



## Article

# The human placenta exhibits a unique transcriptomic void

Sungsam Gong,<sup>1,2</sup> Francesca Gaccioli,<sup>1,2</sup> Irving L.M.H. Aye,<sup>1,2</sup> Giulia Avellino,<sup>1,2</sup> Emma Cook,<sup>1</sup> Andrew R.J. Lawson,<sup>3</sup> Luke M.R. Harvey,<sup>3</sup> Gordon C.S. Smith,<sup>1,2,4</sup> and D. Stephen Charnock-Jones<sup>1,2,4,5,\*</sup>

<sup>1</sup>Department of Obstetrics and Gynaecology, NIHR Cambridge Biomedical Research Centre, University of Cambridge, Cambridge, UK

<sup>2</sup>Centre for Trophoblast Research (CTR), Department of Physiology, Development and Neuroscience, University of Cambridge, Cambridge, UK

<sup>3</sup>Wellcome Trust Sanger Institute, Hinxton, UK

<sup>4</sup>These authors contributed equally

<sup>5</sup>Lead contact

\*Correspondence: [dscj1@cam.ac.uk](mailto:dscj1@cam.ac.uk)

<https://doi.org/10.1016/j.celrep.2023.112800>

## SUMMARY

The human placenta exhibits a unique genomic architecture with an unexpectedly high mutation burden and many uniquely expressed genes. The aim of this study is to identify transcripts that are uniquely absent or depleted in the placenta. Here, we show that 40 of 46 of the other organs have no selectively depleted transcripts and that, of the remaining six, the liver has the largest number, with 26. In contrast, the term placenta has 762 depleted transcripts. Gene Ontology analysis of this depleted set highlighted multiple pathways reflecting known unique elements of placental physiology. For example, transcripts associated with neuronal function are in the depleted set—as expected given the lack of placental innervation. However, this demonstrated overrepresentation of genes involved in mitochondrial function ( $p = 5.8 \times 10^{-10}$ ), including PGC-1 $\alpha$ , the master regulator of mitochondrial biogenesis, and genes involved in polyamine metabolism ( $p = 2.1 \times 10^{-4}$ ).

## INTRODUCTION

The placenta has a key role in the pathogenesis of many major complications of pregnancy, such as preeclampsia (PE) and fetal growth restriction (FGR), termed, collectively, the “great obstetrical syndromes,”<sup>1</sup> which account for a substantial burden of global morbidity and mortality. Progress on predicting and preventing these complications is hampered by a lack of mechanistic understanding of normal and abnormal placental function, and we and others have applied multiple studies using omics methods to try and address this knowledge gap. Published studies of the placenta transcriptome tend to focus on identifying genes differentially regulated in complicated pregnancies. Other studies have compared the placental transcriptome across species<sup>2</sup> and across gestation,<sup>3</sup> but there are fewer studies comparing the placental transcriptome with the transcriptomes of other organs.<sup>4,5</sup> mRNA sequencing (RNA-seq) enables transcriptome profiling of tissues or single cells, and there are a number of studies characterizing the so-called transcriptome “landscape” of tissues of interest. It is now an essential part of large-scale multi-omics studies, such as the Encyclopedia of DNA Elements (ENCODE),<sup>6</sup> the RoadMap Epigenomics Project,<sup>7</sup> and the Functional Annotation of Mammalian Genome (FANTOM5) project.<sup>8</sup> However, the human placenta transcriptome is relatively understudied and absent from large-scale “omics” analyses such as the Genotype-Tissue Expression (GTEx) project.<sup>9</sup>

Pan-tissue comparative analyses generally focus on identifying transcripts that are abundant in a tissue of interest while be-

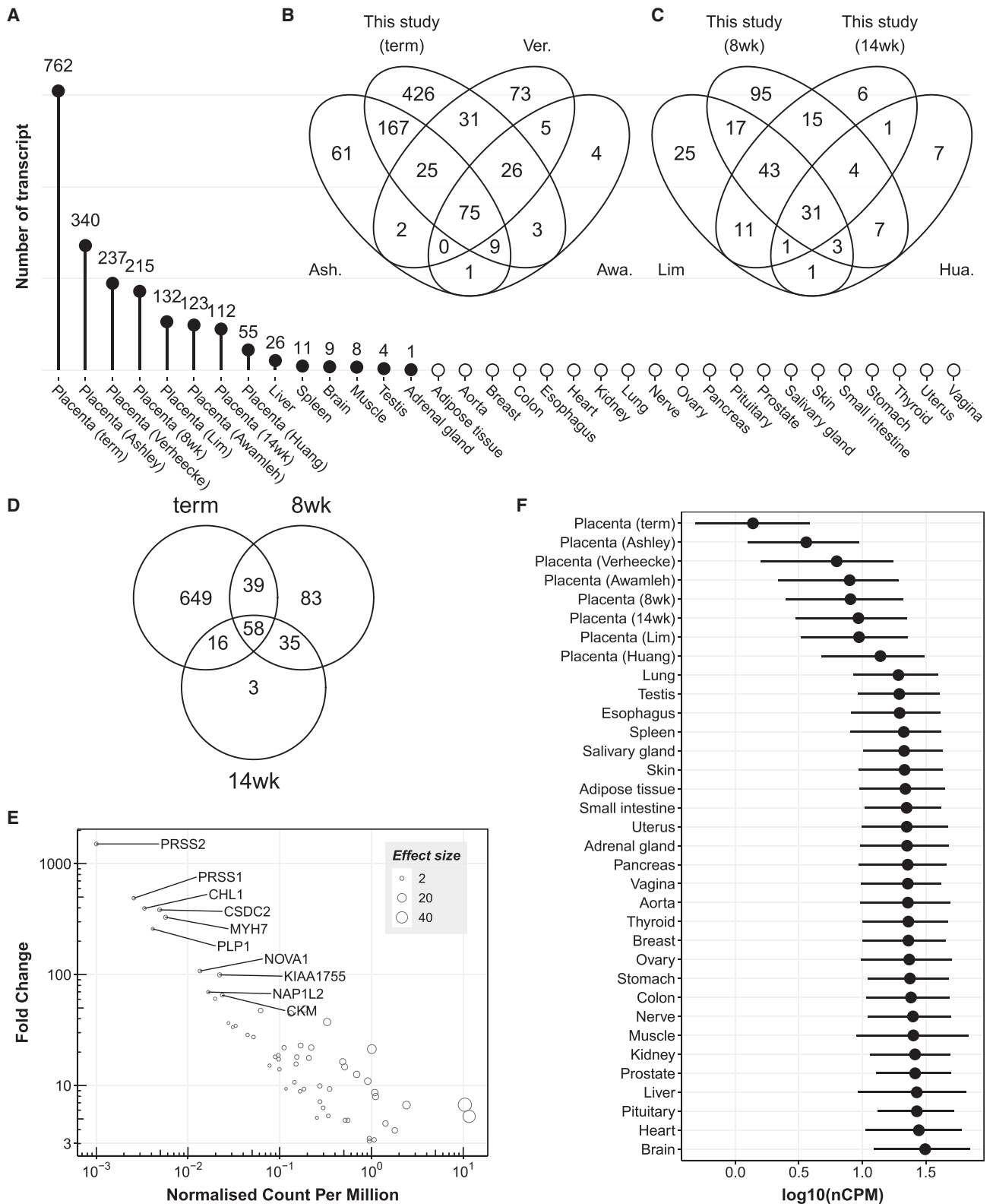
ing absent or depleted in others. Indeed, there are a number of tools and databases that enable “tissue-specific” gene enrichment analysis.<sup>10–12</sup> Studying “tissue-specific” genes provides information about specific functions that define a unique set of characteristics or the “identity” of a tissue of interest. Transcripts that are ubiquitously expressed in multiple tissues, such as housekeeping genes,<sup>13</sup> can be identified, and this gives insight to the functions that all tissues share. In contrast, little attention has been paid to the identification of transcripts that are less abundant, or even absent, in one tissue compared with all others. Here we report transcripts depleted or absent in the human placenta at term and in early gestation compared with 46 other tissues studied in the GTEx project. Functional enrichment analysis of depleted transcripts highlighted pathways that reflect known aspects of placental physiology, such as lack of nervous tissue and unique immunological features. However, these analyses also generated evidence showing that the term human placenta has unique metabolic characteristics, as evidenced by multiple absent transcripts involved in mitochondrial function and polyamine metabolism.

## RESULTS

### Tissue-wide comparison of depleted transcripts

We carried out RNA-seq using 59 human term placentas from the Pregnancy Outcome Prediction (POP) study cohort<sup>14–16</sup> and 14 human placentas from earlier in gestation ( $n = 8$ , 7–8 weeks [8 weeks];  $n = 6$ , 13–14 weeks [14 weeks]).<sup>17</sup> We





(legend on next page)

obtained approximately 38 million reads from each sample (Table S1). We compared the placental transcriptome profile at 8 weeks, 14 weeks, and term with that of 46 tissues from 11,803 samples of GTEx Consortium datasets<sup>9</sup> and investigated which transcripts are absent or depleted in the placenta while being reasonably abundant in other tissues (Table S2). To adjust for differences in the RNA composition across tissues, we applied the following two normalization methods: (1) the median ratio method (DESeq<sup>18</sup> and DESeq2<sup>19</sup>) and the trimmed mean of M-values (TMM<sup>20</sup>) (see STAR Methods for details). For 19,170 eligible protein-coding transcripts, we ranked tissues by their normalized count per million (nCPM) and identified 5,632 and 5,727 transcripts for which the term placenta was ranked 47 (i.e., bottom) based on the DESeq2 and TMM normalization methods, respectively. Then we identified the 762 transcripts that satisfied the following three conditions: (1) nCPM greater than 1 for the tissue of rank 46, (2)  $nCPM_{(\text{rank}=46)}/nCPM_{(\text{rank}=47)}$  greater than 3, and (3) fulfilling both these conditions in the DESeq2 and TMM normalization methods (Table S3). These transcripts are described as being “depleted” hereafter. For early gestational age datasets (8-week and 14-week placenta), we identified 215 and 112 such depleted transcripts, respectively (Tables S4 and S5).

Using the same criteria we applied to the placenta, we sought to identify mRNAs depleted in each of the 46 other tissues available from the GTEx project.<sup>9</sup> Surprisingly, we found that there were ~30 times more transcripts depleted in the placenta than in the liver (26 depleted transcripts), which was the highest among the 46 non-placental tissues. Besides the placenta, only six tissues had one or more depleted transcript: liver, spleen, brain (cerebellar hemisphere), muscle, testis, and adrenal gland (Figure 1A; Table S6). We then used five external placenta RNA-seq datasets generated independently: two early gestational placenta datasets (Lim et al.<sup>21</sup> [n = 4] and Huang et al.<sup>22</sup> [n = 3]) and three term placenta datasets (Verheecke et al.<sup>23</sup> [n = 66], Ashley et al.<sup>24</sup> [n = 4], and Awamleh et al.<sup>25</sup> [n = 21]). These analyses confirmed that the placenta had the most depleted transcripts among the other tissues studied: 132 (Lim), 55 (Huang), 237 (Verheecke), 340 (Ashley), and 123 (Awamleh) (Figure 1A; Table S7). All datasets described above had variable depth of coverage. So, to investigate any possible effect of sequencing depth, we downsampled the reads to 20 million for all samples. We then repeated the analysis with the same methods and criteria as described above (Figure S1A).

The number of samples available in the GTEx dataset was also variable and generally greater than in our placental data. We therefore repeated the analysis using a random selection of 60 samples for each tissue type (Figure S1B). Finally, we repeated the analysis, downsampling the number of reads (20 million) and the number of samples (60) 1,000 times, and the distribution of the number of depleted genes detected is shown in Figure S2. In downsampled datasets, we obtained consistent results showing that the placenta had many more depleted transcripts than other tissues (Figures S1 and S2; Tables S8 and S9). The number of depleted transcripts that were shared among our dataset and other term placenta datasets (Figure 1B) or other first trimester datasets (Figure 1C) was highly significant ( $p < 1 \times 10^{-314}$  for the term placenta and  $p = 7.4 \times 10^{-209}$  for the early placenta; both Fisher’s exact test). In our term and early gestation placental datasets, a total of 883 transcripts were depleted at any gestational age, and among these, 58 were depleted in all three trimesters (Table S10; Figure 1D). Figure 1E shows the 58 genes for which the transcripts are depleted in our three placental datasets ranked by their fold change, with the top 10 genes being annotated, compared with 46 tissues from the GTEx dataset (Table S10).

Among the 762 transcripts depleted in our term placenta dataset, we did not detect any transcripts encoding *PRSS2* (serine protease 2; also known as trypsin 2), while *MT-ND6* (mitochondrially encoded NADH:ubiquinone oxidoreductase core subunit 6) had the highest level of expression (nCPM = 123) but was still at least 3-fold lower than in all other tissues studied (Table S3). Interestingly, the 762 transcripts depleted in our term placenta were also less abundant in our early gestation placenta samples and other placenta datasets studied (Figure 1F). Moreover, the 426 transcripts uniquely depleted in our term placenta data (Figure 1B) were also less abundant in external term placenta datasets, and the majority of them were ranked bottom in the Ashley (335, 79%), Verheecke (245, 58%), and Awamleh (179, 42%) datasets (Figure S3).

