An Epigenetic Analysis of the Human Placenta

Erin Cuffe Macaulay (née Daly)

A thesis submitted for the degree of Doctor of Philosophy at the University of Otago Dunedin, New Zealand

December 2010
Abstract

The human placenta is a highly specialized organ that is responsible for the survival of pregnancy. During its development, placental trophoblast cells invade into the uterine wall to establish a blood supply for the growing fetus. Previous studies have suggested similarities between the invasive phenotypes of trophoblasts and tumour cells; however, a key difference is that trophoblast invasion is under strict control. Given that epigenetic mechanisms have been linked with the silencing of key regulatory genes in cancer, we hypothesized that the epigenetic regulation of first-trimester placental trophoblasts may provide a mechanistic relationship between placental and cancer growth. Further, although the hypomethylated environment within the pseudo-malignant placenta is unique, its role in facilitating placental function is poorly understood. We sought to document placental-specific epigenetic modifications, taking into account that the origin of the placenta is determined during the earliest stages of embryonic development, when the inner-cell mass is first distinguished from the trophectoderm, and when the inner-cell mass further differentiates into the primitive endoderm and the epiblast.

A genome-wide methylation analysis was performed using methylated DNA immunoprecipitation (MeDIP) combined with hybridisation to promoter microarrays to identify differentially methylated gene promoters between first-trimester human placenta and peripheral blood DNA. The promoter methylation of 29 candidate genes was then quantified using Sequenom MassARRAY®. Differential methylation patterns were detected in placental tissues compared to both fetal and adult somatic tissues. The relationship between promoter methylation and gene expression was then assessed using real-time PCR and immunohistochemistry.

The promoter methylation of one gene, KCNH5, was found to be lineage-specific: low in all tissues derived from the extra-embryonic lineages (trophectoderm and primitive endoderm) and very high in tissues derived from the embryonic (epiblast) lineage. The dichotomous promoter methylation of KCNH5 was found to regulate the lineage-specific expression of alternative gene transcripts. Interestingly, the KCNH5 promoter that is used in tissues derived
from the extra-embryonic lineages, and which shows dichotomous methylation, has recently evolved from a SINE retrotransposon that is present in only humans, old world monkeys and apes. To our knowledge, this the first example of a human transcript derived from the insertion of a SINE element. Finally, the lineage origin of the extra-embryonic mesenchyme has been a topic of longstanding debate. The combined epigenetic and expression profiles of KCNH5 in placental villous stroma provide compelling evidence that the extra-embryonic mesenchyme is derived from the primitive endoderm.

Retrotransposons are normally silenced by methylation to prevent genome dysfunction. However, the placenta is becoming increasingly known as a tissue in which retrotransposons are actively transcribed. We observed that the absence of retrotransposon-silencing by methylation permitted the emergence of a placental-specific transcript by allowing the retrotransposon to serve as an alternative promoter for KCNH5. Examination of additional retrotransposon-derived genes in the placenta (INSL4 and ERVWE1) confirmed that dichotomous methylation between embryonic and extra-embryonic lineages is a feature of early development. The finding that the retro-elements in these genes have escaped the normal silencing mechanism suggests that they may have functional roles that are unique to the invasive placentas of humans and recent primates.
My PhD was an exciting adventure thanks to the help and support from many important people. First and foremost, a huge thank you to my supervisor, Professor Ian Morison, who not only gave me the time, the tools and the knowledge to achieve my goals, but who has also been a role model for me as a scientist, and a good friend. Secondly, a special thanks to Rob Weeks, a Senior Research Fellow in the Morison Group, who gave up his precious time to answer my many questions and to teach me new techniques. To Verity Oliver and Jackie Ludgate, thanks for being such great friends and for making the lab a happy place to come to everyday. I also received helpful support from my PhD committee, including Professor Tony Reeve, Professor Mike Legge and Associate Professor Peter Dearden, who asked those thought-provoking questions that kept me on my toes.

Importantly, I’d like to acknowledge the women who donated their precious pregnancy tissues to this project. This study would have not been possible without them. I’d also like to thank the helpful nurses at Dunedin Hospital who treated the patients and their tissues with the upmost respect, and always greeted me with a smile.

I am very grateful to have received generous funding during the course of my PhD. Firstly, I’d like to recognise the Prestigious PhD Scholarship that I was fortunate to receive from the University of Otago. I’d also like to thank the National Research Centre for Growth and Development for funding my research and providing me with generous travel funding to attend conferences and visit labs around the world. To Dr. Stephen Lye and his group at the Mount Sinai Hospital in Toronto, thank you for teaching me so much about the trophoblast! I’d also like to acknowledge those at the Babraham Institute in Cambridge, UK, particularly Dr. Simon Andrews who performed the MeDIP-microarray data analyses for this project. To Professor Wolf Reik, Dr. Myriam Hemberger, Dr. Wendy Dean, Dr. Fatima Santos and Dr. Ray Ng, thank you for the enjoyable visit – you each gave me a fascinating insight into the world of developmental epigenetics during the early stages of my PhD. I am also grateful to Professor Graham Burton, Professor Ashley Moffett and Dr. Vardhman Rakyan for sharing their time with me during my visits to their laboratories. I’d also like to recognise the Royal Society of New Zealand, the Department of Biochemistry, the Functional Genomics, Gene...
Expression and Proteomics Theme and the Division of Health Sciences at the University of Otago for their generous travel awards. And lastly, thank you to the Dunedin School of Medicine for the “Finishing Your PhD” Scholarship that supported me as I wrote this PhD thesis.

There are many people who were not involved in the science-side of this project that deserve special recognition. To James, my fiancé and best friend, who not only kept me company in lab on the weekends, but who has been a stable, loving presence in my life during the ups and downs over the years. Thanks to my neighbour, Leonie, who always leant an ear during our daily rides to University together. To my grandparents, who have given me unconditional love and support even though we are on opposite sides of the world. To my father, who has instilled in me a natural curiosity to explore the world ever since I was a little girl. From the science experiments in his workshop to the cliff-jumping lessons on the ski slopes and the sailboat-capsizing practice on the lake, my father has taught me how to approach life with ambition. And finally I thank my mother, who is resting peacefully now, but who gave me the optimism, strength and unyielding courage to navigate through life, whatever path it may take. She has been with me in spirit every step of the way.
# Table of Contents

Abstract iii
Acknowledgements v
Table of Contents vii
List of Tables xiv
List of Figures xvii
List of Abbreviations xx

## Chapter One: Introduction

1.1 The Human Placenta 1
1.1.1 Placental Function 1
1.1.2 Placental Structure 2

1.2 Development of the Human Placenta 3
1.2.1 Fertilisation and Cell Lineage Specification 3
1.2.2 Blastocyst Implantation 5
1.2.3 Trophoblast Invasion 6
1.2.3.1 The Trophoblast 6
1.2.3.2 The Villous Trophoblast 7
1.2.3.3 The Extra-Villous Trophoblast 7

1.3 Molecular Mechanisms Regulating Trophoblast Invasion 8
1.3.1 Factors from the Maternal Decidua 9
1.3.2 Cell Surface Molecules 9
1.3.3 Effect of Oxygen Tension 10

1.4 Placental Pathologies 11
1.4.1 Pre-eclampsia 11
1.4.2 Placental Insufficiency (Intra-Uterine Growth Restriction) 12
1.4.3 Placenta Acreta, Increta and Percreta 12
1.4.4 Tumours of the Placenta 13

1.5 Epigenetics 13
1.5.1 Histone Modifications 14
1.5.2 DNA methylation 14
1.5.2.1 CpG Dinucleotides 15
1.5.2.2 CpG Islands 15
1.5.2.3 The Mechanism of DNA Methylation 15

1.6 DNA Methylation: Role in Early Human Embryonic Development 17
1.6.1 Epigenetic Reprogramming 17
1.6.1.1 Reprogramming in the Primordial Germ Cells 18
1.6.1.2 Reprogramming in the Embryo 18

1.6.2 Genomic Imprinting 18
1.6.2.1 The Conflict Hypothesis 19
1.6.2.2 Establishing Imprinting Marks 19
### Chapter One: Epigenetics

1.6.3 X Chromosome Inactivation 20
1.6.4 Epigenetic Silencing of Retrotransposons 20

1.7 Epigenetic Modifications: Response to the Environment 21
1.7.1 Pre-natal Epigenetic Modifications 21
1.7.2 Post-natal Epigenetic Modifications 22

1.8 Epigenetics in the Placenta 23
1.8.1 Imprinted Genes in the Placenta 23
1.8.2 Non-Imprinted Genes in the Placenta 24
1.8.3 The Hypomethylated State of the Placenta 24
1.8.3.1 X Chromosome Hypomethylation 25
1.8.3.2 Hypomethylation of Repetitive Elements 25

1.9 Behavioural Similarities of Trophoblasts and Tumours 26

1.10 DNA Methylation: Role in Tumorigenesis 27

1.11 The Epigenetic Regulation of Trophoblasts: Similarities to Cancer 27
1.11.1 Tumour Suppressor Genes 27
1.11.2 Cell Adhesion Molecules 28

1.12 Techniques to Study Genome-wide Methylation Patterns 29
1.12.1 Genome-wide Methylation Analysis by Methylation-Sensitive Restriction Enzymes 29
1.12.2 Analysing Methylation using Microarrays 30

1.13 Aims of this Study 31

### Chapter Two: Materials and Methods

2.1 Collection of Pregnancy Tissues 33
2.1.1 Ethical Approval 33
2.1.2 Sample Collection 33
2.1.2.1 Collection of First-Trimester Placenta 33
2.1.2.2 Collected of Term Placenta 34
2.1.3 Processing of Placental Samples 34
2.1.3.1 Processing of First-Trimester Placenta 34
2.1.3.2 Processing of Term Placenta 35
2.1.4 DNA Extraction from Placental Tissue 35

2.2 Polymerase Chain Reaction Conditions 36
2.2.1 Standard PCR Protocol 36
2.2.2 Standard PCR Cycling Conditions 37
2.2.3 PCR Product Detection 37
2.2.3.1 Agarose Gel Electrophoresis 37

2.3 Fetal Sex Determination for Placental Samples 38
2.3.1 Detecting the Y Chromosome 38
2.3.2 Detecting both X and Y Chromosomes 38

2.4 Maternal Contamination Screening by Microsatellite Analysis 39
2.4.1 Primer Design 39
2.4.2 α32P-labelled PCR 39
2.4.3 Polyacrylamide Gel Electrophoresis 40

2.5 Methylated DNA Immunoprecipitation and Promoter Microarrays 41
2.5.1 Sample Selection and Pooling 43
2.5.1.1 First-Trimester Placenta Samples 43
2.5.1.2 Peripheral Blood Control Samples 43
2.5.2 Methylated DNA Immunoprecipitation (MeDIP) 44
  2.5.2.1 NlaIII Restriction Digest 44
  2.5.2.2 Digested DNA Purification 45
  2.5.2.3 Removal of Total Genomic DNA 45
  2.5.2.4 Immunoprecipitation of Methylated DNA 45
2.5.3 DNA Selection and Ligation 46
  2.5.3.1 Biotinylation of DNA 46
  2.5.3.2 Annealing and Ligation 47
  2.5.3.3 PCR Amplification and Labeling of DSL Products 48
2.5.4 Hybridisation to 20K Gene Promoter Microarrays 49
  2.5.4.1 Blocking of Microarray Slides 49
  2.5.4.2 Hybridisation to Microarray Slides 49
  2.5.4.3 Washing of Microarray Slides 50
  2.5.4.4 Microarray Scanning 50
2.5.5 Microarray Data Analysis 51
  2.5.5.1 Microarray Data Normalisation 51
  2.5.5.2 Microarray Data Analysis #1 51
  2.5.5.3 Microarray Data Analysis #2 51
  2.5.5.4 Microarray Data Analysis #3 52
  2.5.5.5 Selection and Location of Differentially Methylated Candidate Genes 52
2.6 Sequenom MassARRAY® EpiTYPER® Platform for Quantitative Methylation Analysis 53
  2.6.1 Sample Selection for Sequenom 54
  2.6.2 Sequenom Primer Design 54
  2.6.3 Bisulfite Conversion of Genomic DNA 54
  2.6.4 Sequenom PCRs 55
    2.6.4.1 Optimisation of Bisulfite PCRs 55
    2.6.4.2 Sequenom PCR Reaction Conditions 59
  2.6.5 Processing of Sequenom Reactions 59
    2.6.5.1 Treatment with Shrimp Alkaline Phosphatase 59
    2.6.5.2 T-Cleavage and RNA Transcription 60
    2.6.5.3 Conditioning the Reaction Products and Spotting of the Microarrays 60
  2.6.6 Sequenom Data Analysis 61
2.7 Gene Expression Analysis by Reverse-Transcription PCR 63
  2.7.1 RNA Extraction from Tissues 63
  2.7.2 Reverse-Transcription Reaction 64
    2.7.2.1 Testing cDNA before use in RT-PCR 64
    2.7.2.2 Making cDNA Standards for Quantitative RT-PCR 65
  2.7.3 Quantitative RT-PCR using TaqMan® Gene Expression Assays 66
  2.7.4 Calculation of Gene Expression 66
2.8 Transcript-Specific RT-PCRs for KCNH5 67
  2.8.1 Primer Design for Transcript Specific RT-PCRs 67
  2.8.2 Optimisation of Transcript-Specific RT-PCRs 67
  2.8.3 Transcript-Specific RT-PCRs 68
2.9 Western Blot Analysis for KCNH5 Protein 69
  2.9.1 Protein Extraction from Samples 69
    2.9.1.1 Sample Selection 69
    2.9.1.2 Tissue/Cell Lysis and Total Protein Isolation 69
2.9.1.3 Determining Total Protein Concentration using the Bicinchoninic Acid (BCA) Assay

2.9.2 SDS Polyacrylamide Gel Electrophoresis for Western Blot Analysis

2.9.2.1 Pouring the SDS-Polyacrylamide Gels

2.9.2.2 Sample Preparation for use in SDS-Polyacrylamide Gel Electrophoresis

2.9.2.3 Loading and Running SDS-Polyacrylamide Gels

2.9.2.4 Coomassie Staining to Visualise Total Protein

2.9.3 Wet Blot Transfer of Protein to Membrane

2.9.4 Blocking of Endogenous Protein

2.9.5 Primary Antibody Incubation

2.9.6 Secondary Antibody Incubation

2.9.7 Membrane Washing and Labeling

2.9.8 Beta-actin Protein Loading Control

2.10 Immunohistochemistry for KCNH5 Protein in Placenta

2.10.1 Tissue Selection for Immunohistochemistry

2.10.2 Optimisation of Immunohistochemistry

2.10.3 Haematoxylin and Eosin Staining

2.10.4 Immunostaining of Placental Samples

2.10.4.1 Blocking of Endogenous Peroxidase

2.10.4.2 Heat Retrieval of Antigen

2.10.4.3 Antibody Staining

Chapter Three: Collection, Processing and Screening of Placental Tissue

3.1 Collection of Placental Tissue

3.1.1 Medical Terminations of Pregnancy in the First Trimester

3.1.2 Gestational Age of First-Trimester Placental Samples

3.1.3 Collection of Term Placenta

3.2 Processing of Placental Tissue

3.2.1 DNA Extraction from Placental Tissue

3.2.2 Fetal Sex Determination for Placental Samples

3.2.2.1 Detection of the Y Chromosome

3.2.2.2 Detection of the X and Y Chromosomes

3.2.3 Maternal Contamination Screening by X Chromosome Microsatellite Analysis

Chapter Four: Genome-wide Methylation Analysis by Methylated DNA Immunoprecipitation and Promoter Microarrays

4.1 MeDIP-DSL and Promoter Microarrays

4.1.1 Selection of Microarray Platform

4.1.2 Sample Selection and Experimental Design

4.2 MeDIP-DSL Microarray Results

4.2.1 Digestion of Genomic DNA

4.2.2 Immunoprecipitation, Labelling and Amplification of Samples

4.2.3 Scanning of Microarray Slides

4.3 MeDIP-DSL Microarray Data Analyses

4.3.1 Normalisation of Microarray Data

4.3.2 Data Analysis #1: Using Fold-Change Cutoff Values

4.3.3 Data Analysis #2: A Quality-Threshold (QT) Clustering Analysis
1.1 Methodology

1.2 Data Analysis

2.1 Introduction

2.2 Experiment Details

2.3 Results

3.1 Introduction

3.2 Experimental Design

3.3 Results

4.1 Introduction

4.2 Methodology

4.3 Data Analysis

4.3.1 Data Analysis #1: Largest Fold Change Difference

4.3.2 Data Analysis #2: Largest Fold Change Difference

4.3.3 Data Analysis #3: Largest Fold Change Difference

4.4 Differentially Methylated Candidate Genes

4.4.1 Location of Differential Methylation

4.4.2 CpG Density of Candidate Genes

Chapter Five: Quantifying Candidate Gene Methylation by Sequenom MassARRAY® EpiTyper® Analysis

5.1 The Sequenom MassARRAY® EpiTyper® Assay

5.2 Sample Selection and Experimental Design for Sequenom Analysis

5.2.1 Sample Selection and Processing

5.2.2 Selection and Analysis of Differentially Methylated Candidate Genes

5.3 Sequenom Analysis of Candidate Genes from MeDIP Analysis #1

5.3.1 Sequenom Analysis of GABRA6

5.3.2 Sequenom Analysis of C20orf144

5.3.3 Sequenom Analysis of KCTD4

5.3.4 Sequenom Analysis of KRT17

5.3.5 Sequenom Analysis of AKR1C4

5.4 Sequenom Analysis of Candidate Genes from MeDIP Analysis #2

5.4.1 Sequenom Analysis of NBR1

5.4.2 Sequenom Analysis of PCDHA12

5.4.3 Sequenom Analysis of RASSF2

5.4.4 Sequenom Analysis of HINT1

5.5 Sequenom Analysis of Candidate Genes from MeDIP Analysis #3

5.5.1 Sequenom Analysis of OGG1

5.5.2 Sequenom Analysis of ATAD4

5.5.3 Sequenom Analysis of CACNA2D1

5.5.4 Sequenom Analysis of C20orf19

5.5.5 Sequenom Analysis of MRPL36

5.5.6 Sequenom Analysis of COL23A1

5.5.7 Sequenom Analysis of FEM1B

5.5.8 Sequenom Analysis of FRG1

5.5.9 Sequenom Analysis of KIF22

5.5.10 Sequenom Analysis of CHKB

5.5.11 Sequenom Analysis of C6orf114

5.5.12 Sequenom Analysis of BCL9

5.5.13 Sequenom Analysis of MYO5A

5.5.14 Sequenom Analysis of LDB2

5.5.15 Sequenom Analysis of KCNH5

5.5.16 Sequenom Analysis of AICDA

5.5.17 Sequenom Analysis of SGMS1

5.5.18 Sequenom Analysis of ITGA2

5.5.19 Sequenom Analysis of GLT25D1

5.6 Summary of Sequenom Results

Chapter Six: Gene Expression Analysis of Differentially Methylated Genes

6.1 Gene Expression Analysis

6.1.1 Quantitative Reverse-Transcription PCR using TaqMan Gene Expression Assays
6.1.2 The Standard Curve for Quantitative RT-PCR 156

6.2 Gene Expression Analysis using Quantitative RT-PCR 157
6.2.1 Sample Selection and RNA Extraction 157
6.2.2 cDNA for Quantitative RT-PCRs 158
   6.2.2.1 Making cDNA 158
   6.2.2.2 Testing cDNA 158
6.2.3 Examining Reference Genes for use in RT-PCR 160
   6.2.3.1 Trialling Reference Genes 160
   6.2.3.2 Problems with Reference Genes 161
   6.2.3.3 Avoiding Reference Genes: Normalising to Total Input RNA 163

6.3 Gene Expression Results for Differentially Methylated Genes 164
6.3.1 Relative Gene Expression for AKR1C4 164
6.3.2 Relative Gene Expression for AICDA 166
6.3.3 Relative Gene Expression for ATAD4 168
6.3.4 Relative Gene Expression for KCNH5 170

6.4 Analysis of a Closely-Related Gene: KCNH1 172
6.4.1 Sequenom Analysis of KCNH1 173
6.4.2 Gene Expression Analysis for KCNH1 174
6.4.3 Sequenom Analysis of KCNH1 and KCNH5 in Cancerous Tissues 176
6.4.4 Gene Expression Analysis of KCNH1 and KCNH5 in Cancerous Tissues 177

Chapter Seven: Characterisation of KCNH5: Promoters, Transcripts and Proteins 179

7.1 Alternative Transcripts of KCNH5 179
   7.1.1 Transcript-Specific RT-PCRs 179
      7.1.1.1 Selection of Alternative KCNH5 Transcripts 179
      7.1.1.2 Design and Optimisation of Transcript-Specific RT-PCRs 180
      7.1.1.3 Results from the Transcript-Specific RT-PCRs 181
   7.1.2 Sequenom Analysis for the Promoter of KCNH5 Transcript 1b 182

7.2 KCNH5 Protein Analysis by Western Blot 185
   7.2.1 Experimental Design for Western Blot 187
      7.2.1.1 Selection of Antibody for KCNH5 187
      7.2.1.2 Selection of Samples 188
      7.2.1.3 Screening samples for western blot 188
      7.2.1.4 Protein Isolation and the BCA Assay 190
   7.2.2 Results from the Western Blot for KCNH5 190

7.3 Immunohistochemistry to Localise KCNH5 Protein 195
   7.3.1 Results for Immunohistochemistry for KCNH5 in Placenta 197
   7.3.2 Immunohistochemistry for KCNH5 Supports the Origin of Placental Mesenchyme 201

Chapter Eight: Retrotransposon-Derived Transcripts in the Placenta 203

8.1 A Retrotransposon Gave Rise to KCNH5 Transcript 1a 203
   8.1.1 Retrotransposons 203
      8.1.1.1 LTR-retrotransposons 205
      8.1.1.2 (DIRS)-like elements and PLEs 206
      8.1.1.3 LINEs 206
      8.1.1.4 SINEs 206
   8.1.2 Characterisation of the SINE retrotransposon in Transcript 1a 208
8.2 DNA Methylation Analysis of Other Retrotransposon-Derived Genes in the Placenta
8.2.1 ERVWE1 (the Syncytin-1 gene) 211
8.2.2 MIDI 213
8.2.3 EBR 216
8.2.4 PTN 218
8.2.5 INSL4 219
8.2.6 Summary of Results from the Methylation Analysis of Other Retrotransposon-Derived Genes in the Placenta 221

Chapter Nine: Discussion of Results and Future Directions 223
9.1 Epigenetics in the Placenta 223
9.2 Genome-Wide Methylation Analysis of First-Trimester Placenta 225
9.2.1 Sample Collection Results 225
9.2.2 MeDIP-Microarray Results 225
9.2.3 Sequenom Results 226
9.3 Functional Significance of Differentially Methylated Genes 226
9.3.1 Gene Expression Results 226
9.3.2 Results from Methylation and Expression Analysis of KCNH1 227
9.4 Characterisation of KCNH5 227
9.4.1 Transcript-Specific RT-PCR Results 227
9.4.2 Western Blot Results 228
9.4.3 Immunohistochemistry Results 228
9.4.4 The Placental-Specific KCNH5 Transcript is a Result of Retrotransposition 228
9.5 Methylation Analysis of Other Retrotransposon-Derived Genes in the Human Placenta 229
9.6 What is the Role of Retrotransposons in the Placenta? 230
9.7 What is the Epigenetic Link between Trophoblasts and Tumours? 232
9.8 Future Directions for Epigenetic Analyses in the Placenta 233
9.8.1 Methods of Epigenetic Analysis 233
9.8.2 Placental Tissue 235
9.8.3 Future Studies 235
9.9 Conclusions 236

References 239
List of Tables

Chapter Two:

Table 2.1. Primer sequences for sex-determining PCRs. 38
Table 2.2. Primer sequences and PCR conditions for X chromosome microsatellite analysis. 39
Table 2.3. First-trimester placenta sample selection and pooling for methylated DNA immunoprecipitation and promoter microarray assay. 43
Table 2.4. Peripheral blood control sample selection and pooling for methylated DNA immunoprecipitation and promoter microarray assay. 44
Table 2.5. Primer sequences and PCR conditions for genes analysed using the Sequenom MassARRAY® EpiTYPER® platform. 56
Table 2.6. TaqMan® gene expression assays used for gene expression analysis by quantitative real-time PCR. 63
Table 2.7. Primer sequences for the B2M PCR to test cDNA. 64
Table 2.8. Samples used for cDNA standards in quantitative RT-PCR. 65
Table 2.9. Primer Sequences and RT-PCR Conditions for KCNH5’s Transcript-Specific RT-PCRs. 68

Chapter Three:

Table 3.1. First-trimester placental samples collected for methylation analysis. 88

Chapter Four:

Table 4.1. Fold change values for candidate differentially methylated genes identified at the 1.4-fold cutoff in MeDIP microarray analysis #1. 104
Table 4.2. Candidate genes selected from gene sets in the quality threshold (QT) clustering analysis of microarray data at the 1.5-fold threshold. 108
Table 4.3. The genes with the largest fold change difference between peripheral blood and first-trimester placenta. 110
Table 4.4. The Aviva Systems Biology probe sequences for candidate differentially methylated genes. 111

Chapter Five:

Table 5.1. Fetal and adult tissues processed for Sequenom analysis. 115
Table 5.2. Samples used for Sequenom analysis. 116
Table 5.3. Sequenom results showing biologically meaningful genes. 118
Table 5.4. CpG density and Sequenom validation of candidate differentially methylated genes.

Chapter Six:
Table 6.1. Samples selected for gene expression analysis.
# List of Figures

## Chapter One:

**Figure 1.1.** The structure of the human haemochorial placenta and associated pregnancy tissues.  
2  
**Figure 1.2.** Cell division and lineage specification during the earliest stages of human embryonic development.  
4  
**Figure 1.3.** Blastocyst apposition, adhesion and invasion during human embryonic development, adapted from Norwitz *et al.* 2001.  
6  
**Figure 1.4.** The anchoring cell column comprised of the villous and extra-villous trophoblasts.  
8  
**Figure 1.5.** Establishing DNA methylation marks by DNA methyltransferase enzymes, adapted from Waterland and Michels, 2007.  
17

## Chapter Two:

**Figure 2.1.** Methylated DNA immunoprecipitation (MeDIP) using the Aviva Systems Biology DNA selection and ligation (DSL) platform and 20K promoter microarrays.  
42  
**Figure 2.2.** The Sequenom MassARRAY® EpiTYPER® platform (adapted from Ehrich *et al.*, 2005).  
53  
**Figure 2.3.** Primer design for Sequenom MassARRAY® EpiTYPER® platform.  
54  
**Equation 2.1.** The formula used by the Sequenom MassARRAY® EpiTYPER® software to determine the percentage methylation of a fragment with \( n \) CpG sites (Coolen *et al.*, 2007).  
62  
**Equation 2.2.** The equations used to calculate gene expression from quantitative RT-PCR data.  
67  
**Equation 2.3.** The equations used to calculate total protein concentration in samples for the western blot.  
71  
**Figure 2.4.** The EnVision™ Dual Link detection system using DAB substrate for immunohistochemistry on placental tissue to localise KCNH5 protein.  
75

## Chapter Three:

**Figure 3.1.** Photographs of a selection of collected placental samples from terminated pregnancies in the first trimester.  
81  
**Figure 3.2.** Photographs through a microscope of isolated placental tissue containing trophoblast villi.  
82  
**Figure 3.3.** The region of the *SRY* gene targeted to determine the fetal sex of placental samples.  
84
Figure 3.4. Results from the amplification of a region of the SRY gene to detect the Y chromosome in placental samples.

Figure 3.5. The region of the AMEL gene homolog on the X and Y chromosomes targeted to determine the fetal sex of placental samples.

Figure 3.6. Results from the amplification of a region of the AMEL gene to detect the X and Y chromosomes in placental samples.

Figure 3.7. The microsatellite markers on the X chromosome used to screen placental samples for maternal contamination.

Figure 3.8. Results from microsatellite marker ss4915376 at Xq26.2 to screen for maternal contamination in placental samples.

Figure 3.9. Results from microsatellite marker DXS8051 at Xp22.31 to screen for maternal contamination in placental samples.

Figure 3.10. Results from microsatellite marker DXS6803 at Xq21.31 to screen for maternal contamination in placental samples.

Figure 3.11. Results from microsatellite marker DXS8043 at Xq27.3 to screen for maternal contamination in placental samples.

Figure 3.12. Results from microsatellite marker DXS8377 at Xq28 to screen for maternal contamination in placental samples.

Figure 3.13. Results from five microsatellite markers on the X chromosome to screen for maternal contamination in placental samples.

Chapter Four:

Figure 4.1. Optimisation of the NlaIII restriction enzyme on 1 µg genomic DNA in a 20 µL digest visualised on a 0.8% agarose gel.

Figure 4.2. The PCR screening of samples to test the ligation reaction.

Figure 4.3. The PCR screening of samples to test the labelling reaction.

Figure 4.4. The six candidate genes identified in the MeDIP-microarray experiment using a 1.4-fold change threshold cutoff.

Figure 4.5. The quality threshold (QT) clustering analysis (≥ 10 genes per group, R ≥ 0.7) for the MeDIP-microarray data at the 1.5-fold threshold.

Figure 4.6. Genes that show the largest changes between peripheral blood and first-trimester placenta.

Chapter Five:

Figure 5.1. Sequenom results for GABRA6.

Figure 5.2. Sequenom results for C20orf144.

Figure 5.3. Sequenom results for KCTD4.

Figure 5.4. Sequenom results for KRT17.

Figure 5.5. Sequenom results for AKR1C4.

Figure 5.6. Sequenom results for NBR1.

Figure 5.7. Sequenom results for PCDHA12.
Figure 5.8. Sequenom results for RASSF2.
Figure 5.9. Sequenom results for HINT1.
Figure 5.10. Sequenom results for OGG1.
Figure 5.11. Sequenom results for ATAD4.
Figure 5.12. Sequenom results for CACNA2D1.
Figure 5.13. Sequenom results for C20orf19.
Figure 5.14. Sequenom results for MRPL36.
Figure 5.15. Sequenom results for COL23A1.
Figure 5.16. Sequenom results for FEM1B.
Figure 5.17. Sequenom results for FRG1.
Figure 5.18. Sequenom results for KIF22.
Figure 5.19. Sequenom results for CHKB.
Figure 5.20. Sequenom results for C6orf114.
Figure 5.21. Sequenom results for BCL9.
Figure 5.22. Sequenom results for MYO5A.
Figure 5.23. Sequenom results for LDB2.
Figure 5.24. Sequenom results for KCNH5.
Figure 5.25. Sequenom results for AICDA.
Figure 5.26. Sequenom results for SGMS1.
Figure 5.27. Sequenom results for ITGA2.
Figure 5.28. Sequenom results for GLT25D1.

Chapter Six:
Figure 6.1. PCR screening of cDNA for the B2M gene to test the RT reaction.
Figure 6.2. Mean Ct values after RT-PCR for B2M and TOP3A.
Figure 6.3. Expression of AKR1C4 normalised to input RNA compared to mean promoter methylation.
Figure 6.4. Expression of AICDA normalised to input RNA compared to mean promoter methylation.
Figure 6.5. Expression of ATAD4 normalised to input RNA compared to mean promoter methylation.
Figure 6.6. Expression of KCNH5 normalised to input RNA compared to mean promoter methylation.
Figure 6.7. Sequenom results for KCNH1.
Figure 6.8. Expression of KCNH1 normalised to input RNA compared to mean promoter methylation.
Figure 6.9. Sequenom analysis for KCNH1 and KCNH5 in cancerous tissues.
**Figure 6.10.** KCNH1 and KCNH5 expression and promoter methylation in cancerous tissues.

Chapter Seven:

**Figure 7.1.** Genomic arrangement of KCNH5’s alternative transcripts.

**Figure 7.2.** Primer design to identify expression of KCNH5’s alternative transcripts.

**Figure 7.3.** Results from the amplification of cDNA for two transcripts of KCNH5.

**Figure 7.4.** Sequenom results for KCNH5 transcript 1b.

**Figure 7.5.** The six human mRNA sequences for KCNH5 and the protein isoforms they encode.

**Figure 7.6.** Screening samples for the western blot by quantitative RT-PCR.

**Figure 7.7.** Western blot results for KCNH5.

**Figure 7.8.** Arrangement of the voltage-gated potassium channels encoded by the various isoforms of KCNH5.

**Figure 7.9.** KCNH5 protein localised in first-trimester placenta by immunohistochemistry.

**Figure 7.10.** KCNH5 protein localised in second and third trimester placenta by immunohistochemistry.

Chapter Eight:

**Figure 8.1.** Genomic arrangement of the SINE (AluY) retrotransposon located in the region of differential KCNH5 methylation.

**Figure 8.2.** Classification of retrotransposons modified from Wicker et al. 2007.

**Figure 8.3.** The SINE (AluY) retrotransposon located in upstream and in the first 162 bp of KCNH5’s exon 1a.

**Figure 8.4.** Genomic arrangement of ERVWE1 (the Syncytin-1 gene).

**Figure 8.5.** Sequenom results for ERVWE1.

**Figure 8.6.** Genomic arrangement of MID1 adapted from Reiss et al. 2007.

**Figure 8.7.** Sequenom results for MID1.

**Figure 8.8.** Genomic arrangement of EBR adapted from Reiss et al. 2007.

**Figure 8.9.** Sequenom results for EBR.

**Figure 8.10.** Genomic arrangement of PTN adapted from Reiss et al. 2007.

**Figure 8.11.** Sequenom results for PTN.

**Figure 8.12.** Genomic arrangement of INSL4.

**Figure 8.13.** Sequenom results for INSL4.
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>aa</td>
<td>Amino acid</td>
</tr>
<tr>
<td>ABI</td>
<td>Applied Biosystems</td>
</tr>
<tr>
<td>ALL</td>
<td>Acute lymphoblastic leukaemia</td>
</tr>
<tr>
<td>APS</td>
<td>Ammonium persulfate</td>
</tr>
<tr>
<td>ASB</td>
<td>Aviva Systems Biology</td>
</tr>
<tr>
<td>bp</td>
<td>Base pair(s)</td>
</tr>
<tr>
<td>BCA</td>
<td>Bicinchoninic acid</td>
</tr>
<tr>
<td>cDNA</td>
<td>Complementary DNA</td>
</tr>
<tr>
<td>CpG</td>
<td>Cytosine base adjacent to a guanine base separated by a phosphate</td>
</tr>
<tr>
<td>cNBD</td>
<td>Cyclic nucleotide binding domain</td>
</tr>
<tr>
<td>DAB</td>
<td>Diaminobenzidine</td>
</tr>
<tr>
<td>ddH$_2$O</td>
<td>Double distilled water</td>
</tr>
<tr>
<td>DEPC</td>
<td>Diethyl pyrocarbonate</td>
</tr>
<tr>
<td>DMSO</td>
<td>Dimethylsulfoxide</td>
</tr>
<tr>
<td>DNMT</td>
<td>DNA methyltransferase</td>
</tr>
<tr>
<td>DSL</td>
<td>Deoxynucleic acid selection and ligation</td>
</tr>
<tr>
<td>EDTA</td>
<td>Ethylenediaminetetraacetic acid</td>
</tr>
<tr>
<td>FCS</td>
<td>Fetal calf serum</td>
</tr>
<tr>
<td>HERV</td>
<td>Human endogenous retrovirus</td>
</tr>
<tr>
<td>HRP</td>
<td>Horseradish peroxidase</td>
</tr>
<tr>
<td>IVF</td>
<td>In vitro fertilisation</td>
</tr>
<tr>
<td>LCM</td>
<td>Laser capture microdissection</td>
</tr>
<tr>
<td>LINE</td>
<td>Long interspersed nuclear element</td>
</tr>
<tr>
<td>LTR</td>
<td>Long terminal repeat</td>
</tr>
<tr>
<td>MeDIP</td>
<td>Methylated deoxynucleic acid immunoprecipitation</td>
</tr>
<tr>
<td>MMPs</td>
<td>Matrix metalloproteinases</td>
</tr>
<tr>
<td>mRNA</td>
<td>Messenger ribonucleic acid</td>
</tr>
<tr>
<td>NCBI</td>
<td>National Centre for Biotechnology Information</td>
</tr>
<tr>
<td>NEB</td>
<td>New England Biolabs</td>
</tr>
<tr>
<td>PAS</td>
<td>Per-Ant-Sim domain</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Name</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>PBS</td>
<td>Phosphate buffered saline</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase chain reaction</td>
</tr>
<tr>
<td>PMT</td>
<td>Photo-multiplier tube</td>
</tr>
<tr>
<td>QT</td>
<td>Quality threshold</td>
</tr>
<tr>
<td>RNA</td>
<td>Ribonucleic acid</td>
</tr>
<tr>
<td>rpm</td>
<td>Revolutions per minute</td>
</tr>
<tr>
<td>RT-PCR</td>
<td>Reverse transcription polymerase chain reaction</td>
</tr>
<tr>
<td>SAP</td>
<td>Shrimp alkaline phosphatase</td>
</tr>
<tr>
<td>SDS</td>
<td>Sodium dodecyl sulfate</td>
</tr>
<tr>
<td>Sequenom</td>
<td>Sequenom MassARRAY® EpiTYPER® Assay</td>
</tr>
<tr>
<td>SINE</td>
<td>Short interspersed nuclear element</td>
</tr>
<tr>
<td>TAE</td>
<td>Tris-acetate EDTA buffer</td>
</tr>
<tr>
<td>TBE</td>
<td>Tris-borate EDTA buffer</td>
</tr>
<tr>
<td>TBS</td>
<td>Tris-buffered saline</td>
</tr>
<tr>
<td>TE</td>
<td>Tris-EDTA buffer</td>
</tr>
<tr>
<td>TG</td>
<td>Tris-glycine</td>
</tr>
<tr>
<td>Tris</td>
<td>Tris(hydroxymethyl)aminomethane</td>
</tr>
<tr>
<td>UCSC</td>
<td>University of California Santa Cruz (Genome Browser)</td>
</tr>
</tbody>
</table>
Chapter One

Introduction

1.1 The Human Placenta

Every human life relies on the placenta. It is an indispensible human organ, even though its development and function is limited to nine months. During pregnancy, the placenta mediates the transfer of oxygen and nutrients from the mother to the fetus while simultaneously disguising the fetus from the mother’s immunological rejection mechanisms (Cross et al., 1994; Benirschke, 1998). Being the first major organ to develop, the placenta must perform these critical functions for a successful maternal-fetal relationship to be established and for the pregnancy to carry to term. Failure of placental development and function is also associated with multiple placental pathologies, many of which are common yet poorly understood. Although its central role in pregnancy is undoubtedly clear, the human placenta itself is a relatively unexplored yet fascinating transient human tissue.

1.1.1 Placental Function

The most fitting description of the placenta was made by Mossman as “an apposition of parental and fetal tissue for the purposes of physiological exchange” (Mossman, 1991), also cited in (Wooding, 2008). One of the main functions of the placenta is to act as an organ surrogate until the fetal tissues develop. Specifically, the placenta takes on the roles of the fetal lungs, gut and kidneys, as it is responsible for the transport of oxygen and nutrients and the elimination of waste from the fetal circulation (Moll, 1985). These roles are achieved by establishing an intimate vascular connection between maternal and fetal blood, which is critical for the survival of pregnancy (Burton et al., 2009). The second main function of the placenta is to provide an ‘immunological camouflage’, protecting the fetus from the maternal immune system that would otherwise attack the ‘invading’ conceptus (Lala et al., 1983). The placenta consequently acts as a physical barrier between the mother and fetus. It must establish a delicate equilibrium so that the maternal immune system can tolerate the necessary fetal requirements yet continue to protect the mother from excessive fetal intrusion (Moffett and Loke, 2006).
The two main functions of the placenta - simultaneous nourishment and immunological disguise - are in conflict with each other. Handling these discordant functions has led to the development of the placenta’s unique structures, all of which have adapted to suit the different requirements of each group of placental mammals (Wooding, 2008).

### 1.1.2 Placental Structure

Within the eutharian (placental) mammals, three main types of placentas have evolved. These three groups are categorised according to the depth of placental invasion into maternal tissue, and include the non-invasive epitheliochorial placenta, the moderately invasive endotheliochorial placenta and the highly invasive haemochorial placenta (Martin, 2003). It is believed that the haemochorial placenta is the ancestral state of the placenta for all eutherian mammals (Wildman et al., 2006). The human placenta is haemochorial, meaning that the cells of the placenta are in direct contact with maternal blood (Figure 1.1). The cells of the placenta – the trophoblasts – are responsible for establishing the maternal-fetal blood connection and will be discussed in more detail in the following section.

---

**Figure 1.1.** The structure of the human haemochorial placenta and associated pregnancy tissues. The fetus is surrounded by the amnion membrane (blue), which is adjacent to the chorion membrane (green). The umbilical cord (pink) connects the fetus to the placenta (dark red). The trophoblast cells invade into the maternal decidua to remodel maternal spiral arteries and establish a blood supply for the developing fetus.
In addition to the placenta, there are other important tissues that develop during pregnancy, including the chorion, amnion and umbilical cord. The umbilical cord is the connecting membrane from the fetus to the placenta. It contains two arteries and one vein that are responsible for the exchange of nutrients, gases and waste between the placenta and fetus. The umbilical cord is attached to the amnion, the innermost membrane that surrounds and protects the fetus. The chorion membrane is adjacent and adherent to the amnion and is attached to the placenta (Figure 1.1). These structures form in conjunction with the placenta during the early stages of human embryonic development.

1.2 Development of the Human Placenta

The human placenta is a highly specialized organ that undergoes extensive structural changes during pregnancy. These developmental events are critical to functionally-coordinate the placenta with the developing fetal and maternal components (Ohlsson, 1993). Although the placenta grows continuously during the nine months of pregnancy, the most important developmental stages occur during the first trimester. After fertilisation, a series of cell divisions prepare the embryo for implantation; the stage at which placental development begins (Benirschke et al., 2006). Documentation of the earliest stages of human embryonic development is based on studies in animals given the lack of human specimens available from these early stages of gestation (Norwitz et al., 2001; Cross, 2005).

1.2.1 Fertilisation and Cell Lineage Specification

Human embryonic development begins with fertilisation, an event that occurs in the fallopian tube approximately 24 to 48 hours after ovulation. The fertilised ovum, termed the zygote, undergoes a series of cell divisions and becomes the morula; a mass of eight to 16 cells encased by a protective coating called the zona pellucida (Figure 1.2). The morula travels down the fallopian tube and enters the uterine cavity roughly two to three days post fertilisation. Once the morula has grown to a mass of 16 to 32 cells, it becomes the blastocyst, containing a fluid-filled cavity and displaying signs of cell lineage specification.

The first cell fate decision is made when the surface cells of the blastocyst become the trophectoderm, a lineage that is distinguished from the inner cell mass (Figure 1.2) (Enders and Schlafke, 1981). The trophectoderm then gives rise to the trophoblast; the primary cell type that will form the placenta. The inner cell mass undergoes a second wave of lineage specification when it further divides into the primitive endoderm and the epiblast lineages
(Figure 1.2) (Enders et al., 1986). The trophectoderm and primitive endoderm lineages give rise to the extra-embryonic structures such as the placenta, chorion and yolk sac, whereas the epiblast lineage gives rise to the embryonic tissues, including the embryo, amnion and umbilical cord (Enders and Schlafke, 1981; Cross et al., 1994; Benirschke et al., 2006; Rossant, 2007). The lineage origin of the placental stroma, however, remains unclear and will be discussed in more detail in Chapter Seven (Luckett, 1978; Enders and King, 1988). Once these three cell lineages are established - the trophectoderm, primitive endoderm and epiblast - the blastocyst is ready for implantation.

**Figure 1.2.** Cell division and lineage specification during the earliest stages of human embryonic development. After fertilisation, the zygote undergoes a series of cell divisions and becomes the morula, encased by the zona pellucida (grey). The presence of a fluid-filled cavity (yellow) and first cell lineage differentiation event occurs at the blastocyst stage, when the trophectoderm (red) is distinguished from the inner cell mass (purple). The second stage of cell lineage specification occurs when the inner cell mass divides into the primitive endoderm (green) and the epiblast (blue). Of the three established lineages, the trophectoderm and primitive endoderm will give rise to the extra-embryonic structures including the placenta, chorion and early yolk sac. The epiblast will give rise to the embryonic structures including the amnion, umbilical cord and the fetus.
1.2.2 Blastocyst Implantation

Implantation is a complex and coordinated process that requires extensive cross-talk and molecular interaction between the active blastocyst and the receptive maternal endometrium (Enders, 2000; Lim et al., 2002). Upon entry into the uterine cavity, the blastocyst hatches from the protective shell of the zona pellucida, exposing the trophoblast stem cells to the maternal endometrium. The physical alignment and temporal-spatial synchronization of both fetal and maternal tissues is critical in order for the blastocyst to be sequestered into the uterine tissue (Paria et al., 2002). Studies in primates suggest that implantation occurs roughly six to seven days after fertilisation and is likely to involve three stages: blastocyst apposition, adhesion and invasion into maternal decidua (Enders and Lopata, 1999; Norwitz et al., 2001).

Although the blastocyst naturally displays an invasive potential, the uterus is innately programmed to prevent invasion. In order for implantation to occur, there is a critical window of uterine receptivity that allows for blastocyst invasion (Psychoyos, 1986). The ovarian hormones, oestrogen and progesterone, are responsible for activating various molecules to induce uterine receptivity, including uterine growth factors, cytokines, lipid mediators and transcription factors (Salamonsen et al., 2000; Paria et al., 2002). This hormone signalling transforms the pre-receptive uterus into a receptive state, which allows the blastocyst to subsequently attach and invade into the maternal decidua. During implantation, the blastocyst is positioned so that the embryonic pole (the part of the blastocyst containing the inner cell mass) makes the initial contact with the endometrial tissue (Figure 1.3) (Benirschke et al., 2006).

Once the blastocyst comes in contact with the receptive endometrium, it undergoes a stage of adhesion known as the “implantation window” (Edwards, 1988). This is a short yet critical period during which the blastocyst must attach to the uterine wall. The attachment of the blastocyst’s apical membrane to the uterine apical membrane is a mystery to cell biologists, given that apical plasma membranes of epithelial tissues are not normally adhesive (Denker, 1993). Nonetheless, this implantation window allows for the required adhesion of blastocyst-uterine apical membranes. Blastocyst attachment is followed by a much longer phase of invasion by placental trophoblast cells (Figure 1.3). At this stage in development, the trophoblast becomes responsible for the growth and function of the placenta and for the continuation of pregnancy.
Figure 1.3. Blastocyst apposition, adhesion and invasion during human embryonic development, adapted from Norwitz et al. 2001. The fluid-filled blastocyst enters the uterine cavity after exiting the fallopian tube. Hormone signalling induces uterine receptivity, which allows the blastocyst to adhere to the uterine wall. After implantation, trophoblast cells undergo stages of proliferation, migration and invasion into maternal decidua to anchor the developing placenta and establish a trophic connection for the growing embryo.

1.2.3 Trophoblast Invasion

1.2.3.1 The Trophoblast

The trophoblast cell was given its name back in 1889 when Hubrecht characterised it as a cell that does not give rise to the embryo but is fundamental for its nourishment (Hubrecht, 1889). In fact, the term “tropho” is Greek for “nutrition”, a term that precisely describes the cell’s function (Moffett, 2006). Upon implantation, trophoblast cells undergo strictly regulated phases of proliferation, migration and invasion to establish an intimate trophic connection between the mother and fetus (Bischof and Irminger-Finger, 2005; Pollheimer and Knofler, 2005).

Survival of the fetus is hinged upon the ability of the trophoblasts to differentiate into their respective sub-types, and for each sub-type to undergo regulated phenotypic and functional changes. The behaviour of the trophoblast is controlled by many elements in the extracellular matrix as well as cell adhesion molecules and growth factors (Chakraborty et al., 2002;
Pollheimer and Knofler, 2005). As a result of this coordinated and controlled invasion, the placenta can assume a branched villous structure that maximizes contact to maternal blood (Figure 1.3). The mature placenta is therefore able to mediate the biological exchanges necessary between mother and fetus, which is critical for successful pregnancy.

Once the blastocyst has implanted into the uterine wall, the cells of the trophectoderm become the cytotrophoblast cells, which are mono-nucleated proliferative cells that function as the stem cells of the placenta (Figure 1.4) (Cross et al., 2003). After implantation, cytotrophoblasts can differentiate along one of two pathways, resulting in the formation of either the villous or the extra-villous trophoblast (Loke, 1995; Duc-Goiran et al., 1999).

1.2.3.2 The Villous Trophoblast

To form the villous trophoblast, cytotrophoblasts fuse to form the syncytiotrophoblast layer of the placental villous (Figure 1.4). The syncytiotrophoblast is a multi-nucleate, continuous layer of cells that functions as the physical barrier between the mother and the fetus, responsible for mediating gas exchange, substrate transport and waste elimination (Benirschke et al., 2006). The syncytiotrophoblast serves as the endocrine tissue of the placenta as it secretes large amounts of protein hormones, including human chorionic gonadotropin (hCG); the first hormone secreted by the trophoblast cells that is used to diagnose pregnancy (Jameson and Hollenberg, 1993). The syncytiotrophoblast is also responsible for placental villous self-renewal, which involves the shedding of aged nuclei by the formation of syncytial knots and bridges that are eventually cleared away by the maternal circulation (Benirschke et al., 2006).

1.2.3.3 The Extra-Villous Trophoblast

During trophoblast invasion, cytotrophoblast cells detach from the villous basement membrane and differentiate into extra-villous trophoblasts, cells that are responsible for the physical attachment of the placenta to the uterine wall. To achieve this, extra-villous trophoblasts form anchoring cell columns at the villous tips that eventually extend into the maternal decidua (Damsky et al., 1992; Fisher and Damsky, 1993; Aplin et al., 1999) (Figure 1.4). The population of extra-villous trophoblasts is comprised of both proliferative and invasive cells, both of which have spatially and temporally regulated phenotypes (Nishimura et al., 2004). Extra-villous trophoblasts within the cell column are proliferative and non-invasive, whereas the extra-villous trophoblasts that exit the cell column become invasive and non-proliferative (Genbacev et al., 2000).
There are two types of invasive extra-villous trophoblasts: the interstitial and the endovascular trophoblasts. As shown in Figure 1.4, interstitial extra-villous trophoblasts are those cells that lose contact with the anchoring cell column and migrate through maternal decidua, eventually reaching the inner third of the myometrium. Endovascular extra-villous trophoblasts are a sub-set of the interstitial trophoblasts, and are responsible for remodelling maternal spiral arteries to increase blood flow for the developing fetus (Figure 1.4) (Zhou et al., 1997; Benirschke et al., 2006).

**Figure 1.4.** The anchoring cell column comprised of the villous and extra-villous trophoblasts. The original cytotrophoblast (pink) differentiates into either the villous or extra-villous trophoblast. Cytotrophoblasts fuse to form the villous syncytiotrophoblast layer (yellow). Upon contact with maternal decidua, the cytotrophoblast becomes the extra-villous trophoblast (blue, green, purple). First the extra-villous trophoblast displays a proliferative (blue) phenotype and then transitions to display an invasive (green) phenotype. The interstitial extra-villous trophoblast (green) migrates through maternal decidua. The endovascular extra-villous trophoblast (purple) remodels maternal spiral arteries.

### 1.3 Molecular Mechanisms Regulating Trophoblast Invasion

Many different molecules and signalling pathways have been proposed to play a key role in regulating the migration and invasion of the extra-villous trophoblast during placental development. These molecules and pathways were identified using *in vitro* models including...
isolated primary cultures, explant cultures and trophoblast cell lines (Fisher and Damsky, 1993; LaMarca et al., 2005; Pollheimer, 2005). These models have successfully displayed both the two-dimensional and three-dimensional propagation of trophoblast cells and have greatly expanded the knowledge of trophoblast biology. However, although many studies have associated various molecules and pathways with the establishment of the maternal-fetal interface, the exact mechanisms that regulate trophoblast behaviour have yet to be discovered.

1.3.1 Factors from the Maternal Decidua

Prior to invasion, the extra-villous trophoblast cells must change from a proliferative trophoblast to an invasive trophoblast in order for it to detach from the anchoring cell column and invade into the maternal decidua. Interestingly, this ‘phenotype switch’ has been associated with various factors secreted by the maternal decidua (Starkey et al., 1988; Wright et al., 2006). In a study by Wright and colleagues, first trimester placental villous explants were prepared and treated with maternal decidua-cultured media to create an in vivo environment for the extra-villous trophoblast. An immunohistochemical analysis revealed that, when in the presence of decidua-cultured media, markers of the proliferative extra-villous trophoblast phenotype were down-regulated (e.g., Connexin 40, Ki67, EGF-R and α5 integrin), while markers for the invasive extra-villous trophoblast phenotype were up-regulated (e.g., c-erbB2 and α1 integrin) (Wright et al., 2006). These changes in protein expression in the presence of decidua-cultured media suggested that the transition between the proliferative and invasive phenotypes of the extra-villous trophoblast is regulated by factors secreted by the maternal decidua.

1.3.2 Cell Surface Molecules

Upon detachment from the anchoring cell column, the invasion of the interstitial extra-villous trophoblast through the epithelial layer of the endometrium is mediated by proteases and cell adhesion molecules located on the trophoblast cell surface (Damsky et al., 1992; Fisher and Damsky, 1993). The best characterised of these are the matrix metalloproteinases (MMPs), which are zinc-dependent proteases that degrade the extra-cellular matrix as the trophoblast invades through the maternal decidua (Duc-Goiran et al., 1999). Collagenase is one of the main protein-cleaving MMPs and can degrade collagen IV, V, and VII (Baker and Kingdom, 2004). In a study by Fisher et al., the inhibition of collagenase activity stopped the invasion of interstitial extra-villous trophoblasts in vitro, confirming the critical role of MMPs in trophoblast invasion (Fisher and Damsky, 1993). Another molecule, endostatin - a protein fragment of collagen XVIII generated during trophoblast migration - is suggested to be a
negative regulator of trophoblast invasion, responsible for controlling the depth of invasion into maternal decidua (Dixelius et al., 2002; Pollheimer et al., 2005).

In addition to MMPs, various cell adhesion and immunological molecules also play a major role in trophoblast behaviour. Interestingly, during trophoblast invasion, the expression of the trophoblasts’ epithelial cell adhesion receptors is down regulated so that it can migrate through decidua, whereas the expression of its vascular cell adhesion receptors is up regulated so that the trophoblast can adhere to the walls of the uterine arteries (Zhou et al., 1997). Furthermore, as the trophoblast invades into maternal tissue, it expresses class I human leukocyte molecules, specifically HLA-G, in order to suppress the maternal immune system from responding to the developing fetus (Kovats et al., 1990; McMaster et al., 1995). Regarding the maternal immune system, the interaction of maternal macrophages with the invading trophoblast is thought to be important in regulating the critical function of endovascular trophoblast apoptosis (Reister et al., 1999; Reister et al., 2001). The apoptosis of endovascular trophoblasts has been shown to be important for the proper transformation of maternal arteries, whereas apoptosis of interstitial trophoblasts is thought to contribute to shallow trophoblast invasion (Huppertz et al., 2005).

1.3.3 Effect of Oxygen Tension

During extra-villous trophoblast invasion into the uterine wall, the interstitial trophoblasts release vasodilators (such as nitric oxide and carbon monoxide) to prepare the arteriole walls for destruction (Lyall, 2005). Upon contact with a maternal artery, the endovascular trophoblasts form plugs that prevent the artery from spilling blood into the inter-villous space. The formation of these plugs creates a hypoxic environment for the placenta and fetus (Hustin and Schaaps, 1987; Burton et al., 1999). Studies have shown that as the pregnancy progresses, the difference in oxygen tension between the placenta and the maternal circulation decreases, and the oxygen levels in the placenta almost match that of the endometrium by the thirteenth week of gestation (Rodesch et al., 1992). At this time, which is the end of the first trimester of pregnancy, the endovascular trophoblast plugs in the arterial walls loosen and the placenta is exposed to maternal blood flow (Jauniaux et al., 2003).

It is thought that the endovascular trophoblast is able to sense the oxygen tension as it remodels the maternal spiral arteries, ensuring that an adequate blood supply is established for the pregnancy to continue (Caniggia and Winter, 2002). Several trophoblast oxygen-sensing mechanisms have been proposed, many of which include pathways that regulate gene
expression and thus trophoblast behaviour. These trophoblast oxygen-sensing mechanisms commonly rely on the formation of reactive oxygen species combined with redox-sensitive transcription factors to control gene expression during the first trimester of pregnancy; the time during which trophoblast activity is at its peak. The hypoxia inducible factor (HIF) family is the most well characterised transcription factor in the trophoblast that is responsive to hypoxic conditions. Interestingly, the two groups of genes that are affected by the placenta’s hypoxic condition are genes that promote cell survival or genes that are directly involved in implantation and placentation (James et al., 2006).

In addition to the hypoxia-sensitive genes, a large number of other genes have been identified that regulate the behaviour of the trophoblast and the ultimate formation of the placenta (Hemberger and Cross, 2001). However, although various molecules, signalling pathways and genes have been associated with trophoblast differentiation and invasion, no robust molecular mechanism has yet been found to account for the developmental progression of the placenta. Such a discovery could lead to the design of therapies that could treat placental-related pathologies.

