The University of Adelaide

Doctoral Thesis

An Integrative Analysis of the Human Placental Transcriptome

Author:
Sam Buckberry

Primary Supervisor:
Professor Claire Roberts

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

in the

School of Paediatrics and Reproductive Health
Discipline of Obstetrics and Gynaecology

July 2015

Declaration of Authorship

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.

I give consent to this copy of my thesis when deposited in the University Library, being made available for loan and photocopying, subject to the provisions of the Copyright Act 1968.

The author acknowledges that copyright of published works contained within this 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.

Signed: _____

Sam Buckberry

Date: _____

Dedicated to Nana, Mum and Kylie
Three special ladies who have always looked after me

THE UNIVERSITY OF ADELAIDE

Abstract

Discipline of Obstetrics and Gynaecology

Doctor of Philosophy

An Integrative Analysis of the Human Placental Transcriptome

by Sam Buckberry

Pregnancy outcome is inextricably linked to placental development, which is strictly regulated both temporally and spatially by mechanisms that are only partially understood. Although the placenta is absolutely indispensable for fetal development *in utero*, it remains the least understood human tissue. Although the placenta is a shared organ between the mother and fetus, it is of embryonic origin, and therefore its development is largely regulated by the fetal genome.

This overall goal of this research was to investigate three key aspects of human placental gene regulation: (1) The effect of genomic imprinting on gene regulation, (2) the differences in placental gene expression between the sexes, and (3) the co-expression relationships that exist between genes on a transcriptome scale.

Firstly, this research identified a window of epigenetic imprinting plasticity for the long non-coding RNA *H19*, which is heavily implicated in placental development and function. These results suggested that variation in *H19* imprinting may contribute to early programming of placental phenotype and highlighted the need for quantitative and robust methodologies to further elucidate the role of imprinted genes in normal and pathological placental development.

Secondly, by conducting a transcriptome-scale meta-analysis of sex-biased gene expression, this research revealed that 140 genes are differentially expressed between male and female placentae. A majority of these genes are autosomal, many of which are involved in high-level regulatory processes such as gene transcription, cell growth and proliferation and hormonal function. Of particular interest, all genes in the *LHB-CGB* cluster were expressed more highly in female placentas, which includes genes involved in placental development, the maintenance of

pregnancy and maternal immune tolerance of the conceptus. These results demonstrated that sex-biased gene expression in the normal human placenta occurs across the genome and includes genes that are central to growth, development and the maintenance of pregnancy.

Thirdly, by undertaking a comprehensive analysis of human placental gene co-expression using RNA sequencing and the integration of five human and one mouse transcriptome dataset, this research identified clusters of correlated genes, whose patterns of co-expression are highly preserved across human gestation and between human and mouse, subsequently revealing highly conserved molecular networks involved in placental development. Furthermore, by reducing the complexity of the placental transcriptome by summarizing co-expressed genes, this work identified a group of co-expressed genes implicated in preeclampsia and also outlines a novel method for identifying for non-invasive biomarkers of placental development.

In summary, each aspect of this PhD research has provided new insights into how gene expression is regulated in the human placenta and has revealed previously unappreciated aspects of the placental transcriptional landscape.

Acknowledgements

Science is a collaborative effort and the work presented herein is no exception. I would like to express my sincere gratitude to Claire Roberts and Tina Bianco-Miotto, whose expertise, understanding, and patience, added considerably to my PhD experience. I appreciate and respect their vast knowledge and skill in many areas of reproductive and molecular biology, and am deeply thankful for their assistance in writing up the work presented in this thesis. I would also like to thank the other members of my supervisory committee, Stephen Bent for providing guidance and insight regarding bioinformatics analyses and Gus Dekker for expert advice and assistance regarding the clinical aspects of this research project. I also extend my sincere thanks to Hannah Brown for providing scientific guidance.

I would like to acknowledge Shalem Leemaqz, Dan Kortschak, Jessica Laurence, Prabha Andraweera and Amanda Highet for participating in discussions of this work, Dylan McCullough for laboratory technical support, Liying Yan, Ann Meyer and Matthew Poulin of EpigenDx for their pyrosequencing technical assistance, Joel Geoghegan at the Australian Cancer Research Foundation Cancer Genomics Facility for RNA sequencing support and Kartik Shankar for providing raw RNA sequencing data for validation purposes. I am grateful for the valuable and independent mentoring support from Jozef Gecz, and thank Lisa Martin for assisting with proof reading and copy editing. I would also like to acknowledge the other post-graduate students in the CTR lab; Zimin Zhuang, Fleur Spronk, Benjamin Mayne, Rebecca Wilson, Sultana Khoda, Alison Leviton and Ang Zhou.

I recognise that this research would not have been possible without the financial assistance I have received, and am thankful for the Australian Postgraduate Award from the Australian Government Department for Education and Training, and generous PhD top-up scholarships from the Channel 7 Children's research foundation and Healthy Development Adelaide, and the National Health and Medical Research Council (NHMRC). Support for traveling to conferences was generously provided by The Robinson Research Institute and the Society of Reproductive Biology. Funding for this research was provided by NHMRC through project grants (APP1059120 and APP565320) and a NHMRC Senior Research Fellowship awarded to Claire Roberts (APP1020749).

Finally, I would like to thank my family for all their encouragement during my candidature, and I want to extend a special thank you to my wife Kylie for all her understanding, love and support throughout my undergraduate and postgraduate training.

Publications Arising from this Thesis

1. **Buckberry, S.**, Bianco-Miotto, T., Hiendleder, S. & Roberts, C. T. Quantitative allele-specific expression and DNA methylation analysis of H19, IGF2 and IGF2R in the human placenta across gestation reveals H19 imprinting plasticity. ***PLoS One*** 7, e51210 (2012).

2. **Buckberry, S.**, Bianco-Miotto, T. & Roberts, C. T. Imprinted and X-linked non-coding RNAs as potential regulators of human placental function. ***Epigenetics*** 9, 81–89 (2014).

3. **Buckberry, S**., Bent, S. J., Bianco-Miotto, T. & Roberts, C. T. massiR: a method for predicting the sex of samples in gene expression microarray datasets. ***Bioinformatics*** 30, 2084–2085 (2014).

4. **Buckberry, S.**, Bianco-Miotto, T., Bent, S. J., Dekker, G. A. & Roberts, C. T. Integrative transcriptome meta-analysis reveals widespread sex-biased gene expression at the human fetal–maternal interface. ***Molecular Human Reproduction*** 20, 810-819 (2014).

5. **Buckberry, S.** & Roberts, C. T. Why are males more at risk in the womb? ***Australasian Science*** 35, 9, 16–18 (2014). **Not peer reviewed*

6. **Buckberry, S.**, Bianco-Miotto, T., Bent, S. J., Dekker, G. A. & Roberts, C. T. Placental transcriptome co-expression analysis reveals conserved regulatory programs and points toward a preeclampsia gene cluster. ***To be submitted May 2015***.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 The Placenta Forms the Foundation of Development *in utero*

Placental development in humans begins shortly after an embryo implants into the lining of the uterus. Here, placental cells begin their invasion and colonisation of the uterine vessels and form finger-like projections known as villi, which provide maximum contact area with the maternal blood. Placental chorionic villi facilitate exchange of nutrients, gases and wastes between the mother and fetus, thereby forming the foundation for successful pregnancy.

The process of placental trophoblast invasion, which has many similarities with cancer metastasis [1], appears to be strictly controlled in humans, both spatially and temporally, through mechanisms that are only partially understood [2, 3]. Impaired placental invasion has been implicated in several complications of pregnancy such as preeclampsia, intrauterine growth restriction [4] and preterm labour [5, 6]. For example, in preeclampsia, invasion of the maternal spiral arterioles is typically shallow, resulting in poor maternal blood flow to the placenta [4, 7, 8]. Despite extensive research efforts, our understanding of how placental development is regulated at the molecular level remains inadequate, especially given the severity of pathologies arising from abnormal placentation.

Although the placenta is most widely known for mediating fetal–maternal exchange, it also plays a major role in directing the mother's adaptation to pregnancy. To achieve this, the placenta secretes a variety of steroid and peptide hormones that intricately modulate maternal physiology in a way that enables pregnancy to be sustained. For example, human chorionic gonadotropin (hCG)

released from syncytiotrophoblast cells maintains progesterone secretion from the corpus luteum in the ovary during the first trimester, which is essential for the maintenance of pregnancy and fetal development.

Although the placenta is a shared organ between a mother and her fetus, it is an extra-embryonic tissue that originates from the conceptus and is therefore genetically identical to other fetal tissues. Subsequently, placental development and function is primarily governed by the fetal genome in the context of a nutrient supply from the mother.

1.2 Genomic Imprinting in the Placenta

One mechanism of placental genome regulation that has intrigued biologists for decades is the phenomenon of genomic imprinting. This is where allelic expression is dependent upon the parent from whom the allele was inherited. Imprinting affects gene dosage, with the imprinted allele considered repressed and functionally silenced [9, 10]. Imprinting is largely, although not exclusively, observed in eutherian mammals and is thought to have arisen with viviparity and the evolutionary emergence of the chorioallantoic placenta [11–14]. Of particular interest is the fact that more imprinted genes are expressed in the placenta than any other tissue, and the most widely recognised mechanism for maintenance of imprinting is DNA methylation.

The prevailing model of imprinting suggests that DNA methylation imprints are established across the genome shortly after fertilisation [10]. A vast majority of research into the effects of genomic imprinting have centred around the regulation of the *H19-IGF2* locus, which encompasses the two most highly expressed genes in the placenta [15]. Paternally expressed *IGF2* encodes the growth-promoting insulin-like growth factor 2, a potent mitogen involved in the regulation of cell proliferation, growth and development. The reciprocally imprinted, maternally expressed *H19* gene is located approximately 130 kb downstream of *IGF2* on human chromosome 11 and encodes a highly expressed, growth regulating, non-coding RNA that shares regulatory elements with *IGF2* [16]. However, the mechanisms by which *H19* interacts with *IGF2* and regulates placental growth and development are not fully understood.

Studies using animal models have demonstrated the functional importance of imprinting *H19* and *IGF2* genes during intrauterine development [17–20], however the timing of imprinting establishment in humans remains unclear. Limited research on allele-specific expression in the human placenta suggests that imprinting

may be dynamic across gestation. Although some differences in allele-specific expression of imprinted genes between the first trimester and term human placenta have previously been reported [21], there appear to be no studies focused on potential changes across the first trimester; a highly dynamic period of placental growth and differentiation. Thus, there are few or no data on the establishment of imprinting during early gestation, or any data regarding the stability of DNA methylation imprints throughout the first trimester and later gestation.

The primary aim of the first study presented in this thesis was to quantify allele-specific expression and DNA methylation for these reciprocally imprinted genes, which have been widely used to exemplify the phenomenon of imprinting in humans and rodents. The results indicated that the *H19* long non-coding RNA has a high degree of allelic variation between six and ten weeks of gestation, which stabilises at around 11–12 weeks [22]. Given that *H19* is a regulator of placental and embryonic growth [23], these results suggested that the perturbation of *H19* imprinting might significantly influence early programming of placental phenotype.

Just prior to the results of this first study being published in *PLoS One* (Chapter 2), Keniry et al. reported in *Nature Cell Biology* that the growth suppressing ability of *H19* was due to a microRNA (miR-675) encoded within the first H19 exon [24]. They showed that levels of miR-675 increased as gestation progressed, which acted to suppress placental growth towards the end of gestation [24]. Although overall *H19* expression remains largely unchanged throughout gestation, the RNA-binding protein Elavl1 appeared to bind to the *H19* RNA transcript, preventing miR-675 from being excised early in gestation Keniry:2012fi. These results certainly suggested that the functionality of *H19* was due to the actions of miR-675. However, it is important to note that these experiments were conducted in the mouse model and may not be directly translatable to humans. To further explore the conclusions drawn by Keniry et al., I conducted a computational evaluation of their proposed *H19*/miR-675 model in humans. This analysis, which formed part of a review published in *Epigenetics* (Chapter 3), provided several lines of evidence indicating that miR-675 regulation is indeed different in the human placenta [25].

In this review, we also discuss an intriguing cluster of imprinted miRNAs that are expressed almost exclusively in the placenta from the paternal allele. Of particular interest, C19MC miRNAs are detectable in exosomes released from trophoblast cells and in maternal blood, and are implicated in pregnancy complications such as preeclampsia and preterm birth. However, the C19MC cluster is unique to the primate lineage, which imposes serious limitations on model organism research.

Additionally, we review the evidence for X-linked miRNAs as potential drivers of sex differences in placental gene expression. Since the majority of genes are

autosomal, many sexually dimorphic traits are driven by the sex-biased expression of autosomal genes. Much of the scientific literature in the past has attributed the influence of sex hormones to sexual dimorphism; however, in this review, we explore the idea that X-linked miRNA may target many autosomal genes, giving rise to sex differences in autosomal gene expression. While many sex-specific gene expression differences have been appreciated for some time, their phenotypic and clinical implications, particularly in the placenta and in pregnancy complications, remain relatively unexplored. We subsequently developed some of these ideas, which led to the next major study presented in this thesis (Chapter 5).

1.3 Genome-Wide Sex Differences in Placental Gene Expression

During intrauterine development, there are distinct sex differences in fetal growth trajectories and hence in birth weight [26–28], with a sex bias in the prevalence of preterm birth [29, 30], pregnancy complications such as preeclampsia [31, 32] and perinatal death [33]. As fetal growth and development are highly dependent on the exchange efficiency and capacity of the placenta, sex-specific differences in normal and pathological fetal development are most likely due to sex differences in placental function. Several studies have shown a distinct sex bias in the prevalence of placental dysfunction in a spectrum of pregnancy and fetal health conditions associated with abnormal placentation [32, 34–38]. However, the underlying mechanisms that predispose one sex over the other to deviate from a normal course of development remain elusive.

Given our lack of knowledge regarding sex differences in placental genome regulation, there was a crucial need to establish the baseline differences between the sexes in normal development. Since the sex differences in expression observed in other human tissues have been subtle [39–43], it was deemed necessary to undertake a study large enough to have the power to detect small but significant sex differences in placental gene expression.

To address this need, we conducted an integrative meta-analysis of publicly available microarray data to determine the extent of sex-biased placental gene expression. This involved mining gene expression data repositories for raw probe-level data from non-pathological human placental tissue. After extensive curation and filtering, the resulting dataset consisted of more than 300 samples with a total of 9.65 million data points. Analysis of these data revealed that more than 140 genes consistently show significant sex-biased expression, and that a majority of these

genes are autosomal. Of particular interest, we detected higher female expression from all seven genes in the *LHB-CGB* (hCG) cluster, which includes genes involved in placental development, the maintenance of pregnancy and maternal immune tolerance of the conceptus. This demonstrated that sex-biased gene expression in the normal human placenta occurs across the genome. This work was published in the journal *Molecular Human Reproduction* (Chapter 5) [44].

At the beginning of this meta-analysis we encountered a significant problem: a large proportion of publicly available data lacked sample sex information. To overcome this problem, I created an algorithm that can predict the sex of samples in gene expression datasets with a high degree of accuracy. A manuscript describing this algorithm was published in the journal *Bioinformatics* (Chapter 4) [45]. Pleasingly, the findings from this meta-analysis of sex-biased expression attracted worldwide media attention and we were subsequently invited to write an editorial piece describing this work to be published in *Australasian Science*. This short article is presented, as published, in Chapter 6 [46].

1.4 The Underlying Organisation of the Human Placental Transcriptome

Although the microarray meta-analysis generated a substantial amount of new knowledge regarding gene regulation in the placenta, we also became increasingly aware of how little was known about placental genome regulation throughout gestation.

During my PhD candidature, two separate manuscripts were published describing the human placental transcriptome [15, 47]. The main outcome of these two studies, which both utilised RNA sequencing (RNA-Seq) to quantify global gene expression, was the identification of RNA transcripts that are enriched in the placenta compared to other human tissues and placental-specific RNA splice variations [15, 47]. Although these higher resolution analyses surveyed the transcriptome at the level of exon splicing, there were few new insights regarding the organisation of placental gene transcription.

A common feature of a vast majority of previous studies on placental gene expression is that the data are typically summarised at the gene level for between-group comparisons. Consequently, the greatest significance is then attributed to individual genes where the differences between groups reach an appropriate significance

threshold. Although these gene-level analyses have unquestionable utility, the inherent natural organisation of the transcriptome remains largely unexplored.

In other fields of research, more holistic approaches to the analysis of gene expression data have recently begun to reveal unappreciated patterns of transcriptional organisation with regards to lipid metabolism [48], cancer [49], human brain development and neuropathology [50–52], and embryonic development [53]. These systems-level approaches typically involve the identification of groups of genes with highly correlated expression across samples. By focusing on the relationships between genes using a guilt-by-association approach, identifying groups of genes that are expressed in a highly coordinated manner can uncover higher order relationships among genes and their products. Further *post hoc* characterisation of these relationships can then provide insight into the biological functions arising from the underlying transcriptional program.

Considering the valuable insight to be gained by adopting a systems biology approach to analysing the placental transcriptome, the aim of the next study was to closely examine the co-expression relationships between genes in the human placenta. This study, presented in Chapter 7, involved profiling the human placental transcriptome by RNA-Seq and integrating several transcriptome datasets. The results of this comprehensive analysis revealed highly correlated patterns of gene expression that are associated with distinct biological processes, and highlighted a cluster of co-regulated genes implicated in preeclampsia. Furthermore, by drastically reducing the dimensionality of gene expression data through summarising highly correlated genes in this study, we illustrate a potential framework for screening biomarkers of placental development.

1.5 Summary

The PhD research presented here focuses on three key aspects of placental gene expression: (1) the regulation of expression by genomic imprinting, (2) the role of fetal sex in placental genome regulation, and (3) the underlying organisation of the transcriptome across gestation. All three unique aspects of this project have provided novel insights into mechanisms of gene regulation in the placenta and highlighted new avenues of research regarding the role of placental gene expression in pregnancy complications. Furthermore, many of the methods developed during the course of these projects have not been previously applied in the fields of placental biology or obstetric medicine, and subsequently provide a new framework for investigating gene regulation at the fetal–maternal interface.

Bibliography

[1] Murray, M. J. and Lessey, B. A. Embryo implantation and tumor metastasis: common pathways of invasion and angiogenesis. *Seminars in reproductive endocrinology,* 17(3) (1999), pp. 275–290.

[2] Grahamm, C. H. and Lala, P. K. Mechanisms of placental invasion of the uterus and their control. *Biochemistry and Cell Biology,* 70(10-11) (1992), pp. 867–874.

[3] Lyall, F. The human placental bed revisited. *Placenta,* 23(8-9) (2002), pp. 555–562.

[4] Khong, T. Y., De Wolf, F., Robertson, W. B., and Brosens, I. Inadequate maternal vascular response to placentation in pregnancies complicated by pre-eclampsia and by small-for-gestational age infants. *British Journal of Obstetrics and Gynaecology,* 93(10) (1986), pp. 1049–1059.

[5] Kim, Y. M., Chaiworapongsa, T., Gomez, R., Bujold, E., Yoon, B. H., Rotmensch, S., Thaler, H. T., and Romero, R. Failure of physiologic transformation of the spiral arteries in the placental bed in preterm premature rupture of membranes. *Am J Obstet Gynecol,* 187(5) (2002), pp. 1137–1142.

[6] Kim, Y. M., Bujold, E., Chaiworapongsa, T., Gomez, R., Yoon, B. H., Thaler, H. T., Rotmensch, S., and Romero, R. Failure of physiologic transformation of the spiral arteries in patients with preterm labor and intact membranes. *Am J Obstet Gynecol,* 189(4) (2003), pp. 1063–1069.

[7] Rinkenberger, J. L., Cross, J. C., and Werb, Z. Molecular genetics of implantation in the mouse. *Developmental genetics,* 21(1) (1997), pp. 6–20.

[8] Zhou, Y., Genbacev, O., Damsky, C. H., and Fisher, S. J. Oxygen regulates human cytotrophoblast differentiation and invasion: implications for endovascular invasion in normal pregnancy and in pre-eclampsia. *Journal of Reproductive Immunology,* 39(1-2) (1998), pp. 197–213.

[9] Ferguson-Smith, A. and Surani, M. Imprinting and the epigenetic asymmetry between parental genomes. *Science,* 293(5532) (2001), p. 1086.

[10] Reik, W. and Walter, J. Genomic imprinting: parental influence on the genome. *Nature Reviews Genetics,* 2(1) (2001), pp. 21–32.

[11] Monk, D., Arnaud, P., Apostolidou, S., Hills, F. A., Kelsey, G., Stanier, P., Feil, R., and Moore, G. E. Limited evolutionary conservation of imprinting in the human placenta. *Proceedings of the National Academy of Sciences,* 103(17) (2006), pp. 6623–6628.

[12] Haig, D. Altercation of generations: genetic conflicts of pregnancy. *American Journal of Reproductive Immunology,* 35(3) (1996), p. 226.

[13] Haig, D. and Westoby, M. An earlier formulation of the genetic conflict hypothesis of genomic imprinting. *Nature Genetics,* 38(3) (2006), pp. 271–271.

[14] Haig, D. Genetic Conflicts in Human Pregnancy. *The Quarterly Review of Biology,* 68(4) (1993), pp. 495–532.

[15] Saben, J., Zhong, Y., McKelvey, S., Dajani, N. K., Andres, A., Badger, T. M., Gomez-Acevedo, H., and Shankar, K. A comprehensive analysis of the human placenta transcriptome. *Placenta,* 35(2) (2014), pp. 125–131.

[16] Gabory, A., Ripoche, M. A., Yoshimizu, T., and Dandolo, L. The H19 gene: regulation and function of a non-coding RNA. *Cytogenetic and Genome Research,* 113(1-4) (2006), pp. 188–193.

[17] Coan, P. M., Fowden, A. L., Constancia, M., Ferguson-Smith, A., Burton, G. J., and Sibley, C. P. Disproportional effects of Igf2 knockout on placental morphology and diffusional exchange characteristics in the mouse. *The Journal of physiology,* 586(Pt 20) (2008), pp. 5023–5032.

[18] Constancia, M., Angiolini, E., Sandovici, I., Smith, P., Smith, R., Kelsey, G., Dean, W., Ferguson-Smith, A., Sibley, C. P., Reik, W., and Fowden, A. Adaptation of nutrient supply to fetal demand in the mouse involves interaction between the Igf2 gene and placental transporter systems. *Proceedings of the National Academy of Sciences of the United States of America,* 102(52) (2005), pp. 19219–19224.

[19] Leighton, P. A., Ingram, R. S., Eggenschwiler, J., Efstratiadis, A., and Tilghman, S. M. Disruption of imprinting caused by deletion of the H19 gene region in mice. *Nature,* 375(6526) (1995), pp. 34–39.

[20] Sibley, C. P., Coan, P. M., Ferguson-Smith, A., Dean, W., Hughes, J., Smith, P., Reik, W., Burton, G. J., Fowden, A. L., and Constancia, M. Placental-specific insulin-like growth factor 2 (Igf2) regulates the diffusional exchange characteristics of the mouse placenta. *Proceedings of the National Academy of Sciences of the United States of America,* 101(21) (2004), pp. 8204–8208.

[21] Pozharny, Y., Lambertini, L., Ma, Y., Ferrara, L., Litton, C. G., Diplas, A., Jacobs, A. R., Chen, J., Stone, J. L., Wetmur, J., and Lee, M.-J. Genomic loss of imprinting in first-trimester human placenta. *American Journal of Obstetrics and Gynecology,* 202(4) (2010), 391.e1–391.e8.

[22] Buckberry, S., Bianco-Miotto, T., Hiendleder, S., and Roberts, C. T. Quantitative allele-specific expression and DNA methylation analysis of H19, IGF2 and IGF2R in the human placenta across gestation reveals H19 imprinting plasticity. *PloS one,* 7(12) (2012), e51210.

[23] Gabory, A., Jammes, H., and Dandolo, L. The H19 locus: role of an imprinted non-coding RNA in growth and development. *BioEssays,* 32(6) (2010), pp. 473–480.

[24] Keniry, A., Oxley, D., Monnier, P., Kyba, M., Dandolo, L., Smits, G., and Reik, W. The H19 lincRNA is a developmental reservoir of miR-675 that suppresses growth and Igf1r. *Nature cell biology,* 14(7) (2012), pp. 659–665.

[25] Buckberry, S., Bianco-Miotto, T., and Roberts, C. T. Imprinted and X-linked non-coding RNAs as potential regulators of human placental function. *Epigenetics: official journal of the DNA Methylation Society,* 9(1) (2014), pp. 81–89.

[26] Crawford, M. A., Doyle, W., and Meadows, N. Gender differences at birth and differences in fetal growth. *Human Reproduction,* 2(6) (1987), pp. 517–520.

[27] Onis, M. de, Siyam, A., Borghi, E., Onyango, A. W., Piwoz, E., and Garza, C. Comparison of the World Health Organization growth velocity standards with existing US reference data. *Pediatrics,* 128(1) (2011), e18–26.

[28] Bertino, E., Coscia, A., Boni, L., Rossi, C., Martano, C., Giuliani, F., Fabris, C., Spada, E., Zolin, A., and Milani, S. Weight growth velocity of very low birth weight infants: role of gender, gestational age and major morbidities. *Early Human Development,* 85(6) (2009), pp. 339–347.

[29] Brettell, R., Yeh, P. S., and Impey, L. W. M. Examination of the association between male gender and preterm delivery. *European journal of obstetrics, gynecology, and reproductive biology,* 141(2) (2008), pp. 123–126.

[30] Ingemarsson, I. Gender aspects of preterm birth. *British Journal of Obstetrics and Gynaecology,* 110 (2003), pp. 34–38.

[31] Elsmén, E., Källén, K., Marsál, K., and Hellström-Westas, L. Fetal gender and gestational-age-related incidence of pre-eclampsia. *Acta obstetricia et gynecologica Scandinavica,* 85(11) (2006), pp. 1285–1291.

[32] Murji, A., Proctor, L. K., Paterson, A. D., Chitayat, D., Weksberg, R., and Kingdom, J. Male sex bias in placental dysfunction. *American Journal of Medical Genetics Part A,* 158A (2012), pp. 779–783.

[33] Vatten, L. J. and Skjaerven, R. Offspring sex and pregnancy outcome by length of gestation. *Early Human Development,* 76(1) (2004), pp. 47–54.

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

[35] Hodyl, N. A., Wyper, H., Osei-Kumah, A., Scott, N., Murphy, V. E., Gibson, P., Smith, R., and Clifton, V. L. Sex-specific associations between cortisol and birth weight in pregnancies complicated by asthma are not due to differential glucocorticoid receptor expression. *Thorax,* 65(8) (2010), pp. 677–683.

[36] Di Renzo, G., Rosati, A., Sarti, R., and Cruciani, L. Does fetal sex affect pregnancy outcome? *Gender medicine,* 4(1) (2007), pp. 19–30.

[37] Edwards, A., Megens, A., and Peek, M. Sexual origins of placental dysfunction. *The Lancet,* 355 (2000), pp. 203–204.

[38] Ghidini, A. and Salafia, C. M. Gender differences of placental dysfunction in severe prematurity. *BJOG : an international journal of obstetrics and gynaecology,* 112(2) (2005), pp. 140–144.

[39] Trabzuni, D., Ramasamy, A., Imran, S., Walker, R., Smith, C., Weale, M. E., Hardy, J., Ryten, M., and North American Brain Expression Consortium. Widespread sex differences in gene expression and splicing in the adult human brain. *Nature communications,* 4 (2013), p. 2771.

[40] Sood, R., Zehnder, J. L., Druzin, M. L., and Brown, P. O. Gene expression patterns in human placenta. *Proceedings of the National Academy of Sciences,* 103(14) (2006), pp. 5243–5244.

[41] Whitney, A. R., Diehn, M., Popper, S. J., Alizadeh, A. A., Boldrick, J. C., Relman, D. A., and Brown, P. O. Individuality and variation in gene expression patterns in human blood. *Proceedings of the National Academy of Sciences of the United States of America,* 100(4) (2003), pp. 1896–1901.

[42] Zhang, Y. E., Vibranovski, M. D., Landback, P., Marais, G. A. B., and Long, M. Chromosomal redistribution of male-biased genes in mammalian evolution with two bursts of gene gain on the X chromosome. *PLoS biology,* 8(10) (2010).

[43] Cvitic, S., Longtine, M. S., Hackl, H., Wagner, K., Nelson, M. D., Desoye, G., and Hiden, U. The Human Placental Sexome Differs between Trophoblast Epithelium and Villous Vessel Endothelium. *PloS one,* 8(10) (2013), e79233.

[44] Bianco-Miotto, T., Bent, S. J., Dekker, G. A., and Roberts, C. T. Integrative transcriptome meta-analysis reveals widespread sex-biased gene expression at the human fetal–maternal interface. *Molecular Human Reproduction,* (2014).

[45] Buckberry, S., Bent, S. J., Bianco-Miotto, T., and Roberts, C. T. massiR: a method for predicting the sex of samples in gene expression microarray datasets. *Bioinformatics (Oxford, England),* (2014).

[46] Buckberry, S. and Roberts, C. T. Why are males more at risk in the womb? *australasian Science,* 35(9) (2014), pp. 16–18.

[47] Kim, J., Zhao, K., Jiang, P., Lu, Z.-X., Wang, J., Murray, J. C., and Xing, Y. Transcriptome landscape of the human placenta. *BMC Genomics,* 13(1) (2012), p. 115.

[48] Langfelder, P., Castellani, L. W., Zhou, Z., Paul, E., Davis, R., Schadt, E. E., Lusis, A. J., Horvath, S., and Mehrabian, M. A systems genetic analysis of high density lipoprotein metabolism and network preservation across mouse models. *Biochimica et biophysica acta,* 1821(3) (2012), pp. 435–447.

[49] Hu, Y., Wu, G., Rusch, M., Lukes, L., Buetow, K. H., Zhang, J., and Hunter, K. W. Integrated cross-species transcriptional network analysis of metastatic susceptibility. *Proceedings of the National Academy of Sciences of the United States of America,* 109(8) (2012), pp. 3184–3189.

[50] Voineagu, I., Wang, X., Johnston, P., Lowe, J. K., Tian, Y., Horvath, S., Mill, J., Cantor, R. M., Blencowe, B. J., and Geschwind, D. H. Transcriptomic analysis of autistic brain reveals convergent molecular pathology. *Nature,* 474(7351) (2011), pp. 380–384.

[51] Oldham, M. C., Konopka, G., Iwamoto, K., Langfelder, P., Kato, T., Horvath, S., and Geschwind, D. H. Functional organization of the transcriptome in human brain. *Nature Neuroscience,* 11(11) (2008), pp. 1271–1282.

[52] Gupta, S., Ellis, S. E., Ashar, F. N., Moes, A., Bader, J. S., Zhan, J., West, A. B., and Arking, D. E. Transcriptome analysis reveals dysregulation of innate immune response genes and neuronal activity-dependent genes in autism. *Nature communications,* 5 (2014), p. 5748.

[53] Xue, Z., Huang, K., Cai, C., Cai, L., Jiang, C.-y., Feng, Y., Liu, Z., Zeng, Q., Cheng, L., Sun, Y. E., Liu, J.-y., Horvath, S., and Fan, G. Genetic programs in human and mouse early embryos revealed by single-cell RNA sequencing. *Nature,* 500(7464) (2013), pp. 593–597.

Statement of Authorship

Title of Paper	Quantitative Allele-Specific Expression and DNA Methylation Analysis of H19, IGF2 and IGF2R in the Human Placenta Across Gestation Reveals H19 Imprinting Plasticity
Publication Status	✓ Published, ○ Accepted for Publication, ○ Submitted for Publication, ○ Publication style
Publication Details	Buckberry, Sam, Tina Bianco-Miotto, Stefan Hiendleder, and Claire T. Roberts. 2012. "Quantitative Allele-Specific Expression and DNA Methylation Analysis of H19, IGF2 and IGF2R in the Human Placenta Across Gestation Reveals H19 Imprinting Plasticity." PloS One Vol.7, Issue.12, e51210.

Author Contributions

By signing the Statement of Authorship, each author certifies that their stated contribution to the publication is accurate and that permission is granted for the publication to be included in the candidate's thesis.

Name of Principal Author (Candidate)	Sam Buckberry		
Contribution to the Paper	Sam Buckberry conceived, designed and carried out the experiments in this study, performed all statistical analyses, interpreted the results, authored the first draft of the manuscript, incorporated suggestions from co-authors and peer-reviewers and wrote the final manuscript.		
Signature		Date	7/4/15

Name of Co-Author	Tina Bianco-Miotto		
Contribution to the Paper	Tina Bianco-Miotto was involved in conceiving the study, supervised experimental design and laboratory work, contributed valuable insight toward interpreting the results and provided detailed comments for all manuscript drafts.		
Signature		Date	1/4/15

Name of Co-Author	Stefan Hiendleder		
Contribution to the Paper	Stefan Hiendleder was involved in conceiving the study and experimental design, contributed valuable insight toward interpreting the results and provided extensive comments on the first draft and the final manuscript.		
Signature		Date	19/03/2015

Name of Co-Author	Claire T Roberts		
Contribution to the Paper	Claire T Roberts conceived the study, supervised the experimental design, provided all samples and funding for the experiments, contributed valuable insight towards interpreting the results and provided detailed comments for all manuscript drafts.		
Signature		Date	7.4.15

Chapter 2

Quantitative Allele-Specific Expression and DNA Methylation Analysis of *H19*, *IGF2* and *IGF2R* in the Human Placenta Across Gestation Reveals *H19* Imprinting Plasticity

Sam Buckberry, Tina Bianco-Miotto, Stefan Hiendleder and Claire T Roberts

Abstract

Imprinted genes play important roles in placental differentiation, growth and function, with profound effects on fetal development. In humans, *H19* and *IGF2* are imprinted, but imprinting of *IGF2R* remains controversial. The *H19* non-coding RNA is a negative regulator of placental growth and altered placental imprinting of *H19-IGF2* has been associated with pregnancy complications such as preeclampsia, which have been attributed to abnormal first trimester placentation. This suggests that changes inimprinting during the first trimester may precede aberrant placental morphogenesis. To better understand imprinting in the human placenta during early gestation, we quantified allele-specific expression for *H19*, *IGF2* and *IGF2R* in first trimester

(6–12 weeks gestation) and term placentae (37–42 weeks gestation) using pyrosequencing. Expression of *IGF2R* was biallelic, with a mean expression ratio of 49:51 (SD = 0.07), making transient imprinting unlikely. Expression from therepressed *H19* allelesranged from 1–25% and was higher ($P < 0.001$) in first trimester ($13.5 \pm 8.2\%$) compared to term ($3.4 \pm 2.1\%$) placentae. Surprisingly, despite the known coregulation of *H19* and *IGF2*, little variation in expression of the repressed *IGF2* alleles was observed ($2.7 \pm 2.0\%$). To identify regulatory regions that may be responsible for variation in *H19* allelic expression, we quantified DNA methylation in the *H19-IGF2* imprinting control region and *H19* transcription start site (TSS). Unexpectedly, we found positive correlations ($P < 0.01$) between DNA methylation levels and expression of the repressed *H19* allele at 5 CpG's 2000 bp upstream of the *H19* TSS. Additionally, DNA methylation was significantly higher ($P < 0.05$) in first trimester compared with term placentae at 5 CpG's 39–523 bp upstream of the TSS, but was not correlated with *H19* repressed allele expression. Our data suggest that variation in *H19* imprinting may contribute to early programming of placental phenotype and illustrate the need for quantitative and robust methodologies to further elucidate the role of imprinted genes in normal and pathological placental development.

2.1 Introduction

Genomic imprinting refers to parent-of-origin-dependent allele-specific gene expression. Imprinting affects gene dosage, with the imprinted allele considered repressed and functionally silenced [1, 2]. Imprinting is largely, although not exclusively, observed in eutherian mammals and is thought to have arisen with viviparity and the evolutionary emergence of the chorioallantoic placenta [3, 4]. The prevailing hypothesis on the origin of imprinting is based on paternal-maternal conflict and postulates that paternally expressed genes have been selected to maximize fetal resource acquisition from the mother, while maternally expressed genes have been selected to balance resource allocation to current and future offspring [4]. As imprinted genes appear to facilitate this *tug-of-war* between the maternal and paternal genomes, the conflict hypothesis predicts that imprinted genes are involved in fetal and placental growth and development during pregnancy [2, 4, 5].

Studies using animal models have demonstrated the functional importance of imprinting of *H19*, *IGF2* and *IGF2R* genes during intrauterine development [6–

10]. Paternally expressed *IGF2* encodes the growth promoting insulin-like growth factor II, a potent mitogen involved in regulating cell proliferation, growth and development. The reciprocally imprinted, maternally expressed *H19* gene is located approximately 130kb downstream of *IGF2* on human chromosome 11 and encodes a highly expressed, growth regulating, non-coding RNA that shares regulatory elements with *IGF2* [11]. The mechanism by which *H19* interacts with *IGF2* and regulates growth is not fully understood and appears to involve long range interaction of differentially methylated regions and complex loop structures that regulate the activity of parental alleles [12–14]. More recently, *H19* has been identified as a trans regulator of an imprinted gene network for growth and development [15], apparently through miRNAs processed from the H19 transcript [11, 16, 17]. The H19 large intergenic non-coding RNA (lincRNA) is highly expressed in extra-embryonic cell lineages and is a developmental reservoir of miR-675 that suppresses placental growth in the mouse [18]. The IGF2 receptor (*IGF2R*) mediates endocytosis and clearance or activation of a variety of ligands involved in the regulation of cell growth and motility, including insulin-like growth factor II [19–21].

Studies in mice have demonstrated that altered imprinting of *H19*, *IGF2* and *IGF2R* are associated with placental and fetal growth abnormalities [11, 22, 23], some of which are consistent with data from human studies. For example, (epi)mutations in the *H19-IGF2* region are associated with Silver-Russell and Beckwith-Wiedemann syndromes, which manifest *in utero* in severely growth-restricted and overgrowth phenotypes, respectively [24]. Furthermore, altered epigenetic regulation of the *H19-IGF2* region in human placenta has been associated with pregnancy complications such as preeclampsia, which are preceded by placental pathologies [25, 26]. A significant role in placental development has been established for *H19* and *IGF2* in mouse and human, but knowledge on the role of *IGF2R* in human placental development is limited. The *IGF2R* gene is imprinted in all tissues except brain in mouse, but the majority of human samples indicate non-imprinted biallelic expression [3, 27–29]. The minority of samples with imprinted or partially imprinted expression suggested developmental stage-specific transient imprinting. However, the developmental role of rare, transient or partial *IGF2R* imprinting in the human placenta [3, 27, 30–33] remains to be established.

In the human placenta, biallelic expression of imprinted genes, including *H19*, has been observed at higher rates during the first trimester of pregnancy compared to term [25, 34, 35]. Intriguingly, biallelic expression of *H19* in term placentae has been associated with preeclampsia in one study [25], yet subtle variation in *H19* allelic expression in healthy term placentae has also been observed [36]. This

limited research on allele-specific expression in the human placenta suggested that imprinting may be dynamic across gestation with potential plasticity in imprinting beyond blastocyst and implantation stages. Although some differences in allele-specific expression of imprinted genes between the first trimester and term human placenta have been reported [34], there appear to be no studies addressing potential changes across the first trimester, a highly dynamic period of placental growth and differentiation. Thus, there is little or no data on temporal variation in imprinting of these genes across gestation, or if imprinting is stable throughout the first trimester and later gestation. In the present study, we quantified the allelic expression ratio for *H19*, *IGF2* and *IGF2R* and DNA methylation in the *H19-IGF2* imprinting control region across 6–12 weeks of gestation in first trimester placentae and in term placentae between 37–42 weeks of gestation.

2.2 Materials and Methods

2.2.1 Ethics Statement

Ethics approval was granted by the Children, Youth and Women's Health Service Research Ethics Committee (REC2249/2/13), the Central Northern Adelaide Health Service Ethics of Human Research Committee (Approval #2005082) and the University of Adelaide Human Research Ethics Committee (H-137-2006). Written informed consent was obtained from all patients.

2.2.2 Sample Collection

First trimester placental samples ranging from 6–12 weeks of gestation were obtained from elective terminations of pregnancies at the Women's and Children's Hospital, South Australia. The consulting physician determined gestational age by observation and the date of the last menstrual period. Placental villous samples were washed in sterile PBS and snap frozen in liquid nitrogen before being stored at −80 °C. Term placenta samples were collected from pregnancies classified as being uncomplicated by using the criteria described in [37], and were collected and dissected post-delivery at the Lyell McEwin Health Service, South Australia, and incubated in RNAlater solution (Invitrogen) at 4 °C for 24 hours before being stored at −80 °C.

2.2.3 Genotyping

DNA was extracted from placental tissue and parental blood using the Qiagen® DNeasy® blood and tissue kit following the manufacturer's instructions. DNA concentration was determined using the NanoDrop® ND-1000 Spectrophotometer and diluted to 12 ng µL^{-1} with nuclease-free water (Mo Bio Laboratories). Isolated DNA from first trimester placental samples was genotyped for *IGF2* rs680, *IGF2R* rs998075 and *IGF2R* rs1570070 single nucleotide polymorphisms (SNPs) by PCR and High Resolution Melt (HRM) analysis (see Methods S1 S2.2 on page 39). Term placenta and parental DNA SNP genotypes for *H19* rs217727 and *IGF2* rs680 were determined by multiplex PCR and the Sequenom® MassARRAY® system, using the iPLEX® GOLD single base extension reaction on custom arrays at the Australian Genome Research Facility, Brisbane, Australia.

2.2.4 Quantification of Allele Specific Expression

Placental samples were thawed and homogenised with 1 mL TRIzol (Invitrogen) per 100 mg tissue. TRIzol (Invitrogen) extraction was performed according to the manufacturer's guidelines. RNase-free glycogen (Ambion) was added at 25 µg per 1 mL of TRIzol (Invitrogen) to aid in RNA visualisation. RNA samples were DNase treated using the TURBO DNA-free™ kit (Ambion) following the manufacturer's instructions for rigorous treatment. Following DNase treatment, 2 µL of RNA was subjected to PCR with DNA-specific primers (Table S2.1 on page 34). The DNase treatment was determined to be effective if samples showed no amplification after 35 cycles. The concentration of DNase-treated RNA was calculated with the NanoDrop® ND-1000 Spectrophotometer.

First-strand cDNAs were synthesised from 500 ng DNase-treated RNA using the iScript™ cDNA Synthesis Kit (Bio-Rad), following the manufacturer's instructions. Reverse transcriptase was omitted for negative controls and aliquots of the master mix without added RNA were included in PCR experiments to rule out contamination. Following reverse transcription, cDNA was diluted 1:10 with nuclease-free water (Mo Bio Laboratories). Aliquots from five cDNA samples were pooled and serially diluted 5-fold for primer validation and PCR optimisation.

PCR primers flanking SNP regions and pyrosequencing primers were designed using the PSQ™ assay design software (Biotage™). Reverse primers featured 5' biotin modifications and were HPLC purified. All oligonucleotides were synthesised by GeneWorks (Adelaide) and are listed in Table S2.2 on page 35. Each

sample was pyrosequenced in triplicate, with each replicate generated in an independent PCR cycling run. PCR was performed using 10 µL reactions with 2 µL of cDNA, 5 µL SsoFast EvaGreen Supermix (Bio-Rad) and 300 nM of each primer. Cycling conditions were 2 min enzyme activation at 95 °C followed by 40 cycles of 5 s at 95 °C and 20 s at 60 °C. PCR products were sequenced by pyrosequencing using the methods detailed below.

2.2.5 Quantification of DNA Methylation

DNA for methylation analysis was extracted from placental villous tissue by homogenizing 50 mg to 100 mg tissue in 500 µL of TES (10 mM Tris-HCL pH 8.0, 1 mM EDTA, 100 mM NaCl), then adding 300 µg Proteinase K and 30 µL of 20% SDS followed by an overnight incubation at 37 °C. Then 3 M NaCl was added to precipitate proteins and the supernatant was collected by centrifugation. The DNA was pelleted using 2 volumes of absolute ethanol and washed in 70% ethanol, air dried and resuspended in TE pH 8.0 [38].

Each DNA sample was bisulfite treated in triplicate by EpigenDx (Massachusetts, USA) using 500 ng of DNA and a proprietary bisulfite salt solution followed by incubation for 14 hours at 50 °C. Bisulfite treated DNA was purified using Zymogen DNA columns and was eluted with 20 µL of TE pH 8.0, 1 µL of which was used for PCR reactions. The PCR was performed with 0.2 µM of each primer for EpigenDx methylation assays ADS025, ADS596FS and ADS004 with one of the PCR primers being biotinylated for purifying the final PCR product.

2.2.6 Pyrosequencing

The biotinylated PCR products were bound to Streptavidin Sepharose HP (Amersham Biosciences, Sweden), and the Sepharose beads containing the immobilized PCR product were purified, washed and denatured using a 0.2 M NaOH solution and rewashed all using the Pyrosequencing Vacuum Prep Tool (Qiagen) as recommended by the manufacturer. Then 0.2 µM pyrosequencing primer was annealed to the purified single-stranded PCR product. 10 µL of the PCR products were sequenced using the PSQ96 HS System (Biotage AB) following the manufacturer's instructions at EpigenDX Genome and Epigenome Research Facility (Massachusetts, USA). The status of each locus was analyzed individually using QCpG software (Qiagen).

2.2.7 Statistical Analysis

Repressed allele expression differences between gestational age classes for each gene were analyzed using one-way analysis of variance (ANOVA). Differences between first trimester (6–12 weeks of gestation) and term (37–42 weeks of gestation) samples were analyzed using the *t*-test. Differences in allelic expression measured at two loci in the one sample were analyzed using the paired *t*-test. The relationship between repressed allele expression from two genes in the same sample was tested by calculating the Pearson's bivariate correlation coefficient. Differences in levels of DNA methylation between first trimester and term samples were tested for each individual CpG site and for each region using independent *t*-tests. The relationship between repressed allele expression and mean DNA methylation for each region and CpG site was tested using Pearson's correlation. Results were considered significant at $P < 0.05$. All statistical analyses were performed using GraphPad PRISM 5.0.

2.3 Results

2.3.1 Quantification of Allele-Specific Gene Expression in the Human Placenta

DNA samples from placental tissue was genotyped for SNPs *H19* rs217727, *IGF2* rs680, *IGF2R* rs998075 and *IGF2R* rs1970070 to identify heterozygous individuals. Sixty-nine samples in total were heterozygous for at least one of the tested candidate SNPs. The number of heterozygotes identified for each gestational age class is summarized in Table 2.1 on the next page. As parental DNA corresponding to term placenta samples was available for 28 cases, we genotyped maternal and paternal DNA for *H19* and *IGF2* polymorphisms to determine the parental origin of expressed alleles. In all cases with sufficient parental genotype information, *H19* was maternally expressed ($n = 11$) and *IGF2* ($n = 9$) was paternally expressed, as expected (Table S2.3 on page 36).

Relative expression from each *H19*, *IGF2* and *IGF2R* allele was quantified in placenta samples by pyrosequencing of SNP loci. Relative allelic expression levels for *H19*, *IGF2* and *IGF2R* in first trimester and term placenta samples are presented in Figure 2.1 on page 22 and Figure 2.2 on page 23 with each gene showing a unique allele expression profile. Technical replicates obtained from independent PCR reactions showed average standard deviations (SD) of 0.44%

TABLE 2.1: Number of informative heterozygous samples for *H19*, *IGF2* and *IGF2R* for each gestational age class. First trimester samples range from 6–12 weeks, term samples range from 37–42 weeks.

	Number of Heterozygotes		
Gestational Age (weeks)	*H19*	*IGF2*	*IGF2R*
6	1	1	1
7–8	6	5	14
9–10	4	6	8
11–12	2	2	1
37	1	0	NA
38	0	0	NA
39	7	5	NA
40	5	8	NA
41	5	6	NA
42	0	1	NA
Total	**31**	**34**	**24**

for *H19* rs217727, 1.02% for *IGF2* rs680, 3.14% for *IGF2R* rs998075 and 3.84% for *IGF2R* rs1570070, respectively, indicating robust assays with negligible inter-PCR variation. The greater standard deviation for *IGF2R* replicates is likely due to the higher PCR quantification cycle (C_q) required for data acquisition as compared with *H19* and *IGF2* (data not shown). The ratio of repressed allele to predominantly expressed allele is depicted in Figure 2.3 on page 24, where a 0:100 ratio represents no expression from the repressed allele and a 50:50 ratio represents balanced, i.e., bialleleic, non-imprinted expression. Across first trimester gestational age classes, expression from the *H19* repressed allele shows notable inter-individual variation in contrast to the almost uniform monoallelic expression observed for *IGF2* (Figure 2.3 on page 24). *IGF2R* allele-specific expression in first trimester placenta samples showed balanced expression, with some samples potentially showing a slight allelic bias (Figure 2.2 on page 23, Figure 2.3 on page 24).

2.3.2 Biallelic Expression of *IGF2R* in the First Trimester Placenta

Allelic expression ratios for *IGF2R* in first trimester placenta was measured at two SNP loci (rs998075 $n = 16$ and rs1570070 $n = 13$). Five samples were heterozygous for both SNPs, and no significant difference (paired t-test, $P = 0.42$) was detected between the expression ratios for the two SNP loci, indicating that

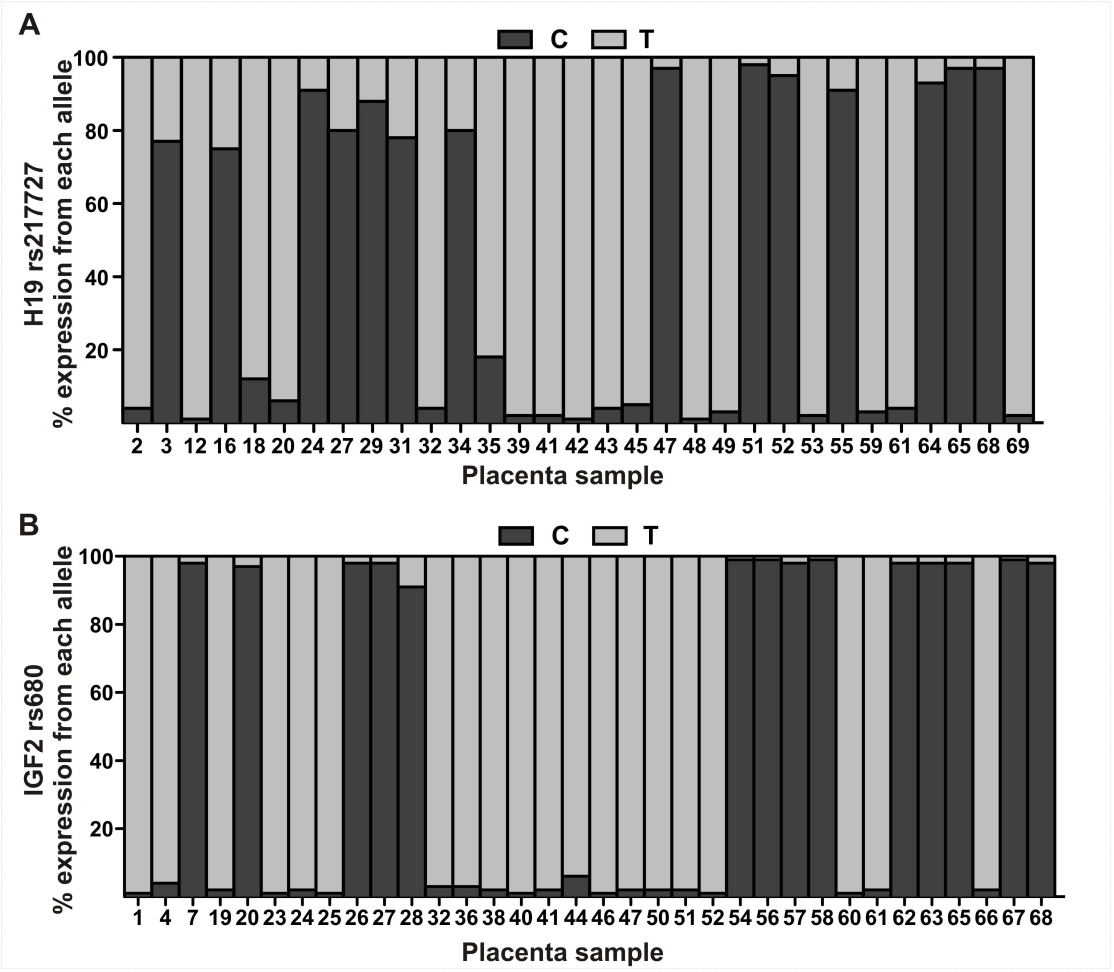


FIGURE 2.1: Relative expression from *H19* and *IGF2* alleles in the human placenta. Shaded bars for (A) *H19* and (B) *IGF2* represent the proportion of expression (%) from each allele. Samples 1–38 are from first trimester (6–12 weeks of gestation) placentae and samples 39–69 are from term (37–42 weeks of gestation) placentae.

both polymorphisms were equivalent in quantifying allele-specific expression (Figure 2.2C). All heterozygous IGF2R samples were therefore combined for analyses, and, when expression was quantified at both loci in one sample, the average allelic ratio of the two loci was used. The results clearly show biallelic *IGF2R* expression in all first trimester placental samples assessed (Figures 2.2A, 2.2B), with a mean allele expression ratio of 49:51 at both SNP loci (rs998075 SD = 7.1%, rs1570070 SD = 6.9%) with expression ratios ranging from 36:64 to 49:51 (Figure 2.3 on page 24). These SNP based *IGF2R* pyrosequencing results provide no evidence for *IGF2R* imprinted expression in the first trimester placenta and thus confirm the non-imprinted status of *IGF2R* throughout gestation.

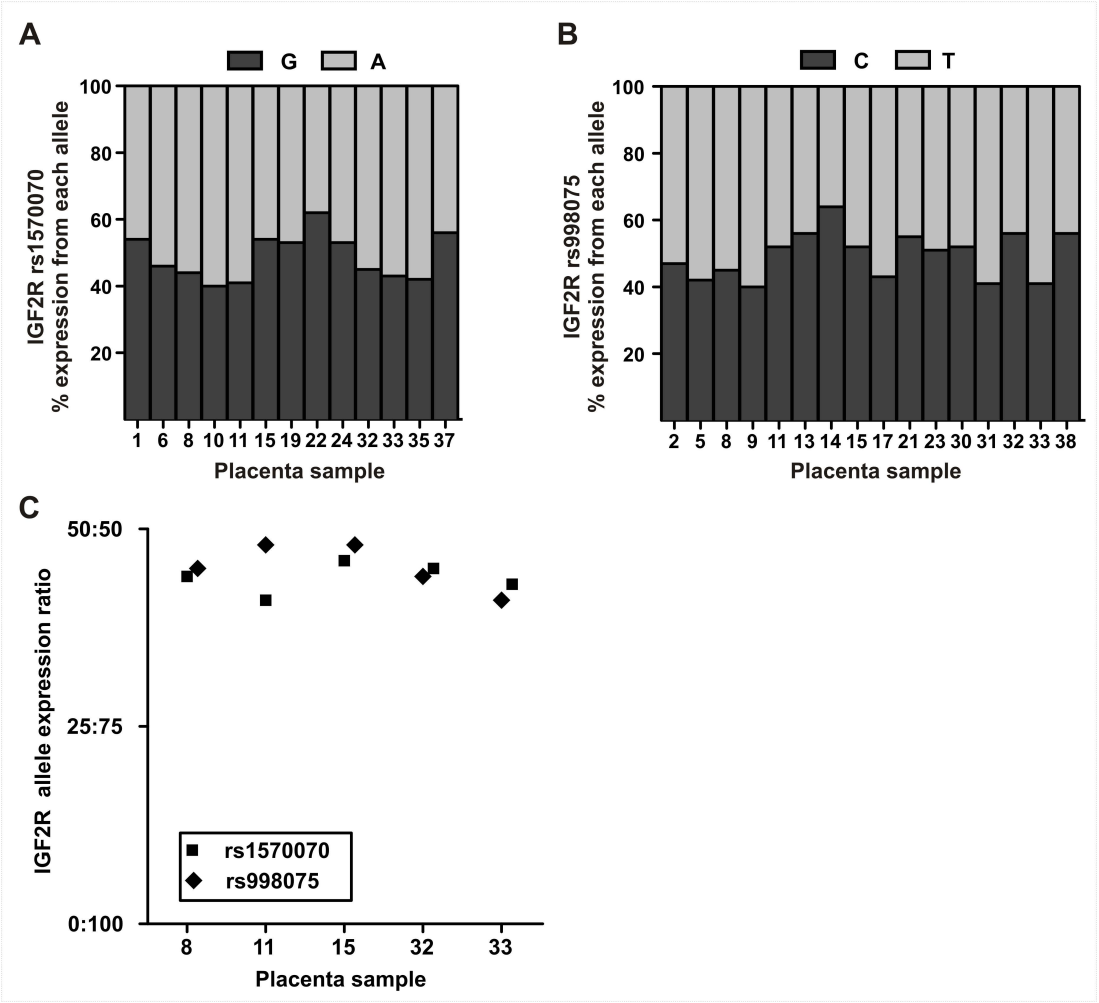


FIGURE 2.2: Relative expression from *IGF2R* alleles in the human placenta. Shaded bars for (A) *IGF2R* rs1570070 and (B) *IGF2R* rs998075 represent the proportion of expression (%) in first trimester placentae. (C) Allelic expression ratios for *IGF2R* measured two SNP loci in the same sample. These paired samples indicate both SNP loci are equivalent (paired t-test, $P = 0.42$) for evaluating *IGF2R* allele-specific expression.

2.3.3 Increased Expression from the *H19* Repressed Allele is Higher in First Trimester Placenta

Expression from the *H19* repressed allele was quantified in 13 first trimester placenta samples obtained at 6–12 weeks of gestation (Figure 2.1A). Mean expression from the repressed allele was 13.5% (SD ±8.2; Figure 2.1A) and ranged from 0.9–24.7% (Figure 2.3 on the following page). Expression of the *H19* repressed allele appeared to decrease with gestational age in the first trimester samples (Figure 2.4A), but we found no significant differences between first trimester gestational age classes. To further test the hypothesis that expression from the repressed *H19* allele decreases across gestation, we then quantified allelic expression in term

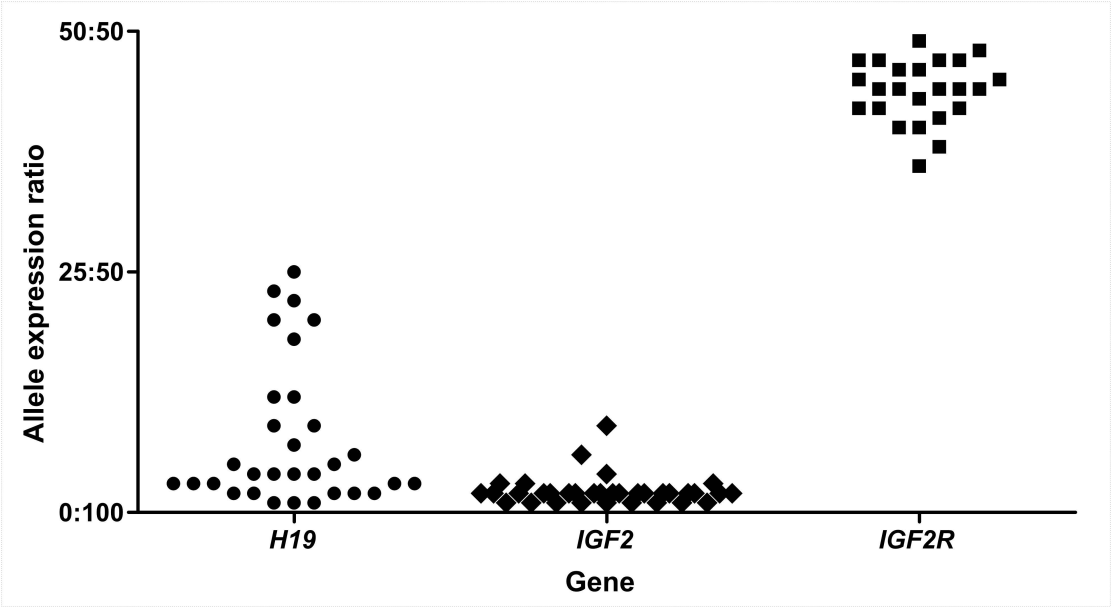


FIGURE 2.3: Allelic expression ratios for *H19*, *IGF2* and *IGF2R* in the human placenta. The 50:50 ratio represents equal expression from both alleles and 0:100 ratio represents expression exclusively from one allele. Each point on the graph represents the allelic expression ratio measured in an individual placental sample. *H19* and *IGF2* samples are from first trimester and term placentae, *IGF2R* samples are all from first trimester placentae.

placenta samples obtained between 37–42 weeks of gestation ($n = 18$). Expression from the repressed *H19* allele at term was significantly lower ($P < 0.001$) than the level of expression observed in first trimester placenta samples (Figure 2.4B and Table 2.2).

TABLE 2.2: Relative levels of repressed allele expression in human first trimester and term placentae.

