The human placenta exhibits a unique transcriptomic void

Sungsam Gong[1,2]
Francesca Gaccioli[1,2]
Irving L.M.H. Aye[1,2]
Giulia Avellino[1,2]
Emma Cook[1]
Andrew R.J. Lawson[3]
Luke M.R. Harvey[3]
Gordon C.S. Smith[1,2,4]
D. Stephen Charnock-Jones[1,2,4*]

[1]Department of Obstetrics and Gynaecology, University of Cambridge, NIHR Cambridge Biomedical Research Centre, Cambridge, United Kingdom
[2]Centre for Trophoblast Research (CTR), Department of Physiology, Development and Neuroscience, University of Cambridge, Cambridge, United Kingdom
[3]Wellcome Trust Sanger Institute, Hinxton, United Kingdom
[4]These authors contributed equally
*Corresponding author: Professor D. Stephen Charnock-Jones, Department of Obstetrics and Gynaecology, University of Cambridge, Box 223 The Rosie Hospital, Cambridge, CB2 0SW, UK. Tel: +44 (0)1223 336875, Fax: +44 (0)1223 215327, e-mail: dscj1@cam.ac.uk

Abstract

We have recently demonstrated that the human placenta exhibits a unique genomic architecture with an unexpectedly high mutation burden(Coorens et al. 2021) and it is also well recognized that the placenta uniquely expresses many genes(Gong et al. 2021). However, the placenta is relatively understudied in systematic comparisons of gene expression in different organs. The aim of the present study was to identify transcripts which were uniquely absent or depleted, comparing the placenta with 46 other human organs. Here we show that 40/46 of the other organs had no transcripts which were selectively depleted and that of the remaining six, the liver had the largest number with 26. In contrast, the term placenta had 762 depleted transcripts. Gene Ontology analysis of this depleted set highlighted multiple pathways reflecting known unique elements of placental physiology. However, analysis of term samples demonstrated massive over representation of genes involved in mitochondrial function (P=5.8x10^{-10}), including PGC-1α - the master regulator of mitochondrial biogenesis, and genes involved in polyamine metabolism (P=2.1x10^{-4}). We conclude that the term placenta exhibits a unique metabolic environment.

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”(Brosens et al. 2011) and which account for a substantial burden of global morbidity and mortality. Progress on predicting and preventing these complications is hampered by lack of mechanistic understanding of normal and abnormal placental function and we and others have applied multiple studies using omic 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(Armstrong et al. 2017) and across gestation(Buckberry et al. 2017) but there are fewer studies comparing the placental transcriptome with the transcriptomes of other organs(Kim et al. 2012; Gong et al. 2021). RNA-Seq enables transcriptome profiling of tissues or single cells and there are a number of studies characterizing so-called the 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)(ENCODE Project Consortium

2012), the RoadMap Epigenomics Project(Roadmap Epigenomics Consortium et al. 2015), and the Functional Annotation of Mammalian Genome (FANTOM5) project(de Rie et al. 2017). However, the human placenta transcriptome is relatively understudied and absent from large-scale "omic" analyses such as the Genotype-Tissue Expression (GTEx) project(GTEx Consortium 2020).

Pan-tissue comparative analyses generally focus on identifying transcripts that are abundant in a tissue of interest while being absent or depleted in others. Indeed, there are a number of tools and databases that enable "tissue-specific" gene enrichment analysis(Jain and Tuteja 2019; Watanabe et al. 2019; Papatheodorou et al. 2020). Studying "tissue-specific" genes provides information about specific functions that define a unique set of characteristics or "identity" of a tissue of interest. Transcripts that are ubiquitously expressed in multiple tissues, such as house-keeping genes(Eisenberg and Levanon 2013) can be identified and this gives insight to the functions that all tissue share. In contrast, little attention has been paid to the identification of transcripts that are less abundant, or even absent, in one tissue compared to 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 which reflect known aspects of placental physiology, such as lack of nervous tissue and unique immunological features. However, these analyses also generated evidence 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

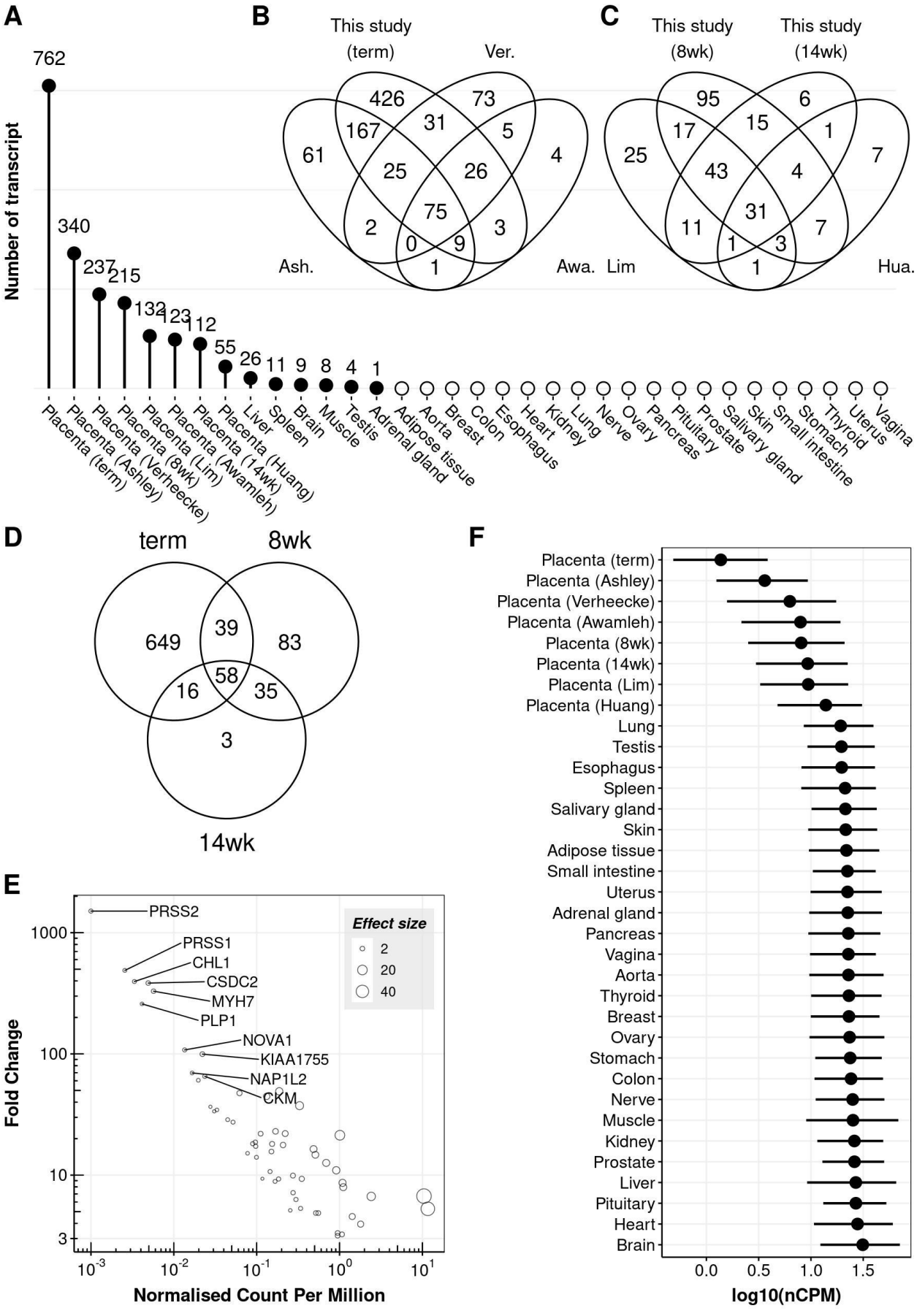
We carried out mRNA sequencing (RNA-seq) using 59 human term placentas from the POP study cohort(Pasupathy et al. 2008; Gaccioli et al. 2016; Gong et al. 2018b) and 14 human placentas from earlier in gestation (n=8, 7-8 weeks (8wk); n=6, 13-14 weeks (14wk))(Prater et al. 2021). We obtained approximately 38 million reads from each sample (**Supplementary Table 1**). We compared the placental transcriptome

profile at 8wk, 14wk and term with that of 46 tissues from 11,803 samples of GTEx Consortium datasets(GTEx Consortium 2020) and investigated which transcripts are absent or depleted in the placenta while being reasonably abundant in other tissues (**Supplementary Table 2**). To adjust for differences in the RNA composition across tissues, we applied the following two normalization methods: 1) the median ratio method (DESeq(Anders and Huber 2010)) and 2) the trimmed mean of M-values (TMM(Robinson and Oshlack 2010)) (see 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 DESeq and TMM normalization methods, respectively. Then we selected the 762 transcripts which satisfied the following three conditions: (1) nCPM >1 for the tissue of rank 46, (2) $nCPM_{(rank=46)}/nCPM_{(rank=47)} > 3$, and (3) fulfilling these conditions both in DESeq and TMM normalization methods (**Supplementary Table 3**) – these transcripts are described as being "depleted" hereafter. For early gestational age datasets (8wk and 14wk placenta) we identified 215 and 112 such depleted transcripts respectively (**Supplementary Tables 4 and 5**).

