic cell models, many molecular mechanisms for the establishment and maintenance
7 of pregnancy have been obtained. In particular, the early stages of pregnancy have a significant
8 impact on pregnancy outcomes [2] [3] [4]. Models of trophoblast-like cells differentiation from
9 stem cells provide insights into the field. In previous studies, this model was compared with the
10 transcriptome of primary cytotrophoblast recovered from term placenta trophoblastic cells ⁴.
11 Due to Placental tissues at different stages of development are quite different, transcriptome
12 data from villi in early pregnancy could provide further insights into this area.

13 In this study, we collected human first-trimester chorionic villus and decidual tissue from the
14 same patient, performed high-throughput RNA sequencing. In particular, we analysis the
15 expression of important trophoblastic cell-specific factors. Next, highly expressed genes in
16 different trophoblastic cell models, including hESC line cells after BMP4 treatment (TB)
17 comparison with H1 [6] and trophectoderm (TE) in comparison with pluripotent epiblast (EPI)
18 cells [7], chorionic villus (CV) in comparison with decidua (DC)), were identified by differential
19 expression gene analysis. The transcription factors enriched in these models were then identified
20 by newly developed tools BART [8] tool, which provides functional interpretations to differential
21 gene expression analysis. Figure 1 shows the experiment and analysis process of this study. Table
22 5 Summary of datasets analyzed in this study.

23

1

2 **Materials and Method**

3 Collection protocol

4 The first-trimester placenta (six to nine-week of gestation from 6 healthy women, confirmed
5 by the embryo size under detection of ultrasound) was collected in the Second Affiliated Hospital
6 of Zhengzhou University. After being separated ex-vivo, the placenta was immediately washed in
7 the ice saline and divided into several parts by a scalpel blade on ice, and then were transformed
8 into cryogenic vials which were filled with 1 ml RNA store beforehand. When disposed of properly,
9 the samples were put into 4°C refrigerator for 24h to let the RNA store immerse them. After 24h
10 the RNA store was abandoned and the samples were sopped up using sterile absorbing paper,
11 then the samples were separately collected into new cryogenic vials and were store into -80°C
12 refrigerator for further investigations.

13

14 HE (Hematoxylin & Eosin) staining

15 Samples collected before were fixed in formalin, embedded in paraffin and sliced up to 4 µm
16 sections, and then were deparaffinized and rehydrated. The deparaffin and rehydration protocols
17 are xylene I for 5 min, xylene II for 5 min, 100% ethanol for 2min, 95% ethanol for 1min, 80%
18 ethanol for 1min, 75% ethanol for 1min, and finally distilled water for 2min. After the process
19 above, the sections were stained in hematoxylin for 5 min and rinsed with tap water, then
20 differentiated in hydrochloric acid and ethanol for the 30s respectively. At last, the sections were
21 soaked into tap water for 5min and sealed in neutral resins with cover glass, then were observed
22 under an ordinary optical microscope, 10 pictures were randomized obtained per section.

23

24 Immunohistochemistry

25 Paraffin-embedding, deparaffinized, and the rehydrating process is the same as HE staining.
26 After these steps, the sections were subjected to antigen retrieval for 3min in a medical pressure
27 cooker with citration solution (pH=6), subsequently treated with endogenous catalase blocker
28 and horse serum to eliminate the interference of endogenous catalase and nonspecific staining.
29 The sections were then incubated with primary antibody. Next day the sections were washed in

1 PBS and then incubated in the secondary antibody (1:200, Abbkine) for 1 hour in room
2 temperature, and after DAPI staining, the sections were observed under the fluorescence
3 microscope, pictures were randomized obtained per section.