### Dynamic changes of depleted transcripts during pregnancy

We further investigated the 883 transcripts depleted in the placenta at any trimester of pregnancy (Figure 1D) to see how they change over time. Among the 215 transcripts depleted in the first trimester placenta, 83 transcripts were not shared with other gestational ages (T1 in Figure 2A), and their abundances

#### Figure 1. The number of absent or depleted transcripts in various tissues

(A) The number of depleted transcripts are shown in placenta samples (ours and other studies) and other somatic tissues. The term, 8-week, and 14-week placentas are from this study. Tissues with open circles represent zero depleted transcripts.

(B–D) Venn diagrams showing the number of transcripts and their overlaps depleted in early gestational placentas (B), term placentas (C), and three gestational stages of the placenta datasets from this study (D). The lists of transcripts depleted in the non-placental and the external placental tissues are available in Tables S6 and S7, respectively. Ver., Verheecke; Ash., Ashley; Awa., Awamleh; Hua., Huang.

(E) Abundance of the transcripts (x axis) relatively depleted in all three trimesters are shown along with their fold change (y axis, calculated as follows:  $nCPM_{(\text{rank}=46)}/nCPM_{(\text{rank}=47; \text{placenta})}$ ). The counts (per million) on the x axis are normalized by the TMM method, and the data are available in Table S10. To avoid fold change being infinite values, a small number (0.001) was added to nCPM of the term placenta. The transcripts with the tenth-highest fold change (*PRSS2* being the top one) are shown with their gene names. The size of a circle represents the effect size (i.e.,  $nCPM_{(\text{rank}=46)} - nCPM_{(\text{rank}=47; \text{placenta})}$ ).

(F) The range of transcript abundance for the 762 genes in the placental tissues and 26 representative somatic tissues of 46 we studied. Dot, median; line, interquartile range (IQR). For display, non-placental tissues shown in (F) were manually selected when there were at least two subregions from the same tissue. For example, we analyzed a total of 13 brain subregions in this study, and the cerebellar hemisphere is shown here to represent the brain. The representative subregions are shown in Table S2.



There were 7 transcripts involved with transition metal ion binding: *BHMT2*, *CYP27A1*, *DTX1*, *FAM20C*, *MMP17*, *NR4A2*, and *TRIM47*. However, no significant GO term was found among the 35 transcripts shared between the first and the second trimester placenta, suggesting a transition of different functional roles at that point.

We found that the second trimester placenta sample had the smallest number of depleted transcripts ( $n = 112$ ) compared with those from the first trimester ( $n = 215$ ) and the term placenta ( $n = 762$ ). Indeed, only 3 transcripts (cilia- and flagella-associated protein 91 [*CFAP91*; also known as *MAATS1*], myomesin 2 [*MYOM2*], and neurotrophic receptor tyrosine kinase 3 [*NTRK3*]) were uniquely depleted in the second trimester (T2 in Figure 2A). However, they all were ranked bottom (i.e., the least abundant compared with non-placental tissues) in the first trimester and close to bottom in the term placenta. However, they were not sufficiently low enough to satisfy the 3-fold threshold for the TMM and DESeq2 normalization methods. For example, for *MYOM2*, the fold changes based on TMM and DESeq2 were 2.9- and 2.8-fold, respectively.

There were 16 transcripts shared between the second and the term placenta, and five of them (adenylate cyclase 1 [*ADCY1*], reelin [*RELN*], cellular communication network factor 3 [*NOV*; currently called *CCN3*], neuronal pentraxin 2 [*NPTX2*], and neuronal pentraxin receptor [*NPTXR*]) were involved with regulation of the nervous system process and modulation of chemical synaptic transmission (T2+T3 in Figure 2B); these functions could be limited in the second trimester and term placentas.

Among the 883 transcripts depleted in any of the three trimesters, 58 are depleted in all three (Table S10; T1+T2+T3 in Figure 2A), and they are associated with various GO terms (T1+T2+T3 in Figure 2B; Table S11). It is unsurprising that genes annotated with the GO terms “astrocyte development,” “gliogenesis,” and “skeletal muscle adaptation” are absent in the placenta. However, because the placenta is a steroidogenic organ, the depletion of genes associated with “positive regulation of steroid hormone metabolic process” is more surprising. The genes annotated with this term include aldo-keto reductase family members (*AKR1C1* and *AKR1C2*). This is consistent with the requirement for placental steroid production because these reductases inactivate steroid hormones,<sup>26</sup> specifically progesterone in the case of *AKR1C1*. Peroxisome proliferator activated receptor gamma (PPARG) is abundant and essential for placental development<sup>27</sup> and function. However, the depletion of PPARG coactivator 1 alpha (*PPARGC1A*; also known as PGC-1 $\alpha$ ) transcripts suggests that the usual coordination between PPARG and PGC-1 $\alpha$  does not occur in the placenta.<sup>28</sup> Of note, PGC-1 $\alpha$  is also directly implicated in regulating mitochondrial biogenesis<sup>29</sup> and regulation of mitochondrial genes (see below).

Several depleted transcripts are annotated with the GO term “extracellular matrix organization.” Trypsins 1 and 2 (PRSS1 and PRSS2) are notable because these were essentially absent from the placenta. Because these proteins are key activators of multiple matrix metalloproteases, this suggests that initiation of matrix remodeling is mediated by other proteases. Transcripts encoding two type IV collagen genes (*COL4A3* and *COL4A4*) were depleted. These collagens are components of basement membranes and form a triple helix (with *COL4A5*). Mutation or

loss of any of these three genes causes Alport’s syndrome.<sup>30</sup> However, lack of the  $\alpha3.\alpha4.\alpha5$ (IV) collagen protomer is without effect in the placenta, in contrast to the other organs affected in Alport’s syndrome. It is likely that the  $\alpha1.\alpha1.\alpha2$ (IV) collagen protomer is sufficient, and, in fact, the placenta has the highest expression of *COL4A1* among the GTEx tissues from our previous study.<sup>4</sup> Keratin filament transcripts (*KRT4*, *KRT5*, and *KRT13*) are also depleted and annotated with the GO term “extracellular matrix organization.” These keratins are characteristic of stratified epithelial surfaces<sup>31</sup> (such as the esophagus, in which the expression level is more than 10,000 times higher), and this difference likely reflects the syncytial nature of the trophoblast epithelial surface.

### Transcripts related to polyamine metabolism

Using KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway analysis of the 649 transcripts depleted only at term, we also noted that genes for arginine and proline metabolism and, hence, the polyamine (putrescine, spermidine, and spermine) metabolic pathway were also significantly overrepresented in the depleted set (Figure S4A). We investigated a total of 15 genes annotated with the GO term “polyamine metabolic process” (GO:0006595) together with its 11 child terms (e.g., polyamine acetylation and deacetylation) and found that 5 of them (spermine oxidase [*SMOX*], polyamine oxidase [*PAOX*], spermidine synthase [*SRM*], spermidine/spermine N1-acetyltransferase family member 2 [*SAT2*], and antizyme inhibitor 2 [*AZIN2*]) were depleted in the term placenta (Figure S4A). Interestingly, the transcript abundance of these 5 genes was also lower in the early gestational placentas, but not sufficiently low to be classified as being depleted (i.e., not below the 3-fold difference). All three placenta datasets (8 weeks, 14 weeks, and term) were clustered differently from 46 somatic tissues.

We have reported previously that placental polyamine metabolism is implicated in sex-related differences in placental complications of human pregnancy.<sup>4</sup> In this study we found that arginine and proline metabolism (and, hence, the polyamine metabolic pathway) is significantly overrepresented in the transcripts depleted in the term placenta and that those transcripts involved with the polyamine pathway were less abundant from the early gestational placentas (Figure S4A). The depleted transcripts are involved with either early steps of polyamine anabolism, such as synthesis of ornithine (via ornithine decarboxylase 1 [*ODC1*]) and spermidine (via *SRM*), or polyamine catabolism, such as *SMOX*, *PAOX*, and *SAT2*. In contrast, some genes related to the polyamine pathway were actually more abundant in the placenta (e.g., *SAT1*, *AZIN1*, *SMS*, and *AMD1*) than non-placental tissues. Mapping these on the polyamine metabolic pathway, they are involved in either methionine metabolism (*SMD1*) or a later step of polyamine anabolism involved with spermine (*SMS* and *SAT1*). Taken together, these data suggest that spermidine may be depleted (because of lower levels of *ODC1* and *SRM*), whereas spermine and its intermediates (e.g., DiAcSpm) could be more abundant (because of higher levels of *SMS* and *SAT1*).

We identified 762 depleted transcripts in the term placenta, which was the highest number among three trimesters, and 649 of them (85%) were uniquely depleted at term (T3 in

Figure 2A). GO analysis showed that these genes are predominantly associated with mitochondrion-specific processes, suggesting that the term placenta has diminished capacity for these functions (T3 in Figure 2B; Table S11). They include genes encoding 12 mitochondrial ribosomal proteins; ATP synthase, H<sup>+</sup> transporting, mitochondrial F1 complex, delta subunit (*ATP5D*); succinate dehydrogenase complex assembly factor 1 (*SDHAF1*); NADH:ubiquinone oxidoreductase complex assembly factors (*NDUFAF3* and *NDUFAF8*) and subunits (*NDUFB7*, *NDUFS7*, and *NDUFS8*); and mitochondrially encoded NADH:ubiquinone oxidoreductase core subunit 6 (*MT-ND6*). Even though these transcripts were not sufficiently low to be classified as being depleted in earlier gestations, most of them were also less abundant than somatic tissues (Figures S4B and S4C).

Having observed significantly overrepresented mitochondrion-related GO terms in the list of depleted transcripts, we examined the proportion of RNA-seq reads from the 19,170 protein-coding genes that mapped to mitochondrial DNA. The term placenta has the lowest percentage (3.4%), followed by the aorta (4.5%) and the 8-week placenta (4.6%) (Figure S5A). In contrast, the heart (left ventricle, 39.7%), kidneys (cortex, 31.3%), and liver (21.1%) expressed the most mitochondrial protein-coding transcripts. We also examined the extent of mitochondrial transcripts, including protein-coding and non-coding transcripts, such as mitochondrial rRNA and tRNA, and found that the term placenta also showed the lowest proportion of reads mapped to mitochondrial DNA (3.7%; Figure S5B).

### Abundance of nuclear-encoded transcripts localized in the mitochondria

Mitochondria contain proteins encoded by nuclear DNA and subsequently imported into the mitochondria as well as those directly encoded by mitochondrial DNA (mtDNA). MitoMiner<sup>32</sup> is a database of protein-coding genes with strong support for mitochondrial localization and, hence, function. Having observed association of mitochondrion-related GO terms in the 762 transcripts depleted in the term placenta, we further investigated how many of these encode mitochondrial proteins defined in MitoMiner and found 84 ( $p = 5.1 \times 10^{-10}$ , Fisher's exact test). For the 234 transcripts depleted earlier in gestation (either first or second trimester; Figure 1D), the number overlapping with MitoMiner is only 9 ( $p = 0.382$ , Fisher's exact test). We then examined the abundance of transcripts encoding the 1,042 genes in MitoMiner across non-placental tissues from GTEx and placentas obtained in all three trimesters (8 weeks, 14 weeks, and at term). We found that the 3 placental tissues clustered on a single branch and were distinct from the other tissues (Figure 3A). The transcript abundance of the 1,042 MitoMiner genes was lowest in the term placenta, followed by the 14-week placenta, while it was most abundant in muscle, followed by the liver and heart (Figure S5C). However, the term placenta also had 9 transcripts whose abundance levels were higher than any other somatic tissues we compared (Figure 3B); two of them, kynurenine 3-monooxygenase and age-related maculopathy susceptibility 2 (KMO and ARMS2, respectively) were measured using qRT-PCR (discussed further below). Interestingly, multidimensional scaling of all 11,876 samples showed profound clustering of all 49 tissues, indicating tissue-specific

expression of the 1,042 genes (Figures 3C and S6). All placental samples were clustered closely together. Next, using whole-genome sequencing (WGS) datasets, we examined mtDNA copy numbers of the term placental tissue ( $n = 80$ ) and compared those of four healthy tissues (endometrium<sup>33</sup> [ $n = 398$ ], blood<sup>34</sup> [ $n = 199$ ], colon<sup>35</sup> [ $n = 568$ ], and liver<sup>36</sup> [ $n = 517$ ]; Table S12; Figure S7) and 21 non-placental tissues from the Cancer Genome Atlas Pan-Cancer Analysis of Whole Genomes (PCAWG) Consortium.<sup>37</sup> We found that the mtDNA copy number was not substantially lower in the placenta than in other tissues we compared.