	% Repressed allele expression		
	First trimester	Term	*P* value
H19	13.5 ± 8.3	3.4 ± 2.0	< 0.0001
IGF2	2.6 ± 2.0	1.9 ± 1.1	0.1734
IGF2R	43.8 ± 3.2	ND	

ND = Not Determined

Expression from the *IGF2* repressed allele contributed on average 2.7% (SD 2.1%, $n = 34$) to total *IGF2* transcript in placenta samples (Figure 2.1B). No significant differences in expression from the *IGF2* repressed allele were observed between first trimester gestational age classes (ANOVA $P > 0.05$) or between first trimester and term (Table 2.2, Figures 2.4C, 2.4D) placentae. These results show that imprinted *IGF2* expression is tightly regulated and stable across gestation.

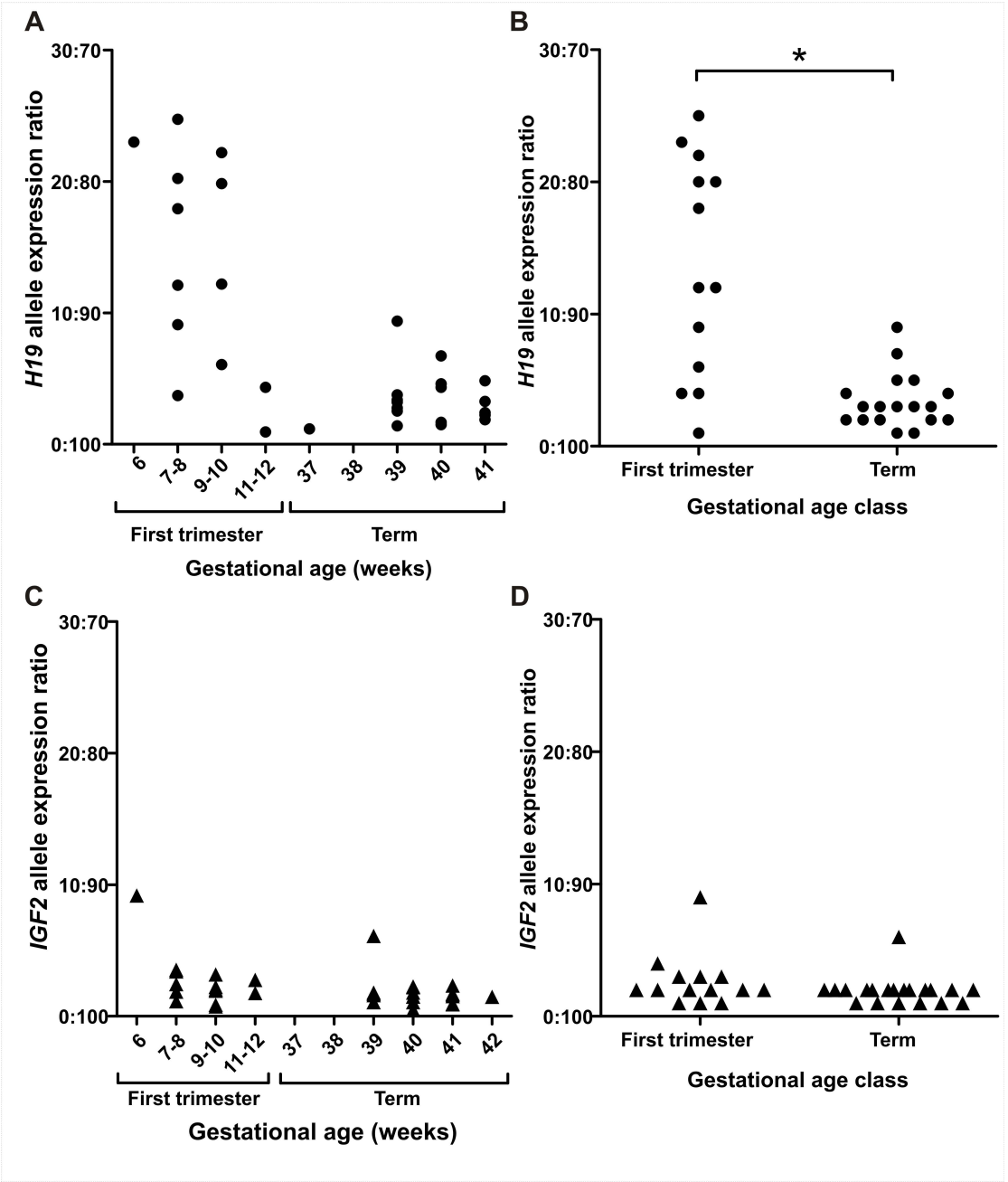


FIGURE 2.4: Ratio of expression from each allele in human first trimester and term placentae measured by pyrosequencing. Each point on the graph represents the allelic expression ratio observed in an individual placental sample. (A) *H19* allelic expression ratio for each gestational age class. (B) Expression from the *H19* repressed allele is significantly higher ($^*P < 0.001$) in first trimester placental samples. (C–D) *IGF2* allelic expression ratios are similar for each gestational age class (C) with no significant difference (D) between first trimester and term. First trimester samples are 6–12 weeks of gestation term samples are 37–42 weeks of gestation.

Eleven placenta samples were heterozygous for both *H19* and *IGF2* polymorphisms, which allowed us to test for a correlation in repressed allele expression for these adjacent co-regulated imprinted genes. We found that expression from the *H19* repressed allele was not correlated ($P = 0.88$, $r = 0.54$) with expression from the *IGF2* repressed allele (Figure 2.5).

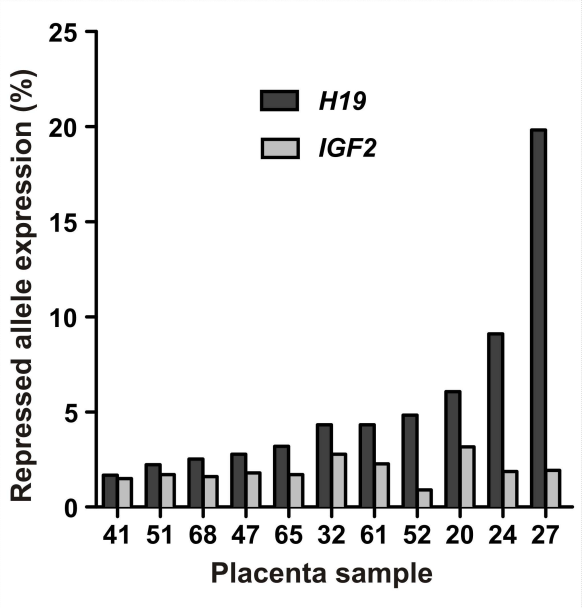


FIGURE 2.5: Relative level of expression from repressed alleles in samples heterozygous for both *H19* rs217727 and *IGF2* rs680. Graph shows increased expression from the *H19* repressed allele is not correlated ($P = 0.09$, $r = 0.54$) with expression from the *IGF2* repressed allele.

2.3.4 Locus-Specific DNA Methylation Differences in the *H19-IGF2* Region Between First Trimester and Term Placentae

To investigate if DNA methylation levels at specific CpG's are correlated with *H19* repressed allele expression, we quantified methylation levels in three regions (Figure 2.6 on the following page) using bisulfite pyrosequencing. The two regions upstream of the transcription start site (TSS) (regions 1 and 2 on Figure 2.6) were selected as they cover or are directly adjacent to sites that have been shown to be differentially methylated [39], and region 3 (Figure 2.6) was selected as it spanned the *H19* promoter region and the TSS.

The first region (denoted 1 in Figure 2.6) encompassed five CpG sites with a mean methylation level of $54.7 \pm 7.8\%$, which would be expected at a differentially methylated imprinted locus. In region 1, there was no significant difference in mean

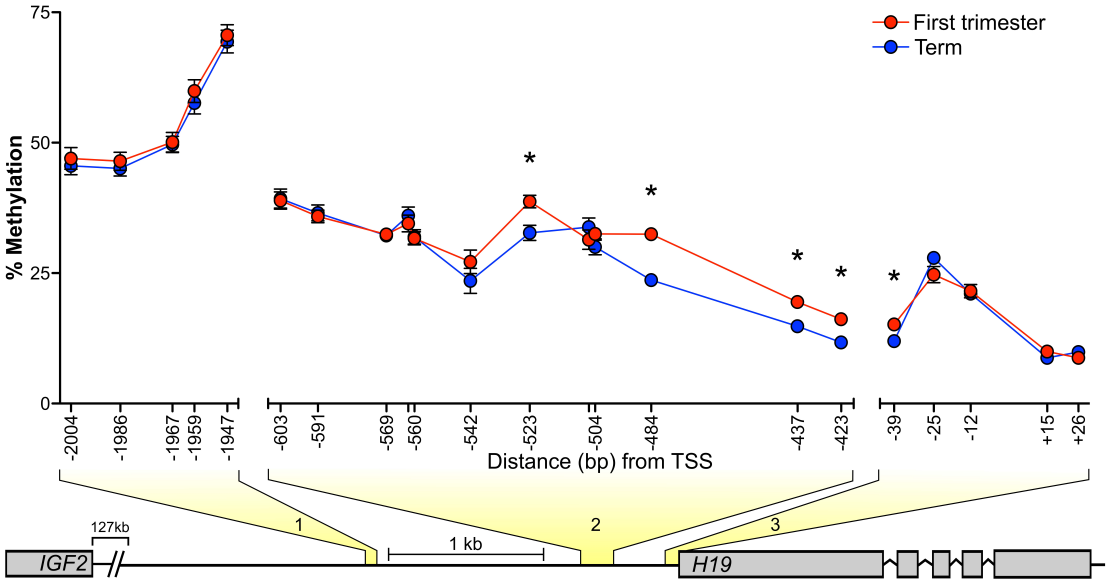


FIGURE 2.6: Placental methylation levels in regions upstream and covering the *H19* transcription start site (TSS). Each genomic region where DNA methylation was measured is highlighted in yellow and numbered 1–3. DNA methylation levels for individual CpG loci are shown for first trimester (red circles, $n = 13$) and term (blue circles, $n = 18$) placentae. Distance (bp) of cytosine nucleotide from *H19* TSS is represented on x-axis. Each data point represents the mean methylation level for the gestational age class. * Indicates a significant difference in methylation levels between first trimester and term placentae at individual CpG sites. Error bars represent SEM and when not present SEM was too low to depict on the graph. The schematic representation below the graph highlights the regions between *H19* and *IGF2* where bisulfite DNA pyrosequencing was performed. Region 1 covers 5 CpG sites (Chr11:2021011-2021070), region 2 covers 12 CpG sites (Chr11:2021011-2021070) and region 3 covers 5 CpG sites (Chr11:2019079-2019145). Genomic coordinates refer to reference assembly GRCh37/hg19.

methylation levels between first trimester ($54.8 \pm 6.9\%$) and term ($53.5 \pm 7.2\%$) placentae or at any of the 5 individual CpG sites (Figure 2.6, Table S2.4 on page 37). The second region assessed (denoted 2 in Figure 2.6), covered 12 CpG sites which showed overall hypomethylation, with a mean methylation level of $30.9 \pm 3.9\%$ in first trimester and $28.9 \pm 5.3\%$ in term placentae. When analyzed independently, 4 of the 12 CpG sites, 3 of which are adjacent to each other, showed significantly higher methylation in first trimester placentae in comparison to term placentae (Figure 2.6). The third region that spanned the *H19* TSS showed mean methylation levels of $16.1 \pm 3.1\%$ in first trimester placentae and $15.5 \pm 3.5\%$ in term placentae. When each CpG site was analyzed individually, the cytosine nucleotide 39 bp upstream from the *H19* TSS (Figure 2.6) showed significantly higher methylation ($P = 0.02$) in first trimester placentae ($15.2 \pm 3.6\%$ vs $12.0 \pm 3.2\%$). Details of DNA methylation levels in first trimester and term

placentae at each individual CpG site and the statistical comparisons between the groups are listed in Table S2.4 on page 37.

2.3.5 *H19* Repressed Allele Expression is Correlated with Higher Levels of DNA Methylation

As distinct variation in expression from the *H19* repressed allele in first trimester placentae was observed, we tested for correlations between the level of repressed allele expression and levels of DNA methylation at CpG's of first trimester placentae. In region 1, a significant positive correlation ($P < 0.001$, $r = 0.65$) was observed between repressed allele expression and the mean methylation level across the region (Figure 2.7A). When each of the 5 CpG sites in this region were analyzed independently for the same correlation, the results remained significant for each site (Table S2.5 on page 38). This correlation was not observed for region 2 (Figure 2.7B, $P = 0.36$, $r = 0.07$) or 3 (Figure 2.7C, $P = 0.47$, $r = 0.05$) or for any individual CpG sites within these regions (Table S2.5).

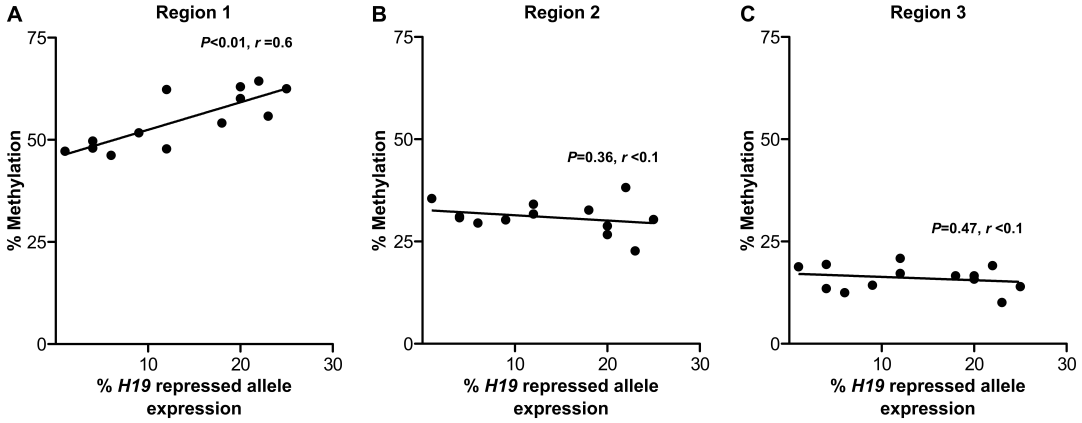


FIGURE 2.7: Levels of *H19* repressed allele expression and DNA methylation in human first trimester placentae. (A) Increased expression from the repressed *H19* allele is positively correlated ($P = 0.0016$, $r = 0.61$) with increased DNA methylation in region 1. (B & C) *H19* repressed allele expression is not correlated with DNA methylation in region 2 ($P = 0.3626$, $r = 0.08$) or region 3 ($P = 0.4791$, $r = 0.04$). Each point on the graph represents individual first trimester placenta samples. Methylation levels in each region represent the average methylation from 5 CpG sites in region 1 (Chr11:2021011-2021070), 12 CpG sites in region 2 (Chr11:2021011-2021070) and 5 CpG sites in region 3 (Chr11:2019079-2019145). Genomic coordinates refer to reference assembly GRCh37/hg19.

2.4 Discussion

Imprinted genes are known to be critically involved in placental development and function. Aberrant patterns of imprinted gene expression are implicated in pregnancy complications such as preeclampsia and intrauterine growth restriction [5, 40–43]. Although the symptoms of these conditions manifest late in pregnancy, their pathogenesis is commonly attributed to compromised first trimester placental development [44]. Previous research on genomic imprinting in the human placenta has focused on the term placenta [25, 32, 36, 41, 45, 46] and data during the first trimester of gestation is limited [27, 34, 35, 47]. In the present study, we investigated the imprinting status (i.e., allele-specific expression) of three genes, *H19*, *IGF2* and *IGF2R*, which have known, but poorly understood, associations with pregnancy complications and placental abnormalities in humans and/or animal models [6–10]. We assessed allele-specific expression of these genes and DNA methylation in the *H19-IGF2* imprinting control region in first trimester (6–12 weeks of gestation) and term (37–42 weeks of gestation) placentae.

We assessed *IGF2R* allele-specific expression, as the imprinting status of this important gene for prenatal growth and development remains controversial in human. We observed balanced expression from both *IGF2R* alleles, and although we did not investigate any potential imprinting mechanisms for this gene, these results suggest *IGF2R* is not imprinted in the first trimester placenta. Imprinting of *IGF2R* has been suggested to be a polymorphic trait in humans, with a small proportion of individuals showing monoallelic expression or partial imprinting [3, 27, 33]. In this study, we assessed more informative samples than previous studies [3, 27, 33, 48] but found no evidence for polymorphic *IGF2R* imprinting in the placenta. Although we observed overall a balanced expression of alleles for *IGF2R*, individual allelic expression ratios ranged from 36:64 to 49:51. This variation may reflect what has been described previously as partial repression or allelic preference [27, 33]. It is presently unclear if this subtle imbalance of *IGF2R* allelic expression is due to genetic variation in allele-specific epigenetic regulation or a parent-of-origin effect.

Allele-specific expression of *H19* showed considerable inter-individual variation, with expression from the repressed (i.e. imprinted) allele contributing up to 25% of total *H19* transcript in the first trimester placenta. In contrast, *IGF2* showed predominantly monoallelic expression and little variation between individuals, with one allele contributing more than 90% of *IGF2* transcript in all investigated samples. This indicated that *IGF2* allele-specific expression is tightly regulated in the first trimester placenta and suggests that *IGF2* imprinting is established early in development and remains stable throughout gestation.

Determining loss of imprinting or biallelic expression of imprinted genes was previously performed by restriction fragment length polymorphism (RFLP) analysis. This method provides a qualitative or semi-quantitative assessment of monoallelic or biallelic expression. In human placentae from uncomplicated pregnancies, *H19* RFLP data showed biallelic expression before 10 weeks of gestation and imprinted expression at term [25, 35]. However, term placentae from preeclamptic pregnancies were reported to display biallelic expression with the RFLP method [25]. This biallelic *H19* expression could indicate a failure to establish correct *H19* imprinting with downstream effects on placental development [25, 35]. The data presented in the current study show that *H19* expression from the imprinted, i.e. repressed, allele can range from 9–22% at 9–10 weeks of gestation, highlighting the potential ambiguity in classifying expression as mono- or biallelic by less sensitive methods. Our data support the view [32, 49] that classification of genes as imprinted or non-imprinted by qualitative methods may be a less meaningful distinction than quantitative measurements of imprinting status based on precise estimates of relative contributions from each allele.

More recently, quantitative PCR and pyrosequencing have been used to evaluate allele-specific expression in placental tissue. By using these highly sensitive methodologies, expression from the "silenced", imprinted, alleles has been generally higher in first trimester placentae [46] with some variation at term [36]. Both the RFLP assay and quantitative allele-specific expression approaches support the concept that repressed allele expression changes through gestation in the placenta, particularly during early pregnancy [25, 34, 35, 46]. Using placental tissue from 6–12 weeks of gestation, we tested the hypothesis that imprinted allele-specific expression changes during the first trimester of pregnancy. We found no significant differences between early and late first trimester allelic expression ratios for *H19*, *IGF2* or *IGF2R*. Although we quantified allelic expression ratios using a highly sensitive technique, the method used for classifying gestational age, our sample size, and the proportion of heterozygotes in each group may have prevented the detection of significant changes across first trimester age groups. When comparing first trimester and term placenta samples for *H19*, we found a significant decrease in the proportion of repressed allele expression at term. Furthermore, these results for *H19* show notable inter-individual variation early during placental development, and more uniformity in allelic expression ratios as gestation progresses. This is a clear demonstration of dynamic change in imprinting status well beyond the blastocyst and implantation stages. However, an alternative explanation for the observed differences in *H19* allelic expression ratios between first trimester and term samples in the present study could be the unbiased sampling of material from elective terminations of pregnancy versus the selected material at term that

came from normal pregnancies only. Placental tissue from elective terminations of pregnancy in first trimester will, by necessity, include those from pregnancies that may have been destined to develop a pregnancy complication e.g. preeclampsia, preterm labour or intrauterine growth restriction which are typically diagnosed later in gestation. Potentially, first trimester placental samples exhibiting expression from the repressed allele may have been destined to retain biallelic expression and associate with preeclampsia later. However, we consider this unlikely given that 8 out of 13 first trimester samples had greater than 10% expression from the repressed allele and preeclampsia occurs in just 8% of women in the community where our samples were collected [50].

Our results show *H19* expression from the repressed allele is not correlated with expression from the *IGF2* repressed allele in the same samples. The prevailing regulatory model of the *H19-IGF2* region based on differential DNA methylation predicts that both genes are not expressed from a single chromosome. Although this model is supported by considerable evidence [12, 51, 52] (and references cited therein), there is also evidence to suggest that this model may be insufficient (reviewed in [53]). The data presented here show higher expression from the repressed *H19* allele is not correlated with any change in *IGF2* repressed allele expression in individual placentae. Additionally, we show that DNA methylation levels at CpG sites (1946–2005 bp upstream of the *H19* TSS) that flank the 6th CTCF binding domain [39], are positively correlated with the level of expression from the *H19* repressed allele, which was unexpected given the prevailing regulatory model. Furthermore, we observed significantly higher DNA methylation in first trimester placentae in the region 422–524 bp upstream of the *H19* TSS that surrounds the differentially methylated region (DMR) [39], despite finding no correlation with *H19* repressed allele expression. This suggests DNA methylation in the DMR decreases progressively throughout gestation with no effect on *H19* allelic regulation. Together, these findings suggest that the methylation dependant enhancer competition model of the *H19-IGF2* locus may not fully explain the patterns of allele-specific expression observed for these genes in the early human placenta, as suggested previously [53]. However, although we assessed DNA methylation at sites within the *H19-IGF2* regulatory region, we did not assess methylation across all the CTCF binding sites upstream of *H19*. Moreover, we did not investigate additional regulatory mechanisms, such as the actions of other non-coding RNA's and repressive histone modifications that are involved in placental-specific imprinting [54–57]. Therefore we are unable to rule out other mechanistic changes that may be influencing *H19* allele specific expression.

An important consideration when using placental tissue for studying genomic imprinting is that this organ arises from multiple extra-embryonic and embryonic cell

lineages. Cells descended from both the inner cell mass and trophectoderm may show major epigenetic differences [58], and as a result, analysis of whole placental villous tissue may not identify cell lineage-specific imprinting effects. In this study, we show a clear imprinting effect for *IGF2* in all heterozygous first trimester placenta samples, which suggests that all cell types composing the placental villi had *IGF2* imprinting mechanisms in place. However, for *H19* we observed notable inter-individual variation in expression from the imprinted allele. This variation could be due to the heterogeneous nature of the placental villous tissue sampled and *H19* lineage-specific imprinting at the single cell level. Cell-specific imprinted gene expression has been proposed as an *all or none* phenomenon in placental cell lines [59], and *H19* biallelic expression has been shown to be specific to extravillous cytotrophoblast cells [47], suggesting there is no intermediate imprinting effect at the single cell level. Therefore, observing variations in relative expression from the imprinted allele in placental tissue may simply reflect the fraction of cells with complete biallelic expression [59]. As first trimester placental tissue sampling is expected to yield a higher proportion of extravillous cytotrophoblast cells than those collected at term, changes in the level of imprinting across gestation may reflect proportional changes in cell lineage populations as the placenta differentiates. This suggests future studies of placental imprinting dynamics should consider the potential influence of placental cell type heterogeneity.

The *H19*, *IGF2* and *IGF2R* genes have key roles in placental development, yet the phenotypic effect of their allele-specific expression across gestation remains unknown. However, the role of *H19* as a regulator of the recently described imprinted gene network suggests potentially significant phenotypic effects [15]. This may depend on differences in gene dosage, but could also involve more complex regulatory effects in *trans*. To date, the normal developmental patterns of imprinted gene expression in the human placenta are poorly understood. As altered patterns of imprinting in term placentae are associated with pregnancy complications, identifying when these abnormal patterns are established may aid in elucidating the origins of placental abnormalities implicated in their aetiology. Our results highlight the requirement for robust and sensitive methods to determine the role of imprinted allele-specific expression in placentae from complicated pregnancies. Undoubtedly, precise methods and comprehensive studies will be required to progress towards understanding the molecular basis of potentially life threatening pregnancy complications in which defective placentation is implicated.

Acknowledgements

We thank Liying Yan, Ann Meyer and Matthew Poulin of EpigenDx for their pyrosequencing technical assistance and expertise. The authors would also like to thank all members of the Claire T Roberts Placental Development Laboratory for helpful discussions.

S2.1 Supporting Information

TABLE S2.1: Genomic DNA specific primers used to detect DNA contamination in RNA samples.

Gene	Location	Direction	Primer Sequence (5′–3′)
IGF2R	Intron	Fwd	GCCTCTTCTTGTTAATTTCCCTGTT
	Exon	Rev	TTCAGTTTCTCCACAGACATTCAA

TABLE S2.2: Details of genes, SNP regions and primers used for quantifying allele-specific expression by pyrosequencing.

Gene	SNP	PCR primer sequence (5′–3′)	Amplicon size (bp)	Pyrosequencing primer (5′–3′)
H19	rs217727	Fwd-CGGCGACTCCATCTTCATG Rev-(B)TCCAGCTCTGGGATGATGTG	75	ATGGCCACCCCCTGCG
IGF2	rs680	Fwd-TGGCCAGTTTACCCTGAAAATTC Rev-(B)TGGACTTGAGTCCCTGAACCA	116	CCTGTGATTTCTGGG
IGF2R	rs998075	Fwd-CTCGGTGTGTGTCTTTCATTGTT Rev-(B)CATATTATGATGGGATGATCCAAC	73	TGTCTTTCATTGTTATAGGG
IGF2R	rs1570070	Fwd-AGCAGCAGGATGTCTCCATAG Rev-(B)TGTATTTCAGTTTCTCCACAGACA	118	CCCAGAGCGGAGGTT

(B) denotes 5′ nucleotide biotin modification

TABLE S2.3: Parental and placental genotypes with placental allele expression ratio for *H19* and *IGF2*.

	H19 rs217727 genotypes			Placenta allele expression ratio		
Sample	Maternal	Paternal	Placenta	C	T	Expression
41	C/T	–	C/T	0.02	0.98	ND
42	C/T	C	C/T	0.01	0.99	Maternal
43	T/T	–	C/T	0.04	0.96	Maternal
45	C/T	–	C/T	0.05	0.95	ND
47	C/C	C/T	C/T	0.97	0.03	Maternal
48	C/T	C	C/T	0.01	0.99	Maternal
49	C/T	C	C/T	0.03	0.97	Maternal
51	C/C	C/T	C/T	0.98	0.02	Maternal
52	C/C	C/T	C/T	0.95	0.05	Maternal
53	C/T	C	C/T	0.02	0.99	Maternal
55	C/C	C/T	C/T	0.91	0.09	Maternal
59	C/T	–	C/T	0.03	0.97	ND
61	C/T	C/T	C/T	0.04	0.96	ND
68	C/C	–	C/T	0.97	0.03	Maternal
69	C/T	C/T	C/T	0.02	0.98	Maternal
	IGF2 rs680 Genotypes			Placenta allele expression ratio		
Sample	Maternal	Paternal	Placenta	C	T	Expression
40	C	C/T	C/T	0.01	0.99	Paternal
41	C	–	C/T	0.02	0.99	Paternal
44	C/T	C/T	C/T	0.06	0.94	ND
46	–	C/T	C/T	0.01	0.99	ND
47	C/T	C/T	C/T	0.02	0.98	ND
50	C/T	–	C/T	0.02	0.98	ND
51	C/T	C/T	C/T	0.02	0.98	ND
52	C	–	C/T	0.01	0.99	Paternal
54	C/T	C/T	C/T	0.99	0.01	ND
56	T	C/T	C/T	0.99	0.01	Paternal
57	C/T	–	C/T	0.98	0.02	ND
58	T	C	C/T	0.99	0.02	Paternal
60	C	C/T	C/T	0.01	1.00	Paternal
61	C/T	C/T	C/T	0.02	0.98	ND
62	–	C/T	C/T	0.98	0.02	ND
63	T	C	C/T	0.98	0.02	Paternal
66	C/T	C/T	C/T	0.02	0.98	ND
67	T	–	C/T	0.99	0.02	Paternal
68	T	–	C/T	0.98	0.02	Paternal

ND = Could not be determined

TABLE S2.4: Parental and placental genotypes with placental allele expression ratio for *H19* and *IGF2*.

			% DNA Methylation		
	Distance (bp) from TSS	Genomic location (GRCh37/hg19)	First trimester N=13, Mean ± SD	Term N=18 Mean ± SD	*P* value
Region 1	−2004	2021069	47.0 ± 7.5	45.6 ± 7.1	0.5899
	−1986	2021051	46.5 ± 6.2	45.1 ± 6.3	0.5527
	−1967	2021032	50.1 ± 6.8	49.7 ± 6.6	0.8717
	−1959	2021024	59.9 ± 7.9	57.6 ± 8.9	0.4636
	−1947	2021012	70.6 ± 7.2	69.4 ± 9.3	0.7001
	Region mean		54.8 ± 6.9	53.5 ± 7.2	0.6065
Region 2	−603	2019668	38.9 ± 6.0	39.3 ± 7.7	0.8885
	−591	2019656	35.9 ± 4.3	36.5 ± 6.5	0.7559
	−569	2019634	32.4 ± 2.5	32.2 ± 4.7	0.8808
	−562	2019627	34.5 ± 5.8	36.0 ± 7.0	0.5248
	−560	2019625	31.7 ± 4.5	32.0 ± 5.9	0.8789
	−542	2019607	27.2 ± 8.1	23.5 ± 10.2	0.2959
	−523	2019588	38.7 ± 4.4	32.7 ± 6.2	**0.0056**
	−504	2019569	31.4 ± 6.6	33.8 ± 7.4	0.3604
	−502	2019567	32.5 ± 4.3	30.0 ± 6.4	0.2361
	−484	2019549	32.5 ± 4.2	23.7 ± 4.6	< **0.0001**
	−437	2019502	19.5 ± 2.6	14.8 ± 2.2	< **0.0001**
	−423	2019488	16.2 ± 2.4	11.7 ± 2.8	< **0.0001**
	Region mean		30.9 ± 3.9	28.9 ± 5.3	0.2377
Region 3	−39	2019144	15.2 ± 3.6	12.0 ± 3.2	**0.0152**
	−25	2019130	24.7 ± 5.5	27.9 ± 4.8	0.0927
	−12	2019117	21.6 ± 4.6	21.1 ± 4.3	0.7552
	+15	2019091	10.0 ± 1.9	8.8 ± 3.2	0.2427
	+26	2019080	8.8 ± 2.9	9.9 ± 4.2	0.4347
	Region mean		16.1 ± 3.1	15.5 ± 3.5	0.6467

TABLE S2.5: Pearson's correlation of *H19* repressed allele expression and DNA methylation levels at individual CpG loci in human first trimester placentae.

	Distance (bp) from TSS	Genomic location (GRCh37/hg19)	P value	R^2 value
Region 1	−2004	2021069	**0.0011**	0.6348
	−1986	2021051	**0.0008**	0.6540
	−1967	2021032	**0.0017**	0.6062
	−1959	2021024	**0.0007**	0.6630
	−1947	2021012	**0.0077**	0.4898
	Region mean		**0.0008**	0.6527
Region 2	−603	2019668	0.1255	0.1999
	−591	2019656	0.2814	0.1045
	−569	2019634	0.3068	0.0945
	−562	2019627	0.1494	0.1793
	−560	2019625	0.2429	0.1216
	−542	2019607	0.7337	0.0110
	−523	2019588	0.6850	0.0155
	−504	2019569	0.5050	0.0414
	−502	2019567	0.5199	0.0386
	−484	2019549	0.4551	0.0517
	−437	2019502	0.3780	0.0712
	−423	2019488	0.4514	0.0525
	Region mean		0.3626	0.0758
Region 3	−39	2019144	0.7795	0.0074
	−25	2019130	0.2029	0.1429
	−12	2019117	0.3004	0.0969
	+15	2019091	0.1589	0.1719
	+26	2019080	0.6730	0.0168
	Region mean		0.4791	0.0465

S2.2 Methods S1

S2.2.1 DNA Genotyping by HRM

First trimester placenta DNA was genotyped for *IGF2* and *IGF2R* SNPs by PCR and high-resolution melt (HRM) analysis. PCR primers (Table S2.6 on the next page) were custom designed using Primer Express (v2.0 Applied Biosystems), with amplicon melting characteristics assessed using DinaMelt [60] and checked for specificity using the UCSC *in silico* PCR tool (`http://genome.ucsc.edu`) with the GRCh37/hg19 reference genome [61, 62]. All HRM oligonucleotides were manufactured by GeneWorks (Adelaide). Gene, primer and amplicon details for the PCR-HRM reactions are listed in Table S2.6. PCR-HRM was performed in 10 µL reactions, including 25 ng of genomic DNA, 250 nM of each forward and reverse primer, 5 µL of either SsoFast EvaGreen Supermix (Bio-Rad) for *IGF2R* rs1570070 or MeltDoctor (Applied Biosystems) for *IGF2R* rs998075 and *IGF2* rs680 using the Corbett Rotor-Gene 6000. Initial denaturation was 98 °C for 2 minutes (SsoFast) or 10 minutes (MeltDoctor), followed by 40–45 cycles of 2-step temperature cycling of 98 °C for 5 seconds and 60 °C for 20 seconds.

Immediately following PCR, samples were held at 50 °C for 30 seconds before temperature ramping at 0.1 °C steps at 2 seconds per step across a melt domain of > 10 °C, which was specific for each amplicon (Table S2.6). HRM curves were analysed using Corbett Rotor-Gene software (Corbett Research version 1.7, build 87). Normalisation was performed using windows of 0.5 °C at least 2 °C before the first melt transition and at least 1 °C after the sample had completely melted. Genotypes were called at 90% confidence by comparison to sequence verified controls using the Corbett Research HRM Software v 1.7 (Figure S2.1A). Control genotypes were confirmed by visually checking sequence chromatograms (Figure S2.1B) generated by the ABI 3130xl genetic analyser at Flinders and SouthPath Sequencing Facility, South Australia.

Table S2.6: Gene, SNP, and HRM genotyping assay details used for genotyping placental DNA.

Gene	SNP	PCR primer sequence (5′–3′)	Amplicon size (bp)	HRM range (°C)
IGF2	rs680	Fwd-TGGCCAGTTTACCCTGAAAATTC Rev-TGGACTTGAGTCCCTGAACCA	116	75–88
IGF2R	rs1570070	Fwd-GCCTCTTCTTGTTAATTTCCCTGTT Rev-TTCAGTTTCTCCACAGACATTCAA	95	67–77
IGF2R	rs998075	Fwd-CTCGGTGTGTGTCTTTCATTGTT Rev-CATATTATGATGGGATGATCCAAC	73	69–81

SNP = Single Nucleotide Polymorphism, HRM = High Resolution Mel

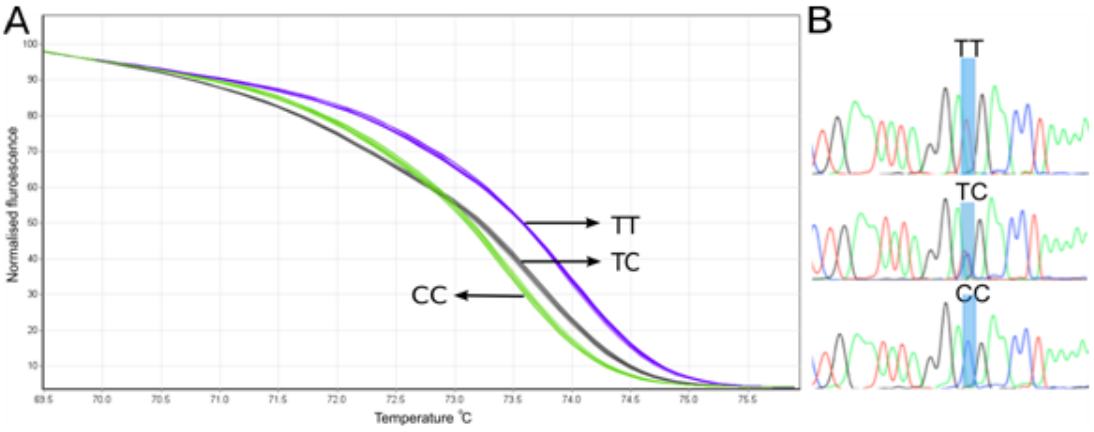


FIGURE S2.1: Genotyping of *IGF2R* rs1570070 by HRM with sequenced controls. (A) Normalised HRM melt plot showing rs1570070 TT, TC and CC genotypes. Samples are grouped into distinct genotype curves. Green and purple curves represent homozygous samples and grey curve represents heterozygous samples. Each curve group includes a sequenced control sample. (B) Chromatogram traces confirm control sample genotypes, polymorphic site highlighted for each genotype.

Bibliography

[1] Ferguson-Smith, A. and Surani, M. Imprinting and the epigenetic asymmetry between parental genomes. *Science,* 293 (2001), p. 1086.

[2] Reik, W. and Walter, J. Genomic imprinting: parental influence on the genome. *Nat. Rev. Genet.,* 21(1) (2001), pp. 21–32.

[3] Monk, D., Arnaud, P., Apostolidou, S., Hills, F. A., Kelsey, G., Stanier, P., Feil, R., and Moore, G. E. Limited evolutionary conservation of imprinting in the human placenta. In: *Proc. Natl. Acad. Sci.,* vol. 103. 2006, pp. 6623–6628.

[4] Haig, D. Altercation of generations: genetic conflicts of pregnancy. *Am J Reprod Immunol,* 35(3) (1996), pp. 226–232.

[5] Haig, D. Genetic Conflicts in Human Pregnancy. *Q Rev. Biol.,* 68 (1993), pp. 495–532.

[6] Coan, P., Fowden, A., Constancia, M., Ferguson-Smith, A., and Burton, G. Disproportional effects of Igf2 knockout on placental morphology and diffusional exchange characteristics in the mouse. *J Physiol,* 586 (2008), pp. 5023–5032.

[7] Constancia, M., Angiolini, E., Sandovici, I., Smith, P., and Smith, R. Adaptation of nutrient supply to fetal demand in the mouse involves interaction between the Igf2 gene and placental transporter systems. In: *Proc. Natl. Acad. Sci.,* vol. 102. 2005, pp. 19219–19224.

[8] Leighton, P., Ingram, R., Eggenschwiler, J., Efstratiadis, A., and Tilghman, S. Disruption of imprinting caused by deletion of the H19 gene region in mice. *Nature,* 375 (1995), pp. 34–39.

[9] Sibley, C., Coan, P., Ferguson-Smith, A., Dean, W., and Hughes, J. Placental-specific insulin-like growth factor 2 (Igf2) regulates the diffusional exchange characteristics of the mouse placenta. In: *Proc. Natl. Acad. Sci.,* vol. 101. 2004, pp. 8204–8208.

[10] Wang, Z., Fung, M., Barlow, D., and Wagner, E. Regulation of embryonic growth and lysosomal targeting by the imprintedIgf2/Mpr gene. *Nature,* 372 (1994), pp. 464–467.

[11] Gabory, A., Ripoche, M., Yoshimizu, T., and Dandolo, L. The *H19* gene: regulation and function of a non-coding RNA. *Cytogenet. Genome Res.,* 113 (2006), pp. 188–193.

[12] Murrell, A., Heeson, S., and Reik, W. Interaction between differentially methylated regions partitions the imprinted genes Igf2 and H19 into parent-specific chromatin loops. *Nat. Genet.,* 36 (2004), pp. 889–893.

[13] Qiu, X., Vu, T., Lu, Q., Ling, J., and Li, T. A complex deoxyribonucleic acid looping configuration associated with the silencing of the maternal Igf2 allele. *Mol. Endocrinol.,* 22 (2008), pp. 1476–1488.

[14] Kurukuti, S., Tiwari, V., Tavoosidana, G., Pugacheva, E., and Murrell, A. CTCF binding at the H19 imprinting control region mediates maternally inherited higher-order chromatin conformation to restrict enhancer access to Igf2. In: *Proc. Natl. Acad. Sci.,* vol. 103. 2006, pp. 10684–10689.

[15] Gabory, A., Jammes, H., and Dandolo, L. The H19 locus: role of an imprinted non-coding RNA in growth and development. *BioEssays,* 32 (2010), pp. 473–480.

[16] Cai, X. and Cullen, B. R. The imprinted H19 noncoding RNA is a primary microRNA precursor. *RNA,* 13(3) (2007), pp. 313–6.

[17] Steck, E., Boeuf, S., Gabler, J., Werth, N., Schnatzer, P., Diederichs, S., and Richter, W. Regulation of H19 and its encoded microRNA-675 in osteoarthritis and under anabolic and catabolic in vitro conditions. *J Mol Med (Berl),* 90(10) (2012), pp. 1185–95.

[18] Keniry, A., Oxley, D., Monnier, P., Kyba, M., Dandolo, L., Smits, G., and Reik, W. The H19 lincRNA is a developmental reservoir of miR-675 that suppresses growth and Igf1r. *Nat Cell Biol,* 14(7) (2012), pp. 659–65.

[19] Harris, L., Crocker, I., Baker, P., Aplin, J., and Westwood, M. IGF2 Actions on Trophoblast in Human Placenta Are Regulated by the Insulin-Like Growth Factor 2 Receptor, Which Can Function as Both a Signaling and Clearance Receptor. *Biol. Reprod.,* 84 (2011), pp. 440–446.

[20] Ghosh, P., Dahms, N., and Kornfeld, S. Mannose 6-phosphate receptors: new twists in the tale. *Nat Rev Mol Cell Biol,* 4 (2003), pp. 202–212.

[21] Hawkes, C. and Kar, S. The insulin-like growth factor-II/mannose-6-phosphate receptor: structure, distribution and function in the central nervous system. *Brain Res Rev,* 44 (2004), pp. 117–140.

[22] Lau, M. M., Stewart, C. E., Liu, Z., Bhatt, H., Rotwein, P., and Stewart, C. L. Loss of the imprinted IGF2/cation-independent mannose 6-phosphate receptor results in fetal overgrowth and perinatal lethality. *Genes Dev,* 8(24) (1994), pp. 2953–63.

[23] Constancia, M., Hemberger, M., Hughes, J., Dean, W., Ferguson-Smith, A., Fundele, R., Stewart, F., Kelsey, G., Fowden, A., Sibley, C., and Reik, W. Placental-specific IGF-II is a major modulator of placental and fetal growth. *Nature,* 417(6892) (2002), pp. 945–948.

[24] Hirasawa, R. and Feil, R. Genomic imprinting and human disease. *Essays Biochem,* 48(1) (2010), pp. 187–200.

[25] Yu, L., Chen, M., Zhao, D., Yi, P., Lu, L., Han, J., Zheng, X., Zhou, Y., and Li, L. The H19 gene imprinting in normal pregnancy and pre-eclampsia. *Placenta,* 30(5) (2009), pp. 443–7.

[26] Gao, W., Li, D., Xiao, Z., Liao, Q., and Yang, H. Detection of global DNA methylation and paternally imprinted H19 gene methylation in preeclamptic placentas. *Hypertens. Res.,* 1(7) (2011), pp. 655–661.

[27] Oudejans, C., Westerman, B., Wouters, D., Gooyer, S., and Leegwater, P. Allelic IGF2R Repression Does Not Correlate with Expression of Antisense RNA in Human Extraembryonic Tissues. *Genomics,* 73 (2001), pp. 331–337.

[28] Hu, J., Balaguru, K., Ivaturi, R., Oruganti, H., and Li, T. Lack of reciprocal genomic imprinting of sense and antisense RNA of mouse insulin-like growth factor II receptor in the central nervous system. *Biochem. Biophys. Res. Commun.,* 257 (1999), pp. 604–608.

[29] Yamasaki, Y., Kayashima, T., Soejima, H., Kinoshita, A., and Yoshiura, K.-I. Neuron-specific relaxation of Igf2r imprinting is associated with neuron-specific histone modifications and lack of its antisense transcript Air. *Hum. Mol. Genet,* 14 (2005), pp. 2511–2520.

[30] Ogawa, O., McNoe, L., Eccles, M., Morison, l., and Reeve, A. Human insulin-like growth factor type I and type II receptors are not imprinted. *Hum. Mol. Genet,* 2 (1993), pp. 2163–2165.

[31] Kalscheuer, V., Mariman, E., Schepens, M., Rehder, H., and Ropers, H. The insulin-like growth factor type-2 receptor gene is imprinted in the mouse but not in humans. *Nat. Genet.,* 5 (1993), pp. 74–78.

[32] Daelemans, C., Ritchie, M., Smits, G., Abu-Amero, S., and Sudbery, I. High-throughput analysis of candidate imprinted genes and allele-specific gene expression in the human term placenta. *BMC Genet,* 11 (2010), p. 25.

[33] Xu, Y., Goodyer, C., Deal, C., and Polychronakos, C. Functional polymorphism in the parental imprinting of the human IGF2R gene. *Biochem. Biophys. Res. Commun,* 197 (1993), pp. 747–754.

[34] Pozharny, Y., Lambertini, L., Ma, Y., Ferrara, L., Litton, C. G., Diplas, A., Jacobs, A. R., Chen, J., Stone, J. L., Wetmur, J., and Lee, M. J. Genomic loss of imprinting in first-trimester human placenta. *Am. J. Obstet. Gynecol,* 202(4) (2010), pp. 391–398.

[35] Jinno, Y., Ikeda, Y., Yun, K., Maw, M., Masuzaki, H., Fukuda, H., Inuzuka, K., Fujishita, A., Ohtani, Y., and Okimoto, T. Establishment of functional imprinting of the H19 gene in human developing placentae. *Nat Genet,* 10(3) (1995), pp. 318–324.

[36] Tabano, S., Colapietro, P., Cetin, I., Grati, F., and Zanutto, S. Epigenetic modulation of the IGF2/H19 imprinted domain in human embryonic and extra-embryonic compartments and its possible role in fetal growth restriction. *Epigenetics,* 5 (2010), pp. 313–324.

[37] McCowan, L., Roberts, C., Dekker, G., Taylor, R., and Chan, E. Risk factors for small-for-gestational-age infants by customised birthweight centiles: data from an international prospective cohort study. *BJOG,* 117 (2010), pp. 1599–1607.

[38] Miller, S., Dykes, D., and Polesky, H. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res,* 16 (1988), p. 1215.

[39] Takai, D., Gonzales, F., Tsai, Y., Thayer, M., and Jones, P. Large scale mapping of methylcytosines in CTCF-binding sites in the human H19 promoter and aberrant hypomethylation in human bladder cancer. *Hum Mol Genet,* 10 (2001), pp. 2619–2626.

[40] Frost, J. and Moore, G. The Importance of Imprinting in the Human Placenta. *PLoS Genet,* 6 (2010), pp. 1001–1015.

[41] Nelissen, E., van Montfoort, A., Dumoulin, J., and Evers, J. Epigenetics and the placenta. *Hum. Reprod. Update,* 17 (2010), pp. 397–417.

[42] Ergaz, Z., Avgil, M., and Ornoy, A. Intrauterine growth restriction – etiology and consequences: What do we know about the human situation and experimental animal models? *Reprod. Toxicol,* 20 (2005), pp. 301–322.

[43] Sibai, B., Dekker, G., and Kupferminc, M. Pre-eclampsia. *Lancet,* 365 (2005), pp. 785–799.

[44] Roberts, C. IFPA Award in Placentology Lecture: Complicated interactions between genes and the environment in placentation, pregnancy outcome and long term health. *Placenta,* 31 Suppl (2010), S47–53.

[45] Wang, Y., Fu, J., Song, W., and Wang, L. Genomic imprinting status of IGF2 and H19 in placentas of fetal growth restriction patients. *J. Genet.,* 89 (2010), pp. 213–216.

[46] Lambertini, L., Diplas, A., Lee, M., Sperling, R., and Chen, J. A sensitive functional assay reveals frequent loss of genomic imprinting in human placenta. *Epigenetics,* 3 (2008), p. 261.

[47] Adam, G., Cui, H., Miller, S., Flam, F., and Ohlsson, R. Allele-specific in situ hybridization (ASISH) analysis: a novel technique which resolves differential allelic usage of H19 within the same cell lineage during human placental development. *Development,* 122 (1996), p. 839.

[48] Killian, J., Nolan, C., Wylie, A., Li, T., and Vu, T. Divergent evolution in M6P/IGF2R imprinting from the Jurassic to the Quaternary. *Hum. Mol. Genet,* 10 (2001), p. 1721.

[49] Pastinen, T. Genome-wide allele-specific analysis: insights into regulatory variation. *Nat. Rev. Genet,* 11 (2010), pp. 533–538.

[50] Andraweera, P., Dekker, G., Thompson, S., and Roberts, C. Single-nucleotide polymorphisms in the KDR gene in pregnancies complicated by gestational hypertensive disorders and small-for-gestational-age infants. *Reprod Sci,* 19 (2012), pp. 547–554.

[51] Kaffer, C., Srivastava, M., Park, K., Ives, E., and Hsieh, S. A transcriptional insulator at the imprinted H19/Igf2 locus. *Genes Dev,* 14 (2000), pp. 1908–1919.

[52] Thorvaldsen, J. and Bartolomei, M. Mothers setting boundaries. *Science,* 288 (2000), pp. 2145–2146.

[53] Arney, K. H19 and Igf2 – enhancing the confusion? *Trends Genet,* 19 (2003), pp. 17–23.

[54] Lewis, A., Mitsuya, K., Umlauf, D., Smith, P., and Dean, W. Imprinting on distal chromosome 7 in the placenta involves repressive histone methylation independent of DNA methylation. *Nat Genet,* 36 (2004), pp. 1291–1295.

[55] Mohammad, F., Pandey, G., Mondal, T., Enroth, S., and Redrup, L. Long noncoding RNA-mediated maintenance of DNA methylation and transcriptional gene silencing. *Development,* 139 (2012), pp. 2792–2803.

[56] Monk, D., Wagschal, A., Arnaud, P., Muller, P., and Parker-Katiraee, L. Comparative analysis of human chromosome 7q21 and mouse proximal chromosome 6 reveals a placental-specific imprinted gene, TFPI2/Tfpi2, which requires EHMT2 and EED for allelic-silencing. *Genome Res,* 18 (2008), pp. 1270–1281.

[57] Okae, H., Hiura, H., Nishida, Y., Funayama, R., and Tanaka, S. Re-investigation and RNA sequencing-based identification of genes with placenta-specific imprinted expression. *Hum Mol Genet,* 21 (2012), pp. 548–558.

[58] Reik, W., Santos, F., Mitsuya, K., Morgan, H., and Dean, W. Epigenetic asymmetry in the mammalian zygote and early embryo: relationship to lineage commitment? *Philos. Trans. R. Soc. Lond. B Biol. Sci,* 358 (2003), p. 1403.

[59] Diplas, A., Hu, J., Lee, M., Ma, Y., and Lee, Y. Demonstration of all-or-none loss of imprinting in mRNA expression in single cells. *Nucleic Acids Res,* 37 (2009), p. 7039.

[60] Markham, N. and Zuker, M. DINAMelt web server for nucleic acid melting prediction. *Nucleic Acids Research,* 33 (2005), W577.

[61] Kuhn, R., Karolchik, D., Zweig, A., Wang, T., and Smith, K. The UCSC genome browser database: update 2009. *Nucleic Acids Research,* 37 (2009), p. D755.

[62] Kent, W. BLAT: the BLAST-like alignment tool. *Genome Research,* 12 (2002), p. 656.

Statement of Authorship

Title of Paper	Imprinted and X-linked non-coding RNAs as potential regulators of human placental function
Publication Status	◉ Published, ○ Accepted for Publication, ○ Submitted for Publication, ○ Publication style
Publication Details	Buckberry, S., Bianco-Miotto, T. & Roberts, C. T. Imprinted and X-linked non-coding RNAs as potential regulators of human placental function. Epigenetics 9, 81–89 (2014).

Author Contributions

By signing the Statement of Authorship, each author certifies that their stated contribution to the publication is accurate and that permission is granted for the publication to be included in the candidate's thesis.

Name of Principal Author (Candidate)	Sam Buckberry		
Contribution to the Paper	Sam Buckberry carried out the literature review, performed the computational analyses, authored the first draft of the manuscript, incorporated suggestions from co-authors and peer-reviewers and wrote the final manuscript.		
Signature		Date	7/4/15

Name of Co-Author	Tina Bianco-Miotto		
Contribution to the Paper	Tina Bianco-Miotto provided detailed comments for all manuscript drafts.		
Signature		Date	1/4/15

Name of Co-Author	Claire T Roberts		
Contribution to the Paper	Claire T Roberts provided detailed comments for all manuscript drafts.		
Signature		Date	7.4.15

Name of Co-Author			
Contribution to the Paper			
Signature		Date	

Chapter 3

Imprinted and X-Linked Non-Coding RNAs as Potential Regulators of Human Placental Function

Sam Buckberry, Tina Bianco-Miotto and Claire T Roberts

Abstract

Pregnancy outcome is inextricably linked to placental development, which is strictly controlled temporally and spatially through mechanisms that are only partially understood. However, increasing evidence suggests non-coding RNAs (ncRNAs) direct and regulate a considerable number of biological processes and therefore may constitute a previously hidden layer of regulatory information in the placenta. Many ncRNAs, including both microRNAs and long non-coding transcripts, show almost exclusive or predominant expression in the placenta compared to other somatic tissues and display altered expression patterns in placentas from complicated pregnancies. In this review we explore the results of recent genome-scale and single gene expression studies using human placental tissue but include studies in the mouse where human data are lacking. Our review focuses on the ncRNAs epigenetically regulated through genomic imprinting or X-chromosome inactivation and includes recent evidence surrounding the *H19* lincRNA, the

imprinted C19MC cluster microRNAs, and X-linked miRNAs associated with pregnancy complications.

3.1 Introduction

The best-known function of the placenta is to mediate fetal-maternal exchange throughout pregnancy but it also plays a major role in directing maternal adaptation to pregnancy by secreting a variety of steroid and peptide hormones that modulate maternal physiology without which pregnancy could not be sustained. The placenta is a unique organ in several respects. Firstly, although the placenta is a shared organ between mother and fetus, it is an extra-embryonic tissue and is therefore primarily regulated by the fetal genome. Secondly, the placenta completely separates from mother and fetus after birth, making it the only truly transient organ. For this reason, the placenta may not be under the same lifetime epigenetic constraints as other somatic tissues. Placental development in humans begins shortly after an embryo implants into the lining of the uterus where it begins a strikingly invasive process that remodels the uterine spiral arterioles to sequester a maternal blood supply to facilitate efficient feto-maternal exchange. This invasive process, which has many similarities with cancer metastasis [1], appears to be strictly controlled both spatially and temporally in humans through mechanisms that are only partially understood [2, 3]. However, emerging evidence, particularly from high-throughput gene expression technologies, suggests non-coding RNA molecules (ncRNAs) direct and regulate a considerable number of biological processes and cellular functions. Therefore ncRNAs may constitute a previously hidden layer of regulatory information in the placenta. In this review, we focus on the imprinted and X-linked ncRNAs, which are typically expressed from only one allele. We explore the regulation of these ncRNAs in the context of human placental development. Examining particular influential genomic regions, a key focus of this review will be the role that ncRNA expression in the placenta plays in pregnancy complications such as preeclampsia that are attributed to abnormal placental development. Although this review is focused on human placental development, studies in the mouse are also discussed where human data are lacking.

3.2 The Placenta is Key to Fetal and Maternal Health

The placenta is part of the conceptus and therefore is genetically identical to the fetus. Its development is initiated at implantation about 5–6 days after conception and follows a dynamic and constantly changing trajectory providing gaseous and nutrient exchange functions between the maternal and fetal circulations to support fetal growth [4]. Impaired placental (trophoblast) invasion has been implicated in several complications of pregnancy such as preeclampsia and intrauterine growth restriction (IUGR) [5] and pre-term labour [6, 7]. For example, in preeclampsia invasion of the spiral arterioles and the maternal decidual stroma is shallow, resulting in poor maternal blood flow to the placenta [5, 8, 9]. Despite huge research efforts, our understanding of the highly complex molecular regulation of both normal and abnormal placentation is still inadequate. However, ncRNAs are emerging as key regulators of development [10, 11], and therefore provide new avenues of inquiry relating to placental differentiation and function. If so, the perturbed regulation of ncRNAs in the placenta may result in one or more of a continuum of pregnancy complications that compromise the health of both mother and infant.

3.3 Classification and Detection of Non-Coding RNA

There are many different classes of ncRNAs, as these molecules vary greatly with regards to sequence length and complexity, splicing isoforms, polyadenylation, regulation and biological function. The most well-characterised class of ncRNAs are the infrastructural RNAs (rRNAs, tRNAs, snRNAs, snoRNAs), which constitute many integral cellular components and are involved in processes such as translation, transcript splicing and higher level regulatory processes including DNA methylation [10]. Other ncRNAs are typically classed based on their sequence length, with RNAs shorter than $\sim$ 200 nucleotides termed short non-coding RNAs (sncRNAs), and those greater than $\sim$ 200 nucleotides are termed long non-coding RNAs (lncRNAs) [11]. The sncRNAs include the microRNAs (miRNAs), piwi interacting RNAs (piRNAs) and the small interfering RNAs (siRNAs). These short RNAs, particularly the miRNAs, have received the most attention to date, and currently dominate the ncRNA literature. However, there has been a steady accumulation of evidence indicating that lncRNA transcripts, as a class, have a diverse

repertoire of biological functions [11] and constitute a significant proportion of total cellular RNA [12].

Although the central dogma of biology has previously allowed little scope for the regulatory capabilities of ncRNA (for a review see ref. [13]), the ability to detect and measure ncRNAs has also hindered progress towards appreciating the gamut of their functional abilities. Detecting ncRNAs in any tissue has posed challenges for several reasons. Firstly, distinguishing if a transcript has protein-coding ability can be difficult as ncRNA transcripts can originate from intronic and untranslated regions of coding transcripts, or can be alternative splice variations that abolish a transcript's coding potential [14, 15]. Secondly, lncRNAs can be transcribed from DNA that spans intergenic and coding regions resulting in transcripts that host protein-coding DNA sequence. Thirdly, many ncRNAs do not end with a poly-A signal [12] which is characteristic of protein coding genes. This difference has profound implications regarding ncRNA detection as many cDNA, SAGE, microarray and RNA-Seq methods rely on poly-A labelling, enrichment or priming. For these reasons among several others (see ref. [16]), ncRNA transcripts can be difficult to discover and measure, which has subsequently hampered our ability to annotate and functionally classify ncRNAs in health and disease.

3.4 The Roles of Non-Coding RNAs in Genomic Imprinting in the Placenta

Imprinted genes are known to have significant effects on placental development and are implicated in many placental pathologies [17–19]. Imprinted genes are expressed in a parent-of-origin-dependent manner, with the imprinted alleles being epigenetically silenced [20, 21]. Genomic imprinting is typically observed in clusters of $\sim$ 2–12 genes, with most of these clusters having at least one lncRNA gene [22]. The epigenetic regulation of imprinting can involve DNA methylation imprints, repressive histone modifications, and complex enhancer competition scenarios involving *cis*-acting lncRNA transcripts [22–25]. Ablation of lncRNAs within imprinting clusters typically results in the loss of imprinting [22], demonstrating that lncRNAs can act as *cis* regulators of autosomal gene expression.

Imprinting is largely, although not exclusively, observed in eutherian mammals and is thought to have arisen with viviparity and the evolutionary emergence of the placenta [26, 27]. The prevailing evolutionary hypothesis of imprinting suggests that paternally-expressed genes have been selected to maximise fetal resource acquisition from the mother, while maternally-expressed genes have been

selected to balance resources allocated to current and future offspring [27]. Since imprinted genes are suggested to facilitate a *tug-of-war* between maternal and paternal genomes, this hypothesis predicts that imprinted genes are heavily involved in fetal and placental growth and development throughout pregnancy [21, 27, 28]. Not surprisingly, more imprinted genes are expressed in the placenta than in any other tissue, with several being placenta specific [29].

Although the exact mechanisms regulating imprinted regions remain unclear, the maintenance of imprints appears to differ between embryonic and extra-embryonic tissues [29]. This suggests that extra embryonic cell lineages, many of which make up the placenta, may employ regulatory mechanisms involving ncRNAs that are not observed in embryonic cell lineages. Despite the fact that much of our understanding of placental imprinting comes from studies in mice [29], the evidence from human research to date suggests that many human placental abnormalities and pregnancy complications are associated with altered imprinting involving ncRNAs.

3.5 The Imprinted *H19* long Non-Coding RNA and miR-675

H19 was one of the first lncRNAs to be discovered and is considered a key regulatory molecule in placental development. *H19* lies within a large imprinted domain (>1 MB), and is predominantly expressed from the maternal chromosome. *H19* placental expression is largely monoallelic [30] and is one of the most highly expressed genes in the human placenta [31]. However, the functional roles of *H19* are only now beginning to emerge.

H19, and the adjacent and reciprocally-imprinted *IGF2* gene, make up one of the most widely studied imprinted genomic regions in humans. Both *H19* and *IGF2* share many *cis*-regulatory elements, with the prevailing regulatory model of this locus indicating a complex interaction of DNA methylation, CTCF binding and enhancer competition scenarios mainly elucidated through targeted deletion and transgenic techniques in murine models [32].

Somewhat consistent with observations in humans, studies in mice have further demonstrated that altered imprinting of *H19* is associated with placental and fetal growth abnormalities [32–34]. For example, (epi)mutations in the *H19-IGF2* region are associated with Silver-Russell and Beckwith-Wiedemann syndromes,

which manifest phenotypically *in utero* as severe growth-restriction and overgrowth, respectively [35]. Furthermore, altered epigenetic regulation of the *H19-IGF2* region in human placentas has been associated with the pregnancy complication preeclampsia, which is attributed to abnormal placental development early in gestation [36, 37]. Biallelic expression of *H19* has been observed at higher rates during the first trimester of pregnancy compared to term [36, 38, 39] with the early first trimester placenta showing patterns of imprinting plasticity [30]. Together, these studies suggest *H19* plays an important regulatory role in early placental development.