1.4 Placental Pathologies

Placental dysfunction is usually associated with failed trophoblast invasion; however, it is also a result of failed decidualisation of the maternal endometrium. Prior to trophoblast invasion and transformation of maternal spiral arteries, the uterine tissue must undergo an initial decidualisation process to prepare itself to accept the impending invasion (Brosens et al., 2002; Bischof and Irminger-Finger, 2005). Decidualisation is a tightly regulated process, controlled by various uterine hormones, cytokines and growth factors. The result is a structurally and chemically modified uterine environment; primed to interact with the extravillous trophoblast and allow its invasion and remodelling of maternal arteries (Chelbi and Vaiman, 2008). Failure of either process - decidualisation and/or trophoblast invasion - results in abnormal placentation. In addition to pregnancy loss, the two most common placental pathologies are pre-eclampsia and placental insufficiency (resulting in intra-uterine growth restriction), although various other disorders have also been reported.

1.4.1 Pre-eclampsia

Pre-eclampsia is a placental disorder that occurs in approximately 5% of pregnancies worldwide (Sibai et al., 1997; Chelbi and Vaiman, 2008). The disorder is caused by shallow
invasion of trophoblasts into maternal spiral arteries (Gerretsen et al., 1981; Kaufmann et al., 2003). This defective trophoblast invasion leads to an increased vascular resistance in the poorly remodelled maternal spiral arteries, which is associated with the clinical symptoms of maternal hypertension and large amounts of protein in the urine. Although the main cause of pre-eclampsia occurs during the first trimester of pregnancy, the clinical symptoms do not present until mid to late gestation, so treatment of the disorder is limited to the premature delivery of the placenta. If no action is taken, pre-eclampsia can be fatal to both the mother and fetus. Although this placental dysfunction is commonly thought to be fetally-induced, it has been reported that maternal factors may also be implicated in the pathogenesis of pre-eclampsia (Roberts and Lain, 2002).

1.4.2 Placental Insufficiency (Intra-Uterine Growth Restriction)

Placental insufficiency commonly results in intra-uterine growth restriction, a condition where the fetus does not achieve its growth potential in utero, resulting in newborns of low birth weight (Ergaz et al., 2005). Although there may be fetal genetic factors contributing to its onset, placental insufficiency is mainly caused by an increase in villous trophoblast apoptosis in response to the hypoxic placental environment and pro-inflammatory cytokines, resulting in limited blood flow and nutrient transfer to the fetus (Holcberg et al., 2001; Ishihara et al., 2002; Baschat, 2004). The increase in trophoblast apoptosis is displayed by the increased occurrence of syncytial knots in the placenta of an intra-uterine growth restricted fetus. Syncytial knots are a useful marker to detect trophoblast apoptosis given that they represent the syncytiotrophoblast shedding dead cellular material from the placental villous (Levy et al., 2002). There is no single defining condition of fetuses that suffer from intra-uterine growth restriction, instead the condition is described as suboptimal fetal growth and disproportionate body features (Scifres and Nelson, 2009).

1.4.3 Placenta Accreta, Increta and Percreta

Other placental pathologies caused by abnormal trophoblast invasion are placenta accreta, increta and percreta, which affect roughly one in 2,500 pregnancies (Breen et al., 1977; Pelosi, 1998). Placenta accreta is when the trophoblasts invade through the endometrium and into the superficial layers of the myometrium, the next layer of the uterine wall. A more serious condition is placenta increta, where the trophoblasts invade even further into the myometrium. However, placenta percreta is the most severe, characterised by trophoblast invasion through the entire uterine wall and attachment onto other maternal organs, such as the bladder or rectum (Al-Serehi et al., 2008).
**1.4.4 Tumours of the Placenta**

Excessive trophoblast invasion has been associated with placental tumours, such as the hydatidiform mole and choriocarcinoma (Bischof and Irminger-Finger, 2005; Benirschke et al., 2006). The hydatidiform mole is a benign placental tumour that presents as deteriorated placental villi. Since the hydatidiform mole is typically the result of a bipaternal (androgenic) conceptus, this pathology is commonly associated with non-viable pregnancies. However, rarely a hydatidiform mole can turn cancerous and develop into a choriocarcinoma; a malignant placental tumour that displays complete destruction of the placental villous bed. Although it appears that the hyper-invasive trophoblast gave rise to these pathologies, it is believed that abnormal trophoblast behaviour is more a consequence of these placental tumours rather than the cause (Bischof and Irminger-Finger, 2005).

Determining the genes that are associated with each of these placental pathologies may be useful in developing targeted therapies or better yet, ways to prevent abnormal placentation. Previous work has identified numerous candidate genes that are linked to placental dysfunction. However, more recent work suggests that examining the chemical modifications to the genome - the field of epigenetics - may be key to unravelling the molecular mechanisms that regulate the development of the placenta, its function and dysfunction.

**1.5 Epigenetics**

The term *epigenetics* is defined as *outside conventional genetics* (Jaenisch and Bird, 2003). Epigenetics refers to chemical modifications to the DNA sequence that influence gene expression but do not change the nucleotide sequence itself. The two main epigenetic mechanisms are histone modifications and DNA methylation. The epigenetic markings on DNA and core histones differentiate between active and inactive chromatin, thus affecting which genes are available for transcription (Li, 2002; Bird and Macleod, 2004). The epigenetic complement to the genome is referred to by some as the ‘epigenome’ (Callinan and Feinberg, 2006).

Although the term ‘epigenetics’ was coined by Conrad Waddington back in 1942, the ‘epigenetic era’ did not begin until 2000 (Junien, 2006). Waddington originally defined epigenetics as ‘the developmental processes by which genotype gives rise to phenotype’ (Waddington, 1942). At the time, Waddington associated the term with epigenesis - when an
undifferentiated zygote gives rise to the developmental programming of an organism - in an attempt to explain the developmental origins of gene expression and signalling pathways (Richards, 2006). Waddington’s definition also included the concept that epigenetic mechanisms are reset at one point in the life cycle. The term has always been used in a developmental context but its application has changed over time.

Currently, the term epigenetics is still used to explain such ‘reprogrammable’ mechanisms, but it is also used to explain changes in gene expression that are not linked to developmental programming. Many areas of epigenetics are focused on the role of the environment in establishing epigenetic marks, their stability and heritability over time, and their implications for genome instability. Although there has been an abundance of interest to investigate the causes of this non-Mendelian form of inheritance, there is still a paucity of information regarding the influence of the environment on epigenetic mechanisms (Youngson and Whitelaw, 2008).

1.5.1 Histone Modifications

The epigenetic modifications to histone proteins are complex and include processes of acetylation, methylation, phosphorylation and ubiquitination (Wade et al., 1997; Jenuwein, 2001). These modifications regulate the conformation of the histone-packaged chromatin, which in turn regulates gene expression. In particular, a large amount of work has been done on lysine acetylation and methylation (Bernstein et al., 2007). Lysine acetylation strongly correlates with chromatin accessibility, thus increasing the transcriptional activity of the genome (Perry and Chalkley, 1982; Lee et al., 1993). The effects of lysine methylation depend on which lysine residue is modified, and can lead to either chromatin activation or repression (Shi et al., 2004). Histone modifications and DNA methylation are co-dependent processes that work together to regulate gene expression.

1.5.2 DNA methylation

DNA methylation involves the covalent addition of a methyl group to the fifth carbon of a cytosine residue. The cytosine residues that undergo methylation are normally located next to a guanine, such that most DNA methylation occurs at CpG dinucleotides. The “p” simply represents the phosphodiester bond between the cytosine and guanine residues.
1.5.2.1 CpG Dinucleotides

In vertebrates, most of the CpG dinucleotides in the genome are methylated, while in plants, non-CpG methylation has shown to have an established functional role (Bird, 2002; Chan et al., 2005). It has been shown that methylated cytosine residues serve as a guide for the establishment and maintenance of other epigenetic marks (Jaenisch and Bird, 2003; Richards, 2006). For example, methylated cytosines can direct post-translational modifications of histone proteins, thereby altering chromatin structure and potentially genome function (Jaenisch and Bird, 2003; Wang et al., 2004). Methylated cytosines, therefore, alter the access to the primary nucleotide sequence, thereby influencing activity at the DNA, nucleosomal and chromosomal level, and ultimately affecting phenotype.

1.5.2.2 CpG Islands

Interestingly, the methylated cytosine in the CpG dinucleotides can undergo spontaneous deamination to yield thymine (Bird, 1980). Over the course of evolution, this has reduced the number of CpG dinucleotides represented in the mammalian genome (Lander et al., 2001). However, regions containing unmethylated cytosines are protected from this evolutionary degradation. These CpG-rich regions are called CpG islands and are often located in the 5’ promoter region of genes. In normal cells, the CpG sites within CpG islands are usually unmethylated, and the downstream gene is available for transcription. However, when promoter CpG islands are methylated, the downstream gene is silenced. Thus the methylation status of promoter CpG islands can be a useful surrogate marker for gene expression (Song et al., 2005), taking into account that the absence of methylation does not always indicate gene expression, rather the potential for gene expression. In addition to promoter regions, CpG islands are also found within highly repetitive regions of DNA, such as transposable elements, where they are frequently methylated to prevent genome dysfunction (Weber et al., 2005; Eckhardt et al., 2006).

1.5.2.3 The Mechanism of DNA Methylation

Epigenetic modification can be seen as a ‘gene switch’ that regulates gene expression. Mechanistically, it is thought that cytosine methylation interferes with transcription factor binding. This interference occurs once the methyl-CpG binding domain protein is recruited to the methylated cytosine, thus blocking the transcription factor, either directly or through subsequent changes in chromatin structure, and preventing gene expression (Nan et al., 1998; Richards, 2006). Further, it is thought that the concentration of CpG dinucleotides within a
promoter region might also affect the likelihood of gene repression by methylation. Weber et al. generated an epigenomic map of DNA methylation, RNA polymerase II occupancy and histone tail modifications for 16,000 gene promoters, and found that the CpG content of a gene’s promoter sequence was in inversely related to its methylation level and silencing potential (Weber et al., 2007).

DNA methylation is established by a group of DNA methyltransferase (DNMT) enzymes. These enzymes obtain a methyl group from an S-adenosyl-L-methionine molecule (the methyl donor) and subsequently catalyse the transfer of the methyl group onto a cytosine residue in the DNA (Razin and Cedar, 1977). DNMT1 is a maintenance enzyme that restores the methylation patterns in dividing cells after DNA replication (Figure 1.5) (Bestor et al., 1988). The two other enzymes, DNMT3a and DNMT3b, are required during embryonic development to initiate de novo methylation; a process during which new methylation patterns are established on unmethylated DNA to set methylation imprints or methylate repeat regions of DNA (Figure 1.5) (Okano et al., 1999; Li, 2002). Another methyltransferase family member, DNMT3L, is a regulatory factor that does not catalyse the addition of a methyl group but assists in the establishment of de novo methylation (Bourc'his et al., 2001; Kareta et al., 2006).

Although the exact mechanisms involved in establishing DNA methylation patterns are unknown, the important role of DNMTs in re-setting and maintaining these patterns is becoming increasingly clear. For example, ICF syndrome (immunodeficiency, centromeric region instability and facial abnormalities) is a disorder in which mutated DNMT3b does not transfer methyl groups to DNA (Hansen et al., 1999). In patients with ICF syndrome, the inactivation of DNMT3b results in hypomethylation of the satellite 2 repeat; a pericentromeric repeat region on chromosomes one, nine and sixteen that is highly methylated in normal cells (Hansen et al., 1999; Hassan et al., 2001). A DNMT3b knockout study detected hypomethylation at other minor satellite centromeric repeat regions in the mouse, confirming that DNMT3b plays an important role in normal development and disease (Okano et al., 1999). It has also been found that environmental factors, such as diet or aging, can influence the activity of DNMT3a and DNMT3b, which can permanently change a gene’s transcriptional status, resulting in genome instability or disease (Figure 1.5) (Waterland and Michels, 2007).
1.6 DNA Methylation: Role in Early Human Embryonic Development

Regulation of gene expression during human embryonic development is a highly specialized process. Most of the knowledge regarding the epigenetic events that control gene expression during these early stages of growth is based on studies in mice.

1.6.1 Epigenetic Reprogramming

During early embryonic development, there are two important epigenetic events during which genome-wide methylation patterns are reprogrammed: the reprogramming of primordial germ cells and the reprogramming of the embryo (Reik et al., 2001). Both events are required to re-set the epigenetic marks in the genome to prevent the transmission of aberrant epigenetic information to subsequent generations.
1.6.1.1 Reprogramming in the Primordial Germ Cells

Early in development, the highly methylated genomes of the primordial germ cells – the cells that will give rise to a sperm or an oocyte – undergo a wave of genome-wide demethylation (Bestor, 2000; Reik et al., 2001; Dean et al., 2003). One consequence of this demethylation event is to re-set the methylation imprints in the maternal and paternal gametes. By the time the primordial germ cells reach the gonads, all of the methylation patterns from imprinted gene regions have been erased (Reik et al., 2001). The methylation marks are then re-established in the sperm and oocyte genomes, yet the timing of these methylation events differs between the gametes. The re-methylation of imprinted genes in the sperm genome occurs soon after the initial demethylation event and is completed shortly after birth (Li et al., 2004). However, re-methylation of imprinted genes in the oocyte genome is a post-natal event, beginning in the maturation stage of an oocyte and finishing shortly before ovulation (Eichenlaub-Ritter et al., 2007; Bourc’his and Proudhon, 2008). Thus, the methylation patterns for imprinted genes are established at different developmental time points during gametogenesis (Lucifero et al., 2004; Song et al., 2009).

1.6.1.2 Reprogramming in the Embryo

Once fertilization occurs, both paternal and maternal genomes are reprogrammed to remove the methylation marks from the gametes. The first event is the active demethylation of the paternal genome (within the male pronucleus), an event that occurs only 6-8 hours after fertilisation and is complete before the first cell division (Mayer et al., 2000; Reik et al., 2001). The maternal genome is demethylated more slowly, between the zygote and blastocyst stage, by passive demethylation (Reik et al., 2001). The imprinted genes, however, are protected from these waves of genome-wide demethylation. Once implantation occurs, de novo methylation patterns are established to non-imprinted genes in the epiblast (which will form the embryo, amnion and umbilical cord) but re-methylation occurs to a lesser degree in the non-imprinted genes in the trophectoderm and primitive endoderm (which will form the chorion and placenta) (Dean et al., 2003). The differential re-establishment of these epigenetic modifications has been shown to play a central role in directing the early lineage commitment of cells (Santos et al., 2002; Reik et al., 2003).

1.6.2 Genomic Imprinting

The paternal and maternal gametes are programmed very differently and, because of the presence of imprinted genes, both are required for normal embryonic development. Imprinted
genes contain specific epigenetic marks that lead to expression of only one parent’s allele of the gene. This is known as ‘parent-of-origin’ specific gene expression, since imprinted genes express either the paternal or maternal allele.

1.6.2.1 The Conflict Hypothesis

Genomic imprinting is thought to be the solution to what is known as the genetic ‘conflict hypothesis’, where the mother wants to ration her genetic resources for subsequent offspring yet the father, in competition with other fathers, wants to maximise the survival of his offspring. This hypothesis is based on work showing that maternally expressed imprinted genes restrict fetal growth and paternally expressed imprinted genes enhance fetal growth (Moore and Haig, 1991). The first two imprinted genes that were discovered in mice display this reciprocal functioning of the paternal and maternal genomes: the paternally expressed \textit{Igf2} gene is a growth promoter, whereas the maternally expressed \textit{Igf2r} gene is a negative regulator of \textit{Igf2} (Haig and Graham, 1991).

Studies in mice have demonstrated the importance of imprinting by showing that embryos containing either two paternal or two maternal genomes do not survive to term (Barton \textit{et al.}, 1984; McGrath and Solter, 1984; Devriendt, 2005). Examination of these embryos showed the androgenotes (containing two paternal genomes) to have poorly formed embryonic tissues and well-developed extra-embryonic components, while the gynogenotes (containing two maternal genomes) displayed well-developed embryonic tissues and poorly formed extra-embryonic components (McGrath and Solter, 1984). The occurrence of the complete hydatidiform mole - a placental pathology that results when a fetus is comprised of two paternal genomes and no maternal genome - is also a reflection of abnormal genomic imprinting (Devriendt, 2005). These studies demonstrate that although both parental genomes are required for embryogenesis, diploidy alone is not sufficient for normal development. Imprinting has therefore evolved to promote the combined genetic interests of the mother and father during reproduction.

1.6.2.2 Establishing Imprinting Marks

There are approximately 40 genes in the human genome that display gene expression from a single parental allele (Morison \textit{et al.}, 2005). Although imprinted genes represent a very small portion of the genome, normal embryonic development is hinged upon maintaining these parent-of-origin specific patterns of DNA methylation. The DNA methyltransferases discussed previously (DNMT1, DNMT3a, DNMT3b and DNMTL) play a key role in
retaining these patterns. A study by Okano and colleagues demonstrated the importance of DNMTs by inactivating Dnmt3a and Dnmt3b in mouse embryonic stem cells. Their results showed extensive genome-wide demethylation, including specific regions of two well-studied imprinted genes, Xist and Igf2 (Okano et al., 1999). These findings, along with other DNA methyltransferase knockout experiments, have demonstrated the importance of maintaining DNA methylation imprints in early embryonic development.

1.6.3 X Chromosome Inactivation

One of the roles of DNA methylation in females is to inactivate one X chromosome during the early stages of embryonic development. Since males (XY) inherit one X chromosome and females (XX) inherit two, one of the X alleles must be silenced in females to control the levels of gene expression coming from the X chromosome. X inactivation is thus referred to as a form of gene dosage compensation in females. To achieve this process, one X chromosome is transcriptionally silenced through a series of epigenetic events, including DNA methylation and histone modifications, which results in a silenced heterochromatic state (Li, 2002; Chow and Heard, 2009). The epigenetic marks that induce X inactivation are established during the preimplantation stages of female embryonic development (van den Berg et al., 2009).

The process of X inactivation involves the accumulation of a large RNA molecule encoded by the gene, XIST, on the X chromosome that is to be inactivated (Brown et al., 1992). After one of the X chromosomes is targeted by XIST, it becomes heavily methylated to prevent gene expression on the inactive X allele. The active X chromosome remains relatively unmethylated, except that it acquires specific methylation modifications to prevent expression of the XIST gene (Payer and Lee, 2008). The active X chromosome will express the TSIX gene, a negative regulator of XIST, in order to maintain XIST silencing and preserve its active state (Lee et al., 1999). Although some genes on the inactive X chromosome are not inactivated or are variably inactivated (Carrel and Willard, 2005), the event of X chromosome inactivation creates a large difference in gene-associated DNA methylation between the sexes.

1.6.4 Epigenetic Silencing of Retrotransposons

The majority of the CpG sites in the human genome that are hypermethylated are located within highly repetitive sequences of DNA (Rodriguez et al., 2008). Some of these repetitive sequences, termed retrotransposons, have the ability to self-replicate and insert themselves at other locations within the genome. Since retrotransposons are known for this disruptive
“copy and paste” behaviour, most of their sequences are silenced by DNA methylation as a way of suppressing their activity (Slotkin and Martienssen, 2007). Studies have shown that after fertilisation and during the epigenetic reprogramming stage of the embryo, imprinted genes and retrotransposons are mostly resistant to the wave of genome-wide demethylation that occurs during this time in development (Reik et al., 2001). DNA methylation is believed to be important to silence the potentially disruptive activity of retrotransposons and maintain genome stability in the developing embryo (Bestor and Tycko, 1996; Walsh et al., 1998; Bestor, 2000).

1.7 Epigenetic Modifications: Response to the Environment

1.7.1 Pre-natal Epigenetic Modifications

The occurrence of epigenetic modifications during fetal development may have relevance to the fetal origins of human disease. Barker and colleagues were first to discover that adult disease might be programmed in utero when they found a significant correlation between low birth weight and risk of developing cardiovascular disease during adult life (Barker et al., 1989). More recent studies have provided evidence to support this notion of “fetal programming”. It is thought that a developing fetus is susceptible to changes in the maternal environment, thus the fetus develops mechanisms to cope with these changes. The mechanisms are either adaptive or disruptive depending on how to promote survival in the altered maternal environment (Padmanabhan, 2007). The exact mechanisms by which fetal programming occurs is not yet known; however, fetal exposure to different hormonal and metabolic conditions in utero can alter cell and organ differentiation, thus permanently affecting the function of these organs later in adult life (Gluckman and Hanson, 2004; Fernandez-Twinn and Ozanne, 2006).

Just as adverse maternal factors can have lasting effects on fetal development, fetal programming also includes the concept of developmental plasticity. This notion explains a cell’s ability to change its behaviour in response to environmental cues (Feinberg, 2007). Developmental plasticity suggests that there is a critical time period during fetal development where humans are able to adapt to their environment. Jansson and Powell propose that a fetus can adapt to a nutrient-poor intrauterine environment in order to promote survival, but if its post-natal environment is nutrient rich, the programmed metabolism to a nutrient-poor environment may lead to adult disease (Jansson and Powell, 2007).
Some studies suggest that developmental plasticity involves epigenetic modifications in utero. These modifications contribute to “adaptive” phenotypes that are established to meet the demands of the post-natal environment (Jaenisch and Bird, 2003; Godfrey et al., 2007). Those phenotypes that match the demands of the post-natal environment might enhance survival, while those phenotypes that do not match the demands of adult life might increase risk of disease. For example, the dietary factors of a pregnant mouse affect the epigenetic programming of her unborn pups. It has been shown that a diet rich in methyl donors, such as folate, methionine, choline and vitamin B₁₂, can affect methylation levels at specific alleles and permanently change gene expression (Wolff et al., 1998; Waterland and Jirtle, 2003; Dolinoy et al., 2006). Epigenetic processes such as DNA methylation are important for the developmental programming of an embryo given that their disruption has been associated with post-natal abnormalities, including imprinting disorders (Jiang et al., 2004; Godfrey et al., 2007).

1.7.2 Post-natal Epigenetic Modifications

In addition to pre-natal stimuli, post-natal factors can also affect epigenetic processes and lead to an increased risk of adult disease. Environmental stimuli, such as diet, smoking and pollution, are thought to effect epigenetic modifications and permanently change gene expression and thus susceptibility to disease (Jacob et al., 1998; Pulling et al., 2004; Bollati et al., 2007; Waterland and Michels, 2007).

Many studies are now trying to investigate other links between epigenetic mechanisms and increased risk for disease. In fact, this concept has become so prevalent that a new field of study has been created, called epigenetic epidemiology. Defined by Waterland and Michels, epigenetic epidemiology is the study of the associations between epigenetic variation and risk of disease (Waterland and Michels, 2007). It is known that epigenetic mechanisms are responsible for two things: gene expression and the potential to alter gene expression. Moreover, once gene expression is epigenetically altered, these changes may be passed onto subsequent generations. Although it cannot be completely separated from genetic epidemiology, the field of epigenetic epidemiology can be used to investigate the epigenetic mechanisms of transgenerational inheritance that influence disease risk (Waterland and Michels, 2007).
Due to its central role in pregnancy, the placenta reflects both normal and pathological processes that occur in both mother and fetus. Investigating the epigenetic mechanisms involved in placental development and function may be significant in linking fetal and maternal health with pre- and postnatal conditions, since many pregnancy-related pathologies stem from complications with the fetal-maternal interaction.

1.8 Epigenetics in the Placenta

The cell-types of the placenta are established and maintained according to the interpretation of epigenetic information. Specifically, epigenetic modifications have been shown to regulate the differentiation of the trophoblast into its respective sub-types, and to regulate the behaviour of these sub-types as they establish the foundation for what will ultimately become the placenta. A large number of epigenetically regulated genes have been identified that are associated with placental development. The epigenetic profiles of these genes, along with other genome-wide methylation patterns, all contribute to the important epigenetic landscape that is required for normal placental development and function (Hemberger, 2007).

1.8.1 Imprinted Genes in the Placenta

The placenta has been labelled as one of the most important sites of imprinted-gene action (Constancia et al., 2004; Coan et al., 2005; Ferguson-Smith et al., 2006). Almost every known imprinted gene is expressed within the placenta, and most are thought to be involved in regulating cell differentiation, transporter activity, placental development and fetal growth (Ferguson-Smith and Surani, 2001; Reik and Walter, 2001; Reik et al., 2003; Ferguson-Smith et al., 2006). As the primary source of resource exchange during embryonic development, the placenta is seen as the critical organ where the interaction to satisfy the opposing demands of the maternal and paternal genomes takes place (Coan et al., 2005). The information published from studies in mice is very useful in understanding the conservation of imprinted gene expression and function in human placental development.

In the placenta, the paternally expressed genes involved in enhancing placental/fetal growth are exposed to the maternally expressed genes that involved in restraining placental/fetal growth. For example, the paternally expressed gene, Igf2, was shown to directly contribute to the placenta’s transfer of nutrients to the fetus. One study deleted the Igf2 transcript in Igf2-expressing trophoblast cells and observed a reduction in placental weight followed by fetal growth restriction (Constancia et al., 2002). Other paternally expressed genes, Peg1, Peg3
and Peg10, are thought to have growth-enhancing effect on the placenta (Lefebvre et al., 1998; Hiby et al., 2001; Ono et al., 2003). One the other hand, the maternally expressed gene, Tsc2, may act as a tumour suppressor to prevent excessive placental growth (Paulsen et al., 2000). Moreover, the maternally expressed gene, Mash2, codes for a transcription factor that is essential for trophoblast differentiation into its progenitor cells, but its expression declines as gestation progresses to either maintain cell proliferation or restrict cell differentiation (Guillemot et al., 1994; Tanaka et al., 1997; Hemberger and Cross, 2001).

There have been approximately 13 chromosomal regions identified in the mouse which, when parental origin imprinting is disturbed, lead to major phenotypic abnormalities in both fetus and placenta. This, taken with the identification of numerous imprinted genes associated with placental development, highlights the importance of imprinted genes in the growth and function of the placenta (Coan et al., 2005; Hemberger, 2007).

1.8.2 Non-Imprinted Genes in the Placenta

In addition to imprinted genes, studies have shown that the epigenetic regulation of non-imprinted genes is also essential for normal placental development and function. In a study by Serman et al., a reduction in global DNA methylation after treatment with the demethylating agent, 5-azacytidine, resulted in significantly lower levels of glycoprotein gene expression, placental weight and trophoblast proliferation in rats (Serman et al., 2007). The treatment completely disrupted placental structure and markedly reduced the function of proliferating trophoblast cells. Assuming that the effects of 5-azacytidine were specifically mediated by demethylation, this study demonstrates the importance of epigenetic regulation during placental development. Other work has shown that modifying the DNA methylation and histone acetylation levels in human placental explants leads to altered expression of genes involved in the synthesis of prostaglandin, a key molecule required for the maintenance of pregnancy (Mitchell, 2006). There are many other epigenetically regulated genes that contribute to placental development - too many to discuss in the context of this thesis - yet these genes are well documented in a review by Hemberger (2007).

1.8.3 The Hypomethylated State of the Placenta

The placenta is considered to be a globally hypomethylated tissue. It has been shown that the placenta contains approximately 20% fewer methylated cytosines than most other tissues (Ehrlich et al., 1982; Gama-Sosa et al., 1983; Fuke et al., 2004). Interestingly, the reduction in DNA methylation in the placenta is not evenly distributed throughout the genome. It is
mostly a result of the hypomethylation of promoter regions across the X chromosome and the hypomethylation of repetitive elements (Cotton et al., 2009). Thus the expression of some of the normally silenced X-linked genes and repetitive elements in the placenta has been associated with the placenta’s unique hypomethylated state (Migeon et al., 1985; Yoder et al., 1997; Seifarth et al., 2005; Kudaka et al., 2008).

### 1.8.3.1 X Chromosome Hypomethylation

In the placenta, the levels of inactive X chromosome methylation are lower than that of the active X chromosome; an epigenetic phenomenon that is opposite to what occurs in somatic tissues (Cotton et al., 2009). In the placenta, the reactivation of the silenced X chromosome (by demethylation) has been associated with trophoblast behaviour; specifically the cessation of trophoblast proliferation and subsequent trophoblast apoptosis that occurs during the remodelling of maternal spiral arteries (Migeon et al., 2005). The expression of a small number of X-linked genes has demonstrated X chromosome reactivation in the human placenta, such as the bi-allelic expression of G6PD, an X-linked gene that is normally expressed from only the active X allele in somatic tissues (Migeon et al., 1985). Nevertheless, the hypomethylated state of the inactive X chromosome in the placenta is not associated with a large increase in the expression of X-linked genes, so it is hypothesised that the expression of the non-coding XIST RNA and heterochromatin conformation is sufficient for the inactive X chromosome in the placenta to maintain its silenced state (Nguyen and Disteche, 2006).

### 1.8.3.2 Hypomethylation of Repetitive Elements

The placenta is becoming increasingly known for harbouring unmethylated repetitive sequences (Reiss et al., 2007). Of the repeat elements, the Alu elements, long interspersed nuclear elements (LINEs) and satellite regions display a reduced level of methylation in the placenta compared to their hypermethylated status in the majority of somatic tissues (Gama-Sosa et al., 1983; Hellmann-Blumberg et al., 1993; Shen et al., 2006). The epigenetic reactivation of repeat sequences in the placenta is demonstrated by the finding that many repetitive elements are transcribed in the placenta that are normally silenced by methylation in somatic tissues (Kato et al., 1987; Wilkinson et al., 1993; Okahara et al., 2004; Kim et al., 2006; Reiss et al., 2007). Furthermore, the expression of these normally deleterious repeat elements suggests a functional role for these sequences in the placenta - a role that currently remains unknown yet will be discussed in Chapters Eight and Nine.
Hypomethylation of repetitive elements is usually linked with genome instability, a trait that is commonly associated with cancer. In fact, the loss of DNA methylation at certain repetitive sequences has been shown to be a hallmark in some cancers (Ehrlich, 2002; Szpakowski et al., 2009). Interestingly, the hypomethylation of repetitive elements in the placenta and in cancer is just one of the many well-documented similarities between placental trophoblasts and tumour cells in cancer.

1.9 Behavioural Similarities of Trophoblasts and Tumours

The behaviour of the trophoblast during placental development has often been compared to the process of tumorigenesis in cancer. Fisher and colleagues were amongst the first to notice the similarity between trophoblast and tumour invasion. Their experiments displayed how the interactions amongst trophoblast cells led to the co-existence of genetically dissimilar fetal and maternal cells; a phenomenon that was very similar to tumour invasion in cancer (Damsky et al., 1992; Fisher and Damsky, 1993). However, they also observed striking differences, noting that trophoblast invasion was tightly regulated and confined to the first of three layers of the endometrium, while tumour invasion was not under spatial or temporal control. In addition, they noted that trophoblast invasion did not disturb the maternal immune system, whereas invading tumour cells did in most cancers (Fisher and Damsky, 1993).

A comprehensive review by Ferretti et al. revealed many striking similarities between trophoblast and cancer cells, comparing the molecular circuits shared between the two (Ferretti et al., 2007). The proliferative, migratory and invasive phenotypes shared by these cells prompted the compilation of various findings from trophoblast and cancer studies in attempt to identify the genes responsible for their similar behaviours. The review depicts various enzymes, hormones, proto-oncogenes, growth factors and many other key molecules expressed by both trophoblast and cancer cells (Ferretti et al., 2007). However, there are the key differences between trophoblasts and tumours, such as the ability of the trophoblast to control the extent of its growth and invasion into maternal tissue (Soundararajan and Rao, 2004). Further investigation into the pathways that regulate the proliferative, migratory and invasive phenotypes of trophoblasts and tumour cells may provide new targets for the treatment of pregnancy-related diseases and potentially even cancer.
1.10 DNA Methylation: Role in Tumorigenesis

The discovery that tumour cells undergo loss of DNA methylation suggested that epigenetic mechanisms might be involved in cancer (Feinberg and Vogelstein, 1983; Ehrlich et al., 1985). Aberrant DNA methylation in cancer was then localised to gene promoter regions when studies found that CpG-rich promoter regions were especially prone to hypermethylation (Baylin et al., 1986; Esteller et al., 2001; Herman and Baylin, 2003; Gebhard et al., 2006). This launched multiple investigations that aimed to find functional gene targets that became aberrantly methylated in cancer.

In 1989, it was proposed that DNA methylation might induce loss of function of tumour suppressor genes in cancer (Greger et al., 1989). Since then, many cancer studies have focused on DNA methylation as the silencing mechanism for tumour suppressor genes. Inactivity of tumour suppressor genes contributes to uncontrolled cell proliferation, resulting in tumour growth. For example, the well-studied tumour suppressor gene in breast cancer, BRCA1, is often inactivated by DNA methylation (Dobrovic and Simpfendorfer, 1997). It has also been suggested that the epigenetic changes that inactivate tumour suppressor genes might even occur at the stem-cell level (Ohm et al., 2007). At present, alterations in DNA methylation, such as promoter hypermethylation, may be as common as mutation in terms of a mechanism of gene silencing in cancer (Weber et al., 2005; Ohm et al., 2007).

1.11 The Epigenetic Regulation of Trophoblasts: Similarities to Cancer

Given that placental trophoblast cells are phenotypically similar to cancer cells, the observation that epigenetic alterations were common in human tumours prompted investigation into similar epigenetic events in the placenta. The discovery of epigenetically regulated tumour suppressor genes and cell adhesion molecules during placental development has taken the behavioural analogy between trophoblasts and tumour cells to the epigenetic level.

1.11.1 Tumour Suppressor Genes

The epigenetic silencing of tumour suppressor genes during the periods of trophoblast invasion supports the concept that the epigenetic regulation of trophoblasts and tumour cells may be related. The tumour-like behaviour of the trophoblast led to the discovery of a hypermethylated tumour suppressor gene, RASSF1A, in the placenta of humans and rhesus
monkeys (Chiu et al., 2007). RASSF1A happens to be the most frequently methylated tumour suppressor gene in all cancers (Pfeifer and Dammann, 2005). Interestingly, RASSF1A was found to be hypermethylated in the placenta throughout gestation, whereas this gene is largely unmethylated other fetal tissues (Chiu et al., 2007). A second tumour suppressor gene, SERPINB5 (also known as Maspin), is differentially expressed in the human placenta as a result of the methylation and acetylation of histone tails (Dokras et al., 2002; Dokras et al., 2006). SERPINB5 is involved in cell invasion, apoptosis and angiogenesis in cancer cells (Zou et al., 1994; Sheng et al., 1996; Zhang et al., 2005). The epigenetically regulated expression of SERPINB5 was inversely correlated with gestation; low expression levels were detected during the first trimester of pregnancy – the period of maximum trophoblast invasion – and expression levels increased as gestation progressed. Another tumour suppressor gene, APC, is highly methylated in the human placenta (Wong et al., 2008). Hypermethylation of APC is a feature of several pre-cancerous and cancerous tissues (Esteller et al., 2000; Deng et al., 2004). Thus the hypermethylation of these genes in the human placenta demonstrates the importance of tumour-suppressor gene silencing for normal placental development.

1.11.2 Cell Adhesion Molecules

The epigenetic regulation of cell adhesion molecules has also strengthened the trophoblast-tumour analogy. The adherens junction molecules, E-cadherin and plakoglobin, play a critical role in placental development as they regulate the differentiation, migration and invasion of the trophoblasts (Coutifaris et al., 1991; Zhou et al., 1997; Aplin et al., 1999). Alterations in the expression of these molecules have been associated with many human cancers, including an invasive cancer of the placenta, the choriocarcinoma (Berx et al., 1998; Nakamura et al., 2005). A study by Rahnama et al. treated a choriocarcinoma cell line with the DNA methyltransferase inhibitor, 5-aza-2’-deoxycytidine, and observed a non-migratory and non-invasive trophoblast phenotype (Rahnama et al., 2006). They also detected a large increase in the gene promoter activity, mRNA levels, and protein levels of both E-cadherin and plakoglobin (Rahnama et al., 2006). To test further, the DNA methyltransferases, DNMT3a and DNMT3b, were knocked out using small interfering RNAs. The combined depletion of these DNMTs resulted in increased promoter activity of E-cadherin and plakoglobin (Rahnama et al., 2006). The results of this study suggested that E-cadherin and plakoglobin activity is dependent on methylation, although promoter methylation was not directly assessed. Importantly, this work demonstrated that normal trophoblast migration and invasion during placental development is regulated by DNA methylation.
1.12 Techniques to Study Genome-wide Methylation Patterns

DNA methylation patterns were originally studied on a gene-by-gene basis. Although the candidate-gene approach greatly contributes to investigating the role of DNA methylation at the single-gene level, it does not provide insight into the genome-wide distribution of methylation modifications. For this, technological advancements have been made to facilitate the study of epigenetics on a global scale (Esteller, 2007). The major ‘upgrade’ to epigenetic research occurred when techniques were developed to profile the methylation modifications in the entire genome, referred to as the ‘methylome’ (Feinberg, 2001). Analysing genome-wide methylation patterns has led to a better understanding of the epigenetic processes involved in both normal and disease states. At the time of this study, the methods that were available to analyse the methylome included those using methylation-sensitive restriction enzymes and microarray-based methods.

1.12.1 Genome-wide Methylation Analysis by Methylation-Sensitive Restriction Enzymes

One of the first widely used tools for global methylation analysis was by restriction landmark genomic scanning (RLGS). This method involves successive digestions of genomic DNA by methylation-sensitive restriction enzymes combined with two-dimensional gel electrophoresis (Hatada et al., 1991; Plass et al., 1996; Costello et al., 2009). The terms “restriction landmark” refer to the use of restriction enzymes that cut frequently within CpG islands; locations in the genome that are highly susceptible to differential methylation thus “landmarks” for epigenetic analysis (Lindsay and Bird, 1987; Dai et al., 2002). In this method, genomic DNA is digested and the ends of the cleaved (unmethylated) fragments are labelled with radioactivity. Fragments are cut by a second restriction enzyme, separated by gel electrophoresis and digested by a third restriction enzyme. Fragments are then separated by electrophoresis in a perpendicular direction. Gel results display the copy number of unmethylated CpG sites, and global methylation can be assessed by comparing the presence of absence of a spot between samples, or by observing changes in spot intensity (Costello et al., 2009). The disadvantages of RLGS are that not all CpG islands contain the restriction sites and it is difficult to map specific fragments to a genomic location. This original tool for global methylation analysis has since been replaced by more sophisticated technologies.

Another method that uses methylation-sensitive restriction enzymes for genome-wide methylation analysis is by amplification of intermethylated sites (AIMS). This is a system of DNA fingerprinting to identify the methylation profile of a cell (Frigola et al., 2002; Jorda et
Briefly, genomic DNA is initially digested by a methylation-sensitive restriction enzyme, which cuts only at unmethylated CpG-sites. The fragments are then cleaved with a methylation-independent isoschizomer of the first enzyme, which cuts the remaining CpG sites in the fragments, all of which are methylated. Adaptors are then ligated onto fragments and sequences are amplified by PCR. Since PCR products contain a methylated CpG-site at either end (due to the differential restriction digest), the banding patterns observed after electrophoresis of products represent the methylated fragments in that sample. The main limitation to the AIMS method is determining the genomic location of a specific fragment - the band must be cut out of the gel and sequenced.

### 1.12.2 Analysing Methylation using Microarrays

More time and cost-effective methods for global methylation analysis became available when CpG-island and promoter microarrays were introduced. In fact, the combination of using methylation-sensitive restriction enzymes with microarray technology changed the field of epigenetics. A widely used method for high-throughput DNA methylation analysis using such system is known as differential methylation hybridisation (DMH) (Huang et al., 1997; Yan et al., 2009). This technique generates a DNA methylation profile by hybridisation fragments from a methylation-sensitive restriction digest onto CpG island microarrays. The advantage of using two methylation-sensitive restriction enzymes is that it reduces the chance of amplifying fragments that result from incomplete digestion, which increases the chance of enriching for fragments that contain purely methylated CpG sites (Yan et al., 2009). However, although DMH and other microarray-based methods have greatly improved the mapping of methylation marks to the genome, more simple and selective methods for enriching for methylated sequences have been developed.

A potential major advance in global methylation analysis occurred when the chromatin immunoprecipitation (ChIP) technique was modified for methylation analysis and combined with microarray technology. An antibody against 5-methylcytosine was developed to immunoprecipitate and thus enrich for methylated DNA sequences (MeDIP) (Reynaud et al., 1992). Once immunoprecipitated, the affinity-purified methylated DNA and non-immunoprecipitated DNA are dye-labelled and co-hybridised to CpG-island or promoter microarrays to identify sequences with high, medium or low levels of methylation (Weber et al., 2005; Keshet et al., 2006; Zhang et al., 2006; Weber et al., 2007). There are various microarray platforms available including those created by NimbleGen, Agilent, Affymetrix, Illumina and Aviva Systems Biology. At the time this technique was introduced, the use of
MeDIP combined with microarray technology showed promise to revolutionise epigenetic research by providing a rapid and effective tool for analysing thousands of methylated CpG sites throughout the genome simultaneously. The MeDIP-microarray technique will be discussed in more detail in Chapter Four.

1.13 Aims of this Study

Although the placenta is a normal tissue, its constituent trophoblast cells may share similar molecular processes with malignant cells. Epigenetic studies of the placenta and cancer have reached a point of convergence where it seemed appropriate to work towards finding a common epigenetic link between the two. Thus the first hypothesis for this study was that the methylation patterns associated with the behaviours of trophoblast and tumours cells may be related. Second, the role of placental-specific epigenetic modifications, such as global DNA hypomethylation, in facilitating placental function is poorly understood. Since epigenetic modifications play a central role in directing the early lineage commitment of cells (Santos et al., 2002; Reik et al., 2003), it was hypothesised that placental-specific DNA methylation differences might reflect the specialised functions of the placenta compared to somatic tissues. Thus it was of interest to document placental-specific methylation patterns in the placenta, taking into account that the origin of the placenta is determined during the earliest stages of embryonic development, when the inner-cell mass is first distinguished from the trophectoderm, and when the inner-cell mass further differentiates into the primitive endoderm or the epiblast.

At the time of this study, the advancements in the field of epigenetics made it possible to explore these hypotheses on the genome-wide scale. It therefore seemed appropriate to investigate the mechanisms involved in regulating placental development by using a large-scale method of epigenetic analysis. Furthermore, the availability of first-trimester placental tissue from women undergoing medical terminations of pregnancy at Dunedin Hospital provided a unique opportunity to study the human placenta at such an early stage in gestation; the stage at which trophoblast invasion is at its peak. Term placenta was also collected in order to compare the epigenetic profiles of early and late stage placental tissue in the hope of identifying any placental-specific epigenetic modifications. In addition to placental tissue, a wide range of normal fetal and adult tissues, and even some cancerous tissues, were available for use in this project.
In this study, a genome-wide assessment of DNA methylation in first-trimester placenta, followed by subsequent confirmatory experiments, was performed to identify methylation changes that may occur in the placenta during different stages of gestation. This epigenetic analysis also sought to identify placental-specific methylation patterns that may be biologically relevant to the development or function of this relatively unexplored yet fascinating transient human tissue.

The main goals of this study are to:

- Examine first-trimester placental tissue for differentially methylated genes that might be involved in regulating the similar behaviours of trophoblasts and tumour cells.
- Determine placental-specific patterns of gene methylation that may be unique to placental development and function.
- Explore the functional significance of candidate differentially methylated genes in the context of placental development, function and dysfunction.

The results from this study may provide a model system for the epigenetic modifications that occur in the placenta. Any discoveries may provide a clue to the nature of the epigenetic modifications that are associated with normal placental function, placental-related pathologies and possibly the origin of adult disease.
Chapter Two

Materials and Methods

2.1 Collection of Pregnancy Tissues

2.1.1 Ethical Approval

Human pregnancy tissues were collected from terminated pregnancies according to the provisions of the Lower South Regional Ethics Committee approval:

Dr. Ian M. Morison, Erin Daly. “Gene activity during human placental development.” Lower South Regional Ethics Committee. 25 June 2007. Project Key: LRS/07/05/017. Localities: Cancer Genetics Laboratory, Department of Biochemistry, University of Otago. Approval was given to collect the products of conception from 5–9 week terminated pregnancies and placentas from full-term pregnancies at Dunedin Hospital.

2.1.2 Sample Collection

2.1.2.1 Collection of First-Trimester Placenta

The nurses at Dunedin Hospital presented our study to women who had elected to have a medical termination of pregnancy (5-9 wks gestation). For their medical procedure, women were given an oral dose of the prostaglandin, mifepristone (200 mg) to commence the abortion. Approximately 48 h later, the women were given a vaginal dose of misoprostol (800 µg) to induce labour. Most women passed their products of conception whole following the administration of misoprostol. After the products were successfully passed, women were asked once again to ensure that they still consented to participate in our study. The whole products of conception were then transferred to the laboratory for processing within 2-3 h of being passed.

A critical question relating to the collection of first-trimester placental tissue was whether the termination procedure could affect the epigenetic state of the tissue. At present, there is no evidence to suggest that the drugs used to induce pregnancy terminations alter DNA methylation. Although it would be useful to investigate, there is a widely accepted dogma
that once DNA methylation markings are established, they are relatively stable. Many believe
that DNA methylation becomes ‘locked in’ during the early stages of embryonic development
(after the epigenetic reprogramming of the embryo), and these marks are maintained
throughout subsequent cell divisions (Jones and Takai, 2001; Reik, 2007). It is important to
note, however, that DNA methylation levels for some oestrogen responsive genes can be
cylical given that some gene promoters display variable methylation levels over a 100-
minute interval (Kangaspeska et al., 2008). In general, because the methylation of the vast
majority of genes is stable, it does not seem likely that the drugs used to induce pregnancy
termination would alter the epigenetic state of the placental tissue collected in this study.

2.1.2.2 Collected of Term Placenta

Full term placentas were collected 3-4 h after delivery and immediately brought to the
laboratory for processing. Term placentas were collected from natural births and caesarean
sections.

2.1.3 Processing of Placental Samples

Pregnancy tissues were processed in the laboratory immediately after they were received. In
very few circumstances, tissue was preserved in 100% ethanol and kept at 4°C for 24-48 h.

2.1.3.1 Processing of First-Trimester Placenta

Most of the samples collected were the complete products of conception containing an intact
gestational sac (chorion) lined with trophoblastic villi. The gestational sac was assigned a
tissue sample identification number, placed in a Petri dish and washed with PBS to remove
residual blood. After measuring and taking photographs, the appearance of the tissue was
recorded. A small incision was made in the sac to locate the fetus, which was usually
contained within the amniotic sac and attached by the umbilical cord. If present, the fetus and
umbilical cord were removed and processed for a separate study.

After removal and processing of the fetus, the gestational sac was dissected. First, trophoblast
villous cells were carefully teased away from the outer chorionic membrane, washed with
PBS buffer and collected in 2 mL microcentrifuge tubes. Where trophoblast villous cells
were removed, the chorionic membrane was left exposed. The chorion was then carefully
teased away from the amnion (internal membrane) and both tissues were collected in separate
2 mL microcentrifuge tubes. Any maternal blood clots present in the products of conception
were also collected. Residual buffer was carefully discarded from all tubes containing tissue.
After dissection, placental tissues were snap frozen in liquid nitrogen and then immediately transferred to -80°C for permanent storage.

2.1.3.2  Processing of Term Placenta

For laboratory records, each full term placenta was assigned a tissue identification number. The placenta was transferred into a large glass dish and all excess fluid surrounding the placenta was decanted. The tissue was then examined and photographed. Roughly 1 cm³ portions of the umbilical cord were cut and transferred to a clean Petri dish. These cut sections were washed in PBS and transferred to 2 mL microcentrifuge tubes.

Placentas were positioned with amniotic side facing up and maternal side facing down. The two membranes of the placenta, the amnion and chorion, were then collected. First, the amniotic sac was stretched upwards to identify clean, healthy sections of membrane. Approximately 10 cm³ sections of the internal membrane (amnion) were carefully scraped off and transferred to a clean Petri dish. Once the amnion was removed, roughly 10 cm³ sections of the exposed external membrane (chorion) were cut out and transferred to another clean Petri dish. The membranous tissues were cut up into smaller sections, washed with PBS buffer and collected into 2 mL microcentrifuge tubes.

Before further processing, a careful inspection of the placental tissue was performed in order to avoid areas containing syncytial knots, maternal blood clots or other unhealthy looking tissue. Placentas were then positioned with amniotic side facing up and maternal side facing down. Approximately 2 cm-wide and 6 cm-long vertical slices were made into the placenta. Slices were removed and laid horizontally in the dish so that the three layers of the placenta (fetal, middle and maternal) could be identified. The top 1 cm of the slice was cut off as the “fetal layer,” the middle 1 cm was removed as the “middle layer” and the bottom 1 cm was taken as the “maternal layer”. Tissue from each layer of the placenta was transferred to a clean Petri dish and cut up into smaller pieces. Tissues were washed with PBS buffer, collected into 2 mL microcentrifuge tubes and immediately snap frozen in liquid nitrogen. Tissues were stored at -80°C.

2.1.4  DNA Extraction from Placental Tissue

Placental tissue was removed from -80°C. A 25 mg portion of tissue was taken and left to thaw to room temperature. DNA was extracted using the QIAamp DNA Mini Kit (Qiagen, Cat. #51306) with minor modifications to the manufacturer’s protocol. All reagents were
Materials and Methods

contained in the kit unless stated otherwise. First, 180 µL Buffer ATL was added to the sample and a pestle was used to break up the tissue. Next, 25 µL Proteinase K (600 mAU/mL) was added followed by incubation at 56°C until the tissue had completely lysed (~3 h). To help with tissue lysis, the sample was removed every 30 min to be vortexed for 30 sec. Once lysis was complete, 200 µL Buffer AL was added followed by incubation at 70°C for 10 min. Next, 200 µL of 100% ethanol was added and the sample was transferred to a QIAamp Spin Column. The sample was centrifuged at 16,000 x g for 1 min. Next, 500 µL Buffer AW1 was added and the sample was centrifuged at 16,000 x g for 1 min. Then, 500 µL Buffer AW2 was added and the sample was spun at 16,000 x g for 3 min, followed by a final spin for 1 min at 16,000 x g. To elute the DNA, 200 µL ddH₂O was added, the column was incubated for 10 min at room temperature and then spun at 16,000 x g for 1 min. The elution step was repeated and combined with the first elution for a total of 400 µL DNA. DNA concentration was determined by measuring 1.5 µL on a NanoDrop ND_1000 Spectrophotometer (NanoDrop Technologies). DNA was stored at 4°C.

2.2 Polymerase Chain Reaction Conditions

2.2.1 Standard PCR Protocol

A standard reaction mix of 20 µL was used for PCRs on genomic DNA:

<table>
<thead>
<tr>
<th>Reagent</th>
<th>Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>GeneAmp® 10x PCR Buffer II (ABI, Cat. #N808-0249)</td>
<td>2.0 µL</td>
</tr>
<tr>
<td>MgCl₂ (25 mM)</td>
<td>1.2 µL</td>
</tr>
<tr>
<td>Forward Primer (10 µM)</td>
<td>1.0 µL</td>
</tr>
<tr>
<td>Reverse Primer (10 µM)</td>
<td>1.0 µL</td>
</tr>
<tr>
<td>dNTPs (2.5 mM)</td>
<td>1.0 µL</td>
</tr>
<tr>
<td>ddH₂O</td>
<td>11.6 µL</td>
</tr>
<tr>
<td>AmpliTaq Gold® DNA Polymerase (ABI, Cat. #/N808-0249)</td>
<td>0.2 µL</td>
</tr>
<tr>
<td>DNA (2.5 ng/µL)</td>
<td>2.0 µL</td>
</tr>
</tbody>
</table>

* For a negative control, 2 µL ddH₂O was used instead of DNA.

The annealing temperature and magnesium chloride concentration were optimised for each PCR. If PCRs were not working after temperature and magnesium chloride optimisation, 5% dimethylsulfoxide (DMSO) was added to the reaction. The primer tables listed in this chapter contain the optimal conditions for each PCR.
2.2.2 Standard PCR Cycling Conditions

The following standard conditions used for PCR cycling were performed on a PTC-200 Peltier Thermal Cycler (MJ Research):

Enzyme activation: 95°C for 10 min
35 cycles of:
   95°C for 30 sec
   (Optimal annealing temperature)°C for 30 sec
   72°C for 30 sec
Final extension: 72°C for 7 min

2.2.3 PCR Product Detection

2.2.3.1 Agarose Gel Electrophoresis

A 2% agarose gel (SeaKem® LE Agarose; Lonza, Cat # 50004) was prepared with 1 x TAE buffer* and left to set for 20 min. PCR products were combined with xylene cyanol loading dye** and loaded into the gel. To measure PCR product size, 5 µL (0.1 µg) of 1 Kb Plus DNA Ladder™ (Invitrogen, Cat. #10787-018) was loaded into the first well of each gel. Products were electrophoresed at 100 V for up to 40 min in 1 x TAE buffer containing ethidium bromide (10 mg/mL). Gels were photographed under UV light in the BioRad GelDoc or Molecular Imager® FX systems to visualise PCR products.

To resolve fragments of similar size, PCR products were electrophoresed in 4% NuSieve® GTG® Agarose gels (Cambrex, Cat #50084) at 80 V for up to 120 min in 1 x TAE buffer containing ethidium bromide (10 mg/mL).

*Tris-acetate (TAE) Buffer (50x) (Sambrook et al., 1989)
242 g Tris
57.1 mL glacial acetic acid
100 mL 0.5 M EDTA (pH 8.0)
Add ddH₂O to make up to 1 L and then dilute 1:50 with ddH₂O before use.

**Xylene Cyanol Loading Dye (Sambrook et al., 1989)
0.25% bromophenol blue
0.25% xylene cyanol FF
15% Ficoll in water
Solution stored at room temperature.
2.3 Fetal Sex Determination for Placental Samples

2.3.1 Detecting the Y Chromosome

This PCR was designed to detect the presence or absence of the Y chromosome in the fetus of every placental sample. Primers were designed previously by Ryuji Fukuzawa (Cancer Genetics Lab, University of Otago) to amplify a region of the Sex-Determining Region Y (SRY) gene. SRY encodes the testis-determining factor and initiates male sex determination (Goodfellow and Lovell-Badge, 1993). The size of the amplified SRY product was 470 bp. After electrophoresis in a 2% agarose gel, the presence of a 470 bp product determined the sample to be male, and no presence of PCR product determined the sample to be female. The SRY primer sequences and PCR conditions are listed in Table 2.1.

2.3.2 Detecting both X and Y Chromosomes

A second PCR was designed to validate results from the SRY PCR for sex-determination. This PCR was used to detect the presence of both X and Y-chromosomes in placental samples. Primers were previously designed by Rob Weeks (Senior Research Fellow in Morison Lab, University of Otago) to amplify intron 1 of the Amelogenin gene, which is located on both sex chromosomes and has a 6 bp deletion on the X chromosome intron sequence (Sembon et al., 2008). Amplification of the X chromosome Amelogenin gene (AMELX) gives rise to a 106 bp product, and amplification of the Y chromosome Amelogenin gene (AMELY) gives rise to a 112 bp product. The use of one primer set to distinguish between sex chromosomes made this a useful PCR for sex-determination. PCR products were run on a 4% NuSieve agarose gel to detect the small (6 bp) difference in fragment lengths. The AMEL primer sequences and PCR conditions are listed in Table 2.1.

<table>
<thead>
<tr>
<th>Lab #</th>
<th>Primer Name</th>
<th>Primer Sequence</th>
<th>Annealing Temperature</th>
</tr>
</thead>
<tbody>
<tr>
<td>3159</td>
<td>SRY_F</td>
<td>GAA TAT TCC CGC TCT CCG GA</td>
<td>62°C</td>
</tr>
<tr>
<td>3160</td>
<td>SRY_R</td>
<td>GCT GGT GCT CCA TTC TTG AG</td>
<td></td>
</tr>
<tr>
<td>2905</td>
<td>AMELFor</td>
<td>CCC TGG GCT CTG TAA AGA ATA GTG</td>
<td>60°C</td>
</tr>
<tr>
<td>2906</td>
<td>AMELRev</td>
<td>ATC AGA GCT TAA ACT GGG AAG CTG</td>
<td></td>
</tr>
</tbody>
</table>
2.4 Maternal Contamination Screening by Microsatellite Analysis

It was critical to determine that placental samples contained purely fetal DNA because maternal DNA might have a different methylation profile. To ensure that each sample was purely fetal, five highly informative X chromosome microsatellite repeats were analysed. Since the sex of the fetal samples had been previously determined in the sex-determining PCRs, the presence of an extra X allele would suggest maternal DNA in the sample.

2.4.1 Primer Design

The most heterozygous X chromosome microsatellite repeats were chosen using the University of California Santa Cruz (UCSC) Genome Browser (Kent et al., 2002) and Invitrogens MapPairs (Invitrogen, 2005). Some of the primer sets used for microsatellite analysis were previously designed by Ryuji Fukuzawa (Cancer Genetics Lab, University of Otago). The microsatellite primer sequences are listed in Table 2.2.

2.4.2 $\alpha^{32}$P-labelled PCR

Microsatellite PCRs were first optimised using the standard PCR optimisation methods (see above) and products were run on a 2% agarose gel. Once PCRs were optimised, radioactivity was incorporated into the PCR products by adding 0.2 µL of $\alpha^{32}$P-dNTP (GE Healthcare, Cat. #AA00005) to the standard PCR mix. Optimised PCR conditions are listed in Table 2.2.

Table 2.2. Primer sequences and PCR conditions for X chromosome microsatellite analysis.