4

5 RNA-seq experiment

6 Total RNA was extracted with Trizol (Tiangen, Beijing) and assessed with Agilent 2100
7 BioAnalyzer (Agilent Technologies, Santa Clara, CA, USA) and Qubit Fluorometer (Invitrogen).
8 Total RNA samples that meet the following requirements were used in subsequent experiments:
9 RNA integrity number (RIN) > 7.0 and a 28S:18S ratio > 1.8. RNA-seq libraries were generated and
10 sequenced by CapitalBio Technology (Beijing, China). The triplicate samples of all assays were
11 constructed an independent library, and do the following sequencing and analysis. The NEB Next
12 Ultra RNA Library Prep Kit for Illumina (NEB) was used to construct the libraries for sequencing.
13 NEB Next Poly(A) mRNA Magnetic Isolation Module (NEB) kit was used to enrich the poly(A)
14 tailed mRNA molecules from 1 µg total RNA. The mRNA was fragmented into ~200 base pair
15 pieces. The first-strand cDNA was synthesized from the mRNA fragments reverse transcriptase
16 and random hexamer primers, and then the second-strand cDNA was synthesized using DNA
17 polymerase I and RNaseH. The end of the cDNA fragment was subjected to an end repair process
18 that included the addition of a single “A” base, followed by ligation of the adapters. Products
19 were purified and enriched by polymerase chain reaction (PCR) to amplify the library DNA. The
20 final libraries were quantified using KAPA Library Quantification kit (KAPA Biosystems, South
21 Africa) and an Agilent 2100 Bioanalyzer. After quantitative reverse transcription-polymerase chain
22 reaction (RT-qPCR) validation, libraries were subjected to paired-end sequencing with pair-end
23 150-base pair reading length on an Illumina NovaSeq 6000.

24

25 RNA-seq data analysis

26 Transcript abundance was quantified using Kallisto (Bray et al., 2016) and gene fold changes
27 were generated by comparing gene expression levels between two groups using the limma R
28 package (Ritchie et al., 2015). Figure S3 shows library size analysis results. P-value or q-value was
29 used to conduct a significance analysis. Parameters for classifying significantly DEGs are >2-fold

differences ($|\log_2FC| > 1$, FC: the fold change of expressions) in the transcript abundance and $q < 0.05$. Gene fold changes were transformed using $\log_2$ and displayed on the x-axis; P-values were corrected using the Benjamini-Hochberg method, transformed using $-\log_{10}$, and displayed on the y-axis. The functional enrichment analysis was performed using g: Profiler (version e99_eg46_p14_f929183) with g: SCS multiple testing correction methods applying a significance threshold of 0.05 (Raudvere et al., 2019). Hierarchical clustering of arbitrary types of objects from a matrix of distances and shows a corresponding dendrogram based on Orange (Demšar et al., 2013). The parameter setting could be found in Figure S8.

9

10 2D mapping

Orange [11], which provided a wrapper for scikit-learn algorithms [12], was used for batch effect remove, filtering (by cells and genes), scaling, normalization, clustering, dimensionality reduction, clustering and visualize cell clusters using, t-SNE, and PCA.

14

15 Identify putative transcription factors regulating differentially expressed genes

We used the transcription factor prediction tool BART¹². BART was run with all default settings, and the provided transcription factor databases. BART was run with all default settings, and the provided transcription factor databases.

19

20

21

1

2 Results

3 The expression level of trophoblast specific transcription factors in Chorionic villus

4 Chorionic villus exhibits typical chorionic villus tissue morphology according to hematoxylin
5 and eosin stain (H&E tissue sections) (Figure 2A). We have previously identified specific
6 transcription factors (including *JUN*, *FOS*, *TFAP2A*, *TFAP2C*, *TEAD1*, *TEAD3*, *TEAD4*, *GATA2*, *GATA3*)
7 in trophoblast cells based on DNAs-seq derived from ENCODE project [14]. The results showed
8 that *JUN*, *FOS*, *TFAP2A*, *TFAP2C*, *TEAD4*, *GATA2*, *GATA3* were most strongly expressed in chorionic
9 villus (Figure 2B). *TEAD1*, *TEAD3* can be detected, though at lower levels than other genes. As a
10 control, the well-studied chorionic villus surface marker *HLA-G* and *KRT7* were also expressed in
11 the chorionic villus (Figure S1A, B).

12

13

14 RNA-Sequencing of first-trimester chorionic villus and decidua

15 Transcriptome data of chorionic villus and decidua from six first-trimester donors were
16 obtained by high-throughput sequencing. Table 1 listed the top 10 highly expressed genes in six
17 human chorionic villus tissues. Among them, human chorionic gonadotropin (hCG) family genes
18 including *CGA*, *CGB5*, *CGB8*, *CGB3*, and pregnancy-specific glycoproteins (PSGs) [16] family genes
19 including *PSG3*, *PSG1* are the most abundantly expressed genes in chorionic villus samples (Table
20 S1, Figure S2A). Table 2 listed the top 10 highly expressed genes in corresponding decidua
21 tissues.