Using the same criteria we applied to the placenta, we sought to identify mRNAs depleted in each of the 46 other tissues. 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** and **Supplementary Table 6**). We then used five external placenta RNA-Seq datasets generated independently (two early gestational placenta datasets: (1) Lim et al.(Lim et al. 2017) (n=4) and (2) Huang et al.(Huang et al. 2018) (n=3) and three term placenta datasets: (3) Verheecke et al.(Verheecke et al. 2018) (n=66), (4) Ashley et al.(Ashley et al. 2021) (n=4), and (5) Awamleh et al.(Awamleh et al. 2019) (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** and **Supplementary Table 7**). All the datasets described above had variable depth of coverage. So, to investigate any possible effect of sequencing depth, we down-sampled the reads to 20 million for all samples. We then

repeated the analysis with the same methods and criteria as described above. In down-sampled datasets, we obtained consistent results that the placenta had many more depleted transcripts than other tissues (**Supplementary Figure 1** and **Supplementary Table 8**). 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 ($P<1x10^{-314}$ for the term placenta and $P=7.4x10^{-209}$ for the early placenta; both Fisher's exact test). In our term and early gestation placental datasets, there was a total of 883 transcripts depleted at any gestational age and among these, 58 were depleted in all three trimesters (**Supplementary Table 9** and **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 to 46 tissues from the GTEx dataset (see **Supplementary Table 9**).

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) whilst *MT-ND6* (mitochondrially encoded NADH:ubiquinone oxidoreductase core subunit 6) had the highest level of expression (nCPM=123), but was still three-fold lower than in all the other tissues studied (**Supplementary Table 3**). Interestingly, the 762 transcripts depleted in our term placenta were also less abundant in our early gestation placenta samples and other placenta data sets 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 majority of them were ranked bottom in the Ashely (335; 79%), Verheecke (245; 58%), and Awamleh (179; 42%) datasets (**Supplementary Figure 2**).



151

Figure 1. The number of absent or depleted transcripts in various tissues. A, The number of depleted transcripts are shown in the placenta samples (our and other studies) and other somatic tissues. The term, 8wk and 14wk 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 **Supplementary Tables 6 and 7**, respectively. Ver. (Verheecke), Ash. (Ashley), Awa. (Awamleh), Hua. (Huang). **E**, Abundances of the transcripts (x-axis) relatively depleted in all three trimesters are shown along with their fold change (y-axis, calculated as follows: $nCPM_{(rank=46)}/nCPM_{(rank=47;placenta)}$). The counts (per million) on x-axis are normalized by TMM method and the data is available in **Supplementary Table 9**. To avoid fold change being infinite values, a small number (0.001) was added to nCPM of the term placenta. The transcripts with the 10 highest fold change (*PRSS2* being the top one) are shown with their gene names. The size of circle represents the effect size (i.e. $nCPM_{(rank=46)}$ - $nCPM_{(rank=47;placenta)}$). **F**, The range of transcript abundances for the 762 genes in the placental tissues and 26 representative somatic tissues out of 46 we studied. Dot: median; line: interquartile range (IQR). For display, non-placental tissues shown in **F**, were manually selected if there are 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 sub-regions are shown in **Supplementary Table 2**.

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 increased as gestation progressed (**Supplementary Figure 3A**). A number of gene ontology (GO) terms were significantly enriched ($P<0.05$) in these 83 transcripts (T1 in **Figure 2B**), including MHC protein complex (*HLA-A*, *HLA-DPA1* and *HLA-DPB1*) and ion channel inhibitor activity (*ANKRD36*, *CAMK2D*, *LYNX1*, and *SCN1B*), suggesting a limited role for these functions in the first trimester.

We found the second trimester placenta sample had the smallest number of depleted transcripts (n=112) compared to 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 to 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 both TMM and DESeq normalization methods. For example, for *MYOM2* the fold changes based on TMM and DESeq were 2.9 and 2.8 fold respectively.

Among the 883 transcripts depleted in any of the three trimesters, 58 are depleted in all three (**Supplementary Table 9** and T1+T2+T3 in **Figure 2A**) and they are associated with various GO terms (T1+T2+T3 in **Figure 2B**). It is unsurprising that genes annotated with the GO terms "astrocyte development, "gliogenesis" and "skeletal muscle adaptation" are absent in the placenta. However, as 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 as these reductases inactivate steroid hormones(Penning et al. 2015), specifically progesterone in the case of *AKR1C1*. Peroxisome Proliferator Activated Receptor Gamma (PPARG) is abundant and is essential for placental development(Valle et al. 2005) and function. However, the depletion of PPARG Coactivator 1 Alpha (*PPARGC1A)* transcripts (also known as PGC-1α) suggests that the usual coordination between PPARG and PGC-1α does not occur in the placenta(Hondares et al. 2006). Of note, PGC-1α is also directly implicated in regulating mitochondrial biogenesis(Wu et al. 1999) and the regulation of mitochondrial genes (see below).

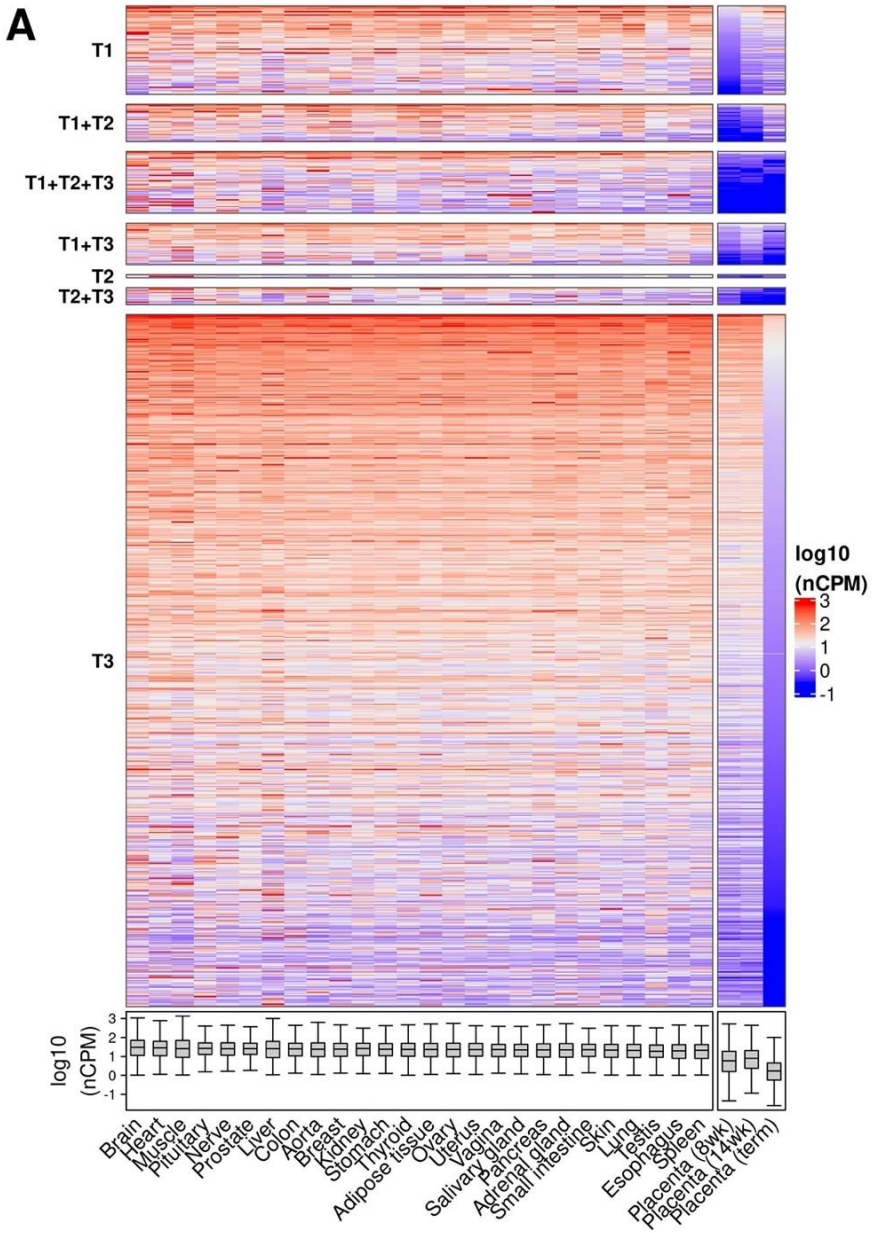
Several depleted transcripts are annotated with the GO term "extracellular matrix organization". Trypsins 1 and 2 (PRSS1 and PRSS2) are notable as these were essentially absent from the placenta. As these proteins are key activators of multiple matrix metalloproteases this suggests that initiation of matrix remodeling is mediated by other proteases. Transcripts encoding two type IV collagen genes (*COL4A3*, *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(Hudson et al. 2003). However, the lack of the α3. α4. α5(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 α1. α1. α2(IV) collagen

