Molecular Human Reproduction, Vol.20, No.8 pp. 810–819, 2014
Advanced Access publication on May 27, 2014 doi:10.1093/molehr/gau035

molecular human reproduction

ORIGINAL RESEARCH

Integrative transcriptome meta-analysis reveals widespread sex-biased gene expression at the human fetal–maternal interface

Sam Buckberry[1], Tina Bianco-Miotto[1,2], Stephen J. Bent[1], Gustaaf A. Dekker[1,3], and Claire T. Roberts[1,*]

[1]The Robinson Research Institute, School of Paediatrics and Reproductive Health, The University of Adelaide, Adelaide 5005, Australia [2]School of Agriculture Food & Wine, The University of Adelaide, Adelaide 5005, Australia [3]Lyell McEwin Hospital, Elizabeth Vale, SA 5112, Australia

*Correspondence address. E-mail: claire.roberts@adelaide.edu.au

Submitted on March 13, 2014; resubmitted on April 23, 2014; accepted on May 1, 2014

ABSTRACT: As males and females share highly similar genomes, the regulation of many sexually dimorphic traits is constrained to occur through sex-biased gene regulation. There is strong evidence that human males and females differ in terms of growth and development *in utero* and that these divergent growth strategies appear to place males at increased risk when in sub-optimal conditions. Since the placenta is the interface of maternal–fetal exchange throughout pregnancy, these developmental differences are most likely orchestrated by differential placental function. To date, progress in this field has been hampered by a lack of genome-wide information on sex differences in placental gene expression. Therefore, our motivation in this study was to characterize sex-biased gene expression in the human placenta. We obtained gene expression data for >300 non-pathological placenta samples from 11 microarray datasets and applied mapping-based array probe re-annotation and inverse-variance meta-analysis methods which showed that >140 genes (false discovery rate (FDR) <0.05) are differentially expressed between male and female placentae. A majority of these genes (>60%) are autosomal, many of which are involved in high-level regulatory processes such as gene transcription, cell growth and proliferation and hormonal function. Of particular interest, we detected higher female expression from all seven genes in the LHB-CGB cluster, which includes genes involved in placental development, the maintenance of pregnancy and maternal immune tolerance of the conceptus. These results demonstrate that sex-biased gene expression in the normal human placenta occurs across the genome and includes genes that are central to growth, development and the maintenance of pregnancy.

Key words: gene expression / microarray / placenta / pregnancy / sex chromosomes

Introduction

Females and males of many species demonstrate numerous differences in morphology and physiology, yet they share highly similar genomes. This suggests that the regulation of many sexually dimorphic traits occurs through sex-specific patterns of gene regulation. Since fetal growth *in utero* is dependent on the capacity of the placenta to facilitate exchange between the mother and fetus, developmental disparities between the sexes are likely orchestrated by differential placental function.

The observation that males grow faster *in utero* and have a greater body length and weight at birth than females with equivalent placental size (Misra *et al.*, 2009) indicates that the male placenta functions more efficiently (Forsén *et al.*, 1999; Eriksson *et al.*, 2010). However, there is a developmental trade-off: a consequence of growing more quickly and being larger *in utero* is that males are left with less reserve placental capacity to draw upon if sub-optimal conditions arise. In turn, this places males at increased risk of under nutrition (Eriksson *et al.*, 2010), which can restrict growth and lower birthweight, both of which have been linked to males' increased risk of adult-onset disorders such as cardiovascular disease (Barker, 2002). A recent study has also shown a distinct male bias in the prevalence of placental dysfunction (Murji *et al.*, 2012), and provides support for the findings of previous studies that showed sex biases in a spectrum of pregnancy complications and fetal health outcomes associated with abnormal placental development (Edwards *et al.*, 2000; Ingemarsson, 2003; Clifton, 2010; Lao *et al.*, 2011; Hadar *et al.*, 2012). Although sex differences in terms of growth, development and predisposition to pregnancy complications are increasingly becoming recognized, the underpinning sex biases in placental gene regulation remain unclear.

Recent efforts using massively parallel sequencing techniques have begun to expand our knowledge of the human placental transcriptional (Kim *et al.*, 2012) and epigenetic landscapes (Schroeder *et al.*, 2013). These studies have revealed that the placenta is unique in several ways, including the expression of placenta-specific genes, placenta-specific alternative splicing and widespread partially DNA methylated domains that regulate gene expression (Kim *et al.*, 2012; Schroeder *et al.*, 2013). Despite advancing our understanding of human placental gene regulation, these studies were not designed to capture the effect of sex, and therefore provide no clues as to underlying sex differences in placental function. An earlier study, which was the first to describe the human placental transcriptome, noted several genes with a sex-biased expression; these were located on both sex chromosomes and the autosomes (Sood *et al.*, 2006). However, given the low number of placental samples assessed, it is unlikely that the study was able to detect the true extent of sex-biased gene expression in the placenta.

In the present study, our aim was to characterize comprehensively the extent of sex-biased gene expression in the human placenta. To achieve this, we took advantage of the vast amount of human placental gene expression microarray data available in public repositories to perform a large-scale gene expression meta-analysis. In order to characterize only normal placental function, we selected samples from microarray datasets where no placental pathology or associated pregnancy complication was indicated. In applying integrative meta-analysis methods, our results demonstrate that sex-biased gene expression in the normal human placenta occurs across the genome and includes genes that are central to placental growth, development and the maintenance of pregnancy.

Materials and Methods

Study selection

We searched the public data repositories Gene Expression Omnibus and ArrayExpress, and the literature, for microarray gene expression datasets containing samples of human placental tissue. Our initial selection criteria required candidate datasets to have at least six individual placenta samples that were collected at the time of delivery. With the focus being on sex differences in normal development, we limited the inclusion of samples to those where no pregnancy or placental pathology was detailed in the associated metadata. For example, if a dataset contained placenta samples from pregnancies featuring pre-eclampsia and normal pregnancy controls, only the control sample arrays were included in the meta-analysis. Additionally, the meta-analysis was limited to studies where the raw, non-normalized, probe-level data were available for all the array probes. Arrays with pooled samples were excluded.

Array pre-processing and quality control

Since data were obtained from multiple microarray platforms, pre-processing methods were tailored to each platform. Affymetrix datasets were pre-processed, log-transformed and normalized using either the robust multi-array average (RMA) or GeneChip-RMA (GC-RMA) method depending on platform using Simpleaffy (Wilson and Miller, 2005). Applied Biosystems arrays were pre-processed using comprehensive R-based microarray analysis (CARMA); probes with a flag value of >100 were removed from the dataset before quantile normalization (Rainer *et al.*, 2006). Illumina bead arrays were pre-processed using Beadarray before quantile normalization (Dunning *et al.*, 2007). Datasets with arrays processed in multiple batches (as detailed in the metadata) were batch corrected using the 'comBat' function in the SVA package (Johnson *et al.*, 2007; Leek *et al.*, 2012). Outliers were eliminated from each dataset (see Supplementary data, Table SII) before re-normalization by checking the distance between arrays and assessing MA plots generated using ArrayQualityMetrics (Kauffmann *et al.*, 2009).