<table>
<thead>
<tr>
<th>Lab #</th>
<th>Primer Name</th>
<th>Primer Sequence</th>
<th>Annealing Temp</th>
<th>Chromosome Location</th>
<th>Heterozygosity</th>
</tr>
</thead>
<tbody>
<tr>
<td>3493</td>
<td>ss4915376_F</td>
<td>GTA AAA ATT TAC GGT TGT NCC AA</td>
<td>55°C</td>
<td>Xq26.2</td>
<td>0.69</td>
</tr>
<tr>
<td>3494</td>
<td>ss4915376_R</td>
<td>TCT CCC TAT CCA ACT CAT GC</td>
<td>55°C</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3161</td>
<td>DXS8051_F</td>
<td>ACC AGA AAT GAG CGA TTA TTG</td>
<td>51.1°C</td>
<td>Xp22.31</td>
<td>0.87</td>
</tr>
<tr>
<td>3162</td>
<td>DXS8051_R</td>
<td>TTT TTG AAC TAA GAA CCT GGA G</td>
<td>51.1°C</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3163</td>
<td>DXS6803_F</td>
<td>GAA ATG TGC TTG GAC AGG AA</td>
<td>55°C</td>
<td>Xq21.31</td>
<td>0.47</td>
</tr>
<tr>
<td>3164</td>
<td>DXS6803_R</td>
<td>CAA AAA GGG ACA TAT GCT ACT T</td>
<td>55°C</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3169</td>
<td>DXS8043_F</td>
<td>AGT TCT CAG AAA CAT TTG AGG C</td>
<td>55°C</td>
<td>Xq27.3</td>
<td>0.81</td>
</tr>
<tr>
<td>3170</td>
<td>DXS8043_R</td>
<td>AAT TAT TGG CAA AGA GTA CAG GCA G</td>
<td>55°C</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3171</td>
<td>DXS8377_F</td>
<td>CAC TTC ATG GCT TAC CAC AG</td>
<td>59°C</td>
<td>Xq28</td>
<td>0.95</td>
</tr>
<tr>
<td>3172</td>
<td>DXS8377_R</td>
<td>GAC CTT TGG AAA GCT AGT GT</td>
<td>59°C</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
2.4.3 Polyacrylamide Gel Electrophoresis

A 6% polyacrylamide/urea* solution was prepared with 1x TBE buffer**. The 6% polyacrylamide gel plug*** was poured and left to set for 10 min. The 6% denaturing polyacrylamide gel**** solution was prepared, poured and left to set for 30 min. PCR products containing radioactivity were diluted 2:1 with 95% formamide loading buffer, denatured at 95°C for 5 min and loaded into the pre-heated (50°C) polyacrylamide gel. Products were electrophoresed at 1500 V for up to 3 h in 1x TBE buffer. Following electrophoresis, the gel was attached to Whitman paper, dried at 80°C overnight and exposed to Kodak X-OMAT BT film.

During the final stages of X chromosome microsatellite analysis, regulations were imposed that made radioactivity (α³²P-dNTP) difficult to obtain so the protocol for microsatellite analysis was modified. The PCRs for the last four placental samples (F135-F138) were set up using the standard PCR protocol without radioactivity. To detect small differences in fragment size, PCR products were electrophoresed in a 4% MetaPhor® agarose gel (Lonza, Cat #50181) at 60 V for up to 3 h in 1x TAE buffer containing ethidium bromide (10 mg/mL). Two samples that were previously screened using radioactivity (F132 and F133) were included as controls to confirm this modified protocol for microsatellite analysis.

---

* 6% Polyacrylamide/Urea Solution
75 mL 40% Acrylamide/Bis (BioRad, Cat #161-0144)
240 g Urea
50 mL 1x TBE buffer
500 mL ddH₂O
Place solution in 37°C water bath to dissolve the urea.

** 5x Tris-borate (TBE) Buffer (Sambrook et al., 1989)
54 g Tris
27.5 g Boric acid
20 mL 0.5 M EDTA (pH 8.0)
Top up to 1 L with ddH₂O and dilute 1:5 with ddH₂O before use.

*** 6% Polyacrylamide Gel Plug
5 mL 6% Polyacrylamide/Urea solution
80 μL 10% Ammonium Persulfate (APS) (Sigma Aldrich, Cat #A3678)
15 μL Temed (Invitrogen UltraPure™ Temed, Cat #15524-010)

**** 6% Polyacrylamide/Urea Gel
45 mL 6% Polyacrylamide/Urea Solution
400 μL 10% Ammonium Persulfate (Sigma Aldrich, Cat #A3678)
20 μL Temed (Invitrogen UltraPure™ Temed, Cat #15524-010)
2.5 Methylated DNA Immunoprecipitation and Promoter Microarrays

At the time, the most appropriate technique available for genome-wide methylation analysis was methylated DNA immunoprecipitation (MeDIP) combined with gene promoter microarrays. This method enables the identification of CpG-rich gene promoter sequences at the genome-wide level. In the MeDIP technique, the monoclonal antibody anti-5-methylcytosine is used to immunoprecipitate methylated cytosine residues in the DNA (Reynaud et al., 1992). The immunoprecipitated (methylated) fraction is then compared to the initial total genomic DNA using a microarray-based method of analysis. The Aviva Systems Biology DNA Selection and Ligation (DSL) 20K promoter microarray platform was chosen because of its highly specific processing of samples and unique gene promoter microarray. First, the binding and ligation of adjacent 20-mer oligo probes create unique templates that correspond to the 40-mer probes spotted on the promoter microarray. Second, the ligated 40-mer oligo templates are PCR-amplified using T3 and T7 primers so that all sequences are amplified at the same efficiency. The MeDIP-DSL microarray platform is displayed in Figure 2.1.
Materials and Methods

Figure 2.1. Methylated DNA immunoprecipitation (MeDIP) using the Aviva Systems Biology DNA selection and ligation (DSL) platform and 20K promoter microarrays. 10 µg genomic DNA is digested by a restriction enzyme (NlaIII) and split into two groups: 0.5 µg is set aside as total genomic DNA and 9.5 µg is treated with anti-5-methylcytosine antibody to immunoprecipitate methylated DNA. The total genomic DNA and immunoprecipitated (methylated) DNA samples are biotinylated and 20-mer oligos are annealed. Adjacent oligos are ligated together, forming 40-mer oligos that serve as the template for PCR amplification. PCR is performed using dye-labelled primers (red for methylated DNA, green for total genomic DNA) and PCR products are co-hybridised to Aviva’s 20k gene promoter microarrays. Spot colour indicated promoter methylation level.
2.5.1 Sample Selection and Pooling

Only female samples were included in this experiment to avoid sex differences caused by the occurrence of X chromosome inactivation by DNA methylation. Samples were pooled within each group to minimize individual sample variation and identify the most common differences in methylation between first-trimester placenta and peripheral blood.

2.5.1.1 First-Trimester Placenta Samples

To investigate potential methylation changes during early gestation, first-trimester placental samples (from only female fetuses) were divided into groups based on gestational age. Samples were pooled within each gestational age group. A total of 15 first-trimester placenta samples were pooled into 3 groups: 5-6 wks (n = 4), 6-7 wks (n = 5) and 7+ wks (n = 6). The groups of first-trimester placenta samples that were pooled on the microarrays are listed in Table 2.3.

<table>
<thead>
<tr>
<th>Sample #</th>
<th>Gestational Age (weeks)</th>
<th>Sample #</th>
<th>Gestational Age (weeks)</th>
<th>Sample #</th>
<th>Gestational Age (weeks)</th>
</tr>
</thead>
<tbody>
<tr>
<td>F105</td>
<td>5.5</td>
<td>F114</td>
<td>6 +3</td>
<td>F121</td>
<td>7</td>
</tr>
<tr>
<td>F106</td>
<td>5 - 6</td>
<td>F117</td>
<td>6</td>
<td>F129</td>
<td>8</td>
</tr>
<tr>
<td>F124</td>
<td>5</td>
<td>F132</td>
<td>6+</td>
<td>F130</td>
<td>7</td>
</tr>
<tr>
<td>F125</td>
<td>5 +3</td>
<td>F135</td>
<td>6 +4</td>
<td>F131</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>F136</td>
<td>6 +4</td>
<td>F133</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>F137</td>
<td>7</td>
</tr>
</tbody>
</table>

2.5.1.2 Peripheral Blood Control Samples

Female post-natal peripheral blood DNA samples were selected as the representative somatic tissue in conjunction with Verity Oliver (Morison Lab, University of Otago). Peripheral blood control DNA was pooled into groups (n = 12 for each of four peripheral blood arrays). The groups of peripheral blood samples that were pooled on the microarrays are displayed in Table 2.4.
Table 2.4. Peripheral blood control sample selection and pooling for methylated DNA immunoprecipitation and promoter microarray assay.

<table>
<thead>
<tr>
<th>Microarray Slide 4: Peripheral blood females</th>
<th>Microarray Slide 5: Peripheral blood females</th>
<th>Microarray Slide 6: Peripheral blood females</th>
<th>Microarray Slide 7: Peripheral blood females</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample #</td>
<td>Age (yrs)</td>
<td>Sample #</td>
<td>Age (yrs)</td>
</tr>
<tr>
<td>X7145</td>
<td>5.2</td>
<td>X7084</td>
<td>5.6</td>
</tr>
<tr>
<td>X7157</td>
<td>5.8</td>
<td>X7056</td>
<td>5.6</td>
</tr>
<tr>
<td>X7149</td>
<td>8.2</td>
<td>X7064</td>
<td>5.2</td>
</tr>
<tr>
<td>X7172</td>
<td>5.5</td>
<td>X7058</td>
<td>4.9</td>
</tr>
<tr>
<td>X7118</td>
<td>7.8</td>
<td>X7076</td>
<td>8.2</td>
</tr>
<tr>
<td>X7128</td>
<td>5.7</td>
<td>X7059</td>
<td>7.1</td>
</tr>
<tr>
<td>X7151</td>
<td>5.5</td>
<td>X7110</td>
<td>7.7</td>
</tr>
<tr>
<td>X7142</td>
<td>7.6</td>
<td>X7095</td>
<td>4.7</td>
</tr>
<tr>
<td>X7181</td>
<td>4.7</td>
<td>X7080</td>
<td>7.9</td>
</tr>
<tr>
<td>X7174</td>
<td>5.4</td>
<td>X7071</td>
<td>5.7</td>
</tr>
<tr>
<td>X7177</td>
<td>7.0</td>
<td>X7098</td>
<td>5.3</td>
</tr>
<tr>
<td>X7154</td>
<td>5.0</td>
<td>X7108</td>
<td>5.5</td>
</tr>
</tbody>
</table>

2.5.2 Methylated DNA Immunoprecipitation (MeDIP)

The Aviva Systems Biology DNA Methylation ChIP-DSL protocol (version 1.4) was followed to perform methylated DNA immunoprecipitation.

2.5.2.1 NlaIII Restriction Digest

Genomic DNA required digestion for optimal antibody binding in the immunoprecipitation step. The NlaIII restriction enzyme (New England Biolabs, Cat. #R0125L) was chosen because of its 4 bp recognition site, which digests DNA into pieces of roughly 250-500 bp. The restriction activity of NlaIII was optimised on 1 µg of peripheral blood DNA and the optimal volume chosen for restriction digest was 20 µL of NlaIII for 10 µg of genomic DNA.

The NlaIII restriction digest of was set up as follows:

Genomic DNA 10 µg
10 x NEBuffer 4 (NEB, Cat. #R0125L) 20 µL
Bovine Serum Albumin (100 µg/mL; NEB, Cat. #R0125L) 10 µL
ddH₂O make up to final volume of 200 µL
NlaIII enzyme (NEB, Cat. #R0125L) 20 µL

DNA was digested at 37°C for 3 h, followed by enzyme inactivation at 65°C for 20 min.
2.5.2.2 *Digested DNA Purification*

Digested DNA was purified using the QIAquick PCR Purification Kit (Qiagen, Cat. #28104) according to the manufacturer’s protocol. Briefly, 1100 µL Buffer PB was added to the DNA sample, which was loaded onto a spin column and centrifuged at 16,000 x g for 1 min. Next, 750 µL Buffer PE was added and the sample was centrifuged at 16,000 x g for 1 min followed by an additional spin 16,000 x g for 1 min. DNA was eluted in 50 µL ddH₂O.

2.5.2.3 *Removal of Total Genomic DNA*

DNA concentration was determined for each elution by measuring 1.5 µL on a NanoDrop ND_1000 Spectrophotometer (NanoDrop Technologies). Once DNA concentration was determined, 500 ng of DNA was set aside as the “Total Genomic DNA” sample that would be compared to the immunoprecipitated sample on the microarray.

2.5.2.4 *Immunoprecipitation of Methylated DNA*

Digested DNA was first denatured at 95°C for 10 min, then immediately put on ice. The volume of each sample was adjusted to 150 µL with PBS buffer containing 0.05% Triton X-100, pH 7.0 (Sigma, Cat #T-9284). Next, 5 µg (5 µL) anti-5-methylcytosine antibody (ASB, Cat. #AVAMM99021) was added and samples were incubated at 4°C overnight on a rotating platform. Following incubation, 40 µL of Zymed Rec-Protein G-Sepharose® 4B Conjugate beads (Invitrogen, Cat #10-1241) were added and samples were incubated for 1 h at room temperature on a rotating platform. Following incubation, samples were carefully transferred into 5 mL PBS in a 15 mL conical centrifuge tube, spun at 2000 x g for 4 min and the supernatant was discarded. This wash step was repeated twice more, followed by a final wash step with 1 mL TE buffer. Next, 50 µL Elution Buffer (ASB, Cat #AK-0503) was added, samples were vortexed to re-suspend beads and incubated for 15 min in a 65°C water bath. During incubation, samples were vortexed briefly every 2 min. Samples were then spun for 30 sec at 16,000 x g and the supernatant was collected. This elution was repeated twice more and supernatants were pooled giving a final volume of 150 µL immunoprecipitated DNA.

Immunoprecipitated DNA was then purified using the QIAquick PCR Purification Kit (Qiagen, Cat. #28104) according to the manufacturer’s protocol (see above). DNA was eluted in 50 µL ddH₂O. To concentrate the DNA, the 50 µL eluate was evaporated in a speed-vacuum for 1 h. A total volume of 20 µL ddH₂O was then added to each sample and left to resuspend for 2 h. The samples were mixed by vortexing and centrifuged for 1 min at
16,000 x g to pellet any remaining protein. The 20 µL supernatant was carefully transferred to a clean tube without disturbing the protein pellet. DNA concentration was determined for each elution by measuring 1.5 µL on a NanoDrop ND_1000 Spectrophotometer (NanoDrop Technologies). The immunoprecipitated DNA was either used immediately for subsequent biotinylation or stored at -80°C.

2.5.3 DNA Selection and Ligation

The following steps used the Aviva Systems Biology ChIP-DSL protocol (version H20K 2.0). All reagents were contained in the DSL-H20K Kit (ASB, Cat #AK-0504) unless stated otherwise.

2.5.3.1 Biotinylation of DNA

PCR tubes were prepared so that one tube was for “Total Genomic DNA” and the other tube was for “MeDIP DNA”. In each tube, 18 µL DNA was combined with 1 µL lamda phage (λ) DNA (100 ng/µL) and 1 µL PHOTOPROBE® Biotin. The samples were incubated in a PCR machine at 95°C for 10 min. Following incubation, 60 µL ddH2O was added to each reaction and samples were transferred into 1.7 mL centrifuge tubes. Subsequently, 80 µL 0.1 M Tris-HCl (pH 9.5) was added and samples were thoroughly vortexed. Next, 160 µL 2-butanol (at room temperature) was added and samples were again thoroughly vortexed, followed by centrifugation at 16,000 x g for 3 min to separate the phases. The upper butanol phase was carefully removed and discarded. This extraction was repeated with an additional 160 µL of 2-butanol, followed by centrifugation at 16,000 x g for 3 min and the butanol layer was discarded. During this extraction, the volume of the aqueous phase decreased to ~60 µL after the second extraction. The aqueous phase was then transferred into a new 1.7 mL centrifuge tube by pipetting from the bottom of the tube.

Biotinylated DNA was then precipitated by adding the following reagents in this order: 10 µL 10 M C2H3O2NH4 (ammonium acetate), 2 µL 1 M MgCl2, 1 µL glycogen, and 200 µL ice cold 100% ethanol. Samples were mixed well and incubated at -20°C for 30 min, then centrifuged (at 4°C) for 20 min at 16,000 x g to pellet the precipitated nucleic acids. The pellet was washed with 0.5 mL cold 70% ethanol and centrifuged (at 4°C) at 16,000 x g for 5 min. The supernatant was then carefully poured off and the pellets were left to air dry at room temperature. Pellets were then re-suspended in 10 µL TE Buffer and stored at 4°C.
2.5.3.2 Annealing and Ligation

Both DSL Wash Buffer and 1 x Taq DNA Ligase Buffer (NEB, Cat. #B0208S) were pre-warmed to 45°C. Streptavidin-coated magnetic beads were washed before use: the stock tube was vortexed well to re-suspend beads and 5 µL of bead slurry (per reaction) was transferred into a PCR tube. Next, 150 µL 2 x Binding Buffer was added and the mixture was pipetted to re-suspend beads. The tube was then placed in a 96-well magnetic plate holder for 2 min. Once the beads had settled to the side of the tube, the buffer was carefully discarded by pipetting without disturbing the beads. This wash was repeated once more. The beads were finally re-suspended in 5 µL 2 x Binding Buffer per reaction tube.

DNA selection and ligation reactions were set up in PCR tubes as follows:

<table>
<thead>
<tr>
<th>Component</th>
<th>Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>Biotinylated DNA</td>
<td>10 µL</td>
</tr>
<tr>
<td>H2OK Oligo Mix</td>
<td>10 µL</td>
</tr>
<tr>
<td>2x Binding Buffer</td>
<td>20 µL</td>
</tr>
</tbody>
</table>

Reaction tubes were placed in a thermal cycler, heated to 95°C for 10 min and then cooled to 45°C for 30 min. Following incubation, 5 µL of pre-washed streptavidin-coated magnetic beads was added to each sample and samples were mixed thoroughly by pipetting. Tubes were then incubated in a thermal cycler at 45°C for 1.5 h. Following incubation, tubes were placed into a 96-well magnetic holder and 180 µL pre-warmed (45°C) DSL Wash Buffer was added to each sample. After mixing by pipette, samples were left for 2 min to allow the beads to settle to the sides of the tube. All liquid was then carefully removed without disturbing the beads. This wash step was repeated twice more using the pre-warmed (45°C) DSL Wash Buffer and a final wash using 150 µL pre-warmed (45°C) 1 x Taq DNA Ligase Buffer (NEB, Cat. #M0208S). Beads were then re-suspended in 40 µL 1 x Taq DNA Ligase Buffer (NEB, Cat. No. B0208S) and 1 µL Taq DNA Ligase (NEB, Cat. #M0208S) was added to each reaction tube. Tubes were then incubated in a thermal cycler at 45°C for 1 h.

Following incubation, tubes were placed in a 96-well magnetic holder and beads were left to settle for 2 min. Supernatants were carefully removed by pipette and 180 µL pre-warmed (45°C) DSL Wash Buffer was added to wash the beads. This wash was repeated and residual solution was carefully removed after the final wash. Beads were re-suspended in 40 µL ddH2O and samples were mixed well by pipetting.
Samples were eluted by heating at 95°C for 10 min and then immediately chilled on ice for 5 min. Following elution, tubes were placed into the 96-well magnetic holder and beads were left to settle for 2 min. Supernatants were carefully transferred into clean 0.6 mL PCR tubes and the magnetic beads were discarded. Samples were stored at -20°C until used in the subsequent PCR amplification step.

2.5.3.3  PCR Amplification and Labeling of DSL Products

Two PCR reactions were performed for all samples. The first PCR was used as a diagnostic tool to test the quality of the DSL reaction. The second PCR was to label the DSL products with fluorescent dyes for later use on microarrays.

The diagnostic PCR was set up as follows:

GeneAmp® 10x PCR Buffer II (ABI, Cat. #N808-0249) 2.5 µL
MgCl₂ (25 mM) 1.5 µL
dNTPs (2.5 mM) 0.5 µL
T7 Primer (30 µM) 0.5 µL
T3 Primer (30 µM) 0.5 µL
Sample 2.0 µL
ddH₂O 17.1 µL
AmpliTaq Gold® DNA Polymerase (ABI, Cat. #/N808-0249) 0.4 µL

The diagnostic PCR was performed on a PTC-200 Peltier Thermal Cycler (MJ Research) using the following cycling conditions:

Enzyme activation: 95°C for 5 min
25 cycles of: 95°C for 30 sec
54°C for 2 min
72°C for 2 min
Final extension: 72°C for 5 min
Hold: 4°C until use

After PCR, 5 µL of PCR product was visualized on a 2% agarose gel stained with ethidium bromide. If a faint band was detected, the labeling PCR was then set up as follows:

GeneAmp® 10x PCR Buffer II (ABI, Cat. #N808-0249) 2.5 µL
MgCl₂ (25 mM) 1.5 µL
dNTPs (2.5 mM) 0.5 µL
T3 Primer (30 µM) 0.5 µL
Add either: 15 µM Atto 550-T7 Primer (Total Genomic DNA only) 1.0 µL
15 µM Atto 647-T7 Primer (MeDIP Genomic DNA only) 1.0 µL
DNA Sample 4.0 µL
ddH₂O 14.6 µL
AmpliTaq Gold® DNA Polymerase (ABI, Cat. #/N808-0249) 0.4 µL

The same PCR cycling conditions were used as the first PCR but for a total of 30 cycles.
After the labeling PCR, 5 µL of PCR product was visualized on a 2% agarose gel stained with ethidium bromide. The remaining labeled PCR samples were protected from light and immediately used for hybridisation to microarrays.

2.5.4 **Hybridisation to 20K Gene Promoter Microarrays**

2.5.4.1 **Blocking of Microarray Slides**

The Hybridization Kit (ASB, Cat. #AK-0505) provided a blocking buffer and three pre-wash solutions. The 1 x Blocking Buffer was pre-warmed to 50°C in a water bath. Each 250 mL PreWash Solution (1, 2 and 3) was aliquoted into five 50 mL transparent conical centrifuge tubes so that each tube contained 50 µL PreWash solution. The H20K microarray slide (ASB, Cat. #AC-0020) was placed into 50 mL PreWash Solution 1 and inverted gently for 5 min at room temperature. The slide was carefully removed and transferred to 50 mL PreWash Solution 2 and mixed gently for 2 min. The slide was removed and placed into the second 50 mL tube containing PreWash Solution 2 and inverted gently for 2 min. The slide was removed and transferred to 50 mL PreWash Solution 3 and inverted gently for 10 min. Next, the slide was removed and placed in 50 mL ddH2O and inverted gently for 1 min. The slide was immediately transferred into 50 mL pre-warmed (50°C) 1 x Blocking Buffer and then incubated in a 50°C water bath for 15 min. The slide were removed and placed into fresh 50 mL ddH2O and inverted gently for 1 min at room temperature. The slide was then transferred into an empty 50 mL conical centrifuge tube (with slide label toward the bottom of tube) and dried by centrifugation at 800 x g for 2 min.

2.5.4.2 **Hybridisation to Microarray Slides**

The blocked microarray slide was placed in a hybridization chamber and a clean cover slip was cut to size and placed on top of the microarray approximately 5 mm from the long edges of the slide. The hybridisation probe was prepared in an amber 50 mL conical centrifuge tube during the 15 min blocking incubation mentioned above. Each microarray slide required 60 µL hybridisation probe, prepared as follows:

- Hybridisation Solution: 30 µL
- Atto-550-labeled DSL sample (Total Genomic sample): 10 µL
- Atto-647-labeled DSL sample (MeDIP sample): 10 µL
- ddH2O: 10 µL
The hybridisation probe was thoroughly vortexed and then heated at 95°C for 5 min in the dark. After incubation, samples were immediately placed on ice for 2 min and then briefly centrifuged at 16,000 x g. Next, 55 µL of the hybridization probe was slowly added to the corner of each cover slip so that it completely filled (by capillary action) the area under the cover slip with no air bubbles. Then, 20 µL ddH₂O was added to the wells of the hybridization chamber and the chamber was assembled and incubated at 50°C overnight.

2.5.4.3 Washing of Microarray Slides

Washing solutions were prepared according to the manufacturer’s protocol. The Wash Solution was pre-warmed to 50°C in a water bath. The hybridisation chamber was disassembled and the microarray slide was carefully transferred (using tweezers) into 50 mL of pre-warmed Wash Solution and gently mixed for 1 min. The slide was then carefully removed making sure that the cover slip was left behind in the Wash Solution. The slide was placed into a second 50 mL of pre-warmed Wash Solution and the tube was incubated at room temperature for 15 min on a vertical rotating platform. The slide was then transferred into 50 mL of 0.2 x SSC solution, followed by gentle mixing for 1 min at room temperature. This step was repeated in the second 50 mL of 0.2 x SSC solution. The slide was then transferred into 50 mL of 0.1 x SSC solution and inverted gently for 1 min. This step was repeated in a second 50 mL of 0.1 x SSC. The slide was then placed into an amber conical centrifuge tube and dried by centrifugation at 800 x g for 5 min. The microarray slide was then ready for scanning and analysis.

2.5.4.4 Microarray Scanning

The microarray slide was scanned using the GenePix® 4000B scanner (Molecular Device) and the GenePix® Pro 6.0 computer software. The slide was placed face down on the scanner and the GenePix program was used to set scanning preferences. The photo-multiplier tube (PMT) gain was set to 660 in both channels (red and green) and a preview scan of the slide was performed. The PMT gain could then be manually adjusted for brightness, contrast, and saturation; but in this study no PMT gain adjustments were made. The scan area was set and a full data scan was performed at resolution of 5 µm with a PMT of 660 for both channels. After scanning, an image alignment was performed by loading the GenePix array list (GAL) file provided by Aviva Systems Biology. Once all spots were correctly aligned, the image was saved as a GenePix results (GPR) file and imported into Microsoft Excel for data analysis.
2.5.5 **Microarray Data Analysis**

Microarray data analysis was performed by Dr. Simon Andrews at Babraham Bioinformatics (Cambridge, United Kingdom). Simon has extensive experience in MeDIP microarray analysis. He was also working as a colleague of Prof. Ian Morison at the Babraham Institute during Ian’s one-year sabbatical in 2008.

2.5.5.1 **Microarray Data Normalisation**

The microarray data set was analysed to detect differences in methylation levels between the pooled array groups of first-trimester placenta and peripheral blood. First, a Loess normalisation was performed on the raw data using the GeneSpring GX® software (Agilent Technologies). Normalised data from the arrays were combined into groups for filtering and statistical analysis. Three separate analyses were performed.

2.5.5.2 **Microarray Data Analysis #1**

The first data analysis was performed on five data sets: three data sets of first-trimester placenta pooled by gestational age (5-6 wks, 6-7 wks, and 7-9 wks) and two data sets of pooled peripheral blood. The two peripheral blood data sets were compared to the three first-trimester placenta experimental groups. After data normalization, +/- 1.4 fold-change cut-offs were set to determine consistently “methylated” or consistently “unmethylated” genes between first-trimester placenta and peripheral blood. This cut-off value was based on previous MeDIP analyses (Farthing *et al*., 2008) and was used as a baseline from which to look for outliers in the overall distribution of the data. Genes were classed as “methylated” if they had a fold change greater than or equal to 1.4 (red/green ≥ 1.4), and genes were classed as “unmethylated” if they had a fold change less than or equal to 0.7 (red/green ≤ 0.7).

2.5.5.3 **Microarray Data Analysis #2**

A quality-threshold (QT) clustering analysis (≥ 10 genes per group, R ≥ 0.7) using the GeneSpring GX® QT clustering algorithm was performed to identify genes that displayed similar increasing or decreasing patterns of methylation. QT clustering is a method for identifying and grouping gene expression patterns in multiple data sets (Heyer *et al*., 1999). The QT clustering analysis was performed on the same microarray data sets that were used in the previous analysis (three first-trimester placenta and two peripheral blood data sets).
2.5.5.4 Microarray Data Analysis #3

Data from two additional peripheral blood microarrays became available from Verity Oliver (Morison Lab, University of Otago), so they were added to the previous peripheral blood data. A third analysis was performed comparing the four peripheral blood data sets to the three first-trimester placenta data sets. In this analysis, the 20 most significantly changing genes between first-trimester placenta and peripheral blood were identified.

2.5.5.5 Selection and Location of Differentially Methylated Candidate Genes

Candidate genes were selected based on the three methods of microarray analysis explained above. To identify the exact region of differential methylation, the 40-mer probe sequences spotted on the microarray were obtained from Aviva Systems Biology. Genomic locations had been revised since the arrays were designed in 2005, so it was necessary to obtain the updated genomic locations of the 40-mer probe sequences for accurate candidate gene validation. The online PromoSer tool (Halees et al., 2003) was used to extract a 1000 bp genomic sequence that was 800 bp upstream and 200 bp downstream from the gene's transcription start site. The probe was usually located within the extracted 1000 bp promoter sequence, in either forward or reverse orientation. If probes were not located within the sequence extracted from PromoSer, they were found using GenBank’s online sequence alignment tool, BLAST (Benson et al., 1999).
2.6 **Sequenom MassARRAY® EpiTYPER® Platform for Quantitative Methylation Analysis**

The Sequenom MassARRAY® EpiTYPER® platform (referred to now on as Sequenom) is a quantitative tool for high-throughput DNA methylation analysis. It has become a widely used technique that relies on MALDI-TOF (Matrix-Assisted Laser Desorption/Ionization Time-of-Flight) Mass Spectrometry to quantify DNA methylation in fragments of bisulfite-converted PCR product that have been reverse-transcribed and cleaved base-specifically (Figure 2.2) (Ehrich et al., 2005). This technique can be used to identity multiple differentially methylated CpG sites in DNA amplimers of up to 600 bp.

![Figure 2.2](image)

**Figure 2.2.** The Sequenom MassARRAY® EpiTYPER® platform (adapted from Ehrich et al., 2005). Genomic DNA is bisulfite-treated and PCR is performed. After PCR, unmethylated cytosines (blue) are converted to thymines and methylated cytosines (red) remain as cytosines. PCR product is reverse-transcribed into single-stranded RNA so that methylated sequences contain guanines at CpG sites and unmethylated sequences contain adenines. Base-specific (U/T) cleavage is performed followed by mass spectrometry. Since guanines are 16 Daltons heavier than adenines, a mass change is observed and DNA methylation is quantified by comparing the different mass spectrometry signals (peak area) of methylated and non-methylated DNA.
2.6.1 Sample Selection for Sequenom

In addition to the first-trimester placenta and peripheral blood samples that were analysed in the MeDIP-microarray experiment, a selection of fetal and adult tissues, including sections from one term placenta, were chosen for examination by Sequenom analysis. All tissue samples included in this study were obtained from the Cancer Genetics Laboratory (University of Otago). DNA was extracted from fetal and adult somatic tissues (listed in Table 5.1) using the QIAamp DNA Mini Kit (Qiagen, Cat. #51306) as described previously (pg. 35).

2.6.2 Sequenom Primer Design

To validate candidate gene methylation by Sequenom, primers were designed to amplify the region corresponding to the 40 bp probe that was spotted on the promoter microarray. Primers were designed for bisulfite-converted DNA using methPrimerDB (Pattyn et al., 2006) or EpiDesigner (http://www.epidesigner.com/). The optimal amplicon length was set to 350 bp. As instructed by the EpiTYPER® manual, a T7-promoter tag was incorporated into the reverse primer for the post-PCR reverse-transcription step and a 10 bp tag was incorporated into the forward primer to balance primer lengths. Primer design for Sequenom is displayed in Figure 2.3.

```
5' ─ CAGTAATACGACTCATATAGGG ─ AGAAGGCT ─ Reverse Gene Primer Sequence ─ 3'

5' ─ AGGAAGAGAG ─ Forward Gene Primer Sequence ─ 3'
```

*Figure 2.3.* Primer design for Sequenom MassARRAY® EpiTYPER® platform. The green sequence on the reverse primer shows the T7-promoter tag and the red sequence acts as a constant 5’ fragment for the RNase A reaction. The blue sequence on the forward primer is incorporated to balance primer lengths in PCR amplification.

2.6.3 Bisulfite Conversion of Genomic DNA

Genomic DNA was bisulfite-converted using the protocol and reagents from the EZ DNA Methylation™ Kit (Zymo Research, Cat. #D5002). Some modifications were made to the protocol as suggested by colleagues at Babraham Bioinformatics (Cambridge, UK). First, 45 µL ddH₂O was added to a PCR tube (one tube per sample). The volume required for 1 µg of DNA was removed from the ddH₂O and the same volume of DNA was then added so that
each 45 µL volume contained 1 µg DNA. Next, 5 µL Dilution Buffer was added and tubes were vortexed and incubated at 37°C for 15 min using a thermal cycler. Then, 100 µL of the prepared CT Conversion Reagent was added to each sample, avoiding exposure to light. Samples were then placed in a thermal cycler for a cycling incubation using Zymo’s “Alternative Cycling Protocol” as stated below:

Denaturation: 95°C for 30 sec
Conversion: 50°C for 15 min
Repeat for 20 cycles

After incubation, samples were immediately placed on ice for 10 min. Samples were transferred into 400 µL Binding Buffer in 1.7 mL centrifuge tubes, mixed by pipetting and transferred into Zymo-Spin™ I columns. After centrifugation at 16,000 x g for 30 sec, 200 µL Wash Buffer was added and samples were centrifuged for 30 sec at 16,000 x g. Next, 200 µL Desulphonation Buffer was added, samples were incubated at room temperature for 15 min and tubes were centrifuged at 16,000 x g for 30 sec. Next, 200 µL Wash Buffer was added and tubes were centrifuged at 16,000 x g for 30 sec. This step was repeated once more. DNA was eluted in ddH₂O to obtain a final volume of 60 µL bisulfite-converted DNA.

2.6.4 Sequenom PCRs

The polymerase chain reaction was used to amplify regions of bisulfite-converted DNA. It has been shown that immediate processing of bisulfite-converted DNA yields optimal PCR product (Christine Cauldrey, AgResearch Ruakura), so PCR was performed directly after the bisulfite conversion of genomic DNA whenever possible. The primer sequences and PCR conditions for all Sequenom PCRs performed in this study are listed in Table 2.5 and do not include the primer tags shown in Figure 2.3.

2.6.4.1 Optimisation of Bisulfite PCRs

Before PCR primers were used, an optimisation of the annealing temperature and magnesium concentration was always performed. It was critical to optimise Sequenom PCRs to maximise the amount of amplified product and minimise the amount of primer-dimer. It has been shown that primer-dimer can interfere with the Sequenom assay since primer sequences can be reverse-transcribed and cleaved, thus contributing to the amount of product (signal) present in the mass spectrometry spectra. Sequenom’s “spectra quality” measures the amount of signal present and acts as the assay’s quality control. The presence of primer-dimer can increase the amount of apparent signal and therefore decrease the accuracy of the assay by
falsely indicating strong product signal when there may be poor overall PCR fragment quality (van den Boom and Ehrich, 2009). PCR optimisation was therefore critical to maintain the accuracy of Sequenom’s spectra quality.

Primer sets were initially tested at a temperature range of 55°C–65°C, and if the optimal product was not detected, primers were tested at a temperature range of 50°C–55°C. PCR products were run on a 2% agarose gel stained with ethidium bromide to determine if sufficient PCR product was present. All primer sets showed optimal performance at 1.5 mM MgCl₂.

Table 2.5. Primer sequences and PCR conditions for genes analysed using the Sequenom MassARRAY® EpiTYPER® platform.

<table>
<thead>
<tr>
<th>Gene Name</th>
<th>Primer Name and Number</th>
<th>Sequence (5’→3’ )</th>
<th>Product Cpgs</th>
<th>Product Length (bp)</th>
<th>Annealing Temp</th>
</tr>
</thead>
<tbody>
<tr>
<td>GABRA6</td>
<td>NM_000811_F (#3738)</td>
<td>AAG AGT TGT GTA GGA ATT TTT TTG TG</td>
<td>4</td>
<td>391</td>
<td>60°C</td>
</tr>
<tr>
<td></td>
<td>NM_000811_R (#3739)</td>
<td>AAA TTT ACA AAA TTC CCC TAA CTC C</td>
<td>4</td>
<td>321</td>
<td>60°C</td>
</tr>
<tr>
<td>C20orf144</td>
<td>NM_080825_F (#3740)</td>
<td>AGG GTT AAA ATT TAT ATT TAA ATT ATA GAA</td>
<td>4</td>
<td>436</td>
<td>59.4°C</td>
</tr>
<tr>
<td></td>
<td>NM_080825_R (#3741)</td>
<td>AAC AAA ACT AAT CAC CAC TAA AAA C</td>
<td>4</td>
<td>436</td>
<td>59.4°C</td>
</tr>
<tr>
<td>KCTD4</td>
<td>NM_198404_F (#3744)</td>
<td>ATA TTT GGA TTT GTT TAA TTG ATA AGA</td>
<td>8</td>
<td>436</td>
<td>59.4°C</td>
</tr>
<tr>
<td></td>
<td>NM_198404_R (#3745)</td>
<td>ACT ATT ATA TTT CCC TTC ATA CTC CTT</td>
<td>8</td>
<td>436</td>
<td>59.4°C</td>
</tr>
<tr>
<td>KRT17</td>
<td>NM_000422_F2 (#3762)</td>
<td>ATT TTI TTA TTA GAT AGG GTT GAG G</td>
<td>13</td>
<td>396</td>
<td>60°C</td>
</tr>
<tr>
<td></td>
<td>NM_000422_R2 (#3763)</td>
<td>TTC CTT TAT AAA CCA CCA AAT AAA C</td>
<td>13</td>
<td>396</td>
<td>60°C</td>
</tr>
<tr>
<td>AKR1C4</td>
<td>NM_001818_F2 (#3764)</td>
<td>ATT TGT TGG ATG AGT TTA TTA GAT AGA TAG</td>
<td>3</td>
<td>433</td>
<td>60°C</td>
</tr>
<tr>
<td></td>
<td>NM_001818_R2 (#3765)</td>
<td>ATA CAT AAA TAC CAA ATC CCA ATA C</td>
<td>3</td>
<td>433</td>
<td>60°C</td>
</tr>
<tr>
<td>NBR1</td>
<td>NM_005899_F (#3750)</td>
<td>GTA GTT TTT TAA GGT TAG TGG TTT G</td>
<td>27</td>
<td>497</td>
<td>60°C</td>
</tr>
<tr>
<td></td>
<td>NM_005899_R (#3751)</td>
<td>TTA CAA AAT TCC TCC CTT AAA CTT C</td>
<td>27</td>
<td>497</td>
<td>60°C</td>
</tr>
<tr>
<td>PCDHA12</td>
<td>NM_018903_F (#3752)</td>
<td>AGA GTG ATT TAG GAA GTG GTT AAA T</td>
<td>10</td>
<td>404</td>
<td>60°C</td>
</tr>
<tr>
<td></td>
<td>NM_018903_R (#3753)</td>
<td>CCC AAA CTA CAA AAA ACA AAA AC</td>
<td>10</td>
<td>404</td>
<td>60°C</td>
</tr>
<tr>
<td>RASSF2</td>
<td>NM_014737_F2 (#3766)</td>
<td>TTT TTT TTA GAT AAT GTT GTG GTT G</td>
<td>12</td>
<td>404</td>
<td>58.3°C</td>
</tr>
<tr>
<td></td>
<td>NM_014737_R2 (#3767)</td>
<td>AAA AAA CCT ATT TAA CCC CTA TCC</td>
<td>12</td>
<td>404</td>
<td>58.3°C</td>
</tr>
<tr>
<td>HINT1</td>
<td>NM_005340_F2 (#3768)</td>
<td>TTT TAG GAG ATA AGA AAG GAG GAT AGA A</td>
<td>26</td>
<td>339</td>
<td>60°C</td>
</tr>
<tr>
<td></td>
<td>NM_005340_R2 (#3769)</td>
<td>ACC TTA ACA ATC TCA TCT ACC ATC TC</td>
<td>26</td>
<td>339</td>
<td>60°C</td>
</tr>
<tr>
<td>Gene Name</td>
<td>Primer Name and Number</td>
<td>Sequence (5’→3’)</td>
<td>Product CpGs</td>
<td>Product Length (bp)</td>
<td>Annealing Temp</td>
</tr>
<tr>
<td>-------------</td>
<td>------------------------</td>
<td>------------------</td>
<td>--------------</td>
<td>---------------------</td>
<td>----------------</td>
</tr>
<tr>
<td>MYO5A</td>
<td>MYO5A_F (#3840)</td>
<td>AGA GGG TTT TTT TTA AAT TTT TTT T</td>
<td>27</td>
<td>323</td>
<td>58.3°C</td>
</tr>
<tr>
<td></td>
<td>MYO5A_R (#3841)</td>
<td>ACA AAA CAT CAT CAT CAA CCT A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CACNA2D1</td>
<td>CACNA2D1_F (#3842)</td>
<td>ATT TTT AGG TAG GTA TTG GAT TTT G</td>
<td>31</td>
<td>424</td>
<td>56.3°C</td>
</tr>
<tr>
<td></td>
<td>CACNA2D1_R (#3843)</td>
<td>AAC AAA CAA AAA AAC TTA TAA AAA C</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BCL9</td>
<td>BCL9_F (#3844)</td>
<td>TTA GTT GAG TTG TTT GAG AGA GAG AGA</td>
<td>4</td>
<td>324</td>
<td>64.5°C</td>
</tr>
<tr>
<td></td>
<td>BCL9_R (#3845)</td>
<td>ACT AAC AAA CCA CAA ACA AAA CAT TC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FRG1</td>
<td>FRG1_F (#3846)</td>
<td>TTT TGT TTT TTT GTA TAT ATA GT</td>
<td>40</td>
<td>405</td>
<td>60°C</td>
</tr>
<tr>
<td></td>
<td>FRG1_R (#3847)</td>
<td>CTA AAT AAC TCA AAA TTC CAC TIC C</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CHKB</td>
<td>CHKB_F (#3848)</td>
<td>GGG TAT TTA TGT AAT TGG AGT TAT</td>
<td>51</td>
<td>491</td>
<td>60°C</td>
</tr>
<tr>
<td></td>
<td>CHKB_R (#3849)</td>
<td>AAA CCA ATA TGT AAT TGC TAC TIC C</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KIF22</td>
<td>KIF22_F (#3850)</td>
<td>TGA GTG GGA TTG AAT TAT AGA TTA AG</td>
<td>18</td>
<td>374</td>
<td>52°C</td>
</tr>
<tr>
<td></td>
<td>KIF22_R (#3851)</td>
<td>ACA CTA AAC TCC TTG TCC CTT TAC T</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FEM1B</td>
<td>FEM1B_F (#3852)</td>
<td>TAG GAG GAG GTT TTA GGA GTT AAG G</td>
<td>33</td>
<td>376</td>
<td>56.3°C</td>
</tr>
<tr>
<td></td>
<td>FEM1B_R (#3853)</td>
<td>AAC TAA AAA AAA CCT ACA AAC TCT C</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OGG1</td>
<td>OGG1_F (#3854)</td>
<td>AGG TTT TTG ATG TTG TGA GTT TTT T</td>
<td>26</td>
<td>405</td>
<td>58.3°C</td>
</tr>
<tr>
<td></td>
<td>OGG1_R (#3855)</td>
<td>ACA CTA AAC TCT CTA CCA ACT ACT C</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C20orf19</td>
<td>C20orf19_F2 (#3878)</td>
<td>TAA TAA ATT TTG GTA AGA AAG GAG AT</td>
<td>42</td>
<td>458</td>
<td>60°C</td>
</tr>
<tr>
<td></td>
<td>C20orf19_R2 (#879)</td>
<td>AAC TCA ACA CAA CCT CTA AAT AAC C</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AICDA</td>
<td>AICDA_F (#3858)</td>
<td>ATT GGA TTT AAA TTA GGA GTA AAG T</td>
<td>2</td>
<td>296</td>
<td>60°C</td>
</tr>
<tr>
<td></td>
<td>AICDA_R (#3859)</td>
<td>AAA AAA AAC CAA TAC AAC CAC CC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GLT25D1</td>
<td>GLT25D1_F (#3860)</td>
<td>GAG GAA ATT GAG GTT AAG GTT TTG T</td>
<td>17</td>
<td>299</td>
<td>58.3°C</td>
</tr>
<tr>
<td></td>
<td>GLT25D1_R (#3861)</td>
<td>CAC CTA CAA CCC TAC CCA AAA C</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MRPL36</td>
<td>MRPL36_F (#3862)</td>
<td>TTT TGG AGT TGA TAA ATT TTA G</td>
<td>22</td>
<td>419</td>
<td>52°C</td>
</tr>
<tr>
<td></td>
<td>MRPL36_R (#3863)</td>
<td>CAT ATT TTT AAA TAA ACA AAA AAA A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C6orf114</td>
<td>C6orf114_F (#3864)</td>
<td>TTT TTA ATT TTT TTG AGG TTT TTT TT</td>
<td>16</td>
<td>350</td>
<td>53.9°C</td>
</tr>
<tr>
<td></td>
<td>C6orf114_R (#3865)</td>
<td>ACT AAA AAA TAA CTT TCC TCT TCT C</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SGMS1</td>
<td>SGMS1_F (#3866)</td>
<td>TTA TTG GGT ATT AAT TTT AAG GTT</td>
<td>12</td>
<td>307</td>
<td>53.9°C</td>
</tr>
<tr>
<td></td>
<td>SGMS1_R (#3867)</td>
<td>TCA AAT AAT AAA AAA CTA AAA AAA C</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gene Name</td>
<td>Primer Name and Number</td>
<td>Sequence (5'→3')</td>
<td>Product CpGs</td>
<td>Product Length (bp)</td>
<td>Annealing Temp</td>
</tr>
<tr>
<td>-----------</td>
<td>------------------------</td>
<td>------------------</td>
<td>--------------</td>
<td>---------------------</td>
<td>---------------</td>
</tr>
<tr>
<td>KCNH5</td>
<td>KCNH5_F (#3868)</td>
<td>TTT GTT TTT TAG GTT GGA GTG TAG T</td>
<td>17</td>
<td>353</td>
<td>60°C</td>
</tr>
<tr>
<td></td>
<td>KCNH5_R (#3869)</td>
<td>AAT ATC TTT CAT TTA CTC AAA AAT T</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>COL23A1</td>
<td>COL23A1_F (#3870)</td>
<td>TTA GTT AGG GGT GGG TTG TAG AAG</td>
<td>30</td>
<td>338</td>
<td>60°C</td>
</tr>
<tr>
<td></td>
<td>COL23A1_R (#3871)</td>
<td>TAC AAA TAA ACC CTA CCA AAA AAA A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATAD4</td>
<td>ATAD4_F (#3872)</td>
<td>GTT AGG AGA GTG GGG TAG GAG TAG T</td>
<td>7</td>
<td>366</td>
<td>60°C</td>
</tr>
<tr>
<td></td>
<td>ATAD4_R (#3873)</td>
<td>ATA CAA CCA AAC ACC CAA AAA TAA A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ITGA2</td>
<td>ITGA2_F2 (#3880)</td>
<td>TTA GGG TAG GAA AGT TTG TTA GGG</td>
<td>24</td>
<td>283</td>
<td>60°C</td>
</tr>
<tr>
<td></td>
<td>ITGA2_R2 (#3881)</td>
<td>AAA ATT TCT AAA CAA CTC CTA CAA C</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LDB2</td>
<td>LDB2_F (#3876)</td>
<td>TTT TTT TGG AGT GGG TTG TTT TT</td>
<td>6</td>
<td>362</td>
<td>58.3°C</td>
</tr>
<tr>
<td></td>
<td>LDB2_R (#3877)</td>
<td>CAT AAA CTA TAT CCT TCC CAA TAC A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KCNH1</td>
<td>KCNH1_F2 (#3900)</td>
<td>GGG AGT TTT TTA GGT TGG TTT TT</td>
<td>43</td>
<td>386</td>
<td>59°C</td>
</tr>
<tr>
<td></td>
<td>KCNH1_R2 (#3928)</td>
<td>CAT AAT CAT CCT CCC AAC AAC TC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KCNH5</td>
<td>P2_KCNH5_F2 (#4012)</td>
<td>GGG TTG GAG TTG GAG TTT GTT</td>
<td>34</td>
<td>382</td>
<td>63.5°C</td>
</tr>
<tr>
<td>(Transcript 1b)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>P2_KCNH5_R2 (#4013)</td>
<td>TCT ACA TCT CTC CCC CTA CTA AAA A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KCNH5</td>
<td>P2_KCNH5_F2 (#4014)</td>
<td>TTT TTT AGT AGG GGG AGA GAT GTA GA</td>
<td>38</td>
<td>445</td>
<td>63.5°C</td>
</tr>
<tr>
<td>(Transcript 1b) PCR 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>P2_KCNH5_R2 (#4015)</td>
<td>ATA CCA CCA ACC CTC TCT TAC C</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ERVWE1</td>
<td>Syncytin_F external (#4018)</td>
<td>TTT GGA TAG TGA ATA TAG AT</td>
<td>7</td>
<td>352</td>
<td>54.3°C</td>
</tr>
<tr>
<td></td>
<td>Syncytin_R (#4020)</td>
<td>GGT AAT AGT AGT TTT TTT TT</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MID1</td>
<td>MID1_F2 (#4027)</td>
<td>AGA TTT GAA ATT AGG TTT GGG TTT G</td>
<td>6</td>
<td>298</td>
<td>63.5°C</td>
</tr>
<tr>
<td></td>
<td>MID1_R2 (#4028)</td>
<td>TAG GAG TTT TGT TGA TGT TTT TGG T</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EBR</td>
<td>EBR_F (#4029)</td>
<td>TTG GTG TTA TAT TTT GGG TTT GTT A</td>
<td>7</td>
<td>352</td>
<td>63.5°C</td>
</tr>
<tr>
<td></td>
<td>EBR_R (#4030)</td>
<td>GAG GAA TTT TGT TTG TAG TTT GTT TTG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PTN</td>
<td>PTN_F (#4033)</td>
<td>GGG AAA TAA GGG ATA AGA GAG ATT TT</td>
<td>8</td>
<td>446</td>
<td>63.5°C</td>
</tr>
<tr>
<td></td>
<td>PTN_R (#4034)</td>
<td>AGG TTG TTT AGT TTT TTA GTT GAT T</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>INSL4</td>
<td>INSL4_F2 (#4023)</td>
<td>TAT TTT TTT TGG TGG TAG TGTG G</td>
<td>4</td>
<td>438</td>
<td>62.4°C</td>
</tr>
<tr>
<td></td>
<td>INSL4_R2 (#4024)</td>
<td>AAA GGG GTG GAA TTT ATA TTA TTA A</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
2.6.4.2 Sequenom PCR Reaction Conditions

Each PCR reaction mix was made up in double volume (10 µL) so that 5 µL could be tested prior to use in Sequenom. Sequenom PCRs were set up as follows:

<table>
<thead>
<tr>
<th>Component</th>
<th>Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>10 x Hot Star Buffer (Qiagen, Cat. #203205)</td>
<td>1.00 µL</td>
</tr>
<tr>
<td>dNTPs (25 mM)</td>
<td>0.08 µL</td>
</tr>
<tr>
<td>Hot Star Taq (Qiagen, Cat. #203205)</td>
<td>0.08 µL</td>
</tr>
<tr>
<td>Forward Primer (1 µM)</td>
<td>2.00 µL</td>
</tr>
<tr>
<td>Reverse Primer (1 µM)</td>
<td>2.00 µL</td>
</tr>
<tr>
<td>ddH2O</td>
<td>2.84 µL</td>
</tr>
<tr>
<td>Bisulfite-converted DNA</td>
<td>2.00 µL</td>
</tr>
</tbody>
</table>

* For a negative control, 1 µL ddH2O was used instead of DNA.

Sequenom PCRs were performed using a C1000™ Thermal Cycler (BioRad) with a ramp rate of 2.0°C/sec. The cycling conditions were as follows:

- **Enzyme activation:** 94°C for 15 min
- **45 cycles of:**
  - 94°C for 20 sec (Optimal annealing temperature)°C for 30 sec
  - 72°C for 1 min
- **Final extension:** 72°C for 3 min
- **Hold:** 4 °C until use

After PCR, 5 µL of product was loaded onto a 2% agarose gel stained with ethidium bromide and electrophoresed at 100 V for 45 min to confirm successful and uncontaminated PCR amplification. The remaining 5 µL of PCR product was stored at 4°C.

2.6.5 Processing of Sequenom Reactions

Slight modifications were made to the following protocols from the EpiTYPER® Application Guide (Sequenom, version 1, March 2006). All reagents used in the protocols were contained within the T-Cleavage MassCLEAVE™ Reagent Kit (Sequenom, Cat. #10129.1).

2.6.5.1 Treatment with Shrimp Alkaline Phosphatase

Shrimp alkaline phosphatase (SAP) cleaves the phosphates of unincorporated dNTPs, so this step removed all unincorporated dNTPs from the PCR products. Before SAP treatment, PCR products were removed from 4 °C, thawed at room temperature and a 12-channel pipette was used to transfer 5 µL of PCR product into a 384-well plate (ABgene, Cat. #AB-1384). The SAP enzyme solution was then prepared on ice for each reaction by diluting 0.3 µL SAP solution with 1.7 µL RNase-free ddH2O. The SAP solution was thoroughly vortexed and
Materials and Methods

60 centrifuged for 10 sec at 100 x g. Next, 2 µL of the SAP solution was added to each reaction in the 384-well plate. The plate was sealed and centrifuged for 1 min at 180 x g. The plate was then incubated in an ABI PRISM® 7900HT (Applied Biosystems) thermal cycler as follows:

<table>
<thead>
<tr>
<th>Enzyme Activity:</th>
<th>37°C for 20 min</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enzyme Deactivation:</td>
<td>85°C for 5 min</td>
</tr>
<tr>
<td>Hold:</td>
<td>4°C until use</td>
</tr>
</tbody>
</table>

2.6.5.2 T-Cleavage and RNA Transcription

During the SAP incubation, the hMC reaction cocktail was prepared in a 15 mL conical centrifuge tube in the order of the following reagents:

<table>
<thead>
<tr>
<th>RNase-free ddH2O</th>
<th>3.21 µL</th>
</tr>
</thead>
<tbody>
<tr>
<td>5 x T7 Polymerase Buffer</td>
<td>0.89 µL</td>
</tr>
<tr>
<td>T-Cleavage Mix</td>
<td>0.22 µL</td>
</tr>
<tr>
<td>DTT (100 mM)</td>
<td>0.22 µL</td>
</tr>
<tr>
<td>T7 RNA &amp; DNA Polymerase</td>
<td>0.40 µL</td>
</tr>
<tr>
<td>RNase A</td>
<td>0.06 µL</td>
</tr>
</tbody>
</table>

The hMC cocktail was vortexed well. A Matrix 12-Channel Electronic Pipette (ThermoScientific, Cat. #2019) was used to transfer 5 µL of the hMC cocktail to each well of a new 384-well plate. Once the SAP incubation was complete, the PCR/SAP plate was removed from the thermal cycler and the sealing film was carefully removed. The 12-channel pipette was then used to transfer 2 µL of the PCR/SAP mix to each well of the new 384-well plate containing 5 µL of the hMC cocktail. The plate was carefully sealed and centrifuged at 180 x g for 1 min and then incubated at 37°C for 3 h. Plates were stored at -20°C until processed.

2.6.5.3 Conditioning the Reaction Products and Spotting of the Microarrays

Plates containing the hMC reaction products were transferred on dry ice to a Sequenom facility for the remaining steps of the protocol to be performed. Two Sequenom facilities in New Zealand were used for this project: 1) AgResearch Centre for Reproduction and Genomics (CRG) at Invermay, Mosgiel, Dunedin, where the protocol was performed by Tim Manley and 2) National Research Centre for Growth and Reproduction at the Liggins Institute, Auckland, where the protocol was performed by Cameron McLean.
Upon arrival at the Sequenom facility, the plate was immediately put in -20°C while the Clean Resin plate was prepared. Clean Resin was spread onto the 384-well Clean Resin plate and excess resin was removed. The plate was left to stand at room temperature for at least 20 min. During this time, the plate containing the hMC reaction products was removed from -20°C, thawed to room temperature and centrifuged for 1 min at 180 x g. The plate was then placed on the liquid handler deck, the sealing film was removed and the program “hMC Water Addition” was run using the liquid handler controller PC computer. This method added 20 µL of nano-pure water to each well of the plate. When the water addition was finished, the plate was re-sealed and centrifuged for 1 min at 180 x g. The sealing film was then removed and plate was placed upside-down onto the Clean Resin plate. The plates were firmly held together and flipped over so that the Clean Resin plate was on top. The Clean Resin plate was tapped to help the Clean Resin fall into each well of the hMC reaction products plate. The plates were then separated and examined to make sure that there was no Clean Resin left on the Clean Resin plate, and that each well of the hMC reaction products plate contained Clean Resin.

The hMC reaction products plate was rotated for 10 min at room temperature. The plate was then centrifuged for 5 min at 3200 x g, and the hMC reaction products were transferred to a SpectroCHIP (Sequenom, Cat. #10117) using the MassARRAY® Nanodispenser. SpectroCHIPs were then placed into the Sequenom Mass-Spectrometer. The hMC reaction products plate layout was defined using the EpiTYPER® software. Once the plate layout was defined, the SpectroCHIPs were processed and data was collected.