protomer is sufficient and in fact the placenta has the highest expression of COL4A1 among the GTEx tissues from our previous study(Gong et al. 2021). Keratin filament transcripts (*KRT4*, *KRT5*, *KRT13*) are also depleted and annotated with the GO term "extracellular matrix organization". These keratins are characteristic of stratified epithelial surfaces(Moll et al. 2008) (such as the esophagus in which the expression level is >10,000 times higher) and this difference likely reflects the syncytial nature of the trophoblast epithelial surface.

We identified 762 depleted transcripts in the term placenta and this was the highest number among three trimesters, and 649 of them (85%) were uniquely depleted at term (T3 in **Figure 2A**). GO analysis showed these genes are predominantly associated with mitochondria-specific processes, suggesting that the term placenta has diminished capacity for these functions (T3 in **Figure 2B**). They include genes encoding 12 mitochondrial ribosomal proteins, ATP synthase, H^+ transporting, mitochondrial F1 complex, delta subunit (*ATP5D*), succinate dehydrogenase complex assembly factor 1 (*SDHAF1*), NADH:ubiquinone oxidoreductase complex assembly factors (*NDUFAF3, NDUFAF8*) and subunits (*NDUFB7, NDUFS7, 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 (**Supplementary Figure 3B**). 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 over-represented (**Supplementary Text and Supplementary Figure 4**).

Having observed significantly over-represented mitochondria-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 8wk placenta (4.6%) (**Supplementary Figure 5A**). In contrast, the heart (left ventricle; 39.7%), the kidney (cortex; 31.3%) and the liver (21.1%) expressed the most mitochondrial protein-coding transcripts. We also examined the extent of mitochondrial transcripts including both the protein-coding and non-coding transcripts, such as mitochondrial

256 rRNA and tRNA, and found the term placenta also showed the lowest proportion of
257 reads mapped to mitochondrial DNA (3.7%, **Supplementary Figure 5B**).


A
T1
T1+T2
T1+T2+T3
T1+T3
T2
T2+T3
T3
log10
(nCPM)
3
2
1
0
-1
log10
(nCPM)
3
2
1
0
-1
Brain
Heart
Muscle
Pituitary
Nerve
Prostate
Liver
Colon
Aorta
Breast
Kidney
Stomach
Thyroid
Adipose tissue
Ovary
Uterus
Vagina
Salivary gland
Pancreas
Adrenal gland
Small intestine
Skin
Lung
Testis
Esophagus
Spleen
Placenta (8wk)
Placenta (14wk)
Placenta (term)

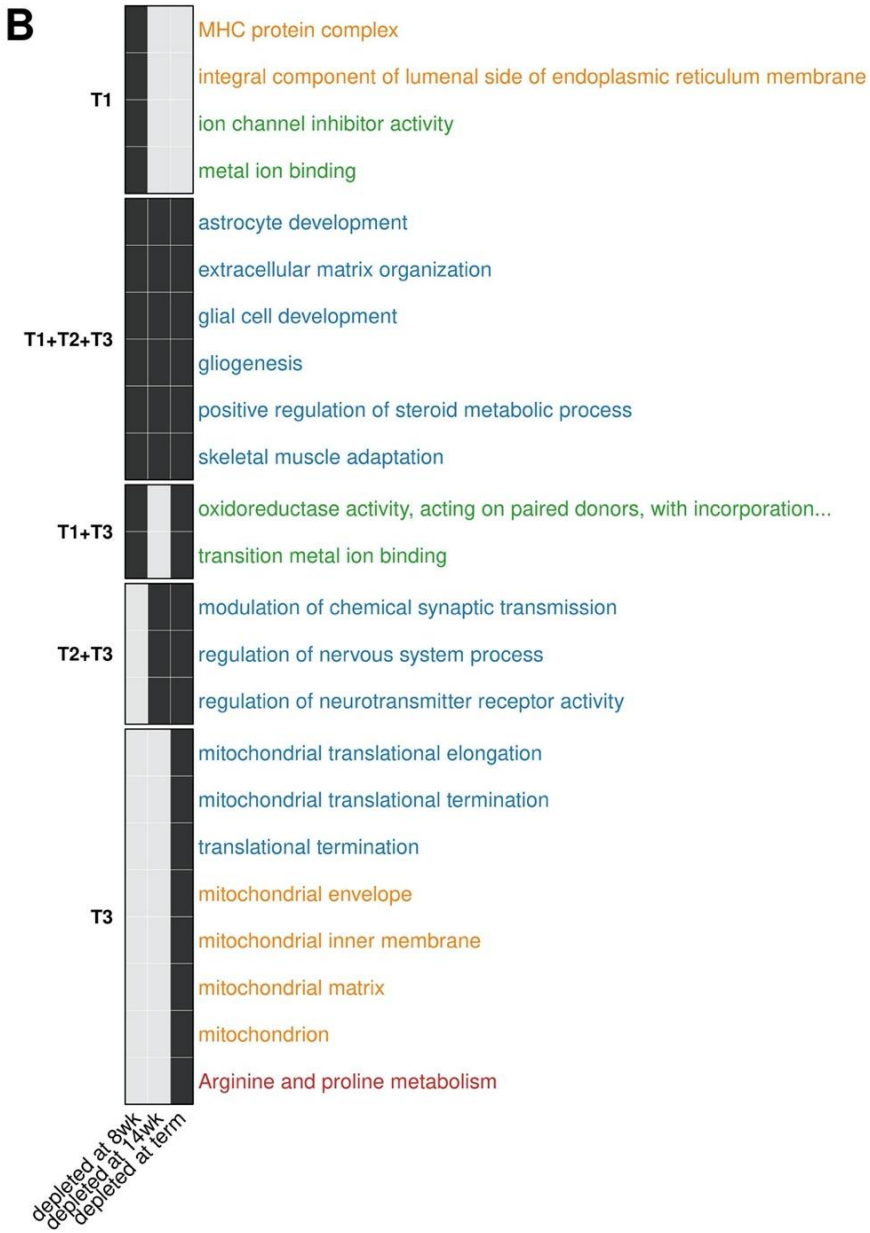

B
T1
MHC protein complex
integral component of lumenal side of endoplasmic reticulum membrane
ion channel inhibitor activity
metal ion binding
T1+T2+T3
astrocyte development
extracellular matrix organization
glial cell development
gliogenesis
positive regulation of steroid metabolic process
skeletal muscle adaptation
T1+T3
oxidoreductase activity, acting on paired donors, with incorporation...
transition metal ion binding
T2+T3
modulation of chemical synaptic transmission
regulation of nervous system process
regulation of neurotransmitter receptor activity
T3
mitochondrial translational elongation
mitochondrial translational termination
translational termination
mitochondrial envelope
mitochondrial inner membrane
mitochondrial matrix
mitochondrion
Arginine and proline metabolism
depleted at 8wk
depleted at 14wk
depleted at term