### Mitochondrial copy number across multiple tissues

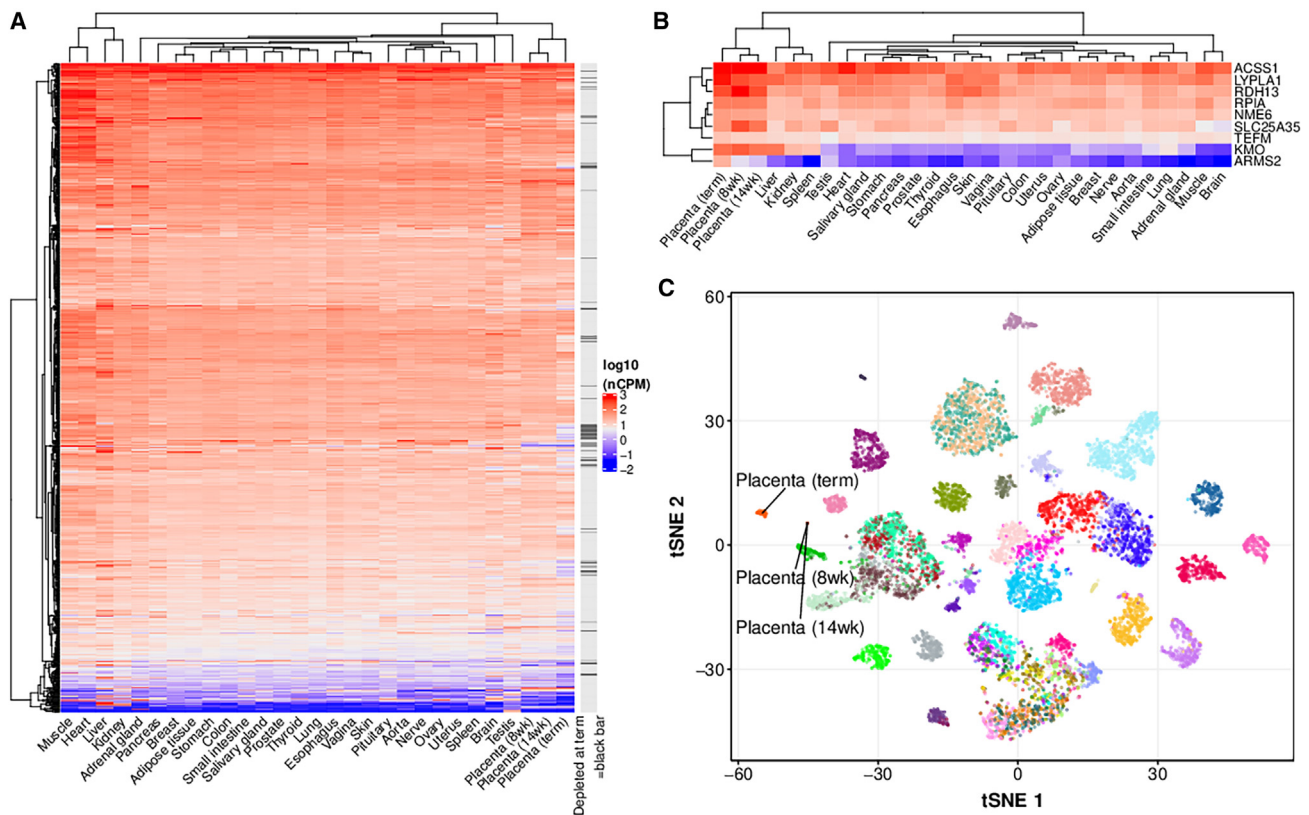
Having observed depleted mitochondrial transcripts in the placenta in the third trimester, we hypothesized that a reduction in these transcripts might be associated with a reduction in the number of mitochondria and that this, in turn, might be reflected in a reduction in mtDNA copy number. We therefore examined mtDNA copy number using the WGS data obtained from the term placental tissue of the POP study cohort ( $n = 80$ ). We then compared the depth of coverage with the following four healthy tissues: endometrium ( $n = 398$ ), blood ( $n = 199$ ), colon ( $n = 568$ ), and liver ( $n = 517$ ). We obtained at least 20 $\times$  coverage for each sample, but this varied across tissues (Figure S7; Table S12). To adjust for differences in sequencing depth, the mtDNA depth of coverage was normalized by the genome depth of coverage, and this measurement was used as the mtDNA copy number.

We compared the mtDNA copy number in the placenta with four other healthy tissues. We found that the median mtDNA copy number of the placenta was 201, which was marginally higher than that of the endometrium (median = 184,  $p = 0.06$ , Mann-Whitney test) but significantly lower than that in blood (median = 250,  $p < 5.4 \times 10^{-6}$ , Mann-Whitney test), colon (median = 520,  $p < 1.0 \times 10^{-44}$ , Mann-Whitney test) and liver (median = 905,  $p < 4.4 \times 10^{-38}$ , Mann-Whitney test) (Figure S7A). To compare the mtDNA copy number in a broader range of tissues, we downloaded mtDNA copy numbers from the 2,157 cancer samples in the Cancer Genome Atlas PCAWG Consortium. Overall, the mtDNA copy number of the placenta was significantly lower than those of 21 non-placental tissues from PCAWG (median = 304,  $p < 4.5 \times 10^{-9}$ , Mann-Whitney test), but the placenta was the seventh lowest of the 22 tissues we studied (Figure S7B).

### Promotor methylation and chromosome bias

To investigate whether CpG island or promotor methylation levels of the depleted genes are different than those of highly expressed placental genes, we compared the methylation levels<sup>38</sup> of 71 protein-coding transcripts specifically enriched in the placenta,<sup>4</sup> which we have reported previously, with those of the depleted transcripts we report here. We did not observe any significant difference in DNA methylation between the enriched and depleted transcripts in the CpG islands associated with these genes (Figure S8A). However, when we studied the promoter regions of the two sets of genes, the median promoter methylation of enriched transcripts (36.9%) was actually higher than that of depleted transcripts (10.2%,  $p = 3.4 \times 10^{-11}$ ).

We investigated the chromosomal locations of the genes encoding depleted transcripts, and we noted that the proportion



**Figure 3. Nuclear-encoded mitochondrial transcripts**

(A) A heatmap representation of the abundance of 1,042 genes in MitoMiner (rows) across 29 tissues (columns). nCPM ( $\log_{10}$  scale) of a transcript is color-coded from red (higher) to blue (lower).

(B) A heatmap showing the abundance ( $\log_{10}$  scale) of 9 highly enriched MitoMiner transcripts in the placenta.

(C) A multidimensional scaling plot (t-distributed stochastic neighbor embedding [tSNE]) of 11,876 samples from 49 tissues using nCPMs (calculated by the TMM method) of 1,042 genes in MitoMiner. Each dot represents one of 11,876 samples (i.e., 11,803 samples from GTEx and 73 from the placenta of the following gestational ages: 8 at 8 weeks, 6 at 14 weeks, and 59 at term), and each color represents a tissue; the full color-coding is shown in Figure S6. For display, non-placental tissues, shown in (A), were manually selected when there were at least two subregions from the same tissue. For example, we analyzed a total of 13 brain subregions in this study, and the cerebellar hemisphere is shown here to represent the brain. The representative subregions are shown in Table S2.

of these genes located on chromosome 19 was higher than expected by chance (odds ratio = 3.18,  $p = 1.5 \times 10^{-25}$ , Fisher's exact test; Figure S8B).

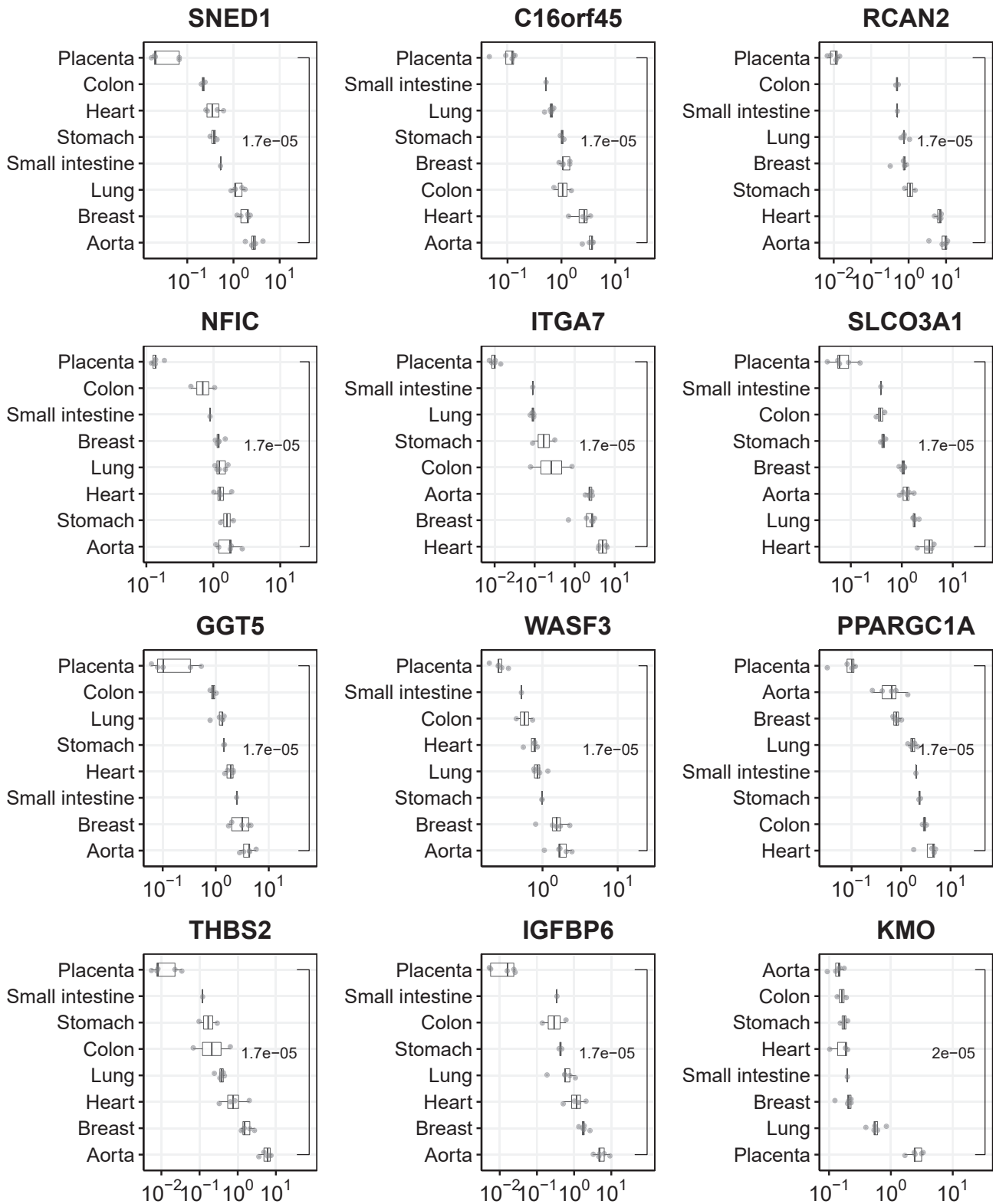
### Validation of placental transcript abundance

To confirm the transcript abundance levels of the placenta and other somatic tissues that we identified based on the RNA-seq dataset, we performed qRT-PCR assays in independent samples. We selected a total of 13 transcripts (11 depleted and 2 mitochondrion-associated transcripts enriched in the term placenta) and measured their mRNA levels in eight human tissues: placenta (term), aorta, heart, breast, lung, stomach, small intestine, and colon (see STAR Methods for details). For each of the 11 depleted transcripts, we confirmed that their abundance was significantly lower in the placenta than in the other tissues ( $p = 1.7 \times 10^{-5}$ , Mann-Whitney test), and the placenta was ranked lowest (Figure 4). We also confirmed that the mRNA level of KMO, one of the enriched transcripts in the placenta, was highest in the placenta, and it was significantly higher ( $p = 2 \times 10^{-5}$ , Mann-Whitney test) than in the other tissues tested

(Figure 4; Table S13). For ARMS2, another enriched target, 19 of the 24 non-placental tissue samples used were not assayable by qPCR (below the limit of detection), whereas all 5 of the placental samples were measurable; this is consistent with the RNA-seq data showing enrichment in the placenta (Figure 3B). Overall, the 13 transcripts selected for validation of either being depleted (11) or enriched (2) from RNA-seq datasets were confirmed using qRT-PCR.

### DISCUSSION

The key finding of the present study is that the human placenta has a unique transcriptome architecture compared with all of the organs studied in the GTEx project. More than 80% of the organs studied in the GTEx project lack even a single uniquely depleted transcript. Of the organs studied in the GTEx project, the liver had the largest number of uniquely depleted transcripts, with 26 depleted or absent. Strikingly, the transcriptome of term placenta had almost 30 times more depleted or absent transcripts than the liver. GO analysis indicated that some of the



(legend on next page)

absent transcripts reflect known unique qualities of placental function. For example, the placenta has no innervation, and this likely reflects the identification by GO analysis of multiple transcripts involved in formation of elements of the nervous system. Similarly, recognition of the allo-immune placenta is essential for normal pregnancy,<sup>39</sup> and this is reflected in unique expression of MHC antigens, which was another GO analysis pathway identified. However, GO analysis identified other pathways that we did not anticipate.

