MONITORING PLACENTAL DEVELOPMENT AND FUNCTION IN THE AGE OF MULTI-OMIC HIGH THROUGHPUT SEQUENCING

Melanie Smith

A thesis submitted for the degree of Doctor of Philosophy

School of Women's and Children's Health
Adelaide Medical School
The University of Adelaide
February 2022

Table of Contents

Papers Arising from this Thesis

1. **Melanie D. Smith**, Katherine Pillman, Tanja Jankovic-Karasoulos, DaleMcAninch, Qianhui Wan, K. Justinian Bogias, Dylan McCullough, Tina Bianco-Miotto, James Breen & Claire T. Roberts (2021): Large-scale transcriptome-wide profiling of microRNAs in human placenta and maternal plasma at early to mid gestation, *RNA Biology*, DOI:10.1080/15476286.2021.1963105

Publications Associated with this Thesis

1. Qianhui Wan, Shalem Yiner-Lee Leemaqz, Stephen M. Pederson, Dylan McCullough, Dale Christopher McAninch, Tanja Jankovic-Karasoulos, **Melanie D. Smith**, K. Justinian Bogias, James Breen, Claire T. Roberts, Tina Bianco-Miotto (2019): Quality Control Measures for Placental Sample Purity in DNA Methylation Array Analyses. *Placenta* 88.
   Author Contribution 10%
2. James Breen, Dale McAninch, Tanja Jankovic-Karasoulos, Dylan McCullough, **Melanie D. Smith**, K. Justinian Bogias, Qianhui Wan, Awais Choudhry, Nhi Hin, Stephen M. Pederson, Tina Bianco-Miotto, Claire T. Roberts (2020): Temporal placental genome wide expression profiles reflect three phases of utero-placental blood flow during early to mid human gestation. *bioRxiv*.
   Author Contribution 10%

Abstract

Pregnancy is arguably the most dangerous time in life for both mother and child. A healthy pregnancy is reliant on a healthy placenta, the organ that orchestrates both maternal adaptations to pregnancy and the transfer of nutrients, hormones, gases and wastes between the maternal and fetal circulations. Defects in placentation, particularly impaired trophoblast invasion, are implicated in many adverse pregnancy outcomes. Complications of pregnancy such as preeclampsia compromise the health of the mother and her fetus both in the immediate and long term. Access to placental chorionic villus tissue is, understandably, restricted making *in vivo* study of the placenta difficult. However, recent high-throughput sequencing technologies and innovative sampling methods are finally enabling scientists to unravel the molecular components of pregnancy and the placenta in real time. This thesis aims to capitalise on these new technologies to develop a process for monitoring pregnancy at early to mid gestation, and for complications at term. In this thesis I use a data-driven, bioinformatics-led approach to analyses of human placental and maternal transcriptomes, taking advantage of an extensive biobank of tissue samples from elective pregnancy terminations (6-23 weeks' gestation) and post parturition (37+ weeks' gestation).

In the introduction, the human placental transcriptome is explained in the context of normal development and function, pregnancy complications, biomarker discovery and sex-differences that impact gene expression. The literature review in Chapter 2 describes the male- and female-specific sex-differences observed in the human placental transcriptome and technical aspects of high-throughput sequencing

approaches and their implications for analysis. Chapter 3 provides a comprehensive profile of microRNAs in early to mid gestation human placenta (6-23 weeks') and matched maternal plasma, where I identified dynamic changes in placenta and maternal plasma microRNA expression during this time. The next two chapters demonstrate the development of a new method and bioinformatics tool for the identification of red blood cell transcripts in human plasma, an issue identified in Chapter 3. Chapter 4 details a novel 20 microRNA signature set identified as differentially abundant between plasma from samples with and without substantiated haemolysis. This method forms the basis of Chapter 5 where I develop *DraculR*, a Shiny/web based tool for the detection of haemolysis in plasma from young women and recommends that assessment of red blood cell contents becomes standard pre-analytical practice. In Chapter 6, I used both sex-specific and sex-naive bioinformatics approaches to identify the placental gene expression differences between preeclamptic and uncomplicated pregnancies. Network analysis established a profile of gene co-expression associated with late-onset preeclampsia. Finally, I synthesise all research themes to describe potential future applications for this work.

Overall, the analyses herein have reinforced the utility of high throughput sequencing data in pregnancy monitoring and provided a window into the potential benefits that a "Precision Medicine"-type approach could have in obstetrics care. I report major technical biases in using high throughput sequencing technologies for monitoring health, such as haemolysis, batch correction and library preparation inconsistencies.

Quality assessment is essential for the establishment of quality reference data, without which our ability to produce robust reference sets render precision health impossible.

Declaration

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in my name, in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission in my name, for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint-award of this degree.

The author acknowledges that copyright of published works contained within the thesis resides with the copyright holder(s) of those works. I also give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library Search and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

I acknowledge the support I have received for my research through the provision of an Australian Government Research Training Program Scholarship.

Melanie D. Smith Date: 8 February 2022

Acknowledgements

I am particularly indebted to my supervisors **Jimmy Breen**, **Claire Roberts** and **Katherine Pillman** for their generous mentorship, guidance, and support.

Thank you, **Jimmy**, for your confidence in me. I have found our every interaction energising, always coming away with a new sense of possibility. When we met, I couldn't have imagined where I am today. I'm glad I shared the journey with you.

Claire, your passion for women's health is a true force of nature. Thank you for your continued and unwavering support. You helped me to see that learning is a journey, not a destination.

Kat, in you I found someone who never allowed me to get away with anything less than a meticulous statement. For that I am truly grateful. You help me to be a better scientist and a better person.

To my lab mates, **Tanja**, **Shalem, Dylan** and **Anya**, you make it a joy to come to work every day. Thank you for your support and your friendship.

To my dear friends in the Bioinformatics Community, **Dan Kortshak**, **Nhi Hin**, **Ning Liu**, **Jacqueline Rehn**, **Qianhui Wan, Pei Qin, Na Sai** and **Justin Bogias**. Thank you for your support and your inspiration.

Most importantly, to my wonderful Family. **Denise**, it's been a long road - but we made it. I promise never to undertake another PhD. For you, **Alexander**, the little man in our lives, I wish all the success in the world. I hope you can see that with time and persistence, anything is possible.

Finally, to the **Women** who, in what must be their most difficult day, choose to support scientific research through placental donation - thank you.

Funding

This research is supported by NIH NICHD R01 HD089685-01 Maternal molecular profiles reflect placental function and development across gestation PI Roberts, an Australian Government Research Training Program (RTP) Scholarship awarded to MDS, a National Health and Medical Research Council Investigator Grant (GNT1174971) awarded to CTR and a Matthew Flinders Professorial Fellowship awarded to CTR and funded by Flinders University. JB is supported by the James & Diana Ramsay Foundation.

Chapter 1

Introduction

The miracle of the human placenta

The placenta begins to develop in humans shortly after conception when, in search of nutrients, the embryo travels through a fallopian tube to the uterus, before first attaching and then implanting into the decidualised endometrium (uterine lining). Attachment triggers the differentiation of the trophectoderm, the trophoblast cell layer of the blastocyst, into an inner cytotrophoblast and an outer multinucleated syncytiotrophoblast mass [1–3]. Rapidly differentiating, the basic cellular organisation of the placenta is in place approximately 18 days post ovulation, making the placenta the first organ of the conceptus to develop and function [1]. Implantation marks the beginning of placental invasion and is instigated when syncytiotrophoblast processes infiltrate between endometrial lumenal epithelial cells and into the decidua (pregnant endometrium). Invasion by extravillous cytotrophoblast cells that have undergone epithelial mesenchymal transition (EMT) is both interstitial and endovascular. Endovascular extravillous cytotrophoblasts invade, colonise and transform the uterine spiral arterioles, remodelling them and rendering them non-responsive to vasoactive molecules in the maternal circulation. This process provides the maternal blood supply to the placenta with uninterrupted perfusion without turbulent flow. Notably, impairments in this process are associated with pregnancy complications, many of which are associated with placental insufficiency, induced by aberrant placental development [4,5].

Colonisation and remodelling of the maternal spiral arterioles is one dimension of placental direction of the mother’s physiological adaptations to pregnancy. Crucial to pregnancy success, these are managed through a delicate interplay between the placenta’s secretion of a number of steroid and peptide hormones and the physiological response from the maternal system, including the secretion of hormones that affect the placenta. In addition to its role in maternal adaptations, the main functions of the human placenta include gas exchange, metabolic transfer and fetal protection from the maternal immune response. Interaction between the fetal and maternal systems occurs primarily at the terminal branches of the chorionic villi within the intervillous space. Here, oxygenated, nutrient rich maternal blood bathes the chorionic villi bringing the molecules that pass through the syncytium before being taken into the fetal system. In this way, the placenta acts as a barrier between the fetal and maternal systems and can help protect the fetus against xenobiotic molecules, infections and maternal disease [2].

Pregnancy complications and sex differences

Functionally, the placenta is the interface between the developing fetus and the mother. All oxygen and nutrients supplied to the fetus from the maternal circulation and all fetal waste products, including carbon dioxide (CO_2) and urea, must first pass through the placenta via the syncytiotrophoblast. How efficiently this exchange occurs has a profound impact on fetal development and the health of both the mother and the fetus throughout pregnancy. Further, there is evidence from epidemiological studies that strongly suggest there is a “fetal origin” that predisposes adults to chronic diseases including heart disease, type 2 diabetes mellitus and obesity [6,7], indicating that for each child, pregnancy health has a lifelong impact on their future wellbeing. During pregnancy, aberrant placental function has been implicated in pregnancy complications including preeclampsia (PE), intrauterine growth restriction (IUGR) [8], gestational diabetes mellitus (GDM) [9] and spontaneous preterm birth (SPTB) [10]. Whilst the placenta has often been seen as an asexual organ, as a product of conception, genotypically the placenta is identical to the fetus with the same XX or XY chromosomes. Given that the incidence of many pregnancy complications exhibits sexual dimorphism, fetal sex must also be given consideration in the investigation of the placenta and its role in pregnancy complications [11–13].

Whilst it is difficult to identify all factors accounting for sex differences in the occurrence of pregnancy complications, there is strong evidence to suggest that male and female bearing pregnancies respond quite differently to maternal stress. When confronted with a maternal stress challenge for example, the male fetus will continue to grow where the female fetus will attenuate growth [12]. This may be the reason that male fetal sex is associated with preterm birth and female fetal sex associated with fetal growth restriction [14]. The male fetus is at higher risk of many adverse pregnancy outcomes including higher rates of GDM [15,16], PE, preterm premature rupture of membranes, abruptio placentae, polyhydramnios, fetal macrosomia, failure to progress through labor, cord prolapse and umbilical cord knots [13,17,18] and supported by systematic review [11].

Bioinformatic analyses of the human placenta

The obvious link between placental development, pregnancy progression and the ongoing health of both the mother and fetus/child clearly indicate a need for placenta research. However, also clear are the myriad ethical and logistical obstacles in performing this research. The use of animal models is limited given the physiological and anatomical differences between humans and model organisms [19]. Placental cell lines, whilst a valuable and important resource are by definition single-sex [20], may not express important placental genes [21] and cannot substitute for complex tissue. Placental explants, again valuable and important, are limited in that they are not fit for long term observation such as placental development over 40 weeks, the period of human gestation and whilst there has been advancement in the creation of placental organoids, there is much work to be done before they are a viable alternative to whole tissue.

Given the importance of understanding the placenta we must next ask “how” we might begin to untangle the complex stages of organ development. In this thesis, we have concentrated on two sides of genomic profiling, that of gene expression and that of post transcriptional gene regulation. In order to profile gene expression we have used bulk RNA extraction from whole tissue, followed by high throughput sequencing of RNA libraries (RNAseq). RNAseq allows us to take a snapshot of gene expression and to compare this snapshot across samples in time, sex and pregnancy pathology. For post transcriptional regulation we have focused on RNA extraction and size fractionation from whole tissue, followed by high throughput sequencing of the small RNA fraction to profile micro RNAs (miRNAseq). MicroRNAs (miRNAs) are short, ~22 nucleotide lengths of single stranded non-coding RNA that, along with the RISK complex, act to repress or silence the translation of targeted messenger RNAs in the cytoplasm [22–24] and are known to be important in placental development [25–27].

Prior to the introduction of oxygenated maternal blood into the intervillous space at ~10-12 weeks’ gestation, the placenta develops in a low, but physiological, oxygen environment. After this time, maternal/fetal exchange is established as oxygenated maternal blood enters the intervillous space and bathes the villi [4]. How the placenta responds to oxidative stress at this important time point forms the impetus for our investigation. The primary aim of the first study presented in this thesis (Chapter 3) was to quantify the miRNA expression of 96 chorionic villus samples taken from non-medical, elective terminations, between 6-23 weeks' gestation along with that of plasma from matched maternal blood samples taken at the time of termination. To establish the effect of oxygen, we first profiled the miRNA transcripts of all samples passing quality control filters before performing a differential expression between the 6-10 weeks' and 11-23 weeks' gestation groups. The results indicate expression of 637 mature miRNAs, including species from important miRNA clusters and an additional 588 candidate *de novo* miRNA predicted using miRDeep2.

Unsupervised clustering of samples using principal component analysis (PCA) identified a clear gestational age gradient. After establishing a baseline of placental miRNAs from early to mid gestation, we further assessed differential expression between the 6-10 weeks' versus 11-23 weeks' gestation groups and identified 374 differentially expressed miRNAs. Interestingly, we identified the relative proportion of transcripts originating from the highly placenta specific chromosome 19 miRNA cluster (C19MC) and of the chromosome 13 (miR-17~92) and paralogs from chromosome 7 (106b~25) and chromosome X (106a~363) decreased after 10 weeks' gestation, whilst the chromosome 14 miRNA cluster (C14MC) relative proportion increased over this same time period, hinting at a possible biomarker of placental oxygenation. We also note that whilst miRNA clusters have been demonstrated to be transcribed by RNA-Pol II as a single polycistron

[28] individual members of these clusters display a range of expression levels which leads to the question - what is regulating the regulators?

Ectopic oxygenation of the placenta could lead to pregnancy loss, a placenta being in a low oxygen state past the introduction of maternal blood could indicate aberrant placental implantation or development. The promise of miRNAs as biomarkers of placental oxygenation given by both the differential expression and the change in proportion of mature transcripts we identified in tissue, lead naturally to the investigation of miRNA abundance in matched maternal plasma samples (hereafter referred to as plasma). Extraction and quantification of miRNA from plasma using high throughput sequencing is difficult and relatively new. However, using the QIAGEN miRNA serum/plasma kit (Qiagen, Hilden, Germany) our investigation identified 790 mature miRNA species in plasma. miRNA is not expressed in plasma but rather miRNA identified here have their origin linked to various systemic compartments [29]. The exception to this is the highly placenta specific miRNA cluster C19MC. We identified that in plasma, abundance of the C19MC members, miR-516b-5p and miR-517a-3p, appear to increase across early to mid gestation showing some promise of the oxygen signature we hypothesised; however, these results were not statistically significant (FDR > 0.05).

in silico haemolysis detection in human plasma

Early in our analysis of the plasma samples we identified that a number of samples were outliers. Further investigation detected an unusual abundance pattern of known red blood cell (RBC) associated miRNA which prompted the quantification of haemolysis in our plasma samples. Haemolysis occurs when red blood cell contents are released into the blood, usually in this context, due to rupture (lysis) of red blood cells (erythrocytes) during extraction or preparation of blood, prior to the separation, through centrifugation, of plasma and other components of the blood sample. Red blood cells are known repositories of miRNAs in the circulatory system [30] and because of this, haemolysis during sample preparation alters the miRNA content of plasma [31].

Without a method to subtract the RBC miRNA signature from our sample data, we set out to understand the impact of haemolysis on our ongoing experiments. These analyses form the basis for experiments leading to Chapter 4 of this thesis. To establish a miRNA plasma abundance baseline, we first quantified sample haemolysis through delta quantification cycle (ΔCq), where expression levels of a known blood cell associated miRNA (miR-451) and a control miRNA (miR-23a) are determined based on raw Cq values and the difference between the two is calculated. Using both data from the current gold standard haemolysis quantification and the high throughput sequencing data from libraries prepared using 154 plasma samples taken from 24 non-pregnant and 130 pregnant women aged 16 to 46 years, we identified transcripts from 1,133 mature miRNAs.

Circulating cell-free miRNAs have long been investigated as promising biomarkers in body fluids such as plasma, serum, urine and saliva. miRNAs identified in plasma are very stable [32] due to being packaged in micro vesicles such as exosomes [33,34] or bound to protein complexes such as argonaute 2 (Ago2), nucleophosmin 1 (NPM 1) and high density lipoprotein (HDL) [35–37]. In the context of placental health, circulating cell-free miRNAs have the potential to become an important biomarker resource, given their minimally invasive accessibility. However, before we, as a community, establish clinically relevant miRNA biomarkers, we must first understand the impact that red blood cell associated miRNA have on our analyses. Previous research conducted into various cancer biomarkers has established that many miRNA thought to be useful biomarkers have subsequently been identified as associated with haemolysis, casting doubt on their biomarker utility [38].

To quantify the effects of haemolysis on miRNA abundance we first counted the number of unique mature miRNAs identified in each of our samples. Then, using an analysis of variance, we identified a significant difference between the haemolysed and non-haemolysed samples (ANOVA; $p\text{-value} < 0.05$), with haemolysis being associated with fewer mature miRNA species detected at a given read depth. Using data from high throughput sequencing, we were able to establish the effect of haemolysis on miRNA read counts previously demonstrated through qPCR [31,39,40]. To establish which miRNA were being affected by haemolysis we performed a differential expression analysis between the samples identified as haemolysed and non-haemolysed using the ΔCq method described above. After filtering for low abundance, we identified 138 miRNAs with a higher abundance in haemolysed compared to non-haemolysed samples and, using these miRNA, established a plasma miRNA haemolysis signature. Further, adapting concepts from previous RT-qPCR analyses, we used a 20 miRNA haemolysis signature set to establish a new measure of the inclusion of red blood cell associated miRNA in human plasma, the “Haemolysis Metric” and subsequently conceived a tool that could be used for *in silico* detection of haemolysis in data from high throughput sequencing of human plasma. This led to the next major study presented in this thesis (Chapter 5), an online tool available to the research community.

After establishing that our plasma samples were affected by red blood cell contents and that this in turn reduced our power to detect differences between our sample groups, we used plasma samples with known haemolysis quantification (ΔCq method) to establish a new method for identifying haemolysis *in silico*. We next incorporated this new method into a publicly available, Shiny web-based tool, DraculR. Hosted at www.shiny.io, DraculR enables any user to upload miRNA sequencing data from human plasma extraction and after filtering and normalisation, return our Haemolysis Metric and a recommendation regarding the downstream analysis of each of the samples being considered. Further, using *a priori* knowledge of miRNA associated with the user condition of interest, any of the 20 miRNA from the haemolysis signature set can be removed from consideration. This step allows for the quantification of haemolysis without the confounding disease or condition associated with a given miRNA.

Sexual dimorphism in the human preeclamptic placenta

Fetal development *in utero* is completely dependent on adequate placental differentiation and function, a lack of which results in placental insufficiency and pregnancy complications [2,8,41]. Placental development is orchestrated by the placental genome, identical to that of the fetus and the interplay between the placental and maternal environments. Males for example grow faster *in utero* with a greater body length and body weight at birth compared to girls of equivalent placental size [42]. This is often considered to indicate that the placenta from a male bearing pregnancy is more efficient than that of the female bearing pregnancy [12,43,44]. This male growth strategy does however have a trade off. When the maternal environment is favourable, growing quickly may be of benefit, but if the maternal environment is under stress, the male fetus is at increased risk of undernutrition because they have less placental reserve capacity than their female counterparts [12].

Preeclampsia is a pregnancy complication characterised by maternal hypertension and fetal growth restriction [45] with a global incidence rate of 4.6% [46]. Late onset preeclampsia, like many pregnancy complications, has an increased incidence in male bearing pregnancies [47]. Given the sex-differences in outcome, to investigate the transcriptional landscape of the preeclamptic placenta, we conducted sex-specific analyses, the final experiment presented in this thesis (Chapter 6).

We created libraries for bulk RNA-seq and miR-seq from placenta samples taken after delivery from uncomplicated pregnancies with those with late onset preeclampsia. Using data from 59 (22 male, 37 female) samples, we further stratified our analysis by fetal sex. We performed a differential expression analysis to test the null hypotheses, independently for placenta from male and female bearing pregnancies, that there was no difference in gene expression between placenta from preeclamptic and control pregnancies. We identified 6 up- and 2 down-regulated genes in the male comparison, with no significant change in the female comparison. The most highly up-regulated gene in the male comparison was Ceruloplasmin (*CP*), a major plasma antioxidant that chelates transition metal ions of copper and prevents the catalysed production of free radicals in the cell [48]. In the context of pregnancy, oxidative stress leading to endothelial dysfunction is suggested to have a causative role in the pathophysiology of preeclampsia [49].

To complement our differential expression analyses and to more closely investigate the co-expression relationship between placental genes and preeclampsia, we adopted the weighted gene co-expression network analysis (WGCNA) approach. Prior to our investigation 35 additional samples previously sequenced in our lab were added to the analysis. WGCNA analysis revealed both sex-specific and conserved highly correlated patterns of gene expression that are associated with distinct biological processes, maternal smoking, maternal age and genes implicated in preeclampsia.

Summary

The PhD research presented here is focused on three key aspects of pregnancy health and development: (1) placental expression of miRNA post transcriptional regulators across the crucial introduction of oxygenated maternal blood, (2) the effect of haemolysis on biomarker identification and (3) sexual dimorphism in the preeclamptic human placenta. All three unique aspects of this project have provided novel insights into gene regulation, biomarker identification and the importance of fetal sex in the investigation of gene expression in placental pathology. Furthermore, we provide a new method for the *in silico* identification of haemolysis in miRseq from human plasma and provide this tool for open community use.

References

1. Boyd JD, Hamilton WJ. The human placenta. Heffer; 1970.

2. Gude NM, Roberts CT, Kalionis B, King RG. Growth and function of the normal human placenta. Thromb Res. 2004;114: 397–407.

3. Soares MJ, Varberg KM, Iqbal K. Hemochorial placentation: development, function, and adaptations. Biol Reprod. 2018;99: 196–211.

4. Jauniaux E, Watson AL, Hempstock J, Bao YP, Skepper JN, Burton GJ. Onset of maternal arterial blood flow and placental oxidative stress: A possible factor in human early pregnancy failure. Am J Pathol. 2000;157: 2111–2122.

5. Ilekis JV, Tsilou E, Fisher S, Abrahams VM, Soares MJ, Cross JC, *et al.* Placental origins of adverse pregnancy outcomes: potential molecular targets: an Executive Workshop Summary of the Eunice Kennedy Shriver National Institute of Child Health and Human Development. American Journal of Obstetrics and Gynecology. Mosby; 2016. pp. S1–S46.

6. Gabory A, Roseboom TJ, Moore T, Moore LG, Junien C. Placental contribution to the origins of sexual dimorphism in health and diseases: Sex chromosomes and epigenetics. Biology of Sex Differences. 2013. doi:10.1186/2042-6410-4-5

7. Tarrade A, Panchenko P, Junien C, Gabory A. Placental contribution to nutritional programming of health and diseases: Epigenetics and sexual dimorphism. Journal of Experimental Biology. Company of Biologists Ltd; 2015. pp. 50–58. doi:10.1242/jeb.110320

8. Khong TY, De Wolf F, Robertson WB, Brosens I. Inadequate maternal vascular response to placentation in pregnancies complicated by pre-eclampsia and by small-for-gestational age infants. Br J Obstet Gynaecol. 1986;93: 1049–1059.

9. Edu A, Teodorescu C, Dobjanschi CG, Socol ZZ, Teodorescu V, Matei A, *et al.* Placenta changes in pregnancy with gestational diabetes. Rom J Morphol Embryol. 2016;57: 507–512.

10. Morgan TK. Role of the Placenta in Preterm Birth: A Review. Am J Perinatol. 2016;33: 258–266.

11. Di Renzo GC, Rosati A, Sarti RD, Cruciani L, Cutuli AM. Does fetal sex affect pregnancy outcome? Gend Med. 2007;4: 19–30.

12. Eriksson JG, Kajantie E, Osmond C, Thornburg K, Barker DJP. Boys live dangerously

in the womb. Am J Hum Biol. 2010;22: 330–335.

13. Verburg PE, Tucker G, Scheil W, Erwich JJHM, Dekker GA, Roberts CT. Sexual Dimorphism in Adverse Pregnancy Outcomes - A Retrospective Australian Population Study 1981-2011. PLoS One. 2016;11: e0158807.
14. Melamed N, Yogev Y, Glezerman M. Fetal gender and pregnancy outcome. J Matern Fetal Neonatal Med. 2010;23: 338–344.
15. Myatt L. Sexual Dimorphism in Regulation of Trophoblast Respiration and Metabolic Reprogramming with Obesity and Gestational Diabetes. Placenta. 2019. p. e10. doi:10.1016/j.placenta.2019.06.037
16. Myatt L, Wang E, Stratton K, Bramer L, Webb-Robertson B-J, Kyle J, *et al.* Changes in Placental Lipidomics with Obesity and Gestational Diabetes: Sexual Dimorphism. Placenta. 2019. pp. e44–e45. doi:10.1016/j.placenta.2019.06.145
17. Sheiner E, Levy A, Katz M, Hershkovitz R, Leron E, Mazor M. Gender does matter in perinatal medicine. Fetal Diagn Ther. 2004;19: 366–369.
18. Elsmén E, Källén K, Maršál K, Hellström-Westas L. Fetal gender and gestational-age-related incidence of pre-eclampsia. Acta Obstet Gynecol Scand. 2006;85: 1285–1291.
19. Schmidt A, Morales-Prieto DM, Pastuschek J, Fröhlich K, Markert UR. Only humans have human placentas: Molecular differences between mice and humans. J Reprod Immunol. 2015. doi:10.1016/j.jri.2015.03.001
20. Shah K, McCormack CE, Bradbury NA. Do you know the sex of your cells? Am J Physiol Cell Physiol. 2014;306: C3–18.
21. Lee CQE, Gardner L, Turco M, Zhao N, Murray MJ, Coleman N, *et al.* What Is Trophoblast? A Combination of Criteria Define Human First-Trimester Trophoblast. Stem Cell Reports. 2016;6: 257–272.
22. Bartel DP. MicroRNAs: Genomics, Biogenesis, Mechanism, and Function. Cell. 2004. pp. 281–297. doi:10.1016/S0092-8674(04)00045-5
23. Friedman RC, Farh KK-H, Burge CB, Bartel DP. Most mammalian mRNAs are conserved targets of microRNAs. Genome Res. 2008;19: 92–105.
24. Ameres SL, Zamore PD. Diversifying microRNA sequence and function. Nature Reviews Molecular Cell Biology. Nature Publishing Group; 2013. pp. 475–488. doi:10.1038/nrm3611

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

26. Morales-Prieto DM, Chaiwangyen W, Ospina-Prieto S, Schneider U, Herrmann J, Gruhn B, *et al.* MicroRNA expression profiles of trophoblastic cells. Placenta. 2012;33: 725–734.

27. Doridot L, Miralles F, Barbaux S, Vaiman D. Trophoblasts, invasion, and microRNA. Front Genet. 2013;4: 248.

28. Bortolin-Cavaille ML, Dance M, Weber M, Cavaille J. C19MC microRNAs are processed from introns of large Pol-II, non-protein-coding transcripts. Nucleic Acids Res. 2009. doi:10.1093/nar/gkp205

29. Zhao C, Sun X, Li L. Biogenesis and function of extracellular miRNAs. ExRNA. 2019. doi:10.1186/s41544-019-0039-4

30. Sun L, Yu Y, Niu B, Wang D. Red Blood Cells as Potential Repositories of MicroRNAs in the Circulatory System. Front Genet. 2020;11: 442.

31. Kirschner MB, Kao SC, Edelman JJ, Armstrong NJ, Vallely MP, van Zandwijk N, *et al.* Haemolysis during sample preparation alters microRNA content of plasma. Pfeffer S, editor. PLoS One. 2011;6: e24145.

32. Mitchell PS, Parkin RK, Kroh EM, Fritz BR, Wyman SK, Pogosova-Agadjanyan EL, *et al.* Circulating microRNAs as stable blood-based markers for cancer detection. Proc Natl Acad Sci U S A. 2008;105: 10513–10518.

33. Cortez MA, Calin GA. MicroRNA identification in plasma and serum: a new tool to diagnose and monitor diseases. Expert Opin Biol Ther. 2009;9: 703–711.

34. Kosaka N, Iguchi H, Ochiya T. Circulating microRNA in body fluid: a new potential biomarker for cancer diagnosis and prognosis. Cancer Sci. 2010;101: 2087–2092.

35. Vickers KC, Palmisano BT, Shoucri BM, Shamburek RD, Remaley AT. MicroRNAs are transported in plasma and delivered to recipient cells by high-density lipoproteins. Nat Cell Biol. 2011;13: 423–433.

36. Arroyo JD, Chevillet JR, Kroh EM, Ruf IK, Pritchard CC, Gibson DF, *et al.* Argonaute2 complexes carry a population of circulating microRNAs independent of vesicles in human plasma. Proceedings of the National Academy of Sciences. 2011;108: 5003–5008.

37. Cortez MA, Bueso-Ramos C, Ferdin J, Lopez-Berestein G, Sood AK, Calin GA. MicroRNAs in body fluids—the mix of hormones and biomarkers. Nat Rev Clin Oncol. 2011;8: 467–477.

38. Pritchard CC, Kroh E, Wood B, Arroyo JD, Dougherty KJ, Miyaji MM, *et al.* Blood cell origin of circulating microRNAs: a cautionary note for cancer biomarker studies. Cancer Prev Res . 2012;5: 492–497.

39. Blondal T, Jensby Nielsen S, Baker A, Andreasen D, Mouritzen P, Wrang Teilum M, *et al.* Assessing sample and miRNA profile quality in serum and plasma or other biofluids. Methods. Academic Press; 2013. pp. 164–169. doi:10.1016/j.ymeth.2012.09.015

40. Kirschner MB, Edelman JJB, Kao SCH, Vallely MP, Van Zandwijk N, Reid G. The impact of hemolysis on cell-free microRNA biomarkers. Front Genet. 2013;4: 94.

41. Roberts CT. IFPA Award in Placentology Lecture: Complicated interactions between genes and the environment in placentation, pregnancy outcome and long term health. Placenta. 2010;31 Suppl: S47–53.

42. 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.

43. 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.

44. Ghidini A, Salafia CM. Gender differences of placental dysfunction in severe prematurity. BJOG. 2005;112: 140–144.

45. Brown MA, Magee LA, Kenny LC, Karumanchi SA, McCarthy FP, Saito S, *et al.* Hypertensive Disorders of Pregnancy: ISSHP Classification, Diagnosis, and Management Recommendations for International Practice. Hypertension. 2018;72: 24–43.

46. Abalos E, Cuesta C, Grosso AL, Chou D, Say L. Global and regional estimates of preeclampsia and eclampsia: a systematic review. Eur J Obstet Gynecol Reprod Biol. 2013;170: 1–7.

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

48. Padayatty SJ, Katz A, Wang Y, Eck P, Kwon O, Lee J-H, *et al.* Vitamin C as an

antioxidant: evaluation of its role in disease prevention. J Am Coll Nutr. 2003;22: 18–35.

49. Siddiqui IA, Jaleel A, Tamimi W, Al Kadri HMF. Role of oxidative stress in the pathogenesis of preeclampsia. Archives of Gynecology and Obstetrics. 2010. pp. 469–474. doi:10.1007/s00404-010-1538-6

Chapter 2

Literature Review

Sexual dimorphism in the human placenta transcriptome

Abstract

Biological sex is an essential variable in all biomedical experiments yet is often absent or unreported on public data repositories. The presence of sex information is even more important in research into pregnancy complications, where male bearing pregnancies are disproportionately impacted. In this review, we present a current state of sex-bias in human placental research, with a specific focus on transcriptomics, and the importance of incorporating fetal sex into study design. Accounting for sex in gene expression studies, and studying sexual dimorphism generally in biomedical research, is likely to improve current disease models and allow for accurate use of Precision Medicine techniques for diagnosis and to inform treatments.

Key words: Placenta, transcriptome, sexual dimorphism

Introduction

Sexual dimorphism in human morphology and physiology is evident even though the human genome, with the exception of the sex chromosomes, is shared between biological males and females. The most obvious genetic difference between the sexes is that of the sex chromosomes (Figure 1A). In fact, simply possessing a Y chromosome predisposes a fetus and its mother to a number of non-preferred pregnancy outcomes [1], many of which are associated with placental insufficiency [2]. Interestingly, the X chromosome exhibits a unique plasticity of expression in placenta with potential for inactivation [X chromosome inactivation (XCI)] or escape from XCI in response to intrauterine conditions which may help to buffer the female fetus from an adverse maternal environment [3].

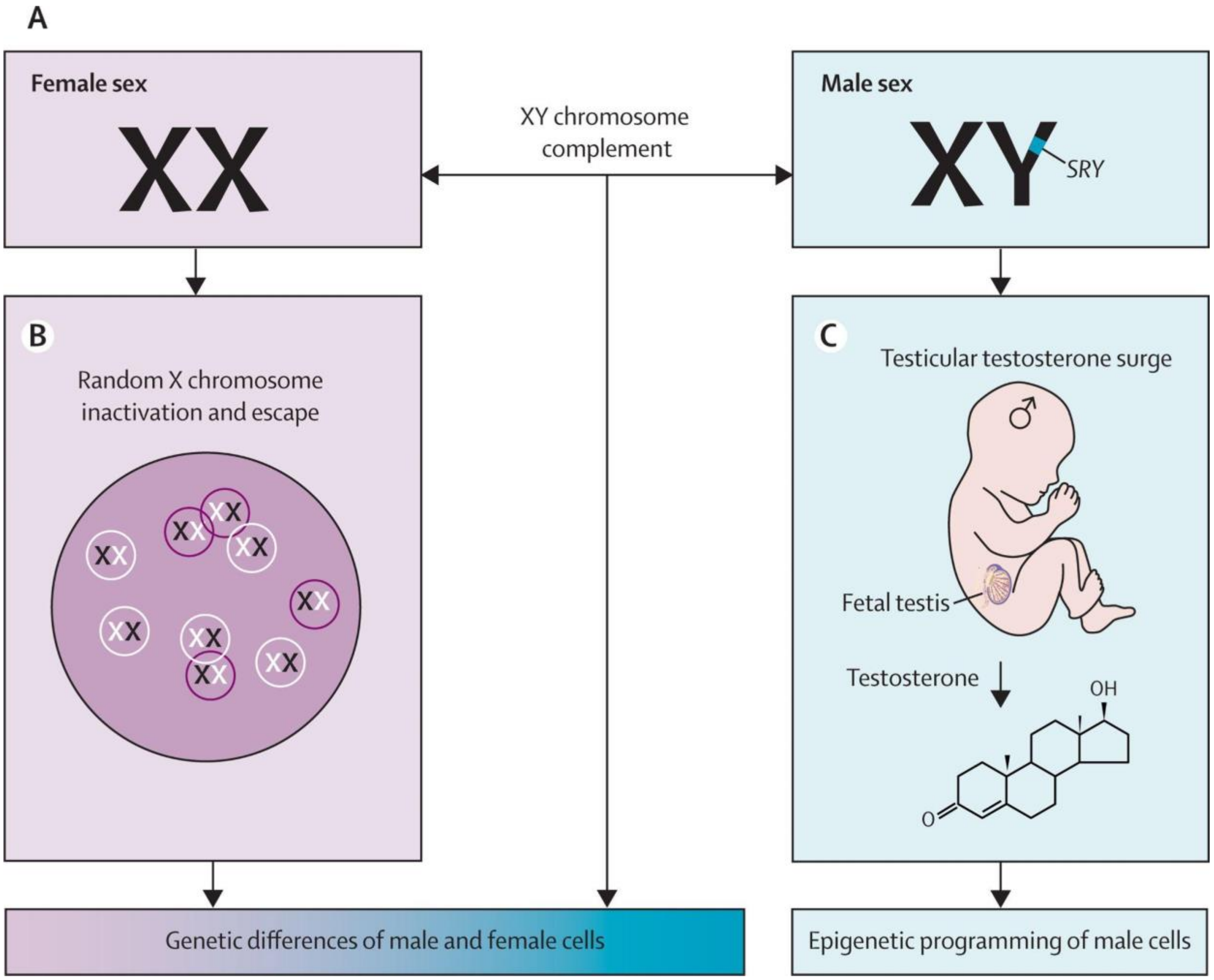


Figure 1. **A**, Genetic sex differences begin with cells carrying a compliment of either XX or XY chromosomes. Genes expressed on the Y chromosome are only present in genotypically male individuals. **B**, Random X inactivation in the placenta results in a mosaicism of active X. Some X-linked genes escape inactivation with evidence of X-linked genes in placenta up-regulated in female bearing pregnancies [4] **C**, The Y chromosome gene *SRY* is present only in genotypically male individuals and is responsible for testis production and resultant testosterone production. [5]

Genes expressed from the Y chromosome are only present in genotypically male individuals whereas genes expressed from the X chromosome are present in both males and females. When considering sex differences in the placenta it would be natural to presume these differences are largely due to the presence of a Y chromosome in the male bearing pregnancy, or the presence of two X chromosomes (Figure 2B), with the resulting dosage compensation and X-inactivation (XCI) in the female bearing pregnancy. However, the differences in gene expression are not restricted to genes on the sex chromosomes, but rather also include gene expression, epigenetic regulation and response to stress from autosomal genes. It is important to note that the active X chromosome in female mammalian cells within the same organisms may be of maternal or paternal origin resulting in a mosaic pattern of activity [6], this includes the placenta which exhibits a patchwork of mosaic, independent genetic units and random XCI [6–8].

Whilst only 4% of the genes found to be differentially expressed between male and female tissues are found on the X chromosome, of the 30 genes found to exhibit consistent sex-biased expression, 22 are known to constitutively escape XCI [9]. This is consistent with Gonzalez *et al.* (2018) who, using RNA sequencing data from late first trimester placenta, identified 58 genes with significantly different expression between males and females (25 X-linked, 15 Y-linked, and 18 autosomal genes). Interestingly, of the X-linked genes found to be up-regulated in placenta from female bearing pregnancies, 59.1% are known to escape XCI [4]. Though the mechanism for XCI is not fully defined, it has been demonstrated that transcriptional suppressor SPEN1-3 (SPEN), which is essential for the initiation of gene silencing on the X chromosome and the loss of which has been associated with deficient XCI, is immediately recruited to the X chromosome upon the up-regulation of Xist and is targeted to the promoters and enhancers of active genes[10]. Whilst differences in sex-chromosome gene expression are expected, it is reasonable to conclude that any sex-specific differences in autosomal gene expression are driven by sex-specific pre- and post-transcriptional elements [11,12].

Sexual dimorphism begins early in development, with the male fetus growing faster than its female counterpart [13,14] with testosterone (Figure 1C) an important hormone in programming sex differences in physiology and susceptibility to diseases that will manifest in adulthood [5]. Males have a greater body length and weight at birth than females [4] with on average, a greater feto placental weight ratio [15]. This has led to the consensus that the male placenta is more efficient than that of the female [16–18]. This rapid growth however, has a downside, with male pregnancies more likely to suffer from placental dysfunction [19] and pregnancy complications associated with placental insufficiency including preeclampsia (PE), intrauterine growth restriction (IUGR) and spontaneous preterm birth (SPTB) [13]. Further to complications in early pregnancy, when considering successful pregnancies that progress to term (ie ≥ 37 weeks') male sex also significantly predisposes to the elongation of gestation with (chromosomally normal) males likely to be born later compared to (chromosomally normal) females [20]. As well as the risks associated with gestation, the developmental origins of health and disease (DOHaD) hypothesis establishes a framework in which the *in utero* environment, mediated by the placenta, influences disease outcomes for mother and child later in life [15,21].

Genomic imprinting is an epigenetic process that involves differential epigenetic marking of the parental chromosomes, including DNA methylation and histone modifications, without altering the genetic sequence. It results in gene expression in a parent-of-origin-specific manner [22]. Differential epigenetic marks on parental chromosomes result in differential accessibility by the transcriptional machinery and thus alleles of some genes having different levels of activity, often to the point of monoallelic expression [23]. Originally described in the insect *Pseudococcus nipae,* imprinting has since been identified in higher organisms with ~200 genes known to be imprinted in the human genome [24]. Whilst dispersed throughout the genome, these genes are often located in clusters [25–27], thought to be a function of the epigenetic marks that give rise to the parent of origin specific silencing. Expression from imprinted genes may show both spatial and/or temporal heterogeneity meaning that certain imprinted genes are only expressed in a single tissue, or at a single time in development [28]. Many imprinted genes are expressed in placenta and are important for placental development [29]. It has been proposed that imprinting has an evolutionary explanation and indeed it appears that in general, maternally derived imprinted genes attenuate fetal growth, whilst paternally derived imprinted genes promote fetal growth [22,29].

This review will introduce known sex differences in human pregnancy and the placenta more broadly, before discussing their implications for precision medicine. We introduce gene expression quantification methods and discuss the impact of sex on the accuracy of experimental results. This leads into how we study the placenta and the limitations of current methods, recent and exciting methods such as single cell sequencing, and finally an investigation into sex chromosomes, X-inactivation (XCI) and placental mosaicism.
Based on the guidelines provided by the Institute of Medicine (US) Committee on Understanding the Biology of Sex and Gender Differences [30], reference to the “sex” of

the tissue/cell line in this review refers to “genotypic sex”, where an individual is either chromosomally XX or XY.

Sex differences in pregnancy outcomes

It is well known that there are sex-specific differences in placental physiology and pregnancy/neonatal outcomes (including maternal) [12–14,18,31–33]. As early as the late 18th century clinicians recorded their observations noting that on average male neonates were larger than female neonates, and importantly, male neonates had a larger variance in birth weight, and suffered poorer neonatal outcomes. Clarke [34] identified an excess mortality rate in male neonates, attributing this to their larger head circumference and subsequent issues with delivery. Meiosis during spermatogenesis, produces equal numbers of X- and Y-bearing gametes, which would suggest the sex ratio at conception, should be equally balanced. What we see, however, consistent across time, is a higher proportion of male births [13,14,34–36]. The sex ratio among abnormal embryos, however, is male-biased, and the sex ratio among normal embryos is female-biased. This may account for the lower sex ratio in the first week or so after conception (due to excess male mortality) [37].

More recently, it has been suggested that the early loss of male fetuses due to environmental exposures results in a stronger birth cohort overall in males, and an increased vulnerability later in development for females. Often referred to as the “culled cohort effect” the assumption is that females rely on adaptive flexibility early in gestation to escape the consequences of early life exposure to adversity. Males however are unable to moderate their growth, and thus are at greater risk of death early in gestation resulting in a stronger cohort at term [38,39]. We also often consider that the placenta from a male bearing pregnancy is more efficient than that of the female bearing pregnancy. In a study of 437 severely premature liveborn neonates, males were found to have higher birth weight centiles whilst having similar feto placental weight ratios [17]. As gestation continues, female births form a greater proportion of total births, due to an excess of male PTBs [40] until post-term is reached at which point male births increase as a proportion of total births [20]. However, these finding have not been replicated in all populations.

The difference in the way male and female bearing pregnancies respond to stressors is thought to explain why male fetal sex is associated with pre-term birth and female fetal sex is associated with fetal growth restriction [41]. In fact, male bearing pregnancies are at higher risk of many adverse pregnancy outcomes including gestational diabetes melitus (GDM), preeclampsia (PE), infection, preterm premature rupture of membranes, abruptio placentas, polyhydramnios, fetal macrosomia, failure to progress through labour, cord prolapse and umbilical cord knots [31,42,43], also supported by systematic review [14].

It is important to note that fetal sex-specific responses to stressors are at least in part modulated directly by the placenta. Sex-specific differences in the cortisol stress response occur before birth, with much higher levels of cortisol secretion for male than for female ovine fetuses [44]. Using a mouse model, researchers have clearly established an effect of sex chromosomes on placental size, with XY placentas being significantly larger than XX placentas and that this difference is independent of androgen signalling [45]. Sexual dimorphism has also been observed in embryonic cells isolated from mice at E10.5 whereby cells respond differently to dietary stressors even before the production of sex hormones indicating the effect of sex chromosomes are most important and hormones are secondary [46].

There is also a sexually dimorphic response by the placenta to maternal influences such as maternal diet and obesity. Investigation into the effect of famine on placenta size found that while famine conditions (starvation) reduced placental size overall, the decrease in placental area was greater for male than female bearing pregnancies [47]. At the opposite end of the scale, higher maternal BMI increased the risk of macrosomia and large for gestational age births (LGA) in male and female neonates while decreasing the risk of small for gestational age births (SGA) only in male bearing pregnancies with GDM predictive for macrosomia in male babies only [48].

Maternal obesity results in placental overgrowth and fetal hypoxia as manifested by normoblastemia (the presence of nucleated red blood cells in the peripheral blood); it is also associated with an increased incidence of chronic villitis (CV) and fetal thrombosis, both more prevalent in female placentas [49]. Examination of placentas from term C-sections with mothers classed as "lean" or "obese" found that male fetuses of lean women had the highest antioxidant activity. This protection is lost with obesity, perhaps contributing to the increased incidence of adverse outcomes in male bearing pregnancies. Obese women carrying a male fetus had the highest placental measurements of protein carbonyls (oxidative stress marker) and nitrotyrosine (reactive nitrogen species) levels while lean women carrying a male fetus had the highest total antioxidant capacity [50]. Placental or chorion trophoblast cells from pregnancies with a male fetus produced more proinflammatory TNFa in response to LPS stimulation and less anti-inflammatory IL-10 and granulocyte colony stimulating factor (G-CSF) compared to cells from women carrying a female fetus [51]. They also produced more prostaglandin synthase (PTGS-2) and less prostaglandin dehydrogenase (PGDH). These data suggest that in the presence of a male fetus the trophoblasts have the potential to generate a more proinflammatory environment [51].

Maternal stress also affects the fetus in a sex-specific manner, with more detrimental impacts observed in male babies [52]. In a study of affluent Scandinavian women, the negative effect on fetal growth from maternal smoking was found to affect the developing male fetus proportionately more than the female fetus. Both male and female newborns of mothers who smoked heavily have lower birth weight than female newborns whose mothers did not smoke while pregnant. However, males, but not females, whose mothers smoked during pregnancy had a smaller head circumference than their unexposed peers [53].

In a retrospective review on the effect of fetal sex on labour and delivery, Eogan *et al.* [54] identified male deliveries as more likely to require oxytocin augmentation, and instrumental vaginal delivery. Male neonates had a larger head circumference and were more likely to have an operative delivery. Female births were more likely to be accompanied by meconium stained liquor, indicative of fetal distress [54].

The influence of fetal sex on pregnancy outcomes may also be mediated by sex-specific maternal adaptations to pregnancy. We should consider the heightened immune response of the mother previously reported in the male bearing pregnancy [55]. One study on normotensive women diagnosed with PE, assessed the association between fetal sex and maternal peripheral microvascular responses (including baseline perfusion, response to hyperemia, post-occlusive reperfusion, and vasodilation) in response to corticotropin-releasing hormone (CRH). Researchers found that at 31-40 weeks' gestation those women carrying a male fetus exhibited increased vasodilation in response to CRH and greater baseline perfusion than those carrying a female fetus. Preeclamptic women with a male fetus demonstrated a significantly reduced vasodilatory response to CRH, reduced baseline perfusion and reduced response to hyperemia compared to normotensive women pregnant with a male fetus. Microvascular function was not different between preeclamptic and normotensive women carrying a female fetus [56].

Despite data illustrating profound differences between the sexes in immune function, sex differences in the pathogenesis of disease are often overlooked in biomedical research. Whilst enforcing editorial policies around reporting possible sex-biases is likely time consuming, incorporating male and female samples into all experimental design may also impact the statistical power of a study [57]. Editorial policies that require authors to report the sex of their cells, animals, and subjects will highlight instances where a sex-bias is present, which will improve our understanding of the pathogenesis of diseases, with the long-term goal of personalizing treatments for males and females in an effort to protect us equally [55]. Since pregnancy only ever occurs in individuals who are genetically female individuals, there is a misconception that pregnancy associated research, such as that in placenta is exempt from sex-specificity. However, sex differences in pregnancy outcomes are increasingly commonly reported, as are differences in the short and long term effects of pregnancy complications on both mother and child.

In a case-control study investigating sex differences associated with placental dysfunction in severe prematurity (437 consecutive deliveries before 32 weeks' gestation of singleton, liveborn, non-anomalous, infants Male = 232; Female = 205) sex differences exist in placental histology of extreme prematurity were tested. No association between fetal sex and lesions of acute inflammation, intraplacental vascular pathology or uteroplacental vascular pathology were identified. Lesions of chronic placental inflammation had a significantly higher score for male compared to female fetuses. The distribution of chronic placental inflammation identified significantly more severe lesions for male compared to female fetuses at the decidua basalis (i.e. the site of implantation of the blastocyst into the endometrial epithelium), than within the placental villi or the amniochorionic membranes (where interstitial trophoblast invasion is minimal). In cases of prematurity <32 weeks' gestation in male bearing pregnancies, placental lesions predominantly indicate a maternal immune rejection of extravillous trophoblasts in the decidua basalis [17].

Commensurate with fetal sex-related differences in pregnancy outcomes, the fundamental phenotypic differences between adult men and women are readily evident. As previously discussed, both the *in utero* environment and pregnancy complications not only affect the developing fetus in a sex-specific manner, but may also have long term effects on one's development and adult health outcomes [32,58]. Antenatal care and research into pregnancy health must include not only the sex of the mother, but also that of the fetus, and therefore the sex of the placenta. This will not only improve the establishment of sex-specific risk factors for pregnancy complications, but it has also been suggested that novel therapies are available if we take note of all the influences that sex has on the molecular systems within a cell. A concept for which the term "sexome" has been coined [59]. One aspect of the sexome we are able to accurately assay, is gene expression, for which there are a number of technologies.

High-throughput sequencing approaches for identifying sex-biased gene expression

Gene expression profiling

There are a number of methods used to profile transcription in human tissues. Four common methods are detailed in Figure 2. Methods such as qPCR (Figure 2A) that use single, targeted oligonucleotides for amplification, are limited in that they constrain the profile to a relatively small number of known, targeted transcripts. Subsequently, the implementation of hybridization technology such as the microarray (Figure 2B), which utilises a predetermined set of oligonucleotide probes, embedded on a chip, to quantitate the gene expression of many transcripts simultaneously and Nanostring, a fluorescence based technology that uses a short chain of fluorescent tags as barcodes for quantification of specific mRNAs of interest [60] (Figure 2C), gave us the ability to accurately profile a significant portion of the genome at one time [61].

High throughput sequencing technologies (Figure 2D), such as those developed by 454 Life Sciences (Roche) [62] and Illumina (formerly Solexa sequencing) [63] had the accuracy of microarray, with the power to identify differentially expressed genes, but also have the ability to detect transcripts with low expression, alternate splice variants and novel transcripts [64]. Unlike targeted arrays or other transcript amplification methods, high-throughput sequencing allows us to remove past limitations and connect sex to our research at a molecular level with a precision never before possible. In their excellent review into current approaches of transcript profiling in the human placenta Yong and Chan (2020) noted that prior to 2015, microarray technology was predominantly used in placental transcriptome profile research but has since been superseded by high-throughput RNA sequencing platforms [65].

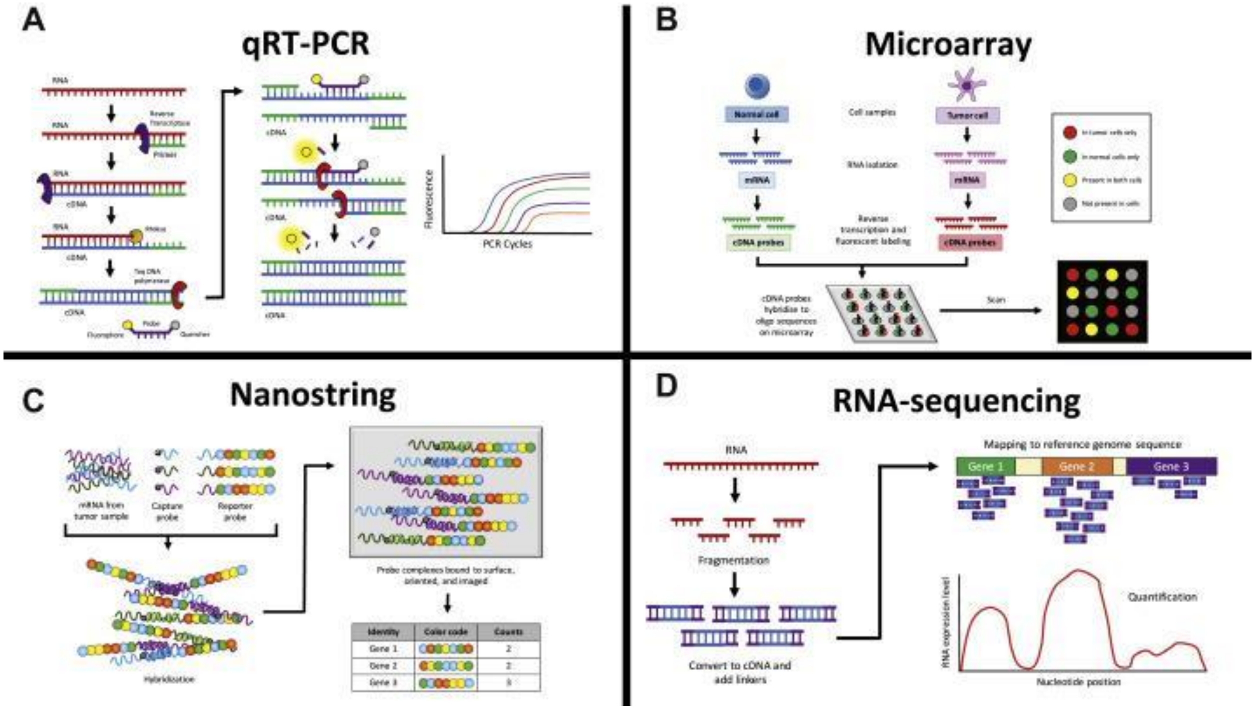


Figure 2. A, qRT-PCR (TaqMan) sample RNA is converted into cDNA and amplified in the presence of target specific oligonucleotides bound to a fluorescent probe and fluorescence quencher. This technique is limited in that it uses single, targeted oligonucleotides for amplification. **B**, for microarray, RNA is extracted from two comparative tissues, here, normal and tumor cells, reverse transcribed and labelled with fluorescent probes (green for normal cDNA and red for tumor cDNA). The cDNAs are then applied to a microarray chip where they bind to complementary sequences from annotated genes. The relative amount of green versus red fluorescence corresponds to the relative expression of genes in normal versus tumor cells. **C**, Nanostring. Extracted RNA is hybridized with target-specific capture probes and reporter probes. The reporter probes consist of 6 distinct positions, each of which may be occupied by one of several fluorophores in precisely ordered fluorescent barcodes. An electric current is used to align the probe-target complexes and the fluorescent barcodes are counted to digitally quantitate each target. **D**, RNA-seq. Extracted RNA is fragmented, reverse transcribed and amplification adapters are ligated. The cDNA is then sequenced using high-throughput sequencing platforms. The resulting sequences are then aligned to a reference genome and quantified to reveal the expression levels of various genes in the sample [66].

Until recently, similarities in regions of DNA such as the pseudo-autosomal homologs found on the X and Y chromosomes, have made analysing sexually dimorphic gene expression by techniques such as qPCR or microarray in these regions impossible. High throughput sequencing allows for single nucleotide precision in the analysis of gene expression not only from the entire exome, but also allows for investigation of non-coding

RNA expression, three-dimensional chromosome structure, locations of enhancers and promoters, alternate splicing events, novel transcripts and DNA methylation. Accurate analyses of high throughput sequencing data requires the development and implementation of specialist algorithms to manage the complexity of aligning these data to the transcriptome. Complexities include alternate splice variants, novel transcripts, and increasingly, long transcripts [67]. Annotating the X-Y homologous regions still poses some limitations but correct mapping and variant calling from NGS data is possible with the implementation of alignment tools such as *X-align* [68].