Recent work suggests that *H19* is a *trans* regulator of an imprinted gene network for growth and development [40] involving miRNAs hosted within the *H19* transcript [32, 41, 42], which may account for some of *H19*'s bioactivity. Most recently, Keniry et al. have described *H19* as a developmental reservoir of miR-675 in the mouse [43]. This study shows the miR-675 microRNA is processed from the first exon of *H19* in a developmental stage specific manner in the placenta. They also showed that levels of miR-675 increased with gestation acting as a placental growth suppressor [43]. Although overall *H19* expression remained unchanged throughout gestation, the RNA-binding protein Elavl1 (also known as *HuR*) appeared to bind to the *H19* transcript preventing excision of miR-675 early in gestation [43]. Elavl1 abundance decreased as gestation progressed, enabling miR-675 to be processed and to act as a placental growth suppressor [43]. Although this study has increased our knowledge of *H19* function in the placenta, it may not accurately portray the situation in humans for several reasons. Firstly, the human and mouse *H19* transcripts show notable sequence divergence. Secondly, a microarray analysis by Sitras et al. found no significant difference in *ELAVL1* expression between first trimester and third trimester human placentas [44], (Figure 3.1A) which is contrary to the observation in mice. However, as suggested by Keniry et al., the excision of miR-675 may also be regulated by additional RNA binding proteins [43]. To examine this possibility, we performed an *in silico* analysis of RNA binding protein domains within the human and mouse *H19* transcripts. We note that the ELAVL1 binding sites that flank the miR-675 locus in mouse are not present in the human transcript (Figure 3.1A). However, we observed that binding domains flanking miR-675 existed for the RNA binding proteins NONO and YBX1 in the human *H19* transcript, and these proteins show a significant decrease in expression as gestation progresses (Figure 3.1B). These differences between mouse and human indicate further work is required to elucidate the true extent of *H19* and miR-675 regulation and functionality in the human placenta. This would require miR-675 expression across human gestation to be evaluated, followed by a careful analysis of how miR-675 excision is repressed in humans. These experiments using

human derived samples will be a fundamental step towards determining why *H19* is implicated in human pregnancy complications attributed to abnormal placental development.

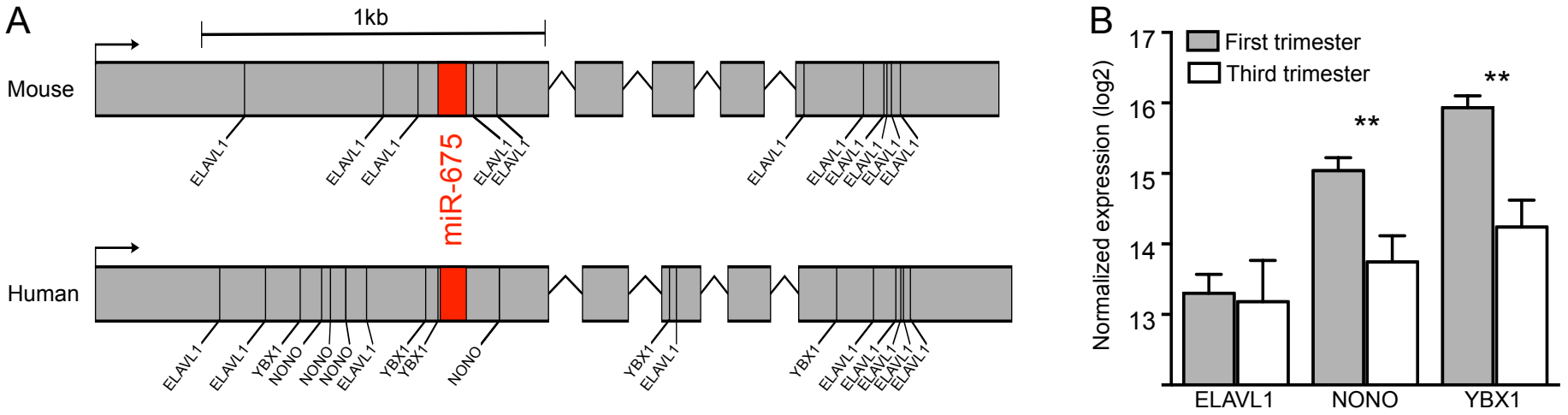


FIGURE 3.1: The *H19* lncRNA transcript and RNA binding proteins in human and mouse. (A) Schematic representation of human and mouse *H19* transcripts indicating locations of RNA binding protein motifs. The ELAVL1 binding motifs that surround the miR-675 locus in the mouse transcript are not present in the human transcript, however binding motifs for the NONO and YBX1 proteins are present in the human transcript. (B) Expression of genes that encode the RNA binding proteins *ELAVL1*, *NONO* and *YBX1* in human first and third trimester placentas. Expression of *NONO* and *YBX1* show a significant decrease in expression as gestation progresses, with *ELAVL1* showing no difference across gestation. Data for RNA binding protein expression differences between first and third trimester were reported in ref. [44] and the figures were generated using normalized array data obtained from the NCBI Gene Expression Omnibus (accession GSE28551). The RNA binding protein recognition sequences were predicted using the RNA binding protein database [45]. Human and mouse *H19* transcript sequences obtained from UCSC reference genomes hg19 and mm10 respectively.

3.6 The Imprinted C19MC miRNA Cluster

An intriguing observation of placental-expressed miRNAs arises from the largest known miRNA cluster discovered to date; C19MC. This cluster, located at human chromosome 19 (19q13.42), features ∼46 miRNA genes transcribed from a ∼100kb region. C19MC is imprinted, with only the paternally-inherited chromosome being expressed [46, 47] predominantly, if not exclusively, in the placenta [48]. Furthermore, C19MC is unique to the primate lineage, excluding model organism research to determine the functions of miRNAs in this cluster [48].

Transcription of C19MC miRNAs can be activated in some cells by treatment with DNA methylation inhibitors indicating that the region is under DNA methylation dependent epigenetic control [46, 49, 50]. Further evidence also suggests that the C19MC miRNAs are excised from a much larger lncRNA, which is transcribed from an RNA Pol II promoter within a CpG-rich region [46, 48]. C19MC miRNAs are also expressed in exosomes released from primary human trophoblast cells and are detectable in the serum of pregnant women [51], highlighting their potential as fetal maternal signalling molecules that may modulate maternal adaptation to pregnancy. Although the precise functional mechanisms of C19MC miRNAs remain largely unknown, the abundance of C19MC transcripts in the placenta, their imprinted regulation, and detection in the maternal circulation, suggest a significant role in placental biology.

Studies of pregnancy complications attributed to abnormal placental development, in particular those focusing on placental gene expression in preeclampsia where a transcriptome-wide method (microarrays or high-throughput RNA sequencing) is employed, have shown differential regulation in the placental expression of some miRNAs in the C19MC family (Table 3.1) [52–55].

Together, these studies have identified 21 miRNAs with increased placental expression in preeclampsia and/or pre-term birth when compared to normal pregnancies, with eight of these miRNAs showing increased expression in at least two studies (Table 3.1). Although empirical evidence is currently lacking for the targets of many C19MC miRNAs, miR-520g and miR-520h have been shown experimentally to repress expression of *VEGF*, an angiogenic gene implicated in preeclampsia [56]. Furthermore, expression of the VEGF receptor gene, *FLT1*, has also shown consistently higher expression in placentas from preeclamptic pregnancies [57–63]. Additionally, the cell cycle inhibitor and apoptosis associated *CDKN1A* (p21) gene is a validated target of several C19MC miRNAs differentially expressed in preeclampsia (Table 3.1), which further implicates these miRNA genes given the links preeclampsia has with perturbed apoptosis [64].

These results are of particular interest for several reasons, particularly with respect to preeclampsia. Firstly, as the C19MC region is imprinted, increased miRNA expression may result from loss of imprinting, and the loss of imprinting of other placental ncRNAs such as *H19* has known associations with preeclampsia [36]. Secondly, C19MC allelic repression is regulated by DNA methylation imprints [46], and alteration to DNA methylation in the placenta is also associated with preeclampsia [65]. Thirdly, at least two miRNAs in the C19MC cluster target the *VEGFA* gene that is closely networked to the *FLT1* gene. Fourthly, at least three C19MC miRNAs have been shown to target the ELAVL1 gene, which may be involved in regulating miR-675 excision from *H19* transcript (see discussion of ref. [43] above).

The differential expression of several C19MC miRNAs in the placenta is also associated with preterm labour [66]. Preterm birth and subsequent preterm delivery allows the investigation of placental gene expression at a much earlier time-point than normal laboured deliveries. As such, the changes in miRNA expression may simply reflect the changing patterns of C19MC miRNAs as gestation progresses, and not be indicative of pathology. However, expression of C19MC miRNAs in cells derived from matched placentas sampled during both first trimester and term is comparable [47], suggesting the aberrant C19MC regulation in preterm birth is not due to developmental stage differences. Together, these findings warrant further inquiry into the biological role of the C19MC miRNAs, particularly in identifying their regulatory potential, as this increased understanding could reveal novel therapeutic targets.

TABLE 3.1: Imprinted or X-linked miRNA differentially expressed in preeclampsia (PE) and pre-term birth (PTB) with validated targets and potential mechanisms

Pregnancy complication	miRNA (cytoband)	Expression in complication vs. control	Experimentally validated target genes [67, 68] detectable in the human placenta [31]	Potential roles and contributing mechanisms
PE	miR-20b [54, 55, 69] (Xq26.2)	Increased	*ARID4B, BAMBI, CDKN1A, CRIM1, ESR1, HIF1A, HIPK3, MYLIP, PPARG, STAT3, VEGFA*	Impairing placental function through suppression of genes (*CRIM1, HIF1A, VEGFA*) that have a role in maintaining endothelial cell function and angiogenesis [70–73]. Repression of genes (*CDKN1A, HIPK3, STAT3*) involved in apoptosis and trophoblast invasion [74–76].
PE	miR-222 [55, 77] (Xq11.3)	Increased	*BBC3, BCL2L11, CDKN1B, CDKN1C, CORO1A, ESR1, FOS, FOXO3, ICAM1, MMP1, PPP2R2A, PTEN, SOD2, SSSCA1, STAT5A, TCEAL1, TNFSF10, TP53*	Down-regulation of genes (*BBC3, BCL2, CORO1A, FOS, FOXO3, TNFSN10, TP53*) that can promote apoptosis [78–83]. Down regulation of genes (*ICAM1*) involved in endothelial cell function [84]. Down regulation of genes (*CDKN1B, CDKN1C*) that regulate cell cycle progression and trophoblast differentiation [85].

Continued on next page

Table 3.1 – continued from previous page

Pregnancy complication	miRNA (cytoband)	Expression in complication vs. control	Experimentally validated target genes [67, 68] detectable in the human placenta [31]	Potential roles and contributing mechanisms
PE	miR-223 [53, 54] (Xq12)	Decreased	*CHUK, Il6, IRS1, LMO2, NFIX, RHOB, STMN1*	Up-regulation of a gene (*RHOB*) involved in apoptosis signaling [86]. Up-regulation of a gene (*IL6*) involved with immune response and inflammation [87].
PE	miR-519b [54, 55] (19q13.42)	Increased	*CDKN1A, ELAVL1*	Alteration of apoptosis signals (*CDKN1A*) [74]. Down-regulation of *ELAVL1*, potentially altering miR-675 excision from the *H19* lincRNA [43].
PE	miR-519e [53, 54] (19q13.42)	Increased	*CDKN1A*	Alteration of apoptosis signals (*CDKN1A*) [74].
PE	miR-520g [54, 55] (19q13.42)	Increased	*VEGFA*	Impaired endothelial cell function and angiogenesis (*VEGFA*) [88].
PE	miR-524 [54, 69] (19q13.42)	Increased	–	–
PE/PTB	miR-517a [53, 66] (19q13.42)	Increased	–	Regulation of apoptosis [89]
PE/PTB	miR-518b [53, 54, 66, 90] (19q13.42)	Increased	–	–

Continued on next page

Table 3.1 – continued from previous page

Pregnancy complication	miRNA (cytoband)	Expression in complication vs. control	Experimentally validated target genes [67, 68] detectable in the human placenta [31]	Potential roles and contributing mechanisms
PE/PTB	miR-520h [54, 90] (19q13.42)	Increased	*ABCG2, CDKN1A, ID1, ID3, SMAD6, VEGFA*	Down-regulation of a gene (*ABCG2*) involved in protecting fetal exposure to xenobiotics ingested by the mother [91]. Alteration of apoptosis signals (*CDKN1A*) [74]. Impaired endothelial cell function and angiogenesis (*VEGFA*) [88].
PE/PTB	miR-526b [54, 66] (19q13.42)	Increased	–	–

3.7 X-Chromosome ncRNAs in Placental Development and Pregnancy Complications

During intrauterine development, there are distinct sex differences in fetal growth trajectories and hence in birth weight [92–94], with a sex bias in the prevalence of preterm birth [95, 96], pregnancy complications such as preeclampsia [97, 98] and perinatal death [99]. As fetal growth and development are highly dependent on the exchange efficiency and capacity of the placenta, sex-specific differences in normal and pathological fetal development are most likely due to sex differences in placental function. Recent work has shown that there is a distinct male sex bias in the prevalence of placental dysfunction [98], which supports the findings of previous studies [100–104] showing sex biases in a spectrum of pregnancy conditions and fetal health associated with abnormal placentation. However, the underlying mechanisms that predispose one sex over the other to deviate from a normal course of fetal development remain unknown.

Since the majority of genes are autosomal, many sexually dimorphic traits are driven by the sex-biased expression of autosomal genes [105]. For decades, much of the scientific literature has solely attributed the influence of sex hormones to sexual dimorphism, yet increasing evidence suggests sex chromosome genes are also implicated in the regulation of autosomal gene expression. (for a review see ref. [105]) While many sex specific gene expression differences have been appreciated for some time, their phenotypic and clinical implications, particularly in the placenta and in pregnancy complications, remain relatively unexplored [106–109].

3.8 X-Linked miRNAs as Potential Drivers of Sex Differences in Placental Gene Expression

When comparing the X chromosome to the 22 autosomes, the human X chromosome (with 140 annotated miRNAs in miRBase [110]) appears to be enriched for miRNA genes when considering its size and genomic content. Only chromosome 1, which features eleven more miRNA genes than the X chromosome, is richer in miRNA content and is ∼100MB larger in size. In contrast, the Y chromosome has only two annotated miRNA genes in a pseudo-autosomal region, which undergoes recombination with the X chromosome.

The observation that X chromosomes have high miRNA gene content highlights the potential of X-linked miRNAs to contribute to sex-biased autosomal gene expression. As X-linked miRNAs can potentially target multiple autosomal genes, sex-biased expression of X linked miRNAs could trigger cascade-like effects, potentially driving sex-biased expression of many autosomal genes.85 Additionally, ∼35 X-linked miRNAs are located within introns of protein-coding genes which are likely to share transcriptional elements with their host genes, potentially resulting in co-regulation. For example, the X-linked gene *CHM* which hosts miR-361 is known to escape X chromosome inactivation (XCI), which could therefore lead to sex-biased expression of miR-361, and through autosomal gene targeting result in sex biased expression of autosomal genes.

3.9 Previous Studies Provide Few Clues to Which miRNAs Escape XCI

Although X-linked miRNAs are potential drivers of sex differences in placental development and function, there is little evidence to suggest which, if any, of these miRNAs actually escape XCI, leading to biallelic (and potentially increased) expression in females. The inactivation of X chromosomes in human extra-embryonic tissues has been a topic of continual research over the past three decades, although the status and extent of XCI in the human placenta remains controversial. Most early studies focused on allele-specific expression from the *G6PD* (Xq28) locus to determine if the X chromosome was randomly inactivated or skewed towards paternal or maternal XCI, and showed that XCI varies notably across samples, with patterns of random and skewed X-inactivation [111–115]. In later attempts to clarify the status of XCI in the human placenta, research shifted to different loci, again showing mixed results of skewed and random XCI [116–121]. Taken together, these results suggest a high degree of XCI heterogeneity in the extra-embryonic tissues of female fetuses. Although only assessing genes on the q-arm of the X chromosome, these studies suggest that when XCI deviates from random, it is the paternal X chromosome that is most often inactivated.

What has become increasingly apparent over the past decade however, is that multiple X chromosome regions escape inactivation, and that these extend beyond chromosomal regions with Y chromosome homologues [106, 122]. As the results of any XCI assay are dependent on the loci under investigation, it now appears spurious to infer the regulation of a whole chromosome (or region) based on the assessment of one or two X chromosome loci. In an attempt to widen the scope of

our understanding of XCI in the human placenta, recent work has demonstrated allele-specific expression profiles of 22 genes spread across the X chromosome [123]. The results of this most comprehensive placental study to date suggest that XCI in the human placenta is random, with localised mosaic patterns of maternal and paternal XCI [123]. However, given the number of samples and the methodological limitations [124–126], placental XCI studies still lack the depth of XCI research conducted using other human tissues. Subsequently, we have very little indication of what ncRNAs, particularly miRNAs, escape XCI and potentially contribute to sex-biased placental gene in normal and complicated pregnancies.

3.10 X-Linked miRNAs Associated with Preeclampsia

Recent studies have shown many X chromosome miRNAs that occur in clusters are differentially regulated in placentas from preeclamptic pregnancies (Figure 3.2). When summarising the results of these studies, increased expression of miR-20b [54, 55, 69] and miR-222 [55, 77] and decreased expression of miR-223 [53, 54] is supported by two or more studies. Of particular interest, these miRNA have been shown experimentally to target multiple genes involved in processes such as apoptosis, angiogenesis and immune response (Table 3.1), all of which are implicated in the pathogenesis of preeclampsia [64, 70, 127].

There is also evidence implicating the miRNA cluster at Xq26.3 which flanks the placenta-specific *PLAG1* gene. This cluster contains six miRNAs, four of which (miR-424, miR-542, miR-450a-1 and miR-450b) have been shown to be differentially regulated in placentas from preeclamptic pregnancies [52–55]. Curiously, at the individual miRNA level, miR-424 is up-regulated in preeclampsia, while miR-542 and miR-450b are down-regulated, and for miR-450a-1 the data are conflicting [53, 55]. miR-424 is an interesting case since it overlaps the transcription start site of MGC16121, a lincRNA that is virtually unstudied. Additionally, miR-424 is a hypoxia-induced regulator of *HIF1A* and involved in angiogenesis [128], highlighting its potential role in preeclampsia.

Although many of these miRNAs are in close proximity to genes shown to escape XCI (Figure 3.2), sample sex information would be required to determine if these X-linked miRNA expression differences are indeed sex related and resulting from differential X chromosome epigenetic regulation. However tenuous the link between XCI and expression differences in preeclampsia, the preliminary evidence discussed here justifies further investigation. Future work focused on delineating

the boundaries of XCI in the human placenta and validating the targets of miRNAs that escape XCI may provide clues to the mechanisms giving rise to sex-biases in placental development and the downstream implications for adverse pregnancy outcomes such as preeclampsia.

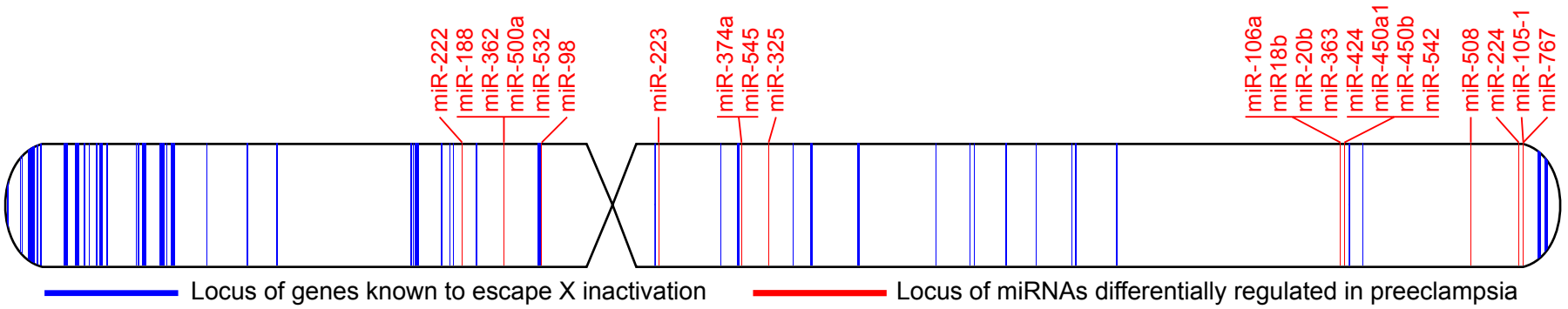


FIGURE 3.2: Idiogram representation of the human X chromosome showing the X-linked miRNAs associated with preeclampsia (red) occur in clusters in close proximity of genes that escape X inactivation (blue). The data for miRNAs showing altered expression in preeclampsia were derived from refs. [52–55, 77, 129] and the data for genes that escape inactivation were adapted from ref. [106] with genomic coordinates converted to hg19 coordinates using the UCSC liftOver tool.

3.11 Conclusions

Non-coding RNAs are increasingly implicated in many developmental and pathological processes; placental development is no exception. Although much research points towards the central role of ncRNAs in placental development and function, large gaps in our knowledge remain. In particular, the ncRNAs under epigenetic regulatory control through mechanisms such as XCI of genomic imprinting appear to be influential players. However, many questions remain regarding the functional actions of these transcripts and whether their change in expression in associated pregnancy complications is a cause or consequence. In either case, increasing our understanding of the epigenetically regulated ncRNAs in normal placental development is essential if these perplexing molecules are ever to be used as diagnostic or predictive biomarkers.

Acknowledgements

SB is supported by a Healthy Development Adelaide & Channel 7 Children's Research Foundation PhD Scholarship and an Australian Postgraduate Award. TBM is supported by the Cancer Council SA and SAHMRI Beat Cancer Project (TBM APP1030945). CTR is supported by a National Health and Medical Research Council (NHMRC) Senior Research Fellowship APP1020749. The research described herein was funded by NHMRC Research Project #565320 (http://www.nhmrc.gov.au/). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Bibliography

[1] Murray, M. J. and Lessey, B. A. Embryo implantation and tumor metastasis: common pathways of invasion and angiogenesis. *Semin Reprod Endocrinol,* 17(3) (1999), pp. 275–90.

[2] Graham, C. H. and Lala, P. K. Mechanisms of placental invasion of the uterus and their control. *Biochem Cell Biol,* 70(10-11) (1992), pp. 867–74.

[3] Lyall, F. The human placental bed revisited. *Placenta,* 23(8-9) (2002), pp. 555–62.

[4] Roberts, C. IFPA Award in Placentology Lecture: Complicated interactions between genes and the environment in placentation, pregnancy outcome and long term health. *Placenta,* 31 Suppl (2010), S47–53.

[5] Khong, T. Y., De Wolf, F., Robertson, W. B., and Brosens, I. Inadequate maternal vascular response to placentation in pregnancies complicated by pre-eclampsia and by small-for-gestational age infants. *BJOG: An International Journal of Obstetrics & Gynaecology,* 93(10) (1986), pp. 1049–1059.

[6] Kim, Y. M., Bujold, E., Chaiworapongsa, T., Gomez, R., Yoon, B. H., Thaler, H. T., Rotmensch, S., and Romero, R. Failure of physiologic transformation of the spiral arteries in patients with preterm labor and intact membranes. *American Journal of Obstetrics and Gynecology,* 189(4) (2003), pp. 1063–1069.

[7] Kim, Y. M., Chaiworapongsa, T., Gomez, R., Bujold, E., Yoon, B. H., Rotmensch, S., Thaler, H. T., and Romero, R. Failure of physiologic transformation of the spiral arteries in the placental bed in preterm premature rupture of membranes. *American Journal of Obstetrics and Gynecology,* 187(5) (2002), pp. 1137–1142.

[8] Rinkenberger, J. L., Cross, J. C., and Werb, Z. Molecular genetics of implantation in the mouse. *Developmental genetics,* 21(1) (1997), pp. 6–20.

[9] Zhou, Y., Genbacev, O., Damsky, C. H., and Fisher, S. J. Oxygen regulates human cytotrophoblast differentiation and invasion: implications for endovascular invasion in normal pregnancy and in pre-eclampsia. *Journal of Reproductive Immunology,* 39(1-2) (1998), pp. 197–213.

[10] Mattick, J. S. and Makunin, I. V. Non-coding RNA. *Hum Mol Genet,* 15 Spec No 1 (2006), R17–29.

[11] Mercer, T. R., Dinger, M. E., and Mattick, J. S. Long non-coding RNAs: insights into functions. *Nat Rev Genet,* 10(3) (2009), pp. 155–9.

[12] Kapranov, P., St Laurent, G., Raz, T., Ozsolak, F., Reynolds, C. P., Sorensen, P. H., Reaman, G., Milos, P., Arceci, R. J., Thompson, J. F., and Triche, T. J. The majority of total nuclear-encoded non-ribosomal RNA in a human cell is 'dark matter' un-annotated RNA. *BMC Biol,* 8 (2010), p. 149.

[13] Mattick, J. S. Challenging the dogma: the hidden layer of non-protein-coding RNAs in complex organisms. *Bioessays,* 25(10) (2003), pp. 930–9.

[14] Moran, V. A., Perera, R. J., and Khalil, A. M. Emerging functional and mechanistic paradigms of mammalian long non-coding RNAs. *Nucleic Acids Res,* 40(14) (2012), pp. 6391–400.

[15] Frankish, A., Mudge, J. M., Thomas, M., and Harrow, J. The importance of identifying alternative splicing in vertebrate genome annotation. *Database (Oxford),* 2012 (2012), bas014.

[16] Mercer, T. R., Gerhardt, D. J., Dinger, M. E., Crawford, J., Trapnell, C., Jeddeloh, J. A., Mattick, J. S., and Rinn, J. L. Targeted RNA sequencing reveals the deep complexity of the human transcriptome. *Nat Biotechnol,* 30(1) (2012), pp. 99–104.

[17] Fowden, A. L., Coan, P. M., Angiolini, E., Burton, G. J., and Constancia, M. Imprinted genes and the epigenetic regulation of placental phenotype. *Prog Biophys Mol Biol,* 106(1) (2011), pp. 281–8.

[18] Frost, J. M. and Moore, G. E. The importance of imprinting in the human placenta. *PLoS Genet,* 6(7) (2010), e1001015.

[19] Bressan, F. F., De Bem, T. H., Perecin, F., Lopes, F. L., Ambrosio, C. E., Meirelles, F. V., and Miglino, M. A. Unearthing the roles of imprinted genes in the placenta. *Placenta,* 30(10) (2009), pp. 823–34.

[20] Ferguson-Smith, A. C. and Surani, M. A. Imprinting and the epigenetic asymmetry between parental genomes. *Science,* 293(5532) (2001), pp. 1086–9.

[21] Reik, W. and Walter, J. Genomic imprinting: parental influence on the genome. *Nat. Rev. Genet.,* 21(1) (2001), pp. 21–32.

[22] Koerner, M. V., Pauler, F. M., Huang, R., and Barlow, D. P. The function of non-coding RNAs in genomic imprinting. *Development,* 136(11) (2009), pp. 1771–83.

[23] Constancia, M., Pickard, B., Kelsey, G., and Reik, W. Imprinting mechanisms. *Genome Res,* 8(9) (1998), pp. 881–900.

[24] Kacem, S. and Feil, R. Chromatin mechanisms in genomic imprinting. *Mamm Genome,* 20(9-10) (2009), pp. 544–56.

[25] Koerner, M. V. and Barlow, D. P. Genomic imprinting-an epigenetic gene-regulatory model. *Curr Opin Genet Dev,* 20(2) (2010), pp. 164–70.

[26] Monk, D., Arnaud, P., Apostolidou, S., Hills, F. A., Kelsey, G., Stanier, P., Feil, R., and Moore, G. E. Limited evolutionary conservation of imprinting in the human placenta. In: *Proc. Natl. Acad. Sci.,* vol. 103. 2006, pp. 6623–6628.

[27] Haig, D. Altercation of generations: genetic conflicts of pregnancy. *Am J Reprod Immunol,* 35(3) (1996), pp. 226–232.

[28] Haig, D. Genetic Conflicts in Human Pregnancy. *Q Rev. Biol.,* 68 (1993), pp. 495–532.

[29] Wagschal, A. and Feil, R. Genomic imprinting in the placenta. *Cytogenet Genome Res,* 113(1-4) (2006), pp. 90–8.

[30] Buckberry, S., Bianco-Miotto, T., Hiendleder, S., and Roberts, C. T. Quantitative allele-specific expression and DNA methylation analysis of H19, IGF2 and IGF2R in the human placenta across gestation reveals H19 imprinting plasticity. *PLoS One,* 7(12) (2012), e51210.

[31] Kim, J., Zhao, K., Jiang, P., Lu, Z., Wang, J., Murray, J. C., and Xing, Y. Transcriptome landscape of the human placenta. *BMC Genomics,* 13 (2012), p. 115.

[32] Gabory, A., Ripoche, M., Yoshimizu, T., and Dandolo, L. The *H19* gene: regulation and function of a non-coding RNA. *Cytogenet. Genome Res.,* 113 (2006), pp. 188–193.

[33] Lau, M. M., Stewart, C. E., Liu, Z., Bhatt, H., Rotwein, P., and Stewart, C. L. Loss of the imprinted IGF2/cation-independent mannose 6-phosphate receptor results in fetal overgrowth and perinatal lethality. *Genes Dev,* 8(24) (1994), pp. 2953–63.

[34] Constancia, M., Hemberger, M., Hughes, J., Dean, W., Ferguson-Smith, A., Fundele, R., Stewart, F., Kelsey, G., Fowden, A., Sibley, C., and Reik, W. Placental-specific IGF-II is a major modulator of placental and fetal growth. *Nature,* 417(6892) (2002), pp. 945–948.

[35] Hirasawa, R. and Feil, R. Genomic imprinting and human disease. *Essays Biochem,* 48(1) (2010), pp. 187–200.

[36] Yu, L., Chen, M., Zhao, D., Yi, P., Lu, L., Han, J., Zheng, X., Zhou, Y., and Li, L. The H19 gene imprinting in normal pregnancy and pre-eclampsia. *Placenta,* 30(5) (2009), pp. 443–7.

[37] Gao, W. L., Li, D., Xiao, Z. X., Liao, Q. P., Yang, H. X., Li, Y. X., Ji, L., and Wang, Y. L. Detection of global DNA methylation and paternally imprinted H19 gene methylation in preeclamptic placentas. *Hypertens Res,* 34(5) (2011), pp. 655–61.

[38] Pozharny, Y., Lambertini, L., Ma, Y., Ferrara, L., Litton, C. G., Diplas, A., Jacobs, A. R., Chen, J., Stone, J. L., Wetmur, J., and Lee, M. J. Genomic loss of imprinting in first-trimester human placenta. *Am. J. Obstet. Gynecol,* 202(4) (2010), pp. 391–398.

[39] Jinno, Y., Ikeda, Y., Yun, K., Maw, M., Masuzaki, H., Fukuda, H., Inuzuka, K., Fujishita, A., Ohtani, Y., and Okimoto, T. Establishment of functional imprinting of the H19 gene in human developing placentae. *Nat Genet,* 10(3) (1995), pp. 318–324.

[40] Gabory, A., Jammes, H., and Dandolo, L. The H19 locus: role of an imprinted non-coding RNA in growth and development. *Bioessays,* 32(6) (2010), pp. 473–80.

[41] Cai, X. and Cullen, B. R. The imprinted H19 noncoding RNA is a primary microRNA precursor. *RNA,* 13(3) (2007), pp. 313–6.

[42] Steck, E., Boeuf, S., Gabler, J., Werth, N., Schnatzer, P., Diederichs, S., and Richter, W. Regulation of H19 and its encoded microRNA-675 in osteoarthritis and under anabolic and catabolic in vitro conditions. *J Mol Med (Berl),* 90(10) (2012), pp. 1185–95.

[43] Keniry, A., Oxley, D., Monnier, P., Kyba, M., Dandolo, L., Smits, G., and Reik, W. The H19 lincRNA is a developmental reservoir of miR-675 that suppresses growth and Igf1r. *Nat Cell Biol,* 14(7) (2012), pp. 659–65.

[44] Sitras, V., Fenton, C., Paulssen, R., Vårtun, A., and Acharya, G. Differences in Gene Expression between First and Third Trimester Human Placenta: A Microarray Study. *PLoS ONE,* 7 (2012), e33294.

[45] Cook, K. B., Kazan, H., Zuberi, K., Morris, Q., and Hughes, T. R. RBPDB: a database of RNA-binding specificities. *Nucleic Acids Res,* 39(Database issue) (2011), pp. D301–8.

[46] Noguer-Dance, M., Abu-Amero, S., Al-Khtib, M., Lefevre, A., Coullin, P., Moore, G. E., and Cavaille, J. The primate-specific microRNA gene cluster (C19MC) is imprinted in the placenta. *Hum Mol Genet,* 19(18) (2010), pp. 3566–82.

[47] Flor, I., Neumann, A., Freter, C., Helmke, B. M., Langenbuch, M., Rippe, V., and Bullerdiek, J. Abundant expression and hemimethylation of C19MC in cell cultures from placenta-derived stromal cells. *Biochem Biophys Res Commun,* 422(3) (2012), pp. 411–6.

[48] Bortolin-Cavaille, M. L., Dance, M., Weber, M., and Cavaille, J. C19MC microRNAs are processed from introns of large Pol-II, non-protein-coding transcripts. *Nucleic Acids Res,* 37(10) (2009), pp. 3464–73.

[49] Saito, Y., Suzuki, H., Tsugawa, H., Nakagawa, I., Matsuzaki, J., Kanai, Y., and Hibi, T. Chromatin remodeling at Alu repeats by epigenetic treatment activates silenced microRNA-512-5p with downregulation of Mcl-1 in human gastric cancer cells. *Oncogene,* 28(30) (2009), pp. 2738–44.

[50] Tsai, K. W., Kao, H. W., Chen, H. C., Chen, S. J., and Lin, W. C. Epigenetic control of the expression of a primate-specific microRNA cluster in human cancer cells. *Epigenetics,* 4(8) (2009), pp. 587–92.

[51] Donker, R. B., Mouillet, J. F., Chu, T., Hubel, C. A., Stolz, D. B., Morelli, A. E., and Sadovsky, Y. The expression profile of C19MC microRNAs in primary human trophoblast cells and exosomes. *Mol Hum Reprod,* 18(8) (2012), pp. 417–24.

[52] Enquobahrie, D. A., Abetew, D. F., Sorensen, T. K., Willoughby, D., Chidambaram, K., and Williams, M. A. Placental microRNA expression in pregnancies complicated by preeclampsia. *Am J Obstet Gynecol,* 204(2) (2011), pages.

[53] Zhu, X. M., Han, T., Sargent, I. L., Yin, G. W., and Yao, Y. Q. Differential expression profile of microRNAs in human placentas from preeclamptic pregnancies vs normal pregnancies. *Am J Obstet Gynecol,* 200(6) (2009), pages.

[54] Ishibashi, O., Ohkuchi, A., Ali, M. M., Kurashina, R., Luo, S. S., Ishikawa, T., Takizawa, T., Hirashima, C., Takahashi, K., Migita, M., Ishikawa, G., Yoneyama, K., Asakura, H., Izumi, A., Matsubara, S., Takeshita, T., and Takizawa, T. Hydroxysteroid (17-beta) dehydrogenase 1 is dysregulated by miR-210 and miR-518c that are aberrantly expressed in preeclamptic placentas: a novel marker for predicting preeclampsia. *Hypertension,* 59(2) (2012), pp. 265–73.

[55] Hu, Y., Li, P., Hao, S., Liu, L., Zhao, J., and Hou, Y. Differential expression of microRNAs in the placentae of Chinese patients with severe pre-eclampsia. *Clin Chem Lab Med,* 47(8) (2009), pp. 923–9.

[56] Ye, W., Lv, Q., Wong, C. K., Hu, S., Fu, C., Hua, Z., Cai, G., Li, G., Yang, B. B., and Zhang, Y. The effect of central loops in miRNA:MRE duplexes on the efficiency of miRNA-mediated gene regulation. *PLoS One,* 3(3) (2008), e1719.

[57] Enquobahrie, D. A., Meller, M., Rice, K., Psaty, B. M., Siscovick, D. S., and Williams, M. A. Differential placental gene expression in preeclampsia. *Am J Obstet Gynecol,* 199(5) (2008), pages.

[58] McCarthy, C., Cotter, F. E., McElwaine, S., Twomey, A., Mooney, E. E., Ryan, F., and Vaughan, J. Altered gene expression patterns in intrauterine growth restriction: potential role of hypoxia. *Am J Obstet Gynecol,* 196(1) (2007), pages.

[59] Nishizawa, H., Pryor-Koishi, K., Kato, T., Kowa, H., Kurahashi, H., and Udagawa, Y. Microarray analysis of differentially expressed fetal genes in placental tissue derived from early and late onset severe pre-eclampsia. *Placenta,* 28(5-6) (2007), pp. 487–97.

[60] Winn, V. D., Gormley, M., Paquet, A. C., Kjaer-Sorensen, K., Kramer, A., Rumer, K. K., Haimov-Kochman, R., Yeh, R. F., Overgaard, M. T., Varki, A., Oxvig, C., and Fisher, S. J. Severe preeclampsia-related changes in gene expression at the maternal-fetal interface include sialic acid-binding immunoglobulin-like lectin-6 and pappalysin-2. *Endocrinology,* 150(1) (2009), pp. 452–62.

[61] Jarvenpaa, J., Vuoristo, J. T., Savolainen, E. R., Ukkola, O., Vaskivuo, T., and Ryynanen, M. Altered expression of angiogenesis-related placental genes in pre-eclampsia associated with intrauterine growth restriction. *Gynecol Endocrinol,* 23(6) (2007), pp. 351–5.

[62] Sitras, V., Paulssen, R., Leirvik, J., Vårtun, A., and Acharya, G. Placental gene expression profile in intrauterine growth restriction due to placental insufficiency. *Reprod Sci,* 16 (2009), pp. 701–711.

[63] Meng, T., Chen, H., Sun, M., Wang, H., Zhao, G., and Wang, X. Identification of differential gene expression profiles in placentas from preeclamptic pregnancies versus normal pregnancies by DNA microarrays. *OMICS,* 16(6) (2012), pp. 301–11.

[64] Levy, R. The role of apoptosis in preeclampsia. *The Israel Medical Association journal: IMAJ,* 7(3) (2005), pp. 178–81.

[65] Novakovic, B. and Saffery, R. The ever growing complexity of placental epigenetics - role in adverse pregnancy outcomes and fetal programming. *Placenta,* 33(12) (2012), pp. 959–70.

[66] Mayor-Lynn, K., Toloubeydokhti, T., Cruz, A. C., and Chegini, N. Expression profile of microRNAs and mRNAs in human placentas from pregnancies complicated by preeclampsia and preterm labor. *Reprod Sci,* 18(1) (2011), pp. 46–56.

[67] Papadopoulos, G. L., Reczko, M., Simossis, V. A., Sethupathy, P., and Hatzigeorgiou, A. G. The database of experimentally supported targets: a functional update of TarBase. *Nucleic acids research,* 37(Database issue) (2009), pp. D155–8.

[68] Hsu, S. D., Lin, F. M., Wu, W. Y., Liang, C., Huang, W. C., Chan, W. L., Tsai, W. T., Chen, G. Z., Lee, C. J., Chiu, C. M., Chien, C. H., Wu, M. C., Huang, C. Y., Tsou, A. P., and Huang, H. D. miRTarBase: a database curates experimentally validated microRNA-target interactions. *Nucleic Acids Research,* 39(Database issue) (2011), pp. D163–9.

[69] Wang, W., Feng, L., Zhang, H., Hachy, S., Satohisa, S., Laurent, L. C., Parast, M., Zheng, J., and Chen, D. B. 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,* 97(6) (2012), E1051–9.

[70] Andraweera, P., Dekker, G., Thompson, S., and Roberts, C. Single-nucleotide polymorphisms in the KDR gene in pregnancies complicated by gestational hypertensive disorders and small-for-gestational-age infants. *Reprod Sci,* 19 (2012), pp. 547–554.

[71] Glienke, J., Sturz, A., Menrad, A., and Thierauch, K. H. CRIM1 is involved in endothelial cell capillary formation in vitro and is expressed in blood vessels in vivo. *Mechanisms of development,* 119(2) (2002), pp. 165–75.

[72] Pugh, C. W. and Ratcliffe, P. J. Regulation of angiogenesis by hypoxia: role of the HIF system. *Nature medicine,* 9(6) (2003), pp. 677–84.

[73] Pringle, K. G., Kind, K. L., Sferruzzi-Perri, A. N., Thompson, J. G., and Roberts, C. T. Beyond oxygen: complex regulation and activity of hypoxia inducible factors in pregnancy. *Human reproduction update,* 16(4) (2010), pp. 415–31.

[74] Cazzalini, O., Scovassi, A. I., Savio, M., Stivala, L. A., and Prosperi, E. Multiple roles of the cell cycle inhibitor p21(CDKN1A) in the DNA damage response. *Mutation Research,* 704(1-3) (2010), pp. 12–20.

[75] Curtin, J. F. and Cotter, T. G. Live and let die: regulatory mechanisms in Fas-mediated apoptosis. *Cellular Signalling,* 15(11) (2003), pp. 983–92.

[76] Poehlmann, T. G., Fitzgerald, J. S., Meissner, A., Wengenmayer, T., Schleussner, E., Friedrich, K., and Markert, U. R. Trophoblast invasion: tuning through LIF, signalling via Stat3. *Placenta,* 26 Suppl A (2005), S37–41.

[77] PengFei, L., YaLi, H., Sha, H., Liu, L., JunLi, Z., and YaYi, H. The expression of microRNA in placenca from severe preeclampsia patients. *Chinese Journal of Practical Gynecology and Obstetrics,* 25 (2009), pp. 911–14.

[78] Thomadaki, H. and Scorilas, A. BCL2 family of apoptosis-related genes: functions and clinical implications in cancer. *Critical Reviews in Clinical Laboratory Sciences,* 43(1) (2006), pp. 1–67.

[79] Chan, K. T., Creed, S. J., and Bear, J. E. Unraveling the enigma: progress towards understanding the coronin family of actin regulators. *Trends in Cell biology,* 21(8) (2011), pp. 481–8.

[80] Preston, G. A., Lyon, T. T., Yin, Y., Lang, J. E., Solomon, G., Annab, L., Srinivasan, D. G., Alcorta, D. A., and Barrett, J. C. Induction of apoptosis by c-Fos protein. *Molecular and cellular biology,* 16(1) (1996), pp. 211–8.

[81] Urbich, C., Knau, A., Fichtlscherer, S., Walter, D. H., Bruhl, T., Potente, M., Hofmann, W. K., Vos, S. de, Zeiher, A. M., and Dimmeler, S. FOXO-dependent expression of the proapoptotic protein Bim: pivotal role for apoptosis signaling in endothelial progenitor cells. *FASEB journal: official publication of the Federation of American Societies for Experimental Biology,* 19(8) (2005), pp. 974–6.

[82] Gonzalvez, F. and Ashkenazi, A. New insights into apoptosis signaling by Apo2L/TRAIL. *Oncogene,* 29(34) (2010), pp. 4752–65.

[83] Halperin, R., Peller, S., Sandbank, J., Bukovsky, I., and Schneider, D. Expression of the p53 gene and apoptosis in gestational trophoblastic disease. *Placenta,* 21(1) (2000), pp. 58–62.

[84] Dye, J. F., Jablenska, R., Donnelly, J. L., Lawrence, L., Leach, L., Clark, P., and Firth, J. A. Phenotype of the endothelium in the human term placenta. *Placenta,* 22(1) (2001), pp. 32–43.

[85] Ullah, Z., Kohn, M. J., Yagi, R., Vassilev, L. T., and DePamphilis, M. L. Differentiation of trophoblast stem cells into giant cells is triggered by p57/Kip2 inhibition of CDK1 activity. *Genes & development,* 22(21) (2008), pp. 3024–36.

[86] Prendergast, G. C. Actin' up: RhoB in cancer and apoptosis. *Nature reviews. Cancer,* 1(2) (2001), pp. 162–8.

[87] Akira, S., Hirano, T., Taga, T., and Kishimoto, T. Biology of multifunctional cytokines: IL 6 and related molecules (IL 1 and TNF). *FASEB Journal,* 4(11) (1990). Official publication of the Federation of American Societies for Experimental Biology, pp. 2860–7.

[88] Andraweera, P. H., Dekker, G. A., and Roberts, C. T. The vascular endothelial growth factor family in adverse pregnancy outcomes. *Human Reproduction Update,* 18 (2012), pp. 436–457.

[89] Yoshitomi, T., Kawakami, K., Enokida, H., Chiyomaru, T., Kagara, I., Tatarano, S., Yoshino, H., Arimura, H., Nishiyama, K., Seki, N., and Nakagawa, M. Restoration of miR-517a expression induces cell apoptosis in bladder cancer cell lines. *Oncology Reports,* 25(6) (2011), pp. 1661–8.

[90] Montenegro, D., Romero, R., Kim, S. S., Tarca, A. L., Draghici, S., Kusanovic, J. P., Kim, J. S., Lee, D. C., Erez, O., Gotsch, F., Hassan, S. S., and Kim, C. J. Expression patterns of microRNAs in the chorioamniotic membranes: a role for microRNAs in human pregnancy and parturition. *J Pathol,* 217(1) (2009), pp. 113–21.

[91] Kolwankar, D., Glover, D. D., Ware, J. A., and Tracy, T. S. Expression and function of ABCB1 and ABCG2 in human placental tissue. *Drug Metabol Dispos,* 33(4) (2005), pp. 524–9.

[92] Crawford, M. A., Doyle, W., and Meadows, N. Gender differences at birth and differences in fetal growth. *Hum Reprod,* 2(6) (1987), pp. 517–20.

[93] Bertino, E., Coscia, A., Boni, L., Rossi, C., Martano, C., Giuliani, F., Fabris, C., Spada, E., Zolin, A., and Milani, S. Weight growth velocity of very low birth weight infants: role of gender, gestational age and major morbidities. *Early Hum Dev,* 85(6) (2009), pp. 339–47.

[94] de Onis, M., Siyam, A., Borghi, E., Onyango, A. W., Piwoz, E., and Garza, C. Comparison of the World Health Organization growth velocity standards with existing US reference data. *Pediatrics,* 128(1) (2011), e18–26.

[95] Brettell, R., Yeh, P. S., and Impey, L. W. Examination of the association between male gender and preterm delivery. *Eur J Obstet Gynecol Reprod Biol,* 141(2) (2008), pp. 123–6.

[96] Ingemarsson, I. Gender aspects of preterm birth. *British Journal of Obstetrics and Gynaecology,* 110 (2003), pp. 34–38.

[97] Elsmen, E., Kallen, K., Marsal, K., and Hellstrom-Westas, L. Fetal gender and gestational-age-related incidence of pre-eclampsia. *Acta Obstet Gynecol Scand,* 85(11) (2006), pp. 1285–91.

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

[99] Vatten, L. J. and Skjaerven, R. Offspring sex and pregnancy outcome by length of gestation. *Early Hum Dev,* 76(1) (2004), pp. 47–54.

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

[101] Hodyl, N. A., Wyper, H., Osei-Kumah, A., Scott, N., Murphy, V. E., Gibson, P., Smith, R., and Clifton, V. L. Sex-specific associations between cortisol and birth weight in pregnancies complicated by asthma are not due to differential glucocorticoid receptor expression. *Thorax,* 65(8) (2010), pp. 677–83.

[102] Di Renzo, G. C., Rosati, A., Sarti, R. D., Cruciani, L., and Cutuli, A. M. Does fetal sex affect pregnancy outcome? *Gend Med,* 4(1) (2007), pp. 19–30.

[103] Edwards, A., Megens, A., Peek, M., and Wallace, E. M. Sexual origins of placental dysfunction. *Lancet,* 355(9199) (2000), pp. 203–4.

[104] Ghidini, A. and Salafia, C. M. Gender differences of placental dysfunction in severe prematurity. *BJOG,* 112(2) (2005), pp. 140–4.

[105] Wijchers, P. J. and Festenstein, R. J. Epigenetic regulation of autosomal gene expression by sex chromosomes. *Trends Genet,* 27(4) (2011), pp. 132–40.

[106] Carrel, L. and Willard, H. F. X-inactivation profile reveals extensive variability in X-linked gene expression in females. *Nature,* 434(7031) (2005), pp. 400–404.

[107] Ober, C., Loisel, D. A., and Gilad, Y. Sex-specific genetic architecture of human disease. *Nat Rev Genet,* 9(12) (2008), pp. 911–22.

[108] Graves, J. A. Review: Sex chromosome evolution and the expression of sex-specific genes in the placenta. *Placenta,* 31 Suppl (2010), S27–32.

[109] Arnold, A. P. and Lusis, A. J. Understanding the sexome: measuring and reporting sex differences in gene systems. *Endocrinology,* 153(6) (2012), pp. 2551–5.

[110] Kozomara, A. and Griffiths-Jones, S. miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res,* 39(Database issue) (2011), pp. D152–7.

[111] Ropers, H. H., Wolff, G., and Hitzeroth, H. W. Preferential X inactivation in human placenta membranes: is the paternal X inactive in early embryonic development of female mammals? *Hum Genet,* 43(3) (1978), pp. 265–73.

[112] Migeon, B. R. and Do, T. T. In search of non-random X inactivation: studies of fetal membranes heterozygous for glucose-6-phosphate dehydrogenase. *Am J Hum Genet,* 31(5) (1979), pp. 581–5.

[113] Harrison, K. B. and Warburton, D. Preferential X-chromosome activity in human female placental tissues. *Cytogenet Cell Genet,* 41(3) (1986), pp. 163–8.

[114] Harrison, K. B. X-chromosome inactivation in the human cytotrophoblast. *Cytogenet Cell Genet,* 52(1-2) (1989), pp. 37–41.

[115] Mohandas, T. K., Passage, M. B., Williams J. W., 3., Sparkes, R. S., Yen, P. H., and Shapiro, L. J. X-chromosome inactivation in cultured cells from human chorionic villi. *Somat Cell Mol Genet,* 15(2) (1989), pp. 131–6.

[116] Goto, T., Wright, E., and Monk, M. Paternal X-chromosome inactivation in human trophoblastic cells. *Mol Hum Reprod,* 3(1) (1997), pp. 77–80.

[117] Looijenga, L. H., Gillis, A. J., Verkerk, A. J., Putten, W. L. van, and Oosterhuis, J. W. Heterogeneous X inactivation in trophoblastic cells of human full-term female placentas. *Am J Hum Genet,* 64(5) (1999), pp. 1445–52.

[118] Zeng, S. M. and Yankowitz, J. X-inactivation patterns in human embryonic and extra-embryonic tissues. *Placenta,* 24(2-3) (2003), pp. 270–5.

[119] Willemsen, R., Bontekoe, C. J., Severijnen, L. A., and Oostra, B. A. Timing of the absence of FMR1 expression in full mutation chorionic villi. *Hum Genet,* 110(6) (2002), pp. 601–5.

[120] Dhara, S. K. and Benvenisty, N. Gene trap as a tool for genome annotation and analysis of X chromosome inactivation in human embryonic stem cells. *Nucleic Acids Res,* 32(13) (2004), pp. 3995–4002.

[121] Uehara, S., Tamura, M., Nata, M., Ji, G., Yaegashi, N., Okamura, K., and Yajima, A. X-chromosome inactivation in the human trophoblast of early pregnancy. *J Hum Genet,* 45(3) (2000), pp. 119–26.

[122] Lee, J. T. Gracefully ageing at 50, X-chromosome inactivation becomes a paradigm for RNA and chromatin control. *Nature Reviews Molecular cell Biology,* 12 (2011), pp. 815–826.

[123] Moreira de Mello, J. C., Araujo, E. S. de, Stabellini, R., Fraga, A. M., Souza, J. E. de, Sumita, D. R., Camargo, A. A., and Pereira, L. V. Random X inactivation and extensive mosaicism in human placenta revealed by analysis of allele-specific gene expression along the X chromosome. *PLoS One,* 5(6) (2010), e10947.

[124] Ge, B., Gurd, S., Gaudin, T., Dore, C., Lepage, P., Harmsen, E., Hudson, T. J., and Pastinen, T. Survey of allelic expression using EST mining. *Genome Res,* 15(11) (2005), pp. 1584–91.

[125] Wang, H. and Elbein, S. C. Detection of allelic imbalance in gene expression using pyrosequencing. *Methods Mol Biol,* 373 (2007), pp. 157–76.

[126] Knight, J. C. Allele-specific gene expression uncovered. *Trends Genet,* 20(3) (2004), pp. 113–6.

[127] Borzychowski, A. M., Sargent, I. L., and Redman, C. W. Inflammation and pre-eclampsia. *Seminars in fetal & neonatal medicine,* 11(5) (2006), pp. 309–16.

[128] Ghosh, G., Subramanian, I. V., Adhikari, N., Zhang, X., Joshi, H. P., Basi, D., Chandrashekhar, Y. S., Hall, J. L., Roy, S., Zeng, Y., and Ramakrishnan, S. Hypoxia-induced microRNA-424 expression in human endothelial cells regulates HIF-alpha isoforms and promotes angiogenesis. *J Clin Invest,* 120(11) (2010), pp. 4141–54.

[129] Lazar, L., Nagy, B., Molvarec, A., Szarka, A., and Rigo J., J. Role of hsa-miR-325 in the etiopathology of preeclampsia. *Mol Med Rep,* 6(3) (2012), pp. 597–600.

Statement of Authorship

Title of Paper	massiR: a method for predicting the sex of samples in gene expression microarray datasets
Publication Status	◉ Published, ○ Accepted for Publication, ○ Submitted for Publication, ○ Publication style
Publication Details	Buckberry, S., Bent, S. J., Bianco-Miotto, T. & Roberts, C. T. massiR: a method for predicting the sex of samples in gene expression microarray datasets. Bioinformatics 30, 2084-2085 (2014).

Author Contributions

By signing the Statement of Authorship, each author certifies that their stated contribution to the publication is accurate and that permission is granted for the publication to be included in the candidate's thesis.

Name of Principal Author (Candidate)	Sam Buckberry
Contribution to the Paper	Sam Buckberry conceived, developed and tested the software algorithm, created the software documentation and authored the first draft of the manuscript, incorporated suggestions from co-authors and peer-reviewers and wrote the final manuscript.
Signature	Date: 7/4/15

Name of Co-Author	Stephen J Bent
Contribution to the Paper	Stephen J Bent was involved in critical discussions and offered valuable insight regarding the implementation of the algorithm and commented on all drafts of the manuscript.
Signature	Date: 25/3/2015

Name of Co-Author	Tina Bianco-Miotto
Contribution to the Paper	Tina Bianco-Miotto was involved in critical discussion regarding algorithm development and commented on all drafts of the manuscript.
Signature	Date: 1/4/15

Name of Co-Author	Claire T Roberts
Contribution to the Paper	Claire T Roberts was involved in critical discussion regarding algorithm development and commented on all drafts of the manuscript.
Signature	Date: 7.4.15

Chapter 4

massiR: a Method for Predicting the Sex of Samples in Gene Expression Microarray Datasets

Sam Buckberry, Stephen J Bent, Tina Bianco-Miotto and Claire T Roberts

Abstract

High-throughput gene expression microarrays are currently the most efficient method for transcriptome-wide expression analyses. Consequently, gene expression data available through public repositories has largely been obtained from microarray experiments. However, the metadata associated with many publicly available expression microarray datasets often lacks sample sex information, therefore limiting the reuse of these data in new analyses or larger meta-analyses where the effect of sex is to be considered. Here we present the *massiR* package, which provides a method for researchers to predict the sex of samples in microarray datasets. Using information from microarray probes representing Y chromosome genes, this package implements unsupervised clustering methods to classify samples into male and female groups, providing an efficient way to identify or confirm the sex of samples in mammalian microarray datasets.

Availability: *massiR* is implemented as a Bioconductor package in R. The package and the vignette can be downloaded at bioconductor.org and are provided under a GPL-2 license.

4.1 Introduction

For over a decade, high-throughput microarray experiments have been generating large volumes of genome-wide expression data and the reporting requirements of many journals have seen that much of these data are made publicly available. Given the substantial value of these accumulated datasets, it is becoming increasingly common to reuse gene expression data to validate new findings or to pose new biological questions. However, the value of microarray datasets is largely dependent on the completeness and accuracy of the associated metadata, which is reliant on diligent reporting by researchers and accurate representation upon submission [1].

Given that the sex of many species is an easily observable and usually unambiguous classification, it is surprising the number of microarray datasets in public repositories that lack the associated sample sex information. Sex-biased gene expression in normal and pathological tissues is well recognized for both sex chromosome and autosomal genes [2, 3]. Sex biases also exist in the prevalence and severity of many common human diseases, such as cardiovascular disease and some cancers [4]. As sex is a potential influencing factor of both pathological and non-pathological phenotypes, gene ex-pression analyses that do not account for sex-specific effects could fail to identify a significant proportion of genes that contribute to the condition under investigation [4]. Therefore, the absence of sample sex information restricts the reuse of gene expression datasets where the researcher intends to factor the effect of sex in reanalysis or reinterpretation, or when intending to include such datasets in larger gene expression meta-analyses.

In this application note we present *massiR* (MicroArray Sample Sex Identifier), a Bioconductor package for predicting the sex of samples in microarray datasets. This method allows researchers to expand their analyses to retrospectively incorporate sex as a variable, generate or confirm sex information associated with publicly available datasets, to accurately predict the sex for samples missing this information or to identify mislabeled samples.

4.2 Methods and Validation

Methods

The *massiR* analysis begins by importing normalized gene expression data using standard methods. The first step extracts the expression values for probes that

correspond to Y chromosome genes. Here the user has the option of using their own list of probes corresponding to Y chromosome genes or using the probe lists included with the package. The included lists correspond to popular microarray platforms and contain identifiers for probes that uniquely map to Y chromosome genes (for details see Supplementary Information).

When the expression values for Y chromosome probes are extracted, the expression variance for each probe across all samples is calculated. This allows the identification of low variance probes that are unlikely to be informative in sex classification. The user has the option of selecting a probe variation threshold so only the most informative probes are used in the classification process, a decision which can be informed by inspecting an easily generated probe variation plot.

To classify samples as either male or female, clustering is performed using the values from the subset of Y chromosome probes by implementing the partitioning around medoids algorithm to perform k-medoids clustering [5], where samples are assigned to one of two clusters. The two clusters are then compared using the probe expression values across all samples in each cluster. Samples within the cluster featuring the highest Y chromosome probe values are classed as male and those amongst the cluster with the lowest Y probe values are classed as female. Results such as sample probe mean, standard deviation and z-scores are returned with the sex predicted for each sample.

The *massiR* package includes functions for generating informative plots of the data at different stages of the analysis, enabling the user to inspect various elements of the data. These include a bar plot of mean probe expression for each sample, a heat map of probe values for each sample and principal component plots of sample clusters The vignette accompanying the *massiR* package provides a concise description of the workflow and detailed examples of how to use all the included functions.

Validation

We tested the sex classification accuracy of the *massiR* package using publicly available gene expression datasets for human and mouse tissues with sample sex information (See Supplementary Information for results). Additionally, we tested the accuracy of sex classification in datasets with skewed sex ratios by randomly selecting male and female samples from five empirical human datasets to create data subsets with a wide range of male/female ratios (Figure 4.1). Assuming sex was correctly reported in the metadata, the results from this testing show that

the correct sex prediction rate is 97.2% (±1.2 SEM) for datasets that contain between 15–85% males. As we observed greater variability in prediction accuracy outside this range (Figure 4.1), we include a function in the *massiR* package for detecting datasets with skewed sex ratios using an implementation of the dip test for unimodality [5, 6]. See the Supplementary Information for details on further testing and results.