2.6.6  **Sequenom Data Analysis**

The data was analysed using the EpiTYPER® software. The data contained the percentage of methylation at individual CpGs sites within each amplicon. A visual diagram (Epigram) was also generated that used colour-coded circles to display the percentage methylation at CpG sites. The EpiTYPER® software uses a formula to determine the percentage methylation of a fragment within an amplicon based on peak heights. The formula calculates a ratio comparing the methylated to unmethylated CpG sites within a fragment as shown in Equation 2.1.
Materials and Methods

**Equation 2.1.** The formula used by the Sequenom MassARRAY® EpiTYPER® software to determine the percentage methylation of a fragment with \( n \) CpG sites (Coolen *et al.*, 2007). The signal to noise ratio is represented by the methylated Peak (Peak\(_{i}\), Peak\(_{2}\), etc) and the unmethylated Peak (Peak\(_{\text{unmethylated}}\)).

\[
\text{Methylation of a fragment with } n \text{ CpG sites} = \frac{\text{Peak}_{1} + \text{Peak}_{2} + \ldots \ldots \text{Peak}_{n}}{\text{Peak}_{\text{unmethylated}} + \text{Peak}_{1} + \text{Peak}_{2} + \ldots \ldots \text{Peak}_{n}}
\]

The EpiTYPER® (.csv) file containing the calculated percentage methylation data was imported into Microsoft Excel. Before methylation values were accepted, data was fitted against specific criteria that were suggested by Christine Cauldrey, a frequent user of Sequenom for methylation analysis (AgResearch Ruakura, Hamilton). First, the spectra quality (peak height and confidence score) for each reaction was examined using the EpiTYPER® program. The Sequenom assay is set to detect a minimum cleaved fragment size of 1500 Daltons (Da) (roughly 4 bases) and a maximum cleaved fragment size of 9000 Da (roughly 27 bases). It has been reported that fragments with a mass of 1700 Da or less are more likely to match with other small fragments and lead to inaccurate peak calls (Coolen *et al.*, 2007), so fragments with a mass of 1700 Da or less were removed from analysis in this study. Next, CpG sites were discarded if they gave no data in greater than thirty percent of the samples. Finally, individual samples were removed if they had no data at thirty percent or more of the CpG sites within the amplicon.

Following filtering to exclude poor quality data, the methylation percentages at each CpG site were averaged using Microsoft Excel to generate the percentage methylation for the amplicon. Although an R-script was created by Coolen and colleagues (Coolen *et al.*, 2007) to calculate average amplicon methylation, the algorithm was not used in this study since a manual calculation allowed the removal of poor quality data by using the filtering criteria explained above (Christine Cauldrey, AgResearch, Ruakura, Hamilton, *personal communication*). The standard deviation, sample count and 95% confidence interval for the mean percentage methylation for each group of tissues was then calculated using Microsoft Excel.
2.7 Gene Expression Analysis by Reverse-Transcription PCR

Quantitative TaqMan® Gene Expression Assays (Applied Biosystems, Cat #4331182) were used to determine if gene expression correlated with promoter methylation data from Sequenom. Four candidate genes were chosen for expression analysis based on their contrasting methylation patterns in first-trimester placental tissue and peripheral blood: AKR1C4, KCNH5, AICDA, and ATAD4. Two housekeeping genes were included as reference (control) genes: B2M and TOP3A. The TaqMan® Gene Expression Assays used in this experiment are listed in Table 2.8.

Table 2.6. TaqMan® gene expression assays used for gene expression analysis by quantitative real-time PCR.

<table>
<thead>
<tr>
<th>Gene Name</th>
<th>TaqMan® Assay ID</th>
<th>TaqMan® Probe Location (Genome Browser Assembly NCBI36/hg18)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AKR1C4</td>
<td>Hs00559542_m1</td>
<td>Chr. 10: 5244957 - 5248756</td>
</tr>
<tr>
<td>KCNH5</td>
<td>Hs00544949_m1</td>
<td>Chr. 14: 62542837 - 62553425</td>
</tr>
<tr>
<td>AICDA</td>
<td>Hs00757808_m1</td>
<td>Chr. 12: 8649078 - 8650875</td>
</tr>
<tr>
<td>ATAD4</td>
<td>Hs00256987_m1</td>
<td>Chr. 17: 43384332 - 43390101</td>
</tr>
<tr>
<td>B2M</td>
<td>Hs00984320_m1</td>
<td>Chr. 15: 42795819 - 42797649</td>
</tr>
<tr>
<td>TOP3A</td>
<td>Hs00172806_m1</td>
<td>Chr. 17: 18129446 - 18134725</td>
</tr>
<tr>
<td>KCNH1</td>
<td>Hs00608142_m1</td>
<td>Chr. 1: 208918280 - 209015509</td>
</tr>
</tbody>
</table>

2.7.1 RNA Extraction from Tissues

RNA was extracted from frozen human tissue samples using the PureLink™ Micro-to-Midi Total RNA Purification System (Invitrogen, Cat #12183-018). The protocol “Purifying RNA from Animal Tissues” was followed and all reagents were contained in the kit unless stated otherwise. RNA Lysis Solution was freshly prepared by adding 1% 2-mercaptoethanol to the required volume of RNA Lysis Solution. A 10-60 mg piece of frozen tissue was transferred to a 1.5 mL tube (on ice) and 0.6 mL of the prepared RNA Lysis Solution (containing 1% 2-mercaptoethanol) was added. Tissue was minced using an RNase-free pestle and then centrifuged at 12000 x g for 2 min at room temperature. The supernatant was then transferred to a 1.5 mL tube and the lysate was passed through a 1 mL syringe with a 20-gauge needle 10 times. The sample was centrifuged at 12000 x g for 2 min and the supernatant was transferred to a clean tube. After addition of 70% ethanol, RNA was purified using a spin column and eluted in 60 uL RNase-free ddH2O. RNA concentration was determined by measuring 1.5 µL on a NanoDrop ND_1000 Spectrophotometer (NanoDrop Technologies). The 60 µL RNA sample was separated into 15 µL aliquots which were stored at -80 °C.
2.7.2 Reverse-Transcription Reaction

Complementary DNA (cDNA) was made from RNA using the protocol and reagents from the High Capacity RNA-to-cDNA Kit (Applied Biosystems, Cat. #4387406). Approximately 2 µg total RNA in up to 9 µL RNase-free ddH$_2$O was used in each reverse-transcription (RT) reaction. Each 20 µL RT reaction was set up as follows:

<table>
<thead>
<tr>
<th>Comp.</th>
<th>Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 x RT Buffer</td>
<td>10 µL</td>
</tr>
<tr>
<td>20 x Enzyme Mix</td>
<td>1 µL</td>
</tr>
<tr>
<td>RNA Sample</td>
<td>Up to 9 µL (2 µg RNA)</td>
</tr>
<tr>
<td>RNase-free ddH$_2$O</td>
<td>Make up to final volume of 20 µL</td>
</tr>
</tbody>
</table>

* For -RT reaction, replace 20 x Enzyme Mix with 1 µL RNase-free ddH$_2$O

Samples with low RNA concentrations (less than 222 ng/µL) were limited by the protocol’s maximum RNA input of 9 µL and therefore had less than 2 µg of RNA in their RT reaction. Reactions were incubated in a thermal cycler using the following conditions:

- Enzyme Activity: 37°C for 60 min
- Enzyme Inactivation: 95°C for 5 min
- Hold: 4°C until use

After the RT reaction, 180 µL of ddH$_2$O was added to each sample to increase the volume of available cDNA for use in quantitative reverse transcriptase PCR (RT-PCR).

2.7.2.1 Testing cDNA before use in RT-PCR

Before use in RT-PCR, each cDNA sample was tested by end-point PCR to validate both reactions with and without reverse transcriptase. A previously designed PCR for the B2M gene (designed by Rob Weeks, Senior Research Fellow in Morison lab) was used to test cDNA quality. Primer sequences for the B2M PCR are listed in Table 2.9.

<table>
<thead>
<tr>
<th>Lab #</th>
<th>Primer Name</th>
<th>Primer Sequence</th>
<th>Product Length (bp)</th>
<th>Annealing Temperature</th>
</tr>
</thead>
<tbody>
<tr>
<td>3524</td>
<td>B2mFor</td>
<td>GAG TAT GCC TGC CGT GTG</td>
<td>109</td>
<td>60 °C</td>
</tr>
<tr>
<td>3525</td>
<td>B2mRev</td>
<td>AAT CCA AAT GCG GCA TCT</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The 20 µL B2M PCR for each sample was set up as follows:

- GeneAmp® 10x PCR Buffer II (ABI, Cat. #N808-0249) 2.0 µL
- MgCl$_2$ (25 mM) 1.2 µL
- Forward Primer (10 µM) 1.0 µL
- Reverse Primer (10 µM) 1.0 µL
Materials and Methods

Materials and Methods

65
dNTPs (2.5 mM) 1.0 µL
dH2O 11.6 µL
AmpliTaq Gold® DNA Polymerase (ABI, Cat. #/N808-0249) 0.2 µL
Diluted cDNA (0.5 µL cDNA + 1.5 µL ddH2O) 2.0 µL
* For a negative control, 2 µL ddH2O was used instead of cDNA.

PCR was performed using a C1000™ Thermal Cycler (BioRad) using the following cycling conditions:

Enzyme activation: 95°C for 10 min
35 cycles of:
   95°C for 30 sec
   60°C for 30 sec
   72°C for 30 sec
Final extension: 72°C for 7 min
Hold: 4°C until use

PCR product was visualized on a 2% agarose gel stained with ethidium bromide.

2.7.2.2 Making cDNA Standards for Quantitative RT-PCR

To determine assay efficiency, each RT-PCR required a set of cDNA standards to produce a standard curve. The cDNA standards needed to be made from a tissue that had moderate to high levels of gene expression, and most importantly the standards needed to produce threshold cycle (Ct) values that covered the range of Ct values in the assay. Since gene expression varied across tissue types, a range of tissues was required to make the appropriate cDNA standards for each assay (Table 2.8).

<table>
<thead>
<tr>
<th>Gene Name</th>
<th>TaqMan® Assay ID</th>
<th>Sample used for cDNA standards</th>
</tr>
</thead>
<tbody>
<tr>
<td>AKR1C4</td>
<td>Hs00559542_m1</td>
<td>Fetal liver (F129)</td>
</tr>
<tr>
<td>KCNH5</td>
<td>Hs00544949_m1</td>
<td>Fetal brain (F96)</td>
</tr>
<tr>
<td>AICDA</td>
<td>Hs00757808_m1</td>
<td>Fetal liver (F129)</td>
</tr>
<tr>
<td>ATAD4</td>
<td>Hs00256987_m1</td>
<td>Fetal liver (F129)</td>
</tr>
<tr>
<td>B2M</td>
<td>Hs00984320_m1</td>
<td>Peripheral blood (x3236)</td>
</tr>
<tr>
<td>TOP3A</td>
<td>Hs00172806_m1</td>
<td>Peripheral blood (x3236)</td>
</tr>
<tr>
<td>KCNH1</td>
<td>Hs00608142_m1</td>
<td>Fetal brain (F96)</td>
</tr>
</tbody>
</table>

Table 2.8. Samples used for cDNA standards in quantitative RT-PCR.

Standards contained the following amounts of cDNA: 45 ng, 22.5 ng, 11.25 ng, 5.63 ng, 2.81 ng, 1.41 ng and 0.70 ng and 0.35 ng.
2.7.3 Quantitative RT-PCR using TaqMan® Gene Expression Assays

RT-PCR was performed with slight modifications to the manufacturer’s protocol. The cDNA samples and PCR mix were dispensed using the CAS-1200™ automated PCR setup robot (Corbett Robotics) into a 384-well Clear Optical Reaction Plate (Applied Biosystems, Cat #4309849). Reactions were performed in triplicate for each sample. Each 10 µL reaction was set up as follows:

- ABsolute™ QPCR ROX Mix (ABgene, Cat #CM-205) 5.5 µL
- 20 x TaqMan® Gene Expression Assay Mix 0.5 µL
- cDNA Sample 4.5 µL

RT-PCR was performed on an ABI PRISM® 7900HT (Applied Biosystems) instrument using a 384-well block. The SDS 2.1 software (Applied Biosystems) was used for quantification during the following RT-PCR cycling conditions:

- Enzyme activation: 95°C for 15 min
- 40 cycles of: 95°C for 15 sec
- 60°C for 60 sec

2.7.4 Calculation of Gene Expression

Data was collected from the ABI PRISM® 7900HT (Applied Biosystems) using the default parameter settings on the instrument. The threshold value was set at 0.2 for all RT-PCRs. The threshold cycle (Ct) values collected during the run represent the cycle number at which the fluorescent signal from the PCR product crossed the threshold value. Since the amount of template approximately doubles after each cycle of PCR, the Ct values are used to calculate the relative amount of template that was produced during the run. Data from the ABI PRISM® 7900HT (Applied Biosystems) instrument was imported into Microsoft Excel for gene expression analysis.

For each triplicate sample, the mean and standard deviation of the Ct values were calculated. The cDNA standards were used to make a standard curve by taking the log of the amount of known cDNA and plotting it against the Ct values. A linear regression line was then fitted to the data to determine the slope and y-intercept of the standard curve, which was used to calculate relative gene expression for every sample. The relative gene expression for each sample was then normalised to the amount of RNA that went into the sample’s RT reaction (Equation 2.2). The inconsistent expression of housekeeping genes across the various tissue
types in this study prevented them being used as reference genes in the normalisation of gene expression. Therefore, this method of “corrected gene expression” was the most accurate way to calculate gene expression for a variety of tissues (Thellin et al., 1999).

\[
\log_{2}^{2}(\text{cDNA}) = \frac{(\text{Mean } C_\text{T} - \text{Y-intercept})}{\text{Slope}}
\]

Relative amount of cDNA expressed (ng) = \(2^{\text{[DNA]}}\)

Corrected expression = \(\frac{\text{Relative amount of cDNA expressed (ng)}}{\text{Amount of RNA into RT reaction (ng)}}\)

**Equation 2.2.** The equations used to calculate gene expression from quantitative RT-PCR data.

### 2.8 Transcript-Specific RT-PCRs for KCNH5

The University of California Santa Cruz (UCSC) Genome Browser (Kent et al., 2002) was used to identify alternative transcripts of KCNH5. Two human mRNA sequences from GenBank were extracted for analysis: BC043409 and AF493798. For simplicity, these transcripts were renamed transcript 1a and transcript 1b, respectively.

#### 2.8.1 Primer Design for Transcript Specific RT-PCRs

To distinguish which tissues express each transcript, two end-point RT-PCRs were designed for cDNA. Forward primers were designed within the unique exon 1 of each transcript and a common reverse primer was designed in exon 4. Primer design for transcript-specific RT-PCR is displayed in Chapter Seven (Figure 7.2).

#### 2.8.2 Optimisation of Transcript-Specific RT-PCRs

Samples that expressed KCNH5 were used to optimise the transcript-specific RT-PCRs. Since it was hypothesized that the transcripts may be expressed in a lineage-specific manner, a tissue from both lineages was chosen to optimise the RT-PCRs. Primer sets were optimised for annealing temperature, magnesium chloride concentration and primer concentration. DMSO was added to the RT-PCR for transcript 1b. The primer sequences and optimised RT-PCR conditions are listed in Table 2.9.
Table 2.9. Primer Sequences and RT-PCR Conditions for KCNH5’s Transcript-Specific RT-PCRs.

<table>
<thead>
<tr>
<th>Transcript Name</th>
<th>Lab #</th>
<th>Primer Name</th>
<th>Primer Sequence</th>
<th>Product Length (bp)</th>
<th>Annealing Temp</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transcript 1a</td>
<td>4008</td>
<td>kcnh5 BC043409_F</td>
<td>CTGGGATTACAGGGC TGA GC</td>
<td>422</td>
<td>64.2°C</td>
</tr>
<tr>
<td></td>
<td>4006</td>
<td>kcnh5 common_R</td>
<td>AGTACACAGGAAACA AGACCAC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transcript 1b</td>
<td>4007</td>
<td>kcnh5 AF493798_F</td>
<td>AAGAAGGGCTGT GGCACC</td>
<td>368</td>
<td>59°C</td>
</tr>
<tr>
<td></td>
<td>4006</td>
<td>kcnh5 common_R</td>
<td>AGTACACAGGAAACA AGACCAC</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

2.8.3 Transcript-Specific RT-PCRs

The standard 20 µL reaction volume was halved because of the limited volume of cDNA. Each 10 µL reaction was set up as follows:

- GeneAmp® 10x PCR Buffer II (ABI, Cat. #N808-0249): 1.0 µL
- MgCl₂ (25 mM): 0.6 µL
- Forward Primer (10 µM): 0.5 µL
- Reverse Primer (10 µM):
  - Transcript 1a: 1.0 µL
  - Transcript 1b: 0.5 µL
- DMSO (Transcript 1b only): 0.5 µL
- dNTPs (2.5 mM): 0.5 µL
- ddH₂O: 5.3 µL
- AmpliTaq Gold® DNA Polymerase (ABI, Cat.#/N808-0249): 0.1 µL
- cDNA: 1.0 µL

* For a negative control, 1 µL ddH₂O was used instead of cDNA.

RT-PCR was performed using a C1000™ Thermal Cycler (BioRad) using the standard PCR cycling conditions (pg. 37) but for 40 cycles. PCR products were visualized on a 2% agarose gel stained with ethidium bromide.
2.9 Western Blot Analysis for KCNH5 Protein

Western blot analysis was used to investigate whether the alternative transcripts of KCNH5 produced two different proteins. This method uses polyacrylamide gel electrophoresis (PAGE) to separate denatured proteins based on size. After electrophoresis, proteins are transferred to a nitrocellulose membrane and the protein of interest is detected on the membrane by labelled antibody and a chemiluminescent detection system.

2.9.1 Protein Extraction from Samples

2.9.1.1 Sample Selection

Tissues expressing high levels of KCNH5 were selected for use in the western blot. These tissues were chosen based on expression data from the TaqMan® Gene Expression Assay, which amplified exons 2 and 3 that are common to both transcripts of KCNH5. Positive and negative control tissues were also selected. The AbCam antibody company recommended HepG2 cell lysate as a positive control for KCNH5. The negative control tissue, adult primary melanocytes, was chosen after screening multiple cell lines and tissues by western blot. Expression of KCNH5 was confirmed in all samples by performing RT-PCR using the TaqMan® Gene Expression Assays for KCNH5 and B2M. Gene expression was normalised to input RNA, and RT-PCR products were also visualised on a 2% agarose gel.

2.9.1.2 Tissue/Cell Lysis and Total Protein Isolation

The following protocol for tissue/cell lysis and total protein isolation was provided by Grace Li (Developmental Genetics Group, University of Otago). Tissue samples and cell pellets were removed from -80°C and kept on dry ice. Individual 1.5 mL tubes were chilled on ice. The RIPA buffer* had been prepared previously (without protease inhibitors) and was stored at 4°C.

The cell lysis buffer was then prepared on ice as follows:

<table>
<thead>
<tr>
<th>Component</th>
<th>Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>20X Complete protease inhibitor</td>
<td>50 µL</td>
</tr>
<tr>
<td>0.1 M PMSF</td>
<td>10µL</td>
</tr>
<tr>
<td>100 uM NaOrthovanodate</td>
<td>10µL</td>
</tr>
<tr>
<td>RIPA buffer</td>
<td>930µL</td>
</tr>
</tbody>
</table>
Approximately 10 mg of each frozen tissue sample was placed in a chilled 1.7 mL centrifuge tube. Cell pellets (of approximately 1x10^6 cells) were placed on ice. Next, 150 µL cell lysis buffer was added to each sample. A plastic pestle was used to break up tissue samples and a pipette was used to break up cell pellets. Additional cell lysis buffer was added to the sample if required. Tissue lysis was considered to be complete when no pieces of tissue or cell pellet were visible in solution.

Samples were incubated on ice for 30 min, and then centrifuged at 16,000 x g for 20 min at 4°C. Samples were placed on ice and 4 µL of supernatant was transferred into a 0.6 mL tube for subsequent use in the bicinchoninic acid (BCA) assay. The remaining supernatant was transferred into a 1.5 mL tube. All protein lysates were stored at -80°C until use.

* Radioimmunoprecipitation Assay (RIPA) buffer
0.5 mL 1 M Tris.Cl pH 8.0
1.5 mL 1 M NaCl
1 mL 10% NP-40
50 mg Sodium deoxycholate
100 µL 10% SDS
6.9 mL ddH₂O

2.9.1.3 Determining Total Protein Concentration using the Bicinchoninic Acid (BCA) Assay

The protein concentration of each sample was determined in order to load the same amount of protein for the western blot. Protein concentrations were determined by including a set of protein standards in the assay that were used to make a standard curve. Protein standards ranging from 0 to 1 mg/mL were made from BSA. Protein standards were loaded into a 96-well plate in triplicates. The 4 µL of protein lysate from each sample was removed from -80°C and left to thaw on ice. Each sample was diluted 1:10 by adding 36 µL ddH₂O, giving a final volume of 40 µL. After vortexing, 20 µL samples were loaded into a 96-well plate in duplicate.

The protein concentration of each sample was determined in a 96-well format by using the BCA™ Protein Assay Kit (Thermo Scientific, Cat #23227) and following the manufacturer’s protocol. Absorbance values measured at 540 nm were entered into Microsoft Excel and values for replicate samples were averaged to determine the mean. A standard curve was generated from the BSA standards by taking the mean standard absorbance value and plotting it against the known protein concentration. A linear regression line was fitted to the data to determine the slope and y-intercept of the standard curve, which was then used to determine
the total protein concentration for each sample, which is shown in Equation 2.3. Protein concentrations were corrected for the 1:10 assay dilution (Equation 2.3).

| Total Protein Concentration (µg/µL) = (Mean absorbance - Y-intercept) / Slope |
|--------------------------------------|---------------------------------------------|
| Corrected Total Protein Concentration (µg/µL) = Total Protein Concentration (µg/µL) * 10 |

Equation 2.3. The equations used to calculate total protein concentration in samples for the western blot.

2.9.2 SDS Polyacrylamide Gel Electrophoresis for Western Blot Analysis

The Mini-PROTEAN® III Electrophoresis Cell System (BioRad, Cat #165-3301) was used to separate proteins based on molecular weight.

2.9.2.1 Pouring the SDS-Polyacrylamide Gels

An 8% Bis-Tris Polyacrylamide Separating Gel* solution was prepared. A 1 mL pipette was used to pour the separating gel mixture between glass gel plates. Care was taken to ensure no bubbles were present in the gel. Once poured, ddH₂O was added on top of the gel and the gel was left to set for 1 h. The ddH₂O was then poured off and a piece of filter paper was used to absorb any residual ddH₂O in between the glass plates. A 5% Bis-Tris Polyacrylamide Stacking Gel** solution was prepared and was poured on top of a pre-set 8% separating gel. A 10-well comb was placed into the stacking gel and the gel was left to set for 1.5 h.

* 8% Bis-Tris Polyacrylamide Separating Gel
2.477 mL ddH₂O
1.16 mL 40% Acrylamide (BioRad, Cat #161-0140)
0.8 mL 2% Bis Solution (BioRad, Cat #161-0142)
1.5 mL 1.5 M Tris (pH 8.8)
60 µL 10% Ammonium persulfate (APS) (Sigma Aldrich, Cat #A3678)
3 µL Temed (Invitrogen UltraPure™ Temed, Cat #15524-010)
Reagents were added in the order listed.

** 5% Bis-Tris Polyacrylamide Stacking Gel
1.049 mL ddH₂O
0.242 mL 40% Acrylamide (BioRad, Cat #161-0140)
167 µL 2% Bis Solution (BioRad, Cat #161-0142)
0.5 mL 1.5 M Tris (pH 8.8)
20 µL 10% Ammonium persulfate (APS) (Sigma Aldrich, Cat #A3678)
2.5 µL Temed (Invitrogen UltraPure™ Temed, Cat #15524-010)
Reagents were added in the order listed.
2.9.2.2 Sample Preparation for use in SDS-Polyacrylamide Gel Electrophoresis

Protein samples were prepared so that 40 µg of total protein was contained in 18 µL ddH2O. Next, 6 µL of 4x protein sample buffer* was added to each sample and then incubated (denatured) at 99°C for 10 min in an Eppendorf Thermomixer gently shaking at 300 rpm. After incubation, samples were stored at -80°C until used.

* 4x Protein Sample Buffer
0.24 M Tris-HCl (pH 6.8)  
0.24 M SDS  
40% Glycerol  
20% B-mercaptoethanol  
0.02% Bromophenol blue  
Make up in ddH2O and store at -20°C.

2.9.2.3 Loading and Running SDS-Polyacrylamide Gels

A 10 x Tris-Glycine* (TG) solution was prepared for the wet blot transfer. A 1 x SDS running buffer** (0.1% SDS) solution was then prepared and poured between gel plates and the 40 µg protein samples were loaded into wells. Two protein markers were loaded at either side of each gel: 1 µL of MagicMark™ XP Western Protein Standard (Invitrogen, Cat #LC5602) and 10 µL of PageRuler™ Prestained Protein Ladder (Fermentas Life Sciences, Cat #SM0671). Gels were run at a constant 100 V for 1.5 h or until the loading dye front reached the bottom of the gel. After electrophoresis, plates were carefully separated. The separating gel was rinsed with ddH2O and placed into a clean plastic dish for subsequent staining with coomassie blue.

* 10 x Tris Glycine (TG)
30.3 g Tris  
144 g Glycine  
Add to 900 mL of ddH2O and dissolve on a hot plate set to 50°C using a magnetic stirrer.

** 1 x SDS Running Buffer
100 mL 10 x TG buffer  
10 mL 10% SDS  
Top up to 1 L with ddH2O and store at room temperature.

2.9.2.4 Coomassie Staining to Visualise Total Protein

Before proceeding with the transfer step, gels were stained with coomassie blue to visualise the protein and to check for equal loading across samples. Gels were placed into a 0.1%
coomassie blue staining solution* for 2 h on a shaking platform, and then into a coomassie destaining solution** for 1 h on a shaking platform.

* 0.1% Coomassie Blue Staining Solution
0.25 g Coomassie Brilliant Blue R-250 (BioRad, Cat #M2839)
125 mL ddH₂O
100 mL Methanol (40%)
25 mL Acetic Acid (10%)

** Coomassie Blue Destaining Solution
225 mL ddH₂O
225 mL Methanol (45%)
50 mL Acetic Acid (10%)

2.9.3 Wet Blot Transfer of Protein to Membrane

During electrophoresis, a 1 x Wet Blot Transfer buffer* was prepared and chilled at -20°C. The gel was rinsed with ddH₂O and a transfer cassette was placed into a plastic container filled with cold 1 x Wet Blot Transfer buffer. Sponges and filter paper were soaked in cold 1 x Wet Blot Transfer buffer. In the transfer cassette, consecutive layers were placed as follows: sponge, two pieces of soaked filter paper, gel, Hybond-C Extra Nitrocellulose membrane (Amersham Biosciences, Cat #RPN303E), two pieces of soaked filter paper and a sponge. After removal of bubbles, the cassette was closed and placed in the transfer gel tank. Cold 1 x Wet Blot Transfer buffer submerged the cassette and an ice pack was placed in the gel tank. The transfer apparatus was placed on a magnetic block and a small magnetic stir bar was put into the tank. The transfer was performed at a constant 225 mA for 2.5 h.

* 1 x Wet Blot Transfer Buffer
100 mL 10 x TG buffer
200 mL Methanol
1 mL 10% SDS
Top up to 1 L with ddH₂O. (Note: Before adding methanol, 300 mL ddH₂O was added to the 10 x TG buffer to prevent the solution from getting too warm.)

2.9.4 Blocking of Endogenous Protein

A 1 x Phosphate Buffered Saline Tween* (PBST) solution was prepared. Following the transfer, the membrane was removed from the cassette and blocked in a 2.5% non-fat milk blocking solution** for 1 h on a shaking platform at 65 rpm.
**1 x Phosphate Buffered Saline Tween (PBST)**
10 tablets 1 x PBS (pH 7.3) (Oxoid, Cat # BR0014G)
5 mL 20% Tween® 20 (0.1%) (Sigma-Aldrich, Cat #P2287)
Top up to 1 L with ddH₂O and dissolve using a magnetic stir bar.

**2.5% Non-fat Milk Blocking Solution**
0.625 g non-fat milk powder
25 mL 1 x PBST

### 2.9.5 Primary Antibody Incubation

The primary antibody solution* was prepared. A rabbit polyclonal antibody raised against a synthetic peptide derived from amino acids #156-205 of KCNH5 (Abcam, Cat #ab32975) was used to localise KCNH5 protein. The membrane was washed once with 1 x PBST and put into a 1:1000 dilution of KCNH5 antibody in a 50 mL conical centrifuge tube. The membrane was incubated overnight on a rotator at 4°C.

*Primary Antibody Solution*
625 µL Fetal Calf Serum (Bio International Ltd, NZ)
24.375 mL 1 x PBST
0.025 g Sodium azide (BDH, Cat #30111)

### 2.9.6 Secondary Antibody Incubation

The membrane was washed three times in 1 x PBST (5 min per wash) on a shaking platform at 65 rpm. The membrane was then put into a 50 mL conical centrifuge tube containing a 1:2000 secondary antibody dilution of horseradish-peroxidase (HRP) labeled anti-rabbit IgG (Sigma Aldrich, Cat #A0545) in 2.5% non-fat milk/PBST. The conical centrifuge tube was incubated at room temperature on a rotator for 1.5 h.

### 2.9.7 Membrane Washing and Labeling

After washing the membrane four times in 1 x PBST (5 mins per wash) on a shaking platform at 65 rpm, a labeling solution was prepared using SuperSignal West Pico Chemiluminescent Substrate (Thermo Scientific, Cat #34080) according to the manufacturer’s protocol. A film cassette, two plastic sheets and a plastic dish were cleaned with 70% ethanol. The membrane was removed from 1 x PBST, drained to remove residual liquid and laid in the plastic dish. The labeling solution was added to the membrane using a pipette so that the entire membrane was covered with labeling solution. The membrane was incubated for 5 min and then the labeling solution was drained off. The membrane was sealed between the clear plastic sheets
and any air bubbles were gently pushed out. The sealed membrane was placed into the cassette and brought to the dark room. Blots were exposed to Kodak X-OMAT BT Film for 1 min, 5 min and 1 h.

2.9.8 Beta-actin Protein Loading Control

To remove previous KCNH5 protein labeling, blots were rinsed twice in ddH₂O (15 min per wash) on a shaking platform at 65 rpm. A 1:10,000 dilution of Beta-actin antibody (AbCam, Cat #ab6276) was prepared in 2.5% FCS/PBST. The primary antibody was incubated overnight at 4°C on a rotator. All of the post-primary antibody steps explained above were followed. A 1:5000 secondary antibody solution of HRP-labeled anti-mouse IgG (Sigma Aldrich, Cat #A9917) was prepared in 2.5% non-fat milk/PBST. Membranes detecting Beta-actin protein were exposed to film for only 10 sec.

2.10 Immunohistochemistry for KCNH5 Protein in Placenta

Immunohistochemistry was performed by Mandy Fischer in the University of Otago’s Histology Unit to localise KCNH5 protein in placental tissue. The EnVision™ Dual Link (Dako, Cat #K4065) detection system and 3,3’-Diaminobenzidine (DAB) substrate was used (Figure 2.4). This indirect yet highly sensitive method involves the binding of a primary antibody to the antigen of interest, followed by a labelled secondary antibody binding to the primary antibody. The secondary antibody is coupled with the streptavidin-horseradish peroxidase enzyme, which reacts with the DAB substrate and produces a brown-coloured stain when the secondary antibody attaches to the primary antibody.

**Figure 2.4.** The EnVision™ Dual Link detection system using DAB substrate for immunohistochemistry on placental tissue to localise KCNH5 protein. The primary antibody binds to the antigen of interest. The secondary antibody, labelled with a streptavidin-horseradish peroxidase enzyme, binds to the primary antibody and a reaction between horseradish-peroxidase and DAB substrate produces a brown-coloured stain to localise the protein of interest.
2.10.1 **Tissue Selection for Immunohistochemistry**

Tissue sections were obtained from paraffin-embedded blocks of placental tissue ranging from 5-40 wks gestation. Previous members of the Morison Lab provided the blocks containing placental tissue from 5-11 wks gestation. Dr. Noelyn Hung (Senior Lecturer, Dunedin School of Medicine) provided the blocks containing placental tissue from 11-40 wks gestation. Multiple tissue sections were cut from each block so that consecutive sections would be used for both KCNH5 staining and haematoxylin and eosin staining.

2.10.2 **Optimisation of Immunohistochemistry**

The rabbit polyclonal antibody for KCNH5 (Abcam, Cat #ab32975) that was used in the western blot was also used to localise KCNH5 protein by immunohistochemistry. A positive control tissue was required for optimisation of the KCNH5 primary antibody. KCNH5 protein is highly expressed in brain tissue (Ludwig et al., 2000). Sections containing fetal brain and sections of placental tissue were used for antibody optimisation. The primary antibody was diluted in Dako Antibody Diluent (Dako, Cat #S0809) and four primary antibody dilutions were tested: 1:125, 1:250, 1500 and 1:1000. Two endogenous protein block solutions were tested and applied separately to tissue sections during primary antibody optimization: the Dako serum-free block (Dako, Cat #X0909) and Bovine Serum Albumin (BSA).

2.10.3 **Haematoxylin and Eosin Staining**

One slide from each tissue sample was stained with Haematoxylin and Eosin. After the slide was rinsed in water, the Gill II Haematoxylin stain (Surgipath, Cat #01520) was applied and left to for 4 min. The slide was rinsed well in running tap water. The slide was then placed into Scott’s Tap Water* for 2 min, followed by a wash in running tap water. The slide was stained with Alcoholic Eosin (Surgipath, Cat #01600) for 1 min and rinsed briefly in tap water, followed by three dehydrating washes in absolute alcohol. The slide was washed three times with xylene for 2 min each wash. A drop of Entellan glue was added to mount a cover slip on top of the stained tissue section.

* **Scott’s Tap Water**
  2 g Potassium bicarbonate
  20 g Magnesium sulphate
  Add ddH₂O to make up to 1 L and thoroughly dissolve.
2.10.4 Immunostaining of Placental Samples

2.10.4.1 Blocking of Endogenous Peroxidase

Following deparaffinisation, slides were incubated in methanolic 3% peroxidase (H₂O₂) for 10 min to block endogenous peroxidase. Slides were then rinsed and soaked in distilled water for 2 min.

2.10.4.2 Heat Retrieval of Antigen

Slides were placed in a microwave-safe container containing citrate buffer (pH 6.0) for 5 min at 70% power. This step was repeated twice more for a total of 15 min of heating in the microwave. After cooling for 30 min, slides were rinsed in distilled water and transferred into a Tris Buffered Saline* (TBS) solution (pH 7.6) for 5 min.

* Tris Buffered Saline (TBS) (Sambrook et al., 1989)
50 mM Tris
150 mM NaCl
Adjust pH to 7.6 using HCl.

2.10.4.3 Antibody Staining

The Dako serum-free protein block (Dako, Cat #X0909) was added to the tissue to fill the area within a wax circle. The Dako block was left on the slide for 10 min and then drained off. Next, 400 µL of a 1:250 dilution of the rabbit polyclonal KCNH5 antibody was added to the slide and incubated for 30 mins. After two 5 min washes in TBS, the secondary antibody, EnVision™ Dual Link (Dako, Cat #K4061), was added and incubated for 30 min. The slide was briefly rinsed with TBS and then soaked in TBS for 5 min. Next, 150 µL of liquid DAB+ Chromogen (Dako, Cat #K3468) substrate was added to the slide for 5 min. After rinsing under distilled water, the slide was counterstained with Haematoxylin solution (Surgipath, Cat #01520).
Chapter Three

Collection, Processing and Screening of Placental Tissue

3.1 Collection of Placental Tissue

3.1.1 Medical Terminations of Pregnancy in the First Trimester

A safe and effective procedure for the medical termination of pregnancy became available in New Zealand after a High Court Judgement in April 2003 (Sparrow, 2004). The protocol included the administration of two drugs, mifepristone and misoprostol, to interrupt pregnancies in women who were no more than 63 days past their last menstrual period. Before it became legal, the procedure was trialled at Wellington Hospital and effectively terminated 94% of pregnancies (Shand et al., 2005). The therapy is now offered at licensed facilities around New Zealand.

Medical staff at Dunedin Hospital administered the approved mifepristone and misoprostol protocol to women undergoing medical terminations of pregnancy in the first trimester. Nurses at Dunedin Hospital presented our study to patients only after they had made their decision to terminate pregnancy. Once the procedure was complete, patients were asked again if they still consented to donate their tissues. Fully consented pregnancy tissues were collected and processed as described previously (Chapter Two). In total, 31 first-trimester placental samples were collected from Dunedin Hospital for use in this study.

3.1.2 Gestational Age of First-Trimester Placental Samples

Nurses provided the gestational age of each sample when the pregnancy tissues were collected. It was important to know the gestational age of each sample, as one of the aims of this project was to identify patterns of methylation in the placenta that may correlate with gestation. The gestational age of each sample was estimated by ultrasound before the pregnancy was terminated. Clinically, the gestational age of a fetus is measured from the first day of a women’s last menstrual cycle to the current date of measurement. Although though the length of the menstrual cycle varies between women (Treloar et al., 1967), gestational age is measured by adding two weeks to the date post conception (ranging 38 to 42 weeks). Since
the terminated pregnancy samples did not always contain a fetus (or an intact fetus), staging according to the Carnegie collection (O'Rahilly et al., 1987) could not be consistently performed across samples. Therefore, the combination of menstrual cycle dating and fetal ultrasound prior to termination of pregnancy was used for this study. Of the 31 samples collected, 6 samples were from pregnancies of 5-6 weeks gestation, 12 samples were of 6-7 weeks gestation, and 13 samples were of 7-9 weeks gestation. The gestational ages of the first-trimester placental samples collected for this study are listed in Table 3.1.

3.1.3 Collection of Term Placenta

Full-term placental tissue was collected as a reference tissue for this study. The methylation profile from term placenta would be compared to the methylation profile from first-trimester placenta to look for similarities or differences in placental methylation. The examination of placental tissues from both early and late stages of gestation would also allow us to identify placental-specific methylation compared to other tissues. Three full term placentas were collected from friends that wished to donate their placentas to this study. Two of the placentas were from natural births and one was collected from a caesarean delivery.

3.2 Processing of Placental Tissue

Before dissection, first-trimester placental tissue was washed in PBS and photographed in a Petri dish. A selection of placental tissue prior to dissection is shown in Figure 3.1. Once the appearance of the tissue was recorded and photographed, tissues were prepared for dissection. Residual maternal blood was washed from placental tissue to prevent maternal contamination of extracted DNA and RNA. Once the placental tissue was clean, the trophoblast villous cells were teased away from the outer chorionic membrane using tweezers and washed in a new solution of PBS. This method of trophoblast isolation was similar to a protocol learned during a visit to the lab of Dr. Stephen Lye (Mount Sinai Hospital, Toronto, Canada), where they performed trophoblast explant cultures from first-trimester trophoblasts. After isolation and washing, tissue containing trophoblast villous cells was examined under a microscope and photographed (Figure 3.2).
Figure 3.1. Photographs of a selection of collected placental samples from terminated pregnancies in the first trimester. Tissues were given a sample number and placed in a Petri dish in PBS buffer. A. Sample F132, gestational age = 6 weeks. B. Sample F135, gestational age = 6 weeks, 4 days. C. Sample F136, gestational age = 6 weeks, 4 days. D. Sample F131, gestational age = 7 weeks. E. Sample F133, gestational age = 7 weeks. F. Sample F129, gestational age = 8 weeks.
Figure 3.2. Photographs through a microscope of isolated placental tissue containing trophoblast villi. Sample F120, gestational age = 7 weeks. A. 1.25 x magnification B. 5 x magnification C. 10 x magnification.

3.2.1 DNA Extraction from Placental Tissue

DNA was extracted from placental samples as described previously (Chapter Two). Colleagues in the Morison Lab suggested Qiagen’s QIAamp DNA Mini Kit for DNA extractions since it had proven to be a rapid and reliable tool to extract DNA from snap frozen human tissue samples. The modification made to the manufacturer’s protocol was to perform two separate elutions of each DNA sample. This was done to maximise DNA yield for use in this project and future studies. Performing two elutions was also done as a precautionary measure in case either stock of DNA became contaminated at any stage. As shown in Table 3.1, performing a second elution increased the total DNA yield for each sample. For most samples, the DNA concentration was higher in the first elution than the second; however in some cases the second elution yielded more DNA than the first.
3.2.2 Fetal Sex Determination for Placental Samples

The fetal sex of each placental sample needed to be determined at the beginning of this project. Since placental tissue is comprised of fetal DNA (Benirschke et al., 2006), the DNA extracted from placental tissue was used for determination of fetal sex. The known fetal sex would be used to screen samples for maternal contamination and to design sex-specific sample groups for subsequent experiments. The sex chromosome arrangement in humans is normally XY in males and XX in females (Morishima et al., 1962). Given that samples would be screened for maternal contamination by examining X chromosome microsatellite repeats, the X chromosome count for each sample needed to be determined before the microsatellite results could be interpreted.

Furthermore, there is a difference in the amount of X chromosome methylation between males and females since one of the two X chromosomes in females is inactivated by DNA methylation (Payer and Lee, 2008). This sex-specific methylation difference meant that placental tissue from female and male fetuses should be separated to avoid methylation differences that may be directly or indirectly associated with X inactivation in females. Since samples were going to be pooled in the subsequent methylation analysis and the peripheral blood comparison tissue was from females, it was necessary to use placentas from female fetuses to allow comparisons of methylation with peripheral blood.

3.2.2.1 Detection of the Y Chromosome

Methods of mammalian sex determination have improved significantly over the years. Previous time-consuming procedures such as cytogenic analysis have progressed to more rapid and highly accurate PCR-based systems (Sembon et al., 2008). To identify male fetuses, a PCR test was designed to detect the presence of the Y chromosome. Specifically, this PCR amplified a region of the Y chromosome that contained the sex-determining region Y (SRY) gene, the gene that initiates male sex determination (Sinclair et al., 1990). The SRY gene is expressed in the fetal gonadal ridge and initiates the development of the testis; however if SRY is not expressed in the fetal gonadal ridge, ovaries will develop instead of testis (Koopman et al., 1990). Known for its role as the ‘initiator’ of sex determination and location on the Y chromosome, the SRY gene was a nice target to identify a Y chromosome in placental samples.
PCR amplification of a 470 bp region of the \( SRY \) gene was performed using primers \#3159 and \#3160 as shown in Figure 3.3.

**Figure 3.3.** The region of the \( SRY \) gene targeted to determine the fetal sex of placental samples. The forward and reverse primers are shown in yellow, and span a region of the exon shown in red. The amplified \( SRY \) PCR product of 470 bp is shown in blue. The coordinates refer to the transcription start site of the \( SRY \) gene.

PCR products were electrophoresed on 2% agarose gels stained with ethidium bromide as described previously (Chapter Two). After visualisation of PCR products under UV light, the presence of a 470 bp band indicated that a sample was male, and absence of a product suggested that a sample was female, as shown in Figure 3.4. The positive and negative control samples were from known male and female DNA samples. The positive control (male sample) displayed a 470 bp band and the negative control (female sample) displayed no band in the gel. These results confirmed that this PCR was amplifying the \( SRY \) gene on the \( Y \) chromosome. Of the 31 placental samples, the \( Y \) chromosome was detected in 16 samples and no product was detected in 15 samples (Figure 3.4).
Figure 3.4. Results from the amplification of a region of the SRY gene to detect the Y chromosome in placental samples. Samples are listed above the well of the gel into which they were loaded and denoted by their fetal “F” sample number. Samples that displayed a 470 bp product were identified as male (♂) and samples with no product were identified as female (♀).

3.2.2.2 Detection of the X and Y Chromosomes

Although amplification of the SRY gene is a useful and accurate method to detect the Y chromosome, the lack of a 470 bp band was not enough evidence to confirm that a sample is female. There is a small chance that lack of a band may be experimental, such as poor quality DNA or failure to add template during PCR setup. Therefore, an additional sex-determining PCR was required to confirm the results from the SRY PCR.

The Amelogenin (AMEL) gene was found to have copies on both the sex chromosomes, and the length of the gene’s non-coding regions differs between the sex alleles (Salido et al., 1992). Although AMEL encodes an extracellular matrix protein involved in the production of
tooth enamel, the gene has been more commonly used as a marker for sex-determination. A study by Sullivan and colleagues developed a reliable PCR-based sex test that involved the amplification of a section of the *AMEL* gene (Sullivan et al., 1993). The PCR yielded fragments of two sizes, 106 bp and 112 bp, from the X and Y homologues of *AMEL*, respectively. When this PCR was designed, it was thought that the primers spanned part of the first intron of *AMEL*, however, genes were not well annotated in 1993 and the current Genome Browser release (GRCh37/hg19) shows the region of amplification to span part of the third intron of *AMEL*. This PCR proved to be a fast and effective method for sex-determination and was unique for its detection of both sex chromosomes. Therefore, the primer sequences designed by Sullivan and colleagues were used in this study to confirm the fetal sex of each placental sample.

PCR amplification of a 106 bp region of the *AMEL* gene on the X chromosome and a 112 bp region of the *AMEL* gene on the Y chromosome using primers #2905 and #2906 is shown in Figure 3.5.

**Figure 3.5.** The region of the *AMEL* gene homolog on the X and Y chromosomes targeted to determine the fetal sex of placental samples. The forward and reverse primers are shown in yellow, and span a region of Intron 3 shown in red. The coordinates refer to the transcription start site for *AMEL*. **A.** The amplified *AMEL* PCR product of 106 bp on the X chromosome is shown in blue. **B.** The amplified *AMEL* PCR product of 112 bp on the Y chromosome is shown in blue.
To resolve the small difference in the sizes between the amplified products from the X and Y chromosome, PCR products were electrophoresed using 4% NuSieve gels. The presence of a single 106 bp band determined a sample to be female (XX). The presence of two differently sized bands, one at 106 bp and one at 112 bp, determined a sample to be male (XY). The control samples from known male and female DNA gave the correct banding patterns in the gel (Figure 3.6). The results for all samples screened using the AMEL PCR are shown in Figure 3.6. Of the 31 placental samples, 16 samples were identified as male and 15 samples as female.

![Figure 3.6](image)

**Figure 3.6.** Results from the amplification of a region of the AMEL gene to detect the X and Y chromosomes in placental samples. Samples are listed above the well of the gel into which they were loaded and denoted by their fetal “F” sample number. Samples that displayed a single product of 106 bp were identified as female (♀) and samples that displayed both products of 106 bp and 112 bp were identified as male (♂).

Results from the two sex-determining PCRs agreed and thus confirmed the fetal sex of each placental sample (Table 3.1).
Table 3.1. First-trimester placental samples collected for methylation analysis.

<table>
<thead>
<tr>
<th>Sample #</th>
<th>Gestational Age (of 40 wks)</th>
<th>Sex of Fetus</th>
<th>DNA Concentration Elution 1 (ng/µL)</th>
<th>DNA Concentration Elution 2 (ng/µL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>F104</td>
<td>5 - 6</td>
<td>Male</td>
<td>560</td>
<td>325</td>
</tr>
<tr>
<td>F105</td>
<td>5.5</td>
<td>Female</td>
<td>454</td>
<td>463</td>
</tr>
<tr>
<td>F106</td>
<td>5 - 6</td>
<td>Female</td>
<td>676</td>
<td>285</td>
</tr>
<tr>
<td>F107</td>
<td>6 +3</td>
<td>Male</td>
<td>565</td>
<td>77</td>
</tr>
<tr>
<td>F108</td>
<td>7 +6</td>
<td>Male</td>
<td>557</td>
<td>92</td>
</tr>
<tr>
<td>F110</td>
<td>6 +2</td>
<td>Male</td>
<td>517</td>
<td>491</td>
</tr>
<tr>
<td>F111</td>
<td>6</td>
<td>Male</td>
<td>540</td>
<td>744</td>
</tr>
<tr>
<td>F112</td>
<td>6 +6</td>
<td>Male</td>
<td>585</td>
<td>239</td>
</tr>
<tr>
<td>F113</td>
<td>6</td>
<td>Male</td>
<td>1255</td>
<td>356</td>
</tr>
<tr>
<td>F114</td>
<td>6 +3</td>
<td>Female</td>
<td>416</td>
<td>423</td>
</tr>
<tr>
<td>F116</td>
<td>8 +1</td>
<td>Male</td>
<td>413</td>
<td>297</td>
</tr>
<tr>
<td>F117</td>
<td>6</td>
<td>Female</td>
<td>695</td>
<td>201</td>
</tr>
<tr>
<td>F118</td>
<td>5 +6</td>
<td>Male</td>
<td>605</td>
<td>389</td>
</tr>
<tr>
<td>F119</td>
<td>8 +1</td>
<td>Male</td>
<td>450</td>
<td>337</td>
</tr>
<tr>
<td>F120</td>
<td>7</td>
<td>Male</td>
<td>417</td>
<td>167</td>
</tr>
<tr>
<td>F121</td>
<td>7</td>
<td>Female</td>
<td>369</td>
<td>160</td>
</tr>
<tr>
<td>F122</td>
<td>7</td>
<td>Male</td>
<td>443</td>
<td>78</td>
</tr>
<tr>
<td>F124</td>
<td>5</td>
<td>Female</td>
<td>249</td>
<td>202</td>
</tr>
<tr>
<td>F125</td>
<td>5 +3</td>
<td>Female</td>
<td>395</td>
<td>687</td>
</tr>
<tr>
<td>F126</td>
<td>7</td>
<td>Male</td>
<td>354</td>
<td>149</td>
</tr>
<tr>
<td>F127</td>
<td>6 +1</td>
<td>Male</td>
<td>343</td>
<td>78</td>
</tr>
<tr>
<td>F128</td>
<td>6 +</td>
<td>Male</td>
<td>309</td>
<td>193</td>
</tr>
<tr>
<td>F129</td>
<td>8</td>
<td>Female</td>
<td>252</td>
<td>156</td>
</tr>
<tr>
<td>F130</td>
<td>7</td>
<td>Female</td>
<td>88</td>
<td>93</td>
</tr>
<tr>
<td>F131</td>
<td>7</td>
<td>Female</td>
<td>319</td>
<td>153</td>
</tr>
<tr>
<td>F132</td>
<td>6 +</td>
<td>Female</td>
<td>313</td>
<td>234</td>
</tr>
<tr>
<td>F133</td>
<td>7</td>
<td>Female</td>
<td>335</td>
<td>302</td>
</tr>
<tr>
<td>F135</td>
<td>6 +4</td>
<td>Female</td>
<td>369</td>
<td>236</td>
</tr>
<tr>
<td>F136</td>
<td>6 +4</td>
<td>Female</td>
<td>282</td>
<td>326</td>
</tr>
<tr>
<td>F137</td>
<td>7</td>
<td>Female</td>
<td>405</td>
<td>82</td>
</tr>
<tr>
<td>F138</td>
<td>8 +4</td>
<td>Male</td>
<td>365</td>
<td>215</td>
</tr>
</tbody>
</table>

3.2.3 Maternal Contamination Screening by X Chromosome Microsatellite Analysis

All samples were screened for microsatellite repeats on the X chromosome to ensure that each sample contained purely fetal DNA. Since the sex of the fetal samples had been previously determined in the sex-determining PCRs, the presence of an extra X allele would suggest maternal DNA in the sample. Subsequent study results would not be accurate if maternal DNA was present, so it was critical to exclude maternal contamination in every sample. Five nucleotide repeats with a high percentage of heterozygosity were chosen in order to detect all possible X chromosome alleles. The microsatellite markers used on the X chromosome are shown (Figure 3.7).
Radioactivity-labelled PCR products (samples F104-F133) were run on a 6% polyacrylamide gel and PCR products without radioactivity (samples F135-F138) were run on a 4% MetaPhor® agarose gel. Maternal contamination was very unlikely if the sample’s banding pattern was as follows: a male sample displayed no more than one band which represented his single X allele, and a female sample displayed no more than two bands which represented her two X alleles. It was accepted for a female sample to display one band since the same microsatellite repeat may be present on both X alleles. Given that these microsatellites are highly heterozygous, there is high probability that maternal DNA would show as an extra band. These microsatellite PCRs were conducted on multiple occasions. Shadow banding was present in some of the polyacrylamide gels but it did not interfere with interpretation of the microsatellite results.
Results from microsatellite marker ss4915376 are shown for samples F104-F133 (Figure 3.8).

**Figure 3.8.** Results from microsatellite marker ss4915376 at Xq26.2 to screen for maternal contamination in placental samples. Samples were labelled with radioactivity and run on a 6% polyacrylamide gel. Sample number and fetal sex is listed above the corresponding well. The number of bands (representing X alleles) was counted and compared to the expected number of X alleles for male (XY) or female (XX) samples. No samples displayed an extra X allele.

Results from microsatellite marker DXS8051 are shown for samples F104-F133 (Figure 3.9).

**Figure 3.9.** Results from microsatellite marker DXS8051 at Xp22.31 to screen for maternal contamination in placental samples. Refer to the legend of Figure 3.8 for details.
Results from microsatellite marker DXS6803 are shown for samples F104-F133 (Figure 3.10).

![Figure 3.10](image1.png)

**Figure 3.10.** Results from microsatellite marker DXS6803 at Xq21.31 to screen for maternal contamination in placental samples. Refer to the legend of Figure 3.8 for details.

Results from microsatellite marker DXS8043 are shown for samples F104-F133 (Figure 3.11).

![Figure 3.11](image2.png)

**Figure 3.11.** Results from microsatellite marker DXS8043 at Xq27.3 to screen for maternal contamination in placental samples. Refer to the legend of Figure 3.8 for details.
Results from microsatellite marker DXS8377 are shown for samples F104-F133 (Figure 3.12).

![Diagram of microsatellite marker DXS8377 at Xq28](image)

**Figure 3.12.** Results from microsatellite marker DXS8377 at Xq28 to screen for maternal contamination in placental samples. Refer to the legend of Figure 3.8 for details.

Results from all five X chromosome microsatellite markers are shown for samples F132-F138 (Figure 3.13). For these samples, the non-radioactive PCR products were visualised on 4% MetaPhor® agarose gels stained with ethidium bromide. Samples F132 and F133 were analysed by the radioactive and non-radioactive methods to confirm that both methods detected the same microsatellite banding patterns. In the non-radioactive gel, the banding patterns of samples F132 and F133 matched their respective banding patterns in the radioactive gel, therefore the non-radioactive method for microsatellite marker detection was confirmed. Similar to the radioactive-PCRs, these non-radioactive PCRs were also conducted multiple times in an attempt to optimise the image quality from the MetaPhor® agarose gels.
Figure 3.13. Results from five microsatellite markers on the X chromosome to screen for maternal contamination in placental samples. Samples F132-F138 were run on a 4% MetaPhor® agarose gel. Sample number and fetal sex is listed above the corresponding well. The number of bands (representing X alleles) was counted and compared to the expected number of X alleles for male (XY) or female (XX) samples. No samples displayed extra X alleles.

The results from examining five X chromosome microsatellite markers showed no evidence of maternal contamination in any of the 31 first-trimester placental DNA samples.
Chapter Four

Genome-wide Methylation Analysis by Methylated DNA Immunoprecipitation and Promoter Microarrays

4.1 MeDIP-DSL and Promoter Microarrays

When this study began in 2007, the use of MeDIP combined with promoter microarrays was the most advanced yet appropriate technique for genome-wide methylation analysis. This method was being used in cancer studies and had proved useful for identifying differentially methylated sequences throughout the genome (Weber et al., 2005; Gebhard et al., 2006; Keshet et al., 2006). In addition, our collaborators at the Babraham Institute (Cambridge, UK) had extensive experience with MeDIP-microarray analysis, so we had a support system in place to handle both the technical and statistical aspects of this new technology. The MeDIP-microarray method was therefore selected as the platform to use for our genome-wide search for differentially methylated gene promoters in first-trimester placenta.

4.1.1 Selection of Microarray Platform

Although there were various MeDIP-microarray platforms available, the Aviva Systems Biology DNA selection and ligation (DSL) assay was chosen because of claims that it had a highly specific and sensitive assay design. As shown in Figure 2.1 (Chapter Two), after the immunoprecipitation of methylated DNA, the DSL platform uniquely selects for methylated sequences using 20,000 sets of paired probes that bind to promoter locations throughout the genome. The subsequent ligation of adjacently bound probes enhances the specificity of the method. Once adjacent probes are ligated, their T7 and T3 overhangs are recognised by primers in the amplification step. The unique amplification of using T7 and T3-primers avoids amplification of random, highly methylated gene sequences throughout the genome. In addition, the use of the same T7 and T3 primer sequences to amplify all probe sequences in this assay achieves similar amplification efficiency for all probes.

The design of the Aviva Systems Biology 20K human gene promoter microarray was another reason that this platform was selected. A promoter-based microarray was preferred over a
CpG-island microarray for the genome-wide methylation analysis of first-trimester placenta. Given that not all promoter regions contain CpG islands, a promoter microarray would identify methylated sequences irrespective of CpG density, thus providing a more extensive investigation of promoter methylation throughout the genome. This project aimed to identify patterns of differential methylation in the placenta that were biologically significant, so gene promoter regions were appropriate locations to examine. Furthermore, it has been reported that some non-CpG island promoters show a significant correlation between methylation and gene expression (Weber et al., 2007). Therefore, selecting a promoter-based microarray platform over a CpG-island platform ensures that the analysed regions are not restricted to CpG-rich promoters.