22 Then, we identify the sex of the cell line based on the recently determined gender-specific
23 transcript markers, *RPS4Y1*, *EIF1AY*, *DDX3Y*, *KDM5D* (Staedtler et al., 2013). The results indicate
24 that 10_4, 11_1, 12_2 is derived from males, and 13_4, 9_4 is derived from the female (Figure
25 S2B). We then compared the differentially expressed genes between the chorionic villus (13_4,
26 9_4, 10_4, 11_1, 12_2) and decidua (10_E,11_E,12_B,13_D,8_C,9_E) groups. Volcano plots are
27 used to display the differential expressions of genes in each group (Figure 2C). Each point in the
28 scatter plot represents a gene; the axes display the significance versus fold-change estimated by
29 the differential expression analysis. Our analysis identified that 1315 genes were up-regulated in

1 chorionic villus compared with decidua and 2272 genes were up-regulated in decidua compared
2 with chorionic villus (log2FoldChange, $p < 0.05$) (Table S3). The 10 top genes (logFC>2, $p < 0.05$)
3 significantly upregulated in chorionic villus compared with decidua are shown in Table 3. The top
4 10 genes (logFC>2, $p < 0.05$) significantly up-regulated in decidua compared with chorionic villus
5 are shown in Table 4. Kisspeptin and its receptors, highly expressed in chorionic villus, play an
6 essential role in the establishment of the maternal-fetal dialogue.

7 Next, we analyzed the expression of trophoblastic cell-specific transcription factors (Knott
8 and Paul, 2014) and putative surface marker genes in chorionic villus (Figure 2D, 2E). Our analysis
9 was based on the system of stem cell differentiation into trophoblast cells (ESCd1 (Yabe et al.,
10 2016), ESCd2 (Krendl et al., 2017)), immortalized trophoblast cells (HTR8/SVneo (Lee et al., 2016),
11 Bewo (Renaud et al., 2015), Jeg3 (Ferreira et al., 2016)), and the chorionic villus collected in this
12 study. For the corresponding gene expression matrix, please refer to Table S2. Interestingly, genes
13 from chorionic gonadotropin (hCG) family genes are highly expressed in both chorionic villus and
14 cell lines. While, pregnancy-specific glycoproteins (PSGs) family genes are highly expressed only
15 in chorionic villus, and at a low expression level in vitro cell. Trophoblast-specific transcription
16 factors are highly expressed in chorionic villus and showed heterogeneous expression level in
17 immortalized cell lines and stem cell differentiation systems. Immunofluorescence results
18 confirmed that *TFAP2C* was detectable in immortalized trophoblastic cell lines JEG3 in vitro
19 (Figure S1C).

20 Enrichment analysis revealed enriched GO terms in the top 500 upregulated genes in
21 chorionic villus (Figure S4) including "reproductive process", "multicellular organismal process",
22 "hormone activity", "animal organ morphogenesis", "female pregnancy", "embryonic
23 morphogenesis", "placenta development", "cis-regulatory region binding"[17] et al. Not
24 surprisingly, enriched GO terms of upregulated genes in the decidua, the primary place for the
25 establishment and communication of pregnancy immune microenvironment, include many
26 immune-related terms such as "regulation of immune system process", "positive regulation of
27 immune system process" and "immune response" et al (Figure S5).

28 KEGG pathway analysis of the top 500 upregulated genes in chorionic villus revealed
29 previously identified trophoblast related pathways "PPAR signaling pathway" (Barak et al., 2008),

(Figure S4). KEGG analysis of the top 500 upregulated genes in decidua shows that immune-related pathways approximately occupying the top ten pathways, including “Antigen processing and presentation”, “Cytokine-cytokine receptor interaction”, and “Th17 cell differentiation” (Figure S5). More detailed analysis results on GO and KEGG enrichment analysis could be found in Supplementary file 1.

PCA and t-SNE function similarly in identifying a small number of samples.

PCA and t-SNE analysis was employed to identify global patterns of the gene expression profile of six chorionic villus and corresponding decidua tissues. The results show an overall similarity result generated by PCA and the t-SNE map (Figure 3A). Both methods correctly clustered 12 samples into two types, namely chorionic villus and decidua. Nevertheless, t-SNE can distinguish these two groups more significantly into different tissue types, and the same type of tissues is more closely clustered together (Figure 3B).