Predicting the sex of samples in datasets lacking sex information

In 7 of the 11 datasets in this meta-analysis, the sample's sex was not identified in either the associated repository metadata or in the associated publication. Therefore, to maximize the number of usable datasets, we used the Bioconductor package *massiR* to predict fetal sex (Buckberry *et al.*, 2014a). This method utilizes expression values for probes that map unambiguously to the Y chromosome and unsupervised clustering of samples based on Y chromosome probes with the highest variance.

We tested this method on placental datasets with known sample sex to determine its accuracy with placental data and to validate the sex of samples in datasets where sex information was detailed in the metadata. For datasets with known sex, this method predicted the correct sex with 100% accuracy in all but one dataset (GSE30032), where the sex of every sample was the opposite of the predicted sex in every case (as detailed in the GEO metadata). This method uses Y chromosome-specific probe information and so given that all samples designated male in the metadata were predicted to be female, and vice versa, we concluded that the metadata were incorrect. Therefore, we used the *massiR* predicted sex in this study.

Re-annotation of microarray datasets and probe summarization

Gene expression data were obtained from various microarray platforms that have different probes targeting the same genes. We therefore annotated each dataset with common gene identifiers to increase cross-platform concordance. We selected the gene identifiers from Ensembl Genes release 69 annotation (Flicek *et al.*, 2012) for probe mapping, which is the genome annotation used in the human GENCODE project (Harrow *et al.*, 2012). Probes from all Illumina and Affymetrix datasets were mapped to the human reference genome (GRCh37.3p) to translate platform-specific, probe-level identifiers to the Ensembl gene level identifiers. Probes were mapped using the Ensembl Functional Genomics Array Mapping Environment, in which individual probes are mapped to both the genome and the cDNA sequence. Alignments were performed by Ensembl using an analysis pipeline which implements the Exonerate sequence comparison and alignment tool (Slater and Birney, 2005). A 1-bp mismatch was permitted between the probe and the genome sequence assembly, and probes that match at 100+ locations (e.g. suspected Alu repeats) are discarded (see permalink for detailed methods http://jan2013.archive.ensembl.org/info/docs/microarray_probe_set_mapping.html).

Probe sequences were unavailable for the Applied Biosystems arrays and so Ensembl gene identifiers supplied by the manufacturer were used to identify target genes. For Applied Biosystems probes with no listed Ensembl identifier, the supplied gene symbol was used to identify the target gene using the HGNC database (Gray *et al.*, 2012) and an Ensembl Gene identifier was subsequently assigned. Any remaining identifiers with GenBank accessions (Benson *et al.*, 2011) were checked for a match against human sequences with sufficient gene information, and then designated an Ensembl Gene Identifier.

Probe mapping and annotation of all datasets (except Applied Biosystems arrays) allowed identification of four cases: (i) probes that map uniquely to a single-gene identifier (one-to-one mapping), (ii) probes that map to multiple gene identifiers (one-to-many mapping), (iii) multiple probes that map to the same gene identifier (many-to-one mapping) and (iv) probes that do not map to any genes in the reference genome. These re-annotation results are summarized in Supplementary data, Table SIII.

When a probe mapped to multiple gene identifiers (Case ii), a new probe identifier was created for each probe-to-gene mapping; this allowed the use of all possible information for each gene in the analysis. For gene identifiers where multiple probes were mapped (Case iii), probe values were summarized into a single representative value per gene identifier within each study, using a fixed inverse-variance model as previously described (Ramasamy *et al.*, 2008). Probes with insufficient information, or that did not map any gene identifier, were removed from the analysis.

Meta-analysis of annotated datasets

To meta-analyze the 11 annotated microarray datasets, we applied the inverse-variance method as detailed in Ramasamy *et al.* (2008) using the *rmeta* package, functions adapted from the *metaGEM* package (https://spiral.imperial.ac.uk/handle/10044/1/4217) and custom R scripts. Briefly, study-specific effect sizes were calculated for each probe within each study by calculating the probe mean and standard deviation, corrected for effect size using Hedges' *g* to account for the number of samples in each group. These study-specific estimates were then combined using a random effects inverse-variance method for each gene identifier to calculate the pooled effect size and standard error. Z-statistics were then calculated for each gene identifier to obtain a nominal *P*-value, which was then corrected using the FDR. After significance testing, the resulting dataset was limited to genes represented by at least three studies for downstream analyses. All data processing and analyses were carried out in the R statistical environment (version 2.15.2).

Prediction of upstream transcription factor regulation

The sex-biased gene set was analyzed for enrichment of transcription factor (TF)-binding sites (TFBS) using the oPOSSUM program, and the JASPAR vertebrate core profiles (Kwon *et al.*, 2012; Mathelier *et al.*, 2013). For each gene, we searched for TFBS motifs in the conserved regions of the 10 kb upstream/downstream sequences using a conservation cut-off of 0.4, a matrix score threshold of 85% and a minimum specificity of 8-bits. The highly enriched TFBSs were identified by ranking TFs using results from Fisher's exact test and z-score rankings.

Resolving CGB/LHB cluster sequence homology

Genes in the LHB-CGB cluster are both functionally and evolutionarily related (Liina Nagirnaja, 2010), and subsequently have a high degree of sequence homology. In such cases, the sequence specificity of each microarray probe is a key determinant in differentiating between the expression of individual genes. We re-annotated all array probes through mapping to a common reference genome, therefore were able to determine which probes mapped uniquely, or mapped to multiple genes, in the LHB-CGB cluster. In this meta-analysis, all probes that represent *LHB* expression mapped uniquely; therefore, it is unlikely that the *LHB* expression results are confounded by non-specific binding with *CGB* gene transcripts. In the case of the *CGB* genes, mapping specificity differed between platforms. Affymetrix probes mapped with low specificity: 10 probes mapped to all *CGB* cluster genes and only 1 probe mapped uniquely (to *CGB7*). However, probes from the Illumina platforms mapped with much higher specificity. Of these, nine probes mapped specifically genes in one of the three classes of CGB protein isoforms (Fig. 6) and three of these probes had single-gene specificity. A majority of samples in this study (76%) were assayed on Illumina platforms and so we have reasonably high confidence that the expression results for *CGB* genes are composed primarily of values from probes with the highest specificity.

Amino acid sequences for *LHB* and *CGB* cluster genes were downloaded from ENSEMBL. Sequences were aligned using MAFFT (v7.130b) with L-INS-i settings, and the tree was calculated with the average distance using percent identity in Jalview (v2.8). Branch lengths represent the percentage mismatch between two nodes.

Gene enrichment and pathway analysis

Enriched biological functions and canonical pathways associated with sex-biased genes were determined using Ingenuity Pathway Analysis (Ingenuity Systems, v18030641).

Results and discussion

Meta-analysis of sex-biased gene expression in the human placenta

This meta-transcriptome analysis of the sex differences in human placental gene expression incorporated 303 samples from 11 microarray datasets generated on six different platforms (Table I). We limited this analysis to non-pathological placental samples to provide the most accurate evaluation of sex differences in relative gene expression in normal human pregnancies at the time the fetus was delivered.