The most striking example was that many transcripts related to mitochondrial function were far less abundant in the term placenta than in other somatic tissues. Interestingly, the mRNA for PPARGC1A (also known as PGC-1 $\alpha$ ), a transcriptional master regulator of mitochondrial biogenesis, was depleted in the placenta in all three trimesters as well as in four external placenta datasets (Table S7). *Ppargc1a*-null mice have decreased expression of mitochondrial genes, especially those encoding various subunits of the electron transport chain,<sup>40,41</sup> suggesting a possible link between its low abundance and the diminished content of mitochondrial transcripts in the placenta. PGC-1 $\alpha$  interacts with a very wide range of transcriptional co-activators, is a key regulator of metabolic homeostasis,<sup>42</sup> and protects cells against oxidative damage by inducing the expression of several ROS (reactive oxygen species)-detoxifying enzymes, such as superoxide dismutase 2 (SOD2). Interestingly, our RNA-seq analysis showed that SOD2 mRNA was the lowest in the early gestation placentas as well as having a low rank (45 of 47) in the term placenta. Because ectopic expression of PGC-1 $\alpha$  reduced levels of ROS,<sup>27,43</sup> it has been suggested that PGC-1 $\alpha$  ensures high energy metabolism and removal of its toxic byproducts at the same time. We recently reported that the placenta has a unique somatic mutation profile,<sup>44</sup> predominantly the SBS18 signature, which is associated with oxidative stress. This could be explained, at least in part, by accumulation of ROS, possibly because of lower levels of PGC-1 $\alpha$  and SOD2 transcripts in the placenta.

The syncytiotrophoblast (STB) is a multinucleated epithelium covering the outer layer of chorionic villi, and it differentiates from cytotrophoblast (CTB). STB mitochondria have different morphological properties compared with CTB (specifically, they are smaller with irregular spherical cristae and a dense matrix),<sup>45,46</sup> and it has been suggested that these changes are related to steroidogenesis of STB.<sup>47,48</sup> Functional studies have shown that STB mitochondria have reduced membrane potential, increased levels of hydrogen peroxide, and lower antioxidant levels and are more sensitive to ROS.<sup>49–51</sup> These characteristics are due to the complement of mitochondrial proteins present and again reflect the features of placental biology that are not found in other tissues.

We have reported previously that placental polyamine metabolism is implicated in placental related complications of human pregnancy.<sup>14</sup> In the current study, we found that genes associated with the KEGG pathway “arginine and proline meta-

bolism” were overrepresented among the depleted transcripts. In this group were transcripts from five genes (SRM, SAT2, SMOX, AZIN2, and PAOX) involved in polyamine metabolism; these were depleted at term but were also less abundant earlier in gestation (Figure S4A). Interestingly, expression levels of some genes in the polyamine pathway (e.g., *SAT1*, *AZIN1*, *SMS*, and *AMD1*) were higher in the placenta than in non-placental tissues. Kajander et al.<sup>52</sup> have reported enzymatic activities of SRM and spermine synthase (SMS) in seven human tissues, and the SMS-to-SRM ratio was highest in the placenta (~5), followed by the kidneys (~3.7). In our comparative analysis of RNA-seq datasets, we confirmed that the SMS-to-SRM ratio at the transcript level was also highest in the term placenta (48.7), followed by the 14-week placenta (5.4) and the 8-week placenta (5.3). The kidney cortex (2.5) appeared to have the highest ratio among 46 non-placenta tissues we compared, with the pancreas (0.17) being the lowest; it was also the lowest from the protein ratio (0.4) reported by Kajander et al.<sup>52</sup> This demonstrates that our tissue-wide comparison of transcripts parallels previous analyses based on measurement of proteins level.

The mechanisms underlying the reduced expression of the selectively depleted transcripts in the placenta remains to be determined. We considered possible differences in the number of mitochondria per nucleus, but the placenta was not an outlier in mitochondrial DNA copy number (Figure S7), and some mitochondrial genes actually had higher levels of expression in the placenta than in other tissues. Something else we considered was differential methylation of the promoter regions of such genes. While it has long been recognized that placental DNA is globally hypomethylated compared with other tissues,<sup>53,54</sup> it varies in a locus-specific manner.<sup>38,55</sup> We observed that the median promoter methylation of enriched transcripts was actually higher than that of depleted transcripts (36.9% vs. 10.2%,  $p = 3.4 \times 10^{-11}$ ; Figure S8A) and runs counter to the typical “high-methylation-low-expression” relationship; this suggests that other mechanisms may be involved, and this is an area for future study.

A higher proportion of depleted transcripts than expected (odds ratio = 3.18,  $p = 1.5 \times 10^{-25}$ ; Figure S8B) are encoded on chromosome 19, and this chromosome has the highest gene density of all human chromosomes and the highest GC and CpG content. It is unusual in that nearly a quarter of its genes fall in 20 tandemly arranged gene families.<sup>56,57</sup> Several of these families have direct roles in pregnancy; for example, the beta subunit of chorionic gonadotropin (CGB; 6 functional genes), the pregnancy-specific glycoproteins (PSGs; 10 functional genes), and the large imprinted cluster of microRNAs (miRNAs), which are largely placenta specific<sup>58</sup> (C19CM; 46 genes). These genes are all highly expressed in the placenta.<sup>4</sup> However, chromosome 19 encodes other unusual gene families. Natural killer cells play an important role in human pregnancy,<sup>59</sup> and their receptors (killer cell immunoglobulin-like receptors [KIRs]) are

#### Figure 4. qPCR validation assays for 14 transcripts

The transcript abundance levels (x axis) were measured in 8 tissues (y axis) which were ranked from the lowest to highest level using the median abundance for each transcript. Each dot represents an individual sample (by taking the mean of the technical triplicate), and each of the boxes shows the median and IQR. The horizontal lines (whiskers) extended from the box represent a range of  $1.5 \times$  IQR from both ends. The p values (Mann-Whitney test) are shown between the placenta and the rest of non-placental tissues. The qPCR data used for the graph are available in Table S13.

highly polymorphic and are in a cluster within the 1-Mb leukocyte receptor complex in chromosome 19.

Finally, transposable elements or their remains account for a surprisingly large fraction of the human genome, and host organisms have evolved numerous strategies to defend themselves against the threat posed by functional endogenous retroelements. One such mechanism involves a very large and rapidly evolving family of transcription factors: Krüppel-associated box domain zinc-finger proteins (KRAB-ZFPs).<sup>60</sup> There are an estimated 352 genes encoding KRAB-ZFPs in the human genome, and 209 of these are located within six clusters on chromosome 19.<sup>61</sup> Thus chromosome 19 has many unusual features, and many of these are related to placental or reproductive function. We now add to its list of unusual features: having an overrepresentation of depleted transcripts in the human placenta.

We investigated whether there was any overlap between the depleted transcripts (762 at term) and those we have identified previously as being dysregulated in preeclampsia or growth restriction (a total of 784).<sup>4</sup> Sixteen transcripts are present on both lists, but this is fewer than would be expected by chance alone—the expected number would be 30. We therefore conclude that the placental dysregulation in PE and growth restriction occurs among genes that are more highly expressed and present in other tissues and that the depleted genes are not upregulated in these conditions.

We compared the RNA-seq data obtained from isolated human trophoblast subpopulations and trophoblast stem cells (Okae et al.<sup>62</sup>) with our data obtained at 8 weeks' gestation. 169 of the 214 depleted transcripts in the first trimester are in the bottom 50% of transcripts listed in Okae et al.<sup>62</sup> when ranked by mean log<sub>2</sub>(fragments per kilobase of transcript per million mapped reads [FPKM]+1). This is a very significant overrepresentation among the weakly expressed transcripts ( $p = 6.47 \times 10^{-43}$ ) and indicates that the transcripts we identified as depleted are also weakly expressed in isolated trophoblast cells.

Clearly, there is a risk that apparent differences in expression could be observed because of batch and other confounding effects, and this could occur at any of a range of levels, including RNA extraction, the sequencing platform employed, the transcriptome size, and the normalization method employed. For example, the results of long RNA-seq are profoundly affected by whether RNA was size selected or selected by oligo(dT)s for selective extraction of mRNA through its polyadenylated tail.<sup>4</sup> We mitigated these risks using multiple approaches. First, we ensured comparable methods of RNA extraction and sequencing between our own samples (mirVana miRNA isolation Kit, Ambion) and GTEx (Tissue miRNA Kit [PreAnalytiX] and miRNeasy Mini Kit [QIAGEN]). Second, we analyzed multiple distinct placental RNA-seq datasets, including those we generated and other publicly available sources. Third, we compared multiple bioinformatics approaches to determine that the results were robust to definitions using different pipelines. Fourth, we showed that downsampling the number of reads, the number of samples, or both the reads and the samples in the GTEx data had very little effect on the number of depleted transcripts identified. Finally, we externally validated some of the key results using qPCR and a separate group of

placental samples and tissue samples obtained from a local tissue bank. Transcriptome complexity varies among tissues, and the placenta has the seventh least complex transcriptome compared with the solid tissue in GTEx.<sup>4</sup> We considered whether this complexity might have some bearing on the number of depleted transcripts detected. We analyzed the GTEx tissue using the same selection criteria, and these tissues have many fewer depleted transcripts. Most have no depleted transcripts at all, but those that do have a range of transcriptome complexities. The brain and testis have the most complex transcriptomes and have 9 and 4 depleted transcripts, respectively. The solid tissue with the least complex transcriptome (the pancreas) has no depleted transcripts. However, the liver (third least complex) has 26 depleted transcripts. Thus, among the GTEx tissues, there is no relationship between transcriptome complexity and the number of depleted transcripts. Therefore, our results cannot be explained by differences in transcriptome complexity. We conclude that these data support a unique transcriptomic void in the human placenta. We speculate that this void might identify transcripts that are dispensable in a rapidly growing and transient organ but not in others. Moreover, neglect of the placenta in a large-scale international consortium resulted in exclusion of one of the human body's most interesting transcriptional landscapes.

#### Limitations of the study

We analyzed sequence data from several previous studies and took multiple approaches to avoid possible batch effects. Nonetheless, the possibility remains that such artifacts might still be present. Having identified a large number of transcripts that are depleted in the placenta, it is a conceptual challenge to determine the biological significance of a consistent reduction of a large population of transcripts. Future studies to investigate the transcript landscape of other rapidly growing but transient tissues, such as the corpus luteum in the ovary, could address whether the observations reported here are similar in other transient tissues.

#### STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- **KEY RESOURCES TABLE**
- **RESOURCE AVAILABILITY**
  - Lead contact
  - Materials availability
  - Data and code availability
- **EXPERIMENTAL MODEL AND SUBJECT DETAILS**
  - Placenta samples
  - Tissue collection for RT-qPCR analysis
  - Study approval
- **METHOD DETAILS**
  - RNA sequencing and data processing
  - Whole-genome sequencing and data processing
  - GTEx data processing
  - Identification of absent or depleted protein-coding transcripts

- RNA extraction for RT-qPCR validation
- Target selection for RT-qPCR validation
- Calculation of mitochondrial copy number
- Gene Ontology analysis
- Dimension reduction
- External datasets used in this study
- Identification of transcripts localized in the mitochondria
- **QUANTIFICATION AND STATISTICAL ANALYSIS**

#### SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.celrep.2023.112800>.

#### ACKNOWLEDGMENTS

This work was supported by the Medical Research Council (United Kingdom; G1100221 and MR/K021133/1) and the National Institute for Health Research (NIHR) Cambridge Biomedical Research Centre (Women's Health theme). I.L.M.H.A. was funded by the Centre for Trophoblast Research, University of Cambridge (CTR) Next Generation Fellowship. G.A. is funded by the CTR PhD scholarship. A.R.J.L. and L.M.R.H. are funded by Wellcome PhD studentships. We would like to thank Katrina Holmes, Josephine Gill, Leah Bibby, Samudra Ranawaka, and Ryan Millar for technical assistance during the study. We would like to thank Dr. Álvaro Cortés-Calabuig for kindly sharing research data. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR, or the Department of Health and Social Care.

#### AUTHOR CONTRIBUTIONS

D.S.C.-J. and G.C.S.S. conceived the experiments. S.G., D.S.C.-J., and G.C.S.S. designed the experiments. F.G., I.L.M.H.A., G.A., and E.C. performed the experiments. S.G. analyzed all sequencing data. E.C. managed sample collection and processing and the biobank in which all samples were stored. A.R.J.L. and L.M.R.H. provided and processed sequence data. All authors contributed to writing the manuscript and approved the final version.

#### DECLARATION OF INTERESTS

D.S.C.-J. reports non-financial support from Roche Diagnostics Ltd. outside the submitted work. G.C.S.S. reports personal fees and non-financial support from Roche Diagnostics Ltd. outside the submitted work. D.S.C.-J. and G.C.S.S. report grants from Sera Prognostics Inc. and non-financial support from Illumina Inc. outside the submitted work. G.C.S.S. has been a paid consultant to GSK (preterm birth) and is a member of a data monitoring committee for GSK trials of RSV vaccination in pregnancy.