Methods for preparing samples for transcriptional profiling vary depending on the method being used. High quality RNA is essential for accurate profiling and requires that tissue samples or cell cultures are rapidly frozen or rapidly stored in an appropriate volume in an adequate amount of an RNA stabilising solution to prevent the degradation of transcripts [69,70]. Specialist analyses such as that of small RNA or other non-coding RNAs will require additional steps such as size fractionation or cross-linking [71]. Analysis of the human placental transcriptome is typically performed using cell lines or placental tissue collected post delivery or post termination of pregnancy. Samples are commonly taken from post delivery placenta or from post termination of pregnancy [65]. Both offer the opportunity to study placental transcription, however, it is reasonable to say that samples taken from whole tissue are physiologically more relevant to *in situ* placenta, and that the use of a single placental cell line negates any opportunity to investigate sex differences given that a cell line originates from a single individual [72].

With the advent of new sequencing technologies, it is now possible to comprehensively characterise to what extent gene expression in the human placenta presents a sex-bias. To better understand sex driven differences in pregnancy outcomes, researchers have focussed on methods to determine the extent of sex-biased gene expression in non-pathological human placental tissue.

Single-cell transcriptomics

Like other human tissues the placenta is a rich tapestry of cell types and spatial heterogeneity that changes dynamically across gestation [73–75]. Unlike bulk RNA-seq profiles, single-cell RNA-seq analysis allows for cell type-specific analysis of the placental transcriptome. Whilst the choice of cell-sorting methods may affect the proportion of cell types that make it through to sequencing, single-cell RNA-seq captures expression data from all but genes with the lowest expression in bulk RNA-seq profiles [76].

To date single-cell analyses using placental tissue tend to not include the sex characteristics of samples [75–78] pool tissue from both sexes such that sex-specific information is lost [79] or specifically target a single fetal sex [80]. All single-cell research converges on the same lineage differentiation pathways previously discussed here with CTB, STB and EVT cells identified in all analyses and the CTB to STB and CTB to EVT differentiation pathways identified (Figure 3A & B). The lack of sex-specific analyses illustrates a potentially important area for the future development of single cell analysis in human placental research.

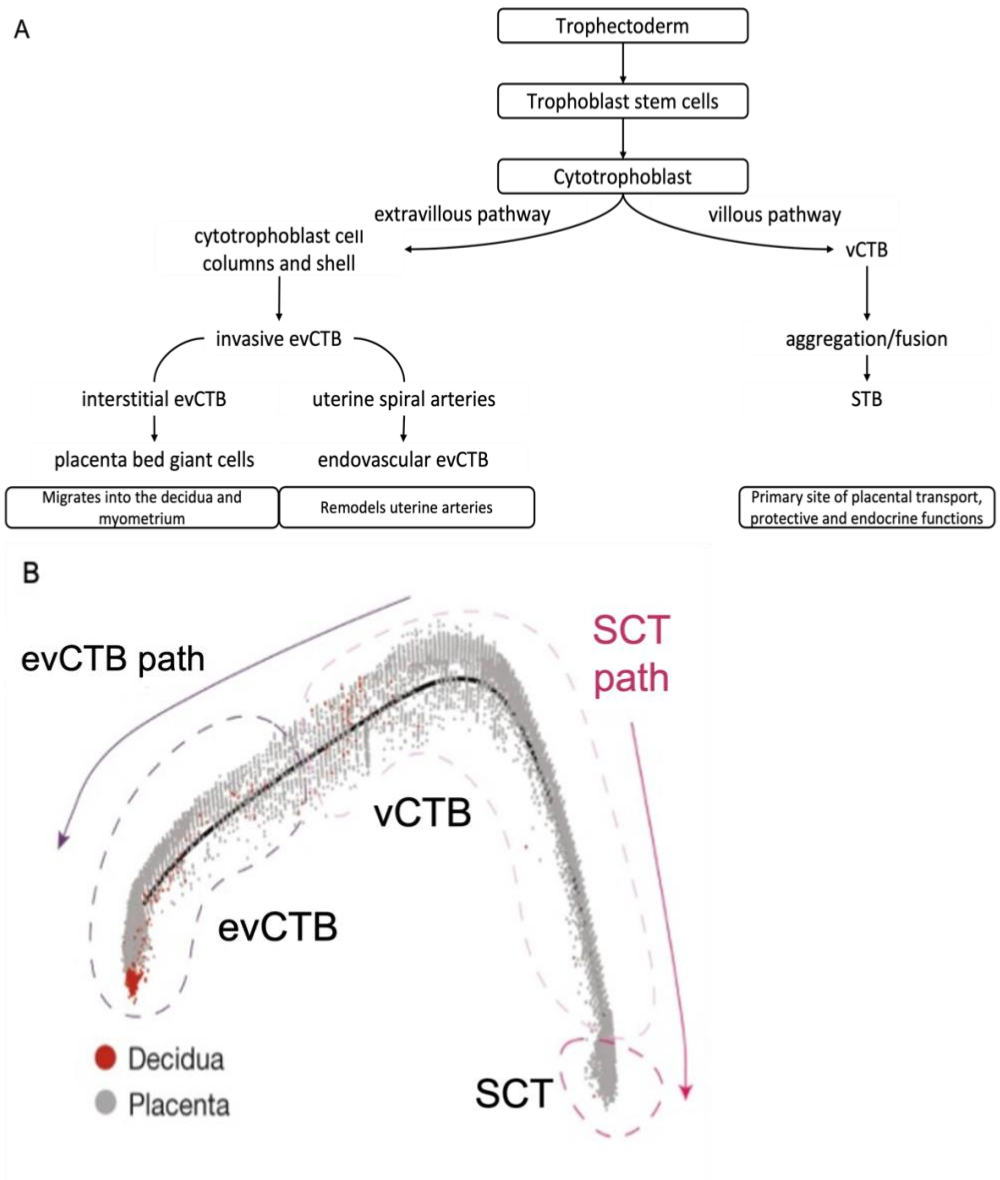


Figure 3: Human trophoblast development pathways. In humans, after successful implantation and initiation of placentation, trophoblast cells develop down two pathways. **A**) Depending on their location within the cytotrophoblast cell column, proliferative cytotrophoblast (evCTB) will migrate from the top of the anchoring villi invading the decidua (forming placental bed giant cells) or the uterine spiral arteries (forming endovascular evCTB), acquiring an invasive phenotype in the extravillous pathway. In the villous pathway, mononucleated villous cytotrophoblast (vCTB) fuse to form a single, uninterrupted syncytium, the syncytiotrophoblast (STB). Figure adapted from [2,81,82]. **B**) Single cell analysis has elucidated a pseudotime ordering of trophoblast differentiation from villous cytotrophoblast (vCTB) through the extravillous pathway (evCTB path) to extravillous trophoblast (EVT), and through the villous pathway (SCT path) to syncytiotrophoblast (SCT) adapted from [75].

Sex-derived transcriptional differences between male and female placenta

Normal pregnancy

In a recent characterization of sex differences in the human transcriptome Oliva *et al.* [9] generated a catalogue of sex differences in gene expression across 44 human tissues identifying 13,294 sex-biased genes across all tissues. The caveat to the study was that it did not include the placenta. Sex-biased gene expression in the normal human placenta occurs across the genome and includes both coding and non-coding transcripts that are central to growth [83] and development of the placenta and the maintenance of pregnancy, with a majority (60%) of sexually dimorphic gene expression found in autosomal regions of the genome [12]. The implications for this are not limited to the biological investigation of pregnancy but affect the methods we use to investigate and interpret the results of gene expression experiments.

How do we define what is “normal” in gene expression? Cleal *et al*. [84] set out to find a set of housekeeping mRNAs in human placenta. Their aim was for these mRNAs to be used to help normalise qPCR data. *UBC*, *TOP1* and *YWHA* were identified as stably expressed in placenta and not affected by extraction technology. The two most stable of these genes were determined to be appropriate to create a normalization factor [84]. However, in 2010 after investigation of a larger dataset, the authors added that whilst the genes they had identified were stable, they were now able to identify sex-differences in expression of these (and other) mRNA, effectively biasing any research question by the introduction of a sex-biased normalisation factor. Following this discovery the authors concluded that fetal sex should be considered in placenta research [85] while others agreeing that fetal sex should be considered as an independent variable in gene expression analysis of human placental villi [83].

Fetal sex differentially affects gene expression in a cell-phenotype dependent manner in cytotrophoblasts, syncytiotrophoblast, arterial and venous endothelial cells taken from human placenta [83]. Investigation of gene expression in human placenta, with respect to sex, shows that genes expressed at higher levels in female placentas include those with roles in immune regulation such as *JAK1, IL2RB, Clusterin, LTBP, CXCL1, IL1RL1* and cytokine mRNA [11,86]. Estrogens, such as 17β-oestradiol (also known as E2) and oestriol, progesterone and testosterone, can mediate many of the sex-based differences in immune responses [87]. In both males and females, the effects of androgens, such as testosterone, on immune function are largely suppressive, thereby leading to decreased T- and B-cell proliferation and decreased immunoglobulin and cytokine production [88]. Whilst the mechanisms leading to sex-driven differential gene expression remain to be fully understood, there are documented methylation differences across multiple cohorts [89] with impacts on genes encoding proteins related to immune function, growth/transcription factor signalling, and transport across cell membranes. These highlight the differences in the epigenetic response to the maternal environment and offer insight into one potential mechanism for differential sex outcomes of pregnancy we see across populations.

Sexually dimorphic gene expression in non-pathological human placenta from male and female bearing pregnancies can be considered a baseline [12] and may help to elucidate mechanisms leading to sex-differences in fetal growth trajectory. However, sexually-dimorphic gene expression in the uncomplicated pregnancy does not address sex-differences in the risk of many diseases of pregnancy associated with placental function [14,17,19,90,91]. For this, we need to establish what, if any, sex-biased gene expression is present in pregnancy with associated complications.

Pregnancy complications

Male and female embryos exhibit sex-specific transcriptional regulation [92], thus male and female development can be viewed as separate processes from conception to birth resulting in variable disease outcomes [93]. The placenta, a product of conception, should be viewed through the same paradigm as fetal development, given that it secretes hormones into maternal circulation affecting maternal adaptation to pregnancy and mediates maternal factors that influence on the fetus. The fact remains: simply being male increases your risk of gestating in a complicated pregnancy [14,43].

The sex-specific risk of a pregnancy complicated by preeclampsia (PE) was originally identified over 50 years ago [94]. PE is an inflammatory state and may influence placental function in a sex-specific manner. While multiple studies support the findings of Toivanen and Hirvonen (1970) [95,96], there have been others that report no association between fetal sex and maternal risk for preeclampsia [97,98] and some finding conversely that female fetal sex increases maternal risk [43,99]. In an attempt to resolve the discrepancy, Jaskolka *et al.* [100] performed a systematic review and meta-analysis concluding that male fetal sex was associated with maternal risk for preeclampsia/eclampsia in only the non-Asian population. This review performed a subgroup analysis using pooled ancestry assigned as non-Asian (mostly white European, two Middle Eastern and one African population) and Asian (unspecified). Further research suggests the interaction between maternal ancestry, fetal sex and PE indicates higher odds of PE among African American women who carry a male fetus compared to women of other ethnic backgrounds [101]. Notably, children diagnosed with autism spectrum disorder, another disorder with a strong male bias, are twice as likely to have gestated in a pregnancy complicated by PE, compared to children with typical development [102]. In a targeted approach, increased expression of *TNF-α, IL-8* and *IL-6* in male compared to female placentas from preeclamptic pregnancies was found. Decreased expression of *VEGF* in male preeclamptic placenta compared to uncomplicated male controls was found but there was no difference identified in female placentas. Interestingly they found that whilst the protein expression of *HIF-1α* was significantly higher in male PE placentas compared to female, and reduced *HIF-1a* activity was found in PE compared to normal placentas, no sex-specific differences in *HIF1-α* activity were found [103]. Although there was no fetal sex data included, a recent systematic review and meta-analysis found that in populations at high altitude (≥2500 m) maternal systolic and diastolic blood pressure at term was higher, and that women are at a higher risk of gestational hypertension, but at a reduced risk of preeclampsia [104].

A number of studies have shown that placental gene expression is different in a sex-specific manner when a mother has asthma during pregnancy. In one such study, cytokine mRNA expression was found to be decreased in female bearing pregnancies, but was unaltered in placentas from male bearing pregnancies complicated by asthma, resulting in reduced fetal growth in females unless the mother continued to use inhaled steroids [86]. Further, Clifton and Murphy [86] found that in pregnancies complicated by maternal asthma but not treated with inhaled steroids, placental glucocorticoid receptor-α (*GR-α),* glucocorticoid receptor-β (*GR-β*) and mineralocorticoid receptor (*MR*) expression decreased in females but increased in male bearing pregnancies. In a further study, there were 6 and 59 genes differentially expressed in placentas from pregnancies affected and unaffected by maternal asthma in male and female bearing pregnancies respectively [105]. Many of the genes identified as sexually dimorphic were associated with growth, inflammation and immune pathways.

Experimental biases that impact sex-related research

Placental cell lines

Research into human placental development and function is hampered both by ethical concerns and the availability of tissue. Consequently, research is often conducted using model organisms or immortalised cell lines, although it is argued that human placentation cannot be accurately modelled in any other animal [106]. The placenta is largely made up of cells derived from the trophectoderm of the preimplantation embryo [2]. Trophoblast cells develop down two specific cellular differentiation pathways, villous and extravillous, each resulting in terminally differentiated cells with specific functions. Early molecular and physiological research elucidated the cellular differentiation (Figure 3A) with more recent single cell sequencing analyses fine tuning the differentiation pathways into a more complex, pseudotime model (Figure 3B) [75].

Trophoblast cells can be isolated from placenta across gestation, but these are difficult to work with because once isolated the trophoblast rapidly differentiate and do not proliferate well *in vitro*. In an attempt to overcome the constraints inherent to isolated trophoblasts there have been many attempts to create appropriate model systems for placental research. These include the transformation of trophoblast cells from early gestation placenta and the development of an experimental system in which human embryonic stem cells (hESC) are driven towards a trophoblast cell type [107–115]. However, whilst these efforts have allowed for closer investigation of the very early trophoblast lineage specification and expansion from pluripotent progenitors along multiple sub-lineage differentiation pathways [107,115,116] there is ongoing discussion as to the true “trophoblast” nature of these supposed trophoblast cells induced from hESC [117].

Given this limitation, immortalised cell lines are often used to investigate molecular and differentiation pathways in the placenta. The use of cell lines however, introduces a number of issues. These may include karyotypic abnormalities and the loss of expression of placenta specific miRNA clusters (HTR8/SVneo cell lines for example, do not express miRNA from placenta specific chromosome 19 miRNA cluster C19MC) and the loss of critical interaction between different cell compartments which provides information to drive specific cell functions [118,119]. Additionally, consider the implied sex-bias which must also by definition limit the investigation into cells originating from a single individual and hence of a given sex [118–120] all of which highlights the importance of primary placental tissue in research.

Interestingly, defining the “sex” of a cell line is complicated as cell lines in culture are known to enter a “de-sexualised” state in which cells may lose their Y or inactive X chromosome (dependent on the initial sample sex), and may compensate by the duplication of the remaining X chromosome [121].

Table 1. Human trophoblast-derived cell lines commonly used for *in vitro* modelling of trophoblast function.

Cell line	Origin	Category	Fetal Sex	Notes	Reference
Be Wo	Choriocarcinoma	Cancer	XY	Spontaneously syncytialises [122]	[123,124]
ED27	First trimester chorionic villus isolated cells	Spontaneous	Unspecified	Human papillomavirus-related endocervical adenocarcinoma cells	[125]
ED31	First trimester chorionic villus isolated cells	Spontaneous	Unspecified		[125]
ED77	First trimester chorionic villus isolated cells	Spontaneous	Unspecified		[126]
HP-A1	Term placenta isolated cells	Immortalised with adenovirus (ori) SV40tsA209	Unspecified		[127]

Table 1 cont. Human trophoblast-derived cell lines commonly used for *in vitro* modelling of trophoblast function.

Cell line	Origin	Category	Fetal Sex	Notes	Reference
HP-A2	Term placenta isolated cells	Immortalised with adenovirus (ori) SV40tsA209	Unspecified		[127]
HP-W1	Term placenta isolated cells	Immortalised with adenovirus (ori) SV40 (wild type)	Unspecified		[127]
HT	Term placenta chorionic villus isolated cells	Spontaneous	XX	Abnormal aneuploid karyotype; tumourigenic (self-limiting); poorly differentiated	[128]
HT-116	First trimester extravillous trophoblast isolated cells	Spontaneous	Unspecified	Low expression of E-cadherin	[129]
HTR-8	First trimester extravillous trophoblast isolated cells	Spontaneous	X [130]	Reaches senescence after 12 passages	[131]
HTR8/SVneo	HTR-8 cells	Immortalised with SV40 T antigen	X [130]	Well characterised, phenotypically accurate; non-tumorigenic	[131]
JAR	Choriocarcinoma	Cancer	XY		[132]
JEG	Choriocarcinoma	Cancer	XY		[133]
IST-1	First trimester villous explant	HPV16 E6/E7	XX	Non-tumorigenic	[134]
NHT	First trimester chorionic villus isolated cells	Spontaneous	XY	Possible to grow in serum-free media	[135]

Table 1 cont. Human trophoblast-derived cell lines commonly used for *in vitro* modelling of trophoblast function.

Cell line	Origin	Category	Fetal Sex	Notes	Reference
NPC	First trimester chorionic villus isolated cells	Spontaneous	XY		[136]
RSVT-2, 2/C	HTR-8 cells	Immortalised with pRSVT	Unspecified		[136]
SGHPL-4	First trimester (unspecified) isolated cells	Immortalised with poly-L-ornithine with pSV3neo	Unspecified	Cytokeratin-negative	[137]
SPA-26	First trimester (unspecified) isolated cells	Immortalised with SV40tsa255	XY		[138]
Swan 71	First trimester (unspecified) isolated cells	Telomerase immortalised (pA317-hTERT)	Unspecified	No expression of CD45 or CD68; low hCG secretion	[139]
TCL-1	Term choriodecidua isolated cells	Immortalised with pZipSV40	Unspecified		[140]
TL	Term chorionic villus cells	Spontaneous	XX	Tumorigenic	[141]

Impact of experimental bias in biomarker identification

With sex-specific differences between male and female bearing placentas well established, it is natural for research to turn to sex-specific biomarker identification. Sex-specific transcriptome differences associated with immune response are found in adult peripheral blood [142], and fetal-sex differences have been identified in hormones measured in the maternal system and are used as biomarkers in current clinical use. For example, one of the first trimester markers used for Down syndrome screening, free b-human chorionic gonadotropin (b-hCG), has been shown to be higher in female than in male bearing pregnancies, resulting in a higher median calculated Down syndrome risk assigned to female bearing pregnancies [143]. Nucleic acids from the fetus detected from either amniotic fluid or chorionic villus biopsy are also used in routine clinical practice to identify fetal aneuploidies [144]. However, due to the increased risk of iatrogenic fetal loss, maternal blood also provides a minimally invasive source of fetally derived nucleic acids [145].

As well as DNA, fetal RNA is detectable in the maternal circulation. Placenta associated miRNA cluster transcripts have also been isolated and quantified from maternal plasma and serum [119,146–150], with changes in expression observed across gestation [151]. In its initial phase this approach was limited to microarray and quantitative PCR [1,146,149] but since the advent of high-throughput sequencing, the use of this technology for larger sample numbers, quantification of rare or low expression transcripts and potential for discovery of novel transcripts has complimented miRNA research. However, these transcripts are often found at low levels and identification of sex differences is difficult. In a study of miRNA from cord blood taken from 8 uncomplicated deliveries (4 female and 4 male) no sexual dimorphism was detected in cord blood miRNA [152], although increasing the number of biological replicates may provide statistical power to determine if sex-differences are present.

Gene expression differences between male and female bearing pregnancies may not provide the biomarker we hope for. In a 2004 immuno-histochemistry study the authors noted maternal serum and cord blood hCG levels were higher in pregnant women with female fetuses than in those carrying male fetuses while there were no sex differences in Ki-67 immunostaining rates of the cytotrophoblast cells (a common measure of cytotrophoblast cell proliferation). The authors concluded the sex-specific differences in maternal serum and cord blood hCG levels ares not associated with cytotrophoblast cell proliferation in the human term placenta and that the sex of the fetus does not seem to affect the regulation of cytotrophoblast cell proliferation [153].

Conclusion

Despite centuries of observation regarding the sexual dimorphism of pregnancy outcomes, and decades of evidence to demonstrate the sex-specific relationships between the environment and placental function, a known factor in pregnancy success, we find ourselves still without a clear understanding of the placental sexual-dimorphism. This may be due to difficulty in accessing high quality RNA from placenta and, in particular, limitations in accessing placental tissue from pregnancy terminations. There are approximately 19,901 protein-coding genes and 15,779 non-coding genes in the human genome [154]. Of these 14250 have been identified in the placenta, with 931 exhibiting sex specific expression [4].

Recent literature highlights that "to ignore the sex of the placenta is no longer sound scientific practice" [90] and that “genetic studies that ignore sex-specific effects in their design and interpretation could fail to identify a significant proportion of the genes that contribute to risk for complex diseases” [155]. Yet a decade later we see very little change in placenta research practices.

References

[1] R.S. Rodosthenous, H.H. Burris, A.P. Sanders, A.C. Just, A.E. Dereix, K. Svensson, M. Solano, M.M. Téllez-Rojo, R.O. Wright, A.A. Baccarelli, Second trimester extracellular microRNAs in maternal blood and fetal growth: An exploratory study, Epigenetics. 12 (2017) 804–810.

[2] N.M. Gude, C.T. Roberts, B. Kalionis, R.G. King, Growth and function of the normal human placenta, Thromb. Res. 114 (2004) 397–407.

[3] B.R. Migeon, J. Axelman, P. Jeppesen, Differential X reactivation in human placental cells: implications for reversal of X inactivation, Am. J. Hum. Genet. 77 (2005) 355–364.

[4] T.L. Gonzalez, T. Sun, A.F. Koeppel, B. Lee, E.T. Wang, C.R. Farber, S.S. Rich, L.W. Sundheimer, R.A. Buttle, Y.-D.I. Chen, J.I. Rotter, S.D. Turner, J. Williams 3rd, M.O. Goodarzi, M.D. Pisarska, Sex differences in the late first trimester human placenta transcriptome, Biol. Sex Differ. 9 (2018) 4.

[5] F. Mauvais-Jarvis, N. Bairey Merz, P.J. Barnes, R.D. Brinton, J.-J. Carrero, D.L. DeMeo, G.J. De Vries, C.N. Epperson, R. Govindan, S.L. Klein, A. Lonardo, P.M. Maki, L.D. McCullough, V. Regitz-Zagrosek, J.G. Regensteiner, J.B. Rubin, K. Sandberg, A. Suzuki, Sex and gender: modifiers of health, disease, and medicine, Lancet. 396 (2020) 565–582.

[6] L.H. Looijenga, A.J. Gillis, A.J. Verkerk, W.L. van Putten, J.W. Oosterhuis, Heterogeneous X inactivation in trophoblastic cells of human full-term female placentas, Am. J. Hum. Genet. 64 (1999) 1445–1452.

[7] T.H.H. Coorens, T.R.W. Oliver, R. Sanghvi, U. Sovio, E. Cook, R. Vento-Tormo, M. Haniffa, M.D. Young, R. Rahbari, N. Sebire, P.J. Campbell, D.S. Charnock-Jones, G.C.S. Smith, S. Behjati, Inherent mosaicism and extensive mutation of human placentas, Nature. 592 (2021) 80–85.

[8] S. Uehara, M. Tamura, M. Nata, G. Ji, N. Yaegashi, K. Okamura, A. Yajima, X-chromosome inactivation in the human trophoblast of early pregnancy, J. Hum. Genet. 45 (2000) 119–126.

[9] M. Oliva, M. Muñoz-Aguirre, S. Kim-Hellmuth, V. Wucher, A.D.H. Gewirtz, D.J. Cotter, P. Parsana, S. Kasela, B. Balliu, A. Viñuela, S.E. Castel, P. Mohammadi, F. Aguet, Y. Zou, E.A. Khramtsova, A.D. Skol, D. Garrido-Martín, F. Reverter, A. Brown, P. Evans, E.R. Gamazon, A. Payne, R. Bonazzola, A.N. Barbeira, A.R. Hamel, A. Martinez-Perez, J.M. Soria, GTEx Consortium, B.L. Pierce, M. Stephens, E. Eskin, E.T. Dermitzakis, A.V. Segrè, H.K. Im, B.E. Engelhardt, K.G. Ardlie, S.B. Montgomery, A.J. Battle, T. Lappalainen, R. Guigó, B.E. Stranger, The impact of sex on gene expression across human tissues, Science. 369 (2020). https://doi.org/10.1126/science.aba3066.

[10] F. Dossin, I. Pinheiro, J.J. Żylicz, J. Roensch, S. Collombet, A. Le Saux, T. Chelmicki,

M. Attia, V. Kapoor, Y. Zhan, F. Dingli, D. Loew, T. Mercher, J. Dekker, E. Heard, SPEN integrates transcriptional and epigenetic control of X-inactivation, Nature. (2020). https://doi.org/10.1038/s41586-020-1974-9.

[11] R. Sood, J.L. Zehnder, M.L. Druzin, P.O. Brown, Gene expression patterns in human placenta, Proc. Natl. Acad. Sci. U. S. A. 103 (2006) 5478–5483.

[12] S. Buckberry, T. Bianco-Miotto, S.J. Bent, Integrative transcriptome meta-analysis reveals widespread sex-biased gene expression at the human fetal–maternal interface, Mol. Hum. Reprod. (2014). https://academic.oup.com/molehr/article-abstract/20/8/810/2459830.

[13] L.J. Vatten, R. Skjaerven, Offspring sex and pregnancy outcome by length of gestation, Early Hum. Dev. 76 (2004) 47–54.

[14] G.C. Di Renzo, A. Rosati, R.D. Sarti, L. Cruciani, A.M. Cutuli, Does fetal sex affect pregnancy outcome?, Gend. Med. 4 (2007) 19–30.

[15] D.P. Misra, C.M. Salafia, R.K. Miller, A.K. Charles, Non-linear and gender-specific relationships among placental growth measures and the fetoplacental weight ratio, Placenta. 30 (2009) 1052–1057.

[16] T. Forsén, J.G. Eriksson, J. Tuomilehto, C. Osmond, D.J. Barker, Growth in utero and during childhood among women who develop coronary heart disease: longitudinal study, BMJ. 319 (1999) 1403–1407.

[17] A. Ghidini, C.M. Salafia, Gender differences of placental dysfunction in severe prematurity, BJOG. 112 (2005) 140–144.

[18] J.G. Eriksson, E. Kajantie, C. Osmond, K. Thornburg, D.J.P. Barker, Boys live dangerously in the womb, Am. J. Hum. Biol. 22 (2010) 330–335.

[19] A. Murji, L.K. Proctor, A.D. Paterson, D. Chitayat, R. Weksberg, J. Kingdom, Male sex bias in placental dysfunction, Am. J. Med. Genet. A. 158A (2012) 779–783.

[20] M.Y. Divon, A. Ferber, H. Nisell, M. Westgren, Male gender predisposes to prolongation of pregnancy, Am. J. Obstet. Gynecol. 187 (2002) 1081–1083.

[21] J.G. Eriksson, Developmental Origins of Health and Disease – from a small body size at birth to epigenetics, Ann. Med. 48 (2016) 456–467.

[22] T. Moore, D. Haig, Genomic imprinting in mammalian development: a parental tug-of-war, Trends Genet. 7 (1991) 45–49.

[23] W. Reik, J. Walter, Genomic imprinting: parental influence on the genome, Nat. Rev. Genet. 2 (2001) 21–32.

[24] V. Tucci, A.R. Isles, G. Kelsey, A.C. Ferguson-Smith, Erice Imprinting Group, Genomic Imprinting and Physiological Processes in Mammals, Cell. 176 (2019) 952–965.

[25] M. Noguer-Dance, S. Abu-Amero, M. Al-Khtib, A. Lefèvre, P. Coullin, G.E. Moore, J. Cavaillé, The primate-specific microRNA gene cluster (C19MC) is imprinted in the placenta, Hum. Mol. Genet. 19 (2010) 3566–3582.

[26] E.C. Malnou, D. Umlauf, M. Mouysset, J. Cavaillé, Imprinted microRNA gene clusters

in the evolution, development, and functions of mammalian placenta, Frontiers in Genetics. 10 (2019). https://doi.org/10.3389/fgene.2018.00706.

[27] A.R. Isles, A.J. Holland, Imprinted genes and mother-offspring interactions, Early Hum. Dev. 81 (2005) 73–77.

[28] M.V. Koerner, D.P. Barlow, Genomic imprinting—an epigenetic gene-regulatory model, Curr. Opin. Genet. Dev. 20 (2010) 164–170.

[29] W. Reik, M. Constância, A. Fowden, N. Anderson, W. Dean, A. Ferguson-Smith, B. Tycko, C. Sibley, Regulation of supply and demand for maternal nutrients in mammals by imprinted genes, J. Physiol. 547 (2003) 35–44.

[30] Institute of Medicine (US) Committee on Understanding the Biology of Sex and Gender Differences, Exploring the Biological Contributions to Human Health: Does Sex Matter?, National Academies Press (US), Washington (DC), 2014.

[31] P.E. Verburg, G. Tucker, W. Scheil, J.J.H.M. Erwich, G.A. Dekker, C.T. Roberts, Sexual Dimorphism in Adverse Pregnancy Outcomes - A Retrospective Australian Population Study 1981-2011, PLoS One. 11 (2016) e0158807.

[32] A. Gabory, T.J. Roseboom, T. Moore, L.G. Moore, C. Junien, Placental contribution to the origins of sexual dimorphism in health and diseases: Sex chromosomes and epigenetics, Biology of Sex Differences. 4 (2013). https://doi.org/10.1186/2042-6410-4-5.

[33] F. Uguz, K. Gezginc, F. Yazici, Are major depression and generalized anxiety disorder associated with intrauterine growth restriction in pregnant women? A case-control study, Gen. Hosp. Psychiatry. 33 (2011) 640.e7–9.

[34] J. Clarke, XVII. Observations on some causes of the excess of the mortality of males above that of females. By Joseph Clarke, M. D. Physician to the Lying-in Hospital at Dublin. Communicated by the Rev. Richard Price, D. D. F. R. S. in a letter to Charles Blagden, M, Philosophical Transactions of the Royal Society of London. 76 (1786) 349–362.

[35] E. Pergament, P.B. Toydemir, M. Fiddler, Sex ratio: a biological perspective of “Sex and the City,” Reprod. Biomed. Online. 5 (2002) 43–46.

[36] W.H. James, Analysing data on the sex ratio of human births by cycle day of conception, Hum. Reprod. 15 (2000) 1206–1208.

[37] S.H. Orzack, J.W. Stubblefield, V.R. Akmaev, P. Colls, S. Munné, T. Scholl, D. Steinsaltz, J.E. Zuckerman, The human sex ratio from conception to birth, Proc. Natl. Acad. Sci. U. S. A. 112 (2015) E2102–11.

[38] C.A. Sandman, L.M. Glynn, E.P. Davis, Is there a viability-vulnerability tradeoff? Sex differences in fetal programming, J. Psychosom. Res. 75 (2013) 327–335.

[39] R. Catalano, T. Bruckner, Secondary sex ratios and male lifespan: damaged or culled cohorts, Proc. Natl. Acad. Sci. U.S.A. 103 (2006) 1639–1643.

[40] J. Zeitlin, M.-J. Saurel-Cubizolles, J. De Mouzon, L. Rivera, P.-Y. Ancel, B. Blondel, M. Kaminski, Fetal sex and preterm birth: are males at greater risk?, Hum. Reprod.

17 (2002) 2762–2768.

[41] N. Melamed, Y. Yogev, M. Glezerman, Fetal gender and pregnancy outcome, J. Matern. Fetal. Neonatal Med. 23 (2010) 338–344.

[42] E. Sheiner, A. Levy, M. Katz, R. Hershkovitz, E. Leron, M. Mazor, Gender does matter in perinatal medicine, Fetal Diagn. Ther. 19 (2004) 366–369.

[43] E. Elsmén, K. Källén, K. Maršál, L. Hellström-Westas, Fetal gender and gestational-age-related incidence of pre-eclampsia, Acta Obstet. Gynecol. Scand. 85 (2006) 1285–1291.

[44] D.A. Giussani, A.J.W. Fletcher, D.S. Gardner, Sex differences in the ovine fetal cortisol response to stress, Pediatr. Res. 69 (2011) 118–122.

[45] H. Ishikawa, A. Rattigan, R. Fundele, P.S. Burgoyne, Effects of sex chromosome dosage on placental size in mice, Biol. Reprod. 69 (2003) 483–488.

[46] C. Penaloza, B. Estevez, S. Orlanski, M. Sikorska, R. Walker, C. Smith, B. Smith, R.A. Lockshin, Z. Zakeri, Sex of the cell dictates its response: differential gene expression and sensitivity to cell death inducing stress in male and female cells, FASEB J. 23 (2009) 1869–1879.

[47] T.J. Roseboom, R.C. Painter, S.R. de Rooij, A.F.M. van Abeelen, M.V.E. Veenendaal, C. Osmond, D.J.P. Barker, Effects of famine on placental size and efficiency, Placenta. 32 (2011) 395–399.

[48] W. Ricart, J. López, J. Mozas, A. Pericot, M.A. Sancho, N. González, M. Balsells, R. Luna, A. Cortázar, P. Navarro, O. Ramírez, B. Flández, L.F. Pallardo, A. Hernández, J. Ampudia, J.M. Fernández-Real, I. Hernández-Aguado, R. Corcoy, Maternal glucose tolerance status influences the risk of macrosomia in male but not in female fetuses, J. Epidemiol. Community Health. 63 (2009) 64–68.

[49] S.M. Leon-Garcia, H.A. Roeder, K.K. Nelson, X. Liao, D.P. Pizzo, L.C. Laurent, M.M. Parast, D.Y. LaCoursiere, Maternal obesity and sex-specific differences in placental pathology, Placenta. 38 (2016) 33–40.

[50] L.S. Evans, L. Myatt, Sexual dimorphism in the effect of maternal obesity on antioxidant defense mechanisms in the human placenta, Placenta. 51 (2017) 64–69.

[51] J. Challis, J. Newnham, F. Petraglia, M. Yeganegi, A. Bocking, Fetal sex and preterm birth, in: Placenta, 2013: pp. 95–99.

[52] T.L. Bale, Sex differences in prenatal epigenetic programing of stress pathways, Stress. 14 (2011) 348–356.

[53] B. Zaren, G. Lindmark, L. Bakketeig, Maternal smoking affects fetal growth more in the male fetus, Paediatr. Perinat. Epidemiol. 14 (2000) 118–126.

[54] M.A. Eogan, M.P. Geary, M.P. O'Connell, D.P. Keane, Effect of fetal sex on labour and delivery: retrospective review, BMJ. 326 (2003) 137.

[55] S.L. Klein, Immune cells have sex and so should journal articles, Endocrinology. 153 (2012) 2544–2550. https://doi.org/10.1210/en.2011-2120.

[56] M.J. Stark, L. Dierkx, V.L. Clifton, I.M.R. Wright, Alterations in the Maternal Peripheral

Microvascular Response in Pregnancies Complicated by Preeclampsia and the Impact of Fetal Sex, J. Soc. Gynecol. Investig. 13 (2006) 573–578.

[57] Institute of Medicine (US) Board on Population Health, Public Health Practice, IMPLICATIONS FOR JOURNALS OF SEX-SPECIFIC REPORTING POLICIES OF JOURNALS, National Academies Press (US), 2012.

[58] A. Tarrade, P. Panchenko, C. Junien, A. Gabory, Placental contribution to nutritional programming of health and diseases: Epigenetics and sexual dimorphism, Journal of Experimental Biology. 218 (2015) 50–58. https://doi.org/10.1242/jeb.110320.

[59] A.P. Arnold, A.J. Lusis, Understanding the sexome: Measuring and reporting sex differences in gene systems, Endocrinology. 153 (2012) 2551–2555. https://doi.org/10.1210/en.2011-2134.

[60] M.H. Veldman-Jones, R. Brant, C. Rooney, C. Geh, H. Emery, C.G. Harbron, M. Wappett, A. Sharpe, M. Dymond, J.C. Barrett, E.A. Harrington, G. Marshall, Evaluating Robustness and Sensitivity of the NanoString Technologies nCounter Platform to Enable Multiplexed Gene Expression Analysis of Clinical Samples, Cancer Res. 75 (2015) 2587–2593.

[61] B. Cox, K. Leavey, U. Nosi, F. Wong, J. Kingdom, Placental transcriptome in development and pathology: Expression, function, and methods of analysis, American Journal of Obstetrics and Gynecology. 213 (2015) S138–S151. https://doi.org/10.1016/j.ajog.2015.07.046.

[62] M. Margulies, M. Egholm, W.E. Altman, S. Attiya, J.S. Bader, L.A. Bemben, J. Berka, M.S. Braverman, Y.-J. Chen, Z. Chen, S.B. Dewell, L. Du, J.M. Fierro, X.V. Gomes, B.C. Godwin, W. He, S. Helgesen, C.H. Ho, G.P. Irzyk, S.C. Jando, M.L.I. Alenquer, T.P. Jarvie, K.B. Jirage, J.-B. Kim, J.R. Knight, J.R. Lanza, J.H. Leamon, S.M. Lefkowitz, M. Lei, J. Li, K.L. Lohman, H. Lu, V.B. Makhijani, K.E. McDade, M.P. McKenna, E.W. Myers, E. Nickerson, J.R. Nobile, R. Plant, B.P. Puc, M.T. Ronan, G.T. Roth, G.J. Sarkis, J.F. Simons, J.W. Simpson, M. Srinivasan, K.R. Tartaro, A. Tomasz, K.A. Vogt, G.A. Volkmer, S.H. Wang, Y. Wang, M.P. Weiner, P. Yu, R.F. Begley, J.M. Rothberg, Genome sequencing in microfabricated high-density picolitre reactors, Nature. 437 (2005) 376–380.

[63] S.T. Bennett, C. Barnes, A. Cox, L. Davies, C. Brown, Toward the 1,000 dollars human genome, Pharmacogenomics. 6 (2005) 373–382.

[64] J.C. Marioni, C.E. Mason, S.M. Mane, M. Stephens, Y. Gilad, RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays, Genome Res. 18 (2008) 1509–1517.

[65] H.E.J. Yong, S.-Y. Chan, Current approaches and developments in transcript profiling of the human placenta, Hum. Reprod. Update. (2020). https://doi.org/10.1093/humupd/dmaa028.

[66] D.S. Rogawski, N.A. Vitanza, A.C. Gauthier, V. Ramaswamy, C. Koschmann, Integrating RNA sequencing into neuro-oncology practice, Transl. Res. 189 (2017)

93–104.

[67] A. Dobin, C.A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, T.R. Gingeras, STAR: ultrafast universal RNA-seq aligner, Bioinformatics. 29 (2013) 15–21.

[68] T.H. Webster, M. Couse, B.M. Grande, E. Karlins, T.N. Phung, P.A. Richmond, W. Whitford, M.A. Wilson, Identifying, understanding, and correcting technical artifacts on the sex chromosomes in next-generation sequencing data, Gigascience. 8 (2019). https://doi.org/10.1093/gigascience/giz074.

[69] I. Fajardy, E. Moitrot, A. Vambergue, M. Vandersippe-Millot, P. Deruelle, J. Rousseaux, Time course analysis of RNA stability in human placenta, BMC Mol. Biol. 10 (2009) 21.

[70] M.L. Jobarteh, S.E. Moore, C. Kennedy, L. Gambling, H.J. McArdle, The effect of delay in collection and processing on RNA integrity in human placenta: experiences from rural Africa, Placenta. 35 (2014) 72–74.

[71] A.M.L. Coenen-Stass, I. Magen, T. Brooks, I.Z. Ben-Dov, L. Greensmith, E. Hornstein, P. Fratta, Evaluation of methodologies for microRNA biomarker detection by next generation sequencing, RNA Biol. 15 (2018) 1133–1145.

[72] K. Shah, C.E. McCormack, N.A. Bradbury, Do you know the sex of your cells?, Am. J. Physiol. Cell Physiol. 306 (2014) C3–18.

[73] J.C. Moreira de Mello, E.S.S. de Araújo, R. Stabellini, A.M. Fraga, J.E.S. de Souza, D.R. Sumita, A.A. Camargo, L.V. Pereira, Random X inactivation and extensive mosaicism in human placenta revealed by analysis of allele-specific gene expression along the X chromosome, PLoS One. 5 (2010) e10947.

[74] T. Chu, D. Handley, K. Bunce, U. Surti, W.A. Hogge, D.G. Peters, Structural and regulatory characterization of the placental epigenome at its maternal interface, PLoS One. 6 (2011) e14723.

[75] R. Vento-Tormo, M. Efremova, R.A. Botting, M.Y. Turco, M. Vento-Tormo, K.B. Meyer, J.-E. Park, E. Stephenson, K. Polański, A. Goncalves, L. Gardner, S. Holmqvist, J. Henriksson, A. Zou, A.M. Sharkey, B. Millar, B. Innes, L. Wood, A. Wilbrey-Clark, R.P. Payne, M.A. Ivarsson, S. Lisgo, A. Filby, D.H. Rowitch, J.N. Bulmer, G.J. Wright, M.J.T. Stubbington, M. Haniffa, A. Moffett, S.A. Teichmann, Single-cell reconstruction of the early maternal–fetal interface in humans, Nature. 563 (2018) 347–353.

[76] M. Pavličev, G.P. Wagner, A.R. Chavan, K. Owens, J. Maziarz, C. Dunn-Fletcher, S.G. Kallapur, L. Muglia, H. Jones, Single-cell transcriptomics of the human placenta: inferring the cell communication network of the maternal-fetal interface, Genome Res. 27 (2017) 349–361.

[77] Y. Liu, X. Fan, R. Wang, X. Lu, Y.-L. Dang, H. Wang, H.-Y. Lin, C. Zhu, H. Ge, J.C. Cross, H. Wang, Single-cell RNA-seq reveals the diversity of trophoblast subtypes and patterns of differentiation in the human placenta, Cell Res. 28 (2018) 819–832.

[78] B. Lv, Q. An, Q. Zeng, X. Zhang, P. Lu, Y. Wang, X. Zhu, Y. Ji, G. Fan, Z. Xue, Single-cell RNA sequencing reveals regulatory mechanism for trophoblast cell-fate divergence in human peri-implantation conceptuses, PLoS Biol. 17 (2019) e3000187.

[79] J.C.H. Tsang, J.S.L. Vong, L. Ji, L.C.Y. Poon, P. Jiang, K.O. Lui, Y.-B. Ni, K.F. To, Y.K.Y. Cheng, R.W.K. Chiu, Y.M.D. Lo, Integrative single-cell and cell-free plasma RNA transcriptomics elucidates placental cellular dynamics, Proc. Natl. Acad. Sci. U. S. A. 114 (2017) E7786–E7795.

[80] H. Suryawanshi, P. Morozov, A. Straus, N. Sahasrabudhe, K.E. A Max, A. Garzia, M. Kustagi, T. Tuschl, Z. Williams, A single-cell survey of the human first-trimester placenta and decidua, Science Advances. 4 (2018). http://advances.sciencemag.org/.

[81] H. Okae, H. Toh, T. Sato, H. Hiura, S. Takahashi, K. Shirane, Y. Kabayama, M. Suyama, H. Sasaki, T. Arima, Derivation of Human Trophoblast Stem Cells, Cell Stem Cell. 22 (2018) 50–63.e6.

[82] C. Vaillancourt, D. Lanoix, F. Le Bellego, G. Daoud, J. Lafond, Involvement of MAPK signalling in human villous trophoblast differentiation, Mini Rev. Med. Chem. 9 (2009) 962–973.

[83] S. Cvitic, M.S. Longtine, H. Hackl, K. Wagner, M.D. Nelson, G. Desoye, U. Hiden, The human placental sexome differs between trophoblast epithelium and villous vessel endothelium, PLoS One. 8 (2013) e79233.

[84] J.K. Cleal, P. Day, M.A. Hanson, R.M. Lewis, Measurement of housekeeping genes in human placenta, Placenta. 30 (2009) 1002–1003.

[85] J.K. Cleal, P.L. Day, M.A. Hanson, R.M. Lewis, Sex differences in the mRNA levels of housekeeping genes in human placenta, Placenta. 31 (2010) 556–557.

[86] V.L. Clifton, V.E. Murphy, Maternal asthma as a model for examining fetal sex-specific effects on maternal physiology and placental mechanisms that regulate human fetal growth, Placenta. 25 Suppl A (2004) S45–52.

[87] A. Bouman, M.J. Heineman, M.M. Faas, Sex hormones and the immune response in humans, Hum. Reprod. Update. 11 (2005) 411–423.

[88] E.N. Fish, The X-files in immunity: Sex-based differences predispose immune responses, Nature Reviews Immunology. 8 (2008) 737–744. https://doi.org/10.1038/nri2394.

[89] E. Martin, L. Smeester, P.A. Bommarito, M.R. Grace, K. Boggess, K. Kuban, M.R. Karagas, C.J. Marsit, T.M. O’Shea, R.C. Fry, Sexual epigenetic dimorphism in the human placenta: Implications for susceptibility during the prenatal period, Epigenomics. 9 (2017) 267–278.

[90] V.L. Clifton, Review: Sex and the human placenta: mediating differential strategies of fetal growth and survival, Placenta. 31 Suppl (2010) S33–9.

[91] A. Edwards, A. Megens, M. Peek, E.M. Wallace, Sexual origins of placental dysfunction, Lancet. 355 (2000) 203–204.

[92] P. Bermejo-Alvarez, D. Rizos, D. Rath, P. Lonergan, A. Gutierrez-Adan, Sex determines the expression level of one third of the actively expressed genes in bovine blastocysts, Proc. Natl. Acad. Sci. U. S. A. 107 (2010) 3394–3399.

[93] D.P. Sundrani, S.S. Roy, A.T. Jadhav, S.R. Joshi, Sex-specific differences and developmental programming for diseases in later life, Reprod. Fertil. Dev. 29 (2017) 2085–2099.

[94] P. Toivanen, T. Hirvonen, Sex ratio of newborns: preponderance of males in toxemia of pregnancy, Science. 170 (1970) 187–188.

[95] W.H. James, Sex ratios of offspring and the causes of placental pathology, Hum. Reprod. 10 (1995) 1403–1406.

[96] M.H. Aliyu, H.M. Salihu, O. ’neil Lynch, A.P. Alio, P.J. Marty, Fetal sex and differential survival in preeclampsia and eclampsia, Arch. Gynecol. Obstet. 285 (2012) 361–365.

[97] M. Makhseed, V.M. Musini, M.A. Ahmed, Association of fetal gender with pregnancy-induced hypertension and pre-eclampsia, Int. J. Gynaecol. Obstet. 63 (1998) 55–56.

[98] C.D. Hsu, T.R. Johnson, F.R. Witter, Fetal gender effect on pre-eclampsia: a retrospective study of Baltimore area in the United States of America, Int. J. Gynaecol. Obstet. 45 (1994) 160–161.

[99] A. Shiozaki, Y. Matsuda, S. Satoh, S. Saito, Impact of fetal sex in pregnancy-induced hypertension and preeclampsia in Japan, J. Reprod. Immunol. 89 (2011) 133–139.

[100] D. Jaskolka, R. Retnakaran, B. Zinman, C.K. Kramer, Fetal sex and maternal risk of pre-eclampsia/eclampsia: a systematic review and meta-analysis, BJOG. 124 (2017) 553–560.

[101] H. Mirzakhani, S.T. Weiss, Fetal sex and risk of preeclampsia: Dose maternal race matter?, J. Matern. Fetal. Neonatal Med. (2020) 1–9.

[102] C.K. Walker, P. Krakowiak, A. Baker, R.L. Hansen, S. Ozonoff, I. Hertz-Picciotto, Preeclampsia, placental insufficiency, and autism spectrum disorder or developmental delay, JAMA Pediatr. 169 (2015) 154–162.

[103] S. Muralimanoharan, A. Maloyan, L. Myatt, Evidence of sexual dimorphism in the placental function with severe preeclampsia, Placenta. 34 (2013) 1183–1189.

[104] I.D. Grant, D.A. Giussani, C.E. Aiken, Blood pressure and hypertensive disorders of pregnancy at high altitude: a systematic review and meta-analysis, Am J Obstet Gynecol MFM. 3 (2021) 100400.

[105] A. Osei-Kumah, R. Smith, I. Jurisica, I. Caniggia, V.L. Clifton, Sex-specific differences in placental global gene expression in pregnancies complicated by asthma, Placenta. 32 (2011) 570–578.

[106] A. Schmidt, D.M. Morales-Prieto, J. Pastuschek, K. Fröhlich, U.R. Markert, Only humans have human placentas: Molecular differences between mice and humans, J. Reprod. Immunol. (2015). https://doi.org/10.1016/j.jri.2015.03.001.

[107] R.-H. Xu, X. Chen, D.S. Li, R. Li, G.C. Addicks, C. Glennon, T.P. Zwaka, J.A. Thomson, BMP4 initiates human embryonic stem cell differentiation to trophoblast,

Nat. Biotechnol. 20 (2002) 1261–1264.

[108] T. Nagamatsu, T. Fujii, T. Ishikawa, T. Kanai, H. Hyodo, T. Yamashita, Y. Osuga, M. Momoeda, S. Kozuma, Y. Taketani, A primary cell culture system for human cytotrophoblasts of proximal cytotrophoblast cell columns enabling in vitro acquisition of the extra-villous phenotype, Placenta. 25 (2004) 153–165.

[109] R. Harun, L. Ruban, M. Matin, J. Draper, N.M. Jenkins, G.C. Liew, P.W. Andrews, T.C. Li, S.M. Laird, H.D.M. Moore, Cytotrophoblast stem cell lines derived from human embryonic stem cells and their capacity to mimic invasive implantation events, Hum. Reprod. 21 (2006) 1349–1358.

[110] J.L. James, P.R. Stone, L.W. Chamley, The isolation and characterization of a population of extravillous trophoblast progenitors from first trimester human placenta, Hum. Reprod. 22 (2007) 2111–2119.

[111] O. Genbacev, M. Donne, M. Kapidzic, M. Gormley, J. Lamb, J. Gilmore, N. Larocque, G. Goldfien, T. Zdravkovic, M.T. McMaster, S.J. Fisher, Establishment of human trophoblast progenitor cell lines from the chorion, Stem Cells. 29 (2011) 1427–1436.

[112] M. Marchand, J.A. Horcajadas, F.J. Esteban, S.L. McElroy, S.J. Fisher, L.C. Giudice, Transcriptomic signature of trophoblast differentiation in a human embryonic stem cell model, Biol. Reprod. 84 (2011) 1258–1271.

[113] T. Takao, K. Asanoma, K. Kato, K. Fukushima, R. Tsunematsu, T. Hirakawa, S. Matsumura, H. Seki, S. Takeda, N. Wake, Isolation and characterization of human trophoblast side-population (SP) cells in primary villous cytotrophoblasts and HTR-8/SVneo cell line, PLoS One. 6 (2011) e21990.

[114] R. Udayashankar, D. Baker, E. Tuckerman, S. Laird, T.C. Li, H.D. Moore, Characterization of invasive trophoblasts generated from human embryonic stem cells, Hum. Reprod. 26 (2011) 398–406.

[115] M. Amita, K. Adachi, A.P. Alexenko, S. Sinha, D.J. Schust, L.C. Schulz, R.M. Roberts, T. Ezashi, Complete and unidirectional conversion of human embryonic stem cells to trophoblast by BMP4, Proc. Natl. Acad. Sci. U. S. A. 110 (2013) E1212–21.

[116] P. Das, T. Ezashi, L.C. Schulz, S.D. Westfall, K.A. Livingston, R. Michael Roberts, Effects of FGF2 and oxygen in the BMP4-driven differentiation of trophoblast from human embryonic stem cells, Stem Cell Research. 1 (2007) 61–74. https://doi.org/10.1016/j.scr.2007.09.004.

[117] R.M. Roberts, K.M. Loh, M. Amita, A.S. Bernardo, K. Adachi, A.P. Alexenko, D.J. Schust, L.C. Schulz, B.P.V.L. Telugu, T. Ezashi, Others, Differentiation of trophoblast cells from human embryonic stem cells: to be or not to be?, Reproduction. 147 (2014) D1–D12.

[118] R.B. Donker, J.F. Mouillet, T. Chu, C.A. Hubel, D.B. Stolz, A.E. Morelli, Y. Sadovsky, The expression profile of C19MC microRNAs in primary human

trophoblast cells and exosomes, Mol. Hum. Reprod. 18 (2012) 417–424.

[119] D.M. Morales-Prieto, S. Ospina-Prieto, W. Chaiwangyen, M. Schoenleben, U.R. Markert, Pregnancy-associated miRNA-clusters, J. Reprod. Immunol. 97 (2013) 51–61.

[120] J.L. Reiter, H.M. Drendel, S. Chakraborty, M.M. Schellinger, M.-J. Lee, G. Mor, Cytogenetic features of human trophoblast cell lines SWAN-71 and 3A-subE, Placenta. 52 (2017) 17–20.

[121] J. Xu, X. Peng, Y. Chen, Y. Zhang, Q. Ma, L. Liang, A.C. Carter, X. Lu, C.-I. Wu, Free-living human cells reconfigure their chromosomes in the evolution back to uni-cellularity, Elife. 6 (2017). https://doi.org/10.7554/eLife.28070.

[122] S. Al-Nasiry, B. Spitz, M. Hanssens, C. Luyten, R. Pijnenborg, Differential effects of inducers of syncytialization and apoptosis on BeWo and JEG-3 choriocarcinoma cells, Hum. Reprod. 21 (2006) 193–201.

[123] R.A. Pattillo, G.O. Gey, The establishment of a cell line of human hormone-synthesizing trophoblastic cells in vitro, Cancer Res. 28 (1968) 1231–1236.

[124] R.A. Pattillo, G.O. Gey, E. Delfs, R.F. Mattingly, Human hormone production in vitro, Science. 159 (1968) 1467–1469.

[125] M. Morgan, D. Kniss, S. McDonnell, Expression of metalloproteinases and their inhibitors in human trophoblast continuous cell lines, Exp. Cell Res. 242 (1998) 18–26.

[126] E.M. Diss, S.G. Gabbe, J.W. Moore, D.A. Kniss, Study of thromboxane and prostacyclin metabolism in an invitro model of first-trimester human trophoblast, Am. J. Obstet. Gynecol. 167 (1992) 1046–1052.

[127] K.J. Lei, Y. Gluzman, C.J. Pan, J.Y. Chou, Immortalization of virus-free human placental cells that express tissue-specific functions, Mol. Endocrinol. 6 (1992) 703–712.

[128] C.K. Ho, S.Y. Li, K.J. Yu, C.C. Wang, H. Chiang, S.Y. Wang, Characterization of a human tumorigenic, poorly differentiated trophoblast cell line, In Vitro Cell. Dev. Biol. Anim. 30A (1994) 415–417.

[129] G. Aboagye-Mathiesen, J. Laugesen, M. Zdravkovic, P. Ebbesen, Isolation and characterization of human placental trophoblast subpopulations from first-trimester chorionic villi, Clin. Diagn. Lab. Immunol. 3 (1996) 14–22.

[130] J. Pastuschek, O. Nonn, R.N. Gutiérrez-Samudio, J.M. Murrieta-Coxca, J. Müller, J. Sanft, B. Huppertz, U.R. Markert, T. Groten, D.M. Morales-Prieto, Molecular characteristics of established trophoblast-derived cell lines, Placenta. 108 (2021) 122–133.

[131] C.H. Graham, T.S. Hawley, R.G. Hawley, J.R. MacDougall, R.S. Kerbel, N. Khoo, P.K. Lala, Establishment and characterization of first trimester human trophoblast cells with extended lifespan, Exp. Cell Res. 206 (1993) 204–211.

[132] R.O. Hussa, M.T. Story, R.A. Pattillo, Regulation of human chorionic gonadotropin

(hCG) secretion by serum and dibutyryl cyclic AMP in malignant trophoblast cells in vitro, J. Clin. Endocrinol. Metab. 40 (1975) 401–405.

[133] P.O. Kohler, W.E. Bridson, Isolation of hormone-producing clonal lines of human choriocarcinoma, J. Clin. Endocrinol. Metab. 32 (1971) 683–687.

[134] I. Shih, T. Wang, T. Wu, R.J. Kurman, J.D. Gearhart, Expression of Mel-CAM in implantation site intermediate trophoblastic cell line, IST-1, limits its migration on uterine smooth muscle cells, J. Cell Sci. 111 (Pt 17) (1998) 2655–2664.

[135] L. Rong-Hao, S. Luo, L.Z. Zhuang, Establishment and characterization of a cytotrophoblast cell line from normal placenta of human origin, Hum. Reprod. 11 (1996) 1328–1333.