The Aviva Systems Biology promoter microarray contained approximately 20,000 oligos that corresponded to the amplified probe sequences generated in the DNA Selection and Ligation (DSL) assay. The oligo sequences on the array were between 800 bp upstream and 200 bp downstream of the gene’s transcription start site; the region where methylation has been shown to affect gene transcription. The Aviva Systems Biology DSL and 20K microarray platform was used to identify differentially methylated gene promoters in first-trimester human placenta.

4.1.2 Sample Selection and Experimental Design

The aim of this study was to observe the most common differences in gene promoter methylation between first-trimester placenta and a representative somatic tissue, peripheral blood. Peripheral blood was selected as the comparison somatic tissue for this experiment because it was being used in a simultaneous MeDIP-microarray experiment examining the effects of in vitro fertilisation (IVF) on DNA methylation (performed by Verity Oliver, previous PhD student in Morison Lab).

Since a major location of epigenetic variation between the sexes is methylation of the inactive X chromosome in females, the only way to avoid this sex-specific methylation difference was to examine the sexes separately. Although it would be of interest to examine both males and females separately, the high cost of performing this MeDIP-microarray prevented such an experiment. Female samples were selected for this MeDIP-DSL microarray experiment for two reasons: 1) the comparison peripheral blood tissue was from female samples, and 2) examining females samples meant that any differential X chromosome methylation could be included in the data analysis.
To minimise the inter-individual variation in methylation, female samples were pooled into groups on the microarrays. The pooling of samples reduced the biological variation within each sample group, thereby allowing identification of the most consistent differences in gene promoter methylation. However, it is important to note that pooling samples also includes the risk of ‘outlier’ samples interfering with gene discovery. For example, if one sample in a pool has extremely different global methylation levels compared to the other samples in that pool, this ‘outlier’ sample can shift the measured global methylation level for the whole pooled group. Thus the effect of one or more outlier samples in a pool can potentially prevent the discovery of true differences in methylation when pooled groups are compared. In this project, the use of multiple groups of pooled placenta would decrease the risk of any outlier samples interfering with gene discovery. In addition, pooling samples substantially improved the affordability of this microarray experiment as it reduced the number of microarrays needed to compare the methylation profiles of first-trimester placenta and peripheral blood.

In order to identify methylation patterns in the placenta that were associated with gestational age, first-trimester placental samples were divided into three pooled groups based on gestational age. For example, a placental sample from a pregnancy of 5 weeks, 6 days gestation was put into the “5-6 week” age group, and a sample from a pregnancy of 6 weeks, 1 day gestation was put into the “6-7 week” age group. The groups of pooled placenta were compared to groups of pooled peripheral blood. Unfortunately, the high cost of the MeDIP-microarray experiment meant that technical replicates of these microarrays could not be performed. Any methylation changes associated with gestational age may actually represent biological or technical variation between these three arrays. However, since the methylation of the majority of genes is expected to be stable during the first trimester, the subsequent data analyses could, to some extent, treat the three placenta microarrays as biological and technical replicates.

### 4.2 MeDIP-DSL Microarray Results

#### 4.2.1 Digestion of Genomic DNA

Pooled DNA was fragmented so that the anti-5-methylcytosine antibody could efficiently bind and immunoprecipitate methylated regions of DNA. The optimal fragment size for successful MeDIP was reported to be 300-600 bp (Jacinto et al., 2008). Some MeDIP experiments use sonication to fragment genomic DNA prior to immunoprecipitation; however
the Aviva Systems Biology protocol suggested a restriction digest using the NlaIII enzyme. The NlaIII restriction enzyme is a frequent 4-bp cutter that should theoretically cleave every 256 base pairs \((4^4 = 256)\), although fragment size will vary. Restriction digest was a more reliable method than sonication to fragment genomic DNA for two reasons. First, NlaIII digestion ensured that DNA would consistently be cut into similar size fragments, while sonication could produce large variation in DNA fragment sizes. Second, the DNA fragments after NlaIII digestion were of optimal size for efficient anti-5-methylcytosine binding. Before being used in this experiment, the NlaIII digest was optimised on 1 µg of genomic DNA as shown in Figure 4.1. After visualising results by gel electrophoresis, the optimal volume chosen for digestion of 1 µg genomic DNA was 2 µL of NlaIII enzyme. The cleaved fragment sizes predominantly ranged from 100 to 1000 base pairs after NlaIII digestion (Figure 4.1).

**Figure 4.1.** Optimisation of the NlaIII restriction enzyme on 1 µg genomic DNA in a 20 µL digest visualised on a 0.8% agarose gel. Lane 1: 1 kb Plus Ladder™, Lane 2: Empty well; Lane 3: 2 µL NlaIII used in the digest, Lane 4: 4 µL NlaIII used in the digest, Lane 5: 0 µL NlaIII used in the digest (untreated genomic DNA).
4.2.2 Immunoprecipitation, Labelling and Amplification of Samples

MeDIP was performed on fragmented DNA as described previously (Chapter Two). The anti-5-methylcytosine antibody was used to immunoprecipitate and enrich for methylated DNA. This monoclonal antibody is derived from clone 33 D3, the most commonly used clone for MeDIP (Reynaud et al., 1992; Habib et al., 1999; Mayer et al., 2000; Fraga et al., 2004; Keshet et al., 2006). Given that this clone has been used in most MeDIP protocols, the source of the anti-5-methylcytosine antibody should therefore have little effect on the immunoprecipitation efficiency.

After immunoprecipitation, the selection and ligation of the ‘Total Genomic’ and ‘MeDIP’ DNA samples was performed with minor modifications to the Aviva Systems Biology protocol described in Chapter Two. Prior to microarray hybridisation, two PCR-based screening methods were used to test the efficiency of the DNA selection and ligation assay. The samples in these screening PCRs (Figures 4.2 and 4.3) include an individual peripheral blood sample, two pools of peripheral blood and three pools of first-trimester placenta. The individual peripheral blood sample was not used in this project.

The first PCR screen was performed to ensure that the 20-mer oligo probes had ligated successfully and formed 40-mer oligos. Each ligated 40-mer also contained an upstream T3 primer tag and a downstream T7 primer tag (universal primer landing sites). By using the probes’ T3 and T7 primer tags for PCR amplification, probe ligation was confirmed if a 100 bp product was visualised on a 2% agarose gel stained with ethidium bromide. As shown in Figure 4.2, all samples displayed a 100 bp band (in varying intensities), which confirmed probe ligation. The varying intensities in PCR product between the ‘Total Genomic’ and ‘MeDIP’ DNA samples represent the substantial decrease in abundant template after enrichment for methylated sequences by MeDIP. Samples were then fluorescently labelled and amplified.
After labelling the ligated oligos with fluorescence, a second PCR screen was performed to ensure that the labelled products were successfully amplified. After visualisation by gel electrophoresis, the presence of a 100 bp product confirmed that the labelled probe sequences had been amplified by PCR (Figure 4.3).
Figure 4.3. The PCR screening of samples to test the labelling reaction. The name of each sample is displayed above to the well into which that product was loaded. A. ‘Total Genomic’ DNA samples. A 100 bp band can be seen in each sample. The band below the 100 bp product is primer-dimer. B. ‘MeDIP’ DNA samples. A 100 bp band can be seen in each sample. The faint band below the 100 bp product in peripheral blood pools #1 and #2 is primer-dimer.

4.2.3 Scanning of Microarray Slides

After hybridisation, the GenePix® scanner was used to scan all microarray slides as described previously (Chapter Two). The scanning of slides at a high resolution of 5 µm ensured that each spot (oligo) on the array would be scanned multiple times. During slide scanning, the photo-multiplier tube (PMT) gain was set at 660 for both green and red channels. Consideration was given to optimizing the PMT gain setting, since it is normally adjusted for expression-based microarrays to compensate for intensity differences and avoid spot-specific saturation (Yang et al., 2001; Dodd et al., 2004; Lyng et al., 2004). However, the MeDIP-DSL microarray had to be treated differently since a much stronger signal was expected from the ‘Total Genomic’ DNA channel than from the ‘MeDIP’ DNA channel. The signal from the ‘Total Genomic’ DNA channel, visible at 550 nm (Atto 550, green dye) was expected to be stronger because every probe in the DSL assay should have found template to bind to in genomic DNA. This sample behaved like an internal control for the oligos on the microarray, where each oligo should give a signal. However, the signal from the ‘MeDIP’ sample, visible
at 647 nm (Atto 647, red dye), was restricted to only those oligos that corresponded to methylated sequences of DNA. Therefore, not all probes in the DSL assay would have template to bind to in the ‘MeDIP’ DNA sample. The default PMT gain setting of 660 was used for slide scanning to prevent altering the channel intensities further.

After scanning, the mean signal intensity of each spot and median background signal of the array were calculated using the GenePix® software. Since the spot sizes varied across the array, the mean signal intensity was determined so that the size of the spot would not influence its final signal. To account for signal distortion from sources such as dust or wash buffer residue, the median background signal was determined using the GenePix® software.

4.3 MeDIP-DSL Microarray Data Analyses

4.3.1 Normalisation of Microarray Data

Microarray data are measures of relative spot intensity. There can be significant differences in signal strength caused by sample labeling and hybridisation efficiency, which can introduce dye bias in the data. Dye bias is also a direct result of the physical difference between the two dyes as they have different concentration vs. intensity response curves, especially at lower concentrations. Data normalization is therefore critical to make data comparable between the two channels on the array, as well as comparable to other arrays.

The MeDIP-microarray data required a unique method of normalisation compared to what is typically used for expression-based microarrays. Since the ‘Total Genomic’ DNA sample was expected to hybridise to every spot on the array and the immunoprecipitated (MeDIP) DNA sample would only hybridise to spots that represented methylated genes, differences in signal intensities were expected to be very large between the two channels. A unique method of MeDIP-microarray data normalisation was devised by Dr. Simon Andrews at the Babraham Institute (Cambridge, UK) (Farthing et al., 2008). I was fortunate to be in contact with Dr. Simon Andrews, who kindly offered to perform the data normalisation and subsequent analyses for this MeDIP-microarray experiment.

Dr. Simon Andrews first performed a Loess normalisation on the raw microarray data. This is a common method used to remove dye bias variation from a two-coloured microarray data, based on the assumption that dye bias is the result of differences in spot intensity (Smyth and
A Loess normalization involves fitting a locally weighted regression line to the data to adjust the intensity value for each data point. The regression line is used to make the intensity coming from both red (Atto-647) and green (Atto-550) channels equal, which represented by a log ratio of zero. The use of Loess normalisation allows for direct and reproducible comparisons to be made between microarray data sets. After the Loess normalisation was performed, the normalised data from the arrays was combined into groups for filtering and statistical analysis.

Three separate analyses were performed on the MeDIP-microarray data. The aim of two data analyses (#1 and #3) was to identify genes that displayed placental-specific promoter methylation. The aim of data analysis #2 was to identify changes in promoter methylation that were associated with the gestational age of the placental tissue. Genes were selected from each analysis and added to a list of candidate differentially methylated genes. The methylation of each of these genes would be validated by Sequenom (Chapter Five).

4.3.2 Data Analysis #1: Using Fold-Change Cutoff Values

The aim of this analysis was to generate a list of candidate genes that were consistently “methylated” or “unmethylated” in the three placenta arrays when compared to the two peripheral blood arrays. After normalisation, the two peripheral blood data sets were combined and used as a single data set by taking the mean fold-change value of each spot from both peripheral blood arrays. The single (combined) peripheral blood data set was then compared to the three first-trimester placenta data sets. Initially, a fold-change cutoff value of 2.0 was set as the threshold for microarray data analysis. This meant that genes were deemed “methylated” if they had a fold change (red / green) that was ≥ 2.0, and genes were deemed “unmethylated if they had a fold change that was ≤ 0.5. However, there were no genes that changed state between placenta and peripheral blood at the 2-fold threshold. At a cutoff value of 1.5-fold, two genes were “methylated” (fold change ≥ 1.5) and one gene was “unmethylated” (fold change ≤ 0.66) in all three placental data sets compared to peripheral blood. At a 1.4-fold threshold, six candidate genes were identified: three genes were deemed “methylated” in placenta (fold change ≥ 1.4) and “unmethylated” in peripheral blood (fold change ≤ 0.7), and three genes were deemed “unmethylated” in placenta and “methylated” in peripheral blood. The six candidate differentially methylated genes identified at the 1.4-fold cutoff are shown in Figure 4.4.
Figure 4.4. The six candidate genes identified in the MeDIP-microarray experiment using a 1.4-fold change threshold cutoff. The four microarray data sets are shown on the x-axis. The y-axis represents fold change. Three genes (GABRA6, CCRL1, and C20orf144) were “unmethylated” in peripheral blood and “methylated” in the three placenta arrays. Three genes (AKR1C4, KRT17 and KCTD4) were “methylated” in peripheral blood and “unmethylated” in the three placenta arrays.

The fold change values for each of the six candidate genes identified at the 1.4-fold cutoff are listed in Table 4.1.

Table 4.1. Fold change values for candidate differentially methylated genes identified at the 1.4-fold cutoff in MeDIP microarray analysis #1.

<table>
<thead>
<tr>
<th>Gene Name</th>
<th>GenBank Accession</th>
<th>Peripheral blood (Slides 4+5)</th>
<th>5-6 wks placenta (Slide 1)</th>
<th>6-7 wks placenta (Slide 2)</th>
<th>7-9 wks placenta (Slide 3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GABRA6</td>
<td>NM_000811</td>
<td>0.57</td>
<td>1.59</td>
<td>1.73</td>
<td>2.16</td>
</tr>
<tr>
<td>C20orf144</td>
<td>NM_080825</td>
<td>0.61</td>
<td>1.65</td>
<td>1.95</td>
<td>1.51</td>
</tr>
<tr>
<td>CCRL1</td>
<td>NM_178445</td>
<td>0.66</td>
<td>1.54</td>
<td>2.07</td>
<td>1.75</td>
</tr>
<tr>
<td>KCTD4</td>
<td>NM_198404</td>
<td>1.50</td>
<td>0.42</td>
<td>0.39</td>
<td>0.22</td>
</tr>
<tr>
<td>KRT17</td>
<td>NM_000422</td>
<td>1.75</td>
<td>0.58</td>
<td>0.5</td>
<td>0.56</td>
</tr>
<tr>
<td>AKR1C4</td>
<td>NM_001818</td>
<td>1.56</td>
<td>0.62</td>
<td>0.71</td>
<td>0.59</td>
</tr>
</tbody>
</table>
4.3.3 Data Analysis #2: A Quality-Threshold (QT) Clustering Analysis

A quality-threshold (QT) clustering analysis (≥ 10 genes per group, R ≥ 0.7) using the GeneSpring QT clustering algorithm was performed to identify genes that displayed similar patterns of promoter methylation that correlated positively or inversely with gestational age. The three pools of placenta were treated individually, so a change in methylation in any of the gestational age pools, or in the peripheral blood pool, was considered a result. At a 1.5-threshold, the QT clustering analysis produced a list of 407 potentially interesting genes that were arranged into 12 gene cluster (pattern) sets (Figure 4.5). Each gene cluster contained genes that displayed similar patterns of changing promoter methylation between any of the four sample groups. The peripheral blood data set was comprised of data from two microarrays (labeled “Con” for control). The three gestational age pools of first-trimester placenta are from individual microarrays (labeled “5” for 5-6 wks, “6” for 6-7 wks, and “7” for 7-9 wks).

The results from the QT clustering analysis were used to identify patterns of changing promoter methylation between first-trimester placenta and peripheral blood. Gene set 1 contained the largest number of genes that displayed similar patterns of changing methylation (106 genes, Figure 4.5). This gene cluster showed a change in promoter methylation in the seven to nine week placenta group, whereas methylation levels appeared more similar in the younger placental groups and peripheral blood.

Two of the twelve gene clusters displayed particularly interesting patterns of methylation changes. Gene sets 4 and 11 displayed promoter methylation patterns that were linearly correlated with development (Figure 4.5), which suggested that the methylation changes might be biologically meaningful. Gene set 4 showed a general linear decrease in placental methylation between five to nine weeks of gestation (Figure 4.5). This decrease in promoter methylation may represent genes that are epigenetically silenced in early placental development, but then become activated (unsilenced) as gestation progresses. Gene set 11 showed a general linear increase in placental methylation between five to nine weeks of gestation (Figure 4.5), which may represent genes that become silenced as the placenta develops.
Figure 4.5. The quality threshold (QT) clustering analysis (≥ 10 genes per group, R ≥ 0.7) for the MeDIP-microarray data at the 1.5-fold threshold. Data from five microarrays generated a list of 407 interesting genes, which were arranged into 12 gene sets (clusters). Each gene set contains genes that have similar patterns of changing promoter methylation between the data sets. The data sets listed on the x-axis are: Con: the combined peripheral blood data (mean value from the two peripheral blood arrays); 5: 5-6 wks placenta; 6: 6-7 wks placenta and 7: 7-9 wks placenta. The y-axis represents fold-change. The coloured lines with each gene set display the changing pattern of promoter methylation for genes within the gene cluster.
A literature search was performed for each gene in gene sets 4 and 11, and individual genes were selected based on their known function. Although this is a form of gene ‘cherry picking’, using the reported gene function as the selection criteria for candidate genes was the most appropriate method to narrow down the large lists of genes that were generated by this clustering analysis.

In gene set 4 (Figure 4.5), the NBR1 gene (neighbour of BRCA1 gene 1) was selected as a candidate gene because of its function and genomic location. NBR1 is positioned next to the well-known tumour suppressor gene, BRCA1, which is epigenetically silenced in breast and ovarian cancer (Brown et al., 1996; Dobrovic and Simpfendorfer, 1997; Dimitrov et al., 2001). It was reported that NBR1 was down regulated in malignant mammary cells compared to normal cells, suggesting a similar functional role as BRCA1 (Dimitrov et al., 2001). NBR1 may be acting as an epigenetically regulated growth regulator during placental development. If NBR1 is involved in the suppression of cell growth, perhaps the decrease in promoter methylation observed during five to nine weeks of gestation could play a role in the regulation of trophoblast invasion that is required for placental development.

The second gene selected from gene set 4 was PCDHA12 (protocadherin alpha 12). PCDHA12 encodes a cadherin-like cell adhesion protein, and it has been reported that this gene is epigenetically silenced in breast cancer (Novak et al., 2008). Since cell adhesion molecules are critical for successful trophoblastic migration through maternal decidua during the first trimester of pregnancy (Rahnama et al., 2006), perhaps the decrease in promoter methylation observed for PCDHA12 correlates with the migratory behaviour of trophoblasts. It would be interesting if PCDHA12 was acting as an epigenetically silenced cell adhesion molecule and therefore not interfering with trophoblast invasion – a critical process required for successful placental development.

In gene set 11 (Figure 4.5), the RASSF2 gene (ras association domain family member 2) was selected because of its role as a tumour suppressor and its reported epigenetic inactivation in many cancers (Vos et al., 2003; Akino et al., 2005; Endoh et al., 2005; Kaira et al., 2007; Cooper et al., 2008; Imai et al., 2008). RASSF2 has been found to function as a suppressor of genes involved in angiogenesis, immune response and metastasis (Imai et al., 2008). Both angiogenesis and the prevention of a maternal immune response are important processes involved in successful placental development (Cross et al., 1994; Zhou et al., 1997; Duc-Goiran et al., 1999; Than et al., 2009). Perhaps the increase in promoter methylation of
RASSF2 in the placenta during five to nine weeks of gestation relates to the increased blood supply that is established between the mother and fetus (through angiogenesis). The epigenetic silencing of RASSF2 may also correlate with the suppression of the maternal immune response—a critical process that is required for a successful pregnancy.

The second gene selected from gene set 11 was HINT1 (histidine triad nucleotide binding protein 1). HINT1 is a tumour suppressor gene that has been reported to be epigenetically silenced in some cancers (Su et al., 2003; Yuan et al., 2004; Wang et al., 2007). If HINT1 is acting as a growth regulator gene in the placenta, then perhaps the increase in placental promoter methylation during five to nine weeks of gestation correlates with the increasing invasive activity of trophoblasts during this time of gestation. HINT1 may be an epigenetically regulated tumour suppressor gene that is involved in both cancer and placental development.

In total, four genes were selected from the QT clustering analysis at the 1.5-fold threshold. The fold change values for these genes in each data set are listed in Table 4.2. Based on the literature, we hypothesized that the promoter methylation of these four genes may regulate the activity of both trophoblasts and tumours.

<table>
<thead>
<tr>
<th>Gene Set #</th>
<th>Gene Name</th>
<th>GenBank Accession</th>
<th>Peripheral blood (Slides 4+5)</th>
<th>5-6 wks placenta (Slide 1)</th>
<th>6-7 wks placenta (Slide 2)</th>
<th>7-9 wks placenta (Slide 3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>NBR1</td>
<td>NM_005899</td>
<td>0.59</td>
<td>1.62</td>
<td>1.06</td>
<td>0.49</td>
</tr>
<tr>
<td>4</td>
<td>PCDHA12</td>
<td>NM_018903</td>
<td>0.57</td>
<td>2.30</td>
<td>1.32</td>
<td>0.64</td>
</tr>
<tr>
<td>11</td>
<td>RASSF2</td>
<td>NM_014737</td>
<td>0.52</td>
<td>0.53</td>
<td>0.88</td>
<td>1.57</td>
</tr>
<tr>
<td>11</td>
<td>HINT1</td>
<td>NM_005340</td>
<td>0.53</td>
<td>0.33</td>
<td>0.62</td>
<td>1.82</td>
</tr>
</tbody>
</table>

### 4.3.4 Data Analysis #3: Largest Fold Change Difference

A third microarray data analysis was performed when data from two additional peripheral blood microarrays became available from Verity Oliver (Morison Lab, University of Otago). The new peripheral blood data sets were added to previous peripheral blood data, and the mean value was taken from the four data sets to generate a single peripheral blood data set. The three gestational age groups of first-trimester placenta were also combined and the mean was taken to create a single placenta data set.
Initially, data were analysed using threshold cutoff values as done in Analysis #1, but this method did not identify any differentially methylated candidate genes. A statistical approach was then taken, however no candidate genes were found even with a p-value as high as 0.1. This analysis was then modified to focus clearly on the difference between the means, thereby ignoring formal statistical testing to identify genes that displayed the largest difference in fold change between the mean of the placenta and peripheral blood data sets. These genes are shown in Figure 4.6. For each of these genes, the mean fold change and fold change difference between peripheral blood and first-trimester placenta are listed in Table 4.3.

**Figure 4.6.** Genes that show the largest changes between peripheral blood and first-trimester placenta. On the x-axis: Peripheral Blood: the mean peripheral blood fold change value taken from four arrays; First-trimester placenta: the mean first-trimester placenta fold change value taken from three arrays. The line drawn at y = 1 represents the 1:1 ratio (red / green = 1).

As shown in Figure 4.6, more than half of the candidate genes identified in this analysis do not span the 1:1 ratio (the line drawn at y = 1). The 1:1 ratio represents when the mean fold change of red / green = 1 within a data set. Given that the 1:1 ratio represents the mean of the data set, this mean, or any ratio, cannot be used to determine a specific methylation status. The candidate genes that have fold change values above the 1:1 ratio in both placenta and
peripheral blood appear more “methylated”, given that their mean fold change in both tissues is $\geq 1$. For example, the mean fold change of $COL23A1$ (Table 4.3) changes from 2.53 (peripheral blood) to 6.17 (first-trimester placenta), indicating that this gene is significantly less methylated in peripheral blood than in first-trimester placenta. Despite the fact that p-values were ignored in this analysis, the candidate genes identified were added to the list of genes that would undergo subsequent validation in the event that any of them were found to be biologically significant.

Table 4.3. The genes with the largest fold change difference between peripheral blood and first-trimester placenta.

<table>
<thead>
<tr>
<th>Gene Name</th>
<th>GenBank Accession</th>
<th>Fold Change (Red / Green)</th>
<th>Absolute value of difference between blood and placenta</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Peripheral blood (Slides 4-8)</td>
<td>First-trimester placenta (Slides 1-3)</td>
</tr>
<tr>
<td>OGG1</td>
<td>NM_016821</td>
<td>2.12</td>
<td>0.97</td>
</tr>
<tr>
<td>ATAD4</td>
<td>NM_024320</td>
<td>0.92</td>
<td>2.95</td>
</tr>
<tr>
<td>CACNA2D1</td>
<td>NM_000722</td>
<td>1.12</td>
<td>2.80</td>
</tr>
<tr>
<td>C20orf19</td>
<td>NM_018474</td>
<td>0.76</td>
<td>1.97</td>
</tr>
<tr>
<td>MRPL36</td>
<td>NM_032479</td>
<td>1.14</td>
<td>2.89</td>
</tr>
<tr>
<td>COL23A1</td>
<td>NM_173465</td>
<td>2.53</td>
<td>6.17</td>
</tr>
<tr>
<td>FEM1B</td>
<td>NM_015322</td>
<td>1.20</td>
<td>2.76</td>
</tr>
<tr>
<td>FRG1</td>
<td>NM_004477</td>
<td>1.34</td>
<td>2.79</td>
</tr>
<tr>
<td>KIF22</td>
<td>NM_007317</td>
<td>0.93</td>
<td>3.18</td>
</tr>
<tr>
<td>CHK1</td>
<td>NM_005198</td>
<td>1.83</td>
<td>0.88</td>
</tr>
<tr>
<td>C6orf114</td>
<td>NM_033069</td>
<td>1.17</td>
<td>3.58</td>
</tr>
<tr>
<td>BCL9</td>
<td>NM_004326</td>
<td>1.34</td>
<td>3.51</td>
</tr>
<tr>
<td>MYO5A</td>
<td>NM_000259</td>
<td>1.23</td>
<td>3.56</td>
</tr>
<tr>
<td>LDB2</td>
<td>NM_001290</td>
<td>2.61</td>
<td>1.26</td>
</tr>
<tr>
<td>KCNH5</td>
<td>NM_172376</td>
<td>3.04</td>
<td>1.32</td>
</tr>
<tr>
<td>AICDA</td>
<td>NM_020661</td>
<td>1.91</td>
<td>0.79</td>
</tr>
<tr>
<td>SGMS1</td>
<td>NM_147156</td>
<td>1.24</td>
<td>4.11</td>
</tr>
<tr>
<td>ITGA2</td>
<td>NM_002203</td>
<td>1.89</td>
<td>0.85</td>
</tr>
<tr>
<td>GLT25D1</td>
<td>NM_024656</td>
<td>0.81</td>
<td>2.79</td>
</tr>
</tbody>
</table>

Taken together, the three separate microarray analyses classified a total of 29 differentially methylated candidate genes between first-trimester placenta and peripheral blood.

### 4.4 Differentially Methylated Candidate Genes

#### 4.4.1 Location of Differential Methylation

The 40-bp microarray probe sequence for each candidate gene was obtained from Aviva Systems Biology (Table 4.4). The probe sequences were used to verify the genomic location of each gene and to extract the gene’s promoter sequence. Since the location of the Aviva probe was anywhere between 800 bp upstream and 200 bp downstream of the gene’s...
transcription start site, the precise location was identified to ensure that subsequent validation assays interrogated the same position.

Table 4.4. The Aviva Systems Biology probe sequences for candidate differentially methylated genes.

<table>
<thead>
<tr>
<th>Gene Name</th>
<th>40 bp probe sequence from Aviva Systems Biology</th>
<th>Chromosome</th>
</tr>
</thead>
<tbody>
<tr>
<td>GABRA6</td>
<td>GCAAGATGCCAATGCTTCTCCAGATTTTCCACAGGGCCCT</td>
<td>5</td>
</tr>
<tr>
<td>C20orf144</td>
<td>CCAAGGCTAAGAATGAGCTGAGTGAGGACCCAGATTACCCTG</td>
<td>20</td>
</tr>
<tr>
<td>CCRL1</td>
<td>GCAACTAGTGAATTTATGACAAATGCTGTGAAGAATTATTA</td>
<td>3</td>
</tr>
<tr>
<td>KC1D4</td>
<td>TCTTCACAACAAACATCCCTCCCTCAGATACAAATATGTG</td>
<td>13</td>
</tr>
<tr>
<td>KR1T7</td>
<td>ACCTGACTGCTCTGCCTTCTCTTCTCCTCTGCTATTGGCT</td>
<td>17</td>
</tr>
<tr>
<td>AKR1C4</td>
<td>CTCCTACTGCTGTTGTAATACACCACATCCTGCTCTCCTCCT</td>
<td>10</td>
</tr>
<tr>
<td>NBR1</td>
<td>CACATCGAATCTTCTTCACCAATACAAATGGCTGTGCTGTG</td>
<td>17</td>
</tr>
<tr>
<td>PCDHA12</td>
<td>CCCAAATGCTGACAAATTTAGATTAAAGAGTGACCCAGGAGA</td>
<td>5</td>
</tr>
<tr>
<td>RASSF2</td>
<td>ACCAGGAGAGAACGAAAGGAGATGGACACACAGGGAGGA</td>
<td>20</td>
</tr>
<tr>
<td>HINT1</td>
<td>CACTACAAAGATACAGCATGAGGAGCAGAGGAGGTGTATACAC</td>
<td>5</td>
</tr>
<tr>
<td>OGG1</td>
<td>CTGGATTTTCACATGGCCCTTCTGACTCCTGCTCGATT</td>
<td>3</td>
</tr>
<tr>
<td>ATAD4</td>
<td>CTCCTGATATGCTGGAAGTCAACAGACACACACAGAGGTG</td>
<td>17</td>
</tr>
<tr>
<td>CACNA2D1</td>
<td>CTAAGAAGCAGGAGACAGCTCCTCGCTCTCCAAGTCTCT</td>
<td>7</td>
</tr>
<tr>
<td>C20orf19</td>
<td>GCAAGCGAGAGACGAAAGAAAGGGAGGGAATGCGCCTTGT</td>
<td>20</td>
</tr>
<tr>
<td>MRPL36</td>
<td>TACTGCCCGCCTCTCGCTGTTCCCGAGCTGTTTACAT</td>
<td>5</td>
</tr>
<tr>
<td>COL23A1</td>
<td>CTCTAAATGTGAGTGCGCAGAGGTCGCTTTCCGCCACGAA</td>
<td>5</td>
</tr>
<tr>
<td>FEM1B</td>
<td>TGAATTTCGCTGCTGCTGAGAGGAGCTGCTGTTTCCAGCAGG</td>
<td>15</td>
</tr>
<tr>
<td>FGK1</td>
<td>TTCCCTGACACACAGTGCGGCAGGAGGAGGAGGAGGAGGAGG</td>
<td>4</td>
</tr>
<tr>
<td>KIF22</td>
<td>GGTGGTCTTACGCGACGCGCTCTCGAGCAGAAGGCAGCGCG</td>
<td>16</td>
</tr>
<tr>
<td>CHK1</td>
<td>GCGAGAGAGAGGCGCTCTCGCTCCTCCTCCAGCAGCAG</td>
<td>22</td>
</tr>
<tr>
<td>C8orf114</td>
<td>GATAGGCGTGCTCAGGCCAGAGAGGAAGGAGGAGGAGGAGG</td>
<td>6</td>
</tr>
<tr>
<td>BCL9</td>
<td>ACTCCACACAGTGCAAGCAGCAGCAGCAGCAGCAGGAGGAGG</td>
<td>1</td>
</tr>
<tr>
<td>MYO5A</td>
<td>CAGGCTGATGATGATGATGATGATGATGATGATGCGAGGG</td>
<td>15</td>
</tr>
<tr>
<td>LDB2</td>
<td>ACAAGGACTGTTGAGCTTCCACCTAAGACTCGAGCAGGCTGTCG</td>
<td>4</td>
</tr>
<tr>
<td>KCNH5</td>
<td>GGGAAGAGAAGACTTCCATCCAGAGAGGAGGGAGGAGGAGG</td>
<td>14</td>
</tr>
<tr>
<td>AICDA</td>
<td>GCAGAGCTGCCCCCTAATTGGTTAACCCTGCTGTACCTCGCC</td>
<td>15</td>
</tr>
<tr>
<td>SGMS1</td>
<td>TGAAGTTTCTGCTCCTCCGCTTGGTGACGCTTAGCCGCTT</td>
<td>10</td>
</tr>
<tr>
<td>ITGA2</td>
<td>ATATGCCACCCCGTCTCCACAGGAGGAGCAAAGCCCTGCCA</td>
<td>5</td>
</tr>
<tr>
<td>GLT25D1</td>
<td>TAACATCCCTTCCACCGCGCTCCCGAGCGCTATGCTGACCTTTA</td>
<td>19</td>
</tr>
</tbody>
</table>

4.4.2 CpG Density of Candidate Genes

The one limitation to the MeDIP technique is that CpG density can create bias in the enrichment for methylated sequences. CpG density has been shown to affect the binding affinity of the anti-5-methylcytosine antibody (Weber et al., 2005; Keshet et al., 2006; Weber et al., 2007). Given that the region of CpG density that usually correlates with biological function is the 1000 bases on either side of the gene’s transcription start site, Saxonov and colleagues classified promoter CpG density into two groups: high-density CpG promoters and low-density CpG promoters (Saxonov et al., 2006). High-density CpG promoters (approximately 70% of all gene promoters) tend to be unmethylated whereas low-density CpG promoters (less than 30% of all promoters) are generally methylated (Keshet et al., 2006; Weber et al., 2007; Ball et al., 2009).
In the MeDIP technique, successful immunoprecipitation by the anti-5-methylcytosine antibody requires the same absolute number of methylated CpG sites regardless of the CpG density. Thus, similar enrichment is expected of low-density CpG promoters that are highly methylated and high-density CpG promoters that have low levels of methylation (Zilberman and Henikoff, 2007). Keshet and colleagues reported that the anti-5-methylcytosine antibody could enrich for methylated sequences in regions that have a CpG density of only two to three percent (Keshet et al., 2006). Weber and colleagues found that regions containing a higher number of methylated CpG sites result in a greater enrichment of methylated sequences (Weber et al., 2005). The anti-5-methylcytosine antibody has been shown to favour enrichment of regions of high CpG densities, such as CpG islands, meaning that regions of low CpG density throughout the genome may be overlooked (Weber et al., 2007). Weber’s group also found that the antibody was insensitive to regions of low CpG density regardless of the actual methylation levels in these regions (Weber et al., 2007).

It is actually the intermediate-density CpG promoters that have shown the best correlation between promoter methylation and gene expression (Weber et al., 2007). After extensive experience with MeDIP followed by candidate gene validation, our collaborators at the Babraham Institute (Cambridge, UK) suggested that gene promoters with an intermediate CpG density of two to nine percent were most likely to be true enrichments of methylated promoter sequences (Farthing et al., 2008). Whilst the effect of CpG density was considered, given that our MeDIP experiments generated a relatively small list of candidate differentially methylated genes, it was decided that no genes would be excluded based on CpG density. Therefore, 29 genes were chosen for subsequent methylation validation by Sequenom MassARRAY®, without consideration of their CpG density. The subsequent validation of candidate gene methylation would be assessed to determine if those promoters of intermediate-density CpG represented true enrichments of methylated sequences.
Chapter Five

Quantifying Candidate Gene Methylation by
Sequenom MassARRAY® EpiTyper® Analysis

5.1 The Sequenom MassARRAY® EpiTyper® Assay

One of the most widely used and reliable methods for quantifying CpG site methylation is the Sequenom MassARRAY® EpiTyper® assay (referred to from here on as Sequenom). This high-throughput technique for DNA methylation analysis uses MALDI-TOF (Matrix-Assisted Laser Desorption/Ionization Time-of-Flight) mass spectrometry to measure CpG site methylation in PCR products (Ehrich et al., 2005). The MassARRAY® MALDI-TOF mass spectrometer can quantify methylation within thousands of PCR fragments per day.

Because of its high throughput, Sequenom is generally favoured over the bisulfite sequencing method to examine methylation within individual genes. Bisulfite sequencing is a method that involves the PCR amplification of bisulfite-converted DNA, followed by cloning and sequencing to examine the methylation patterns of individual alleles (Frommer et al., 1992). Not only is cloning time-consuming and sequencing expensive, but the bisulfite sequencing method is only semi-quantitative. With Sequenom’s ability to quantify CpG site methylation in a time- and cost-effective manner, this technique has become the preferred tool for examining methylation within individual genes.

The Sequenom assay relies on the bisulfite conversion of genomic DNA to differentiate between methylated and unmethylated CpG sites. When genomic DNA is treated with sodium bisulfite, unmethylated cytosines undergo selective deamination and are converted to uracil, while methylated cytosines remain unchanged. Genomic regions of interest are then amplified by PCR, during which uracils are replaced by thymines, resulting in a sequence change of unmethylated cytosines to thymines. The series of C/T changes allows methylated cytosines to be differentiated from unmethylated cytosines. The bisulfite conversion method was initially developed for methylation analysis using sequencing of cloned PCR products.
Sequenom Analysis

(Frommer et al., 1992); however, bisulfite conversion has now become integrated into most methods that examine CpG site methylation.

After PCR is performed on bisulfite converted DNA, a reverse transcriptase reaction is performed to generate single-stranded RNA. As seen in Figure 2.2 (Chapter Two), an unmethylated cytosine is reverse-transcribed into an adenine (A) and a methylated cytosine is reverse-transcribed into a guanine (G). The single-stranded RNA is then cleaved base-specifically, either using a restriction enzyme which cleaves at uracils (T-cleavage) or one that cleaves at cytosines (C-cleavage). The base-specific cleavage results in multiple cleaved fragments that vary in length.

The mass of the cleaved fragments is then measured using the MassARRAY® MALDI-TOF mass spectrometer, which can detect the 16 Dalton mass difference between an A (representing an unmethylated cytosine) and a G (representing a methylated cytosine). Individual CpG site methylation is then quantified by comparing the proportion of products containing a G to those products containing an A at each CpG site. Both of Sequenom’s cleavage reactions (C-cleavage and T-cleavage) have been shown to be equally informative at examining approximately 82% of the CpG sites in a given PCR product (Ehrich et al., 2005). However, for interrogation of PCR products that contain CpG-rich sequences, such as CpG islands, it has been reported that the T-cleavage reaction is more informative than the C-cleavage, since the C-cleavage reaction has been shown to produce many small uninformative fragments due to the high number of cleaved CpG sites (Coolen et al., 2007). Given that this study was examining the methylation of gene promoters (regions of DNA that are known to be CpG-rich), the T-cleavage reaction was selected for the Sequenom assay.

The Sequenom assay can accurately measure cleaved fragments ranging from 1700 to 9000 Daltons (approximately 5 to 27 nucleotides). Fragments that are too short or too long to be accurately quantified are discarded from analysis. For cleaved fragments of optimal mass, the measure of methylation within each fragment is a result of the total mass shift across the fragment, resulting from either one or multiple CpG sites. Although measures of CpG methylation are reliable and reproducible using this method, there are some limitations to the Sequenom assay. One limitation is if there are multiple CpG sites present within a single cleaved fragment: Sequenom cannot determine the exact percentage of methylation at each CpG site, rather it uses an algorithm (Equation 2.1, Chapter Two) to estimate the degree of methylation of each CpG site (Coolen et al., 2007). Furthermore, when two fragments have a
matching (or overlapping) mass, the Sequenom assay cannot distinguish between the two fragments so both fragments are discarded. Additional limitations to the Sequenom assay will be discussed at the end of this chapter.

The Sequenom assay is designed to process information from all cleavage fragments, even those that do not fit the assay’s optimal criteria. This is reflected by Sequenom’s ability to quantify methylation with a confidence range of plus or minus five percent (Ehrich et al., 2005; Ehrich et al., 2007). Based on its reliable, high-throughput and cost-effective assay design, the Sequenom method was selected as the best tool available for validating candidate gene methylation.

### 5.2 Sample Selection and Experimental Design for Sequenom Analysis

#### 5.2.1 Sample Selection and Processing

A main goal of this project was to investigate for placental-specific methylation patterns. Validation of the MeDIP gene methylation results required first-trimester placenta to be compared to peripheral blood, thus 31 first-trimester placenta samples and 24 peripheral blood samples were selected for Sequenom analysis. In addition, several other fetal and adult somatic tissues were selected for Sequenom analysis to identify methylation patterns that were specific to the placenta. A total of 12 adult somatic tissues and 11 fetal somatic tissues were obtained from the Cancer Genetics Laboratory (University of Otago). Six sections from one term placenta (that was collected earlier in this study) were also included. The details for these additional tissues are listed in Table 5.1, and when combined with the first-trimester placenta and peripheral blood samples, a total of 84 samples were selected for Sequenom analysis.

<table>
<thead>
<tr>
<th>Sample #</th>
<th>Tissue Type</th>
<th>Tissue Detail</th>
</tr>
</thead>
<tbody>
<tr>
<td>x351</td>
<td>Adult somatic</td>
<td>Muscle</td>
</tr>
<tr>
<td>x351</td>
<td>Adult somatic</td>
<td>Lymph node</td>
</tr>
<tr>
<td>x351</td>
<td>Adult somatic</td>
<td>Pancreas</td>
</tr>
<tr>
<td>x351</td>
<td>Adult somatic</td>
<td>Heart</td>
</tr>
<tr>
<td>x351</td>
<td>Adult somatic</td>
<td>Lung</td>
</tr>
<tr>
<td>x351</td>
<td>Adult somatic</td>
<td>Thyroid</td>
</tr>
<tr>
<td>x351</td>
<td>Adult somatic</td>
<td>Liver</td>
</tr>
<tr>
<td>x272</td>
<td>Adult somatic</td>
<td>Kidney</td>
</tr>
<tr>
<td>x272</td>
<td>Adult somatic</td>
<td>Spleen</td>
</tr>
<tr>
<td>H111</td>
<td>Adult somatic</td>
<td>Brain</td>
</tr>
<tr>
<td>x5068</td>
<td>Adult somatic</td>
<td>Colon</td>
</tr>
</tbody>
</table>
Table 5.2. Samples used for Sequenom analysis.

<table>
<thead>
<tr>
<th>First-trimester Placenta (n = 31)</th>
<th>Term Placenta (n = 6)</th>
<th>Peripheral Blood (n = 24)</th>
<th>Adult Somatic Tissues (n = 12)</th>
<th>Fetal Somatic Tissues (n = 11)</th>
</tr>
</thead>
<tbody>
<tr>
<td>F104 F122 Umbilical cord</td>
<td>x7056 x7118</td>
<td>x351 Muscle</td>
<td>F96 Brain</td>
<td></td>
</tr>
<tr>
<td>F105 F124 Maternal 1/3</td>
<td>x7058 x7128</td>
<td>x351 Lymph node</td>
<td>F96 Adrenal</td>
<td></td>
</tr>
<tr>
<td>F106 F125 Middle 1/3</td>
<td>x7059 x7142</td>
<td>x351 Pancreas</td>
<td>F96 Stomach</td>
<td></td>
</tr>
<tr>
<td>F107 F126 Fetal 1/3</td>
<td>x7064 x7145</td>
<td>x351 Heart</td>
<td>F96 Umbilical cord</td>
<td></td>
</tr>
<tr>
<td>F108 F127 Chorion</td>
<td>x7071 x7149</td>
<td>x351 Lung</td>
<td>F96 Kidney</td>
<td></td>
</tr>
<tr>
<td>F110 F128 Amnion</td>
<td>x7076 x7151</td>
<td>x351 Thyroid</td>
<td>F74 Vertebrae</td>
<td></td>
</tr>
<tr>
<td>F111 F129</td>
<td>x7080 x7154</td>
<td>x351 Liver</td>
<td>F52 Tongue</td>
<td></td>
</tr>
<tr>
<td>F112 F130</td>
<td>x7084 x7157</td>
<td>x272 Kidney</td>
<td>F52 Thigh muscle</td>
<td></td>
</tr>
<tr>
<td>F113 F131</td>
<td>x7095 x7172</td>
<td>x272 Spleen</td>
<td>F51 Heart</td>
<td></td>
</tr>
<tr>
<td>F114 F132</td>
<td>x7098 x7174</td>
<td>H111 Brain</td>
<td>F51 Eyeball</td>
<td></td>
</tr>
<tr>
<td>F116 F133</td>
<td>x7108 x7177</td>
<td>x5068 Colon</td>
<td>F129 Liver</td>
<td></td>
</tr>
<tr>
<td>F117 F135</td>
<td>x7110 x7181</td>
<td>x8258 Cervix</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F118 F136</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F119 F137</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F120 F138</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F121</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
5.2.2 Selection and Analysis of Differentially Methylated Candidate Genes

Three separate analyses of the MeDIP-microarray data generated a list of 29 differentially methylated candidate genes that required independent validation. As discussed in Chapter Four, candidate genes were selected from each analysis by using threshold values, QT-clustering and the largest fold-change difference between first-trimester placenta and peripheral blood data sets. Given that the MeDIP-microarray experiment was used as a gene discovery tool to identify differentially methylated genes in the placenta, CpG density was not used to exclude candidate genes from undergoing subsequent validation.

As mentioned previously, it was important to amplify the region of differential methylation that corresponded to the Aviva Systems Biology 40-bp probe sequence that was spotted on the MeDIP microarray (Table 4.4, Chapter Four). For some genes, it was not always possible to design functional PCR primers around the probe sequence, so primers were designed as close to the probe sequence as possible. In addition, some genes did not contain CpG sites at the Aviva probe sequence itself. This is explained by the fact that the methylation associated with the probe sequence can be many base pairs away from the actual probe location depending on the size of the immunoprecipitated fragment.

After each Sequenom run, data was filtered using specific criteria that were described previously (Chapter Two) in order to remove any inaccurate measurements of methylation. The criteria for data filtering were suggested by Christine Cauldrey (AgResearch Ruakura, Hamilton) based on her extensive experience in using Sequenom for DNA methylation analysis. In the Sequenom results that follow, the “mean promoter methylation” was determined for each sample by averaging the individual CpG site methylation values within the amplicon. All CpG sites were included in the calculation of mean promoter methylation unless stated otherwise. In addition, single CpG sites were located on individual cleavage fragments unless stated otherwise. A mean promoter methylation value was determined for first-trimester placenta, term placenta, peripheral blood, adult somatic and fetal somatic tissue groups by taking the average of the mean promoter methylation values for each of the samples within the group. The standard deviation and 95% confidence interval of the group mean is displayed by the error bars on the graphs in this chapter.

In order to identify placental-specific methylation patterns, gene promoter methylation in placenta was compared to that in somatic tissues (taken as a group). Although statistical
testing is extremely useful to determine whether small differences between sample groups are significant, the aim of this study was to identify large differences in promoter methylation between the placenta and all somatic tissues. For example, a gene that showed low methylation in the placenta and high methylation in somatic tissues (or vice versa) would be selected for subsequent analysis, as it may be biologically significant. For each of the tissues the 95% confidence interval of the mean was calculated. For those genes that showed large differences between the mean values of placental and somatic tissues, the lack of overlap between the 95% confidence intervals of the mean was used to determine statistical significance.

In the following sections, the Sequenom results for 28 genes are displayed. The results are presented in a standardised format. As shown in Table 5.3, 24 of these genes did not show biologically meaningful changes in methylation. The genes that show meaningful differences are shown on pgs. 1244, 132, 1455 and 1466 (Table 5.3). A description of the structure and results for each gene is presented in the following pages.

**Table 5.3.** Sequenom results showing biologically meaningful genes.

<table>
<thead>
<tr>
<th>Gene Name</th>
<th>Biologically meaningful?</th>
<th>Page #</th>
</tr>
</thead>
<tbody>
<tr>
<td>GABRA6</td>
<td>No</td>
<td>120</td>
</tr>
<tr>
<td>C20orf144</td>
<td>No</td>
<td>121</td>
</tr>
<tr>
<td>KCTD4</td>
<td>No</td>
<td>122</td>
</tr>
<tr>
<td>KRT17</td>
<td>No</td>
<td>123</td>
</tr>
<tr>
<td>AKR1C4</td>
<td>Yes</td>
<td>124</td>
</tr>
<tr>
<td>NBR1</td>
<td>No</td>
<td>126</td>
</tr>
<tr>
<td>PCDHA12</td>
<td>No</td>
<td>127</td>
</tr>
<tr>
<td>RASSF2</td>
<td>No</td>
<td>128</td>
</tr>
<tr>
<td>HINT1</td>
<td>No</td>
<td>129</td>
</tr>
<tr>
<td>OGG1</td>
<td>No</td>
<td>131</td>
</tr>
<tr>
<td>ATAD4</td>
<td>Yes</td>
<td>132</td>
</tr>
<tr>
<td>CACNA2D1</td>
<td>No</td>
<td>133</td>
</tr>
<tr>
<td>C20orf19</td>
<td>No</td>
<td>134</td>
</tr>
<tr>
<td>MRPL36</td>
<td>No</td>
<td>135</td>
</tr>
<tr>
<td>COL23A1</td>
<td>No</td>
<td>136</td>
</tr>
<tr>
<td>FEM1B</td>
<td>No</td>
<td>137</td>
</tr>
<tr>
<td>FRG1</td>
<td>No</td>
<td>138</td>
</tr>
<tr>
<td>KIF22</td>
<td>No</td>
<td>139</td>
</tr>
<tr>
<td>CHK8</td>
<td>No</td>
<td>140</td>
</tr>
<tr>
<td>C6orf114</td>
<td>No</td>
<td>141</td>
</tr>
<tr>
<td>BCL9</td>
<td>No</td>
<td>142</td>
</tr>
<tr>
<td>MYO5A</td>
<td>No</td>
<td>143</td>
</tr>
<tr>
<td>LDB2</td>
<td>No</td>
<td>144</td>
</tr>
<tr>
<td>KCNH5</td>
<td>Yes</td>
<td>145</td>
</tr>
<tr>
<td>AICDA</td>
<td>Yes</td>
<td>146</td>
</tr>
<tr>
<td>SGMS1</td>
<td>No</td>
<td>147</td>
</tr>
<tr>
<td>ITGA2</td>
<td>No</td>
<td>148</td>
</tr>
<tr>
<td>GLT25D1</td>
<td>No</td>
<td>149</td>
</tr>
</tbody>
</table>
5.3 Sequenom Analysis of Candidate Genes from MeDIP Analysis #1

In the first MeDIP-microarray analysis, genes were selected using a 1.4-fold threshold cutoff value, indicating that a gene’s fold-change value (red/green) changed between the methylated state (fold change $\geq 1.4$) and unmethylated state (fold change $\leq 0.7$), between placenta and peripheral blood. Six candidate genes were identified in this analysis (Table 4.1). Although many primer sets were trialled, no successful bisulfite PCR could be achieved for CCRL1 so this candidate gene was not studied.

This initial Sequenom run to validate candidate gene promoter methylation analysed all 84 samples. For the purposes of data presentation, sample results were collated into “tissue groups” and contained the following number of samples in each group: 31 first-trimester placentas, six separate regions of one term placenta, 24 peripheral bloods, 12 adult somatic tissues and 11 fetal somatic tissues.

A standardised summary of Sequenom results is presented for each of the genes in this section and for subsequent genes in this chapter.
5.3.1 Sequenom Analysis of GABRA6

The gamma-aminobutyric acid receptor, alpha 6 (GABRA6) gene encodes the alpha-6 subunit of the GABA<sub>A</sub> receptor. The MeDIP-microarray analysis predicted GABRA6 to be “methylated” in first-trimester placenta and “unmethylated” in peripheral blood. A 350 bp product containing four CpG sites was analysed by Sequenom (Figure 5.1A). One CpG site was excluded from analysis since it gave no data in ≥ 30% of samples. Mean promoter methylation for GABRA6 was determined for samples within each tissue group (Figure 5.1B).

![Image](image.png)

Figure 5.1. Sequenom results for GABRA6. A. Region of GABRA6 analysed by Sequenom. Forward and reverse primers are shown in yellow. NlaIII restriction sites used prior to MeDIP are labelled. The region of methylation quantified by Sequenom is shown in grey. Red circles represent CpG sites and the blue box represents the location of the Aviva Systems Biology 40 bp probe sequence. The coordinates refer to the genomic location with respect to the transcription start site. B. Mean promoter methylation data for each tissue group. Error bars represent the 95% confidence interval of the mean.

Sequenom analysis of GABRA6 did not confirm the differential methylation indicated by the MeDIP-microarray analysis. The methylation of first-trimester placenta (9%) was lower than peripheral blood (21%), the opposite of the MeDIP-microarray prediction. Since this project was looking for large-scale differences in promoter methylation, the variable (low) methylation of GABRA6 displayed by the other tissue groups was not considered significant.
5.3.2 Sequenom Analysis of C20orf144

C20orf144 (chromosome 20 open reading frame 144) is encode the hypothetical protein LOC128864; however it is a predicted gene sequence and is currently under review by NCBI. The MeDIP-microarray data predicted C20orf144 to be “methylated” in first-trimester placenta and “unmethylated” in peripheral blood. A 281 bp product containing four CpG sites was analysed by Sequenom (Figure 5.2A). One CpG site was removed from analysis because it was located on a cleavage fragment that had a mass that was too low to be accurately detected by the MALDI-TOF spectrometer. Mean promoter methylation for C20orf144 was determined for samples within each tissue group (Figure 5.2B).

![Figure 5.2](image.png)

**Figure 5.2.** Sequenom results for C20orf144. For legend, see Figure 5.1.

Sequenom did not confirm the methylation status of C20orf144 that was predicted from the MeDIP-microarray experiment. Methylation of first-trimester placenta (48%) was lower than peripheral blood (72%), which was opposite to the MeDIP-microarray prediction. Even so, the difference in methylation between first-trimester placenta and peripheral blood was not large enough to be considered significant. The other tissue groups displayed variable levels of promoter methylation, but these differences were not considered to be biologically significant.
5.3.3 Sequenom Analysis of KCTD4

The potassium channel tetramerisation domain containing 4 (KCTD4) gene encodes a BTB/POZ domain-containing protein. The MeDIP-microarray analysis indicated that KCTD4 was “unmethylated” in first-trimester placenta and “methylated” in peripheral blood. A 396 bp product containing eight CpG sites was analysed by Sequenom (Figure 5.3A). Three CpG sites had to be excluded from analysis since one site was located on a low-mass cleavage fragment and two sites gave no data in ≥30% of samples. Mean promoter methylation for KCTD4 was determined for samples within each tissue group (Figure 5.3B).

![Sequenom results for KCTD4](image.png)

**Figure 5.3.** Sequenom results for KCTD4. For legend, see Figure 5.1.

Sequenom analysis of KCTD4 did not confirm the differential methylation identified in the MeDIP-microarray analysis. Although promoter methylation in first-trimester placenta (70%) was lower than the methylation of peripheral blood (93%), this difference was not large enough to be considered significant. The low methylation observed in term placenta (10%) was ignored since only one of the six sections of term placenta was included the data analysis; the other sections were removed since ≥30% of their CpG sites gave no data.
### 5.3.4 Sequenom Analysis of KRT17

The keratin 17 (*KRT17*) gene encodes the type I intermediate filament chain keratin 17 protein, which is expressed in epidermal appendages such as the nail bed, hair follicle and sebaceous glands. The MeDIP-microarray data predicted *KRT17* to be “unmethylated” in first-trimester placenta and “methylated” in peripheral blood. A 356 bp product containing 13 CpG sites was analysed by Sequenom (Figure 5.4A). Five CpG sites were excluded from analysis since two sites were located on fragments of high mass and three were located on fragments of low mass, all of which were outside of the accurate detection limits of the mass spectrometer. Mean promoter methylation for *KRT17* was determined for samples within each tissue group (Figure 5.4B).

![Sequenom Analysis Diagram](image)

**Figure 5.4.** Sequenom results for *KRT17*. For legend, see Figure 5.1.

Sequenom did not confirm the differential methylation of *KRT17* identified in the MeDIP-microarray experiment. Although methylation of first-trimester placenta (49%) was lower than peripheral blood (69%), this difference was not large enough to be considered significant. Although slightly higher, the methylation profiles of the adult somatic tissues and fetal somatic tissues were also not considered to be biologically significant.
5.3.5 **Sequenom Analysis of AKR1C4**

The aldo-keto reductase family 1, member C4 (AKR1C4) gene encodes a protein that is part of the aldo-keto reductase superfamily. The MeDIP-microarray data predicted AKR1C4 to be “unmethylated” in first-trimester placenta and “methylated” in peripheral blood. A 393 bp product containing three CpG sites was analysed by Sequenom (Figure 5.5A). One CpG site was removed from analysis because it was located on a fragment that had a mass lower than the spectrometer’s detection limit. Mean promoter methylation for AKR1C4 was determined for samples within each tissue group (Figure 5.5B).

![Figure 5.5. Sequenom results for AKR1C4. For legend, see Figure 5.1.](image)

The Sequenom analysis confirmed the differential methylation of AKR1C4 predicted by the MeDIP-microarray experiment. The methylation of first-trimester placenta (34%) was considered to be significantly lower than that of peripheral blood (85%). Although a larger difference was expected, the differential promoter methylation of AKR1C4 may be biologically significant. The low methylation level of term placenta (37%) also suggested a placental-specific methylation pattern, although the confidence interval for term placenta is large.
5.4 **Sequenom Analysis of Candidate Genes from MeDIP Analysis #2**

In the second MeDIP-microarray analysis, a quality-threshold (QT) clustering analysis was performed to identify genes that displayed similar patterns of promoter methylation that correlated positively or inversely with gestational age. At the 1.5-fold threshold, four genes were selected from two gene clusters based on their known function (Table 4.2).

