Analysis of a putative imprinted locus within the TRAPPC9 intellectual disability gene

by

Rosalind Law

A thesis submitted in conformity with the requirement for the degree of Master of Science

Institute of Medical Science
University of Toronto

© by Rosalind Law 2014
Analysis of a putative imprinted locus within the TRAPPC9 intellectual disability gene

Rosalind Law

Master of Science
Institute of Medical Science
University of Toronto
2014

Individuals were identified carrying heterozygous copy number variations within the autosomal recessive intellectual disability gene TRAPPC9 (chr8q24.3) and lacking apparent mutations on the second allele, thus indicating the involvement of a conserved imprinting mechanism at this locus between mice (15qD3) and humans.

My primary objective was to identify whether a potentially equivalent to mouse Peg13, PEG13, located within an intron of TRAPPC9, was imprinted in various human tissues and cell lines.

To investigate imprinting at this locus, allelic expression of PEG13 was assessed using Sanger sequencing and pyrosequencing. To identify regulatory regions that may be responsible for variations in PEG13 allelic expression, methylation profiles at CpG islands were determined in human tissues and cell lines.

The PEG13 equivalent was found preferentially-expressed from the paternal allele. Its imprinted expression appears to be tissue-specific and conserved in human fetal brain. Expression patterns of PEG13 in human tissues of paralleled KCNK9.
Acknowledgements

I feel incredibly grateful for the opportunities I have experienced and the people I have encountered in the past two years. I am also thankful for those who have guided me through the unpredictable, yet rewarding field of medical research.

First and foremost, thank you to my supervisor Dr. John Vincent for taking a chance and allowing me to become part of your research team. You have been incredibly supportive, patient and understanding throughout my entire graduate experience, and had confidence in me even when I doubted myself. Despite your busy schedule and many commitments, you were always available and approachable. You always have best interests of your lab members and for that I could not have asked for a better mentor.

Thank you to my committee member Dr. Rosanna Weksberg for guiding me throughout my project from start to finish. Your insight and expertise has proven invaluable. I am also appreciative of your support and understanding.

To the Vincent lab post-doctoral fellows – Dr. Kirti Mital, Dr. Arshad Rafiq, Dr. Nasim Vasli and Dr. Hamid Hedari – thank you for sharing your experience, wealth of knowledge and expertise. Thank you for guiding me when I felt lost, empathizing with my frustrations and celebrating every milestone. To my fellow peers, Brian Degagne (MSc candidate) and Taimoor Sheikh (PhD. Candidate), thank you for making time spent in the lab and coursework more enjoyable. Your positivity, encouragement and words of wisdom always brightened my day.

Thank you to our research analyst Anna Mikhailov for always being there to answer my questions and for taking time out of your schedule to help me no matter how trivial the task.

The main techniques employed in this study would not have been possible without the expertise and hospitality of The Krembil Family Epigenetics Lab. Specifically, thank you to Dr. Tarang Khare (PDF), Sasha Ebrahimi (PhD. Candidate) and Ricardo Harripaul (PhD. Candidate) for teaching me the ropes of pyrosequencing and DNA methylation. Your knowledge in epigenetics continues to astound me.

Lastly, I wish to thank my family, specifically my sister Harriet and my parents, for your unwavering patience not only during this degree, but also in all my academic and non-academic endeavours. Thank you for the support and encouragement during the stressful and hard times, and for equally celebrating and rejoicing in the good. Without you, I would not be the person I am today.
Contributors

This project would not have been made possible without our study collaborators Dr. Miho Ishida (UCL, UK) and Dr. Kazuhiko Nakabayashi (NCCHD, Japan). I am grateful for the opportunity to be a small contributor to this project. Their work and unique experimental findings were integral components to this thesis. I would like to extend a special thank you to Dr. Ishida for allowing me to help validate her findings and for providing me with the key component of this project, the fetal tissues.
Table of Contents

Abstract ii
Acknowledgements iii
Contributors iv
Table of Contents v
List of Tables viii
List of Figures ix
List of Appendices x
Abbreviations xi
Declaration of Academic Achievement xiii

CHAPTER 1: Introduction 1
1.1 Neurodevelopment 1
  1.1.1 CNS organogenesis: neurulation 2
  1.1.2 Regional development of the brain 3
  1.1.3 Fundamental events in corticogenesis 4
1.3 Neurodevelopmental disorders 11
  1.3.1 Intellectual disability 11
  1.3.2 Epidemiology 13
  1.3.3 Autism spectrum disorders 14
  1.3.4 Comorbidities associated with ID and ASD 15
1.4 Aetiology of ID and ASD 16
  1.4.1 Environmental factors 17
  1.4.2 Genetic factors 17
  1.4.3 Copy number variations 19
1.5 Epigenetic Mechanisms 21
  1.5.1 DNA methylation and hydroxymethylation 23
  1.5.2 Histone modifications 25
  1.5.3 Long non-coding RNAs 27
  1.5.4 CGIs and transcription 28
1.6 Genomic imprinting 29
  1.6.1 Definition 29
  1.6.2 Functional organization of imprinted genes 31
  1.6.3 Methylation dynamics in development 32
  1.6.4 Conservation of imprinted genes 33
  1.6.5 Regulation of genomic imprinting: Models 35
  1.6.6 Deregulation of Imprinted Genes 38
1.7 Genomic imprinting in the brain 39
  1.7.1 Parent of origin effects in the brain 39
  1.7.2 Identification of imprinted genes 40
  1.7.3 Imprinting in neurodevelopment and brain function 41
1.8 Genomic imprinting at the TRAPPC9 locus 43

CHAPTER 2: Materials, Methods and Techniques 47
2.1 Introduction and study design 47
2.2 Statistical analysis 48
2.3 Sample quality and quantification 48
2.4 Sanger sequencing 48
2.4 Isolation of nucleic acids from whole blood 49
2.5 Human fibroblasts 50
  2.5.1 Culturing and growth conditions 50
  2.5.2 Cryogenic Storage 51
  2.6 cDNA Preparation for reverse transcription 51
2.7 Real-time quantitative PCR 51
  2.7.1 Principle 51
  2.7.2 Assay design and validation 55
  2.7.3 Experimental design 55
  2.7.4 Procedure 56
2.8 Pyrosequencing 56
  2.8.1 Principle of pyrosequencing 56
  2.8.2 Experimental design 57
  2.8.3 Pyrosequencing procedure 59

CHAPTER 3: CNV Analysis in ID and Autism Probands 60
3.1 TRAPPC9 60
  3.1.1 TRAPPC9 function in the CNS 60
  3.1.2 TRAPPC9 and ID 62
  3.1.3 Heterozygous CNVs at the TRAPPC9 locus 63
  3.1.4 Imprinting at the Chr15q13 domain 63
  3.1.5 Imprinting at the Chr8q24.3 domain 65
3.2 Materials and methods 69
  3.2.1 Sample Collection 69
  3.2.2 Whole Gene Sequencing 69
  3.2.3 CNV Analysis 70
3.3 Results 73
  3.3.1 TRAPPC9 CNVs confer phenotypic variation 73
  3.3.2 Validation of CNVs in human lymphocytes 74
  3.3.3 Relevance of characterizing PEG13 in humans 75

CHAPTER 4: Expression analysis of PEG13 80
4.1 Materials and methods: Allelic expression of PEG13 81
  4.1.1 Human fetal tissues 81
  4.1.2 Human fibroblasts and leukocytes 82
  4.1.3 Pyrosequencing validation 83
4.2 Materials and methods: Expression analysis of PEG13 in human tissues 83
  4.2.1 Sample collection 84
  4.2.2 cDNA synthesis 84
  4.2.3 RT-qPCR 84
4.3 Results 86
  4.3.1 PEG13 is a novel paternally expressed gene in human fetal brain 86
  4.3.2 Allelic expression of PEG13 is conserved in human fibroblasts 88
  4.3.3 PEG13 is highly expressed in human brain 95
  4.3.4 PEG13 in Autism and ID probands 95

CHAPTER 5: Methylation Analysis of CpG Islands 98
5.1 Materials and methods 99
5.1.1 Sample collection 99
5.1.2 Bisulfite conversion of DNA 100
5.1.3 Quantification of DNA methylation 102
5.1.4 PCR amplification 102
5.1.5 Pyrosequencing and statistical analysis 103
5.1.6 Combined bisulfite restriction analysis 103

5.2 Results 108
5.2.1 Identification of the PEG13-DMR 108
5.2.2 PEG13-CpG island is a maternal DMR 109
5.2.3 KCNK9-associated CpG island is hypomethylated 111
5.2.4 Methylation in ID and autism probands 111

CHAPTER 6: Summary and conclusions, general discussion and future directions 114
6.1 General discussion 114
6.2 Conclusions 125
6.3 Future directions 125

References 127

APPENDIX 1: Example of pyrograms created from a pyrosequencing program 152
APPENDIX 2: List of studies which have identified TRAPPC9 as a cause of NS-ARID 153
APPENDIX 3: Expression analysis of Trappc9 in mice 154
APPENDIX 4: Allelic expression analysis for PEG13 157
APPENDIX 5: Relative expression of alleles at PEG13 at SNPs rs2270409 160
APPENDIX 6: Identification of the PEG13-DMR 161
List of Tables

Table 1.1 List of neurodevelopmental and neuropsychiatric disorders caused by deregulated imprinting or demonstrate parent-of-origin effects
Table 2.1 List of individual fibroblast cell lines
Table 3.1 Summary of genes imprinted at human chr8q24.3 and the syntenic mouse region, chr15qD3.
Table 3.2 List of primers pairs used to sequence TRAPPC9 exons.
Table 3.3 qPCR primers for CNV validations
Table 3.4 List of ID and autism probands with heterozygous deletions/duplications overlapping TRAPPC9 and the CNV breakpoints according to gene assembly hg19
Table 3.5 List of tissues samples heterozygous for SNP rs4289794
Table 3.6 PCR assays for SNP loci at the putative PEG13 locus
Table 3.7 Oligonucleotides used for quantifying allelic expression of SNPs by pyrosequencing
Table 3.8 Summary of pyrosequencing results for PEG13 expression in human fetal tissues
Table 3.9 Number of informative heterozygous samples for SNP rs2270409 at PEG13
Table 3.10 Number of informative heterozygous samples for SNP rs4289794 at PEG13
Table 3.11 Number of informative heterozygous samples for SNP rs380221 at PEG13
Table 4.1 List and description of individuals from which sperm DNA was collected.
Table 4.2 List and description of control leukocyte DNA
Table 4.3 List of oligonucleotides used for bisulfite pyrosequencing and regions analyzed across all biological samples for sperm, leukocytes and fibroblasts
List of Figures

Figure 1.1  Mechanisms for deregulated imprinting in the brain
Figure 3.1  Schematic illustrating the breakpoints of CNVs spanning the TRAPPC9 locus
Figure 3.2  Schematic representation comparing the mouse and human TRAPPC9/KCNK9/AGO2/PEG13 imprinting cluster.
Figure 3.3  Validating heterozygous CNVs in autism and ID proband by qPCR
Figure 4.1  Schematic diagram of the 17th intron of TRAPPC9 and SNPs used to quantify allelic expression of PEG13
Figure 4.2  Electropherograms showing the imprinting status of PEG13.
Figure 4.3  Relative expressions of PEG13 alleles in human fetal tissues
Figure 4.4  Relative allelic expression of PEG13 in human fibroblasts.
Figure 4.5  Electropherograms showing the allelic bias of PEG13 in fibroblasts at rs2270409
Figure 4.7  Relative expressions of PEG13 alleles in fibroblasts at rs2270409
Figure 4.8  Electropherograms showing the allelic bias of PEG13 at rs4289794
Figure 4.9  Relative expressions of PEG13 alleles in fibroblasts at rs4289794
Figure 4.10  Electropherograms showing the allelic bias of PEG13 at rs380221
Figure 4.11  Relative expressions of PEG13 alleles in fibroblasts at rs380221
Figure 4.12  Electropherograms showing the allelic bias of PEG13
Figure 4.13  Genotyping for PEG13 expression in Autism and ID probands with heterozygous deletions proximal to the PEG13 locus
Figure 4.14  Relative expression of PEG13 and KCNK9 in human tissues and cells by qPCR
Figure 5.1  Bisulfite conversion of DNA
Figure 5.2  Methylation profiles at CpG island derived from bisulfite pyrosequencing in human peripheral blood leukocytes
Figure 5.3  Methylation profiles at CpG island derived from bisulfite pyrosequencing in human fibroblasts
Figure 5.4  Methylation profiles at CpG island derived from bisulfite pyrosequencing in human sperm
Figure 5.5  Schematic of the human TRAPPC9/PEG13 locus
Figure 5.6  Methylation at the KCNK9 CpG island derived from bisulfite pyrosequencing in human leukocytes and fibroblasts
Figure 5.7  Methylation at the PEG13 and KCNK9 CpG islands derived from bisulfite pyrosequencing in Autism and ID proband leukocytes
Figure 6.1  Proposed model for how PEG13 instigates imprinting of KCNK9 in the brain through higher order chromatin looping
## List of Appendices