Figure 2. Abundance and GO analysis of transcripts depleted in the placenta during pregnancy. A, A heatmap representation of the abundance of 883 placenta-depleted transcripts (rows) across 29 tissues including our placenta datasets, i.e. 8wk (T1), 14wk (T2), and term (T3). nCPM ($\log_{10}$ scale) is color-coded from red (higher) to blue (lower) and their values across the 883 transcripts are shown as a boxplot in the bottom. 26 non-placental tissues are shown in descending order of the median CPM and the placenta samples are ordered by gestational age. The 883 transcripts are ordered in descending order of the average CPM across the placental samples. **B**, A summary of the gene ontology (GO) and KEGG pathway analysis. Significantly (adjusted $P<0.05$) overrepresented GO terms and KEGG pathway are shown with the following three sources: biological process (blue); molecular function (green); cellular component (orange); KEGG (red). The black and grey square represent being depleted and not being depleted, respectively, at the corresponding gestational age. To note, GO terms with at least the depth of 5 or more from the root and the intersection size of 3 or more (i.e. the number of overlaps between the query and the number of annotated genes belong to the GO terms in question) were plotted. For KEGG pathways, those with the intersection size of 5 or more were plotted. The full list of significantly overrepresented GO terms is shown in **Supplementary Table 10**. For **A** and **B**, the 883 transcripts are divided by their depleted status among the placenta samples at 8wk, 14wk, and term as shown by the following: T1: depleted at 8wk only; T1+T2: depleted at 8wk and 14wk; T1+T2+T3: depleted at 8wk, 14wk and term; T1+T3: depleted at 8wk and term; T2: depleted at 14wk only; T2+T3: depleted at 14wk and term; T3: depleted at term only. For display, non-placental tissues, shown in **A**, were manually selected if there are 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 sub-regions are show in **Supplementary Table 2**. nCPM: normalized count per million; wk: week of gestation.

Abundance of nuclear-encoded transcripts localized in the mitochondria

Mitochondria contain proteins encoded by nuclear DNA and subsequently imported to the mitochondria as well as those directly encoded by mitochondrial DNA (mtDNA). MitoMiner(Smith and Robinson 2016) is a database of protein coding genes with strong support for mitochondrial localization and hence function. Having observed association of mitochondria-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=5.1x10^{-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 (8wk, 14wk 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 were lowest in the term placenta, followed by the 14wk placenta, while they were the most abundant in the muscle, followed by the liver and the heart (**Supplementary Figure 6)**. 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 (KMO and ARMS2) were measured using RT-qPCR (discussed further below). Interestingly multidimensional scaling of all 11,876 samples showed a profound clustering of all 49 tissues, indicating tissue-specific expression of the 1,042 genes (**Figure 3C** and **Supplementary Figure 7**). 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(Moore et al. 2020) (n=398), blood(Lee-Six et al. 2018) (n=199), colon(Lee-Six et al. 2019) (n=568), and liver(Brunner et al. 2019) (n=517); see **Supplementary Table 11**, **Supplementary Figure 8**) and 21 non-placental tissues from the Cancer Genome Atlas Pan-Cancer Analysis of Whole Genomes (PCAWG) Consortium(Yuan et al. 2020). We found that mtDNA copy number was not substantially lower in the placenta than other tissues we compared (see **Supplementary Text**).

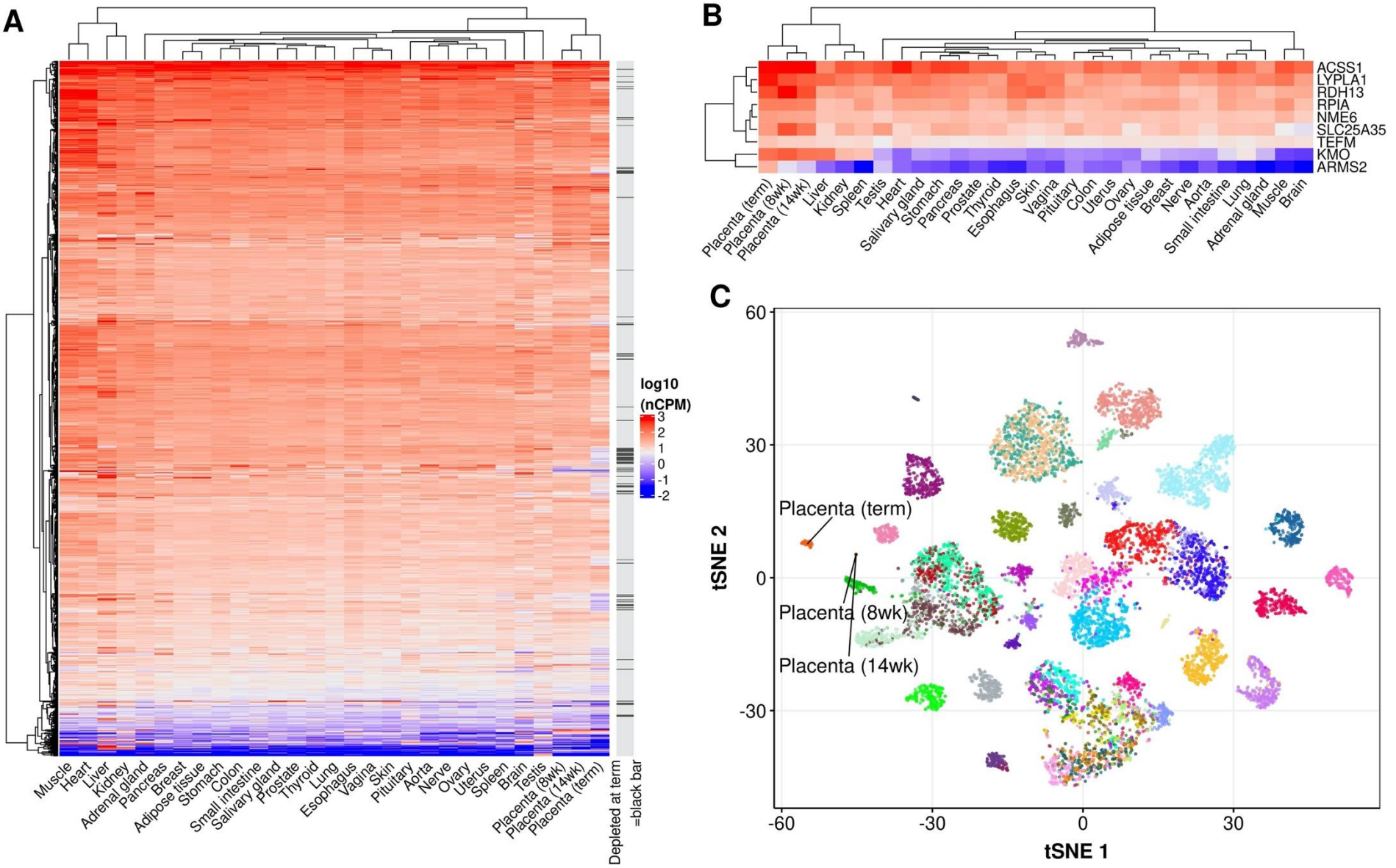

A
log10
(nCPM)
3
2
1
0
-1
-2
Muscle
Heart
Liver
Kidney
Adrenal gland
Pancreas
Breast
Adipose tissue
Stomach
Colon
Small intestine
Salivary gland
Prostate
Thyroid
Lung
Esophagus
Vagina
Skin
Pituitary
Aorta
Nerve
Ovary
Uterus
Spleen
Brain
Testis
Placenta (8wk)
Placenta (14wk)
Placenta (term)
Depleted at term
=black bar
B
ACSS1
LYPLA1
RDH13
RPIA
NME6
SLC25A35
TEFM
KMO
ARMS2
Placenta (term)
Placenta (8wk)
Placenta (14wk)
Liver
Kidney
Spleen
Testis
Heart
Salivary gland
Stomach
Pancreas
Prostate
Thyroid
Esophagus
Skin
Vagina
Pituitary
Colon
Uterus
Ovary
Adipose tissue
Breast
Nerve
Aorta
Small intestine
Lung
Adrenal gland
Muscle
Brain
C
60
30
0
-30
tSNE 2
Placenta (term)
Placenta (8wk)
Placenta (14wk)
-60
-30
0
30
tSNE 1

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 (tSNE) of 11,876 samples from 49 tissues using nCPMs (calculated by 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 8wk, 6 at 14wk and 59 at term) and each color represents a tissue and the full color-coding is shown in **Supplementary Figure 7**. For display, non-placental tissues, shown in **A**, were manually selected if there are 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 sub-regions are show in **Supplementary Table 2**. nCPM: normalized count per million; wk: week of gestation; tSNE: t-distributed stochastic neighbor embedding.