Comparison of the transcriptome of different trophoblast cell models

First, genes up-regulated in different trophoblastic cell models were obtained by differential expression gene analysis, including hESC line cells after BMP4 treatment (TB) comparison with H1⁵ and trophectoderm (TE) in comparison with pluripotent epiblast (EPI) cells⁶, chorionic villus (CV) in comparison with decidua (DC)) (Table S4, 5). We chose the top 200 different expression from each group to predict the transcriptional regulators of trophoblastic cells using a newly developed functional transcriptional regulatory predictive tool [8]. Different cell models showed different enrichment of transcription factors (Figure 4A). *GATA3*, *TRIM28*, and *ESR1* is the most enriched transcription factor in TB, CV and TE respectively (Figure 4B). All the trophoblastic cells from three different sources were enriched *GATA3*, *GRHL2*, *NR3C1*, *EP300*, *TP63*, *BANF1*, *TEAD4*, *ESR1*, *TFAP2A*, *TFAP2C*, *TP53*, *YAP1*, *H2AFX*. These transcription factors can be used as components of the core transcriptional regulatory network of trophoblast cells. In general, CV and TB were relatively close, with more overlapping transcription factors, while TE had less

overlapping transcription factors. The intersection and complement of these data sets are listed in Table 6.

Discussion

In this study, we collected six human chorionic villus and decidua tissue at an early stage of pregnancy and performed transcriptome sequencing based on high-throughput sequencing technology. Although there are public RNA-seq data available for chorionic villus and decidua tissue, this is the first time that the RNA-seq data were obtained from the chorionic villus and decidua which derived from the same patient.

We obtained trophoblastic cell-specific transcription factors based on chromatin accessibility data in previous studies [14]. In this study, we are based on a recently developed transcription factor prediction tool BART on transcriptome data obtained from the same article and process [6]. Based on these two data sets, many transcription factors have been successfully predicted (for example, *GATA3*, *GATA2*, *TEAD1*, *JUND*, *TEAD4*, *FOS*, *TFAP2A*, *JUN*, *TFAP2C* et al). The analysis indicates the reliability of the analysis results of this newly developed tool. The experiment and cost of transcriptome data acquisition are significantly lower than Dnas-seq and ChIP-seq. Therefore, this tool provides a technical guarantee for convenient transcriptional regulation analysis in the future.

We identified genes that were relatively highly expressed in trophoblast cells in different trophoblast cell models. At the same time, we identified transcription factors that these genes might be regulated. This analysis is of certain significance for further exploration of the development of placenta and the occurrence of pregnancy-related diseases in the future.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18

Acknowledgments

All authors contributed to the study conception and design. LYJ conceived the project and completed the core program. LYJ performed the computational analysis. ZY and YLG performed the wet experiment. HGM, RJL, and MJ provide the necessary software and hardware foundation for this research. LYJ wrote the manuscript. All authors analyzed and discussed the results.

We give thanks to Qunying Wei of Department of Obstetrics and Gynecology, the Second Affiliated Hospital of Zhengzhou University for technical support. We also give thanks to Wuhan ServiceBio technology co.LTD for the assistance on the immunohistochemical experiment. Sequencing results from this study have been assigned to European Nucleotide Archive with accession ERX3472341 for chorionic villus and ERX3472299 for decidua.

Reference

1. Cox, B., Leavey, K., Nosi, U., Wong, F., & Kingdom, J. (2015). Placental transcriptome in development and pathology: expression, function, and methods of analysis. *American Journal of Obstetrics & Gynecology*, 213(4), S138–S151. <https://doi.org/10.1016/j.ajog.2015.07.046>
2. Li, H., Liu, Y., Liu, H., & Sun, X. (2020). Effect for Human Genomic Variation During the BMP4-Induced Conversion From Pluripotent Stem Cells to Trophoblast. *Frontiers in Genetics*, 11. <https://doi.org/10.3389/fgene.2020.00230>
3. Telugu, B. P., Adachi, K., Schlitt, J. M., Ezashi, T., Schust, D. J., Roberts, R. M., & Schulz, L. C. (2013). Comparison of extravillous trophoblast cells derived from human embryonic stem cells and from first trimester human placentas. *Placenta*, 34(7), 536–543. <https://doi.org/10.1016/j.placenta.2013.03.016>
4. Jain, A., Ezashi, T., Roberts, R. M., & Tuteja, G. (2017). Deciphering transcriptional regulation in human embryonic stem cells specified towards a trophoblast fate. *Scientific Reports*, 7(1), 17257. <https://doi.org/10.1038/s41598-017-17614-5>
5. Yabe, S., Alexenko, A. P., Amita, M., Yang, Y., Schust, D. J., Sadovsky, Y., ... Roberts, R. M. (2016). Comparison of syncytiotrophoblast generated from human embryonic stem cells and from term placentas. *Proceedings of the National Academy of Sciences of the United States of America*, 113(19), E2598-2607. <https://doi.org/10.1073/pnas.1601630113>
6. Xie, W., Schultz, M. D., Lister, R., Hou, Z., Rajagopal, N., Ray, P., ... Ren, B. (2013). Epigenomic Analysis of Multilineage Differentiation of Human Embryonic Stem Cells. *Cell*, 153(5),