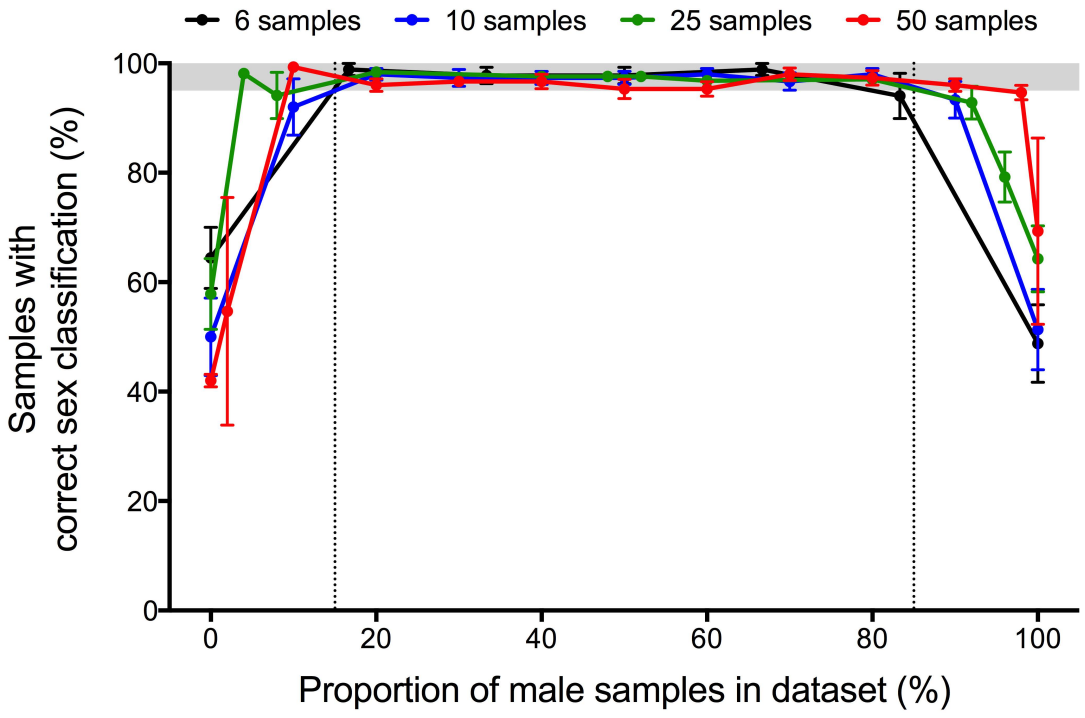


FIGURE 4.1: Sex prediction accuracy of the *massiR* package using human gene expression datasets with a range of male/female ratios. The correct sex prediction rate is 97.2% (±1.2 SEM) for datasets with $> 15\%$ and $< 85\%$ males which is the area between the vertical dotted lines. Points represent mean, vertical bars show the standard error of the mean. The grey band at the top of the plot shows the 95–100% range. These results are a summary of tests conducted using publicly available expression data from human brain, colorectal, kidney, and placenta tissue, and peripheral blood mononuclear cells. The data subsets for each were generated by randomly selecting male and female samples for pre-determined dataset sizes and sex ratios.

4.3 Conclusion

To our knowledge this is the only available software package for predicting the sex of samples in gene expression microarray datasets. This easily implemented method opens the door to both prospective and retrospective gene expression analyses that wish to consider the effect of sex on gene expression.

Acknowledgements

We thank Shalem Leemaqz and Dan Kortschak for their insight and opinion regarding this work.

Funding: SB is supported by a Healthy Development Adelaide & Channel 7 Children's Research Foundation PhD Scholarship and an Australian Postgraduate Award. TB-M is supported by the Cancer Council SA and SAHMRI Beat Cancer Project (TBM APP1030945). CTR is supported by a National Health and Medical Research Council (NHMRC) Senior Research Fellowship APP1020749. The work was funded by NHMRC Research Project #565320 awarded to CTR.

S4.1 Supporting Information

S4.2 Y Chromsome Probe Indentifiers Included with the Massi Package

To identify probes that represent Y chromosome genes, we used the Ensembl mappings of probes for commercially available microarray platforms. We selected this option because Ensembl have independently mapped the probes from numerous platforms to a common reference genome, and the annotation information for many platforms is accessible through the Bioconductor package *biomaRt*. This method allowed us to select probes that map uniquely to Y chromosome genes. A detailed example on how to obtain probe information for commercial microarray platforms is included with the *massiR* vignette. For details on probe mapping methods, see the permalink: http://jan2013.archive.ensembl.org/info/docs/microarray_probe_set_mapping.html.

S4.3 Testing and Validation

We searched the NCBI GEO public repository for gene expression microarray datasets with associated sex information in the metadata for testing purposes (Supplementary Table S4.1). When raw data were available, we preprocessed and normalized the arrays before performing quality assessments using standard methods and Bioconductor packages in R. Any arrays that were deemed to be outliers were removed from the dataset, then the data were re-normalised before predicting the sex of samples using the *massiR* package.

To test the accuracy of this method, we selected ten datasets encompassing multiple microarray platforms and samples derived from various normal and pathological tissues (Supplementary Table S4.1). In 6/10 datasets, this method predicted the sex of the samples with 100% accuracy (Supplementary Table S4.1). However, this validation methodology is dependent on the accuracy of the associated metadata. Given that this prediction method only uses information from Y chromosome probes, we interrogated each dataset to examine probe-specific expression values for each sample to further understand why we encountered a few isolated cases of misclassification (see below). Therefore it is reasonable to suggest that some of these discrepancies may be due to unintended errors in the metadata and not due to misclassification.

TABLE S4.1: Validation results for predicting the sex of samples in microarray datasets using the MASSI package.

							Male samples			Female samples		
GEO accession	Species	Tissue	Platform	No. Samples	No. correctly predicted	Overall Prediction accuracy	No. samples	No. correctly predicted	Prediction accuracy	No. samples	No. Correctly predicted	Prediction accuracy
GSE45330	Human	Blood cells	Illumina Human HT-12 V4	77	76	**98.70%**	33	32	96.97%	44	44	100.00%
GSE29378	Human	Brain	Illumina Human HT-12 V3	63	61	**96.83%**	38	38	100.00%	25	23	92.00%
GSE35896	Human	Colorectal	Affymetrix HG-U133 Plus 2.0	58	57	**98.28%**	27	26	96.30%	31	31	100.00%
GSE25906	Human	Placenta	Illumina Human-6 V2	60	60	**100.00%**	31	31	100.00%	29	29	100.00%
GSE13546	Human	Lung cancer	Affymetrix HG-U133 Plus 2.0	15	15	**100.00%**	3	3	100.00%	12	12	100.00%
GSE20950	Human	Adipose	Affymetrix HG-U133 Plus 2.0	39	39	**100.00%**	12	12	100.00%	27	27	100.00%
GSE14335	Human	Fibroblast cells	Affymetrix HG-U133A 2.0	10	10	**100.00%**	3	3	100.00%	7	7	100.00%
GSE40435	Human	Kidney	Illumina Human HT-12 V4	202	195	**96.53%**	118	113	95.76%	84	82	97.62%
GSE29585	Mouse	Placenta	Affymetrix Mo. Exon 1.0 ST	16	16	**100.00%**	8	8	100.00%	8	8	100.00%
GSE35182	Mouse	Heart	Affymetrix Mo. Gene 1.0 ST	24	24	**100.00%**	12	12	100.00%	12	12	100.00%
Totals				**564**	**553**	**98.05%**	**285**	**278**	**97.54%**	**279**	**275**	**98.57%**

S4.3.1 Samples Classified as Male but Listed as Female in the Metadata

There were four cases across two datasets where samples were predicted as male using this method but listed as female in the metadata (Supplementary Table S4.1). When interrogating the individual Y chromosome probe values, we observed that all of these samples show expression of Y chromosome genes well within the range of all the other male samples in the dataset (Supplementary Figures S4.1 & S4.2).

S4.3.2 Samples Classified as Female but Listed as Male in the Metadata

There were eight cases across three datasets where samples were classified as female using this method but indicated as male in the metadata (Supplementary Table S4.1). In all but one of these cases (Supplementary Figure S4.3) the intensity values for Y chromosome probes was well within the range of female samples, and showed no indication of any Y chromosome gene expression (Supplementary Figures S4.2–S4.4). However, although this infers that several of these samples are female (as predicted), one cannot exclude the possibility that the cells assayed were not expressing Y chromosome genes at that time point.

S4.3.3 Performance with Skewed Sex Ratios

To test the performance of the *massiR* method with datasets with skewed sex ratios, we randomly selected male and female samples from large array datasets to generate random data subsets with a spectrum of sex ratios. This was performed with human brain (GSE29378), colorectal (GSE35896), kidney (GSE40435), placenta tissue (GSE25906) and peripheral blood mononuclear cells (GSE45330). For each dataset, we separated the male and female samples and then randomly selected samples from each group to create datasets of pre-determined sex ratios and sample sizes. For each dataset, we performed this randomized dataset construction process in triplicate. The summarized results for each tissue type are presented in Supplementary Figure S4.5.

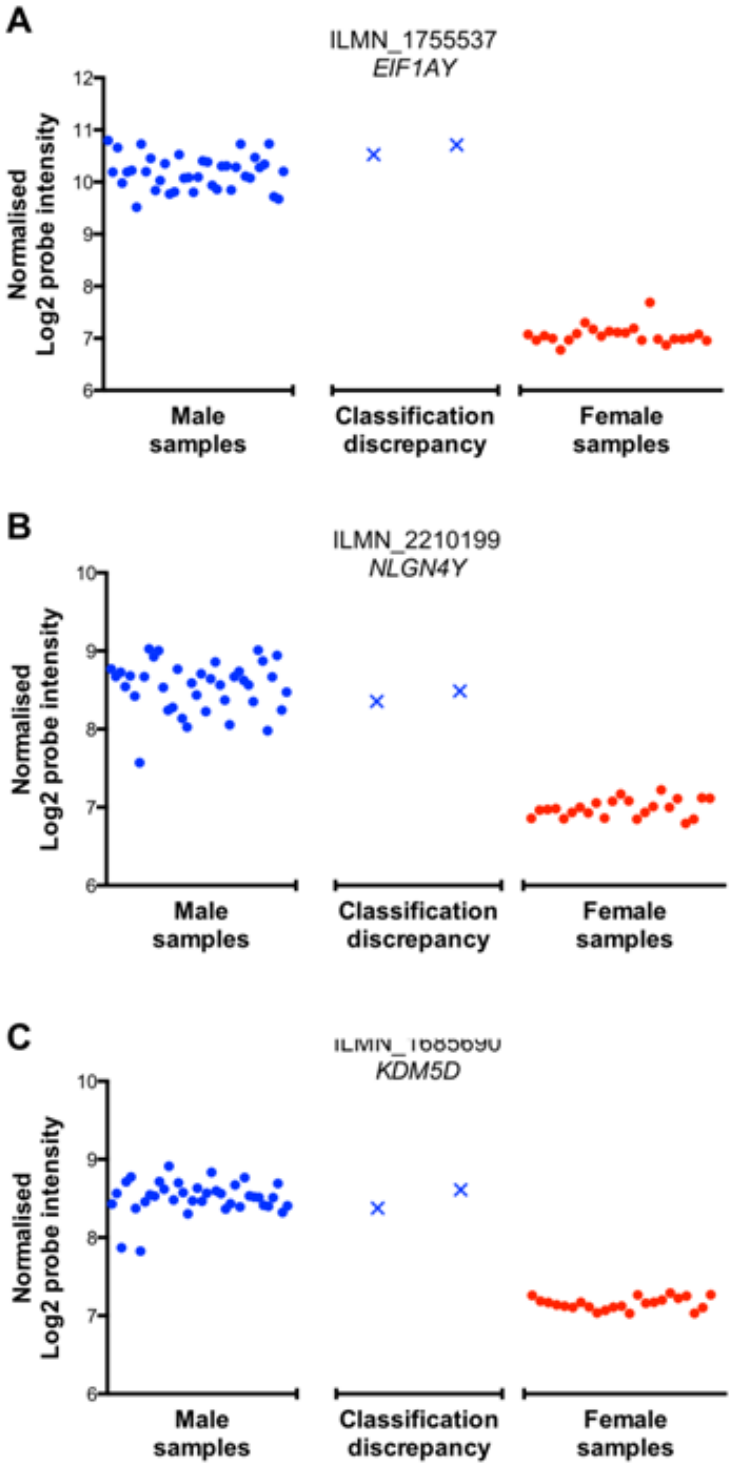


FIGURE S4.1: Prediction and validation of the sex of samples in dataset GSE29378. Plots show microarray probe intensity values for Y chromosome genes *EIF1AY* (**A**), *NLGN4Y* (**B**) and *KDM5D* (**C**). This shows that two samples (blue crosses) listed as female in the metadata show Y chromosome gene expression values comparable to male samples (blue dots), which are distinct from the samples confirmed as female (red dots).

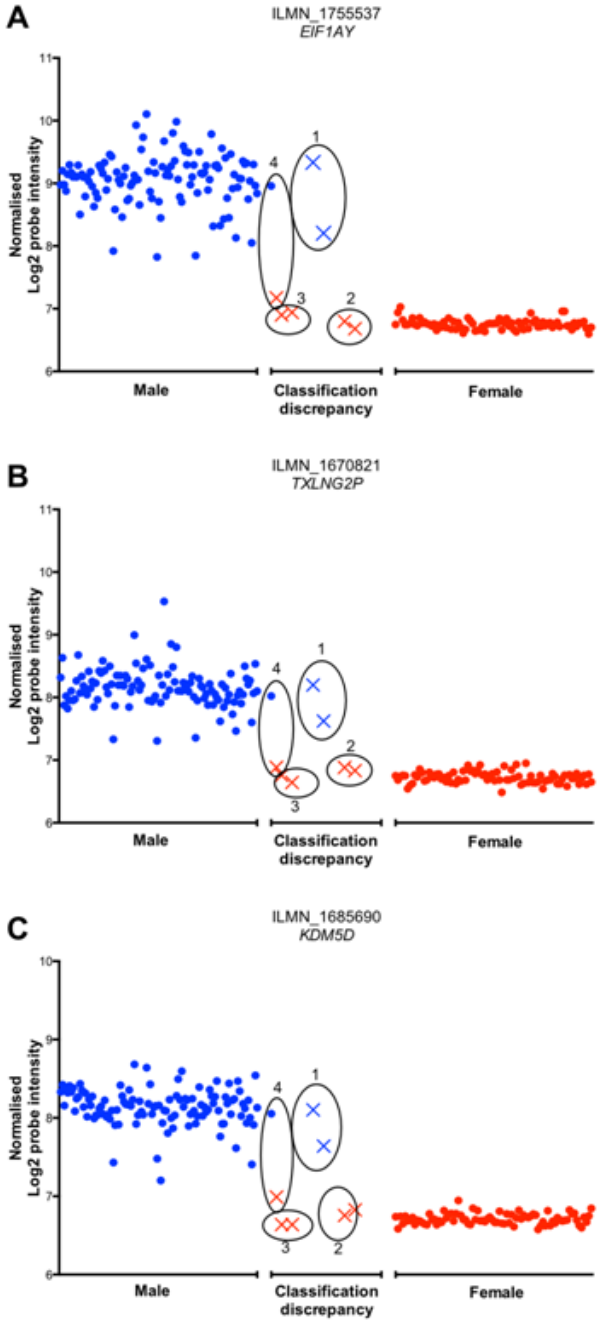


FIGURE S4.2: Prediction and validation of the sex of samples in dataset GSE40435. Plots show microarray probe intensity values for Y chromosome genes *EIF1AY* (**A**), *TXLNGP2* (**B**), *KDM5D* (**C**). Samples are derived from paired tumor and adjacent non-tumor tissue. Dots within the male (blue) and female (red) groups were predicted to be the same sex as listed in the metadata. Samples with discrepant classification are represented by crosses, with the colour corresponding to the predicted sex. The pairs of samples within circle (1–4) were obtained from the same individual. These plots show the misclassification occurred for both samples in three pairs (1–3). In paired group 4, the misclassified sample was derived from the tumor tissue and the correctly classified sample was derived from adjacent normal tissue.

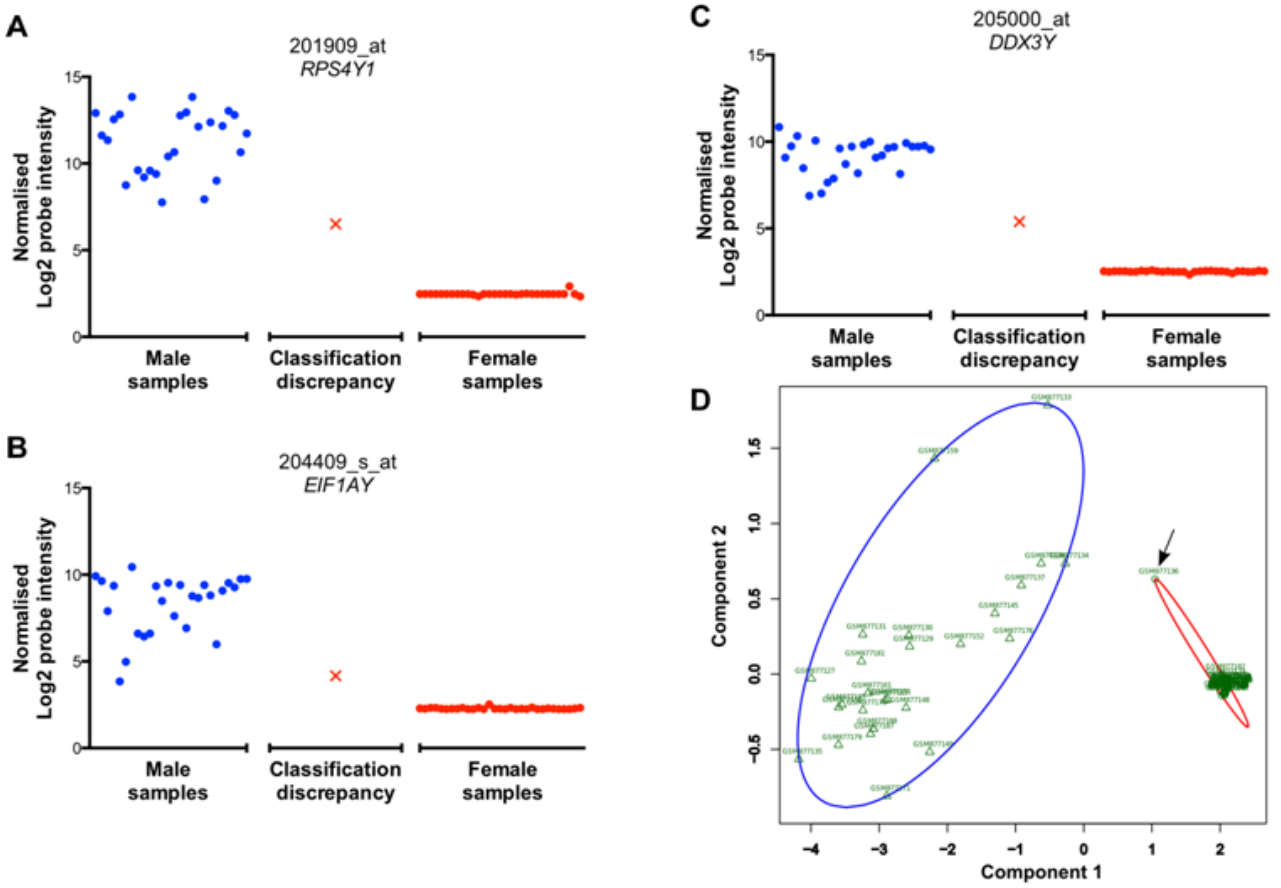


FIGURE S4.3: Prediction and validation of the sex of samples in dataset GSE35896. Plots show probe intensity values for Y chromosome genes *RPS4Y1* (**A**), *EIF1AY* (**B**), *DDX3Y* (**C**). One sample, indicated by the red cross was predicted to be female but listed as male in the metadata. This misclassified sample showed probe intensity values for all three genes greater than all other female samples (red dots), but less than that of males (blue dots), which suggests a genuine misclassification. When inspecting the PCA plot of these samples (**D**), this misclassified sample is plotted distinctly apart from the other female samples, although placed within female cluster.

S4.3.4 Detecting Datasets with Skewed Ratios

The *massiR* package includes a function that aids in detecting if a dataset has a skewed male/female ratio. This function calculates a standardized score for each sample and implements the dip test to test for unimodality. As a relatively sex-balanced dataset would typically show a bi-modal distribution of these standardized scores, the dip statistic is used to predict if a dataset shows a unimodal distribution that would be expected if a vast majority of samples were of one sex. We tested this function using the same randomly generated data as above to develop the guidelines for detecting dataset with skewed sex ratios which are outlined in the *massiR* package vignette (Supplementary Figure S4.6).

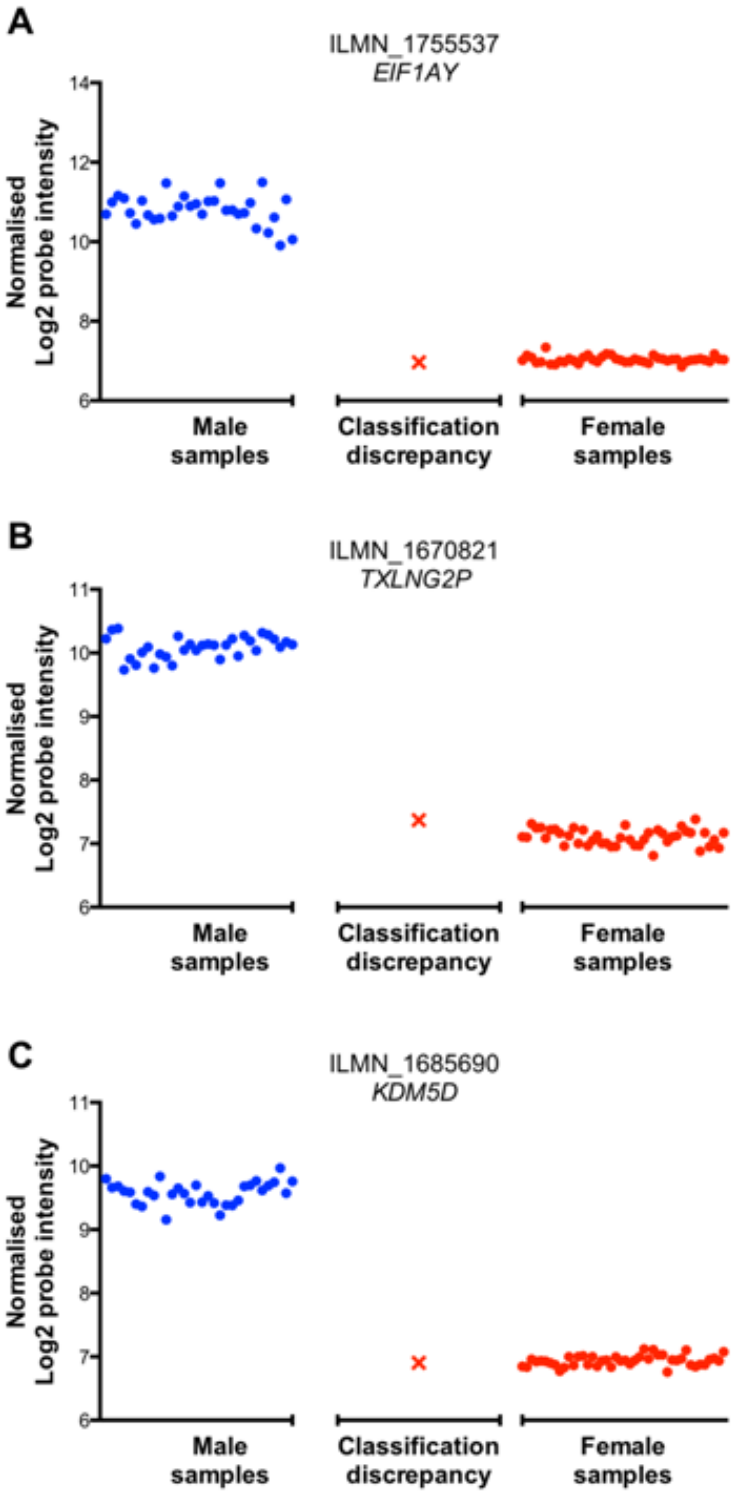


FIGURE S4.4: Prediction and validation of the sex of samples in dataset GSE45330. Plots show probe intensity values for Y chromosome genes *EIF1AY* (**A**), *TXLNGP2* (**B**), *KDM5D* (**C**). One sample, indicated by the red cross was predicted to be female but listed as male in the metadata. The Y chromosome probe intensity values for this sample are in the range of all other female samples (red dots) in this dataset and distinct from all the male samples (blue dots).

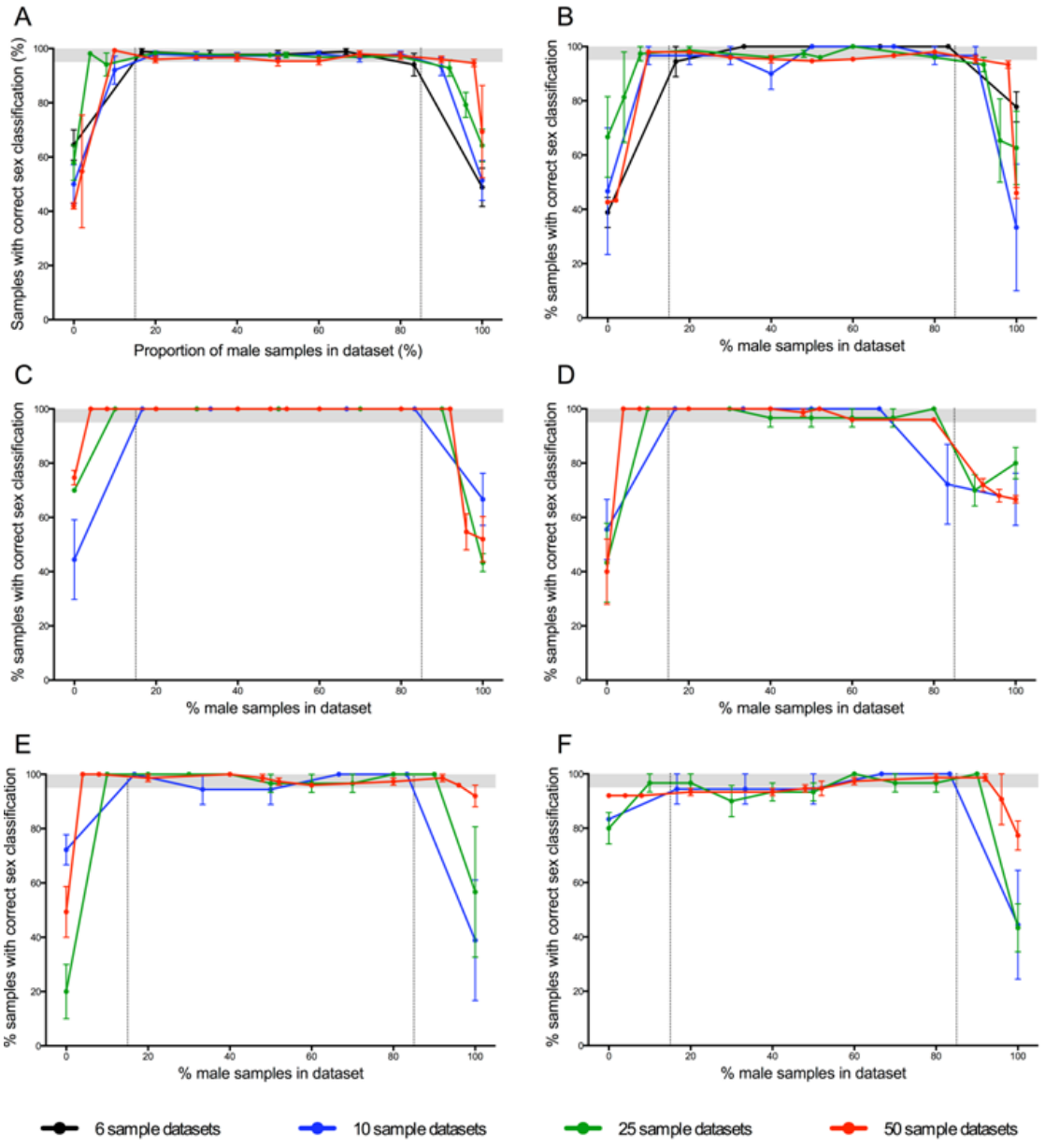
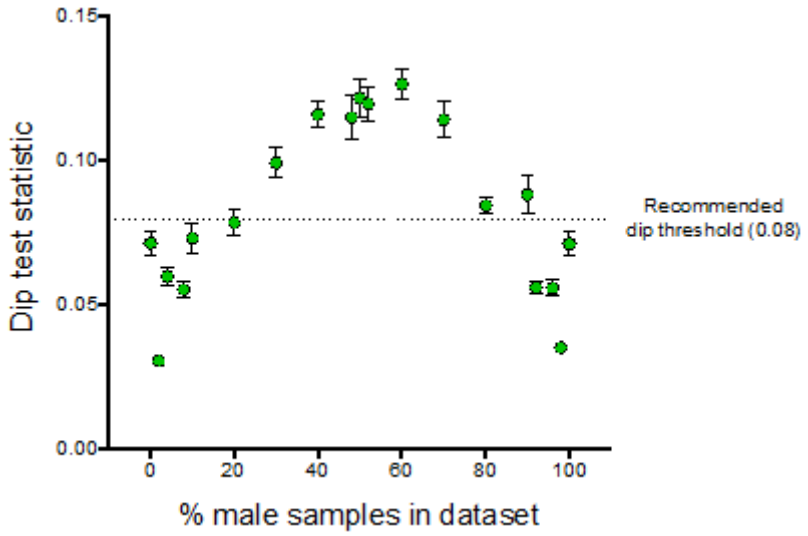


FIGURE S4.5: Sex prediction accuracy of the *massiR* package using five human gene expression datasets and a range of male/female ratios. (**A**) Results summary of the five datasets, (**B**) Kidney tissue (GSE40435), (**C**) placenta tissue (GSE25906), (**D**) colorectal tissue (GSE35896), (**E**) Blood mononucleocytes (GSE45330), (**F**) brain tissue (GSE29378). Points represent mean, vertical bars represent the standard error of the mean. The grey band at the top of the plot shows the 95–100% range. The correct sex prediction rate is 97.2% (±1.2 SEM) for datasets with $> 15\%$ and $< 85\%$ males which is the area between the vertical dotted lines.



[ht]

FIGURE S4.6: The dip test statistic as a method for identifying datasets with a skewed sex ratio. This plot shows the relationship between the dip test statistic as returned by the massi.dip function and the proportion of males in the dataset. This plot summarizes randomly selected sample and data subsets adapted from empirical kidney tissue (GSE40435), placenta tissue (GSE25906), colorectal tissue (GSE35896), Blood mononucleocytes (GSE45330), brain tissue (GSE29378) datasets. Points represent mean, vertical bars represent the standard error of the mean. Datasets with a dip test statistic greater than the threshold (0.08) are unlikely to feature skewed sex ratios that will affect the performance of predicting sample sex using the *massiR* package.

Bibliography

[1] Rung, J. and Brazma, A. Reuse of public genome-wide gene expression data. *Nat. Rev. Genet.,* 14 (2012), pp. 89–99.

[2] Ellegren, H. and Parsch, J. The evolution of sex-biased genes and sex-biased gene expression. *Nat. Rev. Genet.,* 8 (2007), pp. 689–698.

[3] Rinn, J. and Snyder, M. Sexual dimorphism in mammalian gene expression. *Trends in Genetics,* (2005).

[4] Ober, C., Loisel, D. A., and Gilad, Y. Sex-specific genetic architecture of human disease. *Nat Rev Genet,* 9(12) (2008), pp. 911–22.

[5] Hennig, C. *fpc: Flexible procedures for clustering.* R package version 2.1-5. 2013.

[6] Hartigan J. A. Hartigan, P. M. The dip test of unimodality. *Ann Stat,* 13 (1985), pp. 70–84.

Statement of Authorship

Title of Paper	Integrative transcriptome meta-analysis reveals widespread sex-biased gene expression at the human fetal–maternal interface.
Publication Status	◉ Published, ○ Accepted for Publication, ○ Submitted for Publication, ○ Publication style
Publication Details	Buckberry, S., Bianco-Miotto, T., Bent, S. J., Dekker, G. A. & Roberts, C. T. Integrative transcriptome meta-analysis reveals widespread sex-biased gene expression at the human fetal–maternal interface. Molecular Human Reproduction 20, 810-819 (2014).

Author Contributions

By signing the Statement of Authorship, each author certifies that their stated contribution to the publication is accurate and that permission is granted for the publication to be included in the candidate's thesis.

Name of Principal Author (Candidate)	Sam Buckberry		
Contribution to the Paper	Sam Buckberry conceived and designed the meta-analysis, carried out all bioinformatics and statistical analyses, interpreted the results, authored the first draft of the manuscript, incorporated suggestions from co-authors and peer-reviewers and wrote the final manuscript.		
Signature		Date	7/4/15

Name of Co-Author	Tina Bianco-Miotto		
Contribution to the Paper	Tina Bianco-Miotto was involved in study design, contributed valuable insight toward interpreting the results and provided detailed comments for all manuscript drafts.		
Signature		Date	1/4/15

Name of Co-Author	Stephen J Bent		
Contribution to the Paper	Stephen J Bent contributed valuable insight regarding the bioinformatics and statistical methods and provided detailed comments on the first and final manuscript drafts.		
Signature		Date	25/3/2015

Name of Co-Author	Gustaaf A Dekker		
Contribution to the Paper	Gustaaf A Dekker contributed valuable insight towards interpreting the results and commented on the final manuscript.		
Signature		Date	2/4/15

Statement of Authorship

Title of Paper	Integrative transcriptome meta-analysis reveals widespread sex-biased gene expression at the human fetal–maternal interface.
Publication Status	◉ Published, ○ Accepted for Publication, ○ Submitted for Publication, ○ Publication style
Publication Details	Buckberry, S., Bianco-Miotto, T., Bent, S. J., Dekker, G. A. & Roberts, C. T. Integrative transcriptome meta-analysis reveals widespread sex-biased gene expression at the human fetal–maternal interface. Molecular Human Reproduction 20, 810-819 (2014).

Author Contributions

By signing the Statement of Authorship, each author certifies that their stated contribution to the publication is accurate and that permission is granted for the publication to be included in the candidate's thesis.

Name of Principal Author (Candidate)	Sam Buckberry		
Contribution to the Paper	Sam Buckberry conceived and designed the meta-analysis, carried out all bioinformatics and statistical analyses, interpreted the results, authored the first draft of the manuscript, incorporated suggestions from co-authors and peer-reviewers and wrote the final manuscript.		
Signature		Date	7/4/15

Name of Co-Author	Claire T Roberts		
Contribution to the Paper	Claire T Roberts was involved in study design, contributed valuable insight towards interpreting the results and provided detailed comments for all manuscript drafts.		
Signature		Date	7.4.15

Name of Co-Author			
Contribution to the Paper			
Signature		Date	

Name of Co-Author			
Contribution to the Paper			
Signature		Date	

Chapter 5

Integrative Transcriptome Meta-Analysis Reveals Widespread Sex-Biased Gene Expression at the Human Fetal–Maternal Interface

Sam Buckberry, Tina Bianco-Miotto, Stephen J Bent, Gustaaf A Dekker and Claire T Roberts

Abstract

As males and females share highly similar genomes, the regulation of many sexually dimorphic traits is constrained to occur through sex-biased gene regulation. There is strong evidence that human males and females differ in terms of growth and development *in utero*, and that these divergent growth strategies appear to place males at increased risk when in sub-optimal conditions. Since the placenta is the interface of maternal-fetal exchange throughout pregnancy, these developmental differences are most likely orchestrated by differential placental function. To date, progress in this field has been hampered by a lack of genome-wide information on sex differences in placental gene expression. Therefore, our motivation in this study was to characterize sex-biased gene expression in the human placenta. We obtained gene expression data for $>$ 300 non-pathological placenta samples from

11 microarray datasets and applied mapping-based array probe re-annotation and inverse-variance meta-analysis methods which showed that > 140 genes (FDR < 0.05) are differentially expressed between male and female placentae. A majority of these genes ($> 60\%$) are autosomal, many of which are involved in high-level regulatory processes such as gene transcription, cell growth and proliferation, and hormonal function. Of particular interest, we detected higher female expression from all seven genes in the LHB-CGB cluster, which includes genes involved in placental development, the maintenance of pregnancy, and maternal immune tolerance of the conceptus. These results demonstrate that sex-biased gene expression in the normal human placenta occurs across the genome and includes genes that are central to growth, development, and the maintenance of pregnancy.

5.1 Introduction

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

The observation that males grow faster *in utero* and have a greater body length and weight at birth than females with equivalent placental size [1] indicates that the male placenta functions more efficiently [2, 3]. However, there is a developmental trade-off: a consequence of growing more quickly and being larger *in utero* is that males are left with less reserve placental capacity to draw upon if sub-optimal conditions arise. In turn, this places males at increased risk of under nutrition [3], which can restrict growth and lower birth weight, both of which have been linked to males' increased risk of adult-onset disorders such as cardiovascular disease [4]. A recent study has also shown a distinct male bias in the prevalence of placental dysfunction [5], and supports the findings of previous studies showing sex biases in a spectrum of pregnancy complications and fetal health outcomes associated with abnormal placental development [6–10]. Although sex differences in terms of growth, development and predisposition to pregnancy complications are increasingly becoming recognized, the underpinning sex biases in placental gene regulation remain unclear.

Recent efforts using massively parallel sequencing techniques have begun to expand our knowledge of the human placental transcriptional [11] and epigenetic landscapes [12]. These studies have revealed that the placenta is unique in several ways, including the expression of placenta-specific genes, placenta-specific alternative splicing, and widespread partially DNA methylated domains that regulate gene expression [11, 12]. Despite advancing our understanding of human placental gene regulation, these studies were not designed to capture the effect of sex, and therefore provide no clues as to underlying sex differences in placental function. An earlier study, which was the first to describe the human placental transcriptome, noted several genes with a sex-biased expression; these were located on both sex chromosomes and the autosomes [13]. However, given the low number of placental samples assessed, it is unlikely that the study was able to detect the true extent of sex-biased gene expression in the placenta.

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

5.2 Results and Discussion

5.2.1 Meta-Analysis of Sex-Biased Gene Expression in the Human Placenta

This meta-transcriptome analysis of the sex differences in human placental gene expression incorporated 303 samples from 11 microarray datasets generated on six different platforms (Table 5.1). We limited this analysis to non-pathological placental samples to provide the most accurate evaluation of sex differences in relative gene expression in normal human pregnancies at the time the fetus was delivered.

Table 5.1: Details of each datasets included in the meta-analysis

Dataset	GEO Accession	Array manufacturer	Array platform	No. of Samples	Male	Female
1	GSE10588 [14]	Applied Biosystems	Human Genome Survey v2	21	14	7
2	GSE12216 [15]	Applied Biosystems	Human Genome Survey v2	8	5	3
3	GSE18809 [16]	Affymetrix	U133 plus 2	9	3	6
4	GSE24129 [17]	Affymetrix	Human Genome 1 ST	8	5	3
5	GSE25906 [18]	Illumina	Human-6 v2	37	21	16
6	GSE27272 [19]	Illumina	HumanRef-8 v3	51	32	19
7	GSE28551 [20]	Applied Biosystems	Human Genome Survey v2	20	14	6
8	GSE30032 [21]	Illumina	HumanRef-8 v3	54	26	28
9	GSE35574 [22]	Illumina	Human-6 v2	40	23	17
10	GSE36828 (Unpublished)	Illumina	HumanHT-12 v3	47	26	21
11	GSE7434 [23]	Affymetrix	U133 plus 2	8	5	3
			Total	**303**	**174**	**129**

To improve microarray cross-platform concordance and to standardize gene identifiers throughout this meta-analysis, we re-annotated array probes by mapping to a common reference genome. After this re-annotation and summary process, we were able to quantify expression of 31,844 Ensembl genes (hereafter referred to as genes) across the human genome. To confirm the sex of samples and to predict sex when it was not listed in the associated meta-data or publication, we employed an unsupervised clustering technique that classifies the sex of samples in microarray datasets using signal intensity values for probes that map unambiguously to Y-chromosome genes [24].

When limiting the results to genes measurable in at least three studies and with a false discovery rate (FDR) of < 0.05, a total of 142 genes showed significant sex-biased expression. Of these 142 genes, 75 showed higher expression in placentas from female fetuses and 67 genes were more highly expressed in placentas from male fetuses (Figure 5.1). At the FDR of 0.05, we expect 3.75 and 3.35 genes to be false positives in female and male groups, respectively. In the female group, 55 up-regulated genes were autosomal and 20 were X-linked. Of genes significantly up-regulated in the male group, 33 genes were expressed from the autosomes, 16 were expressed from the X chromosome, and 18 were Y chromosome genes (Figure 5.2). We do not consider the Y-linked genes to be differentially expressed; rather these genes are expressed at consistently detectable levels in placentas from male fetuses, and therefore may potentially influence placental phenotype.

The majority of sex-biased genes were autosomal but, as expected, many were located on the sex chromosomes. The X-linked and autosomal genes with the highest level of significance were *HDHD1* and CGB, respectively (Figure 5.1). When inspecting the contribution of individual studies for autosomal gene expression bias, despite there being a lower magnitude of difference, the direction of change was consistent across datasets for many male and female biased genes (Figure 5.3). The results for all genes, the number of studies where they were measurable, and the expression differences with statistics are provided in Supplementary Data File 1.

When comparing these results with previous studies where sex-biased expression has been assessed in other human tissues, genes showing sex-biased expression appear to exhibit that bias with a high degree of tissue specificity. A vast majority of sex-biased genes in the human placenta are not observed to have sex-biased expression in human brain, liver or blood [27–29] (Figure S5.1A). When comparing our results to studies where sex-biased expression was assessed in placental tissue or cells, many of the genes in this study have no previously reported sex expression bias (Figure S5.1B).

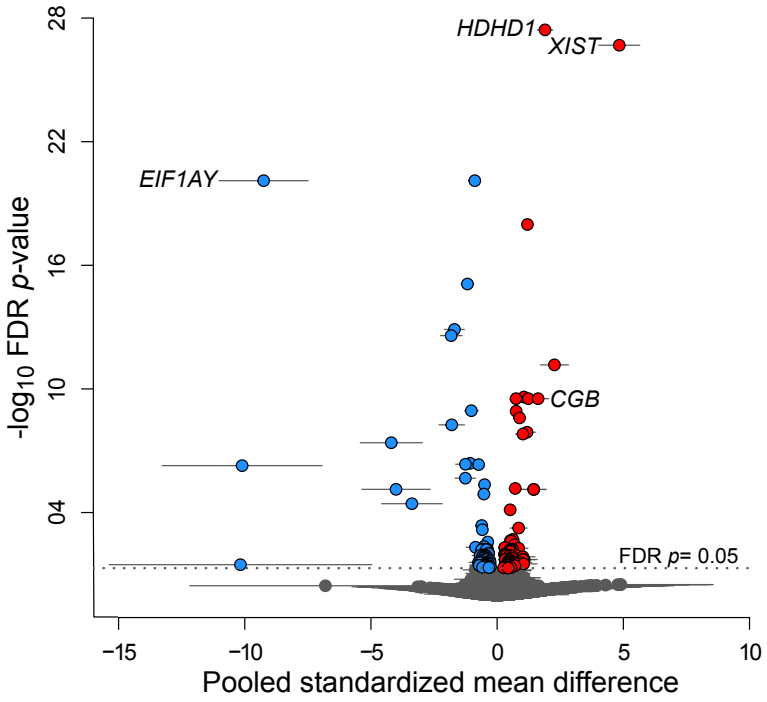


FIGURE 5.1: Volcano plot showing pooled effect size and level of significance for 31,844 Ensembl genes when comparing sex-biased gene expression in the human placenta. Blue dots represent genes with significant male-biased expression and red dots are those with significant female-biased expression. Horizontal bars indicate the 95% confidence interval. Points represent genes detected in at least three studies.

5.2.2 Identification of Potential Transcriptional Regulators of Sex-Biased Gene Expression

To predict transcription factors (TFs) that may be involved in regulating sex-biased gene expression, we searched for conserved transcription factor binding sites (TFBS) in the 10kb of DNA sequence up and downstream of the transcription start sites of sex-biased genes. This was done using oPOSSUM-3 and the JASPAR core motifs [30, 31]. This analysis identified potential binding sites for 166 vertebrate TFs.

Since the results of this analysis are best interpreted using relative rankings [30], we selected the TFs that ranked in the upper quartiles of both z-scores and Fisher scores (Figure S5.2), which limited the initial list to 14 TFs (Table S5.1). In order to further investigate whether these TFs may be involved in regulating sex-biased gene expression, we checked if the genes encoding these TFs were expressed at detectable levels in the human term placenta using publicly available RNA-Seq data [11]. In this comparison, expression data for nine of these TFs were available,

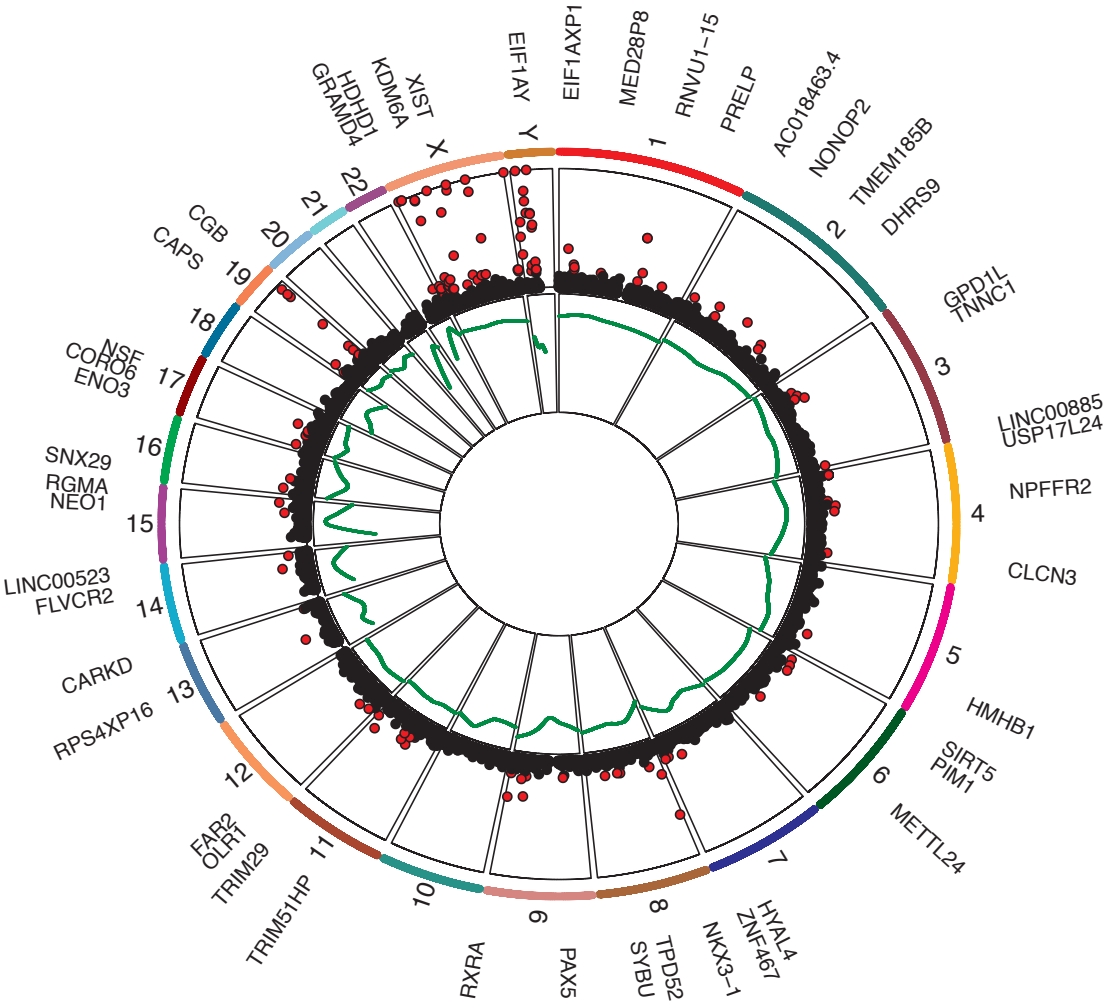


FIGURE 5.2: Circos plot summarizing the meta-analysis of sex-biased gene expression in the human placenta. The outer-most scatterplot track depicts chromosomal location and level of significance ($-\log 10$ FDR p-value). The red points represent genes with FDR p-values < 0.05. The closer points are to the outside of the track, the higher the significance. The inner track is a loess smoothed line plot representing the number of datasets where information was available for each genomic region, ranging from 3 to 11 datasets. Gene labels for selected genes of significance are plotted outside the chromosome highlights. Circos plot was generated using an R implementation of Circos [25, 26].

with seven being expressed at detectable levels and comparable with human adult tissues expression (Figure 5.4A and Figure S5.3).

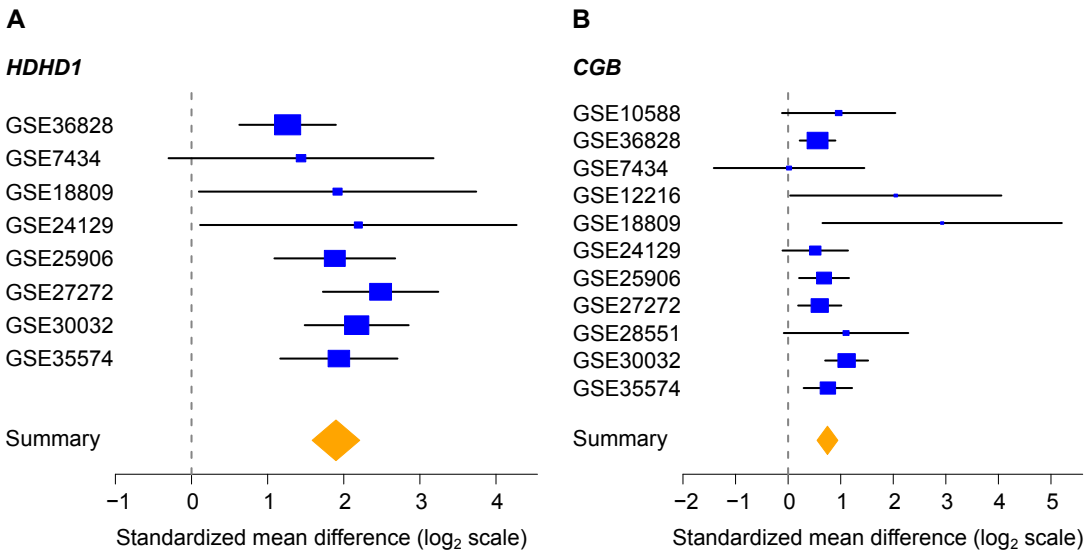


FIGURE 5.3: Forest plots showing the standardized mean difference between males and females for the most statistically significant X-linked gene *HDHD1* (**A**) and autosomal gene *CGB* (**B**). Size of the blue box for each study is proportional to sample size; horizontal lines represent standard error. Yellow diamond represents the gene summary across all studies where the gene was detectable. GEO accession identifiers on the y-axis represent datasets.

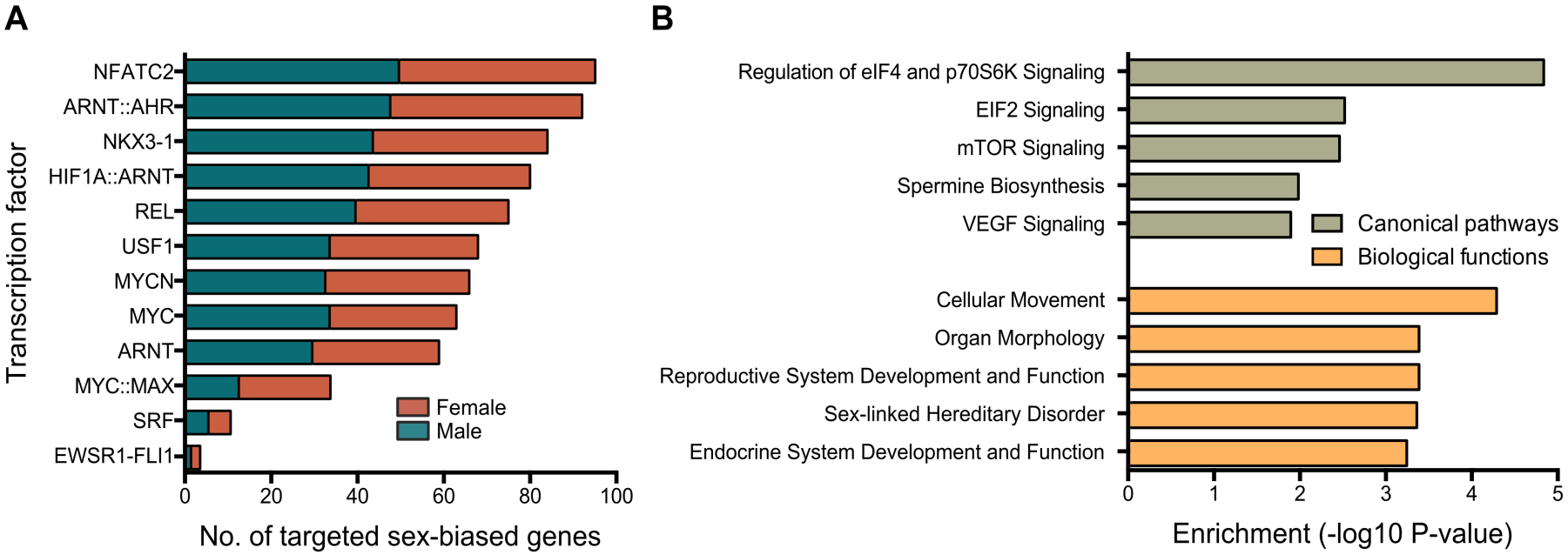


FIGURE 5.4: (**A**) Transcription factors expressed in the human placenta that show enriched binding domains surrounding genes with sex-biased expression. (**B**) Top biological functions and canonical pathways associated with sex-biased gene expression in the human placenta. Functions and pathways were determined using Ingenuity Pathway Analysis.

Expression of *MYCN* is highest in placental tissue when compared to any of the adult tissues (Figure S5.3); this is also the gene that encodes the TF of highest significance in the TF-binding motif analysis (Table S5.1). *NKX3-1*, which encodes a homeobox-containing TF, showed significant female expression bias in this meta-analysis, and significant enrichment in the TF-binding motif analysis. *NKX3-1* expression in the placenta is detectable and comparable with a majority of other adult tissues (Figure S5.3). *NKX3-1* is a tumor suppressor and its expression appears to be strictly regulated by androgens and loss of its expression is associated with prostate cancer development [32]. This suggests the *NKX3-1* female expression bias observed in this study may be due to different androgen profiles in male and female placentas, which in turn may drive sex differences in the transcription of the numerous *NKX3-1* target genes.

RXRA, which encodes a hypoxic responsive hormone receptor and TF, showed a consistent male expression bias in this study. Although falling just below our cut-off criteria for enriched TF binding sites, the binding of RXRA in four different complexes with other proteins was detected in the enrichment analysis (Figure S5.2). In the mouse, *RXRA* knockout placentas exhibit multiple defects, and RXRA antagonists are known to be involved in stimulating hCG production through interaction with *CGB* gene promoters [33] (see results below). *RXRA* was also identified as a target of MYCN in the TF-binding motif analysis, suggesting the *RXRA*-encoded TF may be a significant player in defining the sex differences in gene transcription and placental function.

5.2.3 High-Level Molecular Functions and Pathways are Associated with Sex-Biased Genes

Since pathway analysis is a valuable tool in estimating gene function in different tissues and systems, we applied the list of sex-biased genes to search for molecular pathways and processes statistically enriched with sex-biased genes. Ingenuity Pathway Analysis showed that sex-biased genes are involved with high-level functions such as cellular movement, organ morphology and endocrine function (Figure 4B). Among the top five canonical pathways associated with sex-biased genes were mTOR and VEGF signaling (Figure 5.4B). The mTOR signaling pathway is a key regulator of cell growth and proliferation, and is activated during angiogenesis [34]. The VEGF pathway involves many genes implicated in angiogenesis, placental development and adverse pregnancy outcomes [35]. These suggest that sex-biased expression of genes involved in these placental development pathways could potentially drive differential function of pathways involved in other key placental

processes such as establishing the vascular architecture (angiogenesis), and the proliferation of placental cells.

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

5.2.4 Sex-Biased Expression of X-Linked Genes

In female mammals, one of the two X chromosomes is typically inactivated to compensate for gene dosage differences between the sexes (for review see refs [36, 37]). However, some genes escape X-inactivation (XCI) and are expressed from both X chromosomes in females. Subsequently, those genes that escape XCI potentially contribute to sexually dimorphic traits.

Numerous studies have measured escape from XCI in human cells and tissues, although the extent of escape from XCI in extra-embryonic tissues, including the human placenta, remains controversial [38]. We detected 20 X-linked genes with significant female-biased expression, many of which appear to cluster in distinct chromosomal regions (Figure 5.5). The most significant of these genes was *HDHD1* (Figure 5.3A), which encodes a phosphatase involved in the dephosphorylation of modified RNA nucleotides [39]. Additionally, the long non-coding RNAs *XIST* and *JPX*, which are known to be involved in the mechanisms giving rise to XCI [36, 37], also showed significant female expression bias, as expected.

To assess whether escape from XCI may be the underlying cause of X-linked gene expression bias in this study, we compared our results with a previously published extensive profile of human XCI [40] (Figure 5.5). Of the 20 X-linked genes with female expression bias, XCI profiling information was available for 16, of which 11 had strong evidence of expression from the inactive X chromosome [40]. This suggests that this is most likely to be the primary cause of X-linked female expression bias. The female biased X-linked genes are associated with several biological functions, including conversion of sulfated steroid precursors to estrogens (*STS*), and histone demethylation (*KDM6A*).

Additionally, we observed clusters of X-linked genes with male expression bias including five genes at the pseudo-autosomal Xp22.33 region, and two genes at Xq22.1 in the ARMCX family (Figure 5.5.) The ARMCX3 and ARMCX6 genes

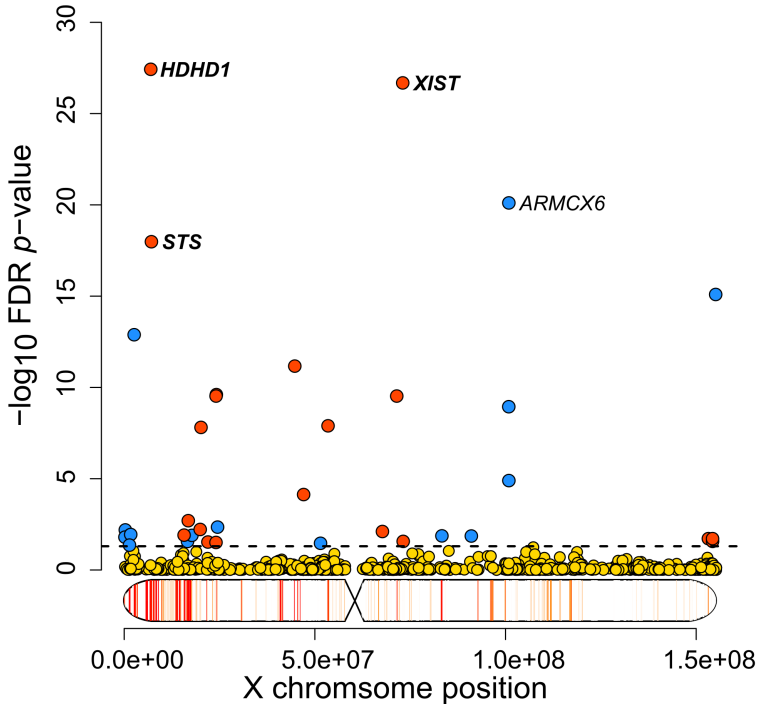


FIGURE 5.5: Sex-biased expression of X-linked genes in the human placenta. Red points indicate genes showing significantly higher expression in female samples, blue points represent genes expressed significantly higher in male samples. Heat map below plot represents the level of expression from the inactive X chromosome observed in [40].

are thought to have originated during the evolution of placental mammals, and are known to be involved in mitochondrial regulation [41].

Taken together, the X-linked genes comprise a considerable proportion of highly significant sex-biased genes detected in this study, and have biological functions relating to hormone regulation, and higher order regulatory mechanisms such as RNA modification and histone methylation. Given that the sex chromosomes define the difference between the sexes at a cellular level, sex chromosome genes with expression biases are clearly potential drivers or regulators of sex-biased autosomal gene expression.

5.2.5 LHB-CGB Cluster Genes Show Female Expression Bias

Among the sex-biased autosomal genes, the LHB-CGB cluster of seven genes on chromosome 19 showed the most significant female expression bias (Figure 5.6). This contiguous cluster consists of the *LHB* gene that encodes the beta-subunit of

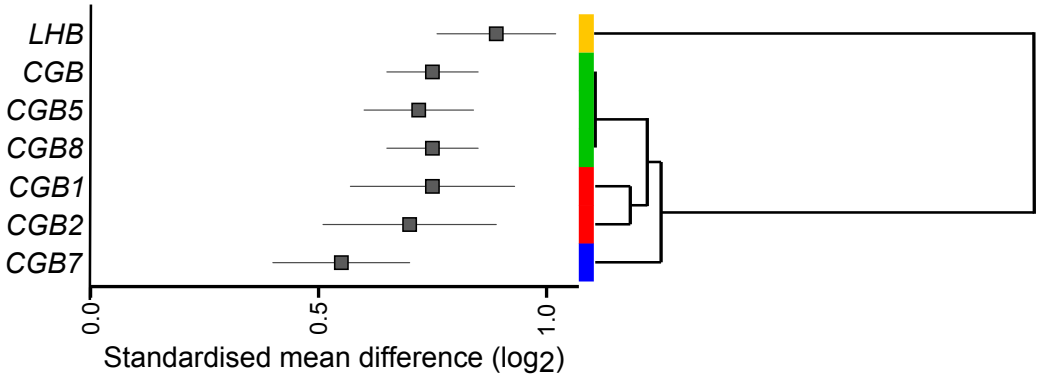


FIGURE 5.6: Female-biased expression of *LHB* and *CGB* cluster genes. Boxes represent mean expression difference for each gene, and bars represent 95% confidence interval. FDR p-value < 0.05 for all genes. Dendrogram shows the average distance between genes using percent identity of the amino acid sequences, and branch lengths represent the percentage mismatch between two nodes. Colored bars depict functional groupings; Yellow represents *LHB*, green represents identical *CGB* protein isoforms, red represents pseudogenes, blue represents a divergent CGB isoform. Note that *CGB*, *CGB5* and *CGB8* have identical amino acid sequences.

luteinizing hormone (LH), four chorionic gonadotropin (hCG) beta-subunit coding genes (*CGB*, *CGB5*, *CGB7*, and *CGB8*) and two pseudogenes (*CGB1* and *CGB2*) [42]. The four *CGB* genes encoding hCG beta-subunits can be grouped into two classes based on protein sequence; *CGB*, *CGB5*, and *CGB8* encode identical amino acid sequences, while *CGB7* encodes a variant peptide (Figure 5.6 and Figure S5.4).