[136] N.K.S. Khoo, J.F. Bechberger, T. Shepherd, S.L. Bond, K.R. McCrae, G. Scot Hamilton, P.K. Lala, SV40 Tag transformation of the normal invasive trophoblast results in a premalignant phenotype. I. Mechanisms responsible for hyperinvasivess and resistance to anti-invasive action of TGFβ, International Journal of Cancer. 77 (1998) 429–439. https://doi.org/10.1002/(sici)1097-0215(19980729)77:3<429::aid-ijc20>3.0.co;2-1.

[137] M.Y. Choy, I.T. Manyonda, The phagocytic activity of human first trimester extravillous trophoblast, Hum. Reprod. 13 (1998) 2941–2949.

[138] J.Y. Chou, Establishment of clonal human placental cells synthesizing human choriogonadotropin, Proceedings of the National Academy of Sciences. 75 (1978) 1854–1858. https://doi.org/10.1073/pnas.75.4.1854.

[139] S.L. Straszewski-Chavez, V.M. Abrahams, A.B. Alvero, P.B. Aldo, Y. Ma, S. Guller, R. Romero, G. Mor, The isolation and characterization of a novel telomerase immortalized first trimester trophoblast cell line, Swan 71, Placenta. 30 (2009) 939–948.

[140] M.P. Lewis, M. Clements, S. Takeda, P.L. Kirby, H. Seki, L.B. Lonsdale, M.H. Sullivan, M.G. Elder, J.O. White, Partial characterization of an immortalized human trophoblast cell-line, TCL-1, which possesses a CSF-1 autocrine loop, Placenta. 17 (1996) 137–146.

[141] C.K. Ho, H. Chiang, S.Y. Li, C.C. Yuan, H.T. Ng, Establishment and characterization of a tumorigenic trophoblast-like cell line from a human placenta, Cancer Res. 47 (1987) 3220–3224.

[142] R. Jansen, S. Batista, A.I. Brooks, J.A. Tischfield, G. Willemsen, G. van Grootheest, J.-J. Hottenga, Y. Milaneschi, H. Mbarek, V. Madar, W. Peyrot, J.M. Vink, C.L. Verweij, E.J.C. de Geus, J.H. Smit, F.A. Wright, P.F. Sullivan, D.I. Boomsma, B.W.J.H. Penninx, Sex differences in the human peripheral blood transcriptome, BMC Genomics. 15 (2014) 33.

[143] Y. Yaron, I. Wolman, M.J. Kupferminc, Y. Ochshorn, A. Many, A. Orr-Urtreger, Effect of fetal gender on first trimester markers and on Down syndrome screening, Prenat. Diagn. 21 (2001) 1027–1030.

[144] P.J. Willems, H. Dierickx, E. Vandenakker, D. Bekedam, N. Segers, K. Deboulle, A. Vereecken, The first 3,000 Non-Invasive Prenatal Tests (NIPT) with the Harmony test in Belgium and the Netherlands, Facts Views Vis Obgyn. 6 (2014) 7–12.

[145] M. Ehrich, C. Deciu, T. Zwiefelhofer, J.A. Tynan, L. Cagasan, R. Tim, V. Lu, R. McCullough, E. McCarthy, A.O.H. Nygren, J. Dean, L. Tang, D. Hutchison, T. Lu, H. Wang, V. Angkachatchai, P. Oeth, C.R. Cantor, A. Bombard, D. van den Boom, Noninvasive detection of fetal trisomy 21 by sequencing of DNA in maternal blood: a study in a clinical setting, Am. J. Obstet. Gynecol. 204 (2011) 205.e1–11.

[146] S.S.C. Chim, T.K.F. Shing, E.C.W. Hung, T.-Y. Leung, T.-K. Lau, R.W.K. Chiu, Y.M.D. Lo, Detection and characterization of placental microRNAs in maternal plasma, Clin. Chem. 54 (2008) 482–490.

[147] S.-S. Luo, O. Ishibashi, G. Ishikawa, T. Ishikawa, A. Katayama, T. Mishima, T. Takizawa, T. Shigihara, T. Goto, A. Izumi, A. Ohkuchi, S. Matsubara, T. Takeshita, T. Takizawa, Human Villous Trophoblasts Express and Secrete Placenta-Specific MicroRNAs into Maternal Circulation via Exosomes1, Biol. Reprod. 81 (2009) 717–729.

[148] Q. Ge, H. Li, Q. Yang, J. Lu, J. Tu, Y. Bai, Z. Lu, Sequencing circulating miRNA in maternal plasma with modified library preparation, Clin. Chim. Acta. 412 (2011) 1989–1994.

[149] K. Miura, S. Miura, K. Yamasaki, A. Higashijima, A. Kinoshita, K.-I. Yoshiura, H. Masuzaki, Identification of pregnancy-associated microRNAs in maternal plasma, Clin. Chem. 56 (2010) 1767–1771.

[150] Z. Williams, I.Z. Ben-Dov, R. Elias, A. Mihailovic, M. Brown, Z. Rosenwaks, T. Tuschl, Comprehensive profiling of circulating microRNA via small RNA sequencing of cDNA libraries reveals biomarker potential and limitations, Proceedings of the National Academy of Sciences. (2013). https://doi.org/10.1073/pnas.1214046110.

[151] S. Gilad, E. Meiri, Y. Yogev, S. Benjamin, D. Lebanony, N. Yerushalmi, H. Benjamin, M. Kushnir, H. Cholakh, N. Melamed, Z. Bentwich, M. Hod, Y. Goren, A. Chajut, Serum microRNAs are promising novel biomarkers, PLoS One. 3 (2008) e3148.

[152] G.P. Brennan, D.M. Vitsios, S. Casey, A.-M. Looney, B. Hallberg, D.C. Henshall, G.B. Boylan, D.M. Murray, C. Mooney, RNA-sequencing analysis of umbilical cord plasma microRNAs from healthy newborns, PLoS One. 13 (2018) e0207952.

[153] M. Gol, B. Tuna, E. Dogan, B. Gulekli, M. Bagci, S. Altunyurt, U. Saygili, Does fetal gender affect cytotrophoblast cell activity in the human term placenta? Correlation with maternal hCG levels, Acta Obstet. Gynecol. Scand. 83 (2004) 711–715.

[154] J. Harrow, A. Frankish, J.M. Gonzalez, E. Tapanari, M. Diekhans, F. Kokocinski, B.L. Aken, D. Barrell, A. Zadissa, S. Searle, I. Barnes, A. Bignell, V. Boychenko, T. Hunt, M. Kay, G. Mukherjee, J. Rajan, G. Despacio-Reyes, G. Saunders, C. Steward, R. Harte, M. Lin, C. Howald, A. Tanzer, T. Derrien, J. Chrast, N. Walters, S.

Balasubramanian, B. Pei, M. Tress, J.M. Rodriguez, I. Ezkurdia, J. van Baren, M. Brent, D. Haussler, M. Kellis, A. Valencia, A. Reymond, M. Gerstein, R. Guigó, T.J. Hubbard, GENCODE: the reference human genome annotation for The ENCODE Project, Genome Res. 22 (2012) 1760–1774.

[155] C. Ober, D.A. Loisel, Y. Gilad, Sex-specific genetic architecture of human disease, Nature Reviews Genetics. 9 (2008) 911–922. https://doi.org/10.1038/nrg2415.

Chapter 3

Large-scale transcriptome-wide profiling of microRNAs in human placenta and maternal plasma at early to mid gestation

Statement of Authorship

Title of Paper	Large-scale transcriptome-wide profiling of microRNAs in human placenta and maternal plasma at early to mid gestation
Publication Status	☒ Published ☐ Accepted for Publication ☐ Submitted for Publication ☐ Unpublished and Unsubmitted work written in manuscript style
Publication Details	Melanie D. Smith, Katherine Pillman, Tanja Jankovic-Karasoulos, DaleMcAninch, Qianhui Wan, K. Justinian Bogias, Dylan McCullough, Tina Bianco-Miotto, JamesBreen & Claire T. Roberts (2021): Large-scale transcriptome-wide profiling of microRNAsin human placenta and maternal plasma at early to mid gestation, RNA Biology, DOI:10.1080/15476286.2021.1963105

Principal Author

Name of Principal Author (Candidate)	Melanie Smith
Contribution to the Paper	- All bioinformatics analyses - Drafting of manuscript - Producing figures / visualisation of data - Corresponding author on manuscript - Management of deposition of transcriptome data into public database (GEO)
Overall percentage (%)	90%
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.

Signature		Date	06/09/2021

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);

ii. permission is granted for the candidate in include the publication in the thesis; and

iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Katherine Pillman
Contribution to the Paper	- Expert miRNA analyses support - Development of experimental plans - Editing of manuscript drafts

Signature		Date	6/9/2021

Name of Co-Author	Tanja Jankovic-Karasoulos
Contribution to the Paper	- Performed experiments - Editing of manuscript drafts

Signature		Date	6/9/2021

Please cut and paste additional co-author panels here as required.

Name of Co-Author	Dale McAninch		
Contribution to the Paper	- Performed experiments - Editing of manuscript drafts		
Signature		Date	06/09/2021

Name of Co-Author	Qianhui Wan		
Contribution to the Paper	- Assisted with manuscript drafting		
Signature		Date	06/09/2021

Name of Co-Author	K.Justinian Bogias		
Contribution to the Paper	- Assisted with manuscript drafting		
Signature		Date	06/09/2021

Name of Co-Author	Dylan McCullough		
Contribution to the Paper	- Performed experiments		
Signature		Date	07/09/2021

Name of Co-Author	Tina Bianco-Miotto		
Contribution to the Paper	- Conceived and developed experimental plans - Editing of manuscript drafts		
Signature		Date	06/09/2021

Name of Co-Author	James Breen		
Contribution to the Paper	- Conceived and developed experimental plans - Editing of manuscript drafts		
Signature		Date	06/09/2021

Name of Co-Author	Claire T Roberts		
Contribution to the Paper	- Conceived and developed experimental plans - Editing of manuscript drafts		
Signature		Date	06/09/2021

RNA BIOLOGY
https://doi.org/10.1080/15476286.2021.1963105

Taylor & Francis
Taylor & Francis Group

RESEARCH PAPER

OPEN ACCESS Check for updates

Large-scale transcriptome-wide profiling of microRNAs in human placenta and maternal plasma at early to mid gestation

Melanie D. Smith [a,b,c], Katherine Pillman [d], Tanja Jankovic-Karasoulos [a,b,c], Dale McAninch [b], Qianhui Wan [c], K. Justinian Bogias [a,b], Dylan McCullough [c], Tina Bianco-Miotto [a,e], James Breen [a,b,f], and Claire T. Roberts [a,b,c]

[a]Robinson Research Institute, University of Adelaide, Adelaide, SA, Australia; [b]Adelaide Medical School, University of Adelaide, Adelaide, SA, Australia; [c]Flinders Health and Medical Research Institute, Flinders University, Bedford Park, SA, Australia; [d]Centre for Cancer Biology, University of South Australia/SA Pathology, Adelaide, SA, Australia; [e]School of Agriculture Food and Wine, Waite Research Institute, University of Adelaide, Adelaide, SA, Australia; [f]South Australian Genomics Centre, South Australian Health & Medical Research Institute, Adelaide, SA, Australia

ABSTRACT

MicroRNAs (miRNAs) are increasingly seen as important regulators of placental development and opportunistic biomarker targets. Given the difficulty in obtaining samples from early gestation and subsequent paucity of the same, investigation of the role of miRNAs in early gestation human placenta has been limited. To address this, we generated miRNA profiles using 96 placentas from presumed normal pregnancies, across early gestation, in combination with matched profiles from maternal plasma. Placenta samples range from 6 to 23 weeks' gestation, a time period that includes placenta from the early, relatively low but physiological (6–10 weeks' gestation) oxygen environment, and later, physiologically normal oxygen environment (11–23 weeks' gestation).

We identified 637 miRNAs with expression in 86 samples (after removing poor quality samples), showing a clear gestational age gradient from 6 to 23 weeks' gestation. We identified 374 differentially expressed (DE) miRNAs between placentas from 6–10 weeks' versus 11–23 weeks' gestation. We see a clear gestational age group bias in miRNA clusters C19MC, C14MC, miR-17 ~ 92 and paralogs, regions that also include many DE miRNAs. Proportional change in expression of placenta-specific miRNA clusters was reflected in maternal plasma.

The presumed introduction of oxygenated maternal blood into the placenta (between ~10 and 12 weeks' gestation) changes the miRNA profile of the chorionic villus, particularly in placenta-specific miRNA clusters. Data presented here comprise a clinically important reference set for studying early placenta development and may underpin the generation of minimally invasive methods for monitoring placental health.

ARTICLE HISTORY
Received 11 November 2020
Revised 15 July 2021
Accepted 29 July 2021

KEYWORDS
miRNA; placenta; pregnancy; C19MC; C14MC; DLK1-Dl03; miR-17~92

Introduction

The human placenta is a complex, rapidly developing and highly regulated organ shared between two genetically unique individuals, the mother and her foetus. Its function is vital for the transfer of nutrients, gases and wastes between maternal and foetal circulations, with dysfunction leading to placental insufficiencies and pregnancy complications [1,2]. Essential placental functions are mediated by the syncytiotrophoblast, large multinucleated cells covering the chorionic villi and providing a large surface area for bi-directional exchange of molecules between foetal and maternal circulations.

Early development of the human placenta occurs in a relatively low, but physiological, oxygen environment due to the occluding presence of extravillous cytotrophoblasts (EVTs) in the uterine spiral arterioles. The placenta transitions to that which is usually considered a physiologically normal, oxygen environment over several weeks from 10 weeks' gestation when EVTs become dislodged permitting maternal blood to flow into the placental intervillous space [1,3]. This period of rising oxygen tension is a critical time for placental development, as an appropriate response to the accompanying burst of oxidative stress is crucial to the ongoing success of the pregnancy [3].

Robust characterization of important regulatory mechanisms surrounding normal placental development can give insight into pregnancy health, and the potential development of pregnancy complications [2,4]. Identification of biomarkers of ectopic oxygenation or placental dysfunction leading to pregnancy complications is an important step in monitoring pregnancy health.

MicroRNAs (miRNAs) are an important class of RNA molecules that can help us to understand regulation of developmental pathways within human tissues. As such they are prime targets for biomarker research. miRNAs are short, approximately 20–24 nucleotide (nt) lengths of single-stranded, non-coding ribonucleic acid (RNA) originating in non-coding regions of the genome [5] and comprehensively reviewed in Bartel (2018) [6]. They most commonly function

CONTACT Melanie D. Smith ✉ melanie.smith@adelaide.edu.au;Robinson Research Institute, University of Adelaide, Adelaide, SA 5005, Australia; Adelaide Medical School, University of Adelaide, Adelaide SA 5005, Australia; Flinders Health and Medical Research Institute, Flinders University, Bedford Park SA 5042, Australia

Supplemental data for this article can be accessed here.

as post-transcriptional regulators of gene expression either by translational repression or targeted degradation by cleavage of mRNA transcripts in the cytoplasm [7,8]. miRNAs target mRNA regulatory sequences that are often located in the 3′ untranslated region (UTR) of the mRNA transcript with potential for multiple conserved target sites within a single mRNA [5,9]. Approximately half of known mammalian miRNAs are found within intergenic regions of the genome, with the remaining miRNAs mostly transcribed from the open reading frame of other genes, including long non-coding RNA (lncRNA) [10–12]. Adding to this complexity, miRNA may be transcribed as a single, or polycistronic unit [11].

The critical role of the placenta in successful pregnancy is clear, and there is also evidence linking developmental programming of chronic adult disease such as heart disease, diabetes and obesity with the placental phenotype *in utero* [4,13,14]. However, there are currently no non-invasive tests to determine which women are likely to develop pregnancy complications in routine use, nor is there a comprehensive reference set of miRNA expression during early pregnancy that can be used for development of pregnancy health biomarkers. In this study, we performed high-throughput sequencing to generate robust miRNA expression profiles for 96 placentas and matched maternal plasma samples between 6–23 weeks' gestation, creating a comprehensive set of miRNA expression profiles for presumably healthy placental function. We identified differential miRNA expression between 6–10 and 11–23 weeks' gestation which could reflect physiological changes occurring in this important time period. Using placenta-associated miRNA clusters found on chromosome 14, 19, 13, 7 and X, we also investigated how maternal plasma reflects placental profiles to enable future biomarker discoveries.

Methods

Sample collection

Placental chorionic villous tissue samples were obtained with informed, written consent from women undergoing elective terminations of otherwise healthy pregnancies. Samples of peripheral blood (6–9 mL) were collected from the same women into standard EDTA blood tubes at the time of termination and stored on ice until processed. Whole blood underwent centrifugation at 800 x g for 15 min at 4°C before plasma removal and then spun for a further 15 min. All samples were stored at −80°C until further processing. Prior to termination, accurate gestational age was determined using transvaginal ultrasonography. Termination samples were collected from the Pregnancy Advisory Centre (PAC), Woodville, South Australia. All placenta samples were assessed for gross morphology before inclusion in the study.

RNA extraction and library preparation

All placenta tissues were collected post termination and processed as soon as possible after collection and placed into RNALater (Thermo Fisher) within 15 min. Total RNA was isolated using a modified protocol of the RNeasy mini plus kit (Qiagen, Hilden, Germany) whereby the RW1 buffer, a proprietary component of the RNeasy Kit that eliminates small RNAs, is replaced with 100% ethanol. The sample is then subjected to size fractionation for small RNAs before sequencing. For plasma, miRNA was isolated from 200 µL plasma samples using the QIAGEN miRNA serum/plasma kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. All samples were stored at −80°C until further processing. miRNA real-time qPCR for the assessment of haemolysis was conducted by QIAGEN Genomic Services (Qiagen, Hilden, Germany). 2 µl RNA was reverse transcribed in 10 µL reactions using miRCURY LNA Kit (QIAGEN version5). Each RT was performed including an artificial RNA spike-in (UniSp6). cDNA was diluted 50 x and assayed in 10 µl PCR reactions according to the protocol for miRCURY LNA miRNA PCR; each miRNA was assayed once by qPCR using assays for miR-23a, miR-30 c, miR-103, miR-142-3p, and miR-451. In addition to these miRNA assays, the RNA spike-ins were assayed. The amplification was performed in a LightCycler® 480 Real-Time PCR System (Roche) in 384 well plates. The amplification curves were analysed using in-house software, both for the determination of the Cq (by the 2nd derivative method) and for melting curve analysis.

All placenta samples were genotyped for foetal sex using high resolution melt curve analysis of the gene that encodes amelogenin. Amelogenin is found on both the X (AMELX) and Y (AMELY) chromosomes, the X allele features a 3 bp deletion in exon 3 allowing the identification of samples with only X chromosomes (female) or X and Y chromosomes (male) prior to sequencing. Forward primer 5′-CCCTGGGCTCTGTAAAGAATAGTG-3′, reverse primer 5′-ATCAGAGCTTAAACTGGGAAGCTG-3. qPCR was performed using SSOFast EvaGreen Supermix (Bio-Rad, CA), primers at 250 nM final concentration, 5 ng of DNA per reaction, on a Bio-Rad CFX384 Real-Time PCR System. Cycling conditions: initial denaturation 98°C 30s, 40 cycles of 98°C for 5s and 60°C for 5s. High resolution melt curve analysis was performed from 65°C to 85°C with a 0.2°C increment every 10s. Melt curve between 65°C and 70°C was analysed using Bio-Rad Precision melt software (Bio-Rad, CA) to identify sex genotypes. Library preparation and sequencing was performed by Qiagen (Valencia, CA) using the QIAseq miRNA Library Kit and QIAseq miRNA 48 Index IL kits as per manufacturer's instructions. Amplified cDNA libraries underwent single-end sequencing by synthesis (Illumina 1.9).

High-throughput sequencing analysis

Alignment and analysis of miRNA sequencing data were processed from raw FASTQ files using the *bcbio-nextgen* pipeline [15]. Briefly, adapter detection and trimming were performed using Atropos [16]. Alignment was performed using STAR [17] using the human reference genome build GRCh37[18],(https://www.ncbi.nlm.nih.gov/grc. Quality control metrics were assessed using FastQC[19] (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) to check for per base sequence quality, sequence length distribution and duplication levels, and summarized using multiQC [20]. miRNA were clustered and collapsed using SeqCluster [21], and individual miRNAs detected using SeqBuster [15]. Annotation was performed using the miraligner [15] with miRBase version 21.0 [22,23] (http://www.mirbase.org) as the reference database.

Differential expression analysis

All profile and expression analyses were conducted in the R statistical environment (v.3.3.2), using the *edgeR* (v.3.16.5) and *limma* [24] (v.3.30.11) R/Bioconductor packages. *edgeR* [25] was used to filter miRNA with low expression and normalize for library composition bias. Batch effects, introduced by sequencing samples in multiple sequencing runs, were evident after normalization and corrected using a *limma* batch correction method [24]. All samples were then normalized using the Trimmed Mean of M values (TMM). For plasma, $\log_2$ transformed counts per million (CPM) were 1 to 1 paired with placenta data prior to creation of plasma figures. Briefly, after creating design matrices for both wanted (gestational age and tissue type) and unwanted (sample independent) effects, the *limma* method remove Batch Effects was employed to regress out the effect of patient derived variation from the matched plasma count.

Sample-weights and log transformation was performed using *limma* package [26] with the voom function used to estimate the mean-variance relationship between individual observations and then applied to the normalized log-count data. Differential expression analysis including, moderated F-statistic evaluation, adjusted p-value estimation and Log Fold Change (LFC) analysis were performed using a moderated t-test [27] with Benjamini-Hochberg (BH) multiple hypothesis test corrections [28]. The data were first pooled per gestational age group and then compared. After adjustment, expression of miRNAs was considered significantly different at FDR ≤ 0.05. The workflow, source code and input files associated with this research available at (https://github.com/mxhp75/earlyPlacentamiRNA_SeqProfile.git).

Results

Characterizing a comprehensive miRNA profile of chorionic villous tissue across early gestation to characterize the human placental miRNA profile across early to mid gestation, Illumina NextSeq 75 bp single-end read sequencing was performed on miRNA libraries from chorionic villous samples obtained from 96 singleton pregnancies following elective pregnancy terminations (44 females and 52 males bearing pregnancies). An average of 18.72 million reads were sequenced per sample (range ~10–29 million reads; Additional file 1: Table S1). miRNA counts were generated by mapping to human miRBase V21 [22]. Libraries with low sequencing depth were removed, including samples PAC0131 (14,595 reads) and PAC0071 (removed due to low mapping after initial read count of 1,819,847 reads). Subsequent unsupervised clustering (Additional file 2: Figure S1) using principal component analysis (PCA) showed eight additional samples (PAC0041, PAC0039, PAC0045, PAC0008, PAC0035, PAC0024, PAC0034, PAC0006) that segregated from the remaining 86 samples, most likely due to the presence of adjacent non placental villous tissue. This hypothesis was tested using matched DNA methylation profiles which confirmed the presence of decidual tissue [29], leading to the removal of these samples from subsequent analyses.

After filtering, miRNAs were aligned to the human GRCh37 reference genome and counts produced from 4,665 miRBase annotations [22], identifying an initial set of 2,032 expressed miRNAs. To remove sequencing noise, miRNAs with <5 reads across all samples were removed. A total of 1,422 miRNAs were identified with non-zero expression means across all 86 samples, representing a standard known miRBase miRNA expression set for early to mid gestation human placenta. After accounting for sequencing batch effects, very low abundance miRNAs were discarded by converting to $\log_2$ CPM values and removing miRNAs with expression <2.5 CPM in ≥29 samples (see methods), leaving a total of 637 robust miRNAs for downstream profile analyses. An additional 588 candidate *de novo* miRNA sequences were predicted through miRDeep2 analysis.

Of the 637 identified miRNAs, the top ten miRNAs with the highest mean expression across all samples in placenta were miR-30d-5p, miR-125a-5p, miR-517a-3p, miR-199a-3p, miR-26b-5p, miR-26a-5p, let-7a-5p, miR-21-5p, miR-126-3p and miR-516b-5p (Table 1a). For a full list of miRNAs identified and profiled in this study see Additional File 1: Table S2. A principal component analysis (PCA) of all 637 miRNAs (Figure 1) indicates a clear gradient across early gestation. The PCA identified 23.2% variance accounted for in PC1, with the top 10 miRNAs contributing to PCA dimension one (PC1) being: miR-519 c-3p, miR-328-3p, miR-519b-3p, miR-20a-5p, miR-19b-3p, miR-501-3p, miR-485-5p, miR-3605-3p and miR-515-5p, miR-106a-5p (Additional file 1: Table S3). Interestingly both −3p and −5p arms of the miR-20a transcript, previously associated with placental angiogenesis [30], contributed highly to the gestational age change associated with PC1.

Differentially expressed miRNAs across early gestation

Given the importance of oxygen tension in orchestrating the growth and differentiation of the placenta between 6 and 23 weeks' gestation, we separated the 86 samples in our cohort by their gestation relative to the presumed introduction of oxygenated maternal blood flow into the intervillous space. Using prior knowledge we designated that ≤10 weeks' (6–10 weeks') gestation be considered pre- and >10 weeks' (11–23 weeks') gestation be considered post-initiation of maternal blood flow into the intervillous space [3]. This enabled us to identify changes in miRNA expression across this time-point, and to independently assess highly expressed miRNA in each group. When miRNAs from all 86 samples were analysed according to expression prior to and post 10 weeks' gestation, after accounting for potentially confounding factors such as maternal age, maternal BMI, maternal smoking status, we identified 13 constitutively expressed miRNAs with consistent high expression ($\log_2$ CPM 14.35–15.60) in both gestation groups (Table 1b). Notably, 4 of the top 20 most highly

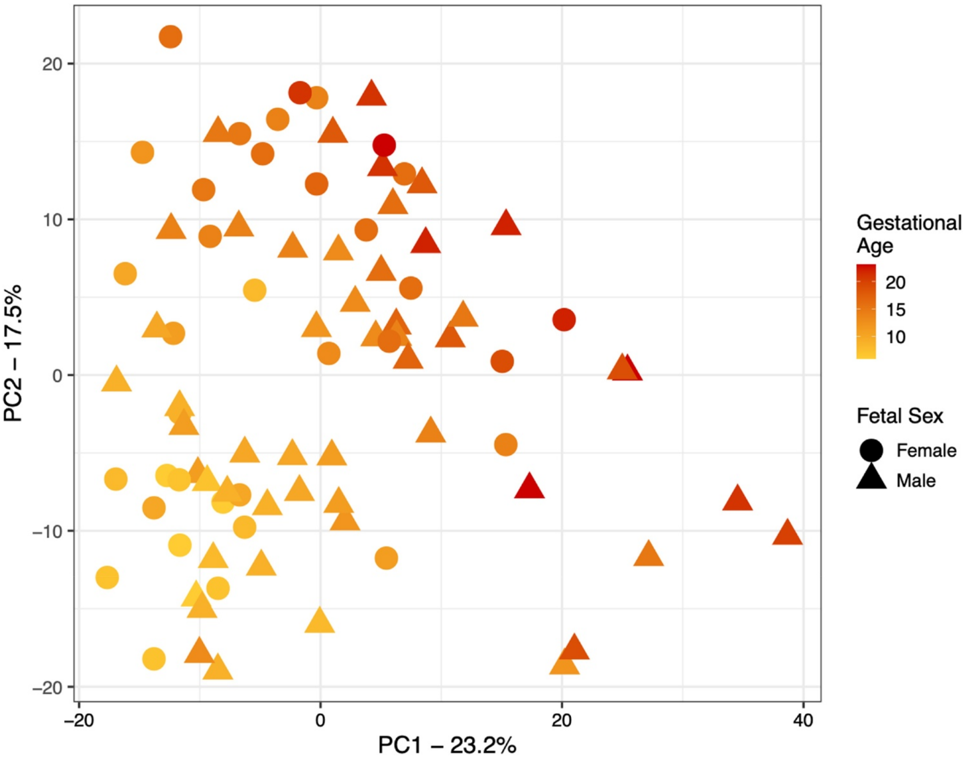


Figure 1. Unsupervised clustering of placental chorionic villous samples after batch correction. The PCA plot indicates a clear gestational age gradient from early (left) to later gestation (right) in the first dimension. This trend reflects the gestational age signature identified in our miRNA expression data.

Table 1. miRNA with highest expression in early gestation placenta. Global miRNA expression was calculated using the average expression across early gestation (6–23 weeks') and log_2 transformed. (**a**) Top ten miRNA with highest average global expression. (**b**) Constitutively highly expressed miRNA with membership in the top ten by highest average expression if calculated independently for the 6–10 weeks' and 11–23 weeks' gestation groups.

a

miRNA	miRNA Family	Avg. Global Expression (log_2 CPM)
miR-30d-5p	miR-30-5p	15.60
miR-125a-5p	miR-125-5p	15.51
miR-517a-3p	miR-517-5p	15.45
miR-199a-3p	miR-199-3p	15.43
miR-26b-5p	miR-26-5p	15.25
miR-26a-5p	miR-26-5p	14.84
let-7a-5p	let-7-5p/98-5p	14.47
miR-21-5p	miR-21-5p/590-5p	14.39
miR-126-3p	None listed	14.31
miR-516b-5p	miR-516b-5p	14.30

b

miRNA	miRNA Family	Avg. Expression (log_2 CPM)	
		6–10 weeks'	11–23 weeks'
miR-30d-5p	miR-30d-5p	15.53	15.64
miR-125a-5p	miR-125-5p	14.93	15.74
miR-517a-3p	miR-517-5p	15.52	15.40
miR-199a-3p	miR-199-3p	15.21	15.56
miR-26b-5p	miR-26-5p^	15.27	15.21
let-7a-5p	let-7-5p/98-5p	13.59	14.83
miR-21-5p	miR-21-5p/590-5p	14.20	14.59
miR-126-3p	None listed	13.91	14.56
miR-516b-5p	miR-516b-5p	13.23	14.31
miR-16-5p	miR-15-5p/16-5p/195-5p/424-5p/497-5p	14.34	14.32
miR-516a-5p	miR-516a-5p	14.40	14.11
miR-143-3p	miR-143-3p	13.60	14.52
miR-26a-5p	miR-26-5p^	14.54	14.99

^shared seed region

expressed miRNAs are from the placenta-specific chromosome 19 cluster, previously identified as being highly expressed in placenta [31–33].

After identifying the highly expressed miRNAs in placenta, we then investigated the change in expression that occurs after 10 weeks' gestation once maternal blood flow into the placenta is initiated and presumably when oxygen tension begins to rise. Differential expression analysis identified 374 significantly (FDR <0.05) different miRNAs, with 163 down-regulated and 211 up-regulated in the 11–23 weeks' gestation placenta compared to 6–10 weeks' gestation (Figure 2a; Additional file 1: Table S4). Strongly downregulated miRNAs (negative log_2FC) include miR-4483, miR-129-5p, miR-124-3p and miR-122-5p, and highest up-regulated (positive log_2FC) include miR-4645-3p, miR-195-3p, miR-137, miR-139-3p, miR-6715b-3p, miR-3927-3p and miR-1269b. Differentially expressed miRNAs had predominantly low expression (Figure 2b) with an average log_2 expression >10 only observed in miR-9-5p (log_2FC of −1.43, Figure 2c) and let-7a-5p (log_2FC of 1.22, Figure 2d). We see that for both let-7 and miR-9, the 5′ and 3′ arms follow a similar expression trajectory. Further examples are provided in Additional file 2, Figure S2. These mature fragments were originally transcribed as a single transcript before further processing (Figure 2e).

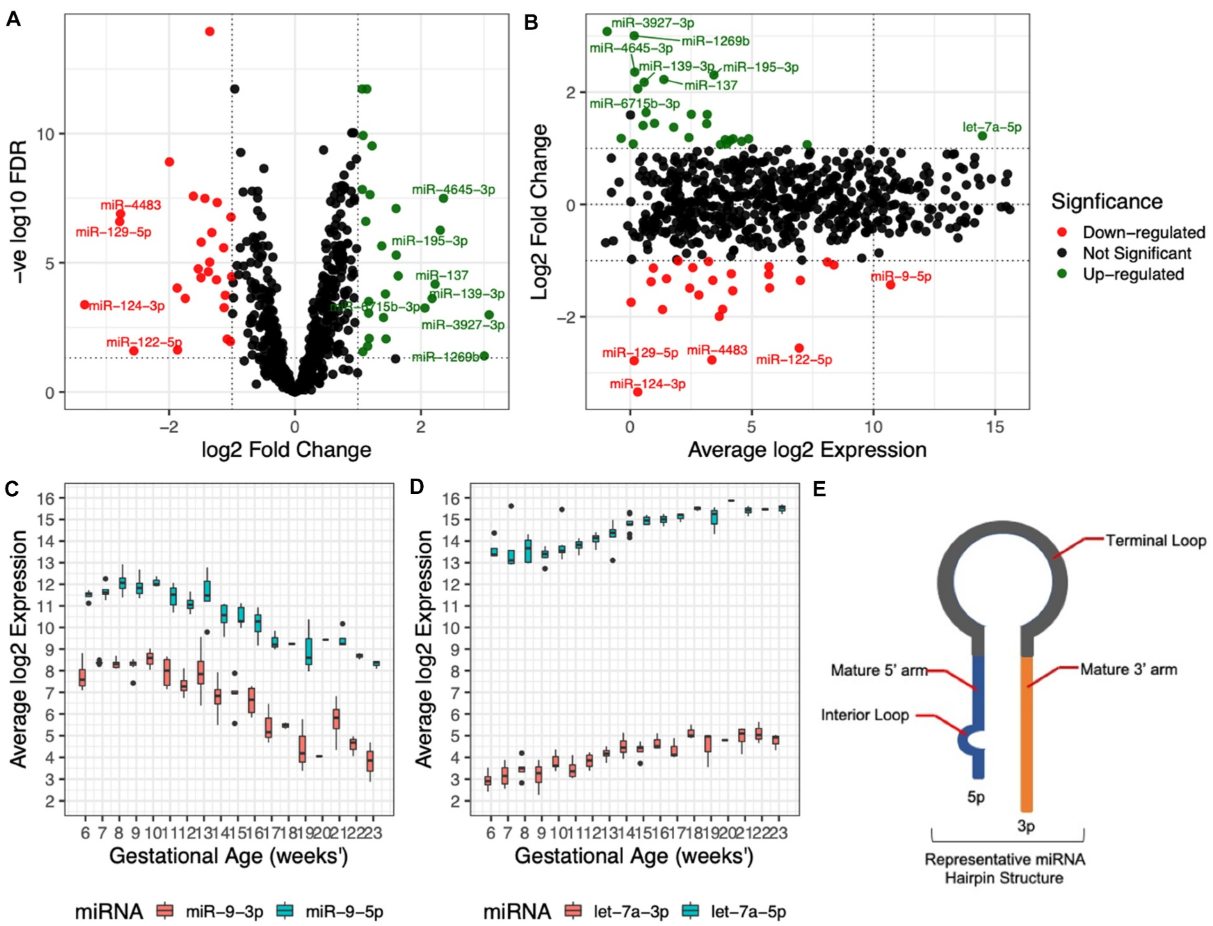


Figure 2. Differential expression analysis between 6–10 weeks' and 11–23 weeks' gestation placenta. (a) Volcano plot of differential expression. Linear regression identified 26 down-regulated with a $\log_2$FC <-1 (red) and 26 up-regulated with a $\log_2$FC >1 (green) miRNA between gestation groups. (b) MA plot of $\log_2$ fold change as a function of $\log_2$ average expression. Box plots of average expression as a function of gestational age for miR-9-5p/3p (c) and let-7a-5p/3p (d) show a clear changing pattern of expression from early to mid gestation. (e) Schematic of a representative miRNA hairpin structure indicates the origin of the 5′ and 3′ arms of the mature miRNA transcript.

Conserved placenta-associated miRNA clusters display variable expression between early and mid gestation

In vertebrate species miRNAs are significantly enriched in clusters, with ~50% of miRNA clusters being the result of random or non-local duplications. The placenta-associated clusters analysed herein included chromosome 19 miRNA cluster commonly known as C19MC (19q13.41) [34], chromosome 14 cluster (C14MC; 14q.32) [35], and chromosome 13 cluster (miR-17 ~ 92; 13q31.3) with its paralogs the chromosome X cluster miR-106a~363 (Xq26.2), and chromosome 7 cluster miR-106b~25 (7q22.1) [36]. We identified the expression of 48 mature miRNAs from C19MC, 77 from C14MC and 20 from miR-17 ~ 92 and paralogs. Of the 48 mature miRNAs identified in the C19MC cluster, 34 miRNA had significant (FDR <0.05) differential expression, that were down regulated in 11–23 weeks' compared to 6–10 weeks' gestation placenta (Additional file 1: Table S5). C19MC cluster expression, as a proportion of total miRNA transcripts, was reduced after 10 weeks' gestation (Figure 3a). Interestingly, while miRNAs in this cluster have been demonstrated to be transcribed by RNA-Pol II as a single polycistron [37], the individual members display a range of expression levels (Figure 3b).

The C14MC cluster, that contains 77 mature miRNAs, had 56 miRNAs with significant (FDR <0.05) differential expression. 55 of these miRNAs were up-regulated, while only one miRNA (miR-410-5p; FDR 1.99×10^{-3}) was down-regulated in the 11–23 weeks' gestation group (Additional file 1: Table S6). C14MC-member transcript expression, as a proportion of total expression, increased across early to mid gestation in placenta, further differing from C19MC finding by being more centred around the group mean (Figure 4a). Similar to C19MC cluster members, we found a range of expression levels amongst the C14MC members (Figure 4b) that are also thought to be transcribed as a single polycistronic transcript, suggesting more complex regulation of primary miRNA transcripts processing [38]. The miRNA miR-412-5p, previously observed as highly expressed in first trimester placenta [39], is notable in that it appears to have a bimodal distribution of expression (Figure 4c) that, upon further investigation, was not found to be related to sample characteristics such as foetal sex, maternal age, maternal BMI or maternal smoking status.

Of the 20 mature miRNAs in the miR-17 ~ 92 cluster and its paralogs, 17 had significant (FDR <0.05) differential

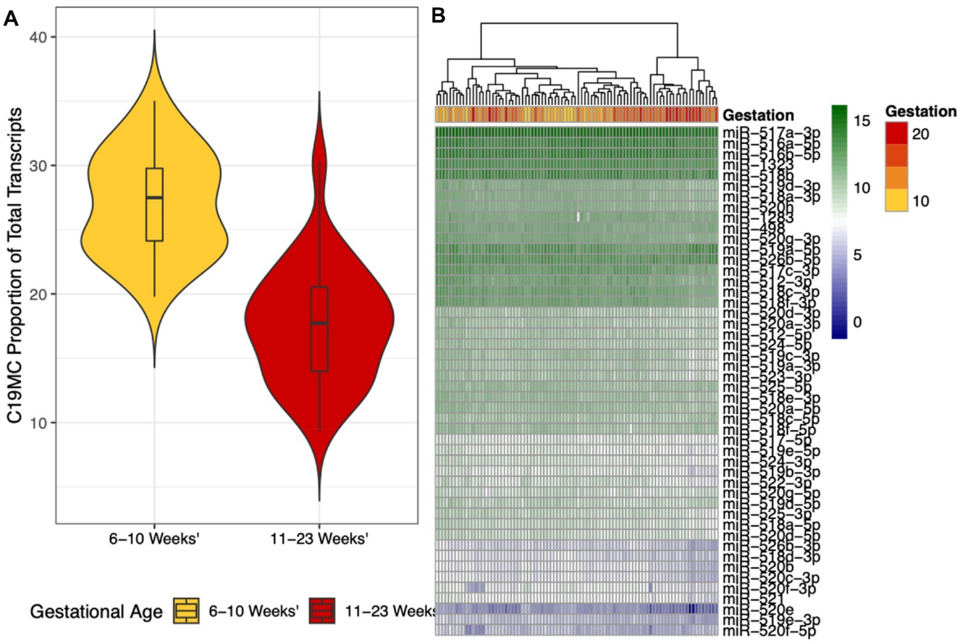


Figure 3. The proportion of C19MC transcripts decreased after 10 weeks' gestation in placenta. The proportion of combined C19MC member expression as a function of total miRNA expression is decreased in 11–23 weeks' compared to 6–10 weeks' gestation (violin plot) (a) with expression of individual cluster members tending to decrease across early to mid gestation (b).

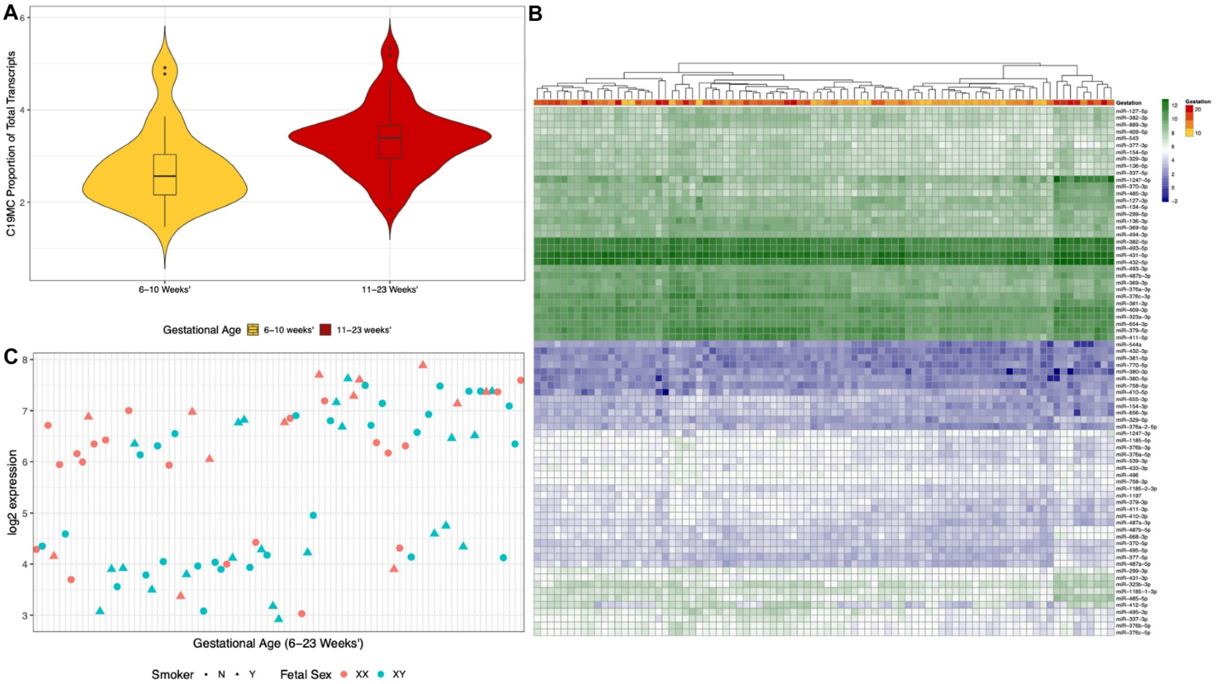


Figure 4. The proportion of C14MC transcripts increased in 11–23 weeks' compared to 6–10 weeks' gestation placenta. The proportion of C14MC member expression as a function of total miRNA expression is increased after 10 weeks' gestation compared to 6–10 weeks' gestation (violin plot) (a). Individual cluster members' expression is increased across early gestation in 55 of the 56 miRNAs (b). miR-412-5p appears to have a bimodal distribution which is unexplained by foetal sex, maternal age or maternal smoking status (c).

expression with 15 being down- and 2 being up-regulated in the 11–23 weeks' gestation group compared to the 6–10 weeks' (Additional file 1: Table S7). Investigation into the proportion of placenta expression originating from the miR-17 ~ 92 cluster and its paralogs revealed a pattern similar to that of the C19MC member transcripts with the proportion of transcripts identified in placenta tissue as a function of total transcripts decreasing across early gestation (Figure 5(a–c)). We identified a variation in expression patterns between each of the three paralogs (Figure 5d), with chromosome 13 expression decreasing across early to mid gestation.

Whilst we find significant (FDR <0.05) differential expression in all of the above miRNA clusters, the $\log_2$FC between 6–10 weeks' and 11–23 weeks' gestation groups is small (C19MC $\log_2$FC −1.02 to −0.19; C14MC $\log_2$FC 0.23 to 0.94 and $\log_2$FC 0.67; miR-17 ~ 92 $\log_2$FC −0.2 to −1.13 and $\log_2$ FC 0.84 to 0.89), with changes in the proportional expression of each cluster as a function of total miRNA expression suggesting an important regulatory mechanism for these clusters in placenta.

Characterizing a comprehensive profile of placental and endogenous miRNA in maternal plasma across early gestation

To characterize the miRNA profile across early to mid gestation in maternal plasma, we sequenced 96 1 to 1 matched maternal plasma samples (hereafter referred to as plasma) from women sampled at the time of pregnancy termination using the same sequencing protocol as detailed for the chorionic villous tissue. Plasma and tissue samples from each termination are assigned a common unique identifier, with a suffix used to separate sample types. miRNA sequencing resulted in approximately 17.77 million 75 bp single-end reads per sample in plasma (range 9,986,347–41,344,361 reads: Additional file 3: Table S8) across two sequencing runs. After initial quality control and sequence analysis, we identified a number of samples with aberrant expression of miR-451a and miR-16 indicating potential for red blood cell miRNA contamination. We performed a subsequent analysis using the delta-delta CT method (ΔΔCt) to calculate the relative fold change of gene expression of samples using quantitative polymerase chain reaction (qPCR). The ΔΔCt value of miR-23a minus miR-451 is examined, as miR-23a is known to be stable in samples affected by haemolysis whilst miR-451 is known to vary with contamination. This method identified 11 samples (PAC0033, PAC0034, PAC0041, PAC0048, PAC0050, PAC0051, PAC0054, PAC0056, PAC0062, PAC0084 and PAC0105) as more likely to contain haemolysis contamination, a common issue in plasma research. Despite potential issues involving haemolysis, we aimed to establish the presence of placenta-specific miRNAs within maternal plasma by including all available samples regardless of whether they had matched placenta data or not.

Alignment and filtering were performed using the same protocol as for chorionic villous tissue. The number of unique miRNA species identified per sample varied, ranging from 1022 miRNA species in PAC0077 to 355 miRNA species in PAC0010. Prior to filtering we identified 1775 unique mature miRNAs across all plasma samples, with a core set of 790 high confidence miRNAs present after filtering. Of these, 531 were also present in placenta samples, 259 unique to plasma, and 106 miRNAs unique to placenta (Additional file 3: Table S9). Unsupervised clustering (PCA) of all 790 high confidence miRNAs identified in plasma shows no gestational age gradient for maternal plasma from 6 to 23 weeks' gestation (Additional file 2: Figure S3).

Given the variability of miRNA in plasma, we hypothesized that the detection of placental miRNAs in plasma may be restricted to miRNA present in high concentrations in matched placenta. We therefore investigated whether 10 miRNAs (Table 1a) with the highest abundance in placenta were present in plasma and found that all ten miRNA were at high abundance levels in matched plasma (Figure 6, Table 2a). Many of the most highly expressed plasma miRNA are red blood cell (RBC) associated.

Of the 790 miRNAs identified in plasma, the top ten miRNAs with the highest global expression were miR-16-5p, miR-486-5p, miR-92a-3p, let-7a-5p, let-7b-5p, miR-451a, let-7 f-5p, miR-25-3p, let-7i-5p, miR-93-5p (Table 2b). These were also all highly expressed in placental tissue (average global expression of 8.95–14.47 Log_2 normalized CPM). miR-141, miR-200b, miR-139 and miR-184, previously identified in third trimester plasma, were also identified at high abundance in our early gestation plasma, with miR-141-3p abundance increasing across early gestation, as shown in previous studies [40].

We investigated plasma levels of the 21 miRNAs identified as both differentially and highly expressed in the matched placentas (Table 3). Whilst all 21 placental miRNAs were also detected in plasma, their origin is not necessarily placenta-specific, with cellular enrichment linked to various systemic compartments [41]. For example, five of these miRNAs have cellular enrichment in immune response, another three miRNAs in pluripotent and embryonic stem cells, and one miRNA is enriched in extra-embryonic stem cells. The latter may originate from the foetal membranes or other extra-embryonic tissue such as chorionic villi.

We further investigated the abundance of plasma miRNAs from three miRNA clusters previously identified as highly placenta-specific or placenta associated [34–36]. We found 44 mature miRNA from the C19MC cluster, 48 miRNA from the C14MC cluster and 20 miRNA from the miR-17 ~ 92 cluster and its paralogs in plasma. These cluster miRNAs have previously been the target of biomarker research in pregnancy health but to our knowledge these are not being used clinically.

The change in relative proportion of expression by miRNA cluster members in plasma across the 10–11 weeks' gestation threshold is difficult to detect with C14MC and C19MC members each representing less than 1% of the miRNA abundance quantified in plasma. As such the proportional change identified in placenta is not entirely reflected in plasma. Abundance levels of the miR-17 ~ 92 cluster and its paralogs represent approximately 0.03% of the total miRNA identified in plasma.

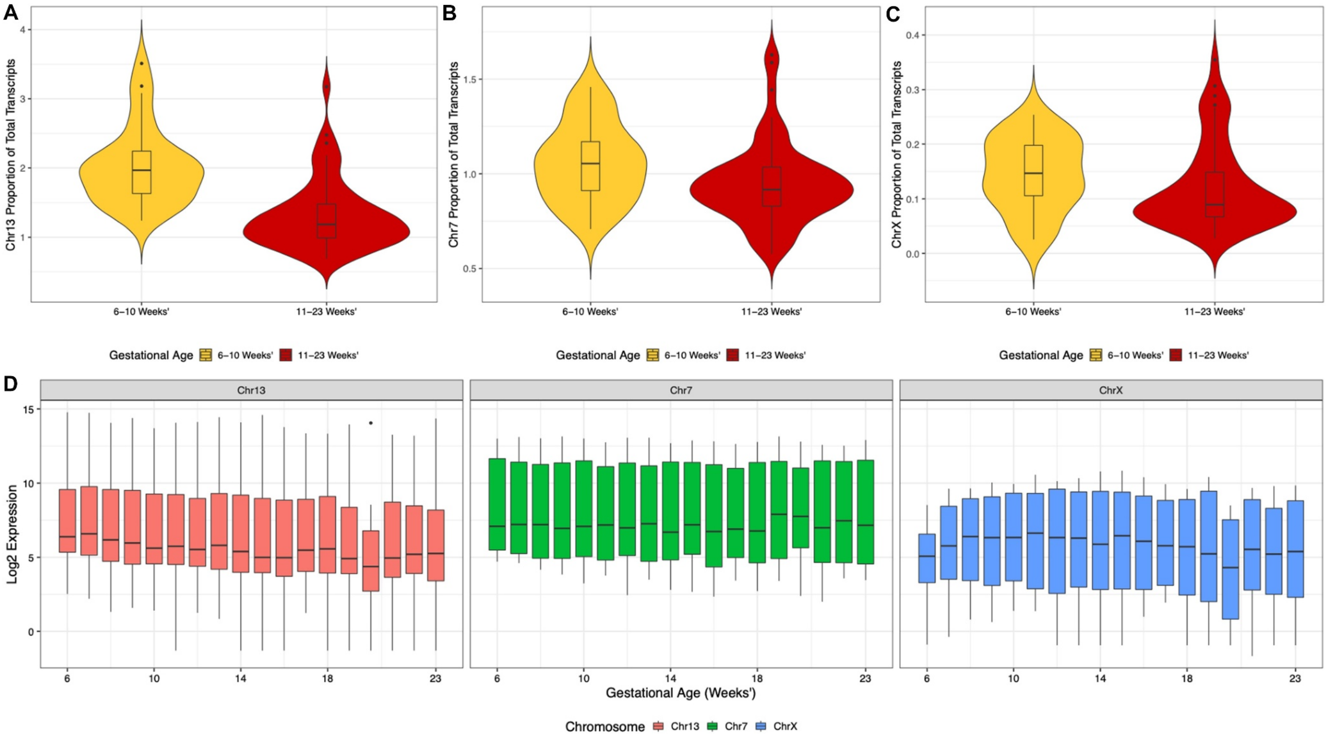


Figure 5. The proportion of miR-17 ~ 92 (Chr 13) and paralogs 106b~25 (Chr 7) and 106a~363 (Chr X) transcripts decreased in 11–23 weeks' compared to 6–10 weeks' gestation placenta. Violin plot showing decreased proportion of miR-17 ~ 92 member expression as a function of total miRNA expression after 10 weeks' gestation (a, b, c) with a general trend of decreased expression of individual cluster members across early gestation (d). Although chromosome 13 and chromosome X appear correlated, expression of miR-17 ~ 92 (chromosome 13) and its paralogs miR-106a~363 (chromosome 7) and miR-106b~25 (X chromosome) does not appear to follow a common expression pattern in placental tissue as we would expect if all 3 clusters are identically regulated by C-Myc.

Discussion

In this study, we have assessed the miRNA profile of 96 human placenta and matched plasma samples in early to mid gestation (6–23 weeks'), generating a placental tissue dataset using high-throughput sequencing that is substantially larger than previous reports that used microarray and qPCR profiling. We derive robust and comprehensive profiles that define miRNA expression at high-resolution, applying stringent bioinformatic protocols for the analysis of sequencing data, akin to large atlas studies.

Differentially expressed miRNAs reflect the dynamically changing utero-placental environment during early to mid gestation

In order to determine the potential association between placental miRNA expression and pregnancy health, particularly the expression of highly abundant miRNAs across early gestation, we examined more closely a period in early gestation involving a steep rise in oxygen tension in the intervillous space (IVS). Rising oxygen tension occurs between 10 and 12 weeks' gestation, at the onset of maternal blood flow into the IVS, resulting in an oxidative burst in placenta [3]. Previous studies have addressed miRNA response to hypoxia and the introduction of oxygen [42–46], including how ectopic oxygen tension may contribute to diseases of pregnancy [42,47–49]; and reviewed in [50]. The number and extensive gestation range of our samples enabled us to analyse the differential expression of miRNAs prior to, during and post the critical 10–11 weeks' gestational time point. This has potential to lead to identification of placental mechanisms for the remediation of oxidative stress at this crucial time in pregnancy.

That the placenta develops initially in a condition of low (but physiological) oxygen tension has been known for some time [3], and it has been suggested that this is important for the protection of the early foetus from the effects of oxidative stress such as tissue damage from oxygen free radicals [51]. miRNAs are uniquely poised to play a significant role in the mediation of gene expression throughout the early change of oxygen tension in the placenta by employing post transcriptional regulatory mechanisms to rapidly and precisely fine tune the placental environment in response to oxidative stress. We identified upregulation of let-7b/e-5p, miR-21, miR-23a/b-3p, miR-24-3p and miR-199a-5p, also associated with *HIF1a*, in the 6–10 weeks' gestation group. All of these have previously been shown to be upregulated by hypoxia. We also found lower abundance of let-7a and miR-101-3p in our 6–10 weeks' gestation group, both of which have previously been shown to be downregulated by hypoxia [52–56]; and reviewed in [57]. miRNAs identified in this study likely play an important role in the mitigation and regulation of hypoxia in the placenta. This may prove important to our understanding of diseases of pregnancy such as preeclampsia and early pregnancy loss.

The importance of miRNA function in the early placenta is not limited to the mitigation of oxidative stress but also in

Table 2. **miRNA by average expression in maternal plasma**. Average miRNA expression was calculated using the average normalized CPM expression across early to mid gestation (6–23 weeks') and log_2 transformed. **(a)** All top 10 miRNA with highest abundance in placenta were also detected at high abundance in plasma. Two miRNA (miR-517a-3p and miR-516b-5p) originate from the C19MC. **(b**) Top 10 miRNA by abundance in maternal plasma. RBC-associated miRNA identified herein have been previously identified in plasma samples including adult plasma and umbilical cord plasma but have not previously been profiled extensively in maternal plasma in early to mid gestation.

a			b		
miRNA	miRNA Family	Avg. Global Expression (log_2 CPM)	miRNA	miRNA Family	Avg. Global Expression (log_2 CPM)
miR-30d-5p	miR-30-5p	13.35	miR-16-5p$^+$	miR-15-5p/16-5p/195-5p/424-5p/497-5p	18.23
miR-125a-5p	miR-125-5p	10.59	miR-486-5p$^+$	miR-486-5p	17.46
miR-517a-3p^	miR-517-5p	4.71	miR-92a-3p$^+$	miR-25-3p/32-5p/92-3p/363-3p/367-3p	16.22
miR-199a-3p	miR-199-3p	10.62	let-7a-5p	let-7-5p/98-5p	15.70
miR-26b-5p	miR-26-5p	12.84	let-7b-5p	let-7-5p/98-5p	15.55
miR-26a-5p	miR-26-5p	12.70	miR-451a$^+$	miR-451	14.42
let-7a-5p	let-7-5p/98-5p	15.70	let-7 f-5p	let-7-5p/98-5p	14.15
miR-21-5p	miR-21-5p/590-5p	12.87	miR-25-3p	miR-25-3p/32-5p/92-3p/363-3p/367-3p	13.97
miR-126-3p	None listed	13.41	let-7i-5p	let-7-5p/98-5p	13.92
miR-516b-5p^	miR-516b-5p	4.27	miR-93-5p	miR-17-5p/20-5p/93-5p/106-5p/519-3p	13.58
^ C19MC				$^+$ RBC	

Table 3. **The abundance in plasma of miRNA differentially expressed in matched placenta**. Average miRNA expression was calculated using the average CPM expression across early to mid gestation (6–23 weeks') and log_2 transformed.