Validation of placental transcript abundances

To confirm the transcript abundance levels of the placenta and other somatic tissues that we identified based on RNA-Seq dataset, we performed RT-qPCR assays in independent samples. We selected a total of 13 transcripts (11 depleted and 2 mitochondria-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 Methods for details). For each of the 11 depleted transcripts, we confirmed that their abundances were significantly lower in the placenta than the other tissues (P=1.7x10^{-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 the highest in the placenta and it was significantly higher (P=2x10^{-5}, Mann-Whitney test) than the other tissues tested (**Figure 4**). 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 RT-qPCR.

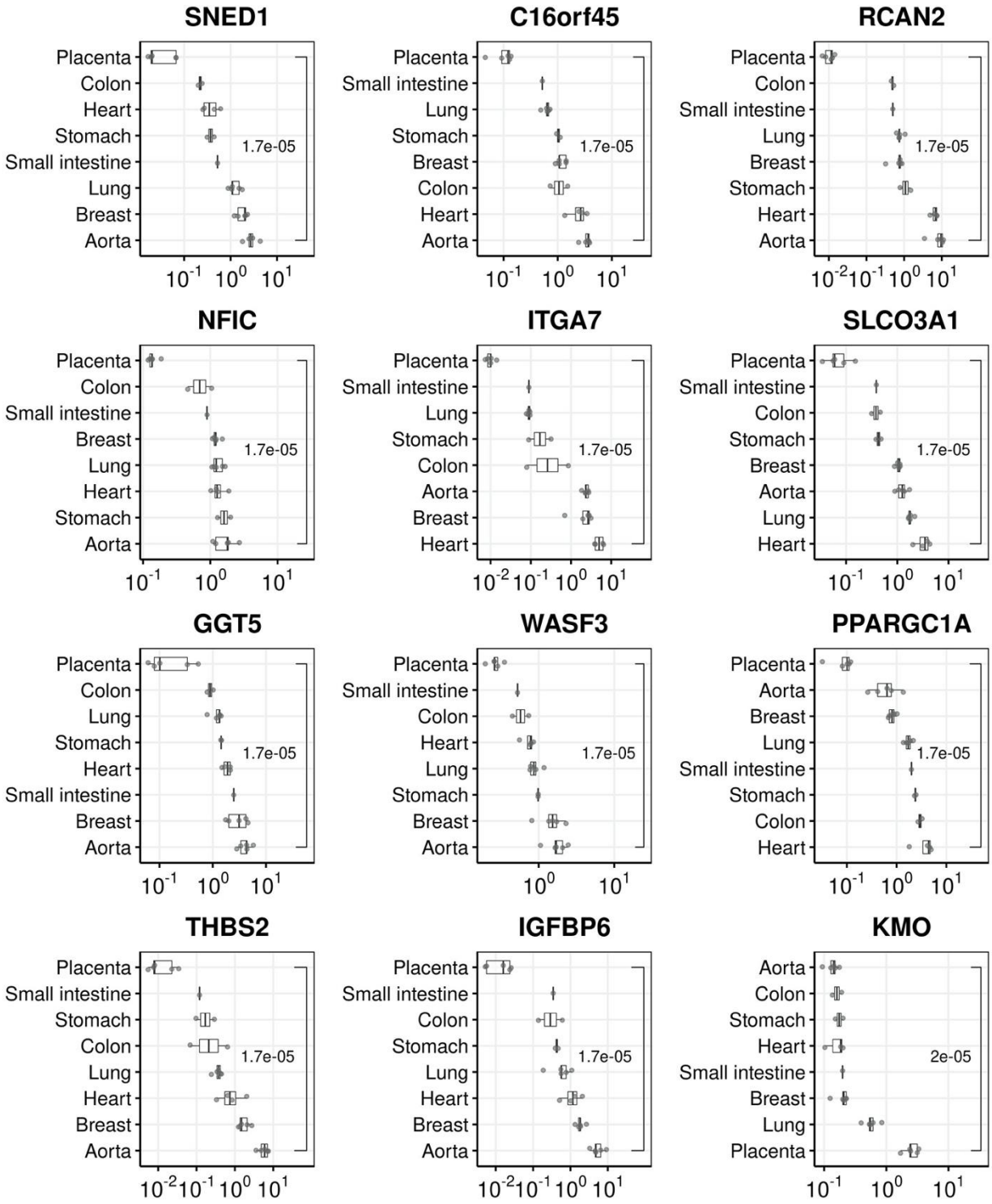


Figure 4. qPCR validation assays for 14 transcripts. The transcript abundance levels (x-axis) were measured in 8 tissues (y-axis) which were ranked from the lowest to highest level using the median abundance for each transcript. Each dot represents an individual sample (by taking the mean of the technical triplicate) and each of the boxes shows the median and interquartile range (IQR). The horizontal lines (whiskers) extended from the box represent a range of 1.5*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 **Supplementary Table 12**.

Discussion

The key finding of the present study is that the human placenta has a unique transcriptome architecture compared to 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. Gene Ontology analysis indicated that some of the absent transcripts reflect known unique qualities of placental function. For example, the placenta has no innervation and this likely reflects the identification by Gene Ontology 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(Moffett et al. 2017) and this is reflected in unique expression of MHC antigens, which was another Gene Ontology analysis pathway identified. However, Gene Ontology analysis identified other pathways which 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 other somatic tissues. Interestingly, the mRNA for PPARGC1A (Peroxisome proliferator-activated receptor gamma coactivator 1-alpha; also known as PGC-1α), a transcriptional master regulator of mitochondrial biogenesis, was depleted in the placenta in all three trimesters, as well as in four external placenta datasets (**Supplementary Table 7**). *Ppargc1a*-null mice have decreased expression of mitochondrial genes, especially those encoding various subunits of the electron transport chain (Austin and St-Pierre 2012; Vernier and Giguère 2021), suggesting a possible link between its low abundance and the diminished content of mitochondrial transcripts in the placenta. PGC-1α interacts with a very wide range of transcriptional co-activators and is a key regulator of metabolic homeostasis(Miller et al. 2019) 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 (45/47) in the term placenta. As ectopic expression of PGC-1α reduced levels of ROS(Valle et al. 2005; St-Pierre et al. 2006), it has been suggested that PGC-1α ensures high energy metabolism and removal of its toxic by-

products at the same time. We recently reported that the placenta has a unique somatic mutation profile(Coorens et al. 2021), predominantly the SBS18 signature which is associated with oxidative stress. This could be explained, at least in part, by accumulation of ROS possibly due to lower level of PGC-1α 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). The STB mitochondria have different morphological properties compared to CTB, specifically, they are smaller with irregular spherical cristae and a dense matrix(Holland et al. 2017; Fisher et al. 2020) and it has been suggested that these changes are related to steroidogenesis of STB(Martínez et al. 1997; Martinez et al. 2015). Functional studies have shown that STB mitochondria have reduced membrane potential, increased levels of hydrogen peroxide, lower antioxidant level and are more sensitive to ROS(Watson et al. 1998; Bustamante et al. 2014; Schoots et al. 2018). 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 previously reported that placental polyamine metabolism is implicated in placentally related complications of human pregnancy(Gong et al. 2018b). In the current study we found that genes associated with the KEGG pathway “arginine and proline metabolism” were over-represented among the depleted transcripts. Within 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 (see **Supplementary Text** and **Supplementary Figure 4**). Interestingly, expression levels of some genes in the polyamine pathway (e.g., *SAT1, AZIN1, SMS,* and *AMD1*) were higher in the placenta than non-placental tissues. Kajander et al.(Kajander et al. 1989) reported enzymatic activities of spermidine synthase (SRM) and spermine synthase (SMS) in seven human tissues, and the SMS-to-SRM ratio was the highest in the placenta (~5) followed by the kidney (~3.7). In our comparative analysis of RNA-Seq datasets, we confirmed that the SMS-to-SRM ratio at their transcript level was also the highest at the term placenta (48.7), followed by the 14wk placenta (5.4), and the 8wk 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. 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. One 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 to other tissues(Ehrlich et al. 1982; Fuke et al. 2004), it varies in locus-specific manner(Chatterjee et al. 2016; Gong et al. 2018a). We previously reported 71 protein-coding transcripts specifically enriched in the placenta(Gong et al. 2021), and we compared their promotor methylation levels(Gong et al. 2018a) with those of the transcripts depleted in the placenta we report here. However, we did not observe any significant difference in DNA methylation between the enriched and depleted transcripts in the CPG islands associated with these genes (**Supplementary Figure 9**). Moreover, 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=3.4x10^{-11}). This observation runs counter to the typical “high-methylation - low-expression” relationship and suggests that other mechanisms may be involved, and this is an area for future study. We considered possible differences in the number of mitochondria per nucleus but the placenta was not an outlier in mitochondrial DNA copy number (**Supplementary Figure 8**) and some mitochondrial genes actually had higher levels of expression in the placenta than in other tissues.