The *CGB*-encoded hCG hormone is the important embryonic signal for maternal recognition of pregnancy in primates. Indeed it is essential for the prolongation of corpus luteal function and hence progesterone synthesis until the placenta takes over. The many functions of hCG relating to placental growth, invasion, angiogenesis, and the regulation of maternal immune tolerance of the placenta and fetus are well-described (for reviews see [43, 44]). Lowered *CGB* expression in the placenta has also been observed in miscarriages, and is higher in ectopic, molar, and growth-restricted pregnancies [45, 46]. The *LHB*-encoded luteinizing hormone (LH) is primarily expressed in the pituitary gland, and is widely known for its action in the gonads to induce sex steroid synthesis and gametogenesis (see reference [47] for review). However, LHB is also expressed at appreciable levels in the human placenta (Figure S5.5), which is a feature that appears to be conserved across therian mammals [48].

Both hCG and LH hormones bind to the same transmembrane receptor (LHCGR), which is known to induce multiple signals including cyclic adenosine monophosphate (cAMP) [49]. hCG is also known to regulate *VEGF* and its receptors [50], which are heavily implicated in placental development and adverse pregnancy outcomes [35]. Meta-analysis profiling of placental gene expression in preeclampsia indicates that up-regulation of *LHB* contributes to the gene expression signature of preeclampsia [51], while up-regulation of *LHB* and *CGB* in the placenta are associated with intrauterine growth restriction [46].

Taken together, these results provide substantial evidence for female-biased expression of the hormone-coding *LHB* and *CGB* genes in the human placenta. This suggests that, through the actions of *LHB* and *CGB* genes, female fetuses may invest more in placental growth and vasculogenesis, while males invest these resources in body growth. Indeed, the ratio of birthweight to placental weight in male human infants is higher than for females [7] suggesting that to maintain a high growth rate the male fetus extracts maximal nutrients from the placenta with little reserve capacity if adversity strikes. Perturbed expression of *LHB* and *CGB* is also associated with preeclampsia [51] and intrauterine growth restriction [46] where placental pathology is implicated. This indicates fetal sex-specific risks for these conditions could be partially attributable to differential regulation of gene networks involving these genes.

5.3 Conclusions

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

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

The results presented here also demonstrate sex-biased expression for many autosomal genes, including genes encoding the LH and hCG hormones. Given that LH and hCG have a potent ability in promoting placental growth and vasculogenesis,

these results suggest that female fetuses invest more in extra-embryonic tissue development than males. Since mothers can allocate limited resources to a fetus in utero, these findings support the conjecture that males invest more resources in body growth and development (embryonic tissues) at the expense of investing less in the development of extra-embryonic tissues [3, 6]. This may be a key reason as to why there is a male bias in the incidence of placental dysfunction [5], and in pregnancy complications where placental pathology is implicated [51–53].

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

5.4 Materials and Methods

5.4.1 Study Selection

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

5.4.2 Array Pre-Processing and Quality Control

Since data were obtained from multiple microarray platforms, pre-processing methods were tailored to each platform. Affymetrix datasets were pre-processed, log-transformed and normalized using either the robust multi-array average (RMA) or

GeneChip-RMA (GC-RMA) method depending on platform using Simpleaffy [54]. Applied Biosystems arrays were pre-processed using comprehensive R-based microarray analysis (CARMA); probes with a flag value of > 100 were removed from the dataset before quantile normalization [55]. Illumina bead arrays were pre-processed using Beadarray before quantile normalization [56]. Datasets with arrays processed in multiple batches (as detailed in the meta-data), were batch-corrected using the 'comBat' function in the SVA package [57, 58]. Outliers were eliminated from each dataset (see table S5.2) before re-normalization by checking the distance between arrays and assessing MA plots generated using ArrayQualityMetrics [59].

5.4.3 Predicting The Sex of Samples in Datasets Lacking Sex Information

In 7 of the 11 datasets in this meta-analysis, the sample's sex was not identified in either the associated repository meta-data or in the associated publication. Therefore, to maximize the number of usable datasets, we used the Bioconductor package *massiR* to predict fetal sex [24]. This method utilizes expression values for probes that map unambiguously to the Y chromosome, and unsupervised clustering of samples based on Y chromosome probes with the highest variance.

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

5.4.4 Re-Annotation of Microarray Datasets and Probe Summarization

Gene expression data were obtained from various microarray platforms that have different probes targeting the same genes. We therefore annotated each dataset with common gene identifiers to increase cross-platform concordance. We selected

the gene identifiers from Ensembl Genes release 69 annotation [60] for probe mapping, which is the genome annotation used in the human GENCODE project [61]. Probes from all Illumina and Affymetrix datasets were mapped to the human reference genome (GRCh37.3p) to translate platform-specific, probe-level identifiers to the Ensembl gene level identifiers. Probes were mapped using the Ensembl Functional Genomics Array Mapping Environment, in which individual probes are mapped to both the genome and the cDNA sequence. Alignments were performed by Ensembl using an analysis pipeline which implements the Exonerate sequence comparison and alignment tool [62]. A 1 bp mismatch was permitted between the probe and the genome sequence assembly, and probes that match at 100+ locations (e.g. suspected Alu repeats) are discarded (see permalink for detailed methods http://jan2013.archive.ensembl.org/info/docs/microarray_probe_set_mapping.html).

Probe sequences were unavailable for the Applied Biosystems arrays, so Ensembl gene identifiers supplied by the manufacturer were used to identify target genes. For Applied Biosystems probes with no listed Ensembl identifier, the supplied gene symbol was used to identify the target gene using the HGNC database [63], and an Ensembl Gene identifier was subsequently assigned. Any remaining identifiers with GenBank accessions [64] were checked for a match against human sequences with sufficient gene information, and then designated an Ensembl Gene Identifier.

Probe mapping and annotation of all datasets (except Applied Biosystems arrays) allowed identification of four cases: (1) probes that map uniquely to a single gene identifier (one-to-one mapping), (2) probes that map to multiple gene identifiers (one-to-many mapping), (3) multiple probes that map to the same gene identifier (many-to-one mapping), and (4) probes that do not map to any genes in the reference genome. These re-annotation results are summarized in Table S5.3.

When a probe mapped to multiple gene identifiers, (case 2) a new probe identifier was created for each probe-to-gene mapping; this allowed the use of all possible information for each gene in the analysis. For gene identifiers where multiple probes were mapped, (case 3) probe values were summarized into a single representative value per gene identifier within each study, using a fixed inverse-variance model as previously described [65]. Probes with insufficient information, or that did not map any gene identifier, were removed from the analysis.

5.4.5 Meta-Analysis of Annotated Datasets

To meta-analyze the 11 annotated microarray datasets, we applied the inverse-variance method as detailed in [65] using the *rmeta* package, functions adapted from the *metaGEM* package (https://spiral.imperial.ac.uk/handle/10044/1/4217) and custom R scripts. Briefly, study-specific effect sizes were calculated for each probe within each study by calculating the probe mean and standard deviation, corrected for effect size using Hedges' g to account for the number of samples in each group. These study-specific estimates were then combined using a random effects inverse-variance method for each gene identifier to calculate the pooled effect size and standard error. Z-statistics were then calculated for each gene identifier to obtain a nominal p-value, which was then corrected using the False Discovery Rate (FDR). After significance testing, the resulting dataset was limited to genes represented by at least three studies for downstream analyses. All data processing and analyses were carried out in the R statistical environment (version 2.15.2).

5.4.6 Prediction of Upstream Transcription Factor Regulation

The sex-biased gene set was analyzed for enrichment of TFBSs using the oPOSSUM program, and the JASPAR vertebrate core profiles [30, 31]. For each gene, we searched for TF binding motifs in the conserved regions of the 10kb upstream/-downstream sequences using a conservation cut-off of 0.4, a matrix score threshold of 85% and a minimum specificity of 8-bits. The highly enriched TFBSs were identified by ranking transcription factors using results from Fisher's exact test and z-score rankings.

5.4.7 Resolving CGB/LHB Cluster Sequence Homology

Genes in the LHB-CGB cluster are both functionally and evolutionarily related [42], and subsequently have a high degree of sequence homology. In such cases, the sequence specificity of each microarray probe is a key determinant in differentiating between the expression of individual genes. We re-annotated all array probes through mapping to a common reference genome, therefore were able to determine which probes mapped uniquely, or mapped to multiple genes, in the LHB-CGB cluster. In this meta-analysis, all probes that represent *LHB* expression mapped uniquely; therefore it is unlikely that the *LHB* expression results are

confounded by non-specific binding with *CGB* gene transcripts. In the case of the *CGB* genes, mapping specificity differed between platforms. Affymetrix probes mapped with low specificity: 10 probes mapped to all *CGB* cluster genes, and only one probe mapped uniquely (to *CGB7*). However, probes from the Illumina platforms mapped with much higher specificity. Of these, nine probes mapped specifically genes in one of the three classes of CGB protein isoforms (Figure 5.6), and three of these probes had single-gene specificity. A majority of samples in this study (76%) were assayed on Illumina platforms, so we have reasonably high confidence that the expression results for *CGB* genes are composed primarily of values from probes with the highest specificity.

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

5.4.8 Gene Enrichment and Pathway Analysis

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

Funding

SB is supported by a Healthy Development Adelaide and Channel 7 Children's Research Foundation PhD Scholarship and an Australian Postgraduate Award. TB-M is supported by the Cancer Council SA and SAHMRI Beat Cancer Project (TBM APP1030945). CTR is supported by a National Health and Medical Research Council (NHMRC) Senior Research Fellowship APP1020749 (http://www.nhmrc.gov.au). This project was funded in part by NHMRC Project APP1059120 awarded to CTR, TB-M and SJB. The funders had no role in study design, analysis, decision to publish, or preparation of the manuscript.

Acknowledgements

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

S5.1 Supplementary Information

This file contains a summary of the results of this meta-analysis. It includes details for all genes including Ensembl gene ID, HUGO gene symbol, number of studies with probes mapped to each gene, standardized mean difference summary for each gene and associated standard error, p-values and FDR p-values.

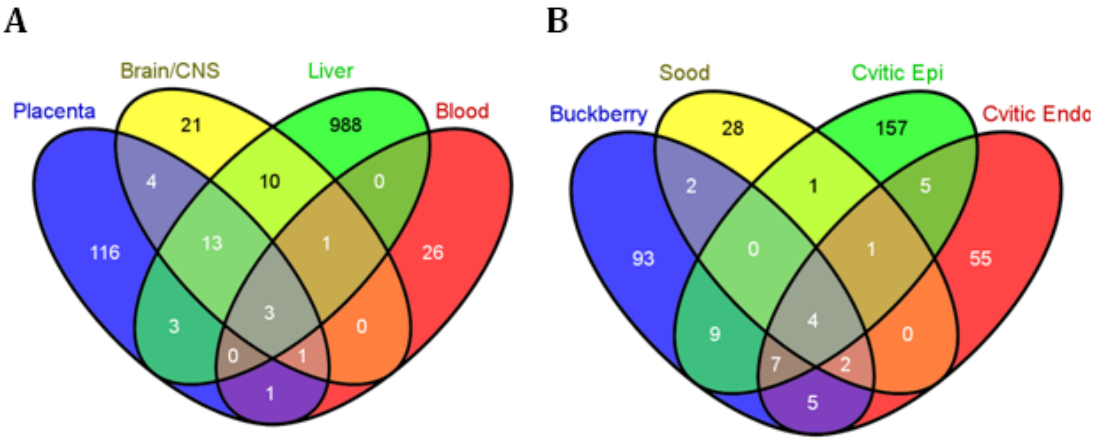


FIGURE S5.1: (**A**) Venn diagram showing the overlap of sex-biased genes in different human tissue/cell types. This comparison of the results from previously published work [27–29] suggests sex-biased gene expression is largely regulated in a tissue specific manner. (**B**) Comparison of differentially expressed genes in this meta-analysis and previously studies investigating sex-biased gene expression in placental tissue [13], and placental epithelium and endothelial tissue [66]. A majority of overlapping genes are located on the Y chromosome. Venn diagrams created using the published lists of differentially expressed genes in each study. Gene identifiers from each study were converted to common identifiers for this comparison.

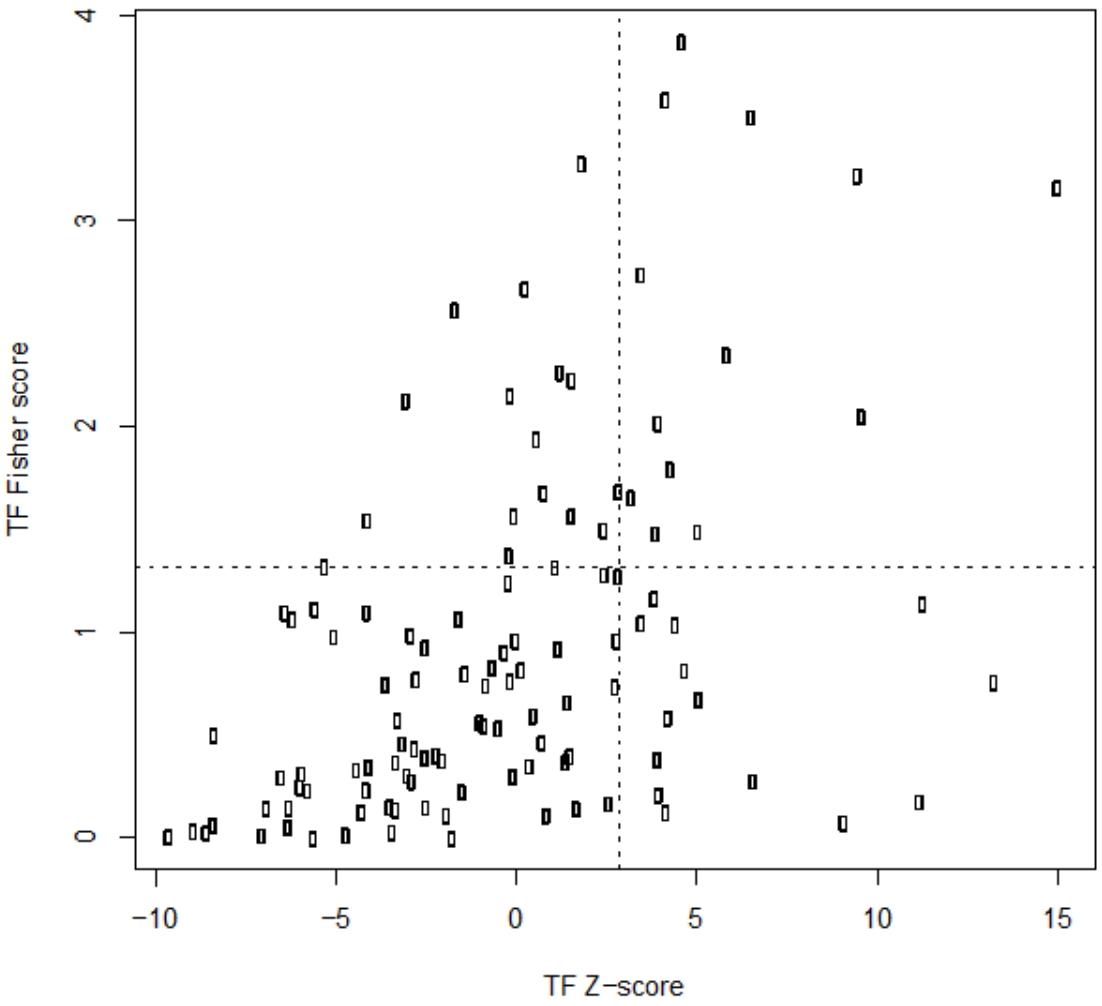


FIGURE S5.2: Statistical measures for transcription factor binding motifs flanking transcription start sites of genes with sex-biased expression in the human placenta. Each point on the graph represents a transcription factor with experimentally validated binding motif. Red points represent transcription factors in the upper quartile (dashed lines) of both z-scores and Fisher score rankings. Z-scores on the X-axis are indicative of the likelihood that the number of TFBS nucleotides detected for the sex-biased genes is significant as compared with the number of TFBS nucleotides detected for the entire gene set in the meta-analysis. For each transcription factor, the Fisher scores are the negative natural log of the probability that the number of hits vs. non-hits for the sex-biased genes could have occurred by random chance based on the hits vs. non-hits for the entire background gene set.

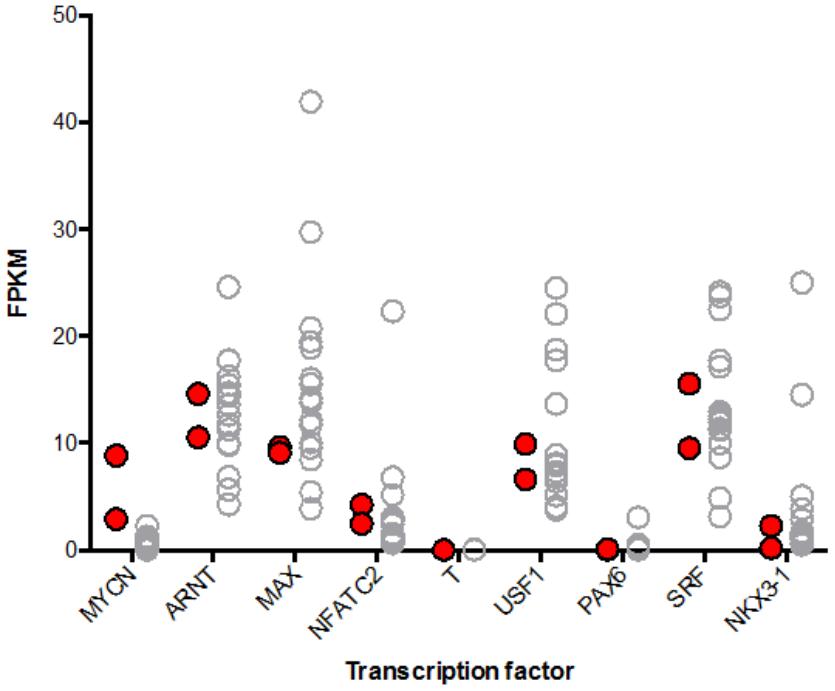


FIGURE S5.3: Comparative gene expression levels of transcription factors in the human term placenta (chorion and decidua as red dots) and sixteen adult human tissues (grey circles). Gene expression represented as Fragments per Kilobase per Million mapped reads (FPKM). Genes with FPKM > 0.1 were considered detectable. This figure was created using RNA-Seq data adapted from Kim et al. and the Illumina human body map dataset [11].

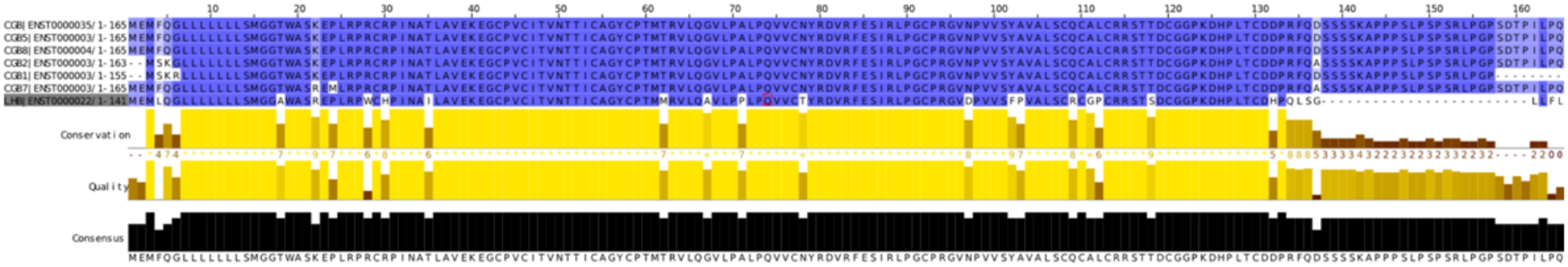


FIGURE S5.4: LHB-CGB gene cluster amino acid sequence alignment. Amino acid sequences for each gene were downloaded from ENSEMBL and aligned using MAFFT (v7.130b) with L-INS-i settings.

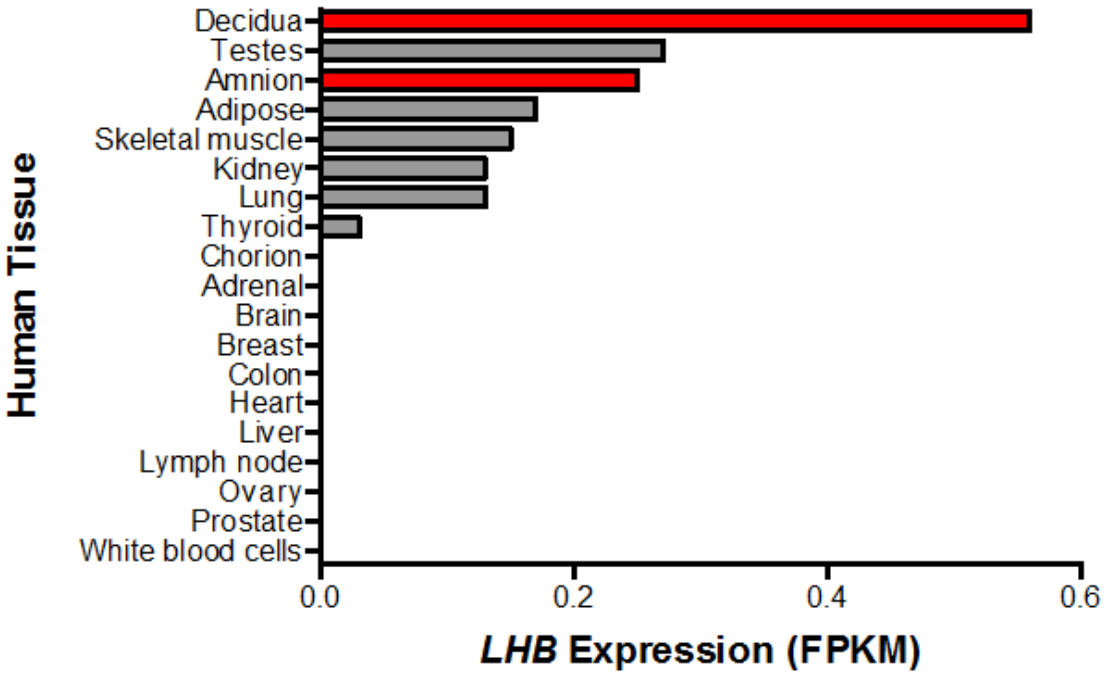


FIGURE S5.5: *LHB* expression in multiple human tissues, with extra-embryonic tissues highlighted in red. This figure was created using RNA-Seq data from Kim et al. and the Illumina human body map dataset [11]. Note that expression data from the pituitary gland is not included in this dataset.

Table S5.1: Transcription factors with enriched binding sites surrounding genes with sex biased expression.

Transcription factor	Class	Family	Associated gene ID	Expression in human placenta	Sex-biased gene hits	FDR
MYCN	Zipper-Type	Helix-Loop-Helix	ENSG00000134323	Detected	65	4.80E-49
MYC	Zipper-Type	Helix-Loop-Helix	–	No data	62	1.53E-20
ARNT::AHR	Zipper-Type	Helix-Loop-Helix	–	No data	91	3.19E-20
ARNT	Zipper-Type	Helix-Loop-Helix	ENSG00000143437	Detected	58	3.98E-10
MYC::MAX	Zipper-Type	Helix-Loop-Helix	ENSG00000125952	Detected	33	1.97E-08
NFATC2	Ig-fold	Rel	ENSG00000101096	Detected	94	1.45E-06
T	Beta-Hairpin-Ribbon	T	ENSG00000164458	Not detected	23	1.18E-05
USF1	Zipper-Type	Helix-Loop-Helix	ENSG00000158773	Detected	67	4.83E-05
HIF1A::ARNT	Zipper-Type	Helix-Loop-Helix	–	No data	79	7.89E-05
REL	Ig-fold	Rel	–	No data	74	1.73E-04
EWSR1-FLI1	Winged Helix-Turn-Helix	Ets	–	No data	3	2.23E-04
PAX6	Helix-Turn-Helix	Homeo	ENSG00000007372	Not detected	15	9.88E-04
SRF	Other Alpha-Helix	MADS	ENSG00000112658	Detected	10	2.38E-03
NKX3-1	Helix-Turn-Helix	Homeo	ENSG00000167034	Detected	83	7.07E-03
MYCN	Zipper-Type	Helix-Loop-Helix	ENSG00000134323	Detected	65	4.80E-49

TABLE S5.2: Number of arrays at each step of the selection process for the meta analysis.

Dataset Accession	No. of arrays in dataset	No. of arrays from non-pathological samples	Number of arrays failed QC	No. of arrays included in meta-analysis
GSE10588	43	21	0	21
GSE12216	16	8	0	8
GSE18809	10	10	1	9
GSE24129	24	8	0	8
GSE25906	60	37	0	37
GSE27272	54	54	3	51
GSE28551	37	21	1	20
GSE30032	57	57	3	54
GSE35574	94	40	0	40
GSE36828	48	48	1	47
GSE7434	10	10	2	8
Total	**453**	**314**	**11**	**303**

Table S5.3: Re-annotation and gene summarization details for each microarray platfoms.

Manufacturer	Platform	Total probes	Probes mapped to Ensembl genes	Probes with one-to-one mapping	ID's after gene expanding	Total Ensembl genes after Summarisation
Applied Biosystems	Human Genome Survey 2	32,878	16,265	16,185	16,345	14,391
Affymetrix	Human Gene 1 ST Array	32,321	28,126	24,760	36,260	27,928
Affymetrix	Human Genome U133 Plus 2	54,675	29,206	28,451	30,378	15,895
Illumina	Human-6 2	48,701	26,952	24,810	30,321	22,699
Illumina	HumanHT-12 3	48,804	31,211	28,630	35,444	22,903
Illumina	HumanRef-8 3	24,526	22,346	20,989	24,193	18,007

Table S5.4: Top 5 Ingenutiy canonical pathways and biological functions enriched for sex-biased genes.

	Enriched pathway/function	$-\log$ (p-value)	No. of molecules	Molecules with male biased expression	Molecules with female biased expression
Canonical pathways	Regulation of eIF4 and p70S6K Signaling	4.81	7	EIF1AY, RPS4Y1, PPP2R3B, RPS4Y2	RPS4X, EIF1AX, EIF2S3
	EIF2 Signaling	2.49	5	RPS4Y1, RPS4Y2	RPS4X, EIF1AX, EIF2S3
	mTOR Signaling	2.42	5	RPS4Y1, RPS6KA6, PPP2R3B, RPS4Y2	RPS4X
	Spermine Biosynthesis	1.97	1	–	SMS
	VEGF Signaling	1.87	3	EIF1AY	EIF1AX, EIF2S3
Biological functions	Cellular Movement	4.27	3	–	ANGPT2, HSPE1, NKX3-1
	Organ Morphology	3.37	2	–	LHB, NKX3-1
	Reproductive System Development and Function	3.37	2	–	LHB, NKX3-1
	Sex-linked Hereditary Disorder	3.36	7	CA2, USP9Y	CA5B, HSD17B10, NAA10, SMS, STS,
	Endocrine System Development and Function	3.22	2	–	HSD17B10, DHRS9

Bibliography

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

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

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

[4] Barker, D. J. P. Fetal programming of coronary heart disease. *Trends Endocrinol Metab,* 13 (2002), pp. 364–368.

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

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

[7] Edwards, A., Megens, A., Peek, M., and Wallace, E. M. Sexual origins of placental dysfunction. *Lancet,* 355(9199) (2000), pp. 203–4.

[8] Ingemarsson, I. Gender aspects of preterm birth. *British Journal of Obstetrics and Gynaecology,* 110 (2003), pp. 34–38.

[9] Lao, T. T., Sahota, D. S., Suen, S. S. H., and Law, L. W. The impact of fetal gender on preterm birth in a southern Chinese population. *J Matern Fetal Neonatal Med,* 24 (2011), pp. 1440–1443.

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

[11] Kim, J., Zhao, K., Jiang, P., Lu, Z., Wang, J., Murray, J. C., and Xing, Y. Transcriptome landscape of the human placenta. *BMC Genomics,* 13 (2012), p. 115.

[12] Schroeder, D. I., Blair, J. D., Lott, P., Yu, H. O. K., Hong, D., Crary F. Ashwood, P., Walker, C., Korf, I., and Robinson, W. P. The human placenta methylome. In: *Proceedings of the National Academy of Sciences,* vol. 110. 2013, pp. 6037–6042.

[13] Sood, R., Zehnder, J. L., Druzin, M. L., and Brown, P. O. Gene expression patterns in human placenta. In: *Proceedings of the National Academy of Sciences,* vol. 103. 2006, pp. 5243–5244.

[14] Sitras, V., Paulssen, R., Leirvik, J., Vårtun, A., and Acharya, G. Placental gene expression profile in intrauterine growth restriction due to placental insufficiency. *Reprod Sci,* 16 (2009), pp. 701–711.

[15] Sitras, V., Paulssen, R. H., Grønaas H. Leirvik, J., Hanssen, T. A., Vårtun, A., and Acharya, G. Differential placental gene expression in severe preeclampsia. *Placenta,* 30 (2009), pp. 424–433.

[16] Chim, S. S. C., Lee, W. S., Ting, Y. H., Chan, O. K., Lee, S. W. Y., and Leung, T. Y. Systematic identification of spontaneous preterm birth-associated RNA transcripts in maternal plasma. *PLoS ONE,* 7 (2012), e34328.

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

[18] Tsai, S., Hardison, N. E., James, A. H., Motsinger-Reif, A. A., Bischoff, S. R., Thames, B. H., and Piedrahita, J. A. Transcriptional profiling of human placentas from pregnancies complicated by preeclampsia reveals disregulation of sialic acid acetylesterase and immune signalling pathways. *Placenta,* 32 (2011), pp. 175–182.

[19] Votavova, H., Dostalova Merkerova, M., Fejglova, K., Vasikova, A., Krejcik, Z., Pastorkova, A., Tabashidze, N., Topinka, J., Veleminsky, M., and Sram, R. J. Transcriptome alterations in maternal and fetal cells induced by tobacco smoke. *Placenta,* 32 (2011), pp. 763–770.

[20] Sitras, V., Fenton, C., Paulssen, R., Vårtun, A., and Acharya, G. Differences in Gene Expression between First and Third Trimester Human Placenta: A Microarray Study. *PLoS ONE,* 7 (2012), e33294.

[21] Votavova, H., Dostalova Merkerova, M., Krejcik, Z., Fejglova, K., Vasikova, A., Pastorkova, A., Tabashidze, N., Topinka, J., Balascak, I., and Sram, R. J. Deregulation of gene expression induced by environmental tobacco smoke exposure in pregnancy. *Nicotine Tob Res,* 14 (2012), pp. 1073–1082.

[22] Guo, L., Tsai, S. Q., Hardison, N. E., James, A. H., Motsinger-Reif, A. A., Thames, B., Stone, E. A., Deng, C., and Piedrahita, J. A. Differentially expressed microRNAs and affected biological pathways revealed by modulated modularity clustering (MMC) analysis of human preeclamptic and IUGR placentas. *Placenta,* 34 (2013), pp. 599–605.

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

[24] Buckberry, S., Bent, S. J., Bianco-Miotto, T., and Roberts, C. T. massiR: a method for predicting the sex of samples in gene expression microarray datasets. *Bioinformatics,* (2014).

[25] Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S. J., and Marra, M. A. Circos: an information aesthetic for comparative genomics. *Genome Res,* 19 (2009), pp. 1639–1645.

[26] Zhang, H., Meltzer, P., and Davis, S. RCircos: an R package for Circos 2D track plots. *BMC Bioinformatics,* 14 (2013), p. 244.

[27] Zhang, Y., Klein, K., Sugathan, A., Nassery, N., Dombkowski, A., Zanger, U. M., and Waxman, D. J. Transcriptional profiling of human liver identifies sex-biased genes associated with polygenic dyslipidemia and coronary artery disease. *PLoS ONE,* 6 (2011), e23506.

[28] Whitney, A. R., Diehn, M., Popper, S. J., Alizadeh, A. A., Boldrick, J. C., Relman, D. A., and Brown, P. O. Individuality and variation in gene expression patterns in human blood. In: *Proc Natl Acad Sci USA,* vol. 100. 2003, pp. 1896–1901.

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

[30] Kwon, A. T., Arenillas, D. J., Hunt, R. W., and Wasserman, W. W. oPOSSUM-3: Advanced Analysis of Regulatory Motif Over-Representation Across Genes or ChIP-Seq Datasets. *G3,* 2 (2012), pp. 987–1002.

[31] Mathelier, A., Zhao, X., Zhang, A. W., Parcy, F., Worsley-Hunt, R., Arenillas, D. J., Buchman, S., Chen, C. Y., Chou, A., and Ienasescu, H. JASPAR 2014: an extensively expanded and updated open-access database of transcription factor binding profiles. *Nucleic Acids Res,* 42 (2013), pp. D142–D147.

[32] Meeks, J. J. and Schaeffer, E. M. Genetic Regulation of Prostate Development. *Journal of Andrology,* 32 (2011), pp. 210–217.

[33] Barak, Y., Sadovsky, Y., and Shalom-Barak, T. PPAR Signaling in Placental Development and Function. *PPAR Res,* 142082 (2008).

[34] Laplante, M. and Sabatini, D. M. mTOR Signaling. *Cold Spring Harbor Perspectives in Biology,* 4 (2012), a011593–a011593.

[35] Andraweera, P., Dekker, G., Thompson, S., and Roberts, C. Single-nucleotide polymorphisms in the KDR gene in pregnancies complicated by gestational hypertensive disorders and small-for-gestational-age infants. *Reprod Sci,* 19 (2012), pp. 547–554.

[36] Lee, J. T. Gracefully ageing at 50, X-chromosome inactivation becomes a paradigm for RNA and chromatin control. *Nature Reviews Molecular cell Biology,* 12 (2011), pp. 815–826.

[37] Augui, S., Nora, E. P., and Heard, E. Regulation of X-chromosome inactivation by the X-inactivation centre. *Nat Rev Genet,* 12 (2011), pp. 429–442.

[38] Buckberry, S., Bianco-Miotto, T., and Roberts, C. T. Imprinted and X-linked non-coding RNAs as potential regulators of human placental function. *Epigenetics,* 9 (2014). Advanced access publication March 22, pp. 81–89.

[39] Preumont, A., Rzem, R., Vertommen, D., and Van Schaftingen, E. HDHD1, which is often deleted in X-linked ichthyosis, encodes a pseudouridine-5'-phosphatase. *Biochem J,* 431 (2010), pp. 237–244.

[40] Carrel, L. and Willard, H. F. X-inactivation profile reveals extensive variability in X-linked gene expression in females. *Nature,* 434(7031) (2005), pp. 400–404.

[41] López-Domńech G. Serrat, R., Mirra, S., D'Aniello, S., Somorjai, I., Abad, A., Vitureira, N., García-Arumí, E., Alonso, M. T., and Rodriguez-Prados, M. The Eutherian Armcx genes regulate mitochondrial trafficking in neurons and interact with Miro and Trak2. *Nat Commun,* 3 (2012), p. 814.

[42] Liina, N. Genomics and genetics of gonadotropin beta-subunit genes: Unique FSHB and duplicated LHB/CGB loci. *Mol Cell Endocrinol,* 329 (2010), pp. 4–16.

[43] Norris, W., Nevers, T., Sharma, S., and Kalkunte, S. Review: hCG, preeclampsia and regulatory T cells. *Placenta,* 32(Suppl 2) (2011), S182–5.

[44] Bansal, A. S., Bora, S. A., Saso, S., Smith, J. R., Johnson, M. R., and Thum, M.-Y. Mechanism of human chorionic gonadotrophin-mediated immunomodulation in pregnancy. *Expert Rev Clin Immunol,* 8 (2012), pp. 747–753.

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

[46] McCarthy, C., Cotter, F. E., McElwaine, S., Twomey, A., Mooney, E. E., Ryan, F., and Vaughan, J. Altered gene expression patterns in intrauterine growth restriction: potential role of hypoxia. *Am J Obstet Gynecol,* 196(1) (2007), pages.

[47] Henke, A. and Gromoll, J. New insights into the evolution of chorionic gonadotrophin. *Mol Cell Endocrinol,* 291 (2008), pp. 11–19.

[48] Menzies, B. R., Pask, A. J., and Renfree, M. B. Placental expression of pituitary hormones is an ancestral feature of therian mammals. *Evodevo,* 2 (2011), p. 16.

[49] Ryu, K. S., Gilchrist, R. L., Koo, Y. B., Ji, I., and Ji, T. H. Gene, interaction, signal generation, signal divergence and signal transduction of the LH/CG receptor. *International Journal of Gynecology & Obstetrics,* 60 (1998), S9–S20.

[50] Brouillet, S., Hoffmann, P., Chauvet, S., Salomon, A., Chamboredon, S., Sergent, F., Benharouga, M., Feige, J. J., and Alfaidy, N. Revisiting the role of hCG: new regulation of the angiogenic factor EG-VEGF and its receptors. *Cellular and Molecular Life Sciences,* 69 (2012), pp. 1537–1550.

[51] Kleinrouweler, C. E., van Uitert, M., Moerland, P. D., Ris-Stalpers, C. v. J. A. M., and Afink, G. B. Differentially expressed genes in the pre-eclamptic placenta: a systematic review and meta-analysis. *PLoS ONE,* 8 (2013), e68991.

[52] Vatten, L. J. and Skjaerven, R. Offspring sex and pregnancy outcome by length of gestation. *Early Hum Dev,* 76(1) (2004), pp. 47–54.

[53] Di Renzo, G., Rosati, A., Sarti, R., and Cruciani, L. Does fetal sex affect pregnancy outcome? *Gend Med,* 4 (2007), pp. 19–30.

[54] Wilson, C. L. and Miller, C. J. Simpleaffy: a BioConductor package for Affymetrix Quality Control and data analysis. *Bioinformatics,* 21 (2005), pp. 3683–3685.

[55] Rainer, J., Sanchez-Cabo, F., Stocker, G., Sturn, A., and Trajanoski, Z. CARMAweb: comprehensive R- and bioconductor-based web service for microarray data analysis. *Nucleic Acids Res,* 34 (2006), W498–503.

[56] Dunning, M. J., Smith, M. L., Ritchie, M. E., and Tavaré, S. beadarray: R classes and methods for Illumina bead-based data. *Bioinformatics,* 23 (2007), pp. 2183–2184.

[57] Johnson, W. E., Li, C., and Rabinovic, A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics,* 8 (2007), pp. 118–127.

[58] Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E., and Storey, J. D. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics,* 28 (2012), pp. 882–883.

[59] Kauffmann, A., Gentleman, R., and Huber, W. arrayQualityMetrics–a bioconductor package for quality assessment of microarray data. *Bioinformatics,* 25 (2009), pp. 415–416.

[60] Flicek, P., Ahmed, I., Amode, M. R., Barrell, D., Beal, K., Brent, S., Carvalho-Silva, D., Clapham, P., Coates, G., and Fairley, S. Ensembl 2013. *Nucleic Acids Res,* 41 (2012), pp. D48–D55.

[61] Harrow, J., Frankish, A., Gonzalez, J. M., Tapanari, E., Diekhans, M., Kokocinski, F., Aken, B. L., Barrell, D., and Zadissa A. nad Searle, S. GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res,* 22 (2012), pp. 1760–1774.

[62] Slater, G. S. C. and Birney, E. Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics,* 6 (2005), p. 31.

[63] Gray, K. A., Daugherty, L. C., Gordon, S. M., Seal, R. L., Wright, M. W., and Bruford, E. A. Genenames.org: the HGNC resources in 2013. *Nucleic Acids Res,* 41 (2012), pp. D545–D552.

[64] Benson, D. A., Karsch-Mizrachi, I., Clark, K., Lipman, D. J., Ostell, J., and Sayers, E. W. GenBank. *Nucleic Acids Res,* 40 (2011), pp. D48–D53.

[65] Ramasamy, A., Mondry, A., Holmes, C. C., and Altman, D. G. Key issues in conducting a meta-analysis of gene expression microarray datasets. *PLoS Med,* 5 (2008), e184.

[66] Cvitic, S., Longtine M. S. Hackl, H., Wagner, K., Nelson, M. D., and Desoye G. Hiden, U. The Human Placental Sexome Differs between Trophoblast Epithelium and Villous Vessel Endothelium. *PLoS ONE,* 8 (2013), e79233.

Statement of Authorship

Title of Paper	Why are males more at risk in the womb?
Publication Status	◉ Published, ○ Accepted for Publication, ○ Submitted for Publication, ○ Publication style
Publication Details	Buckberry, S. & Roberts, C. T. Why are males more at risk in the womb? Australasian Science 35, 9, 16-18 (2014). * Not peer reviewed

Author Contributions

By signing the Statement of Authorship, each author certifies that their stated contribution to the publication is accurate and that permission is granted for the publication to be included in the candidate's thesis.

Name of Principal Author (Candidate)	Sam Buckberry		
Contribution to the Paper	Sam Buckberry wrote the first draft of the manuscript, incorporated comments and wrote the final manuscript.		
Signature		Date	7/4/15

Name of Co-Author	Claire T Roberts		
Contribution to the Paper	Claire T Roberts commented extensively on the first draft, contributed valuable insight and co-authored the final manuscript.		
Signature		Date	7.4.15

Name of Co-Author			
Contribution to the Paper			
Signature		Date	

Name of Co-Author			
Contribution to the Paper			
Signature		Date	

Chapter 6

Buckberry, S. and Roberts, C. T. (2014) Why are males more at risk in the womb? *Australasian Science*, 35(9), 16-18.

Statement of Authorship

Title of Paper	Placental transcriptome co-expression analysis reveals conserved regulatory programs and points toward a preeclampsia gene cluster.
Publication Status	○ Published, ○ Accepted for Publication, ○ Submitted for Publication, ◉ Publication style
Publication Details	Buckberry, S., Bianco-Miotto, T., Bent, S. J., Dekker, G. A. & Roberts, C. T. Placental transcriptome co-expression analysis reveals conserved regulatory programs and points toward a preeclampsia gene cluster. (2015).

Author Contributions

By signing the Statement of Authorship, each author certifies that their stated contribution to the publication is accurate and that permission is granted for the publication to be included in the candidate's thesis.

Name of Principal Author (Candidate)	Sam Buckberry
Contribution to the Paper	Sam Buckberry conceived, designed and carried out the experiments in this study, performed all bioinformatics and statistical analyses, interpreted the results, authored the first draft of the manuscript, incorporated suggestions from co-authors and wrote the final manuscript.
Signature	Date: 7/4/15

Name of Co-Author	Tina Bianco-Miotto
Contribution to the Paper	Tina Bianco-Miotto was involved in conceiving the study, supervised experimental design, contributed valuable insight toward interpreting the results and provided detailed comments for all manuscript drafts.
Signature	Date: 1/4/15

Name of Co-Author	Stephen J Bent
Contribution to the Paper	Stephen J Bent was involved in conceiving the study, supervised experimental design and contributed valuable insight toward the analysis methodology and provided detailed comments for all manuscript drafts.
Signature	Date: 25/3/2015

Name of Co-Author	Gustaaf A Dekker
Contribution to the Paper	Gustaaf A Dekker was involved in sample procurement, clinical evaluation of study participants, study design and commented on the final manuscript.
Signature	Date: 1-4-15

Statement of Authorship

Title of Paper	Placental transcriptome co-expression analysis reveals conserved regulatory programs and points toward a preeclampsia gene cluster.
Publication Status	○ Published, ○ Accepted for Publication, ○ Submitted for Publication, ◉ Publication style
Publication Details	Buckberry, S., Bianco-Miotto, T., Bent, S. J., Dekker, G. A. & Roberts, C. T. Placental transcriptome co-expression analysis reveals conserved regulatory programs and points toward a preeclampsia gene cluster. (2015).

Author Contributions

By signing the Statement of Authorship, each author certifies that their stated contribution to the publication is accurate and that permission is granted for the publication to be included in the candidate's thesis.

Name of Principal Author (Candidate)	Sam Buckberry		
Contribution to the Paper	Sam Buckberry conceived, designed and carried out the experiments in this study, performed all bioinformatics and statistical analyses, interpreted the results, authored the first draft of the manuscript, incorporated suggestions from co-authors and wrote the final manuscript.		
Signature		Date	7/4/15

Name of Co-Author	Claire T Roberts		
Contribution to the Paper	Claire T Roberts conceived the study, supervised the experimental design, provided all samples and funding for the experiments, contributed valuable insight towards interpreting the results and provided detailed comments for all manuscript drafts.		
Signature		Date	7.4.15

Name of Co-Author			
Contribution to the Paper			
Signature		Date	

Name of Co-Author			
Contribution to the Paper			
Signature		Date	

Chapter 7

Placental Transcriptome Co-Expression Analysis Reveals Conserved Regulatory Programs and Points Toward a Preeclampsia Gene Cluster

Sam Buckberry, Tina Bianco-Miotto, Stephen J Bent, Gustaaf A Dekker and Claire T Roberts

Abstract

Mammalian development *in utero* is absolutely dependent on proper placental development, which is ultimately governed by the molecular instructions encoded in the placental genome. The regulation of the placental genome can be directly studied by exploring the underlying organisation of the placental transcriptome through a systematic analysis of gene-wise co-expression relationships. In this study, we performed a comprehensive analysis of human placental gene co-expression using RNA sequencing and the integration of multiple transcriptome datasets spanning human gestation. We identified modules of co-expressed genes that are highly preserved across gestation, and between human and mouse, revealing highly conserved molecular networks involved in placental development. Our analyses identified a cluster of genes implicated in preeclampsia that show highly correlated patterns of expression suggesting regulation by a common set

of factors. Furthermore, by summarising co-expressed gene modules, we demonstrate a novel way of screening for biomarkers of placental gene expression and development. Together, our findings provide a new framework for studying gene expression in the placenta and reveal previously unappreciated aspects of the placental transcriptional landscape.

7.1 Introduction

The placenta is the first human organ to start developing once the embryo implants into to the mother's uterus shortly after conception. At implantation, placental trophoblast cells begin to invade into the lining of the uterus, where they colonise and transform the mother's spiral arterioles and additionally extra-embryonic tissue establishes its own placental network of blood vessels. Together these processes facilitate the exchange of all nutrients, gases and waste throughout pregnancy. However, despite the placenta's indispensable role in intrauterine mammalian development, the placenta remains the least understood human tissue [1].

Normal placental function is dependent on appropriate growth and development of its structural components, which are underpinned by the fine-tuned regulation of gene expression. Consequently, alterations to placental gene regulation are thought to be a major contributor to pregnancy pathologies. Several studies aimed at elucidating the molecular basis of placental development have utilised high-throughput gene expression technologies, such as RNA sequencing (RNA-Seq) and microarrays, and show that the placenta undergoes global shifts in gene expression between the first and third trimesters [2]. They also show that placentas from pre-eclamptic pregnancies feature a distinct expression signature [3], and that some of these expression differences arise approximately six months before the condition manifests [4]. Recently, two placental transcriptome studies employing RNA-Seq have described the breadth of gene expression in the human placenta and show that the placenta exhibits unique patterns of exon splicing and greater than four-fold enrichment for > 800 genes compared to other human tissues [5, 6].

Despite these efforts, progress towards developing accurate markers of healthy and pathological pregnancy has been slow. Likewise, the spectrum of environmental factors influencing placental development remain unclear. This slow progress can be attributed, in part, to the inherent difficulties in obtaining placental tissue from multiple time points prior to birth, a limited understanding of how placental

development is influenced by environmental factors and maternal physiology, and the lack of suitable animal models for studying human pregnancy pathology. This paucity of knowledge was recently recognised by the National Institute of Child Health and Human Development (NICHD) and the National Institutes of Health (NIH) in the United States, which subsequently prompted the inception of the *Human Placenta Project* [1, 7].

A common feature in previous studies on placental gene regulation is that expression data are typically summarised at the gene level for between-group comparisons, widely known as *differential expression*. With differential expression, the greatest significance is attributed to individual genes where the differences between groups reach an appropriate significance threshold. Although differential expression analyses have unquestionable utility, the inherent natural organisation of the transcriptome remains largely unexplored. Conversely, more holistic methods that consider the gene-wise relationships in gene expression data have cast new light on previously unappreciated patterns of transcriptional organisation with regards to lipid metabolism [8], cancer [9], human brain development and neuropathology [10–12], and embryonic development [13]. These co-expression 'systems' approaches identify groups of genes where expression levels are highly correlated across samples. By leveraging the inter-individual expression variability between biological samples, such strategies enable the identification of higher-order relationships among genes. Further *post hoc* characterisation of these relationships then has the ability to provide insight into the biological processes arising from the underlying transcriptional program.

To gain a new perspective on placental genome regulation, we performed a comprehensive analysis of placental gene co-expression. Our results reveal distinct groups of genes that show highly correlated patterns of gene expression and are associated with specific biological processes and placental pathologies. By drastically reducing the dimensionality of gene expression data through summarising highly correlated genes, we illustrate a potential framework for screening biomarkers of placental gene expression and development.

7.2 Results

7.2.1 Constructing a Weighted Human Placental Co-Expression Gene Network

To explore patterns of gene co-expression in the healthy human term placenta, we performed single-strand 100-base paired-end total RNA-Seq for 16 samples at an average depth of 38 million uniquely mapped reads per sample. By summarising the RNA-Seq reads by counting the number of overlaps with known genes, we detected 15,861 genes above the threshold of > 1 read count per million, which we show is an accurate threshold of detection based on quantification of spiked synthetic RNAs (Figure S7.1 on page 164).

To integrate gene-level expression profiles into a higher-order systems level framework, normalised gene expression values were used to perform a weighted gene co-expression network analysis (WGCNA) [14]. To construct the gene-wise network, we first calculated Pearson's correlation matrix, then raised this matrix to a power to weight strong correlations at the expense of weaker ones, thus resulting in a weighted network (see Methods). To identify groups of genes with highly correlated patterns of expression, these data were then transformed into a topological overlap matrix of 'connection-strengths' [14]. This was then used as input for unsupervised hierarchical clustering, where we employed a dynamic tree-cutting algorithm [15] to group tree branches into 13 distinct clusters of highly connected genes, which we refer to as *modules* (Figure 7.1).

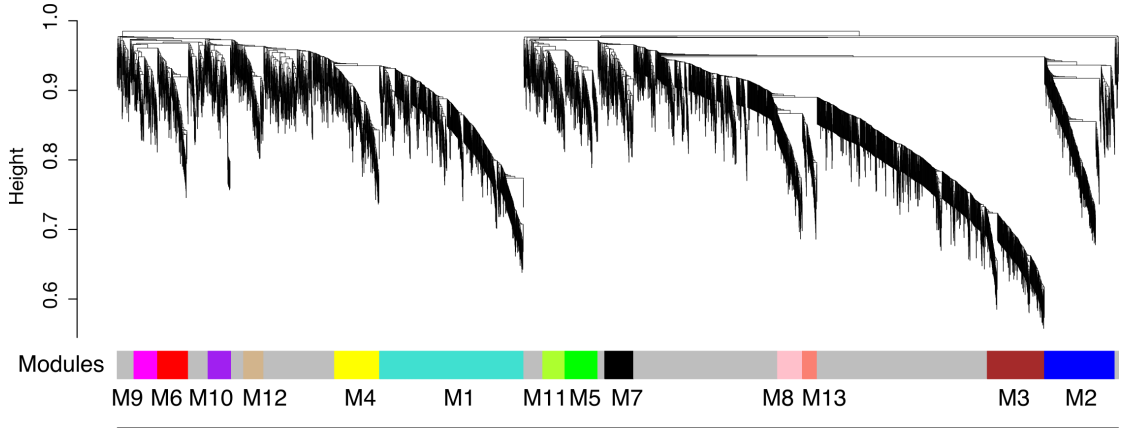


FIGURE 7.1: Weighted gene co-expression network analysis of the human placenta reveals distinct clusters of co-expressed genes. Average linkage hierarchical clustering dendrogram of genes based on gene expression topological overlap. Modules of co-expressed genes were assigned colours and identifiers M1–M13, which are represented in the horizontal bar below the dendrogram.

Each module was then summarised by calculating the module *eigengene* for each sample, which is the first principal component of gene expression values for the module. Therefore, the eigengene represents a weighted average of gene expression. For each gene, we then define its membership in each module as the absolute correlation between the gene's expression and the module's eigengene, and represent this correlation as *kME* [14]. Genes are assigned to modules if they have an absolute $kME > 0.7$. Note that by quantifying membership through correlation, module membership for each gene is no longer binary and allows genes to be members of more than one module (Figure S7.2), thus connecting modules in a network.

The proportion of gene expression variation explained by each eigengene ranged between 39.1% (M10) and 79.6% (M3) (Table 7.1). This demonstrates that even for large modules such as M3 (844 genes), a significant proportion of variance can be captured by a single representative value. For each gene module, the top hub genes ($kME > 0.9$) are reported in Table 7.1, and genes with a $kME > 0.7$ for each module are listed in Table S1 (See Supplementary Data File). The plots in Figure 7.2 demonstrate the high correlation of the top ten most connected genes for modules M2 and M3, and how gene variance is accurately reflected by the module eigengene. A gene ontology analysis showed co-expression modules are enriched for distinct biological processes and molecular functions (Table S2 in Supplementary Data File).

TABLE 7.1: Co-expression module characteristics.

Module	No. of genes	Variance explained by eigengene	Top ten hub genes ($kME > 0.9$)
M1	740	44.6%	*ZNF845, ZNF808, GPR160, GIN1, ATP5J, ZNF567, ANAPC10, C8orf59, MRPS36, RBM7*
M2	262	48.9%	*EPHA10, ARIH2OS, TUBD1, FLJ42102, KIAA0101, RPL13AP20, CD96, PDE6A, GGT8P, SLC35F1*
M3	844	79.6%	*NOTCH3, PLXND1, PALM, CSPG4, ARHGEF17, DCHS1, MARK4, KIRREL, LTBP4, AXL*
M4	566	51.5%	*HMMR, CASC5, DEPDC1, CDK1, KIF15, CCNA2, AIM1, TTK, ESCO2, EXO1*
M5	116	45.5%	*ATP2A1, C11orf35, P2RY2, CCDC33, ASIC3, KIFC2, IL17REL, CLIC3, MTVR2, RBBP8NL*
M6	88	51.4%	*HN1, ASAP3, SLC12A8, ASPHD2, B3GNT7, IL17RE, PRG2, NOG, IL2RB, PIPOX*
M7	112	41.5%	*SNORD114-29, CDH11, FAM198B, SNORD114-7, SNORD114-10, FKBP7, SNORD114-14, C4orf32, SNORD114-26, SNORD113-2*
M8	390	68.1%	*SBF1, ULK1, STRA6, DOT1L, BCAR1, TMEM184A, B3GNT8, SLC25A22, C19orf71, INTS1*
M9	79	44.5%	*SELL, S100A12, LRRK2, CYTIP, MNDA, ACSL1, FPR2, TGFA, LOC100505806, TMEM71*
M10	110	39.1%	*MTHFS, TTTY15, RPS4Y1, TXLNG2P, TTTY10, KDM5D, UTY, EIF1AY, ZFY, PRKY*
M11	112	43.1%	*PGAP3, GPR137, PRR5, ARTN, C10orf10, C7orf43, ALDH4A1, EFS, RELL2, ADIRF*
M12	81	51.7%	*PVRL4, ARHGEF4, NDRG1, INHBA, SYDE1, INHA, MIR210HG, C8orf58, SIGLEC6, PDZD7*
M13	414	71.0%	*FAM195B, FBXL15, BRAT1, AKAP2, SCAND1, EME2, CCDC85B, C19orf60, PGLS, TSR3*

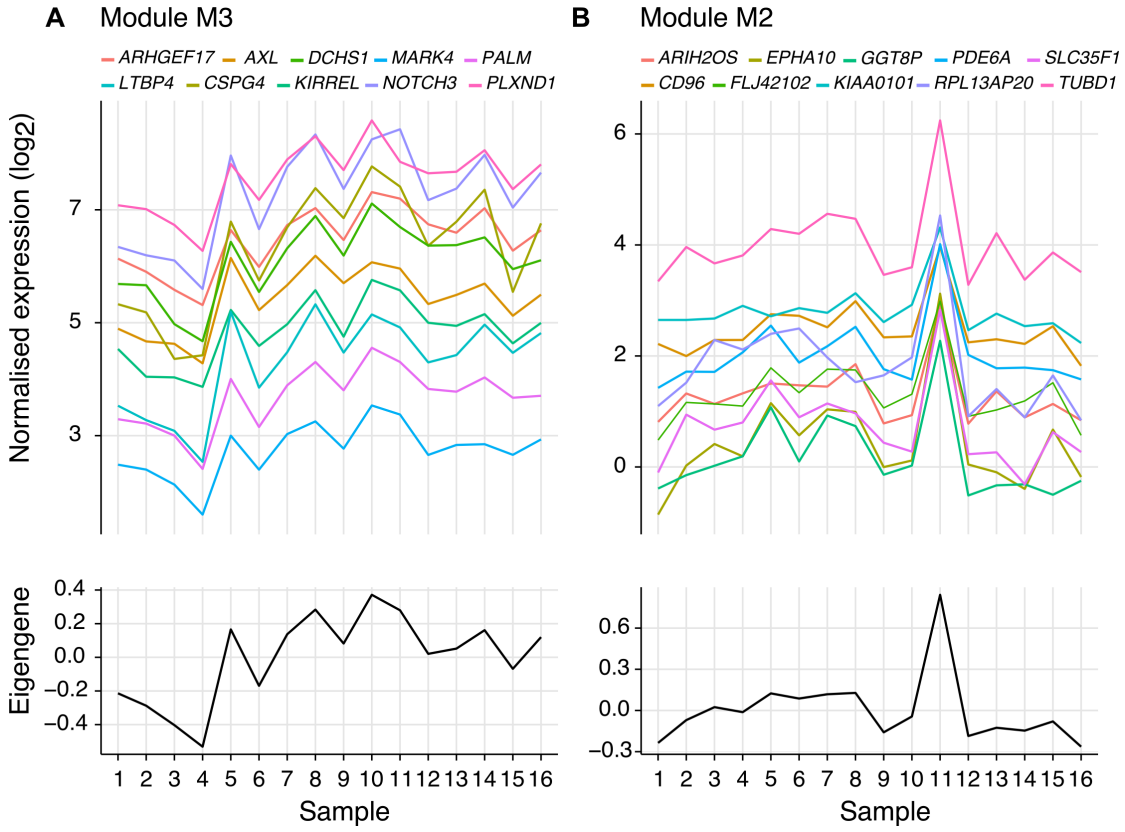


FIGURE 7.2: Gene–eigengene correlations identify module hub genes that are consistently co-expressed in the human placenta. The upper line plots show the top ten genes with the highest module membership (kME) for modules M3 (A) and M2 (B). Each continuous line represents a gene, with different genes showing a similar variability of expression across samples on the x-axis.

7.2.2 Co-Expression Modules are Reproducible

To evaluate the reproducibility of these gene modules in the third trimester placenta, we obtained raw RNA-Seq data from a previously published study on the human placental transcriptome [5] and tested whether the density and connectivity patterns of gene modules we defined in our reference dataset were preserved. To quantify reproducibility, we applied a preservation permutation test [16] to summarise evidence that the network topology is preserved in independent test sets and report the Z_{summary} statistic to summarise module preservation. In this independent third trimester dataset, 4/14 modules show highly significant preservation scores ($Z_{\text{summary}} > 10$), and 8/14 were at least moderately preserved ($Z_{\text{summary}} > 5$) despite a lower depth of sequencing [5] (Figure 7.3).