Placenta			Maternal Plasma		
Placenta DE miRNA	**log_2FC**	**AveExp (log_2)**	**AveExp (log_2)**	**Compartment of origin***	**Cellular enrichment^**
let-7a-5p	1.24	14.27	15.70	brain, epididymis, spinal cord	Mesodermal cells
miR-542-5p	1.11	7.07	0.66	epididymis, small intestine, lymph node	None listed
miR-222-3p	1.00	9.66	6.51	bladder, epididymis, prostate	Endothelial cell, endothelial cell of vascular tree, blood vessel endothelial
miR-1-3p	1.00	7.20	4.45	muscle, myocardium	Muscle, skeletal muscle, mesodermal cell
miR-199b-5p	0.97	10.24	2.72	colon, epididymis, oesophagus, skin, thyroid, vein	Fibroblast, mesodermal cell
miR-99a-5p	0.98	8.57	5.92	Arachnoid mater, brain, epididymis, prostate, spinal cord	Mast cell, leukocyte, myeloid leukocyte
miR-223-3p	0.99	10.93	13.28	vein	Leukocyte, haematopoietic cell, myeloid leukocyte
miR-221-3p	0.94	13.13	9.65	Epididymis, prostate	Endothelial cell, ciliated epithelial cell, muscle cell
miR-100-5p	0.98	12.83	5.58	Arachnoid mater, nerve, spinal cord, brain	Muscle cell, smooth muscle cell, vascular associated smooth muscle cell
miR-181a-5p	0.94	13.06	10.09	Arachnoid mater, brain, dura mater, spinal cord	Leukocyte, haematopoietic cell, B cell
miR-363-3p	0.94	9.18	9.26	Epididymis, vein	Leukocyte, haematopoietic cell, lymphocyte
miR-122-5p	−2.53	6.74	13.09	Liver, vein	Hepatocyte, endodermal cell
miR-9-5p	−1.42	10.51	2.19	Brain, spinal cord	Neural cell, neuroectodermal cell, ectodermal cell
miR-9-3p	−1.37	6.79	−2.25	Brain, dura mater, spinal cord	None listed
miR-372-3p	−1.08	8.17	−2.98	No profile	Pluripotent stem cell, embryonic stem cell, induced pluripotent stem cell, depleted in extraembryonic stem cell
miR-371a-5p	−1.01	7.90	−2.81	Constitutive expression	Pluripotent stem cell, embryonic stem cell, induced pluripotent stem cell, depleted in extraembryonic stem cell
miR-520 f-3p	−0.98	6.86	−3.07	spleen	Pluripotent stem cell, embryonic stem cell, induced pluripotent stem cell, depleted in extraembryonic stem cell
miR-934	−0.93	9.32	−2.44	Constitutive expression, highest in gallbladder and pleura	Epithelial cell, extraembryonic cell, endo-epithelial cell
miR-20a-5p	−0.85	10.02	9.73	Constitutive expression, highest in vein	Pluripotent stem cell, epithelial cell, endothelial cell
miR-17-5p	−0.77	6.80	8.77	Constitutive expression, highest in vein	Epithelial cell, pluripotent stem cell, endothelial cell
miR-373-3p	−0.68	6.74	−2.57	kidney	Pluripotent stem cell, embryonic stem cell, induced pluripotent stem cell

* [83]
^ [84]

cellular proliferation, trophoblast invasion and cellular differentiation, which are important to the placenta and also to the success of the pregnancy. let-7a-5p expression in first trimester placental explants has been implicated in a reduction in cytotrophoblast proliferation. In our differential expression analysis we identified let-7a-5p as up-regulated in the 11–23 weeks' gestation compared to the 6–10 weeks' gestation group. Our detailed week-by-week expression data suggests an incremental increase across early to mid gestation which together with previous term data is consistent with let-7a-5p's proposed role in regulation of placental growth by the reduction of cytotrophoblast proliferation [58]. The miR-34 family is known as a tumour suppressor miRNA family due to their synergistic effects with p53 [59]. Consistent with this, we

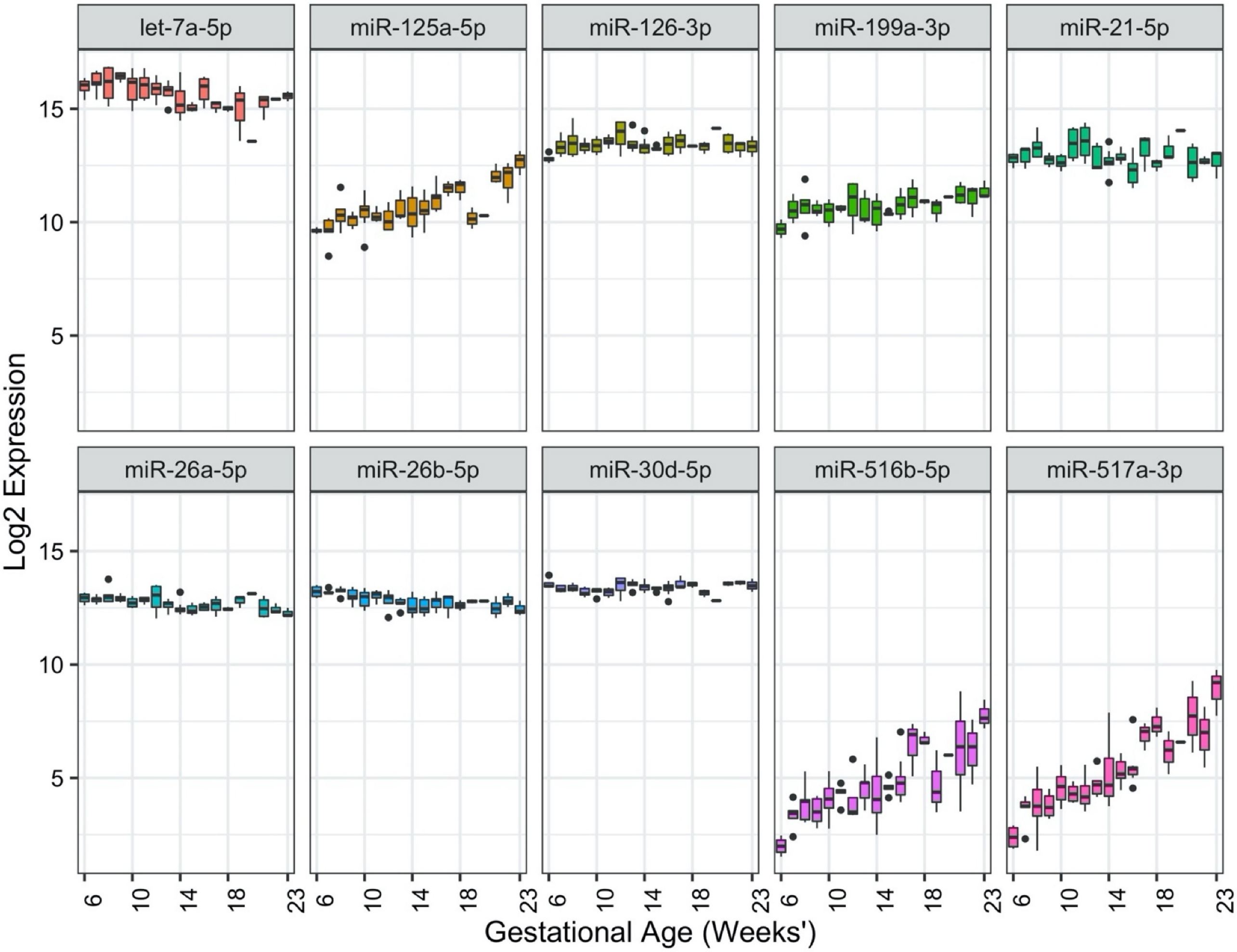


Figure 6. Abundance of maternal plasma miRNAs, identified as highly expressed in chorionic villous tissue, from 6–23 weeks' gestation. All ten miRNAs identified as the most highly expressed in chorionic villous tissue were identified as highly abundant in matched plasma samples. Abundance of the C19MC miRNA members, miR-516b-5p and miR-517a-3p, appear to increase across early to mid gestation.

identified up-regulation of all six miR-34 family members reflecting a reduction in cellular proliferation as the pregnancy progresses and placental growth trajectory begins to slow. Similarly, miR-378a-5p was previously found to enhance cell survival and promote trophoblast migration and invasion [60]. Our data showed significant down-regulation of the miR-34 family in 11–23 weeks' gestation compared to the 6–10 weeks' gestation placenta potentially linking down-regulation of these miRNAs with the reduced trophoblast migration as pregnancy progresses. miR-200 family members are also known to have an established role in placental function, in particular in the cellular transition from epithelial to mesenchymal (EMT) phenotype which is a prerequisite for extravillous trophoblast invasion of the maternal decidua. miR-200 family members play a well-known role in EMT through a feedback loop with the ZEB family of transcription factors. We identified seven members of the miR-200 family, six of which are down-regulated in the 11–23 weeks' gestation group. The miR-200 family is known to be highly expressed in epithelial cell types [61] and thus its down-regulation is consistent with development of the vascularized mesenchymal core in chorionic villi in the second trimester.

Proportional change in expression miRNA clusters play a role in placenta gene regulation during development

miRNA clusters are known to play an important role in the placenta, with these regions thought to have co-evolved and to act cooperatively to repress target genes [62]. Interestingly, while miRNA clusters are typically transcribed as a single polycistronic transcript, we identified large intra-sample variances in the expression levels of cluster members (Additional file 2, Figure S4) hinting at additional post transcriptional regulatory mechanisms. There are a number of miRNA clusters expressed either exclusively, or preferentially in placenta [31,63,64]. In particular, the C19MC, C14MC and miR-17 ~ 92 clusters have been shown to regulate phenotypic and functional diversity in frequently used cell lines and isolated primary trophoblasts, and in the immune system, by which they may play a role in immune tolerance to paternal antigens [32,65,66].

C19MC is expressed almost exclusively in the placenta but also in certain tumours and undifferentiated cells [34,67]. C19MC is a large placenta specific imprinted cluster located on chromosome 19q13.41 and mono-allelically expressed from the paternally inherited chromosome [68]. This cluster has been difficult to study as its expression is both placenta and primate specific [68] and hence lacks an ortholog in mouse [35]. Importantly, C19MC members are expressed at lower than trophoblast levels in JEG-3 and JAR cells, two cell lines derived from choriocarcinoma [69], and are not expressed in the commonly used HTR8/SVneo cell line [64,67] highlighting the need for primary placental tissue in research. Consistent with our findings, previous research has shown that C19MC constitutes a large proportion of total miRNA transcripts (15% at term), with our data additionally showing a gestational decline from 27% at 6–10 weeks' gestation to 23% at 11–23 weeks' gestation [33,67]. In addition, there was a significant (FDR < 0.05) down-regulation in 34 C19MC cluster members in 11–23 weeks' placenta compared to the 6–10 weeks' gestation. Interestingly, a previous report showed that 46 of the 47 miRNAs from the C19MC cluster were up-regulated in third trimester villous trophoblasts compared to first trimester [32]. It is possible that this miRNA cluster exhibits a dynamic gestational expression pattern but perhaps more likely is the possibility that differences between studies can be explained by differences in expression between assessment of trophoblast versus whole chorionic villous tissue. The latter is more physiologically relevant and fits our question. However, this also introduces the question of whether the changes we see are due at least in part to the change in cell type proportions seen in the placenta throughout gestation [70].

A recent miR-517a/c *in situ* hybridization analysis localized expression from these C19MC miRs to cytotrophoblast and syncytiotrophoblast in chorionic villi, proximal cytotrophoblasts in anchoring villi, decreased expression in distal cytotrophoblasts with further reduction in extra-villous cytotrophoblasts. This finding highlights variable expression of the C19MC cluster in different placental cell types, and shows loss of expression of this cluster as trophoblasts differentiate to extravillous phenotypes post EMT [71]. Another report concurs showing higher expression of C19MC members in villous trophoblasts compared to extravillous trophoblasts suggestive of a role for the C19MC cluster miRNA in attenuation of EVT migration through the direct targeting of mRNA transcripts related to cellular movement [72]. Mouillet *et al.* (2015) suggested a role for C19MC in cellular differentiation or the maintenance of pluripotency [73]. Further investigation into the proportional change in expression from pregnancies complicated by placental pathology may help to identify miRNA species with biomarker potential in plasma. However, low expression levels in some maternal plasma samples have frustrated our efforts in this regard.

C14MC, located at the imprinted DLK1-DIO3 domain on human chromosome 14q32 is the largest known human miRNA cluster comprising 52 miRNA genes expressed from the maternally inherited chromosome [64]. The C14MC cluster is found exclusively in eutherian mammals and has been suggested to be essential to the evolution of this lineage, with expression in humans demonstrating a strong bias towards brain, placenta and some embryonic tissues [63]. Our analysis identified expression from 84 mature miRNAs in the C14MC cluster with the proportion of member transcripts increased in the 11–23 weeks' gestation group compared to 6–10 weeks' gestation. Previous studies investigating C14MC expression, in both whole villous tissue and in primary cytotrophoblasts, have reported a decrease in expression from first to third trimester. Using whole villous tissue, Gu *et al.* (2013) [74] identified 11 miRNA from this cluster down-regulated in third compared to first trimester, and proposed a link between both C19MC and C14MC in immune suppression and innate/adaptive immune response in the mother. In primary cytotrophoblast, Morales-Prieto *et al.* (2012) [32] identified 34 miRNA from this cluster which were also down-regulated in third compared to first trimester [64,74]. Whilst the discrepancy between our data and Morales-Prieto *et al.* (2013) can be explained by the use of different biological source (single cell type versus whole tissue), that between our data and Gu *et al.* (2013) is not so clear but may be due to differences in technology, the number of cluster members identified, or that our samples range up to 23 weeks' gestation and do not include term. It is plausible that rather than a steady decline from early gestation to term, the expression of C14MC members increases across early gestation, before decreasing at term. Consistent with a recent report by Wommack *et al.* (2018) [75] who profiled circulating miRNAs in maternal plasma and found an inverse relationship between placenta-specific clusters C19MC and C14MC we found this in our tissue samples.

Finally, in the analysis of placenta-associated clusters we investigated the expression of miR-17 ~ 92 and paralogs miR-106a~363 and miR-106b~25, identifying 19 mature miRNAs. The proportion of total expression from miR-17 ~ 92 members and paralogs are less than 1% per cluster, with all three clusters showing an overall decrease in the proportion of total expression across early to mid gestation. Using the miRNA sequences provided by Kumar *et al.* (2013) [65], we determined the 5′ or 3′ origin of the transcripts by cross-checking the given sequences against miRBase v21 [23,65], and were thus able to directly compare their microarray data with our sequencing data (detailed in Additional File 3: Table S11).

Plasma miRNA sequencing offers a potential source of non-invasive biomarkers for detecting placenta health

In addition to establishing the placental miRNA profile across early to mid gestation and, given the potential for placental health biomarker discovery using our access to matched maternal plasma, we also aimed to analyse miRNA profiles in maternal plasma sampled in women at the same time as placenta. The identification of placenta-derived, or pregnancy-associated miRNA in plasma offers a potential window into minimally invasive monitoring of pregnancy progression.

Working with miRNA in plasma is known to be difficult due to potential inclusion of red blood cells or other cellular debris [76–78]. Red blood cell contamination may mask any signal we hoped to detect from the placenta as haemolysis may perturb miRNA species and count distributions due to overloading of red blood cell miRNAs [76,78,79]. In our study we see evidence that 11 of our samples were likely affected by haemolysis. This common difficulty with plasma data, led us to assess the detection of miRNAs of interest revealing that the 10 most abundant placental miRNAs were reflected by high abundance levels in matched plasma. Furthermore, these plasma miRNAs in our study show strong concordance with previous miRNA plasma studies [77,80].

Differential expression and clear gestational age gradient clustering patterns in placenta miRNAs between 6–10 and 11–23 weeks' gestation were not particularly reflected by maternal plasma, indicating that plasma does not reflect the changes in the miRNA profile in placenta across early to mid gestation. miRNAs that were abundant in placenta were not always found to be abundant in plasma. In particular, miRNAs enriched in Pluripotent stem cell/embryonic stem cell were of low abundance in plasma compared to placenta. With the exception of placenta-specific miRNAs, cellular enrichment could be linked to various other systemic compartments. For example, five miRNAs have cellular enrichment in immune response cells, whilst another three miRNAs have enrichment in pluripotent and embryonic stem cells, and one miRNA is enriched in extra-embryonic stem cells which may originate from the foetal membranes or other extra-embryonic tissue such as chorionic villi. Of particular interest, miR-99a-5p enrichment includes enrichment in mast cells which are involved in inflammatory response and allergic reactions. We see an apparent increase in abundance of miR-99a-5p across early to mid gestation, and some additional increase in male- compared with female-bearing pregnancies, which may be indicative of the increased maternal immune response to the male conceptus [81].

Complementary to the investigation of miRNA clusters in placenta, we also investigated the expression of three highly placenta-associated miRNA clusters and confirmed the abundance of these cluster miRNAs in plasma, with the proportion of cluster member expression revealing biomarker potential in plasma. One pathway for the transfer of placental molecules into the maternal circulation is via exosomes. It has previously been reported that members of the C19MC are the predominant miRNA species expressed in exosomes released from primary human trophoblasts [67]. We were able to confirm the presence of 39 C19MC, 48 C14MC and 20 miR-17 ~ 92 miRNA in matched plasma, with their presence showing promise for the use of miRNA as biomarkers for placental health. Perhaps more importantly, week-by-week examination of miRNAs across early gestation, as described in this study, are key elements for developing pregnancy surveillance measures in the future, especially when combined with disease-specific profiles associated with preeclampsia, gestational diabetes and preterm birth. Our investigation here of miRNAs as biomarkers of placental, and by proxy pregnancy health have identified many areas for further investigation. With the implementation of machine learning algorithms, cell free DNA work and placenta specific analysis standards, much more can be expected from this field.

Conclusion

Appropriate placental growth and development is essential for pregnancy success and miRNAs play roles in mediating these. We investigated placental and maternal plasma miRNAs using matched samples taken after elective termination of otherwise normal pregnancies and found 374 significantly (FDR <0.05) differentially expressed miRNA between 6–10 and 11–23 weeks' gestation placenta.

The research conducted herein is the most comprehensive attempt to accurately profile the miRNA landscape of the human placenta, in part, because of the use of miRNA sequencing rather than microarray or qPCR. This dataset represents unprecedented access to a large number of human placenta samples from 6–23 weeks of gestation. Together these data are an important reference set for miRNA expression of early to mid gestation placental development and function, at time points rarely seen in previous placental sequencing work. These provide an important resource for the placental biology field. Further investigation using maternal plasma in early gestation samples from pregnancies with known outcomes will provide further insights into the search for miRNAs as biomarkers of pregnancy health and disease.

Acknowledgements

We wish to acknowledge the generosity of the women who donated their placenta for our research. Without them, this research would not be possible. We also acknowledge valuable input from QIAGEN Genomic Services.

Funding

This research is supported by NIH NICHD R01 HD089685-01 Maternal molecular profiles reflect placental function and development across gestation PI Roberts, an Australian Government Research Training Program (RTP) Scholarship awarded to MDS, a National Health and Medical Research Council Investigator Grant (GNT1174971) awarded to CTR and a Matthew Flinders Professorial Fellowship awarded to CTR and funded by Flinders University. JB is supported by the James & Diana Ramsay Foundation.

Authors contributions

CTR created the concept. MS, CTR, JB and TBM conceived and developed experimental plans. TJK, DMcA and DMcC performed experiments. MS and JB analysed the sequencing data with help from KB and QW. Manuscript written by MS, JB and CTR with help from TBM and KP. All authors read and approved the final version of the manuscript.

Author information

[1]Robinson Research Institute, [2]Adelaide Medical School, [3]Flinders Health and Medical Research Institute, Flinders University, [4]The University of South Australia, [5]South Australian Health & Medical Research Institute (SAHMRI), [6]School of Agriculture Food and Wine; University of Adelaide.

Availability of data and materials

The dataset(s) supporting the conclusion of this article are available in NCBI's Gene Expression Omnibus [82] and are accessible through GEO Series accession number GSE151362 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE151362).

Declarations

Ethics approval for the collection of placenta tissue and blood from women undergoing elective pregnancy termination between 6–23 weeks' gestation was provided under HREC/16/TQEH/33, by The Queen Elizabeth Hospital Human Research Ethics Committee (TQEH/LMH/MH).

Disclosure statement

All authors declare no competing interests.

ORCID

Melanie D. Smith http://orcid.org/0000-0001-7016-8245
Katherine Pillman http://orcid.org/0000-0002-5869-889X
Tanja Jankovic-Karasoulos http://orcid.org/0000-0003-4293-359X
Qianhui Wan http://orcid.org/0000-0003-4207-4940
K. Justinian Bogias http://orcid.org/0000-0002-5776-2052
Tina Bianco-Miotto http://orcid.org/0000-0002-8431-5338
James Breen http://orcid.org/0000-0001-6184-0925
Claire T. Roberts http://orcid.org/0000-0002-9250-2192

References

[1] Gude NM, Roberts CT, Kalionis B, et al. Growth and function of the normal human placenta. Thromb Res. 2004;114(5–6):397–407.
[2] Sibai B, Dekker G, Kupferminc M. Pre-eclampsia. Lancet. 2005;365(9461):785–799.
[3] Jauniaux E, Watson AL, Hempstock J, et al. Onset of maternal arterial blood flow and placental oxidative stress: a possible factor in human early pregnancy failure. Am J Pathol. 2000;157(6):2111–2122.
[4] Henriksen T, Clausen T. The fetal origins hypothesis: placental insufficiency and inheritance versus maternal malnutrition in well-nourished populations. Acta Obstet Gynecol Scand. 2002;81 (2):112–114.
[5] Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. [Internet] Cell. 2004;116(2):281–297. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0092867404000455
[6] Bartel DP. Metazoan MicroRNAs. Cell. 2018;173:20–51.
[7] Winter J, Jung S, Keller S, et al. Many roads to maturity: microRNA biogenesis pathways and their regulation. [Internet] Nat Cell Biol. 2009;11(3):228–234. Available from: http://www.nature.com/articles/ncb0309-228
[8] Ameres SL, Zamore PD. Diversifying microRNA sequence and function. [Internet] Nat Rev Mol Cell Biol. 2013;14(8):475–488. Available from: http://www.nature.com/articles/nrm3611
[9] Bartel DP. MicroRNAs: target recognition and regulatory functions. Cell. 2009;136(2):215–233.
[10] Saini HK, Griffiths-Jones S, Enright AJ. Genomic analysis of human microRNA transcripts. Proc Natl Acad Sci USA. 2007 Nov 6;104(45):17719–24.doi:10.1073/pnas.0703890104
[11] Kim VN, Han J, Siomi MC. Biogenesis of small RNAs in animals. [Internet] Nat Rev Mol Cell Biol. 2009;10(2):126–139.
[12] Dykes IM, Emanueli C. Transcriptional and post-transcriptional gene regulation by long non-coding RNA. [Internet] Genom Proteom Bioinform. 2017;15(3):177–186. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5487525
[13] Morley R. Fetal origins of adult disease. [Internet] Semin Fetal Neonatal Med. 2006;11(2):73–78.
[14] Burton GJ, Fowden AL, Thornburg KL. Placental origins of chronic disease. Physiol Rev. 2016;96(4):1509–1565.
[15] Pantano L, Estivill X, Martí E. SeqBuster, a bioinformatic tool for the processing and analysis of small RNAs datasets, reveals ubiquitous miRNA modifications in human embryonic cells. Nucleic Acids Res. [Internet] 2009; 38. Available from: https://watermark.silverchair.com/gkp1127.pdf?token=AQECAHi208BE49Ooan9kkhW_Ercy7Dm3ZL_9Cf3qfKAc485ysgAAAbQwggGwBgkqhkiG9w0BBwagggGhMIIBnQIBADCCAZYGCSqGSIb3DQEHATAeBglghkgBZQMEAS4wEQQM1TUUngRW8zyuVolBAgEQgIIBZ7gj8AvZ9tPmZnDNEGCVulK4NZusRxNE3WbfR5NrHrMCKeH
[16] Didion JP, Martin M, Collins FS. Atropos: specific, sensitive, and speedy trimming of sequencing reads. PeerJ. 2017;5:e3720.
[17] Dobin A, Davis CA, Schlesinger F, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29:15–21.
[18] Church DM, Schneider VA, Graves T, et al. Modernizing reference genome assemblies. PLoS Biol. 2011;9(7):e1001091.
[19] Andrews S FastQC: a quality control tool for high throughput sequence data. [Internet]. 2010; Available from: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/
[20] Ewels P, Magnusson M, Lundin S, et al. MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics. 2016;32(19):3047–3048.
[21] Pantano L, Estivill X, Martí E. A non-biased framework for the annotation and classification of the non-mirna small RNA transcriptome. Bioinformatics. 2011;27(22):3202–3203.
[22] Griffiths-Jones S, Grocock RJ, van Dongen S, et al. miRBase: microRNA sequences, targets and gene nomenclature. Nucleic Acids Res. 2006;34(90001):D140–4.
[23] Kozomara A, Griffiths-Jones S. MiRBase: annotating high confidence microRNAs using deep sequencing data. Nucleic Acids Res. Internet 2014;42(D1):D68–D73.
[24] Ritchie ME, Phipson B, Wu D, et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43(7):e47.
[25] Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2009;26(1):139–140.
[26] Law CW, Chen Y, Shi W, et al. Voom: precision weights unlock linear model analysis tools for RNA-seq read counts. Genome Biol. 2014;15(2):R29.
[27] Smyth GK. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Stat Appl Genet Mol Biol. 2004;3(1):1–25.
[28] Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Series B Stat Methodol. 1995;57:289–300.
[29] Wan Q, Yiner-Lee Leemaqz S, Pederson SM, et al. Quality control measures for placental sample purity in DNA methylation array analyses. Placenta. 2019;88:8–11.
[30] Wang W, Feng L, Zhang H, et al. Preeclampsia up-regulates angiogenesis-associated MicroRNA (i.e. miR-17, −20a, and −20b) That target Ephrin-B2 and EPHB4 in human placenta. J Clin Endocrinol Metab. 2012;97(6):E1051–E1059.
[31] Luo -S-S, Ishibashi O, Ishikawa G, et al. Human villous trophoblasts express and secrete placenta-specific MicroRNAs into maternal circulation via exosomes1. Biol Reprod. 2009;81(4):717–729.
[32] Morales-Prieto DM, Chaiwangyen W, Ospina-Prieto S, et al. MicroRNA expression profiles of trophoblastic cells. Placenta. 2012;33(9):725–734.
[33] Carreras-Badosa G, Bonmat A, Ortega FJ, et al. Dysregulation of placental miRNA in maternal obesity is associated with pre-and postnatal growth. J Clin Endocrinol Metab. 2017;102(7):2584–2594.
[34] Bentwich I, Avniel A, Karov Y, et al. Identification of hundreds of conserved and nonconserved human microRNAs. Nat Genet. 2005;37(7):766–770.
[35] Morales-Prieto DM, Ospina-Prieto S, Schmidt A, et al. Elsevier trophoblast research award lecture: origin, evolution and future of placenta miRNAs. Placenta. [Internet] 2014; 35:S39–S45.
[36] Olive V, Jiang I, He L. Mir-17-92, a cluster of miRNAs in the midst of the cancer network. Int J Biochem Cell Biol. 2010;42(8):1348–1354.
[37] Bortolin-Cavaille ML, Dance M, Weber M, et al. C19MC microRNAs are processed from introns of large Pol-II, non-protein-coding transcripts. Nucleic Acids Res. [Internet] 2009;37(10):3464–3473.
[38] Seitz H, Royo H, Bortolin ML, et al. A large imprinted microRNA gene cluster at the mouse Dlk1-Gtl2 domain. Genome Res. 2004;14(9):1741–1748.
[39] Schmidt JK, Block LN, Golos TG. Defining the rhesus macaque placental miRNAome: conservation of expression of placental miRNA clusters between the macaque and human. Placenta. 2018;65:55–64.

[40] Chim SSC, Shing TKF, Hung ECW, et al. Detection and characterization of placental microRNAs in maternal plasma. Clin Chem. 2008;54(3):482–490.

[41] Zhao C, Sun X, Li L. Biogenesis and function of extracellular miRNAs. [Internet] ExRNA. 2019;1(1). DOI:10.1186/s41544-019-0039-4.

[42] Mouillet JF, Chu T, Hubel CA, et al. The levels of hypoxia-regulated microRNAs in plasma of pregnant women with fetal growth restriction. Placenta. 2010;31(9):781–784.

[43] Mouillet JF, Chu T, Nelson DM, et al. MiR-205 silences MED1 in hypoxic primary human trophoblasts. FASEB J. 2010;24 (6):2030–2039.

[44] Shen J, Xia W, Khotskaya YB, et al. EGFR modulates microRNA maturation in response to hypoxia through phosphorylation of AGO2. Nature. 2013;497(7449):383–387.

[45] Camps C, Saini HK, Mole DR, et al. Integrated analysis of microRNA and mRNA expression and association with HIF binding reveals the complexity of microRNA expression regulation under hypoxia. Mol Cancer. 2014;13(1):28.

[46] Mori A, Nishi H, Sasaki T, et al. HLA-G expression is regulated by miR-365 in trophoblasts under hypoxic conditions. Placenta. 2016;45:37–41.

[47] Chelbi ST, Vaiman D. Genetic and epigenetic factors contribute to the onset of preeclampsia. Mol Cell Endocrinol. 2008;282(1–2):120–129.

[48] Cronqvist T, Saljé K, Familari M, et al. Syncytiotrophoblast vesicles show altered micro-RNA and haemoglobin content after Ex-vivo perfusion of placentas with haemoglobin to mimic preeclampsia. PLoS One. 2014;9(2):e90020.

[49] Li Q, Pan Z, Wang X, et al. miR-125b-1-3p inhibits trophoblast cell invasion by targeting sphingosine-1-phosphate receptor 1 in preeclampsia. Biochem Biophys Res Commun. 2014;453 (1):57–63.

[50] Poston L, Raijmakers MTM. Trophoblast oxidative stress, antioxidants and pregnancy outcome—A review. Placenta. 2004;25:S72–8.

[51] Burton GJ, Hempstock J, Jauniaux E. Oxygen, early embryonic metabolism and free radical-mediated embryopathies. Reprod Biomed Online. 2003;6(1):84–96.

[52] Hua Z, Lv Q, Ye W, et al. MiRNA-directed regulation of VEGF and other angiogenic factors under hypoxia. PLoS One. 2006;1(1):e116.

[53] Hebert C, Norris K, Scheper MA, et al. High mobility group A2 is a target for miRNA-98 in head and neck squamous cell carcinoma. Mol Cancer. 2007;6(1):5.

[54] Kulshreshtha R, Ferracin M, Wojcik SE, et al. A microRNA signature of hypoxia. Mol Cell Biol. 2007;27(5):1859–1867.

[55] Rane S, He M, Sayed D, et al. Downregulation of MiR-199a derepresses hypoxia-inducible factor-1α and sirtuin 1 and recapitulates hypoxia preconditioning in cardiac myocytes. Circ Res. 2009;104(7):879–886.

[56] Gonsalves CS, Kalra VK. Hypoxia-mediated expression of 5-lipoxygenase-activating protein involves HIF-1alpha and NF-kappaB and microRNAs 135a and 199a-5p. J Immunol. 2010;184(7):3878–3888.

[57] Nallamshetty S, Chan SY, Loscalzo J. Hypoxia: a master regulator of microRNA biogenesis and activity. Free Radic Biol Med. 2013;64:20–30.

[58] Forbes K, Farrokhnia F, Aplin JD, et al. Dicer-dependent miRNAs provide an endogenous restraint on cytotrophoblast proliferation. Placenta. [Internet] 2012;33(7):581–585.

[59] Zhang L, Liao Y, Tang L. MicroRNA-34 family: a potential tumor suppressor and therapeutic candidate in cancer. J Exp Clin Cancer Res. 2019;38(1):53.

[60] Luo L, Ye G, Nadeem L, et al. MicroRNA-378a-5p promotes trophoblast cell survival, migration and invasion by targeting Nodal. J Cell Sci. 2012;125:3124–3132.

[61] Humphries B, Yang C. The microRNA-200 family: small molecules with novel roles in cancer development, progression and therapy. Oncotarget. 2015;6(9):6472–6498.

[62] Wang Y, Luo J, Zhang H, et al. MicroRNAs in the same clusters evolve to coordinately regulate functionally related genes. Mol Biol Evol. 2016;33(9):2232–2247.

[63] Glazov EA, McWilliam S, Barris WC, et al. Origin, evolution, and biological role of miRNA cluster in DLK-DIO3 genomic region in placental mammals. Mol Biol Evol. 2008;25(5):939–948.

[64] Morales-Prieto DM, Ospina-Prieto S, Chaiwangyen W, et al. Pregnancy-associated miRNA-clusters. J Reprod Immunol. 2013;97(1):51–61.

[65] Kumar P, Luo Y, Tudela C, et al. The c-Myc-Regulated MicroRNA-17 92 (miR-17 92) and miR-106a 363 clusters Target hCYP19A1 and hGCM1 to inhibit human trophoblast differentiation. Mol Cell Biol. 2013;33(9):1782–1796.

[66] Kuo G, Wu CY, Yang HY. MiR-17-92 cluster and immunity. [Internet] J Formosan Med Assoc. 2019; DOI:10.1016/j.jfma.2018.04.013.

[67] Donker RB, Mouillet JF, Chu T, et al. The expression profile of C19MC microRNAs in primary human trophoblast cells and exosomes. Mol Hum Reprod. 2012;18(8):417–424.

[68] Noguer-Dance M, Abu-Amero S, Al-Khtib M, et al. The primate-specific microRNA gene cluster (C19MC) is imprinted in the placenta. Hum Mol Genet. 2010;19(18):3566–3582.

[69] Lee CQE, Gardner L, Turco M, et al. What is trophoblast? A combination of criteria define human first-trimester trophoblast. Stem Cell Reports. 2016;6(2):257–272.

[70] Burton GJ, Fowden AL. The placenta: a multifaceted, transient organ. Philos Trans R Soc Lond B Biol Sci. 2015;370 (1663):20140066.

[71] Mong EF, Yang Y, Akat KM, et al. Chromosome 19 microRNA cluster enhances cell reprogramming by inhibiting epithelial-to-mesenchymal transition. Sci Rep. 2020;10(1):3029.

[72] Xie L, Mouillet JF, Chu T, et al. C19MC microRNAs regulate the migration of human trophoblasts. Endocrinology. 2014;155 (12):4975–4985.

[73] Mouillet J-FJ-F, Ouyang Y, Coyne CB, et al. MicroRNAs in placental health and disease. Am J Obstet Gynecol. 2015;213(4):S163–72.

[74] Gu Y, Sun J, Groome LJ, et al. Differential miRNA expression profiles between the first and third trimester human placentas. AJP. 2013;304:E836–43.

[75] Wommack JC, Trzeciakowski JP, Miranda RC, et al. Micro RNA clusters in maternal plasma are associated with preterm birth and infant outcomes. PLoS One. 2018;13(6):e0199029.

[76] Pritchard CC, Cheng HH, Tewari M. MicroRNA profiling: approaches and considerations. [Internet] Nat Rev Genet. 2012;13(5):358–369. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4517822

[77] Blondal T, Jensby Nielsen S, Baker A, et al. Assessing sample and miRNA profile quality in serum and plasma or other biofluids. [Internet] Methods. 2013;59(1):164–169. Available from: https://www.sciencedirect.com/science/article/pii/S1046202312002551

[78] Cheng HH, Yi HS, Kim Y, et al. Plasma processing conditions substantially influence circulating microRNA biomarker levels. PLoS One. 2013;8(6):e64795.

[79] Kirschner MB, Kao SC, Edelman JJ, et al. Haemolysis during sample preparation alters microRNA content of plasma. PLoS One. 2011;6(9):e24145.

[80] Brennan GP, Vitsios DM, Casey S, et al. RNA-sequencing analysis of umbilical cord plasma microRNAs from healthy newborns. PLoS One. 2018;13(12):e0207952.

[81] Clifton VL. Review: sex and the human placenta: mediating differential strategies of fetal growth and survival. Placenta. 2010;31(Suppl):S33–9.

[82] Edgar R, Domrachev M, Lash AE. Gene expression omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res. 2002;30(1):207–210.

[83] Ludwig N, Leidinger P, Becker K, et al. Distribution of miRNA expression across human tissues. Nucleic Acids Res. 2016;44 (8):3865–3877.

[84] Lizio M, Harshbarger J, Abugessaisa I, et al. Update of the FANTOM web resource: high resolution transcriptome of diverse cell types in mammals. Nucleic Acids Res. 2017;45(D1):D737–43.

Chapter 4

Haemolysis detection in microRNA-seq from clinical plasma samples

Statement of Authorship

Title of Paper	Haemolysis detection in microRNA-seq from clinical plasma samples
Publication Status	☐ Published ☐ Accepted for Publication ☐ Submitted for Publication ☒ Unpublished and Unsubmitted w ork w ritten in manuscript style
Publication Details	Haemolysis detection in microRNA-seq from clinical plasma samples Melanie D. Smith, Shalem Y. Leemaqz, Tanja Jankovic-Karasoulos, Dale McAninch, Dylan McCullough, James Breen, Claire T. Roberts, Katherine Pillman

Principal Author

Name of Principal Author (Candidate)	Melanie Smith		
Contribution to the Paper	- All bioinformatics analyses - Drafting of manuscript - Producing figures / visualisation of data		
Overall percentage (%)	90%		
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.		
Signature		Date	13/09/2021

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);

ii. permission is granted for the candidate in include the publication in the thesis; and

iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Shalem Y. Leemaqz		
Contribution to the Paper	- Statistical support - Helped design experiments		
Signature		Date	13/9/2021

Name of Co-Author	Tanja Jankovic-Karasoulos		
Contribution to the Paper	- Performed experiments - Editing of manuscript drafts		
Signature		Date	13/9/2021

Please cut and paste additional co-author panels here as required.

Name of Co-Author	Dale McAninch		
Contribution to the Paper	- Performed experiments		
Signature		Date	13/09/2021

Name of Co-Author	Dylan McCullough		
Contribution to the Paper	- Performed experiments		
Signature		Date	13/09/2021

Name of Co-Author	James Breen		
Contribution to the Paper	- Conceived and developed experimental plans - Editing of manuscript drafts		
Signature		Date	13/09/2021

Name of Co-Author	Claire T Roberts		
Contribution to the Paper	- Conceived and developed experimental plans - Editing of manuscript drafts		
Signature		Date	13/09/2021

Name of Co-Author	Katherine Pillman		
Contribution to the Paper	- Expert miRNA analyses support - Development of experimental plans - Editing of manuscript drafts		
Signature		Date	13/9/2021

Haemolysis detection in microRNA-seq from clinical plasma samples

Melanie D. Smith[1,2]*, Shalem Y. Leemaqz[4], Tanja Jankovic-Karasoulos[4], Dale McAninch[1,2], Dylan McCullough[4], James Breen[1,2,3], Claire T. Roberts[4], Katherine Pillman[5]*

[1] Robinson Research Institute, University of Adelaide, Adelaide SA 5005, Australia
[2] Adelaide Medical School, University of Adelaide, Adelaide SA 5005, Australia
[3] South Australian Genomics Centre, South Australian Health & Medical Research Institute, Adelaide SA 5000, Australia
[4] Flinders Health and Medical Research Institute, Flinders University, Bedford Park SA
[5] Centre for Cancer Biology, University of South Australia/SA Pathology, Adelaide SA 5000, Australia

* Correspondence: melanie.smith@adelaide.edu.au, katherine.pillman@unisa.edu.au

Note: Supplementary Figures and Tables are hosted on Figshare

(https://figshare.com/projects/Chapter_4_-_Supplementary/122501)

Abstract

Background

The abundance of cell-free microRNA (miRNA) has been measured in many body fluids, including blood plasma, which has been proposed as a source with novel, minimally invasive biomarker potential for several diseases. While methods for the quantification of plasma miRNAs continue to improve, there is no consensus on an optimal reference miRNA or to what extent haemolysis, not visible to the naked eye, may affect plasma miRNA content. Here we propose a new method for the detection of haemolysis in miRNA high-throughput sequencing data from libraries prepared using human plasma.

Methods

To establish a miRNA haemolysis signature in plasma we first identified differentially expressed miRNAs between samples with known haemolysis status and selected miRNA with statistically significant higher abundance in our haemolysed group. We further refined this miRNA signature such that only high confidence miRNA remained. Given there may be both technical and biological reasons for differential abundance of signature miRNA, and to ensure the method developed here was relevant outside of our specific context, that is women of reproductive age, we tested for significant differences between pregnant and non-pregnant groups.

Results

Here we report a novel 20 miRNA signature (miR-106b-3p, miR-140-3p, miR-142-5p, miR-532-5p, miR-17-5p, miR-19b-3p, miR-30c-5p, miR-324-5p, miR-192-5p, miR-660-5p, miR-186-5p, miR-425-5p, miR-25-3p, miR-363-3p, miR-183-5p, miR-451a, miR-182-5p, miR-191-5p, miR-194-5p, miR-20b-5p) that can be used to identify the contribution of haemolysis, *in silico,* to miRNAs quantified in high throughput miRNA sequencing data.

Conclusion

Given the potential for haemolysis contamination, we recommend that assay for haemolysis detection become standard pre-analytical practice and provide here a simple method for haemolysis detection.

Introduction

MicroRNAs (miRNAs) represent a class of short, ~22 nt single stranded non-coding RNA transcripts found in the cytoplasm of most cells that act as post transcriptional regulators of gene expression [1,2]. In their canonical action, miRNAs mediate the expression of specific messenger RNA (mRNA) targets by binding to the 3`-untranslated region (UTR) of transcripts and either repressing translation or marking them for degradation [3]. Importantly, miRNAs demonstrate tissue specific, temporal and spatial expression specificity and are known regulators of development, with most mammalian mRNAs harbouring conserved targets of one or many miRNAs [1,2,4]. In the canonical miRNA pathway, target specificity requires exact nucleotide sequence complementarity between the miRNA ‘seed’ region (the first 2-7 bases at the 5` end of the mature miRNA transcript) and the 3`-UTR of the mRNA.

Whilst miRNA expression is both temporally and spatially tissue-specific, they are also identified beyond the cells in which they were synthesised, in various body fluids including urine, saliva and blood plasma [5]. Circulating cell-free miRNAs identified in plasma are packaged in micro vesicles such as exosomes [6,7] or bound to protein complexes such as argonaute 2 (Ago2), nucleophosmin 1 (NPM 1) and high density lipoprotein (HDL) [8–10], making them exceptionally stable [5]. This stability, coupled with their minimally invasive accessibility, has suggested circulating cell-free miRNAs as an important resource for the identification of novel biomarkers.

Whilst much progress has been made in the search for novel miRNA biomarkers of disease processes [11,12], results of this research are often inconsistent or even contradictory [5,13]. There are many reasons for this, including variations in enrichment, extraction and quantification methods, variation between individuals, lack of consensus regarding optimal reference miRNA for normalisation and the difficulty in quantifying both the amount and quality of RNA transcripts from blood plasma samples [14,15]. An important but often overlooked factor, is the potential for sample haemolysis during blood collection or sample preparation which results in miRNA from lysed red blood cells being spilled into and retained within the plasma sample to be assayed [14].

The issue of haemolysis contamination altering the miRNA content of plasma and the potential for confounding biomarker discovery has been reported previously [14,16,17]. Using RT-qPCR Kirschner and colleagues (2011) showed that contamination of plasma samples with the miRNA content of red blood cells changed the abundance of both miR-16 and miR-451. This in turn altered the relative abundance of potential biomarkers for mesothelioma and coronary artery disease including miR-92a and miR-15. Using the same technique Pritchard *et al.* (2012) demonstrated in plasma, that 46 of the then known 79 circulating miRNA cancer biomarkers were highly expressed in more than one blood cell type, noting that the effects of sample specific blood cell counts and haemolysis can alter the miRNA biomarker levels in a single patient sample up to 50-fold. As a result the authors emphasised caution in classifying blood cell associated miRNAs as biomarkers given the alternate interpretation possible.

Haemolysis is associated with either blood collection or RNA extraction and sample preparation. Thus, despite differences between the quantification methods, high throughput sequencing data used in our study is equally susceptible to the confounding effects of sample haemolysis on miRNA abundance levels in plasma as RT-qPCR. In the assessment of haemolysis contamination in plasma samples, there are currently two gold standard approaches: 1. Delta quantification cycle (ΔCq), where expression levels of a known blood cell associated miRNA (miR-451) and a control miRNA (miR-23a) are determined based on raw Cq values and the difference between the two is calculated, and 2. Spectrophotometry, where absorbance is measured at 414 nm with the use of a spectrophotometer. In the case of ΔCq assessment, miR-451 is known to vary and miR-

23a is known to be invariant in plasma affected by haemolysis [14,15]. Using spectrophotometry, haemolysis is quantified by assessing the presence of cell free haemoglobin by measuring the absorbance at 414 nm, the absorbance maximum of free haemoglobin [18,19]. Both methods however, require access to sufficient amounts of the original plasma sample, laboratory equipment and further laboratory tests. Free access to a web-tool that can perform *in silico* assessment of red blood cell contamination in human plasma would be of exceptional value to the research community.

Whilst it is well established that haemolysis frequently occurs during extraction or processing of blood samples, the assessment of red blood cell contamination is rarely mentioned in publications. It is even more rare that the results of any such testing are present in the metadata with publicly available sequencing data. There is currently no publicly available tool for analysis of haemolysis where there is no access to the physical plasma sample. Although the theory underlying identification of haemolysis in plasma is relatively straight-forward, surprisingly, this has never before been extrapolated into a data-only approach. The paucity of haemolysis information in the context of publicly available datasets combined with the lack of tools to identify affected datasets after the fact, substantially limits the usefulness and reproducibility of this material. Further, it increases the risk that results obtained may unwittingly represent blood-cell based phenomena rather than signatures of the pathology of interest.

In this study, we assessed miRNA abundance in high throughput sequencing data from libraries prepared using human plasma from pregnant and non-pregnant women of reproductive age. Using a set of samples with confirmed haemolysis (ΔCq (miR-23a-miR-451)), we established a set of 20 miRNAs differentially abundant between plasma from samples with and without substantiated haemolysis. Using the expression values of these 20 miRNAs as a ‘signature’ of haemolysis, we calculated the difference between the mean normalised expression levels of these miRNAs compared to those of all other miRNAs (as a ‘background’ set). This produced a quantitative metric which represents the strength of the evidence of haemolysis in an individual sample. When this metric is interpreted in the context of other samples, it can be used to notify the researcher about sample(s) that display substantial evidence of haemolysis. The researcher may consider discarding these samples from further analyses or using caution in their interpretation. We consulted the EMBL-EBI Expression Atlas (ebi.ac.uk) to ensure all signature miRNAs are identified in multiple human tissues (male and female) and have no known developmental stage association. For ease of application we have developed this method into a web based Shiny/R application, DraculR (a tool that allows a user to upload and assess haemolysis in high-throughput plasma miRNA-seq data), for use by the research community (as detailed in Chapter 5).

Results

Illumina NextSeq 75 bp single-end read sequencing was performed on miRNA libraries from 154 plasma samples taken from 24 non-pregnant and 130 pregnant women aged 16 to 46 years (Qiagen, Hilden, Germany). Prior to sequencing, RT-qPCR was used to analyse ΔCq (miR-23a-miR-451), where the ratio of miR-23a to miR-451 (or ΔCq (miR-23a-miR-451) ≥ 7) correlates with the degree of haemolysis. We identified 14 plasma samples with a ΔCq of 7 or above (Supplementary Table 1). An average of ~2.9 million reads were sequenced per sample (range ~0.25-18.6 million reads; Supplementary Table 2). Thirty one libraries with < 1 million reads were considered to be unreliable due to low sequencing output and were removed from further analyses. There was no difference in the proportion of haemolysed and non-haemolysed data in the exclusion of samples due to low library size (Fisher's exact test p-value = 0.7). Sequence alignment was performed using BWA [20] to the human genome (version GRCh38) and miRNA read counts were generated by mapping to human miRBase v22 [21,22] identifying 1,133 mature miRNAs.

To analyse the effects of haemolysis on miRNA expression data from next generation sequencing, we first determined the number of unique mature miRNAs identified in each of our samples and analysed the data relative to read depth. Using an analysis of variance (ANOVA; p-value < 0.05) we identified a significant difference between the haemolysed and non-haemolysed samples, with haemolysis being frequently associated with fewer mature miRNA species detected at a given read depth (Figure 1).

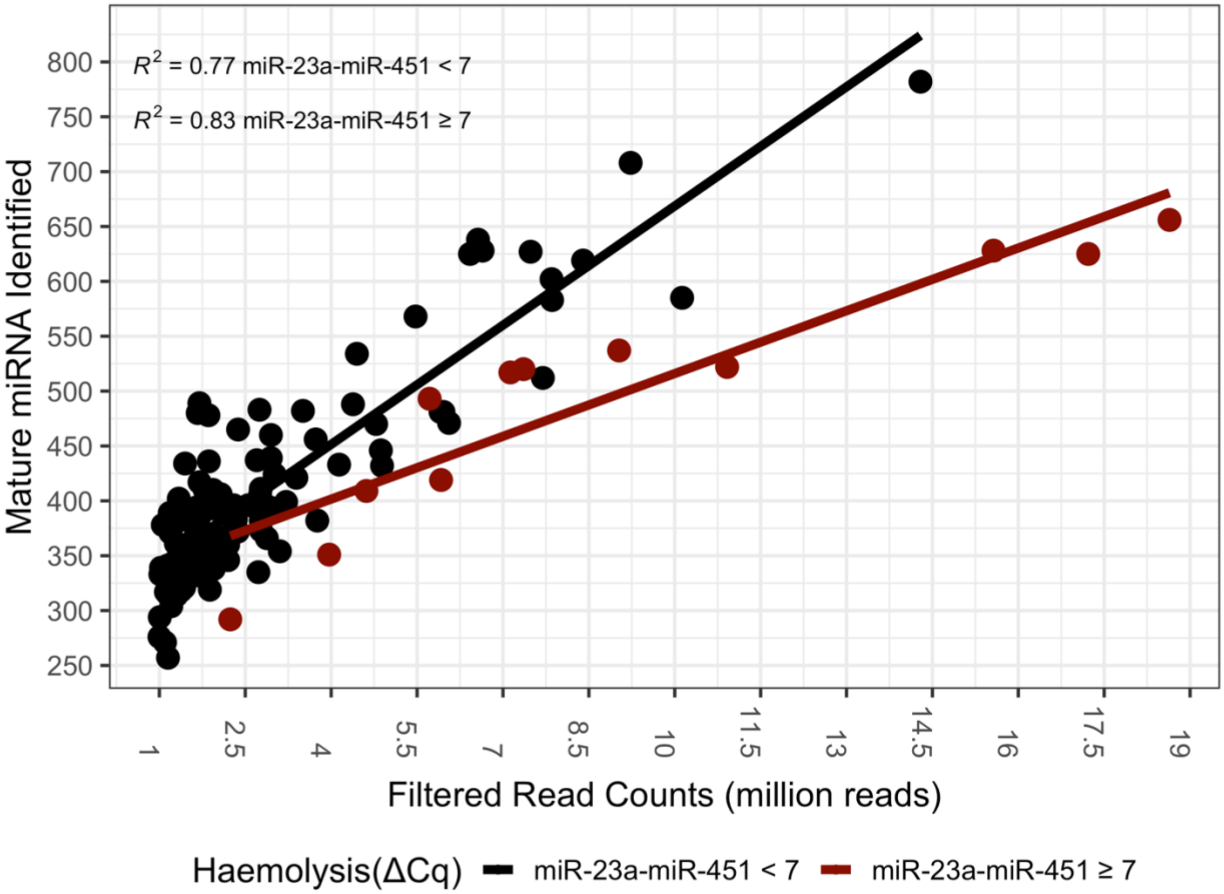


Figure 1. The number of mature miRNA species identified in an individual sample increases with read depth for both haemolysed and non-haemolysed samples however, the number of mature miRNA species identified for a given read depth is significantly lower in samples affected by haemolysis (blue) when compared to a non-haemolysed sample (red) of equal read depth.

microRNA haemolysis signature set

To ensure that miRNAs identified here were representative of those found in a broad set of plasma samples, we first filtered to discard miRNAs of low abundance. After filtering, 189 highly abundant miRNA remained. Differential expression analysis comparing miRNA read counts identified 138 miRNAs with a higher abundance in haemolysed compared to non-haemolysed samples (statistically significant differentially expressed miRNA, false discovery rate (FDR) < 0.05, with a $\log_2$ fold change (logFC) > 0)) (Supplementary Figure 1; Supplementary Figure 2(a,b)). We further ranked the differentially expressed miRNAs based on logFC, FDR and abundance levels and subset the list such that only miRNAs which had a logFC > 0.9 and were in the top 60 percent of each of the FDR and abundance rank criteria remained. A high confidence set of 20 miRNA remained, indicative of a haemolysis signature (Table 1).

For *in silico* assay of haemolysis in our samples, we further removed miRNAs associated with pregnancy to avoid confounding miRNA associated with haemolysis and pregnancy. Differential expression analysis of miRNA read counts from pregnancy and non-pregnancy samples identified 127 miRNAs (FDR < 0.05) that were significantly differentially expressed between the groups (Supplementary Figure 3(a,b)). Strikingly, one of our first observations highlighted the importance of including haemolysis analysis as an adjunct in our study: miR-451, which is the sole haemolysis signature miRNA used in the current ΔCq (miR-23a-miR-451) gold standard method for haemolysis detection was discovered to be highly correlated with pregnancy status, indicating a strong confounding factor in pregnancy studies when haemolysis levels are estimated using RT-qPCR alone. Accordingly, miR-451 was removed from calculations hereafter along with 9 other miRNAs that were differentially expressed between the pregnant and non-pregnant groups from the core set of haemolysis signature miRNA. This resulted in 10 miRNAs remaining for evaluation of haemolysis levels.

Table 1. 20 miRNAs with a general-use plasma haemolysis signature set. To remove confounding effects within our pregnancy-specific dataset, we identified a subset of 10 abundant miRNA which are invariant with respect to pregnancy.

miRNA	logFC	Average Expression ($\log_2$ CPM)	Adjusted p-value	Pregnancy Assoc.
miR-106b-3p	1.589	8.731	8.61E-15	no
miR-140-3p	1.073	10.098	2.75E-13	no
miR-142-5p	0.962	10.651	4.96E-12	no
miR-532-5p	1.288	7.237	4.96E-12	no
miR-17-5p	0.952	7.892	7.84E-12	no
miR-19b-3p	1.128	8.696	1.93E-09	no
miR-30c-5p	0.95	7.325	2.48E-09	no
miR-324-5p	1.304	7.186	2.50E-09	no
miR-192-5p	0.941	8.944	1.37E-08	no
miR-660-5p	1.305	7.62	3.45E-10	no
miR-186-5p	1.228	8.052	2.75E-13	yes
miR-425-5p	1.282	11.246	4.96E-12	yes
miR-25-3p	1.212	12.939	1.26E-11	yes
miR-363-3p	1.237	7.882	4.52E-11	yes
miR-183-5p	1.55	9.382	9.34E-11	yes
miR-451a	1.372	13.002	3.65E-10	yes
miR-182-5p	1.341	10.585	2.48E-09	yes
miR-191-5p	0.929	11.79	4.68E-09	yes
miR-194-5p	0.937	7.679	1.85E-08	yes
miR-20b-5p	0.932	7.43	1.96E-08	yes

Incorporating concepts from previous RT-qPCR analyses of haemolysis, we established a new measure of the inclusion of red blood cell associated miRNA in human plasma. After establishing the 20 miRNA signature associated with RBC content inclusion, we determined the geometric mean of the distribution of miRNA read counts as an appropriate measure of abundance and summary statistic. Using this summary statistic, this method calculates a ‘Haemolysis Metric’, defined as the difference between the geometric means of the normalised abundance levels of the haemolysis miRNA signature set compared to that of all other miRNAs (the ‘background’ set). Note that in the case of a condition:control study, to reduce the risk of confounding the Haemolysis Metric with experimental variables, the signature set should be reduced to exclude any miRNA known to be differentially expressed between groups. In this case, the geometric mean of the reduced signature set will be calculated, as defined in (1).