<table>
<thead>
<tr>
<th>Appendix</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Appendix 1</td>
<td>Example of pyrograms</td>
</tr>
<tr>
<td>Appendix 2</td>
<td>List of studies which have identified TRAPPC9 as a cause of NS-ARID</td>
</tr>
<tr>
<td>Appendix 3</td>
<td>Allelic expression of TRAPPC9 in mice</td>
</tr>
<tr>
<td>Appendix 4</td>
<td>Electropherograms showing PEG13 expression in fetal tissues</td>
</tr>
<tr>
<td>Appendix 5</td>
<td>Relative expression of alleles at PEG13 at SNP rs2270409</td>
</tr>
<tr>
<td>Appendix 6</td>
<td>Identification of the PEG13-DMR.</td>
</tr>
</tbody>
</table>
## Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASD</td>
<td>Autism Spectrum Disorder</td>
</tr>
<tr>
<td>AS</td>
<td>Angelman Syndrome</td>
</tr>
<tr>
<td>ASE</td>
<td>Allele-Specific Expression</td>
</tr>
<tr>
<td>ASM</td>
<td>Allele-Specific Methylation</td>
</tr>
<tr>
<td>BWS</td>
<td>Beckwith-Wiedmann Syndrome</td>
</tr>
<tr>
<td>cDNA</td>
<td>Complementary DNA</td>
</tr>
<tr>
<td>CGI</td>
<td>CpG Island</td>
</tr>
<tr>
<td>ChIP</td>
<td>Chromatin Immunoprecipitation</td>
</tr>
<tr>
<td>CNV</td>
<td>Copy Number Variation</td>
</tr>
<tr>
<td>COBRA</td>
<td>Combined Bisulfite Restriction Analysis</td>
</tr>
<tr>
<td>CTCF</td>
<td>Ccctc-Binding Factor (Zinc Finger Protein)</td>
</tr>
<tr>
<td>CpG</td>
<td>Cytosine-phosphate-Guanine dinucleotide</td>
</tr>
<tr>
<td>CpH</td>
<td>Cytosine-phosphate-(adenine, thymine, cytosine)</td>
</tr>
<tr>
<td>DMEM</td>
<td>Dulbecco’s Modified Eagle Medium Nutrient Mixture</td>
</tr>
<tr>
<td>DMR</td>
<td>Differentially Methylated Region</td>
</tr>
<tr>
<td>DMSO</td>
<td>Dimethyl sulfoxide</td>
</tr>
<tr>
<td>DSM-V</td>
<td>The Diagnostic and Statistical Manual of Mental Disorders 5th Ed.</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic Acid</td>
</tr>
<tr>
<td>DNMT</td>
<td>DNA Methyltransferase</td>
</tr>
<tr>
<td>dNTP</td>
<td>Dinucleotide Tri-Phosphate</td>
</tr>
<tr>
<td>EDTA</td>
<td>Ethylenediaminetetraacetic Acid</td>
</tr>
<tr>
<td>FBS</td>
<td>Fetal Bovine Serum</td>
</tr>
<tr>
<td>GAPDH</td>
<td>Glyceraldehyde-3-Phosphate Dehydrogenase</td>
</tr>
<tr>
<td>HDAC1</td>
<td>Histone Deacetylase 1</td>
</tr>
<tr>
<td>IC</td>
<td>Imprinting Centre</td>
</tr>
<tr>
<td>ID</td>
<td>Intellectual Disability</td>
</tr>
<tr>
<td>IGF2</td>
<td>Insulin-Like Growth Factor 2</td>
</tr>
<tr>
<td>KCNK9</td>
<td>Potassium Channel Subfamily K Member 9</td>
</tr>
<tr>
<td>lncRNA</td>
<td>Long Non-coding RNA</td>
</tr>
<tr>
<td>MeCP2</td>
<td>Methyl-CpG Binding Protein</td>
</tr>
<tr>
<td>5mC</td>
<td>5-methyl Cytosine</td>
</tr>
<tr>
<td>5hmC</td>
<td>5-hydroxymethyl Cytosine</td>
</tr>
<tr>
<td>miRNA</td>
<td>Micro Ribonucleic Acid</td>
</tr>
<tr>
<td>ncRNA</td>
<td>Non-coding Ribonucleic Acid</td>
</tr>
<tr>
<td>NS-ID</td>
<td>Non-syndromic Intellectual Disability</td>
</tr>
<tr>
<td>NDD</td>
<td>Neurodevelopmental Disorder</td>
</tr>
<tr>
<td>NS-ARID</td>
<td>Non-syndromic Autosomal Recessive Intellectual Disability</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase Chain Reaction</td>
</tr>
<tr>
<td>PEG13</td>
<td>Paternally-expressed Gene 13</td>
</tr>
<tr>
<td>PGC</td>
<td>Primordial Germ Cell(s)</td>
</tr>
<tr>
<td>PTM</td>
<td>Post-translational Modification</td>
</tr>
<tr>
<td>PWS</td>
<td>Prader-Willi Syndrome</td>
</tr>
<tr>
<td>RT-qPCR</td>
<td>Real-time Quantitative PCR</td>
</tr>
<tr>
<td>RGC</td>
<td>Radial Glial Cell</td>
</tr>
<tr>
<td>RNA</td>
<td>Ribonucleic Acid</td>
</tr>
<tr>
<td>RT-PCR</td>
<td>Reverse Transcription PCR</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>SAM</td>
<td>S-Adenosyl Methionine</td>
</tr>
<tr>
<td>S-ID</td>
<td>Syndromic Intellectual Disability</td>
</tr>
<tr>
<td>siRNA</td>
<td>Small Interfering RNA</td>
</tr>
<tr>
<td>TRAPPC9</td>
<td>Trafficking Particle Protein Complex 9</td>
</tr>
<tr>
<td>TSS</td>
<td>Transcriptional Start Site</td>
</tr>
</tbody>
</table>
Declaration of Academic Achievement

All work in this thesis was completed by me except the following. Tissue processing of human fetal tissues and downstream applications, specifically nucleic acid extraction, RT-PCR and genotyping was conducted by Dr. Miho Ishida from University College London. Identification of the $PEG13$-DMR in fetal tissues was performed by Dr. Kazu Nakabayashi from the National Research Institute for Child Health and Development (Tokyo, Japan). DNA and RNA for fibroblast trios were provided by Dr. Rosanna Weksberg (Weksberg Lab, SickKids Hospital, Toronto). Sperm DNA was provided by The Krembil Family Epigenetics Lab (Centre for Addiction and Mental Health, Toronto). CNV validation and $TRAPPC9$ sequencing for Proband 1 was performed by Liana Kaufman (IMS, University of Toronto).
Chapter 1: Introduction

1.1 Neurodevelopment

Temporal and spatially regulated genetic programs, along with gene-environment interactions, underpin normal brain function and govern neurodevelopment from embryogenesis to adulthood (Cannon et al., 2003; Vaccarino et al., 2001; Walsh et al., 2008, review). As one of the first major organs to form, brain volume and gross brain regions – including substructures and various nuclei – are primarily established *in utero* (Kretschnmann et al., 1986), and mature and develop according to unique temporal schedules (Rice & Barone, 2000). Organizational and functional changes continue to occur during the first two years of postnatal life as the brain reaches adult size and synaptic contacts undergo further establishment and fine-tuning (Kretschnmann et al., 1986).

Considered the defining feature of mammalian evolution, the cerebral cortex with its six-layered neocortex is responsible for higher cognitive and executive functions including memory and learning (Casey et al., 2005). The cerebral cortex develops in the fetal period when most neurons acquire their spatial orientation, morphology and intercellular connectivity, where each cortical region follows different developmental trajectories e.g. visual cortex versus prefrontal cortex (Bhardwaj et al., 2006; Bourgeois, 1997; Huttenlocher & Dabholkar, 1997). Additional age-related changes occur well into young adulthood, making the cortex the last brain region to complete maturation (Caviness et al., 1996; Chugani et al., 1987; Giedd et al., 1999; Sowell et al., 2007). Animal studies, along with human postmortem and neuroimaging studies, have unraveled the conserved developmental programs of cortical development in mice and humans (Casey et al., 2005; Huttenlocher et al., 1982; Shaw et al., 2008). Although these processes differ in time scales among mammals (i.e. days in rodents versus weeks to months in humans) the sequence of events remain similar (Rice & Barone, 2000), with the foundation of neuronal architecture primarily established in fetal development (Clancy et al., 2007; Romijn et al., 1991). A complete review of neurodevelopment and the
associated molecular components are not within the scope of this thesis. However, to understand the critical periods and timing of neurodevelopment, I will briefly outline the main events.

1.1.1 CNS organogenesis: Neurulation

Neurodevelopment begins at the onset of organogenesis with the formation of the notochord. This cellular rod derives from proliferating ectodermal cells that migrated between the hypoblast and epiblast layers of the embryonic disc, known as the primitive streak, during gastrulation (Halacheva et al., 2011). Through a process termed neural induction, the notochord marks the primitive axis of the early embryo by patterning the surface of the ectoderm (the outer of the three germ layers) with a cadre of diffusible signaling molecules and extraembryonic interactions, which outline the anteroposterior axis of the neuroepithelium (Imuta et al., 2014; Beddington et al., 1994). Subsequently, multipotent stem cells between the notochord and epiblast differentiate into pseudostratified columnar epithelium, leading to the formation of the neural plate and signifying the start of neurulation (Tropepe et al., 2001).

The first morphological evidence of the human central nervous system (CNS) is discernible on gestational day (GD) 16 with the genesis of the neural plate (O’Rahilly & Gardner, 1979). On GD18, the sides of the neural plate thicken and elevate as neural folds, creating an invagination known as the neural groove (Copp et al., 2003). Eventually the neural folds meet and fuse together at the midline to close the neural plate (GD22) (Smith & Schoenwolf, 1997; Yamaguchi & Miura, 2013), beginning and ending with the closure of the rostral (anterior) neuropore and caudal (posterior) neuropore, respectively (Copp et al., 1990; O’Rahilly & Gardner, 1979). By the end of the third week of gestation (GD 26-28), a hollow neural tube is formed (Rice & Barone, 2000). The neural tube has apical-basal polarity, with the apical side facing inward and the basal side facing outward, which represents the pial surface of the developing brain (Götz & Huttner, 2005). A single layer of proliferating undifferentiated neuroepithelial cells, or neuroprogenitor cells, line the apical side of the
neural tube and will eventually become neurons or non-neuronal (glial) cells. Meanwhile, the cavity of the neural tube itself forms the ventricular system of the brain and comprises the lateral (first and second), third and fourth ventricles (Rice & Barone, 2000). Secondary neurulation also marks the emergence and subsequent segregation of neural crest cells from the neural tube, a population of cells lying between the neuroectoderm and the neural tube (Bronner-Fraser, 1994). Later in fetal development, neural crest cells differentiate and give rise to most elements of the peripheral nervous system including the ganglia, cranial nerves, Schwann cells and sensory neurons (Rice & Barone, 2000; Bronner-Fraser, 1994). The neural tube itself separates from the overlying neuroectoderm and develops into the spinal cord and gross regions of the brain.

1.1.2 Regional development of the brain

On GD 28, three primary enlargements (or vesicles) bud from the anterior portion of the neural tube in a caudal-to-rostral gradient, beginning with the rhombencephalon (hindbrain), followed by the mesencephalon (midbrain), and lastly the prosencephalon (forebrain) (Rice & Barone, 2000). The most rostral of the primary vesicles, the prosencephalon, gives rise to two secondary expansions known as the diencephalon (the optic vesicles and hypothalamus) and the telecephalon (Müller & O’Rahilly, 1987). The telencephalon forms the basis of the cerebral hemispheres and cortex. Similarly, two secondary outgrowths emerge from the rhombencephalon called the metencephalon (the pons and cerebellum) and the myelencephalon (the medulla oblongata) (Koop et al., 1986). The mesencephalon remains as a single unit and eventually develops into the brainstem, which includes the tectum and tegmentum. By mid-gestation, the neural tube resembles a globular-shaped brain with discernible features, including convolutions (gyri and sulci), and enlarged brain regions such as the cerebral hemispheres and cortex (Rice & Barone, 2000). The cerebral cortex arises from the anterior end of the neural tube because of proliferating and differentiating neural progenitor cells. This process, termed corticogenesis, can be defined by five successive but coordinated, overlapping
stages: (1) neurogenesis, characterized by neuroblasts differentiating and proliferating into both neuronal and non-neuronal cells; (2) cortical neurons migrating to laminar position; (3) synaptogenesis; (4) gliogenesis and myelination; and lastly (5) apoptosis, which trims and sculpts synaptic contacts (Rice & Barone, 2000). The relative time scale and length for corticogenesis is summarized in Figure 1.1.

1.1.3 Fundamental events in corticogenesis

Neurogenesis

Neurogenesis describes the process by which all cell types in the brain are derived from progenitor cells (Malik et al., 2013). Regional differences exist concerning the timing of neurogenesis and the rate of cytogenic proliferation. For the cortex, neurogenesis begins in the embryonic period at ~human GD 43 (Rakic, 1995). All cells of the CNS, regardless of function or morphology, originate from a layer of germinal neuroepithelial cells lining the lateral ventricles of the neural tube (Farkas & Huttner, 2008; O'Rourke et al., 1995). These neuroepithelial cells, otherwise classified as stem cells, exhibit self-renewing properties and are considered multipotent because of their ability to have neurogenic or gliogenic cell fates.

The layer of neuroepithelial progenitor cells thickens as neural and glial precursor cells are generated, creating a multilayer tissue that becomes the cerebral wall (Sidman & Rakic, 1973). A region closest to the ventricular surface of the neural tube known as the ventricular zone (VZ) separates from the pial surface; this region is the primary source of cortical neurons (Noctor et al., 2002). At the VZ, two horizontal neuronal layers form. First, the preplate, which arises above the VZ in the mid-embryonic period, contains the “pioneering” neurons (Götz & Huttner, 2005). Lastly, the subventricular zone (SVZ) forms between the VZ and the preplate late in the third trimester (rat GD17), and serves as another site for neuron production toward the end of fetal development (Rakic,
1975); the SVZ is the only germinal zone to persist postnatally (Martínez-Cerdeño et al., 2006; Meyer et al., 2000). Completion of the VZ proliferative zone signifies the onset of neurogenesis.

The pool of progenitor cells is composed primarily of radial glial cells (RGCs) and have two roles in neurodevelopment, serving as both progenitor cells (Malatesta et al., 2000) and as guides for migrating neurons (Rakic, 1972). These cells are highly polarized along their apical-basal axis, and depending on the stage in neurodevelopment, maintain contact with the ventricular surface (apical) and the pial surface (basal) via their long ascending and descending processes (Pinto & Götz, 2007). As cortical progenitor cells arise, their nuclei migrate along their apical-basal axis according to their progression in the cell cycle, resulting in changes to their cell morphology (McConnell, 1995, 1991; Sauer et al., 1935). In DNA replication, or S phase, nuclei are located in the outer third or half of the VZ, but rapidly descend to the ventricular surface in G2 phase to complete mitosis (McConnell, 1995). With entry into G1 phase, however, nuclei move outward to the pial surface where daughter cells can either reenter the cell cycle as a neural stem or progenitor cell or exit the VZ and differentiate.

Prior to neurogenesis, most progenitor cells divide symmetrically and duplicate the population of founder cells with each mitotic cycle, expanding the proliferative capacity (and size) of the developing cortex (Noctor et al, 2004). These rapid divisions generate two daughter cells with identical cell fates and remain in the VZ to undergo further divisions (Noctor et al., 2004). However, as neurodevelopment continues, asymmetric divisions steadily increase until following closure of the neural tube and later in neurogenesis, nearly all cells divide asymmetrically by two main mechanisms (Haubensak et al., 2003). First, neurogenic divisions produce another self-renewing radial glial progenitor that re-enters the mitotic cycle and a second, different cell type such as a non-stem progenitor cell or a neuron (Noctor et al., 2004; Götz et al., 2002). Secondly, asymmetric progenitor divisions can generate a self-renewing RGC and an intermediate progenitor (IP) cell that migrates to
the SVZ - all IP cells undergo symmetric, neurogenic divisions to either amplify the progenitor pool or produce neurons that may adopt a neuronal or glial cell fate (Nocter et al., 2004).

Following each subsequent division, precursor cells lose their multipotency and become progressively restricted from generating neuronal subtypes that have been born earlier in neurodevelopment. (Quian et al., 1998; Lillien et al., 1997). The coordinated process of neuronal fate specification is determined by an intrinsic timing mechanism at each developmental stage (Shen et al., 2006), thus enabling each of the emerging layers to contain a unique population of neuronal cells. Stem cell differentiation arises from the expression of transcription factors and signaling molecules (Vaccarino et al., 2001), which regulate a combination of cellular programs (Tropepe et al., 2001) that involve transforming growth factor B (TGF-B) signaling, and external cues such as fibroblast growth factors (Fgf) (Götz & Huttner, 2005). Beginning with the first asymmetric divisions, neuroblasts leave the SVZ and VZ and differentiate as they migrate to their final cortical positions. 