- 1 1134–1148. <https://doi.org/10.1016/j.cell.2013.04.022>
- 2 7. Blakeley, P., Fogarty, N. M. E., del Valle, I., Wamaitha, S. E., Hu, T. X., Elder, K., ... Niakan, K. K.
- 3 (2015). Defining the three cell lineages of the human blastocyst by single-cell RNA-seq.
- 4 *Development*, 142(18), 3151–3165. <https://doi.org/10.1242/dev.123547>
- 5 8. Wang, Z., Civelek, M., Miller, C. L., Sheffield, N. C., Guertin, M. J., & Zang, C. (2018). BART: a
- 6 transcription factor prediction tool with query gene sets or epigenomic profiles.
- 7 *Bioinformatics*, 34(16), 2867–2869. <https://doi.org/10.1093/bioinformatics/bty194>
- 8 9. Bray, N. L., Pimentel, H., Melsted, P., & Pachter, L. (2016). Near-optimal probabilistic
- 9 RNA-seq quantification. *Nature Biotechnology*, 34(5), 525–527.
- 10 <https://doi.org/10.1038/nbt.3519>
- 11 10. Raudvere, U., Kolberg, L., Kuzmin, I., Arak, T., Adler, P., Peterson, H., & Vilo, J. (2019).
- 12 g:Profiler: a web server for functional enrichment analysis and conversions of gene lists
- 13 (2019 update). *Nucleic Acids Research*, 47(W1), W191–W198.
- 14 <https://doi.org/10.1093/nar/gkz369>
- 15 11. Demšar, J., Curk, T., Erjavec, A., Gorup, Č., Hočevar, T., Milutinovič, M., ... Zupan, B. (2013).
- 16 Orange: Data Mining Toolbox in Python. *Journal of Machine Learning Research*, 14,
- 17 2349–2353.
- 18 12. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... Duchesnay, É.
- 19 (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12,
- 20 2825–2830.
- 21 13. Ma, W., Wang, Z., Zhang, Y., Magee, N. E., Chen, Y., & Zang, C. (2020). BARTweb: a web
- 22 server for transcription factor association analysis. *bioRxiv*, 2020.02.17.952838.

- 1 <https://doi.org/10.1101/2020.02.17.952838>
- 2 14. Liu, Y., Ding, D., Liu, H., & Sun, X. (2017). The accessible chromatin landscape during
- 3 conversion of human embryonic stem cells to trophoblast by bone morphogenetic protein
- 4 4. *Biology of Reproduction*, 96(6), 1267–1278. <https://doi.org/10.1093/biolre/iox028>
- 5 15. Theofanakis, C., Drakakis, P., Besharat, A., & Loutradis, D. (2017). Human Chorionic
- 6 Gonadotropin: The Pregnancy Hormone and More. *International Journal of Molecular*
- 7 *Sciences*, 18(5), 1059. <https://doi.org/10.3390/ijms18051059>
- 8 16. Moore, T., & Dveksler, G. S. (2014). Pregnancy-specific glycoproteins: complex gene families
- 9 regulating maternal-fetal interactions. *International Journal of Developmental Biology*,
- 10 58(2-3-4), 273–280. <https://doi.org/10.1387/ijdb.130329gd>
- 11 17. Abdulghani, M., Jain, A., & Tuteja, G. (2019). Genome-wide identification of enhancer
- 12 elements in the placenta. *Placenta*, 79, 72–77.
- 13 <https://doi.org/10.1016/j.placenta.2018.09.003>
- 14 18. Lee, B., Kroener, L. L., Xu, N., Wang, E. T., Banks, A., Williams, J., ... Pisarska, M. D. (2016).
- 15 Function and Hormonal Regulation of GATA3 in Human First Trimester Placentation. *Biology*
- 16 *of Reproduction*, 95(5). <https://doi.org/10.1095/biolreprod.116.141861>
- 17 19. Ferreira, L. M. R., Meissner, T. B., Mikkelsen, T. S., Mallard, W., O'Donnell, C. W., Tilburgs,
- 18 T., ... Strominger, J. L. (2016). A distant trophoblast-specific enhancer controls HLA-G
- 19 expression at the maternal–fetal interface. *Proceedings of the National Academy of*
- 20 *Sciences*, 113(19), 5364–5369. <https://doi.org/10.1073/pnas.1602886113>
- 21 20. Yabe, S., Alexenko, A. P., Amita, M., Yang, Y., Schust, D. J., Sadovsky, Y., ... Roberts, R. M.
- 22 (2016). Comparison of syncytiotrophoblast generated from human embryonic stem cells