Received: September 22, 2022

Revised: February 8, 2023

Accepted: June 26, 2023

Published: July 14, 2023

#### REFERENCES

1. Brosens, I., Pijnenborg, R., Vercruyse, L., and Romero, R. (2011). The "Great Obstetrical Syndromes" are associated with disorders of deep placentation. *Am. J. Obstet. Gynecol.* *204*, 193–201. <https://doi.org/10.1016/j.ajog.2010.08.009>.
2. Armstrong, D.L., McGowen, M.R., Weckle, A., Pantham, P., Caravas, J., Agnew, D., Benirschke, K., Savage-Rumbaugh, S., Nevo, E., Kim, C.J., et al. (2017). The core transcriptome of mammalian placentas and the divergence of expression with placental shape. *Placenta* *57*, 71–78. <https://doi.org/10.1016/j.placenta.2017.04.015>.
3. Buckberry, S., Bianco-Miotto, T., Bent, S.J., Clifton, V., Shoubridge, C., Shankar, K., and Roberts, C.T. (2017). Placental transcriptome co-expression analysis reveals conserved regulatory programs across gestation. *BMC Genom.* *18*, 10. <https://doi.org/10.1186/s12864-016-3384-9>.
4. Gong, S., Gaccioli, F., Dopierala, J., Sovio, U., Cook, E., Volders, P.J., Martens, L., Kirk, P.D.W., Richardson, S., Smith, G.C.S., and Charnock-Jones, D.S. (2021). The RNA landscape of the human placenta in health and disease. *Nat. Commun.* *12*, 2639. <https://doi.org/10.1038/s41467-021-22695-y>.
5. Kim, J., Zhao, K., Jiang, P., Lu, Z.x., Wang, J., Murray, J.C., and Xing, Y. (2012). Transcriptome landscape of the human placenta. *BMC Genom.* *13*, 115. <https://doi.org/10.1186/1471-2164-13-115>.
6. ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature* *489*, 57–74. <https://doi.org/10.1038/nature11247>.
7. Roadmap Epigenomics Consortium; Kundaje, A., Meuleman, W., Ernst, J., Bilenyk, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., et al. (2015). Integrative analysis of 111 reference human epigenomes. *Nature* *518*, 317–330. <https://doi.org/10.1038/nature14248>.
8. de Rie, D., Abugessaisa, I., Alam, T., Arner, E., Arner, P., Ashoor, H., Åström, G., Babina, M., Bertin, N., Burroughs, A.M., et al. (2017). An integrated expression atlas of miRNAs and their promoters in human and mouse. *Nat. Biotechnol.* *35*, 872–878. <https://doi.org/10.1038/nbt.3947>.
9. GTEx Consortium (2020). The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* *369*, 1318–1330. <https://doi.org/10.1126/science.aaz1776>.
10. Jain, A., and Tuteja, G. (2019). TissueEnrich: Tissue-specific gene enrichment analysis. *Bioinformatics* *35*, 1966–1967. <https://doi.org/10.1093/bioinformatics/bty890>.
11. Papatheodorou, I., Moreno, P., Manning, J., Fuentes, A.M.P., George, N., Fexova, S., Fonseca, N.A., Füllgrabe, A., Green, M., Huang, N., et al. (2020). Expression Atlas update: from tissues to single cells. *Nucleic Acids Res.* *48*, D77–D83. <https://doi.org/10.1093/nar/gkz947>.
12. Watanabe, K., Umičević Mirkov, M., de Leeuw, C.A., van den Heuvel, M.P., and Posthuma, D. (2019). Genetic mapping of cell type specificity for complex traits. *Nat. Commun.* *10*, 3222. <https://doi.org/10.1038/s41467-019-11181-1>.
13. Eisenberg, E., and Levanon, E.Y. (2013). Human housekeeping genes, revisited. *Trends Genet.* *29*, 569–574. <https://doi.org/10.1016/j.tig.2013.05.010>.
14. Gong, S., Sovio, U., Aye, I.L., Gaccioli, F., Dopierala, J., Johnson, M.D., Wood, A.M., Cook, E., Jenkins, B.J., Kouman, A., et al. (2018). Placental polyamine metabolism differs by fetal sex, fetal growth restriction, and preeclampsia. *JCI Insight* *3*, e120723. <https://doi.org/10.1172/jci.insight.120723>.
15. Gaccioli, F., Lager, S., Sovio, U., Charnock-Jones, D.S., and Smith, G.C. (2017). The pregnancy outcome prediction (POP) study: Investigating the relationship between serial prenatal ultrasonography, biomarkers, placental phenotype and adverse pregnancy outcomes. *Placenta* *59*, S17–S25. <https://doi.org/10.1016/j.placenta.2016.10.011>.
16. Pasupathy, D., Dacey, A., Cook, E., Charnock-Jones, D.S., White, I.R., and Smith, G.C.S. (2008). Study protocol. A prospective cohort study of unselected primiparous women: the pregnancy outcome prediction study. *BMC Pregnancy Childbirth* *8*, 51. <https://doi.org/10.1186/1471-2393-8-51>.
17. Prater, M., Hamilton, R.S., Wa Yung, H., Sharkey, A.M., Robson, P., Abd Hamid, N.E., Jauniaux, E., Charnock-Jones, D.S., Burton, G.J., and Cindrova-Davies, T. (2021). RNA-Seq reveals changes in human placental metabolism, transport and endocrinology across the first-second trimester transition. *Biol. Open* *10*, bio058222. <https://doi.org/10.1242/bio.058222>.
18. Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biol.* *11*, R106. <https://doi.org/10.1186/gb-2010-11-10-r106>.

19. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* *15*, 550. <https://doi.org/10.1186/s13059-014-0550-8>.
20. Robinson, M.D., and Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* *11*, R25. <https://doi.org/10.1186/gb-2010-11-3-r25>.
21. Lim, Y.C., Li, J., Ni, Y., Liang, Q., Zhang, J., Yeo, G.S.H., Lyu, J., Jin, S., and Ding, C. (2017). A complex association between DNA methylation and gene expression in human placenta at first and third trimesters. *PLoS One* *12*, e0181155. <https://doi.org/10.1371/journal.pone.0181155>.
22. Huang, Z., Du, G., Huang, X., Han, L., Han, X., Xu, B., Zhang, Y., Yu, M., Qin, Y., Xia, Y., et al. (2018). The enhancer RNA Inc-SLC4A1-1 epigenetically regulates unexplained recurrent pregnancy loss (URPL) by activating CXCL8 and NF- $\kappa$ B pathway. *EBioMedicine* *38*, 162–170. <https://doi.org/10.1016/j.ebiom.2018.11.015>.
23. Verheecke, M., Cortès Calabuig, A., Finalet Ferreira, J., Brys, V., Van Bree, R., Verbist, G., Everaert, T., Leemans, L., Gziri, M.M., Boere, I., et al. (2018). Genetic and microscopic assessment of the human chemotherapy-exposed placenta reveals possible pathways contributive to fetal growth restriction. *Placenta* *64*, 61–70. <https://doi.org/10.1016/j.placenta.2018.03.002>.
24. Ashley, B., Simner, C., Manousopoulou, A., Jenkinson, C., Hey, F., Frost, J.M., Rezwani, F.I., White, C.H., Lofthouse, E.M., Hyde, E., et al. (2022). Placental uptake and metabolism of 25(OH)vitamin D determine its activity within the fetoplacental unit. *Elife* *11*, e71094. <https://doi.org/10.7554/eLife.71094>.
25. Awamleh, Z., Gloor, G.B., and Han, V.K.M. (2019). Placental microRNAs in pregnancies with early onset intrauterine growth restriction and preeclampsia: potential impact on gene expression and pathophysiology. *BMC Med. Genomics* *12*, 91. <https://doi.org/10.1186/s12920-019-0548-x>.
26. Penning, T.M., Chen, M., and Jin, Y. (2015). Promiscuity and diversity in 3-ketosteroid reductases. *J. Steroid Biochem. Mol. Biol.* *151*, 93–101. <https://doi.org/10.1016/j.jsbmb.2014.12.003>.
27. Valle, I., Alvarez-Barrientos, A., Arza, E., Lamas, S., and Monsalve, M. (2005). PGC-1 $\alpha$  regulates the mitochondrial antioxidant defense system in vascular endothelial cells. *Cardiovasc. Res.* *66*, 562–573. <https://doi.org/10.1016/j.cardiores.2005.01.026>.
28. Hondares, E., Mora, O., Yubero, P., Rodríguez de la Concepción, M., Iglesias, R., Giralt, M., and Villarroya, F. (2006). Thiazolidinediones and rexinoids induce peroxisome proliferator-activated receptor-coactivator (PGC)-1 $\alpha$  gene transcription: an autoregulatory loop controls PGC-1 $\alpha$  expression in adipocytes via peroxisome proliferator-activated receptor-gamma coactivation. *Endocrinology* *147*, 2829–2838. <https://doi.org/10.1210/en.2006-0070>.
29. Wu, Z., Puigserver, P., Andersson, U., Zhang, C., Adelman, G., Mootha, V., Troy, A., Cinti, S., Lowell, B., Scarpulla, R.C., and Spiegelman, B.M. (1999). Mechanisms controlling mitochondrial biogenesis and respiration through the thermogenic coactivator PGC-1. *Cell* *98*, 115–124. [https://doi.org/10.1016/S0092-8674\(00\)80611-X](https://doi.org/10.1016/S0092-8674(00)80611-X).
30. Hudson, B.G., Tryggvason, K., Sundaramoorthy, M., and Neilson, E.G. (2003). Alport's syndrome, Goodpasture's syndrome, and type IV collagen. *N. Engl. J. Med.* *348*, 2543–2556. <https://doi.org/10.1056/NEJMra022296>.
31. Moll, R., Divo, M., and Langbein, L. (2008). The human keratins: biology and pathology. *Histochem. Cell Biol.* *129*, 705–733. <https://doi.org/10.1007/s00418-008-0435-6>.
32. Smith, A.C., and Robinson, A.J. (2016). MitoMiner v3.1, an update on the mitochondrial proteomics database. *Nucleic Acids Res.* *44*, D1258–D1261. <https://doi.org/10.1093/nar/gkv1001>.
33. Moore, L., Leongamornlert, D., Coorens, T.H.H., Sanders, M.A., Ellis, P., Dento, S.C., Dawson, K.J., Butler, T., Rahbari, R., Mitchell, T.J., et al. (2020). The mutational landscape of normal human endometrial epithelium. *Nature* *580*, 640–646. <https://doi.org/10.1038/s41586-020-2214-z>.
34. Lee-Six, H., Øbro, N.F., Shepherd, M.S., Grossmann, S., Dawson, K., Belmonte, M., Osborne, R.J., Huntly, B.J.P., Martincorena, I., Anderson, E., et al. (2018). Population dynamics of normal human blood inferred from somatic mutations. *Nature* *561*, 473–478. <https://doi.org/10.1038/s41586-018-0497-0>.
35. Lee-Six, H., Olafsson, S., Ellis, P., Osborne, R.J., Sanders, M.A., Moore, L., Georgakopoulos, N., Torrente, F., Noorani, A., Goddard, M., et al. (2019). The landscape of somatic mutation in normal colorectal epithelial cells. *Nature* *574*, 532–537. <https://doi.org/10.1038/s41586-019-1672-7>.
36. Brunner, S.F., Roberts, N.D., Wylie, L.A., Moore, L., Aitken, S.J., Davies, S.E., Sanders, M.A., Ellis, P., Alder, C., Hooks, Y., et al. (2019). Somatic mutations and clonal dynamics in healthy and cirrhotic human liver. *Nature* *574*, 538–542. <https://doi.org/10.1038/s41586-019-1670-9>.
37. Yuan, Y., Ju, Y.S., Kim, Y., Li, J., Wang, Y., Yoon, C.J., Yang, Y., Martincorena, I., Creighton, C.J., Weinstein, J.N., et al. (2020). Comprehensive molecular characterization of mitochondrial genomes in human cancers. *Nat. Genet.* *52*, 342–352. <https://doi.org/10.1038/s41588-019-0557-x>.
38. Gong, S., Johnson, M.D., Dopierala, J., Gaccioli, F., Sovio, U., Constância, M., Smith, G.C., and Charnock-Jones, D.S. (2018). Genome-wide oxidative bisulfite sequencing identifies sex-specific methylation differences in the human placenta. *Epigenetics* *13*, 228–239. <https://doi.org/10.1080/15592294.2018.1429857>.
39. Moffett, A., Chazara, O., and Colucci, F. (2017). Maternal allo-recognition of the fetus. *Fertil. Steril.* *107*, 1269–1272. <https://doi.org/10.1016/j.fertnstert.2017.05.001>.
40. Austin, S., and St-Pierre, J. (2012). PGC1 $\alpha$  and mitochondrial metabolism—emerging concepts and relevance in ageing and neurodegenerative disorders. *J. Cell Sci.* *125* (Pt 21), 4963–4971. <https://doi.org/10.1242/jcs.113662>.
41. Vernier, M., and Giguère, V. (2021). Aging, senescence and mitochondria: the PGC-1/ERR axis. *J. Mol. Endocrinol.* *66*, R1–R14. <https://doi.org/10.1530/JME-20-0196>.
42. Miller, K.N., Clark, J.P., and Anderson, R.M. (2019). Mitochondrial regulator PGC-1 $\alpha$ —Modulating the modulator. *Curr. Opin. Endocr. Metab. Res.* *5*, 37–44. <https://doi.org/10.1016/j.coemr.2019.02.002>.
43. St-Pierre, J., Drori, S., Uldry, M., Silvaggi, J.M., Rhee, J., Jäger, S., Handschin, C., Zheng, K., Lin, J., Yang, W., et al. (2006). Suppression of reactive oxygen species and neurodegeneration by the PGC-1 transcriptional coactivators. *Cell* *127*, 397–408. <https://doi.org/10.1016/j.cell.2006.09.024>.
44. Coorens, T.H.H., Oliver, T.R.W., Sanghvi, R., Sovio, U., Cook, E., Vento-Tormo, R., Haniffa, M., Young, M.D., Rahbari, R., Sebire, N., et al. (2021). Inherent mosaicism and extensive mutation of human placentas. *Nature* *592*, 80–85. <https://doi.org/10.1038/s41586-021-03345-1>.
45. Holland, O., Dekker Nitert, M., Gallo, L.A., Vejzovic, M., Fisher, J.J., and Perkins, A.V. (2017). Review: Placental mitochondrial function and structure in gestational disorders. *Placenta* *54*, 2–9. <https://doi.org/10.1016/j.placenta.2016.12.012>.
46. Fisher, J.J., Bartho, L.A., Perkins, A.V., and Holland, O.J. (2020). Placental mitochondria and reactive oxygen species in the physiology and pathophysiology of pregnancy. *Clin. Exp. Pharmacol. Physiol.* *47*, 176–184. <https://doi.org/10.1111/1440-1681.13172>.
47. Martínez, F., Kiriakidou, M., and Strauss, J.F. (1997). Structural and functional changes in mitochondria associated with trophoblast differentiation: methods to isolate enriched preparations of syncytiotrophoblast mitochondria. *Endocrinology* *138*, 2172–2183. <https://doi.org/10.1210/endo.138.5.5133>.
48. Martínez, F., Olvera-Sánchez, S., Esparza-Perusquia, M., Gomez-Chang, E., and Flores-Herrera, O. (2015). Multiple functions of syncytiotrophoblast mitochondria. *Steroids* *103*, 11–22. <https://doi.org/10.1016/j.steroids.2015.09.006>.
49. Bustamante, J., Ramírez-Vélez, R., Czerniczyniec, A., Cicerchia, D., Aguilar de Plata, A.C., and Lores-Arnaiz, S. (2014). Oxygen metabolism in

