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

Declaration of Authorship i

Abstract iii

Acknowledgements v

Publications Arising from this Thesis vi

Contents vii

List of Figures xii

List of Tables xiv

1 Introduction 1
- 1.1 The Placenta Forms the Foundation of Development *in utero* 1
- 1.2 Genomic Imprinting in the Placenta 2
- 1.3 Genome-Wide Sex Differences in Placental Gene Expression 4
- 1.4 The Underlying Organisation of the Human Placental Transcriptome 5
- 1.5 Summary 6

Bibliography 7

2 Quantitative Allele-Specific Expression and DNA Methylation Analysis of *H19*, *IGF2* and *IGF2R* in the Human Placenta Across Gestation Reveals *H19* Imprinting Plasticity 14
Sam Buckberry, Tina Bianco-Miotto, Stefan Hiendleder and Claire T Roberts
- 2.1 Introduction 15
- 2.2 Materials and Methods 17
  - 2.2.1 Ethics Statement 17
  - 2.2.2 Sample Collection 17
  - 2.2.3 Genotyping 18
  - 2.2.4 Quantification of Allele Specific Expression 18
  - 2.2.5 Quantification of DNA Methylation 19

2.2.6 Pyrosequencing . . . 19
2.2.7 Statistical Analysis . . . 20
2.3 Results . . . 20
2.3.1 Quantification of Allele-Specific Gene Expression in the Human Placenta . . . 20
2.3.2 Biallelic Expression of *IGF2R* in the First Trimester Placenta 21
2.3.3 Increased Expression from the *H19* Repressed Allele is Higher in First Trimester Placenta . . . 23
2.3.4 Locus-Specific DNA Methylation Differences in the *H19-IGF2* Region Between First Trimester and Term Placentae . 26
2.3.5 *H19* Repressed Allele Expression is Correlated with Higher Levels of DNA Methylation . . . 28
2.4 Discussion . . . 29
S2.1 Supporting Information . . . 34
S2.2 Methods S1 . . . 39
S2.2.1 DNA Genotyping by HRM . . . 39

Bibliography **42**

3 Imprinted and X-Linked Non-Coding RNAs as Potential Regulators of Human Placental Function **49**
Sam Buckberry, Tina Bianco-Miotto and Claire T Roberts

3.1 Introduction . . . 50
3.2 The Placenta is Key to Fetal and Maternal Health . . . 51
3.3 Classification and Detection of Non-Coding RNA . . . 51
3.4 The Roles of Non-Coding RNAs in Genomic Imprinting in the Placenta . . . 52
3.5 The Imprinted *H19* long Non-Coding RNA and miR-675 . . . 53
3.6 The Imprinted C19MC miRNA Cluster . . . 57
3.7 X-Chromosome ncRNAs in Placental Development and Pregnancy Complications . . . 62
3.8 X-Linked miRNAs as Potential Drivers of Sex Differences in Placental Gene Expression . . . 62
3.9 Previous Studies Provide Few Clues to Which miRNAs Escape XCI 63
3.10 X-Linked miRNAs Associated with Preeclampsia . . . 64
3.11 Conclusions . . . 67

Bibliography **68**

4 *massiR*: a Method for Predicting the Sex of Samples in Gene Expression Microarray Datasets **81**
Sam Buckberry, Stephen J Bent, Tina Bianco-Miotto and Claire T Roberts

4.1 Introduction . . . 82
4.2 Methods and Validation . . . 82
4.3 Conclusion . . . 84

S4.1 Supporting Information . . . 86
S4.2 Y Chromsome Probe Indentifiers Included with the Massi Package . 86
S4.3 Testing and Validation . . . 86
S4.3.1 Samples Classified as Male but Listed as Female in the Metadata . . . 88
S4.3.2 Samples Classified as Female but Listed as Male in the Metadata . . . 88
S4.3.3 Performance with Skewed Sex Ratios . . . 88
S4.3.4 Detecting Datasets with Skewed Ratios . . . 92

Bibliography **96**

5 Integrative Transcriptome Meta-Analysis Reveals Widespread Sex-Biased Gene Expression at the Human Fetal–Maternal Interface **99**
Sam Buckberry, Tina Bianco-Miotto, Stephen J Bent, Gustaaf A Dekker and Claire T Roberts