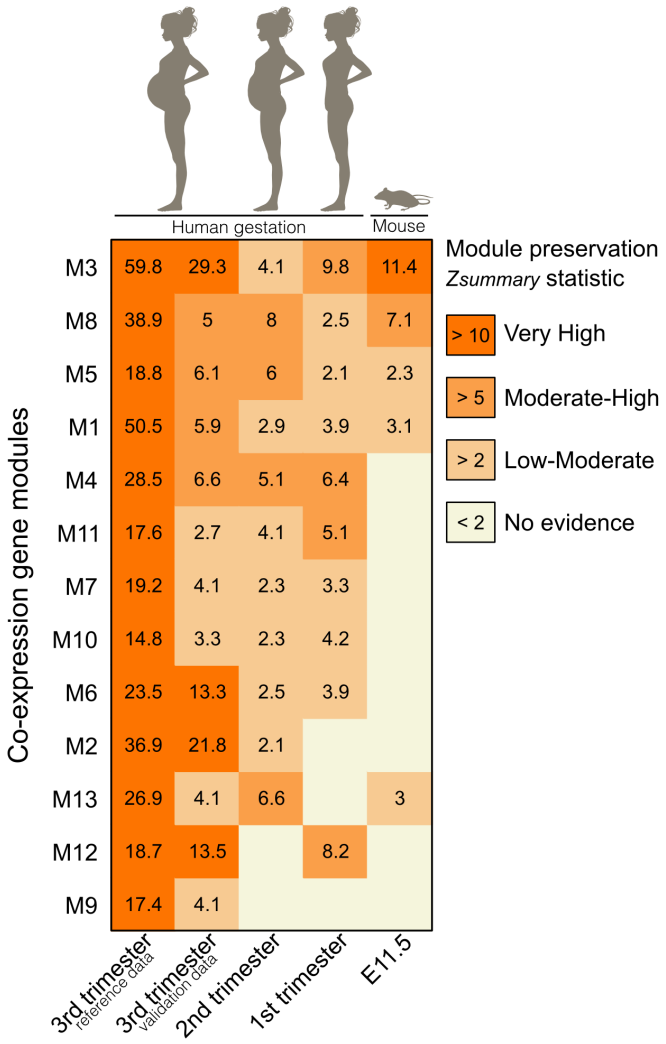


FIGURE 7.3: Preservation heat map of co-expression gene modules in independent datasets shows level of module preservation in the human placenta across human gestation and in mid gestation mouse placenta (E11.5). Colours represent four classes of co-expression preservation as represented by Z-score summary of preservation statistics. $Z_{\text{summary}} > 10$ indicates high level of evidence for module preservation, $Z_{\text{summary}}5-10$ indicates moderate–high preservation, $Z_{\text{summary}}2-5$ indicates low-moderate preservation, and $Z_{\text{summary}} < 2$ indicates no evidence for preservation. Numbers within cells are the Z-score summary statistic. Third trimester reference is data column (far right) represents results from running permutation tests using the data collected in this study. Other columns show the permutation test preservation statistics for previously published placenta transcriptome data [2, 5, 17, 18].

7.2.3 Key Co-Expression Modules are Preserved Across Human Gestation and Conserved in the Mouse

Given that the human placenta undergoes significant growth and remodelling throughout the nine months of gestation [19], we reasoned that if particular co-expression modules were involved in core placental functions, then these modules would be reproducible using gene expression data from earlier gestational time points. To test this hypothesis, we obtained microarray gene expression data from placental tissue collected during the first [2] and second trimesters [17]. Although these datasets contain expression data for substantially fewer genes after filtering and annotation (57.6% and 63.9% of detectable genes in the RNA-Seq dataset, respectively), the module preservation statistics indicate that a majority of modules are nevertheless preserved across gestation at a low to moderate level of significance (Figure 7.3). In particular, M4 shows moderate preservation ($Z_{\text{summary}} > 5$) across all gestational time points, indicating a conserved pattern of gene regulation throughout human gestation. In contrast, the M2 module is highly preserved in the third trimester datasets ($Z_{\text{summary}} > 10$) with little to no evidence of preservation during the first or second trimesters, suggesting M2 genes constitute a molecular program more specific to third trimester placental functions.

As the mouse is the most widely utilised model for studying placental development, we next asked whether the co-expression gene modules were conserved between human and mouse. To achieve this, we obtained raw RNA-Seq data for 23 mid-gestation (E11.5) mouse placenta samples [18] and showed that 5/14 had some degree of evidence for module preservation ($Z_{\text{summary}} > 2$), with M3 showing a highly significant preservation score ($Z_{\text{summary}} > 10$) (Figure 7.3). To further validate the conservation of co-expression between human and mouse, we assembled an independent and unsupervised *de novo* mouse co-expression network using the same methods as our human dataset. By counting the overlapping genes for each module and performing Fisher exact tests, we show that five human modules have at least one mouse counterpart (Bonferonni corrected $p < 0.05$, Figure 7.4). As predicted from the human–mouse Z_{summary} statistics, M3 showed the highest degree of overlap with a mouse module (Bonferroni $p = 2.78 \times 10^{-20}$) and a highly significant *kME* correlation (Pearson's $r = 0.4$, $p = 2.6 \times 10^{-102}$).

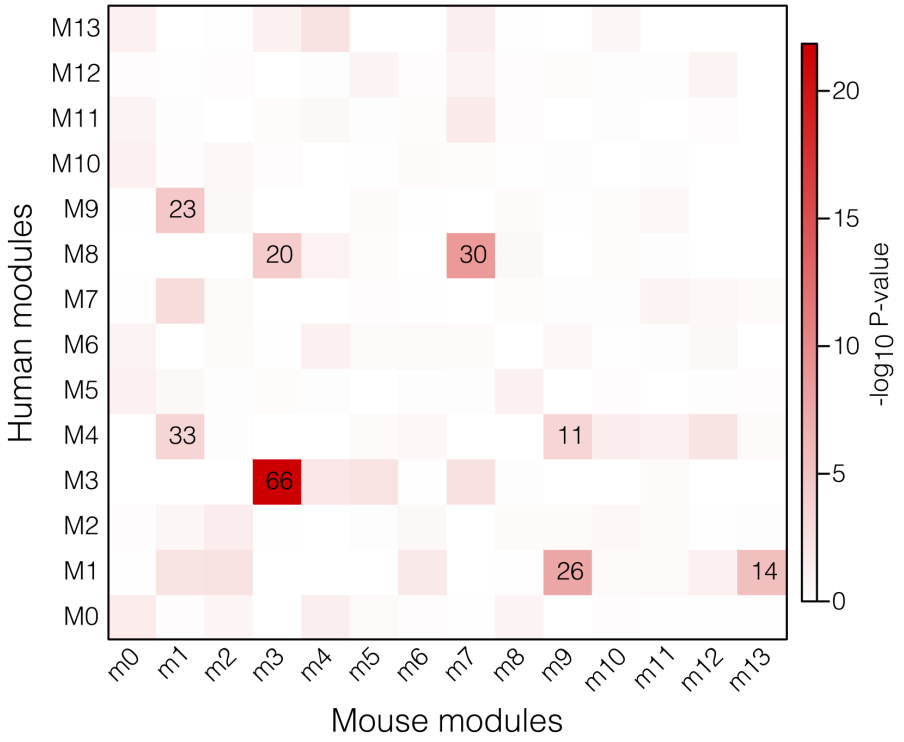


FIGURE 7.4: Overlap between weighted gene co-expression network modules for human and mouse placenta. Heat map colours represent Fisher exact test $-\log_{10}$ p-values. Numbers within cells represent the number of overlapping genes with Bonferroni $p < 0.05$ and shows five human co-expression modules (M1, M3, M4, M8 and M9) have a significant corresponding module in the mouse.

7.2.4 Preserved Modules Feature a Core Set of Transcription Factor Motifs

As M3 genes appear to constitute a highly conserved transcriptional network, we tested for the enrichment of known transcription factor (TF)-binding motifs in the sequence flanking M3 transcription start sites. Of the TFs predicted to target M3 genes (Figure 7.5A), *ZNF423* and *EBF1* were both detectable in the placenta and members of the M3 module ($kME = 0.85$ and $kME = 0.78$, respectively), and highly correlated with the M3 eigengene (Figure 7.5C). ZNF423 has previously been reported to interact with EBF1 [20–23]. Here we show a majority of M3 genes with ZNF423-binding motifs also feature EBF1 motifs (Figure 7.5B), and the density of these motifs is greatest immediately upstream of M3 transcription start sites (Figure 7.5D). These multiple lines of evidence suggest ZNF423 and EBF1 are key regulators of M3 gene transcription. When we performed the same enrichment tests for all other modules, ZNF423 and EBF1 were predicted to target a high proportion of genes within other co-expression modules (Table S7.1 on page 168).

Further inquiry revealed that the most highly preserved modules across human gestation, and between human and mouse (M1, M3-5, M8), feature a core set of TF-binding motifs (Figure S7.3 on page 166), suggesting these co-expressed genes share common regulatory factors and have a high degree of upstream sequence similarity.

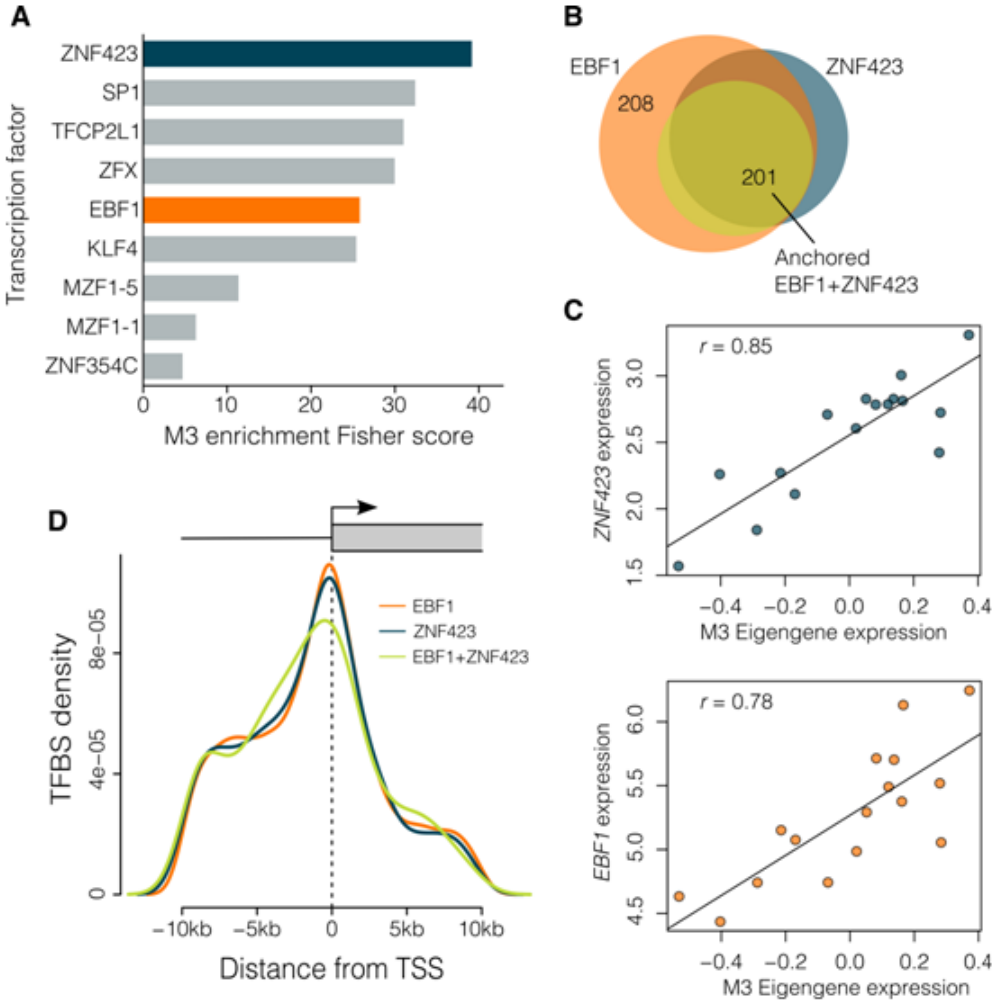


FIGURE 7.5: EBF1 and ZNF423 are potential upstream regulators of M3 gene expression. (A) Enrichment test for TF-binding motifs in the 10kb up- and down-stream of transcription start sites identify two TFs, ZNF423 (blue) and EBF1 (orange), that are members of the M3 module. (B) EBF1 and ZNF423 are predicted to target many of the same M3 genes. Circles in the Venn diagram represent the number of genes targeted by TFs and their overlap – EBF1 (orange), ZNF423 (blue), and when both have motifs directly adjacent to each other (anchored analysis, yellow). (C) ZNF423 and EBF1 expression is highly correlated with M3 eigengene expression. Points represent individual samples. (D) TF-binding motif density is greatest immediately upstream of M3 transcription start sites. Coloured lines represent the density TF motifs for EBF1 (orange), ZNF423 (blue) and the combination of both (green).

7.2.5 Modules of Co-Expressed Genes are Implicated in Pregnancy Complications

The origins of several pregnancy pathologies, such as preterm birth (PTB) and preeclampsia (PE) are largely attributed to abnormal placental development [24–26]. If co-expression modules constitute gene networks involved in placental development, we reasoned that if a particular module underpinned key placental processes, it may be enriched for genes implicated in pregnancy complications. To address this question, we obtained a curated gene list from the PTB gene database [27], and a set of meta-analysis-validated differentially expressed genes in PE [3], and tested our co-expression gene modules for enrichment of genes implicated in these pathologies (Figure 7.6). M9 was statistically enriched for genes associated with PTB (OR = 3.4, FDR = 0.03), but more strikingly, modules M11 and M12 showed significant enrichment for PE-related genes (M11 OR = 16.6, FDR = 2.1×10^{-3}; M12 OR = 101.3, FDR = 1.2×10^{-16}). Notably, three M12 intramodular hub genes (*PVRL4*, *INHBA* and *INHA*) have consistently been shown to be up-regulated in PE [3]. This provided the first line of evidence that M12 gene co-expression genes may be altered in PE.

To further validate the finding that M12 was enriched for genes differentially expressed in PE, we obtained additional independent microarray expression data from a recent study on early-onset PE ($n = 16$) [28] and tested for differences in M12 gene expression. First, a rotation gene set test [29] showed that M12 genes are significantly up-regulated in the PE placenta ($p = 0.021$), providing a second line of evidence for the involvement of M12 in preeclampsia (Figure 7.7A). Following this, we calculated the first principal component for M12 genes in this dataset to obtain an eigengene measure, and showed that M12 eigengene expression is significantly different (t-test, $p = 1.7 \times 10^{-4}$) between PE and control (Figure 7.7B). This demonstrates the robust nature of the eigengene for testing for differences in gene regulation between control and PE pregnancies. Furthermore, the transcriptional regulator INSM1 that is functionally related to placentation [30] was predicted to target a majority of M12 genes (Table S7.3 on page 166), in particular those genes dysregrulated in PE. Moreover, INSM1 is implicated in regulating *LEP* [31], which showed the highest expression difference between PE and controls (Figure 7.7A). Together, these results strongly implicate M12 co-expressed genes in PE and suggest that the mechanisms regulating M12 co-expression may be altered in PE. Thus, we demonstrate a new framework for investigating placental genome regulation in this pregnancy pathology.

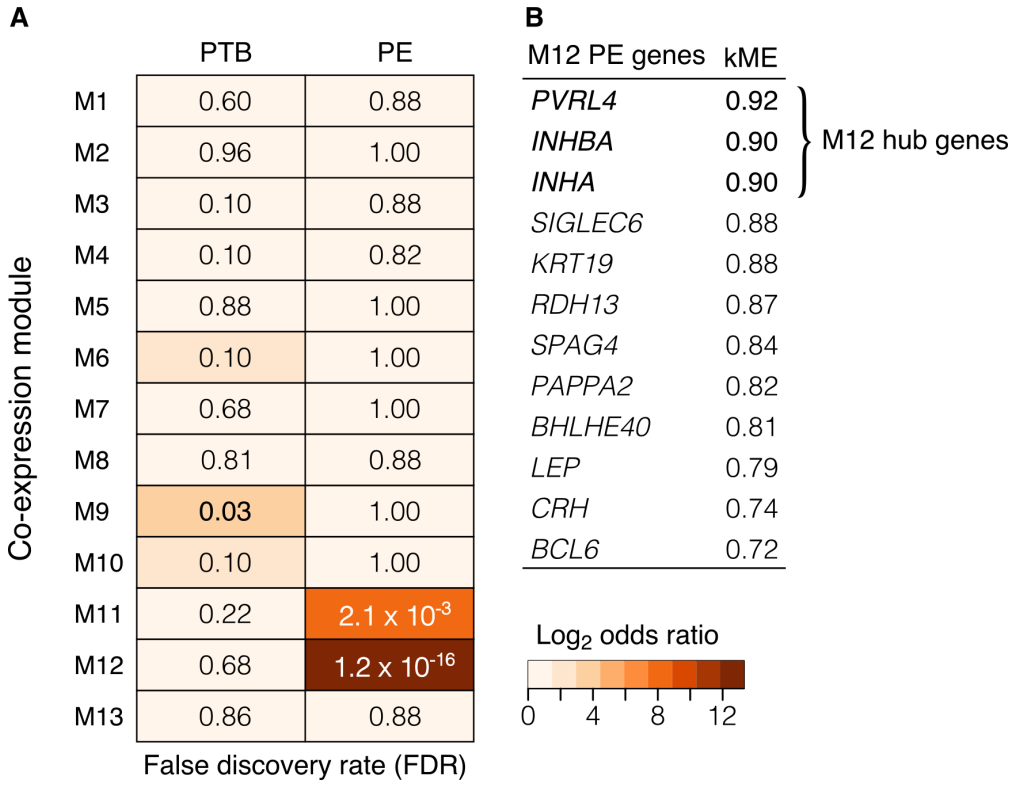


FIGURE 7.6: M12 is enriched for genes that show a meta-signature for preeclampsia. (A) Heat map table shows the statistical enrichment (FDR) of module genes in preterm birth (PTB) and preeclampsia (PE), and cell colours represent log2 odds ratio. (B) M12 genes implicated in PE and their module membership (*kME*). M12 intramodular hubs are in **bold**.

7.2.6 Eigengenes can be Used to Screen for Non-Invasive Markers of Placental Gene Expression

One objective of the recently established *Human Placenta Project* is to improve current methods and develop new technologies for real-time assessment of placental development across pregnancy [1, 7]. Up to this point, we have demonstrated that the organisation of the placental transcriptome can be summarised through identifying modules of functionally related co-expressed genes, and module gene expression is accurately reflected by the eigengene. Therefore, given the dimension-reducing ability and explanatory power of eigengenes, we postulated that they could be used to screen for markers of placental gene expression.

In this study, we had access to placenta-matched maternal blood samples taken at 15 weeks of gestation. To explore the concept of developing markers of placental gene expression, we assayed a panel of biomarkers in maternal blood and screened for eigengene correlates. After reducing the gene modules to those highly preserved ($Z_{\text{summary}} < 10$) in the third trimester reference and validation datasets, and at

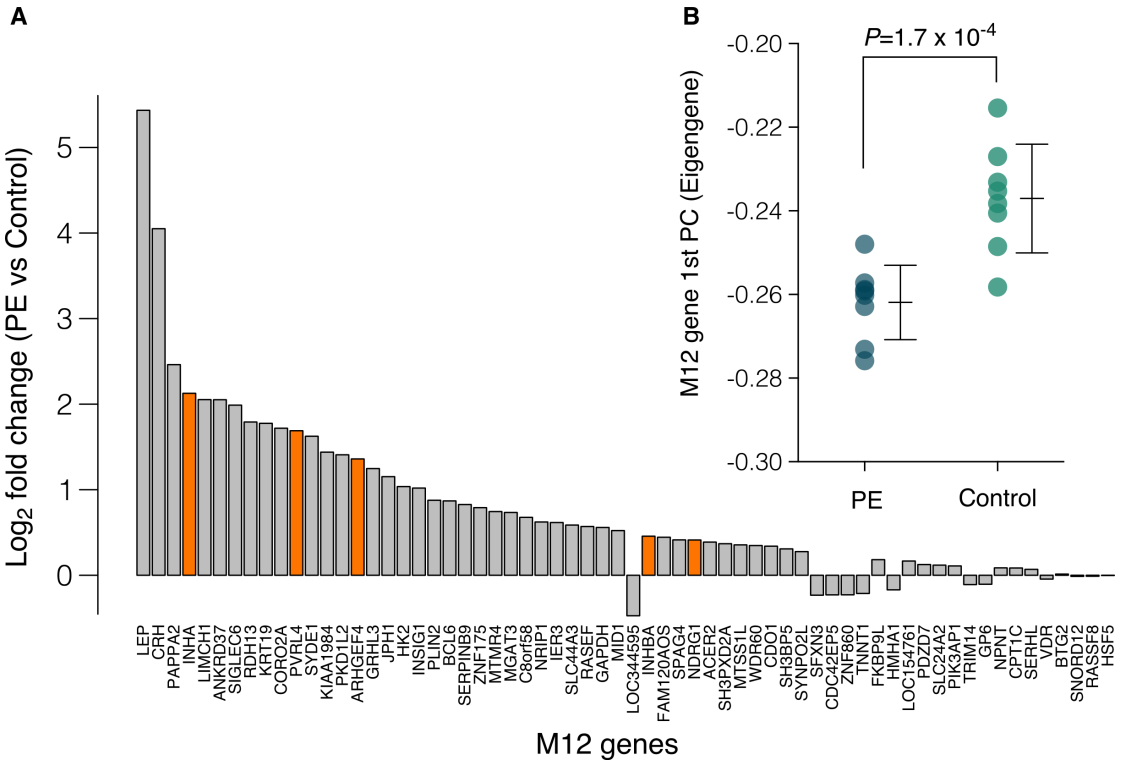


FIGURE 7.7: M12 genes are significantly up-regulated in preeclampsia placentas. (A) Bar plot showing the $\log_2$ fold-change between preeclampsia and control placentas. Orange bars represent M12 hub genes. (B) The M12 eigengene (first principal component) is significantly different between preeclampsia and control placentas.

least low preservation in the first trimester placenta ($Z_{\text{summary}} > 2$), this screen identified maternal blood Caspase-3 as a potential marker (FDR = 0.03, Pearson's $r = 0.76$) of M12 eigengene expression (Figure 7.8A). Closer inspection revealed that the M3 hub gene *NDGR1* is the key driver of this correlation (Figure 7.8B–C). These results demonstrate that by clustering highly correlated genes, it is possible to drastically reduce the dimensionality of gene expression data, which subsequently increases the statistical power for screening biomarkers of placental gene expression.

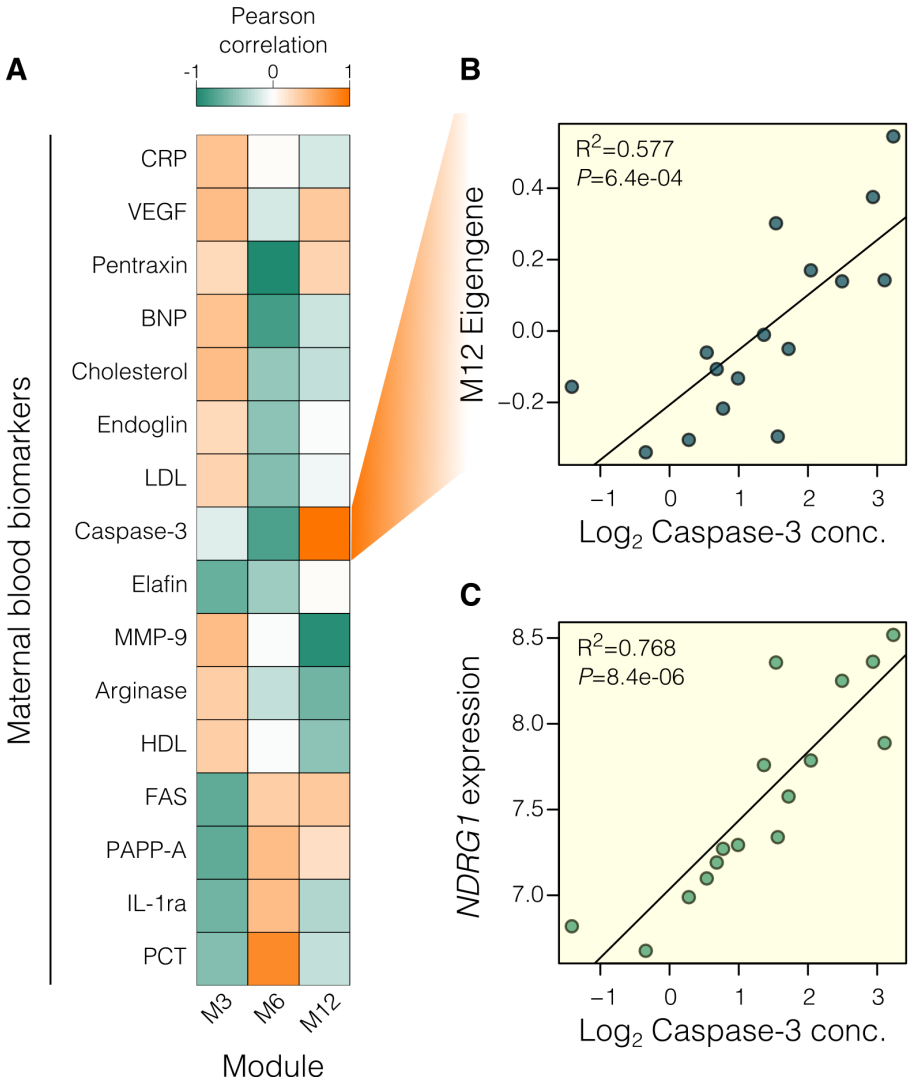


FIGURE 7.8: Module eigengenes can be used to screen for non-invasive biomarkers of placental gene expression. (A) Heatmap showing Pearson correlations between maternal blood biomarkers at 15 weeks of gestation and term placenta co-expression module eigengenes. (B) Caspase-3.13a in maternal blood is a strong predictor of both M12 eigengene expression (B), and the M12 hub gene *NDRG1* (C).

7.3 Discussion

By conducting the first comprehensive co-expression network analysis of the human placental transcriptome, we reveal previously unappreciated aspects of transcriptional organisation at the fetal-maternal interface. This analysis entailed the integration of multiple gene expression datasets and curated databases, which enabled the testing of specific hypotheses regarding placental genome regulation.

Our results demonstrate that a large proportion of the placental transcriptome is organised into distinct modules of co-expressed genes, some of which are preserved across gestation, and conserved between human and mouse. The reproducibility of these networks, constructed from independent datasets and different platforms (RNA-Seq and microarrays) suggest a fundamental modular organisation of the placental transcriptome. Moreover, our cross-species analysis demonstrates that certain aspects of human placental transcriptional organisation are highly preserved in the mouse, indicating the evolutionary conservation of molecular processes which drive mammalian placental development.

When comparing the *de novo* human and mouse networks, five genes were identified as M3/m3 intramodular hub genes ($kME > 0.9$) in both species (*ARHGEF17*, *DOCK6*, *MLK1*, *OSBPL7*, and *PRR12*), demonstrating a high degree of interspecies module reproducibility. These hub genes are centrally located within the M3 module and may be critical components of the network. Of particular interest, *DOCK6* mutations in humans are associated with extreme placental angiopathy and a severely abnormal placental phenotype [32], while *DOCK6* expression is reported to be down-regulated in placentas from growth-restricted fetuses [33]. Similarly, *OSBPL7*, an oxysterol-binding protein, is also reported to be differentially expressed in placentas from preeclamptic pregnancies [34]. Together, these results indicate that M3 co-expressed genes have significant involvement in placental development.

Investigation of the TFs that potentially regulate co-expression revealed that the most preserved modules are predicted to be regulated by a core set of transcription factors, including the M3 genes *EBF1* and *ZNF423*, which potentially target a high proportion of genes in the most highly preserved modules. Although the effects of ZNF423 and EBF1 on placental gene regulation remain largely unexplored, ZNF423 appears to be a multi-functional transcription factor associated with B cell regulation, retinoic acid signalling, notch signalling, DNA damage response pathways, adipogenesis and cancer [20]. Furthermore, homozygous mutation in the homologous gene in mice (*Zfp423*) results in smaller ataxic pups who die shortly after birth [35]. This indicates a critical role for ZNF423 in development. EBF1 can act as both a transcriptional activator and repressor and has known roles in tumour suppression [36]. When EBF1 binds DNA directly as a dimer, it can activate transcription via p300-CBP co-activation [36]. In other contexts, the same DNA binding dimer in conjunction with ZNF423 can recruit the nucleosome remodelling and deacetylase (NuRD) complex, triggering EBF1-mediated transcriptional repression [36]. The observation that EBF1 and ZNF423 are co-expressed in the placenta and members of the M3 module, and their widespread

targeting potential across modules of co-expressed genes indicates that these TFs are candidate key regulators of transcription in the placenta.

The identification of M12 being enriched for genes implicated in PE highlights a new method for identifying genes that may respond to the pathology, or may indeed underlie its aetiology. This guilt-by-association approach, clustered genes implicated in PE (M12) in a completely unsupervised manner, suggesting expression differences in these genes are driven by a set of common factors. The observation that several M12 hub genes are up-regulated in PE, and show highly correlated patterns of expression, implies that expression of other genes within this module is likely driven by the same underlying factors. Viewed in this way, further investigation of the involvement of M12 genes and their upstream regulators in placental development may prove to be a valuable way of generating new hypotheses regarding the placental origins of PE.

The intramodular M12 hub gene *PVRL4*, which is up-regulated in PE [3], is expressed more highly in the placenta compared to other human tissues [37]. *PVRL4* is a potent mediator of epithelial cell colony formation [38] and is also highly expressed in ovarian cancer tissue [39]. Furthermore, cleaved PVRL4 is elevated in the serum of patients with ovarian cancer and is correlated with *PVRL4* expression [39], suggesting that maternal serum PVRL4 may hold potential as a biomarker of PE.

Screening of maternal blood biomarkers for eigengene correlates revealed Caspase-3 as a potential surrogate marker for M12 gene expression, and the M12 intramodular hub gene *NDRG1* was most significantly associated with maternal Caspase-3. Intriguingly, NDRG1 is necessary for p53-mediated caspase activation and apoptosis [40]. This apoptosis pathway is implicated in trophoblast apoptosis and placental pathologies [41], highlighting a new avenue of inquiry regarding the potential relationship between maternal Caspase-3 and M12 genes. However, given the temporal separation between maternal blood sampling and placental delivery, the conclusions that can be drawn from their relationship are limited and require further investigation. Nevertheless, we view this screening analysis as a proof-of-concept, which demonstrates the power of using eigengenes as a screening tool in the search for non-invasive markers of placental development and function.

Several new questions arise from this comprehensive co-expression network analysis. Firstly, are patterns of co-expression altered in placental pathologies? Our analysis of independent expression datasets from PE placentas provide compelling preliminary evidence that M12 genes are up-regulated in PE, which warrants further investigation into the regulators of M12 genes. Secondly, what genetic and environmental factors influence co-expression? A comprehensive assessment of

genotypes and environmental factors such as maternal diet has the potential to reveal drivers of placental expression variation. Thirdly, does silencing of hub genes shift module co-expression and influence placental cell phenotype and behaviour? Functional studies aimed toward elucidating the biological function of co-expression modules may yield new insights into how placental development is regulated.

In summary, we show that a weighted gene co-expression network analysis can provide novel insights into the functional organisation of the placental transcriptome. To the best of our knowledge, the networks described herein have not been described previously, and emphasise that these findings could not be revealed through conventional gene-level summaries or differential expression experiments. In typical differential expression analyses, emphasis is placed on genes where the expression differences reach an appropriate level of significance. Although such experiments have contributed significantly to our understanding of placental genome regulation, the significance of each gene is typically determined in isolation, subsequently failing to connect genes in a manner that reflects the functional organisation of the transcriptome. By connecting genes in a manner that reflects underlying genome regulatory programs, we have exposed previously unappreciated aspects of the placental transcriptional landscape and provide a framework for multiple avenues of *post hoc* inquiry.

7.4 Methods

Ethics Statement

Ethics approval was granted by the Central Northern Adelaide Health Service Ethics of Human Research Committee (Approval #2005082) and the University of Adelaide Human Research Ethics Committee (H-137-2006). Written, informed consent was obtained from all patients.

Sample Collection

Third trimester placenta samples were collected from pregnancies classified as being uncomplicated by using the criteria described in [42], and were collected and dissected post-delivery at the Lyell McEwin Health Service, South Australia. Samples were incubated in RNAlater solution (Invitrogen) at 4 °C for 24 hours

before being stored at −80 °C. Full sample details are listed in Table S7.1 on page 168.

RNA Sequencing

RNA was extracted from 16 placental samples using TRIzol following the manufacturer's protocol. All samples were spiked with 96 External RNA Controls Consortium (ERCC) ExFold RNA transcripts. Ribosomal RNAs were depleted from samples using Ribo-Zero Gold and sequencing libraries were prepared using Illumina®TruSeq®Stranded Total RNA Sample Preparation kits. Sequencing was performed on the Illumina Hi-Seq 2500 using a 100bp paired-end protocol at the Australian Cancer Genomics Facility in Adelaide.

Sequence adapters were trimmed using AdapterRemoval with options `--trimns, --minlength 20`. Trimmed RNA-Seq reads were aligned to known UCSC hg19 genes and the hg19 genome using Bowtie 2 v2.1.0 and TopHat v2.0.9 with options `--library-type=fr-firststrand --mate-inner-dist -20 --mate-std-dev 180`. UCSC hg19 reference genome and transcriptome was obtained through Illumina iGenomes link.

Aligned RNA-Seq reads were summarised using the summarizeOverlaps algorithm with the options `overlapMode=''Union'', ignoreStrand=FALSE, singleEnd=FALSE, fragments=TRUE` [43] to generate a table of unique read counts per gene for each sample. Only genes > 1FPKM were retained (15,861 genes) and count data were transformed and quantile-normalised using the Voom method [44] to produce a numeric matrix of normalised expression values on the log2 scale.

Network Construction

To construct the network of co-expressed genes, we selected the most variable upper third of genes in the placental RNA-Seq dataset using the Weighted Gene Co-expression Network Analysis methods implemented in the WGCNA R package [14]. Briefly, gene expression values were used to construct a signed co-expression network by computing a Pearson's correlation matrix, which is then used to compute an adjacency matrix by raising the correlation matrix to a power. We chose a power of eight, which was determined by plotting scale-free fit and mean connectivity as a function of power (Figure S7.4 on page 167) using the scale-free topology criteria outlined in [45]. By raising the absolute value of the correlation

to a power, the construction of co-expression networks emphasises high correlations at the expense of low correlations [14]. The interconnectedness (topological overlap) of each gene pair was calculated using the adjacency matrix, which was then used as input for average linkage hierarchical clustering.

Gene modules were then defined as branches of the resulting clustering tree, with the branches cut into defined modules using the dynamic tree-cut algorithm [15]. Gene modules were then summarised by calculating module eigengenes, which are defined as the first principal components of the module expression profiles. As module eigengenes capture the maximum amount of variation of gene expression within a module, the eigengene is considered a representative value (or weighted average) of module gene expression [14]. For each module, the gene membership value (kME) is defined as the correlation between the standardised gene expression values for each gene and the module eigengene for each sample [14]. We assigned genes to modules if they had a high module membership defined as $kME > 0.7$, and genes with a value below this threshold were assigned to the M0 (grey) module. Note that using this method allows genes to be members of more than one module.

Module Preservation

To evaluate the preservation of human third trimester placenta gene modules in independent placenta gene expression datasets, we used the WGCNA modulePreservation function to generate module preservation statistics [14]. These methods test whether the density and connectivity patterns of gene modules defined in our reference dataset are preserved in independent datasets. We used the Z_{summary} statistic to summarise the evidence for significant module preservation compared to a random sample of all network genes reiterated over 100 permutations per dataset. We adopted the thresholds suggested by Langfelder et al [8], who indicate $Z_{\text{summary}} < 2$ implies no evidence for module preservation, $2 < Z_{\text{summary}} < 10$ implies weak to moderate preservation, and $Z_{\text{summary}} > 10$ implies strong evidence for module preservation.

RNA-Seq Validation Dataset

We used the raw RNA-Seq reads from 20 human third trimester placenta samples as previously described in a separate analysis of the human placental transcriptome [5]. In this current study, RNA-Seq reads were aligned to the human reference genome and UCSC known genes (hg19) using Tophat 2 with the options `--library-type=fr-unstranded --segment-length=18`. For the mouse

expression data, we obtained RNA-Seq fastq files for 23 samples from the NCBI short read archive (SRA062227). Reads were aligned to mm10 genome and UCSC known genes using Tophat2 with the options `--library-type=fr-unstranded --read-mismatches 3 --read-edit-dist 3`. Alignment bam files were summarised to obtain the number of unique read counts per gene using the summarizeOverlaps function in the genomicAlignments R package [43] with the options `ignore.strand=TRUE, paired=FALSE, mode=''union''` followed by $\log_2$ counts per million transformation and quantile normalisation. To enable the comparison of human and mouse datasets, mouse gene identifiers were converted to orthologous human gene identifiers using Ensembl Biomart and the biomaRt R package. Genes with no corresponding human gene were removed from the analysis.

Microarray Validation Datasets

For second trimester placenta, Affymetrix CEL files for 27 samples (GSE5999) were pre-processed, background subtracted and normalised using the robust multi-average (RMA) algorithm [46]. Pre-processed and normalised data from 16 first trimester placenta samples (GSE2551) and third trimester preeclampsia samples (GSE44711) were downloaded directly from NCBI GEO. Only probes that mapped uniquely to human genes were retained. In cases where multiple probes mapped to the same gene, we selected the probe with the highest mean expression. Differential expression testing of GSE44771 was performed using linear models and a rotation gene set test [29].

Gene Ontology

Gene lists for each module were tested for enrichment of gene ontology (GO) terms using Fisher tests to compute p-values for statistical over-representation of GO terms. These were compared with all the detectable genes (15,861) in our placental gene expression dataset [47].

Transcription Factor Motif Enrichment

The genes within each co-expression gene module were analysed for enrichment of transcription factor (TF)-binding sites (TFBS) against a background gene set of all detectable genes in the placenta dataset (15,861) using the oPOSSUM program and the JASPAR vertebrate core profiles [48, 49]. For each gene, we searched for

TFBS motifs in the conserved regions of the 10kb upstream/downstream sequences using a conservation cut-off of 0.4, a matrix score threshold of 85% and a minimum specificity of 8-bits. The highly enriched TFBSs were identified by ranking TFs using results from Fisher tests and Z-score rankings.

Maternal Blood Biomarkers

A panel of 47 maternal serum biomarkers (15 weeks gestation) were assayed by Alere Discovery (San Diego) to screen for molecules that may influence or be indicative of placental function, as described previously [50]. We removed redundant and potentially uninformative biomarkers using the linear correlation filter in the FSelector R package. We then selected the top three hub genes from modules with absolute eigengene-biomarker correlations > 0.7, and performed univariate linear regression analyses to identify hub genes that are correlated with the maternal blood biomarkers.

S7.1 Supporting Information

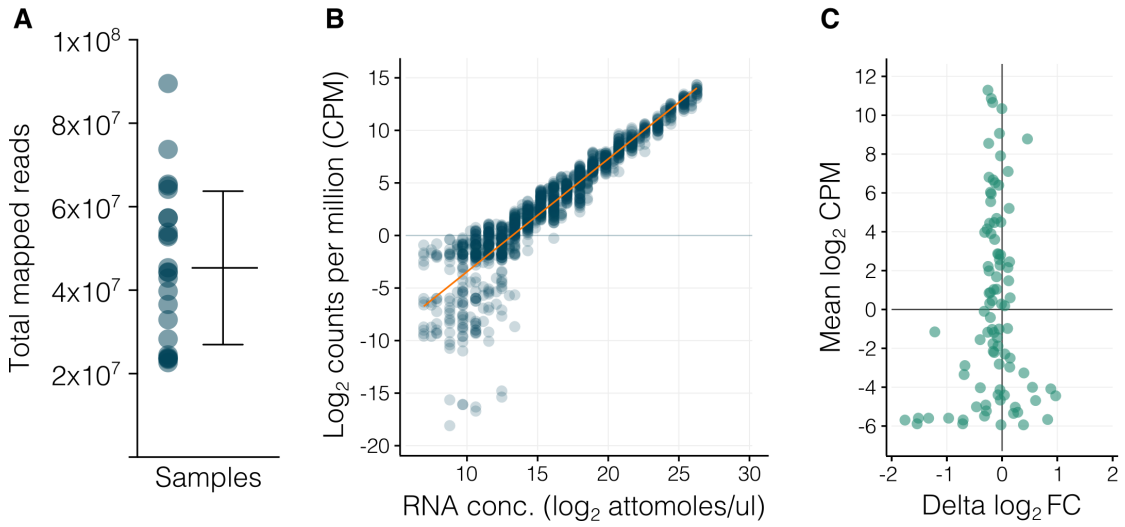


FIGURE S7.1: RNA-Seq metrics and quality control. (A) Number of mapped reads per sample. (B) Absolute concentration of ERCC spike-in RNA transcripts is highly correlated with normalised expression above 1 count per million (CPM). (C) Delta fold change (absolute fold difference – the detected fold difference) for ERCC spike-in RNA transcripts (x-axis) versus normalised expression level.

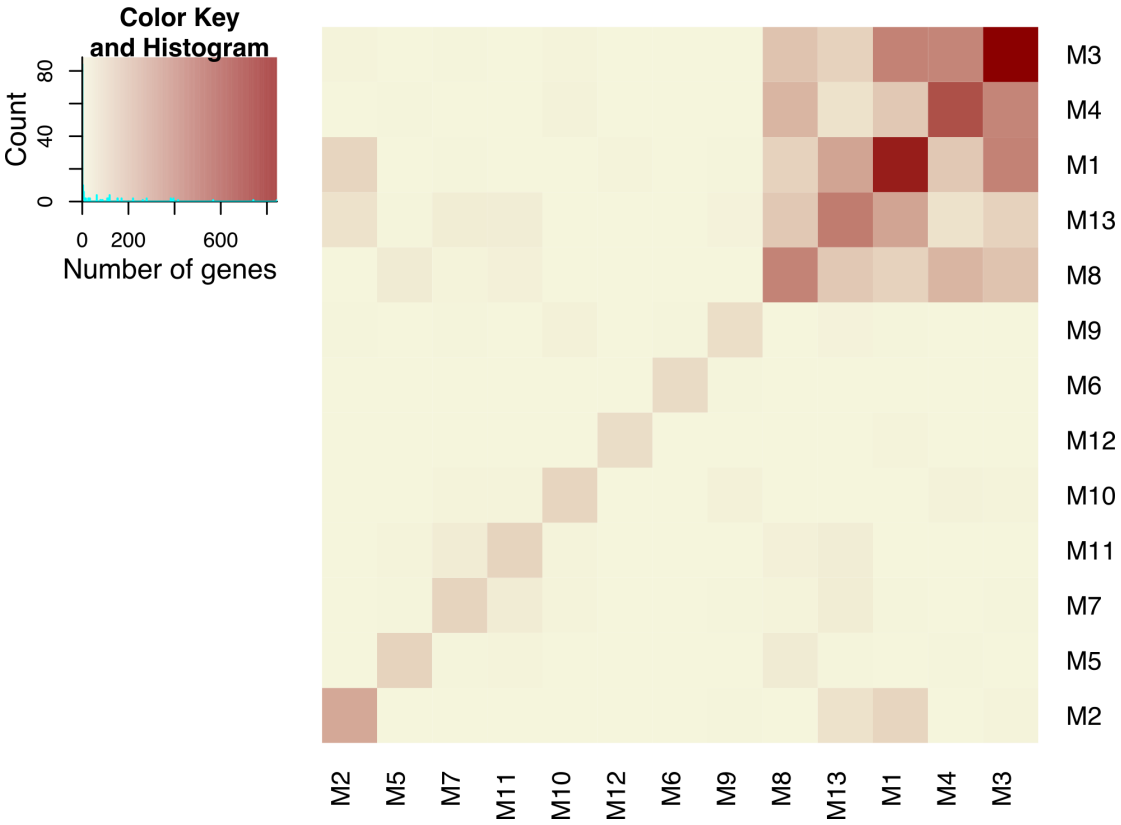
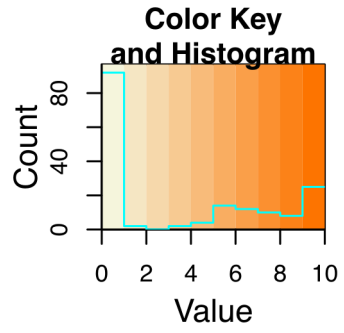


FIGURE S7.2: Heat map showing the gene overlap between modules. Colour intensity represents the number of overlapping genes between two modules.



Transcription factor ID	Human gene ID	Placenta detected (FPKM > 1)
Zfp423	*ZNF423*	Yes
SP1	*SP1*	Yes
Tcfcp2l1	*TFCP2L1*	Yes
INSM1	*INSM1*	No
Zfx	*ZFX*	Yes
EBF1	*EBF1*	Yes
Klf4	*KLF4*	Yes
MZF1_5-13	*MZF1*	Yes
MZF1_1-4	*MZF1*	Yes
ZNF354C	*ZNF354C*	Yes

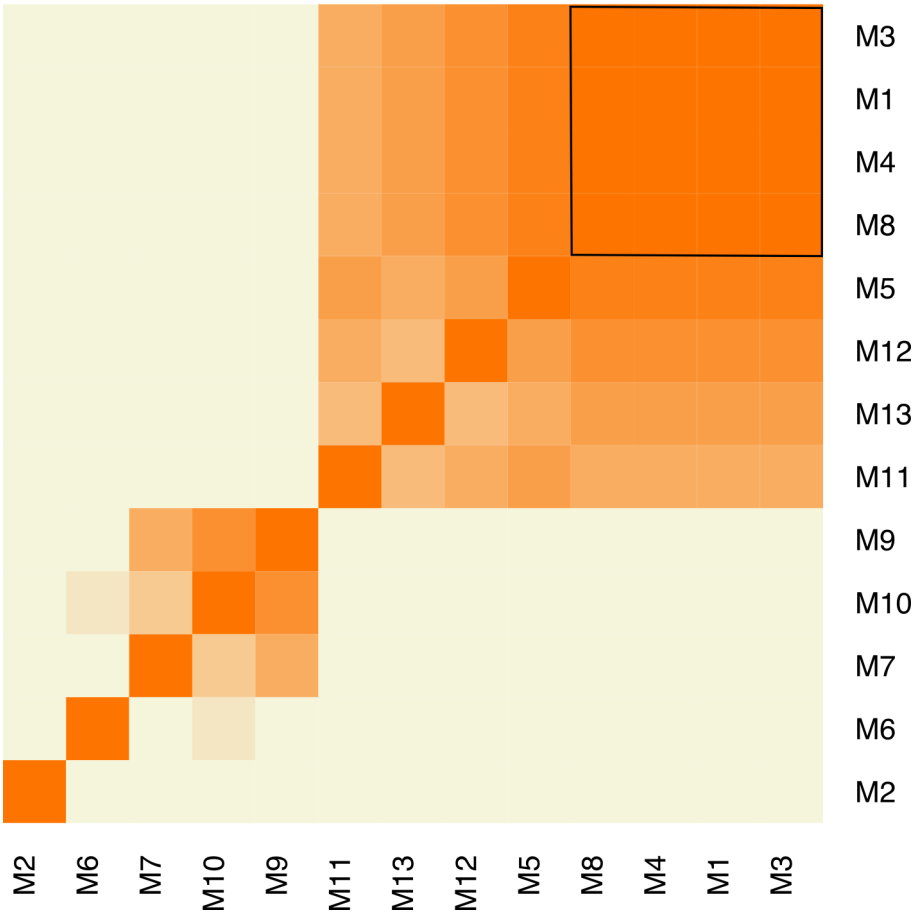


FIGURE S7.3: Heat map showing the number of overlapping top ten transcription factors predicted to regulate each co-expression module. The same top ten transcription factors are predicted in M1, M3, M4 and M8 and are shown in the table above heat map.

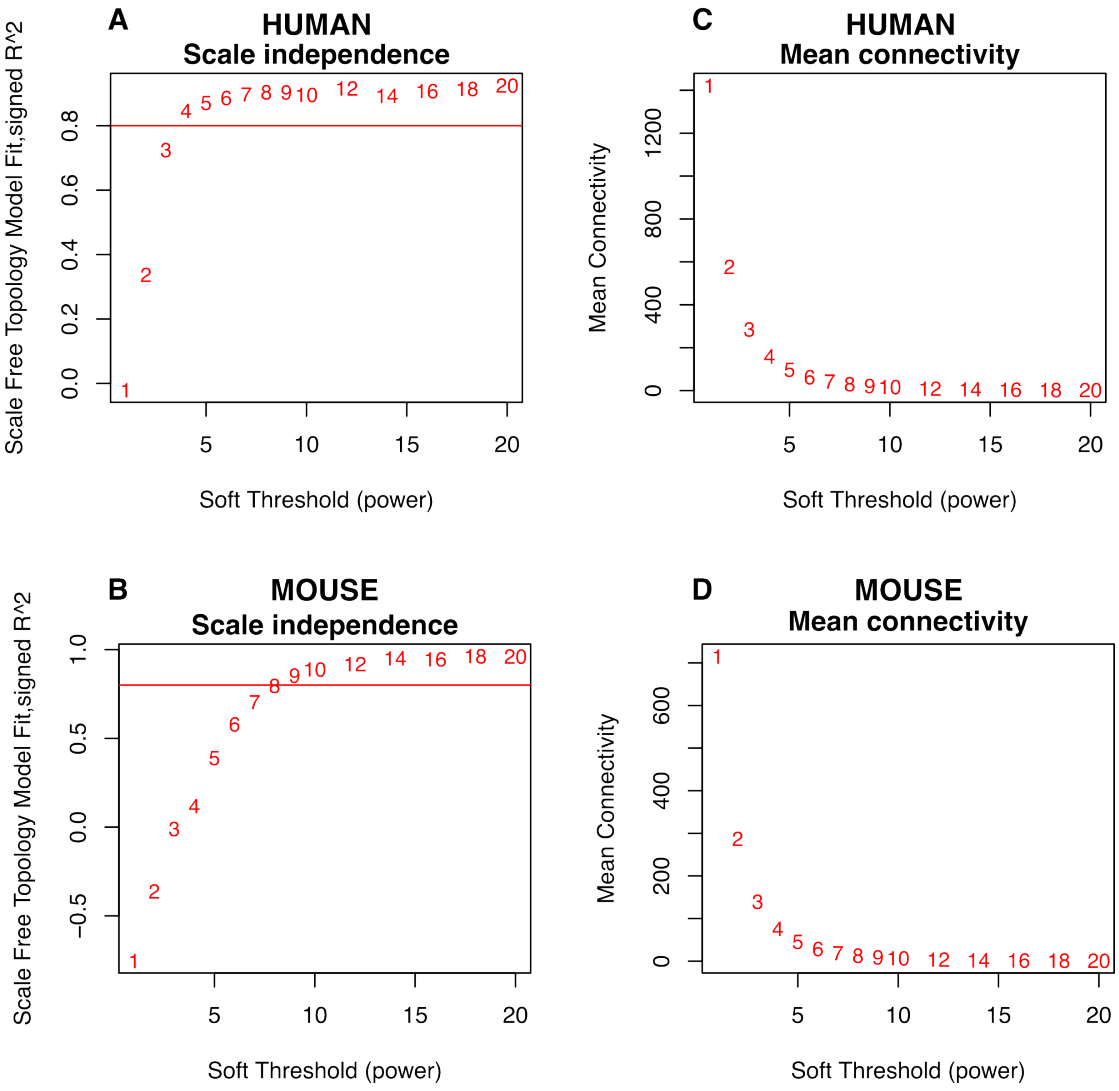


FIGURE S7.4: Summary network indices (y-axes) as functions of the soft-thresholding power (x-axes) for human and mouse. Numbers in the plots indicate the corresponding soft-thresholding powers. Plots indicate that approximate scale-free topology is attained around the soft-thresholding power of eight for both datasets. Because the summary connectivity measures decline steeply with increasing soft-thresholding power, it is advantageous to choose the lowest power that satisfies the approximate scale-free topology criterion.

TABLE S7.1: Sample characteristics.

Sample ID	Mother's BMI	Fetal sex	Fetal birth weight (g)	Head circumference (cm)	Length (cm)
1	24.4	Female	3140	33.3	47.9
2	21.9	Female	3050	33.5	47.8
3	24.8	Male	3565	34.5	50.4
4	20.9	Male	3150	34.8	49.6
5	25.2	Female	4010	36	53
6	26.9	Male	3990	35.8	50.8
7	21.6	Male	3464	34.5	49.5
8	19	Female	3705	35	51
9	18.9	Female	3250	33.5	50.5
10	18	Female	3860	35.7	50.2
11	26.3	Male	3875	36	50.2
12	25.7	Male	3600	36	49.8
13	24.5	Male	4200	36.5	52.9
14	24.6	Female	3550	35.3	49.2
15	25.4	Female	3120	33	48
16	24	Male	3680	34.5	48.5
Mean	23.3	–	3575.6	34.9	50.0
SD	2.8	–	358.1	1.1	1.6

Bibliography

[1] Guttmacher, A. E., Maddox, Y. T., and Spong, C. Y. The Human Placenta Project: placental structure, development, and function in real time. *Placenta,* 35(5) (2014), pp. 303–304.

[2] Sitras, V., Fenton, C., Paulssen, R., Vårtun, A., and Acharya, G. Differences in Gene Expression between First and Third Trimester Human Placenta: A Microarray Study. *PloS one,* 7(3) (2012), e33294.

[3] Kleinrouweler, C. E., Uitert, M. van, Moerland, P. D., Ris-Stalpers, C., Post, J. A. M. van der, and Afink, G. B. Differentially expressed genes in the pre-eclamptic placenta: a systematic review and meta-analysis. *PloS one,* 8(7) (2013), e68991.

[4] Founds, S., Conley, Y., Lyons-Weiler, J., Jeyabalan, A., Allen Hogge, W., and Conrad, K. Altered global gene expression in first trimester placentas of women destined to develop preeclampsia. *Placenta,* 30(1) (2009), pp. 15–24.

[5] Saben, J., Zhong, Y., McKelvey, S., Dajani, N. K., Andres, A., Badger, T. M., Gomez-Acevedo, H., and Shankar, K. A comprehensive analysis of the human placenta transcriptome. *Placenta,* 35(2) (2014), pp. 125–131.

[6] Kim, J., Zhao, K., Jiang, P., Lu, Z.-X., Wang, J., Murray, J. C., and Xing, Y. Transcriptome landscape of the human placenta. *BMC Genomics,* 13(1) (2012), p. 115.

[7] Sadovsky, Y., Clifton, V. L., and Burton, G. J. Invigorating placental research through the "Human Placenta Project". *Placenta,* 35(8) (2014), pp. 527–.

[8] Langfelder, P., Castellani, L. W., Zhou, Z., Paul, E., Davis, R., Schadt, E. E., Lusis, A. J., Horvath, S., and Mehrabian, M. A systems genetic analysis of high density lipoprotein metabolism and network preservation across mouse models. *Biochimica et biophysica acta,* 1821(3) (2012), pp. 435–447.

[9] Hu, Y., Wu, G., Rusch, M., Lukes, L., Buetow, K. H., Zhang, J., and Hunter, K. W. Integrated cross-species transcriptional network analysis of metastatic susceptibility. *Proceedings of the National Academy of Sciences of the United States of America,* 109(8) (2012), pp. 3184–3189.

[10] Voineagu, I., Wang, X., Johnston, P., Lowe, J. K., Tian, Y., Horvath, S., Mill, J., Cantor, R. M., Blencowe, B. J., and Geschwind, D. H. Transcriptomic analysis of autistic brain reveals convergent molecular pathology. *Nature,* 474(7351) (2011), pp. 380–384.

[11] Oldham, M. C., Konopka, G., Iwamoto, K., Langfelder, P., Kato, T., Horvath, S., and Geschwind, D. H. Functional organization of the transcriptome in human brain. *Nature Neuroscience,* 11(11) (2008), pp. 1271–1282.

[12] Gupta, S., Ellis, S. E., Ashar, F. N., Moes, A., Bader, J. S., Zhan, J., West, A. B., and Arking, D. E. Transcriptome analysis reveals dysregulation of innate immune response genes and neuronal activity-dependent genes in autism. *Nature communications,* 5 (2014), p. 5748.

[13] Xue, Z., Huang, K., Cai, C., Cai, L., Jiang, C.-y., Feng, Y., Liu, Z., Zeng, Q., Cheng, L., Sun, Y. E., Liu, J.-y., Horvath, S., and Fan, G. Genetic programs in human and mouse early embryos revealed by single-cell RNA sequencing. *Nature,* 500(7464) (2013), pp. 593–597.

[14] Langfelder, P. and Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC bioinformatics,* 9 (2008), p. 559.

[15] Langfelder, P., Zhang, B., and Horvath, S. Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. *Bioinformatics (Oxford, England),* 24(5) (2008), pp. 719–720.

[16] Langfelder, P., Luo, R., Oldham, M. C., and Horvath, S. Is my network module preserved and reproducible? *PLoS computational biology,* 7(1) (2011), e1001057.

[17] Winn, V. D., Haimov-Kochman, R., Paquet, A. C., Yang, Y. J., Madhusudhan, M. S., Gormley, M., Feng, K.-T. V., Bernlohr, D. A., McDonagh, S., Pereira, L., Sali, A., and Fisher, S. J. Gene expression profiling of the human maternal-fetal interface reveals dramatic changes between midgestation and term. *Endocrinology,* 148(3) (2007), pp. 1059–1079.

[18] Finn, E. H., Smith, C. L., Rodriguez, J., Sidow, A., and Baker, J. C. Maternal Bias and Escape from X Chromosome Imprinting in the Midgestation Mouse Placenta. *Developmental Biology,* (2014).

[19] Gude, N. M., Roberts, C. T., Kalionis, B., and King, R. G. Growth and function of the normal human placenta. *Thrombosis Research,* 114(5-6) (2004), pp. 397–407.

[20] Harder, L., Puller, A.-C., and Horstmann, M. A. ZNF423: Transcriptional modulation in development and cancer. *Molecular & Cellular Oncology,* 1(3) (2014), e969655.

[21] Harder, L., Eschenburg, G., Zech, A., Kriebitzsch, N., Otto, B., Streichert, T., Behlich, A.-S., Dierck, K., Klingler, B., Hansen, A., Stanulla, M., Zimmermann, M., Kremmer, E., Stocking, C., and Horstmann, M. A. Aberrant ZNF423 impedes B cell differentiation and is linked to adverse outcome of ETV6-RUNX1 negative B precursor acute lymphoblastic leukemia. *The Journal of experimental medicine,* 210(11) (2013), pp. 2289–2304.

[22] Tsai, R. Y. and Reed, R. R. Identification of DNA recognition sequences and protein interaction domains of the multiple-Zn-finger protein Roaz. *Molecular and cellular biology,* 18(11) (1998), pp. 6447–6456.

[23] Wang, M. M. and Reed, R. R. Molecular cloning of the olfactory neuronal transcription factor Olf-1 by genetic selection in yeast. *Nature,* 364(6433) (1993), pp. 121–126.

[24] Khong, T. Y., De Wolf, F., Robertson, W. B., and Brosens, I. Inadequate maternal vascular response to placentation in pregnancies complicated by pre-eclampsia and by small-for-gestational age infants. *British Journal of Obstetrics and Gynaecology,* 93(10) (1986), pp. 1049–1059.

[25] Kim, Y. M., Bujold, E., Chaiworapongsa, T., Gomez, R., Yoon, B. H., Thaler, H. T., Rotmensch, S., and Romero, R. Failure of physiologic transformation of the spiral arteries in patients with preterm labor and intact membranes. *Am J Obstet Gynecol,* 189(4) (2003), pp. 1063–1069.

[26] Kim, Y. M., Chaiworapongsa, T., Gomez, R., Bujold, E., Yoon, B. H., Rotmensch, S., Thaler, H. T., and Romero, R. Failure of physiologic transformation of the spiral arteries in the placental bed in preterm premature rupture of membranes. *Am J Obstet Gynecol,* 187(5) (2002), pp. 1137–1142.

[27] Uzun, A., Laliberte, A., Parker, J., Andrew, C., Winterrowd, E., Sharma, S., Istrail, S., and Padbury, J. F. dbPTB: a database for preterm birth. *Database : the journal of biological databases and curation,* 2012 (2012), bar069–bar069.

[28] Blair, J. D., Yuen, R. K. C., Lim, B. K., McFadden, D. E., Dadelszen, P. von, and Robinson, W. P. Widespread DNA hypomethylation at gene enhancer regions in placentas associated with early-onset pre-eclampsia. *Molecular Human Reproduction,* 19(10) (2013), pp. 697–708.

[29] Wu, D., Lim, E., Vaillant, F., Asselin-Labat, M.-L., Visvader, J. E., and Smyth, G. K. ROAST: rotation gene set tests for complex microarray experiments. *Bioinformatics (Oxford, England),* 26(17) (2010), pp. 2176–2182.

[30] Crosley, E. J., Elliot, M. G., Christians, J. K., and Crespi, B. J. Placental invasion, preeclampsia risk and adaptive molecular evolution at the origin of the great apes: evidence from genome-wide analyses. *Placenta,* 34(2) (2013), pp. 127–132.

[31] Maymó, J. L., Pérez Pérez, A., Gambino, Y., Calvo, J. C., Sánchez-Margalet, V., and Varone, C. L. Review: Leptin gene expression in the placenta–regulation of a key hormone in trophoblast proliferation and survival. *Placenta,* 32 Suppl 2 (2011), S146–53.

[32] Lehman, A., Stittrich, A.-B., Glusman, G., Zong, Z., Li, H., Eydoux, P., Senger, C., Lyons, C., Roach, J. C., and Patel, M. Diffuse angiopathy in Adams-Oliver syndrome associated with truncating DOCK6 mutations. *American Journal of Medical Genetics Part A,* 164A(10) (2014), pp. 2656–2662.

[33] Sitras, V., Paulssen, R., Leirvik, J., Vårtun, A., and Acharya, G. Placental gene expression profile in intrauterine growth restriction due to placental insufficiency. *Reproductive sciences (Thousand Oaks, Calif.),* 16(7) (2009), pp. 701–711.

[34] Løset, M., Mundal, S. B., Johnson, M. P., Fenstad, M. H., Freed, K. A., Lian, I. A., Eide, I. P., Bjørge, L., Blangero, J., Moses, E. K., and Austgulen, R. A transcriptional profile of the decidua in preeclampsia. *Am J Obstet Gynecol,* 204(1) (2011), 84.e1–27.