To improve microarray cross-platform concordance and to standardize gene identifiers throughout this meta-analysis, we re-annotated array probes by mapping to a common reference genome. After this re-annotation and summary process, we were able to quantify expression of 31 844 Ensembl genes (hereafter referred to as genes) across the human genome. To confirm the sex of samples and to predict sex when it was not listed in the associated metadata or publication, we employed an unsupervised clustering technique that classifies the sex of samples in microarray datasets using signal intensity values for probes that map unambiguously to Y chromosome genes (Buckberry *et al.*, 2014a).

When limiting the results to genes measurable in at least three studies and with an FDR of <0.05, a total of 142 genes showed significant sex-biased expression. Of these 142 genes, 75 showed higher expression in placentas from female fetuses and 67 genes were more highly expressed in placentas from male fetuses (Fig. 1). At the FDR of 0.05, we expect 3.75 and 3.35 genes to be false positives in female and male groups, respectively. In the female group, 55 up-regulated genes were autosomal and 20 were X-linked. Of genes significantly up-regulated in the male group, 33 genes were expressed from the autosomes, 16 were expressed from the X chromosome and 18 were Y chromosome genes (Fig. 2). We do not consider the Y-linked genes to be differentially expressed; rather these genes are expressed at consistently detectable levels in placentas from male fetuses, and therefore may potentially influence placental phenotype.

The majority of sex-biased genes were autosomal but, as expected, many were located on the sex chromosomes. The X-linked and autosomal genes with the highest level of significance were *HDHD1* and *CGB*, respectively (Fig. 1). When inspecting the contribution of individual studies for autosomal gene expression bias, despite there being a lower magnitude of difference, the direction of change was consistent across datasets for many male and female-biased genes (Fig. 3). The results for all genes, the number of studies where they were measurable and the expression differences with statistics are provided in Supplementary data, File 1.

When comparing these results with previous studies where sex-biased expression has been assessed in other human tissues, genes

Table I Details of each dataset included in the meta-analysis.

Dataset	GEO accession	Array manufacturer	Array platform	No. of samples	Male	Female
1	GSE10588 (Sitras *et al.*, 2009a)	Applied Biosystems	Human Genome Survey v2	21	14	7
2	GSE12216 (Sitras *et al.*, 2009b)	Applied Biosystems	Human Genome Survey v2	8	5	3
3	GSE18809 (Chim *et al.*, 2012)	Affymetrix	U133 plus 2	9	3	6
4	GSE24129 (Nishizawa *et al.*, 2011)	Affymetrix	Human Genome 1 ST	8	5	3
5	GSE25906 (Tsai *et al.*, 2011)	Illumina	Human-6 v2	37	21	16
6	GSE27272 (Votavova *et al.*, 2011)	Illumina	HumanRef-8 v3	51	32	19
7	GSE28551 (Sitras *et al.*, 2012)	Applied Biosystems	Human Genome Survey v2	20	14	6
8	GSE30032 (Votavova *et al.*, 2012)	Illumina	HumanRef-8 v3	54	26	28
9	GSE35574 (Guo *et al.*, 2013)	Illumina	Human-6 v2	40	23	17
10	GSE36828 (Unpublished)	Illumina	HumanHT-12 v3	47	26	21
11	GSE7434 (Huuskonen *et al.*, 2008)	Affymetrix	U133 plus 2	8	5	3
			Total	303	174	129

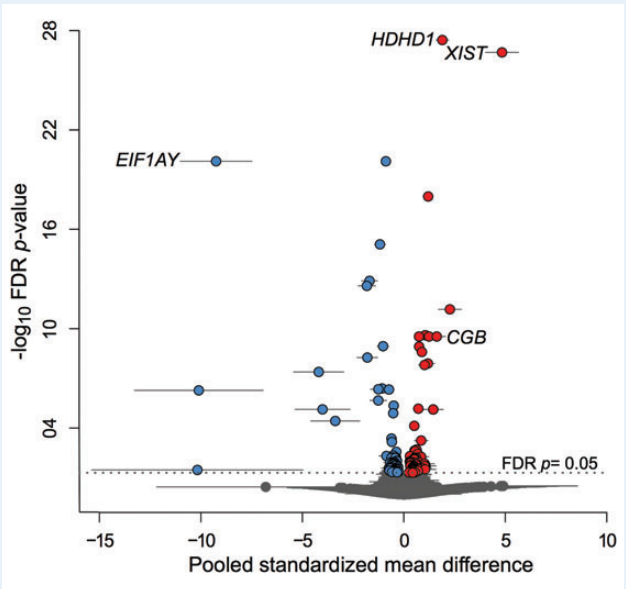

Figure 1 Volcano plot showing pooled effect size and false discovery rate (FDR) *P*-values for 31 844 Ensembl genes when comparing sex-biased gene expression in the human placenta. Blue dots represent genes with significant male-biased expression and red dots are those with significant female-biased expression. Horizontal bars indicate the 95% confidence interval. Points represent genes detected in at least three studies.

showing sex-biased expression appear to exhibit that bias with a high degree of tissue specificity. A vast majority of sex-biased genes in the human placenta are not observed to have sex-biased expression in human brain, liver or blood (Whitney *et al.*, 2003; Zhang *et al.*, 2011; Trabzuni *et al.*, 2013) (Supplementary data, Fig. S1A). When comparing our results to studies where sex-biased expression was assessed in placental tissue or cells, many of the genes in this study have no previously reported sex expression bias (Sood *et al.*, 2006; Cvitic *et al.*, 2013) (Supplementary data, Fig. S1B).

Identification of potential transcriptional regulators of sex-biased gene expression

To predict TFs that may be involved in regulating sex-biased gene expression, we searched for conserved TFBSs in the 10 kb of DNA sequence up and downstream of the transcription start sites of sex-biased genes. This was done using oPOSSUM-3 and the JASPAR core motifs (Kwon *et al.*, 2012; Mathelier *et al.*, 2013). This analysis identified potential binding sites for 166 vertebrate TFs.

Since the results of this analysis are best interpreted using relative rankings (Kwon *et al.*, 2012), we selected the TFs that ranked in the upper quartiles of both z-scores and Fisher scores (Supplementary data, Fig. S2), which limited the initial list to 14 TFs (Supplementary data, Table SI). In order to further investigate whether these TFs may be involved in regulating sex-biased gene expression, we checked if the genes encoding these TFs were expressed at detectable levels in the human term placenta using publicly available RNA-Seq data (Kim *et al.*, 2012). In this comparison, expression data for nine of these TFs were available, with seven being expressed at detectable levels and comparable with human adult tissues expression (Fig. 4A and Supplementary data, Fig. S3).