Let
Z = miRNA gene set, $\log_2$ CPM counts
x = 1, 2, 3,...,p_1 where p_1 = the number of miRNA in Reduced signature set
y = 1, 2, 3,...,p_2 where p_2 = the number of miRNA in Background
i = 1, 2, 3,...,n where n = the sample size after filtering

$$Haemolysis\ Metric = \sqrt[P_1]{\prod_{x=1}^{P_1} Z_{x_i}} - \sqrt[P_2]{\prod_{y=1}^{P_2} Z_{y_i}} \qquad (1)$$

Prior to establishing a threshold for the new Haemolysis Metric we measured the linear dependence between the new Haemolysis Metric and the ΔCq (miR-23a-miR-451) metric by performing a Pearson’s correlation. Our results indicated a Pearson’s rho of 0.64 (p=1.14e-15). With confidence in the correlation, to establish a threshold for the Haemolysis Metric we compared the results of the ΔCq (miR-23a-miR-451) and summary statistic methods directly. Briefly, we compared the Haemolysis Metric to the ΔCq (miR-23a-miR-451) results for matched samples and established a cut off criteria for inclusion into the Clear (no haemolysis detected) and Caution (haemolysis detected) groups

(Figure 2a). We chose a threshold of ≥ 1.9 for the assignment of “Caution” to individual samples based on the minimum summary statistic difference of samples assayed using the ΔCq (miR-23a-miR-451) metric of ≥ 7 (Figure 2a) and the minimal overlap between the distribution of the Haemolysis Metric in haemolysed compared to non-haemolysed samples (Figure 2b). Where a sample is assigned “Caution”, researchers are advised to consider removing the sample, or to continue with caution. Given the correlation of the two metrics is imperfect, samples with a Haemolysis Metric close to the 1.9 cut-off may be interrogated further prior to any decision to retain or remove. Of the 121 samples assayed, 25 samples met the criteria for Caution. Of these, 12 were previously determined as haemolysed or borderline using the ΔCq (miR-23a-miR-451) assay. We found that all samples identified as ΔCq ≥ 7 (Figure 2a, scarlet) are above criteria for the Haemolysis Metric (Figure 2a, horizontal grey bar; threshold ≥ 1.9). Further, we identified 13 samples with a Haemolysis Metric ≥ 1.9 not included in the ΔCq (miR-23a-miR-451) criteria.

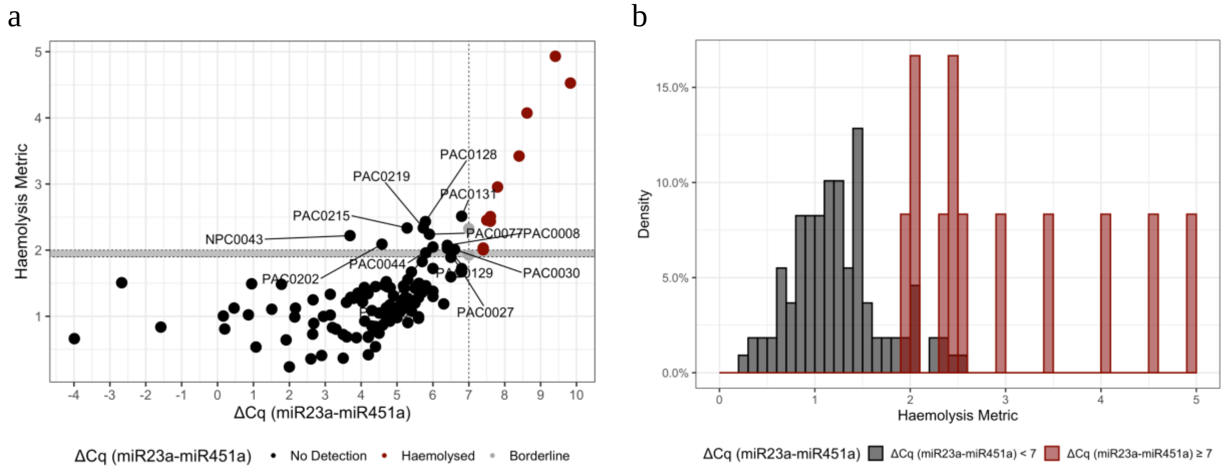


Figure 2: (a) A comparison of the derived Haemolysis Metric and the ΔCq measure of haemolysis shows a clear correlation. We identified 13 samples (named) that we suggest should be discarded or used with caution in further analysis. (b) Histogram of Haemolysis Metric values from the 121 samples in our experiment, coloured according to their ΔCq (miR-23a-miR-451) classification indicate a minimum Haemolysis Metric of ≥ 1.9 for samples previously identified as haemolysed.

Discussion

Through an analysis of differential miRNA expression in samples whose haemolysis levels were known, we identified a novel 20 miRNA signature indicative of haemolysis. Given our hypothesis that plasma samples contaminated with red blood cell (RBC) content would contain proportionally higher levels of many red blood cell-associated miRNA, not just miR-451, we established a method using a group of background miRNAs as a reference. Accordingly, as a group, signature miRNAs (microRNAs which are abundant in red blood cells) are shown to be more highly abundant in the presence of RBCs. The degree of this change can be used as a measure of RBC content contamination and quantified by comparing the geometric means of the expressions of RBC signature miRNAs to that of the background set of miRNAs. We further established that where a comparison between conditions is considered, for example in a biomarker discovery experiment, any miRNA known to be associated with the condition for which the biomarker is proposed should be removed to prevent confounding between the condition of interest and the quantification of red blood cell associated miRNA inclusion.

Our experimental results demonstrate that it is possible to identify a haemolysis signature *in silico*, avoiding the effort and expense of lab validation, and in situations where blood plasma samples are exhausted, otherwise unavailable or cost-prohibitive to assay using current gold standard approaches. Given the limited access to physical samples associated with publicly available data, the Haemolysis Metric technique introduced here, provides the basis for the development of a publicly available tool (detailed in 5).

Among the haemolysis miRNA signature is miR-451a (previously named miR-451), commonly associated with RBC contamination and used in the calculation of ΔCq (miR-23a-miR-451). However, we removed miR-451a together with 9 other miRNAs from our calculation of distribution difference due to changes in miRNA abundance associated with pregnancy. During pregnancy total blood volume increases varying between 20% to 100% above pre-pregnancy levels. This change however is not uniform across all blood components as plasma volume increases proportionally more than the red blood cell mass [23]. This is an important consideration and highlights the limitation of the current gold standard approach that uses two miRNAs rather than a larger signature set to calculate a measure of contamination. If, as in this example, the abundance of a miRNA used to determine the ΔCq is also affected by the condition or pathology under investigation the issue is two-fold. Firstly, you may identify miR-451a as being differentially abundant in the pathology of interest and propose its use as a biomarker only to find that it is confounded with haemolysis. Secondly, you may, using the ΔCq calculation, classify samples as haemolysed when the change in miR-451a abundance is more appropriately associated with the pathology of interest. By establishing a larger signature set of miRNAs to detect haemolysis in small RNA sequencing from human plasma we hope to provide a resource to the community with flexibility and redundancy included in the design to buffer against the issue of confounding conditions of interest with the measure of haemolysis.

We found limited overlap between the miRNAs identified as useful for the detection of haemolysis and those previously reported as markers of haemolysis contamination

[14,17,24]. It is important to note however, that our research question differed from that of the above studies, as did our study methodology. The most important technical difference is in the quantification of miRNAs. The expression values used here are taken from a high throughput sequencing experiment, rather than RT-qPCR used previously. The limitations of RT-qPCR to investigate which, if any, miRNAs are affected by haemolysis has been identified previously [25]. Given that high throughput sequencing allows for quantification of all known miRNA species and that RT-qPCR is targeted, our experiment was able to identify differential abundance in miRNAs not quantified in Kirschner *et al.* (2011) [14], Pritchard *et al.* (2012) [17] or McDonald *et al.* (2011) [24]. There was however greater overlap in the miRNAs identified as up-regulated in the haemolysis group than those in the final miRNA Haemolysis Metric signature set. MiR-16, miR-486-5p and miR-92a-3p were significantly upregulated in the haemolysed group, but not included in the signature set as they failed to pass filtering criteria for logFC and expression. Secondary to the technical differences introduced by the use of different miRNA quantification technologies, it is important to not that all plasma samples used here to establish which miRNAs are affected by haemolysis were taken from adult women of reproductive age. No sex or age information was included with either study compared, although it is likely these samples included specimens from men and women. To account for the potential bias introduced using data from only female and all reproductively aged volunteers, we ensured all miRNAs included here have previously been identified in multiple tissues in both male and female samples, and are not affected by developmental stage. Investigation using a cohort of mixed age and sex is warranted and may help to further determine which miRNAs are affected by haemolysis.

Interestingly, all signature miRNAs, with the exception of miR-325-5p, have previously been reported as prognostically valuable plasma or serum biomarkers. In this small sampling of recent miRNA biomarker research, we identified several instances where more than one of our haemolysis signature miRNA were identified as disease biomarkers for the same condition in the same experiment [26–28] which, given our findings, and those of previous haemolysis research, further call into question their validity as biomarkers of disease or condition. In conjunction with our research we found many miRNAs as suggested circulating biomarkers for multiple disease states. For example, miR-122 was given biomarker potential in liver disease, lung cancer and myasthenia gravis [26,29,30], and miR-660 was given biomarker potential in Alzheimer's disease, breast cancer and lung cancer [27,31,32] respectively. These miRNAs may represent effective biomarkers, or may simply highlight RBC contamination or be indicative of a general state of inflammation.

Data contained in this study were obtained from two cohorts of female volunteers of reproductive age. Whilst we are working in a relatively narrow experimental domain, we have generalised this method such that removal (from the signature miRNA set) of domain specific miRNA is built in, providing a framework that allows use within research conducted in any human plasma context. Our results highlight that ignoring the issue of miRNA from red blood cells leaves researchers open to the risk that newly discovered miRNA disease biomarkers could in fact be biomarkers of haemolysis. In future research, a repeat experiment with samples taken from male and female individuals across a wider age range would expand and strengthen our understanding of the impact of haemolysis on discovery of biomarkers. Our research both recommends and enables tests for haemolysis to become standard pre-analytical practice.

Methods

Sample collection

Peripheral blood (9 mL) was collected with informed, written consent from women undergoing elective terminations of otherwise healthy pregnancies. Blood was collected into standard EDTA blood tubes pre-termination and stored on ice until processed. Whole blood underwent centrifugation at 800 x g for 15 minutes at 4°C before plasma removal and then spun for a further 15 minutes to ensure any remaining cellular debris, including cell membranes from lysed red blood cells, was removed. All samples were stored at -80°C until further processing. Prior to termination, accurate gestational age was determined using transvaginal ultrasonography. Termination samples were collected from the Pregnancy Advisory Centre (PAC), Woodville, South Australia. Blood was also collected with informed, written consent from non-pregnant volunteers at the Adelaide Medical School. Following collection, blood tubes were stored on ice until processing. Whole blood underwent centrifugation at 1015 x g for 10 minutes at 4°C. Approximately 4-6 mL plasma was collected in 2 mL aliquots. 500 µL plasma (the supernatant) were aliquoted into clean tubes and the pellet containing any remaining blood cells at the bottom of the tube was discarded. All samples were stored at -80°C until further processing.

RNA extraction and library preparation

MicroRNA was isolated from 200 µL plasma samples using the QIAGEN miRNA serum/plasma kit (Qiagen, Hilden, Germany) according to the manufacturer’s instructions. All samples were stored at -80°C until further processing.

Library preparation and sequencing was performed by Qiagen (Valencia, CA) using the QIAseq miRNA Library Kit and QIAseq miRNA 48 Index IL kits as per manufacturer’s instructions. Amplified cDNA libraries underwent single-end sequencing by synthesis (Illumina 1.9).

Haemolysis Detection

Plasma samples were examined for haemolysis based on the expression levels of two miRNAs: miR-451 and miR-23a. miR-451 (recently renamed miR-451a) is known to be highly expressed in red blood cells, whereas miR-23a is known to maintain stable abundance levels in plasma. After RNA extraction and cDNA synthesis, the delta quantification cycle (Cq) values for miR-23a-miR-451 were calculated independently for each sample. The evaluation of expression levels was performed based on raw Cq values. According to the Qiagen protocol for haemolysis detection (Qiagen, Hilden, Germany) using the ΔΔCq method, samples with a ΔCq <7 for these two miRNAs were considered as clear of contamination; a ΔCq >7 was considered contaminated; a ΔCq =7 was considered borderline.

miRNA annotation and abundance

Technical variability from library amplification was removed by using umi_tools to collapse mappable reads with the same UMI barcode. Adapter detection and trimming were performed using Atropos [33]. Alignment performed using BWA version 0.7.17-r1188 (GRCh38) and output in FASTQ format [20]. Quality control metrics were assessed using FastQC [34] (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) to check for per base sequence quality, sequence length distribution and duplication levels. Quality control metrics were reported using multiQC [35]. Counts for mature miRNAs were calculated using an in-house script [36] with miRBase version 22.0 [21,22] (http://www.mirbase.org) as the reference miRNA database.

Analysis of potential confounding factors

All profile and expression analyses were conducted in the R statistical environment (v.4.0.2), using the *edgeR* (v.3.16.5) [37] and *limma* (v.3.30.11) [38] R/Bioconductor packages. Prior to conducting the differential expression analysis between haemolysed and non-haemolysed expression data we considered the effect of participant characteristics such as sex, age, smoking, pregnancy status and ethnicity. Sex was not considered here as all samples were taken from female participants, and likewise age was confined to women of reproductive age. Of the remaining factors only pregnancy status was considered with miRNA identified as differentially expressed between pregnant and non-pregnant samples removed from the final set of haemolysis signature miRNA. Sequencing batch was identified and included in all regression models.

Feature calculation/Identification of haemolysis miRNA signature

Prior to defining a collection of haemolysis informative miRNAs, pre-filtering steps were undertaken: 1) mature miRNA with fewer than five reads were reduced to zero independently for each sample, 2) miRNA with fewer than 40 counts per million (CPM) in the haemolysed group (n = 12) were removed from further consideration. This was done to ensure only highly abundant miRNA likely to be present in the majority of our samples, and in datasets not considered here, remain. The Trimmed Means of M values (TMM) normalisation method was used to correct for differences in the underlying distribution of miRNA expression [39]. Next, we used *limma* [38] to obtain the fold change of each miRNA between the haemolysed and non-haemolysed groups to identify miRNAs that are more abundant in the plasma affected by haemolysis. To ensure the haemolysis miRNA signature was robust we took the intersection of the 60 miRNAs from each category of highest expression and lowest adjusted p-value and miRNAs with a logFC > 0.9 revealing a set of twenty high confidence miRNAs. To further refine the set of haemolysis informative miRNAs we used *limma* to calculate the fold change for each miRNA between the pregnant and non-pregnant samples not affected by haemolysis and removed any miRNA common to both the haemolysis and the pregnancy comparison. The workflow, source code and input files associated with this research are available at (https://github.com/mxhp75/haemolysis_maternaPlasma).

Classification - Haemolysis Metric

To classify the data coming from samples as haemolysed, borderline or unaffected we first determined if the sample in question belongs to the non-pregnant or pregnant group.

For samples in the non-pregnant group we subset the miRNA read count table into miRNA from the high confidence haemolysis informative miRNA (n=20) and all others (n=169). Using this data partition we calculated the geometric mean of the distribution of read counts using the *psych* package (v1.8.12) [40] and next subtracted the geometric mean of the counts of "other" miRNA from that of the "haemolysis informative" miRNA. For samples from the pregnant group we performed the calculation of the difference between the geometric mean of each group in the same way however, before calculating the geometric mean of the "haemolysis informative" group we first removed and discarded the miRNA which were associated with pregnancy.

Shiny App for classification

A web based Shiny/R application, DraculR, was developed to facilitate use of this method. More detailed information regarding the application is available in Chapter 5.

Ethics

Ethics approval for the collection of blood from women undergoing elective pregnancy termination between 6–23 weeks' gestation was provided under HREC/16/TQEH/33, by The Queen Elizabeth Hospital Human Research Ethics Committee (TQEH/LMH/MH). Blood from women forming the general population group was collected after informed consent with ethics approval provided under HREC/H/021/2005, by The University of Adelaide Human Research Ethics Committee.

Author contribution

CTR created the concept. MS, KP and JB conceived and developed experimental plans. TJK, DMcA and DMcC performed experiments. MS and KP analysed the sequencing data with statistical support from SYL. Manuscript written by MS and KP. All authors read and approved the final version of the manuscript.

References

1. Bartel DP. MicroRNAs: Genomics, Biogenesis, Mechanism, and Function. Cell. 2004. pp. 281–297. doi:10.1016/S0092-8674(04)00045-5
2. Bartel DP. Metazoan MicroRNAs. Cell. 2018;173: 20–51.
3. Guo H, Ingolia NT, Weissman JS, Bartel DP. Mammalian microRNAs predominantly act to decrease target mRNA levels. Nature. 2010;466: 835–840.
4. Friedman RC, Farh KK-H, Burge CB, Bartel DP. Most mammalian mRNAs are conserved targets of microRNAs. Genome Res. 2008;19: 92–105.
5. Mitchell PS, Parkin RK, Kroh EM, Fritz BR, Wyman SK, Pogosova-Agadjanyan EL, *et al.* Circulating microRNAs as stable blood-based markers for cancer detection. Proc Natl Acad Sci U S A. 2008;105: 10513–10518.
6. Cortez MA, Calin GA. MicroRNA identification in plasma and serum: a new tool to diagnose and monitor diseases. Expert Opin Biol Ther. 2009;9: 703–711.
7. Kosaka N, Iguchi H, Ochiya T. Circulating microRNA in body fluid: a new potential biomarker for cancer diagnosis and prognosis. Cancer Sci. 2010;101: 2087–2092.
8. Vickers KC, Palmisano BT, Shoucri BM, Shamburek RD, Remaley AT. MicroRNAs are transported in plasma and delivered to recipient cells by high-density lipoproteins. Nat Cell Biol. 2011;13: 423–433.
9. Arroyo JD, Chevillet JR, Kroh EM, Ruf IK, Pritchard CC, Gibson DF, *et al.* Argonaute2 complexes carry a population of circulating microRNAs independent of vesicles in human plasma. Proceedings of the National Academy of Sciences. 2011;108: 5003–5008.
10. Cortez MA, Bueso-Ramos C, Ferdin J, Lopez-Berestein G, Sood AK, Calin GA. MicroRNAs in body fluids—the mix of hormones and biomarkers. Nat Rev Clin Oncol. 2011;8: 467–477.
11. Seco-Cervera M, González-Rodríguez D, Ibáñez-Cabellos JS, Peiró-Chova L, Pallardó FV, García-Giménez JL. Small RNA-seq analysis of circulating miRNAs to identify phenotypic variability in Friedreich's ataxia patients. Sci Data. 2018;5: 180021.
12. Ali SA, Gandhi R, Potla P, Keshavarzi S, Espin-Garcia O, Shestopaloff K, *et al.* Sequencing identifies a distinct signature of circulating microRNAs in early radiographic knee osteoarthritis. Osteoarthritis Cartilage. 2020;28: 1471–1481.
13. Witwer KW. Circulating microRNA biomarker studies: pitfalls and potential solutions. Clin Chem. 2015;61: 56–63.
14. Kirschner MB, Kao SC, Edelman JJ, Armstrong NJ, Vallely MP, van Zandwijk N, *et al.* Haemolysis during Sample Preparation Alters microRNA Content of Plasma. Pfeffer S, editor. PLoS One. 2011;6: e24145.
15. Cheng HH, Yi HS, Kim Y, Kroh EM, Chien JW, Eaton KD, *et al.* Plasma Processing Conditions Substantially Influence Circulating microRNA Biomarker Levels. Kiechl S, editor. PLoS One. 2013;8: e64795.

16. Kirschner MB, van Zandwijk N, Reid G. Cell-free microRNAs: potential biomarkers in need of standardized reporting. Front Genet. 2013;4: 56.
17. Pritchard CC, Kroh E, Wood B, Arroyo JD, Dougherty KJ, Miyaji MM, *et al.* Blood cell origin of circulating microRNAs: a cautionary note for cancer biomarker studies. Cancer Prev Res . 2012;5: 492–497.
18. Livak KJ, Schmittgen TD. Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2−ΔΔCT Method. Methods. 2001. pp. 402–408. doi:10.1006/meth.2001.1262
19. Wong C-H, Song C, Heng K-S, Kee IHC, Tien S-L, Kumarasinghe P, *et al.* Plasma free hemoglobin: a novel diagnostic test for assessment of the depth of burn injury. Plast Reconstr Surg. 2006;117: 1206–1213.
20. Li H, Durbin R. Fast and accurate long-read alignment with Burrows–Wheeler transform. Bioinformatics. 2010;26: 589–595.
21. Griffiths-Jones S, Grocock RJ, van Dongen S, Bateman A, Enright AJ. miRBase: microRNA sequences, targets and gene nomenclature. Nucleic Acids Res. 2006;34: D140–4.
22. Kozomara A, Griffiths-Jones S. MiRBase: Annotating high confidence microRNAs using deep sequencing data. Nucleic Acids Res. 2014;42. doi:10.1093/nar/gkt1181
23. Sanghavi M, Rutherford JD. Cardiovascular physiology of pregnancy. Circulation. 2014;130: 1003–1008.
24. McDonald JS, Milosevic D, Reddi HV, Grebe SK, Algeciras-Schimnich A. Analysis of circulating microRNA: preanalytical and analytical challenges. Clin Chem. 2011;57: 833–840.
25. Kirschner MB, Edelman JJB, Kao SCH, Vallely MP, Van Zandwijk N, Reid G. The impact of hemolysis on cell-free microRNA biomarkers. Front Genet. 2013;4: 94.
26. Nogales-Gadea G, Ramos-Fransi A, Suárez-Calvet X, Navas M, Rojas-García R, Mosquera JL, *et al.* Analysis of serum miRNA profiles of myasthenia gravis patients. PLoS One. 2014;9: e91927.
27. Guo R, Fan G, Zhang J, Wu C, Du Y, Ye H, *et al.* A 9-microRNA Signature in Serum Serves as a Noninvasive Biomarker in Early Diagnosis of Alzheimer's Disease. J Alzheimers Dis. 2017;60: 1365–1377.
28. Poroyko V, Mirzapoiazova T, Nam A, Mambetsariev I, Mambetsariev B, Wu X, *et al.* Exosomal miRNAs species in the blood of small cell and non-small cell lung cancer patients. Oncotarget. 2018;9: 19793–19806.
29. Wang X, He Y, Mackowiak B, Gao B. MicroRNAs as regulators, biomarkers and therapeutic targets in liver diseases. Gut. 2021;70: 784–795.
30. Wozniak MB, Scelo G, Muller DC, Mukeria A, Zaridze D, Brennan P. Circulating MicroRNAs as Non-Invasive Biomarkers for Early Detection of Non-Small-Cell Lung Cancer. PLoS One. 2015;10: e0125026.
31. Zou X, Li M, Huang Z, Zhou X, Liu Q, Xia T, *et al.* Circulating miR-532-502 cluster

derived from chromosome X as biomarkers for diagnosis of breast cancer. Gene. 2020;722: 144104.

32. Boeri M, Verri C, Conte D, Roz L, Modena P, Facchinetti F, *et al.* MicroRNA signatures in tissues and plasma predict development and prognosis of computed tomography detected lung cancer. Proc Natl Acad Sci U S A. 2011;108: 3713–3718.
33. Didion JP, Martin M, Collins FS. Atropos: specific, sensitive, and speedy trimming of sequencing reads. PeerJ. 2017;5: e3720.
34. Andrews S. FastQC: a quality control tool for high throughput sequence data. 2010. Available: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/
35. Ewels P, Magnusson M, Lundin S, Käller M. MultiQC: Summarize analysis results for multiple tools and samples in a single report. Bioinformatics. 2016;32: 3047–3048.
36. Saunders K, Bert AG, Dredge BK, Toubia J, Gregory PA, Pillman KA, *et al.* Insufficiently complex unique-molecular identifiers (UMIs) distort small RNA sequencing. Sci Rep. 2020;10: 14593.
37. Robinson MD, McCarthy DJ, Smyth GK. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2009;26: 139–140.
38. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, *et al.* Limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43: e47.
39. Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol. 2010;11: R25.
40. Revelle W. psych: Procedures for psychological, psychometric, and personality research. R package version. 2018;1. Available: http://www2.uaem.mx/r-mirror/web/packages/psych/

Supplementary Information Legends

Note: Supplementary Figures and Tables are hosted on Figshare

(https://figshare.com/projects/Chapter_4_-_Supplementary/122501)

Supplementary figure 1: Haemolysis signature feature selection. Raw single-end reads from smallRNA-seq libraries are preprocessed using a range of Unix- and python-based computational tools to quantify miRNA expression in each library. Data quality is ensured through quality control steps throughout the workflow. Concurrently with sequencing, ΔCq (miR-23a-miR-451) was assessed by RT-qPCR and incorporated into the differential expression analysis.

Supplementary Figure 2: (a) Volcano plot of differential expression. Linear regression identified 138 miRNA which were more highly abundant in haemolysed compared to non-haemolysed samples with FDR < 0.05 (green). (b) MA plot (M (log ratio) and A (mean average)) of Log_2 fold change as a function of Log_2 average expression indicates most miRNA have an average expression < 10 Log_2 CPM. miR-451a and miR-16-5p, both highly red blood cell associated, are highly expressed and more highly abundant in the haemolysed group (green).

Supplementary Figure 3: (a) Volcano plot of differential expression. Linear regression 104 miRNA (FDR < 0.05) which were more highly abundant in the pregnant population compared to non-pregnant samples (red). Haemolysis Metric signature miRNAs (labelled) (b) MA plot (M (log ratio) and A (mean average)) of Log_2 fold change as a function of Log_2 average expression indicates most miRNA have an average expression < 10 Log_2 CPM. Unsurprisingly, the most differentially expressed miRNA are miR-517a-3p, miR-517b-3p, miR-516b-5p, miR-518b, all part of the highly placenta associated chromosome 19 miRNA cluster.

Supplementary Table 1: RT-qPCR Cq data for miR-23a-3p, miR-451a and ΔCq (miR-23a-miR-451).

Chapter 5

DraculR: A web based application for *in silico* haemolysis detection in high throughput small RNA sequencing data

Statement of Authorship

Title of Paper	DraculR: A web based application for in silico haemolysis detection in high throughput small RNA sequencing data
Publication Status	☐ Published ☐ Accepted for Publication ☐ Submitted for Publication ☒ Unpublished and Unsubmitted work written in manuscript style
Publication Details	DraculR: A web based application for in silico haemolysis detection in high throughput small RNA sequencing data Melanie D. Smith, Shalem Y. Leemaqz, Tanja Jankovic-Karasoulos, Dale McAninch, Dylan McCullough, James Breen, Claire T. Roberts, Katherine Pillman

Principal Author

Name of Principal Author (Candidate)	Melanie Smith		
Contribution to the Paper	- All bioinformatics analyses - Drafting of manuscript - Producing figures / visualisation of data		
Overall percentage (%)	90%		
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.		
Signature		Date	13/09/2021

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);

ii. permission is granted for the candidate in include the publication in the thesis; and

iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	Shalem Y. Leemaqz		
Contribution to the Paper	- Statistical support - Helped design experiments		
Signature		Date	13/9/2021

Name of Co-Author	Tanja Jankovic-Karasoulos		
Contribution to the Paper	- Performed experiments - Editing of manuscript drafts		
Signature		Date	13/9/2021

Please cut and paste additional co-author panels here as required.

Name of Co-Author	Dale McAninch		
Contribution to the Paper	- Performed experiments		
Signature		Date	13/09/2021

Name of Co-Author	Dylan McCullough		
Contribution to the Paper	- Performed experiments		
Signature		Date	13/09/2021

Name of Co-Author	James Breen		
Contribution to the Paper	- Conceived and developed experimental plans - Editing of manuscript drafts		
Signature		Date	13/09/2021

Name of Co-Author	Claire T Roberts		
Contribution to the Paper	- Conceived and developed experimental plans - Editing of manuscript drafts		
Signature		Date	13/09/2021

Name of Co-Author	Katherine Pillman		
Contribution to the Paper	- Expert miRNA analyses support - Development of experimental plans - Editing of manuscript drafts		
Signature		Date	13/9/2021

DraculR: A web based application for *in silico* haemolysis detection in high throughput small RNA sequencing data

Melanie D. Smith[1,2,3]*, Shalem Y. Leemaqz[1,2,4], Tanja Jankovic-Karasoulos[1,2,4], Dylan McCullough[1,2,4], Dale McAninch[2], James Breen[1,2,3]*, Claire T. Roberts[1,2,4], Katherine Pillman[5]*

[1] Robinson Research Institute, University of Adelaide, Adelaide SA 5005, Australia
[2] Adelaide Medical School, University of Adelaide, Adelaide SA 5005, Australia
[3] South Australian Health & Medical Research Institute, Adelaide SA 5000, Australia
[4] Flinders Health and Medical Research Institute, Flinders University, Bedford Park SA 5042, Australia
[5] Centre for Cancer Biology, University of South Australia/SA Pathology, Adelaide SA 5000, Australia

* Correspondence: melanie.smith@adelaide.edu.au & Katherine.Pillman@unisa.edu.au

Abstract

Motivation: Often medical biologists find they have blood plasma samples that are known or suspected to have been subjected to haemolysis, the rupture and subsequent release of red blood cell contents, including microRNAs (miRNAs), into surrounding fluid such as plasma. Individuals new to plasma research may be unaware of the issue entirely. The inclusion of red blood cell derived miRNA transcripts in downstream analyses, including but not limited to the analysis of small RNA quantification through high throughput sequencing data, introduces a source of error that is difficult to identify *post hoc* and may lead to spurious results. This error arises when miRNAs usually contained within red blood cells remain in the plasma after centrifugation, effectively combining the two otherwise independent biological compartments.

In reality, patient blood samples frequently include lysed red blood cells. The combination of plasma derived and red blood cell derived miRNA artificially increases the number of red blood cell associated transcripts in the plasma sample and alters the relative abundance of non-red blood cell miRNA. Importantly, this altered abundance may impact research outcomes including those of novel biomarker discovery. miRNAs are an important research focus given their potential as minimally invasive biomarkers. This potential comes in part from their multi-compartment origin, and the long-lived nature of miRNA transcripts in plasma, giving researchers a functional window for tissues that are otherwise difficult or disadvantageous to sample. Where access to the physical plasma sample is possible, it is prudent to assess any potential bias introduced by haemolysis.

Where access to a physical sample is not possible, our tool will provide an *in silico* approach to haemolysis prediction.

Results: We present DraculR, an interactive Shiny/R application that enables a user to upload microRNA expression data from short read sequencing of human plasma as a raw read counts table and interactively calculate a metric that indicates the degree of haemolysis contamination.

Availability and implementation: DraculR and its tutorial are freely available from (https://mxhp75.shinyapps.io/shinyVamp/)

Introduction

MicroRNAs have long been identified in human plasma and, given their stability in this medium, have strong potential as biomarkers. Whilst there are multiple techniques for quantifying the abundance of miRNAs in plasma, high throughput sequencing has the advantage of detecting both known and novel (ie. putative) miRNAs with single base resolution. This fine resolution enables distinction between variants differing by a single nucleotide as well as isomiRs of differing lengths [1]. With this in mind many researchers are implementing high throughput sequencing technologies to identify and quantify the abundance of plasma miRNAs [2–10].

The quantification of RNA transcripts using high throughput sequencing technologies expresses the value of each annotated transcript relative to the total number of sequence reads for a given sample [1]. Given that plasma miRNAs can originate from different cell compartments, accurate profiling of plasma miRNAs requires that miRNAs present in a given sample be plasma derived and not from another blood source such as red blood cells. Haemolysis occurs when red blood cells are sheared during extraction or handling, effectively spilling their contents, including miRNAs, into the volume of blood drawn [11–15]. The effect of this leakage is to change the plasma expression profile through the presence of red blood cell-associated miRNAs, and in turn to alter the global normalisation of sequence counts [1].

Both the increase in relative abundance of red blood cell-associated miRNAs and the aberrant normalisation of libraries have potential to impact the profile analysis of miRNA in plasma [11,12,14] yet assessment of haemolysis is rarely reported. Data quality checks prior to downstream analysis of high throughput sequencing data should include an assessment of haemolysis in the plasma sample from which the sequencing library was produced. For the assessment of haemolysed plasma samples, there are currently two gold standard approaches: 1. Delta quantification cycle (ΔCq) where expression levels of variable miR-451a and invariant miR-23a are determined based on raw Cq values and the difference between the two calculated and 2. Spectrophotometry, based on absorbance maximum of free haemoglobin measured at 414 nm [11,16] [17,18]. However, quantification of haemolysis through current gold standard methods such as ΔCq or absorption are not always possible, presenting a need for an alternative method that does not rely on access to the original plasma sample. In this work we present DraculR, a data driven approach for the assessment of haemolysis confounding *in silico*. The DraculR tool enables the user to upload self generated or publicly available high throughput miRNA sequencing data for assessment and returns both visual and tabular recommendations for downstream analysis of read count data.

Materials and methods

DraculR is an interactive, Shiny/R web based tool, for the *in silico* assessment of haemolysis contributions to small RNA sequencing libraries prepared from human plasma. This “Haemolysis metric” takes advantage of the increased proportion of red blood cell-associated miRNAs (identified in Chapter 4) in the presence of haemolysis contamination. Analogous to the ΔCq (miR-23a-miR-451) method, which determines the difference between the abundance of two miRNA, one known to vary and one known to be invariant in the presence of haemolysis, we present the Haemolysis metric, the difference between a set of miRNAs known to vary and a set of miRNAs known to be invariant in the presence of haemolysis. The Haemolysis metric is calculated as the sample specific difference in geometric means of the normalised gene expression values between two sets of microRNAs: 1) 20 miRNAs identified as indicative of haemolysis ('signature set'), and 2) all other microRNAs ('background'). In this case, the geometric mean of the reduced signature set will be calculated, as defined in (1).

Let
Z = miRNA gene set, $\log_2$ CPM counts
x = 1, 2, 3,...,p_1 where p_1 = the number of miRNA in Reduced signature set
y = 1, 2, 3,...,p_2 where p_2 = the number of miRNA in Background
i = 1, 2, 3,...,n where n = the sample size after filtering

$$Haemolysis\ Metric = \sqrt[P_1]{\prod_{x=1}^{P_1} Z_{x_i}} - \sqrt[P_2]{\prod_{y=1}^{P_2} Z_{y_i}}$$

(1)

In Figure 1 we present an example of samples considered as "Clear" (Figure 1a) and as "Caution" (Figure 1b). In the first example the distance between the geometric mean of the "Background" miRNA compared to that of the "Classifier" miRNA is very small suggesting the two sets of miRNA belong to the same distribution. In the second example the distance between the geometric mean of the "Background" miRNA compared to that of the "Classifier" miRNA is larger than the Clear example. Furthermore, this difference is greater than the 1.9 cut-off established in Chapter 4, and thus classified as "Caution". The

“Background” (light blue) and “Classifier” (scarlet) distributions appear to be independent suggesting that red blood cell associated miRNA have been added to the pool of miRNA isolated in the plasma. In the case of sample *Control_33* (Figure 1b), we would recommend removing the sample data from further analysis; however, where a decision is made to retain samples, the issue of haemolysis should be noted.

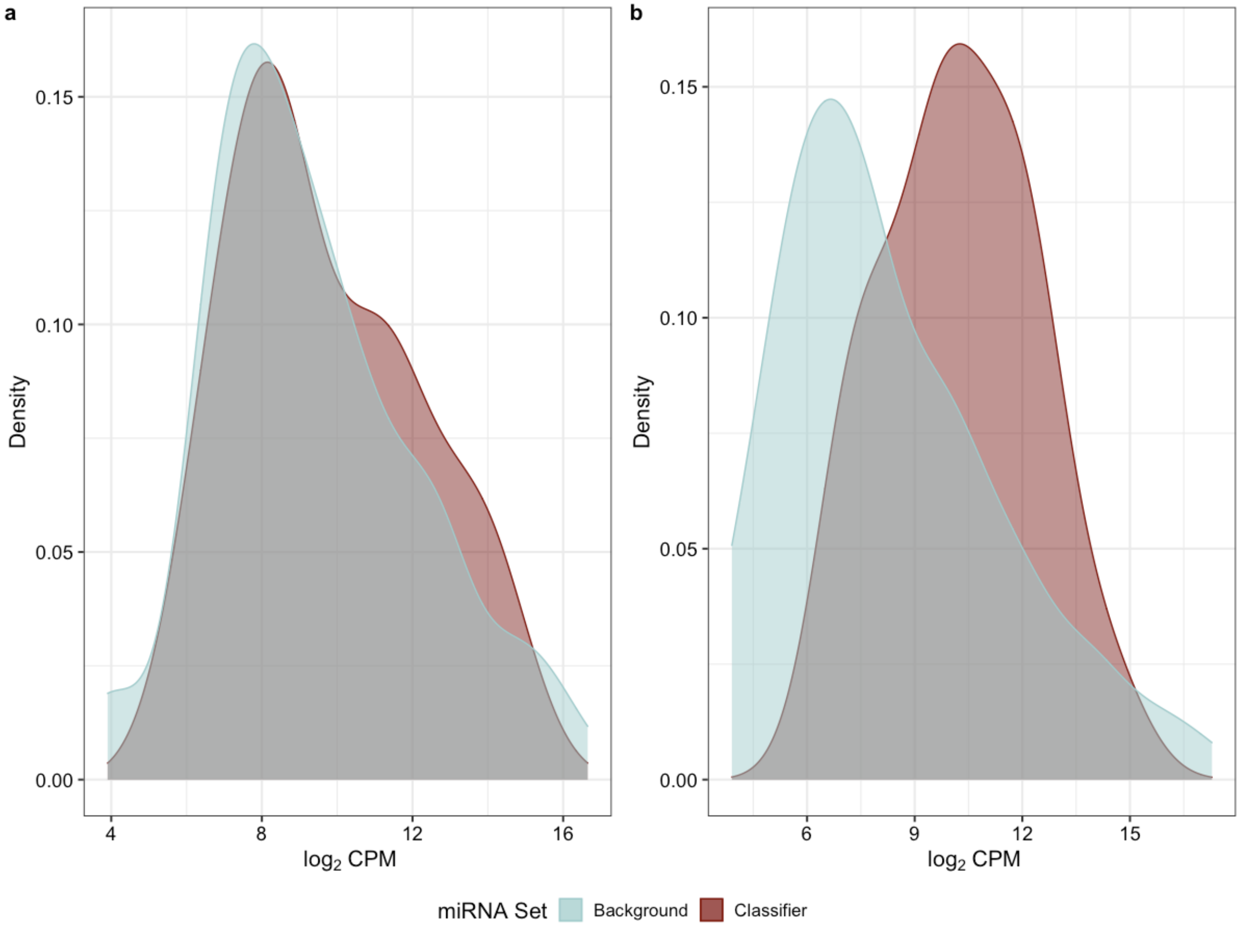


Figure 1. Patient_6 (a) is marked as “Clear” indicating no evidence for haemolysis in this sample. The distance between the geometric mean of “Background” and “Classifier” miRNA is small. Control_33 (b) is marked as “Caution” indicating that we found evidence suggestive of haemolysis. The geometric mean of “Background” and “Classifier” miRNA is ≥1.9 cut-off.

Application

By using the Haemolysis metric calculation one can assess their sample for evidence of haemolysis and obtain recommendations for new, user defined, individual samples as clear for use or use with caution. We have shown that when red blood cell-associated miRNA transcripts are retained in a plasma sample post centrifugation, and subsequently incorporated into the sequencing library, the relative abundance of these miRNAs is increased. This increase is evidenced by the increase in the geometric mean of signature haemolysis miRNAs away from that of the background miRNA giving the background miRNA a smaller relative expression than would be expected from a pure plasma sample taken from the same individual. Using high throughput sequencing data, DraculR is designed to visualize and analyse the distribution of miRNA counts from our haemolysis signature set and compare this with the distribution of counts from background miRNAs. Once calculated, the sample specific Haemolysis metric for user defined data is returned in tabular and graphical format for download and assessment. Samples with a Haemolysis metric ≥ 1.9, our recommended threshold, indicate haemolysis as evidenced by the red blood cell-associated miRNA retention and are identified as '*Caution'.* We recommend removal or at a minimum further investigation of any samples that return a Haemolysis metric above the threshold set here prior to use in any downstream analysis.

Public data example

To illustrate the utility of the application we downloaded four publicly available human plasma high throughput sequencing miRNA datasets from NCBI GEO [19]. The datasets used here were GSE153813, GSE118038, GSE105052, GSE151341 [20–22]. Where the publication associated with the given dataset included miRNA differentially expressed between the conditions being considered, and these miRNA correspond with our haemolysis signature set, these miRNA were dropped from the Haemolysis metric calculation. In each example we were able to detect evidence of haemolysis in multiple samples (Table 1).

Table 1. Publicly available data were assessed for haemolysis using the DraculR method identifying multiple samples to use with caution in each dataset. No haemolysis information was included with the original dataset.

Dataset	Experimental Context	Samples	Caution	Differentially abundant miRNA	Publication
GSE153813	Patient:Control Profile miRNA expression at each stage of menstrual cycle; endometriosis	9	3	NA	NA
GSE118038	Patient:Control Prostate cancer biomarker	70	32	miR-4732-3p, let-7a, miR-26b-5p, miR-98-5p, miR-30c-5p* miR-21-5p	[22]
GSE105052	Patient:Control Friedreich's ataxia	42	3	miR-128-3p, miR-625-3p, miR-130b-5p, miR-151a-5p, miR-330-3p, miR-323a-3p, miR-142-3p	[20]
GSE151341	Patient:Control Early radiographic knee osteoarthritis biomarker	91	4	miR-335-3p, miR-199a-5p, miR-671-3p, miR-1260b, miR-191-3p, miR-335-5p, miR-543	[21]

* miRNA associated with Haemolysis Metric signature

DraculR provides a visual representation of the results from our selected public data sets in the form of a histogram (Figure 2). In this histogram the difference between the geometric means can be seen against a background of data with a known haemolysis quantification based on the ΔCq (miR-23a-miR-451) method. Samples are coloured to indicate either “Clear” (light blue) or “Caution” (scarlet) with background data coloured to indicate the ΔCq (miR-23a-miR-451) result of either “Clear (ΔCq)” (grey with a light blue highlight) for samples with ΔCq < 7 or “Haemolysed (ΔCq)” (grey with a scarlet highlight) for samples with ΔCq ≥ 7.

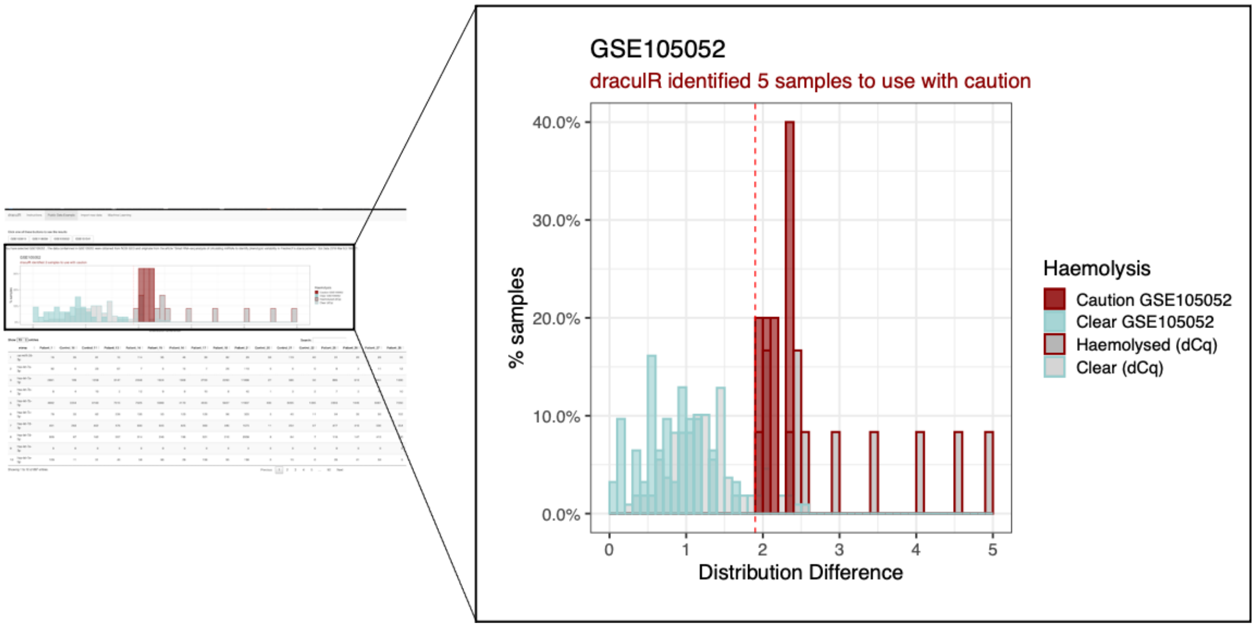
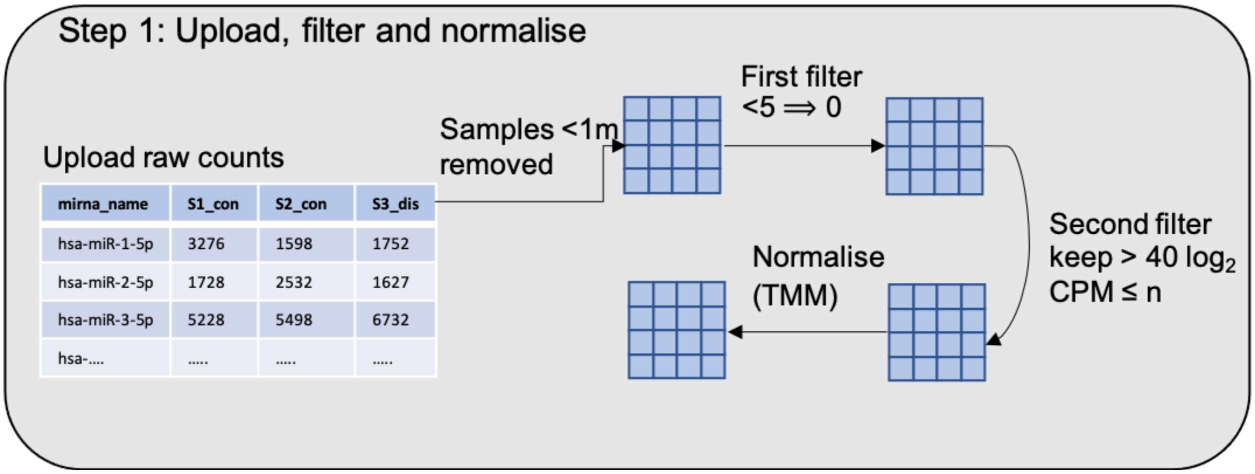


Figure 2. DraculR uses public data to illustrate the potential for unidentified haemolysis with potential to confound biomarker analysis. Here, data pertaining to GSE105052 was retrieved from NCBI Geo for analysis. The screen shot above shows an example where five samples were identified to be used with caution. All data are presented against a background of haemolysed samples assessed using the ΔCq method.

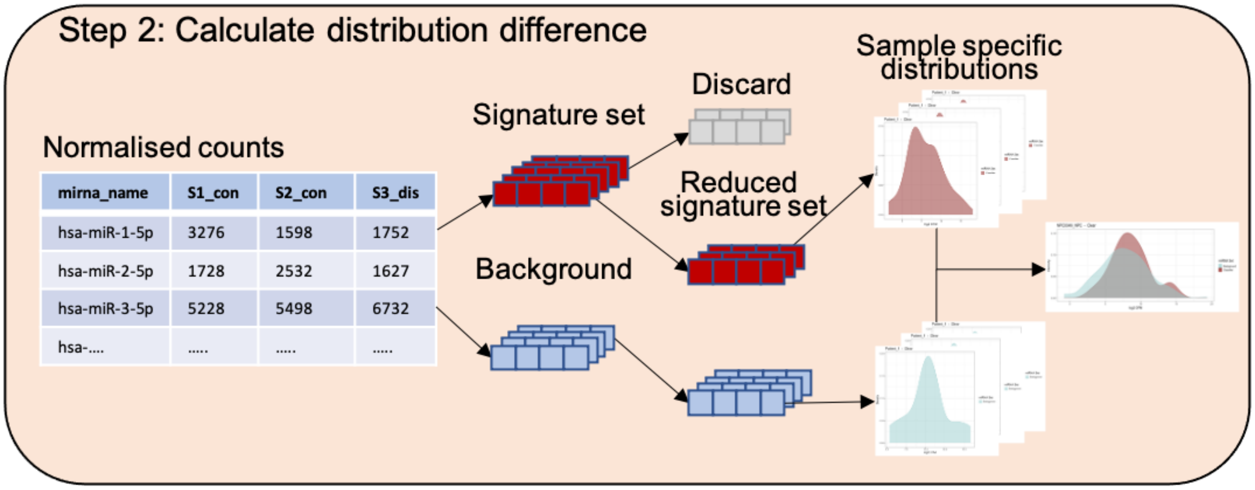
Inputting data into DraculR

DraculR allows users to upload a raw, high throughput sequencing counts table for analysis. Whilst normalisation is locked to the Trimmed Mean of M method (TMM) previously recommended in the *edgeR* workflow [23], the user controls features such as filtering for low expression (Figure 3) and refining the haemolysis signature set based on *a priori* knowledge of miRNAs that may be differentially expressed in the comparison of interest (Figure 4). The purpose of removing miRNAs with a known association to the research hypothesis is to help ensure any issues with haemolysis do not confound the research hypothesis. Note that samples with total miRNA read counts < 1 million are considered to be poorly sequenced and are removed for quality control.



mirna_name	S1_con	S2_con	S3_dis
hsa-miR-1-5p	3276	1598	1752
hsa-miR-2-5p	1728	2532	1627
hsa-miR-3-5p	5228	5498	6732
hsa-....	.....	.....	.....

Figure 3. Import a raw counts table generated by high throughput miRNA sequencing of human plasma libraries. These data will be filtered according to user specified requirements (n = number of samples in the smallest group of interest) and normalised using the Trimmed Mean of M (TMM) method as previously recommended in the *edgeR* workflow [23].

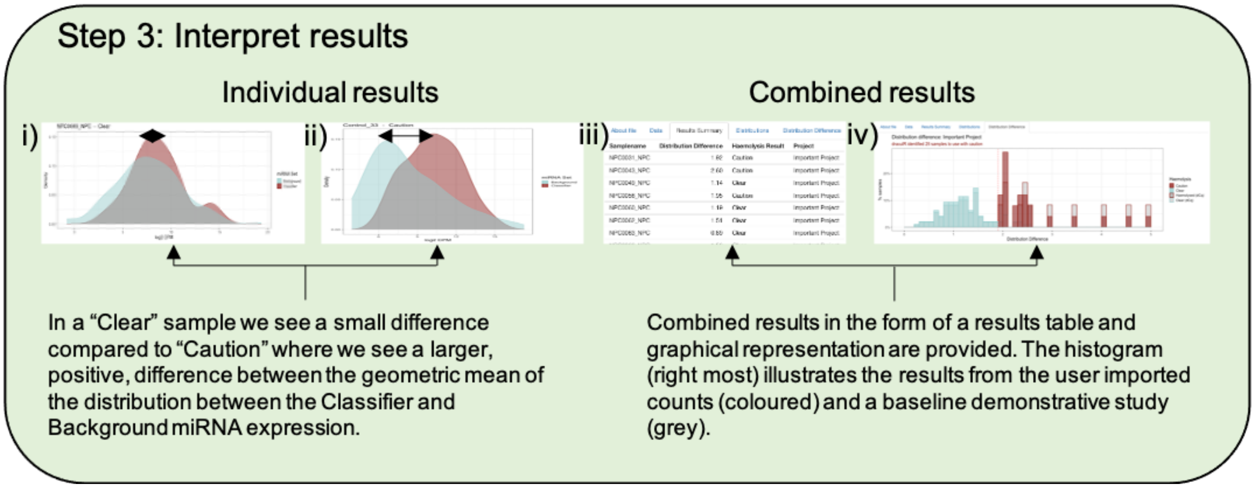


mirna_name	S1_con	S2_con	S3_dis
hsa-miR-1-5p	3276	1598	1752
hsa-miR-2-5p	1728	2532	1627
hsa-miR-3-5p	5228	5498	6732
hsa-....	.....	.....	.....

Figure 4. The distribution difference between the background and signature miRNA counts is calculated on an individual sample basis allowing the user to upload one to many samples as required. In the case of *a priori* knowledge of miRNA differentially abundant between a tested condition/control paradigm the user may choose to reduce the signature miRNA such that they do not include miRNA of interest (recommended).

Visualisation and interpretation

An essential feature of DraculR is that it allows users to visualise and assess the values obtained in the results, through sample specific and consolidated graphics including density plots, histograms and tables (Figure 5). These features help the user decide on the level of haemolysis that may be affecting their analysis by providing a new quality metric. Using this metric the user may choose to remove samples from downstream analyses. However, whether samples with a Haemolysis metric above the suggested threshold are removed or retained, the new information may be informative in the analysis of their miRNA sequencing data.


Step 3: Interpret results

Individual results

Combined results

In a "Clear" sample we see a small difference compared to "Caution" where we see a larger, positive, difference between the geometric mean of the distribution between the Classifier and Background miRNA expression.

Combined results in the form of a results table and graphical representation are provided. The histogram (right most) illustrates the results from the user imported counts (coloured) and a baseline demonstrative study (grey).

Figure 5. Graphical results in the form of a density plot of individual distributions (i, ii) and a histogram of combined distribution differences (iv) are provided along with a combined table of results (iii). The user is provided with both a metric describing the amount of haemolysis and a recommendation of caution if appropriate (iii).

Conclusion

We have developed a Shiny/R web based application that, for the first time, allows users to detect and address the issue of haemolysis in plasma miRNA sequencing data. This problem is common when, either through use of public data, exhaustion of sample, or exhaustion of funds, it is not possible to assess haemolysis using one of the current gold standard approaches (being delta quantification cycle (Cq) values for miR-23a-miR-451 or Spectrophotometry for haemoglobin estimation). The application is easy to use and applicable to high throughput sequencing data from human plasma. The method is robust for cases where a treatment:control style analysis is undertaken and the user has *a priori* knowledge of miRNA that are anticipated to be differentially abundant between groups. Whilst a probabilistic quantification of contamination risk is not possible based on the dataset used here, we plan future work drawing on the methods used by the Shah *et al.* 2016 that will include serial dilution and miRNA quantification of haemolysed plasma samples to validate and further refine our method. DraculR adds value to the growing resource of public data shared by plasma researchers by enabling *in silico* analysis of haemolysis confounding post sequencing. The detection of haemolysis using our Haemolysis metric enables the user to identify and potentially discard low quality samples which are otherwise unknown to be affected by haemolysis. This enables an additional quality metric and the subsequent increased confidence in the use of high throughput miRNA sequencing data for which no haemolysis information is available.

Funding

This research is supported by NIH NICHD R01 HD089685-01 Maternal molecular profiles reflect placental function and development across gestation PI Roberts, an Australian Government Research Training Program (RTP) Scholarship awarded to MDS, a National Health and Medical Research Council Investigator Grant (GNT1174971) awarded to CTR and a Matthew Flinders Professorial Fellowship awarded to CTR and funded by Flinders University. JB is supported by the James & Diana Ramsay Foundation.

Ethics

Ethics approval for the collection of blood from women undergoing elective pregnancy termination between 6–23 weeks' gestation was provided under HREC/16/TQEH/33, by The Queen Elizabeth Hospital Human Research Ethics Committee (TQEH/LMH/MH). Blood from women forming the general population group was collected after informed consent with ethics approval provided under HREC/H/021/2005, by The University of Adelaide Human Research Ethics Committee.

References

1. Pritchard CC, Cheng HH, Tewari M. MicroRNA profiling: Approaches and considerations. Nature Reviews Genetics. NIH Public Access; 2012. pp. 358–369. doi:10.1038/nrg3198

2. Guo R, Fan G, Zhang J, Wu C, Du Y, Ye H, *et al.* A 9-microRNA Signature in Serum Serves as a Noninvasive Biomarker in Early Diagnosis of Alzheimer's Disease. J Alzheimers Dis. 2017;60: 1365–1377.