This Sequenom run was performed at the same time as the previous run, which analysed all 84 samples. Samples contained the following number of samples in each group: 31 first-trimester placentas, six separate regions of one term placenta, 24 peripheral bloods, 12 adult somatic tissues and 11 fetal somatic tissues.
5.4.1  **Sequenom Analysis of NBR1**

The neighbour of BRCA1 gene 1 (*NBR1*) gene was selected because of its location and function that was discussed in Chapter Four. The gene cluster generated from the MeDIP-microarray clustering analysis displayed a general linear decrease in *NBR1* promoter methylation in placenta from five to nine weeks of gestation, while the methylation of peripheral blood was predicted to be low. A 457 bp product containing 27 CpG sites was analysed by Sequenom (Figure 5.6A). Three CpG sites were excluded from analysis since they were located on low-mass cleavage fragments. Seven CpG sites were located on fragments that contained one or two other CpG sites. Mean promoter methylation for *NBR1* was determined for samples within each tissue group (Figure 5.6B).

![Sequenom Analysis Diagram](image)

**Figure 5.6.** Sequenom results for *NBR1*. For legend, see Figure 5.1.

Sequenom did not confirm the pattern of decreasing *NBR1* methylation that was indicated from the MeDIP-microarray clustering analysis. Although the methylation of peripheral blood (6%) was low, the methylation of first-trimester placenta in the gestational age groups of five to six weeks (3%), six to seven weeks (2%) and seven to nine weeks (3%) did not display the general linear decrease in methylation that was predicted.
5.4.2 Sequenom Analysis of PCDHA12

The protocadherin alpha 12 (PCDHA12) gene encodes a cadherin-like cell adhesion protein. This gene was selected from the clustering analysis because of its reported function discussed in Chapter Four. PCDHA12 was selected from the same gene cluster as NBR1, which displayed a general linear decrease in promoter methylation in placenta from five to nine weeks of gestation, while the methylation of peripheral blood was predicted to be low. A 364 bp product containing 10 CpG sites was analysed by Sequenom (Figure 5.7A). Six CpG sites were removed from analysis since four sites were located on low-mass cleavage fragments and two sites were located on a high-mass cleavage fragment. Mean promoter methylation for PCDHA12 was determined for samples within each tissue group (Figure 5.7B).

![Sequenom results for PCDHA12](image)

**Figure 5.7.** Sequenom results for PCDHA12. For legend, see Figure 5.1.

The Sequenom analysis did not confirm the pattern of decreasing PCDHA12 methylation that was indicated from the MeDIP-microarray clustering analysis. Although the methylation of peripheral blood (24%) was lower than all first-trimester placentas, the methylation of the five to six weeks (46%), six to seven weeks (36%) and seven to nine weeks (43%) gestational age groups did not display the general linear decrease in methylation that was predicted.
5.4.3 Sequenom Analysis of RASSF2

The ras association domain family member 2 (RASSF2) gene was selected from the MeDIP-microarray clustering analysis because of its reported function discussed in Chapter Four. RASSF2 was selection from a gene cluster that displayed a general linear increase in promoter methylation in first-trimester placenta during five to nine weeks of gestation, while the methylation of peripheral blood was predicted to be low. A 363 bp product containing 12 CpG sites was analysed by Sequenom (Figure 5.8A). Four CpG sites were excluded from analysis because three sites were located on high-mass fragments and one site was located on a low-mass fragment. Mean promoter methylation for RASSF2 was determined for samples within each tissue group (Figure 5.8B).

Sequenom did not confirm the pattern of increasing RASSF2 methylation that was indicated from the MeDIP-microarray clustering analysis. The methylation of first-trimester placenta in the gestational age groups of five to six weeks (3%), six to seven weeks (2%) and seven to nine weeks (3%) did not display any general linear increase as gestation progressed, and the methylation of peripheral blood (12%) was higher than all groups of placenta. The promoter methylation of the other tissue groups was not biologically significant.
5.4.4 **Sequenom Analysis of HINT1**

The histidine triad nucleotide binding protein 1 (*HINT1*) gene was selected from the MeDIP-microarray clustering analysis because of its reported function discussed in Chapter Four. *HINT1* was selected from the same gene cluster as *RASSF1*, which displayed a general linear increase in promoter methylation in placenta from five to nine weeks of gestation, while the methylation of peripheral blood was predicted to be low. A 299 bp product containing 26 CpG sites was analysed by Sequenom (Figure 5.9A). Twelve CpG sites were removed from analysis as 11 sites were on high-mass cleavage fragments and one site was on a low-mass cleavage fragment. Mean promoter methylation for *HINT1* was determined for samples within each tissue group (Figure 5.9B).

![Figure 5.9](image)

**Figure 5.9.** Sequenom results for *HINT1*. For legend, see Figure 5.1.

The Sequenom analysis did not confirm the pattern of increasing *HINT1* methylation that was indicated from the MeDIP-microarray clustering analysis. Although the methylation of peripheral blood (4%) was low, the methylation of the five to six weeks (8%), six to seven weeks (6%) and seven to nine weeks (7%) gestational age groups did not display the general linear increase in methylation that was predicted.
5.5 **Sequenom Analysis of Candidate Genes from MeDIP Analysis #3**

In the third MeDIP-microarray analysis, genes that displayed the largest difference in fold change between the first-trimester placenta and peripheral blood data sets were selected (Table 4.3). This analysis was performed when data became available from two additional peripheral blood microarrays. Data were combined within each tissue group and the mean was taken to generate a single data set for first-trimester placenta and a single data set for peripheral blood. In this analysis, nineteen genes that displayed the largest difference in fold change were selected for validation (Table 4.3).

A total of 39 samples were analysed in this Sequenom run. As done previously, samples were grouped by tissue type and contained the following number of samples in each group: 17 first-trimester placentas, six separate regions of one term placenta, 6 peripheral bloods, 5 adult somatic tissues (heart, liver, kidney, brain and spleen) and 5 fetal somatic tissues (heart, liver, kidney, brain and adrenal).
5.5.1 **Sequenom Analysis of OGG1**

The 8-oxoguanine DNA glycosylase (*OGG1*) gene encodes an enzyme that catalyses the excision of 8-oxoguanine, a mutagenic base by-product caused by exposure to reactive oxygen. The MeDIP-microarray data predicted *OGG1* to be less methylated (fold change = 0.97) in first-trimester placenta and more methylated (2.12) in peripheral blood. A 364 bp product containing 26 CpG sites was analysed by Sequenom (Figure 5.10A). Four CpG sites were removed from analysis; two sites were located on a high-mass fragment, one site was located on a low-mass fragment, and one site gave no data in ≥ 30% of samples. Mean promoter methylation for *OGG1* was determined for samples within each tissue group (Figure 5.10B).

**Figure 5.10.** Sequenom results for *OGG1*. For legend, see Figure 5.1.

Sequenom did not confirm the differential methylation status of *OGG1* that was predicted from the MeDIP-microarray data analysis. Although the methylation of first-trimester placenta (9%) was lower than that in peripheral blood (12%), the difference was not large enough to be considered significant. The other tissue groups displayed similar low levels of promoter methylation.
5.5.2 Sequenom Analysis of ATAD4

The proline rich 15-like (PRR15L or ATAD4) gene encodes a protein that is part of the ATPase family. The MeDIP-microarray analysis predicted ATAD4 to be “methylated” in first-trimester placenta (2.95) and “unmethylated” in peripheral blood (0.92). A 325 bp product containing seven CpG sites was analysed by Sequenom (Figure 5.11A). One CpG site was removed from analysis because it was located on a fragment that had a mass below the accurate detection limits of the spectrometer. Mean promoter methylation for ATAD4 was determined for samples within each tissue group (Figure 5.11B).

![Diagram A](image)

![Diagram B](image)

**Figure 5.11.** Sequenom results for ATAD4. For legend, see Figure 5.1.

The Sequenom analysis did not validate the differential methylation of ATAD4 predicted by the MeDIP-microarray analysis. The methylation of first-trimester placenta (27%) was significantly lower than that of peripheral blood (80%), the opposite of what was predicted. However, this difference in promoter methylation was one of the largest quantified amongst all of the candidate genes, so perhaps the differences in promoter methylation of ATAD4 may be biologically relevant.
5.5.3 **Sequenom Analysis of CACNA2D1**

The calcium channel, voltage-dependent, alpha 2/delta subunit 1 (CACNA2D1) gene encodes a protein in the voltage-dependent calcium channel complex. The MeDIP-microarray analysis predicted CACNA2D1 to be more methylated in first-trimester placenta (2.8) and less methylated in peripheral blood (1.12). A 383 bp product containing 29 CpG sites was analysed by Sequenom (Figure 5.12A). Nineteen CpG sites were removed from analysis; two sites were located on low-mass cleavage fragments, 13 sites were located on high-mass cleavage fragments and four sites gave no data in ≥ 30% of samples. Mean promoter methylation for CACNA2D1 was determined for samples within each tissue group (Figure 5.12B).

**Figure 5.12.** Sequenom results for CACNA2D1. For legend, see Figure 5.1.

Sequenom did not confirm the methylation status of CACNA2D1 that was predicted from the MeDIP-microarray data analysis. Although the methylation of peripheral blood (7%) was low, the methylation of first-trimester placenta was no higher (8%), so differential methylation was not validated between these two groups. All other tissue groups displayed similar low levels of promoter methylation.
5.5.4  **Sequenom Analysis of C20orf19**

The polo-like kinase 1 substrate 1 (*PLK1S1*, also known as *C20orf19*) gene encodes the centrosomal protein kizuna. The MeDIP-microarray analysis predicted *C20orf19* to be “methylated” in first-trimester placenta (1.97) and “unmethylated” in peripheral blood (0.76). A 418 bp product containing 42 CpG sites was analysed by Sequenom (Figure 5.13A). Fifteen CpG sites were removed from analysis; six sites were located on low-mass fragments, eight sites were located on high-mass fragments and one site gave no data in ≥ 30% of samples. Twenty-five of the analysed CpG sites were located on fragments that contained one, two or three other CpGs. Mean promoter methylation for *C20orf19* was determined for samples within each tissue group (Figure 5.13B).

**Figure 5.13.** Sequenom results for *C20orf19*. For legend, see Figure 5.1.

The Sequenom analysis did not confirm the methylation status of *C20orf19* that was predicted from the MeDIP-microarray data analysis. Although the methylation of peripheral blood (6%) was low, the methylation of first-trimester placenta was not much higher (7%), so differential methylation was not validated between these two groups. The other tissue groups displayed similar low levels of promoter methylation, especially in fetal somatic tissues (1%).
5.5.5 **Sequenom Analysis of MRPL36**

The mitochondrial ribosomal protein L36 (*MRPL36*) gene encodes the 39S subunit protein of a mitochondrial ribosome protein. The MeDIP-microarray analysis predicted *MRPL36* to be more methylated in first-trimester placenta (2.89) and less methylated in peripheral blood (1.14). A 378 bp product containing 22 CpG sites was analysed by Sequenom (Figure 5.14A). Seven CpG sites were removed from analysis; two CpG sites were located on low-mass cleavage fragments and five CpG sites were located on high-mass cleavage fragments. Mean promoter methylation for *MRPL36* was determined for samples within each tissue group (Figure 5.14B).

**Figure 5.14.** Sequenom results for *MRPL36*. For legend, see Figure 5.1.

Sequenom did not validate the differential methylation of *MRPL36* predicted by the MeDIP-microarray analysis. Although the methylation of peripheral blood (7%) was low, the methylation of first-trimester placenta was lower (4%), so differential methylation was not validated between these two groups. The other tissue groups displayed similar low levels of promoter methylation.
5.5.6 Sequenom Analysis of COL23A1

The collagen, type XXIII, alpha 1 (COL23A1) gene encodes one of the transmembrane collagen proteins. The MeDIP-microarray analysis predicted COL23A1 to be more methylated in first-trimester placenta (6.17) and less methylated in peripheral blood (2.53). A 298 bp product containing 30 CpG sites was analysed by Sequenom (Figure 5.15A). Twenty CpG sites were removed from analysis since they were located cleavage fragments that each had a mass that was higher than the spectrometer’s detection limit. Seven of the analysed CpG sites were located on fragments that contained one, two or three other CpG sites. Mean promoter methylation for COL23A1 was determined for samples within each tissue group (Figure 5.15B).

![Graph A](image1)

![Graph B](image2)

**Figure 5.15.** Sequenom results for COL23A1. For legend, see Figure 5.1.

The Sequenom analysis did not validate the differential methylation of COL23A1 predicted by the MeDIP-microarray analysis. Although the methylation of first-trimester placenta (6%) was higher than that of peripheral blood (4%), the difference in promoter methylation was so small that it was considered insignificant. COL23A1 promoter methylation for the other tissue groups was also very low.
5.5.7 **Sequenom Analysis of FEM1B**

The fem-1 homolog b (*FEM1B*) gene is thought to encode a protein, however the gene sequence is provisional and is currently under review by NCBI. The MeDIP-microarray analysis predicted *FEM1B* to be more methylated in first-trimester placenta (2.76) and less methylated in peripheral blood (1.2). A 336 bp product containing 33 CpG sites was analysed by Sequenom (Figure 5.16A). Twenty-one CpG sites were removed from analysis; one CpG site was located on a low-mass cleavage fragment; 15 CpG sites were located on high-mass cleavage fragments, and five CpG sites gave no data in ≥ 30% of samples. Ten of the analysed CpG sites were located on fragments that contained one other CpG site. Mean promoter methylation for *FEM1B* was determined for samples within each tissue group (Figure 5.16B).

**Figure 5.16.** Sequenom results for *FEM1B*. For legend, see Figure 5.1.

Sequenom did not confirm the differential methylation of *FEM1B* that was indicated by the MeDIP-microarray analysis. The methylation of peripheral blood (3%) was almost the same as that of first-trimester placenta (4%), both of which were very lowly methylated. *FEM1B* promoter methylation for the other tissue groups was also very low.
5.5.8  Sequenom Analysis of FRG1

The FSHD region gene 1 (FRG1) gene encodes a protein that localises to the nucleolus and has been reported to be involved in facioscapulohumeral muscular dystrophy (FSHD). The MeDIP-microarray analysis predicted FRG1 to be more methylated in first-trimester placenta (2.79) and less methylated in peripheral blood (1.34). A 365 bp product containing 38 CpG sites was analysed by Sequenom (Figure 5.17A). Eleven CpG sites were removed from analysis; three CpG sites were located on a low-mass cleavage fragments and eight CpG sites were located on high-mass cleavage fragments. Nineteen of the analysed CpG sites were located on fragments that contained one, two or three other CpG sites. Mean promoter methylation for FRG1 was determined for samples within each tissue group (Figure 5.17B).

![Figure 5.17. Sequenom results for FRG1. For legend, see Figure 5.1.](image_url)

The Sequenom analysis did not validate the differential methylation of FRG1 that was indicated by the MeDIP-microarray analysis. The methylation of peripheral blood (5%) was equal to that of first-trimester placenta (5%), so no differences in methylation were detected. FRG1 promoter methylation for the other tissue groups was also very low.
5.5.9 **Sequenom Analysis of KIF22**

The kinesin family member 22 (KIF22) gene encodes a protein in the kinesin-like protein family, a group of proteins involved in transporting organelles within a cell and moving chromosomes during cell division. The MeDIP-microarray analysis predicted KIF22 to be “methylated” in first-trimester placenta (3.18) and “unmethylated” in peripheral blood (0.93). A 333 bp product containing 18 CpG sites was analysed by Sequenom (Figure 5.18A). Six CpG sites were removed from analysis, as they were all located on cleavage fragments that were below the accurate detection limits of the mass spectrometer. Seven of the analysed CpG sites were located on fragments that contained one or two other CpG sites. Mean promoter methylation for KIF22 was determined for samples within each tissue group (Figure 5.18B).

![Figure 5.18](image)

**Figure 5.18.** Sequenom results for KIF22. For legend, see Figure 5.1.

Sequenom did not confirm the differential methylation of KIF22 that was indicated by the MeDIP-microarray analysis. Although the methylation of peripheral blood was low (15%), the methylation of first-trimester placenta (11%) was even lower, demonstrating the opposite pattern to what was predicted by the microarray analysis. KIF22 promoter methylation was relatively low across the other tissue groups.
5.5.10 Sequenom Analysis of CHKB

The choline kinase beta (CHKB) gene encodes a protein that catalyses the phosphorylation of choline, the first of the enzymes involved in the biosynthesis of phosphatidylcholine in animal cells. The MeDIP-microarray analysis predicted CHKB to be “unmethylated” in first-trimester placenta (0.88) and “methylated” in peripheral blood (1.83). A 450 bp product containing 51 CpG sites was analysed by Sequenom (Figure 5.19A). Fourteen CpG sites were removed from analysis; two CpG sites were located on a low-mass cleavage fragments, 11 CpG sites were located on high-mass cleavage fragments and one CpG site gave no data in $\geq 30\%$ of samples. Twenty-seven of the analysed CpG sites were located on fragments that contained one, two or three other CpG sites. Mean promoter methylation for CHKB was determined for samples within each tissue group (Figure 5.19B).

![Figure 5.19](image)

**Figure 5.19.** Sequenom results for CHKB. For legend, see Figure 5.1.

The Sequenom analysis did not validate the differential methylation of CHKB that was indicated by the MeDIP-microarray analysis. Although the methylation of first-trimester placenta (7%) was low, the methylation of peripheral blood (8%) was not significantly higher. CHKB promoter methylation for the other tissue groups was also very low.
5.5.11 Sequenom Analysis of C6orf114

The chromosome 6 open reading frame 114 (C6orf114) gene is thought to encode the hypothetical protein LOC85411; however it is a predicted gene sequence and is currently under review by NCBI. The MeDIP-microarray analysis predicted C6orf114 to be more methylated in first-trimester placenta (3.58) and less methylated in peripheral blood (1.17). A 309 bp product containing 16 CpG sites was analysed by Sequenom (Figure 5.20A). Four CpG sites were removed from analysis since they were located on cleavage fragments that had masses above the accurate detection limits of the spectrometer. Eight of the analysed CpG sites were on fragments that contained one other CpG sites. Mean promoter methylation for C6orf114 was determined for samples within each tissue group (Figure 5.20B).

---

**Figure 5.20.** Sequenom results for C6orf114. For legend, see Figure 5.1.

Sequenom did not confirm the differential methylation of C6orf114 that was predicted by the MeDIP-microarray analysis. Although the methylation of peripheral blood (3%) was lower than that of first-trimester placenta (8%), the difference was not significant and therefore differential methylation was not validated. C6orf114 promoter methylation was low across the other tissue groups.
The B-cell CLL/lymphoma 9 (BCL9) gene encodes a protein that is associated with B-cell acute lymphoblastic leukemia, yet its function is unknown. The MeDIP-microarray analysis predicted BCL9 to be more methylated in first-trimester placenta (3.51) and less methylated in peripheral blood (1.34). A 283 bp product containing four CpG sites was analysed by Sequenom (Figure 5.21A). Two of the CpG sites were removed from analysis since they were located on a cleavage fragment that was too high in mass for accurate detection by the spectrometer. Mean promoter methylation for BCL9 was determined for samples within each tissue group (Figure 5.21B).

![Figure 5.21. Sequenom results for BCL9. For legend, see Figure 5.1.](image)

The Sequenom analysis did not validate the differential methylation of BCL9 that was indicated by the MeDIP-microarray analysis. Although the methylation of first-trimester placenta (8%) was higher than that of peripheral blood (6%), the difference was not large enough to be considered significant. BCL9 promoter methylation for the other tissue groups was also very low.
5.5.13 *Sequenom Analysis of MYO5A*

The myosin VA (*MYO5A*) gene encodes one of three myosin V heavy-chain genes and is part of the myosin gene superfamily. The MeDIP-microarray analysis predicted *MYO5A* to be more methylated in first-trimester placenta (3.56) and less methylated in peripheral blood (1.23). A 282 bp product containing 27 CpG sites was analysed by Sequenom (Figure 5.22A). Twenty CpG sites were removed from analysis; 17 CpG sites were located on a high-mass cleavage fragments and three CpG sites gave no data in ≥ 30% of samples. Four of the analysed CpG sites were located on fragments that contained one other CpG sites. Mean promoter methylation for *MYO5A* was determined for samples within each tissue group (Figure 5.22B).

![Figure 5.22](image)

**Figure 5.22.** Sequenom results for *MYO5A*. For legend, see Figure 5.1.

Sequenom did not confirm the differential methylation of *MYO5A* that was predicted by the MeDIP-microarray analysis. The methylation of first-trimester placenta (1%) was equal to that of peripheral blood (1%). *MYO5A* promoter methylation was quantified at 1% across all sample groups, so there was no differential methylation detected in this analysis.
5.5.14 Sequenom Analysis of LDB2

The LIM domain binding 2 (LDB2) gene encodes a protein that interacts with the LIM domains of nuclear proteins and is capable of binding to a variety of transcription factors. The MeDIP-microarray analysis predicted LDB2 to be less methylated in first-trimester placenta (1.26) and more methylated in peripheral blood (2.61). A 221 bp product containing five CpG sites was analysed by Sequenom (Figure 5.23A). One CpG site was removed from analysis as it was located on a cleavage fragment that had a mass too high to be accurately detected by the spectrometer. Two of the analysed CpG sites were located on a fragment that contained one other CpG site. Mean promoter methylation for LDB2 was determined for samples within each tissue group (Figure 5.23B).

**Figure 5.23.** Sequenom results for LDB2. For legend, see Figure 5.1.

The Sequenom analysis did not validate the differential methylation of LDB2 that was indicated by the MeDIP-microarray analysis. The methylation of first-trimester placenta (13%) was higher than that of peripheral blood (2%), the opposite result to what was predicted by the microarray analysis. Furthermore, even with an opposite result, the difference was not large enough to be considered significant.
5.5.15  **Sequenom Analysis of KCNH5**

The potassium voltage-gated channel, subfamily H (eag-related), member 5 (*KCNH5*) gene encodes the pore-forming (alpha) subunit of a voltage-gated potassium channel. The MeDIP-microarray analysis predicted *KCNH5* to be less methylated in first-trimester placenta (1.32) and more methylated in peripheral blood (3.04). A 313 bp product containing 17 CpG sites was analysed by Sequenom (Figure 5.24A). One CpG sites was removed from analysis since it gave no data in any sample. Eight of the analysed CpG sites were located on fragments that contained one or three other CpG sites. Mean promoter methylation for *KCNH5* was determined for samples within each tissue group (Figure 5.24B).

**Figure 5.24.** Sequenom results for *KCNH5*. For legend, see Figure 5.1.

The Sequenom analysis confirmed the differential methylation of *KCNH5* that was predicted by the MeDIP-microarray analysis. The methylation of first-trimester placenta (15%) was significantly lower than that of peripheral blood (86%) and of the adult somatic and fetal somatic tissues. *KCNH5* promoter methylation was different in the sections of term placenta; the maternal, middle and fetal sections and chorion had very low methylation, while umbilical cord and amnion had very high methylation.
5.5.16 **Sequenom Analysis of AICDA**

The activation-induced cytidine deaminase (*AICDA*) gene encodes an RNA-editing deaminase protein. The MeDIP-microarray analysis predicted *AICDA* to be “unmethylated” in first-trimester placenta (0.79) and “methylated” in peripheral blood (1.91). A 256 bp product containing two CpG sites was analysed by Sequenom (Figure 5.25A). Mean promoter methylation for *AICDA* was determined for samples within each tissue group (Figure 5.25B).

![Figure 5.25](image)

**Figure 5.25.** Sequenom results for *AICDA*. For legend, see Figure 5.1.

The Sequenom analysis confirmed the differential methylation of *AICDA* that was predicted by the MeDIP-microarray analysis. The methylation of *AICDA* in first-trimester placenta (37%) was considered to be significantly lower than that peripheral blood (86%). Although the difference in methylation between first-trimester placenta and peripheral blood was not as large as it was for other genes (i.e., *KCNH5*), it still may have biological significance. Given the moderate levels of methylation in term placenta (63%), adult somatic tissues (68%) and fetal somatic tissues (52%), there are no apparent patterns of placental-specific methylation for *AICDA*. 
5.5.17 **Sequenom Analysis of SGMS1**

The sphingomyelin synthase 1 (SGMS1) gene encodes a five-pass transmembrane protein. The MeDIP-microarray analysis predicted SGMS1 to be more methylated in first-trimester placenta (4.11) and less methylated in peripheral blood (1.24). A 266 bp product containing 12 CpG sites was analysed by Sequenom (Figure 5.26A). Three CpG sites were removed from analysis since they were located on a cleavage fragment that had a mass that was too high to be accurately detected by the spectrometer. Two of the analysed CpG sites were located on a fragment that contained one other CpG site. Mean promoter methylation for SGMS1 was determined for samples within each tissue group (Figure 5.26B).

![Figure 5.26. Sequenom results for SGMS1. For legend, see Figure 5.1.](image)

Sequenom did not validate the differential methylation of SGMS1 that was indicated by the MeDIP-microarray analysis. Although the methylation of first-trimester placenta (2%) was one percent higher than that of peripheral blood (1%), this difference was so small that it was considered insignificant. All other tissue groups displayed extremely low promoter methylation for SGMS1.
5.5.18 **Sequenom Analysis of ITGA2**

The integrin, alpha 2 (*ITGA2*) gene encodes a protein that is part of the integrin alpha chain family and is involved in cell adhesion and cell-surface mediated signalling. The MeDIP-microarray analysis predicted *ITGA2* to be “unmethylated” in first-trimester placenta (0.85) and “methylated” in peripheral blood (1.89). A 243 bp product containing 23 CpG sites was analysed by Sequenom (Figure 5.27A). Eleven CpG sites were removed from analysis; ten CpG sites were located on high-mass cleavage fragments and one CpG site was located on a low-mass cleavage fragment. Eight of the analysed CpG sites were located on fragments that contained one, two or three other CpG sites. Mean promoter methylation for *ITGA2* was determined for samples within each tissue group (Figure 5.27B).

![Figure 5.27. Sequenom results for ITGA2. For legend, see Figure 5.1.](image)

The Sequenom analysis did not confirm the differential methylation of *ITGA2* that was indicated by the MeDIP-microarray analysis. Although the methylation of first-trimester placenta (6%) was low (unmethylated), the methylation of peripheral blood (5%) was even lower. There were no significant differences in *ITGA2* promoter methylation detected between any of the tissue groups.
5.5.19 Sequenom Analysis of GLT25D1

The glycosyltransferase 25 domain containing 1 (GLT25D1) gene encodes the procollagen galactosyltransferase 1 protein. The MeDIP-microarray analysis predicted GLT25D1 to be “methylated” in first-trimester placenta (2.79) and “unmethylated” in peripheral blood (0.81). A 258 bp product containing 17 CpG sites was analysed by Sequenom (Figure 5.28A). Four CpG sites were removed from analysis since they were located on a cleavage fragment that was too large in mass to be accurately detected by the spectrometer. Seven of the analysed CpG sites were located on fragments that contained one or two other CpG sites. Mean promoter methylation for GLT25D1 was determined for samples within each tissue group (Figure 5.28B).

Sequenom did not validate the differential methylation of GLT25D1 that was indicated by the MeDIP-microarray analysis. Although the methylation of the methylation of peripheral blood (3%) was low (unmethylated), the methylation of first-trimester placenta (4%) was not significantly different. GLT25D1 promoter methylation was low for all tissue groups in this analysis; no differential methylation was detected.

Figure 5.28. Sequenom results for GLT25D1. For legend, see Figure 5.1.
5.6 Summary of Sequenom Results

In this study, Sequenom assays were successfully designed for 28 of the 29 candidate differentially methylated genes. It was expected that Sequenom would confirm some, if not all, of the differential methylation patterns that were predicted for these candidate genes.

Given that CpG density was not used to exclude candidate genes from undergoing validation, it was anticipated that the methylation of genes that did not have the ‘optimal’ CpG density might not be confirmed. Our colleagues at the Babraham Institute (Cambridge, UK) suggested that a MeDIP-microarray analysis would most likely detect true enrichments of methylated sequences in regions where CpG density is two to nine percent (10 to 45 CpG sites in 500 bp of sequence) (Farthing et al., 2008). Farthing’s group suggested that this optimal CpG density would increase the rate of candidate gene validation, however they established this threshold while using other criteria to determine their list of differentially methylated genes. In their study, for a gene to be identified as a candidate, a minimum of five microarray probes within each gene promoter region was required to have reached this optimal CpG density (Farthing et al., 2008). This was a major difference to the Aviva Systems Biology single-probe platform used in this study; a difference that allowed the Babraham group to avoid variation caused by individual probes on each microarray. By using these strict criteria, Farthing’s group was able to successfully validate 88% to 100% of their candidate genes that were identified from their MeDIP-microarray data (Farthing et al., 2008). Due to the small list of candidate genes that was generated from the MeDIP-microarray data analyses, all of the candidate genes - regardless of CpG density - were included in the Sequenom analysis in the event that any of them were found to be significant.

Since this study aimed to identify patterns of gene promoter methylation that correlated with biological function, only large-scale differences in methylation between first-trimester placenta and peripheral were considered to be significant. There was no set value used to determine if a difference in methylation was significant; the differences were assessed individually for each gene. If a cutoff value were used to determine whether a difference in methylation was significant, some differences that may be biologically relevant could be overlooked. For example, there is no evidence that a gene promoter that is 90% methylated induces more epigenetic silencing than a promoter that is 80% methylated. Furthermore, it remains unclear whether a single methylated CpG site or multiple methylated CpG sites are responsible for the epigenetic silencing of genes. Therefore, the Sequenom results from this
study were assessed by eye and no statistical analyses were performed. The methylation levels of the other tissue groups (term placenta, adult somatic tissues and fetal somatic tissues) were also examined in order to identify patterns of methylation that were specific to the placenta.

Of the 28 candidate genes, the MeDIP-microarray predicted differential methylation between first-trimester placenta and peripheral blood was confirmed in three genes: \textit{ARK1C4}, \textit{KCNH5} and \textit{AICDA} (Table 5.4). Surprisingly, only one of these three genes had a CpG density that was within the optimal range of two to nine percent; \textit{KCNH5} had a CpG density of 4.4%, while \textit{AKR1C4} and \textit{AICDA} had CpG densities of only 1.0% and 0.4%, respectively (Table 5.4). Of the ten genes examined by Sequenom that had an optimal CpG density, \textit{KCNH5} was the only gene that was validated by Sequenom.

**Table 5.4.** CpG density and Sequenom validation of candidate differentially methylated genes.

<table>
<thead>
<tr>
<th>Gene Name</th>
<th>Chromosome</th>
<th># promoter Cpgs (-800 to +200 bp around TSS)</th>
<th>Cpg density (%)</th>
<th>Optimal Cpg density? (2 to 9%)</th>
<th>Confirmed by Sequenom?</th>
</tr>
</thead>
<tbody>
<tr>
<td>GABRA6</td>
<td>5</td>
<td>6</td>
<td>1.2</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>C2orf144</td>
<td>20</td>
<td>17</td>
<td>3.4</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>KCTD4</td>
<td>13</td>
<td>13</td>
<td>2.6</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>KRT17</td>
<td>17</td>
<td>31</td>
<td>6.2</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>AKR1C4</td>
<td>10</td>
<td>5</td>
<td>1.0</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>NBR1</td>
<td>17</td>
<td>44</td>
<td>8.8</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>PCDA12</td>
<td>5</td>
<td>18</td>
<td>3.6</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>RASSF2</td>
<td>20</td>
<td>44</td>
<td>8.8</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>HINT1</td>
<td>5</td>
<td>36</td>
<td>7.2</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>OGG1</td>
<td>3</td>
<td>74</td>
<td>14.8</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>ATAD4</td>
<td>17</td>
<td>17</td>
<td>3.4</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>CACNA2D1</td>
<td>7</td>
<td>95</td>
<td>19.0</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>C2orf19</td>
<td>20</td>
<td>72</td>
<td>14.4</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>MRPL36</td>
<td>5</td>
<td>71</td>
<td>14.2</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>COL23A1</td>
<td>5</td>
<td>51</td>
<td>10.2</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>SEM1B</td>
<td>4</td>
<td>69</td>
<td>13.8</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>FRG1</td>
<td>16</td>
<td>75</td>
<td>1.5</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>KIF22</td>
<td>22</td>
<td>101</td>
<td>20.2</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>CHK1</td>
<td>5</td>
<td>69</td>
<td>12.8</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>BCL9</td>
<td>1</td>
<td>11</td>
<td>2.2</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>MYO5A</td>
<td>15</td>
<td>150</td>
<td>30.0</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>LDB1</td>
<td>4</td>
<td>8</td>
<td>1.6</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>KCNH5</td>
<td>14</td>
<td>22</td>
<td>4.4</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>AICDA</td>
<td>12</td>
<td>2</td>
<td>0.4</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>SGMS1</td>
<td>10</td>
<td>93</td>
<td>18.6</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>ITGA2</td>
<td>5</td>
<td>65</td>
<td>13.0</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>GLT25D1</td>
<td>19</td>
<td>61</td>
<td>12.2</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>

Although the differential methylation of one gene, \textit{ATAD4}, did not validate the MeDIP-microarray analysis prediction, the difference in methylation measured between first-trimester placenta (27%) and peripheral blood (80%) was so large that it was considered to be
biologically significant. The *ATAD4* promoter region had a CpG density of 3.4% (Table 5.4), which was within the optimal range of CpG density; however, the differential methylation measured by Sequenom was opposite to the MeDIP-microarray prediction. It is unclear how Sequenom can give an opposite result to the MeDIP-microarray prediction. Nevertheless, the differential promoter methylation of *ATAD4* was large enough that it may have biological significance.

The high failure rate for candidate gene validation in this study highlights a few areas of weakness in the MeDIP-microarray assay. Firstly, the anti-5-methylcytosine antibody used in MeDIP may have immunoprecipitated methylated sequences differently between sample groups, thus leading to variation in the enrichment of methylated DNA. Although this seems unlikely, the entire MeDIP-microarray experiment relies on the immunoprecipitation step to be consistent across sample groups, and any variation in MeDIP may hide true candidate genes or recognise false candidate genes. Secondly, although samples were pooled into groups on each microarray to minimise inter-individual variation, there was only one microarray used for each gestational age group of first-trimester placenta. Although these three placenta pools were combined into one group in some of the analyses, the lack of biological replicates for any microarray experiment can result in large variations in microarray data. Variations in microarray data can easily lead to the identification of false candidate genes.

Although it is unlikely, the Sequenom method may have also contributed to the low validation rate of candidate genes. In this assay, bias can be introduced as early as the bisulfite conversion step, since the majority of bisulfite converted DNA is degraded during the conversion reaction. This can greatly reduce the number of gene copies available for PCR amplification, which can lead to PCR bias for any given sample. To correct for PCR bias, Christine Cauldrey (AgResearch Ruakura, Hamilton) suggested the pooling of PCR products prior to using them in Sequenom, thus PCR bias should not have been a limiting factor in this study. However, another limitation to the Sequenom assay is when sequences of high CpG density are examined. When sequences are CpG rich, there is a high chance that most of the cleavage fragments will contain multiple CpG sites, which can result in an inaccurate estimation of the true methylation coming from the individual CpG sites in the fragment. In this study, 13 candidate genes contained high-density CpG promoters (CpG density > 9%), all of which failed to be validated by Sequenom.
Based on the results of others who have used Sequenom to validate genes from MeDIP-microarray data, less than five percent of the differentially methylated candidate genes are validated when using this combination of techniques (Gordon Research Conference: Epigenetics, personal communications). In this study, examination of candidate gene methylation using Sequenom confirmed the differential promoter methylation between first-trimester placenta and peripheral blood for three genes: ARK1C4, KCNH5 and AICDA, giving a validation rate of 10.7% (3/28). A fourth gene, ATAD4, also displayed biologically significant differences in promoter methylation even though it failed to validate the MeDIP-microarray data analysis prediction. These four genes were selected for subsequent gene expression analysis to examine whether their differential promoter methylation regulates gene expression.
Chapter Six

Gene Expression Analysis of Differentially Methylated Genes

6.1 Gene Expression Analysis

The relationship between promoter methylation and gene expression was first reported in 1980 when Sutter and colleagues observed a striking inverse correlation between the level of DNA methylation in a specific genomic region and the amount of messenger RNA (mRNA) produced from that region (Sutter and Doerfler, 1980). It later became clear that the presence or absence of 5-methylcytosine in a gene’s promoter region could repress or permit gene expression, respectively (Razin and Riggs, 1980; Doerfler, 1981; Ehrlich and Wang, 1981; Kruczek and Doerfler, 1983). Although studies have yet to determine the exact mechanisms involved in 5-methylcytosine-induced transcriptional silencing, the biological function of DNA methylation as a repressor of gene expression is now a widely accepted phenomenon.

At this stage in the study, it was necessary to determine whether the differential promoter methylation of four candidate genes (AKR1C4, KCNH5, AICDA, and ATAD4) was inversely correlated with the expression levels of these genes. Sequenom analysis of each of these genes showed low methylation in first-trimester placenta and high methylation in peripheral blood (Chapter Five). The differences in methylation between these two tissues were large enough to be considered significant; however, in order to validate the biological significance of the differential promoter methylation, it was necessary to quantify the amount of mRNA produced by these genes. For each of these genes, it was expected that the low levels of promoter methylation in first-trimester placenta would correspond to detectable levels of gene expression, and the high promoter methylation in peripheral blood would be associated with no, or very low levels, of gene expression.

6.1.1 Quantitative Reverse-Transcription PCR using TaqMan Gene Expression Assays

For gene expression analysis, the fast and reliable method of quantitative reverse-transcription PCR (referred to from here on as RT-PCR) was selected. RT-PCR is an in vitro technique to
amplify individual sequences of mRNA from as little as a single cell (Rappolee et al., 1988). Messenger RNA, in the form of complementary DNA (cDNA), can be quantified by using PCR followed by normalisation of the data against an internal standard to determine relative amounts of mRNA (Wang et al., 1989). Given its highly quantitative characteristics, quantitative RT-PCR has become one of the most widely used methods to measure gene-specific mRNA levels in a single sample or across a group of samples (Wang et al., 1989; Orlando et al., 1998; Bustin, 2000; Bustin et al., 2005; Nolan et al., 2006).

In this study, the Applied Biosystem’s TaqMan® Gene Expression Assays were used for quantitative RT-PCR. The TaqMan method is well known for its fast, reliable and sequence-specific design to measure the abundance of amplified product during each cycle of RT-PCR (Morris et al., 1996; Livak, 1999; Guiver et al., 2000; Leutenegger et al., 2001). The Applied Biosystem’s TaqMan® Gene Expression Assay are designed to span two exons of a gene sequence. Given that some genes produce splice-variants from their first exon (personal communication, Rob Weeks, Senior Research Fellow, Morison lab), any TaqMan assay that spanned a gene’s first intron was avoided when assays were selected for this analysis. The only assay for which it was not possible to avoid the first intron was for ATAD4, since this gene has only two exons.

6.1.2 The Standard Curve for Quantitative RT-PCR

To determine gene expression levels for individual samples, a standard curve was generated from a set of cDNA standards that were included in every RT-PCR. It was critical for each gene-specific assay to include cDNA standards that were made from a tissue that had moderate to high levels of gene expression, otherwise it would be impossible to generate an accurate standard curve. In addition, the cDNA standards needed to produce Ct values that covered the range of Ct values produced in the assay. If the range of Ct values were different to that of the samples, the assay might inaccurately quantify gene expression. The linear regression line fitted to data from the cDNA standards and was then used to determine the relative amount of expressed cDNA for each sample in the assay (Equation 2.2).
6.2 Gene Expression Analysis using Quantitative RT-PCR

6.2.1 Sample Selection and RNA Extraction

Thirty-four samples were selected for gene expression analysis from the five “tissue groups” that were used in Sequenom. All samples except for peripheral blood had been included in the previous Sequenom analysis. Peripheral blood RNA was obtained from Rob Weeks (Senior Research Fellow, Morison lab). RNA was extracted from all other tissues, which are shown in Table 6.1. RNA quality was screened when concentration was determined. The number of samples in each tissue group include: 14 first-trimester placentas, six sections from one term placenta, three peripheral bloods, five adult somatic and five fetal somatic tissues.

Table 6.1. Samples selected for gene expression analysis.

<table>
<thead>
<tr>
<th>Sample #</th>
<th>Tissue Type</th>
<th>Tissue Detail</th>
<th>RNA Concentration (ng/µL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>F105</td>
<td>First-trimester placenta</td>
<td>5.5 / 40 wks</td>
<td>148.9</td>
</tr>
<tr>
<td>F107</td>
<td>First-trimester placenta</td>
<td>6 +3 / 40 wks</td>
<td>63.4</td>
</tr>
<tr>
<td>F110</td>
<td>First-trimester placenta</td>
<td>6 +2 / 40 wks</td>
<td>112.1</td>
</tr>
<tr>
<td>F116</td>
<td>First-trimester placenta</td>
<td>8 +1 / 40 wks</td>
<td>167.4</td>
</tr>
<tr>
<td>F117</td>
<td>First-trimester placenta</td>
<td>6 / 40 wks</td>
<td>92.3</td>
</tr>
<tr>
<td>F118</td>
<td>First-trimester placenta</td>
<td>5 +6 / 40 wks</td>
<td>125.4</td>
</tr>
<tr>
<td>F119</td>
<td>First-trimester placenta</td>
<td>8 +1 / 40 wks</td>
<td>139.3</td>
</tr>
<tr>
<td>F124</td>
<td>First-trimester placenta</td>
<td>5 / 40 wks</td>
<td>1883.7</td>
</tr>
<tr>
<td>F125</td>
<td>First-trimester placenta</td>
<td>5 +3 / 40 wks</td>
<td>1035</td>
</tr>
<tr>
<td>F126</td>
<td>First-trimester placenta</td>
<td>7 / 40 wks</td>
<td>149.1</td>
</tr>
<tr>
<td>F127</td>
<td>First-trimester placenta</td>
<td>6 +1 / 40 wks</td>
<td>268.8</td>
</tr>
<tr>
<td>F129</td>
<td>First-trimester placenta</td>
<td>8 / 40 wks</td>
<td>129.3</td>
</tr>
<tr>
<td>F135</td>
<td>First-trimester placenta</td>
<td>6 +4 / 40 wks</td>
<td>452.7</td>
</tr>
<tr>
<td>F138</td>
<td>First-trimester placenta</td>
<td>8 +4 / 40 wks</td>
<td>143.1</td>
</tr>
<tr>
<td>x8311</td>
<td>Umbilical cord</td>
<td>Term placenta</td>
<td>11.8</td>
</tr>
<tr>
<td>x8311</td>
<td>Maternal 1/3</td>
<td>Term placenta</td>
<td>7.3</td>
</tr>
<tr>
<td>x8311</td>
<td>Middle 1/3</td>
<td>Term placenta</td>
<td>11.1</td>
</tr>
<tr>
<td>x8311</td>
<td>Fetal 1/3</td>
<td>Term placenta</td>
<td>55.8</td>
</tr>
<tr>
<td>x8311</td>
<td>Chorion</td>
<td>Term placenta</td>
<td>15.6</td>
</tr>
<tr>
<td>x8311</td>
<td>Amnion</td>
<td>Term placenta</td>
<td>34.5</td>
</tr>
<tr>
<td>x3235</td>
<td>Peripheral blood</td>
<td>Lymphocytes</td>
<td>94.0</td>
</tr>
<tr>
<td>x3236</td>
<td>Peripheral blood</td>
<td>Lymphocytes</td>
<td>111.6</td>
</tr>
<tr>
<td>x5810</td>
<td>Peripheral blood</td>
<td>Lymphocytes</td>
<td>62.1</td>
</tr>
<tr>
<td>x351</td>
<td>Adult somatic</td>
<td>Heart</td>
<td>11.5</td>
</tr>
<tr>
<td>x351</td>
<td>Adult somatic</td>
<td>Liver</td>
<td>39.3</td>
</tr>
<tr>
<td>x272</td>
<td>Adult somatic</td>
<td>Kidney</td>
<td>37.6</td>
</tr>
<tr>
<td>H111</td>
<td>Adult somatic</td>
<td>Brain</td>
<td>59.0</td>
</tr>
<tr>
<td>x272</td>
<td>Adult somatic</td>
<td>Spleen</td>
<td>288</td>
</tr>
<tr>
<td>F67</td>
<td>Fetal somatic</td>
<td>Fetal heart</td>
<td>62.8</td>
</tr>
<tr>
<td>F129</td>
<td>Fetal somatic</td>
<td>Fetal liver</td>
<td>237.5</td>
</tr>
<tr>
<td>F77</td>
<td>Fetal somatic</td>
<td>Fetal kidney</td>
<td>198.6</td>
</tr>
<tr>
<td>F96</td>
<td>Fetal somatic</td>
<td>Fetal brain</td>
<td>259.4</td>
</tr>
<tr>
<td>F51</td>
<td>Fetal somatic</td>
<td>Fetal adrenal</td>
<td>180.8</td>
</tr>
</tbody>
</table>
6.2.2  cDNA for Quantitative RT-PCRs

6.2.2.1  Making cDNA

Complementary DNA was made from RNA as described previously (Chapter Two). In order to have enough cDNA for all RT-PCRs, 180 µL of ddH₂O was added to increase the final volume of each cDNA sample.

6.2.2.2  Testing cDNA

It was critical to test each cDNA sample before it went into an RT-PCR reaction to ensure that good quality cDNA was present in all samples. If cDNA samples were not screened, a negative result after RT-PCR may be caused by the lack of input cDNA instead of absence of gene expression for that sample. During the reverse transcription reaction for each sample, two tubes were set up; one with the reverse transcriptase enzyme and one without so that each sample would have a negative control. To test cDNA quality, end-point PCR was performed on all samples using a previously designed PCR for the beta-2-microglobulin (B2M) gene. Visualisation of a 109 bp PCR product after gel electrophoresis confirmed that the reaction with reverse transcriptase was successful, and the lack of a 109 bp band confirmed the absence of PCR product in the reaction without reverse transcriptase (Figure 6.1).
Figure 6.1. PCR screening of cDNA for the \( B2M \) gene to test the RT reaction. The name of each sample is listed above the well into which it was loaded. A positive (+) and negative (-) RT reaction was performed for each sample. Water was included as a negative control.

Varying intensities of the 109 bp product was detected in all +RT samples (Figure 6.1). Given that many samples had less than the optimal 2 \( \mu \)g of RNA put into their RT reaction, the amount of starting cDNA template available for PCR differed between samples. Differences in product intensities might also be a result of the differences in \( B2M \) mRNA
template produced by each tissue, as the amounts of gene transcripts will vary between tissues. The faint banding patterns observed in some of the −RT samples are most likely a result of genomic DNA. Since the non-template control (H2O) is negative, the bands in the −RT lanes must be derived from DNA in the RNA samples. Further, the unique banding pattern in the +RT reactions strongly suggests that these product are derived from RNA.

6.2.3 Examining Reference Genes for use in RT-PCR

Most methods of quantitative gene expression analysis include internal standards, or reference genes, which have an assumed stable expression within a cell type. The expression of a reference transcript can be used to normalise the expression of another gene transcript. These reference genes are commonly referred to as “housekeeping genes”; genes that carry out the processes required for cell survival and are therefore expected to be expressed at a constant level within a cell (Thellin et al., 1999).

6.2.3.1 Trialling Reference Genes

In this gene expression analysis, it was unclear whether a reference gene could be used, given the variety of tissue types that were to be examined. However, it seemed necessary to trial some reference genes and examine the consistency of their expression across the various tissues. Two commonly used housekeeping genes were tested: beta-2-microglobulin (B2M) and topoisomerase (DNA) III alpha (TOP3A). Both of these genes had been used as reference genes in previous expression studies in our lab group. B2M was included in the expression analysis of the Testin (TES) gene in childhood acute lymphoblastic leukaemia (Weeks et al., 2010), and TOP3A was used in gene expression studies of colorectal cancer (Anjomshoaa et al., 2008). Given that the TaqMan® Gene Expression Assays for these two genes were readily available, it was both time and cost-effective to trial these reference genes before any others.

To examine the consistency of these genes’ expression levels across tissues, quantitative RT-PCR was performed for B2M and TOP3A for all samples in triplicate. The mean C_t values for B2M and TOP3A were compared across the tissues (Figure 6.2).
Figure 6.2. Mean C<sub>t</sub> values after RT-PCR for <i>B2M</i> and <i>TOP3A</i>. The tissue samples are listed along the x-axis. The “F” tissues correspond to individual first-trimester placenta samples. The mean threshold cycle (C<sub>t</sub>) values are listed on the y-axis. The C<sub>t</sub> values for <i>B2M</i> are displayed in blue, and the C<sub>t</sub> values for <i>TOP3A</i> are displayed in red. Error bars represent the 95% confidence interval of the mean.

As shown in Figure 6.2, the mean C<sub>t</sub> values for <i>B2M</i> and <i>TOP3A</i> were not consistent across the tissues examined in this study. Although cDNA concentration varied between most samples, the samples that did have similar cDNA concentrations (from 2 µg of input RNA) did not show consistent C<sub>t</sub> values for <i>B2M</i> or <i>TOP3A</i> (Figure 6.2). For example, the first-trimester placenta samples, F124, F125, F127 and F135, were generated with same amount of input RNA, and given that they are all the same tissue type, it was thought that these samples would display similar C<sub>t</sub> values for any reference gene. However, the mean C<sub>t</sub> values for <i>B2M</i> for these samples ranged from 19 to 27, and the mean C<sub>t</sub> values for <i>TOP3A</i> for these samples ranged from 25 to 29 (Figure 6.2). Since a one-cycle difference represents a two-fold difference in the amount of generated PCR product, these differing C<sub>t</sub> values represent a large difference in the abundance of <i>B2M</i> and <i>TOP3A</i> transcripts in these first-trimester placenta samples. For example, the 8-cycle difference detected between these samples in the <i>B2M</i> PCR represents a $2^8 = 256$-fold difference in <i>B2M</i> expression. The lack of C<sub>t</sub> value consistency indicated that neither <i>B2M</i> nor <i>TOP3A</i> could be used as a reference gene in this study.

6.2.3.2 Problems with Reference Genes

The wide range of tissues in this study was the main reason for why a reference gene could not be used for normalising gene expression. However, even within a single tissue type,
Reference genes have been reported to show inconsistent expression, especially when a gross dissection of the tissue was performed (Fink et al., 2000; Sugiyama et al., 2002). Gross tissue dissection has the potential for selecting of variety of cell subpopulations within the tissue, all of which may have different gene expression profiles. Combining various cell types (and their different gene expression profiles) can lead to inaccurate quantification of gene transcript levels after quantitative RT-PCR. Although a gross dissection of first-trimester placental villi was carefully performed in this study, and all samples were screened for potential contamination from maternal tissue, there is a chance that samples may contain other cell populations.

Techniques have been developed to select for individual cell subpopulations within heterogeneous tissues, such as laser capture microdissection (LCM), a technique in which a laser selects a specific cell from a tissue without damaging the contents of the cell (Emmert-Buck et al., 1996). The LCM method has been shown to greatly improve the accuracy of gene expression analyses in laser-dissected versus gross-dissected samples (Walch et al., 2001). Although using LCM was considered at the time of sample preparation, the time and cost involved for selecting the large number of cells that were needed for DNA and RNA analysis meant that this technology was not completely necessary. Given that a wide range of gene expression profiles from the various tissues would be examined in this analysis, performing a gross dissection for all tissues was thought to pose no major problems.

Irrespective of dissection method, the use of a valid reference gene to determine gene expression currently remains to be one of the largest problems when using quantitative RT-PCR. Even some of the most commonly used reference genes have been reported to show inconsistent expression both within the same tissue type and between different tissue types (Thellin et al., 1999; Bustin, 2000; Suzuki et al., 2000; Vandesompele et al., 2002). In one study, the expression profiles of ten commonly used reference genes were examined across a variety of human tissues and none were deemed suitable to be used as a single reference gene (Vandesompele et al., 2002). Thus alternative methods for normalising expression across various tissue types have been suggested, such as using the geometric average of multiple reference genes (Vandesompele et al., 2002). However, even though multiple reference genes are used, it is unlikely that this approach can truly normalise gene expression across multiple tissue types. Another method is to normalise to total input RNA, but there is a well-documented discrepancy between the amount rRNA and mRNA in total RNA (Solanas et al.,
Due to the wide range of tissues being analysed in this study, it seemed highly unlikely that any one gene would be consistently expressed. Other references genes could have been trialled to determine a mean reference gene expression value, but the wide variety of tissues and thus gene expression profiles was likely to introduce large error in the data normalisation. Thus it was decided that using reference genes would not be appropriate for normalising gene expression in this study. This decision was reinforced by the MIQE guidelines (Minimum Information for publication of Quantitative real-time PCR experiments), which state that although using a reference gene is the most common method of normalising gene expression data, their expression must be experimentally validated across the tissue types that are being analysed (Bustin et al., 2009). In this study, the expression of B2M and TOP3A was inconsistent across tissue types and therefore these genes were not experimentally validated as reference genes. Although the expression of B2M and TOP3A were not consistent in this analysis, the results do not indicate that these genes are not useful reference genes for other studies. In fact, the mirroring patterns of mean Ct values across most of the tissues (Figure 6.2) represents that these genes are expressed in similar proportions to each other within the tissues.

6.2.3.3 Avoiding Reference Genes: Normalising to Total Input RNA

Given the wide range of tissue types, it was decided that relative gene expression would be normalised to the amount of total input RNA. The amount of input RNA was the amount of total RNA that went into the RT reaction to generate cDNA. This has been reported by some to be an accurate normalisation method, especially when two or more housekeeping genes do not show consistent expression in multiple samples (Bustin, 2002). As noted above, however, rRNA which is the most abundant RNA is a relatively poor surrogate measure of mRNA (Vandesompele et al., 2002). Once relative gene expression was normalised to input RNA, the normalised expression for first-trimester placenta samples and peripheral blood samples were averaged to determine the mean normalised expression value for these two tissue groups. Given that the gene expression profiles would vary for all of the other tissues in the study, the normalised expression for the six sections of term placenta and all adult and fetal somatic tissues were assessed individually.
6.3 Gene Expression Results for Differentially Methylated Genes

6.3.1 Relative Gene Expression for AKR1C4

The aldo-keto reductase family 1, member C4 (AKR1C4) gene was selected for gene expression analysis because of the significant difference in promoter methylation between first-trimester placenta (34%) and peripheral blood (85%) (Figure 5.5B). The AKR1C4 gene encodes an enzyme that catalyzes the bioreduction of chlordecone in the liver (Entrez Gene, NCBI). Based on the known expression of AKR1C4 in liver, the adult liver and fetal liver samples were trialled as the cDNA standards for this assay. The fetal liver sample produced the suitable range of C\(_t\) values for this assay; therefore this sample was chosen to make the cDNA standards and to generate a standard curve. Once relative gene expression was determined and AKR1C4 expression was normalised to total input RNA, gene expression was compared to promoter methylation for each tissue in the assay (Figure 6.3).
It was expected that AKR1C4 expression would be inversely correlated with the levels of promoter methylation. As seen in Figure 6.3, AKR1C4 was not expressed in first-trimester placenta, a tissue with low promoter methylation. AKR1C4 expression was detected in peripheral blood, a tissue with high levels of promoter methylation. The inverse correlation between promoter methylation and gene expression was thus not demonstrated by either first-trimester placenta or peripheral blood. In addition, AKR1C4 expression in fetal liver and fetal adrenal tissue corresponded to moderate and high levels of promoter methylation, respectively, which again did not confirm an inverse relationship between promoter methylation and gene expression. The five samples that displayed no promoter methylation...
data (term placenta fetal 1/3 and chorion, adult liver, fetal kidney and fetal brain) could not be assessed since their methylation data was removed during Sequenom analysis.

The CpG methylation in a promoter region that contains a CpG island is more likely to be involved in regulating gene transcription than the CpG methylation in a promoter that does not contain a CpG island (Song et al., 2005; Eckhardt et al., 2006). Although the degree of promoter methylation that is required to silence gene expression is unknown, the absence of a CpG island in a gene’s promoter region suggests that promoter CpG methylation may not be regulating gene transcription. Thus the methylation of the two CpG sites that were included in the Sequenom analysis for AKR1C4 (Figure 5.5A) most likely does not regulate the expression of this gene. Furthermore, the Genome Browser shows no alternative start sites for AKR1C4, so it is not likely for an alternative promoter to be regulating the expression of this gene (Kent et al., 2002). Overall, given that the expression patterns of AKR1C4 displayed by most tissues did not confirm the predicted inverse relationship between promoter methylation and gene expression, AKR1C4 was excluded from further analysis.

### 6.3.2 Relative Gene Expression for AICDA

The activation-induced cytidine deaminase (AICDA) gene was selected for gene expression analysis because of the significant difference in promoter methylation between first-trimester placenta (37%) and peripheral blood (86%) (Figure 5.25B). Given that AICDA encodes a RNA-editing deaminase protein that is involved in many cell functions (Entrez Gene, NCBI), it was unknown which tissue would make the most appropriate cDNA standards for this expression assay. After trialling a variety of tissues, fetal liver was found to be the highest expressing sample and produced the suitable range of Ct values for this assay; therefore this sample was used to make the cDNA standards and generate a standard curve. Once relative gene expression was determined and AICDA expression was normalised to total input RNA, gene expression was compared to promoter methylation for each tissue (Figure 6.4).
Figure 6.4. Expression of AICDA normalised to input RNA compared to mean promoter methylation. All tissues are listed above the shared x-axis. A. Gene expression normalised to input RNA is displayed in blue. This y-axis represents normalised gene expression using a log scale to display the range of values. B. The mean promoter methylation (quantified by Sequenom) is displayed in red. The y-axis represents mean promoter methylation (%).

Once again, it was expected that patterns of AICDA expression would correspond inversely with patterns of AICDA promoter methylation. As seen in Figure 6.4, AICDA expression was detected in first-trimester placenta, a tissue with low levels of promoter methylation. Higher levels of AICDA were expressed in peripheral blood, a tissue with high promoter methylation, thus the inverse relationship was not confirmed. In most term placental tissues and all fetal
somatic tissues, AICDA expression was detected although moderate to high levels of promoter methylation was present in these tissues (Figure 6.4).