- human placenta mitochondria. *J. Bioenerg. Biomembr.* **46**, 459–469. <https://doi.org/10.1007/s10863-014-9572-x>.
50. Watson, A.L., Skepper, J.N., Jauniaux, E., and Burton, G.J. (1998). Susceptibility of human placental syncytiotrophoblastic mitochondria to oxygen-mediated damage in relation to gestational age. *J. Clin. Endocrinol. Metab.* **83**, 1697–1705. <https://doi.org/10.1210/jcem.83.5.4830>.
  51. Schoots, M.H., Gordijn, S.J., Scherjon, S.A., van Goor, H., and Hillebrands, J.-L. (2018). Oxidative stress in placental pathology. *Placenta* **69**, 153–161. <https://doi.org/10.1016/j.placenta.2018.03.003>.
  52. Kajander, E.O., Kauppinen, L.I., Pajula, R.L., Karkola, K., and Eloranta, T.O. (1989). Purification and partial characterization of human polyamine synthases. *Biochem. J.* **259**, 879–886. <https://doi.org/10.1042/bj2590879>.
  53. Fuke, C., Shimabukuro, M., Petronis, A., Sugimoto, J., Oda, T., Miura, K., Miyazaki, T., Ogura, C., Okazaki, Y., and Jinno, Y. (2004). Age related changes in 5-methylcytosine content in human peripheral leukocytes and placentas: an HPLC-based study. *Ann. Hum. Genet.* **68** (Pt 3), 196–204. <https://doi.org/10.1046/j.1529-8817.2004.00081.x>.
  54. Ehrlich, M., Gama-Sosa, M.A., Huang, L.H., Midgett, R.M., Kuo, K.C., McCune, R.A., and Gehrke, C. (1982). Amount and distribution of 5-methylcytosine in human DNA from different types of tissues of cells. *Nucleic Acids Res.* **10**, 2709–2721. <https://doi.org/10.1093/nar/10.8.2709>.
  55. Chatterjee, A., Macaulay, E.C., Rodger, E.J., Stockwell, P.A., Parry, M.F., Roberts, H.E., Slatter, T.L., Hung, N.A., Devenish, C.J., and Morison, I.M. (2016). Placental hypomethylation is more pronounced in genomic loci devoid of retroelements. *G3 (Bethesda)* **6**, 1911–1921. <https://doi.org/10.1534/g3.116.030379>.
  56. Grimwood, J., Gordon, L.A., Olsen, A., Terry, A., Schmutz, J., Lamerdin, J., Hellsten, U., Goodstein, D., Couronne, O., Tran-Gyamfi, M., et al. (2004). The DNA sequence and biology of human chromosome 19. *Nature* **428**, 529–535. <https://doi.org/10.1038/nature02399>.
  57. Harris, R.A., Raveendran, M., Worley, K.C., and Rogers, J. (2020). Unusual sequence characteristics of human chromosome 19 are conserved across 11 nonhuman primates. *BMC Evol. Biol.* **20**, 33. <https://doi.org/10.1186/s12862-020-1595-9>.
  58. Donker, R.B., Mouillet, J.F., Chu, T., Hubel, C.A., Stolz, D.B., Morelli, A.E., and Sadovsky, Y. (2012). The expression profile of C19MC microRNAs in primary human trophoblast cells and exosomes. *Mol. Hum. Reprod.* **18**, 417–424. <https://doi.org/10.1093/molehr/gas013>.
  59. Colucci, F. (2019). The immunological code of pregnancy. *Science* **365**, 862–863. <https://doi.org/10.1126/science.aaw1300>.
  60. Bruno, M., Mahgoub, M., and Macfarlan, T.S. (2019). The Arms Race Between KRAB-Zinc Finger Proteins and Endogenous Retroelements and Its Impact on Mammals. *Annu. Rev. Genet.* **53**, 393–416. <https://doi.org/10.1146/annurev-genet-112618-043717>.
  61. Yang, P., Wang, Y., and Macfarlan, T.S. (2017). The Role of KRAB-ZFPs in Transposable Element Repression and Mammalian Evolution. *Trends Genet.* **33**, 871–881. <https://doi.org/10.1016/j.tig.2017.08.006>.
  62. Okae, H., Toh, H., Sato, T., Hiura, H., Takahashi, S., Shirane, K., Kabayama, Y., Suyama, M., Sasaki, H., and Arima, T. (2018). Derivation of human trophoblast stem cells. *Cell Stem Cell* **22**, 50–63.e6. <https://doi.org/10.1016/j.stem.2017.11.004>.
  63. Gong, S. (2023). The human placenta exhibits a unique transcriptomic void. Preprint at Zenodo. <https://doi.org/10.5281/zenodo.7938169>.
  64. Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. j.* **17**, 10. <https://doi.org/10.14806/ej.17.1.200>.
  65. Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, R36. <https://doi.org/10.1186/gb-2013-14-4-r36>.
  66. Cunningham, F., Allen, J.E., Allen, J., Alvarez-Jarreta, J., Amode, M.R., Armean, I.M., Austine-Orimoloye, O., Azov, A.G., Barnes, I., Bennett, R., et al. (2022). Ensembl 2022. *Nucleic Acids Res.* **50**, D988–D995. <https://doi.org/10.1093/nar/gkab1049>.
  67. Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M., and Li, H. (2021). Twelve years of SAMtools and BCFtools. *GigaScience* **10**, giab008. <https://doi.org/10.1093/gigascience/giab008>.
  68. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>.
  69. Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930. <https://doi.org/10.1093/bioinformatics/btt656>.
  70. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842. <https://doi.org/10.1093/bioinformatics/btq033>.
  71. Raudvere, U., Kolberg, L., Kuzmin, I., Arak, T., Adler, P., Peterson, H., and Vilo, J. (2019). Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res.* **47**, W191–W198. <https://doi.org/10.1093/nar/gkz369>.
  72. Kolberg, L., Raudvere, U., Kuzmin, I., Vilo, J., and Peterson, H. (2020). gprofiler2 – an R package for gene list functional enrichment analysis and namespace conversion toolset g:Profiler. *F1000Res.* **9**. <https://doi.org/10.12688/f1000research.24956.2>.
  73. Krijthe, J.H. (2015). Rtsne: T-Distributed Stochastic Neighbor Embedding Using Barnes-Hut (CRAN).
  74. de Goffau, M.C., Lager, S., Sovio, U., Gaccioli, F., Cook, E., Peacock, S.J., Parkhill, J., Charnock-Jones, D.S., and Smith, G.C.S. (2019). Human placenta has no microbiome but can contain potential pathogens. *Nature* **572**, 329–334. <https://doi.org/10.1038/s41586-019-1451-5>.
  75. Yang, L., Duff, M.O., Graveley, B.R., Carmichael, G.G., and Chen, L.-L. (2011). Genomewide characterization of non-polyadenylated RNAs. *Genome Biol.* **12**, R16. <https://doi.org/10.1186/gb-2011-12-2-r16>.
  76. Robinson, D.G., and Storey, J.D. (2014). subSeq: determining appropriate sequencing depth through efficient read subsampling. *Bioinformatics* **30**, 3424–3426. <https://doi.org/10.1093/bioinformatics/btu552>.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<b>Critical commercial assays</b>		
ARMS2 TaqMan q-PCR assay	ThermoFisher Scientific	Hs01394203_m1
C16orf45 TaqMan q-PCR assay	ThermoFisher Scientific	Hs01014981_m1
CDC34 TaqMan q-PCR assay	ThermoFisher Scientific	Hs07288692_g1
GGT5 TaqMan q-PCR assay	ThermoFisher Scientific	Hs00897715_m1
IGFBP6 TaqMan q-PCR assay	ThermoFisher Scientific	Hs00181853_m1
ITGA7 TaqMan q-PCR assay	ThermoFisher Scientific	Hs01056475_m1
KMO TaqMan q-PCR assay	ThermoFisher Scientific	Hs00175738_m1
NFIC TaqMan q-PCR assay	ThermoFisher Scientific	Hs00232157_m1
PGC1A TaqMan q-PCR assay	ThermoFisher Scientific	Hs00173304_m1
RCAN2 TaqMan q-PCR assay	ThermoFisher Scientific	Hs00195165_m1
SCLO3A1 TaqMan q-PCR assay	ThermoFisher Scientific	Hs00203184_m1
SNED TaqMan q-PCR assay	ThermoFisher Scientific	Hs00966449_m1
TBP TaqMan q-PCR assay	ThermoFisher Scientific	Hs00427620_m1
THBS2 TaqMan q-PCR assay	ThermoFisher Scientific	Hs01568063_m1
WASF3 TaqMan q-PCR assay	ThermoFisher Scientific	Hs00903488_m1
Lysing Matrix S tubes	MP Biomedical	116925050
RNeasy Plus Mini Kit	Qiagen	74134
High-capacity RNA-to-cDNA kit	ThermoFisher Scientific	4387406
TaqMan Multiplex Master Mix	ThermoFisher Scientific	4486295
RNAlater	ThermoFisher Scientific	AM7021
mirVana miRNA Isolation Kit	ThermoFisher Scientific	AM1560
DNA-free DNA Removal Kit	ThermoFisher Scientific	AM1906
TruSeq Stranded mRNA Library Prep Kit	Illumina	20020595
<b>Biological samples</b>		
Full term human placental samples from the Pregnancy Outcome Prediction study (POPs)	Gong et al., <sup>4</sup> Gaccioli et al., <sup>15</sup> Pasupathy et al. <sup>16</sup>	N/A
First and second trimester human placental samples, previously described.	Prater et al. <sup>17</sup>	N/A
Human aorta, breast, colon, heart, lung, small intestine and stomach	This study	N/A
<b>Deposited data</b>		
Term placenta RNA-Seq data	The EGA ( <a href="https://ega-archive.org/">https://ega-archive.org/</a> )	EGAD00001006304
<b>Software and algorithms</b>		
Code used in this study <sup>63</sup>	Zenodo	<a href="https://doi.org/10.5281/zenodo.7938169">https://doi.org/10.5281/zenodo.7938169</a>
R v3.6.1	<a href="https://www.r-project.org/">https://www.r-project.org/</a>	<a href="http://cran.r-project.org/src/base/R-3/R-3.6.1.tar.gz">http://cran.r-project.org/src/base/R-3/R-3.6.1.tar.gz</a>
cutadapt v1.16	Martin <sup>64</sup>	<a href="https://github.com/marcelm/cutadapt">https://github.com/marcelm/cutadapt</a>
python v3.6.1	<a href="https://www.python.org/">https://www.python.org/</a>	<a href="https://www.python.org/downloads/release/python-361/">https://www.python.org/downloads/release/python-361/</a>
TopHat2 v2.0.12	Kim et al. <sup>65</sup>	<a href="https://ccb.jhu.edu/software/tophat/index.shtml">https://ccb.jhu.edu/software/tophat/index.shtml</a>
Ensembl v88	Cunningham et al. <sup>66</sup>	<a href="https://ftp.ensembl.org/pub/release-88/gtf/homo_sapiens/Homo_sapiens.GRCh38.88.gtf.gz">https://ftp.ensembl.org/pub/release-88/gtf/homo_sapiens/Homo_sapiens.GRCh38.88.gtf.gz</a>
samtools (v1.2-24-g016c62b)	Danecek et al. <sup>67</sup>	<a href="https://github.com/samtools/samtools">https://github.com/samtools/samtools</a>
collapse_annotation.py	GitHub	<a href="https://github.com/broadinstitute/gtex-pipeline/blob/master/gene_model/collapse_annotation.py">https://github.com/broadinstitute/gtex-pipeline/blob/master/gene_model/collapse_annotation.py</a>
bwa v0.7.17-r1188	Li and Durbin <sup>68</sup>	<a href="https://github.com/lh3/bwa/releases/tag/v0.7.17">https://github.com/lh3/bwa/releases/tag/v0.7.17</a>