3. Zheng K, Chen D-S, Wu Y-Q, Xu X-J, Zhang H, Chen C-F, *et al.* MicroRNA expression profile in RAW264.7 cells in response to Brucella melitensis infection. Int J Biol Sci. 2012;8: 1013–1022.

4. Furci L, Schena E, Miotto P, Cirillo DM. Alteration of human macrophages microRNA expression profile upon infection with Mycobacterium tuberculosis. Int J Mycobacteriol. 2013;2: 128–134.

5. Sánchez-Mora C, Soler Artigas M, Garcia-Martínez I, Pagerols M, Rovira P, Richarte V, *et al.* Epigenetic signature for attention-deficit/hyperactivity disorder: identification of miR-26b-5p, miR-185-5p, and miR-191-5p as potential biomarkers in peripheral blood mononuclear cells. Neuropsychopharmacology. 2019;44: 890–897.

6. Jin X, Chen Y, Chen H, Fei S, Chen D, Cai X, *et al.* Evaluation of Tumor-Derived Exosomal miRNA as Potential Diagnostic Biomarkers for Early-Stage Non-Small Cell Lung Cancer Using Next-Generation Sequencing. Clin Cancer Res. 2017;23: 5311–5319.

7. Coenen-Stass AML, Magen I, Brooks T, Ben-Dov IZ, Greensmith L, Hornstein E, *et al.* Evaluation of methodologies for microRNA biomarker detection by next generation sequencing. RNA Biol. 2018;15: 1133–1145.

8. Williams Z, Ben-Dov IZ, Elias R, Mihailovic A, Brown M, Rosenwaks Z, *et al.* Comprehensive profiling of circulating microRNA via small RNA sequencing of cDNA libraries reveals biomarker potential and limitations. Proceedings of the National Academy of Sciences. 2013. doi:10.1073/pnas.1214046110

9. Keller A, Leidinger P, Steinmeyer F, Stähler C, Franke A, Hemmrich-Stanisak G, *et al.* Comprehensive analysis of microRNA profiles in multiple sclerosis including next-generation sequencing. Mult Scler. 2014;20: 295–303.

10. Saini J, Bandyopadhyay B, Pandey AD, Ramachandran VG, Das S, Sood V, *et al.* High-Throughput RNA Sequencing Analysis of Plasma Samples Reveals Circulating

microRNA Signatures with Biomarker Potential in Dengue Disease Progression. mSystems. 2020;5. doi:10.1128/mSystems.00724-20

11. Kirschner MB, Kao SC, Edelman JJ, Armstrong NJ, Vallely MP, van Zandwijk N, *et al.* Haemolysis during Sample Preparation Alters microRNA Content of Plasma. Pfeffer S, editor. PLoS One. 2011;6: e24145.

12. Pritchard CC, Kroh E, Wood B, Arroyo JD, Dougherty KJ, Miyaji MM, *et al.* Blood cell origin of circulating microRNAs: a cautionary note for cancer biomarker studies. Cancer Prev Res . 2012;5: 492–497.

13. Blondal T, Jensby Nielsen S, Baker A, Andreasen D, Mouritzen P, Wrang Teilum M, *et al.* Assessing sample and miRNA profile quality in serum and plasma or other biofluids. Methods. Academic Press; 2013. pp. 164–169. doi:10.1016/j.ymeth.2012.09.015

14. Kirschner MB, Edelman JJB, Kao SCH, Vallely MP, Van Zandwijk N, Reid G. The impact of hemolysis on cell-free microRNA biomarkers. Front Genet. 2013;4: 94.

15. Sun L, Yu Y, Niu B, Wang D. Red Blood Cells as Potential Repositories of MicroRNAs in the Circulatory System. Front Genet. 2020;11: 442.

16. Cheng HH, Yi HS, Kim Y, Kroh EM, Chien JW, Eaton KD, *et al.* Plasma Processing Conditions Substantially Influence Circulating microRNA Biomarker Levels. Kiechl S, editor. PLoS One. 2013;8: e64795.

17. Livak KJ, Schmittgen TD. Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the $2^{-\Delta\Delta CT}$ Method. Methods. 2001. pp. 402–408. doi:10.1006/meth.2001.1262

18. Wong C-H, Song C, Heng K-S, Kee IHC, Tien S-L, Kumarasinghe P, *et al.* Plasma free hemoglobin: a novel diagnostic test for assessment of the depth of burn injury. Plast Reconstr Surg. 2006;117: 1206–1213.

19. Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res. 2002;30: 207–210.

20. Seco-Cervera M, González-Rodríguez D, Ibáñez-Cabellos JS, Peiró-Chova L, Pallardó FV, García-Giménez JL. Small RNA-seq analysis of circulating miRNAs to identify phenotypic variability in Friedreich's ataxia patients. Sci Data. 2018;5: 180021.

21. Ali SA, Gandhi R, Potla P, Keshavarzi S, Espin-Garcia O, Shestopaloff K, *et al.* Sequencing identifies a distinct signature of circulating microRNAs in early radiographic knee osteoarthritis. Osteoarthritis Cartilage. 2020;28: 1471–1481.

22. Giglio S, De Nunzio C, Cirombella R, Stoppacciaro A, Faruq O, Volinia S, *et al.* A preliminary study of micro-RNAs as minimally invasive biomarkers for the diagnosis of prostate cancer patients. J Exp Clin Cancer Res. 2021;40: 79.

23. Robinson MD, McCarthy DJ, Smyth GK. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2009;26: 139–140.

Chapter 6

Sex-specific analysis of placental gene expression in late-onset preeclampsia

Statement of Authorship

Title of Paper	Sex-specific analysis of placental gene expression in late-onset preeclampsia	
Publication Status	☐ Published ☐ Submitted for Publication	☐ Accepted for Publication ☒ Unpublished and Unsubmitted work written in manuscript style
Publication Details	Melanie D. Smith, James Breen, K. Justinian Bogias, Qianhui Wan, Tanja Jankovic-Karasoulos, Dale McAninch, Dylan McCullough, Tina Bianco-Miotto, Anya L. Arthurs, Katherine A. Pillman, Claire T. Roberts: Sex-specific analysis of placental gene expression in late-onset preeclampsia	

Principal Author

Name of Principal Author (Candidate)	Melanie Smith		
Contribution to the Paper	- All bioinformatics analyses - Drafting of manuscript - Producing figures / visualisation of data - Corresponding author on manuscript - Management of deposition of transcriptome data into public database (GEO)		
Overall percentage (%)	80%		
Certification:	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.		
Signature		Date	13/09/2021

Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i. the candidate's stated contribution to the publication is accurate (as detailed above);

ii. permission is granted for the candidate in include the publication in the thesis; and

iii. the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

Name of Co-Author	James Breen		
Contribution to the Paper	- Conceived and developed experimental plans - WGCNA Workflow - Editing of manuscript drafts		
Signature		Date	13/09/2021

Name of Co-Author	K.Justinian Bogias		
Contribution to the Paper	- Run test scripts - Assisted with manuscript drafting		
Signature		Date	13/09/2021

Please cut and paste additional co-author panels here as required.

Name of Co-Author	Qianhui Wan		
Contribution to the Paper	- Assisted with manuscript drafting		
Signature		Date	13/09/2021

Name of Co-Author	Tanja Jankovic-Karasoulos		
Contribution to the Paper	- Performed experiments		
Signature		Date	13/9/2021

Name of Co-Author	Dale McAninch		
Contribution to the Paper	- Performed experiments - Assisted with experimental design		
Signature		Date	13/09/2021

Name of Co-Author	Dylan McCullough		
Contribution to the Paper	- Performed experiments		
Signature		Date	13/09/2021

Name of Co-Author	Tina Bianco-Miotto		
Contribution to the Paper	- Conceived and developed experimental plans - Editing of manuscript drafts		
Signature		Date	13/09/2021

Name of Co-Author	Anya L. Arthurs		
Contribution to the Paper	- Assisted with drafting		
Signature		Date	13/09/2021

Name of Co-Author	Katherine Pillman		
Contribution to the Paper	- Expert miRNA analyses support - Development of experimental plans - Editing of manuscript drafts		
Signature		Date	13/9/2021

Name of Co-Author	Claire T Roberts		
Contribution to the Paper	- Conceived and developed experimental plans - Editing of manuscript drafts		
Signature		Date	13/09/2021

Sex-specific analysis of placental gene expression in late-onset preeclampsia

Melanie D. Smith[1,2,3], James Breen[1,2,4], K. Justinian Bogias[1,2], Tanja Jankovic-Karasoulos[1,2,3], Dale McAninch[2], Dylan McCullough[3], Tina Bianco-Miotto[1,5], Anya L. Arthurs[3], Katherine A. Pillman[6], Claire T. Roberts[1,2,3]

[1] Robinson Research Institute, University of Adelaide, Adelaide SA 5005, Australia
[2] Adelaide Medical School, University of Adelaide, Adelaide SA 5005, Australia
[3] Flinders Health and Medical Research Institute, Flinders University, Bedford Park SA
[4] South Australian Genomics Centre, South Australian Health & Medical Research Institute, Adelaide SA 5000, Australia
[5] School of Agriculture Food and Wine, Waite Research Institute, University of Adelaide, Adelaide SA 5005, Australia
[6] Centre for Cancer Biology, University of South Australia/SA Pathology, Adelaide SA 5000, Australia

Note: Supplementary Figures and Tables are hosted on Figshare

(https://figshare.com/projects/Chapter_6_-_Supplementary_Files/122459)

Abstract

Background

Preeclampsia (PE), a hypertensive disorder of pregnancy, remains the leading cause of maternal and perinatal mortality and morbidity worldwide. Preeclampsia occurs more often in male bearing pregnancy, and at the extremes of maternal age. However, this may be population dependent and can be divided into two distinct phenotypes; early (<34 weeks gestation) and late-onset (>34 weeks gestation) PE. Here we look specifically at late-onset PE using placenta from male bearing pregnancies as a focus to determine specific gene expression networks associated with disease.

Methods

We performed Total RNA sequencing on human placental chorionic villus tissue from uncomplicated and complicated (late-onset preeclampsia) pregnancies at term. Stratifying by fetal sex, we assessed differences in the placental transcriptome (post parturition) between uncomplicated and PE pregnancies. Finally, we used Weighted Gene Co-expression Network Analysis (WGCNA) to compare sex-naive molecular signatures in the context of preeclampsia.

Results

Compared to uncomplicated male-bearing placenta, male PE placentas show transcriptional changes suggesting an increased level of oxidative stress including an upregulation of ceruloplasmin (*CP*) and *DZIP1*. We also show up-regulation of

TMEM150C, associated with the baroreceptor reflex mediating sensing of blood pressure changes and down regulation of *NR3C2*, associated with early-onset hypertension exacerbated in pregnancy. Further, we identified a module gene set associated with anaerobic metabolism of glucose in PE.

Conclusion

Our results highlight the importance of sex-specific analyses in placenta and provide 25 candidate genes for further investigation of PE pathophysiology in male bearing pregnancies. A sex naive analysis using a network co-expression approach identified a module of genes statistically correlated with PE with enrichment for genes in the anaerobic metabolism of glucose.

Introduction

Preeclampsia (PE) is a hypertensive disorder of pregnancy affecting 3-8% of pregnancies worldwide [1–3]. Characterised by *de novo* maternal hypertension with multiorgan involvement detected after the 20th week of pregnancy, it remains one of the most common complications in pregnant women [3,4]. The immediate health risk associated with PE is well known, with the disorder responsible for maternal, fetal and neonatal morbidity and mortality worldwide [5]. Less well known are the long term health effects of PE on both the mother and the child. Both are at higher risk of developing cardiovascular disease and metabolic syndromes, including type II diabetes, later in life [6–12]. Far from being purely physiological, there is also a neurological axis to long term health. For the infant, there are potential long term effects of PE on brain development and function [13] that are associated with autism [14,15], intellectual disability [16], epilepsy [17] and schizophrenia [18,19]. For the mother, neurological effects can include a predisposition to psychosis [20], anxiety [21–23] and depression [24,25].

Preeclampsia is a syndrome composed of a cluster of symptoms and is broadly classified into two major clinical subtypes: early- and late-onset, with a gestational age cutoff of 34 weeks [26–29]. Early-onset PE (EOPE) is characterised by shallow placentation and impaired maternal spiral arteriole remodelling. Impaired remodelling alters maternal blood flow into the placenta and hence causes damage to the chorionic villi and their covering of syncytiotrophoblast [30]. Resultant debris or cellular particles dislodged from the placenta, into the maternal circulation, are suggested to, at least in part, trigger the maternal inflammatory response [31]. By contrast, late-onset PE (LOPE) is considered to be a maternal disorder, resulting from an underlying maternal physiological predisposition and is more likely to be associated with a morphologically normal placenta, larger placental volume and normal fetal growth [27]. Although the presenting features overlap, there is no gold standard for the clinical differentiation between preeclampsia subtypes with current clinical thresholds leaving important uncertainties in precision diagnosis and hence in our understanding of the underlying molecular pathogenesis [32].

Previously, many experimental attempts to identify distinct molecular signatures in PE using microarray technology have failed to find reproducible results, due to small sample sizes and a failure to distinguish between early- and late-onset disease [33–40]. Some do not classify EOPE and LOPE [33,34,36,38], others pool classified samples [37], focus on only EOPE [40] or compare between EOPE and LOPE, but not controls [39], with none including fetal sex. Others, operating under the hypothesis that preeclampsia is a syndrome driven by the deregulation of different molecular pathways, have used aggregated microarray datasets and unsupervised clustering methods to detect multiple subtypes of PE with no significant effect of fetal sex on their analysis [41]. Importantly, the lack of any fetal sex signal may reside in the use of targeted microarray, and highlights the advantages of RNA sequencing (RNA-seq).

RNA-seq, unlike targeted microarrays, allows for profiling of the whole transcriptome (i.e. Total RNA-seq), including coding and non-coding transcripts, and provides a much more detailed view of the placental transcriptional landscape [42]. More recently, studies like that by Ren *et al.* (2021) [43] analysed RNA-seq data from EOPE and LOPE and have identified distinct molecular processes at work separating the two conditions. However, while this study takes into account clinical subtype, it does not consider fetal sex, a factor we consider to be important in the etiology of disease. Inconsistent methods such as failure to distinguish EOPE and LOPE, pooling of samples, and the loss of power induced by these and other issues such as small sample size and microarray limitations have meant our understanding of the origins and maternal–placental contributions to the pathology of PE remain unclear. Traditional methods have established a number of mechanisms underlying PE. These include angiogenesis [35–37], immune function [26,33,34,43,44], oxygen sensing [14,40] and metabolism [43] it is not clear whether these molecular pathways are common to different clinical subtypes.

Late-onset PE, thought to originate through interactions between placental senescence and a maternal genetic predisposition to cardiovascular and metabolic disease [32] is generally considered a maternal disorder. However, resolution still requires delivery of the placenta indicating that the placenta is playing a role in pathology of the disease. Note that rarely PE is first diagnosed during labour or within days after delivery [45]. In the present study, we used high throughput Total RNA-seq to establish a sex-specific comparison of the transcriptional landscape of placenta with and without LOPE. The resultant transcription profiles were first assessed for differential expression and then, using a systems biology approach, we performed a weighted gene correlation analysis (WGCNA) to establish networks of co-expressed genes associated with PE, as well as maternal and neonatal characteristics. Both sex-specific and sex-naive approaches highlight the importance of anaerobic metabolism of glucose, oxidative stress and immune response genes in the context of LOPE. In addition, the sex-specific differential expression analysis identified a number of genes with potential to be causal in the etiology of LOPE which further suggests the importance of clinical subtyping in the search for interventional pathways.

Results

Maternal patient information

Data and placentas from the prospective cohort **Sc**reening f**O**r **P**regnancy **E**ndpoints (SCOPE) study (recruitment between September 2004 and December 2008 in Australia) and the **S**creening **T**ests to Predict Poor **O**utcomes of **P**regnancy (STOP) study (recruitment between March 2015 and December 2017) were used. The hospital from which the SCOPE and STOP participants were recruited serves a socioeconomically disadvantaged population. In both cohorts**,** Nulliparous ostensibly low risk women with singleton pregnancies were recruited. Maternal characteristics are shown in Table 1. In an attempt to simplify the language used herein, I will refer to the sub-division of placenta counts data used in these analyses as: Control male, PE male, Control female and PE female where "Control" refers to women who had uncomplicated pregnancies and "PE" refers to women who had a pregnancy complicated by preeclampsia.

Table 1. Maternal characteristics

	Normotensive		Preeclamptic	
	Male infant (n = 18)	Female infant (n = 33)	Male infant (n = 9)	Female infant (n = 14)
Maternal age	24.5 (20.2-26)	24 (22-28)	22 (20-23)	25.5 (20.2-30.5)
First trimester BMI	25.2 (23.6-26.8)	24.6 (22.1-25.5)	29.2 (21.9-36.2)	29.5 (28-31.2)
Gestational Age (weeks)	40.4 (39.5-41)	40.4 (39.9-41.1)	39.7 (39-40)	38.4 (37.40-40.5)
Birth weight (g)	3575 (3316-3869)	3550 (3240-3825)	3510 (3075-3580)	2934 (2657-3261)

BMI, Body mass index
DE = 58 samples
WGCNA = 80 samples (includes 7 repeat measures)

Sex-specific gene-level differential expression analyses

To investigate gene expression in human placentas from uncomplicated and preeclamptic pregnancies, we performed double-stranded Total RNA-seq on 61 (23 male, 38 female) samples with an average of ~40.2 million paired end reads per sample. Of the initial dataset, 2 samples were removed due to lack of sequencing coverage, and 1 was excluded as the pregnancy was complicated by EOPE, leaving a total of 58 (21 male, 37 female) samples for downstream analyses. Read pairs were aligned to human genome GRCh38 [46] using STAR [47] and RNA-seq data was summarised to the gene level using featureCounts [48] from which we detected 19,493 genes with an official gene symbol. After filtering for genes with low expression (<2 counts per million in <5 samples), 12,911 genes remained for downstream analyses.

Sex differences in the risk of a pregnancy being complicated by PE were identified more than 50 years ago [49], although there have been others that report no association between fetal sex and maternal risk for preeclampsia [50,51]. While there is evidence to suggest a different etiology of disease in the male versus the female bearing pregnancy [52], a recent review and meta-analysis determined that a male bias was not present in all populations [53]. To investigate the sex-specific differences in placental gene expression, in the context of preeclampsia, we used sex and outcome as terms in the regression model whilst controlling for maternal age and maternal BMI. Genes on the Y chromosome were removed as these cannot be tested in females and genes on the X chromosome were retained in the dataset to ensure important X chromosome genes were included in the differential expression (DE) analyses. Often, both Y and X chromosome genes are removed to reduce confounding by sex differences in gene expression however, given the importance of X chromosome genes in placental development, X chromosome genes were retained.

DE for each gene was performed using linear models comparing between LOPE and uncomplicated, adjusting for maternal age and BMI. Custom contrasts were constructed to obtain the DE estimates between LOPE and uncomplicated males and females. Using 12,911 genes after filtering, we identified 25 genes that were differentially expressed (FDR <0.05) in the analysis of the male bearing samples (Table 2, Figure 1a, Supplementary Figure S1), but no significantly differentially expressed genes (FDR <0.05) in the female comparison (Figure 1b). Ceruloplasmin (*CP*), a copper binding glycoprotein, was the most highly up-regulated gene in male PE compared to male Control. All differentially expressed genes were annotated as protein coding, with 2 up-regulated (*ZNF300*, *DZIP1*) and 3 down-regulated genes (*ZNF79*, *TRIM29*, *NR3C2*) being transcription factors (TF).

Table 2: Differential expression analysis identified 21 up- and 4 down-regulated genes (FDR <0.05) in placentas from male bearing PE pregnancies. Genes are ordered in the table by $\log_2$ fold change (logFC).

Up-regulated (Higher in PE)

Gene Name	Ensembl ID	Gene Biotype	logFC	AveExpr	FDR
CP	ENSG00000047457	Protein Coding	2.335	3.58	0.014
ZNF300*	ENSG00000145908	Protein Coding	1.682	0.44	0.026
GSTM3	ENSG00000134202	Protein Coding	1.613	3.30	0.036
SGK2	ENSG00000101049	Protein Coding	1.608	1.05	0.043
COL6A3	ENSG00000163359	Protein Coding	1.424	7.44	0.014
SLC26A6	ENSG00000225697	Protein Coding	1.369	3.27	0.046
IGFBP7	ENSG00000163453	Protein Coding	1.312	5.86	0.049
TMEM150C	ENSG00000249242	Protein Coding	1.309	3.28	0.017
SLC18A2	ENSG00000165646	Protein Coding	1.102	2.55	0.013
DZIP1*	ENSG00000134874	Protein Coding	1.087	3.12	0.009
CRMP1	ENSG00000072832	Protein Coding	1.081	2.27	0.046
ROBO1	ENSG00000169855	Protein Coding	1.014	4.21	0.039
TMEM144	ENSG00000164124	Protein Coding	1.100	2.74	0.017
CD44	ENSG00000026508	Protein Coding	0.986	6.71	0.043
WWOX	ENSG00000186153	Protein Coding	0.830	2.84	0.026
C1orf21	ENSG00000116667	Protein Coding	0.794	4.99	0.043
EHHADH	ENSG00000113790	Protein Coding	0.783	2.67	0.039
RERG	ENSG00000134533	Protein Coding	0.756	4.47	0.039
TSPAN12	ENSG00000106025	Protein Coding	0.750	4.07	0.039
HSD17B12	ENSG00000149084	Protein Coding	0.626	4.77	0.009
SLC2A13	ENSG00000151229	Protein Coding	0.600	4.02	0.461

Down-regulated (Lower in PE)

Gene Name	Ensembl ID	Gene Biotype	logFC	AveExpr	FDR
ZNF79*	ENSG00000196152	Protein Coding	-0.587	3.77	0.049
TRIM29*	ENSG00000137699	Protein Coding	-0.934	6.51	0.014
GPR158	ENSG00000151025	Protein Coding	-1.042	3.00	0.017
NR3C2*	ENSG00000151623	Protein Coding	-1.240	4.57	0.014

*Transcription factor

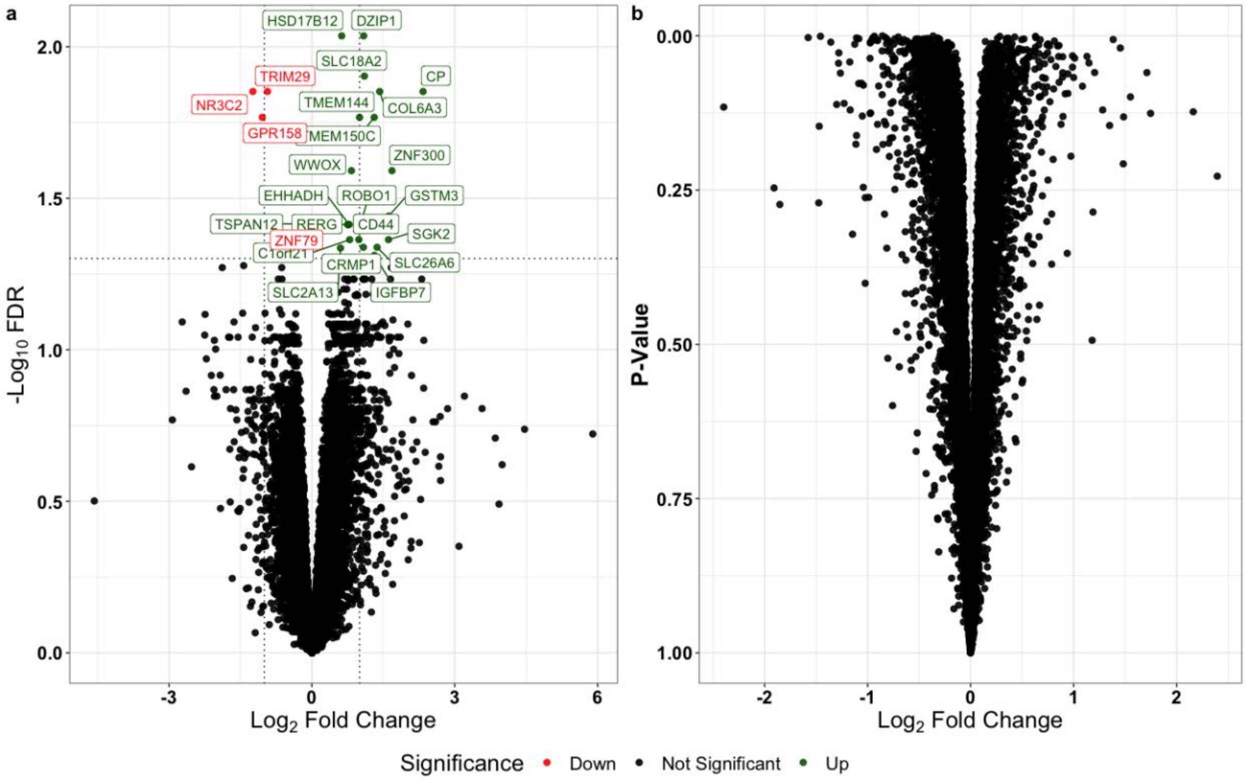


Figure 1: Differentially expressed genes (FDR<0.05; indicated as green (up-) and red (down-regulated in PE) in placenta from preeclamptic and uncomplicated, male- and female-bearing pregnancies. The volcano plots indicate the level of change ($\log_2$ fold change) in log transformed, normalised data for male (a) and female (b) comparisons. Note significance (y-axis) is given here as the adjusted p-value in (a) and the un-adjusted p-value in (b).

Given the small number of significantly DE genes, rather than perform a GO or KEGG functional enrichment test, we performed a gene set enrichment analysis (GSEA) using the full, ranked, DE table [56,57]. GSEA is used to test for enrichment between the ranked list of genes taken from the DE table and curated gene sets, groups of genes that share a common biological function, chromosomal location or regulation. Enrichment in a given gene set can be identified where the members of the DE table, rather than be randomly distributed, are primarily found at the top (up-regulated in a gene set) or the bottom (down-regulated in a gene set) of a curated set [56]. Rather than focus on individual genes, this method identified groups of genes that were enriched for biological pathways. We identified 3 up and 1 down-regulated Hallmark pathways (FDR q-value <0.05) with normalised enrichment score (NES) values >|2|. These were E2F Targets, G2M Checkpoint, Epithelial Mesenchymal Transition and TNFA Signalling Via NFKB down-regulated (Supplementary Figure S2). We also observed an enrichment within the Immunologic gene sets with 3 up gene sets associated with immune related pathways (FDR q-value <0.01) with normalised enrichment score (NES) values >|2|. Gene sets included predominantly those associated with a response to vaccination (Supplementary Figure S3) with an important parallel in placenta given the interaction between mother and fetus who are two biologically distinct organisms. These data suggest broadscale up regulation in differentiation, proliferation and immune response pathways in PE males compared to uncomplicated.

Sex-specific differential expression analysis failed to identify any DE genes in the female PE/Control analysis. Given that statistical power in the female analysis was insufficient to draw independent conclusions, we decided to perform a weighted gene co-expression network analysis using a systems biology approach with the combined male and female data.

Constructing a Co-Expression Network of the human placenta

Given the small sex differences identified, we decided to add sequence data from a previous investigation to our current dataset and perform a sex-naive weighted gene co-expression analysis (WGCNA) [58] using placenta from both male and female bearing pregnancies. We have chosen to combine the two datasets, rather than use the second dataset for validation in a secondary analysis, due to the small sample number, and because the second dataset contains repeat measures for a number of samples used in the initial differential expression analysis. WGCNA is a systems biology approach to analysing gene expression data that is useful for revealing broader biological changes between groups. Compared with DE gene analysis which reveals significant individual genes, WGCNA reveals how the overall gene expression patterns can be represented as groups of genes ("modules") with distinct biological significance. By examining how these modules of genes change in expression between LOPE and uncomplicated samples, we can start to see what biological processes or pathways may be altered in LOPE relative to uncomplicated pregnancies. Briefly, 34 additional samples (18 male, 16 female) were incorporated (Mayne *et al.* Unpublished), with 9 samples removed due to a lack of sequencing coverage, and 3 samples removed due to quality concerns, leaving a combined total of 80 (30 male and 50 female) samples available for downstream analysis. After removing mitochondrial (MT) and Y chromosome genes, filtering for genes with both low expression and low variability between samples (low expression classed as <1.5 counts per million in 23 samples; low variability genes considered as those with a standard deviation above the 40th quantile, ie in the lowest 60% of variability) and accounting for batch effects, 11,724 genes remained. WGCNA was then run on normalised, batch corrected gene expression values. The grey module was composed of unassigned genes from across the network. After module assignment, the grey module genes were removed from further consideration.

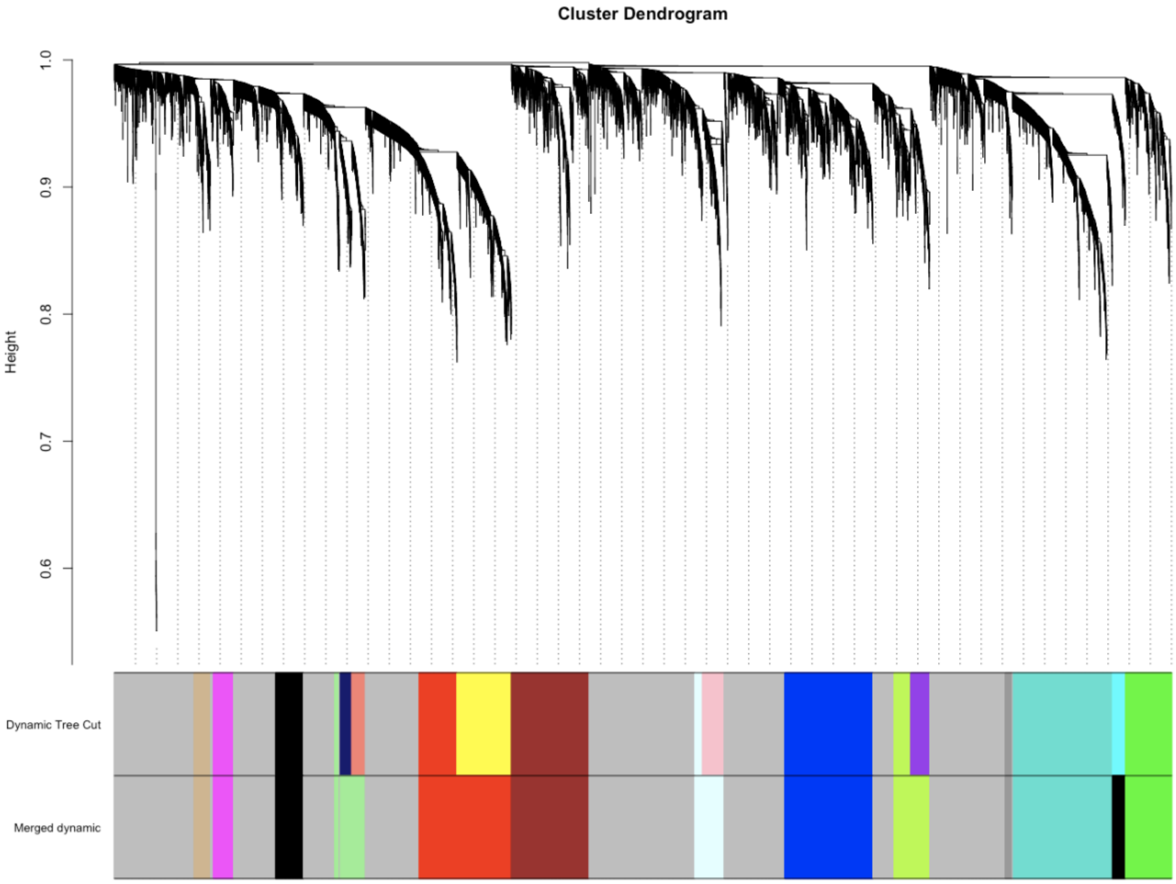


Figure 2: Hierarchical clustering dendrogram of genes from human placentas from uncomplicated and PE placentas reveals 14 distinct clusters of co-expressed genes (coloured segments in the bottom panel titled Merged Dynamic). A dynamic tree cutting method was used for initial hierarchical clustering module detection of the dendrogram (represented as the Dynamic Tree Cut colour band). Next, modules with an eigengene correlation above 0.75 were merged resulting in the final gene module assignments (represented as Merged dynamic colour band).

For each identified module, a summary eigengene was calculated, representing the weighted average of gene expression for a given module. For each gene, module membership was calculated as the absolute correlation between the gene's expression and the module eigengene. Genes with the highest correlation, represented as *kME*, reported in Table 3. The module eigengene was then correlated with biological or phenotypic traits of interest (summarised in Figure 3). Importantly for our context here, the tan module was significantly (p-value <0.05) correlated with pregnancy outcome. This indicates that genes from the tan module are up-regulated in LOPE. The black module we see is correlated with fetal sex. This indicates that genes from the black module are up-regulated in males compared to females. Both tan and black modules were negatively correlated with the associated trait. Both blue and red modules were correlated with fetal sex although neither were correlated with tan or black. The red module genes, negatively correlated, were down-regulated in females with the positively correlated blue module genes up-regulated in females. The turquoise module was negatively correlated with maternal age suggesting that genes in the module are down-regulated in placenta from pregnancies in older women. It is important to note there that between all module and eigengene correlations noted here, the correlation values associated with each trait is small. This indicates a limitation with this experiment which will require further investigation before more robust claims can be made.

Intramodular connectivity (kIM), a measure of connectivity of nodes to other nodes within the same module, was also calculated, here we consider the genes with highest kIM to be hub genes. kIM quantifies module membership (centrality) and identifies highly connected genes which may be key drivers of biology. Genes with the highest intramodular connectivity are reported in Table 3. The red and lightgreen modules have the highest intramodular connectivity (Supplementary Figure S4), with the red module being enriched for response to endoplasmic reticulum stress and glycosylation. The tan module was correlated with pregnancy outcome groups, the black, blue and red modules were correlated with fetal sex and the turquoise module correlated with maternal age (significant at p-value <0.05). Among the hub genes described in Table 3, *NECTIN4* [59], *CALM1* [60], *INHBA* [38,61], *LYN* [61], *NDRG1* [61,62], *CCSAP* [63], *LRRC1* [64] and

SLC2A1 [38], 9 of the ten top hub genes associated with the tan (PE) module have previously been associated with PE. This provides further confidence in the robust nature of the current study. Hub genes from the remaining modules include *RBM26, PLXND1* and *TMED4* which may be associated with fetal sex; *CHCHD2* and *PSMA3-AS1* which may be associated with maternal age and maternal smoking status; and *TOP2A, BRD4, NCKAP1L, NIPBL, EXTL3, CTBP1* for which there was no maternal or fetal characteristic association.

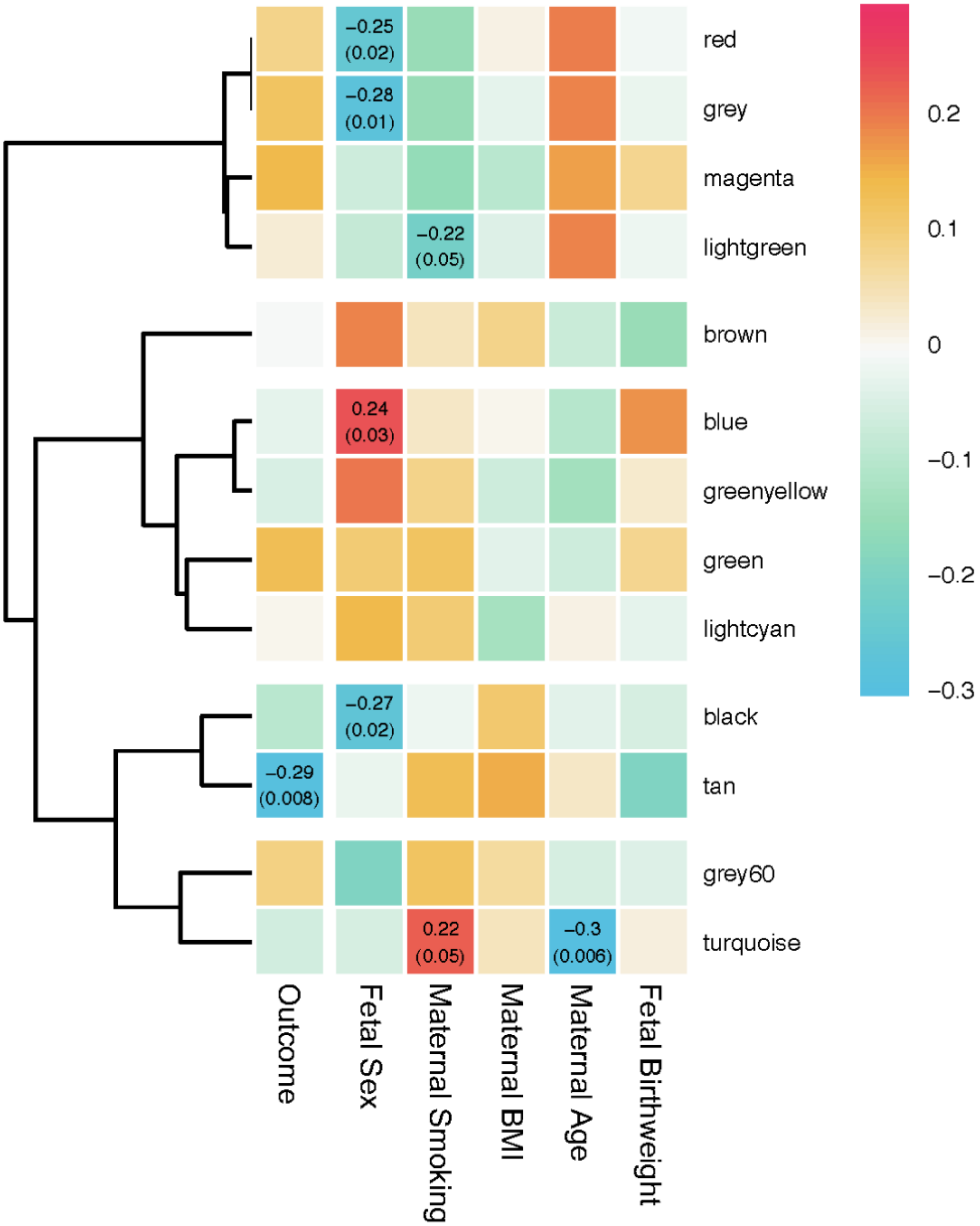


Figure 3: **Gene co-expression modules across PE and Control placentas from male and female bearing pregnancy at term.** Values shown in cells are hybrid Pearson-robust correlations between the overall gene expression in a module (summarised using the eigengene) and pregnancy outcome (PE, Uncomplicated), maternal characteristic (smoking) and fetal sex (male, female) coded as binary variables and maternal characteristics (BMI, age) and neonatal characteristic (birth weight) coded as continuous variables. Each cell with significant values reports the Pearson's correlation (p-values in brackets) which results from correlating the module eigengenes (rows) with the character traits (columns). Gene co-expression analysis identified 13 modules. The tan module is significantly (p-value <0.05) correlated with pregnancy outcome; black, blue, grey and red are correlated with fetal sex; turquoise was correlated with maternal age.

Table 3: Co-expression module characteristics.

Module	No. of genes	Variance explained by eigengene	Highest intramodular connectivity (*kIM*)	Top ten hub genes (*kME >0.8*)
black	460	0.44	*RBM26*	*TMF1*[a], *RBM26, TRPM7, PPP6R3, SCAPER, USP15, GOLGA4, ZNF567*[a], *ZBED5*[a], *TRIP11*
blue	984	0.40	*PLXND1*	*LIX1L, RASA3, RAB3IL1, DNAAF9, ARHGAP35*[a], *NFIX*[a], *CSF1, LTBP2, ADAMTS2, PLXND1*
brown	865	0.25	*TOP2A*	*TOP2A, NCAPG, RACGAP1, LMNB1, BUB1, CDK1, ASPM, DLGAP5, ATAD2, WDR76*
green	514	0.40	*BRD4*	*PRR12*[a], *KMT2D, KMT2B*[a], *BRD4, SETD1B, SKI*[a], *ANKRD11, KDM6B, ZMIZ1, ZFHX3*[a]
greenyellow	399	0.44	*NCKAP1L*	*DOCK2, ITGAM, SASH3, LAPTM5, NCKAP1L, PTPRC, SYK, CTSS, ITGB2, HCK*
grey	5300	0.13	*N/A*	*TBC1D1, NECTIN3, CDC42SE1, KAT5, SLC5A6, HSD3B1, C4orf19, MFSD2A, AZIN1, CYP19A*
grey60	76	0.54	*NIPBL*	*NIPBL, INO80D, PUM2, PRRC2C, SEC63, CDK12, TOP1, KMT2C, NF1, ASH1L*[a]
lightcyan	324	0.30	*EXTL3*	*PTK2B, EXTL3, SLC2A10, REPS2, APOL6, SOX4*[a], *FN1, SERPINE2, SH3PXD2B, SLC43A3*
lightgreen	331	0.60	*CHCHD2*	*CHCHD2, COX8A, PSMB3, PSMB4, COX6A1, ELOB, MRPL54, MRPS7, ATP5F1B, UBL5*
magenta	225	0.43	*CTBP1*[a]	*CTBP1*[a], *CTBP1-AS, FBXW5, HSF1*[a], *ASB6, MAP2K2, PKP3, SHARPIN, CPTP, TPRA1*
red	1030	0.56	*TMED4*	*CYP19A1, RNF11, TMED4, HSD3B1, ATP6V1E1, CHMP2B, SELENOT, TMEM87A, GHITM, TM9SF2*

Table 3 cont.: Co-expression module characteristics.

Module	No. of genes	Variance explained by eigengene	Highest intramodular connectivity (*kIM*)	Top ten hub genes (*kME >0.8*)
tan	180	0.47	*NECTIN4*	*NECTIN4, CALM1, INHBA, LYN, NDRG1, CCSAP, FAM120A, LRRC1, SYDE1, SLC2A1*
turquoise	1106	0.47	*PSMA3-AS1*	*PSMA3-AS1, NKTR, EOLA2-DT, LUC7L3, MATR3, FTX, N4BP2L2-IT2, MALAT1, L3MBTL1*[a], *NEAT1*

[a] Transcription factor

Using this gene co-expression approach, module membership as described by gene co-expression and connectivity data was used to create a network (Figure 4). Co-expression connectivity indicates the connection strength among the nodes in the network. Nodes (genes) with adjacency <0.25 and less than 4 edges to other nodes were removed leaving a network containing nodes primarily from red, turquoise, green, lightgreen, greenyellow, blue and tan modules. Within the network, the tan module appears to form two distinct subclusters, with one subcluster (bottom left tan subcluster) contained 9 of the 10 top hub genes except for *NDRG1.* The second subcluster (top right tan subcluster) contained *NDRG1* and also *FLT1* both of which have repeatedly been shown to be associated with PE [35,37,61,62,65].

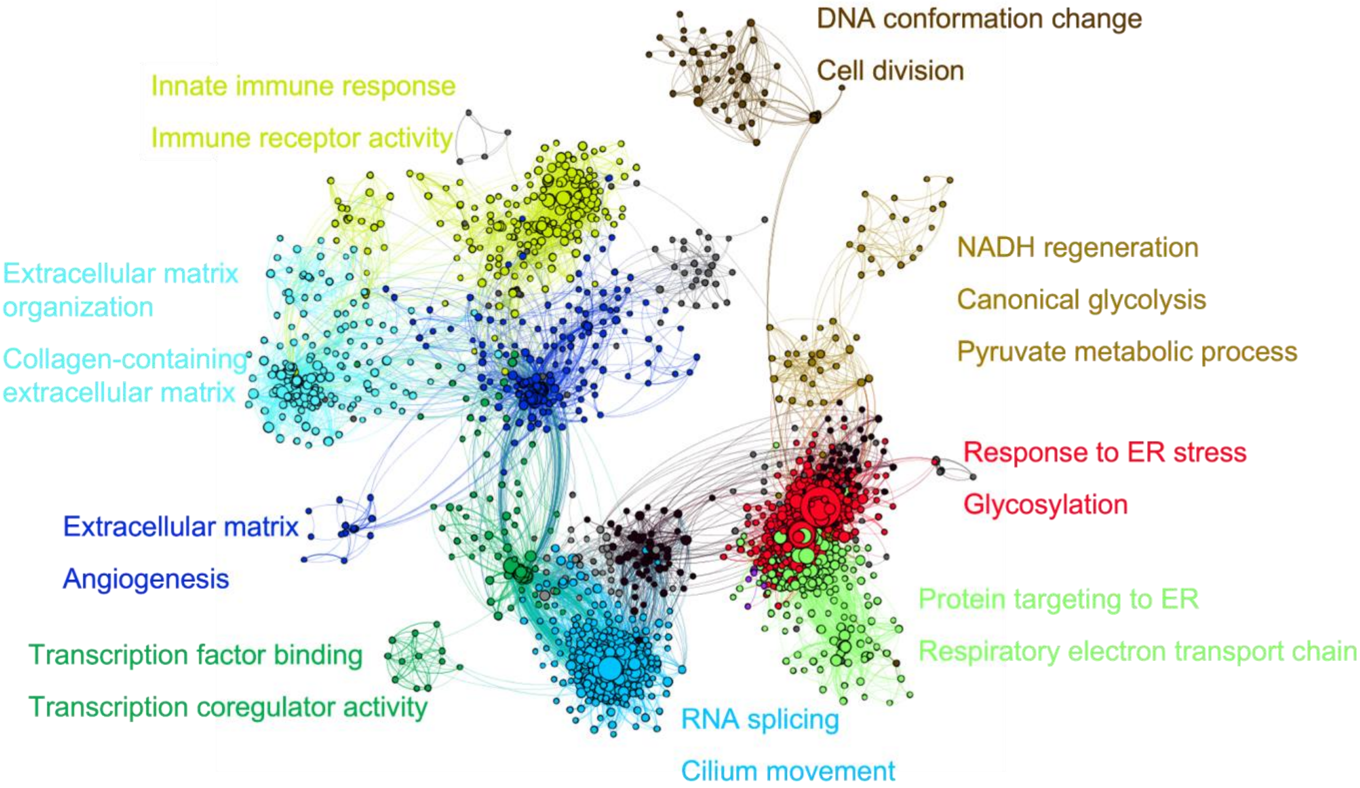


Figure 4. **Gene networks identified through co-expression analyses**. Each node (circle) in the network represents a single gene. Node size is proportional to the degree of connectivity (i.e. the number of connected nodes). Edges (lines connecting each circle) represent co-expression between two genes, with edge weight proportional to the strength of co-expression. The co-expression network is a signed adjacency matrix constructed from RNA-seq data from placenta taken from uncomplicated (Control, n=57) and complicated (PE, n=23) pregnancies, post delivery. Only nodes with at least four connections are shown. Here a "module" of genes represents a group of genes with similar expression patterns in term placenta. We identified 13 modules using a signed, hierarchical clustering and branch cutting method. The signed modules contain genes for which changes in expression are up- or down-regulated in the same direction for all genes in the module. The most significant GO term enrichment is annotated by module colour.

Genes identified as differentially expressed between PE and Control in male bearing placentas were located in the blue, brown and red modules (Figure 5). Notably, the WGCNA modules were created from both male and female bearing pregnancies and as such gene expression correlations may not completely overlap the DE fold change direction from the male comparison. This further highlights the regulatory differences between the placenta of male and female bearing pregnancies. Typically, in a control-treatment variable such as the outcome described here, we would expect to see genes in any one module, with a common fold change direction.

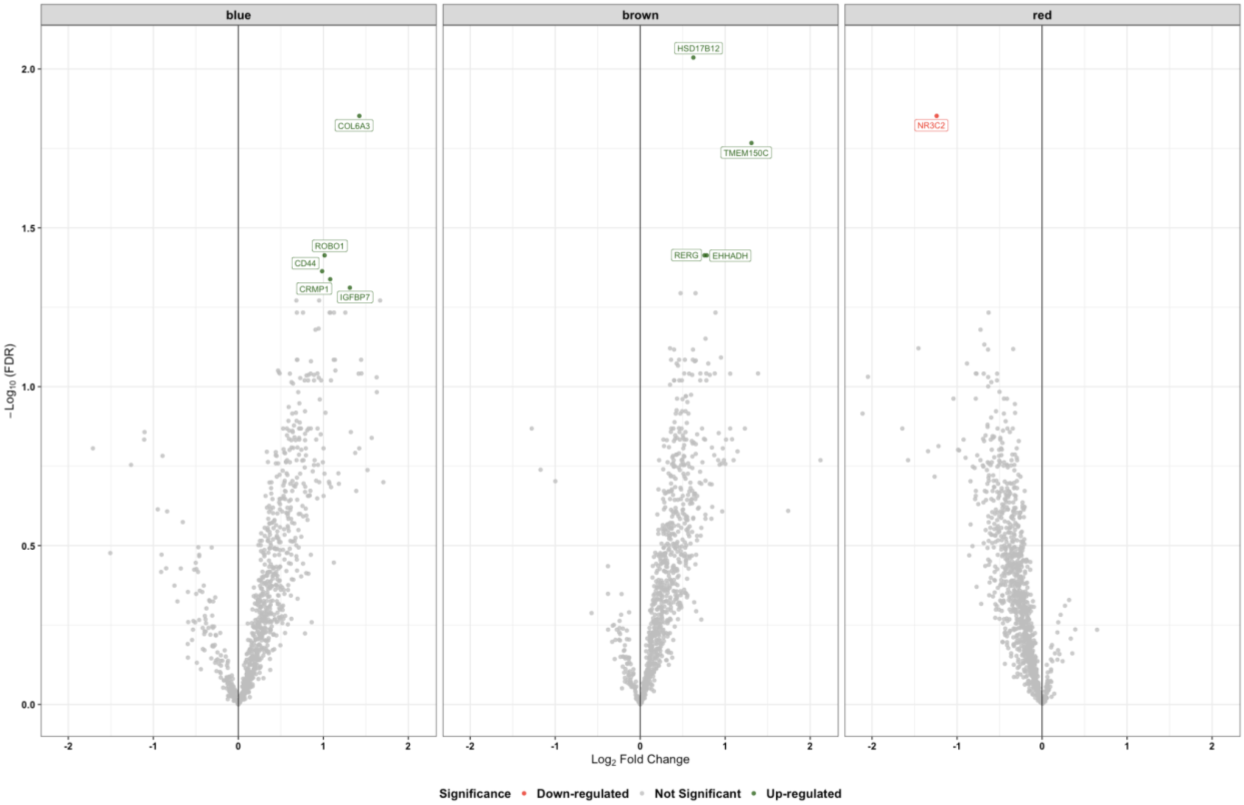
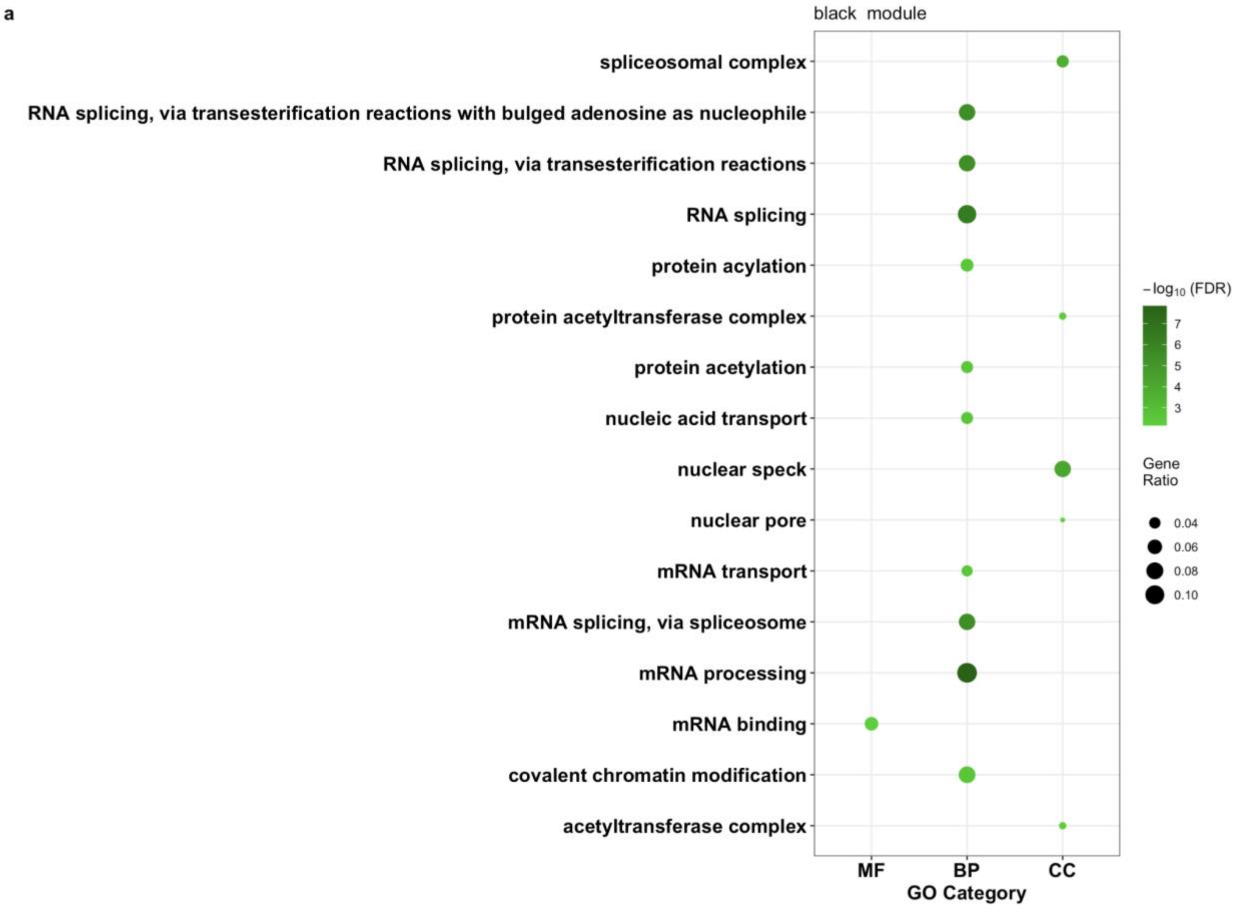


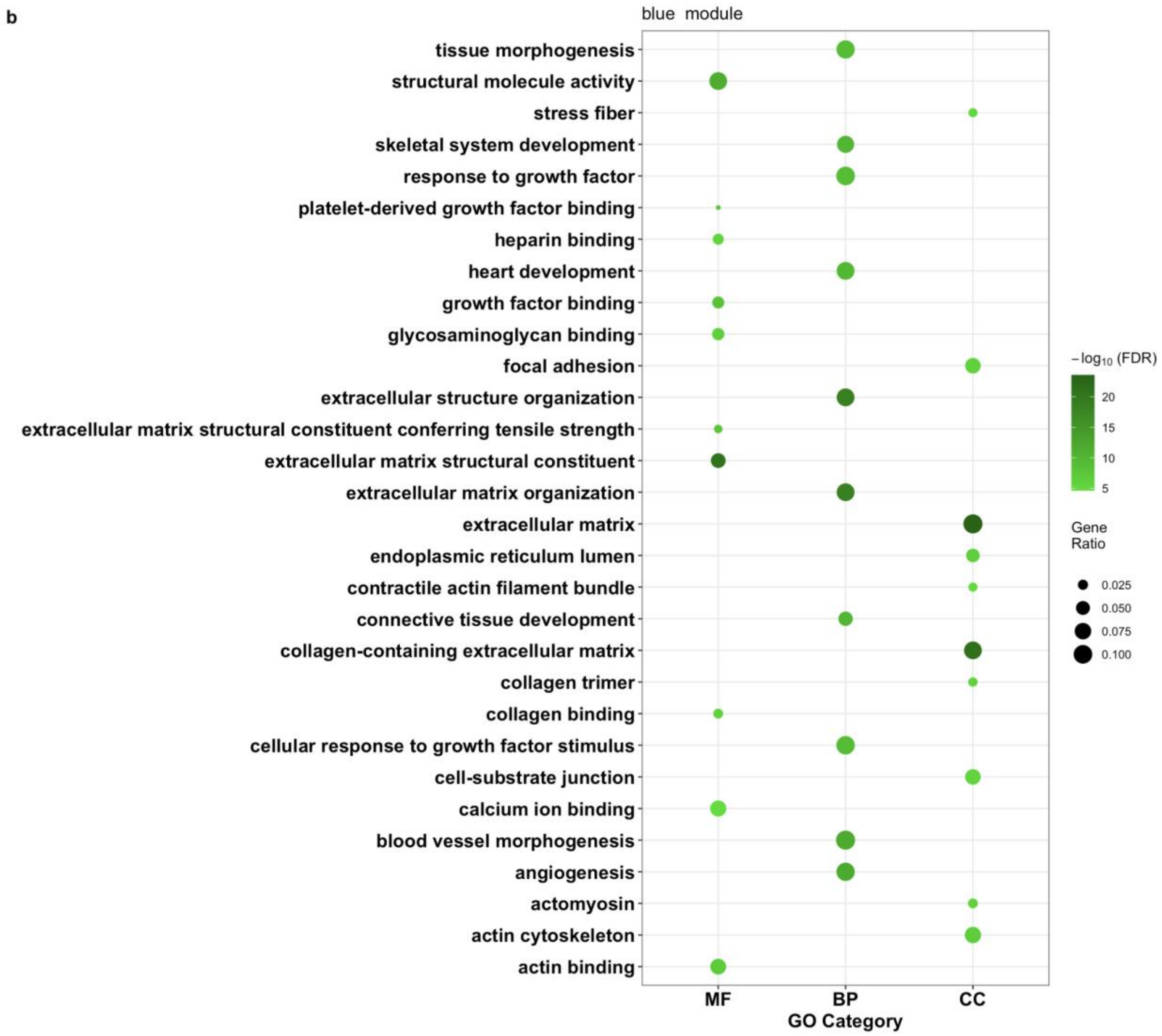
Figure 5: Genes identified as differentially expressed between PE and control placentas (male bearing). Genes up-regulated in the male comparison are found in the blue and brown modules, with down-regulated genes found in the red module. The sign of the fold change is not common to each module highlighting gene regulation differences between the male and female bearing placenta.