Expression of *MYCN* is highest in placental tissue when compared with any of the adult tissues (Supplementary data, Fig. S3); this is also the gene that encodes the TF of highest significance in the TF-binding motif analysis (Supplementary data, Table SI). *NKX3-1*, which encodes a homeobox-containing TF, showed significant female expression bias in this meta-analysis and significant enrichment in the TF-binding motif analysis. *NKX3-1* expression in the placenta is detectable and comparable with a majority of other adult tissues (Supplementary data, Fig. S3). *NKX3-1* is a tumor suppressor and its expression appears to be strictly regulated by androgens and loss of its expression is associated with prostate cancer development (Meeks and Schaeffer, 2011). This suggests the *NKX3-1* female expression bias observed in this study may be due to different androgen profiles in male and female placentas, which in turn may drive sex differences in the transcription of the numerous *NKX3-1* target genes.

RXRA, which encodes a hypoxic responsive hormone receptor and TF, showed a consistent male expression bias in this study. Although falling just below our cut-off criteria for enriched TF-binding sites, the binding

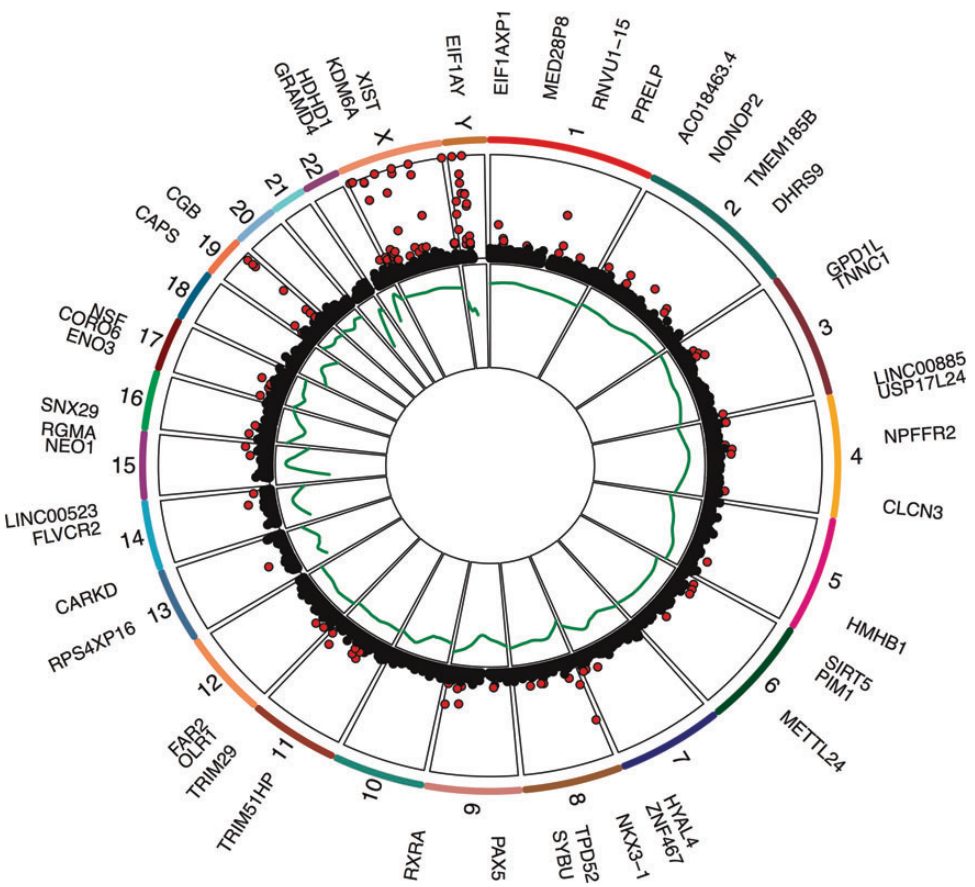

Figure 2 Circos plot summarizing the meta-analysis of sex-biased gene expression in the human placenta. The outer-most scatterplot track depicts chromosomal location and FDR *P*-values). The red points represent genes with FDR *P*-values <0.05. The closer points are to the outside of the track, the higher the significance. The inner track is a loess smoothed line plot representing the number of datasets where information was available for each genomic region, ranging from 3 to 11 datasets. Gene labels for selected genes of significance are plotted outside the chromosome highlights. Circos plot was generated using an R implementation of Circos (Krzywinski *et al.*, 2009; Zhang *et al.*, 2013).

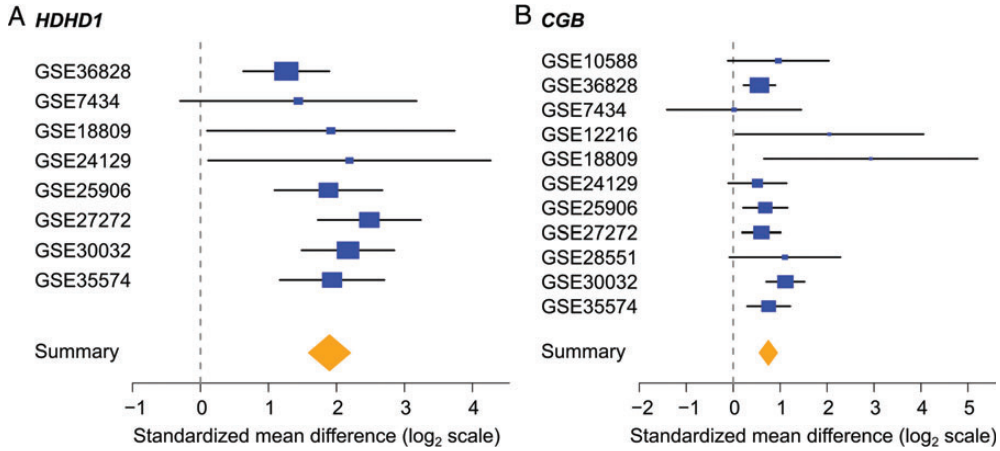

Figure 3 Forest plots showing the standardized mean difference between males and females for the most statistically significant X-linked gene *HDHD1* (**A**) and autosomal gene *CGB* (**B**). Size of the blue box for each study is proportional to sample size; horizontal lines represent standard error. Yellow diamond represents the gene summary across all studies where the gene was detectable. GEO accession identifiers on the *y*-axis represent datasets.

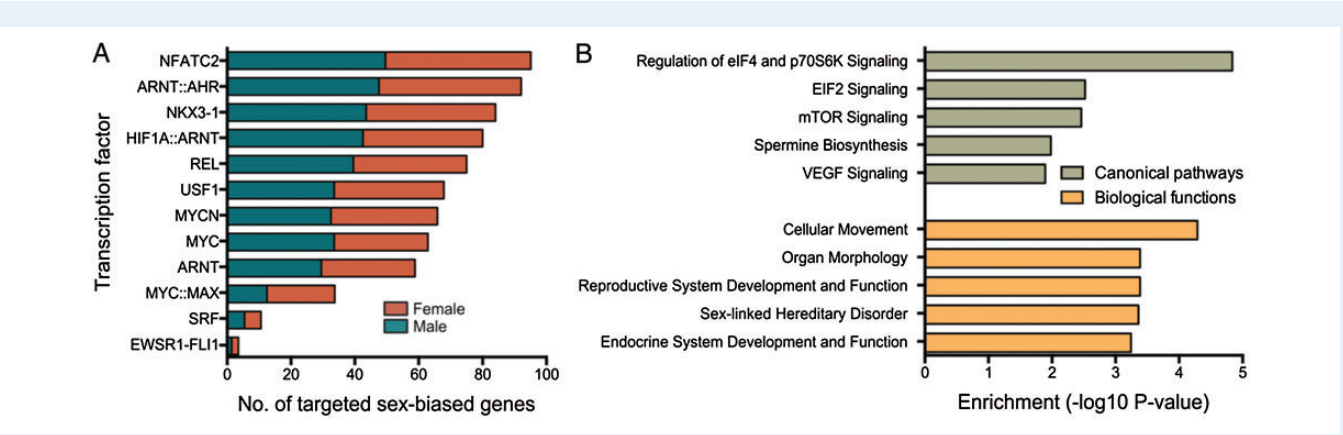

Figure 4 (**A**) Transcription factors expressed in the human placenta that show enriched binding domains surrounding genes with sex-biased expression. (**B**) Top biological functions and canonical pathways associated with sex-biased gene expression in the human placenta. Functions and pathways were determined using Ingenuity Pathway Analysis.