Neuronal migration

The cerebral cortex expands as post-mitotic neurons migrate from their germinal sites toward the pial (outer) surface where they undergo their final differentiation and become arranged into the six cortical layers (O’Rourke et al., 1992; Hatten, 1990). A neuron’s laminar position is determined by its time of origin, as neocortical layers form in an “inside-out-fashion”. As such, neurons in each layer share similar morphological and functional properties (McConnell et al., 1991). Around the 7th week of gestation, the first wave of neurons to migrate out of the VZ form a transient layer termed the preplate (Meyer et al., 2000; O’Rourke et al., 1995; 1992). The preplate neurons are loosely packed and are the first to differentiate, mature and receive synaptic contacts (Supèr et al., 1998); consequently, these neurons are considered the primitive functional cortex. Shortly thereafter, a subsequent wave of postmitotic neurons insinuates themselves into the preplate, splitting the layer
into a superficial marginal zone (MZ) and a deeper subplate (SP) layer (Meyer et al., 2000). Neurons in the MZ coalesce into a distinct monolayer, termed the cortical plate (CP), at 8-9 weeks gestation in humans (E12-E14 in rats) (Rakic, 1972). The CP grows and thickens as newly generated neurons migrate past earlier-generated cells to settle in progressively more superficial layers. Eventually, a six-layered neocortex is generated, classified as layers I (most superficial, outward pial layer) through IV (the deepest, and closest to the geometric center of the brain), and a subplate (layer VII) (Boulder Committee, 1970; Angevine & Sidman, 1961).

Gliogenesis and myelination

In parallel to corticogenesis, myelination and gliogenesis undergo regional and temporal maturation. Glial cells – oligodendrocytes and astrocytes – are similar to neural cells in that they arise from multipotent cells in the VZ and SVZ (Parnavelas et al., 1999). In CNS development, gliogenesis follows neurogenesis as progenitor cells acquire glia-like characteristics: an initial period of asymmetrical division, followed by rapid cell division and increased motility (Quian et al., 2000). Consequently, glioblasts are generated later in embryogenesis (Parnavelas et al., 1999), and continue to differentiate and mature in postnatal life. Accordingly, retroviral injections in rats showed that astrocytes are mainly formed in the third trimester and in postnatal development, while oligodendrocyts are generated exclusively in the postnatal period (Skoff et al., 1976; Sauvegot & Stiles, 2002). Mature oligodendrocytes appear simultaneously with myelination events, which occur exclusively in the postnatal period (Skoff et al., 1976).

Synaptogenesis

The majority of neurons are developed in gestation, however synapse formation and reorganization primarily occurs during the perinatal and early postnatal life (Huttenlocher, 1990). In mammals, synaptogenesis occurs as a heterochronous event across all cortical regions (Bourgeois, 1997;
Huttenlocher & Dabholkar, 1997), and age-related changes have also been reported (Huttenlocher et al., 1982; Huttenlocher et al., 1979).

The initial synapses form above and below the cortical plate within the MZ and SP shortly following neuronal migration and coincide with neurogenesis (Huttenlocher & Dabholkar, 1997; Balslev et al., 1996). Studies in human frontal cortex show evidence for synaptogenesis beginning as early as 6 months in gestation, which is the peak of neurogenesis and neuronal migration (Huttenlocher & Dabholkar, 1997; Huttenlocher et al., 1982). During this period, neurons originate from dendritic shafts of neurons (Bourgeois, 1997). Approximately two months before birth, however, synaptic density starts to increase primarily by the formation of dendritic spines, followed by an exponential phase in synaptic density growth between postnatal two to four months, and plateaus at approximately three years (Huttenlocher & Dabholkar, 1997; Huttenlocher et al., 1982). This period of neuron overproduction facilitates experience-dependent organization of synaptic contacts from the third year until puberty, resulting in functional and experienced-based processes (Bourgeois, 1997; Diamond et al., 1964). From adolescence to adulthood, synaptic density begins to decline due to a phase of rapid neuronal apoptosis, known as synaptic pruning, which reduces 60% of the original neuronal population and a subsequent decrease in the number of synapses (Huttenlocher & Dabholkar, 1997; Huttenlocher et al., 1982).

Events in CNS synaptogenesis have primarily been elucidated by in vitro studies using cultured neurons. The initial synaptic contacts are formed on the dendritic shafts of neurons, where dendritic filopodia of growth cones (Niell et al., 2004) extend toward their cellular targets as a result of changes to cell surface adhesion molecules and dynamic changes to the cytoskeleton (Mitchison & Kirschner, 1988). Upon reaching their synaptic contacts, axons and dendrites accrue different synaptic and stabilizing protein complexes, depending on whether the synapses are inhibitory or
excitatory (Friedman et al., 2000; Washbourne et al., 2002; Yoda & Davis, 2003). Glial cells have also been shown to enhance synapse formation and activity in cultured RGCs (Ullian et al., 20001).

Apoptosis

Programmed cell death, or apoptosis, systematically removes large numbers of neurons in some structures and serves to eliminate and trim synapses while strengthening others in neurodevelopment and early life. Unlike synaptogenesis, however, cells undergoing apoptosis in the brain are targeted synchronously as they are removed from the surrounding tissue (Huttenlocher & Dabholkar, 1997). A 40% decline in synaptic density occurs between puberty and adolescence (Petanjek et al., 2011), and is accompanied by decreases in neuronal density (Bourgeois & Rakic, 1993; Huttenlocher, 1978). In adult life, synaptic density stabilizes and implies the loss of neural plasticity (Huttenlocher, 1978).
Figure 1.1 Schematic representation of human neurodevelopment and events incorticogenesis. Corticogenesis comprises five overlapping, but sequential stages: (1) Neurogenesis, where undifferentiated neuronal and non-neuronal precursor cells proliferate and follow cell lineages; (2) Neuronal migration, differentiating neurons are guided towards the final laminar positions; (3) Synaptogenesis, the establishment of synaptic contacts; (4) Gliogenesis and myelination; and (5) Apoptosis, which trims and fine-tunes synaptic contacts.
1.3 Neurodevelopmental disorders

The critical period between post-conception to early postnatal life has a heightened sensitivity to genetic and environmental influences, and any deviations from strictly controlled developmental programs underlie an increased susceptibility for neurodevelopmental disorders (NDDs). By definition, NDDs are a group of conditions that begin in development and are characterized by developmental deficits (American Psychiatric Association (APA), 2013). Such deficits may manifest as impairments in personal, social, academic or occupational functioning. Autism spectrum disorders (ASD) and intellectual disability (ID) are among the most common and prevalent NDDs.

1.3.1 Intellectual disability

Intellectual disability (intellectual developmental disorder; ID), formerly known as mental retardation, encompasses deficits in both general mental abilities and practical, social and/or conceptual adaptive functions (APA, 2013). Signs and symptoms for ID are apparent before adulthood, with disorder onset characterized by the failure to reach expected milestones in early growth and development (APA, 2013). ID, for the most part, is a non-progressive disorder. However for certain genetic disorders there is a period of regression followed by stabilization, such as in Rett syndrome (Chahrour et al., 2007; Amir et al., 1999), while in others there is a progressive decline in intellectual function throughout the life course (Brereton et al., 2006; Carr, 2005). Individuals often cannot meet personal and social responsibility in one or more aspects of daily life such as occupational and academic pursuits as compared to peers of the same age, gender and socioeconomic background (APA, 2013).

Historically the diagnosis of ID has relied strictly on standardized tests (e.g. Wechsler scales and Vineland Adaptive Behaviour Scale) and placed an emphasis on intellectual ability (APA, 1994). ID was previously defined by an intellectual quotient (IQ) score of approximately two standard
deviations (SD) below the population (IQ≤70), and the classification of disorder severity (i.e. mild, moderate, severe, profound) was based solely on IQ score (APA, 1994; International Classification of Diseases, Version 10 (ICD10); World Health Organization, 1992). However, while intellectual deficits are central to ID, there is now recognition that this disorder requires a more flexible clinical scope and comprehensive approach for diagnosis. Recent revisions to The Diagnostic and Statistical Manual of Mental Disorders (5th ed.; DSM-5; APA, 2013) involve changes to ID terminology, as well as to diagnostic criteria that places equal importance on standardized tests and individualized clinical evaluations for measuring both intelligence and adaptive functions. When possible, interviews from guardians and the affected individual may also be included to make more accurate diagnoses (APA, 2013). The criteria for classifying ID severity currently focus on deficits in adaptive functioning, which is better suited for individuals who are functionally nonverbal or have communication impairments and cannot articulate their internal state of mind.

ID may also be subdivided according to the presence or absence of additional clinical symptoms. The term “syndromic ID” (S-ID) describes cases with obvious patterns of physiological, neuroradiological, and metabolic abnormalities. Common features include, but are not limited to, growth retardation; developmental anomalies of the brain (e.g. dysgenesis of the corpus callosum, lissencephaly and microcephaly); distinctive facial dysmorphologies (Sharp et al., 2007; Andrieux et al., 2010; Feero et al., 2012). Conversely, by definition, non-syndromic ID (NS-ID) or idiopathic ID has no apparent physical manifestations or associated disorder symptoms, making impaired intellectual function the sole diagnostic feature. However, because certain NS-ID individuals have specific symptoms such as epilepsy, mild microcephaly and mild behavioural disorders, the distinction between S-ID and NS-ID may sometimes be indistinguishable (Jamra et al., 2011).
1.3.2 Epidemiology

ID is a health care challenge and a socioeconomic burden in both developing and developed countries, owing in part to limited interventional methods and treatment plans. ID affects 1-3% of the global population (Chechlacz & Gleeson, 2003; Harris, 2006; Leonard & Wen, 2002), making it the leading neurodevelopmental disorder worldwide. A recent meta-analysis reported a prevalence of 10.37/1000 for all severities of ID (Maulik et al., 2010). Among those affected, mild ID (49 < IQ < 70) has the highest frequency (85%), followed by moderate ID (34 < IQ <50) at 9.0%, while severe (19 < IQ < 35) and profound (IQ < 20) ID is estimated to affect 4.0% and 2.0% of individuals, respectively (King et al., 2009; Stevenson et al., 2000). While the prevalence of severe ID remains relatively consistent across studies, the estimated prevalence rates for mild ID differs considerably, with some studies reporting that ID affects as few as 1 in 10,000 people (Rauch et al., 2012) to as many as 60 in 10,000 people (Roeleveld et al., 1997). Such marked differences may be attributed to inconsistencies across study methodologies with regards to study location, sample size, method of ascertainment and diagnostic criteria (e.g. ICD-10 vs. DSM-IV), and/or the statistical methods used (Leonard et al., 2003; Roeleveld et al., 1997). Despite the inaccuracy in measuring true prevalence rates, the associations between risk factors for ID are well-defined.

Ethnocultural, socio-economic and demographic factors positively correlate with susceptibility for ID. Indeed, ID consistently affects more males than females, with male-to-female ratios ranging from 1.5 to 3.1 (Leonard et al., 2003, 2011), and is generally ascribed to X-linked genetic factors (Ropers & Hamel, 2005; Roeleveld et al., 1997). Concerning ethnicity, African-American and Aboriginal children have an increased risked for ID, specifically for mild-to-moderate ID as compared to Caucasians and other major ethnic groups, which is concomitant to socioeconomic factors (Leonard et al., 2003). Indeed, epidemiological studies indicate that poverty has a direct association with risk for ID, as demonstrated by higher prevalence rates (i.e. two- to threefold) in developing countries as
compared to industrialized countries (Maulik & Darmstadt, 2007; Mercadante et al., 2009). High levels of consanguinity, infections, prenatal malnutrition, environmental toxins, adverse prenatal environment and limited access to interventions and preventative measures have contributed to the increased prevalence rates of ID in low-income countries (Patel et al., 2007; Durkin, 2002). Interestingly, while individuals with ID were once expected to have greater rates of mortality than the general population, their overall life expectancy has increased particularly in developed countries due to a decrease in social stigmatization, improved accessibility to mental health services and an increased knowledge of preventative measures (Janicki et al., 1999). However, in terms of health care conditions, the quality of life remains subpar for many affected individuals partly because an increased risk for concurrent disorders and comorbidities, including autism, often accompanies and aggravates ID pathophysiology, and life expectancy is significantly lower than for the non-ID population (Bittles et al., 2002)

1.3.3 Autism Spectrum Disorders

Autism is a behavioural syndrome, although unlike previous suppositions (Kanner, 1943), autism is not a distinct condition but rather a continuum of closely related developmental disorders collectively referred to as the “Autism spectrum disorders” (ASD). The new diagnostic criteria by the DSM-5 (APA, 2013) reflects this notion as autistic disorder (autism), Asperger’s disorder, childhood disintegrative disorder (CDD), pervasive developmental disorder not-otherwise-specified (PDD-NOS) and Rett Syndrome now fall under the same broad diagnostic category as opposed to separate entities with unique diagnostic criteria (APA, 2013). As such, ASDs are clinically defined by three core symptomatic domains: (1) Persistent deficits in social communication and social interaction; (2) Communication deviance; and (3) Restricted and repetitive patterns of behavior, interests, or activities (APA, 2013). Symptoms manifest as impairments in everyday functioning such as occupational and social domains. ASDs differ in the severity of symptoms, pattern of onset,
development of language, deterioration in skills once they have developed, and cognitive
development (Matson & Boisjoli, 2008). Autistic symptomatology typically emerges in the early
developmental period generally between 12 to 24 months of life when individuals experience a
regression following a normal developmental trajectory (Dawson et al., 2000; Maestro et al., 2002).
The broadening of autism nosology proves difficult for clinicians to make accurate, differential
diagnoses between ASD cases and ID cases without ASD (Matson & Shoemaker, 2009).

ASDs affect four times more males than females (Werling & Geschwind, 2013). Since the inception
of ASD nosology, the prevalence and incidence (limited) rates for all collective ASDs has greatly
increased and are now considered the second most common NDD behind ID. Recent estimates place
ASD frequency at a range that falls between 60 to 110 per 10,000 (Baird et al., 2006; Baxter et al.,
2014; Bolton et al., 2004; Chakrabarti et al., 2005). Given that ASDs appear to be increasingly
prevalent, a debate exists as to whether there is a true increase in the incidence of ASD in the general
population, or rather due to increased awareness, changes to autism inclusion criteria, or different
study methodologies (Matson & Kozlowski, 2011; Rutter et al., 2000).