5.1 Introduction . . . 100
5.2 Results and Discussion . . . 101
5.2.1 Meta-Analysis of Sex-Biased Gene Expression in the Human Placenta . . . 101
5.2.2 Identification of Potential Transcriptional Regulators of Sex-Biased Gene Expression . . . 104
5.2.3 High-Level Molecular Functions and Pathways are Associated with Sex-Biased Genes . . . 108
5.2.4 Sex-Biased Expression of X-Linked Genes . . . 109
5.2.5 LHB-CGB Cluster Genes Show Female Expression Bias . . . 110
5.3 Conclusions . . . 112
5.4 Materials and Methods . . . 113
5.4.1 Study Selection . . . 113
5.4.2 Array Pre-Processing and Quality Control . . . 113
5.4.3 Predicting The Sex of Samples in Datasets Lacking Sex Information . . . 114
5.4.4 Re-Annotation of Microarray Datasets and Probe Summarization . . . 114
5.4.5 Meta-Analysis of Annotated Datasets . . . 116
5.4.6 Prediction of Upstream Transcription Factor Regulation . . . 116
5.4.7 Resolving CGB/LHB Cluster Sequence Homology . . . 116
5.4.8 Gene Enrichment and Pathway Analysis . . . 117
S5.1 Supplementary Information . . . 119

Bibliography **128**

6 Why are Male Babies More at Risk in the Womb? **136**
Sam Buckberry and Claire Roberts

7 Placental Transcriptome Co-Expression Analysis Reveals Conserved Regulatory Programs and Points Toward a Preeclampsia Gene Cluster **142**
Sam Buckberry, Tina Bianco-Miotto, Stephen J Bent, Gustaaf A Dekker and Claire T Roberts

7.1 Introduction . . . 143
7.2 Results . . . 145
7.2.1 Constructing a Weighted Human Placental Co-Expression Gene Network . . . 145
7.2.2 Co-Expression Modules are Reproducible . . . 148
7.2.3 Key Co-Expression Modules are Preserved Across Human Gestation and Conserved in the Mouse . . . 150
7.2.4 Preserved Modules Feature a Core Set of Transcription Factor Motifs . . . 151
7.2.5 Modules of Co-Expressed Genes are Implicated in Pregnancy Complications . . . 153
7.2.6 Eigengenes can be Used to Screen for Non-Invasive Markers of Placental Gene Expression . . . 154
7.3 Discussion . . . 156
7.4 Methods . . . 159
S7.1 Supporting Information . . . 164

Bibliography **169**

8 General Discussion **175**
8.1 Overall Significance . . . 175
8.1.1 Epigenetic Plasticity in the Human Placenta at 6–10 Weeks Gestation . . . 175
8.1.2 Widespread Sex-Biased Gene Expression in the Human Placenta . . . 176
8.1.3 Sex-Biased Expression of Genes Encoding hCG . . . 176
8.1.4 Conservation of Gene Regulatory Programs across Gestation 176
8.1.5 Evidence of Co-Regulation of Genes Implicated in Preeclampsia . . . 177
8.2 Contributions to the Field . . . 177
8.2.1 Detailed Reference of Placental Sex-Biased Gene Expression 177
8.2.2 The ***massiR*** Software Package . . . 177
8.2.3 Comprehensive Placental Transcriptome Dataset . . . 178
8.2.4 A New Framework for Screening Biomarkers of Placental Development and Function . . . 178
8.3 Problems Encountered and Limitations . . . 179
8.3.1 Genomic Imprinting in the Placenta . . . 179
8.3.2 Transcriptome Meta-Analysis . . . 179
8.3.3 Transcriptome Co-Expression Analysis . . . 180
8.4 Future Directions . . . 180
8.5 Conclusion . . . 182

Bibliography **183**

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 **185**

B Publication Format: Imprinted and X-Linked Non-Coding RNAs as Potential Regulators of Human Placental Function **197**

C Publication Format: Integrative Transcriptome Meta-Analysis Reveals Widespread Sex-Biased Gene Expression at the Human Fetal–Maternal Interface **207**

D Publication Format: *massiR*: a Method for Predicting the Sex of Samples in Gene Expression Microarray Datasets **218**

E Bioconductor software manual: *massiR:* MicroArray Sample Sex Identifier **221**

F Publication Format: Why are Male Babies More at Risk in the Womb? **239**

List of Figures

2.1 Relative expression from *H19* and *IGF2* alleles in the human placenta. 22
2.2 Relative expression from *IGF2R* alleles in the human placenta. . . . 23
2.3 Allelic expression ratios for *H19*, *IGF2* and *IGF2R* in the human placenta. . . . 24
2.4 Ratio of expression from each allele in human first trimester and term placentae measured by pyrosequencing. . . . 25
2.5 Relative level of expression from repressed alleles in samples heterozygous for both *H19* rs217727 and *IGF2* rs680. . . . 26
2.6 lacental methylation levels in regions upstream and covering the *H19* transcription start site (TSS). . . . 27
2.7 Levels of *H19* repressed allele expression and DNA methylation in human first trimester placentae. . . . 28
S2.1 Genotyping of *IGF2R* rs1570070 by HRM with sequenced controls. 41