of RXRA in four different complexes with other proteins was detected in the enrichment analysis (Supplementary data, Fig. S2). In the mouse, *RXRA* knockout placentas exhibit multiple defects, and RXRA antagonists are known to be involved in stimulating human chorionic gonadotrophin (hCG) production through interaction with *CGB* gene promoters (Barak *et al.*, 2008) (see results below). *RXRA* was also identified as a target of MYCN in the TF-binding motif analysis, suggesting the *RXRA*-encoded TF may be a significant player in defining the sex differences in gene transcription and placental function.

High-level molecular functions and pathways are associated with sex-biased genes

Since pathway analysis is a valuable tool in estimating gene function in different tissues and systems, we applied the list of sex-biased genes to search for molecular pathways and processes statistically enriched with sex-biased genes. Ingenuity Pathway Analysis showed that sex-biased genes are involved with high-level functions such as cellular movement, organ morphology and endocrine function (Fig. 4B). Among the top five canonical pathways associated with sex-biased genes were mTOR and VEGF signaling (Fig. 4B). The mTOR signaling pathway is a key regulator of cell growth and proliferation, and is activated during angiogenesis (Laplante and Sabatini, 2012). The VEGF pathway involves many genes implicated in angiogenesis, placental development and adverse pregnancy outcomes (Andraweera *et al.*, 2012). These suggest that sex-biased expression of genes involved in these placental development pathways could potentially drive differential function of pathways involved in other key placental processes such as establishing the vascular architecture (angiogenesis) and the proliferation of placental cells.

The list of sex-biased genes was also enriched for genes involved in eIF2 and eIF4 signaling pathways (involving several X-linked genes), which are chiefly involved in regulating protein translation. Taken together, sex-biased genes appear to be involved in numerous high-level regulatory processes that could have a multi-factorial influence on developmental processes contributing to sex differences in placental function and hence fetal wellbeing.

Sex-biased expression of X-linked genes

In female mammals, one of the two X chromosomes is typically inactivated to compensate for gene dosage differences between the sexes (for review see refs Lee, 2011; Augui *et al.*, 2011). However, some genes escape X-inactivation (XCI) and are expressed from both X chromosomes in females. Subsequently, those genes that escape XCI potentially contribute to sexually dimorphic traits.

Numerous studies have measured escape from XCI in human cells and tissues, although the extent of escape from XCI in extra-embryonic tissues, including the human placenta, remains controversial (Buckberry *et al.*, 2014b). We detected 20 X-linked genes with significant female-biased expression, many of which appear to cluster in distinct chromosomal regions (Fig. 5). The most significant of these genes was *HDHD1* (Fig. 3A), which encodes a phosphatase involved in the dephosphorylation of modified RNA nucleotides (Preumont *et al.*, 2010). Additionally, the long non-coding RNAs *XIST* and *JPX*, which are known to be involved in the mechanisms giving rise to XCI (Augui *et al.*, 2011; Lee, 2011) also showed significant female expression bias, as expected.

To assess whether escape from XCI may be the underlying cause of X-linked gene expression bias in this study, we compared our results with a previously published extensive profile of human XCI (Carrel and Willard, 2005) (Fig. 5). Of the 20 X-linked genes with female expression bias, XCI profiling information was available for 16, of which 11 had strong evidence of expression from the inactive X chromosome (Carrel and Willard, 2005). This suggests that this is most likely to be the primary cause of X-linked female expression bias. The female-biased X-linked genes are associated with several biological functions, including conversion of sulfated steroid precursors to estrogens (*STS*) and histone demethylation (*KDM6A*).

Additionally, we observed clusters of X-linked genes with male expression bias including five genes at the pseudo-autosomal Xp22.33 region and two genes at Xq22.1 in the ARMCX family (Fig. 5). The ARMCX3 and ARMCX6 genes are thought to have originated during the evolution of placental mammals and are known to be involved in mitochondrial regulation (López-Doménech *et al.*, 2012).

Taken together, the X-linked genes comprise a considerable proportion of highly significant sex-biased genes detected in this study, and have biological functions relating to hormone regulation, and higher order regulatory

mechanisms such as RNA modification and histone methylation. Given that the sex chromosomes define the difference between the sexes at a cellular level, sex chromosome genes with expression biases are clearly potential drivers or regulators of sex-biased autosomal gene expression.

LHB-CGB cluster genes show female expression bias

Among the sex-biased autosomal genes, the LHB-CGB cluster of seven genes on chromosome 19 showed the most significant female expression bias (Fig. 6). This contiguous cluster consists of the *LHB* gene that encodes the beta-subunit of luteinizing hormone (LH), four hCG beta-subunit coding genes (*CGB*, *CGB5*, *CGB7* and *CGB8*) and two pseudogenes (*CGB1* and *CGB2*) (Liina Nagirnaja, 2010). The four *CGB* genes encoding hCG beta-subunits can be grouped into two classes based on protein sequence: *CGB*, *CGB5* and *CGB8* encode identical amino acid sequences, while *CGB7* encodes a variant peptide (Fig. 6 and Supplementary data, Fig. S4).

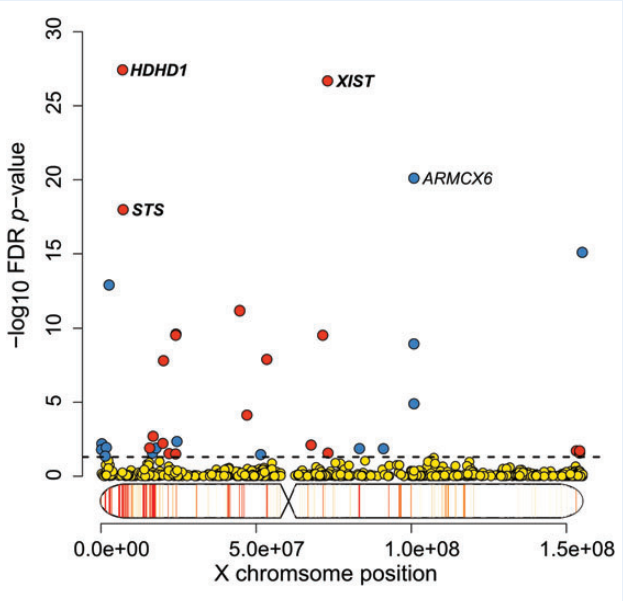

Figure 5 Sex-biased expression of X-linked genes in the human placenta. Red points indicate genes showing significantly higher expression in female samples and blue points represent genes expressed significantly higher in male samples. Heat map below plot represents the level of expression from the inactive X chromosome observed in Carrel and Willard (2005).