1.3.4 Comorbidities associated with ID and ASD

Comorbidities and co-occurring disorders are common to ID and ASD, and include metabolic (Curtin
et al., 2010; Melville et al., 2007; Rimmer & Kiyoshi, 2006), epileptic (Bowley & Kerr, 2000;
Danielsson et al., 2005; Tuchman et al., 2002) and neuropsychiatric disorders (Leyfer et al., 2007).
Cross-sectional and longitudinal studies employing comparison groups consistently report higher
prevalence rates for comorbid psychopathology, DSM disorder, or disability in both children and
adults affected by ID or ASD than the general population (Bakken et al., 2010; Bradley et al., 2004;
Brereton, 2006; Leyfer et al., 2007; LoVullo et al., 2009; Simonoff et al., 2008; Tsankanikos et al.,
2006; Deb et al., 2001). ID and ASD share co-occurring disorders, including, but not limited to
schizophrenia (McCarthy et al., 2010), attention deficit hyperactivity disorder symptoms (Gadow et
al., 2004; Kenneth et al., 2006); obsessive compulsive disorders (McDougle et al., 1995), mood disorders (Dekker & Koot, 2003; Kim et al., 2000), anxiety disorders (Gillott et al., 2001), and major depressive disorders (Leyfer et al., 2006; Ghaziuddin et al., 2002). Most notably, ASD is also frequently diagnosed with ID, although people with ASD fall along the entire spectrum of intellectual ability. In addition, while not a defining feature, ID has been reported in approximately 70% of autistic cases (Larson et al., 2001; Bradley et al., 2004; Matson & Rivet, 2008). As certain authors note, however, recent twin studies estimate that the association between ASD and ID may be considerably lower at 40-45% (Yeargin et al., 2003); this discrepancy has been accredited to clinical ascertainment bias (Chakrabarti & Fombonne, 2005; Hoeskstra et al., 2009).

Dissimilarities exist concerning the rate and pattern of comorbid disorders between ID and ASD. Indeed, the degree of severity for psychopathologies is far greater in individuals with ASD as compared to those with ID (Matson et al., 2008; Bradley et al., 2004). Accordingly, individuals with ASD are more likely to present multiple comorbidities, and the frequency is more pronounced in people with ASD and severe ID than adults only affected by severe ID (Matson et al., 2008). Estimates for a co-occurring psychopathology are as high as 37% in children (Dekker & Koot, 2003) and 60.4% in adults with ID (Deb et al., 2001), as compared to 73% for ASD (Brereton, 2006). Despite these differences, the considerable overlap between these developmental disorders, particularly concerning symptomatology, comorbidities and descriptive characteristics, may partly be attributed to the overlapping etiological factors (Kwok & Cheung, 2007).

1.4 Aetiology of ID and ASD

As previously surmised, neurodevelopment presents windows of vulnerability to environmental and genetic perturbations, both of which have been positively associated with altered cognitive and neurological phenotypes. Because ID and ASD arise from a complex interplay between neurobiological, psychosocial, genetic and environmental factors (Mitchell, 2011; Heikura et al.,
the resulting clinical and genetic heterogeneity of these disorders poses challenges when determining specific etiological causes.

1.4.1 Environmental factors

Epidemiological and clinical studies lend support to environmental contributions to ID and ASD aetiology. These include maternal risk factors during peri- and pre-conception such as parental age (Krakowiak et al., 2012; Leonard et al., 2006; Meyer et al., 2012; Williams et al., 2008), as well as obstetric and perinatal complications (Gardener et al., 2009; Kolevzon et al., 2007; O’Dwyer, 1997). Other congenital contributors include infectious diseases (Gilad et al., 2007; Odeberg et al., 2007) and metabolic problems (Katz & Lazcano-Ponce, 2008). Prospective studies have found associations between the development of autism, developmental delay, and intellectual impairment in infants following intrauterine exposure to environmental neurotoxins in early prenatal life (Canfield et al., 2003; Durkin et al., 2000; Gillberg & Soderstrom, 2003; Grandjean & Landrigan, 2006; Stanwood et al., 2001). The impact of environmental exposures to ASD and ID remains part of ongoing research; however, genetic susceptibility is also commonly involved in aetiology of the disorder (Edelson & Saudino, 2009).

1.4.2 Genetic factors

The genetic contributions to ID and ASD susceptibility originate from population-based twin and family studies (Ronald & Hoekstra, 2011, review). A seminal study of 21 twin pairs by Folstein & Rutter (1977) observed a concordance rate of 36% for stringent autism and 82% in monozygotic (MZ) twins and 10% in dizygotic (DZ) twins for a slightly broad autistic phenotype and other cognitive disorders including ID. Subsequent twin studies applying broader diagnostic spectrums with equivalent (Steffenburg et al., 1989) and larger sample sizes (Bailey et al., 1995; Mason-Brothers et., 1985; Taniai et al., 2008; Rosenberg et al., 2009) found comparable results, where concordance rates ranged from 88% to 96% for MZ twins. Extreme autistic traits also demonstrate a
heritability component as observed in various twin samples (Edelson & Saudino, 2009; Hoekstra et al., 2007; Ronald et al., 2010) and have also been shown to be intergenerational (Constantino & Todd, 2005). Furthermore, non-autistic relatives of individuals with ASD share similar but milder cognitive profiles (Bailey et al., 1998, review; Bolton et al., 1994; Piven et al., 1997).

Genetics causes for ID and ASD range from straightforward to complex and heterogeneous. The most frequently diagnosed disorders with genetic etiology arise from cytogenic anomalies such as aneuploidies (e.g. Down’s syndrome and Turner’s syndrome), chromosomal rearrangements (e.g. balanced or unbalanced translocations and inversions), and common microdeletion syndromes (Rauch et al., 2006). An obvious pattern of male inheritance and the availability of family pedigrees for linkage analysis have led to the identification of over 100 genes in X-linked ID disorders, which accounts for the majority of all monogenic forms of ID (Lubs et al., 2012). The most common single-gene disorders associated with ASD and ID are Rett’s Syndrome with MECP2 (Amir et al., 1999) and Fragile X Syndrome with FMR1 (Loesch et al., 2002; Yudkin et al., 2014). Moreover, homozygosity mapping in consanguineous ID families carrying rare homozygous mutations (truncating, missense or nonsense) has led to the discovery of highly penetrant genes such as MAN1B1 (Rafiq et al., 2010) and TRAPPC9 (Mir et al., 2009). To date, ~ 40 autosomal recessive genes have been implicated in non-syndromic autosomal recessive ID (NS-ARID) (Musante & Ropers, 2014, review).

Genomic aberrations, which are generally detectable by conventional cytogenetic techniques or molecular karyotyping, have clear associations between the genetic factor and phenotypic profile (Sagoo et al., 2008), and explain causation for ~40% and 20% of ID and ASD cases respectively (Devlin & Scherer, 2012; Rauch et al., 2006). However, for most idiopathic cases, the association between pathology and etiology proves difficult, as the contributing genetic factors may be multigenic or not well defined. Accordingly, the advent of next-generation sequencing (NGS)
technology, genome-wide association studies (GWAS) and microarray platforms have led to the identification of numerous susceptibility loci and submicroscopic genetic variants implicated in these NDDs, among which include global and locus-specific structural variants.

1.4.3 Copy number variations

Copy number variations (CNVs) are segments of DNA ranging from 1 kilobase (kb) to several megabases (Mb) in size that cause quantitative changes to the DNA copy number, and encompass deletions (losses) and gains (duplications or insertional transpositions). These structural variants follow normal Mendelian inheritance patterns (Komura et al., 2006) or can be de novo in origin (Sebat et al., 2007), and reside within or span several genes, multiallelic regions, or complex structural rearrangements (Fredman et al., 2004; Sebat et al., 2004; Redon et al., 2006). CNVs account for 12% of the genome (Redon et al., 2006) and exhibit non-uniform distribution as certain regions are more susceptible for genomic rearrangements (Hastings et al., 2009; Stankiewicz & Luski, 2010). For example, regions harbouring repetitive sequences such as low-copy repeats (LCRs) and microsatellites have an increased occurrence of nonallelic-homologous recombination, which facilitate the formation of CNVs (Gu et al., 2008; Nguyen et al., 2006). Additionally, CNVs have enrichment at specific protein-coding genes where they affect gene dosage, and within telomeric and centromeric regions (Nguyen et al., 2006).

Numerous microarray-based studies have found an association between CNVs and disorder susceptibility for ID and ASD (Cooper et al., 2012; Devlin & Scherer, 2012; Iourov et al. 2012; Pinto et al., 2010; Marshall et al., 2008; Qiao et al., 2010; Sebat et al., 2007; Ullmann et al., 2007). Individuals with ASD have a higher frequency of CNVs as compared to the general population, and it is estimated that rare de novo and highly penetrant CNVs contribute to 5% of ASD cases (Devlin & Scherer, 2012). These CNVs have the tendency to overlap with genes of neurological relevance, specifically those involved in synaptic structure and function. Family and case studies have
confirmed the role of candidate genes in disorder etiology, among which include synaptic scaffold proteins \((SHANK1, SHANK2, SHANK3)\) (Berkel et al., 2010; Durand et al., 2006), proteins from the ubiquitin pathway (Glessner et al., 2009), postsynaptic adhesion molecules such as neuroligins \(NLGN3, NLGN2\) and \(NLGN4\) (Jamain et al., 2003; Laummonnier et al., 2004) and neurexins \(NRXN1\) and \(NRXN2\) (Gauthier et al., 2011). It is also not coincidental that most of the genes or CNV regions identified are implicated in ID (Berkel et al., 2010; Noor et al., 2010). Moreover, recurrent CNVs in patients sharing common phenotypic traits are categorized as microdeletion or microduplication syndromes. Common syndromes associated with ASD and ID include del/dup on16p11.2 (Weiss et al., 2008), hemizygous del on 7q11.23 which is associated with Williams-Beuren Syndrome (Osborne et al., 2001; Sanders et al., 2011) and del on 15q13.3 (Sharp et al., 2008).

Observations that pathological CNVs can be inherited from unaffected parents (Devlin & Scherer, 2012) and those with \textit{de novo} CNVs suffer from more severe ID and ASD phenotypes as compared to healthy individuals (Pinto et al., 2010), support the notion that different CNVs exhibit varying degrees of penetrance. Essential to this study is the fact that the occurrence of positive or negative sequelae depends on a) the genomic position of the CNV; and thus b) how the CNV exerts its effects on gene transcription and/or expression, in addition to the cell or tissue type in which the gene is expressed (Ramocki et al., 2008). Genetic syndromes with reciprocal CNVs (i.e. deletions and duplications in the same region) can result in different, but related phenotypes of varying severities. For instance, patients carrying hemizygous deletions (del) on 22q11.2 have developmental delay and CNS defects (Botto et al., 2003; Ensenauer et al., 2003; Shaikh et al., 2000), while those carrying duplications (dup) display more variable phenotypes (Wentzel et al., 2008). As indicated, CNVs can have direct or indirect influences on gene dosage and expression (Cook & Scherer, 2008). Direct mechanisms involve aberrations to an entire gene(s): duplications are associated with increased gene expression, while deletions to one allele cause decreased expression (Duker et al., 2010).
Homozygous deletions disrupt genes or create gene fusions and result in loss of gene function (Cerveira et al., 2006; Kleinjan & Heyningen, 2005; Potocki et al., 2007). Losses or gains outside of coding sequences (e.g. intergenic or intragenic regions) can disrupt cis-regulatory elements (DuBose et al., 2011). Alternatively, a CNV can alter the expression of dosage-sensitive genes by position effect, whereby the breakpoint boundaries affect the chromatin environment in which the gene resides or disrupts the interaction between regulatory elements controlling gene expression (Kleinjan & van Heyningen, 1998, review). For instance, in Potocki-Shaffer syndrome (PSS), a heterozygous deletion on chr11p11.2-p13 located downstream of the \textit{ALX4} gene was suggested to disrupt activity of a proximal enhancer via position effect and cause haploinsufficiency (Wakui et al., 2005). For genes that exhibit allelic expression, such as imprinted genes, downstream effects may also be contingent on whether the CNV resides on the maternal or the paternal chromosome (Girirajan et al., 2012; Martins-Taylor et al., 2014). However, it is important to mention that not all CNVs overlapping gene-rich regions are pathological.

Presently, the precise downstream effects of CNVs in ID and ASD at the functional and molecular levels remain poorly understood. Moreover, given that not all CNVs lead to direct changes in gene expression, supports the involvement of epigenetic mechanisms

1.5 Epigenetic Mechanisms

Epigenetic regulation refers to changes in gene expression without alterations to the underlying genetic sequence by virtue of reversible chemical modifications to chromatin. Under both physiological and pathological conditions, epigenetic processes mediate many cellular processes through changes to the transcriptome. Epigenetic modifications are involved in gene regulation and gene silencing (Berger, 2007), as well as cell differentiation (Kim et al., 2010), DNA replication (Sarraf & Stancheva, 2004), centromere inactivation (Simpson & Sullivan, 2010), and position effect variegation (PEV) (Festenstein et al., 1996; Reute & Spierer, 1992). The epigenome is dynamic in
response to intrinsic and extrinsic cues, integrating signals from developmental programs, hormones, environmental stimuli and stochastic events (Jaenisch & Bird, 2003). As such, different cell types possess unique epigenetic signatures according to their developmental trajectory, function and genotype (Milosavljevic, 2011).

Epigenetic mechanisms involve the interplay between chromatin modifications and chromatin remodeling, which interact with other regulatory proteins and non-coding RNAs (ncRNAs), specifically microRNAs (miRNAs), long ncRNAs, and small interfering RNAs (siRNAs). Different classes of histone- and DNA-modifying enzymes tightly regulate acquisition and removal of chromatin modifications, namely DNA methylation and posttranslational modifications (PTM) of histone proteins, both temporally and spatially (Reik, 2007). The concerted effort of these epigenetic modifications influences chromatin architecture.

Chromatin remodeling is the foundation of epigenetic regulation and describes the interconversion of chromatin to active (relaxed) or repressive (condensed) states, otherwise known as euchromatin and heterochromatin respectively (Arrighi & Hsu, 1971; Babu & Verma, 1987; Simmons, 2007). The manner by which nucleosomes interact with DNA underlies the conformation of chromatin. Each nucleosome contains ~147 bases of DNA wrapped twice around an octameric complex composed of dimers of H2A, H2B, H3 and H4; neighboring nucleosomes are separated by 10-50 base pairs of unlinked DNA (Luger et al., 1997). The epigenetic marks acquired by DNA and histone proteins affect the positioning of nucleosomes, thereby directing the accessibility of transcriptional machinery and protein complexes to their conjugate sequences, namely regulatory regions such as gene promoters, insulators, enhancers, and even within gene bodies (Berger, 2007; Reik, 2007). Inherent to the DNA-nucleosome model is the fact that nucleosomes recognize and exhibit different affinities for particular DNA sequences in cis (Segal et al, 2006). As such, transcriptionally active start sites are depleted of nucleosomes at the 5’end and 3’ untranslated region (UTR), enabling for the
assembly and disassembly of polymerase and transcription factors (Schones et al., 2008; Segal et al., 2006). These epigenetic marks guide the recruitment of ATP-dependent remodeling complexes (e.g. SWI/SNF2; Shen et al. 2000), methyl-CpG-binding proteins (e.g. MECP2; Fuks et al., 2003), and nuclear scaffolding proteins to the site of interest which facilitate changes in transcriptional activity (Berger, 2007). Taken together, epigenetic regulation requires the cross talk between DNA methylation, histone modifications, and in certain cases, ncRNAs, all of which will be addressed in the following sections with an emphasis on DNA methylation.