We noted that the proportion of genes encoding depleted transcripts from chromosome 19 was higher than expected by chance (Odds Ratio=3.18, P=1.5x10^{-25}; Fisher’s exact test) (**Supplementary Figure 10**). 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(Grimwood et al. 2004; Harris et al. 2020). 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 (PSG, 10 functional

448 genes) and the large imprinted cluster of miRNAs which are largely placenta
449 specific(Donker et al. 2012) (C19CM, 46 genes). These genes are all highly
450 expressed in the placenta(Gong et al. 2021). However, chromosome 19 encodes
451 other unusual gene families. Natural killer cells play an important role in human
452 pregnancy(Colucci 2019) and their receptors (killer cell immunoglobulin-like receptors,
453 KIRs.) are highly polymorphic and are in a cluster within the 1 Mb leukocyte receptor
454 complex in Chromosome 19.
455
456 Finally, transposable elements or their remains account for a surprisingly large
457 fraction of the human genome and host organisms have evolved numerous strategies
458 to defend themselves against the threat posed by functional endogenous
459 retroelements. One such mechanism involves the very large and rapidly evolving
460 family of transcription factors, the Krüppel-associated box domain zinc finger proteins
461 (KRAB-ZFPs)(Bruno et al. 2019). There are an estimated 352 genes encoding KRAB-
462 ZFPs in the human genome, and 209 of these are located within six clusters on
463 chromosome 19(Yang et al. 2017). Thus chromosome 19 has many unusual features
464 and many of these are related to placental or reproductive function. We now add to its
465 list of unusual features – that of having an over representation of depleted transcripts
466 in the human placenta.
467
468 Clearly, there is a risk that apparent differences in expression could be observed due
469 to batch effects and this could occur at any of a range of levels, including RNA
470 extraction, the sequencing platform employed, the transcriptome size, and the
471 normalization method employed. For example, the results of long RNA-seq are
472 profoundly affected by whether RNA was size selected or selected by oligo dTs for
473 selective extraction of mRNA through its polyadenylated tail(Gong et al. 2021). We
474 mitigated these risks using multiple approaches. First, we ensured comparable
475 methods of RNA extraction and sequencing between our own samples (mirVana
476 miRNA isolation Kit, Ambion) and GTEx (Tissue miRNA Kit, PreAnalytix® and
477 miRNeasy Mini Kit, Qiagen). Second, we analyzed multiple distinct placental RNA-
478 seq datasets, including those we generated and using other publicly available
479 sources. Third, we compared multiple bioinformatic approaches to determine that the
480 results were robust to definitions using different pipelines. Finally, we externally
481 validated some of the key results using qPCR and a separate group of placental

samples and tissue samples obtained from a local tissue bank. 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 transient organ – but not in others. Moreover, the neglect of the placenta in a large-scale international consortium resulted in the exclusion of one of the human body's most interesting transcriptional landscapes.

Materials and Methods

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(Pasupathy et al. 2008; Gaccioli et al. 2016). 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 analysed in this study - 60 samples were used for RNA-Seq (**Supplementary Table 1**) and 80 for WGS (**Supplementary Table 11**). One of the RNA-Seq samples was dropped from the analysis due to the presence of decidual contamination(Gong et al. 2021) 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°C until analysis. These samples have previously been described in full(Prater et al. 2021).

RNA sequencing and data processing

The POP study placental biopsies were collected within 30 minutes of delivery and flash frozen in RNAlater (ThermoFisher). For each biopsy, total placental RNA was extracted from approximately 5 mg of tissue using the "mirVana miRNA Isolation Kit" (Ambion) which efficiently isolates all RNAs longer than 10 nucleotides in length, followed by DNase treatment ("DNA-free DNA Removal Kit", Ambion). RNA quality was assessed with the Agilent Bioanalyzer and all the samples with RIN values ≥ 7.0 were used in the downstream experiments. RNA-libraries were prepared from 1µg 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 (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):
tophat2 -p 32 --library-type fr-firststrand --output-dir $OUTPUT_DIR --max-multihits 10 --prefilter-multihits -transcriptome-index=$TR_INDEX $BOWTIE2INDEX $TRIMMED_FASTQ

The transcriptome index above was built using transcript annotation from Ensembl v88 (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:

546 *tophat2 -p 32 --library-type fr-firststrand --output-dir $OUTPUT_DIR --raw-juncs*
547 *$MERGED_JUNCTION $UNMAPPED_FASTQ*
548
549 For each sample, the initial and second mapped reads were merged by *samtools*
550 (v1.2-24-g016c62b):
551 *samtools merge $MERGED_BAM $FIST_MAP_BAM $SECOND_MAP_BAM*
552
553 Before gene-level quantification of read counts, we pre-processed the transcript
554 annotation file (Gencode v26) using the '*collapse_annotation.py*' python script
555 available from the following GTEx github site: https://github.com/broadinstitute/gtex-
556 pipeline/tree/master/gene_model.
557 *python3 collapse_annotation.py $GENCODE_26_GTF*
558 *$PROCESSED_GENCODE_26_GTF*
559
560 Finally, we quantified sequencing reads at the gene-level using *featureCounts* tool of
561 *subread* package (v1.5.1):
562 *featureCounts -T 32 -a $PROCESSED_GENCODE_26_GTF -Q 10 -s 2 -p -C -o*
563 *$GENE_COUNT_OUTPUT $MERGED_BAM*
564

565 *Tissue collection for RT-qPCR analysis*

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

Approximately 35 mg of frozen tissues were homogenized by bead beating for 20 s at a speed of 4.5 ms^{-1} 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).