The *CGB*-encoded hCG hormone is the important embryonic signal for maternal recognition of pregnancy in primates. Indeed, it is essential for the prolongation of corpus luteal function and hence progesterone synthesis until the placenta takes over. The many functions of hCG relating to placental growth, invasion, angiogenesis and the regulation of maternal immune tolerance of the placenta and fetus are well described (for reviews see Norris *et al.*, 2011; Bansal *et al.*, 2012). Lowered *CGB* expression in the placenta has also been observed in miscarriages, and is higher in ectopic, molar and growth-restricted pregnancies (McCarthy *et al.*, 2007; Rull *et al.*, 2008). The *LHB*-encoded LH is primarily expressed in the pituitary gland, and is widely known for its action in the gonads to induce sex steroid synthesis and gametogenesis (see Henke and Gromoll, 2008). However, *LHB* is also expressed at appreciable levels in the human placenta (Supplementary data, Fig. S5), which is a feature that appears to be conserved across therian mammals (Menzies *et al.*, 2011).

Both hCG and LH hormones bind to the same transmembrane receptor (LHCGR), which is known to induce multiple signals including cyclic adenosine monophosphate (Ryu *et al.*, 1998). hCG is also known to regulate *VEGF* and its receptors (Brouillet *et al.*, 2012), which are heavily implicated in placental development and adverse pregnancy outcomes (Andraweera *et al.*, 2012). Meta-analysis profiling of placental gene expression in pre-eclampsia indicates that up-regulation of *LHB* contributes to the gene expression signature of pre-eclampsia (Kleinrouweler *et al.*, 2013), while up-regulation of *LHB* and *CGB* in the placenta are associated with intrauterine growth restriction (McCarthy *et al.*, 2007).

Taken together, these results provide substantial evidence for female-biased expression of the hormone-coding *LHB* and *CGB* genes in the human placenta. This suggests that, through the actions of *LHB* and *CGB* genes, female fetuses may invest more in placental growth and

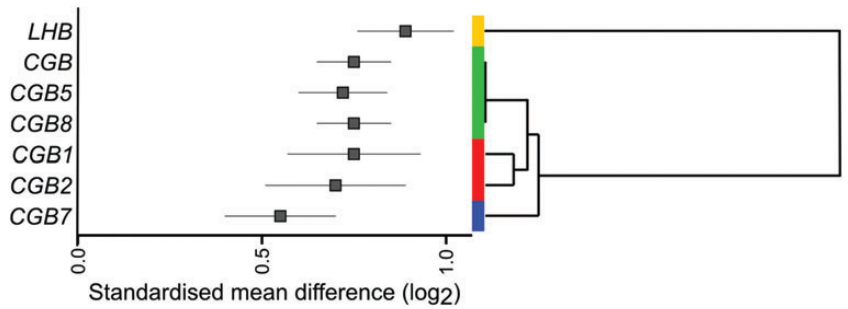

Figure 6 Female-biased expression of *LHB* and *CGB* cluster genes. Boxes represent mean expression difference for each gene and bars represent 95% confidence interval. FDR *P*-value <0.05 for all genes. Dendrogram shows the average distance between genes using percent identity of the amino acid sequences, and branch lengths represent the percentage mismatch between two nodes. Colored bars depict functional groupings; yellow represents *LHB*, green represents identical *CGB* protein isoforms, red represents pseudogenes and blue represents a divergent CGB isoform. Note that *CGB*, *CGB5* and *CGB8* have identical amino acid sequences.

vasculogenesis, while males invest these resources in body growth. Indeed, the ratio of birthweight to placental weight in male human infants is higher than for females (Edwards *et al.*, 2000), suggesting that to maintain a high growth rate the male fetus extracts maximal nutrients from the placenta with little reserve capacity if adversity strikes. Perturbed expression of *LHB* and *CGB* is also associated with pre-eclampsia (Kleinrouweler *et al.*, 2013) and intrauterine growth restriction (McCarthy *et al.*, 2007) where placental pathology is implicated. This indicates that fetal sex-specific risks for these conditions could be partially attributable to differential regulation of gene networks involving these genes.

Conclusions

In this study, we have characterized the gene expression profiles of human male and female placentas from non-pathological term pregnancies. Using an integrative meta-analytical approach, we show that sex-biased gene expression is genome wide, with many genes showing sex-biased expression patterns not observed in other human tissues.

Female-biased expression of X-linked genes appears largely to be the result of escape from XCI, including genes with high-level regulatory functions. As the mechanisms regulating X chromosome regulation are non-hormonal, this is a clear demonstration of sex-biased gene expression that is not directly regulated by the sex hormones.

The results presented here also demonstrate sex-biased expression for many autosomal genes, including genes encoding the LH and hCG hormones. Given that LH and hCG have a potent ability in promoting placental growth and vasculogenesis, these results suggest that female fetuses invest more in extra-embryonic tissue development than males. Since mothers can allocate limited resources to a fetus *in utero*, these findings support the conjecture that males invest more resources in body growth and development (embryonic tissues) at the expense of investing less in the development of extra-embryonic tissues (Clifton, 2010; Eriksson *et al.*, 2010). This may be a key reason as to why there is a male bias in the incidence of placental dysfunction (Murji *et al.*, 2012) and in pregnancy complications where placental pathology is implicated (Vatten and Skjaerven, 2004; Di Renzo *et al.*, 2007; Kleinrouweler *et al.*, 2013).

This study has extended current knowledge surrounding sex-biased gene expression in the human placenta. Having observed widespread sex-biased gene expression in non-pathological tissues, and that the influence of sex is not always considered in gene expression studies, these results highlight the importance of the effect of sex in understanding the natural, sex-based gene expression differences in normal and pathological tissues. This consideration is crucial to begin elucidating the factors that may contribute to the etiology of developmental and chronic adult-onset diseases in which sex biases exist both in terms of incidence and severity.

Supplementary data

Supplementary data are available at http://molehr.oxfordjournals.org/.

Acknowledgements

We wish to thank all of the individuals involved in collecting the primary data used in this meta-analysis and for making it publicly available.

Authors' roles

S.B. conceived and designed the study, carried out all bioinformatics and statistical analyses and wrote the manuscript. T.B.M. and S.J.B. were involved in study design, critical discussion and commented on the manuscript. G.A.D. provided critical discussion and commented on the manuscript. C.T.R. was involved in study design, critical discussion and wrote the manuscript.