1.5.1 DNA methylation and hydroxymethylation

DNA methylation is a fundamental epigenetic modification in mammals with essential roles in embryonic development and gene regulation (Monk et al., 1987), while conferring genomic stability for X chromosome inactivation (Okamoto et al., 2004; Heard et al., 2001) and allele-specific expression for imprinted genes (Li et al., 1993). Deviations from normal methylation patterns have been associated with various human diseases, most notably in cancer (Baylin et al., 2005), imprinting disorders (Paulsen et al., 2001), and neuropsychiatric disorders (Feng & Fan, 2009). Additionally, global loss of methylation causes embryonic lethality in mammals (Li et al., 1992).

DNA methylation, often concentrated at repetitive regions and transposons in most eukaryotes, involves the addition of methyl groups to the five position of carbon rings of cytosine residues (5mC) and accounts for 1% (5x10^7) of total nucleotide bases in the mammalian genome (Ehrlich et al., 1982). The post-replicative addition of a methyl group to cytosine occurs through the action of DNA methyltransferases (DNMTs) (Bestor, 2000), which use S-adenosyl L-methionine (SAM) as a methyl donor (Chiang et al., 1996). Under oxidative conditions, 5mC marks undergo partial conversion to 5-hydroxymethylcytosine (5hmC) by ten-eleven translocation (TET) dioxygenases, which transfer a hydroxyl group to 5mC (Hee al., 2011). DNA methylation patterns are mainly established by de novo DNA methyltransferases DNMT3a and DNMT3b (Okano et al., 1999) and maintained by DNMT1 in
DNA replication (Bestor et al., 1998), which preferentially binds to hemimethylated DNA (Song et al., 2011; 2012). Fidelity of DNA methylation relies not only on DNMT1 activity, but also on the localization of different DNMT3 isoforms such as DNMT3L and other chromatin binding proteins to methylated DNA in certain contexts (Jones & Gangning, 2009). Changes in DNA methylation are introduced in DNA replication either due to de novo methylation or 5mC to Uracil transitions; nonetheless, the overall methylation status (e.g. hypomethylated; hypermethylated) of genomic elements essentially remains conserved following many cell divisions (Riggs et al., 2004; Ushijima et al., 2003). Meanwhile, removal of methylated marks can occur actively or passively. Passive demethylation takes place in the absence of maintenance DNMT1 during DNA replication. Conversely, evidence for active demethylation pathways come from non-replicating cells where loss of methylation was observed at specific loci in neurons and genome-wide in post-fertilized sperm in embryogenesis (Guo et al., 2011; Lee et al., 2002). The mechanisms by which active demethylation occurs are not well elucidated, although recent studies have proposed the involvement of TET enzymes and 5hmC as a reaction intermediate (Kangaspeska et al., 2008; Métivier et al., 2008).

In mammals, methylation preferentially occurs at cytosine-guanine dinucleotides (CpG) (Bird, 2002). Globally, CpGs are non-uniformly distributed in the genome and often methylated, which accounts for ~70% of total methylation. Most CpGs, however, are predominantly concentrated within CpG islands (CGI), genomic regions an average of 1000 bp in length distinguished by a high GC content (>50%) and an elevated frequency of CpG sites (≥ 0.6, the ratio of the number of methylated cytosines to the total number of methylated and unmethylated cytosine residues) (Bird et al., 1985; Rollins et al., 2005). Over half of CGIs coincide with the promoters of annotated genes (Saxonov et al., 2006; Illingsworth et al., 2010), mostly located upstream of transcriptional start sites (TSS) of housekeeping genes and enriched near genes restricted to tissue-specific expression in vertebrates (Gardiner-Garden & Frommer, 1987; Weberr et al., 2007). In general, methylation status is inversely
correlated to gene expression, whereby hypomethylation and hypermethylation are associated with
gene transcription and gene repression respectively (Bogadanovic & Veenstra, 2009); however,
certain CGIs retain their unmethylated status regardless of the transcriptional activity of their
associated gene (Antequera & Bird, 1988; Saxonov et al., 2006). Moreover, methylation is not limited
to CGIs. Sequences up to 2 kb outside of CGIs known as CGI shores display tissue-specific
methylation and have the ability to regulate gene expression (Doi et al., 2009; Irizarry et al., 2009).
By contrast, repetitive elements, namely LINE-1, SINE and Alu sequences, in addition to gene
bodies, are extensively methylated in human somatic tissues (Weisenberger et al., 2005). Methylation
within gene bodies has also been associated with gene repression (Bird, 2002; Wolf et al., 1984). It is
interesting to note that methylation can also exist in non-CG contexts (CpH, where H = A, T, C)
(Lister et al., 2013). Overall, non-CG methylation is a rare stochastic event in most human somatic
tissues (Lister et al., 2009), but more common in cultured pluripotent stem cells including ES cells
(Ziller et al., 2011) and the mouse germ line (Smith et al., 2012; Tomizawa et al., 2011), where it has
been associated with cell differentiation and proliferation. The brain is an exception: CpH
methylation represents a key epigenetic mark in the mammalian CNS with neuron-specific functions
(Lister et al., 2013). As previously indicated, crosstalk exists between DNA methylation and histone
modifications.

1.5.2 Histone modifications

Coordinated histone modifications bring about changes to chromatin architecture and define
chromatin domains (Cheung et al., 2000; Rice et al., 2003), otherwise known as the histone code
(Jenuwein & Allis, 2001). Each core histone, specifically H3 and H4, possesses two conserved N-
terminal amino “tails” which are highly basic due to arginine (R) and lysine (K) residues (Zheng et
al., 2003; Luger et al., 1997). Protruding from nucleosomes, histone tails undergo diverse covalent
modifications including acetylation, methylation, ubiquitylation, phosphorylation and sumoylation
(Peterson & Laniel, 2004) and are carried out by histone modifying enzymes (Kouzarides, 2007). Such modifications lead to corresponding changes in chromatin packaging, thereby affecting the accessibility of various chromatin binding proteins and RNA polymerase (RNAP) to the region of interest (Berger, 2002; Strahl & Allis, 2000). The best characterized modifications are acetylation and methylation of lysine and arginine residues, which are mediated by histone acetyltransferases (HATs), histone deacetylases (HDACs), and histone methyltransferases (Garcia-Cao et al., 2003; Peters et al., 2001; Zhang et al., 2001). Histone acetylation is restricted to lysine residues, which relaxes chromatin conformation by affecting the electrostatic interactions between negatively-charged DNA and positively-charged histone tails (Grunstein, 1997). By contrast, lysine residues can be found either mono-, di- or tri-methylated (Strahl et al., 1999), while arginine residues can be mono- or di-methylated (Kouzarides, 2007). Unlike acetylation, methylation is associated with both open and closed chromatin conformations and does not exert its effects by charge differences to histone tails.

Chromatin-immunoprecipitation (i.e. ChIP-chip) and ChiP-seq analyses has revealed different patterns of histone modifications demarcating functional elements in the genome (Zhou et al., 2011), and thus diverse regulatory consequences. Histone modifications present at gene promoters correlate with the level of transcriptional activity. Indeed, promoter regions of active genes show enrichment for tri- and monomethylation of lysine 4 of histone 3 (H3K4) surrounding TSS (Bernstein et al., 2002; Santos-Rosa et al., 2002; Young et al., 2011) in addition to augmented histone acetylation, which has been well-associated with gene expression and enhanced transcriptional activity (Allfrey et al., 1966; Hebbes et al., 1988). Collectively, these marks lead to open chromatin and reduce nucleosome occupancy. By contrast, trimethylations of H3K9, H3K27 and H3K79 are all linked to inactive promoters and heterochromatin (Barski et al., 2007; Volpe et al., 2002). It is important to note that while H3K27me3 is considered a repressive mark, it may be present at actively expressed
loci (Robertson et al., 2009). The histone profiles at enhancers (defined as cis regulatory elements that bind transcription factors and RNAP to facilitate activation of distal gene promoters) are cell-specific, and may display a combination of mono- or di-methylation of H3K4 and/or acetylation of lysine 27 of H3 (H3K27ac) (Heintzman et al., 2007; Kim et al., 2010), in addition to enrichment for the histone acetyltransferase p300 (Creyghton et al., 2010). Insulator elements, which serve as antagonists for gene expression by blocking enhancer-promoter interactions (Phillips & Corces, 2009), invariably require the binding of CCCTC binding factor (CTCF) to mediate chromatin looping and formation of higher order chromatin (Hou et al., 2008; Splinter et al., 2006). Special regions in the genome, known as bivalent domains, harbour both repressive (e.g. methylation of H3K27) and active (e.g. methylation of H3K4) histone modifications either at a single or adjacent nucleosomes (Bradley et al., 2006). Bivalent domains are located at genes “poised” for expression, which enables quick transitions from condensed to relaxed chromatin states for developmental genes in ES cells, as well as for some imprinted genes (Mikkelsen et al., 2007; Sanz et al., 2008). Moreover, histone modifications can work synergistically with long ncRNAs.

1.5.3 Long non-coding RNAs

Long ncRNAs, RNA transcripts defined as longer than 200 bp in length (Strahl et al., 1999), are organized in functionally demarcated regions, intervening with, or interspersed between coding and non-coding transcripts, as well as intronic and lying antisense to some imprinted and protein-coding genes (Bertone et al., 2004; Imanishi et al., 2004; Ota et al., 2004; Kapranov et al., 2007; Mercer et al., 2009). Many long ncRNAs are highly conserved between mice and humans and associated with epigenetic hallmarks of regulatory regions (Guttman et al., 2009). Supporting the biological relevance of these transcripts, intronic long ncRNAs exhibit restricted expression at specific developmental time points (Dinger et al., 2008; Ng et al., 2011), cell type or lineage (Dinger et al., 2007; Mercer et al., 2010), as well as subcellular compartmentalization as evidenced in mammalian
brain (Mercer et al., 2008; Sone et al., 2007). Most notably, long ncRNAs have temporal, regional and sub-cellular expression patterns in adult mouse and human brains, and parallel expression to proximal brain-specific protein-coding genes (Mercer et al., 2008; Mercer et al., 2010; Ponjavic et al., 2009). Such stringent expression profiles may be intrinsic to the regulatory function of long ncRNAs.

Indeed, long ncRNAs have an emerging role in transcriptional regulation of protein-coding genes by various epigenetic mechanisms, serving as guides and scaffolds for various chromatin-binding proteins. Many long ncRNAs have been implicated in chromatin organization by the recruitment of chromatin remodelling complexes (Khalil et al., 2009; Bertani et al., 2011). For example, the HOXC antisense long ncRNA HOTAIR recruits Poycomb Repressive Complex 2 (PRC2) and binds to LSD1, a histone demethylase, to promote heterochromatin formation and repression of the HOXD locus in trans (Rinn et al., 2007; Tasi et al., 2010). Many long ncRNAs function in large-scale changes to chromatin architecture, namely X chromosome inactivation which includes including Xist and Tsix (Tian et al., 2010; Wutz et al., 2002). Additionally, long ncRNAs have been shown to act as enhancers for gene expression (Orom et al., 2010; Orom & Sheikhattar et al., 2011), as evidenced by the expression of ncRNAs at enhancers enriched in methylation for H3K4 in mouse neurons (Kim et al., 2010). The mechanisms by which long ncRNAs influence enhancer activity is currently unknown. However, taken together, these findings support the role of long ncRNAs in locus-specific and allele-specific expression (Lee, 2009).

1.5.4 CGIs and transcription

The quantity and distribution of CGIs is well-conserved between mice and humans, including those corresponding to gene promoters, in addition to intergenic (within a gene) and intragenic (between genes) CGIs (Illingworth et al., 2010). As previously noted, ~50% of CGIs serve as intrinsic promoters at annotated genes and often lack methylation. Exceptions include CGIs associated with
imprinted genes and genes on the inactive X-chromosome, which acquire allele-specific methylation marks during development and remain transcriptionally silent throughout the life span (Illingsworth et al., 2008; Weber et al., 2007). However, the remaining half represent orphan CGIs, which are defined as intragenic or intergenic CGIs of unknown function (Illingsworth et al., 2010). Significantly, ~40% of orphan CGIs exhibit functional epigenetic marks characteristic of active gene promoters, including TSS, trimethylation of H3K4 and a general lack of methylation (Illingsworth et al., 2010). ChIP studies corroborate these findings, showing that ~21% of human orphan CGIs overlap with RNAP II peaks (Illingsworth et al., 2010), indicating the presence of an associated transcript. Indeed, intragenic CGIs located within introns or overlapping with exons of protein-coding genes serve as promoters for ncRNAs. Most notably, the paternally-expressed mouse long ncRNA, Air, is expressed from its CGI-promoter located in exon 2 of the protein-coding gene, Ifg2r (Sleutal et al., 2002). Moreover, orphan CGIs also exhibit cell- and tissue-specific methylation (Illingsworth et al., 2010; Maunakea et al., 2010). Taken together, DNA methylation, histone modifications and IncRNAs work synergistically to influence transient and long-term changes in gene expression, including genomic imprinting.

1.6 Genomic imprinting

1.6.1 Definition

For autosomal genes, genomic imprinting describes the process by which one parental allele is preferentially or predominantly expressed while the other is selectively silenced by epigenetic mechanisms (McGrath & Solter, 1984; Surani et al., 1990). The requirement for asymmetric functional contributions from the paternal and maternal genomes was first elucidated by nuclear transfer experiments in uniparental diploid (UPD) mouse embryos (Barton et al., 1984; Surani et al., 1984). Androgenetic (AG) embryos (paternal disomy) were embryonic lethal and failed to support embryonic development, while parthenogenetic (PG) embryos (maternal disomy) did not develop
extraembryonic tissues. Subsequent studies in AG and PG mouse chimeras (PG and AG embryos transfected with wild-type cells and carrying a lacZ transgene) circumvented this lethal phenotype and found that AG cells concentrated to mesodermal derivatives, while PG cells accumulated in ectodermal derivatives, including the epidermis and the brain (Surani et al., 1990). Taken together, these studies pointed to the competing interests of the parental genomes, in addition to the spatial and temporal regulation of genomic imprinting.

Imprinted genes are often tissue-specific and developmentally-regulated (Dünzinger et al., 2007; Smit et al., 2005). For instance, the paternally imprinted (or maternally expressed) mouse zinc-finger protein, Zim1, is highly expressed during embryogenesis, particularly in developing limb buds, but transcribed from both alleles in neonatal and adult brains (Kim et al., 1999). Paternal imprinting of the ubiquitin ligase, UBE3A, is restricted to the brain (Rougeulle et al., 1997). Imprinted genes also exhibit species-specificity, whereby those imprinted in eutherian mammals and marsupials for placental efficiency and X chromosome inactivation are not imprinted in monotremes, birds or amphibians (Gehring et al., 2009). A specific example is Commd1 in the mouse, which is paternally imprinted and contains a reciprocally expressed U2af1-rs1; in contrast, COMMD1 in humans shows expression from both parental alleles and does not contain a U2af1-rs1 homologue (Zhang et al., 2006). Additionally, genomic imprinting can be restricted to a gene isoform (Smit et al., 2005; Arnaud et al., 2003). Most splice variants of the human GRB10 gene are paternally expressed in fetal brain; however, one isoform is exclusively transcribed from the maternal allele in skeletal muscle and placenta while the remaining transcripts are biallelically expressed in all other fetal somatic tissues (Blagitko et al., 2000). Moreover, the mouse insulin-like growth factor II receptor (Igf2r) and its antisense transcript Air are reciprocally-imprinted in cortical glial cells and fibroblasts, but not in neurons where Igf2r is transcribed from both alleles and Air is not expressed (Yamasaki et al., 2005), thereby illustrating that imprinting can also be restricted to a cell type. Such strict regulation may be
attributed to the essential biological roles of genomic imprinting and the resulting severe disorders that arise from their dysregulation.