[35] Cheng, L. E., Zhang, J., and Reed, R. R. The transcription factor Zfp423/OAZ is required for cerebellar development and CNS midline patterning. *Developmental Biology,* 307(1) (2007), pp. 43–52.

[36] Liao, D. Emerging Roles of the EBF Family of Transcription Factors in Tumor Suppression. *Molecular Cancer Research,* (2009).

[37] Brancati, F., Fortugno, P., Bottillo, I., Lopez, M., Josselin, E., Boudghene-Stambouli, O., Agolini, E., Bernardini, L., Bellacchio, E., Iannicelli, M., Rossi, A., Dib-Lachachi, A., Stuppia, L., Palka, G., Mundlos, S., Stricker, S., Kornak, U., Zambruno, G., and Dallapiccola, B. Mutations in PVRL4, encoding cell adhesion molecule nectin-4, cause ectodermal dysplasia-syndactyly syndrome. *American journal of human genetics,* 87(2) (2010), pp. 265–273.

[38] Pavlova, N. N., Pallasch, C., Elia, A. E., Braun, C. J., Westbrook, T. F., Hemann, M., and Elledge, S. J. A role for PVRL4-driven cell-cell interactions in tumorigenesis. *eLife,* 2 (2013), e00358.

[39] Derycke, M. S., Pambuccian, S. E., Gilks, C. B., Kalloger, S. E., Ghidouche, A., Lopez, M., Bliss, R. L., Geller, M. A., Argenta, P. A., Harrington, K. M., and Skubitz, A. P. N. Nectin 4 overexpression in ovarian cancer tissues and serum: potential role as a serum biomarker. *American journal of clinical pathology,* 134(5) (2010), pp. 835–845.

[40] Stein, S., Thomas, E. K., Herzog, B., Westfall, M. D., Rocheleau, J. V., Jackson, R. S., Wang, M., and Liang, P. NDRG1 is necessary for p53-dependent apoptosis. *Journal of Biological Chemistry,* 279(47) (2004), pp. 48930–48940.

[41] Heazell, A. E. P. and Crocker, I. P. Live and let die - regulation of villous trophoblast apoptosis in normal and abnormal pregnancies. *Placenta,* 29(9) (2008), pp. 772–783.

[42] McCowan, L. M. E., Roberts, C. T., Dekker, G. A., Taylor, R. S., Chan, E. H. Y., Kenny, L. C., Baker, P. N., Moss-Morris, R., Chappell, L. C., North, R. A., and SCOPE consortium. Risk factors for small-for-gestational-age infants by customised birthweight centiles: data from an international prospective cohort study. *BJOG : an international journal of obstetrics and gynaecology,* 117(13) (2010), pp. 1599–1607.

[43] Lawrence, M., Huber, W., Pagès, H., Aboyoun, P., Carlson, M., Gentleman, R., Morgan, M. T., and Carey, V. J. Software for computing and annotating genomic ranges. *PLoS computational biology,* 9(8) (2013), e1003118.

[44] Law, C. W., Chen, Y., Shi, W., and Smyth, G. K. Voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome biology,* 15(2) (2014), R29.

[45] Zhang, B. and Horvath, S. A general framework for weighted gene co-expression network analysis. *Statistical applications in genetics and molecular biology,* 4 (2005), Article17.

[46] Irizarry, R. A., Bolstad, B. M., Collin, F., Cope, L. M., Hobbs, B., and Speed, T. P. Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Research,* 31(4) (2003), e15.

[47] Falcon, S. and Gentleman, R. Using GOstats to test gene lists for GO term association. *Bioinformatics (Oxford, England),* 23(2) (2007), pp. 257–258.

[48] Kwon, A. T., Arenillas, D. J., Hunt, R. W., and Wasserman, W. W. oPOSSUM-3: Advanced Analysis of Regulatory Motif Over-Representation Across Genes or ChIP-Seq Datasets. *G3,* 2(Spetember) (2012), pp. 987–1002.

[49] Mathelier, A., Zhao, X., Zhang, A. W., Parcy, F., Worsley-Hunt, R., Arenillas, D. J., Buchman, S., Chen, C. y., Chou, A., Ienasescu, H., Lim, J., Shyr, C., Tan, G., Zhou, M., Lenhard, B., Sandelin, A., and Wasserman, W. W. JASPAR 2014: an extensively expanded and updated open-access database of transcription factor binding profiles. *Nucleic Acids Research,* 42 (2013), pp. D142–D147.

[50] Kenny, L. C., Black, M. A., Poston, L., Taylor, R., Myers, J. E., Baker, P. N., McCowan, L. M., Simpson, N. A. B., Dekker, G. A., Roberts, C. T., Rodems, K., Noland, B., Raymundo, M., Walker, J. J., and North, R. A. Early pregnancy prediction of preeclampsia in nulliparous women, combining clinical risk and biomarkers: the Screening for Pregnancy Endpoints (SCOPE) international cohort study. *Hypertension,* 64(3) (2014), pp. 644–652.

Chapter 8

General Discussion

The primary aim of this project was to further our understanding of three aspects of placental genome regulation: (1) the establishment of DNA methylation during first trimester gestation; (2) the effect of fetal sex on placental gene expression; and (3) the underlying organisation of the human placental transcriptome. The work presented in this thesis has provided novel insight into the establishment of DNA methylation imprints during the first trimester; the first comprehensive assessment of sex differences in the human placental transcriptome; and the first integrative analysis of placental gene co-expression.

8.1 Overall Significance

8.1.1 Epigenetic Plasticity in the Human Placenta at 6–10 Weeks Gestation

The research presented in Chapter 2 on genomic epigenetic imprinting during development of the human first trimester placenta highlighted a window of epigenetic plasticity between six and ten weeks of gestation [1]. These results indicate that imprinting at the *H19* locus is not established as early as was previously thought and may contribute to early programming of placental phenotype. Furthermore, this study was the first to employ a highly quantitative measure of allele-specific expression for these genes in the human placenta. Our demonstration of significant allelic variation in *H19* expression, and the observation of consistent biallelic *IGF2R* expression, illustrated the need for robust methodologies to determine the role of imprinted genes in normal and pathological placental development.

8.1.2 Widespread Sex-Biased Gene Expression in the Human Placenta

Our transcriptome meta-analysis [2] revealed that the extent of sex-biased expression in the placenta is more extensive than that described in previous reports [3, 4]. Our results demonstrate that a vast majority of sex-biased genes in the human placenta have not been observed to show sex biases in studies of other human tissues [5–7]. Importantly, given the significant sex biases observed for healthy placental ('control') tissue, the results of this study highlight the need to incorporate fetal sex as a covariate in any studies of placental gene expression. We consider these results to be highly robust, since this study incorporated data from samples collected in Europe, USA and Asia using a variety of microarray platforms from three different manufacturers.

8.1.3 Sex-Biased Expression of Genes Encoding hCG

As human chorionic gonadotropin (hCG) promotes placental growth and vasculogenesis, our observation of consistently higher female expression of genes encoding hCG suggests that female fetuses invest more in extra-embryonic tissue development than males. Since a mother has finite resources to allocate to a fetus *in utero*, these findings support the supposition that males invest more resources in the growth and development of embryonic tissues at the expense of investing less in the development of extra-embryonic tissues [8, 9]. This may be one contributing factor to the male bias in the incidence of placental dysfunction [10] and pregnancy complications where placental pathology is implicated [11–13].

8.1.4 Conservation of Gene Regulatory Programs across Gestation

To investigate transcriptome-wide gene co-expression we assembled 16 human term placental transcriptome datasets (37–40 weeks of gestation) and identified distinct clusters of genes that are expressed in a highly coordinated manner. By integrating these data with additional placental transcriptome datasets from both human and mouse, we show that distinct patterns of co-expression are preserved across human gestation (8–40 weeks). These patterns are highly conserved between human and mouse, subsequently revealing previously unappreciated molecular networks involved in placental development. To the best of our knowledge, this is the first detailed investigation into human placental gene co-expression. Our findings

provide a new foundation for investigations into placental genome regulation by revealing the underlying organisation of the human placental transcriptome.

8.1.5 Evidence of Co-Regulation of Genes Implicated in Preeclampsia

Our transcriptome co-expression analyses also identified a cluster of genes heavily implicated in preeclampsia, which show highly correlated patterns of expression. This suggests regulation by a common set of factors. Numerous studies have identified several genes that are differentially expressed in placentas from normal pregnancies compared to those affected by preeclampsia, however the results have been mixed and largely lack consensus. In our study we demonstrate that many genes previously found to be associated with preeclampsia show highly coordinated patterns of expression. This indicates that specific factors (currently unknown), which modulate gene expression, may be perturbed in preeclampsia. These results therefore provide a new framework for investigating altered gene expression in the placentas from preeclamptic pregnancies.

8.2 Contributions to the Field

8.2.1 Detailed Reference of Placental Sex-Biased Gene Expression

The transcriptome meta-analysis presented in Chapter 5 represents the most comprehensive assessment of sex-biased expression in the non-pathological placenta to date. This serves as a baseline measure of healthy sex-biased gene expression that will enable more in-depth research into sex biases in placental pathologies. Complete annotated results of this study have been made available as supplementary data to this study and serve as a resource for the wider research community.

8.2.2 The *massiR* Software Package

The software package *massiR*, developed during this project, implements algorithms for predicting sample sex from high-throughput gene expression datasets and constitutes a valuable contribution to the wider research community. By enabling *post hoc* identification of sample sex, this tool increases the useability of

thousands of publicly available datasets for researchers wishing to test for sex differences in expression, or to include sex as a covariate in statistical analyses. The *massiR* package is also a valuable validation tool for newly generated datasets, which can aid in detecting common laboratory errors such as tube mislabelling. The *massiR* package is currently hosted on Bioconductor, the world's leading bioinformatics software repository, and is currently downloaded more than 200 times per month.

8.2.3 Comprehensive Placental Transcriptome Dataset

The placental co-expression analysis presented in this thesis (Chapter 7) provides both the resources and the framework for further investigations into the organisation of placental gene regulation. Upon publication of this work, a fully annotated transcriptome assembly, along with comprehensive co-expression results, will be released into the public domain. This will constitute a valuable community resource for researchers conducting further analyses into placental gene regulation.

8.2.4 A New Framework for Screening Biomarkers of Placental Development and Function

In summarising the placental transcriptome by grouping co-expressed genes, we demonstrated a method that significantly reduces the dimensionality of transcriptome-scale data. The results of the study presented in Chapter 7 show how hundreds of functionally related co-expressed genes can be accurately decomposed to a single value with a high degree of information being retained. One powerful implementation of these summarised gene expression measures is to use them to screen for correlates of gene expression without having to impose severe adjustments for testing multiple hypotheses. This subsequently opens the door for high-throughput screening of biomarkers of placental gene expression. Given recent initiatives to pursue non-invasive methods of monitoring placental development in real-time, the approach outlined in Chapter 7 constitutes an important step toward surrogate markers of placental development.

8.3 Problems Encountered and Limitations

8.3.1 Genomic Imprinting in the Placenta

The major limitation with allele-specific expression and methylation analyses was the limited number of loci under investigation. In this first study, we adopted a candidate gene approach, which focused on two genes that have been widely utilised to investigate parent-of-origin allele-specific expression and have known effects on placental phenotype. However, current estimates indicate that the number of imprinted genes expressed in the placenta is in the hundreds, which may be regulated by more than one epigenetic mechanism (DNA methylation and histone modifications). A secondary limitation relates to the study of first trimester placental tissue from elective terminations of pregnancy. Approximately one in five pregnancies feature a complication that could be partially or fully attributed to improper placental development. For this reason, we have little idea if these placentas were obtained from pregnancies that were destined to feature an obstetric complication. Ideally, analysis of first trimester chorionic villous samples from ongoing pregnancies for which outcome is known would provide the best information but also shed light on first trimester origins of pregnancy complications.

8.3.2 Transcriptome Meta-Analysis

Regarding the sex-biased expression analysis, the lack of appropriate software for conducting a microarray meta-analysis with raw probe-level data on this scale meant a significant effort was required to develop robust analysis pipelines for normalising, summarising and ultimately meta-analysing all of the gene expression data.

There were also issues encountered in the use of publicly available data, relating to metadata availability. Firstly, the majority of samples lacked sex information. Although this was overcome by creating an algorithm for predicting sample sex (Chapter 4), it was challenging to obtain sufficient empirical datasets with confirmed sex information in order to validate the algorithm. Given the diversity of the human population, we would have also been able to conduct a much more comprehensive analysis if more detailed metadata were included with publicly available expression data (e.g. body mass index, ethnicity, age, parity, smoking status, etc.). Secondly, authors do not always make placental gene expression array data publicly available in its raw format – if they had, we would have been able to conduct a larger and more powerful study. Finally, the annotation of many

probes on most, if not all, microarray platforms is often ambiguous and we had to exercise caution in deciding which probes to include in the analysis. Although we re-annotated all the arrays by mapping probes back to a common reference genome using a relatively conservative approach, there were still some ambiguities, which raised issues when interpreting the results. However, the shift in the field towards RNA sequencing to quantify global gene expression will help to overcome some of these issues if researchers are more diligent with metadata reporting.

8.3.3 Transcriptome Co-Expression Analysis

At the start of this PhD, RNA-Seq had only just begun to be widely utilised for the quantification of gene expression. For this reason, many algorithms for assembling and quantifying RNA transcripts were immature, and there remains no consensus on the most appropriate way(s) to reliably quantify expression. A large proportion of time during my candidature was therefore spent learning about the algorithms used for RNA-Seq mapping and quantification; optimising the mapping, transcript assembly and quantification pipelines; developing methods of RNA-Seq quality control; and extensively testing the analytical tool parameters. Although time-consuming, a vast amount of experience was gained.

One weakness of our RNA-Seq co-expression analysis relates to the number of biological replicates in our experiment ($n = 16$). A very recent study has suggested that thousands of samples may be required for a gold-standard weighted-gene co-expression network analysis [14]. It was not feasible for us to conduct a study on this scale due to the prohibitive cost of RNA-Seq. However, by integrating multiple datasets from previously published studies, we were able to validate and extend our findings, which undoubtedly strengthened our results. Regardless, this approach is largely reliant on the diligence and transparency of other researchers in terms of their sample collection, accurate metadata reporting, and their use of appropriate quality controls.

8.4 Future Directions

In recent years, undeniable progress has been made towards obtaining a clearer understanding of how genomic information orchestrates the complex processes of mammalian development and cellular function. However, this progress is somewhat paradoxical; the more we learn, the more we realise that there is much we still do not know. The research presented here has cast new light upon the role

of genome regulation in placental development and related pregnancy pathologies, yet many important questions remain.

Firstly, to what degree is the placental epigenome modulated throughout gestation? To what extent does this influence placental phenotype and fetal development? Although the research presented here has increased our understanding of gene expression patterns maintained in the placenta across gestation, we possess limited knowledge of the epigenetic mechanisms that govern such precise regulation. Future work focused on profiling the placental epigenome throughout gestation will undoubtedly help to delineate the role of the epigenome in regulating placental development and function. Moreover, identifying periods of dynamic epigenetic change during placental development may allow us to determine sensitive periods where the regulatory framework for the remainder of gestation is established. This may be crucial to determine how the maternal and external environments can induce stable epigenetic change in the placenta, and how this influences fetal development and health *in utero* and beyond.

Secondly, coupling RNA-Seq with the epigenome profiles outlined above will allow a detailed investigation of genomic imprinting in the placenta across gestation. To date, there have been no comprehensive genome-wide assessments of imprinted gene regulation across gestation. Given that imprinted genes are known to have a profound influence on placental phenotype and fetal development, the information gained from such a study would be incredibly valuable to placental researchers and the wider genomics community.

Thirdly, to what extent are rodent models useful for understanding human placental development? One component of this PhD research revealed a highly conserved transcriptional program between human and mouse, suggesting the existence of common regulatory mechanisms in the placenta. However, several modules of co-regulated genes were not conserved in the mouse, particularly that implicated in preeclampsia. This indicates that these molecular networks underpin functions that may be more specific to human placental development and to human (or primate) pregnancy complications that do not occur in other mammalian orders. Given that the placenta is one of the most highly diverse tissues among mammals, future studies focused on elucidating the specific aspects of conservation and divergence between human and mouse will be required. This will allow us to carefully evaluate the conditions under which rodent models are useful for providing insight into human placental development, and which conditions have limited utility.

Finally, one of the biggest challenges in placental biology is to develop non-invasive methods of monitoring placental development and function in real-time. In Chapter 7, we demonstrated how to screen for surrogates of placental gene expression

by reducing the complexity of the transcriptome. A series of carefully designed experiments would need to be carried out to pursue this idea further. Firstly, full transcriptome profiling is needed, using many more placental samples from normal and pathological pregnancies to generate a highly robust gene co-expression network. Secondly, we would need to measure a high number of biomarkers in placental-matched maternal blood samples and develop methods for combining biomarkers (maternal SNPs, proteins, enzymes, placenta-derived RNAs) that most accurately predict the level of co-expressed genes. Thirdly, through a series of cell and animal model experiments, we would need to determine the regulators and hub genes most critical to the gene networks in order to refine biomarker screening around these results. The integration of genome–epigenome–transcriptome data with clinical phenotypes and biomarkers presents unique challenges that must be overcome if we are to transition into a functioning field of genomic–obstetric medicine. Although such projects are ambitious, the information gained would be of immense value to the discipline of obstetrics and gynaecology.

8.5 Conclusion

The sequence of the human genome encodes the genetic instructions underpinning human physiology. Just over a decade ago, the International Human Genome Sequencing Consortium released the first highly accurate build of the human genome [15], but this was not the end – it was a new beginning. The discovery that the human genome encodes only 20,000–25,000 protein-coding genes meant it rapidly became apparent that physiological processes hinge as much upon genome regulation as genome content.

The work presented in this thesis has explored several aspects of human placental genome regulation and demonstrates the complex and multi-faceted landscape of the human placental transcriptome. We have demonstrated that important changes occur in the regulation of imprinted genes during the first trimester, that fetal sex has a profound influence on placental gene regulation, and that the placental transcriptome is organised into functional modules of co-expressed genes. The work published from this thesis will contribute to the advancement of research into how the placental genome influences fetal development *in utero*. Furthermore, the analytical framework established in this thesis may be beneficial to develop non-invasive methods of monitoring placental development and function in real-time.

Bibliography

[1] Buckberry, S., Bianco-Miotto, T., Hiendleder, S., and Roberts, C. T. Quantitative allele-specific expression and DNA methylation analysis of H19, IGF2 and IGF2R in the human placenta across gestation reveals H19 imprinting plasticity. *PloS one,* 7(12) (2012), e51210.

[2] Bianco-Miotto, T., Bent, S. J., Dekker, G. A., and Roberts, C. T. Integrative transcriptome meta-analysis reveals widespread sex-biased gene expression at the human fetal–maternal interface. *Molecular Human Reproduction,* (2014).

[3] Sood, R., Zehnder, J. L., Druzin, M. L., and Brown, P. O. Gene expression patterns in human placenta. *Proceedings of the National Academy of Sciences,* 103(14) (2006), pp. 5243–5244.

[4] Cvitic, S., Longtine, M. S., Hackl, H., Wagner, K., Nelson, M. D., Desoye, G., and Hiden, U. The Human Placental Sexome Differs between Trophoblast Epithelium and Villous Vessel Endothelium. *PloS one,* 8(10) (2013), e79233.

[5] Trabzuni, D., Ramasamy, A., Imran, S., Walker, R., Smith, C., Weale, M. E., Hardy, J., Ryten, M., and North American Brain Expression Consortium. Widespread sex differences in gene expression and splicing in the adult human brain. *Nature communications,* 4 (2013), p. 2771.

[6] Zhang, Y., Klein, K., Sugathan, A., Nassery, N., Dombkowski, A., Zanger, U. M., and Waxman, D. J. Transcriptional profiling of human liver identifies sex-biased genes associated with polygenic dyslipidemia and coronary artery disease. *PloS one,* 6(8) (2011), e23506.

[7] Whitney, A. R., Diehn, M., Popper, S. J., Alizadeh, A. A., Boldrick, J. C., Relman, D. A., and Brown, P. O. Individuality and variation in gene expression patterns in human blood. *Proceedings of the National Academy of Sciences of the United States of America,* 100(4) (2003), pp. 1896–1901.

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

[9] Eriksson, J. G., Kajantie, E., Osmond, C., Thornburg, K., and Barker, D. J. P. Boys live dangerously in the womb. *American journal of human biology : the official journal of the Human Biology Council,* 22(3) (2010), pp. 330–335.

[10] Murji, A., Proctor, L. K., Paterson, A. D., Chitayat, D., Weksberg, R., and Kingdom, J. Male sex bias in placental dysfunction. *American Journal of Medical Genetics Part A,* 158A (2012), pp. 779–783.

[11] Vatten, L. J. and Skjaerven, R. Offspring sex and pregnancy outcome by length of gestation. *Early Human Development,* 76(1) (2004), pp. 47–54.

[12] Di Renzo, G., Rosati, A., Sarti, R., and Cruciani, L. Does fetal sex affect pregnancy outcome? *Gender medicine,* 4(1) (2007), pp. 19–30.

[13] Kleinrouweler, C. E., Uitert, M. van, Moerland, P. D., Ris-Stalpers, C., Post, J. A. M. van der, and Afink, G. B. Differentially expressed genes in the pre-eclamptic placenta: a systematic review and meta-analysis. *PloS one,* 8(7) (2013), e68991.

[14] Ballouz, S., Verleyen, W., and Gillis, J. Guidance for RNA-seq co-expression network construction and analysis: safety in numbers. *Bioinformatics (Oxford, England),* (2015).

[15] Consortium, I. H. G. S. Finishing the euchromatic sequence of the human genome. *Nature,* 431(7011) (2004), pp. 931–945.

Appendix A

Publication Format: Quantitative Allele-Specific Expression and DNA Methylation Analysis of *H19*, *IGF2* and *IGF2R* in the Human Placenta Across Gestation Reveals *H19* Imprinting Plasticity

OPEN ACCESS Freely available online

PLOS ONE

Quantitative Allele-Specific Expression and DNA Methylation Analysis of *H19, IGF2* and *IGF2R* in the Human Placenta across Gestation Reveals *H19* Imprinting Plasticity

Sam Buckberry[1], Tina Bianco-Miotto[1,2], Stefan Hiendleder[1,3], Claire T. Roberts[1]*

1 The Robinson Institute, Research Centre for Reproductive Health, School of Paediatrics and Reproductive Health, The University of Adelaide, Adelaide, South Australia, Australia, **2** The Robinson Institute, Research Centre for Early Origins of Health and Disease, School of Paediatrics and Reproductive Health, The University of Adelaide, Adelaide, South Australia, Australia, **3** JS Davies Epigenetics and Genetics Group, School of Animal and Veterinary Sciences, The University of Adelaide, Adelaide, South Australia, Australia

Abstract

Imprinted genes play important roles in placental differentiation, growth and function, with profound effects on fetal development. In humans, *H19* and *IGF2* are imprinted, but imprinting of *IGF2R* remains controversial. The H19 non-coding RNA is a negative regulator of placental growth and altered placental imprinting of *H19-IGF2* has been associated with pregnancy complications such as preeclampsia, which have been attributed to abnormal first trimester placentation. This suggests that changes in imprinting during the first trimester may precede aberrant placental morphogenesis. To better understand imprinting in the human placenta during early gestation, we quantified allele-specific expression for *H19, IGF2* and *IGF2R* in first trimester (6–12 weeks gestation) and term placentae (37–42 weeks gestation) using pyrosequencing. Expression of *IGF2R* was biallelic, with a mean expression ratio of 49:51 (SD = 0.07), making transient imprinting unlikely. Expression from the repressed *H19* alleles ranged from 1–25% and was higher ($P<0.001$) in first trimester (13.5±8.2%) compared to term (3.4±2.1%) placentae. Surprisingly, despite the known co-regulation of *H19* and *IGF2*, little variation in expression of the repressed *IGF2* alleles was observed (2.7±2.0%). To identify regulatory regions that may be responsible for variation in *H19* allelic expression, we quantified DNA methylation in the *H19-IGF2* imprinting control region and *H19* transcription start site (TSS). Unexpectedly, we found positive correlations ($P<0.01$) between DNA methylation levels and expression of the repressed *H19* allele at 5 CpG's 2000 bp upstream of the *H19* TSS. Additionally, DNA methylation was significantly higher ($P<0.05$) in first trimester compared with term placentae at 5 CpG's 39–523 bp upstream of the TSS, but was not correlated with *H19* repressed allele expression. Our data suggest that variation in *H19* imprinting may contribute to early programming of placental phenotype and illustrate the need for quantitative and robust methodologies to further elucidate the role of imprinted genes in normal and pathological placental development.

Citation: Buckberry S, Bianco-Miotto T, Hiendleder S, Roberts CT (2012) Quantitative Allele-Specific Expression and DNA Methylation Analysis of *H19, IGF2* and *IGF2R* in the Human Placenta across Gestation Reveals *H19* Imprinting Plasticity. PLoS ONE 7(12): e51210. doi:10.1371/journal.pone.0051210

Editor: Cees Oudejans, VU University Medical Center, The Netherlands

Received June 6, 2012; **Accepted** November 2, 2012; **Published** December 5, 2012

Funding: SB is supported by Australian Postgraduate Award and a Healthy Development Adelaide and Channel 7 Children's Research Foundation PhD scholarship. TBM is supported by a W. Bruce Hall Cancer Council of South Australia Fellowship. SH is a JS Davies Fellow. CTR is supported by a National Health and Medical Research Council (NHMRC) Senior Research Fellowship APP1020749. The research described herein was funded by NHMRC Research Project #565320 (http://www.nhmrc.gov.au/). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

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

Introduction

Genomic imprinting refers to parent-of-origin-dependent allele-specific gene expression. Imprinting affects gene dosage, with the imprinted allele considered repressed and functionally silenced [1,2]. Imprinting is largely, although not exclusively, observed in eutherian mammals and is thought to have arisen with viviparity and the evolutionary emergence of the chorioallantoic placenta [3,4]. The prevailing hypothesis on the origin of imprinting is based on paternal-maternal conflict and postulates that paternally expressed genes have been selected to maximize fetal resource acquisition from the mother, while maternally expressed genes have been selected to balance resource allocation to current and future offspring [4]. As imprinted genes appear to facilitate this *tug-of-war* between the maternal and paternal genomes, the conflict hypothesis predicts that imprinted genes are involved in fetal and placental growth and development during pregnancy [2,4,5].

Studies using animal models have demonstrated the functional importance of imprinting of *H19*, *IGF2* and *IGF2R* genes during intrauterine development [6,7,8,9,10]. Paternally expressed *IGF2* encodes the growth promoting insulin-like growth factor II, a potent mitogen involved in regulating cell proliferation, growth and development. The reciprocally imprinted, maternally expressed *H19* gene is located approximately 130 kb downstream of *IGF2* on human chromosome 11 and encodes a highly expressed, growth regulating, non-coding RNA that shares regulatory elements with *IGF2* [11]. The mechanism by which *H19* interacts with *IGF2* and regulates growth is not fully understood and

appears to involve long range interaction of differentially methylated regions and complex loop structures that regulate the activity of parental alleles [12,13,14]. More recently, *H19* has been identified as a trans regulator of an imprinted gene network for growth and development [15], apparently through miRNAs processed from the *H19* transcript [11,16,17]. The H19 large intergenic non-coding RNA (lincRNA) is highly expressed in extra-embryonic cell lineages and is a developmental reservoir of miR-675 that suppresses placental growth in the mouse [18]. The IGF2 receptor (*IGF2R*) mediates endocytosis and clearance or activation of a variety of ligands involved in the regulation of cell growth and motility, including insulin-like growth factor II [19,20,21].

Studies in mice have demonstrated that altered imprinting of *H19, IGF2* and *IGF2R* are associated with placental and fetal growth abnormalities [11,22,23], some of which are consistent with data from human studies. For example, (epi)mutations in the *H19-IGF2* region are associated with Silver-Russell and Beckwith-Wiedemann syndromes, which manifest *in utero* in severely growth-restricted and overgrowth phenotypes, respectively [24]. Furthermore, altered epigenetic regulation of the *H19-IGF2* region in human placenta has been associated with pregnancy complications such as preeclampsia, which are preceded by placental pathologies [25,26]. A significant role in placental development has been established for *H19* and *IGF2* in mouse and human, but knowledge on the role of *IGF2R* in human placental development is limited. The *IGF2R* gene is imprinted in all tissues except brain in mouse, but the majority of human samples indicate non-imprinted biallelic expression [3,27,28,29]. The minority of samples with imprinted or partially imprinted expression suggested developmental stage-specific transient imprinting. However, the developmental role of rare, transient or partial *IGF2R* imprinting in the human placenta [3,27,30,31,32,33] remains to be established.

In the human placenta, biallelic expression of imprinted genes, including *H19*, has been observed at higher rates during the first trimester of pregnancy compared to term [25,34,35]. Intriguingly, biallelic expression of *H19* in term placentae has been associated with preeclampsia in one study [25], yet subtle variation in *H19* allelic expression in healthy term placentae has also been observed [36]. This limited research on allele-specific expression in the human placenta suggested that imprinting may be dynamic across gestation with potential plasticity in imprinting beyond blastocyst and implantation stages. Although some differences in allele-specific expression of imprinted genes between the first trimester and term human placenta have been reported [34], there appear to be no studies addressing potential changes across the first trimester, a highly dynamic period of placental growth and differentiation. Thus, there is little or no data on temporal variation in imprinting of these genes across gestation, or if imprinting is stable throughout the first trimester and later gestation. In the present study, we quantified the allelic expression ratio for *H19, IGF2* and *IGF2R* and DNA methylation in the *H19-IGF2* imprinting control region across 6–12 weeks of gestation in first trimester placentae and in term placentae between 37–42 weeks of gestation.

Materials and Methods

Ethics Statement

Ethics approval was granted by the Children, Youth and Women's Health Service Research Ethics Committee (REC2249/2/13), the Central Northern Adelaide Health Service Ethics of Human Research Committee (Approval #2005082) and the University of Adelaide Human Research Ethics Committee (H-137-2006). Written informed consent was obtained from all patients.

Sample Collection

First trimester placental samples ranging from 6–12 weeks of gestation were obtained from elective terminations of pregnancies at the Women's and Children's Hospital, South Australia. The consulting physician determined gestational age by observation and the date of the last menstrual period. Placental villous samples were washed in sterile PBS and snap frozen in liquid nitrogen before being stored at -80°C. Term placenta samples were collected from pregnancies classified as being uncomplicated by using the criteria described in [37], and were collected and dissected post-delivery at the Lyell McEwin Health Service, South Australia, and incubated in RNAlater solution (Invitrogen) at 4°C for 24 hours before being stored at -80°C.

Genotyping

DNA was extracted from placental tissue and parental blood using the Qiagen® DNeasy® blood and tissue kit following the manufacturer's instructions. DNA concentration was determined using the NanoDrop® ND-1000 Spectrophotometer and diluted to 12.5 ng/μL with nuclease-free water (Mo Bio Laboratories). Isolated DNA from first trimester placental samples was genotyped for *IGF2* rs680, *IGF2R* rs998075 and *IGF2R* rs1570070 single nucleotide polymorphisms (SNPs) by PCR and High Resolution Melt (HRM) analysis (see Methods S1). Term placenta and parental DNA SNP genotypes for *H19* rs217727 and *IGF2* rs680 were determined by multiplex PCR and the Sequenom® MassARRAY® system, using the iPLEX® GOLD single base extension reaction on custom arrays at the Australian Genome Research Facility, Brisbane, Australia.

Quantification of Allele Specific Expression

Placental samples were thawed and homogenised with 1 mL TRIzol (Invitrogen) per 100 mg tissue. TRIzol (Invitrogen) extraction was performed according to the manufacturer's guidelines. RNase-free glycogen (Ambion) was added at 25 μg per 1 mL of TRIzol (Invitrogen) to aid in RNA visualisation. RNA samples were DNase treated using the TURBO DNA-free™ kit (Ambion) following the manufacturer's instructions for rigorous treatment. Following DNase treatment, 2 μL of RNA was subjected to PCR with DNA-specific primers (Table S1). The DNase treatment was determined to be effective if samples showed no amplification after 35 cycles. The concentration of DNase-treated RNA was calculated with the NanoDrop® ND-1000 Spectrophotometer.

First-strand cDNAs were synthesised from 500 ng DNase-treated RNA using the iScript™ cDNA Synthesis Kit (Bio-Rad), following the manufacturer's instructions. Reverse transcriptase was omitted for negative controls and aliquots of the master mix without added RNA were included in PCR experiments to rule out contamination. Following reverse transcription, cDNA was diluted 1:10 with nuclease-free water (Mo Bio Laboratories). Aliquots from five cDNA samples were pooled and serially diluted 5-fold for primer validation and PCR optimisation.

PCR primers flanking SNP regions and pyrosequencing primers were designed using the PSQ™ assay design software (Biotage™). Reverse primers featured 5′ biotin modifications and were HPLC purified. All oligonucleotides were synthesised by GeneWorks (Adelaide) and are listed in Table S2. Each sample was pyrosequenced in triplicate, with each replicate generated in an independent PCR cycling run. PCR was performed using 10 μL

reactions with 2 μL of cDNA, 5 μL SsoFast EvaGreen Supermix (Bio-Rad) and 300 nM of each primer. Cycling conditions were 2 min enzyme activation at 95°C followed by 40 cycles of 5 sec at 95°C and 20 sec at 60°C. PCR products were sequenced by pyrosequencing using the methods detailed below.

Quantification of DNA Methylation

DNA for methylation analysis was extracted from placental villous tissue by homogenizing 50–100 mg tissue in 500 μl of TES (10 mM Tris-HCL pH8.0, 1 mM EDTA, 100 mM NaCl), then adding 300 μg Proteinase K and 30 μl of 20% SDS followed by an overnight incubation at 37°C. Then 3 M NaCl was added to precipitate proteins and the supernatant was collected by centrifugation. The DNA was pelleted using 2 volumes of absolute ethanol and washed in 70% ethanol, air dried and resuspended in TE pH 8.0 [38].

Each DNA sample was bisulfite treated in triplicate by EpigenDx (Massachusetts, USA) using 500 ng of DNA and a proprietary bisulfite salt solution followed by incubation for 14 hours at 50°C. Bisulfite treated DNA was purified using Zymogen DNA columns and was eluted with 20 μl of TE pH 8.0, 1 μl of which was used for PCR reactions. The PCR was performed with 0.2 μM of each primer for EpigenDx methylation assays ADS025, ADS596FS and ADS004 with one of the PCR primers being biotinylated for purifying the final PCR product.

Pyrosequencing

The biotinylated PCR products were bound to Streptavidin Sepharose HP (Amersham Biosciences, Sweden), and the Sepharose beads containing the immobilized PCR product were purified, washed and denatured using a 0.2 M NaOH solution and rewashed all using the Pyrosequencing Vacuum Prep Tool (Qiagen) as recommended by the manufacturer. Then 0.2 μM pyrosequencing primer was annealed to the purified single-stranded PCR product. 10 μl of the PCR products were sequenced using the PSQ96 HS System (Biotage AB) following the manufacturer's instructions at EpigenDX Genome and Epigenome Research Facility (Massachusetts, USA). The status of each locus was analyzed individually using QCpG software (Qiagen).

Statistical Analysis

Repressed allele expression differences between gestational age classes for each gene were analyzed using one-way analysis of variance (ANOVA). Differences between first trimester (6–12 weeks of gestation) and term (37–42 weeks of gestation) samples were analyzed using the t-test. Differences in allelic expression measured at two loci in the one sample were analyzed using the paired t-test. The relationship between repressed allele expression from two genes in the same sample was tested by calculating the Pearson's bivariate correlation coefficient. Differences in levels of DNA methylation between first trimester and term samples were tested for each individual CpG site and for each region using independent t-tests. The relationship between repressed allele expression and mean DNA methylation for each region and CpG site was tested using Pearson's correlation. Results were considered significant at $P<0.05$. All statistical analyses were performed using GraphPad PRISM 5.0.

Results

Quantification of Allele-specific Gene Expression in the Human Placenta

DNA samples from placental tissue was genotyped for SNPs *H19* rs217727, *IGF2* rs680, *IGF2R* rs998075 and *IGF2*R rs1970070 to identify heterozygous individuals. Sixty-nine samples in total were heterozygous for at least one of the tested candidate SNPs. The number of heterozygotes identified for each gestational age class is summarized in Table 1. As parental DNA corresponding to term placenta samples was available for 28 cases, we genotyped maternal and paternal DNA for *H19* and *IGF2* polymorphisms to determine the parental origin of expressed alleles. In all cases with sufficient parental genotype information, *H19* was maternally expressed (n = 11) and *IGF2* (n = 9) was paternally expressed, as expected (Table S3).

Relative expression from each *H19*, *IGF2* and *IGF2R* allele was quantified in placenta samples by pyrosequencing of SNP loci. Relative allelic expression levels for *H19*, *IGF2* and *IGF2R* in first trimester and term placenta samples are presented in Figures 1 and 2 with each gene showing a unique allele expression profile. Technical replicates obtained from independent PCR reactions showed average standard deviations (SD) of 0.44% for *H19* rs217727, 1.02% for *IGF2* rs680, 3.14% for *IGF2R* rs998075 and 3.84% for *IGF2R* rs1570070, respectively, indicating robust assays with negligible inter-PCR variation. The greater standard deviation for *IGF2R* replicates is likely due to the higher PCR quantification cycle (Cq) required for data acquisition as compared with *H19* and *IGF2* (data not shown). The ratio of repressed allele to predominantly expressed allele is depicted in Figure 3, where a 0:100 ratio represents no expression from the repressed allele and a 50:50 ratio represents balanced, i.e., bialleleic, non-imprinted expression. Across first trimester gestational age classes, expression from the *H19* repressed allele shows notable inter-individual variation in contrast to the almost uniform monoallelic expression observed for *IGF2* (Figure 3). *IGF2R* allele-specific expression in first trimester placenta samples showed balanced expression, with some samples potentially showing a slight allelic bias (Figures 2, 3).

Table 1. Number of informative heterozygous samples for *H19*, *IGF2* and *IGF2*R for each gestational age class.

	Number of Heterozygotes		
Gestational Age (weeks)	***H19***	***IGF2***	***IGF2R***
6	1	1	1
7–8	6	5	14
9–10	4	6	8
11–12	2	2	1
37	1	0	NA
38	0	0	NA
39	7	5	NA
40	5	8	NA
41	5	6	NA
42	0	1	NA
Total	**31**	**34**	**24**

First trimester samples range from 6–12 weeks, term samples range from 37–42 weeks.

doi:10.1371/journal.pone.0051210.t001

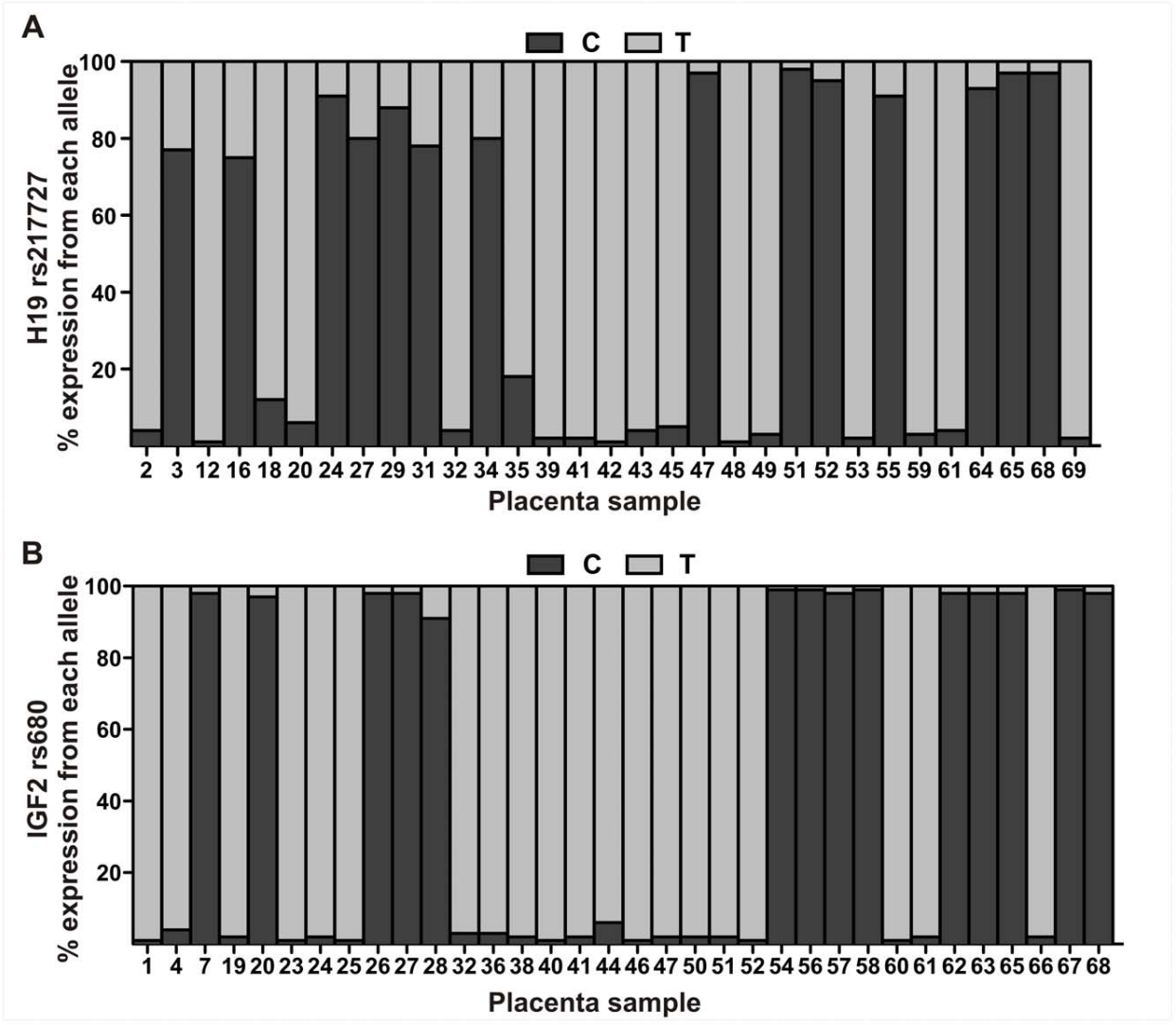


Figure 1. Relative expression from *H19* and *IGF2* alleles in the human placenta. Shaded bars for (A) *H19* and (B) *IGF2* represent the proportion of expression (%) from each allele. Samples 1–38 are from first trimester (6–12 weeks of gestation) placentae and samples 39–69 are from term (37–42 weeks of gestation) placentae.
doi:10.1371/journal.pone.0051210.g001

Biallelic Expression of *IGF2R* in the First Trimester Placenta

Allelic expression ratios for *IGF2R* in first trimester placenta was measured at two SNP loci (rs998075 n = 16 and rs1570070 n = 13). Five samples were heterozygous for both SNPs, and no significant difference (paired *t*-test, $P = 0.42$) was detected between the expression ratios for the two SNP loci, indicating that both polymorphisms were equivalent in quantifying allele-specific expression (Figure 2C). All heterozygous *IGF2R* samples were therefore combined for analyses, and, when expression was quantified at both loci in one sample, the average allelic ratio of the two loci was used. The results clearly show biallelic *IGF2R* expression in all first trimester placental samples assessed (Figures 2A, B), with a mean allele expression ratio of 49:51 at both SNP loci (rs998075 SD = 7.1%, rs1570070 SD = 6.9%) with expression ratios ranging from 36:64 to 49:51 (Figure 3). These SNP based *IGF2R* pyrosequencing results provide no evidence for *IGF2R* imprinted expression in the first trimester placenta and thus confirm the non-imprinted status of *IGF2R* throughout gestation.

Increased Expression from the *H19* Repressed Allele is Higher in First Trimester Placenta

Expression from the *H19* repressed allele was quantified in 13 first trimester placenta samples obtained at 6–12 weeks of gestation (Figure 1A). Mean expression from the repressed allele was 13.5% (SD±8.2; Figure 1A) and ranged from 0.9–24.7% (Figure 3). Expression of the *H19* repressed allele appeared to decrease with gestational age in the first trimester samples (Figure 4A), but we found no significant differences between first trimester gestational age classes. To further test the hypothesis that expression from the repressed *H19* allele decreases across gestation, we then quantified allelic expression in term placenta samples obtained between 37–42 weeks of gestation (n = 18). Expression from the repressed *H19* allele at term was significantly lower ($P<0.001$) than the level of expression

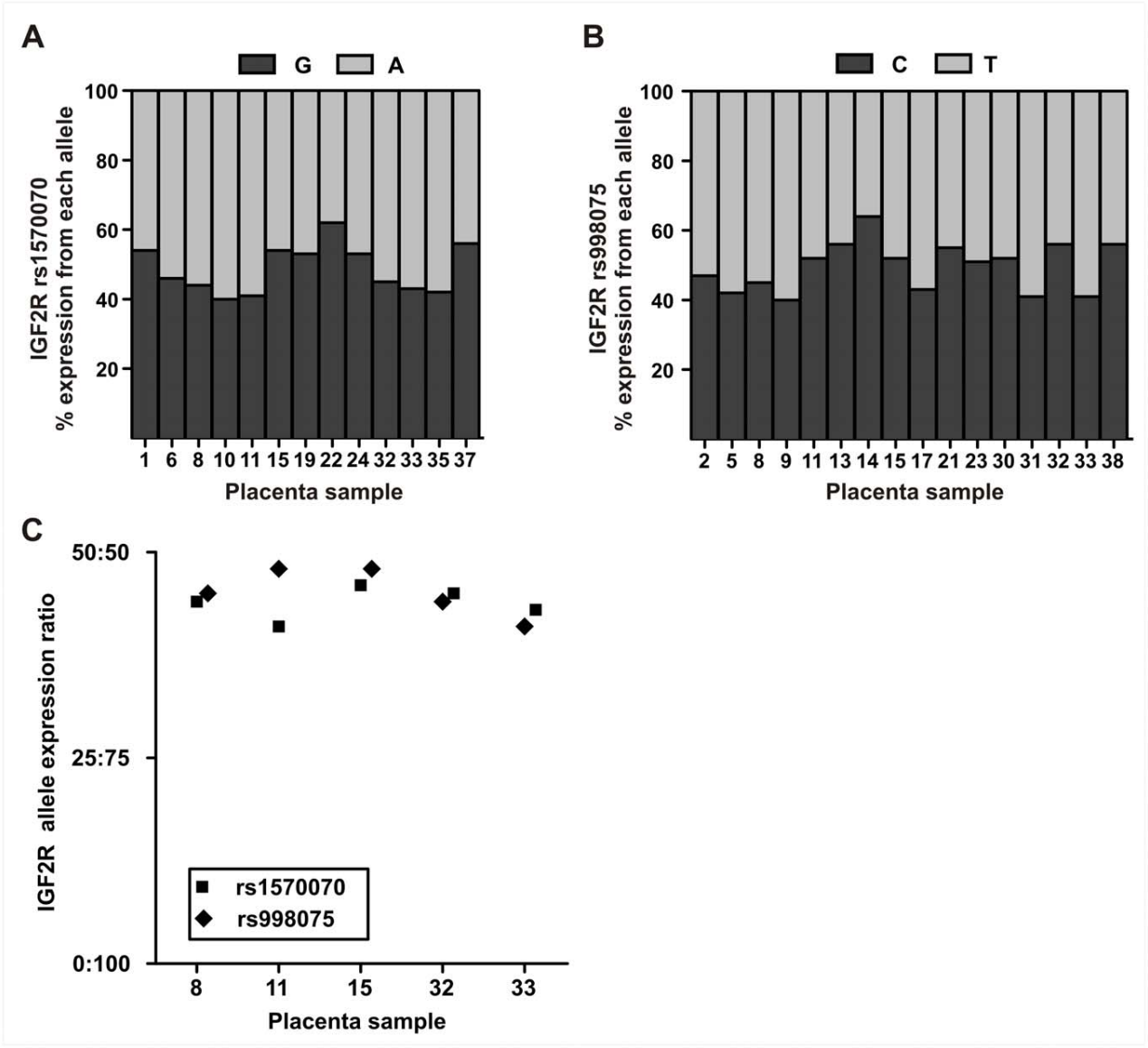


Figure 2. Relative expression from *IGF2R* alleles in the human placenta. Shaded bars for (A) *IGF2R* rs1570070 and (B) *IGF2R* rs998075 represent the proportion of expression (%) in first trimester placentae. (C) Allelic expression ratios for *IGF2R* measured two SNP loci in the same sample. These paired samples indicate both SNP loci are equivalent (paired *t*-test, $P = 0.42$) for evaluating *IGF2R* allele-specific expression.
doi:10.1371/journal.pone.0051210.g002

observed in first trimester placenta samples (Figure 4B and Table 2).

Expression from the *IGF2* repressed allele contributed on average 2.7% (SD 2.1%, n = 34) to total *IGF2* transcript in placenta samples (Figure 1B). No significant differences in expression from the *IGF2* repressed allele were observed between first trimester gestational age classes (ANOVA $P > 0.05$) or between first trimester and term (Table 2, Figures 4C, D) placentae. These results show that imprinted *IGF2* expression is tightly regulated and stable across gestation.

Eleven placenta samples were heterozygous for both *H19* and *IGF2* polymorphisms, which allowed us to test for a correlation in repressed allele expression for these adjacent co-regulated imprinted genes. We found that expression from the *H19* repressed allele was not correlated ($P = 0.88$, $r = 0.54$) with expression from the *IGF2* repressed allele (Figure 5).

Locus-specific DNA Methylation Differences in the *H19-IGF2* Region between First Trimester and Term Placentae

To investigate if DNA methylation levels at specific CpG's are correlated with *H19* repressed allele expression, we quantified methylation levels in three regions (Figure 6) using bisulfite pyrosequencing. The two regions upstream of the transcription start site (TSS) (regions 1 and 2 on Figure 6) were selected as they cover or are directly adjacent to sites that have been shown to be differentially methylated [39], and region 3 (Figure 6) was selected as it spanned the *H19* promoter region and the TSS.

The first region (denoted 1 in Figure 6) encompassed five CpG sites with a mean methylation level of $54.7 \pm 7.8\%$, which would be

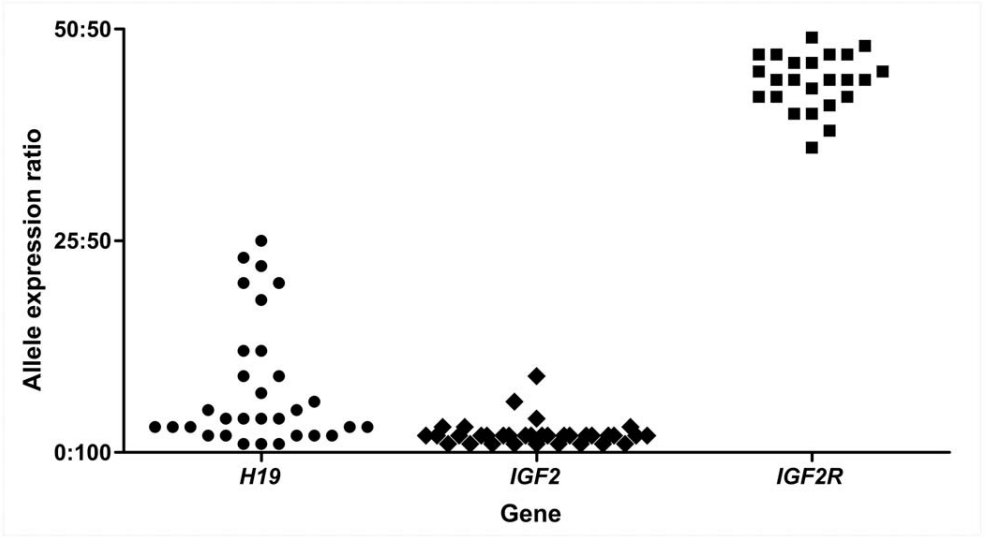


Figure 3. Allelic expression ratios for *H19, IGF2* and *IGF2R* in the human placenta. The 50:50 ratio represents equal expression from both alleles and 0:100 ratio represents expression exclusively from one allele. Each point on the graph represents the allelic expression ratio measured in an individual placental sample. *H19* and *IGF2* samples are from first trimester and term placentae, *IGF2R* samples are all from first trimester placentae.
doi:10.1371/journal.pone.0051210.g003

expected at a differentially methylated imprinted locus. In region 1, there was no significant difference in mean methylation levels between first trimester (54.8±6.9%) and term (53.5±7.2%) placentae or at any of the 5 individual CpG sites (Figure 6, Table S4). The second region assessed (denoted 2 in Figure 6), covered 12 CpG sites which showed overall hypomethylation, with a mean methylation level of 30.9±3.9% in first trimester and 28.9±5.3% in term placentae. When analyzed independently, 4 of the 12 CpG sites, 3 of which are adjacent to each other, showed significantly higher methylation in first trimester placentae in comparison to term placentae (Figure 6). The third region that spanned the *H19* TSS showed mean methylation levels of 16.1±3.1% in first trimester placentae and 15.5±3.5% in term placentae. When each CpG site was analyzed individually, the cytosine nucleotide 39 bp upstream from the *H19* TSS (Figure 6) showed significantly higher methylation ($P = 0.02$) in first trimester placentae (15.2±3.6% vs 12.0±3.2%). Details of DNA methylation levels in first trimester and term placentae at each individual CpG site and the statistical comparisons between the groups are listed in Table S4.

H19 Repressed Allele Expression is Correlated with Higher Levels of DNA Methylation

As distinct variation in expression from the *H19* repressed allele in first trimester placentae was observed, we tested for correlations between the level of repressed allele expression and levels of DNA methylation at CpG's of first trimester placentae. In region 1, a significant positive correlation ($P<0.001$, $r = 0.65$) was observed between repressed allele expression and the mean methylation level across the region (Figure 7A). When each of the 5 CpG sites in this region were analyzed independently for the same correlation, the results remained significant for each site (Table S5). This correlation was not observed for region 2 (Figure 7B, $P = 0.36$, $r = 0.07$) or 3 (Figure 7C, $P = 0.47$, $r = 0.05$) or for any individual CpG sites within these regions (Table S5).

Discussion

Imprinted genes are known to be critically involved in placental development and function. Aberrant patterns of imprinted gene expression are implicated in pregnancy complications such as preeclampsia and intrauterine growth restriction [5,40,41,42,43]. Although the symptoms of these conditions manifest late in pregnancy, their pathogenesis is commonly attributed to compromised first trimester placental development [44]. Previous research on genomic imprinting in the human placenta has focused on the term placenta [25,32,36,41,45,46] and data during the first trimester of gestation is limited [27,34,35,47]. In the present study, we investigated the imprinting status (i.e., allele-specific expression) of three genes, *H19*, *IGF2* and *IGF2R*, which have known, but poorly understood, associations with pregnancy complications and placental abnormalities in humans and/or animal models [6,7,8,9,10]. We assessed allele-specific expression of these genes and DNA methylation in the *H19-IGF2* imprinting control region in first trimester (6–12 weeks of gestation) and term (37–42 weeks of gestation) placentae.

We assessed *IGF2R* allele-specific expression, as the imprinting status of this important gene for prenatal growth and development remains controversial in human. We observed balanced expression from both *IGF2R* alleles, and although we did not investigate any potential imprinting mechanisms for this gene, these results suggest *IGF2R* is not imprinted in the first trimester placenta. Imprinting of *IGF2R* has been suggested to be a polymorphic trait in humans, with a small proportion of individuals showing monoallelic expression or partial imprinting [3,27,33]. In this study, we assessed more informative samples than previous studies [3,27,33,48] but found no evidence for polymorphic *IGF2R* imprinting in the placenta. Although we observed overall a balanced expression of alleles for *IGF2R*, individual allelic expression ratios ranged from 36:64 to 49:51. This variation may reflect what has been described previously as partial repression or allelic preference [27,33]. It is presently unclear if this subtle imbalance of *IGF2R* allelic expression is due to genetic

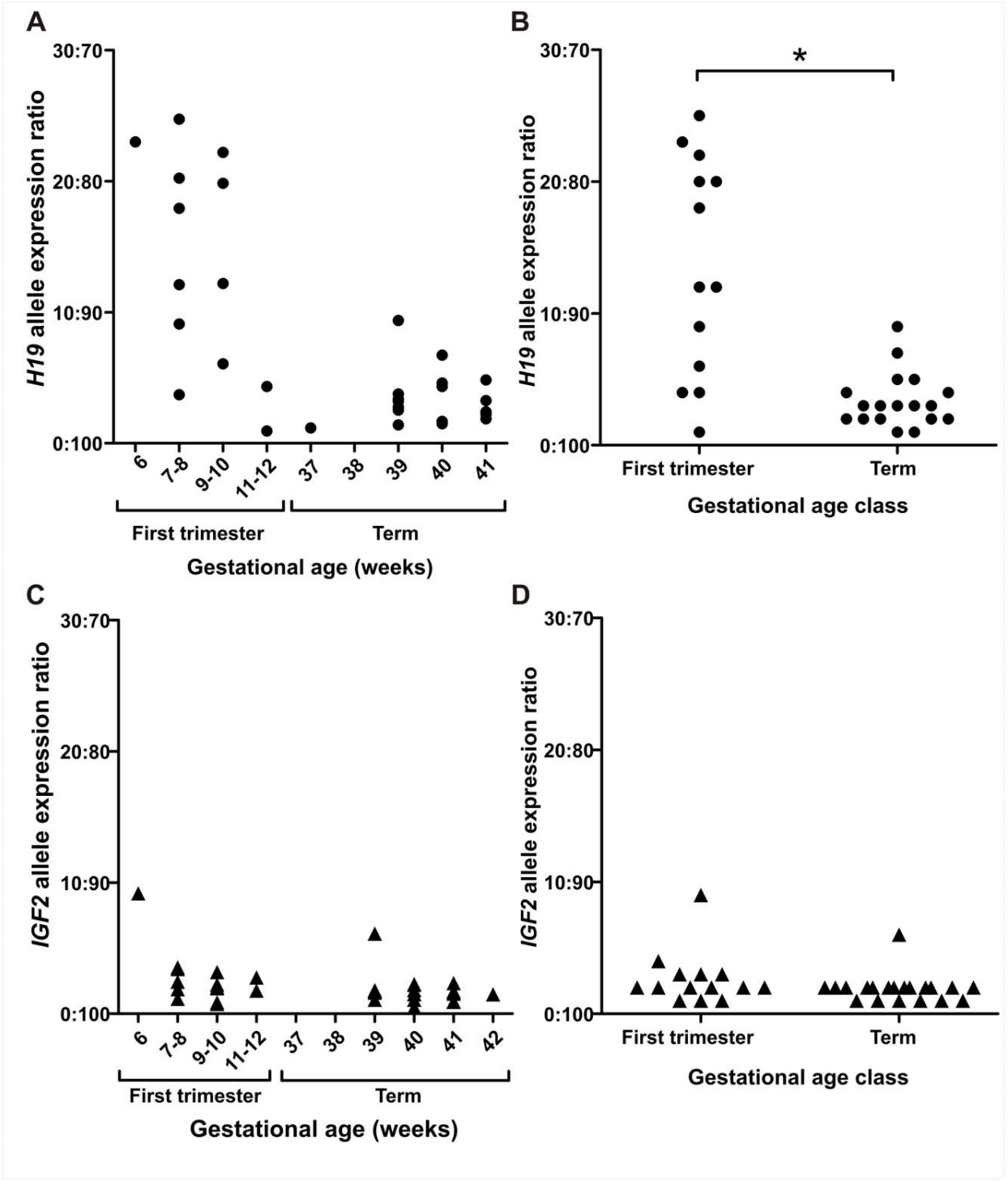


Figure 4. Ratio of expression from each allele in human first trimester and term placentae measured by pyrosequencing. Each point on the graph represents the allelic expression ratio observed in an individual placental sample. (**A**) *H19* allelic expression ratio for each gestational age class. (**B**) Expression from the *H19* repressed allele is significantly higher (*$P<0.001$) in first trimester placental samples. (**C-D**) *IGF2* allelic expression ratios are similar for each gestational age class (**C**) with no significant difference (**D**) between first trimester and term. First trimester samples are 6–12 weeks of gestation term samples are 37–42 weeks of gestation.
doi:10.1371/journal.pone.0051210.g004

variation in allele-specific epigenetic regulation or a parent-of-origin effect.

Allele-specific expression of *H19* showed considerable inter-individual variation, with expression from the repressed (i.e. imprinted) allele contributing up to 25% of total *H19* transcript

Table 2. Relative levels of repressed allele expression in human first trimester and term placentae.

	% Repressed allele expression		
	First trimester	**Term**	**P value**
H19	13.5±8.3	3.4±2.0	<0.0001
IGF2	2.6±2.0	1.9±1.1	0.1734
IGF2R	43.8±3.2	ND	–

ND = Not Determined.
doi:10.1371/journal.pone.0051210.t002

in the first trimester placenta. In contrast, *IGF2* showed predominantly monoallelic expression and little variation between individuals, with one allele contributing more than 90% of *IGF2* transcript in all investigated samples. This indicated that *IGF2* allele-specific expression is tightly regulated in the first trimester placenta and suggests that *IGF2* imprinting is established early in development and remains stable throughout gestation.