Funding

S.B. is supported by a Healthy Development Adelaide and Channel 7 Children's Research Foundation PhD Scholarship and an Australian Postgraduate Award. T.B.-M. is supported by the Cancer Council SA and SAHMRI Beat Cancer Project (TBM APP1030945). C.T.R. is supported by a National Health and Medical Research Council (NHMRC) Senior Research Fellowship APP1020749 (http://www.nhmrc.gov.au). This project was funded in part by NHMRC Project APP1059120 awarded to C.T.R., T.B.-M. and S.J.B. The funders had no role in study design, analysis, decision to publish or preparation of the manuscript. Funding to pay the Open Access publication charges for this article was provided by NHMRC Project APP1059120.

Conflict of interest

None of the authors has any conflicts of interest to declare.

References

Andraweera PH, Dekker GA, Roberts CT. The vascular endothelial growth factor family in adverse pregnancy outcomes. *Hum Reprod Update* 2012; **18**:436–457.

Augui S, Nora EP, Heard E. Regulation of X-chromosome inactivation by the X-inactivation centre. *Nat Rev Genet* 2011;**12**:429–442.

Bansal AS, Bora SA, Saso S, Smith JR, Johnson MR, Thum M-Y. Mechanism of human chorionic gonadotrophin-mediated immunomodulation in pregnancy. *Expert Rev Clin Immunol* 2012;**8**:747–753.

Barak Y, Sadovsky Y, Shalom-Barak T. PPAR signaling in placental development and function. *PPAR Res* 2008;**2008**:142082.

Barker DJP. Fetal programming of coronary heart disease. *Trends Endocrinol Metab* 2002;**13**:364–368.

Benson DA, Karsch-Mizrachi I, Clark K, Lipman DJ, Ostell J, Sayers EW. GenBank. *Nucleic Acids Res* 2011;**40**:D48–D53.

Brouillet S, Hoffmann P, Chauvet S, Salomon A, Chamboredon S, Sergent F, Benharouga M, Feige JJ, Alfaidy N. Revisiting the role of hCG: new regulation of the angiogenic factor EG-VEGF and its receptors. *Cell Mol Life Sci* 2012;**69**:1537–1550.

Buckberry S, Bent SJ, Bianco-Miotto T, Roberts CT. massiR: a method for predicting the sex of samples in gene expression microarray datasets. *Bioinformatics*. 2014a; doi:10.1093/bioinformatics/btu161.

Buckberry S, Bianco-Miotto T, Roberts CT. Imprinted and X-linked non-coding RNAs as potential regulators of human placental function. *Epigenetics* 2014b;**9**:81–89.

Carrel L, Willard HF. X-inactivation profile reveals extensive variability in X-linked gene expression in females. *Nature* 2005;**434**:400–404.

Chim SSC, Lee WS, Ting YH, Chan OK, Lee SWY, Leung TY. Systematic identification of spontaneous preterm birth-associated RNA transcripts in maternal plasma. *PLoS One* 2012;**7**:e34328.

Clifton VL. Review: sex and the human placenta: mediating differential strategies of fetal growth and survival. *Placenta* 2010;**31**(Suppl.):S33–S39.

Cvitic S, Longtine MS, Hackl H, Wagner K, Nelson MD, Desoye G, Hiden U. The human placental sexome differs between trophoblast epithelium and villous vessel endothelium. *PLoS One* 2013;**8**:e79233.

Di Renzo G, Rosati A, Sarti R, Cruciani L. Does fetal sex affect pregnancy outcome? *Gend Med* 2007;**4**:19–30.

Dunning MJ, Smith ML, Ritchie ME, Tavaré S. beadarray: R classes and methods for Illumina bead-based data. *Bioinformatics* 2007; **23**:2183–2184.

Edwards A, Megens A, Peek M. Sexual origins of placental dysfunction. *Lancet* 2000;**355**:203–204.

Eriksson JG, Kajantie E, Osmond C, Thornburg K, Barker DJP. Boys live dangerously in the womb. *Am J Hum Biol* 2010;**22**:330–335.

Flicek P, Ahmed I, Amode MR, Barrell D, Beal K, Brent S, Carvalho-Silva D, Clapham P, Coates G, Fairley S *et al.* Ensembl 2013. *Nucleic Acids Res* 2012;**41**:D48–D55.

Forsén T, Eriksson JG, Tuomilehto J, Osmond C, Barker DJ. Growth in utero and during childhood among women who develop coronary heart disease: longitudinal study. *BMJ* 1999;**319**:1403–1407.

Gray KA, Daugherty LC, Gordon SM, Seal RL, Wright MW, Bruford EA. Genenames.org: the HGNC resources in 2013. *Nucleic Acids Res* 2012; **41**:D545–D552.

Guo L, Tsai SQ, Hardison NE, James AH, Motsinger-Reif AA, Thames B, Stone EA, Deng C, Piedrahita JA. Differentially expressed microRNAs and affected biological pathways revealed by modulated modularity clustering (MMC) analysis of human preeclamptic and IUGR placentas. *Placenta* 2013;**34**:599–605.

Hadar E, Melamed N, Sharon-Weiner M, Hazan S, Rabinerson D, Glezerman M, Yogev Y. The association between stillbirth and fetal gender. *J Matern Fetal Neonatal Med* 2012;**25**:158–161.

Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F, Aken BL, Barrell D, Zadissa A, Searle S *et al.* GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res* 2012;**22**:1760–1774.

Henke A, Gromoll J. New insights into the evolution of chorionic gonadotrophin. *Mol Cell Endocrinol* 2008;**291**:11–19.

Huuskonen P, Storvik M, Reinisalo M, Honkakoski P, Rysä J, Hakkola J, Pasanen M. Microarray analysis of the global alterations in the gene expression in the placentas from cigarette-smoking mothers. *Clin Pharmacol Ther* 2008;**83**:542–550.

Ingemarsson I. Gender aspects of preterm birth. *Br J Obstet Gynaecol* 2003; **110**:34–38.

Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* 2007;**8**:118–127.

Kauffmann A, Gentleman R, Huber W. arrayQualityMetrics—a bioconductor package for quality assessment of microarray data. *Bioinformatics* 2009;**25**:415–416.

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

Kleinrouweler CE, van Uitert M, Moerland PD, Ris-Stalpers C, van der Post JAM, Afink GB. Differentially expressed genes in the pre-eclamptic placenta: a systematic review and meta-analysis. *PLoS One* 2013;**8**:e68991.

Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. Circos: an information aesthetic for comparative genomics. *Genome Res* 2009;**19**:1639–1645.

Kwon AT, Arenillas DJ, Hunt RW, Wasserman WW. oPOSSUM-3: advanced analysis of regulatory motif over-representation across genes or ChIP-Seq datasets 2012;G3 2 : 987–1002.

Lao TT, Sahota DS, Suen SSH, Law LW. The impact of fetal gender on preterm birth in a southern Chinese population. *J Matern Fetal Neonatal Med* 2011;**24**:1440–1443.

Laplante M, Sabatini DM. mTOR signaling. *Cold Spring Harbor Perspect Biol* 2012;**4**. doi:10.1101/cshperspect.a011593.