1.6.2 Functional organization of imprinted genes

Genomic imprints are heritable epigenetic modifications established in germ cells on parental chromosomes (Challet et al., 1991), such that following fertilization differential marks are maintained throughout development and propagated to somatic tissues, conferring allele specific expression (ASE).

Imprinted genes are arranged in chromosomal clusters throughout the genome and respond to regulatory signals in cis from imprinting centers (ICs) which can range up to several bp in size, are CG-rich and often correspond to CGIs. How cells discriminate between parental alleles at imprinted loci is not well understood, however differential DNA methylation at regulatory elements has been identified as a key epigenetic regulator for differential expression. Observations for the inheritance of parent-specific methylation patterns (Bartolomei et al., 1993; Silva et al., 1998; Stöger et al., 1993), in conjunction with experimental evidence showing mice deficient in DNMT activity with loss of monoallelic expression (Li et al., 1993), illustrate the functional role of DNA methylation in eliciting ASE. Accordingly, imprinted expression is often controlled by differentially methylated regions (DMRs) which harbour parent-dependent allele-specific methylation (ASM) in addition to other epigenetic modifications such as histone signatures (Reik & Walter et al., 2001). ICs or DMRs disseminate gene regulation over long distances (up to 1 Mb) and control expression of many loci independent of the parent of origin (Reik & Walter, 2001). However, while DMRs are synonymous with imprinted genes, it is important to mention that not all DMRs function as ICs (Rakyan et al., 2008). For example, IGF2R is biallelically expressed in human tissues despite the presence of a DMR, and its tissue-specific imprinted expression is dependent upon parent-specific histone modifications (Vu et al., 2004). Moreover, evolutionarily-conserved genes such as those involved in
olfaction possess DMRs despite their biallelic expression. Coordinated gene expression within imprinting clusters is therefore mainly dependent upon (1) differential epigenetic marks at regulatory elements and (2) interactions between neighboring genes and their control sequences.

1.6.3 Methylation dynamics in development

Erasure of genomic imprints is essential for resetting parental imprints at each generation (Morgan et al., 2005) and occurs in mammalian development, whereby the germ line undergoes two waves of genome-wide epigenetic reprogramming. Primordial germ cell (PGCs) are subjected to demethylation and includes erasure of parental imprinting marks, which are subsequently re-established later in gametogenesis according to the sex of the embryo (Feng et al., 2010). The second wave of demethylation occurs at post-fertilization; however, methylated imprinted marks are maintained throughout embryogenesis and read in somatic tissues in postnatal life (Khatib et al., 2007; Ueda et al., 2000). The timescale and extent of epigenetic reconfiguration differs according to species and cell type e.g. sperm and oocytes.

Erasure and re-establishment

In the developing embryo, primordial germ cells (PGCs) are highly methylated and, because of imprinted genes, display parent-specific epigenetic marks (Hajkova et al., 2000). Proliferating PGCs migrate to the gonadal ridges where they undergo rapid genome-wide demethylation, which results in the loss of DMRs at imprinted genes, and is concomitant to the loss of H3K9me2 (Lees-Murdock et al., 2003; Lee et al., 2002; Seki et al., 2007). Once in the gonads, PGCs commit to a germ cell fate according to whether the gonadal ridge develops into an ovary or testis, by which time oocytes and sperm enter into meiotic and mitotic arrest, respectively. De novo methylation by DNMT1 (Jajkova et al., 2002; Okano et al., 1999; Webster et al., 2005) occurs in the prospermatogonia stage in males (Kato et al., 2007), and after birth in maturing oocytes prior to ovulation in females (Morgan et al.
Here, parent-specific methylation is established according to the sex of the embryo, however methylated marks are dependent on DNMT3A and its cofactor DNMT3L (Demars & Gicquel, 2012).

### Maintenance and propagation

Following fertilization, the genome undergoes a second wave of demethylation followed by another wave of *de novo* methylation in the preimplantation embryo (Santos et al., 2002); however imprinted genes along with some other sequences retain their gametic methylation (Reik & Walter, 2001). Hours after fertilization, most sequences in the paternal genome are actively demethylated in the zygote just prior to DNA replication (Mayer et al., 2000; Oswald et al., 2000). In contrast, the maternal methylome undergoes passive demethylation during cleavage (Santos et al., 2002; Oswald et al., 2000). Genome-wide *de novo* methylation is believed to occur after implantation in the inner cell mass of the blastocyst by DNMT3a and DNMT3b (Okano et al., 1999), and is essential for establishing somatic methylation patterns (Ratnam et al., 2002). After fertilization, methylation is maintained by DNMT1o (oocyte variant) in the zygote, and then by DNMT1s in embryonic and adult tissues (Howell et al., 2001).

### 1.6.4 Conservation of imprinted genes

Genomic imprinting is limited to a small proportion of genes: ~ 100 and ~50 imprinted genes have been identified in mice ([http://www.har.mrc.ac.uk/services](http://www.har.mrc.ac.uk/services)) and humans ([http://igc.otago.ac.nz/home.html](http://igc.otago.ac.nz/home.html)), respectively, and account for 0.5-1.0% of 20,000 protein-coding genes in the genome. Cross-species comparisons have revealed conservation for imprinting across various phylogenies including higher plants and mammals (Feil & Berger, 2008), and maintenance of syntenic clusters and paralogues amongst eutherians and marsupials (Renfree et al., 2009), as well as between amphibians, fish and chickens (Dünzinger et al., 2007). A main school of thought is that genomic imprinting likely evolved from duplication and translocation events of preimprinted chromosomal regions in lower vertebrates (Dünzinger et al., 2007; Walter & Paulsen, 2003;
Yokomine et al., 2005). The evolutionary conservation for imprinting can partly be explained by the parental conflict hypothesis (Moore & Haig, 1991), which corroborates observations for conflicting interests of maternal and paternal genes in maternal resource allocation to developing offspring (Moore & Haig, 1991): paternal genes promote growth of his offspring while maternal genes limit growth by conserving her resources for survival while providing nourishment to her current and subsequent litters. As such, the hypothesis postulates that the function of an imprinted gene should correlate with the allele from which it is expressed (Moore & Haig, 1991).

In line with this school of thought is the observation that a disproportionate number (~38%) of imprinted genes are dedicated to embryonic development in placental mammals (Coan et al., 2005; Monk et al., 2006; Reik et al., 2004). Extraembryonic tissues are enriched in genes affecting placental growth along with its structural tissues (Constancia et al., 2002; Li et al., 1999; Mayer et al., 2000), and in regulating the provision of maternal nutrients to the fetus (Dao et al., 1998; Mizuno et al., 2002). However, although mice and humans exhibit remarkable concordance for imprinted genes themselves and their syntenic regions, as well as parental origins for the active or silenced allele (Tycko et al., 2002), only 30% of imprinted genes overlap between mice and humans (Okazaki et al., 2002; Monk et al., 2005; Morison et al., 2006; Su et al., 2004). The discordance can be attributed to loss of imprinting status (e.g. biallelic in human vs. monoallelic in mouse) or absence of orthologous genes in humans (Garg et al., 2012; Suzuki et al., 2011; Morcos et al., 2011). This has been primarily observed for imprinted genes in the placenta (Monk et al., 2006), which is not surprising given that most humans pregnancies are singletons as opposed to pregnancies of multiple births, which promotes intralitter competition and places pressure on placental efficiency (Reik et al. 2003). Consequently, the limited conservation of orthologues between mice and humans in the placenta suggests that those which remain conserved in humans are not only remaining evolutionarily critical, but are also required in postnatal adaptation and maternal behaviour as opposed to resource
provision and fetal growth (Keverne, 2001; Lefebvre et al., 1998; Li et al., 1999). Accordingly, it is estimated that ~20% of imprinted genes are involved in neurological functions (Renfree et al., 2009).

1.6.5 Regulation of genomic imprinting

Differential epigenetic marks, namely DNA methylation, of key regulatory elements associated with imprinted genes are inherently recognized by the appropriate transcriptional machinery, thereby resulting in differential gene expression (Hutter et al., 2006). ICRs can differ in their mode of action according to the imprinting cluster; nonetheless, common mechanisms have emerged through studies in mouse chimeras and transgenic mice carrying allele-specific deletions or duplications within these regulatory elements (Fitzpatrick et al., 2002; Lin et al., 2003; Thorvaldsen et al., 1998). Such mechanisms include sex-specific modifications of promoter sequences, silencers, enhancers, boundary elements, and overlapping antisense transcripts at imprinted loci (Constância et al., 1998; Reik & Walter, 2001).

Promoter methylation

Transcriptional silencing at one allele is commonly achieved by methylation at promoter regions and TSSs which tend to be CpG rich (Weber et al., 2007). Hypermethylation of CpG dinucleotides can inhibit gene expression by direct or indirect mechanisms. First, methylated CpG dinucleotides can prevent transcription factors from binding directly to DNA sequences (Watt & Molloy, 1988; Iguchi-Ariga & Schaffer, 1989). Alternatively, methylated DNA can attract transcriptional inhibitors such as methyl-CpG binding proteins (MBDs) (Hendrich & Bird, 1998). These MBDs can recognize methylated DNA, leading to the assembly of repression complexes comprising of transcriptional repressors, HDACs, and DNA methyltransferases (Hendrich & Bird, 1998). In vitro, MeCP2 was shown to localize to a transcriptional–repression domain and was associated with the co-repressor mSin3a (Nan et al., 1999). The subsequent condensation of DNA around histone cores leads to
heterochromatin formation, making the region inaccessible to transcription factors and results in gene repression (Yamashita et al., 2005).

The Insulator Model

Some imprinting clusters are regulated by insulators, which are DNA boundary elements that inhibit the interaction between two regulatory elements, such as promoters and enhancers (Reinhart et al., 2002). The best characterized imprinting clusters for this model is located on human 11p15.5, which harbours two imprinted domains controlled by two ICs referred to as IC1 and IC2 (Bartolomei et al., 1991). IC1 exhibits asymmetric methylation between the parental alleles and regulates monoallelic expression of the paternally-expressed gene IGF2 and the maternally-expressed non-coding RNA H19, and is located 2 kb upstream of the H19 promoter (Bartolomei et al., 1991). As such, the methylated status of this insulator influences interactions between the H19 and IGF2 promoters and their shared enhancer elements which lie downstream of H19 (Murrell et al., 2004). Indeed, while IC1 is methylated in the paternal allele, its unmethylated status on the maternal allele facilitates the binding of the insulator protein CTCF to the element (Szabo et al., 2000); this prevents both de novo methylation at the IC and enhancers from interacting with the IGF2 promoter, resulting in H19 expression and IGF2 suppression (Hark et al., 2000; Bell et al., 2000). Binding of CTCF facilitates higher chromatin structure and creates distinct chromatin boundaries between the IGF2 and H19 genes on the maternal allele (Li et al., 2008; Murrell et al., 2004; Tabano et al., 2010). Conversely on the paternal allele, the methylated IC inhibits CTCF-binding and enables for interactions between the IGF2 enhancer and downstream enhancers (Li et al., 2008). Loss of the insulator results in maternal expression of IGF2 in specific tissues (Thorvaidsen et al., 1998).

The Non-coding RNA Model

The majority of imprinting domains contain overlapping antisense transcripts, most of which encode short and long ncRNAs. In general, these closely juxtaposed transcripts are reciprocally-imprinted
relative to the surrounding sense gene, such that if the ncRNA is expressed on the paternal allele, the sense gene will be expressed from the maternal allele (Beechey et al., 2001). Most of the ncRNAs identified thus far are themselves imprinted and paternally-expressed, and have been shown to regulate imprinting clusters (Reik & Walter, 2001).

The promoter regions of some antisense transcripts are CpG-rich or CGIs and therefore serve as DMRs; the methylated state of an ncRNAs promoter determines the downstream effects of the transcript on the expression of its overlapping protein-coding genes. Indeed, ncRNAs act as long-range cis silencing factors, demonstrating spatial control within an imprinting domain and can be located within promoter regions of sense genes or within intronic regions (Royo & Cavaillé, 2008). Notable imprinting clusters that express IncRNAs include the Gnas locus (Li et al., 2000), the Dlk1/Gtl2 locus (Lin et al., 2003) and the Snrpn locus (Tsai et al., 2005; Rodriguez et al., 2005), though the best described are the Igf2r and Kcnql1 imprinting clusters and their overlapping ncRNA transcripts, Airn and Kcnq1ot1, which originate in the introns of their respective sense genes.

The Igf2r imprinting cluster comprises three maternally-expressed genes, Igf2r and solute carriers, Slc22a2 and Slc22a3, and one paternally-expressed gene, the macro long ncRNA Airn, amongst a few other non-imprinted genes (Sleutals et al., 2003b). The IC for this imprinting cluster also serves as the promoter for Airn (Sleutals et al., 2002a). On the paternal allele, the unmethylated IC results in Airn expression, which overlaps the promoters of protein-coding genes in an antisense direction and results in the subsequent suppression of surrounding genes (Lyle et al., 2000; Sleutal et al., 2003b). Conversely, the IC is hypermethylated on the maternal allele, which prevents the transcription of Airn and allows for the transcription of the protein-coding genes Igf2r, Slc22a2, and Slc22a3 (Sleutals et al., 2002a). Parallel to Igf2r, the Kcnq1 locus also contains several maternally-expressed genes and one paternally-expressed gene encoding a long ncRNA, Kcnq1ot1, whose promoter also
acts as an IC (Macini-Dinardo et al., 2006). The IC, termed KvDMR1, is hypermethylated on the maternal allele which suppresses the ncRNA Kcnq1ot1 and activates adjacent imprinted protein-coding genes (Thakur et al., 2004). Kcnq1ot1 is hypothesized to target repressive chromatin modifying complexes to imprinted genes for silencing, and is not the only mechanism thought to regulate imprinting at the Kcnq locus (Pandey et al., 2008). Deletion of both of these DMRs leads to loss of expression of the antisense transcripts and loss of imprinting of the sense genes (Yamasaki et al., 2005).