As discussed previously, there is no CpG island in the promoter of AICDA; thus the methylation of the two CpG sites that were analysed by Sequenom (Figure 5.25A) most likely does not regulate the transcription of this gene. There is also no evidence for an alternative start site for AICDA on the Genome Browser (Kent et al., 2002), suggesting that there is not likely to be an alternative promoter regulating the expression of this gene. Given that the expression patterns of AICDA in most tissues did not confirm the predicted inverse relationship with promoter methylation, AICDA was excluded from further analysis.

### 6.3.3 Relative Gene Expression for ATAD4

The proline rich 15-like (PRR15L or ATAD4) gene was selected for gene expression analysis because of the significant difference in promoter methylation between first-trimester placenta (27%) and peripheral blood (80%) (Figure 5.11B). Given that ATAD4 encodes an ATPase protein that is supposedly present in all cells (Entrez Gene, NCBI), it was unknown which tissue would be most appropriate to use for the cDNA standards. Fetal liver was chosen to make the cDNA standards as it produced a suitable range of C\textsubscript{T} values for this assay. Once relative gene expression was determined and ATAD4 expression was normalised to total input RNA, gene expression was compared to promoter methylation levels in each tissue (Figure 6.5).
expression normalised to input RNA is displayed in blue. This y-axis represents normalised gene expression using a log scale to display the range of values. B. The mean promoter methylation (quantified by Sequenom) is displayed in red. The y-axis represents mean promoter methylation (%).

Figure 6.5. Expression of ATAD4 normalised to input RNA compared to mean promoter methylation. All tissues are listed above the shared x-axis. A. Gene expression normalised to input RNA is displayed in blue. This y-axis represents normalised gene expression using a log scale to display the range of values. B. The mean promoter methylation (quantified by Sequenom) is displayed in red. The y-axis represents mean promoter methylation (%).

The expression of ATAD4 was predicted to display an inverse correlation with promoter methylation. As seen in Figure 6.5, expression of ATAD4 was not detected in first-trimester placenta, a tissue with low promoter methylation. ATAD4 expression was not present in peripheral blood, a tissue with high levels of promoter methylation, thus an inverse correlation between methylation and expression was observed in this tissue. Fetal heart, fetal liver and fetal kidney all displayed ATAD4 expression that inversely correlated with low
promoter methylation. However, the highest expressing sample, adult heart, correlated with moderate levels of ATAD4 methylation, a result that did not show an inverse relationship.

The presence of CpG islands and alternative promoters was also investigated for ATAD4. As was the case for the previous two genes, there is no CpG island present in the promoter of ATAD4; thus the methylation of the six CpG sites that were analysed by Sequenom (Figure 5.11A) most likely does not regulate the transcription of this gene. There is also no evidence for an alternative start site for this gene on the Genome Browser (Kent et al., 2002), suggesting that there is not likely to be an alternative promoter regulating ATAD4 expression. Given that a selection of tissues, especially first-trimester placenta, did not display the inverse relationship between promoter methylation and gene expression, ATAD4 was excluded from this study.

6.3.4 Relative Gene Expression for KCNH5

The potassium voltage-gated channel, subfamily H (eag-related), member 5 (KCNH5) gene was selected for gene expression analysis because of the significant difference in promoter methylation between first-trimester placenta (15%) and peripheral blood (86%) (Figure 5.24B). Given that KCNH5 encodes a protein that is highly expressed in brain (Entrez Gene, NCBI), the fetal brain sample was used to create a set of cDNA standards that produced a suitable range of Ct values for this assay. Once relative gene expression was determined and KCNH5 expression was normalised to total input RNA, gene expression was compared to promoter methylation levels in each tissue (Figure 6.6).
The results from the expression analysis of *KCNH5* were exciting. When promoter methylation was plotted for each sample, a striking dichotomy in promoter methylation was observed: samples derived from the extra-embryonic lineages (placenta and chorion) displayed very low levels of promoter methylation, whereas samples derived from the embryonic lineage (amnion, umbilical cord and all adult somatic and fetal somatic tissues) had very high levels of methylation (Figure 6.6). The placenta and chorion are derived from the trophectoderm and primitive endoderm (extra-embryonic lineages), while the amnion, umbilical cord and all adult and fetal somatic tissues are derived from the epiblast (embryonic
lineage) (Benirschke et al., 2006). The large difference in methylation of the adjacent and adherent chorion (13%) and amnion (86%) membranes also highlighted the lineage-specificity of KCNH5 promoter methylation.

It was expected that the patterns of KCNH5 expression would inversely mimic the dichotomous patterns of promoter methylation in these tissues. As seen in Figure 6.6, KCNH5 expression was detected in first-trimester placenta and two section of term placenta, which confirmed an inverse relationship with the low level of promoter methylation in tissues derived from the extra-embryonic lineages. However, variable levels of KCNH5 expression were detected in the highly methylated epiblast-derived tissues, displayed by amnion, adult brain, and fetal brain, heart, kidney and adrenal (Figure 6.6). It was expected that tissues with very high levels of promoter methylation would not express KCNH5.

Given this lineage-specific contrast in promoter methylation, the observation that the highly methylated epiblast-derived tissues expressed KCNH5 suggested that the promoter region examined by Sequenom might not be regulating KCNH5 expression in these tissues. According to the Genome Brower, there is an alternative start site for KCNH5 (Kent et al., 2002). It was thus hypothesised that KCNH5 expression may be regulated by this alternative promoter in tissues derived from the embryonic (epiblast) lineage. The methylation of the 16 CpG sites analysed by Sequenom (Figure 5.24A) may therefore be regulating KCNH5 transcription in tissues derived from the extra-embryonic lineages, but not in tissues derived from the epiblast. Further experiments were required to determine if an alternative promoter was in fact regulating KCNH5 expression in epiblast-derived tissues (Chapter Seven).

### 6.4 Analysis of a Closely-Related Gene: KCNH1

Given the intriguing patterns of KCNH5 methylation and expression, it was of interest to learn more about the potential function of KCNH5 in the developing placenta. At the time of this study, there was a paucity of literature documenting the function of KCNH5, and no literature reporting on KCNH5 methylation or expression in the placenta. However, a closely related potassium channel gene with well-reported function provided a platform from which to speculate about the possible roles of KCNH5 in the placenta. The potassium voltage-gated channel, subfamily H (eag-related), member 1 (KCNH1) gene, is roughly 70% homologous to KCNH5 (Ludwig et al., 2000; Bauer and Schwarz, 2001). Like KCNH5, the KCNH1 gene
also encodes a pore-forming (alpha) subunit of a voltage-gated potassium channel that is highly expressed in the brain (Ju and Wray, 2002).

Interestingly, KCNH1 (also known as Eagl) has been associated with tumour development in humans and animals (Hemmerlein et al., 2006; Ding et al., 2007). Hemmerlein and colleagues found that KCNH1 was over-expressed in cancerous tumours, and that decreasing KCNH1 expression in cancerous cell lines resulted in reduced tumour cell proliferation (Hemmerlein et al., 2006). Ding and colleagues reported that KCNH1 was aberrantly expressed in colorectal cancer and may be a useful biomarker for diagnosing the disease (Ding et al., 2007). Although the function of KCNH1 is better understood than that of KCNH5, both of these genes are thought to be involved in cell cycle regulation and tumour progression in cancer (Pardo et al., 1999; Wadhwa et al., 2009).

Since the processes involved in placental development share many of the molecular mechanisms involved in tumourigenesis, the reported role of KCNH1 in tumour development supported a hypothesis that KCNH5 (and potentially KCNH1) may have a regulatory role involved in placental development. At this stage in the project, it seemed necessary to examine KCNH1 in first-trimester placenta to explore any similarities that this gene may share with the epigenetic and expression profiles of KCNH5. If the patterns of promoter methylation and gene expression of these genes were similar, it may be likely that these genes also share functional roles in the developing placenta, and possibly in cancer.

6.4.1 Sequenom Analysis of KCNH1

The promoter methylation of KCNH1 was quantified using Sequenom. The following samples were analysed: 17 first-trimester placentas, six separate regions of one term placenta, 9 peripheral bloods, 5 adult somatic tissues (heart, liver, kidney, brain and spleen) and 5 fetal somatic tissues (heart, liver, kidney, brain and adrenal). To quantify the promoter methylation for KCNH1, a 331 bp product containing 43 CpG sites was analysed by Sequenom (Figure 6.7A). Five CpG sites were excluded from analysis; one CpG site was located on a fragment of low mass, three CpG sites were located on fragments of high mass and one CpG site gave no data in ≥30% of samples. The mean promoter methylation for KCNH1 was averaged within the tissue groups of first-trimester placenta and peripheral blood, whereas the values were displayed individually for all other samples to examine whether lineage-specific methylation was present at the promoter of KCNH1 (Figure 6.7B).
Figure 6.7. Sequenom results for KCNH1. A. The region of KCNH1 analysed by Sequenom. Forward and reverse primers are shown in yellow. The region of methylation quantified by Sequenom is shown in grey. Red circles represent CpG sites and the green box represents the first gene exon. B. The mean promoter methylation data for all samples. Error bars represent the 95% confidence interval of the mean for the tissue groups of first-trimester placenta (n = 17) and peripheral blood (n = 9).

As shown in Figure 6.7B, the mean promoter methylation for KCNH1 was very low in all samples. There was no significant difference in methylation between any of the tissues.

6.4.2 Gene Expression Analysis for KCNH1

The low level of promoter methylation in all tissues was expected to correlate inversely with the patterns of expression for KCNH1. Gene expression analysis was performed for KCNH1 using a TaqMan gene expression assay, and gene expression was normalised to total input RNA. Given that KCNH1 is highly expressed in brain, the fetal brain sample was used to create a set of cDNA standards for this assay. Once relative gene expression was determined and KCNH1 expression was normalised to total input RNA, gene expression was compared to promoter methylation levels in the various tissue groups. The patterns of gene expression compared to promoter methylation for KCNH1 are displayed (Figure 6.8).
As seen in Figure 6.8, KCNH1 was highly expressed in both adult and fetal brain samples, which confirmed the previously reported expression of this gene in brain tissue. The low levels of promoter methylation in tissues that expressed KCNH1 suggest that the absence of methylation in these tissues is permissive of gene expression. However, the absence of hypermethylation in the non-expressing tissues suggests that promoter methylation does not regulate KCNH1 expression in these tissues. KCNH1 displayed no expression in most of the
placental tissues (except chorion). Given that \( KCNH1 \) expression was not similar to \( KCNH5 \) expression in the placenta, \( KCNH1 \) became a less interesting gene to pursue.

6.4.3 \textit{Sequenom Analysis of KCNH1 and KCNH5 in Cancerous Tissues}

One of the main goals of this study was to identify differentially methylated genes that may be regulating the similar behaviours of trophoblast and tumour cells. Given the suggested functional roles of \( KCNH1 \) and \( KCNH5 \) in cancer (Pardo \textit{et al.}, 1999; Wadhwa \textit{et al.}, 2009), it was of interest to examine the promoter methylation and expression of these two genes in a selection of cancerous tissues. DNA and RNA were obtained from five cancerous tissue samples that were provided by the Cancer Genetics Lab (University of Otago). The five tissue samples included colon cancer, pancreatic cancer, breast cancer, acute lymphoblastic leukaemia (ALL) and Wilms tumour (childhood kidney cancer). First, mean promoter methylation of \( KCNH1 \) and \( KCNH5 \) were compared in the five cancer samples (Figure 6.9).

![Figure 6.9. Sequenom analysis for KCNH1 and KCNH5 in cancerous tissues. Mean promoter methylation was determined for each sample and displayed in red for KCNH1 and in pink for KCNH5.](image)

As shown in Figure 6.9, the promoter methylation of \( KCNH1 \) was low in all cancerous tissues except for colon cancer (51%). The low levels of \( KCNH1 \) promoter methylation may be associated with the up-regulated patterns of \( KCNH1 \) expression that was reported in some
cancers (Hemmerlein et al., 2006; Ding et al., 2007). The mean promoter methylation for 
*KCNH5*, however, was significantly higher than *KCNH1* in all of the cancerous tissues 
(Figure 6.9). The high levels of *KCNH5* methylation suggested that *KCNH5* might not be 
expressed in the cancerous tissues.

### 6.4.4 Gene Expression Analysis of *KCNH1* and *KCNH5* in Cancerous Tissues

The promoter methylation and normalised expression for both *KCNH1* and *KCNH5* were 
compared in the five cancerous tissue samples (Figure 6.10).

**Figure 6.10.** *KCNH1* and *KCNH5* expression and promoter methylation in cancerous tissues. Sample 
names are listed above the x-axis. In both graphs, the top y-axis represents normalised gene expression 
using a log scale to display the range of values. The bottom y-axis represents mean promoter methylation 
(%) that was quantified by Sequenom. **A.** Results for *KCNH1*. Gene expression normalised to input 
RNA is displayed in blue and promoter methylation is displayed in red. **B.** Results for *KCNH5*. Gene 
expression normalised to input RNA is displayed in striped blue and promoter methylation is displayed in 
striped red.

As seen in Figure 6.10A, expression of *KCNH1* was detected in all cancerous samples except 
in acute lymphoblastic leukaemia (ALL). *KCNH1* expression displayed an inverse correlation 
to the promoter methylation, but only in three of the five cancerous tissues: pancreatic cancer, 
breast cancer and Wilms tumour (Figure 6.10A). It is unexpected that colon cancer would
show both expression and promoter methylation. One explanation for the lack of an inverse correlation in the colon cancer sample may be that this tissue is highly heterogeneous; thus a variety of cell subpopulations in these tissues may express KCNH1 differently. In order to accurately quantify promoter methylation and gene expression in a heterogeneous tissue sample, the various cell subpopulations would need to be isolated; otherwise the levels of methylation and expression represent the average of expressing and non-expressing cells. Importantly, there are other factors that may also be regulating KCNH1 expression, such as histone modifications, the availability of gene transcription factors and silencing microRNAs. However, in the context of this PhD study, the extent to which promoter methylation regulates KCNH1 expression remains unclear.

In contrast to KCNH1, expression of KCNH5 was rarely detected in the cancerous tissues (Figure 6.10B). The high levels of promoter methylation in the non-expressing samples may be silencing KCNH5 expression in these tissues. Expression of KCNH5 was observed at low levels in colon cancer, and at slightly higher levels in Wilms tumour (Figure 6.10B). In both of these tissues, the high level of KCNH5 promoter methylation did not equate to an inverse correlation with KCNH5 expression. The variable levels of KCNH5 expression despite high levels of promoter methylation are similar to the results from the previous KCNH5 analysis, suggesting that there may be an alternative promoter regulating the expression of KCNH5 in these epiblast-derived cancerous tissues.
Chapter Seven

Characterisation of \textit{KCNH5}: Promoters, Transcripts and Proteins

7.1 Alternative Transcripts of \textit{KCNH5}

Although the function of \textit{KCNH5} in the placenta has yet to be determined, the intriguing results from the promoter methylation and expression analyses prompted further investigation into the transcriptional regulation of \textit{KCNH5}. The lineage-specific patterns of promoter methylation combined with the variable patterns of \textit{KCNH5} expression in epiblast-derived tissues suggested that an alternative promoter might be regulating the expression of \textit{KCNH5} in these tissues. If an alternative promoter were being used in the epiblast-derived tissues, then the region of methylation that was quantified by Sequenom might not correspond to the expression values, which may explain the lack of an association between \textit{KCNH5} promoter methylation and expression that is displayed in Chapter Six (Figure 6.6).

Although there is no literature documenting the lineage-specific methylation and expression of \textit{KCNH5}, the University of California Santa Cruz (UCSC) Genome Browser shows that alternative splicing of \textit{KCNH5} gives rise to various gene transcripts (Kent \textit{et al.}, 2002). A major difference between the displayed transcripts for \textit{KCNH5} is that they use alternative first exons, suggesting that alternative promoters may be regulating gene transcription at these different start sites. Thus it was hypothesized that the promoter regulating \textit{KCNH5} expression in tissues derived from the extra-embryonic lineages (placenta and chorion) may not be regulating the expression of \textit{KCNH5} in tissues derived from the epiblast. To test this hypothesis, RT-PCR was used to test tissues for \textit{KCNH5} transcript-specific expression.

7.1.1 Transcript-Specific RT-PCRs

7.1.1.1 Selection of Alternative \textit{KCNH5} Transcripts

Six human mRNA sequences for \textit{KCNH5} are available from NCBI’s GenBank (Benson \textit{et al.}, 2009). With respect to their transcription start sites, these transcripts can be classified into two groups of transcripts, which are simplified to transcript “1a” and “1b” in the following
sections. The genomic arrangement of these transcripts’ alternative first exons is displayed in Figure 7.1. The region of dichotomous methylation that was quantified by Sequenom is located upstream and into the beginning of exon 1a, and the alignment of these transcripts starts at exon two (Figure 7.1).

**Figure 7.1.** Genomic arrangement of KCNH5’s alternative transcripts. Coloured boxes represent gene exons. The region of methylation that was quantified by Sequenom is displayed by the red lollipops.

### 7.1.1.2 Design and Optimisation of Transcript-Specific RT-PCRs

Two end-point RT-PCRs were designed so that the beginning of transcript 1a could be distinguished from the beginning of transcript 1b in tissues that expressed KCNH5. Primer design for the transcript-specific RT-PCRs is displayed in Figure 7.2. The forward primers were designed within the unique first exon of each transcript and a common reverse primer was designed within exon four (Figure 7.2).

**Figure 7.2.** Primer design to identify expression of KCNH5’s alternative transcripts. Coloured boxes represent gene exons. The region of methylation that was quantified by Sequenom is displayed by the red lollipops. The forward primers are shown in blue and orange and the common reverse primer is shown in purple.
Optimisation of the transcript-specific RT-PCRs was performed using samples that were known to express *KCNH5*. Since the expression of these transcripts was hypothesised to be lineage-specific, tissues derived from an extra-embryonic lineage (first-trimester placenta, F124) and the embryonic (epiblast) lineage (adult brain, H111) were used in PCR optimisation.

### 7.1.1.3 Results from the Transcript-Specific RT-PCRs

Once optimised, the two end-point RT-PCRs were performed on samples that showed significant *KCNH5* expression in the previous real-time TaqMan expression assay described in Chapter Six (Figure 6.6). The samples included in these transcript-specific PCRs included five first-trimester placentas (F110, F116, F117, F124 and F125), amnion (term), fetal brain, fetal adrenal, fetal kidney, fetal heart and adult brain. Products were visualised on a 2% agarose gel after electrophoresis, shown in Figure 7.3. The presence of a 423 bp product confirmed expression of transcript 1a and the presence of a 369 bp product confirmed expression of transcript 1b.

![Figure 7.3](image-url) **Figure 7.3.** Results from the amplification of cDNA for two transcripts of *KCNH5*. The top gel contains PCR products from the transcript 1a PCR and the bottom gel contains PCR products from the transcript 1b PCR.

As seen in Figure 7.3, transcript-specific RT-PCR showed exclusive expression from exon 1a in first-trimester placenta, and exclusive expression from exon 1b in amnion and all somatic tissues. These results confirmed that the lineage-specific promoter methylation of *KCNH5* regulates the lineage-specific expression of alternative transcripts 1a and 1b. There will be discussion about the potential models for the transcriptional regulation of *KCNH5* on pages 177 and 178. At this stage in the project, it seemed necessary to examine whether a reciprocal
pattern of methylation was present between the promoters regulating transcription at exon 1a and exon 1b.

7.1.2 Sequenom Analysis for the Promoter of KCNH5 Transcript 1b

The promoter methylation of transcript 1b was quantified using Sequenom. The same samples that were included in previous Sequenom runs were also examined in this analysis. Given the large number of CpG sites located upstream of exon 1b, two sets of PCR primers were designed to span the promoter region of transcript 1b. Primers were designed using the same Sequenom primer parameters described in Chapter Two. The first PCR analysed a 382 bp product containing 34 CpG sites and the second PCR analysed a 445 bp product containing 38 CpG sites was (Figure 7.4A). A total of 24 CpG sites were excluded from the analysis of both amplicons; 10 CpG sites were located on fragments of high mass, four CpG sites were located on fragments of low mass and ten CpG sites gave no data in ≥ 30% of samples.

The mean promoter methylation values from these two amplicons were averaged to determine the mean promoter methylation for the 801 bp region. During data analysis, the two sections of term placenta (maternal 1/3 and middle 1/3) were removed from analysis since they gave no data in ≥ 30% of the analysed CpG sites. The mean promoter methylation of KCNH5 transcript 1b was averaged for the groups of first-trimester placenta and peripheral blood, whilst all other samples were kept separate in order to examine their individual methylation levels (Figure 7.4B).
**Figure 7.4.** Sequenom results for KCNH5 transcript 1b. **A.** The region of KCNH5 transcript 1b analysed by Sequenom. Forward and reverse primers are shown in yellow for the first PCR and in pink for the second PCR. The region of methylation quantified by Sequenom is shown in grey. Red circles represent CpG sites and the green box represents exon 1b. **B.** The mean promoter methylation data for all samples. Error bars represent the 95% confidence interval of the mean for the tissue groups of first-trimester placenta (n = 17) and peripheral blood (n = 9).

As shown in Figure 7.4B, the reciprocal promoter methylation dichotomy was not displayed between KCNH5 transcript 1a and 1b. The mean promoter methylation of transcript 1b was very low in all samples and there was no pattern of lineage-specific methylation (Figure 7.4B). Nevertheless, the fact that all epiblast-derived tissues displayed low levels of promoter methylation for transcript 1b confirmed that this unmethylated promoter permits transcription at exon 1b in these tissues.

The results from this Sequenom analysis suggested two potential models to explain why transcription occurs at exon 1b in epiblast-derived tissues. One model is that transcriptional regulation of KCNH5 is entirely epigenetic, so the high level of transcript 1a promoter methylation prevents transcription at exon 1a in epiblast-derived tissues. This model would suggest that in tissues derived from the extra-embryonic lineages, lack of methylation of this promoter permits transcription of transcript 1a, whereas in tissues derived the epiblast,
transcription occurs from exon 1b by default. A second model predicts that there are exon 1a- and 1b-specific transcription factors that are present or absent in the various tissues. This would suggest that exon 1a-specific transcription factors are expressed exclusively in the placenta, and exon 1b-specific transcription factors are present in only epiblast-derived tissues. It is more likely that the epigenetic model explains KCNH5 transcription in epiblast-derived tissues, given that transcript 1a promoter methylation would prevent transcription and the downstream exon (1b) would be the next initiation site available for transcription. However, the mechanisms involved in regulating KCNH5 transcription remain unknown.
7.2 KCNH5 Protein Analysis by Western Blot

Returning to the original hypothesis that differentially methylated genes may be important in placental development and function, it was necessary to perform an initial characterisation of the protein encoded by KCNH5; an ether-à-go-go voltage-gated potassium channel. The locus of this gene was first discovered in Drosophila melanogaster when mutant flies displayed a rhythmic leg-shaking behaviour when put under anaesthesia. The gene locus associated with this behaviour was subsequently named ether-à-go-go (EAG) (Kaplan and Trout, 1969). Further homology screening identified two similar sequences, the EAG-related (ERG) gene and EAG-like (ELK) gene; thus giving rise to three subfamilies of ether-à-go-go potassium channel genes: EAG, ERG and ELK (Warmke and Ganetzky, 1994). The EAG subfamily is made up of two closely related voltage-gated potassium channel genes that were initially discovered in rat: Eag1 (KCNHI) (Ludwig et al., 1994) and Eag2 (KCNH5) (Saganich et al., 1999).

Voltage-gated potassium channels are involved in the regulation of cell cycle and proliferation (Bruggemann et al., 1997; Pardo et al., 1998). These ion channels play a central role in controlling the activity of excitable cells, such as regulating the action potential, membrane potential and frequency of cell firing (Ganetzky and Wu, 1983; Yellen, 2002; Jeng et al., 2005). However, voltage-gated potassium channels also have important functions in non-excitable cells, where they regulate intracellular calcium concentration and hence control cell growth and differentiation (Deutsch and Chen, 1993). The role of potassium channels in cell-to-cell interactions suggests that these ion channels are also involved in cell migration and invasive growth (Schwab, 2001). Furthermore, decreased expression of voltage-gated potassium channel genes is reported to interrupt cell growth and proliferation in some invasive cancers (Wonderlin and Strobl, 1996), whereas over-expression has been shown to transform a normal cell into a pre-cancerous cell (Pardo et al., 1999; Wadhwa et al., 2009).

At this stage in the study, it was necessary to examine whether transcripts 1a and 1b of KCNH5 encode different isoforms of KCNH5 protein. An analysis of KCNH5 protein would provide additional confirmation that these transcripts are expressed lineage-specifically. In addition, any differences detected in KCNH5 protein may suggest a functional difference for KCNH5 in the tissues. Although a possible functional role for KCNH5 protein encoded by
transcript 1b has been reported in the literature, the function of transcript 1a in the placenta is unknown.

To initially characterise KCNH5 protein, the amino acid sequence encoded by each of the six human mRNA sequences for KCNH5 was obtained from NCBI’s GenBank (Benson et al., 2009). Two of these mRNA sequences begin translation at the non-coding exon 1a, whereas four mRNA sequences are translated using the protein-coding exon 1b, which encodes 58 amino acids. Once the amino acid sequences were obtained, the MacVector program was then used to predict the molecular weight (in kilodaltons, kDa) of each protein isoform for KCNH5 (MacVector, 2010). For simplicity, the isoforms derived from the mRNA sequences that use the non-coding exon 1a are termed “isoform 1a1” and “isoform 1a2”, whereas the isoforms derived from the mRNA sequences that use the protein-coding exon 1b are termed “isoform 1b1” and “isoform 1b2”. The mRNA sequences and the isoforms they encode are displayed in Figure 7.5.

![Figure 7.5](image)

**Figure 7.5.** The six human mRNA sequences for KCNH5 and the protein isoforms they encode. Coloured boxes represent coding exons. A. Isoform 1a1. B. Isoform 1a2 has a premature stop 46 amino acids before the end of exon seven. C. Isoform 1b1. D. Isoform 1b2 encoded by three human mRNA sequences.

There are significant differences between the predicted isoforms of KCNH5 protein. It is important to note, however, that these protein predictions are based on a very small number of mRNAs. Isoforms 1a1 and 1a2 have lower molecular weights than isoforms 1b1 and 1b2 due to their use of the non-coding exon 1a and putative termination of translation at exon seven (Figure 7.5A and 7.5B). Further, in isoform 1a2, a premature stop is predicted to terminate
translation 46 amino acids before the end of exon seven. It is unclear why the reported transcripts that use exon 1a terminate translation at or before the end of exon seven. These transcripts suggest the presence of placental-specific factors that induce early termination of translation. Isoforms 1b₁ and 1b₂ terminate translation at exon nine and eleven, respectively, therefore these isoforms have higher molecular weights (Figure 7.5C and 7.5D). Given the lineage-specific use of exon 1a and exon 1b in KCNH5 transcripts, it was expected that either isoform 1a₁ or 1a₂ would be detected in placenta, whereas either isoform 1b₁ or 1b₂ would be present in epiblast-derived tissues.

A western blot was used to determine which tissues expressed which isoform of KCNH5. The western blot is an electrophoretic system to transfer proteins that were separated by size in a polyacrylamide gel to a nitrocellulose membrane, followed by probing the membrane with a primary and secondary antibody to detect the protein of interest (Renart et al., 1979; Towbin et al., 1979; Burnette, 1981). The proteins were first separated by size in the polyacrylamide gel. Although this technique is not quantitative, western blotting is one of the most commonly used techniques to determine the molecular weight of a protein of interest.

Three independent studies have previously performed a western blot analysis for KCNH5 protein (Ludwig et al., 2000; Jeng et al., 2005; Wadhwa et al., 2009). Each of these studies localised KCNH5 protein in tissues derived from the epiblast and detected protein that ranged from 100 kDa to 120 kDa. Taking into account post-translational modification of protein and experimental variation, these results confirmed the expression of KCNH5 isoform 1b₂ in epiblast-derived tissues, which has a predicted molecular weight of 112 kDa (Figure 7.5D). At present, there is no reported study that has performed a western blot for KCNH5 protein in the human placenta.

7.2.1 Experimental Design for Western Blot

7.2.1.1 Selection of Antibody for KCNH5

A primary antibody was selected that would detect all isoforms of KCNH5 (Abcam, Cat #ab32975). This rabbit polyclonal antibody was raised against a synthetic peptide derived from amino acids #156-205 of KCNH5, an amino acid sequence that is present in all isoforms of KCNH5.
7.2.1.2 Selection of Samples

A selection of first-trimester placenta and adult brain samples that express high levels of KCNH5 were selected for the western blot. The antibody company (AbCam) suggested using the human liver carcinoma cell line, HepG2, as the positive control tissue. HepG2 cells were obtained from Rob Weeks (Senior Research Fellow, Morison Lab). The negative control tissue, adult primary melanocytes, was chosen after screening multiple cell lines and tissues by western blot. Cell lysate from adult primary melanocytes was kindly donated by the Developmental Genetics Group (University of Otago).

7.2.1.3 Screening samples for western blot

It was important to confirm KCNH5 expression in the samples for the western blot. This was especially necessary for the positive and negative control tissues, as their levels of KCNH5 expression had not yet been quantified. The control tissues, HepG2 cells (positive) and adult primary melanocytes (negative), thus needed their “control” status validated for the western blot.

RT-PCR was performed using the TaqMan® Gene Expression Assays to detect expression of KCNH5 and B2M. The B2M gene, a commonly used reference gene, was included in this experiment to verify the RT-PCR results for KCNH5. Given that B2M expression is expected in most tissues, presence of a B2M product after RT-PCR confirmed that cDNA was of adequate quality. This would confirm that any negative result in the KCNH5 assay was valid.

RT-PCR for KCNH5 and B2M was performed on cDNA from the tissues that were included in the western blot: first-trimester placenta (F116), adult brain (H111), HepG2 cells (positive control) and primary adult melanocytes (negative control). A cDNA sample from yeast was included as a non-template control. A negative control for the RT reaction (-RT) was also included, which represents an RNA sample that was not treated with the reverse transcriptase enzyme during cDNA synthesis, thus no template should be present in this sample for PCR amplification. The fetal brain sample (F96) was used to make a set of cDNA standards and generate a standard curve. After RT-PCR, the relative expression of KCNH5 and B2M was determined for each sample using the standard curve, and relative gene expression was normalised to total input RNA using the calculations shown in Chapter Two (Equation 2.2). The expression of B2M and KCNH5 normalised to input RNA is shown in Figure 7.6.
Figure 7.6. Screening samples for the western blot by quantitative RT-PCR. Tissues are listed on the x-axis. The y-axis represents normalised gene expression using a log scale to display the range of values. Error bars represent the 95% confidence interval of the mean C\textsubscript{T} value that was used to determine relative gene expression (samples were screened in triplicates). A. Gene expression normalised to input RNA for \textit{B2M}. B. Gene expression normalised to input RNA for \textit{KCNH5}.

As seen in Figure 7.6A, expression of \textit{B2M} was detected in all tissues. The levels of \textit{B2M} expression ranged widely between tissues, yet this was not surprising as this same result had been observed when a variety of tissues were screened for \textit{B2M} expression in Chapter Six. The presence \textit{B2M} product in adult primary melanocytes was the most important result, as it confirmed that the absence of \textit{KCNH5} expression in this tissue was not attributed to poor quality cDNA (Figure 7.6B). Primary adult melanocytes was therefore a suitable negative control tissue for the western blot. It was surprising that HepG2 cells, the suggested positive
control tissue, expressed lower levels of KCNH5 than first-trimester placenta and brain (Figure 7.6B). However, the fact that expression was detected in HepG2 cells confirmed that this tissue could be used as the positive control in the western blot.

7.2.1.4 Protein Isolation and the BCA Assay

Protein was isolated from the selected tissues for the western blot using the method of tissue lysis and protein isolation described previously in Chapter Two. The total protein concentration of each sample was then determined by using the bicinchoninic acid (BCA) assay. The BCA assay is a biochemical assay that relies on the reaction between bicinchoninic acid and protein in a highly alkaline solution. The reaction produces a purple coloured solution that deepens in intensity as the concentration of protein increases (Smith et al., 1985). A set of protein standards was included in the BCA assay to generate a standard curve, which was used to determine the total protein concentration of each sample in the assay as shown in Equation 2.3 (Chapter Two).

7.2.2 Results from the Western Blot for KCNH5

Given the lineage-specific expression of transcripts 1α and 1β, it was hypothesised that the western blot results would detect isoform 1α1 or 1α2 in first-trimester placenta and isoform 1β1 or 1β2 in adult brain and HepG2 cells. After 40 µg total protein was electrophoresed in an SDS-PAGE gel, protein was transferred from the gel onto a membrane and the KCNH5 antibody was used to detect all isoforms of KCNH5 that are shown in Figure 7.5. Subsequently, detection of the loading control protein, Beta-actin, was performed on the same membrane using an antibody that detected a 42 kDa Beta-actin protein. The western blots for KCNH5 and Beta-actin are shown in Figure 7.7.
Figure 7.7. Western blot results for KCNH5. Samples are listed above the protein bands to which they correspond. The molecular weight marker (in kDa) is displayed on the left. A. KCNH5 protein. B. Beta-actin protein, the loading control.

The results from the western blot confirmed that first-trimester placenta and brain express different isoforms of KCNH5 protein. As shown in Figure 7.7A, two bands were observed in first-trimester placenta at ~58 kDa and ~53 kDa, while a much larger band was detected in brain at ~170 kDa. Furthermore, the intensities of the bands for KCNH5 in first-trimester placenta are similar to the band intensities in brain, suggesting that the levels of KCNH5 protein in these tissues are similar. There may be less protein in HepG2 cell lysate, which was also suggested by the results from the TaqMan expression assay for KCNH5 shown in Figure 7.6B. The loading control protein, Beta-actin, was detected in all samples at the predicted band size of 42 kDa (Figure 7.7B). The bands for Beta-actin were of similar intensities for all samples, which confirmed that similar amounts of protein for each sample (40 µg) were loaded into the polyacrylamide gel for this western blot.

The two bands in first-trimester placenta at ~58 kDa and ~53 kDa are consistent with the predicted molecular weights of isoforms 1a₁ (49 kDa) and 1a₂ (43 kDa), respectively, taking
into account that post-translational modifications most likely produced slightly heavier proteins. Post-translational modification is a common event that occurs during protein synthesis and involves the chemical modification of proteins at one or more sites. These modifications include the addition of various chemicals to the amino acid residues as well as the full cleavage of the peptide backbone, all of which can change the size, structure and cellular functions of the final processed protein (Walsh, 2005). In a western blot, post-translational modification is a common explanation for why a larger protein is detected than what is predicted by its amino acid sequence (Marino et al., 2010).

According to the Universal Protein Research Knowledgebase (UniProtKB), KCNH5 contains two sites at which post-translational modifications can occur (Tan et al., 2008; Jain et al., 2009; Consortium, 2010). There is a potential glycosylation site at residue 403 (encoded by exon seven), which is present in all isoforms of KCNH5 (Figure 7.5). There is also an ubiquitination site at residue 785 of KCNH5 (encoded by exon eleven), which is present in only isoform 1b2 (Figure 7.5D). Glycosylation is the most common of all post-translational protein modifications, and involves an enzymatic reaction to attach a glycosyl (carbohydrate) group to one of four amino acid residues in the protein peptide (Fukuda et al., 1989; Lis and Sharon, 1993; Arnold et al., 2007). Ubiquitination, however, involves the covalent attachment of the 76-amino acid protein, ubiquitin, to the polypeptide chain, which acts as a molecular tag to direct the protein to the proteasome for degradation (Hershko et al., 1986; Hershko and Ciechanover, 1998; Fang and Weissman, 2004).

Post-translational modification of protein results in a higher molecular weight. A single glycosylation event will increase the molecular weight of a protein by 2 to 3 kDa (Dr. Liz Ledgerwood, Senior Research Fellow, University of Otago, personal communication). Thus the predicted shift in molecular weight as a result of a single glycosylation event would support the slightly heavier isoforms of 1a1 and 1a2 that were detected in first-trimester placenta (Figure 7.7A). The exact increase in molecular weight caused by a single ubiquitination event is unknown, however, the addition of a 76-amino acid protein (ubiquitin) to KCNH5 isoform 1b2 could substantially increase the molecular weight of this protein. Nevertheless, given that the process of ubiquitination directs the protein for degradation, an ubiquitinated protein usually results as a smear of bands on a western blot instead of a higher molecular weight protein (Tan et al., 2008).
KCNH5 protein in brain was observed at ~170 kDa (Figure 7.7A), which was heavier than the predicted size of 112 kDa (isoform 1b_1). This large difference in size between the predicted and detected molecular weight of KCNH5 may be due to post-translational modification of this large protein; however, it seems unlikely that one glycosylation event would produce such a large protein, and an ubiquitination event would probably display a smear of bands on the blot. The KCNH5 antibody cross-reacting with another protein may also cause the unexpectedly high molecular weight of KCNH5 in brain; however, the presence of a single band in each sample lane suggests that this antibody is specific to KCNH5 (Dr. Liz Ledgerwood, Senior Research Fellow, University of Otago, personal communication). A more likely explanation for the larger isoform of KCNH5 detected in brain may be that the 8% polyacrylamide gel had inaccurately detected the size of this protein. Two previous studies that performed a western blot for KCNH5 on brain used lower percentage polyacrylamide gels at 6% and 7.5%, and detected bands at 100 kDa and 120 kDa, respectively (Ludwig et al., 2000; Jeng et al., 2005). A lower percentage (<8%) polyacrylamide gel may have detected a molecular weight that was more consistent with the predicted 112 kDa for isoform 1b_2; however, in this study, it was more important to run all protein samples on the same gel to detect all isoforms simultaneously. Therefore, the 8% gel was used in this experiment to detect the wide-range of molecular weights for the isoforms of KCNH5.

Protein in the HepG2 cell lysate (positive control) was observed at ~85 kDa, which was heavier than the 60 kDa band shown in the antibody company’s western blot image (AbCam), yet consistent with the band sizes of ~80 kDa shown by three other antibody companies (Aviva, Sigma, and LifeSpan BioSciences). These companies, including AbCam, distributed antibodies for KCNH5 that were raised against amino acid residues that are present in all KCNH5 isoforms; therefore their antibodies should detect all isoforms of KCNH5. The Sigma antibody company was contacted regarding its varying predictions for KCNH5 protein, since they also showed a ~28 kDa band to be present with the ~80 kDa band in their HepG2 cell lysate western blot image. Sigma replied by saying that the multiple isoforms of KCNH5 (Figure 5.5) make it highly possible for numerous splice variants of KCNH5 to exist. The presence of splice variants may thus be the reason for why these antibodies detect a variety of different molecular weight bands after a western blot for KCNH5. Nevertheless, the protein detected in HepG2 cell lysate is most likely isoform 1b_1, with a predicted molecular weight at 69 kDa, and post-translational modification (i.e., glycosylation) may have produced the slightly heavier protein.
Although all of these antibody companies suggested using HepG2 cell lysate as the positive control for KCNH5, the results from this western blot indicate that it may not have been the best choice for a control tissue. Previous studies have confirmed the molecular weight of KCNH5 isoform 1b₂ in EPI-derived tissues: Jeng and colleagues performed a western blot for KCNH5 on rat brain homogenates and found a band at 120 kDa, stating that the protein may be slightly heavier as a result of post-translation modification in neurons (Jeng et al., 2005). Ludwig and colleagues localised KCNH5 in regions of rat brain and found protein in the 100 kDa range (Ludwig et al., 2000). Finally, Wadhwa and colleagues detected a band at 112 kDa for KCNH5 protein in human renal cell carcinoma tumours (Wadhwa et al., 2009). Since the molecular weight of KCNH5 in brain was confirmed by Jeng et al and Ludwig et al, the adult brain sample may have been a better choice for the positive control tissue in this experiment. Nevertheless, the ~85 kDa band detected in HepG2 cell lysate was consistent with the ~80 kDa band shown by three antibody companies. Most importantly, the absence of a band in adult primary melanocytes (negative control) confirms that this western blot is accurately detecting KCNH5.

In summary, the western blot protein analysis for KCNH5 established that isoforms 1a₁ and 1a₂ are expressed in first-trimester placenta, isoform 1b₂ is expressed in brain and isoform 1b₁ is expressed in HepG2 cell lysate. The consistency in the sizes of isoforms 1a₁ and 1a₂ in first-trimester placenta also confirms that the KCNH5 transcripts that use exon 1a do in fact terminate translation at or before the end of exon seven, yet the reason for this premature termination is unknown. Based on the lineage-specific methylation and expression profiles for transcripts 1a and 1b of KCNH5, the results from the western blot confirm that tissues derived from the extra-embryonic lineages begin transcription of KCNH5 at exon 1a, while tissues derived from the embryonic (epiblast) lineage use exon 1b.
7.3 Immunohistochemistry to Localise KCNH5 Protein

To continue investigation into the possible functional role of KCNH5 in the placenta, immunohistochemistry was selected to localise KCNH5 protein in placental tissue. Examining the pattern of KCNH5 expression may help in determining the function of this protein in the placenta. Although there are no reports of immunostaining for KCNH5 in placenta, a small number of studies have localised the expression of closely related potassium channel genes in mouse embryonic tissues. The expression of an EAG-related (ERG) gene, KCNH2 (also known as MERG1A), was differentially localised in mouse oocytes and blastocysts, suggesting that the location of KCNH2 expression is developmentally regulated (Winston et al., 2004). Another study localised the expression of EAG- and ERG-like potassium channels in mouse embryonic tissues, suggesting that these proteins may function like a molecular clock to regulate important cellular decisions during the earliest stages of embryogenesis (Day et al., 2001). The results from these studies suggest that potassium channels may be a key regulator of the cellular processes that are required for mouse embryonic development, a role that may translate to a similar function in human embryonic development.

The structure and cellular location of voltage-gated potassium channels, including KCNH5, have been well characterised in epiblast-derived tissues. Voltage-gated potassium channels are transmembrane proteins comprised of four alpha subunits (Stevens et al., 2009). Each subunit is made up of several distinct components, including six membrane-spanning domains (called S1 to S6), a pore region and intracellular N- and C-terminal domains (Yellen, 2002). The first four inter-membrane regions, S1 to S4, are responsible for the voltage-sensing which regulates the opening of the channel, while the pore forming (P) region acts as a filter through which the ions can selectively pass (Yusaf et al., 1996; Bezanilla, 2002). The N-terminal region of KCNH5 has been shown to be involved in regulating the current of the channel (Schonherr et al., 2002). The N-terminal also contains a Per-Ant-Sim (PAS) domain; a region that is common to many proteins and is known to regulate protein-protein interactions (Morais Cabral et al., 1998; Wray, 2004). The C-terminal region of KCNH5 contains a cyclic nucleotide-binding domain (cNBD), which is thought to interact with the PAS domain to regulate the activity of this potassium channel (Stevens et al., 2009). In order of size, the predicted structures of the four isoforms of KCNH5 are shown in Figure 7.8.
Due to the differing amino acid sequences between the four isoforms of KCNH5, there are large differences in the predicted structures of this protein between epiblast-derived tissues and the placenta. The largest and most well characterised isoform 1b$_2$ (988 aa, Figure 7.8A) is localised to the cell membrane in epiblast-derived tissues. This large isoform contains all of the necessary sub-domains that are required for a fully functional KCNH5 protein. The shorter isoform 1b$_1$ (611 aa, Figure 7.8B), also in epiblast-derived tissues, is missing 377 aa in the N-terminal region, resulting in a shorter N-terminal domain with a truncated cNBD region. Given that the cNBD region is thought to be involved in regulating KCNH5 activity...
by interacting with the PAS region (Stevens et al., 2009), it is unclear whether a shortened cNBD region will affect the function of isoform 1b1.

The shorter isoforms of KCNH5 (1a1 and 1a2) are missing many of the amino acid residues that encode the various sub-domains of KCNH5. As seen in Figure 7.8C, isoform 1a1 has a shortened C-terminal which is a result of the missing 58 amino acids that are present in the epiblast-derived isoforms 1b1 and 1b2. This missing sequence also affects the PAS region in the C-terminal. Isoform 1a1 is also missing one of the six transmembrane regions (S6) and part of the pore forming (P) region, as well as the entire N-terminal, which includes the cNBD region. Isoform 1a2 has the same predicted structure as isoform 1a1 yet this isoform is missing an even larger portion of the pore-forming region (Figure 7.8D). Based on the missing amino acid residues in isoforms 1a2 and 1a2, it is unclear whether these alpha subunits will form a tetramer and localise to the cell membrane (Mark Wareing, Senior Scientist, University of Manchester, UK, personal communication). Furthermore, many of the missing amino acid residues are known to encode essential components for KCNH5 function in epiblast-derived tissues, therefore it is unknown whether this protein will have a different function, or any function, in the placenta.

At this stage in the project, the aim was not to characterise the various sub-domains of KCNH5, rather it was to investigate the possible roles of the alternative isoforms of KCNH5 in the placenta. Immunohistochemistry was thus performed as a means to generate clues for the possible functional role of KCNH5 in the placentas.

7.3.1 Results for Immunohistochemistry for KCNH5 in Placenta

Immunohistochemistry was performed on sections from paraffin-embedded blocks of placental tissue using the same antibody as the western blot (detects all isoforms of KCNH5). Blocks of placental tissue ranging from five to 40 weeks of gestation were selected in order to examine the patterns of KCNH5 expression that were associated with placental development. A tissue sample containing fetal brain tissue was selected as the positive control tissue since KCNH5 protein is highly expressed in the brain (Ludwig et al., 2000). To aid with histological analysis, haematoxylin and eosin staining was performed concurrently with the antibody staining for each sample. The slides of placenta stained for KCNH5 were examined and photographed using a light microscope (Figure 7.9).
In first-trimester placenta, *KCNH5* was consistently expressed in the mesenchymal (M) layer of the placental villous core (Figure 7.9A-D), specifically in the stromal and endothelial cells. The placental villous mesenchyme is comprised of connective tissue fibres, blood vessels and placental macrophages, known as Hofbauer cells, all within an extracellular matrix (Castellucci *et al.*, 1980; Castellucci and Kaufmann, 1982; King, 1987). The Hofbauer macrophages are a major cell type in first-trimester villous mesenchyme and are thought to be involved in preventing the fetus from maternal pathogens in the earlier stages of pregnancy (Zaccheo *et al.*, 1989). *KCNH5* was not detected in the villous cytотrophoblast (CTB) or in the syncytiotrophoblast (STB) layers. These results were confirmed by Professor Graham Burton, the director of the Centre for Trophoblast Research (Cambridge, UK), who examined these images and agreed that *KCNH5* protein was localised to the villous mesenchyme rather than to any of the trophoblast subpopulations.
Given that the first-trimester placenta specifically expresses transcript 1a (Figure 7.3), this immunostaining must represent isoforms 1a₁ and/or 1a₂. These isoforms were localised within the cytoplasm of the placental villous mesenchyme rather than at the cell membrane. Given that functional voltage-gated potassium channels are located in a cell membrane (Wray, 2004), it is unclear whether these smaller isoforms of KCNH5 encode functional proteins in the placenta. However, although the mesenchymal staining appears to be cytoplasmic, KCNH5 may actually be located within the membranes of the cells’ organelles. Further experiments would be required to determine if KCNH5 was located within organelle membranes. Nevertheless, the aim of this experiment was to generate clues about the possible function of KCNH5 in the placenta. Since there are no extra-villous trophoblast cells present in the tissue sections prepared for immunohistochemistry, it is unclear whether KCNH5 is involved in the invasive activity of the trophoblasts during placental development. This protein may also encode a non-membrane potassium channel or it simply may be a non-functional protein. Although the function of this protein is unclear, immunohistochemistry confirmed expression of KCNH5 in first-trimester placenta.

To continue investigation into the functional role of KCNH5 in the placenta, it was of interest to examine placental tissue from later stages of gestation. Not only would this confirm the expression of KCNH5 in later stage placenta, but also any changes in the expression patterns may provide additional clues for the function of these smaller placental-specific isoforms 1a₁ and 1a₂. Immunohistochemistry was performed on blocks of placental tissue from the second and third trimesters of pregnancy. The samples were provided by Dr. Noelyn Hung (Senior Lecturer, Dunedin School of Medicine). The slides stained for KCNH5 were examined and photographed using a light microscope (Figure 7.10).
In the second and third trimesters of pregnancy, KCNH5 was continuously expressed in placental villous mesenchyme (Figure 7.10A-C). These results confirm that isoforms 1a₁ and 1a₂ of KCNH5 are expressed in the placenta throughout pregnancy. KCNH5 protein was also detected in the amnion of term placenta (Figure 7.10D). The amnion, comprised of a layer of cuboidal epithelium cells and an underlying layer of connective tissue, is a membrane that fused with the connective tissue of the underlying chorionic plate (Reith, 1965). Interestingly, the adjacent and adherent amnion and chorion membranes are derived from different cell lineages; the amnion is derived from the epiblast (embryonic) lineage, while the chorion is derived from the extra-embryonic lineages. Detection of KCNH5 in the amnion therefore represents expression of KCNH5 transcript 1b; the transcript expressed in epiblast-derived tissues. This result validated the results from the gene expression analysis in Chapter Six, which detected high levels of KCNH5 expression in the amnion by quantitative RT-PCR (Figure 6.6).
7.3.2 Immunohistochemistry for KCNH5 Supports the Origin of Placental Mesenchyme

The results from the immunohistochemistry for KCNH5 may reveal the origin of the placental villous mesenchyme - a topic of longstanding debate. Although it is agreed that villous mesenchyme is not derived from the trophoderm, some believe that mesenchyme is derived from the epiblast while others believe it stems from the primitive endoderm. The trophoderm was ruled out as the lineage origin of villous mesenchyme when studies of mesenchymal development in the rhesus monkey showed that the trophoderm only gives rise to subpopulations of the trophoblasts (Benirschke et al., 2006). At present, some believe that the villous mesenchyme is epiblast-derived and develops from the caudal end of the primitive streak; a structure that forms during early mammalian development that gives rise to the germ layers (Luckett, 1978; Downs, 2009). Others report that villous mesenchyme is derived from the primitive endoderm after a two-step process; where the primitive endoderm first forms a network structure, or reticulum, which then differentiates into the villous mesenchymal cells of the placenta (Enders and King, 1988; Bianchi et al., 1993). Overall, there is a paucity of literature supporting either the epiblast or the primitive endoderm as the lineage origin of the placental villous mesenchyme.

In this study, the combined epigenetic and expression profiles of KCNH5 may provide the answer to this 40-year debate. Given the lineage-specific expression of KCNH5’s alternative transcripts 1a and 1b (Figure 7.3), total placental transcription occurs exclusively from the exon 1a promoter (Figure 7.1). Thus the expressing villous mesenchymal cells in the placenta must use the exon 1a promoter. Since all adult somatic and embryonic tissues, including umbilical cord and amnion, show dense methylation of the exon 1a promoter (Figure 6.6), this lineage-specific methylation dichotomy supports that the placental villous mesenchyme cannot be derived from the epiblast. Instead, the epigenetic profile is consistent with the alternative hypothesis that the villous mesenchyme is derived from the primitive endoderm.

The difference between the epigenetic states of the trophoderm and primitive endoderm versus the epiblast has previously been demonstrated. In comparison to embryonic tissues, undermethylation of repetitive DNA sequences (minor satellite sequences and Mus interspersed family, (MIF) sequences) was found to be a feature of trophoderm and primitive endoderm-derived tissues in the mouse (Chapman et al., 1984). The hypomethylation was not restricted to a certain class of repeat elements or a specific chromosomal location, instead it was suggested to be a feature of the entire genome in
trophectoderm-derived and primitive endoderm-derived tissues (Chapman et al., 1984). Similarly, the epigenetic processes involved in X inactivation in the mouse were reported to be different between the epiblast and the trophoderm and primitive endoderm, where X inactivation is random in the epiblast, but imprinted in the trophoderm and primitive endoderm (Monk and Harper, 1979; Mak et al., 2004; Okamoto et al., 2004). Thus hypomethylation is a feature of tissues derived from the trophoderm and primitive endoderm, and higher levels of methylation are found in tissues derived from the epiblast.

The fact that epigenetic differences have been previously identified between the extra-embryonic lineages (trophoderm and primitive endoderm) and the epiblast supports the dichotomous methylation found between these lineages at KCNH5’s exon 1a promoter region. In summary, the combined epigenetic and expression profiles for KCNH5 in the placenta strongly suggest that the placental villous mesenchyme may be derived from the primitive endoderm.
Chapter Eight

Retrotransposon-Derived Transcripts in the Placenta

8.1 A Retrotransposon Gave Rise to $\textit{KCNH5}$ Transcript 1a

The placental-specific expression of $\textit{KCNH5}$ transcript 1a prompted further analysis into the region of differential methylation that was regulating the expression of this unique transcript. The most exciting moment of this project occurred when subsequent analysis revealed a short interspersed nuclear element (SINE) retrotransposon situated precisely in the region of differential methylation. The genomic location of this SINE element aligned to the region that was quantified by Sequenom, spanning the 17 CpG sites in the promoter region and beginning of exon 1a, as shown in Figure 8.1. According to the Genome Browser, this SINE retrotransposon is located in the last 147 bp of the upstream promoter and the first 162 bp of exon 1a (GRCh37 assembly chr14:63,568,599-63,568,907) (Kent et al., 2002).

![Figure 8.1](image)

**Figure 8.1.** Genomic arrangement of the SINE ($\textit{AluY}$) retrotransposon located in the region of differential $\textit{KCNH5}$ methylation. The SINE retrotransposon is shown in blue and is located on the sense (+) strand. Exon 1a is shown in green and is located on the antisense (-) strand. The region of differential methylation quantified by Sequenom is shown in grey. The red circles represent CpG sites. The enhancer element (A box) and directing element (B box) within the SINE RNA polymerase III promoter are represented by the black boxes. The locations of the $\textit{AluI}$ restriction sites (encoded by the $\textit{Alu}$ element) are shown by the black vertical lines.

8.1.1 Retrotransposons

Retrotransposons are mobile genetic elements that rely on RNA intermediates to propagate themselves throughout eukaryotic genomes (Weiner et al., 1986; Deininger et al., 2003). Their mode of transposition resembles a “copy and paste” behaviour whereby the
Retrotransposon transcribes itself into RNA, which is then reverse-transcribed into the DNA and inserted back into the genome (Finnegan, 1989; Dombroski et al., 1994). Retrotransposons are different to DNA transposons, a class of transposable elements that exhibit a simple “cut and paste” behaviour using DNA intermediates for repositioning in the genome (Finnegan, 1989). Due to the retrotransposons’ self-encoded reverse transcriptase enzyme, these mobile elements have substantially increased the size of most eukaryotic genomes over time and are thus responsible for a high proportion of the repetitive regions in large genomes (SanMiguel et al., 1996; Wicker et al., 2007). These replicating sequences, which have successfully expanded the genome, are sometimes referred to as “junk DNA”.

Retrotransposons can be classified into groups according to their sequence, structure and ability to self-transpose. A review published in Nature Genetics by Wicker et al. explains the most recent guidelines to classify retrotransposons, which are grouped into five ‘orders’: long-terminal repeat (LTR) retrotransposons, *Dictyostelium* intermediate repeat sequence (DIRS)-like elements, Penelope-like elements (PLEs), long interspersed nuclear elements (LINEs) and short interspersed nuclear elements (SINEs) (Wicker et al., 2007). A modified version of this classification system for retrotransposons is shown in Figure 8.2. Each group of retrotransposons contains sub-groups that share a common mode of replication but have different features. A main difference between these sub-groups is in their target site duplication (TSD) region; a short repeat region created upon insertion of the retroelement on both flanking ends of the inserted sequence. In this classification system, there is no conservation in DNA sequence between superfamilies and very few conserved elements at the protein level (Wicker et al., 2007).
Retrotransposons

in the Placenta

8.1.1.1 LTR-retrotransposons

LTR-retrotransposons can range from a few hundred to many thousands of base pairs. The actual “long terminal repeat” sequence at the two flanking ends of these retrotransposons can range from a few hundred to up to five thousand base pairs (Wicker et al., 2007). LTR-retrotransposons contain sequences that encode the essential proteins required for the packaging, reverse transcription and integration of the LTR-retrotransposon back into the genome. These proteins, namely the capsid protein (GAG), aspartic protease (AP), integrase (INT), reverse transcriptase (RT) and RNase H (RH) are displayed in Figure 8.2A. The retrovirus and endogenous retrovirus (ERV) LTR-retrotransposons contain an additional protein in their sequences, the envelope protein (ENV), which allows for their viral-like behaviour (Figure 8.2A) (Seelamgari et al., 2004; Bannert and Kurth, 2006). A retrovirus can be classed as an endogenous retrovirus once it has integrated itself into the germ line and can
be passed on to subsequent generations (Belshaw et al., 2004). All sub-groups of LTR-retrotransposons produce a target site duplication region of four to six base pairs after they have transposed (Wicker et al., 2007). These LTR-retrotransposons comprise around eight percent of the human genome (McCarthy and McDonald, 2004).