Lee JT. Gracefully ageing at 50, X-chromosome inactivation becomes a paradigm for RNA and chromatin control. *Nat Rev Mol Cell Biol* 2011; **12**:815–826.

Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* 2012;**28**:882–883.

Liina Nagirnaja Krluphmgml. Genomics and genetics of gonadotropin beta-subunit genes: unique FSHB and duplicated LHB/CGB loci. *Mol Cell Endocrinol* 2010;**329**:4–16.

López-Doménech G, Serrat R, Mirra S, D'Aniello S, Somorjai I, Abad A, Vitureira N, García-Arumí E, Alonso MT, Rodriguez-Prados M *et al.* The Eutherian Armcx genes regulate mitochondrial trafficking in neurons and interact with Miro and Trak2. *Nat Commun* 2012;**3**:814.

Mathelier A, Zhao X, Zhang AW, Parcy F, Worsley-Hunt R, Arenillas DJ, Buchman S, Chen CY, Chou A, Ienasescu H *et al.* JASPAR 2014: an extensively expanded and updated open-access database of transcription factor binding profiles. *Nucleic Acids Res* 2013;**42**:D142–D147.

McCarthy C, Cotter FE, McElwaine S, Twomey A, Mooney EE, Ryan F, Vaughan J. Altered gene expression patterns in intrauterine growth restriction: potential role of hypoxia. *Am J Obstet Gynecol* 2007;**196**:70.e1–76.

Meeks JJ, Schaeffer EM. Genetic regulation of prostate development. *J Androl* 2011;**32**:210–217.

Menzies BR, Pask AJ, Renfree MB. Placental expression of pituitary hormones is an ancestral feature of therian mammals. *Evodevo* 2011;**2**:16.

Misra DP, Salafia CM, Miller RK, Charles AK. Non-linear and gender-specific relationships among placental growth measures and the fetoplacental weight ratio. *Placenta* 2009;**30**:1052–1057.

Murji A, Proctor LK, Paterson AD, Chitayat D, Weksberg R, Kingdom J. Male sex bias in placental dysfunction. *Am J Med Genet* 2012;**158A**:779–783.

Nishizawa H, Ota S, Suzuki M, Kato T, Sekiya T, Kurahashi H, Udagawa Y. Comparative gene expression profiling of placentas from patients with severe pre-eclampsia and unexplained fetal growth restriction. *Reprod Biol Endocrinol* 2011;**9**:107.

Norris W, Nevers T, Sharma S, Kalkunte S. Review: hCG, preeclampsia and regulatory T cells. *Placenta* 2011;**32**(Suppl. 2):S182–S185.

Preumont A, Rzem R, Vertommen D, Van Schaftingen E. HDHD1, which is often deleted in X-linked ichthyosis, encodes a pseudouridine-5′-phosphatase. *Biochem J* 2010;**431**:237–244.

Rainer J, Sanchez-Cabo F, Stocker G, Sturn A, Trajanoski Z. CARMAweb: comprehensive R- and bioconductor-based web service for microarray data analysis. *Nucleic Acids Res* 2006;**34**:W498–W503.

Ramasamy A, Mondry A, Holmes CC, Altman DG. Key issues in conducting a meta-analysis of gene expression microarray datasets. *PLoS Med* 2008;**5**:e184.

Rull K, Hallast P, Uusküla L, Jackson J, Punab M, Salumets A, Campbell RK, Laan M. Fine-scale quantification of HCG beta gene transcription in human trophoblastic and non-malignant non-trophoblastic tissues. *Mol Hum Reprod* 2008;**14**:23–31.

Ryu KS, Gilchrist RL, Koo YB, Ji I, Ji TH. Gene, interaction, signal generation, signal divergence and signal transduction of the LH/CG receptor. *Int J Gynecol Obstet* 1998;**60**:S9–S20.

Schroeder DI, Blair JD, Lott P, Yu HOK, Hong D, Crary F, Ashwood P, Walker C, Korf I, Robinson WP *et al.* The human placenta methylome. *Proc Natl Acad Sci* 2013;**110**:6037–6042.

Sitras V, Paulssen R, Leirvik J, Vårtun A, Acharya G. Placental gene expression profile in intrauterine growth restriction due to placental insufficiency. *Reprod Sci* 2009a;**16**:701–711.

Sitras V, Paulssen RH, Grønaas H, Leirvik J, Hanssen TA, Vårtun A, Acharya G. Differential placental gene expression in severe preeclampsia. *Placenta* 2009b;**30**:424–433.

Sitras V, Fenton C, Paulssen R, Vårtun A, Acharya G. Differences in gene expression between first and third trimester human placenta: a microarray study. *PLoS One* 2012;**7**:e33294.

Slater GSC, Birney E. Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics* 2005;**6**:31.

Sood R, Zehnder JL, Druzin ML, Brown PO. Gene expression patterns in human placenta. *Proc Natl Acad Sci* 2006;**103**:5243–5244.

Trabzuni D, Ramasamy A, Imran S, Walker R, Smith C, Weale ME, Hardy J, Ryten M, North American Brain Expression Consortium. Widespread sex differences in gene expression and splicing in the adult human brain. *Nat Commun* 2013;**4**:2771.

Tsai S, Hardison NE, James AH, Motsinger-Reif AA, Bischoff SR, Thames BH, Piedrahita JA. Transcriptional profiling of human placentas from pregnancies complicated by preeclampsia reveals disregulation of sialic acid acetylesterase and immune signalling pathways. *Placenta* 2011; **32**:175–182.

Vatten LJ, Skjaerven R. Offspring sex and pregnancy outcome by length of gestation. *Early Hum Dev* 2004;**76**:47–54.

Votavova H, Dostalova Merkerova M, Fejglova K, Vasikova A, Krejcik Z, Pastorkova A, Tabashidze N, Topinka J, Veleminsky M, Sram RJ *et al*. Transcriptome alterations in maternal and fetal cells induced by tobacco smoke. *Placenta* 2011;**32**:763–770.

Votavova H, Dostalova Merkerova M, Krejcik Z, Fejglova K, Vasikova A, Pastorkova A, Tabashidze N, Topinka J, Balascak I, Sram RJ *et al*. Deregulation of gene expression induced by environmental tobacco smoke exposure in pregnancy. *Nicotine Tob Res* 2012;**14**:1073–1082.

Whitney AR, Diehn M, Popper SJ, Alizadeh AA, Boldrick JC, Relman DA, Brown PO. Individuality and variation in gene expression patterns in human blood. *Proc Natl Acad Sci USA* 2003;**100**:1896–1901.

Wilson CL, Miller CJ. Simpleaffy: a BioConductor package for Affymetrix Quality Control and data analysis. *Bioinformatics* 2005;**21**:3683–3685.

Zhang Y, Klein K, Sugathan A, Nassery N, Dombkowski A, Zanger UM, Waxman DJ. Transcriptional profiling of human liver identifies sex-biased genes associated with polygenic dyslipidemia and coronary artery disease. *PLoS One* 2011;**6**:e23506.

Zhang H, Meltzer P, Davis S. RCircos: an R package for Circos 2D track plots. *BMC Bioinformatics* 2013;**14**:244.