3.1 The *H19* lncRNA transcript and RNA binding proteins in human and mouse. . . . 56
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). . 66

4.1 Sex prediction accuracy of the *massiR* package using human gene expression datasets with a range of male/female ratios. . . . 84
S4.1 Prediction and validation of the sex of samples in dataset GSE29378. 89
S4.2 Prediction and validation of the sex of samples in dataset GSE40435. 90
S4.3 Prediction and validation of the sex of samples in dataset GSE35896. 91
S4.4 Prediction and validation of the sex of samples in dataset GSE45330. 93
S4.5 Sex prediction accuracy of the *massiR* package using five human gene expression datasets and a range of male/female ratios. . . . 94
S4.6 The dip test statistic as a method for identifying datasets with a skewed sex ratio. . . . 95

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. . . . 104
5.2 Circos plot summarizing the meta-analysis of sex-biased gene expression in the human placenta. . . . 105

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**). 106
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. . 107
5.5 Sex-biased expression of X-linked genes in the human placenta. . . . 110
5.6 Female-biased expression of *LHB* and *CGB* cluster genes. 111
S5.1 (**A**) Venn diagram showing the overlap of sex-biased genes in different human tissue/cell types. (**B**) Comparison of differentially expressed genes in this meta-analysis and previously studies investigating sex-biased gene expression in placental tissue, and placental epithelium and endothelial tissue. 119
S5.2 Statistical measures for transcription factor binding motifs flanking transcription start sites of genes with sex-biased expression in the human placenta. 120
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). 121
S5.4 LHB-CGB gene cluster amino acid sequence alignment. 122
S5.5 *LHB* expression in multiple human tissues, with extra-embryonic tissues highlighted in red. 123

7.1 Weighted gene co-expression network analysis of the human placenta reveals distinct clusters of co-expressed genes. 145
7.2 Gene–eigengene correlations identify module hub genes that are consistently co-expressed in the human placenta. 148
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). 149
7.4 Overlap between weighted gene co-expression network modules for human and mouse placenta. 151
7.5 EBF1 and ZNF423 are potential upstream regulators of M3 gene expression. 152
7.6 M12 is enriched for genes that show a meta-signature for preeclampsia. 154
7.7 M12 genes are significantly up-regulated in preeclampsia placentas. 155
7.8 Module eigengenes can be used to screen for non-invasive biomarkers of placental gene expression. 156
S7.1 RNA-Seq metrics and quality control. 164
S7.2 Heat map showing the gene overlap between modules. 165
S7.3 Heat map showing the number of overlapping top ten transcription factors predicted to regulate each co-expression module. 166
S7.4 Summary network indices (y-axes) as functions of the soft-thresholding power (x-axes) for human and mouse. 167

List of Tables

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. 21
2.2 Relative levels of repressed allele expression in human first trimester and term placentae. 24
S2.1 Genomic DNA specific primers used to detect DNA contamination in RNA samples. 34
S2.2 Details of genes, SNP regions and primers used for quantifying allele-specific expression by pyrosequencing. 35
S2.3 Parental and placental genotypes with placental allele expression ratio for *H19* and *IGF2*. 36
S2.4 Parental and placental genotypes with placental allele expression ratio for *H19* and *IGF2*. 37
S2.5 Pearson's correlation of *H19* repressed allele expression and DNA methylation levels at individual CpG loci in human first trimester placentae. 38
S2.6 Gene, SNP, and HRM genotyping assay details used for genotyping placental DNA. 40

3.1 Imprinted or X-linked miRNA differentially expressed in preeclampsia (PE) and pre-term birth (PTB) with validated targets and potential mechanisms 59

S4.1 Validation results for predicting the sex of samples in microarray datasets using the MASSI package. 87

5.1 Details of each datasets included in the meta-analysis 102
S5.1 Transcription factors with enriched binding sites surrounding genes with sex biased expression. 124
S5.2 Number of arrays at each step of the selection process for the meta analysis. 125
S5.3 Re-annotation and gene summarization details for each microarray platfoms. 126
S5.4 Top 5 Ingenutiy canonical pathways and biological functions enriched for sex-biased genes. 127

7.1 Co-expression module characteristics. 147
S7.1 Sample characteristics. 168