- 1 and from term placentas. *Proceedings of the National Academy of Sciences*, 113(19),
- 2 E2598–E2607. <https://doi.org/10.1073/pnas.1601630113>
- 3

1 Table 1 Top 10 genes that are highly expressed in the expression profile of villus tissue. Rows
2 represent genes, columns represent samples, and values show the number of mapped reads.

	10_4	11_1	12_2	13_4	9_4	8_4
<i>CGA</i>	1420947	1250455	976873	1373181	2600673	1869716
<i>CGB5</i>	559944	526901	362622	526781	1054370	803139
<i>CGB8</i>	552831	507120	386121	806300	1067900	757055
<i>COL3A1</i>	486944	115352	92798	90939	139125	111177
<i>KISS1</i>	474058	296932	169689	429739	416541	263044
<i>CGB3</i>	412359	374997	455725	862308	1315630	1059610
<i>COL1A1</i>	357175	73687	55306	84492	108167	94342
<i>TFPI2</i>	305102	262126	174158	94923	209330	167988
<i>MT-CO1</i>	300391	267588	253982	308825	308529	268550
<i>COL1A2</i>	227665	38999	32772	39383	50489	42886

3

4

5 Table 2 Top 10 genes that are highly expressed in the expression profile of decidua. Rows
6 represent genes, columns represent samples, and values show the number of mapped reads.

	10_E	11_E	12_B	13_D	8_C	9_E
<i>PAEP</i>	707159	225495	553927	332730	317333	716464
<i>MT-CO1</i>	402678	380707	397730	315579	310992	375237
<i>EEF1A1</i>	274212	194122	314425	233226	203655	228912
<i>FBLN1</i>	148393	46328	95584	46766	46616	108020
<i>MT-ND4</i>	127854	100057	153752	90144	108484	126805
<i>MT-CO3</i>	118360	84003	113272	77937	63105	121600
<i>MT-ND5</i>	112658	86210	65977	83994	96545	104019
<i>MT-CYB</i>	108860	74152	99149	66198	63807	81149
<i>GPX3</i>	103827	113570	280574	105654	71261	78752
<i>IGFBP5</i>	88947	49685	30649	33556	25273	108734

7

8

1 Table 3 The 10 top genes significantly upregulated in chorionic villus compared with decidua
2 (logFC>2, p<0.05). Every row of the table represents a gene; the columns display the estimated
3 measures of differential expression.

gene_symbol	logFC	AveExpr	t	P.Value	adj.P.Val
<i>DLK1</i>	-13.9637	2.418015	-32.4705	3.16E-14	1.83E-11
<i>KISS1</i>	-12.7453	7.639202	-20.4116	1.49E-11	8.85E-10
<i>PSG8</i>	-12.6297	2.410962	-17.86	8.54E-11	3.14E-09
<i>CGA</i>	-12.4301	10.01645	-19.6388	2.47E-11	1.28E-09
<i>XAGE3</i>	-12.4281	0.951458	-25.0237	1.01E-12	1.46E-10
<i>CGB5</i>	-12.3711	8.724329	-17.6354	1.01E-10	3.50E-09
<i>CGB2</i>	-12.3182	3.340725	-8.13134	1.43E-06	6.77E-06
<i>DUSP9</i>	-12.311	2.240281	-19.5565	2.61E-11	1.30E-09
<i>CGB3</i>	-12.2089	8.946367	-17.1212	1.48E-10	4.74E-09
<i>CGB7</i>	-12.1083	3.189653	-13.386	3.46E-09	5.30E-08

4
5 Table 4 The 10 top genes significantly upregulated in chorionic villus compared with decidua
6 (logFC>2, p<0.05). Every row of the table represents a gene; the columns display the estimated
7 measures of differential expression.