Gene co-expression analysis defines highly correlated PE module enriched for processes involved in the anaerobic metabolism of glucose

After identifying neatly defined modules of co-expressed genes, we then used gene ontology (GO) enrichment analysis to identify genes enriched for functional attributes (Figure 6). Tan module genes (correlated with PE) were enriched for processes involved in the anaerobic metabolism of glucose, known as glycolysis. These included polyol metabolic process, NADH regeneration, NAD metabolic process, canonical glycolysis, glucose catabolic process to pyruvate, glycolytic process through glucose-6-phosphate and pyruvate metabolic process. The anaerobic context is relevant here given the association between hypoxia/reperfusion and the pathogenesis of PE. Hallmark and KEGG analysis identified estrogen response early, glycolysis, hypoxia MTORC1 signalling and glycolysis gluconeogenesis pathways, respectively. Again the main association was between tan module genes and hypoxia and glycolytic processes. Interestingly, mTORC1 promotes cell growth primarily through the activation of key anabolic processes [66]. Estrogen response early has been associated with tumor metastasis [67] and its expression is mediated by 17β-estradiol in breast cancer [68]. Importantly in the context of PE, 17β-estradiol or agonists of estrogen receptors (ER), can exert acute vasodilatory actions in placenta [69].

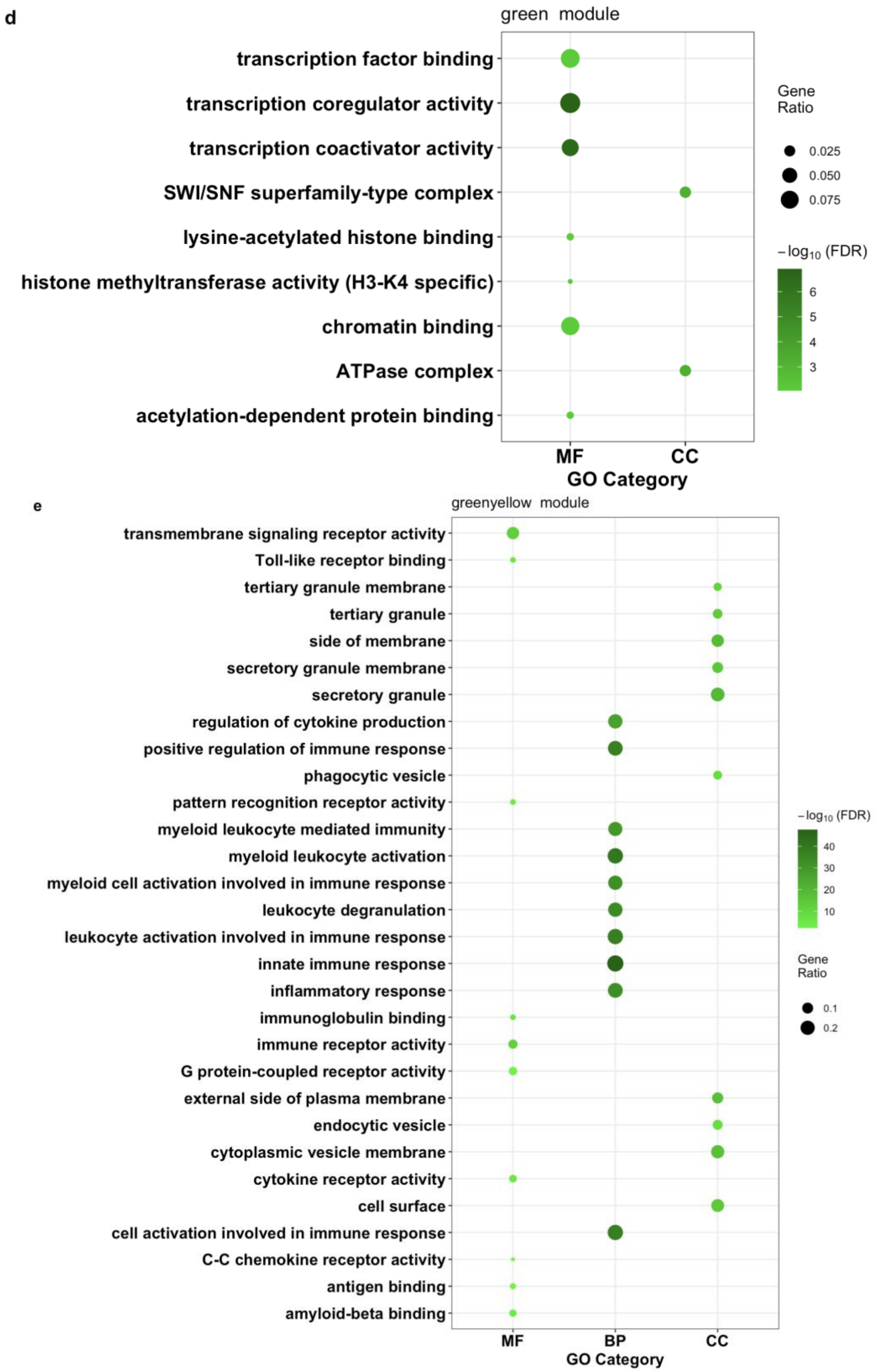
The lightgreen and turquoise modules, which are both correlated with maternal age and smoking, were significantly enriched for Biological Processes including hematopoietic progenitor cell differentiation and regulation of morphogenesis of epithelium, and with mRNA processing and RNA splicing, respectively. Further, Cellular components ribosomal subunit and mitochondrial inner membrane and Molecular function structural component of ribosome and electron transfer activity. The black and blue modules (both correlated with fetal sex) were significantly enriched in terms relating to mRNA processing and RNA splicing (black) and growth and development including cellular response to growth factor stimulus, regulation of growth and Molecular Functions including growth factor binding. This finding is interesting given that the male fetus responds differently to maternal stress, including smoking [70], and male and female fetal growth patterns differ with a resultant increase in feto-placental weight ratio in males [71].



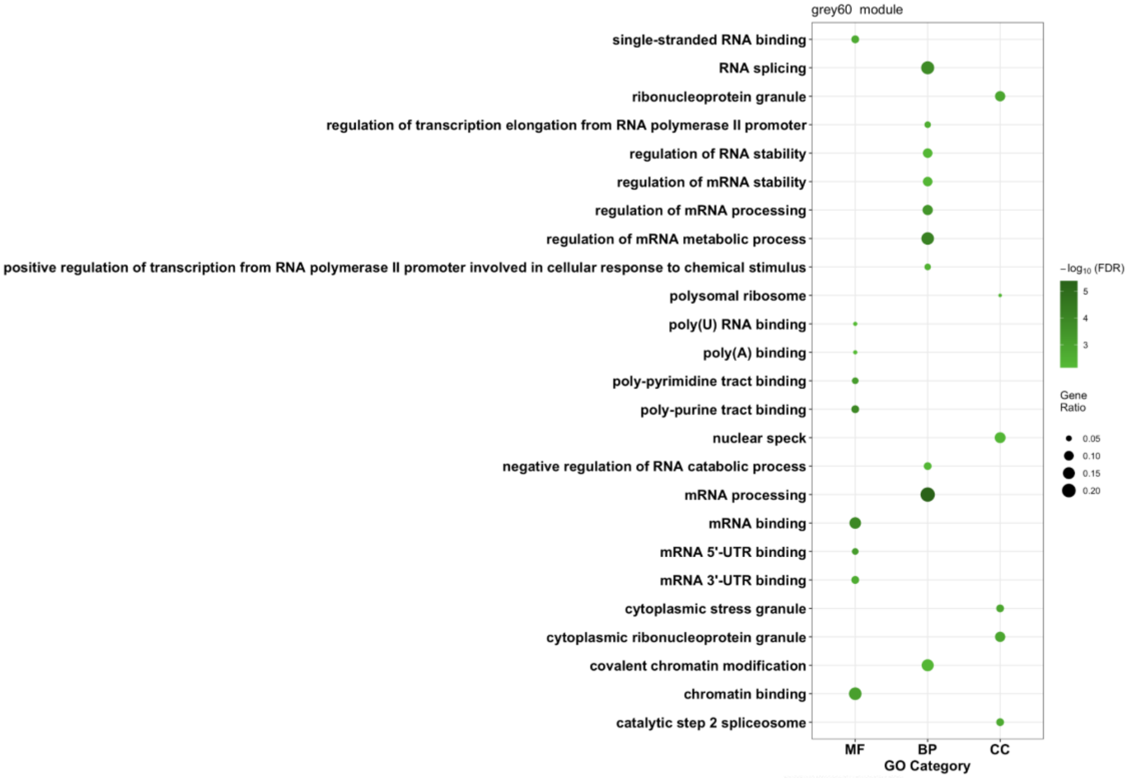

b
blue module
tissue morphogenesis
structural molecule activity
stress fiber
skeletal system development
response to growth factor
platelet-derived growth factor binding
heparin binding
heart development
growth factor binding
glycosaminoglycan binding
focal adhesion
extracellular structure organization
extracellular matrix structural constituent conferring tensile strength
extracellular matrix structural constituent
extracellular matrix organization
extracellular matrix
endoplasmic reticulum lumen
contractile actin filament bundle
connective tissue development
collagen-containing extracellular matrix
collagen trimer
collagen binding
cellular response to growth factor stimulus
cell-substrate junction
calcium ion binding
blood vessel morphogenesis
angiogenesis
actomyosin
actin cytoskeleton
actin binding
MF
BP
CC
GO Category
$-\log_{10}$ (FDR)
20
15
10
5
Gene Ratio
0.025
0.050
0.075
0.100

c

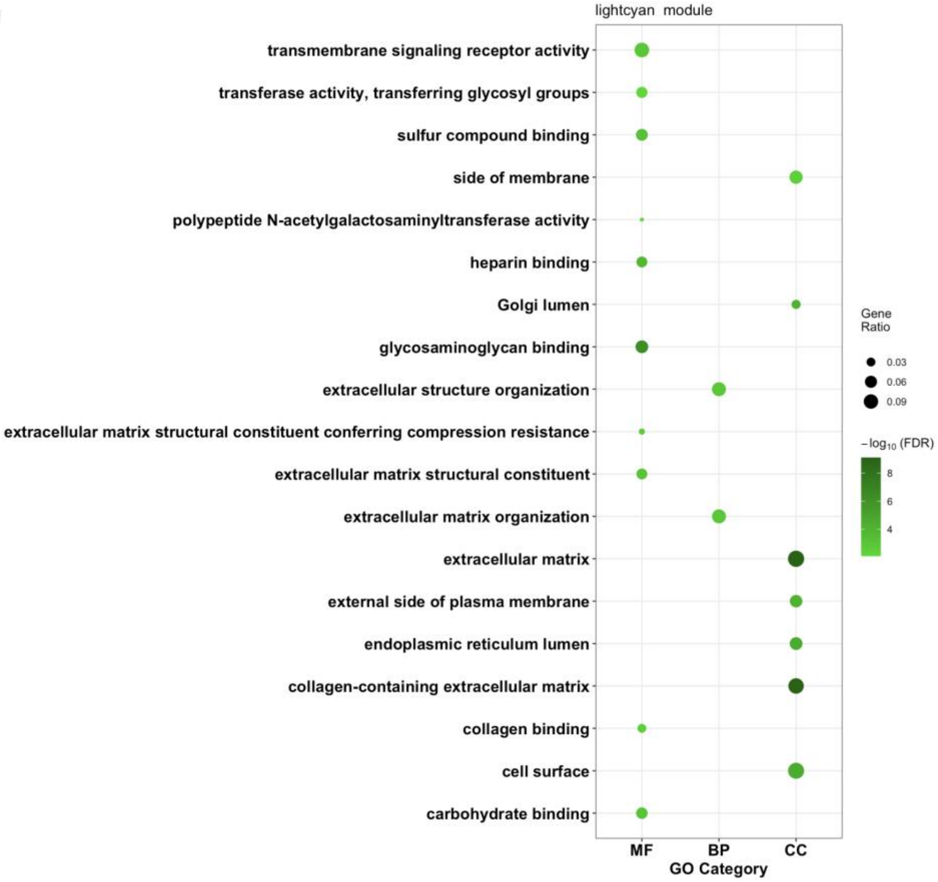

brown module
spindle
sister chromatid segregation
single-stranded DNA helicase activity
single-stranded DNA binding
protein-DNA complex
protein heterodimerization activity
organelle fission
nucleosome
nuclear division
nuclear chromosome segregation
mitotic sister chromatid segregation
mitotic nuclear division
kinetochore
helicase activity
DNA-dependent ATPase activity
DNA secondary structure binding
DNA replication origin binding
DNA replication
DNA packaging complex
DNA helicase activity
DNA conformation change
condensed chromosome, centromeric region
condensed chromosome kinetochore
condensed chromosome
chromosome, centromeric region
chromosome segregation
chromosomal region
cell division
catalytic activity, acting on DNA
ATPase activity
MF
BP
CC
GO Category
Gene Ratio
0.03
0.06
0.09
0.12
−log10 (FDR)
20
15
10
5


d
green module
transcription factor binding
transcription coregulator activity
transcription coactivator activity
SWI/SNF superfamily-type complex
lysine-acetylated histone binding
histone methyltransferase activity (H3-K4 specific)
chromatin binding
ATPase complex
acetylation-dependent protein binding
MF
CC
GO Category
Gene Ratio
0.025
0.050
0.075
−log10 (FDR)
6
5
4
3
e
greenyellow module
transmembrane signaling receptor activity
Toll-like receptor binding
tertiary granule membrane
tertiary granule
side of membrane
secretory granule membrane
secretory granule
regulation of cytokine production
positive regulation of immune response
phagocytic vesicle
pattern recognition receptor activity
myeloid leukocyte mediated immunity
myeloid leukocyte activation
myeloid cell activation involved in immune response
leukocyte degranulation
leukocyte activation involved in immune response
innate immune response
inflammatory response
immunoglobulin binding
immune receptor activity
G protein-coupled receptor activity
external side of plasma membrane
endocytic vesicle
cytoplasmic vesicle membrane
cytokine receptor activity
cell surface
cell activation involved in immune response
C-C chemokine receptor activity
antigen binding
amyloid-beta binding
MF
BP
CC
GO Category
−log10 (FDR)
40
30
20
10
Gene Ratio
0.1
0.2

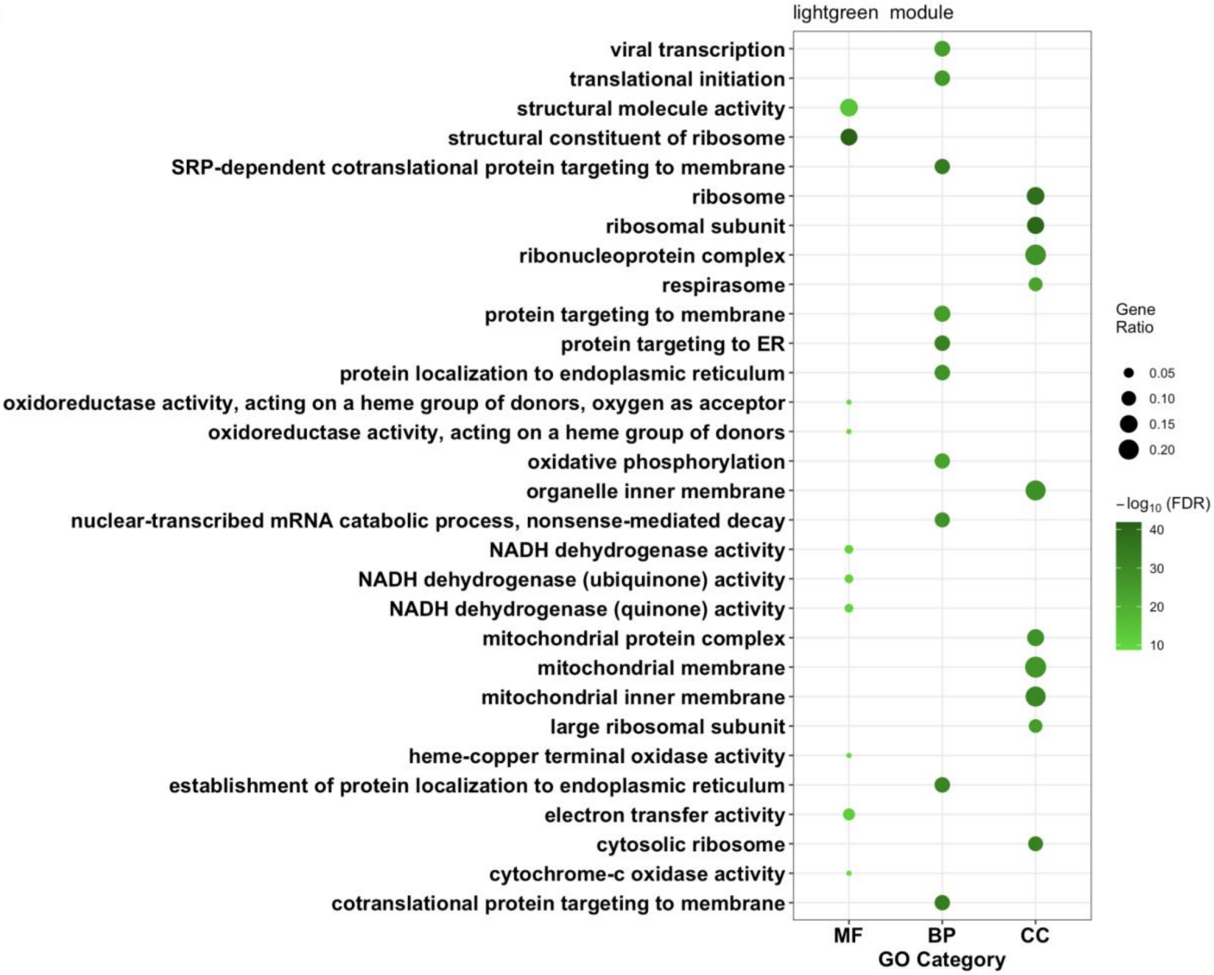
f


grey60 module
single-stranded RNA binding
RNA splicing
ribonucleoprotein granule
regulation of transcription elongation from RNA polymerase II promoter
regulation of RNA stability
regulation of mRNA stability
regulation of mRNA processing
regulation of mRNA metabolic process
positive regulation of transcription from RNA polymerase II promoter involved in cellular response to chemical stimulus
polysomal ribosome
poly(U) RNA binding
poly(A) binding
poly-pyrimidine tract binding
poly-purine tract binding
nuclear speck
negative regulation of RNA catabolic process
mRNA processing
mRNA binding
mRNA 5'-UTR binding
mRNA 3'-UTR binding
cytoplasmic stress granule
cytoplasmic ribonucleoprotein granule
covalent chromatin modification
chromatin binding
catalytic step 2 spliceosome
MF
BP
CC
GO Category
$-\log_{10}$ (FDR)
5
4
3
Gene Ratio
0.05
0.10
0.15
0.20

g


lightcyan module
transmembrane signaling receptor activity
transferase activity, transferring glycosyl groups
sulfur compound binding
side of membrane
polypeptide N-acetylgalactosaminyltransferase activity
heparin binding
Golgi lumen
glycosaminoglycan binding
extracellular structure organization
extracellular matrix structural constituent conferring compression resistance
extracellular matrix structural constituent
extracellular matrix organization
extracellular matrix
external side of plasma membrane
endoplasmic reticulum lumen
collagen-containing extracellular matrix
collagen binding
cell surface
carbohydrate binding
MF
BP
CC
GO Category
Gene Ratio
0.03
0.06
0.09
$-\log_{10}$ (FDR)
8
6
4

h


lightgreen module
viral transcription
translational initiation
structural molecule activity
structural constituent of ribosome
SRP-dependent cotranslational protein targeting to membrane
ribosome
ribosomal subunit
ribonucleoprotein complex
respirasome
protein targeting to membrane
protein targeting to ER
protein localization to endoplasmic reticulum
oxidoreductase activity, acting on a heme group of donors, oxygen as acceptor
oxidoreductase activity, acting on a heme group of donors
oxidative phosphorylation
organelle inner membrane
nuclear-transcribed mRNA catabolic process, nonsense-mediated decay
NADH dehydrogenase activity
NADH dehydrogenase (ubiquinone) activity
NADH dehydrogenase (quinone) activity
mitochondrial protein complex
mitochondrial membrane
mitochondrial inner membrane
large ribosomal subunit
heme-copper terminal oxidase activity
establishment of protein localization to endoplasmic reticulum
electron transfer activity
cytosolic ribosome
cytochrome-c oxidase activity
cotranslational protein targeting to membrane
MF
BP
CC
GO Category
Gene Ratio
0.05
0.10
0.15
0.20
$-\log_{10}$ (FDR)
40
30
20
10

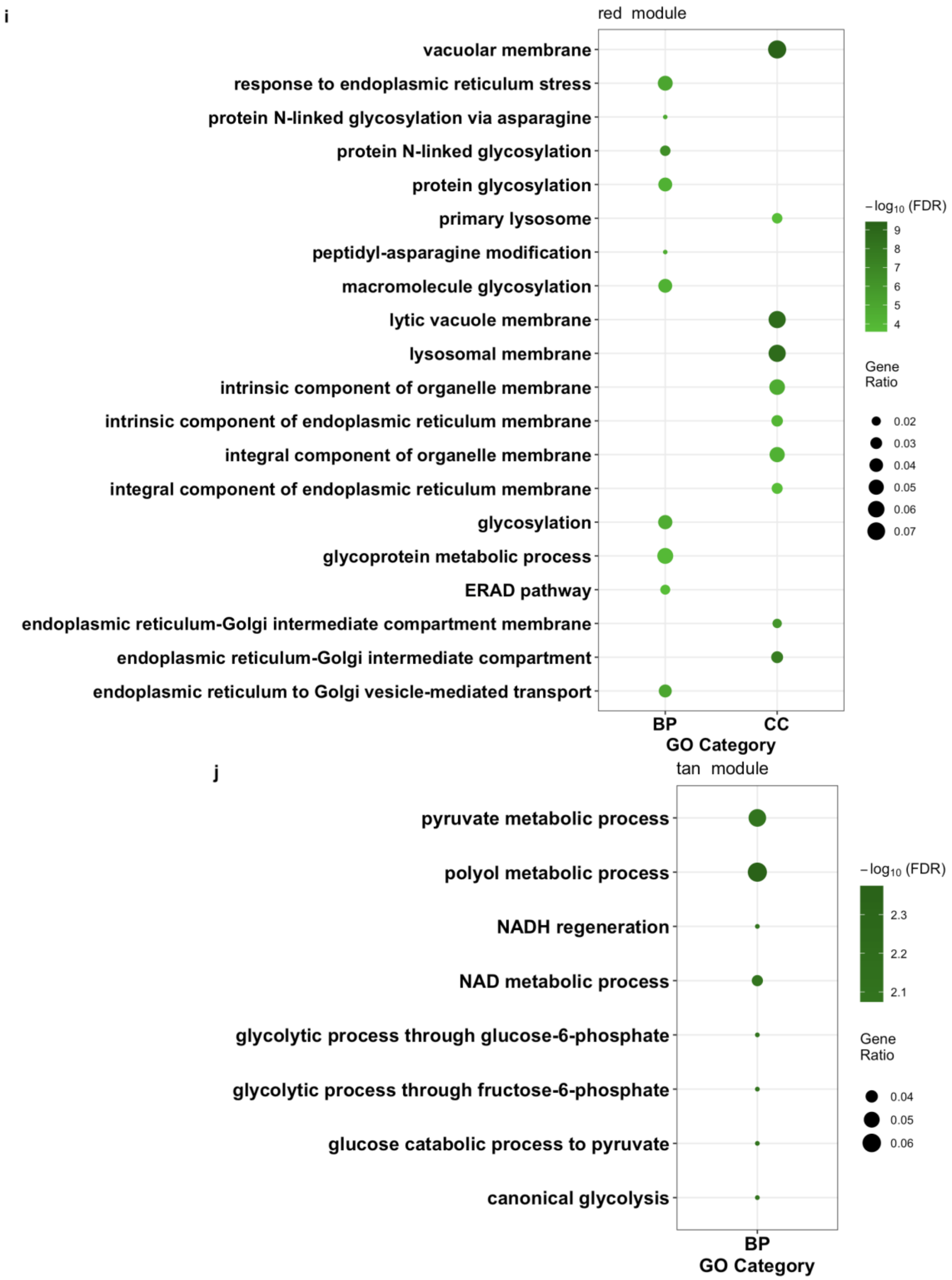


Figure 6. Top Gene Ontology (GO) pathways of WGCNA module genes. Modules with significant Go terms include black (**a**), blue (**b**), brown (**c**), green (**d**), greenyellow (**e**), grey60 (**f**), lightcyan (**g**), lightgreen (**h**), red (**i**) and tan (**j**).

After identifying enriched GO terms, gene set enrichment analysis (GSEA) was used to determine whether, using the module genes as individual *a priori* gene sets, modules provided a better understanding of the underlying biological processes associated with each module. Using KEGG pathway and Hallmark gene sets from the Molecular Signature Database (MSigDB), we found that PE associated tan module genes were enriched in the Glycolysis Gluconeogenesis KEGG pathway and Hypoxia, Glycolysis, MTORC1 Signalling and Estrogen Response Early Hallmark pathways, suggesting a potential role in PE disease progression. Turquoise, the largest module, contained 1106 genes and was enriched for RNA splicing and Cilium movement. This module was correlated with maternal smoking and maternal age in which alternative splicing and Ciliary beat frequency have been implicated [72–74]. The lightgreen module was also correlated with maternal smoking. Lightgreen module genes were enriched for Adipogenesis, Oxidative phosphorylation and Ribosome. Red, the most highly interconnected module, blue and black modules were correlated with fetal sex. These modules were enriched for Response to ER stress, Glycosylation; Extracellular matrix, Angiogenesis and G2M checkpoint respectively. The red and black modules were closely interconnected with black forming two subclusters, one of which was more closely connected to the red cluster, while blue formed two main subclusters with the larger being more closely connected to greenyellow, lightcyan and green clusters. The greenyellow module was enriched for the largest number of KEGG pathways, and both KEGG and Hallmark pathways tended to be associated with an immune or inflammatory response. The green module also formed two subclusters and was enriched for pathways including Transcription factor binding and Transcription factor coregulator activity. This is not surprising given that four of the ten hub genes (Table 2) were transcription factors. The brown module, with the lowest intermodule connectivity, was enriched for Cell division. In a recent single cell study, *TOP2A,* the gene with the highest kIM in the brown module was shown to be highly expressed in proliferative subclusters of multiple placental cells [75]. Lightcyan was interconnected with both the blue and greenyellow modules and enriched for pathways including Extracellular matrix organisation and Collagen containing extracellular matrix. Magenta, one of the smallest modules, genes were highly interconnected with red module genes. In particular, *CTBP1* and *CTBP1-AS,* the two top hub genes. There was no significant enrichment in this module. Lastly, the smallest module, grey60 was enriched for the KEGG pathway Spliceosome. Details of all module gene pathway enrichments are provided in Supplementary Tables 1 and 2.

Discussion

How genetic factors influence the etiology of both essential and gestational hypertension are an ongoing field of research [76,77]. PE is one of the leading causes of maternal death [1,2], yet our understanding of its pathogenesis and prediction is currently incomplete. Studies often focus on EOPE, or pooled groups without clinical subtyping, and rarely with a fetal sex-specific focus.

There is an emerging consensus that the development of PE is influenced by multiple modifiers including maternal and fetal genes and environmental factors [4,77,78]. LOPE, for example, differs from EOPE in that EOPE is often, but not exclusively, more severe and largely originates from impaired first trimester placentation. LOPE disease, by contrast, may be caused by a cascade of placental and systemic inflammation and oxidative stress triggered by a maternal predisposition to cardiovascular and metabolic syndrome-like disorders [79].

Differential expression analysis between LOPE and uncomplicated placenta identified 21 up and 4 down-regulated genes in males. We found no genes with statistical significance in the female comparison. In addition to the sex-specific differential expression analyses, we used the pooled male and female data to perform a weighted gene correlation analysis (WGCNA). WGCNA identified 13 co-expression modules, among which the tan module was correlated with PE, the black, blue and red modules were correlated with fetal sex and the turquoise module was correlated with maternal age.

Differential expression identified signs of oxidative stress in preeclamptic placentas

During our differential expression analysis, we identified Ceruloplasmin (*CP*), a copper binding glycoprotein that has antioxidative activity due to its inhibition of metal ion-catalysed oxidation of lipids [80], as the most highly up-regulated gene when comparing PE/Control males. Previous reports of *CP* have focussed on the maternal system and show increased abundance in maternal serum in pregnancies complicated with PE [81] and in the meta analysis [82]. Increased *CP* is indicative of the stressed state of the PE placenta. Unsurprisingly, maternal serum copper (Cu) and *CP* have been associated with PE severity [83]. In an analysis of maternal plasma (Wilcoxon rank sum test with continuity correction, Wilson and Leemaqz, Unpublished) we found Cu to be significantly (p-value <0.05) higher (~100 µg/L) in women with a pregnancy complicated by PE. In a recent single cell placental study, *CP* was shown to be expressed in extravillous trophoblasts (EVT) [84]. During early placental development, EVTs migrate from the top of the anchoring villi invading the decidua (forming placental bed giant cells) or the uterine spiral arteries (forming endovascular evCTB) [84]. Increased *CP* from invading EVTs may form part of the underlying mechanism for endothelial dysfunction seen in LOPE. Whilst intriguing, it is difficult to say whether the increased placental *CP* we found in placenta tissue and maternal plasma Cu are part of the cause or effect of PE.

Placenta-derived oxidative and nitrosative stresses are considered major molecular determinants of maternal PE [85]. Here we show up-regulation of known hypoxia markers such as *EHHADH* [86]. Glutathione S-transferases, such as Glutathione S-Transferase Mu 3 (*GSTM3*), are part of a large antioxidant enzyme family with a role in detoxification processes. We found *GSTM3* to be up-regulated in placenta from PE males, consistent with the increase in oxidative stress associated with PE. Yet another indication of the male placenta under stress was the up-regulation of DAZ Interacting Protein 1 (*DZIP1*). Up-regulation of *DZIP1* appears to be important for the formation of stress granules (dense aggregations in the cytosol composed of proteins and RNAs that appear when the cell is under stress) during the stress response [87]. Linking this stress to PE, a recent systematic review found an association between zinc and hypertensive disorders of pregnancy [88] with previous research showing that modulating the expression level of *DZIP1* may affect the accumulation of zinc (Zn^{2+}) in the brain [89], which is in turn associated with eclampsia/seizure. *EHHADH* clustered with genes in the brown module which was enriched with pathways including DNA conformation and Cell division, but was not correlated with any of our maternal characteristics or pregnancy outcome. *CP, GSTM3 and DZIP1* were assigned to the grey module suggesting no significant correlation in their gene expression. This may be due to the sex-naive nature of the network co-expression approach. Further analyses using larger sample numbers with a sex-specific approach may help to further elucidate our findings here.

Dysregulated angiogenesis genes associated with late-onset preeclampsia

Preeclampsia is often characterized by dysfunctional endothelium and impaired angiogenesis, resulting in disturbed vasculature of the placenta. Late-onset preeclampsia, has been suggested to be exacerbated by a maternal predisposition to endothelial dysfunction [3]. Inadequate invasion and aberrant angiogenesis also appear to play roles in the pathophysiology of LOPE [90,91]. The molecular causes, however, are still being investigated. *ROBO1*, up-regulated in PE males placentas, in conjunction with *Slit* family proteins, is regulated by hypoxia and has previously been shown to be dysregulated in PE placental trophoblast and endothelial cells [90]. Further, *Slit2*, once bound to *ROBO1*, helps to control vascular network formation by functioning as an attractant to promote the directional migration in vascular endothelial cells *in vitro [91]*.

Vascular formation may also be affected by down-regulation of the zinc finger *ZNF79* which has been implicated in hereditary haemorrhagic telangiectasia, vascular dysplasia, characterized by widespread dermal, mucosal and visceral telangiectases and recurrent haemorrhage [92]. *CD44*, a critical inhibitor of angiogenesis with CD44/FKBPL ratio suggested as a potential PE risk stratification measure [93], and *TSPAN12*, implicated in vascular homeostasis [94] were also up-regulated in PE male placentas. *ROBO1, Slit2* and *CD44* were strongly co-expressed and found in the blue module genes.

Factors associated with proliferation, migration and invasion show causal potential in late-onset preeclampsia

Given its proliferative and invasive nature and the suppression of immune responses, the placenta has been likened to a controlled form of cancer [95]. Proliferation, migration and invasion of trophoblast cells, essential to the function and growth of the placenta, is also implicated in our analyses. We identified up-regulation (in PE males) of invasion suppressor *CRMP1* [96,97], proliferation inhibitor *HSD17B12* [98] and Ras-like and estrogen-regulated growth inhibitor (*RERG*). *RERG,* through regulation by *ERβ*, has been shown to enhance the proliferation of endometriotic cells [99]. Further, we found down-regulation of *GPR158*, the over expression of which is associated with stimulation of cell proliferation [100]. Our network approach identified *CRMP1* assigned to the blue module, *HSD17B12* and *RERG* assigned to the brown module and *GPR158* was found in the grey module.

We also show up-regulation (in PE males) of *WWOX* which, in association with complement component Cq1, is critical to invasion by extravillous cytotrophoblasts [101]. Members of the transmembrane (TMEM) protein family are well characterised in cancer including in the formation of metastases and in the mechanisms leading to cancer cell dissemination such as migration and extracellular matrix remodelling [102]. In a recent analysis, *TMEM150C,* which we found to be up-regulated in PE male placentas, was associated with gestational age, but not with PE [103], again highlighting the need to include fetal sex in placental analyses. A second TMEM member, *TMEM144*, a carbohydrate transporter not well studied in the placenta, harbours a human endogenous retrovirus 5′ long terminal repeat (HERV-E LTR) which, at least in part due to hypomethylation in this unique tissue, functions as an alternative gene promoter in placenta [104]. Consistent with the "placenta as a controlled cancer" paradigm, we found *TRIM29* to be down-regulated in male PE placenta. *TRIM29* appears to have a paradoxical expression pattern. In tissue such as breast or prostate, where *TRIM29* is readily detectable, expression is decreased in cancer tissue, but for organs such as stomach or colon/rectum, where expression is low or undetected, there appears to be an upregulation in cancer tissue [105]. Much like the pattern shown in breast or prostate cancers where *TRIM29* is expressed in healthy tissue, here we found that *TRIM29*, usually detectable in placenta, was down-regulated in male PE placenta. *WWOX*, *TMEM144* and *TRIM29* were found in the grey module with *TMEM150C* assigned to brown. That our DE genes have variously been assigned into different modules highlights both the varied pathways involved in PE and perhaps also the change between sex-specific and sex-naive analysis.

Zinc Finger Protein 300 (*ZNF300*), typically up-regulated in placenta from female bearing pregnancies, was up-regulated in placenta from PE males (FDR <0.05). This transcript encodes a TF that has competitive binding sites to Early Growth Response 1 (*Egr1*), a zinc finger TF elevated in the first trimester placenta of pregnancies that later develop PE [54,106]. Interestingly, up-regulation of *ZNF300* and the subsequent competitive binding with *Egr1,* has been suggested as a factor in the reduced risk of PE in female bearing pregnancies [54]. Unlike the DE comparison, our network approach included a filter for genes with low variance (described in the results). This filter for low variance removed *ZNF300* from the WGCNA analysis. However, our sex-specific DE results suggest that this gene does vary in the male placenta again highlighting that sex differences in expression and regulation of this gene are more complex and deserve further scrutiny.

We also found up-regulation of the collagen gene, *COL6A3,* which encodes the α3 chain of collagen VI, previously associated with increased collagen in placenta from mothers who smoke during pregnancy [107] but, that was previously found not to be dysregulated in PE [108,109]. We identified *COL6A3* up-regulated in PE males compared to Control males. Previously, Collagen VI deficiency has been implicated in mechanisms causing the dystrophic changes in Ullrich congenital muscular dystrophy [110], with complete reduction altering the fibronectin network architecture in fibroblast cultures [111] and reducing cell proliferation [112]. Presence of soluble collagen VI prevents apoptosis in serum-starved fibroblasts [113] through activation of tyrosine phosphorylation of focal adhesion proteins and was suggested to act as a matrix-derived sensor that allows for rapid reconstitution of a tissue defect by activating nearby mesenchymal cells. Insulin-like growth factor-binding protein 7 (IGFBP7)*,* stimulated adhesion of human umbilical vein endothelial cells to type IV collagen substrate inducing morphological changes [114]. Increased activity of *IGFBP7* is associated with cellular senescence, tissue aging, and obesity, with higher concentrations associated with diastolic dysfunction [115]. These complex mechanisms affecting differentiation, regeneration, and apoptosis may also apply in placenta and may be involved in the pathological pathways associated with either the cause or effect of PE.

Mineralocorticoid receptor link between early-onset and gestational hypertension

Abnormalities in the mineralocorticoid receptor (Nuclear Receptor Subfamily 3 Group C Member 2 (*NR3C2*)) effector mechanism, possibly through sodium (Na+) retention, play a role in the etiology of PE [116]. A functional polymorphism of the NR3C2 gene has been associated with risk of early-onset hypertension [76], with variations in this gene associated with aldosterone mediated sodium channel activation affecting salt/water balance associated with early-onset hypertension exacerbated in pregnancy [117]. Interestingly, we identified an up-regulation of NR3C2 in male, but not female, placenta from preeclamptic pregnancies, with no associated polymorphism as previously described. The serum glucocorticoid dependent kinase (*SGK1-3*) family has been identified as critical to the insertion of Na+ channel proteins into the cell membrane in the kidney [118]. All three *SGKs* are involved in the antinatriuretic action of IGF1 and insulin, both elevated in pregnancy. This suggests that dysregulation of *NR3C2*, without protein conformational changes, and the regulation of Na+, may have a causal role in LOPE in male bearing pregnancy. It also further establishes the potential importance of this mechanism in the understanding of PE, PE subtypes, and sex differences in pregnancy outcomes.

The products of solute carrier genes (SLC), transport sugars, amino acids, and inorganic ions across cell membranes undoubtedly play roles in the regulation of the transfer of these materials across the feto-maternal interface throughout pregnancy. We identified up-regulation of three SLC genes in PE male placentas (*SLC26A6, SLC18A2, SLC2A13*) which are associated with transport of Cl−/HCO3− [119], monoamines [120] and glucose (facilitated transport) [121], respectively. Often studied in the brain, altered gene expression, and subsequent dysregulation of protein abundance of the SLC family may play a role in maternal kidney function [122] and the functional role of kidney and placenta in the volume abnormalities (extracellular fluid) observed in PE [123].

We have attempted here to overcome the limitations often seen in studies of PE by focusing on sex-specific analyses of LOPE, specifically. However, our initial power calculations revealed that we would require 75 samples to have 95% confidence in finding ~500 differentially expressed genes. Given that we tested DE independently by fetal sex as a multi-group comparison, a total of 150 samples would be optimal. Previously, Tejera *et al.* (2013) [61] combined microarray data from 5 independent projects to identify 1146 significant genes in an analysis of preeclampsia in placenta tissue. In future work, the sample size should be expanded and the collection of maternal information extended enabling a more detailed investigation of the intricate molecular signature of LOPE. Also, although we identified many genes and pathways worthy of further investigation, no *in vivo* or *in vitro* verification studies have yet been completed.

Whilst placenta flow defects inferred from uterine artery Doppler flow analyses, can be detected as early as 12 weeks' gestation in women who go on to develop PE [124], in the case of LOPE (also known as "maternal preeclampsia"), the placenta is often of an appropriate size for gestation. Genome-wide expression profiling in the context of LOPE, and with the added context of fetal sex, allows for highly specific unravelling of the biological functions involved in this complex disease. With early and late-onset preeclampsia increasingly acknowledged as different pathophysiological processes leading to a common presentation [125], the difference in their underlying molecular bases remains elusive. As with many complex disorders, accurate subtyping is the first step in the path to both understanding the disease, and in providing clinical care and intervention. The different associations identified here between placental TF and response to oxidative stress in the male and female bearing pregnancy warrant further study directed not just at the accurate subtyping of PE, but also in how PE subtypes interact with fetal sex.

Methods

Study Participants

Patient samples were taken from the Adelaide arm of two prospective cohort studies. The first was the **SC**reening f**O**r **P**regnancy Endpoints (SCOPE) study [126,127]. SCOPE is a prospective, international, multicentre cohort study of healthy, nulliparous, singleton-bearing women which aimed to develop screening tools to predict risk for pregnancy complications. Our cohort was recruited from the Lyell McEwin Hospital, Adelaide, Australia between November 2004 and September 2008. The second was the **S**creening **T**ests to predict poor **O**utcomes of **P**regnancy (STOP) study [128]. This is another prospective multicentre cohort study of healthy, nulliparous, singleton-bearing women. Recruitment occurred across three Hospitals in Adelaide, South Australia (Lyell McEwin Hospital, Elizabeth Vale; Modbury Hospital, Modbury; and Women's and Children's

Hospital, North Adelaide) from 2015 to 2018. Women were excluded from participation if they had ≥3 miscarriages or ≥3 terminations of pregnancy, major fetal anomalies, pre-existing hypertension on medication, Type I or Type II diabetes mellitus, renal disease, systemic lupus erythematosus, antiphospholipid syndrome, known major uterine anomaly or previous cervical cone biopsy. Fetal and maternal outcomes were obtained directly from clinical records [129].

Sample collection

Term placental tissue was collected from women recruited as part of either the SCOPE or STOP Studies. Post delivery, term placenta samples were collected from pregnancies classified as being either uncomplicated, using the criteria described in [130] or as preeclamptic, using criteria described in [1,131]. Placentas were collected and dissected within 3 hours from birth and stored at 4 ◦C for 24 h in RNALater (Thermo Fisher) and stored at −80 ◦C.

RNA extraction and library preparation

RNA was extracted from each placental villus sample using the QIAGEN RNeasy Plus Mini Kit (QIAGEN, Hilden, Germany) following the manufacturer’s protocols with the exception of substituting the RW1 buffer for 70% ethanol for better retention of smaller RNA fragments. Sequencing libraries were prepared using Illumina TruSeq Stranded Total RNA Sample Preparation kits and all the ribosomal RNA was depleted using Ribo-Zero Gold. Sequencing was performed on the Illumina Hi-Seq 2500 using a 100bp paired-end protocol at the Flinders University Genomics Facility, SA.

Note: Source code for all analyses conducted here is available at (https://github.com/mxhp75/late_onset_preeclampsia_analysis)

Differential expression analysis

Differential expression analyses were conducted in the R statistical environment (v.4.0.5), using the *edgeR* (v.3.16.5) [28] and *limma* (v.3.30.11) [29] packages. Briefly, *edgeR* was used to filter mRNA with low expression and normalise for library composition bias. All samples were then normalised using the Trimmed Mean of M values (TMM). Sample-weights and log transformation was performed using the *limma* package with the voom function used to estimate the mean-variance relationship between individual observations and then applied to the normalised log-counts data. Differential expression analysis including, moderated F-statistic evaluation, adjusted p-value estimation and $\log_2$ fold change analysis were performed using a moderated t-test [30] with Benjamini-Hochberg (BH) multiple hypothesis test corrections [31]. After adjustment, gene expression was considered significantly different at FDR ≤ 0.05.

Weighted genome co-expression network analysis

Weighted gene correlation network analysis (WGCNA) was carried out using the R package WGCNA [58] and incorporates two sets of term RNA-seq samples. Low count and low variability genes were removed using *edgeR* [132], batch effects removed using *limma* [133].

To construct the gene-wise network, we first created a symmetric correlation matrix (the adjacency matrix) where each cell contains the Pearson's correlation between the expression of a pair of genes, and then raising each correlation value to a soft thresholding power. The adjacency matrix considers how related the expression of two genes are to each other. Genes with low connectivity do not have much influence in the network and act as noise, so it is desirable to remove them. Using the connection strengths between genes in the adjacency matrix, we next transformed the data into a topological overlap matrix (TOM). The TOM considers how related each pair of genes is to the rest of the network. Genes have high topological overlap if they are connected to the same group of genes within the network [134]. From the TOM we calculated a dissimilarity matrix as a measure of distance between genes (dissimilarity matrix = 1-TOM). Finally, the values from the dissimilarity matrix were used to identify modules of co-expressed genes through unsupervised hierarchical clustering. The signed network and adjacency matrices were constructed using a soft power threshold of 12 and modules with a minimum size of 50 generated and merged at distance cut-off of 0.25. Network graphs were constructed and visualised using Gephi (https://gephi.org/).

Gene set enrichment analysis

Gene set enrichment analysis (GSEA) of genes from the differential expression analysis was carried out using the GSEA software ([56], http://www.broad.mit.edu/gsea/) and Molecular Signature Database (MSigDB v6.2) Hallmark and Immunologic (C7) gene sets. For the WGCNA analysis, module genes were tested for enrichment within Gene Ontology and Molecular Signature Database (MSigDB v6.2) Hallmark and KEGG gene sets [56] using the WGCNA R package AnRichment (https://horvath.genetics.ucla.edu/html/CoexpressionNetwork/GeneAnnotation/) and clusterProfiler [135].

Ethics Statement

This study used data obtained through two independent cohorts. Ethics approval for the Screening for Obstetric Perinatal Endpoints (SCOPE) Study was granted by the Central Northern Adelaide Health Service Ethics of Human Research Committee, ethics number REC 1714/5/2008 with all participants providing written, informed consent. Ethics for the Screening Tests to identify poor Outcomes in Pregnancy (STOP) Study was granted by the Northern Adelaide Local Health Network (LMH/MH/PHC) for the Lyell McEwin Hospital and by the WCHN Human Research Ethics Committee for the Women's and Children's Hospital, ethics number HREC/14/WCHN/90, Site Specific Approval SSA/14/NALHN/7 with all participants providing written, informed consent.

Author contribution

CTR created the concept. MS, JB and KP conceived and developed bioinformatic experimental plans. TJK, DMcA and DMcC performed molecular experiments. MS analysed the sequencing data. KJB assisted with WGCNA. Manuscript written by MS. All authors read and approved the final version of the manuscript.

References

1. Brown MA, Magee LA, Kenny LC, Karumanchi SA, McCarthy FP, Saito S, *et al.* Hypertensive Disorders of Pregnancy: ISSHP Classification, Diagnosis, and Management Recommendations for International Practice. Hypertension. 2018;72: 24–43.

2. Abalos E, Cuesta C, Grosso AL, Chou D, Say L. Global and regional estimates of preeclampsia and eclampsia: a systematic review. Eur J Obstet Gynecol Reprod Biol. 2013;170: 1–7.

3. Steegers EAP, von Dadelszen P, Duvekot JJ, Pijnenborg R. Pre-eclampsia. Lancet. 2010;376: 631–644.

4. Sibai B, Dekker G, Kupferminc M. Pre-eclampsia. Lancet. Elsevier; 2005. pp. 785–799.

5. the Guidelines Review Committee WGA by, Others. WHO recommendations for prevention and treatment of pre-eclampsia and eclampsia. Geneva: World Health Organization. 2011.

6. Barker DJ, Gluckman PD, Godfrey KM, Harding JE, Owens JA, Robinson JS. Fetal nutrition and cardiovascular disease in adult life. Lancet. 1993;341: 938–941.

7. Barker DJ. Fetal origins of coronary heart disease. BMJ. 1995;311: 171–174.

8. Barker DJP. The developmental origins of adult disease. J Am Coll Nutr. 2004;23: 588S–595S.

9. Bellamy L, Casas J-P, Hingorani AD, Williams DJ. Pre-eclampsia and risk of cardiovascular disease and cancer in later life: systematic review and meta-analysis. BMJ. 2007. p. 974. doi:10.1136/bmj.39335.385301.be

10. Andraweera PH, Lassi ZS. Cardiovascular Risk Factors in Offspring of Preeclamptic Pregnancies-Systematic Review and Meta-Analysis. J Pediatr. 2019;208: 104–113.e6.

11. Andraweera PH, Gatford KL, Care AS, Bianco-Miotto T, Lassi ZS, Dekker GA, *et al.* Mechanisms linking exposure to preeclampsia in utero and the risk for cardiovascular disease. J Dev Orig Health Dis. 2020;11: 235–242.

12. Plummer MD, Andraweera PH, Garrett A, Leemaqz S, Wittwer M, Aldridge E, *et al.*

Hypertensive disorders of pregnancy and later cardiovascular disease risk in mothers and children. J Dev Orig Health Dis. 2021;12: 555–560.

13. Liu X, Zhao W, Liu H, Kang Y, Ye C, Gu W, *et al.* Developmental and Functional Brain Impairment in Offspring from Preeclampsia-Like Rats. Mol Neurobiol. 2016;53: 1009–1019.
14. Walker CK, Krakowiak P, Baker A, Hansen RL, Ozonoff S, Hertz-Picciotto I. Preeclampsia, placental insufficiency, and autism spectrum disorder or developmental delay. JAMA Pediatr. 2015;169: 154–162.
15. Dachew BA, Mamun A, Maravilla JC, Alati R. Pre-eclampsia and the risk of autism-spectrum disorder in offspring: meta-analysis. Br J Psychiatry. 2018;212: 142–147.
16. Griffith MI, Mann JR, McDermott S. The risk of intellectual disability in children born to mothers with preeclampsia or eclampsia with partial mediation by low birth weight. Hypertens Pregnancy. 2011;30: 108–115.
17. Wu CS, Sun Y, Vestergaard M, Christensen J, Ness RB, Haggerty CL, *et al.* Preeclampsia and risk for epilepsy in offspring. Pediatrics. 2008;122: 1072–1078.
18. Dalman C, Allebeck P, Cullberg J, Grunewald C, Köster M. Obstetric complications and the risk of schizophrenia: a longitudinal study of a national birth cohort. Arch Gen Psychiatry. 1999;56: 234–240.
19. Ursini G, Punzi G, Chen Q, Marenco S, Robinson JF, Porcelli A, *et al.* Convergence of placenta biology and genetic risk for schizophrenia. Nat Med. 2018;24: 792–801.
20. Verdoux H, Sutter AL. Obstetric complications, maternal psychopathology, and the risk of psychosis. Risk and Protective Factors in Schizophrenia. 2002. pp. 105–112. doi:10.1007/978-3-642-57516-7_9
21. Abedian Z, Soltani N, Mokhber N, Esmaily H. Depression and anxiety in pregnancy and postpartum in women with mild and severe preeclampsia. Iran J Nurs Midwifery Res. 2015;20: 454–459.
22. Paarlberg KM, Vingerhoets AJ, Passchier J, Dekker GA, Heinen AG, van Geijn HP. Psychosocial predictors of low birthweight: a prospective study. Br J Obstet Gynaecol. 1999;106: 834–841.
23. Wainstock T, Anteby E, Glasser S, Shoham-Vardi I, Lerner-Geva L. The association between prenatal maternal objective stress, perceived stress, preterm birth and low birthweight. J Matern Fetal Neonatal Med. 2013;26: 973–977.

24. Ciesielski TH, Marsit CJ, Williams SM. Maternal psychiatric disease and epigenetic evidence suggest a common biology for poor fetal growth. BMC Pregnancy Childbirth. 2015;15: 192.

25. Uguz F, Gezginc K, Yazici F. Are major depression and generalized anxiety disorder associated with intrauterine growth restriction in pregnant women? A case-control study. Gen Hosp Psychiatry. 2011;33: 640.e7–9.

26. Aneman I, Pienaar D, Suvakov S, Simic TP, Garovic VD, McClements L. Mechanisms of Key Innate Immune Cells in Early- and Late-Onset Preeclampsia. Front Immunol. 2020;11: 1864.

27. Raymond D, Peterson E. A critical review of early-onset and late-onset preeclampsia. Obstet Gynecol Surv. 2011;66: 497–506.

28. Herzog EM, Eggink AJ, Reijnierse A, Kerkhof MAM, de Krijger RR, Roks AJM, *et al.* Impact of early- and late-onset preeclampsia on features of placental and newborn vascular health. Placenta. 2017;49: 72–79.

29. Herzog EM, Eggink AJ, Willemsen SP, Slieker RC, Wijnands KPJ, Felix JF, *et al.* Early- and late-onset preeclampsia and the tissue-specific epigenome of the placenta and newborn. Placenta. 2017;58. doi:10.1016/j.placenta.2017.08.070

30. Carrasco-Wong I, Aguilera-Olguín M, Escalona-Rivano R, Chiarello DI, Barragán-Zúñiga LJ, Sosa-Macías M, *et al.* Syncytiotrophoblast stress in early onset preeclampsia: The issues perpetuating the syndrome. Placenta. 2021. doi:10.1016/j.placenta.2021.05.002

31. Redman CWG, Tannetta DS, Dragovic RA, Gardiner C, Southcombe JH, Collett GP, *et al.* Review: Does size matter? Placental debris and the pathophysiology of pre-eclampsia. Placenta. 2012;33 Suppl: S48–54.

32. Burton GJ, Redman CW, Roberts JM, Moffett A. Pre-eclampsia: pathophysiology and clinical implications. BMJ. 2019;366: l2381.

33. Tsai S, Hardison NE, James AH, Motsinger-Reif AA, Bischoff SR, Thames BH, *et al.* 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.

34. Enquobahrie DA, Meller M, Rice K, Psaty BM, Siscovick DS, Williams MA. Differential placental gene expression in preeclampsia. Am J Obstet Gynecol. 2008;199: 566.e1–11.

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

36. Kang JH, Song H, Yoon JA, Park DY, Kim SH, Lee KJ, *et al.* Preeclampsia leads to dysregulation of various signaling pathways in placenta. J Hypertens. 2011;29: 928–936.

37. Nishizawa H, Pryor-Koishi K, Kato T, Kowa H, Kurahashi H, Udagawa Y. Microarray analysis of differentially expressed fetal genes in placental tissue derived from early and late onset severe pre-eclampsia. Placenta. 2007;28: 487–497.

38. Moslehi R, Mills JL, Signore C, Kumar A, Ambroggio X, Dzutsev A. Integrative transcriptome analysis reveals dysregulation of canonical cancer molecular pathways in placenta leading to preeclampsia. Scientific Reports. 2013. doi:10.1038/srep02407

39. Liu J, Song G, Meng T, Zhao G. Identification of Differentially Expressed Genes and Signaling Pathways in Placenta Tissue of Early-Onset and Late-Onset Pre-Eclamptic Pregnancies by Integrated Bioinformatics Analysis. Medical Science Monitor. 2020. doi:10.12659/msm.921997

40. Seeho SKM, Park JH, McCracken SA, Eileen D. M. Gallery, Morris JM. Dysregulation of oxygen sensing/responsive pathways in pregnancies complicated by idiopathic intrauterine growth restriction and early onset preeclampsia. Placenta. 2015. pp. A50–A51. doi:10.1016/j.placenta.2015.07.339

41. Leavey K, Bainbridge SA, Cox. Large scale aggregate microarray analysis reveals three distinct molecular subclasses of human preeclampsia. Gleicher N, editor. PLoS One. 2015;10: e0116508.

42. Zhang W, Yu Y, Hertwig F, Thierry-Mieg J, Zhang W, Thierry-Mieg D, *et al.* Comparison of RNA-seq and microarray-based models for clinical endpoint prediction. Genome Biol. 2015;16: 133.

43. Ren Z, Gao Y, Gao Y, Liang G, Chen Q, Jiang S, *et al.* Distinct placental molecular processes associated with early-onset and late-onset preeclampsia. Theranostics. 2021;11: 5028–5044.