(Continued on next page)

<b>Continued</b>		
REAGENT or RESOURCE	SOURCE	IDENTIFIER
featureCounts v1.5.1	Liao et al. <sup>69</sup>	<a href="https://sourceforge.net/projects/subread/files/subread-1.5.1/">https://sourceforge.net/projects/subread/files/subread-1.5.1/</a>
DESeq2 v.1.26	Anders et al. <sup>18</sup> and Love et al. <sup>19</sup>	<a href="https://bioconductor.org/packages/3.10/bioc/html/DESeq2.html">https://bioconductor.org/packages/3.10/bioc/html/DESeq2.html</a>
edgeR v.3.28.1	Robinson and Oshlack <sup>20</sup>	<a href="https://bioconductor.org/packages/3.10/bioc/html/edgeR.html">https://bioconductor.org/packages/3.10/bioc/html/edgeR.html</a>
bedtools v.2.29	Quinlan and Hall <sup>70</sup>	<a href="https://github.com/arq5x/bedtools2/releases/tag/v2.29.0">https://github.com/arq5x/bedtools2/releases/tag/v2.29.0</a>
g:Profiler e103_eg50_p15_68c0e33	Raudvere et al. <sup>71</sup>	<a href="https://biit.cs.ut.ee/gprofiler">https://biit.cs.ut.ee/gprofiler</a>
gprofiler2 v0.19	Kolberg et al. <sup>72</sup>	<a href="https://cran.r-project.org/src/contrib/Archive/gprofiler2/gprofiler2_0.1.9.tar.gz">https://cran.r-project.org/src/contrib/Archive/gprofiler2/gprofiler2_0.1.9.tar.gz</a>
Rtsne v0.15	Krijthe <sup>73</sup>	<a href="https://cran.r-project.org/src/contrib/Archive/Rtsne/Rtsne_0.15.tar.gz">https://cran.r-project.org/src/contrib/Archive/Rtsne/Rtsne_0.15.tar.gz</a>
MitoMiner v4	Smith and Robinson <sup>32</sup>	<a href="http://mitominer.mrc-mbu.cam.ac.uk/release-4.0/mitocarta.do">http://mitominer.mrc-mbu.cam.ac.uk/release-4.0/mitocarta.do</a>
BioRender (for the Graphical Abstract)	N/A	<a href="http://www.biorender.com">www.biorender.com</a>
<b>Other</b>		
Early gestation placenta RNA-Seq data	<a href="https://www.ebi.ac.uk/ena">https://www.ebi.ac.uk/ena</a>	PRJEB38810
Term placenta WGS data	<a href="https://ega-archive.org/">https://ega-archive.org/</a>	EGAD00001004198
GTEX gene-level count v8.p2	<a href="https://gtexportal.org/">https://gtexportal.org/</a>	<a href="https://storage.googleapis.com/gtex_analysis_v8/rna_seq_data/GTEX_Analysis_2017-06-05_v8_RNASeQCv1.1.9_gene_reads.gct.gz">https://storage.googleapis.com/gtex_analysis_v8/rna_seq_data/GTEX_Analysis_2017-06-05_v8_RNASeQCv1.1.9_gene_reads.gct.gz</a>
GTEX gene model v8	<a href="https://github.com/broadinstitute/gtex-pipeline/releases/tag/gtex_v8">https://github.com/broadinstitute/gtex-pipeline/releases/tag/gtex_v8</a>	<a href="https://github.com/broadinstitute/gtex-pipeline/tree/master/gene_model">https://github.com/broadinstitute/gtex-pipeline/tree/master/gene_model</a>
Cancer Genome Atlas Pan-Cancer Analysis of Whole Genomes (PCAWG)	Yuan et al. <sup>37</sup>	<a href="https://ibl.mdanderson.org/tcma/download/TCMA-CopyNumber.tsv.zip">https://ibl.mdanderson.org/tcma/download/TCMA-CopyNumber.tsv.zip</a>
WGS datasets of healthy human tissues	N/A	<a href="https://ega-archive.org/">https://ega-archive.org/</a>
blood	Lee-Six et al. <sup>34</sup>	EGAD00001004086
colon	Lee-Six et al. <sup>35</sup>	EGAD00001004192
endometrium	Moore et al. <sup>33</sup>	EGAD00001004547
liver	Brunner et al. <sup>36</sup>	EGAD00001004578
External placental RNA-seq datasets	N/A	<a href="https://www.ebi.ac.uk/ena">https://www.ebi.ac.uk/ena</a>
Lim	Lim et al. <sup>21</sup>	PRJNA386110
Huang <sup>22</sup>	Huang et al. <sup>22</sup>	PRJNA499121
Ashley	Ashley et al. <sup>24</sup>	PRJNA704615
Awamleh	Awamleh et al. <sup>25</sup>	PRJNA472249
Verheecke	Verheecke et al. <sup>23</sup>	Data obtained directly from the authors

## RESOURCE AVAILABILITY

### Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, D. Stephen Charnock-Jones ([dscj1@cam.ac.uk](mailto:dscj1@cam.ac.uk)).

### Materials availability

This study did not generate new unique reagents.

### Data and code availability

- The term placenta RNA-seq data have been deposited in the European Genome-phenome Archive (EGA, <https://ega-archive.org/>) with the following accession number: EGAD00001006304. This paper analyzes existing, publicly available data. These accession numbers for the datasets are listed in the [key resources table](#).

- All original code and processed datasets have been deposited at Zenodo (<https://doi.org/10.5281/zenodo.7938169>) and they are publicly available from the gitlab codebase (<https://gitlab.com/sunggong/pops-placenta-void-2023>).
- Any additional information required to reanalyze the data reported in this paper is available from the [lead contact](#) upon request.

## EXPERIMENTAL MODEL AND SUBJECT DETAILS

### Placenta samples

All the full-term placenta samples were obtained from the Pregnancy Outcome Prediction (POP) study, a prospective cohort study of nulliparous women attending the Rosie Hospital, Cambridge (UK) for their dating ultrasound scan between January 14, 2008, and July 31, 2012. The study has been previously described in detail.<sup>15,16</sup> Ethical approval for the study was given by the Cambridgeshire 2 Research Ethics Committee (reference number 07/H0308/163) and all participants provided written informed consent. A total of 128 unique placental samples were analyzed in this study - 60 samples were used for RNA-Seq (Table S1) and 80 for WGS (Table S12). One of the RNA-Seq samples was dropped from the analysis due to the presence of decidual contamination<sup>4</sup> and 12 placentas were analyzed using both methods.

First and second trimester tissue samples were collected with informed written patient consent and approval of the Joint University College London/University College London Hospital Committees on the Ethics of Human Research (05/Q0505/82) from 7-8 wGA (n=8) and 13-14 wGA (n=6) uncomplicated pregnancies. Gestational age was confirmed by ultrasound measurement of the crown-rump length of the embryo. All samples were collected from patients undergoing surgical pregnancy termination under general anesthesia for psycho-social reasons. Villous samples were obtained under transabdominal ultrasound guidance from the central region of the placenta using a chorionic villus sampling (CVS) technique. All samples were snap-frozen immediately in liquid nitrogen and stored at  $-80^{\circ}\text{C}$  until analysis. These samples have previously been described in full.<sup>17</sup>

### Tissue collection for RT-qPCR analysis

Placental tissues for qPCR validation were collected from healthy women with normal term pregnancies and scheduled for delivery by elective cesarean section. Participants were consented for research sample collection as part of the surgical procedure, with further permission for storage and transfer of materials to the biobank given under approval 07/MRE05/44. Analysis was performed as part of the Cambridge Blood and Stem Cell Biobank REC ID 18/EE/0199. Human aorta, lung and heart (left ventricle) used in the research study was obtained from the Papworth Hospital Research Tissue Bank. Written consent was obtained for all tissue samples using Papworth Hospital Research Tissue Bank's ethical approval (East of England - Cambridge East Research Ethics Committee) under approval 18/EE/0269. Human colon, stomach and small intestine were obtained from Cambridge University Hospitals Human Research Tissue Bank under approval 04/Q1604/21 and breast tissues from the Institute of Metabolic Sciences.

### Study approval

The full-term placenta samples used for RNAseq were obtained from the Pregnancy Outcome Prediction (POP) study. Ethical approval for the study was given by the Cambridgeshire 2 Research Ethics Committee (reference number 07/H0308/163). First and second trimester tissue samples were collected with approval of the Joint University College London/University College London Hospital Committees on the Ethics of Human Research (05/Q0505/82). Placental tissue used for PCR validation was collected with further permission for storage and transfer of materials to the biobank given under approval 07/MRE05/44 and analysis was performed as part of the Cambridge Blood and Stem Cell Biobank REC ID 18/EE/0199. Human aorta, lung and left ventricle used in the study was obtained from the Papworth Hospital Research Tissue Bank with ethical approval (East of England - Cambridge East Research Ethics Committee) under approval 18/EE/0269. Human colon, stomach and small bowel were obtained from Cambridge University Hospitals Human Research Tissue Bank under approval 04/Q1604/21. All participants provided written informed consent.

## METHOD DETAILS

### RNA sequencing and data processing

The POP study placental biopsies were collected within 30 minutes of delivery and flash frozen in RNeasy Lysis Buffer (Qiagen). For each biopsy, total placental RNA was extracted from approximately 5 mg of tissue using the "mirVana miRNA Isolation Kit" (ThermoFisher) which efficiently isolates all RNAs longer than 10 nucleotides in length, followed by DNase treatment ("DNA-free DNA Removal Kit", ThermoFisher). RNA quality was assessed with the Agilent Bioanalyzer and all the samples with RIN values  $\geq 7.0$  were used in the downstream experiments. RNA-libraries were prepared from 1  $\mu\text{g}$  of total placental RNA with the TruSeq Stranded mRNA Library Prep Kit (Illumina) which captures polyA-tailed transcripts by oligo-dT beads, then pooled and sequenced (single-end, 50bp) using a Single End V4 cluster kit and Illumina HiSeq2500. RNA was also extracted from human first and second trimester placental villi using the RNeasy Plus Universal Mini Kit (Qiagen). Libraries were made using the Illumina TruSeq Stranded mRNA Library Kit according to the manufacturer's instructions.

The adaptor sequences and poor-quality bases were trimmed using *cutadapt* (v1.16)<sup>64</sup> (with python v3.6.1) with the following command:

```
cutadapt -j 32 -a AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC -q 20 -O 8 -m 20 -o $TRIMMED_FASTQ $INPUT_FASTQ
```

The quality-assured trimmed short reads were mapped to the GRCh38 version of human genome reference using *TopHat2* (v2.0.12)<sup>65</sup>:

```
tophat2 -p 32 --library-type fr-firststrand --output-dir $OUTPUT_DIR --max-multihits 10 --pre-filter-multihits --transcriptome-index=$TR_INDEX $BOWTIE2INDEX $TRIMMED_FASTQ
```

The transcriptome index above was built using transcript annotation from Ensembl v88<sup>66</sup> (equivalent to Gencode v26). We applied so-called two-pass (or two-scan) alignment protocol to rescue unmapped reads from the initial mapping by re-aligning unmapped reads toward the exon-intron junctions detected in the first-mapping:

```
tophat2 -p 32 --library-type fr-firststrand --output-dir $OUTPUT_DIR --raw-juncs $MERGED_JUNCTION $UNMAPPED_FASTQ
```

For each sample, the initial and second mapped reads were merged by *samtools* (v1.2-24-g016c62b)<sup>67</sup>:

```
samtools merge $MERGED_BAM $FIST_MAP_BAM $SECOND_MAP_BAM
```

Before gene-level quantification of read counts, we pre-processed the transcript annotation file (Gencode v26) using the '*collapse\_annotation.py*' python script available from the following GTEx github site: [https://github.com/broadinstitute/gtex-pipeline/tree/master/gene\\_model](https://github.com/broadinstitute/gtex-pipeline/tree/master/gene_model).

```
python3 collapse_annotation.py $GENCODE_26_GTF $PROCESSED_GENCODE_26_GTF
```

Finally, we quantified sequencing reads at the gene-level using *featureCounts*<sup>69</sup> tool of *subread* package (v1.5.1):

```
featureCounts -T 32 -a $PROCESSED_GENCODE_26_GTF -Q 10 -s 2 -p -C -o $GENE_COUNT_OUTPUT $MERGED_BAM
```

### Whole-genome sequencing and data processing

The whole genome sequencing dataset of the placenta (n=80) was from 'cohort1' (babies delivered by pre-labor Caesarean section) described in our previous report,<sup>74</sup> which was also based on the POP study. Detailed description of the experimental protocol is available in the original paper.