1.6.6 Deregulation of imprinted genes

The balance of paternally and maternally imprinted genes is critical to normal development and any deviations from the imprinting process in gametogenesis increases the risk lead to imprinting disorders and abnormal developmental outcomes (Chamberlain et al., 2010). Genetic and environmental influences determine the outcomes of epigenetic programming, and defects encompassing either factor, in addition to epigenetic causes, have been associated with explicit imprinting disorders (Lucifero et al., 2004). However, imprinting disorders differ from sex-linked disorders concerning inheritance, and thus, the resulting phenotypes in offspring. Indeed, the most striking consequence of parent-of-origin-effects (POE) is bias in the inheritance of traits, such that maternal traits are exclusively inherited down the matriline and paternal traits are inherited exclusively down the patriline (Reik et al., 2001). Consequently imprinted traits show non-mendelian patterns of inheritance, result in some traits skipping generations, and effect individuals of either sex equally (Reik & Walter, 2001). In other words, the effects of genomic imprinting on autosomal genes is indifferent to the sex of the offspring (Davies et al., 2004), but the parental origin of the allele is critical. This differs from sex-linked traits, whereby the trait depends on the sex of the offspring and consequently affects males and females differently. Thus taken together, the phenotype in offspring resulting from an imprinting disorder or POE depends on the allele from which it is inherited.
Accordingly, directional effects caused by deregulation of imprinted genes have been well-documented at human chromosome 11p15.5 and the orthologous mouse region found on distal chromosome 7, which contains the paternally-expressed insulin-like growth factor 2 (IGF2/igf2) and maternally-expressed genes H19/h19, CDKN1C/cdkn1c, ASCL2/ascl2, TSSC3/tssc3 and KCNQ1/kcnq1 (Kim et al., 1999; Fowden et al., 2006). Accordingly, mutations derived from the paternal alleles are associated with Beckwith-Wiedemann Syndrome (BWS), a prenatal overgrowth syndrome characterized by gigantism, large organs and increased risk for childhood cancer (Ping et al., 1989; Elliott & Maher, 1994), whereas mutations originating from the maternal allele result in Russell-Silver Syndrome, a disorder characterized by intrauterine growth retardation, low birth weight (LBW) and congenital hemihypertrophy (Patton, 1988). Such mutations include inherited or sporadic uniparental disomy (Cooper et al., 2005; Slatter et al., 1994), parental isodisomy (Cooper et al., 2005), chromosomal deletions (Niemitz et al., 2004), as well as abnormal DNA methylation (Catchpole et al., 1997). In addition, while the significance for imprinting and the underlying mechanisms has been well-elucidated in development and fetal processes, emerging evidence in the CNS suggests that imprinting is a critical epigenetic process for neurodevelopment and brain function (Keverne, 2001).

1.7 Evidence for Imprinting in the CNS

Parallel work in humans and mice provide evidence for imprinting and parent of origin effects in neurodevelopment, cognition and behaviour (Wang et al., 2010; Wilkinson et al., 2007; Ideraabdullah et al., 2009).

1.7.1 Parent of origin effects in the brain

The most prominent evidence for imprinted genes in brain growth comes from AG-N and PG-N mouse chimeras. Seminal studies by Allen et al. (1995) and Keverne et al. (1996) found that brains
derived from uniparental chimeras have different phenotypes concerning cell localization and brain size. Most notably, PG and AG cells displayed reciprocal patterns of distribution in specific regions of the brain: PG cells were found concentrated in the cortex but selectively excluded from the hypothalamus; the opposite arrangement was seen with AG cells. These sex-specific patterns persisted throughout the lifespan for both AG and PG animals and were readily observed in early fetal development (~E11) and maintained in adult life (Allen et al., 1995). In addition to reciprocal spatial distributions, neurodevelopmental abnormalities were observed in the mouse chimeras. AG-N chimeras, despite having increased body weight, had reduced brain weight, and a reduced brain/body ratio relative to control chimeras. This phenotype became more pronounced as the degree of AG chimerism (greater contribution from AG cells) increased. Conversely, PG-N chimeras had reduced body weight, and when normalized for body weight, an increased brain/body ratio (specifically a larger forebrain) Given that brain size is normally correlated to body size, this paradoxical effect suggests the maternal and paternal genomes have functional differences in the allocation of resources to the brain and body development. The study authors hypothesized that these observed neuroanatomical differences were the result of parent-specific biases to gene expression and thus brain function (Allen et al., 1995; Keverne et al., 1996).

1.7.2 Identification of imprinted genes

Genome-wide analyses have identified imprinted genes in mouse brain and mapped their respective expression patterns (Gregg et al, 2011; Gregg et al., 2010; Wang et al., 2008). Using whole transcriptome sequencing (RNA-seq) and allele-specific SNPs, Gregg and colleagues (2010) identified parent-of-origin effects in over 1300 loci (both protein-coding and ncRNAs) in embryonic and adult brains. Most notably, 204 putative imprinted clusters were identified, and 52% of these candidate clusters containing both coding and putative noncoding loci. This is particularly interesting, since as previously indicated, non-coding genes can regulate imprinting clusters (Pandey
et al., 2008; Sleutels et al., 2002). Moreover, concomitant to previous studies, these genes showed parental bias in the expression of individual genes and of specific transcript isoforms, developmental period, and different brain regions, revealing a dichotomy regarding the influence of the paternal genomes in brain function where the maternal genome was the main contributor to the embryonic brain in development (Gregg et al., 2010). However, unlike the observations made in the mouse chimera studies (Keverne et al., 1996; Allen et al., 1995), the paternal genome had marked influences in the adult brain, namely in the hypothalamus, cerebellum and cortex. Individual functional studies demonstrate that imprinted genes have diverse roles in the brain, including cell cycle regulation (Arima et al., 2005), regulatory ncRNAs (Sleutals et al., 2002), transcription factors (Keverne et al., 1996), protein trafficking (Greer et al., 2010) and RNA processing (Runte et al., 2001), cellular functions which underlie proper neurodevelopmental and behavioural programs. The genes identified to date exhibit both dynamic spatial and temporal regulation (Davies et al., 2004). For instance, maternal expression of H19 is detectable in human fetal brain (6-12 weeks gestation) where it is also highly expressed, but becomes downregulated and restricted to the pons and globus pallidus in adulthood (Pham et al., 1998). Conversely, mouse Igf2 undergoes monoallelic expression in the adult hypothalamus and pons (Pham et al., 1998). Alternatively, imprinted expression of the snoRNA SNRPN remains imprinted in adult brain from development (Runte et al., 2001). Imprinting for some genes may also be restricted to the brain, as demonstrated by UBE3A, with expression exclusively from the maternal allele (Vu et al., 1997).

1.7.3 Imprinting in neurodevelopment and brain function

Previous work on mouse models with targeted mutations at imprinted loci has been associated with cognitive and behavioural perturbations (review by Isles et al., 2000). PG chimeras demonstrated significantly more aggressive behaviour in adult life (Allen et al., 1995). Mice with maternal deficiency for Ube3a, an ubiquitin protein ligase with monoallelic expression restricted to mitral
cells of the olfactory bulb, as well as in hippocampal and Purkinje neurons, had impaired foetal growth, motor dysfunction and learning deficits (Albrecht et al., 1997; Rouguelle et al., 1997). Additionally, the absence of Gabrb3 expression in mice results in learning and memory deficits, in conjunction to poor motor skills and hyperactivity (Delorey et al., 1998). Lefevbre et al., (1998) showed that female mice deficient in the maternally imprinted gene, Mest (known as Peg1), had abnormal maternal behaviour, in addition to embryonic growth retardation and decreased life expectancy in the postnatal period.

<table>
<thead>
<tr>
<th>Syndrome/Disorder</th>
<th>Parental allele</th>
<th>Common chromosomal defect/de novo or inherited</th>
<th>Disrupted or candidate genes</th>
<th>Clinical phenotype</th>
<th>Reference (first author, year)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Angelman Syndrome</td>
<td>M</td>
<td>Chr15q11-13 Maternal deletion/de novo or inherited</td>
<td>UBE3A, ZNF127, NDN, SNRPN, IPW, GABA_A receptor subtypes</td>
<td>Absent speech  Severe learning disability Dysmorphic facial features Happy affect Inappropriate laughter Epileptic seizures</td>
<td>Buiting, 1995, Hogart, 2009, Knoll, 1989, Sutcliffe, 1994</td>
</tr>
<tr>
<td>Beckwith-Widemann Syndrome</td>
<td>M/P</td>
<td>Chr11p15.5 Abnormal methylation</td>
<td>IGF2, CDKN1C</td>
<td>Macroglossia Macrosomia Organomegaly</td>
<td>Sparago, 2004</td>
</tr>
<tr>
<td>Russell-Silver Syndrome</td>
<td>M</td>
<td>Chr11p15 (disruption to IC)</td>
<td>Hypomethylation at H19 and IGF2</td>
<td>Intrauterine and postnatal growth retardation Relative macrocephaly</td>
<td>Giquel, 2005</td>
</tr>
<tr>
<td>Schizophrenia</td>
<td>P/M</td>
<td>15q13-q14</td>
<td>Genes within this region</td>
<td>Disorganized speech and thinking Paranoia Auditory hallucinations</td>
<td>Freedman, 1997, Ingason, 2011</td>
</tr>
<tr>
<td>Autism</td>
<td>M</td>
<td>15q11-q13</td>
<td>Genes within this region</td>
<td>Developmental delay Social and communication deficits Stereotypies</td>
<td>Cook, 1997, Schroer, 1998</td>
</tr>
</tbody>
</table>

Table 1.1. Imprinting and neuropsychiatric disorders in humans. Explicit imprinting disorders, neuropsychiatric and neurodevelopmental disorders, and their associated genomic defects and clinical
phenotypes. The imprinting clusters and imprinting mechanisms governing their respective chromosomal regions are well-elucidated. *M = maternal and P = paternal; UPD = Uniparental disomy

Work in humans on the patterns of inheritance for several cognitive and neurological disorders widely acknowledges the importance of imprinted genes in the brain. Indeed, sexual dimorphisms in human brain function are most evident in explicit imprinting syndromes, with characterization of specific genetic factors in mouse models. Such conditions, which may be associated with disrupted neurological function and abnormalities, manifest according to the parental origin of the genetic aberration, as summarized in Table 1.1. For example, cytogenetic defects at chromosome 15q11-q13 lead to the neurogenetic disorders Prader-Willi (PWS) or Angelman Syndrome (AS), depending on whether the mutation is located on the maternal or the paternal allele. PWS, characterized by infantile hypotonia, obesity, developmental delay and ID, can arise from one of three mechanisms on the paternal allele: mutations at the IC, maternal UPD caused by non-disjunction, and paternal deletions, all of which involve loss of expression of paternally-derived genes (Cassidy et al., 2000; Sutcliffe et al., 1994; Veltman et al., 2004). PWS individuals have neuroanatomical variations, including anomalies to the anterior pituitary gland, ventriculomegaly and abnormal cortical development (Miller et al., 1996; 2007; Yoshii et al., 2002). In contrast, AS is caused by reciprocal aberrations to PWS, specifically maternal deletions, paternal UPD, or mutations in the genes exclusively or preferentially expressed form the maternal allele, specifically UBE3A (Kishiro et al., 1997; Knoll et al., 1989; Malcolm et al., 1991; Malzac et al., 1998). Individuals with AS often have ID, seizures, motor dysfunction and a happy disposition (Dyken et al., 1992; 2000), concomitant to microcephaly with mild cerebral atrophy marked by the loss of Purkinje and granule cells (Dorries et al., 1988; Jay et al., 1991). Additionally, certain regions demonstrate linkage of behavioural traits with the parent-of-origin (Francks et al., 2003; Francks et al., 2011).
1.8 Genomic at the TRAPPC9 locus

The trafficking protein particle complex 9 (TRAPPC9; chr8q24.3) gene has been implicated in autosomal recessive ID (Kakar et al., 2012; Marangi et al., 2012; Jamra et al., 2011; Koifman et al., 2010; Phillipe et al., 2009; Mir et al., 2009; Mochida et al., 2009). Homozygous frameshift or nonsense mutations in TRAPPC9 cause nonsense-mediated mRNA decay and protein truncation; the resulting decrease in protein expression adversely affects NF-KB signaling and vesicular trafficking, pathways which involved in synaptogenesis and neurogenesis (Hu et al., 2005; Mattson & Camoandola, 2001; Zong et al., 2012). We have recently identified a subset of individuals carrying copy number variations (CNVs), some of which are from unaffected parents while others are de novo in origin. Although the relationship of these unique deletions and duplications in autistic and ID etiology remain unknown, there is evidence indicating the involvement of an imprinting mechanism that may be conserved between mice and humans.

Genomic imprinting at mouse distal chromosome 15 has been examined by our group based on the observation that the syntenic region in mouse contains the known paternally expressed gene, Peg13, located within an intron of Trappc9 (Figure 4.2). Accordingly, allelic expression of Trappc9 was assessed by our collaborator Dr. Kazuhiko Nakabayashi in offspring derived from reciprocally crossed C57BL/6 x JF1 mice (to account for strain differences) in the embryonic period (15.5 dpc; days post coitum), at postnatal day (PD) 0, and at 8 and 24 weeks. Exclusive or preferential expression of Trappc9 from the maternal allele (paternally imprinted) was restricted to a tissue, developmental period and isoform (Appendix 3), as would be expected for an imprinted gene. In the CNS, Trappc9 exhibited maternal-specific expression in fetal brain, in addition to the cerebrum, cerebellum, olfactory bulb, neurons and neural progenitor cells (NPCs); however, Trappc9 was expressed equally from both alleles in glia, fibroblasts and leukocytes. Allelic imbalance of Trappc9 persisted until 8 weeks postnatal in mice and subsequently disappeared, at which point the gene
became biallelically expressed in all brain tissues analyzed. Trappc9 resides between two transcripts with preferential maternal expression: the downstream (and proximal) two-pore potassium channel, Kcnk9, (Ruf et al., 2007) and the upstream catalytic component of the RNA-inducing silencing complex (RISC), Ago2 (also known as Eif2c2) (Gregg et al., 2010; Morita et al., 2007), which binds to Dicer and is involved in miRNA biogenesis (Chendrimada et al., 2005). A paternally-expressed non-coding RNA (ncRNA) Peg13 on the negative strand lies in the 16th intron of Trappc9, and is postulated to regulate imprinting and its associated CGI serves as a DMR (Smith et al., 2003; Davies et al., 2004).

While the imprinting status of TRAPPC9 in humans has not previously been studied, EIF2C2/AGO2 is believed not to be imprinted, however, differential expression (paternal imprinting) of KCNK9 (MIM:605874) remains conserved in humans (Ruf et al., 2007; Luedi et al., 2007), in addition to the hypomethylated profile of its promoter in brain and blood (Ruf et al., 2007). Of clinical relevance, a mutation to the active maternal copy of KCNK9 has been linked to Birk-Barel syndrome (Barel et al., 2008). It is currently not known what mechanism is governing KCNK9 imprinting. Two transcripts – AK307073 and AK748239 – flanking a CGI is presumed to encode one continuous transcript on the negative strand and lies in intron 17 of TRAPPC9. This purported ncRNA shares no sequence conservation with mouse PEG13; however there is evidence suggesting that it, too, may be paternally imprinted and serve as the PEG13 equivalent in humans.