gene_symbol	logFC	AveExpr	t	P.Value	adj.P.Val
<i>PRL</i>	11.00277	0.195864	11.69084	1.89E-08	1.98E-07
<i>TMEM252</i>	10.61131	0.015903	44.4591	4.67E-16	1.45E-12
<i>SLIT1</i>	10.51424	2.056775	14.47267	1.29E-09	2.46E-08
<i>GNLY</i>	10.41486	5.48565	20.63787	1.29E-11	8.11E-10
<i>CHRD1</i>	10.40644	2.649069	19.10613	3.54E-11	1.64E-09
<i>GZMA</i>	10.32547	1.86351	13.22149	4.05E-09	5.94E-08
<i>SLC18A2</i>	10.28437	3.2659	13.22472	4.04E-09	5.93E-08
<i>GABRQ</i>	10.02802	2.47774	22.60054	3.90E-12	3.46E-10
<i>SLPI</i>	9.958351	2.749848	13.8515	2.25E-09	3.77E-08
<i>RORB</i>	9.903442	2.928284	21.57621	7.19E-12	5.38E-10

1 Table 5 Summary of datasets analyzed in this study.

Cell lines	GEO number	Description	Reference
HTR8/SVneo	GSE85995	Function and hormonal regulation of GATA3 in human first-trimester placentation	[18]
JEG3	GSE79779	Transcriptional profiling of JEG3 cells with HLA-G ablation via deletion of Enhancer L	[19]
hESC line cells after BMP4 treatment	GSE73017	Comparison of syncytiotrophoblast differentiated from human pluripotent stem cells and that derived from primary cytotrophoblast recovered from term placentae	[20]
hESC line cells after BMP4 treatment	SRP000941	H1 cells were differentiated to trophoblast-like cells	[6]
Trophectoderm	GSE66507	Single-Cell RNA-seq Defines the Three Cell Lineages of the Human Blastocyst	[7]
Chorionic villus	ERX3472341	chorionic villus in early-gestation stage	This study
Decidua	ERX3472299	decidua in early-gestation stage	This study

2

3

4

5

- 1 Table 6 TFs enriched in different trophoblast cell models (TB: hESC line cells after BMP4
- 2 treatment, TE: Trophectoderm, CV: Chorionic villus).

Cell Model	TF name
[TB]:	<i>AR,TBX5,HOXB13,TCF4,FOS,CDX2,EBF3,ZNF750,HAND2,JUN,PRMT1,BATF,EBNA3C,RAD21,HNF4A,HNF1B,TLE3,BATF3,MEIS1,SMC3,ARID1A,GRIP1,DLX2,MYOG,TTF1,JARID2,SP1,EPAS1,SP1B,PPARG,HNF1A,STAT3,PAX3,ZIM3,LYL1,FOSL1,LHX2,HOXA9,PGBD3,SATB1,SMC1A,NFKB2,PGR,ZNF92,HMG20A,NUP98-HOXA9,SIX2,SOX10,PROX1,IRF4</i>
[TE]:	<i>FOXN1,ZSCAN2,MEN1,ZNF592,GRHL1,RARG,CTBP1,NCAPG,GATAD2B,WDHD1,ZNF687,TCF7L2,SPDEF,RAC3,KDM5A,KLF5,ZXDC,BCL3,ELK4,RCOR1,ZNF75A,EHF,SREBF1,GATAD1,NELFA,ELK1,NELFE,NCAPG2,ELF2,ZNF131,DPF2,APOBEC3B,RXRA,PAF1,ZNF701,HOXA2,DNMT3A,HOXA1,RBP2,NR5A2,MTA3,HCFC1,ZNF711,FEV,NR2C2,CPSF3L,E2F7,KMT2D,GMEB1,EMSY,FOXO2,GMEB2,ELL2,CASP8AP2,GTTF2,KLF10,CREB3L4,FOXG1,KLF4,SMARCC2,AUTS2,DDX21,ETV7,PHF2,MBD3,NFAT5,ETV5,LIN9,BRCA1,INO80,BRD4,CHD1,NR1H2,MXD3,E2F2,DLX1,AFF4</i>
[CV]:	<i>TRIM28,BRDU,ZNF486,YBX1,BAHD1,ZNF274,ZNF680,PLRG1,POLR3D,ZNF649,ZNF586,LRWD1,CENPC,ZNF136,SOX2,NOTCH3,ZNF140,CENPT,ZNF75D,ZNF207,CRY1,ATF4,ZNF781,ZNF287,NANOG,PRDM12,ZNF248,TSC22D4,ZTA,PARP1,SFPQ,BMI1,CBX3,LMTK3,LARP7,POU5F1,EOMES,SMAD3,EHMT2,SMAD2,ZBTB48,PTRF,DACOR1,OTX2,ZFP57,ASCL1,LMNA,ZNF331,ATF7IP,ZNF823,ZNF17,CDCA2,SMN1,ZFP28,ZNF445,CDK2,TERC,ILK,TET2,EED</i>
[TB] and [CV]:	<i>GATA2,MAML3,TP73,GATA4,TEAD1,CEBPB,GATA6,NKX2-71,PHOX2B,FOXA2,DUX4,HOTAIR,CHAT,NME2,SMARCA2,EZH2,REST,ZNF808,MCM2,CTCF,SMAD1,ZNF512,CEBPG,SUMO1,SOX11,SIRT1,ZNF31</i>
[TB] and [TE] and [CV]:	<i>GATA3,GRHL2,NR3C1,EP300,TP63,BANF1,TEAD4,ESR1,TFAP2A,TFAP2C,TP53,YAP1,H2AF</i> X
[TB] and [TE]:	<i>PR,JUND,FOSL2,AHR,ZNF217,GREB1,FOXA1,SMARCC1,HES2,WWTR1</i>