The sequencing files were converted from CRAM format to FASTQ using *samtools* (v1.7-15-g9ce8c64):

```
samtools fastq -F 0x200 $INPUT_CRAM -1 $FASTQ_R1 -2 $FASTQ_R2
```

The adaptor sequences and poor-quality bases were trimmed using *cutadapt* v1.16 (with python v3.6.1) with the following command:

```
cutadapt -j 32 -a AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC -A AGATCGGAAGAGCGTCGTGTAGGAAAGAGTGTAG ATCTCGGTGGTCGCCGATCATT -q 20 -O 8 -m 20 -o $TRIMMED_R1 -p $TRIMMED_R2 $FASTQ_R1 $FASTQ_R2
```

The quality-assured trimmed short reads were mapped to the GRCh38 version of human genome reference using *bwa* (v0.7.17-r1188)<sup>68</sup>:

```
bwa mem -M -t 32 \
-R
"@RG\tID:$ID\tPL:illumina\tPU:run\tLB:$ID\tSM:$Barcode\tCN:CamObsGynae" \
$GRCh38_GENOME_FASTA $TRIMMED_R1 $TRIMMED_R2 \
| samtools view -Sb - > $BAM_FILE
```

The whole genome sequencing dataset of the 1,682 healthy normal tissues (the endometrium, blood, colon, and liver) was generated at the Wellcome Trust Sanger Institute.<sup>33</sup>

### GTEx data processing

We compared our placenta RNA-Seq datasets with 46 somatic tissues from GTEx (v8.p2).<sup>9</sup> To select eligible samples from GTEx RNA-Seq datasets, we used the same filtering conditions applied to our previous study<sup>4</sup>: (a) RNA integrity number (SMRIN)  $\geq 6$ , (b) mapping rate (SMMAPRT)  $\geq 0.9$ , (c) exonic mapping rate (SMEXNCRT)  $\geq 0.75$ , and (d)  $\geq 20$  qualifying samples per tissue. Five tissues (out of 54) were dropped after applying the aforementioned filters: the kidney (medulla), the fallopian tube, the cervix (endocervix), the cervix (ectocervix), and the bladder. We further removed the following three non-solid tissues: the whole blood, cultured fibroblast cells, and EBV-transformed lymphocytes cells. Finally, a total of 11,803 samples were selected from 46 somatic tissues. Table S2 shows the number of samples across the 46 tissues we considered. In our initial analysis, we used 4,454 samples from 20 somatic tissues from GTEx with the following modified criteria: (a) the RNA integrity number (SMRIN)  $\geq 6$ , (b) mapping rate to genome (SMMAPRT)  $> 0.8$ , (c) mapping rate to exon (SMEXNCRT)  $> 0.8$ , (d)  $\geq 10$  qualifying samples of both sexes (i.e. at least 20 samples per tissue), and (e) manual selection of tissue sub-types if two or more were available for the same tissue. We considered 56,156 genes from the gene-level quantification information available from the following file: GTEx\_Analysis\_2017-06-05\_v8\_RNASeQCv1.1.9\_gene\_reads.gct.gz.

### Identification of absent or depleted protein-coding transcripts

We made a count matrix of 56,156 genes by 11,876 samples (i.e. 11,803 samples from GTEx and 73 from the placenta of the following gestational ages: 8 at 8wk, 6 at 14wk and 59 at term), then filtered out genes of the following conditions: (1) the sum of read count across samples is zero ( $n=238$ ), (2) non-polyadenylated RNAs (e.g. transcripts of major histones) reported from the study of Yang et al.<sup>75</sup> ( $n=90$ ), and (3) transcripts which are not annotated as ‘protein-coding’ as per Ensembl v88 ( $n=36,658$ ). After filtering, a total of 19,170 protein-coding genes were considered. To adjust differences in the composition of RNA populations across multiple tissues, we applied the following two normalization methods to the count matrix (a dimension of 19,170 x 11,876): (1) the median ratio method implemented in the ‘*estimateSizeFactors*’ function of DESeq2 (v.1.26) Bioconductor package,<sup>18,19</sup> and (2) the trimmed mean of M-values (TMM), available from ‘*calcNormFactors*’ function of edgeR (v.3.28.1) Bioconductor package.<sup>20</sup> Then, we built two matrices of normalized count per million (nCPM), each of which has a dimension of 19,170 x 11,876, using the ‘*fpm*’ and ‘*cpm*’ functions of DESeq2 and edgeR package, respectively. The columns (i.e. 11,876 samples) of the matrix were reduced to a size of 47 columns (i.e. 46 tissues from GTEx and 1 placental tissue either from early gestation or full term) by taking the mean of nCPM across samples of the same tissue. Using the placenta at term, we identified 5,632 and 5,727 genes for which the placenta was ranked 47 (i.e. bottom) based on DESeq2 and TMM normalization methods, respectively. Finally, we selected 762 of them (Table S3) which satisfied the following three conditions: (1)  $nCPM > 1$  for the tissue ranked 46, (2)  $nCPM_{(rank=46)}/nCPM_{(rank=47)} > 3$ , and (3) fulfilling aforementioned conditions both in DESeq2 and TMM normalization methods. In the down-sampling analysis, we applied binomial sampling, using ‘*rbinom*’ in R, to each column of the count matrix with the subsampling probability being 20 million divided by the sum of each column – a similar approach was previously introduced.<sup>76</sup> This is equivalent to randomly choosing each individual mapped read with the same subsampling probability, so that the final number of down-sampled reads becomes 20 million per sample, which is the minimum number of sequencing reads in our placenta RNA-Seq dataset (Table S1). Then, we applied the same approach of finding absent or depleted transcripts, as described earlier, to the down-sampled count matrix.

### RNA extraction for RT-qPCR validation

Approximately 35 mg of frozen tissues were homogenized by bead beating for 20 s at a speed of 4.5 ms<sup>-1</sup> on a FastPrep24 sample disruption system with Lysing Matrix S tubes (MP Biomedicals, Santa Ana, CA). Total RNA was isolated with the RNeasy Plus Mini Kit (Qiagen) and 200 ng of total RNA from each sample was reverse transcribed using the High-capacity RNA-to-cDNA kit (ThermoFisher Scientific). The qPCR reactions were prepared using TaqMan Multiplex Master Mix (ThermoFisher Scientific).

### Target selection for RT-qPCR validation

We selected a total of 13 transcripts to measure their abundance levels using qPCR. Eleven targets (NFIC, SNED1, GGT5, THBS2, C16orf45, ITGA7, WASF3, IGFBP6, RCAN2, SLCO3A1, and PPARGC1A) were selected by comparing our placenta RNA-Seq datasets to 20 non-placental tissues from GTEx (i.e. an initial analysis) based on the following criteria: 1) depleted in the placenta across three gestational ages (as described above), 2) fold-change (i.e.  $nCPM_{(rank=20)}/nCPM_{(rank=21; term-placenta)} > 5$ , and 3) top 10 by the effect size (i.e.  $nCPM_{(rank=20)} - nCPM_{(rank=21; term-placenta)}$ ). PGC-1 $\alpha$  (PPARGC1A) while it was depleted it was not within the top 10, but it was included considering its important role in regulating mitochondria. Two targets (KMO and ARMS2) were selected by comparing our placenta RNA-Seq datasets to 46 non-placental tissues from GTEx based on the following criteria: 1) annotated in MitoMiner,<sup>32</sup> 2) highest nCPM in the placenta across three gestational ages (i.e. ranked within top 3), and 3)  $nCPM_{(rank=4)} > 1$ . There were 9 transcripts (KMO, ARMS2, LYPLA1, RDH13, RPIA, SLC25A35, ACSS1, NME6, and TEFM) satisfying these conditions and we selected top two by the average fold change of the term placenta compared to 46 non-placental tissues (i.e.  $nCPM_{(term-placenta)}/nCPM_{(non-placenta)}$ ). The following 13 predesigned TaqMan assays were used: NFIC (Hs00232157\_m1), SNED1 (Hs00966449\_m1), GGT5 (Hs00897715\_m1), THBS2 (Hs01568063\_m1), C16orf45 (Hs01014981\_m1), ITGA7 (Hs01056475\_m1), WASF3 (Hs00903488\_m1), IGFBP6 (Hs00181853\_m1), RCAN2 (Hs00195165\_m1), SLCO3A1 (Hs00203184\_m1), PPARGC1A (Hs00173304\_m1), ARMS2 (Hs01394203\_m1), and KMO (Hs00175738\_m1). The above target genes were normalized to the geometric mean of CDC34 (Hs00362082\_m1) and TBP (Hs00427620\_m1).

### Calculation of mitochondrial copy number

The mitochondrial copy number ( $Copy_{mt}$ ) was calculated as the ratio of mitochondrial depth of coverage ( $Cov_{mt}$ ) over the average genome depth of coverage ( $Cov_g$ ). It is formally defined as following:

$$Copy_{mt} = \frac{Cov_{mt}}{Cov_g}, \text{ where } Cov_{mt} = \frac{N_{mt} * L_{mt}}{MT} \text{ and } Cov_g = \frac{N_g * L_g}{G}$$

The depth of coverage ( $Cov$ ) above is defined as the number of mapped bases, which is the number of reads ( $N$ ) multiplied with the length of the mapped reads ( $L$ ), divided by the haploid size of genome ( $G$ ) or mitochondrion ( $MT$ ). We calculated the depth of coverage from the BAM files of whole genome sequencing datasets using *bedtool* (v.2.29)<sup>70</sup> with the following command:

```
bedtools genomecov -ibam $BAM_FILE > ${BAM_FILE%.bam}.cov.txt.
```

### Gene Ontology analysis

The gene ontology analysis was performed using g:Profiler<sup>71</sup> (<https://biit.cs.ut.ee/gprofiler>; version e103\_eg50\_p15\_68c0e33) with FDR multiple-testing correction method applying significance threshold of 0.05. We used the *gprofiler2* R package<sup>72</sup> (v.0.19; <https://cran.r-project.org/web/packages/gprofiler2/>), a R client for the g:Profiler tools, with the 19,170 protein-coding genes (see above) as a list of background genes.

### Dimension reduction

We used *Rtsne* (<https://cran.r-project.org/web/packages/Rtsne>; v0.15)<sup>73</sup> with a default option for the main dimension reduction method as shown in Figure 3C.

### External datasets used in this study

#### *mtDNA copy numbers in cancer*

we downloaded mtDNA copy numbers of the Cancer Genome Atlas Pan-Cancer Analysis of Whole Genomes (PCAWG) Consortium<sup>37</sup> from the following URL: <https://ibl.mdanderson.org/tcma/download/TCMA-CopyNumber.tsv.zip>. The dataset was downloaded as of the 9<sup>th</sup> of March 2020.

#### *WGS datasets of healthy human tissues*

we used WGS datasets of healthy human tissues from the European Genome-phenome Archive (EGA, <https://ega-archive.org/>) with the following accession numbers: EGAD00001004086 (blood), EGAD00001004192 (colon), EGAD00001004547 (endometrium), and EGAD00001004578 (liver).

#### *Placental RNA-seq datasets*

we downloaded four placenta RNA-Seq datasets from the European Nucleotide Archive (ENA, <https://www.ebi.ac.uk/ena>) with the following accession numbers: PRJNA386110 (Lim), PRJNA499121 (Huang), PRJNA704615 (Ashley), and PRJNA472249 (Awamleh). The RNA-Seq dataset from the Verheeecke's study was obtained personally from one of the authors.

### Identification of transcripts localized in the mitochondria

We downloaded MitoMiner (v4),<sup>32</sup> a dataset of mitochondrial localization, from the following URL: <http://mitominer.mrc-mbu.cam.ac.uk/release-4.0/mitocarta.do>. We selected genes of the following conditions: (1) not encoded in mitochondrial chromosome, (2) "Known mitochondrial" or "Predicted mitochondrial" as types of evidence, and (3) one of the 19,156 eligible protein-coding genes described above. After filtering, we considered 1,042 protein-coding genes.

### QUANTIFICATION AND STATISTICAL ANALYSIS

All statistical analyses were performed using R v3.6.1. Details of statistical analysis for each study are included in the respective figure legend. All N-values in figure legends were the number of independent biological replicates. All tests were two-sided and  $P < 0.05$  was defined as statistically significant.