Determining loss of imprinting or biallelic expression of imprinted genes was previously performed by restriction fragment length polymorphism (RFLP) analysis. This method provides a qualitative or semi-quantitative assessment of monoallelic or biallelic expression. In human placentae from uncomplicated pregnancies, *H19* RFLP data showed biallelic expression before 10 weeks of gestation and imprinted expression at term [25,35]. However, term placentae from preeclamptic pregnancies were reported to display biallelic expression with the RFLP method [25]. This biallelic *H19* expression could indicate a failure to establish correct *H19* imprinting with downstream effects on placental development [25,35]. The data presented in the current study show that *H19* expression from the imprinted, i.e. repressed, allele can range from 9–22% at 9–10 weeks of gestation, highlighting the potential ambiguity in classifying expression as mono- or biallelic by less sensitive methods. Our data support the view [32,49] that classification of genes as imprinted or non-imprinted by qualitative methods may be a less meaningful distinction than quantitative measurements of imprinting status based on precise estimates of relative contributions from each allele.

More recently, quantitative PCR and pyrosequencing have been used to evaluate allele-specific expression in placental tissue. By using these highly sensitive methodologies, expression from the "silenced", imprinted, alleles has been generally higher in first trimester placentae [46] with some variation at term [36]. Both the RFLP assay and quantitative allele-specific expression approaches support the concept that repressed allele expression changes through gestation in the placenta, particularly during early pregnancy [25,34,35,46]. Using placental tissue from 6–12 weeks of gestation, we tested the hypothesis that imprinted allele-specific expression changes during the first trimester of pregnancy. We found no significant differences between early and late first trimester allelic expression ratios for *H19*, *IGF2* or *IGF2R*. Although we quantified allelic expression ratios using a highly sensitive technique, the method used for classifying gestational age, our sample size, and the proportion of heterozygotes in each group may have prevented the detection of significant changes across first trimester age groups. When comparing first trimester and term placenta samples for *H19*, we found a significant decrease in the proportion of repressed allele expression at term. Furthermore, these results for *H19* show notable inter-individual variation early

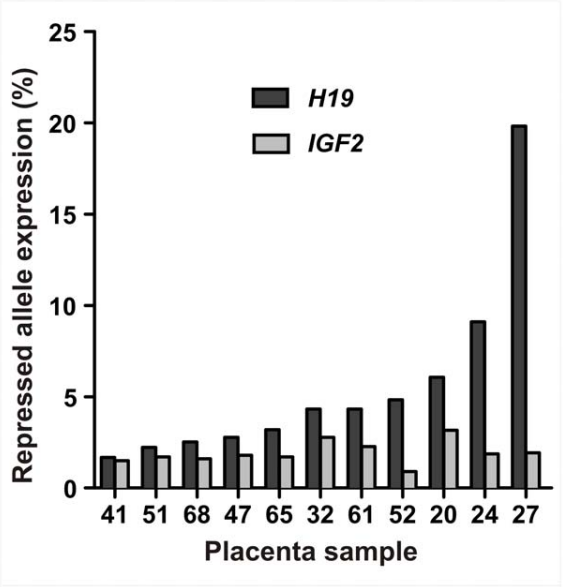


Figure 5. Relative level of expression from repressed alleles in samples heterozygous for both *H19* rs217727 and *IGF2* rs680. Graph shows increased expression from the *H19* repressed allele is not correlated ($P = 0.09$, $r = 0.54$) with expression from the *IGF2* repressed allele.
doi:10.1371/journal.pone.0051210.g005

during placental development, and more uniformity in allelic expression ratios as gestation progresses. This is a clear demonstration of dynamic change in imprinting status well beyond the blastocyst and implantation stages. However, an alternative explanation for the observed differences in *H19* allelic expression ratios between first trimester and term samples in the present study could be the unbiased sampling of material from elective terminations of pregnancy versus the selected material at term that came from normal pregnancies only. Placental tissue from elective terminations of pregnancy in first trimester will, by necessity, include those from pregnancies that may have been destined to develop a pregnancy complication e.g. preeclampsia, preterm labour or intrauterine growth restriction which are typically diagnosed later in gestation. Potentially, first trimester placental samples exhibiting expression from the repressed allele may have been destined to retain biallelic expression and associate with preeclampsia later. However, we consider this unlikely given that 8 out of 13 first trimester samples had greater than 10% expression from the repressed allele and preeclampsia occurs in just 8% of women in the community where our samples were collected [50].

Our results show *H19* expression from the repressed allele is not correlated with expression from the *IGF2* repressed allele in the same samples. The prevailing regulatory model of the *H19-IGF2* region based on differential DNA methylation predicts that both genes are not expressed from a single chromosome. Although this model is supported by considerable evidence [12,51,52] (and references cited therein), there is also evidence to suggest that this model may be insufficient (reviewed in [53]). The data presented here show higher expression from the repressed *H19* allele is not correlated with any change in *IGF2* repressed allele expression in individual placentae. Additionally, we show that DNA methylation levels at CpG sites (1946–2005 bp upstream of the *H19* TSS) that flank the 6th CTCF binding domain [39], are positively correlated with the level of expression from the *H19* repressed allele, which

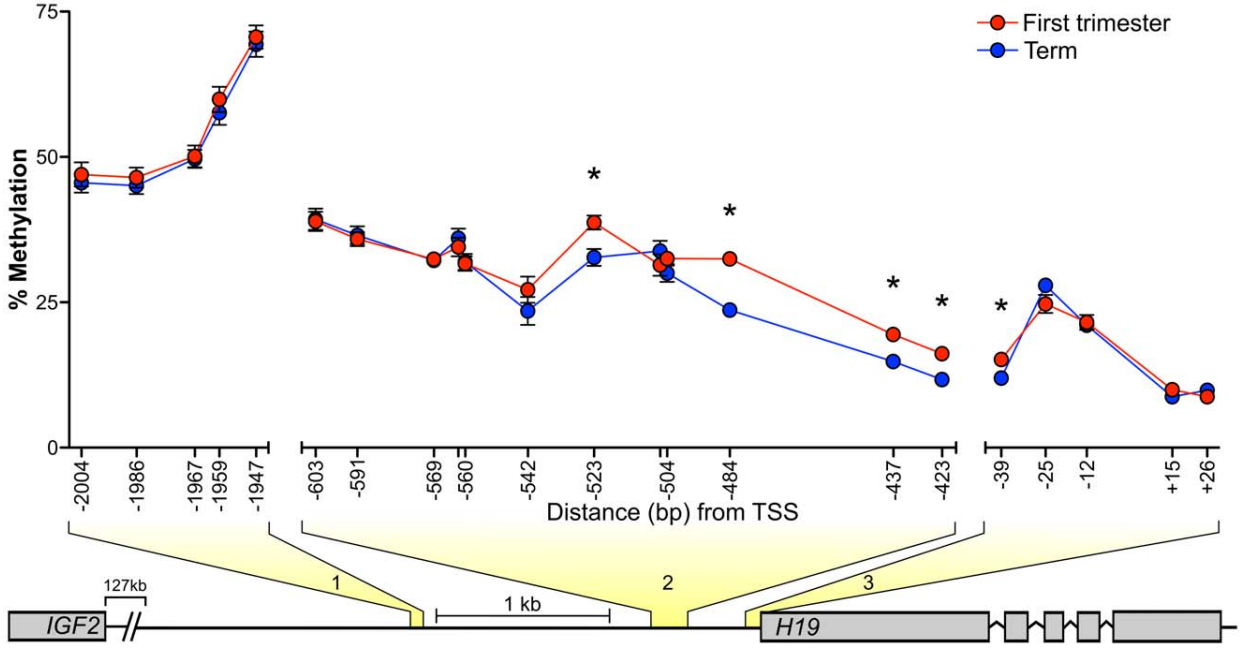


Figure 6. Placental methylation levels in regions upstream and covering the *H19* transcription start site (TSS). Each genomic region where DNA methylation was measured is highlighted in yellow and numbered 1–3. DNA methylation levels for individual CpG loci are shown for first trimester (red circles, n = 13) and term (blue circles, n = 18) placentae. Distance (bp) of cytosine nucleotide from *H19* TSS is represented on x-axis. Each data point represents the mean methylation level for the gestational age class. * Indicates a significant difference in methylation levels between first trimester and term placentae at individual CpG sites. Error bars represent SEM and when not present SEM was too low to depict on the graph. The schematic representation below the graph highlights the regions between *H19* and *IGF2* where bisulfite DNA pyrosequencing was performed. Region 1 covers 5 CpG sites (Chr11:2021011–2021070), region 2 covers 12 CpG sites (Chr11:2021011–2021070) and region 3 covers 5 CpG sites (Chr11:2019079–2019145). Genomic coordinates refer to reference assembly GRCh37/hg19.
doi:10.1371/journal.pone.0051210.g006

was unexpected given the prevailing regulatory model. Furthermore, we observed significantly higher DNA methylation in first trimester placentae in the region 422–524 bp upstream of the *H19* TSS that surrounds the differentially methylated region (DMR) [39], despite finding no correlation with *H19* repressed allele expression. This suggests DNA methylation in the DMR decreases

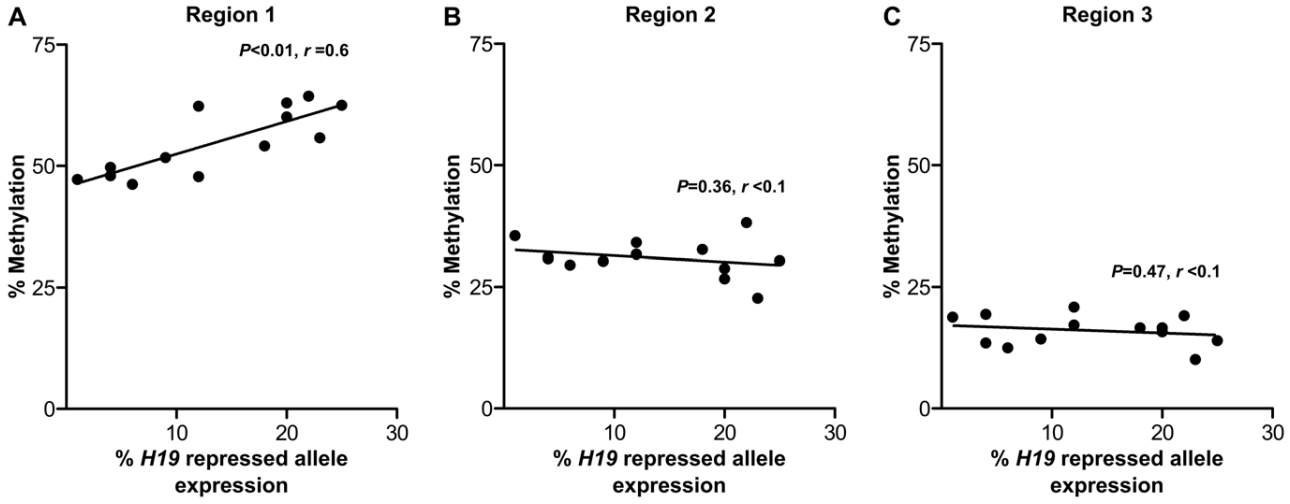


Figure 7. Levels of *H19* repressed allele expression and DNA methylation in human first trimester placentae. (A) Increased expression from the repressed *H19* allele is positively correlated ($P = 0.0016$, $r = 0.61$) with increased DNA methylation in region 1. (B & C) *H19* repressed allele expression is not correlated with DNA methylation in region 2 ($P = 0.3626$, $r = 0.08$) or region 3 ($P = 0.4791$, $r = 0.04$). Each point on the graph represents individual first trimester placenta samples. Methylation levels in each region represent the average methylation from 5 CpG sites in region 1 (Chr11:2021011–2021070), 12 CpG sites in region 2 (Chr11:2021011–2021070) and 5 CpG sites in region 3 (Chr11:2019079–2019145). Genomic coordinates refer to reference assembly GRCh37/hg19.
doi:10.1371/journal.pone.0051210.g007

progressively throughout gestation with no effect on *H19* allelic regulation. Together, these findings suggest that the methylation dependant enhancer competition model of the *H19-IGF2* locus may not fully explain the patterns of allele-specific expression observed for these genes in the early human placenta, as suggested previously [53]. However, although we assessed DNA methylation at sites within the *H19-IGF2* regulatory region, we did not assess methylation across all the CTCF binding sites upstream of *H19*. Moreover, we did not investigate additional regulatory mechanisms, such as the actions of other non-coding RNA's and repressive histone modifications that are involved in placental-specific imprinting [54,55,56,57]. Therefore we are unable to rule out other mechanistic changes that may be influencing *H19* allele specific expression.

An important consideration when using placental tissue for studying genomic imprinting is that this organ arises from multiple extra-embryonic and embryonic cell lineages. Cells descended from both the inner cell mass and trophectoderm may show major epigenetic differences [58], and as a result, analysis of whole placental villous tissue may not identify cell lineage-specific imprinting effects. In this study, we show a clear imprinting effect for *IGF2* in all heterozygous first trimester placenta samples, which suggests that all cell types composing the placental villi had *IGF2* imprinting mechanisms in place. However, for *H19* we observed notable inter-individual variation in expression from the imprinted allele. This variation could be due to the heterogeneous nature of the placental villous tissue sampled and *H19* lineage-specific imprinting at the single cell level. Cell-specific imprinted gene expression has been proposed as an *all or none* phenomenon in placental cell lines [59], and *H19* biallelic expression has been shown to be specific to extravillous cytotrophoblast cells [47], suggesting there is no intermediate imprinting effect at the single cell level. Therefore, observing variations in relative expression from the imprinted allele in placental tissue may simply reflect the fraction of cells with complete biallelic expression [59]. As first trimester placental tissue sampling is expected to yield a higher proportion of extravillous cytotrophoblast cells than those collected at term, changes in the level of imprinting across gestation may reflect proportional changes in cell lineage populations as the placenta differentiates. This suggests future studies of placental imprinting dynamics should consider the potential influence of placental cell type heterogeneity.

The *H19, IGF2* and *IGF2R* genes have key roles in placental development, yet the phenotypic effect of their allele-specific expression across gestation remains unknown. However, the role of *H19* as a regulator of the recently described imprinted gene network suggests potentially significant phenotypic effects [15]. This may depend on differences in gene dosage, but could also involve more complex regulatory effects in *trans*. To date, the normal developmental patterns of imprinted gene expression in the human placenta are poorly understood. As altered patterns of imprinting in term placentae are associated with pregnancy complications, identifying when these abnormal patterns are established may aid in elucidating the origins of placental abnormalities implicated in their aetiology. Our results highlight the requirement for robust and sensitive methods to determine the role of imprinted allele-specific expression in placentae from complicated pregnancies. Undoubtedly, precise methods and comprehensive studies will be required to progress towards understanding the molecular basis of potentially life threatening pregnancy complications in which defective placentation is implicated.

Supporting Information

Table S1 Genomic DNA specific primers used to detect DNA contamination in RNA samples.
(PDF)

Table S2 Details of genes, SNP regions and primers used for quantifying allele-specific expression by pyrosequencing.
(PDF)

Table S3 Parental and placental genotypes with placental allele expression ratio for *H19* and *IGF2*.
(PDF)

Table S4 Comparison of DNA methylation levels at individual CpG loci in 3 regions upstream and surrounding the *H19* transcription start site in human first trimester and term placentae.
(PDF)

Table S5 Pearson's correlation of *H19* repressed allele expression and DNA methylation levels at individual CpG loci in human first trimester placentae.
(PDF)

Methods S1 Genotyping single nucleotide polymorphisms by PCR and High Resolution Melt (HRM) analysis.
(PDF)

Acknowledgments

We thank Liying Yan, Ann Meyer and Matthew Poulin of EpigenDx for their pyrosequencing technical assistance and expertise. The authors would also like to thank all members of the Claire T Roberts Placental Development Laboratory for helpful discussions.

Author Contributions

Conceived and designed the experiments: SB TBM SH CTR. Performed the experiments: SB TBM. Analyzed the data: SB TBM. Contributed reagents/materials/analysis tools: CTR. Wrote the paper: SB TBM SH CTR.

References

1. Ferguson-Smith AC, Surani MA (2001) Imprinting and the epigenetic asymmetry between parental genomes. Science 293: 1086.
2. Reik W, Walter J (2001) Genomic imprinting: parental influence on the genome. Nat. Rev. Genet. 2: 21–32.
3. Monk D, Arnaud P, Apostolidou S, Hills FA, Kelsey G, et al. (2006) Limited evolutionary conservation of imprinting in the human placenta. Proc. Natl. Acad. Sci. 103: 6623–6628.
4. Haig D (1996) Altercation of generations: genetic conflicts of pregnancy. Am. J. Reprod. Immunol. 35: 226.
5. Haig D (1993) Genetic Conflicts in Human Pregnancy. Q Rev. Biol. 68: 495–532.
6. Coan PM, Fowden AL, Constancia M, Ferguson-Smith AC, Burton GJ, et al. (2008) Disproportional effects of Igf2 knockout on placental morphology and diffusional exchange characteristics in the mouse. J Physiol 586: 5023–5032.
7. Constancia M, Angiolini E, Sandovici I, Smith P, Smith R, et al. (2005) Adaptation of nutrient supply to fetal demand in the mouse involves interaction between the Igf2 gene and placental transporter systems. Proc. Natl. Acad. Sci. 102: 19219–19224.
8. Leighton PA, Ingram RS, Eggenschwiler J, Efstratiadis A, Tilghman SM (1995) Disruption of imprinting caused by deletion of the H19 gene region in mice. Nature 375: 34–39.
9. Sibley CP, Coan PM, Ferguson-Smith AC, Dean W, Hughes J, et al. (2004) Placental-specific insulin-like growth factor 2 (Igf2) regulates the diffusional

exchange characteristics of the mouse placenta. Proc. Natl. Acad. Sci. 101: 8204–8208.
10. Wang ZQ, Fung MR, Barlow DP, Wagner EF (1994) Regulation of embryonic growth and lysosomal targeting by the imprintedIgf2/Mpr gene. Nature 372: 464–467.
11. Gabory A, Ripoche MA, Yoshimizu T, Dandolo L (2006) The *H19* gene: regulation and function of a non-coding RNA. Cytogenet. Genome Res. 113: 188–193.
12. Murrell A, Heeson S, Reik W (2004) Interaction between differentially methylated regions partitions the imprinted genes Igf2 and H19 into parent-specific chromatin loops. Nat. Genet. 36: 889–893.
13. Qiu X, Vu TH, Lu Q, Ling JQ, Li T, et al. (2008) A complex deoxyribonucleic acid looping configuration associated with the silencing of the maternal Igf2 allele. Mol. Endocrinol. 22: 1476–1488.
14. Kurukuti S, Tiwari VK, Tavoosidana G, Pugacheva E, Murrell A, et al. (2006) CTCF binding at the H19 imprinting control region mediates maternally inherited higher-order chromatin conformation to restrict enhancer access to Igf2. Proc. Natl. Acad. Sci. 103: 10684–10689.
15. Gabory A, Jammes Hln, Dandolo L (2010) The H19 locus: role of an imprinted non-coding RNA in growth and development. BioEssays 32: 473–480.
16. Cai X, Cullen BR (2007) The imprinted H19 noncoding RNA is a primary microRNA precursor. RNA 13: 313–316.
17. Steck E, Boeuf S, Gabler J, Werth N, Schnatzer P, et al. (2012) Regulation of H19 and its encoded microRNA-675 in osteoarthritis and under anabolic and catabolic in vitro conditions. J. Mol. Med. 90: 1185–1195.
18. Keniry A, Oxley D, Monnier P, Kyba M, Dandolo L, et al. (2012) The H19 lincRNA is a developmental reservoir of miR-675 that suppresses growth and Igf1r. Nat Cell Biol 14: 659–665.
19. Harris LK, Crocker IP, Baker PN, Aplin JD, Westwood M (2011) IGF2 Actions on Trophoblast in Human Placenta Are Regulated by the Insulin-Like Growth Factor 2 Receptor, Which Can Function as Both a Signaling and Clearance Receptor. Biol. Reprod. 84: 440–446.
20. Ghosh P, Dahms NM, Kornfeld S (2003) Mannose 6-phosphate receptors: new twists in the tale. Nat Rev Mol Cell Biol. 4: 202–212.
21. Hawkes C, Kar S (2004) The insulin-like growth factor-II/mannose-6-phosphate receptor: structure, distribution and function in the central nervous system. Brain Res Rev. 44: 117–140.
22. Lau M, Stewart C, Liu Z, Bhatt H, Rotwein P, et al. (1994) Loss of the imprinted IGF2/cation-independent mannose 6-phosphate receptor results in fetal overgrowth and perinatal lethality. Genes Dev.8: 2953.
23. Constancia M, Hemberger M, Hughes J, Dean W, Ferguson-Smith A, et al. (2002) Placental-specific IGF-II is a major modulator of placental and fetal growth. Nature 417: 945–948.
24. Hirasawa R, Feil R (2010) Genomic imprinting and human disease. Essays Biochem 48: 187–200.
25. Yu L, Chen M, Zhao D, Yi P, Lu L, et al. (2009) The H19 Gene Imprinting in Normal Pregnancy and Pre-eclampsia. Placenta 30: 443–447.
26. Gao W, Li D, Xiao Z, Liao Q, Yang H, et al. (2011) Detection of global DNA methylation and paternally imprinted H19 gene methylation in preeclamptic placentas. Hypertens. Res. 1–7: 655–661.
27. Oudejans CBM, Westerman B, Wouters D, Gooyer S, Leegwater PAJ, et al. (2001) Allelic IGF2R Repression Does Not Correlate with Expression of Antisense RNA in Human Extraembryonic Tissues. Genomics 73: 331–337.
28. Hu JF, Balaguru KA, Ivaturi RD, Oruganti H, Li T, et al. (1999) Lack of reciprocal genomic imprinting of sense and antisense RNA of mouse insulin-like growth factor II receptor in the central nervous system. Biochem. Biophys. Res. Commun. 257: 604–608.
29. Yamasaki Y, Kayashima T, Soejima H, Kinoshita A, Yoshiura K-I, et al. (2005) Neuron-specific relaxation of Igf2r imprinting is associated with neuron-specific histone modifications and lack of its antisense transcript Air. Hum. Mol. Genet. 14: 2511–2520.
30. Ogawa O, McNoe LA, Eccles MR, Morison lM, Reeve AE (1993) Human insulin-like growth factor type I and type II receptors are not imprinted. Hum. Mol. Genet. 2: 2163–2165.
31. Kalscheuer VM, Mariman EC, Schepens MT, Rehder H, Ropers HH (1993) The insulin–like growth factor type–2 receptor gene is imprinted in the mouse but not in humans. Nat. Genet. 5: 74–78.
32. Daelemans C, Ritchie ME, Smits G, Abu-Amero S, Sudbery IM, et al. (2010) High-throughput analysis of candidate imprinted genes and allele-specific gene expression in the human term placenta. BMC Genet. 11: 25.
33. Xu YQ, Goodyer CG, Deal C, Polychronakos C (1993) Functional polymorphism in the parental imprinting of the human IGF2R gene. Biochem. Biophys. Res. Commun. 197: 747–754.
34. Pozharny Y, Lambertini L, Ma Y, Ferrara L, Litton CG, et al. (2010) Genomic loss of imprinting in first-trimester human placenta. Am. J. Obstet. Gynecol. 202: 391.e391–391.e398.
35. Jinno Y, Ikeda Y, Yun K, Maw M, Masuzaki H, et al. (1995) Establishment of functional imprinting of the H19 gene in human developing placentae. Nat. Genet. 10: 318–324.
36. Tabano S, Colapietro P, Cetin I, Grati FR, Zanutto S, et al. (2010) Epigenetic modulation of the IGF2/H19 imprinted domain in human embryonic and extra-embryonic compartments and its possible role in fetal growth restriction. Epigenetics 5: 313–324.
37. McCowan LME, Roberts CT, Dekker GA, Taylor RS, Chan EHY, et al. (2010) Risk factors for small-for-gestational-age infants by customised birthweight centiles: data from an international prospective cohort study. BJOG 117: 1599–1607.
38. Miller SA, Dykes DD, Polesky HF (1988) A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res. 16: 1215.
39. Takai D, Gonzales FA, Tsai YC, Thayer MJ, Jones PA (2001) Large scale mapping of methylcytosines in CTCF-binding sites in the human H19 promoter and aberrant hypomethylation in human bladder cancer. Hum Mol Genet 10: 2619–2626.
40. Frost JM, Moore GE (2010) The Importance of Imprinting in the Human Placenta. PLoS Genet. 6: e1001015.
41. Nelissen ECM, van Montfoort APA, Dumoulin JCM, Evers JLH (2010) Epigenetics and the placenta. Hum. Reprod. Update 17: 397–417.
42. Ergaz Z, Avgil M, Ornoy A (2005) Intrauterine growth restriction - etiology and consequences: What do we know about the human situation and experimental animal models? Reprod. Toxicol. 20: 301–322.
43. Sibai B, Dekker G, Kupferminc M (2005) Pre-eclampsia. Lancet 365: 785–799.
44. Roberts CT (2010) IFPA Award in Placentology Lecture: Complicated interactions between genes and the environment in placentation, pregnancy outcome and long term health. Placenta 31: S47–S53.
45. Wang Y, Fu J, Song W, Wang L (2010) Genomic imprinting status of IGF2 and H19 in placentas of fetal growth restriction patients. J. Genet. 89: 213–216.
46. Lambertini L, Diplas A, Lee M, Sperling R, Chen J, et al. (2008) A sensitive functional assay reveals frequent loss of genomic imprinting in human placenta. Epigenetics 3: 261.
47. Adam GI, Cui H, Miller SJ, Flam F, Ohlsson R (1996) Allele-specific in situ hybridization (ASISH) analysis: a novel technique which resolves differential allelic usage of H19 within the same cell lineage during human placental development. Development 122: 839.
48. Killian JK, Nolan CM, Wylie AA, Li T, Vu TH, et al. (2001) Divergent evolution in M6P/IGF2R imprinting from the Jurassic to the Quaternary. Hum. Mol. Genet. 10: 1721.
49. Pastinen T (2010) Genome-wide allele-specific analysis: insights into regulatory variation. Nat. Rev. Genet. 11: 533–538.
50. Andraweera PH, Dekker GA, Thompson SD, Roberts CT (2012) Single-nucleotide polymorphisms in the KDR gene in pregnancies complicated by gestational hypertensive disorders and small-for-gestational-age infants. Reprod Sci.19: 547–554.
51. Kaffer CR, Srivastava M, Park KY, Ives E, Hsieh S, et al. (2000) A transcriptional insulator at the imprinted H19/Igf2 locus. Genes Dev. 14: 1908–1919.
52. Thorvaldsen JL, Bartolomei MS (2000) Mothers setting boundaries. Science 288: 2145–2146.
53. Arney KL (2003) H19 and Igf2 - enhancing the confusion? Trends Genet. 19: 17–23.
54. Lewis A, Mitsuya K, Umlauf D, Smith P, Dean W, et al. (2004) Imprinting on distal chromosome 7 in the placenta involves repressive histone methylation independent of DNA methylation. Nat Genet. 36: 1291–1295.
55. Mohammad F, Pandey GK, Mondal T, Enroth S, Redrup L, et al. (2012) Long noncoding RNA-mediated maintenance of DNA methylation and transcriptional gene silencing. Development. 139: 2792–2803.
56. Monk D, Wagschal A, Arnaud P, Muller PS, Parker-Katiraee L, et al. (2008) Comparative analysis of human chromosome 7q21 and mouse proximal chromosome 6 reveals a placental-specific imprinted gene, TFPI2/Tfpi2, which requires EHMT2 and EED for allelic-silencing. Genome Res. 18: 1270–1281.
57. Okae H, Hiura H, Nishida Y, Funayama R, Tanaka S, et al. (2012) Re-investigation and RNA sequencing-based identification of genes with placenta-specific imprinted expression. Hum Mol Genet. 21: 548–558.
58. Reik W, Santos F, Mitsuya K, Morgan H, Dean W (2003) Epigenetic asymmetry in the mammalian zygote and early embryo: relationship to lineage commitment? Philos. Trans. R. Soc. Lond. B Biol. Sci. 358: 1403.
59. Diplas AI, Hu J, Lee MJ, Ma YY, Lee YL, et al. (2009) Demonstration of all-or-none loss of imprinting in mRNA expression in single cells. Nucleic Acids Res. 37: 7039.

Appendix B

Publication Format: Imprinted and X-Linked Non-Coding RNAs as Potential Regulators of Human Placental Function

Buckberry, S., Bianco-Miotto, T. and Roberts, C.T. (2014). Imprinted and X-linked non-coding RNAs as potential regulators of human placental function. *Epigenetics*, 9(1), 81–89.

Appendix C

Publication Format: Integrative Transcriptome Meta-Analysis Reveals Widespread Sex-Biased Gene Expression at the Human Fetal–Maternal Interface

Buckberry, S., Bianco-Miotto, T., Bent, S.J., Dekker, G.A. and Roberts, C.T. (2014). Integrative transcriptome metaanalysis reveals widespread sex-biased gene expression at the human fetal–maternal interface. *Molecular Human Reproduction,* 20(8), 810–819.

Appendix D

Publication Format: *massiR*: a Method for Predicting the Sex of Samples in Gene Expression Microarray Datasets

Buckberry, S., Bent, S.J., Bianco-Miotto, and Roberts, C.T. (2014). massiR: a method for predicting the sex of samples in gene expression microarray datasets. *Bioinformatics*, 30(14), pp. 2084–2085.

Appendix E

Bioconductor software manual: *massiR:* MicroArray Sample Sex Identifier

massiR: MicroArray Sample Sex Identifier

Sam Buckberry

October 13, 2014

Contents

1 The Problem

Given that the sex of many species is an easily observable and usually unambiguous classification, it is surprising the number of microarray data sets in public repositories that lack the associated sample sex information. Sex-biased gene expression in normal and pathological tissues is a well recognized for both sex chromosome and autosomal genes. Sex biases also exist in the prevalence and severity of many common human diseases, such as cardiovascular disease and some cancers. As sex is a potential influencing factor of both pathological and non-pathological phenotypes, gene expression analyses that do not account for sex-specific effects could fail to identify a significant proportion of genes that contribute the condition under investigation. Therefore, the absence of sample sex information restricts the reuse of gene expression data sets where the researcher intends to factor the effect of sex in reanalysis or reinterpretation, or when intending to include such data sets in larger gene expression meta-analyses.

This is why we developed massiR, a package for predicting the sex of samples in microarray data sets. This package allows researchers to expand their analyses to retrospectively incorporate sex as a variable, generate or confirm sex information associated with publicly available data sets, to accurately predict the sex of samples missing sex information, or as a simple sanity check for your own microarray gene expression data.

2 Importing data and beginning the analysis

The massiR analysis begins by importing standard gene expression data of normalized and log transformed probe values. The gene expression data can be in the form of a data.frame object and have the sample identifiers as the column names and the probe identifiers as the row names, or as an ExpressionSet object. The identifiers for probes corresponding to Y chromosome genes must be as a data.frame object with the probe identifiers as row.names.

To load the included test massiR gene expression data:

```
> library(massiR)
> data(massi.test.dataset)
```

The included gene expression data is composed of 60 samples and 1026 probes as a data.frame object.

To load the test Y chromosome probes corresponding to the included data:

```
> data(massi.test.probes)
```

The Included list of Y chromosome probes contains probe identifiers as row.names in the data.frame class.

3 Extracting the Y chromosome probe data

The first step of the massiR analysis involves extracting the expression values for probes that correspond to Y chromosome genes. The user has the option of using their own list of probes corresponding to Y chromosome genes or using the probe lists included with the package. The included lists correspond to popular microarray platforms and contain identifiers for probes that map uniquely to Y chromosome genes. See section 8 for detials on using the included probes and section 9 for details on obtaining Y chromosome probes easily from Ensembl Biomart.

When the expression values for Y chromosome probes are extracted, the expression variance for each probe across all samples is calculated. This allows the identification of low variance probes, which are unlikely to be informative in sex classification. The user has the option of selecting a probe variation threshold, so only the most informative probes are used in the classification process. Deciding on a probe variation threshold can be informed by inspecting a probe variation plot (Figure 1) generated by the massi.y.plot function. In our experience, using the most variable 25-50% of probes (typically 10-40 probes, depending on platform) produces good results.

To extract data corresponding to Y chromosome probes from the test data set and look at a probe variation plot:

```
> massi.y.out <- massi_y(massi.test.dataset, massi.test.probes)

> massi_y_plot(massi.y.out)
```

This plot (Figure 1) is output to the R graphics device.

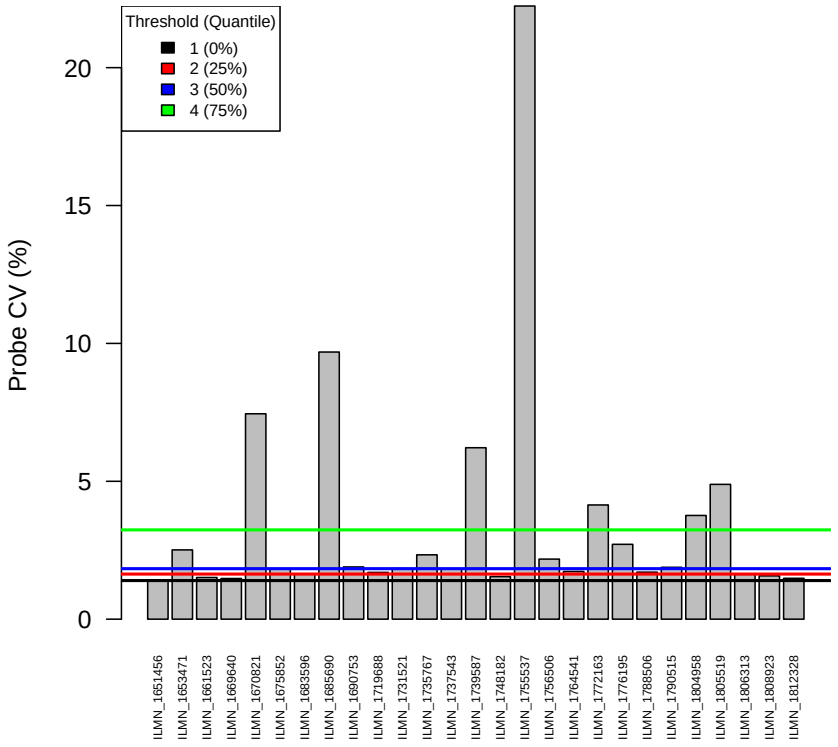


Figure 1: Expression variation (CV) of Y chromosome probes across all samples

After viewing the probe variation plot, a decision can be made regarding which probes to use in the clustering step. The massiR package includes methods for selecting probe variation thresholds based on quantiles. The threshold can be determined by quantiles of probe variance (CV): 1=All probes, 2=Upper 75%, 3=Upper 50%, 4=Upper 25%. It is highly recommended that probe CV plot generated using the massi_y_plot function be inspected to inform threshold choice (Figure 1). The default threshold value is 3.

Once a probe threshold has been decided upon, run the massi_select function. This will return a data.frame with the samples as columns and the subset of selected y chromosome probes as row names.

```
> massi.select.out <-
+   massi_select(massi.test.dataset, massi.test.probes, threshold=4)
```

Check the output for the first 5 samples:

```
> head(massi.select.out)[,1:5]
```

```
                   S1       S2       S3       S4       S5
ILMN_1670821 5.746427 5.686032 6.307110 6.179258 6.594808
ILMN_1685690 5.459125 5.567289 6.919465 6.789817 6.559376
ILMN_1739587 5.883483 5.764190 6.441775 6.438789 6.707278
ILMN_1755537 5.882456 5.831844 8.133164 8.052959 8.298985
ILMN_1772163 5.696833 5.680091 5.907170 6.017871 6.465122
ILMN_1804958 5.815093 5.654395 5.929610 6.104089 5.868732
```

4 Predicting the sex of samples

To classify samples as either male or female, clustering is performed using the values from the subset of Y chromosome probes by implementing the partitioning around medoids algorithm which performs k-medoids clustering (Hennig 2013), where samples are assigned to one of two clusters. The two clusters are then compared using the probe expression values across all samples in each cluster. Samples within the cluster featuring the highest Y chromosome probe values are classed as male and those within the cluster with the lowest Y probe values classed as female. Results such sample probe mean, standard deviation and z-scores are reported in a table together with the sex predicted for each sample.

To predict the sex of the samples using massi_cluster:

```
> results <- massi_cluster(massi.select.out)
```

Extract the results for each sample from the returned list:

```
> sample.results <- data.frame(results[[2]])
```

```
> head(sample.results)
```

```
   ID mean_y_probes_value y_probes_sd    z_score    sex
1  S1            5.911089   0.4572756 -0.6453629 female
2 S10            6.749520   0.8418586  0.7773050   male
3 S11            5.689586   0.4750484 -1.1074329 female
4 S12            6.702993   0.7894613  0.7045705   male
5 S13            5.838450   0.6759924 -0.7193198 female
6 S14            5.819845   0.6593184 -0.7524047 female
```

As you can see, it is a relatively straightforward procedure to produce a table with the predicted sex of each sample with some basic metrics.

5 Visualizing the massiR analysis data

The massiR package includes a function which allows various aspects of the data used in the analysis to be visualized. These plots enable to used to inspect sample and clustering characteristics which could aid in identifying problematic samples and outliers.

To run the massi.plot function with the output from the massi_select and massi_cluster functions:

```
> massi_cluster_plot(massi.select.out, results)
```

This function will generate a heat map with dendrogram of Y chromosome probes as rows and individual samples in columns (Figure 2), a bar plot of mean values and standard deviation from the subset of Y chromosome probes used in K-medoids clustering (Figure 3), with the bars colored with respect to predicted sex and a principal component plot showing clusters (Figure 4). These plots can aid the user in identifying sample outliers or probes that may not be informative in the clustering step.

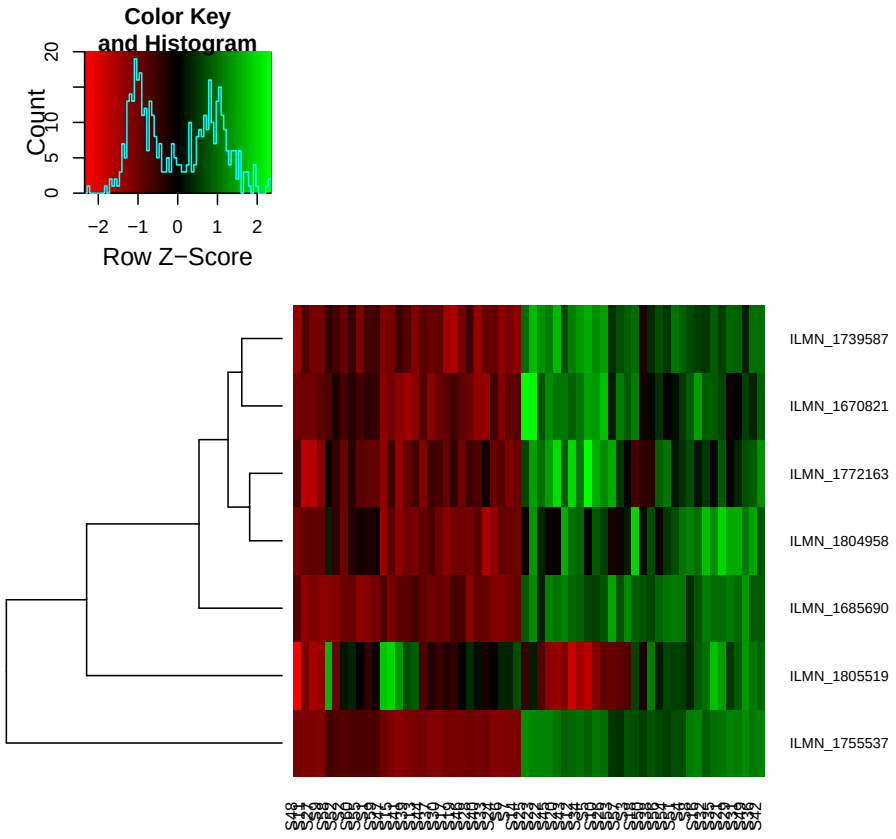


Figure 2: Heat map with dendrogram of Y chromosome probes as rows and individual samples in columns. Notice that the values for the probe in the fifth row are reasonably variable but do not show the same pattern seen with other probes. Therefore viewing the heatmap may help identify problematic probes.

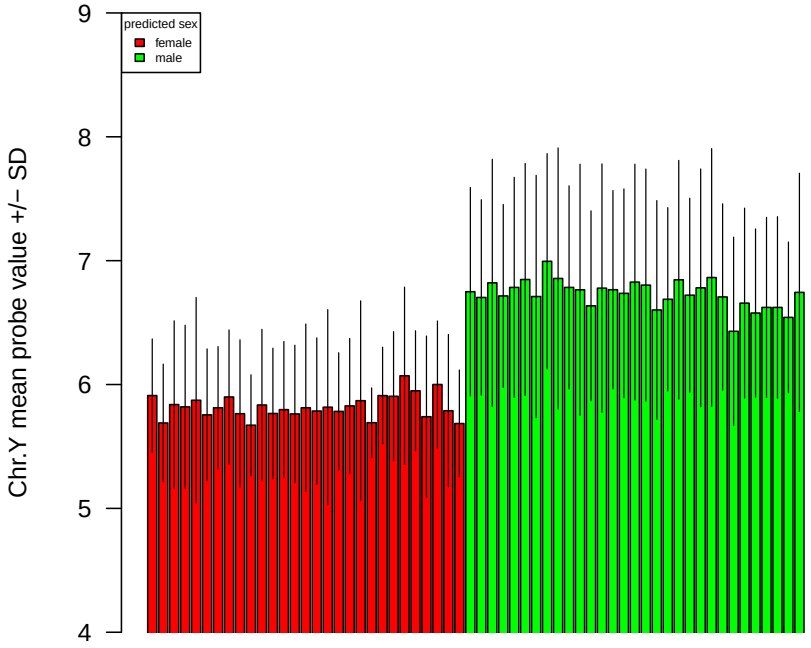


Figure 3: Mean values of the subset of Y chromosome probes used in K-medoids clustering. The bar colors represent clusters, which were assigned as female (red) and male (green)

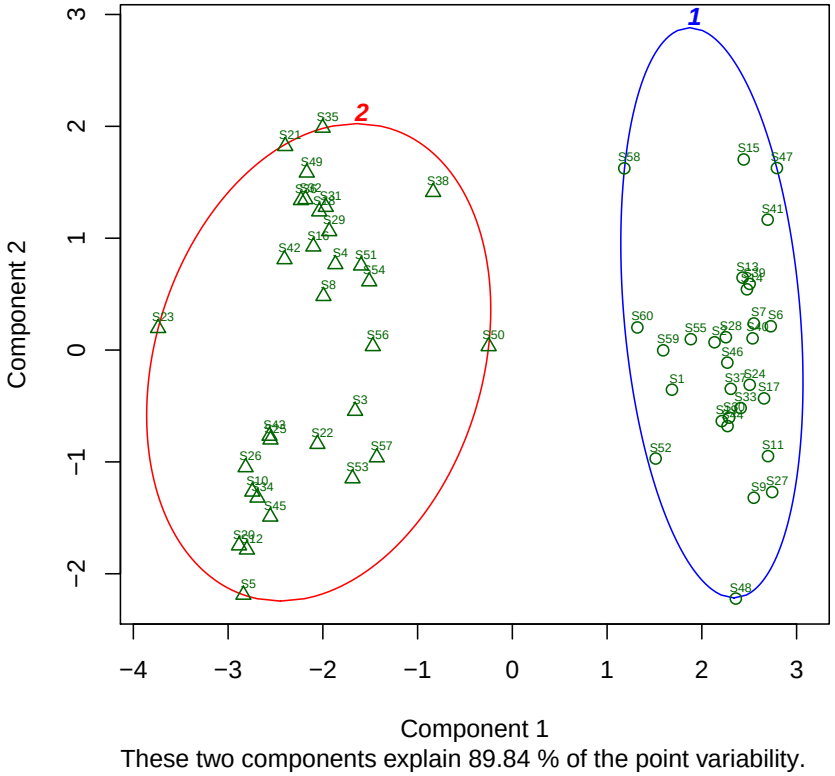


Figure 4: Principal component plot of male and female clusters

6 Check for potential sex bias using the dip test

The massiR method for predicting the sex of samples is >97% accurate for data sets with 6 or more samples and with at least of 15% of either males or females. Outside of this range, this method still performs well in most cases. As there is no guarantee that publicly available data sets will fall within these limits, the function massi.dip can be used to test if the data set might have a male/female ratio that might affect performance.

The massiR method was tested using empirical data sets for five different human tissues. Individual data subsets were randomly generated for each tissue data set ranging from 6-50 samples and with a wide-range of Male/Female ratios. The results of this testing suggest for data sets with >10 samples a dip statistic >0.08 is indicative of at least 15% of males or females in the data set.

The massi_dip function calculates z-scores for each sample and implements the dip test to test for unimodality (Maechler 2013). As a relatively balanced dataset would typically show a bi-modal distribution of the z-scores, the dip statistic is then used to predict if a dataset shows a unimodal distribution that would be expected if a vast majority of samples were of one sex.

To use massi_dip function, which calculates the dip statistic using the data output from the massi_select function:

```
> dip.result <- massi_dip(massi.select.out)
```

This returns the message: dip test statistic is >0.08. This suggests that the proportion of male and female samples in this data set is relatively balanced

Visually inspecting this distribution as a density plot (figure 5) or a histogram plot (figure 6) enables the user to see if there is the expected bi-modal distribution (as there should be distinct distributions for each sex).

To produce a density plot and histogram of sample z-scores:

```
> dip.result <- massi_dip(massi.select.out)
> plot(dip.result[[3]])
```

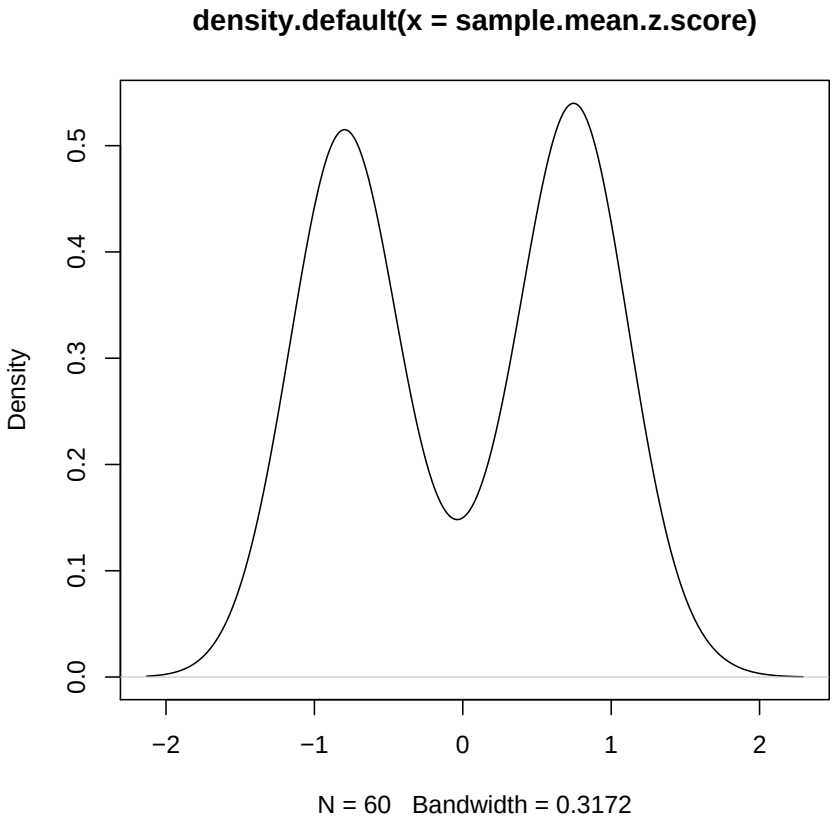


Figure 5: Density plot of mean y chromosome probe z-scores

```
> dip.result <- massi_dip(massi.select.out)
> hist(x=dip.result[[2]], breaks=20)
```

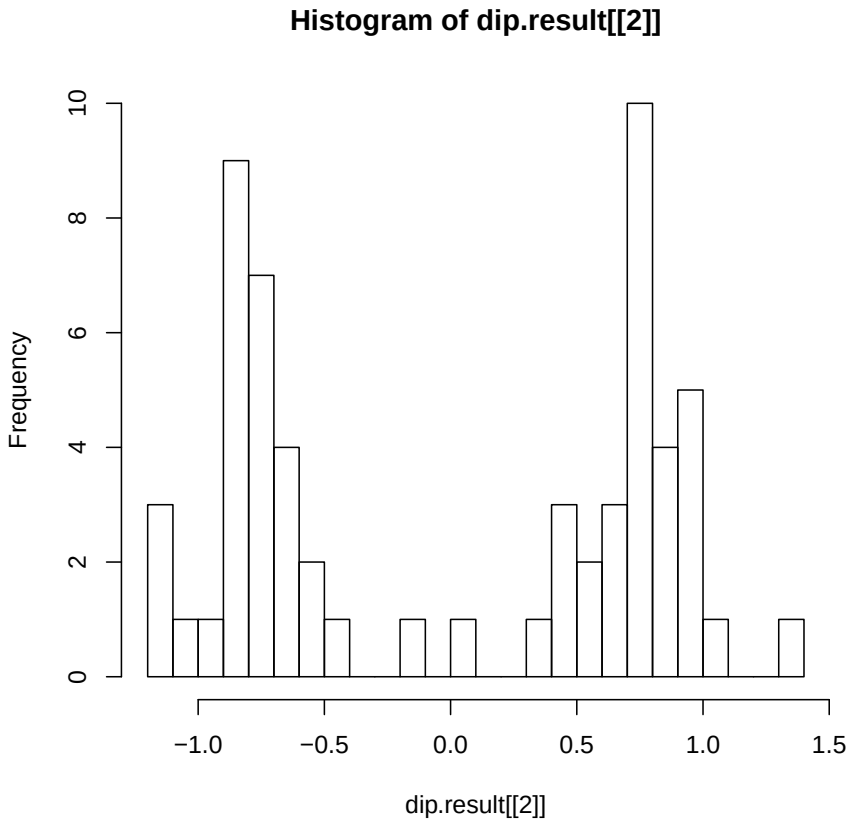


Figure 6: Histogram of mean y chromosome probe z-scores

If the data set was has a sex bias that may influence the accuracy of the massiR sex prediction, then the massi_dip function is likely to return a dip statistic of <0.08. For example, if we are to use the massiR test data set to generate a subset to 20 samples composed of 10% males, we will see that the dip statistic returned is <0.08.

To create this female skewed bias:

get the sample id's for the male and female samples:

```
> male.ids <-
+   subset(sample.results$ID,
+                    subset=sample.results$sex=="male")
> female.ids <-
+   subset(sample.results$ID,
+          subset=sample.results$sex=="female")
```

Create a data subsest of 20 samples with 10% males:

```
> bias.subset.ids <- c(female.ids[1:18], male.ids[1:2])
> bias.subset <- massi.select.out[bias.subset.ids]
```

Use the massi.dip function to test for sex-biased data set:

```
> bias.dip <- massi_dip(bias.subset)
```

Please note that a dip >0.08 is a good indication that there is not a sex bias present that will affect the accuracy of the massiR method. However, and dip statistic <0.08 may still be returned for data sets with >15% males or female or data sets that a suitable for massiR analysis, therefore the results of the massi_dip function should be interpreted with caution and in light of the massi_cluster results.

7 Performing massiR analysis with an ExpressionSet object

The massiR pipeline allows the input of expression data in the class ExpressionSet. Here is an example of how to use data in the ExpressionSet class in a massiR analysis and how to put the results back into the ExpressionSet:

Load the example ExpressionSet data included with the massiR package:

```
> data(massi.eset, massi.test.probes)
```

Using massiR with an ExpressionSet is the same as using a data.frame as in the above example:

```
> eset.select.out <-
+   massi_select(massi.eset, massi.test.probes)
> eset.results <-
+   massi_cluster(eset.select.out)
```

Now to get the massi.cluster results and add them to the ExpressionSet:

```
> # Get the sex for each sample from the massi_cluster results
> eset.sample.results <-
+   data.frame(eset.results[[2]])
> sexData <-
+   data.frame(eset.sample.results[c("ID", "sex")])
> # Extract the order of samples in the ExpressionSet and match with results
> eset.names <-
+   colnames(exprs(massi.eset))
> # match the sample order in massiR results to the same as the ExpressionSet object
> sexData <- sexData[match(eset.names, sexData$ID),]
> # create an annotatedDataFrame to add to ExpressionSet
> pData <- new("AnnotatedDataFrame", data = sexData)
> # add the annotatedDataFrame to the Expressionset as phenoData
> phenoData(massi.eset) <- pData
```

Check the phenoData is in the ExpressionSet and double check that all sample id's from the massiR analysis match the sample identifiers in the ExpressionSet.

```
> # check the phenodata is now within the ExpressionSet
> phenoData(massi.eset)

An object of class 'AnnotatedDataFrame'
  rowNames: 1 12 ... 57 (60 total)
  varLabels: ID sex
  varMetadata: labelDescription

> # check that all phenodata id's match expressionSet column names.
> # This must return "TRUE"
> all(massi.eset$ID == colnames(exprs(massi.eset)))

[1] TRUE
```

8 Using the included massiR Y chromosome probe lists

The massiR package includes lists of Y chromosome probes for widely used Illumina and Affymetrix human gene expression platforms. If you wish to use one of the included probe lists, for example the Illumina human v2 probes:

Load the massiR included probe lists:

```
>  data(y.probes)
```

Check the names of the platforms for the probe lists.

```
>   names(y.probes)
```

```
[1] "illumina_humanwg_6_v1" "illumina_humanwg_6_v2" "illumina_humanwg_6_v1"
[4] "illumina_humanht_12"   "affy_hugene_1_0_st_v1" "affy_hg_u133_plus_2"
```

To get probe list into format for massiR analysis:

```
>   illumina.v2.probes <- data.frame(y.probes["illumina_humanwg_6_v2"])
```

The names of the probe lists correspond to Ensembl biomart attribute names. For instructions on obtaining probe identifiers for other platforms, see the section "Using biomaRt to obtain y chromosome probe lists"

9 Using biomaRt to obtain y chromosome probe lists

Obtaining y chromosome probes lists for many microarray platforms is relatively easy using the biomaRt package (Durinik et al. 2005 and Durinik et al. 2009). This method is recommended because Ensembl have mapped probe sequences to reference genomes for many platforms and this allows ambiguous and non-specific probes to be removed. For details on probe mapping methods, see <http://jan2013.archive.ensembl.org/info/docs/microarray_probe_set_mapping.html>

For example, you can download the probes corresponding to the massiR test data set and obtain the Entrez gene id and genomic positions and convert these into a format for a massiR analysis:

Use the biomaRt package to download genomic regions and Entrez gene id's for Illumina v2 probes:

```
> library(biomaRt)
> mart <- useMart('ensembl', dataset="hsapiens_gene_ensembl")
> filters <- listFilters(mart)
> attributes <- listAttributes(mart)
> gene.attributes <-
+  getBM(mart=mart, values=TRUE,
+        filters=c("with_illumina_humanwg_6_v2"),
+        attributes= c("illumina_humanwg_6_v2", "entrezgene",
+                      "chromosome_name", "start_position",
+                      "end_position", "strand"))
```

Remove the probes mapped to multiple genomic regions:

```
> unique.probe <-
+   subset(gene.attributes, subset=!duplicated(gene.attributes[,1]))
```

Select the probes that correspond to y chromosome genes:

```
> y.unique <-
+   subset(unique.probe, subset=unique.probe$chromosome_name == "Y")
```

Get the probe id's as row.names in the format for massiR analysis:

```
> illumina.v2.probes <-
+   data.frame(row.names=y.unique$illumina_humanwg_6_v2)
```

This is a straightforwd way of obtaining Y chromosome probes for many microarray platforms that is independent of platform manufacturer annotations and is highly reccomended.

10 References

Christian Hennig (2013) fpc: Flexible procedures for clustering. R package version 2.1-6. http://CRAN.R-project.org/package=fpc

Martin Maechler (2013) diptest: Hartigan's dip test statistic for unimodality - corrected code. R package version 0.75-5. http://CRAN.R-project.org/package=diptest

Gregory R. Warnes, Ben Bolker, Lodewijk Bonebakker, Robert Gentleman, Wolfgang Huber, Andy Liaw, Thomas Lumley, Martin Maechler, Arni Magnusson, Steffen Moeller, Marc Schwartz and Bill Venables (2013). gplots: Various R programming tools for plotting data. R package version 2.12.1. http://CRAN.R-project.org/package=gplots

Steffen Durinck, Paul T. Spellman, Ewan Birney and Wolfgang Huber (2009) Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. Nature Protocols 4, 1184-1191.

Steffen Durinck, Yves Moreau, Arek Kasprzyk, Sean Davis, Bart De Moor, Alvis Brazma and Wolfgang Huber (2005) BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. Bioinformatics 21, 3439-3440.

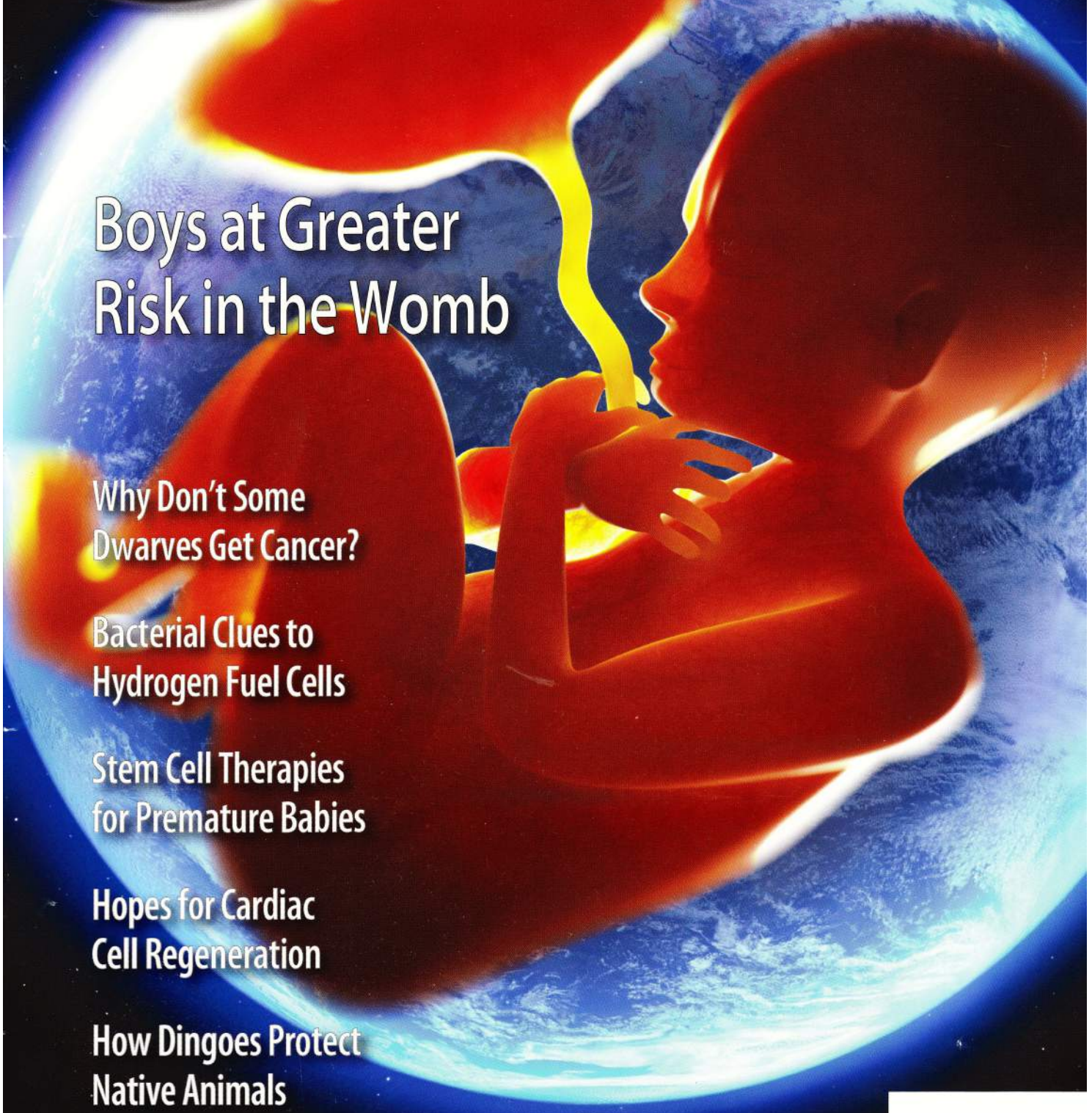
Appendix F

Publication Format: Why are Male Babies More at Risk in the Womb?

Bomb-Proof Buildings • Killer Vines Strangling Rainforests • The Truth About A2 Milk • Is Mental Illness Normal?

Australasian Science

Volume 35 | Number 9
NOVEMBER 2014 | $8.95



Boys at Greater Risk in the Womb

Why Don't Some Dwarves Get Cancer?

Bacterial Clues to Hydrogen Fuel Cells

Stem Cell Therapies for Premature Babies

Hopes for Cardiac Cell Regeneration

How Dingoes Protect Native Animals

ISSN 1442-679X

Crimes of Sleepwalkers • Love Potion Ethics • Who Will Store Our Nuclear Waste?

Buckberry, S. and Roberts, C. T. (2014) Why are males more at risk in the womb? *Australasian Science*, 35(9), 16-18.