44. Founds SA, Conley YP, Lyons-Weiler JF, Jeyabalan A, Hogge WA, Conrad KP. Altered global gene expression in first trimester placentas of women destined to develop preeclampsia. Placenta. 2009;30: 15–24.

45. Sibai BM, Stella CL. Diagnosis and management of atypical preeclampsia-eclampsia. Am J Obstet Gynecol. 2009;200: 481.e1–7.

46. Schneider VA, Graves-Lindsay T, Howe K, Bouk N, Chen H-C, Kitts PA, *et al.* Evaluation of GRCh38 and de novo haploid genome assemblies demonstrates the enduring quality of the reference assembly. bioRxiv. 2016. p. 072116. doi:10.1101/072116

47. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, *et al.* STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29: 15–21.

48. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 2014;30: 923–930.

49. Toivanen P, Hirvonen T. Sex ratio of newborns: preponderance of males in toxemia of pregnancy. Science. 1970;170: 187–188.

50. Makhseed M, Musini VM, Ahmed MA. Association of fetal gender with pregnancy-induced hypertension and pre-eclampsia. Int J Gynaecol Obstet. 1998;63: 55–56.

51. Hsu CD, Johnson TR, Witter FR. Fetal gender effect on pre-eclampsia: a retrospective study of Baltimore area in the United States of America. Int J Gynaecol Obstet. 1994;45: 160–161.

52. Elsmén E, Källén K, Maršál K, Hellström-Westas L. Fetal gender and gestational-age-related incidence of pre-eclampsia. Acta Obstet Gynecol Scand. 2006;85: 1285–1291.

53. Jaskolka D, Retnakaran R, Zinman B, Kramer CK. Fetal sex and maternal risk of pre-eclampsia/eclampsia: a systematic review and meta-analysis. BJOG. 2017;124: 553–560.

54. Xue L, Qiu H, Ma J, Guo M, Li W. ZNF300, a recently identified human transcription factor, activates the human IL-2Rβ promoter through the overlapping ZNF300/EGR1 binding site. Cell Mol Biol Lett. 2010;15: 530–540.

55. Ladd-Acosta C, Andrews SV, Bakulski KM, Feinberg JI, Tryggvadottir R, Yao R, *et al.* Placenta DNA methylation at ZNF300 is associated with fetal sex and placental morphology. bioRxiv. 2021. p. 2021.03.05.433992. doi:10.1101/2021.03.05.433992

56. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005;102: 15545–15550.

57. Mootha VK, Lindgren CM, Eriksson K-F, Subramanian A, Sihag S, Lehar J, *et al.* PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately

downregulated in human diabetes. Nat Genet. 2003;34: 267–273.

58. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008;9: 559.

59. Ito M, Nishizawa H, Tsutsumi M, Kato A, Sakabe Y, Noda Y, *et al.* Potential role for nectin-4 in the pathogenesis of pre-eclampsia: a molecular genetic study. BMC Med Genet. 2018;19: 166.

60. Gong L, Zhao H, Cui Y, Li X. Transcriptome analysis of placentae reveals HELLP syndrome exhibits a greater extent of placental metabolic dysfunction than preeclampsia. Hypertens Pregnancy. 2021;40: 134–143.

61. Tejera E, Bernardes J, Rebelo I. Co-expression network analysis and genetic algorithms for gene prioritization in preeclampsia. BMC Med Genomics. 2013;6: 51.

62. Fu Y, Wei J, Dai X, Ye Y. Increased NDRG1 expression attenuate trophoblast invasion through ERK/MMP-9 pathway in preeclampsia. Placenta. 2017;51: 76–81.

63. Kaartokallio T, Cervera A, Kyllönen A, Laivuori K, Kere J, Laivuori H. Gene expression profiling of pre-eclamptic placentae by RNA sequencing. Sci Rep. 2015;5. doi:10.1038/srep14107

64. Kang Q, Li W, Xiao J, Yu N, Fan L, Sha M, *et al.* Integrated analysis of multiple microarray studies to identify novel gene signatures in preeclampsia. Placenta. 2021;105: 104–118.

65. Choi S-J, Oh S-Y, Kim J-H, Sadovsky Y, Roh C-R. Increased expression of N-myc downstream-regulated gene 1 (NDRG1) in placentas from pregnancies complicated by intrauterine growth restriction or preeclampsia. Am J Obstet Gynecol. 2007;196: 45.e1–7.

66. Ben-Sahra I, Manning BD. mTORC1 signaling and the metabolic control of cell growth. Curr Opin Cell Biol. 2017;45: 72–82.

67. Han W, Hu C, Fan Z-J, Shen G-L. Transcript levels of keratin 1/5/6/14/15/16/17 as potential prognostic indicators in melanoma patients. Sci Rep. 2021;11: 1023.

68. Mori H, Saeki K, Chang G, Hsu P-Y, Wang J, Wu X, *et al.* Estrogen-Induced Cell Cycle Arrest and Apoptosis as an Unexpected Outcome of Aromatase Inhibitor-Resistance in ER+ Breast Cancer. 2021 [cited 1 Sep 2021]. doi:10.21203/rs.3.rs-236742/v1

69. Corcoran JJ, Nicholson C, Sweeney M, Charnock JC, Robson SC, Westwood M, *et*

al. Human uterine and placental arteries exhibit tissue-specific acute responses to 17β-estradiol and estrogen-receptor-specific agonists. Mol Hum Reprod. 2014;20: 433–441.

70. Zaren B, Lindmark G, Bakketeig L. Maternal smoking affects fetal growth more in the male fetus. Paediatr Perinat Epidemiol. 2000;14: 118–126.
71. 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.
72. Bailey KL, Bonasera SJ, Wilderdyke M, Hanisch BW, Pavlik JA, DeVasure J, *et al.* Aging causes a slowing in ciliary beat frequency, mediated by PKCε. Am J Physiol Lung Cell Mol Physiol. 2014;306: L584–9.
73. Deschênes M, Chabot B. The emerging role of alternative splicing in senescence and aging. Aging Cell. 2017;16: 918–933.
74. Parker MM, Chase RP, Lamb A, Reyes A, Saferali A, Yun JH, *et al.* RNA sequencing identifies novel non-coding RNA and exon-specific effects associated with cigarette smoking. BMC Med Genomics. 2017;10: 58.
75. Suryawanshi H, Morozov P, Straus A, Sahasrabudhe N, A Max KE, Garzia A, *et al.* A single-cell survey of the human first-trimester placenta and decidua. Science Advances. 2018;4. Available: http://advances.sciencemag.org/
76. Martinez F, Mansego ML, Escudero JC, Redon J, Chaves FJ. Association of a mineralocorticoid receptor gene polymorphism with hypertension in a Spanish population. Am J Hypertens. 2009;22: 649–655.
77. Cross JC. The genetics of pre-eclampsia: a feto-placental or maternal problem? Clin Genet. 2003;64: 96–103.
78. Roberts JM, Cooper DW. Pathogenesis and genetics of pre-eclampsia. Lancet. 2001;357: 53–56.
79. Roberts JM, Hubel CA. The two stage model of preeclampsia: variations on the theme. Placenta. 2009;30 Suppl A: S32–7.
80. Sarkar J, Seshadri V, Tripoulas NA, Ketterer ME, Fox PL. Role of ceruloplasmin in macrophage iron efflux during hypoxia. J Biol Chem. 2003;278: 44018–44024.
81. Nikolic A, Cabarkapa V, Novakov Mikic A, Jakovljević A, Stosic Z. Ceruloplasmin and antioxidative enzymes in pre-eclampsia. J Matern Fetal Neonatal Med. 2016;29:

2987–2993.

82. Bellos I, Papantoniou N, Pergialiotis V. Serum ceruloplasmin levels in preeclampsia: a meta-analysis. J Matern Fetal Neonatal Med. 2018;31: 2342–2348.

83. Sak S, Barut M, Çelik H, Incebiyik A, Ağaçayak E, Uyanikoglu H, *et al.* Copper and ceruloplasmin levels are closely related to the severity of preeclampsia. J Matern Fetal Neonatal Med. 2020;33: 96–102.

84. Vento-Tormo R, Efremova M, Botting RA, Turco MY, Vento-Tormo M, Meyer KB, *et al.* Single-cell reconstruction of the early maternal–fetal interface in humans. Nature. 2018;563: 347–353.

85. Aouache R, Biquard L, Vaiman D, Miralles F. Oxidative Stress in Preeclampsia and Placental Diseases. Int J Mol Sci. 2018;19. doi:10.3390/ijms19051496

86. Shin E-K, Lee D, Jeung E-B. The acyl-CoA dehydrogenase, very long chain (ACADVL) expression under hypoxic condition in human placental cell line (BeWo). World Congress on Reproductive Biology 2014. BioScientifica; 2014. doi:10.1530/repabs.1.P266

87. Shigunov P, Sotelo-Silveira J, Stimamiglio MA, Kuligovski C, Irigoín F, Badano JL, *et al.* Ribonomic analysis of human DZIP1 reveals its involvement in ribonucleoprotein complexes and stress granules. BMC Mol Biol. 2014;15: 12.

88. Wilson RL, Grieger JA, Bianco-Miotto T, Roberts CT. Association between Maternal Zinc Status, Dietary Zinc Intake and Pregnancy Complications: A Systematic Review. Nutrients. 2016;8. doi:10.3390/nu8100641

89. Schweigel-Röntgen M, Kolisek M. SLC41 transporters--molecular identification and functional role. Bevensee MO, editor. Curr Top Membr. 2014;73: 383–410.

90. Liao W-X, Laurent LC, Agent S, Hodges J, Chen D-B. Human placental expression of SLIT/ROBO signaling cues: effects of preeclampsia and hypoxia. Biol Reprod. 2012;86: 111.

91. Wang B, Xiao Y, Ding BB, Zhang N, Yuan XB, Gui L, *et al.* Induction of tumor angiogenesis by Slit-Robo signaling and inhibition of cancer growth by blocking Robo activity. Cancer Cell. 2003;4: 19–29.

92. McDonald MT, Papenberg KA, Ghosh S, Glatfelter AA, Biesecker BB, Helmbold EA, *et al.* A disease locus for hereditary haemorrhagic telangiectasia maps to chromosome 9q33-34. Nat Genet. 1994;6: 197–204.

93. Todd N, McNally R, Alqudah A, Jerotic D, Suvakov S, Obradovic D, *et al.* Role of A Novel Angiogenesis FKBPL-CD44 Pathway in Preeclampsia Risk Stratification and Mesenchymal Stem Cell Treatment. J Clin Endocrinol Metab. 2021;106: 26–41.

94. Zhang Y, Wang J, Ding Y, Zhang J, Xu Y, Xu J, *et al.* Migrasome and Tetraspanins in Vascular Homeostasis: Concept, Present, and Future. Front Cell Dev Biol. 2020;8: 438.

95. Costanzo V, Bardelli A, Siena S, Abrignani S. Exploring the links between cancer and placenta development. Open Biol. 2018;8. doi:10.1098/rsob.180081

96. Huang L, Li Y, Wang C, Li N, Hou Y, Chang R, *et al.* Overexpression of Collapsin Response Mediator Protein 1 Inhibits Human Trophoblast Cells Proliferation, Migration, and Invasion. Reprod Sci. 2019;26: 954–960.

97. Qiao C, Wang C, Jin F, Zheng D, Liu C. Expression of collapsin response mediator protein 1 in placenta of normal gestation and link to early-onset preeclampsia. Reprod Sci. 2015;22: 495–501.

98. Yang Y, Fang X, Yang R, Yu H, Jiang P, Sun B, *et al.* MiR-152 Regulates Apoptosis and Triglyceride Production in MECs via Targeting ACAA2 and HSD17B12 Genes. Sci Rep. 2018;8: 417.

99. Monsivais D, Dyson MT, Yin P, Coon JS, Navarro A, Feng G, *et al.* ERβ- and prostaglandin E2-regulated pathways integrate cell proliferation via Ras-like and estrogen-regulated growth inhibitor in endometriosis. Mol Endocrinol. 2014;28: 1304–1315.

100. Patel N, Itakura T, Gonzalez JM Jr, Schwartz SG, Fini ME. GPR158, an orphan member of G protein-coupled receptor Family C: glucocorticoid-stimulated expression and novel nuclear role. PLoS One. 2013;8: e57843.

101. Singh J, Ahmed A, Girardi G. Role of complement component C1q in the onset of preeclampsia in mice. Hypertension. 2011;58: 716–724.

102. Marx S, Dal Maso T, Chen J-W, Bury M, Wouters J, Michiels C, *et al.* Transmembrane (TMEM) protein family members: Poorly characterized even if essential for the metastatic process. Semin Cancer Biol. 2020;60: 96–106.

103. Lian IA, Langaas M, Moses E, Johansson Å. Differential gene expression at the maternal-fetal interface in preeclampsia is influenced by gestational age. PLoS One. 2013;8: e69848.

104. Reiss D, Zhang Y, Mager DL. Widely variable endogenous retroviral methylation

levels in human placenta. Nucleic Acids Res. 2007;35: 4743–4754.

105. Kosaka Y, Inoue H, Ohmachi T, Yokoe T, Matsumoto T, Mimori K, *et al.* Tripartite motif-containing 29 (TRIM29) is a novel marker for lymph node metastasis in gastric cancer. Ann Surg Oncol. 2007;14: 2543–2549.

106. Akutagawa O, Nishi H, Kyo S, Higuma C, Inoue M, Isaka K. Early growth response-1 mediates up-regulation of telomerase in placenta. Placenta. 2007;28: 920–927.

107. Bruchova H, Vasikova A, Merkerova M, Milcova A, Topinka J, Balascak I, *et al.* Effect of maternal tobacco smoke exposure on the placental transcriptome. Placenta. 2010;31: 186–191.

108. He P, Shao D, Ye M, Zhang G. Analysis of gene expression identifies candidate markers and pathways in pre-eclampsia. J Obstet Gynaecol. 2015;35: 578–584.

109. Maeda KJ, Showmaker KC, Johnson AC, Garrett MR, Sasser JM. Spontaneous superimposed preeclampsia: chronology and expression unveiled by temporal transcriptomic analysis. Physiol Genomics. 2019;51: 342–355.

110. Demir E, Sabatelli P, Allamand V, Ferreiro A, Moghadaszadeh B, Makrelouf M, *et al.* Mutations in COL6A3 cause severe and mild phenotypes of Ullrich congenital muscular dystrophy. Am J Hum Genet. 2002;70: 1446–1458.

111. Sabatelli P, Bonaldo P, Lattanzi G, Braghetta P, Bergamin N, Capanni C, *et al.* Collagen VI deficiency affects the organization of fibronectin in the extracellular matrix of cultured fibroblasts. Matrix Biol. 2001;20: 475–486.

112. Atkinson JC, Rühl M, Becker J, Ackermann R, Schuppan D. Collagen VI Regulates Normal and Transformed Mesenchymal Cell Proliferationin Vitro. Exp Cell Res. 1996;228: 283–291.

113. Rühl M, Johannsen M, Atkinson J, Manski D, Sahin E, Somasundaram R, *et al.* Soluble collagen VI induces tyrosine phosphorylation of paxillin and focal adhesion kinase and activates the MAP kinase erk2 in fibroblasts. Exp Cell Res. 1999;250: 548–557.

114. Akaogi K, Okabe Y, Sato J, Nagashima Y, Yasumitsu H, Sugahara K, *et al.* Specific accumulation of tumor-derived adhesion factor in tumor blood vessels and in capillary tube-like structures of cultured vascular endothelial cells. Proc Natl Acad Sci U S A. 1996;93: 8384–8389.

115. Januzzi JL Jr, Packer M, Claggett B, Liu J, Shah AM, Zile MR, *et al.* IGFBP7 (Insulin-Like Growth Factor-Binding Protein-7) and Neprilysin Inhibition in Patients

With Heart Failure. Circ Heart Fail. 2018;11: e005133.

116. Armanini D, Zennaro CM, Martella L, Scali M, Pratesi C, Grella PV, *et al.* Mineralocorticoid effector mechanism in preeclampsia. J Clin Endocrinol Metab. 1992;74: 946–949.

117. Geller DS, Farhi A, Pinkerton N, Fradley M, Moritz M, Spitzer A, *et al.* Activating mineralocorticoid receptor mutation in hypertension exacerbated by pregnancy. Science. 2000. pp. 119–123.

118. Gamper N, Fillon S, Feng Y, Friedrich B, Lang PA, Henke G, *et al.* K+ channel activation by all three isoforms of serum- and glucocorticoid-dependent protein kinase SGK. Pflugers Arch. 2002;445: 60–66.

119. Singh AK, Riederer B, Chen M, Xiao F, Krabbenhöft A, Engelhardt R, *et al.* The switch of intestinal Slc26 exchangers from anion absorptive to HCOFormula secretory mode is dependent on CFTR anion channel function. Am J Physiol Cell Physiol. 2010;298: C1057–65.

120. Schwab SG, Franke PE, Hoefgen B, Guttenthaler V, Lichtermann D, Trixler M, *et al.* Association of DNA polymorphisms in the synaptic vesicular amine transporter gene (SLC18A2) with alcohol and nicotine dependence. Neuropsychopharmacology. 2005;30: 2263–2268.

121. Uldry M, Ibberson M, Horisberger JD, Chatton JY, Riederer BM, Thorens B. Identification of a mammalian H(+)-myo-inositol symporter expressed predominantly in the brain. EMBO J. 2001;20: 4467–4477.

122. Gilbert JS, Ryan MJ, LaMarca BB, Sedeek M, Murphy SR, Granger JP. Pathophysiology of hypertension during preeclampsia: linking placental ischemia with endothelial dysfunction. Am J Physiol Heart Circ Physiol. 2008;294: H541–50.

123. Puschett JB. The role of excessive volume expansion in the pathogenesis of preeclampsia. Med Hypotheses. 2006;67: 1125–1132.

124. Plasencia W, Maiz N, Poon L, Yu C, Nicolaides KH. Uterine artery Doppler at 11 0 to 13 6 weeks and 21 0 to 24 6 weeks in the prediction of pre-eclampsia. Ultrasound in Obstetrics and Gynecology. 2008. pp. 138–146. doi:10.1002/uog.5402

125. Roberts JM, Rich-Edwards JW, McElrath TF, Garmire L, Myatt L, Global Pregnancy Collaboration. Subtypes of Preeclampsia: Recognition and Determining Clinical Usefulness. Hypertension. 2021;77: 1430–1441.

126. North RA, McCowan LME, Dekker GA, Poston L, Chan EHY, Stewart AW, *et al.*

Clinical risk prediction for pre-eclampsia in nulliparous women: development of model in international prospective cohort. BMJ. 2011;342: d1875.

127. Kenny LC, Black MA, Poston L, Taylor R, Myers JE, Baker PN, *et al.* Early pregnancy prediction of preeclampsia in nulliparous women, combining clinical risk and biomarkers: the Screening for Pregnancy Endpoints (SCOPE) international cohort study. Hypertension. 2014;64: 644–652.

128. Mohammed H, Roberts CT, Grzeskowiak LE, Giles LC, Dekker GA, Marshall HS. Safety and protective effects of maternal influenza vaccination on pregnancy and birth outcomes: A prospective cohort study. EClinicalMedicine. 2020;26: 100522.

129. Verburg PE, Roberts CT, McBean E, Mulder ME, Leemaqz S, Erwich JJHM, *et al.* Peripheral maternal haemodynamics across pregnancy in hypertensive disorders of pregnancy. Pregnancy Hypertens. 2019;16: 89–96.

130. McCowan LME, Roberts CT, Dekker GA, Taylor RS, Chan EHY, Kenny LC, *et al.* Risk factors for small-for-gestational-age infants by customised birthweight centiles: data from an international prospective cohort study. BJOG. 2010;117: 1599–1607.

131. Tranquilli AL, Dekker G, Magee L, Roberts J, Sibai BM, Steyn W, *et al.* The classification, diagnosis and management of the hypertensive disorders of pregnancy: A revised statement from the ISSHP. Pregnancy Hypertens. 2014;4: 97–104.

132. Robinson MD, McCarthy DJ, Smyth GK. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2009;26: 139–140.

133. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, *et al.* Limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43: e47.

134. Yip AM, Horvath S. Gene network interconnectedness and the generalized topological overlap measure. BMC Bioinformatics. 2007;8: 22.

135. Yu G, Wang L-G, Han Y, He Q-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS. 2012;16: 284–287.

Supplementary Information Legends

Note: Supplementary Figures and Tables are hosted on Figshare

(https://figshare.com/projects/Chapter_6_-_Supplementary_Files/122459)

Figure S1. Box plots indicate IQR (box), median (centre line), mean (centre diamond), upper and lower whiskers represent scores outside the middle 50% of centred $\log_2$ CPMs. Genes up-regulated in placenta from male bearing pregnancies complicated by PE (**i-xxi**) and genes down-regulated in placenta from male bearing pregnancies complicated by PE (**xxii-xxv**).

Figure 2. Gene set enrichment analyses (GSEA) of placenta genes associated with late-onset preeclampsia. Following differential expression analysis genes were ranked according to the product of the sign of the log fold change and the negative log 10 FDR (1). Hallmark pathways enriched with up-regulated genes (**i-iii**) and down-regulated genes (**iv**).

(1) GSEA rank = (sign(logFC))*(-$\log_{10}$(FDR))

Figure S3. Gene set enrichment analyses (GSEA) of placenta genes associated with late-onset preeclampsia. Following differential expression analysis genes were ranked according to the product of the sign of the log fold change and the negative log 10 FDR (1). Immunologic gene sets enriched with up-regulated genes (**i-iii**)

Figure S4. Intramodular connectivity is a measure of module membership (centrality). The red and lightgreen modules have the highest intramodular connectivity indicating that genes within these modules are highly connected and may be key drivers of placental biology.

Table S1. Full Gene Ontology (GO) pathways of WGCNA modules.

Table S2. Hallmark and KEGG pathways enrichment of WGCNA modules.

Chapter 7

Discussion and Future Directions

The placenta, that most essential yet transient organ, remains something of a medical mystery. In many cultures, the placenta has been seen as "mother*,* grandmother and sibling"*,* treated with reverence and prepared for ritual burial [1]. Today, modern research, such as that conducted throughout my candidature, attempts to understand the placenta, not only at a functional level, but to unravel its development and function down to the finest genomic scales. While the techniques and insights included in the following discussion are set in the contextual world of pregnancy and the placenta, the implications are much broader. Both micro- and messenger RNAs are ubiquitous in human tissues. Whilst the placenta has some special considerations including its unique sympatric relationship with both the fetus and the mother, its molecular functions can be probed using cross-disciplinary techniques. The methodological approach set out in Figure 1 illustrates the importance of defining a tissue specific regulatory landscape and the focus on technical considerations that are important to novel biomarker discovery. This strategy is essential, not only to our understanding of the placenta and its role in both maternal and fetal health, but is also applicable to research across precision medicine and precision health.

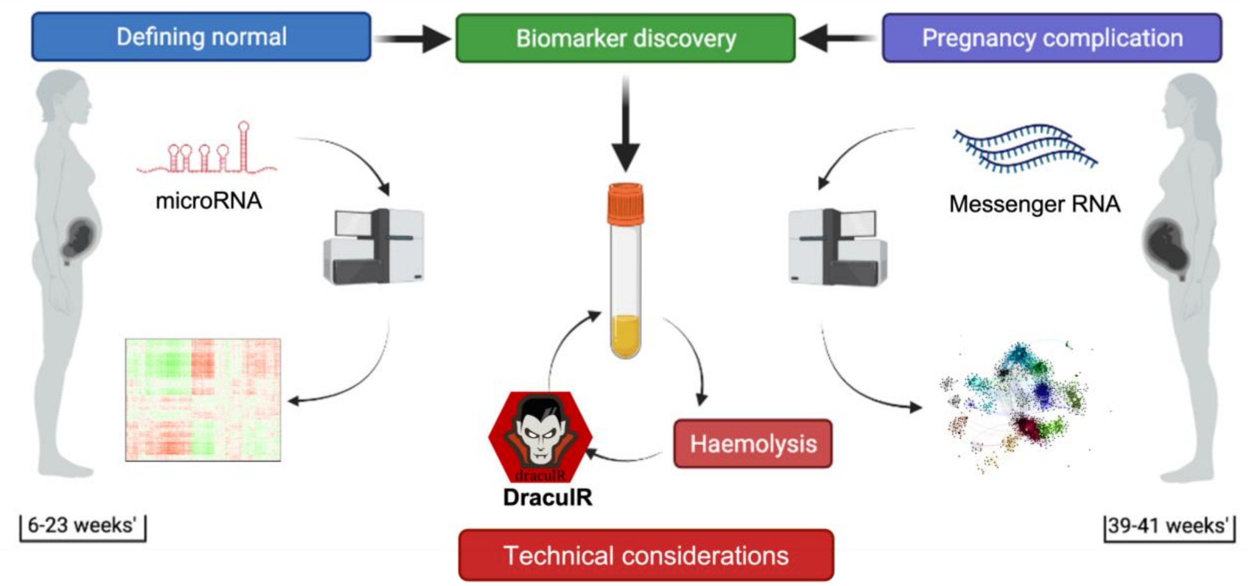


Figure 1. In the early stages of my PhD I characterised a comprehensive profile of miRNAs in early to mid-gestation (6-23 weeks') placenta and matched maternal plasma (Left). This initial miRNA characterisation led to the identification of red blood cell inclusion within the maternal plasma samples. This issue prompted the creation of a new Shiny/web based tool *DraculR*, designed to enable detection of red blood cell inclusion in miRNA sequencing from plasma (Centre). Bulk RNA sequencing was used to identify genes highly correlated with late onset preeclampsia (Right). (Created with Biorender).

This research in this thesis is divided into a number of experimental chapters that investigate miRNA and mRNA expression in the placenta and their technical biases that impact biomarker discovery for pregnancy monitoring (Figure 1). Chapter 2 contains the comprehensive analysis of miRNAs in placenta and matched maternal plasma, using high throughput sequencing data, from our biobank of tissue representing presumably 'normal' human pregnancy from 6-23 weeks' gestation. Importantly, we demonstrated that placental miRNAs are present in maternal plasma, and indicative of oxygen transfer across the placenta-maternal axis. Technical limitations of maternal plasma data identified in Chapter 2 were then further explored in Chapters 3 and 4. These chapters detail a new method developed specifically to detect red blood cell inclusion in miRNA sequencing data from human plasma, and then provides a tool freely available to the research community for this purpose.

After defining this ‘normal’ profile of miRNA expression, and identifying whether maternal plasma could reflect the placental profile, I then shifted to investigating gene expression (mRNA-seq) changes identified from complicated pregnancies (Figure 1), in this case late onset preeclampsia in placenta. Importantly, the results of this chapter highlight the need for sex-specific analysis of the placenta and provide intriguing insights into the varied etiological pathways we need to understand in order to generate true individualised medical insight into preeclampsia. Further, they provide a novel 25 gene set which, through future investigation, will provide further insight into the complex and dynamic etiology of late onset preeclampsia. Interestingly, it highlights the importance of anaerobic metabolic processes as well as oxidative stress in this disease. This work was part of a wider study aiming to establish a multi-omic placental dataset of global importance, and led to the detection of important technical limitations in novel biomarker discovery in the context of pregnancy [2,3].

The regulatory profile of placenta in early to mid gestation is dynamic

As the most comprehensive investigation of placenta miRNAs to date, Chapter 3 provides the most detailed profile of early to mid-gestation placental villus miRNAs in human tissue and matched maternal plasma produced to date. These analyses were undertaken as part of a wider project aimed at novel biomarker discovery. However, before we are able to investigate biomarkers of health and disease we first need to profile what we can consider to be the “normal” signature of placental development. A “normal” profile in this study was determined through the use of important samples taken from women undergoing surgical, elective pregnancy terminations, enabling me to specifically identify miRNAs that were expressed within the placenta within a presumably uncomplicated environment. Through careful analysis using conservative protocols developed in cancer studies, I identified a clear signature of gestational age across samples (using unsupervised clustering techniques) and found that the introduction of oxygenated maternal blood into the intervillous space, roughly between ~10-12 weeks’ gestation [4], fundamentally changed the miRNA profile of the chorionic villus, particularly in placenta-specific miRNA clusters. I identified a clear gestational age group bias in miRNA clusters C19MC, C14MC, miR-17~92 and paralogs with a general decrease in expression seen in C19MC and miR-17~92, and a general increase in expression seen in C14MC members. These clusters expressed either exclusively, or preferentially in placenta [5–7] provide excellent biomarker potential.

Biomarker potential aside, the level of miRNA-derived regulatory information described in this chapter represents a considerable resource for placental biologists. The unique transcriptional and epigenetic profiles of the placenta are vastly different from most human tissues and often are far more similar to samples taken from human cancers [8–10]. Structures such as Partially Methylated Domains (PMDs) for example, are found throughout the placental genome [11–13] and are the hallmark of a unique regulatory profile. While work described here is not intended to address the specific impact of miRNAs on matched mRNA-seq samples from the sample timepoints, as described in

Breen *et al.* [2], the data offers a significant opportunity to identify how placental miRNAs target specific genes.

The identification of a novel disease biomarker in human clinical data is strongly impacted upon by both the quality/technically-derived differences between samples. This may highlight the limitations of miRNA and RNA sequencing data in the pursuit of minimally or non-invasive biomarkers of pregnancy health. My analyses that seek to accurately profile placental health and development through analysis of placental nucleic acids identified in maternal plasma is driven by this need to inform clinical interventions. However, even in single cell sequencing where there is a focus on a single sample at extremely high resolution, the focus is on the implications of gene expression implications on cell to cell signalling, functionally distinct cell types or cellular differentiation [14–16]. Establishing this baseline of normal is crucial for determining patterns of disease as placental insufficiency, induced by aberrant placental development, is implicated in many complications of pregnancy [4,17]. While Chapter 2 focussed on defining what is “normal” in the context of human placenta and maternal plasma miRNAs, the need for medical intervention arises when pregnancy complications occur. Chapter 5 moves away from normal pregnancy and into the context of pregnancy complications, in particular, late onset preeclampsia in which we used sex-specific and sex-naive approaches to identify genes and pathways in disease pathogenesis and response.

Technical biases can impact the development of biomarkers for precision medicine

MicroRNA expression changes during the development of the placenta between 6-23 weeks' gestation and the effects observed around the rise in oxygen tension at the introduction of oxygenated maternal blood described in detail (see Chapter 3), whilst not previously described in such detail, is not in of itself surprising. Nor is the discovery of fetally-derived molecules in the maternal circulation [18,19]. Recent work by several groups has focussed on the expression level of highly placenta associated miRNA clusters from chromosomes 14 and 19 [7,20–22]. Here we were able to demonstrate in placenta that the proportion of highly placenta associated miRNAs changed at the introduction of oxygenated maternal blood and more importantly, that this proportional change was reflected in maternal plasma. Unfortunately, sample quality issues associated with haemolysis failed to allow the depth of analysis planned and it is evident given the limitations identified here that the bioinformatics community requires more and more sophisticated tools to support our common goals [23,24]. Although limiting my initial research findings, this issue led to the creation of a new bioinformatics tool (Chapter 4), highlighting the complexities of working with clinical fluids and has inspired future research efforts to overcome this limitation.

While research into the effects of haemolysis on miRNA quantification and normalisation continues, there is no current, publicly available method for the *in silico* identification of haemolysis in high-throughput sequencing data. Given that red blood cells are known miRNA repositories, without a robust method of haemolysis detection, measurements of miRNA abundance as disease biomarkers are limited [25]. Previous research has identified haemolysis as an issue in biomarker discovery [25–27]. The myriad of changes identified in maternal plasma miRNAs in association with haemolysis highlight the importance of quality primary data and the need to understand where data complexity originates. After filtering for low expression, we identified 138 miRNAs with a higher abundance in haemolysed compared to non-haemolysed samples (FDR <0.05: $\log_2$FC >0). Importantly, we identified miR-451, the sole haemolysis signature miRNA used in the current ΔCq (miR-23a-miR-451) gold standard method for haemolysis detection [25,28], to be highly correlated with pregnancy status. Starting early in pregnancy, maternal haemodynamics change. This change includes an increase in plasma volume, and red blood cell mass [29,30] with the former increasing proportionally more than the latter, with resultant hemodilution [31]. This means that the gold standard for haemolysis detection using miRNA abundance in plasma is therefore not applicable to pregnant women. The tool developed here (detailed in Chapter 4) uses a 20 miRNA signature with the larger miRNA set allowing for the removal of individual or multiple miRNAs that are known or suspected to be affected by a disease or condition. This in turn gives greater confidence in haemolysis detection across a wide range of research questions.

Deriving this haemolysis signature is additionally important to a wide array of clinical applications involving plasma samples. To facilitate the extended use of the technique we also present a Shiny/R web-based application DraculR which can be used to identify haemolysis contamination in human plasma samples. DraculR was specifically designed to give the user a simple tool for inspection of miRNA sequencing data from human plasma extraction, allowing non-computational wet lab and clinical researchers to identify potential haemolysis in a sample before proceeding further with downstream bioinformatics analyses. Using an existing matrix of miRNA counts taken from high throughput sequencing, DraculR returns a simple Haemolysis Metric and provides a clear recommendation regarding the downstream analysis of each of the samples being considered. A simple, but important, quality check is rarely reported in the literature relating to miRNAs in human plasma or serum and its implementation using this or other approaches will improve the quality of research in this field. As discussed previously, we discovered that miR-451, the haemolysis associated miRNA currently used in the ΔCq (miR-23a-miR-451) gold standard method for haemolysis detection, varies as maternal haemodynamics change throughout pregnancy [32]. Likewise, haemodynamic changes may alter results obtained using spectrophotometry for haemolysis detection in a pregnancy context. Whilst this highlights the need for alternative measures, for example the use of a modifiable signature set, it also highlights a limitation of this method as all samples used in the identification of signature miRNA were taken from women of child bearing age. Given this, the method we derived for the assessment of haemolysis detection may not generalise outside of this group. Future work in this area is planned and will include the incorporation of plasma samples from male and female individuals and across a wider age range. Serial dilutions in a similar context to the method used by Shah *et al.* [33] are planned to validate the method proposed here with the intention to make the DraculaR tool freely available to the research community.

What are the implications for personalised treatment in obstetric healthcare?

When thinking about pregnancy health, the placenta and its interaction with maternal physiology is arguably the most important component. Whilst the fetal and maternal bloodstreams do not mix, interactions are mediated between these compartments via components such as oxygen tension, hormones and extracellular vesicles (e.g. exosomes, microvesicles and apoptotic bodies [34]). As detailed in Chapter 3 and 4, this cross-talk provides a potential source of biomarkers for assessing placental health using biological material obtained from the mother. The best approach to biomarker discovery remains to be decided. However, the use of maternal plasma for the realtime assessment of placental health and function may profoundly impact an individual woman's obstetric care. In Chapter 3 I focussed on the utility of miRNA sequencing of small RNA fragments extracted from maternal plasma.

Whilst a placenta signal was present within maternal blood, in particular that of placenta associated miRNA clusters from chromosomes 14 and 19, additional genomics-approaches could be used to identify placenta-specific patterns within maternal blood. With a change to miRNA and mRNA sample preparations [35,36], the enrichment for placenta-derived exosomes may provide greater clarity. Exosomes are derived from placental specific cells and exported into maternal blood, making them perfect targets for biomarker discovery [35–38]. In particular, exosomal enrichment would eliminate the issue of red blood cell content inclusion identified in Chapter 4. Additionally, a number of studies in reproductive health have utilised the application of cell-free DNA (cfDNA) to determine tissue-specific biomarkers [39,40]. Using this approach, genomic DNA or DNA methylation patterns identified from genomic DNA are shed into blood which are then able to be identified and linked to the original cell-type or tissue. The utility of this particular approach, compared to miRNAs, is the specificity to the tissue or even cell-type whereas outside of miRNA clusters that are highly specific to the placenta (e.g. C19MC and C14MC miRNA clusters), the presence of a singular miRNA or mRNA could be tied to any tissue. DNA extraction and library preparation of cfDNA is simple and can be applied in a high-throughput setting with a peripheral blood sample [41–43]. Despite this cfDNA and cfDNA methylation approaches can only distinguish genetic markers rather than identifying an expressed component, so might be only applicable for diseases with a high genetic component.

We found the transcriptional landscape of the human placenta to be remarkably conserved in the context of preeclamptic, uncomplicated, male- and female-bearing pregnancy. But perhaps this is to be expected given that in all instances the organ is still performing the main function of the placenta, the transfer of oxygen, nutrients, hormones and wastes, albeit less efficiently in the case of preeclampsia [44,45]. Importantly, the results of Chapter 5 indicate the presence of biologically relevant metabolic, oxidative stress and immune gene expression responses to preeclampsia in both sex-specific and sex-naive analyses. This is consistent with findings from previous research which identified many of the placental genes demonstrating sexual dimorphism, for example in the context of maternal asthma, to be associated with growth, inflammation and immune pathways [46].

With the implementation of first DNA and then RNA (cDNA library) high throughput sequencing technologies [47–49], biomedical science has become increasingly data-driven [50]. This is true today with the advent of new complementary RNA sequencing technologies such as single cell transcriptomics and spatial transcriptomics [51,52]. To be human interpretable, this high-dimensional data requires dimensional reduction and network-based techniques, such as gene co-expression analysis [53,54]. Weighted gene co-expression analysis (WGCNA) constructs modules of genes with highly correlated expression patterns which can be used to identify biologically relevant molecular pathways and modules correlations with sample characteristics using module eigengenes [55,56]. Using a sex-naive approach we performed weighted gene co-expression analysis to identify correlated gene modules using RNA sequencing gene expression data from term placenta. My analysis identified 13 gene co-expression modules, one of which was significantly correlated with late-onset preeclampsia. When investigating a disease characterised by placental perfusion and oxidative stress, enrichment for hypoxia and anaerobic metabolic process pathways is particularly relevant [57,58]. Interestingly the LOPE signature module was enriched for genes in pathways associated with oxidative stress and glucose metabolism, previously implicated in both early- and late-onset preeclampsia [59–62].

Confluence in the diversity of clinical presentation and many lines of scientific enquiry converge on the consensus that preeclampsia occurs through multiple pathophysiological pathways and is not one single disorder [63,64]. Whilst late-onset preeclampsia is often considered to be a maternal disorder [65], it is important to remember that in this, and most other subtypes of preeclampsia, the placenta is both necessary and sufficient to cause the disease, and delivery of the placenta is the only curative treatment [66]. Targeted studies such as the one presented here and others [46,58,67] are designed to interrogate how characteristics such as fetal sex or disease subtype impact the molecular pathways implicated in disease pathogenesis. I argue that in the age of personalised medicine we must focus on the diversity of presentations to understand the importance of our findings in the pathophysiology of this complex disorder. Because of this, further studies elucidating the molecular mechanisms of preeclampsia from the perspective of the placenta are essential and will provide novel opportunities for clinical intervention.

Sex stratification in particular had interesting implications for genes such as zinc finger protein 300 (*ZNF300*) previously shown to be more lowly expressed in female placenta and to indicate lower preeclampsia risk [68,69] and Collagen Type VI Alpha 3 Chain (*COL6A3*) previously found to not be dysregulated by preeclampsia [70,71] although both *COL6A1* and *COL1A1* were implicated in early onset preeclampsia again highlighting the multi pathway nature of preeclampsia pathogenesis. That my findings are somewhat contrary to those of others has many implications, but highlights two main concepts. Firstly, in the placenta, sex matters (as described further in Chapter 2) and secondly, accurate subtyping of placental disease should be an essential consideration; a feedforward loop whereby research identifies finer gradations of molecular signatures of disease, which further informs the search for interventions in clinical practice. The strengths of my approach included 1) the inclusion of a single preeclampsia subtype; 2) stratification by fetal sex; 3) tissue specific optimisation of numerous data analyses approaches [53,72,73] and 4) my access to data from a large number of human samples.

Due to the difficulty and risk associated with obtaining human placenta samples across the ~40 weeks of human gestation, research is often conducted in immortalised cell lines or animal models. However, as so aptly put by Schmidt (2015) [74], only humans have a human placenta. While important for molecular work, cell lines bring with them limitations in the form of karyotypic abnormalities and aberrant expression when compared to whole tissue [7,75]. To ameliorate these research limitations, all placental work here was conducted using chorionic villus samples taken directly from human placental tissue. This tissue is obtained either after the elective termination of pregnancy, or after delivery of the neonate, and is obtained after informed consent.

Overall, these chapters have described a comprehensive effort to characterise miRNAs in human placenta and maternal plasma, identified and attempted to resolve technical limitations in biomarker discovery and provided a unique sex-specific investigation into the molecular signature of late onset preeclampsia. The results further our understanding of the effect of oxygen and late onset preeclampsia in human placenta at the molecular level and are a valuable resource for hypothesis generation for future work. Additionally we demonstrated that technically rigorous, bioinformatics-focussed strategies are instrumental in elucidating unique molecular insights in complex datasets taken from clinical samples. Collectively, the work here represents an initial contribution towards the ultimate goal of acquiring the depth of understanding required to provide truly individual therapeutic benefits based on a precision medicine model using a data-driven, bioinformatics-led approach.

References

1. Croft Long E. The Placenta in Lore and Legend. Bull Med Libr Assoc. [cited 1 Sep 2021]. Available: https://core.ac.uk/download/pdf/74316194.pdf

2. Breen J, McAninch D, Jankovic-Karasoulos T, McCullough D, Smith MD, Bogias KJ, *et al.* Temporal placental genome wide expression profiles reflect three phases of utero-placental blood flow during early to mid human gestation. bioRxiv. medRxiv; 2020. doi:10.1101/2020.06.25.20139709

3. Wan Q, Yiner-Lee Leemaqz S, Pederson SM, McCullough D, McAninch DC, Jankovic-Karasoulos T, *et al.* Quality control measures for placental sample purity in DNA methylation array analyses. Placenta. 2019;88: 8–11.

4. Jauniaux E, Watson AL, Hempstock J, Bao YP, Skepper JN, Burton GJ. Onset of maternal arterial blood flow and placental oxidative stress: A possible factor in human early pregnancy failure. Am J Pathol. 2000;157: 2111–2122.

5. Glazov EA, McWilliam S, Barris WC, Dalrymple BP. Origin, evolution, and biological role of miRNA cluster in DLK-DIO3 genomic region in placental mammals. Mol Biol Evol. 2008;25: 939–948.

6. Luo S-S, Ishibashi O, Ishikawa G, Ishikawa T, Katayama A, Mishima T, *et al.* Human Villous Trophoblasts Express and Secrete Placenta-Specific MicroRNAs into Maternal Circulation via Exosomes1. Biol Reprod. 2009;81: 717–729.

7. Morales-Prieto DM, Ospina-Prieto S, Chaiwangyen W, Schoenleben M, Markert UR. Pregnancy-associated miRNA-clusters. J Reprod Immunol. 2013;97: 51–61.

8. Koslowski M, Sahin U, Mitnacht-Kraus R, Seitz G, Huber C, Türeci O. A placenta-specific gene ectopically activated in many human cancers is essentially involved in malignant cell processes. Cancer Res. 2007;67: 9528–9534.

9. Costanzo V, Bardelli A, Siena S, Abrignani S. Exploring the links between cancer and placenta development. Open Biol. 2018;8. doi:10.1098/rsob.180081

10. Vidal DO, Ramão A, Pinheiro DG, Muys BR, Lorenzi JCC, de Pádua Alves C, *et al.* Highly expressed placental miRNAs control key biological processes in human cancer cell lines. Oncotarget. 2018;9: 23554–23563.

11. Schroeder DI, LaSalle JM. How has the study of the human placenta aided our understanding of partially methylated genes? Epigenomics. 2013;5: 645–654.

12. Schroeder DI, Blair JD, Lott P, Yu HOK, Hong D, Crary F, *et al.* The human placenta methylome. Proc Natl Acad Sci U S A. 2013;110: 6037–6042.

13. Nordor AV, Nehar-Belaid D, Richon S, Klatzmann D, Bellet D, Dangles-Marie V, *et al.* The early pregnancy placenta foreshadows DNA methylation alterations of solid tumors. Epigenetics. 2017;12: 793–803.

14. Pavličev M, Wagner GP, Chavan AR, Owens K, Maziarz J, Dunn-Fletcher C, *et al.* Single-cell transcriptomics of the human placenta: inferring the cell communication network of the maternal-fetal interface. Genome Res. 2017;27: 349–361.

15. Nelson AC, Mould AW, Bikoff EK, Robertson EJ. Single-cell RNA-seq reveals cell type-specific transcriptional signatures at the maternal-foetal interface during pregnancy. Nat Commun. 2016;7: 11414.

16. Vento-Tormo R, Efremova M, Botting RA, Turco MY, Vento-Tormo M, Meyer KB, *et al.* Single-cell reconstruction of the early maternal–fetal interface in humans. Nature. 2018;563: 347–353.

17. Ilekis JV, Tsilou E, Fisher S, Abrahams VM, Soares MJ, Cross JC, *et al.* Placental origins of adverse pregnancy outcomes: potential molecular targets: an Executive Workshop Summary of the Eunice Kennedy Shriver National Institute of Child Health and Human Development. American Journal of Obstetrics and Gynecology. Mosby; 2016. pp. S1–S46.

18. Hui WWI, Jiang P, Tong YK, Lee W-S, Cheng YKY, Maria I New, *et al.* Universal Haplotype-Based Noninvasive Prenatal Testing for Single Gene Diseases. Clinical Chemistry. 2017. pp. 513–524. doi:10.1373/clinchem.2016.268375

19. Lo YMD, Chan KCA, Sun H, Chen EZ, Jiang P, Lun FMF, *et al.* Maternal plasma DNA sequencing reveals the genome-wide genetic and mutational profile of the fetus. Sci Transl Med. 2010;2: 61ra91.

20. Wommack JC, Trzeciakowski JP, Miranda RC, Stowe RP, Ruiz RJ. Micro RNA clusters in maternal plasma are associated with preterm birth and infant outcomes. PLoS One. 2018;13: e0199029.

21. Inno R, Kikas T, Lillepea K, Laan M. Coordinated Expressional Landscape of the Human Placental miRNome and Transcriptome. Front Cell Dev Biol. 2021;9: 697947.

22. Pastuschek J, Nonn O, Gutiérrez-Samudio RN, Murrieta-Coxca JM, Müller J, Sanft J, *et al.* Molecular characteristics of established trophoblast-derived cell lines. Placenta. 2021;108: 122–133.

23. Stein L. Creating a bioinformatics nation. Nature. 2002;417: 119–120.

24. Ison J, Rapacki K, Ménager H, Kalaš M, Rydza E, Chmura P, *et al.* Tools and data services registry: a community effort to document bioinformatics resources. Nucleic Acids Res. 2016;44: D38–47.

25. Kirschner MB, Kao SC, Edelman JJ, Armstrong NJ, Vallely MP, van Zandwijk N, *et al.* Haemolysis during Sample Preparation Alters microRNA Content of Plasma. Pfeffer S, editor. PLoS One. 2011;6: e24145.

26. Kirschner MB, van Zandwijk N, Reid G. Cell-free microRNAs: potential biomarkers in need of standardized reporting. Front Genet. 2013;4: 56.

27. Pritchard CC, Kroh E, Wood B, Arroyo JD, Dougherty KJ, Miyaji MM, *et al.* Blood cell origin of circulating microRNAs: a cautionary note for cancer biomarker studies. Cancer Prev Res . 2012;5: 492–497.

28. Livak KJ, Schmittgen TD. Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the $2^{-\Delta\Delta CT}$ Method. Methods. 2001. pp. 402–408. doi:10.1006/meth.2001.1262

29. Pritchard JA. CHANGES IN THE BLOOD VOLUME DURING PREGNANCY AND DELIVERY. Anesthesiology. 1965;26: 393–399.

30. Chesley LC. Plasma and red cell volumes during pregnancy. Am J Obstet Gynecol. 1972;112: 440–450.

31. Sanghavi M, Rutherford JD. Cardiovascular physiology of pregnancy. Circulation. 2014;130: 1003–1008.

32. Verburg PE, Roberts CT, McBean E, Mulder ME, Leemaqz S, Erwich JJHM, *et al.* Peripheral maternal haemodynamics across pregnancy in hypertensive disorders of pregnancy. Pregnancy Hypertens. 2019;16: 89–96.

33. Shah JS, Soon PS, Marsh DJ. Comparison of methodologies to detect low levels of hemolysis in serum for accurate assessment of serum microRNAs. PLoS One. 2016. doi:10.1371/journal.pone.0153200

34. Buca D, Bologna G, D'Amico A, Cugini S, Musca F, Febbo M, *et al.* Extracellular Vesicles in Feto-Maternal Crosstalk and Pregnancy Disorders. Int J Mol Sci. 2020;21. doi:10.3390/ijms21062120

35. Tang MKS, Wong AST. Exosomes: Emerging biomarkers and targets for ovarian cancer. Cancer Lett. 2015;367: 26–33.

36. Li H, Ouyang Y, Sadovsky E, Parks WT, Chu T, Sadovsky Y. Unique microRNA Signals in Plasma Exosomes from Pregnancies Complicated by Preeclampsia. Hypertension. 2020;75: 762–771.

37. Pan BT, Johnstone RM. Fate of the transferrin receptor during maturation of sheep reticulocytes in vitro: selective externalization of the receptor. Cell. 1983;33: 967–978.

38. Johnstone RM, Adam M, Hammond JR, Orr L, Turbide C. Vesicle formation during reticulocyte maturation. Association of plasma membrane activities with released vesicles (exosomes). J Biol Chem. 1987;262: 9412–9420.

39. Caggiano C, Celona B, Garton F, Mefford J, Black BL, Henderson R, *et al.* Comprehensive cell type decomposition of circulating cell-free DNA with CelFiE. Nat Commun. 2021;12: 2717.

40. Jensen TJ, Kim SK, Zhu Z, Chin C, Gebhard C, Lu T, *et al.* Whole genome bisulfite sequencing of cell-free DNA and its cellular contributors uncovers placenta hypomethylated domains. Genome Biol. 2015;16: 78.

41. Maron JL, Johnson KL, Bianchi DW. Cell-Free Fetal DNA Plasma Extraction and Real-Time Polymerase Chain Reaction Quantification. Methods in Molecular Medicine^{TM}. 2007. pp. 51–63. doi:10.1007/978-1-59745-298-4_6

42. Devonshire AS, Whale AS, Gutteridge A, Jones G, Cowen S, Foy CA, *et al.* Towards standardisation of cell-free DNA measurement in plasma: controls for extraction efficiency, fragment size bias and quantification. Anal Bioanal Chem. 2014;406: 6499–6512.

43. Pedini P, Graiet H, Laget L, Filosa L, Chatron J, Cherouat N, *et al.* Qualitative and quantitative comparison of cell-free DNA and cell-free fetal DNA isolation by four (semi-)automated extraction methods: impact in two clinical applications: chimerism quantification and noninvasive prenatal diagnosis. Journal of Translational Medicine. 2021. doi:10.1186/s12967-020-02671-8

44. Gude NM, Roberts CT, Kalionis B, King RG. Growth and function of the normal human placenta. Thromb Res. 2004;114: 397–407.

45. Walker CK, Krakowiak P, Baker A, Hansen RL, Ozonoff S, Hertz-Picciotto I. Preeclampsia, placental insufficiency, and autism spectrum disorder or developmental delay. JAMA Pediatr. 2015;169: 154–162.

46. Osei-Kumah A, Smith R, Jurisica I, Caniggia I, Clifton VL. Sex-specific differences in placental global gene expression in pregnancies complicated by asthma. Placenta. 2011;32: 570–578.

47. Maxam AM, Gilbert W. A new method for sequencing DNA. Proc Natl Acad Sci U S

A. 1977;74: 560–564.

48. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A. 1977;74: 5463–5467.

49. Lowe R, Shirley N, Bleackley M, Dolan S, Shafee T. Transcriptomics technologies. PLoS Comput Biol. 2017;13: e1005457.

50. National Research Council, Division on Earth and Life Studies, Board on Life Sciences, Committee on A Framework for Developing a New Taxonomy of Disease. Toward Precision Medicine: Building a Knowledge Network for Biomedical Research and a New Taxonomy of Disease. National Academies Press; 2012.

51. Aldridge S, Teichmann SA. Single cell transcriptomics comes of age. Nat Commun. 2020;11: 4307.

52. Rao A, Barkley D, França GS, Yanai I. Exploring tissue architecture using spatial transcriptomics. Nature. 2021;596: 211–220.

53. Yip AM, Horvath S. Gene network interconnectedness and the generalized topological overlap measure. BMC Bioinformatics. 2007;8: 22.

54. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008;9: 559.

55. Cox B, Tsamou M, Vrijens K, Neven KY, Winckelmans E, de Kok TM, *et al.* A Co-expression Analysis of the Placental Transcriptome in Association With Maternal Pre-pregnancy BMI and Newborn Birth Weight. Front Genet. 2019;10: 354.

56. Li X, He Y, Hao C, Li X, Li X. Weighted gene correlation network analysis reveals novel regulatory modules associated with recurrent early pregnancy loss. Biosci Rep. 2020;40. doi:10.1042/BSR20193938

57. Guerby P, Tasta O, Swiader A, Pont F, Bujold E, Parant O, *et al.* Role of oxidative stress in the dysfunction of the placental endothelial nitric oxide synthase in

preeclampsia. Redox Biol. 2021;40: 101861.

58. Liu J, Song G, Meng T, Zhao G. Identification of Differentially Expressed Genes and Signaling Pathways in Placenta Tissue of Early-Onset and Late-Onset Pre-Eclamptic Pregnancies by Integrated Bioinformatics Analysis. Medical Science Monitor. 2020. doi:10.12659/msm.921997

59. Joffe GM, Esterlitz JR, Levine RJ, Clemens JD, Ewell MG, Sibai BM, *et al.* The relationship between abnormal glucose tolerance and hypertensive disorders of pregnancy in healthy nulliparous women. Calcium for Preeclampsia Prevention (CPEP) Study Group. Am J Obstet Gynecol. 1998;179: 1032–1037.

60. Weissgerber TL, Mudd LM. Preeclampsia and diabetes. Curr Diab Rep. 2015;15: 9.

61. Aouache R, Biquard L, Vaiman D, Miralles F. Oxidative Stress in Preeclampsia and Placental Diseases. Int J Mol Sci. 2018;19. doi:10.3390/ijms19051496

62. Taravati A, Tohidi F. Comprehensive analysis of oxidative stress markers and antioxidants status in preeclampsia. Taiwan J Obstet Gynecol. 2018;57: 779–790.

63. Burton GJ, Redman CW, Roberts JM, Moffett A. Pre-eclampsia: pathophysiology and clinical implications. BMJ. 2019;366: l2381.

64. Roberts JM, Bell MJ. If we know so much about preeclampsia, why haven’t we cured the disease? Journal of Reproductive Immunology. 2013. pp. 1–9. doi:10.1016/j.jri.2013.05.003

65. Raymond D, Peterson E. A critical review of early-onset and late-onset preeclampsia. Obstet Gynecol Surv. 2011;66: 497–506.

66. Powe CE, Levine RJ, Karumanchi SA. Preeclampsia, a disease of the maternal endothelium: the role of antiangiogenic factors and implications for later cardiovascular disease. Circulation. 2011;123: 2856–2869.

67. Herzog EM, Eggink AJ, Reijnierse A, Kerkhof MAM, de Krijger RR, Roks AJM, *et al.*

Impact of early- and late-onset preeclampsia on features of placental and newborn vascular health. Placenta. 2017;49: 72–79.

68. Xue L, Qiu H, Ma J, Guo M, Li W. ZNF300, a recently identified human transcription factor, activates the human IL-2Rβ promoter through the overlapping ZNF300/EGR1 binding site. Cell Mol Biol Lett. 2010;15: 530–540.

69. Akutagawa O, Nishi H, Kyo S, Higuma C, Inoue M, Isaka K. Early growth response-1 mediates up-regulation of telomerase in placenta. Placenta. 2007;28: 920–927.

70. Maeda KJ, Showmaker KC, Johnson AC, Garrett MR, Sasser JM. Spontaneous superimposed preeclampsia: chronology and expression unveiled by temporal transcriptomic analysis. Physiol Genomics. 2019;51: 342–355.

71. He P, Shao D, Ye M, Zhang G. Analysis of gene expression identifies candidate markers and pathways in pre-eclampsia. J Obstet Gynaecol. 2015;35: 578–584.

72. Robinson MD, McCarthy DJ, Smyth GK. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2009;26: 139–140.

73. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, *et al.* Limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43: e47.

74. Schmidt A, Morales-Prieto DM, Pastuschek J, Fröhlich K, Markert UR. Only humans have human placentas: Molecular differences between mice and humans. J Reprod Immunol. 2015. doi:10.1016/j.jri.2015.03.001

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