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(de Goffau et al. 2019), 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
AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGGTCGCCGTA
TCATT -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):

```
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(Moore et al. 2020).

GTEx data processing

We compared our placenta RNA-Seq datasets with 46 somatic tissues from GTEx (v8.p2). To select eligible samples from GTEx RNA-Seq datasets, we used the same filtering conditions applied to our previous study(Gong et al. 2021): (a) RNA integrity number (SMRIN) ≥6, (b) mapping rate (SMMAPRT) ≥0.9, (c) exonic mapping rate (SMEXNCRT) ≥0.75, and (d) ≥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. **Supplementary Table 2** 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) 6, (b) mapping rate to genome (SMMAPRT) >0.8, (c) mapping rate to exon (SMEXNCRT) >0.8, (d) ≥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.(Yang et al. 2011) (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(Anders and Huber 2010), and (2) the trimmed mean of M-values (TMM), available from '*calcNormFactors*' function of edgeR (v.3.28.1) Bioconductor package(Robinson and Oshlack 2010). 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 DESeq and TMM normalization methods, respectively. Finally, we selected 762 of them (**Supplementary Table 3**) 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 DESeq 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(Robinson and Storey 2014). 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 (**Supplementary Table 1**). Then, we applied the same approach of finding absent or depleted transcripts, as described earlier, to the down-sampled count matrix.

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. $nCPM_{(rank=20)}/nCPM_{(rank=21;term\text{-}placenta)}$) > 5, and 3) top 10 by the effect size (i.e. $nCPM_{(rank=20)} - nCPM_{(rank=21;term\text{-}placenta)}$). PGC-1α (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(Smith and Robinson 2016), 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. $nCPM_{(term\text{-}placenta)}/nCPM_{(non\text{-}placenta)}$). 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:

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

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* 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(Raudvere et al. 2019) (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 (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) 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(Yuan et al. 2020) from the following URL: https://ibl.mdanderson.org/tcma/download/TCMA-CopyNumber.tsv.zip. The dataset was downloaded as of the 9th 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 Verheecke's study was obtained personally from one of the authors.

740 *Identification of transcripts localized in the mitochondria*

741 We downloaded MitoMiner (v4), a dataset of mitochondrial localization, from the
742 following URL: http://mitominer.mrc-mbu.cam.ac.uk/release-4.0/mitocarta.do. We
743 selected genes of the following conditions: (1) not encoded in mitochondrial
744 chromosome, (2) "Known mitochondrial" or "Predicted mitochondrial" as types of
745 evidence, and (3) one of the 19,156 eligible protein-coding genes described above.
746 After filtering, we considered 1,042 protein-coding genes.
747

748 *Code Availability*

749 Codes used in this study is available in the Methods section and at
750 https://gitlab.com/sunggong/pops-placenta-mt-2020.
751

752 **Data access**

753 The term placenta RNA-Seq data have been deposited in the European Genome-
754 phenome Archive (EGA, https://ega-archive.org/) with the following accession
755 number: EGAD00001006304. The early gestation placenta RNA-seq data have been
756 deposited in the European Nucleotide Archive (ENA, https://www.ebi.ac.uk/ena) with
757 the following accession number: PRJEB38810. The term placenta WGS data have
758 been deposited in the EGA with the following accession number: EGAD00001004198.
759 Correspondence and requests for materials should be addressed to D.S.C-J.
760 (dscj1@cam.ac.uk)
761

762 **Competing interest statement**

763 D.S.C-J. reports non-financial support from Roche Diagnostics Ltd, outside the
764 submitted work; G.C.S.S. reports personal fees and non-financial support from Roche
765 Diagnostics Ltd, outside the submitted work; D.S.C-J. and G.C.S.S. report grants from
766 Sera Prognostics Inc, non-financial support from Illumina Inc, outside the submitted
767 work. S.G, F.G, I.A, G.A, E.C, A.R.J.L. and L.M.R.H. have nothing to disclose.
768

769

Acknowledgements

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.A. is funded by the Centre for Trophoblast Research (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, G.C.S.S, designed the experiments. F.G, I.A, G.A. and E.C. performed the experiments. S.G. analyzed all the 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.

References

Anders S, Huber W. 2010. Differential expression analysis for sequence count data. *Genome Biol* **11**: R106.

Armstrong DL, McGowen MR, Weckle A, Pantham P, Caravas J, Agnew D, Benirschke K, Savage-Rumbaugh S, Nevo E, Kim CJ, et al. 2017. The core transcriptome of mammalian placentas and the divergence of expression with placental shape. *Placenta* **57**: 71–78.

Ashley B, Simner C, Manousopoulou A, Jenkinson C, Hey F, Frost JM, Rezwan FI, White CH, Lofthouse EM, Hyde E, et al. 2021. Placental uptake and metabolism as determinants of pregnancy vitamin D status. *BioRxiv*.

Austin S, St-Pierre J. 2012. PGC1α and mitochondrial metabolism--emerging concepts and relevance in ageing and neurodegenerative disorders. *J Cell Sci* **125**: 4963–4971.

Awamleh Z, Gloor GB, Han VKM. 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.

Brosens I, Pijnenborg R, Vercruysse L, Romero R. 2011. The "Great Obstetrical Syndromes" are associated with disorders of deep placentation. *Am J Obstet Gynecol* **204**: 193–201.

Brunner SF, Roberts ND, Wylie LA, Moore L, Aitken SJ, Davies SE, Sanders MA, Ellis P, Alder C, Hooks Y, et al. 2019. Somatic mutations and clonal dynamics in healthy and cirrhotic human liver. *Nature* **574**: 538–542.

Bruno M, Mahgoub M, Macfarlan TS. 2019. The Arms Race Between KRAB-Zinc Finger Proteins and Endogenous Retroelements and Its Impact on Mammals. *Annu Rev Genet* **53**: 393–416.

Buckberry S, Bianco-Miotto T, Bent SJ, Clifton V, Shoubridge C, Shankar K, Roberts CT. 2017. Placental transcriptome co-expression analysis reveals conserved regulatory programs across gestation. *BMC Genomics* **18**: 10.

Bustamante J, Ramírez-Vélez R, Czerniczyniec A, Cicerchia D, Aguilar de Plata AC, Lores-Arnaiz S. 2014. Oxygen metabolism in human placenta mitochondria. *J Bioenerg Biomembr* **46**: 459–469.

Chatterjee A, Macaulay EC, Rodger EJ, Stockwell PA, Parry MF, Roberts HE, Slatter TL, Hung NA, Devenish CJ, Morison IM. 2016. Placental hypomethylation is more pronounced in genomic loci devoid of retroelements. *G3 (Bethesda)* **6**: 1911–1921.

Colucci F. 2019. The immunological code of pregnancy. *Science* **365**: 862–863.

Coorens THH, Oliver TRW, Sanghvi R, Sovio U, Cook E, Vento-Tormo R, Haniffa M, Young MD, Rahbari R, Sebire N, et al. 2021. Inherent mosaicism and extensive mutation of human placentas. *Nature* **592**: 80–85.

de Goffau MC, Lager S, Sovio U, Gaccioli F, Cook E, Peacock SJ, Parkhill J, Charnock-Jones DS, Smith GCS. 2019. Human placenta has no microbiome but can contain potential pathogens. *Nature* **572**: 329–334.

de Rie D, Abugessaisa I, Alam T, Arner E, Arner P, Ashoor H, Åström G, Babina M, Bertin N, Burroughs AM, et al. 2017. An integrated expression atlas of miRNAs and their promoters in human and mouse. *Nat Biotechnol* **35**: 872–878.

Donker RB, Mouillet JF, Chu T, Hubel CA, Stolz DB, Morelli AE, Sadovsky Y. 2012. The expression profile of C19MC microRNAs in primary human trophoblast cells and exosomes. *Mol Hum Reprod* **18**: 417–424.

Ehrlich M, Gama-Sosa MA, Huang LH, Midgett RM, Kuo KC, McCune RA, 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.

Eisenberg E, Levanon EY. 2013. Human housekeeping genes, revisited. *Trends Genet* **29**: 569–574.

ENCODE Project Consortium. 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**: 57–74.

Fisher JJ, Bartho LA, Perkins AV, Holland OJ. 2020. Placental mitochondria and reactive oxygen species in the physiology and pathophysiology of pregnancy. *Clin Exp Pharmacol Physiol* **47**: 176–184.

Fuke C, Shimabukuro M, Petronis A, Sugimoto J, Oda T, Miura K, Miyazaki T, Ogura C, Okazaki Y, Jinno Y. 2004. Age related changes in 5-methylcytosine content in human peripheral leukocytes and placentas: an HPLC-based study. *Ann Hum Genet* **68**: 196–204.