8.1.1.2 (DIRS)-like elements and PLEs

The more recently classified retrotransposons are the Dictyostelium intermediate repeat sequence (DIRS)-like elements and the Penelope-like elements (PLEs). The (DIRS)-like elements contain a tyrosine recombinase (YR) instead of an integrase (INT) (Figure 8.2B), which results in no formation of a target site duplication region upon insertion of this retroelement (Wicker et al., 2007). In addition, the terminal or flanking regions of the DIRS-like elements have been found to be inverted or interrupted, suggesting that the mechanisms of insertion of this retroelement superfamily are unique (Cappello et al., 1985). The Penelope-like elements (PLEs) encode only two proteins, a reverse transcriptase and an endonuclease. The long terminal repeat region in PLEs can also be in a sense or antisense orientation (Figure 8.2C) (Evgen'ev et al., 1997; Evgen'ev and Arkhipova, 2005; Wicker et al., 2007).

8.1.1.3 LINEs

The long interspersed nuclear elements (LINEs) do not contain LTRs in their flanking regions. LINEs are sub-classified into five major superfamilies based on the different proteins they encode, however all LINEs encode at least one reverse transcriptase (RT) protein and one endonuclease, either a normal endonuclease (EN) or an apurinic endonuclease (APE) (Figure 8.2D) (Wicker et al., 2007). For retrotransposition, LINEs rely on an RNA polymerase II promoter that is located within the LINE sequence (Singer, 1982). LINE retrotransposons can reach up to several kilobases in length and, combined with their ability to self-transpose, they now comprise roughly 17% of the human genome (Cordaux and Batzer, 2009).

8.1.1.4 SINEs

SINE retrotransposons are repetitive regions of DNA that are less than 500 bp long (Singer, 1982; Okada, 1991; Schmid, 1996). These short sequences do not contain long terminal repeat regions nor encode any functional protein, however they do contain an internal RNA polymerase III promoter that allows the SINE to be expressed. The RNA polymerase III promoter is located at the 5’ end of the SINE sequence and the transcription termination
signal is the poly-T tail at the 3’ end (Kramerov and Vassetzky, 2005). Once expressed, a SINE relies on a LINE-encoded reverse transcriptase enzyme for successful transposition and reintegration into the genome (Dewannieux et al., 2003; Kramerov and Vassetzky, 2005). SINE retrotransposons can be further classified into sub-groups based on the origin of their RNA polymerase III promoter; a sequence located in at the beginning or ‘head’ of the SINE sequence (Wicker et al., 2007). The SINE RNA polymerase III promoter can be of tRNA, 7SL RNA or 5S RNA origin (Kramerov and Vassetzky, 2005) (Figure 8.2E).

Of the SINE retrotransposons, the Alu elements are the most commonly transposed. These elements are named for their repetitive sequences that harbour the recognition site for the AluI restriction enzyme (Houck et al., 1979). Alu elements contain an RNA polymerase III promoter that originates from the 7SL RNA gene (Ullu et al., 1982). The Alu RNA polymerase III promoter is comprised of an enhancing element and a directing element, referred to as the “A box” and the “B box”, respectively (Perez-Stable et al., 1984). Both elements are required to initiate transcription of RNA polymerase III. The enhancing element (A box) is located just downstream of the transcription start site and is responsible for transcriptional efficiency. The directing element (B box) is approximately 50 bp downstream of the enhancing element and determines the accuracy of initiation (Roy et al., 2000). The SINE retrotransposon in the promoter and first exon (1a) of KCNH5 transcript 1a is an Alu element and its genomic arrangement is shown in Figure 8.1.

Most Alu elements are approximately 300 base pairs long and are located within introns and 3’ untranslated regions, and do not contain any coding sequences (Batzer and Deininger, 2002). Like other SINEs, Alu elements rely on the reverse transcriptase activity encoded by LINEs for amplification and transposition. It is reported that over one million copies of Alu elements have spread throughout the human genome, which makes these sequences the most abundant of all repetitive elements, comprising roughly ten percent of the human genome (Lander et al., 2001; Batzer and Deininger, 2002).

Being the most successfully transposed of the SINEs, Alu elements are thought to have the largest mutagenic threat to the human genome (Bennett et al., 2008). Alu elements have been found to accumulate in gene-rich regions, whereas other retrotransposons are more randomly distributed throughout the genome (Lander et al., 2001). In order to maintain normal genome function, Alu elements and other transposons are silenced by DNA methylation (Yoder et al., 1997; Goll and Bestor, 2005). Interestingly, one-third of the human genome’s CpG sites that
are susceptible to DNA methylation are located within *Alu* elements (Schmid, 1991). Most of these CpG-rich *Alu* elements are highly methylated in somatic tissues, and also in the placenta, to prevent their expression in these tissues (Gama-Sosa *et al.*, 1983; Schmid, 1991). Furthermore, the demethylation of *Alu* elements has been associated with aging, cancer and other processes caused by genome disruption and instability (Ehrlich, 2002; Rodriguez *et al.*, 2008; Szpakowski *et al.*, 2009).

### 8.1.2 Characterisation of the SINE retrotransposon in Transcript 1a

The SINE retrotransposon located in the promoter and first exon of *KCNH5* transcript 1a is an *AluY* element, a sub-class of the *Alu* elements (Figure 8.1). The *AluY* elements are thought to be the youngest sub-class of the *Alu* elements, as their divergence from the two older classes (*AluS* and *AluJ*) occurred only 25 million years ago (Batzer *et al.*, 1996; Batzer and Deininger, 2002). However, all *Alu* elements are considered to be evolutionarily young. They exist only in primate genomes due to their recent emergence in the last 65 million years (Daniels and Deininger, 1985).

According to the conservation track on the UCSC’s Genome Browser, the *AluY* element in *KCNH5* transcript 1a is only present in humans, old world monkeys (rhesus macaque) and apes (orang-utan and chimpanzee) (Kent *et al.*, 2002). This suggested that the transposition event for this *AluY* element occurred approximately 25 to 30 million years ago, which is consistent with the emergence of the *AluY* sub-class (Fleagle, 1999).

The discovery of this SINE (*AluY*) retrotransposon in the region of differential *KCNH5* methylation was fascinating given that transposable elements are normally silenced by DNA methylation to maintain normal genome activity (Yoder *et al.*, 1997). This SINE element has uniquely escaped the normal methylation-induced silencing in tissues derived from the extra-embryonic lineages (placenta and chorion). Escape from silencing has thus allowed this retrotransposon to give rise to an alternative promoter and first exon (1a), resulting in the unique *KCNH5* transcript 1a expressed in the placenta. The differential methylation of this recently evolved SINE (*AluY*) element therefore regulates the lineage-specific expression of *KCNH5*’s alternative transcripts, as shown in Figure 8.3. To our knowledge, this is also one of very few examples of a human transcript initiated by a SINE element (Romanish *et al.*, 2009).
Given that most Alu elements are methylated in the placenta (Gama-Sosa et al., 1983), the escape from methylation by the KCNH5 promoter combined with the exclusive expression of transcript 1a in the placenta (and not chorion) suggests that this unique AluY-derived transcript may have a functional role that is unique to the placentas of humans, old world monkeys and apes. This will be discussed in more detail in Chapter Nine. Interestingly, KCNH5 transcript 1a is not the only retrotransposon-derived gene in the placenta. Recently, a small number of other placental-specific gene transcripts have been identified that use an alternative promoter derived from a retrotransposon (Bieche et al., 2003; Reiss et al., 2007; Rawn and Cross, 2008).
8.2 DNA Methylation Analysis of Other Retrotransposon-Derived Genes in the Placenta

Among the previously reported genes that show placental-specific expression, a disproportionate number of these genes have retroviral elements in their regulatory regions (Bieche et al., 2003; Medstrand et al., 2005; Rawn and Cross, 2008). Most of these genes use the retrovirus’ flanking LTR sequences to initiate transcription. In fact, of all the documented LTR-derived tissue-specific genes, 40% are expressed in the placenta (Reiss et al., 2007). Some believe that the hypomethylated state of the placenta has contributed to the demethylation, and subsequent activation, of retrotransposons in this tissue (Okahara et al., 2004). However, recent work suggests that retrotransposon demethylation is independent of the placenta’s hypomethylated environment and instead represents a directed selection of placental-specific genes that are important for placental function and evolution (Dunlap et al., 2006; Reiss et al., 2007). There is a paucity of literature documenting the epigenetic regulation of retrotransposons in the placenta; therefore this is unexplored territory for both epigeneticists and placental pathologists.

At this stage in the project, it was of interest to determine whether other placental-specific retrotransposon-derived genes displayed the same methylation dichotomy between tissues that were derived from embryonic versus extra-embryonic lineages. Examining the methylation status of other retrotransposon-derived gene promoters across a wide range of tissues may improve our understanding of how DNA methylation acts to selectively silence or permit the expression of retrotransposons during the earliest stages of cell lineage specification. Based on few reports in the literature, five retrotransposon-derived genes in the placenta were selected for methylation analysis using Sequenom MassARRAY: ERVWE1, EBR, MID1, PTN and INSL4. These genes will be discussed in more detail in the following sections.

The Sequenom assay, including primer design, sample processing and data analysis, was performed using the same parameters and protocols that have been used for Sequenom throughout this study (described in Chapter Two). The mean promoter methylation was determined for each sample. Then, these means were averaged for the samples of first-trimester placenta and peripheral blood, whilst the means for all other samples were kept separate in order to examine their individual methylation levels.
8.2.1 **ERVWE1 (the Syncytin-1 gene)**

The most well characterised retrotransposon-derived gene in the placenta is *ERVWE1*, which encodes the Syncytin-1 gene. *ERVWE1* is derived from the envelope protein (ENV) of a human endogenous retrovirus and plays an important role in the placenta by regulating the cell-to-cell fusion required for the differentiation of cytotrophoblast into syncytiotrophoblast (Blond *et al.*, 2000; Mi *et al.*, 2000). This gene is critical for successful placental development and function, since inhibition of *ERVWE1* results in a substantial decrease in cytotrophoblast fusion and differentiation *in vitro* (Frendo *et al.*, 2003). Interestingly, the expression of *ERVWE1* is regulated by the 5' LTR of this human endogenous retrovirus, which is located approximately eight kilobases upstream of the ENV gene sequence, as shown in Figure 8.4 (Cheng *et al.*, 2004).

![Genomic arrangement of ERVWE1 (the Syncytin-1 gene). The human endogenous retrovirus (HERV), family W, locus E1 is shown by the white box (HERV-WE1). The flanking 5' and 3' LTRs are shown in grey. The envelope protein (ENV) is displayed in blue and encodes the ERVWE1 protein. The region of quantified methylation is shown by the red lollipops, which is located approximately 8 kb upstream of the ENV sequence and regulates the expression of ERVWE1.](image)

**Figure 8.4.** Genomic arrangement of *ERVWE1* (the Syncytin-1 gene). The human endogenous retrovirus (HERV), family W, locus E1 is shown by the white box (HERV-WE1). The flanking 5' and 3' LTRs are shown in grey. The envelope protein (ENV) is displayed in blue and encodes the ERVWE1 protein. The region of quantified methylation is shown by the red lollipops, which is located approximately 8 kb upstream of the ENV sequence and regulates the expression of *ERVWE1*.

Matouskova and colleagues have examined the methylation of the *ERVWE1* 5' LTR-derived promoter. Using bisulfite sequencing, they measured the promoter methylation in two term placentas and found that it was half that of somatic tissues and a cell line (Matouskova *et al.*, 2006). This difference in methylation between the placenta and epiblast-derived tissues was not as striking as the difference detected between tissues for the SINE-derived *KCNH5* transcript 1a. This may be due to the small sample size in Matouskova’s study or their use of an older and less quantitative technique to measure CpG site methylation. Therefore, *ERVWE1* was selected for Sequenom analysis to examine its promoter methylation across a wider range of tissues derived from the extra-embryonic lineages and the epiblast. Sequenom analysis would determine whether the difference in methylation is established at the earliest stages of cell lineage specification.
The methylation of the *ERVWE1* LTR-derived promoter was quantified by Sequenom. The LTR-internal and *ERVWE1*-specific external primer sequences designed by Matouskova *et al.* were adapted for use in the Sequenom assay by adding the primer tags described in Chapter Two (Matouskova *et al.*, 2006). The same samples that were included in previous Sequenom runs were examined in this analysis. A 352 bp product containing seven CpG sites was analysed by Sequenom (Figure 8.5A). One CpG site had to be excluded from analysis since it gave no data in any of the samples. During data analysis, the fetal 1/3 sample of term placenta was removed from analysis since it gave no data in ≥ 30% of the analysed CpG sites. The mean promoter methylation for *ERVWE1* is shown in Figure 8.5B.

![Figure 8.5. Sequenom results for *ERVWE1*. A. The region of methylation quantified by Sequenom is shown in grey. Red circles represent CpG sites and the blue box represents the 5’ LTR of the human endogenous retrovirus (HERV). The external *ERVWE1*-specific forward primer and internal 5’ LTR-specific reverse primer for the PCR are shown in yellow. B. The mean promoter methylation data for all samples. Error bars represent the 95% confidence interval of the mean for the tissue groups of first-trimester placenta (n = 17) and peripheral blood (n = 9).](image)

As shown in Figure 8.5B, the *ERVWE1* 5’ LTR-derived promoter showed lower methylation in tissues derived from the extra-embryonic lineages (first-trimester placenta, term placenta and term chorion) and higher methylation in all epiblast-derived tissues. The difference between trophectoderm-derived chorion (28%) and epiblast-derived amnion (63%)
methylation was not as large as it was for $KCNH5$. Analysing additional samples of chorion and amnion would provide a more accurate difference in methylation between these tissues. Nevertheless, given that $ERVWE1$ is expressed in only the placenta (Blond et al., 2000; Mi et al., 2000; Cheng et al., 2004), the epigenetic and expression profiles of $ERVWE1$ are similar to that of $KCNH5$ transcript 1a. The methylation dichotomy observed for the $ERVWE1$ LTR-derived promoter of suggests that hypomethylation permits expression of this gene in the placenta, whereas methylation prevents expression of this gene in epiblast-derived tissues.

The next three genes to be discussed ($MID1$, $EBR$ and $PTN$) were selected based on a report by Reiss and colleagues (Reiss et al., 2007). Each of these genes uses a promoter derived from the 5’ LTR of a human endogenous retrovirus to regulate their placental-specific expression. Using bisulfite sequencing, Reiss and colleagues found these three LTR-derived promoters to be hypomethylated in two term placentas and hypermethylated in peripheral blood (Reiss et al., 2007). It was of interest to examine the promoter methylation of these genes in a wider range of tissues in order to determine whether this methylation difference was lineage-specific. Thus, these three genes were selected for Sequenom analysis. For each gene, the primer sequences from Reiss et al. were used to locate the region of differential methylation and Sequenom primers were designed to examine this same site. The samples that were analysed in previous Sequenom runs were also included in this analysis.

### 8.2.2 $MID1$

The midline 1 gene, $MID1$, encodes a microtubule-associated protein that is a member of the RING finger family (Schweiger et al., 1999). When mutated, $MID1$ has been implicated in Opitz syndrome; a disorder involving perturbed formation of midline structures, which results in malformations such as hypertelorism and cleft palate (Opitz, 1987). $MID1$ has five alternative promoters to express tissue-specific gene transcripts, one of which is expressed specifically in the placenta and fetal kidney, another is expressed in only adipose tissue and the remaining three promoters are used ubiquitously (Landry and Mager, 2002; Reiss et al., 2007). The transcript expressed in placenta uses a promoter that is derived from the 5’ LTR of a human endogenous retrovirus family E (HERV-E). The genomic arrangement of $MID1$ and its alternative transcripts is shown in Figure 8.6.
Figure 8.6. Genomic arrangement of *MID1* adapted from Reiss et al. 2007. Coloured boxes display gene exons. The human endogenous retrovirus (HERV), family E, is shown by the white box and its flanking 5’ and 3’ LTRs are shown in grey. This endogenous retrovirus is located 30 kb downstream and 13 kb upstream of two ubiquitously expressed exons, 1c and 1e. The region of quantified methylation is shown by the red lollipops.

A Sequenom assay was designed to quantify the methylation of the 5’ LTR-derived promoter that is used to express *MID1* in the placenta. A 298 bp product containing six CpG sites was analysed by Sequenom (Figure 8.7A). Four CpG sites had to be excluded from analysis since one site gave no data in ≥ 30% of samples, one site was on a fragment of high mass and two sites were on fragments of low mass. The mean promoter methylation for the placental transcript of *MID1* is shown in Figure 8.7B.
As shown in Figure 8.7B, the methylation of the MID1 5' LTR-derived promoter was low in all tissues derived from the extra-embryonic lineages (first-trimester and term placenta and chorion). This confirmed the use of this LTR-derived promoter in the placenta. However, variable methylation was detected in epiblast-derived tissues. This may be due to the removal of four CpG sites during data analysis, which left only two CpG sites to calculate promoter methylation. Further, there was little difference between chorion (6%) and amnion (18%) methylation, which was unexpected given that these tissues are derived from different lineages. Umbilical cord (26%) also displayed lower methylation than expected. These results are based on only one sample of each tissue type, other than the groups of first-trimester placenta (n = 17) and peripheral blood (n = 9). Analysis of additional samples would be useful to confirm whether the methylation patterns are lineage-specific or only tissue-specific. Nevertheless, the adult somatic tissues displayed the highest level of promoter methylation, supporting the idea that this LTR-derived promoter is not used in these tissues.
(Figure 8.7). Although fetal somatic tissues showed variable levels of promoter methylation, the low methylation in fetal kidney (21%) is consistent with the finding that fetal kidney expresses the LTR-derived transcript of MID1 (Landry and Mager, 2002).

### 8.2.3 EBR

The endothelin B receptor gene (EBR) encodes a G-protein coupled receptor for endothelin, a protein that is involved in vascular homeostasis (Arai et al., 1993; Agapitov and Haynes, 2002). In the placenta, endothelins are important for maintaining fetal-maternal blood circulation (Handwerger, 1995). Interestingly, a human endogenous retrovirus, family E (HERV-E) located approximately 57.7 kb upstream of the native first exon of EBR has been shown to give rise to a placental-specific transcript, regulated by a promoter derived from its 5' LTR (Medstrand et al., 2001). This LTR-derived transcript encodes approximately 15% of EBR transcripts in the placenta (Landry and Mager, 2003). The genomic arrangements of the alternative transcripts of EBR are shown in Figure 8.8.

A Sequenom assay was designed to quantify the methylation of the 5' LTR-derived promoter that is used to express EBR in the placenta. A 352 bp product containing seven CpG sites was analysed by Sequenom (Figure 8.9A). Two CpG sites were excluded from analysis since they were located on fragments of low mass, both of which were outside the accurate detection range.
limits of the mass spectrometer. During data analysis, the fetal 1/3 sample of term placenta was removed since it gave no data in ≥ 30% of the analysed CpG sites. The mean promoter methylation for the placental-specific transcript of EBR is shown in Figure 8.9B.

As shown in Figure 8.9B, a significant difference in methylation was detected between tissues derived from the extra-embryonic lineages and tissues derived from the epiblast. All placental samples and chorion displayed low levels of methylation, while all epiblast-derived tissues (except amnion) displayed higher levels of methylation. Surprisingly, there was no contrast between chorion (17%) and amnion (13%) methylation, however, this was also the result for the MIDI LTR-derived promoter and may reflect inaccuracy because of the small number of analysable CpG sites after the Sequenom run. Nevertheless, the promoter methylation in peripheral blood (94%) was consistent with the peripheral blood methylation (93%) determined by Reiss and colleagues (Reiss et al., 2007).
8.2.4 **PTN**

The pleiotrophin gene, *PTN*, encodes a heparin-binding cytokine protein that is involved in a wide range of functions including cell differentiation and angiogenesis (Deuel *et al.*, 2002). *PTN* has also been identified as a proto-oncogene due to its high level of expression in a variety of human tumours and involvement in tumour growth and angiogenesis (Kadomatsu and Muramatsu, 2004). There are two transcripts of *PTN*, one of which is expressed specifically in the placenta using a promoter derived from the 5’ LTR of an endogenous retrovirus (Schulte *et al.*, 1996). However, Reiss and colleagues mention that the *PTN* transcript derived from the native (non LTR) promoter has also been identified in the placenta, according to the GenBank (Reiss *et al.*, 2007; Benson *et al.*, 2009). The genomic arrangement of *PTN* transcripts is shown in Figure 8.10.

![Diagram of genomic arrangement of PTN](image)

**Figure 8.10.** Genomic arrangement of *PTN* adapted from Reiss *et al.* 2007. Coloured boxes display gene exons. The human endogenous retrovirus (HERV), family E, is shown by the white box and its flanking 5’ and 3’ LTRs are shown in grey. This endogenous retrovirus is located 70 kb downstream and 7 kb upstream of the exons 1a and 2, which are used to express *PTN* in other tissues. The region of quantified methylation is shown by the red lollipops.

A Sequenom assay was designed to quantify the methylation of the 5’ LTR-derived promoter that is used to express *PTN* in the placenta. A 446 bp product containing eight CpG sites was analysed by Sequenom (Figure 8.11A). Four CpG sites had to be excluded from analysis since two sites were on fragments of low mass, one site was on a fragment of high mass and one site gave no data in ≥30% of samples. During data analysis, the samples of term placenta maternal 1/3, fetal 1/3, umbilical cord and adult brain were removed since they gave no data in ≥30% of CpG sites. The mean promoter methylation for the placental-specific transcript of *PTN* is shown in Figure 8.11B.
Figure 8.11. Sequenom results for PTN. A. The region of methylation quantified by Sequenom is shown in grey. Red circles represent CpG sites and the blue box represents the 5’ LTR of the human endogenous retrovirus (HERV). Forward and reverse primers for the PCR are shown in yellow. The primers used by Reiss et al. are shown in pink. B. The mean promoter methylation data for all samples. Error bars represent the 95% confidence interval of the mean for the tissue groups of first-trimester placenta (n = 14) and peripheral blood (n = 9).

Although some samples had to be removed during analysis, a general lineage-specific pattern of methylation was detected for this LTR-derived transcript of PTN (Figure 8.11B). Tissues derived from the extra-embryonic lineages (placenta and chorion) displayed low levels of methylation, whereas tissues derived from the epiblast (except amnion) showed higher levels of promoter methylation. No significant difference was detected between chorion (18%) and amnion (20%) methylation. However, as explained previously, the small number of analysable CpG sites precludes over-interpretation of the measures of methylation in this assay. The moderate to high levels of methylation in adult and fetal somatic tissues is consistent with the findings from other genes for these epiblast-derived tissues.

8.2.5 INSL4

The final gene selected for Sequenom analysis was early placental insulin-like 4, INSL4 (also known as EPIL) a gene that encodes an insulin-like hormone peptide that is placental-specific
Retrotransposons in the Placenta

(Chassin et al., 1995; Bellet et al., 1997). Although this gene is relatively uncharacterised, INSL4 is believed to be important for placental apoptosis, an important process required for normal placental development (Millar et al., 2005). This gene is also associated with fetal growth restriction since its expression was substantially higher in growth-restricted fetuses compared to fetuses of normal growth (Giudice et al., 1995). Interestingly, Bieche and colleagues discovered that a human endogenous retrovirus (family K) was located in the promoter region of INSL4, and its 3’ LTR was responsible for regulating INSL4 expression (Bieche et al., 2003). This HERV-K is located approximately 400 bp upstream of the first of two coding exons. Similar to KCNH5, the conservation of this HERV-K suggested that INSL4 was a primate-specific gene (Bieche et al., 2003). The genomic arrangement of INSL4 is shown in Figure 8.12.

![Genomic arrangement of INSL4](image)

**Figure 8.12.** Genomic arrangement of INSL4. Coloured boxes display gene exons. The human endogenous retrovirus (HERV), family K, is shown by the white box and its flanking 5’ and 3’ LTRs are shown in grey. The region of quantified methylation is shown by the red lollipops.

Although Beiche et al. performed a deletion analysis to confirm the regulatory activity of the INSL4 promoter (Bieche et al., 2003), there are no reports on the methylation status of this retrotransposon-derived promoter. Thus, a Sequenom assay was designed to quantify the methylation of the INSL4 5’ LTR-derived promoter. A 438 bp product containing four CpG sites was analysed by Sequenom (Figure 8.13A). During data analysis, the term placenta fetal 1/3 sample was removed since it gave no data in ≥ 30% of CpG sites. The mean promoter methylation for the placental-specific transcript of INSL4 is shown in Figure 8.13B.
As shown in Figure 8.13B, the INSL4 promoter showed low methylation in tissues derived from extra-embryonic lineages and high methylation in epiblast-derived tissues. Similar to KCNH5 transcript 1a, the dichotomous methylation was highlighted by the difference between the trophectoderm-derived chorion (13%) and epiblast-derived amnion (93%). However, although this difference suggests lineage-specific methylation, the small number of CpG sites analysed means that this assay should be re-designed and also repeated with additional samples before the data is over-interpreted. It would be fascinating if the methylation of the INSL4 LTR-derived promoter was truly lineage-specific and potentially responsible for regulating the expression of this placental-specific gene.

8.2.6 Summary of Results from the Methylation Analysis of Other Retrotransposon-Derived Genes in the Placenta

Of the five placental-specific LTR-derived genes that were examined by Sequenom, two genes, ERVWE1 and INSL4, displayed a level of promoter methylation in extra-embryonic
tissues (including the chorion) that was dichotomous with that in embryonic tissues (including the amnion). However, given that only one sample of chorion and amnion was analysed, repeating Sequenom on additional samples would confirm that this lineage-specific methylation dichotomy is real. If this result is accurate, then the difference between chorionic and amniotic methylation indicates that the methylation does not reflect the placenta itself, rather it reveals the origin of these tissues with respect to embryonic or extra-embryonic lineage commitment. Yet the placental-specific use of these LTR-derived promoters suggests that these genes may have a functional significance in the placenta.

Although the other three genes, *MID1*, *EBR* and *PTN*, displayed a general pattern of lineage-specific methylation, the contrast in promoter methylation was not present between the chorion and amnion. It would be useful to repeat Sequenom analysis on these three genes with additional samples of chorion and amnion, in addition to re-designing the PCR to incorporate more analysable CpG sites. However, if the results are real for these three genes, then the methylation may be tissue-specific instead of lineage-specific, suggesting that retrotransposon methylation is established at a later stage in development (during tissue differentiation). Nevertheless, each of these genes displayed low placental methylation, which is consistent with the use of these LTR-derived promoters as alternative gene transcription sites in the placenta.

In summary, the differential expression of these retrotransposon-derived genes relies on methylation-induced silencing of the foreign retroviral DNA in embryonic tissues, and escape from this mechanism in extra-embryonic tissues. The potential functional significance of these epigenetically “active” retrotransposons in the placenta will be discussed in more detail in Chapter Nine.
Chapter Nine

Discussion of Results and Future Directions

9.1 Epigenetics in the Placenta

Epigenetic modifications play a key role in directing the lineage commitment of cells during the earliest stages of mammalian development (Santos et al., 2002; Reik et al., 2003). The differentiation of the extra-embryonic lineages from the embryonic lineage is a two-step process; the first cell fate decision is made when the inner-cell mass is distinguished from the trophectoderm, and the second is when the inner-cell mass divides into the primitive endoderm and epiblast. The extra-embryonic lineages (trophectoderm and primitive endoderm) will give rise to the placenta and chorion, whereas the embryonic lineage (epiblast) will give rise to the amnion, umbilical cord, and the embryo (Enders and Schlafke, 1981; Cross et al., 1994; Benirschke et al., 2006; Rossant, 2007). Differences between the epigenetic states of the extra-embryonic and embryonic lineages have previously been demonstrated (Monk and Harper, 1979; Chapman et al., 1984; Mak et al., 2004; Okamoto et al., 2004). Although the exact role of the differential epigenetic modification between the lineages remains unclear, the growth and survival of the embryo relies on the formation of the placenta; the first and most important organ to develop.

This PhD study was prompted in part by the observation that the development of the human placenta is similar to the process of tumour invasion in cancer, where trophoblast cells behave like tumour cells as they invade maternal tissue to establish the maternal-fetal interface that is required for the survival of pregnancy (Damsky et al., 1992; Fisher and Damsky, 1993). Given that epigenetic events have been linked with the origin of some cancers (Baylin et al., 1986; Esteller et al., 2001; Herman and Baylin, 2003; Gebhard et al., 2006; Esteller, 2007), this study sought to investigate the epigenetic mechanisms regulating placental development to provide a mechanistic relationship between placental and cancer growth. Previous studies have documented the epigenetic silencing of tumour-suppressor genes in placenta as well as other key regulatory genes involved in cancer (Dokras et al., 2006; Rahnama et al., 2006; Chiu et al., 2007; Novakovic et al., 2008; Wong et al., 2008). However, when this study
began in 2007, the extent of tumour-associated epigenetic modification in the placenta was unknown.

Although the behaviour of trophoblast and tumour cells is remarkably similar, there are striking differences between their mechanisms of invasion (Fisher and Damsky, 1993; Soundararajan and Rao, 2004; Ferretti et al., 2007). The ‘controlled’ invasion of the trophoblast, ceasing after the adequate remodelling of maternal spiral arteries, has led some to refer to the placenta as a “well-behaved tumour” (Soundararajan and Rao, 2004). Thus one of the goals of this project was to identify epigenetically regulated genes that were associated with the cessation of trophoblast invasion at the end of the first trimester of pregnancy, with the hopes of translating any findings to tumour invasion in cancer. Given that few studies have access to human placental tissue from the first trimester of pregnancy, the availability of such precious tissues from Dunedin Hospital provided a special opportunity to study epigenetic modifications during this early stage in placental development.

Furthermore, placental-specific epigenetic modifications have been associated with the unique functions of the placenta – functions that include the simultaneous nourishment and protection of the fetus throughout pregnancy (Constancia et al., 2004; Coan et al., 2005; Ferguson-Smith et al., 2006; Hemberger, 2007). Previous studies have demonstrated the importance of epigenetic modification by showing that epigenetic aberrations interrupt normal placental and fetal growth (Rahnama et al., 2006; Hemberger, 2007; Serman et al., 2007). Furthermore, although the hypomethylated state of the placenta is not well understood, many believe that the low levels of methylation in this tissue are functionally significant (Migeon et al., 1985; Reiss et al., 2007; Lambertini et al., 2008; Cotton et al., 2009). However, most studies are based on animal models and cell lines, so the epigenetic environment within the human placenta remains relatively unexplored.

During the course of this study, the methylation patterns in the human placenta were examined to search for tumour-associated methylation and placental-specific methylation patterns that may establish a link between trophoblast and tumour invasion, and placental function and dysfunction. It was a fascinating journey - one that revealed new facets of epigenetic modifications both in the placenta and during the earliest stages of human embryonic development.
9.2 Genome-Wide Methylation Analysis of First-Trimester Placenta

At the beginning of this study, a genome-wide DNA methylation analysis was performed as a “gene discovery tool” to identify differentially methylated genes that may be regulating trophoblast invasion during the first trimester of pregnancy.

9.2.1 Sample Collection Results

Placental tissue was collected from women undergoing medical terminations of pregnancy in the first trimester (five to nine weeks gestation). DNA was extracted from placental villous cells (containing trophoblasts) and the fetal sex of each placental sample was determined to account for sex-specific methylation differences due to X chromosome inactivation in females. Of the 31 collected placental samples, 16 samples were from female fetuses and 15 samples were from male fetuses. Examination of X chromosome microsatellite repeats confirmed that maternal DNA was not present in any of the placental DNA samples.

9.2.2 MeDIP-Microarray Results

When this project began in 2007, the most popular and advanced technique for genome-wide methylation analysis was methylated DNA immunoprecipitation (MeDIP) combined with microarray technology. This method was selected for this study to determine the most common differences in gene promoter methylation between first-trimester placenta and peripheral blood DNA (the representative somatic tissue). Since X inactivation in females creates a large difference in total methylation between the sexes, placental samples from male and female fetuses needed to be separated. Due to cost restrictions, only female placental samples were chosen for this experiment because 1) the comparison tissue (peripheral blood) was females and 2) examining females samples meant that any differential X chromosome methylation could also be including in the data analysis. Samples were also pooled into groups to reduce the biological variation within each sample group and identify the most consistent differences in gene promoter methylation between first-trimester placenta and peripheral blood DNA. Furthermore, in order to identify methylation patterns in the placenta that were associated with gestational age and thus varying degrees of trophoblast invasion, first-trimester placental samples were split into three pooled groups based on gestational age. Three independent MeDIP-microarray analyses identified a total of 29 candidate differentially methylated genes between the groups of first-trimester placenta and peripheral blood DNA.
9.2.3 **Sequenom Results**

The Sequenom MassARRAY® EpiTyper® assay was used to confirm the methylation status of the candidate genes from the MeDIP-microarray experiment. Although Sequenom assays were successfully designed for 28 of the 29 candidate differentially methylated genes, the predicted methylation differences between first-trimester placenta and peripheral blood were confirmed in only three genes: **ARK1C4**, **KCNH5** and **AICDA**. A fourth gene, **ATAD4**, displayed a large difference in promoter methylation between first-trimester placenta and peripheral blood even though it failed to validate the MeDIP-microarray data analysis prediction. Interestingly, the promoter methylation of **KCNH5** was found to be lineage-specific: low in tissues derived from the extra-embryonic lineages (trophectoderm and primitive endoderm) and high in tissues derived from the epiblast. This methylation dichotomy was highlighted by the extreme contrast in promoter methylation between the adjacent and adherent chorion (13%) and amnion (86%) membranes.

9.3 **Functional Significance of Differentially Methylated Genes**

The levels of gene expression were examined using TaqMan® Gene Expression Assays to examine whether the differential promoter methylation of **AKR1C4**, **KCNH5**, **AICDA**, and **ATAD4** was inversely correlated with the expression levels of these genes. Since promoter methylation is usually associated with gene silencing, it was expected that the low levels of gene promoter methylation would correspond to detectable levels of gene expression, and high promoter methylation would be associated with no, or very low levels, of gene expression. Due to the wide range of tissues in these experiments, relative gene expression was normalised to total input RNA.

9.3.1 **Gene Expression Results**

Of the four genes examined, only one, **KCNH5**, displayed meaningful patterns of promoter methylation compared to gene expression. The expected inverse correlation was detected between **KCNH5** promoter methylation and expression in tissues derived from the extra-embryonic lineages (first-trimester and term placenta). Interestingly, the variable levels of **KCNH5** expression in the highly methylated epiblast-derived tissues combined with the presence of an alternative start site for **KCNH5** on the Genome Browser suggested that expression of this gene in epiblast-derived tissues might be regulated by an alternative promoter (Kent et al., 2002).
9.3.2 Results from Methylation and Expression Analysis of KCNH1

*KCNH5* (also known as *EAG2*) encodes a voltage-gated potassium channel and is involved in the regulation of cell cycle and proliferation (Pardo *et al.*, 1999; Wray, 2004). Although there is no literature documenting the function of *KCNH5* in the placenta, *KCNH5* is approximately 70% homologous to *KCNH1* (Bauer and Schwarz, 2001), and both of these genes are thought to be involved in cell cycle regulation and tumour progression in cancer (Pardo *et al.*, 1999; Hemmerlein *et al.*, 2006; Ding *et al.*, 2007; Wadhwa *et al.*, 2009). At this stage in the study, it was thought that if the patterns of promoter methylation and gene expression for *KCNH5* and *KCNH1* were similar, then these genes might also have similar functional roles in the developing placenta, and possibly in cancer. Thus, the Sequenom and TaqMan assays were used to quantify the promoter methylation and expression levels of *KCNH1*, respectively. When the promoter methylation of *KCNH1* was found to be low in all tissues, and *KCNH1* expression was not similar to *KCNH5* expression in the placenta or in cancerous tissues, *KCNH1* was excluded from further analysis.

9.4 Characterisation of KCNH5

Since it was reported that alternative splicing of *KCNH5* gives rise to various gene transcripts that use alternative first exons (Kent *et al.*, 2002), it was thought that the promoter regulating *KCNH5* expression in tissues derived from the extra-embryonic lineages (placenta and chorion) might not be regulating the expression of *KCNH5* in tissues derived from the epiblast.

9.4.1 Transcript-Specific RT-PCR Results

A transcript-specific RT-PCR was designed to distinguish the putative first exons of alternative transcripts of *KCNH5* (termed exon 1a and exon 1b for simplicity). The dichotomous methylation was located at the promoter of exon 1a (the upstream exon). Results from the transcript-specific RT-PCR showed exclusive transcription from exon 1a in placenta, and exclusive transcription from exon 1b in somatic tissues and amnion. These results confirmed that *KCNH5* promoter methylation regulates the lineage-specific expression of alternative transcripts: in tissues derived from extra-embryonic lineages, non-methylation of the upstream promoter permits transcription at exon 1a, whereas in tissues derived the embryonic (epiblast) lineage, transcription occurs from the downstream exon 1b. Subsequent
Sequenom analysis revealed that the promoter of exon 1b is unmethylated in all tissues, which confirmed that KCNH5 transcription in epiblast-derived tissues is permitted at this site.

9.4.2 **Western Blot Results**

Although the functional role for KCNH5 protein encoded by transcripts that use exon 1b has been reported in the literature (Saganich et al., 1999; Ludwig et al., 2000; Ju and Wray, 2002; Jeng et al., 2005), the function of the placental-specific transcripts that use exon 1a is unknown. Therefore KCNH5 protein was characterised in this study using an antibody that detects all isoforms. Western blot analysis confirmed that tissues derived from the extra-embryonic lineages use exon 1a for KCNH5 transcription, while tissues derived from the embryonic (epiblast) lineage initiate transcription at exon 1b. These results also confirmed that the levels of KCNH5 protein in placenta are similar to those in somatic tissues.

9.4.3 **Immunohistochemistry Results**

Immunohistochemistry on human placenta revealed that expression of KCNH5 was restricted to the mesenchymal cells of the villous stroma. There has been longstanding debate over the origin of the villous mesenchyme. Although it is agreed that mesenchyme is not derived from the trophectoderm, some believe it is epiblast-derived and develops from the caudal end of the primitive streak (Luckett, 1978; Downs, 2009), while others report that mesenchyme is derived form the primitive endoderm (Enders and King, 1988; Bianchi et al., 1993). Given that total placental KCNH5 transcription occurs exclusively from the unmethylated exon 1a promoter, the expressing villous mesenchymal cells must have the same unmethylated status at exon 1a and hence initiate KCNH5 transcription at this site. Since all tissues derived from the epiblast show dense methylation at the exon 1a promoter, it is likely that the villous stroma is not derived from the epiblast. Instead, the epigenetic and expression profiles for KCNH5 suggest that the villous mesenchyme is derived from the primitive endoderm. These results are supported by the previously reported epigenetic differences between the extra-embryonic lineages (trophectoderm and primitive endoderm) and embryonic lineage (epiblast) (Monk and Harper, 1979; Chapman et al., 1984; Mak et al., 2004; Okamoto et al., 2004).

9.4.4 **The Placental-Specific KCNH5 Transcript is a Result of Retrotransposition**

Further analysis of the region of differential KCNH5 methylation revealed that 147 bp of the upstream promoter and the first 162 bp of exon 1a has recently evolved from a SINE (AluY) retrotransposon (GRCh37 assembly chr14:63,568,599-63,568,907) (Kent et al., 2002). This
The AluY element is only present in humans, old world monkeys and apes (i.e. macaque, orangutan, and chimpanzee) suggesting that the transposition event occurred approximately 25 to 30 million years ago (Fleagle, 1999). Although other gene promoters have evolved from retro-elements, this placental-specific transcript of KCNH5 appears to be the first example of a human transcript derived from the insertion of a SINE element. Given that most Alu elements are methylated in placenta (Gama-Sosa et al., 1983; Schmid, 1991), the avoidance of methylation by the KCNH5 promoter suggests that this unique AluY-derived transcript maintains a functional role in the placentas of humans, old world monkeys and apes.

9.5 Methylation Analysis of Other Retrotransposon-Derived Genes in the Human Placenta

Since it has been previously reported that some genes show placental specific expression, and that a disproportionate number of these genes have retroviral elements in their regulatory regions (Bieche et al., 2003; Medstrand et al., 2005; Rawn and Cross, 2008), it was of interest to determine whether other genes showed a methylation dichotomy between tissues that were derived from embryonic versus extra-embryonic lineages. Five retrotransposon-derived genes in the placenta were selected for methylation analysis based on reports in the literature: ERVWE1, EBR, MID1, PTN and INSL4. Of these five genes, two displayed a lineage-specific methylation dichotomy that was similar to KCNH5.

ERVWE1 (the Syncytin-1 gene) is the most well studied retrotransposon-derived gene in the placenta, known for its critical role in the differentiation of the cytotrophoblast into syncytiotrophoblast (Blond et al., 2000; Mi et al., 2000; Frendo et al., 2003). Prior to this study, the methylation of ERVWE1 was reported to be approximately half that of somatic tissues and a cell line (Matouskova et al., 2006). Sequenom analysis confirmed the findings of Matouskova et al. although an even larger difference in ERVWE1 methylation was detected between the placenta and somatic tissues. Importantly, ERVWE1 promoter methylation was found to be low in extra-embryonic tissues (first-trimester placenta, term placenta and term chorion) and high in all epiblast-derived tissues.

INSL4 (insulin-like 4) is a primate-specific retrotransposon-derived gene and its promoter contains an endogenous retrovirus (HERV) element that is postulated to mediate its placental-specific expression (Chassin et al., 1995; Bellet et al., 1997; Bieche et al., 2003). Results from the Sequenom analysis determined that INSL4 promoter methylation was low in tissues
derived from extra-embryonic lineages and high in epiblast-derived tissues. Similar to $KCNH5$, the dichotomous methylation of $INSL4$ was highlighted by the difference between the trophectoderm-derived chorion (13%) and epiblast-derived amnion (93%).

For these three genes ($KCNH5$, $ERVWE1$ and $INSL4$), the level of promoter methylation in extra-embryonic tissues (including the chorion) is dichotomous with that in embryonic tissues (including the amnion). Importantly, the difference between chorionic and amniotic methylation indicates that the difference does not reflect the placenta itself, rather it reveals the origin of these tissues with respect to embryonic or extra-embryonic lineage commitment. Thus the expression of these three genes presumably reflects methylation-induced silencing of the exogenous DNA in epiblast-derived tissues, and escape from this mechanism in tissues derived from the extra-embryonic lineages. Furthermore, the epigenetic switch that silences expression of these retro-element-derived transcripts in the embryo occurs during the second cell fate decision, when the epiblast and primitive endoderm lineages differentiate from the inner-cell mass.

Sequenom analysis of $MID1$, $EBR$ and $PTN$ revealed that retrotransposon methylation in these genes may be established at a later stage in development or the methylation may be tissue-specific instead of lineage-specific. Although repeating this analysis may be useful, the low placental methylation of these retrotransposon-derived genes confirmed the use of their LTR-derived promoters as alternative gene transcription sites in the placenta. Overall, the results from this study suggest that the epigenetic reactivation of retrotransposons (either lineage-specifically or tissue-specifically) is the mechanism by which the placenta can selectively express these unique gene transcripts.

### 9.6 What is the Role of Retrotransposons in the Placenta?

Retrotransposons are sometimes referred to as “junk DNA” given their highly repetitive, non-functional sequences and “copy and paste” behaviour to increase their occupancy throughout the genome (Weiner et al., 1986; Finnegan, 1989). These deleterious sequences are usually silenced in most tissues by DNA methylation to prevent genome dysfunction (Yoder et al., 1997; Goll and Bestor, 2005). The placenta, however, is starting to attract new attention as a site where retrotransposon-activation may be of some advantage to genome function.
An increasing number of retrotransposons that are actively transcribed in the placenta have been identified, and a high proportion of these serve as alternative promoters to genes (Kato et al., 1987; Wilkinson et al., 1993; Schulte et al., 1996; Medstrand et al., 2001; Bieche et al., 2003; Landry and Mager, 2003; Okahara et al., 2004; Andersson et al., 2005; Medstrand et al., 2005; Kim et al., 2006; Yi et al., 2006). The majority of the transcribed retrotransposons in the placenta are endogenous retroviruses, and transcription is initiated in the LTR-region of the retroviral sequence. In fact, the human placenta has been noted for expressing more LTR-derived promoters than any other tissue; 40% of all known LTR-derived tissue-specific transcripts are expressed in the human placenta (Medstrand et al., 2005; Reiss et al., 2007). Even some imprinted genes in the placenta have arisen from exogenous DNA sequences, such as PEG10 and PEG11; two retrotransposon-derived paternally expressed imprinted genes that are important for placental development and function (Suzuki et al., 2007; Sekita et al., 2008).

This vast number of active retrotransposons in the placenta has led some to believe that the evolution of placenta-specific promoters by ancient retroviral infection has led to the unique invasive phenotype of the primate hemochorial placenta (Bieche et al., 2003; Cohen and Bischof, 2007; Rawn and Cross, 2008). This theory is supported by a selection of LTR-derived genes, which are specifically expressed in the placenta, as their functions range from preventing apoptosis to inducing cell proliferation and migration (Brandt et al., 2005; Magarinos et al., 2007; Heiss et al., 2008). Currently, it is unclear whether the hypomethylated state of the placenta has led to the activation of LTR-derived promoters in this tissue, or whether the use of LTR-derived promoters is an independent event to select for placental-specific gene transcripts that are of some advantage to placental function and evolution.

Some believe that the hypomethylated state of the placenta has contributed to the activation of retrotransposons, compared to the majority of other tissues in which these sequences are silenced by methylation (Yoder et al., 1997; Seifarth et al., 2005; Kudaka et al., 2008). However, a study by Reiss et al. examined the promoter methylation of nine LTR-derived genes in the placenta to investigate whether hypomethylation was a feature of the entire retroviral sequence or if it was specific to the LTR-derived promoter region (Reiss et al., 2007). The results of this study suggested that the methylation levels across the sequence of an endogenous retrovirus are variable, and the absence of methylation in the LTR region permits the use of this region as a promoter for transcription in the placenta. Furthermore,
Reiss et al. suggest that the hypomethylation of retrotransposon-derived genes in the placenta has parallels with tissue-specific gene promoter hypomethylation rather than being a consequence of the genome-wide hypomethylation in the placenta (Reiss et al., 2007).

9.7 What is the Epigenetic Link between Trophoblasts and Tumours?

One of the main goals of this study was to identify a link between the epigenetic mechanisms regulating the invasion of trophoblast and tumour cells. Although the KCNH5 transcript expressed in epiblast-derived tissues is thought to be involved in tumour progression in cancer (Pardo et al., 1999; Wadhwa et al., 2009), the function of the SINE-derived KCNH5 transcript in the placenta remains unknown. Since the levels of KCNH5 promoter methylation and expression in the placenta did not change significantly during the course of gestation, it is unlikely that this gene is involved in the regulation of trophoblast invasion. In addition, the localisation of KCNH5 in the placental villous mesenchyme and not in any of the trophoblast subpopulations suggests that this gene is not involved in the invasive behaviour of the trophoblast.

Interestingly, the loss of DNA methylation at certain repetitive sequences has been shown to be a hallmark in some cancers (Ehrlich, 2002; Rodriguez et al., 2008; Szpakowski et al., 2009). In particular, studies have shown Alu elements to be methylated in normal somatic tissues while unmethylated in some cancers (Rubin et al., 1994; Ehrlich, 2002; Rodriguez et al., 2008), suggesting transposon-induced genomic instability in tumour cells. Perhaps this unmethylated SINE (AluY)-derived transcript of KCNH5 has a tumour-associated function in the placenta that is not associated with invasion. Further analysis is required to determine the function of the KCNH5 transcript in the primate placenta and potentially in cancer.

After this study commenced, Novakovic et al. performed a genome-wide investigation in purified first-trimester cytotrophoblast cells and term placenta to identify methylation patterns in the placenta that were similar to those in tumour cells (Novakovic et al., 2008). They shared a similar hypothesis, expecting to find epigenetic similarities to account for the behavioural similarities of trophoblasts and tumours. In their study, a comprehensive list of tissue-specific differentially methylation regions (tDMRs) compiled by Rakyan et al. was used to identify tumour-associated genes that were enriched for methylation in the placenta (Rakyan et al., 2008). Surprisingly, after data was filtered and ontologically classified, only eight candidate genes were identified, two of which had previously been reported as
differentially methylated in placenta (APC and RASSF1A) (Chiu et al., 2007; Wong et al., 2008).

In their genome-wide search, Novakovic et al. identified hypermethylated promoters of four genes that functioned as negative regulators of Wnt signalling, a pathway that plays a central role in cellular function and specialisation (Novakovic et al., 2008). Interestingly, the promoter hypermethylation of these four genes was present in human and baboon placentas, but not in the mouse placenta, which suggested an increased occurrence of Wnt signalling in primate trophoblast invasion. Furthermore, this study confirmed previous work showing that hypermethylation of the APC tumour suppressor gene was a feature in the human placenta but not in the mouse placenta (Carter, 2007; Novakovic et al., 2008; Wong et al., 2008).

Although the behavioural phenotypes of trophoblasts and tumours are similar, the study by Novakovic et al. demonstrated that there is little evidence for the methylation-induced silencing of numerous tumour suppressor genes in the placenta that is characteristic of most cancers (Novakovic et al., 2008; Rakyan et al., 2008). Instead, tumour-associated methylation in the placenta is specific to a small subset of genes. Importantly, of the few genes they identified, the presence of tumour-associated hypermethylation in human and baboon placenta but not in the mouse placenta suggests that the epigenetic events that regulate the development of the invasive primate placenta are comparable to those that regulate tumour invasion in cancer (Novakovic et al., 2008).

9.8 Future Directions for Epigenetic Analyses in the Placenta

9.8.1 Methods of Epigenetic Analysis

Advancements in technology have greatly improved the tools available for genome-wide epigenetic analysis. When this study began, MeDIP combined with microarrays was the most advanced and appropriate tool for genome-wide methylation analysis. However, MeDIP has now been superseded by methylation sequencing techniques that allow genome-wide quantification of DNA methylation. As discussed in Chapter Four (4.4.2), many investigators report that candidate genes discovered by MeDIP have a low rate of successful validation. Some investigators have suggested that the poor specificity of the anti-5-methylcytosine antibody can interfere with its enrichment of methylated sequences since the antibody precipitates methylated cytosines irrespective of the proportion of methylated cytosines within
a given fragment (Prof. Ian Morison, PhD Supervisor, *personal communication*). For example, the presence of any methylation, even very low levels in proportion to CpG density, may be sufficient for immunoprecipitation, leading to poor distinction between sequences with high and low methylation. These predictions were supported in this study by the high failure rate of candidate genes when they were independently assessed using Sequenom. Colleagues at the Babraham Institute (Cambridge, UK) had to introduce stringent criteria in order to validate over 80% of their candidate genes identified from a MeDIP-microarray experiment (Farthing et al., 2008). Furthermore, although samples were pooled into groups in this study to minimise inter-individual variation, there were few technical replicates for the gestational groups of first-trimester placenta (one microarray per age group). Therefore, there was minimal opportunity to compensate for technical variation in microarray data, resulting in the identification of false candidate genes.

For future genome-wide methylation analyses, the latest and most advanced technology is second-generation bisulfite sequencing. This high throughput technique examines DNA methylation throughout the genome at the single nucleotide level (Meissner et al., 2005; Meissner et al., 2008). Second-generation bisulfite sequencing has recently transformed the process of genome-wide sequencing by using a random, non-targeted approach that avoids bias compared to other methods that rely on specific primer sets to amplify products for sequencing (Meissner et al., 2005). A particular method known as “reduced representation” bisulfite-sequencing (RRBS) has improved the technique for global methylation sequencing by enriching DNA fragments that contain CpG sites (Meissner et al., 2008; Smith et al., 2009). This technology has not only minimised the amount of post-sequencing data to analyse, but has substantially reduced the cost of sequencing the genome.

When this study commenced in 2007, it was tempting to utilise the second-generation sequencing technology for global methylation analysis; however, there was no computer software available at the time to interpret the large amounts of sequencing data. It was not until mid-2009 when the first bisulfite-sequencing mapping program (BSMAP) became freely available (Xi and Li, 2009). In addition, the cost of performing second-generation sequencing on bisulfite-converted DNA was too expensive at the time. At present, however, this new technology is being used for the first time in New Zealand in our laboratory at the University of Otago.
9.8.2 Placental Tissue

In addition to advances in genome-wide technology, a more sophisticated method of tissue dissection may improve the accuracy of candidate gene discovery. Although a gross dissection of first-trimester placental villi was carefully performed, it is likely that the placental DNA and RNA samples were extracted from various subpopulations of trophoblast cells, each of which may have unique epigenetic and gene expression profiles. In fact, Novakovic et al. identified cell-type epigenetic variation in their study, where the methylation patterns of the four genes regulating Wnt signalling were specific to the cell-type of the placenta (i.e., cytotrophoblast, extra-villous trophoblast, term placental cells) (Novakovic et al., 2008). Combining epigenetic and genetic profiles from a mixed cell population can lead to inaccurate quantification of promoter methylation and gene expression. Although the use of laser capture microdissection was considered for this study, the time and cost involved for selecting the large number of cells to obtain DNA and RNA from multiple tissue samples was considered prohibitive. However, it would be wise to use laser capture microdissection in the future in order to examine the unique methylation and expression profiles of candidate genes within the various subpopulation of placental trophoblast cells.

9.8.3 Future Studies

The results from this project have created a new platform from which to explore the epigenetic modifications in the placenta. Given the expanding list of retrotransposon-derived genes that are expressed in this unique extra-embryonic tissue, it is now important to investigate why the placenta allows for such potentially deleterious sequences to be expressed. It would be fascinating to identify other retrotransposon-derived genes in the placenta that display the lineage-specific methylation dichotomy, to examine the role of DNA methylation in regulating retrotransposon-derived genes and to understand the functional significance of these genes with respect to placental function and dysfunction. The results from this study have created a new working hypothesis, which states that these retrotransposon-derived placental-specific genes have some functional advantage that has evolved over time and contributed to the development, structure and function of the human placenta. Although the Syncytin-1 gene (ERVWE1) is a key example of a retrotransposon-derived gene that has an important functional role in the placenta, the aim of this new hypothesis is to expand beyond Syncytin-1 and identify other functionally significant genes in the placenta that have evolved from exogenous DNA sequences.
To test this hypothesis, second-generation bisulfite sequencing technology would be the most appropriate and effective tool to perform a genome-wide search for differentially methylated retrotransposon-derived genes in the placenta. A candidate-gene approach would also be useful, given that some retrotransposon-derived genes in the placenta have not yet been epigenetically analysed in a wide range of tissues, such as aromatase cytochrome P450 (CYP19) and Leptin (LEP) (Bi et al., 1997; van de Lagemaat et al., 2003; Rawn and Cross, 2008). Both of these LTR-derived genes are functionally important in the placenta; CYP19 is involved in the generation of oestrogens to regulate placental growth and parturition and is highly expressed in the syncytiotrophoblast layer of the human placenta (Fournet-Dulguerov et al., 1987; Simpson et al., 1994; Furbass et al., 2008), and LEP is important as it promotes trophoblast cell proliferation and survival (Magarinos et al., 2007). It would be interesting if the placental-specific expression of these genes, and any other genes that were identified after genome-wide analysis, were regulated by epigenetic events that were established during the early stages of cell lineage commitment, as shown by KCNH5, INSL4 and ERVWE1. It would also be fascinating to examine cancerous tissues for differentially methylated retrotransposon-derived genes in order to determine if DNA methylation is the common mechanism that directs the unsilencing of these selfish sequences in both the placenta and in cancer.

Future work investigating the role of retrotransposons in the placenta will be highly important to the fields of epigenetics and placental pathology. Firstly, the field of placental epigenetics is relatively unexplored, so understanding the role of DNA methylation in the placenta will greatly increase our knowledge of gene regulation in this unique tissue. Secondly, investigating the functional significance of these placental-specific genes may lead to a better understanding placental function and dysfunction. Given that successful pregnancy requires a functional placenta, the results from this research may provide the elusive solutions to placental disorders that commonly result in pathologies such as growth-restricted children, pre-term births or loss of pregnancy due to placental failure.

9.9 Conclusions

In this study, a genome-wide DNA methylation analysis was performed on first-trimester human placenta to identify epigenetic modifications that are potentially associated with the analogous phenotypes of trophoblasts and tumour cells. The differential methylation of one gene, KCNH5, led to the discovery of a placental-specific transcript that was derived from the insertion of a SINE retrotransposon. Not only did the identification of this gene add to the
expanding list of retrotransposon-derived genes in the placenta, but also the lineage-specific methylation of *KCNH5* revealed the epigenetic mechanism that establishes differential retrotransposon silencing between the extra-embryonic and embryonic lineages during the early stages of development. In this project, the examination of other placental-specific retrotransposon-derived genes in a wide range of tissues revealed that DNA methylation selectively silences or permits the expression of retrotransposons during the earliest stages of cell lineage specification, specifically when the epiblast and primitive endoderm differentiate from the inner-cell mass. Interestingly, the placenta is becoming increasingly known as a tissue in which these normally deleterious sequences are actively expressed. This is strikingly similar to events in cancer, where retrotransposons become unsilenced and cause genome instability, resulting in disease. Although the functional significance of these retrotransposon-derived genes in the placenta is not yet clear, their evolutionary conservation in the placentas of human and some primates suggest that they may have contributed to the highly invasive phenotype of the haemochorial placentas in these species. Overall, the results from this PhD study demonstrate that the human placenta is a relatively untapped source of epigenetic information, with the potential to reveal the important epigenetic events that occur during the earliest stages of human embryonic development that may also be implicated in disease.
References


Burnette WN (1981) "Western blotting": electrophoretic transfer of proteins from sodium dodecyl sulfate--polyacrylamide gels to unmodified nitrocellulose and radiographic detection with antibody and radiodinated protein A. Anal Biochem. 112: 195-203.


MacVector I (2010) 11.1.0