1 Figure legend

2

3 Figure 1 The experiment and analysis workflow of this study.

4 Figure 2 (A) Tissue section, hematoxylin staining in chorionic villus; (B) Immunofluorescence
5 staining of the trophoblastic specific transcription factor in chorionic villus; (C) Scatter plots of
6 differentially expressed genes between chorionic villus and decidua. Every point in the plot
7 represents a gene. Red points indicate significantly up-regulated genes, and blue points indicate
8 down-regulated genes; (D) Hierarchical clustering and heatmap displaying differentially expressed
9 genes between chorionic villus and decidua. Every row of the heatmap represents a gene, every
10 column represents a sample, and every cell displays normalized gene expression values;

11

12 Figure 3 (A) Projection of 12 samples (including chorionic villus and decidua) in a 2D-map using
13 PCA: Each point represents a sample which is colored according to the (sub)tissue label; (B)
14 Projection of 12 samples (including chorionic villus and decidua) in a 2D-map using t-SNE
15 (Perplexity=11; metric was set as Euclidean); (C) Expression of pregnancy-related surface marker
16 genes and transcription factors (D) in different cell models of chorionic villus (CV) based on
17 RNA-seq data analysis.

18

19 Figure 4 (A)Venn diagram of top 100 enriched transcription factors in different trophoblastic cell
20 models by BART; (B) Top enriched transcription factors in different trophoblastic cell models (TB,
21 CV, and TE respectively) by BART; Area under the ROC curve (AUC) is calculated for each dataset;
22 AUC are grouped by the factor, and Wilcoxon test is performed for each factor compared with all
23 datasets as background. cumulative distributions show significantly higher AUC for TF.

24

1	Supplementary table
2	
3	Table S1 gene expression matrix of trophoblastic cell-specific transcription factors and putative
4	surface marker genes in chorionic villus;
5	Table S2 Gene expression profiles of 6 Villus and decidua derived by this study;
6	Table S3 differentially expressed genes between chorionic villus vs decidua;
7	Table S4 differentially expressed genes between TB and H1;
8	Table S5 differentially expressed genes between TE and EPI.
9	
10	
11	Supplementary file
12	
13	Supplementary file 1: GO and KEGG enrichment analysis of differentially expressed genes
14	between chorionic villus and decidua;
15	
16	Supplementary figure
17	Figure S1 (A) Immunostaining of HLA-G in chorionic villus; (B) Immunofluorescence of
18	trophoblast-specific surface markers KRT7 and HLA-G in Chorionic villus; (C) Immunofluorescence
19	results of TFAP2C in immortalized trophoblastic cell lines (JEG3) in vitro.
20	Figure S2 (A) Track of the gene expression signal (intensity on Y-axis) at human chorionic
21	gonadotropin (hCG) family and pregnancy-specific glycoproteins (PSGs) in chorionic villus tissues
22	in the human genome; (B) Expression level of tissues-specifically expressed genes, RPS4Y1, EIF1AY,
23	DDX3Y and KDM5D in different chorionic villus tissues;
24	Figure S3 Library Size Analysis results. The figure contains an interactive bar chart that displays
25	the total number of reads mapped to each RNA-seq sample in the dataset.
26	

Figure1

Figure 3

Figure 4