Gaccioli F, Lager S, Sovio U, Charnock-Jones DS, Smith GCS. 2016. The pregnancy outcome prediction (POP) study: Investigating the relationship between serial prenatal ultrasonography, biomarkers, placental phenotype and adverse pregnancy outcomes. *Placenta* **0**.

Gong S, Gaccioli F, Dopierala J, Sovio U, Cook E, Volders P-J, Martens L, Kirk PDW, Richardson S, Smith GCS, et al. 2021. The RNA landscape of the human placenta in health and disease. *Nat Commun* **12**: 2639.

Gong S, Johnson MD, Dopierala J, Gaccioli F, Sovio U, Constância M, Smith GC, Charnock-Jones DS. 2018a. Genome-wide oxidative bisulfite sequencing identifies sex-specific methylation differences in the human placenta. *Epigenetics* **13**: 228–239.

Gong S, Sovio U, Aye IL, Gaccioli F, Dopierala J, Johnson MD, Wood AM, Cook E, Jenkins BJ, Koulman A, et al. 2018b. Placental polyamine metabolism differs by fetal sex, fetal growth restriction, and preeclampsia. *JCI Insight* **3**.

Grimwood J, Gordon LA, 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.

GTEx Consortium. 2020. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**: 1318–1330.

Harris RA, Raveendran M, Worley KC, Rogers J. 2020. Unusual sequence characteristics of human chromosome 19 are conserved across 11 nonhuman primates. *BMC Evol Biol* **20**: 33.

Holland O, Dekker Nitert M, Gallo LA, Vejzovic M, Fisher JJ, Perkins AV. 2017. Review: Placental mitochondrial function and structure in gestational disorders. *Placenta* **54**: 2–9.

Hondares E, Mora O, Yubero P, Rodriguez de la Concepción M, Iglesias R, Giralt M, Villarroya F. 2006. Thiazolidinediones and rexinoids induce peroxisome proliferator-activated receptor-coactivator (PGC)-1alpha gene transcription: an autoregulatory loop controls PGC-1alpha expression in adipocytes via peroxisome proliferator-activated receptor-gamma coactivation. *Endocrinology* **147**: 2829–2838.

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 lnc-SLC4A1-1 epigenetically regulates unexplained recurrent pregnancy loss (URPL) by activating CXCL8 and NF-kB pathway. *EBioMedicine* **38**: 162–170.

Hudson BG, Tryggvason K, Sundaramoorthy M, Neilson EG. 2003. Alport's syndrome, Goodpasture's syndrome, and type IV collagen. *N Engl J Med* **348**: 2543–2556.

Jain A, Tuteja G. 2019. TissueEnrich: Tissue-specific gene enrichment analysis. *Bioinformatics* **35**: 1966–1967.

Kajander EO, Kauppinen LI, Pajula RL, Karkola K, Eloranta TO. 1989. Purification and partial characterization of human polyamine synthases. *Biochem J* **259**: 879–886.

Kim J, Zhao K, Jiang P, Lu Z, Wang J, Murray JC, Xing Y. 2012. Transcriptome landscape of the human placenta. *BMC Genomics* **13**: 115.

Lee-Six H, Øbro NF, Shepherd MS, Grossmann S, Dawson K, Belmonte M, Osborne RJ,

892 Huntly BJP, Martincorena I, Anderson E, et al. 2018. Population dynamics of normal
893 human blood inferred from somatic mutations. *Nature* **561**: 473–478.
894 Lee-Six H, Olafsson S, Ellis P, Osborne RJ, Sanders MA, Moore L, Georgakopoulos N,
895 Torrente F, Noorani A, Goddard M, et al. 2019. The landscape of somatic mutation in
896 normal colorectal epithelial cells. *Nature* **574**: 532–537.
897 Lim YC, Li J, Ni Y, Liang Q, Zhang J, Yeo GSH, Lyu J, Jin S, Ding C. 2017. A complex
898 association between DNA methylation and gene expression in human placenta at first
899 and third trimesters. *PLoS One* **12**: e0181155.
900 Martínez F, Kiriakidou M, Strauss JF. 1997. Structural and functional changes in mitochondria
901 associated with trophoblast differentiation: methods to isolate enriched preparations of
902 syncytiotrophoblast mitochondria. *Endocrinology* **138**: 2172–2183.
903 Martinez F, Olvera-Sanchez S, Esparza-Perusquia M, Gomez-Chang E, Flores-Herrera O.
904 2015. Multiple functions of syncytiotrophoblast mitochondria. *Steroids* **103**: 11–22.
905 Miller KN, Clark JP, Anderson RM. 2019. Mitochondrial regulator PGC-1a-Modulating the
906 modulator. *Current Opinion in Endocrine and Metabolic Research* **5**: 37–44.
907 Moffett A, Chazara O, Colucci F. 2017. Maternal allo-recognition of the fetus. *Fertil Steril*
908 **107**: 1269–1272.
909 Moll R, Divo M, Langbein L. 2008. The human keratins: biology and pathology. *Histochem*
910 *Cell Biol* **129**: 705–733.
911 Moore L, Leongamornlert D, Coorens THH, Sanders MA, Ellis P, Dentro SC, Dawson KJ,
912 Butler T, Rahbari R, Mitchell TJ, et al. 2020. The mutational landscape of normal
913 human endometrial epithelium. *Nature* **580**: 640–646.
914 Papatheodorou I, Moreno P, Manning J, Fuentes AM-P, George N, Fexova S, Fonseca NA,
915 Füllgrabe A, Green M, Huang N, et al. 2020. Expression Atlas update: from tissues to
916 single cells. *Nucleic Acids Res* **48**: D77–D83.
917 Pasupathy D, Dacey A, Cook E, Charnock-Jones DS, White IR, Smith GCS. 2008. Study
918 protocol. A prospective cohort study of unselected primiparous women: the pregnancy
919 outcome prediction study. *BMC Pregnancy Childbirth* **8**: 51.
920 Penning TM, Chen M, Jin Y. 2015. Promiscuity and diversity in 3-ketosteroid reductases. *J*
921 *Steroid Biochem Mol Biol* **151**: 93–101.
922 Prater M, Hamilton RS, Wa Yung H, Sharkey AM, Robson P, Abd Hamid NE, Jauniaux E,
923 Charnock-Jones DS, Burton GJ, Cindrova-Davies T. 2021. RNA-Seq reveals changes
924 in human placental metabolism, transport and endocrinology across the first-second
925 trimester transition. *Biol Open* **10**.
926 Raudvere U, Kolberg L, Kuzmin I, Arak T, Adler P, Peterson H, Vilo J. 2019. g:Profiler: a
927 web server for functional enrichment analysis and conversions of gene lists (2019
928 update). *Nucleic Acids Res* **47**: W191–W198.
929 Roadmap Epigenomics Consortium, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A,
930 Heravi-Moussavi A, Kheradpour P, Zhang Z, Wang J, et al. 2015. Integrative analysis
931 of 111 reference human epigenomes. *Nature* **518**: 317–330.
932 Robinson DG, Storey JD. 2014. subSeq: determining appropriate sequencing depth through
933 efficient read subsampling. *Bioinformatics* **30**: 3424–3426.
934 Robinson MD, Oshlack A. 2010. A scaling normalization method for differential expression
935 analysis of RNA-seq data. *Genome Biol* **11**: R25.
936 Schoots MH, Gordijn SJ, Scherjon SA, van Goor H, Hillebrands J-L. 2018. Oxidative stress in
937 placental pathology. *Placenta* **69**: 153–161.
938 Smith AC, Robinson AJ. 2016. MitoMiner v3.1, an update on the mitochondrial proteomics
939 database. *Nucleic Acids Res* **44**: D1258-61.
940 St-Pierre J, Drori S, Uldry M, Silvaggi JM, Rhee J, Jäger S, Handschin C, Zheng K, Lin J,
941 Yang W, et al. 2006. Suppression of reactive oxygen species and neurodegeneration by
942 the PGC-1 transcriptional coactivators. *Cell* **127**: 397–408.

943 Valle I, Alvarez-Barrientos A, Arza E, Lamas S, Monsalve M. 2005. PGC-1alpha regulates the
944 mitochondrial antioxidant defense system in vascular endothelial cells. *Cardiovasc Res*
945 **66**: 562–573.
946 Verheecke M, Cortès Calabuig A, Finalet Ferreiro J, Brys V, Van Bree R, Verbist G, Everaert
947 T, Leemans L, Gziri MM, Boere I, et al. 2018. Genetic and microscopic assessment of
948 the human chemotherapy-exposed placenta reveals possible pathways contributive to
949 fetal growth restriction. *Placenta* **64**: 61–70.
950 Vernier M, Giguère V. 2021. Aging, senescence and mitochondria: the PGC-1/ERR axis. *J*
951 *Mol Endocrinol* **66**: R1–R14.
952 Watanabe K, Umićević Mirkov M, de Leeuw CA, van den Heuvel MP, Posthuma D. 2019.
953 Genetic mapping of cell type specificity for complex traits. *Nat Commun* **10**: 3222.
954 Watson AL, Skepper JN, Jauniaux E, Burton GJ. 1998. Susceptibility of human placental
955 syncytiotrophoblastic mitochondria to oxygen-mediated damage in relation to
956 gestational age. *J Clin Endocrinol Metab* **83**: 1697–1705.
957 Wu Z, Puigserver P, Andersson U, Zhang C, Adelmant G, Mootha V, Troy A, Cinti S, Lowell
958 B, Scarpulla RC, et al. 1999. Mechanisms controlling mitochondrial biogenesis and
959 respiration through the thermogenic coactivator PGC-1. *Cell* **98**: 115–124.
960 Yang L, Duff MO, Graveley BR, Carmichael GG, Chen L-L. 2011. Genomewide
961 characterization of non-polyadenylated RNAs. *Genome Biol* **12**: R16.
962 Yang P, Wang Y, Macfarlan TS. 2017. The Role of KRAB-ZFPs in Transposable Element
963 Repression and Mammalian Evolution. *Trends Genet* **33**: 871–881.
964 Yuan Y, Ju YS, Kim Y, Li J, Wang Y, Yoon CJ, Yang Y, Martincorena I, Creighton CJ,
965 Weinstein JN, et al. 2020. Comprehensive molecular characterization of mitochondrial
966 genomes in human cancers. *Nat Genet* **52**: 342–352.
967