I proposed that the heterozygous deletions in our probands disrupt an imprinted locus by influencing epigenetic mechanisms, specifically DNA methylation and the expression of a regulatory long non-coding RNA (lncRNA), in an allele-specific manner. The primary objective of this study was to identify whether a putative long ncRNA of unknown function, PEG13, located within an intronic region of TRAPPC9 was imprinted in human fetal brain, and in turn, determine its regulatory role on
the expression of proximal imprinted genes, specifically the maternally-expressed gene *KCNK9*. The putative *PEG13* locus was studied by:

(A) Analyzing allelic expression of *PEG13* in humans

(B) Region-specific DNA methylation at CGIs at *PEG13* and *KCNK9*

(C) To determine the relevance of the heterozygous CNVs at *TRAPPC9* in ID and ASD aetiology

**Figure 1.2 Mechanism of deregulated imprinting in the brain.** Balanced contributions from the paternal and maternal genomes is required for normal brain function and development. Disruptions to this balance can lead to neurodevelopmental disorders by: (1) abnormal expression of the active allele caused by mutations and/or (2) loss of imprinting mechanisms, such as aberrant methylation patterns at the DMR.
Chapter 2: Materials, Methods and Techniques

Introduction and study design

This section will outline the shared sample collection protocols and techniques used to generate data, as well as their purpose in the overall study design to avoid repetition. Details, including the methods employed and data analysis, in addition to differences and deviations from the general methodologies, will be addressed and expanded upon in each of the appropriate study chapters.

A relevant institutional ethics board approved all sample collections. Written informed consent was obtained from all participants or their parents. Total DNA and/or RNA were collected from human fetal tissues, fibroblasts, sperm and peripheral blood. Traditional genetic methods were applied for identifying the putative PEG13 locus, specifically polymerase chain reaction (PCR), Sanger sequencing and pyrosequencing technology (Chapter 5). Analysis of methylation patterns was carried out using bisulfite pyrosequencing (Chapter 6). Lastly, real-time quantitative polymerase chain reaction (RT-qPCR) was used to validate CNVs and for gene expression profiling (Chapter 4 and 5).
2.1 Statistical Analysis

This study primarily used quantitative methods. Allelic expression and methylation levels were expressed quantitatively in percentage ranging from 0% to 100%. Pyrosequencing data for the fetal tissues was performed in triplicate and the mean and S.D. of the technical replicates was calculated. For data from bisulfite pyrosequencing, each CpG site was represented as a box and whisker plot to show the distribution of methylation levels across ten samples (i.e. leukocytes, sperm and fibroblasts). Statistical analysis (i.e. S.D., median, mean) for allele frequency and methylation was performed by GraphPad Prism. Similarly, gene expression and DNA copy number were quantifiable variables and normalized for variations among technical replicates using an internal control prior to the statistical analysis.

2.2 Sample Quality and Quantification

Verification of nucleic acid integrity and concentration were assessed using the The NanoDrop® ND-1000 spectrophotometer (Thermo Scientific) and by gel electrophoresis. An optical density (OD) 260/280 ratio of 1.8 to 2.0 and an OD 260/230 ratio greater than 2.0 was considered acceptable nucleic acid quality. RNA was stored at -80°C and DNA at 4°C for long-term storage.

2.3 Sanger Sequencing

Prior to sequencing, PCR products were cleaned using an equivalent volume of microCLEAN reagent (Microzone). Samples were diluted according to sequencing guidelines and submitted to The Center for Applied Genomics (TCAG; Toronto, ON).

2.4 Isolation of Nucleic Acids from Whole Blood

Anticoagulated blood was fractionated by centrifugation at 1000 x g for 15 min and the supernatant was removed. Contaminants and residual erythrocytes were lysed and the remaining leukocyte-rich fraction was pelleted by centrifugation.
DNA was extracted from peripheral blood leukocytes using a high salt method. Briefly, leukocytes were re-suspended in equal volume of 2x lysis buffer (20 mM Tris-Cl pH 7.6, 20 mM KCl, 20 mM MgCl₂ and 2% Triton-X-100; 10⁷ cells/ml), incubated on ice for 5 min, and pelleted by centrifugation at 1350 x g for 15 min at 4°C. The supernatant was discarded and nuclei pellets were resuspended in 5ml of lysis buffer (10 mM Tris-Cl, 10 mM KCl, 10 mM MgCl₂•6H₂O, 2mM EDTA and 0.4 NaCl). Protein was denatured by 10% SDS and the suspension was incubated at 55°C for 20 min and cooled to RT. NaCl was added to the digested product and mixed vigorously for 15 min followed by centrifugation at 16,000 x g for 15 min at 4°C. The DNA was precipitated in pure EtOH and the spooled DNA was soaked in 70% EtOH for 10 min. The DNA was air dried for 10 min and incubated at 55°C in 500 μl TE buffer (10 mM Tris-HCl, 0.5 mM EDTA, pH 7.6) until DNA was dissolved. The resulting DNA was stored at 4°C.

The RNeasy kit (Qiagen) was used for RNA isolation from peripheral blood leukocytes according to manufacturer’s protocol. To ensure the absence of DNA contamination, RNA was treated to DNase I per 10 μg of template (Life Technologies).

Cell Lines

Peripheral leukocytes were transformed to lymphoblasts using the Epstein - Barr virus (EBV) to create immortal cell lines for test subjects. From whole blood, isolated white blood cells were washed with RPMI-1640 growth media, inoculated with EBV and grown at 37°C until confluency was reached. Lymphoblasts were fed 15% RPMI streptomycin-penicillin weekly. The transformed lymphoblasts were transferred to a 2 mL flask, cryopreserved in an isopropanol chamber (Mr. Frosty™) and stored in liquid nitrogen (N₂).
2.5 Human Fibroblasts

Certain classes (5%) of autosomal genes exhibit random monoallelic or preferential allelic expression not associated with genomic imprinting (Gimelbrant et al., 2007); moreover, few imprinted genes in mice have been validated in human cell lines (Morcos et al., 2011). Investigating allelic expression in fibroblasts serves as a cost-effective, yet straight-forward method for validating the imprinting status at a novel locus. Based on an earlier study which performed a genome-wide analysis of imprinted genes in human cells lines (Morcos et al., 2011), fibroblast cell lines (n=10) were obtained from the NIGMS Human Genetic Cell Repository (Coriell Institute, New Jersey) on recommendation by Dr. Tomi Pastinen (McGill University) to validate and quantify allelic expression of three SNPs found around the PEG13 region: rs2270409, rs3802217 and rs4289794 (Table 2.1). These SNPs were biallelic in lymphoblasts and other SNPs around TRAPPC9 were biallelic as well. Fibroblasts were derived from individuals with rare metabolic syndromes but no neurodevelopmental disorders, and were informative (i.e. heterozygous) for SNPs around the purported PEG13 locus. An equal number of males (n=5) and females (n=5) were selected and ranged from 4 months to 54 years in age.

2.5.1 Culturing and Growth Conditions

Individual fibroblasts cell lines were initially sub-cultured in T25 flasks containing 5ml of fibroblastgrowth media (Dulbecco’s Modified Eagle Medium; DMEM Gibco®; 10% fetal bovine serum (FBS) and 5% Streptomycin-Penicillin), and incubated at 37°C, 5% CO₂ overnight. Cells were washed in 5 ml Phosphate Buffer Saline (PBS; Sigma-Aldrich) and detached from the flask surface by Trypsin/EDTA for 2 min (Sigma-Aldrich). Fibroblasts were subsequently passaged to T75 flasks and cultured in 15 ml growth media at at 37°C, 5% CO₂; media wash changed every 4-5 days. When cells reached confluency (seeding density ~ 1.0 - 1.4 x10⁴ cells/cm²), RNA and DNA was extracted according to manufacturer’s protocol (Machery-Nagel, Nucleospin® RNA and Cultured Cells or
Tissue). Cells were harvested and pelleted by centrifugation at 1000 x g for 15 min prior to nucleic acid extraction.

2.5.2 Cryogenic storage

Cells were washed in PBS and passaged as described above (Section 3.5.1). Aliquots (5 ml each with 5.0 x 10^5/ml) were pelleted by centrifugation at 100 x g for 10 min (10°C). The fibroblast pellet was re-suspended in 1 ml freeze media (5% DMSO, 1 ml; DMEM per aliquot), distributed in cryovials and frozen for cryogenic storage in liquid N_2.

2.6 cDNA preparation for reverse transcription

First-strand cDNA was synthesized from 1 μg of RNA using Superscript III™, a modified Moloney Murine Leukemia Virus Reverse Transcriptase (M-MLV RT), and random primers (10mM) in a 20μl reaction volume according to the manufacturer’s guidelines (Invitrogen). Reverse transcriptase and RNA were omitted for two separate aliquots to confirm the absence of DNA contamination. The reaction mixture was incubated at 25°C for 5 min followed by 50°C for 60 min, and then inactivated by heating at 70°C for 25 min. cDNA was stored at -20°C.

2.7 Real-time quantitative PCR

2.7.1 Principle

Real-time quantitative PCR (RT-qPCR) involves the measurement of PCR product accumulation in real-time and represents the standard method for quantifying gene expression and copy number within targeted regions across multiple samples. RT-qPCR serves as an instrumental tool for validating high throughput assays such as genome-wide microarrays, in addition to providing complementary evidence for phenotypic studies because of its reproducibility and high sensitivity.

The molecular basis of RT-qPCR exploits the kinetics of PCR by measuring the fluorescent signal generated by various detection chemistries, including dyes, which may intercalate with double-
stranded (ds) DNA (SYBR® green), hydrolysis probes (TaqMan®) or molecular beacon probes. During the lag or linear phase, fluorescence is at background levels as the initial amplicons are produced after 10-15 cycles (Schefe et al., 2006). Eventually, as amplicon production enters the exponential stage, the fluorescence intensity is significantly higher than baseline and becomes detectable. The cycle at which this occurs, otherwise known as the quantitative endpoint of the amplification target ($\Delta C_T$ value), correlates to the amount of PCR product produced. The quantity of PCR product doubles each cycle in the log-linear phase and eventually plateaus as the amount of reagent and template diminishes, and the fluorescence signal is no longer detectable or quantifiable.

As previously indicated, experimental analysis use the $C_T$ value, defined as the PCR cycle at which the fluorescent signal crosses an arbitrary threshold, which determines the amount of PCR product amplified in each reaction well (Schmittgen & Livak, 2008). Therefore, the $C_T$ value is inversely proportional to the amount of amplified product – a high $\Delta C_T$ represents low expression while highly expressed genes have low $\Delta C_T$ values; the same principle is applied for quantifying genomic copy number variation (CNVs) and for genotyping SNPs. In turn, the $C_T$ value depends on the amount of template present at the beginning of the PCR reaction, as the more starting material, the quicker the fluorescence reaches threshold.
### Table 2.1 List and description of individual fibroblast cell lines

Fibroblasts (n=10) were derived from individuals with rare, metabolic syndromes and no history of neurodevelopmental or neuropsychiatric disorders. Sample selection included an equal proportion of males and females and age ranged from 4 months to 54 years (Coriell Institute). Raw SNP data (Illumina IM Human Array) was provided by Dr. Tomi Pastinen (McGill University) for SNPs within the *PEG13* region for each individual. SNPs demonstrating allelic imbalance (AI) had a ratio > 0.25 or x < - 0.25 (expression ratio of the paternal allele versus the maternal allele) and a value of 0 indicated biallelic expression. Pyrosequencing was used to quantify AI at SNPs rs2270409, rs4289794 and rs3802217 in each individual * = Passage number prior to subculturing; mo = months; NA = Not available; PDL = Population doubling level.

<table>
<thead>
<tr>
<th>Catalog No</th>
<th>Sample description</th>
<th>Age (yr)</th>
<th>Sex</th>
<th>Origin</th>
<th>Passage #*</th>
<th>PDL</th>
<th>rs2270409</th>
<th>rs4289794</th>
<th>rs3802217</th>
</tr>
</thead>
<tbody>
<tr>
<td>GM00042</td>
<td>Hurler Syndrome</td>
<td>25</td>
<td>M</td>
<td>Caucasian</td>
<td>4</td>
<td>0</td>
<td>0.3624</td>
<td>0.3596</td>
<td></td>
</tr>
<tr>
<td>GM00519</td>
<td>Mucopolysaccharidosis Type VI</td>
<td>4</td>
<td>F</td>
<td>Caucasian</td>
<td>3</td>
<td>7.87</td>
<td>-0.4273</td>
<td>-0.3762</td>
<td>-0.3701</td>
</tr>
<tr>
<td>GM01096</td>
<td>Mucopolysaccharidosis Type IIIA</td>
<td>NA</td>
<td>M</td>
<td>Caucasian</td>
<td>6</td>
<td>6</td>
<td>0.3624</td>
<td>0.3596</td>
<td></td>
</tr>
<tr>
<td>GM01376</td>
<td>Homocystinuria</td>
<td>NA</td>
<td>M</td>
<td>Caucasian</td>
<td>6</td>
<td>0</td>
<td>-0.375</td>
<td>-0.3788</td>
<td>-0.3994</td>
</tr>
<tr>
<td>GM01391</td>
<td>Hurler Syndrome</td>
<td>9 mo</td>
<td>F</td>
<td>Caucasian</td>
<td>6</td>
<td>0</td>
<td>0.3336</td>
<td>-0.4402</td>
<td>-0.4202</td>
</tr>
<tr>
<td>GM01653</td>
<td>apparently healthy non=fetal tissue</td>
<td>37</td>
<td>M</td>
<td>Caucasian</td>
<td>12</td>
<td>5.64</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>GM02175</td>
<td>huntington disease</td>
<td>55</td>
<td>M</td>
<td>Caucasian</td>
<td>8</td>
<td>0</td>
<td>-0.3978</td>
<td>-0.3229</td>
<td>-0.3135</td>
</tr>
<tr>
<td>GM02455</td>
<td>Mucopolysaccharidosis Type IVB</td>
<td>6</td>
<td>F</td>
<td>Caucasian</td>
<td>3</td>
<td>0</td>
<td>-0.3434</td>
<td>0</td>
<td>-0.3641</td>
</tr>
<tr>
<td>GM03073</td>
<td>Adenine monophosphate deaminase 1</td>
<td>38</td>
<td>F</td>
<td>Caucasian</td>
<td>5</td>
<td>0</td>
<td>-0.3183</td>
<td>-0.4294</td>
<td>-0.4267</td>
</tr>
<tr>
<td>GM09256</td>
<td>Von Hippel-Lindau Syndrome</td>
<td>54</td>
<td>F</td>
<td>Caucasian</td>
<td>14</td>
<td>0</td>
<td>0.4578</td>
<td>0.4186</td>
<td>0</td>
</tr>
</tbody>
</table>
Data for RT-qPCR can be obtained via two methods. The first, absolute quantification, determines the total number of copies expressed in a system, which calibrates data against a standard curve. Absolute quantification is not always necessary because information regarding the fold-change is often sufficient. Alternatively, relative quantification involves measuring gene expression against a reference gene otherwise known as an internal control. This method utilizes reference genes in order to normalize mRNA levels between different samples, enabling for an accurate comparison of gene expression or gene copy number across samples. The internal control is usually a housekeeping gene expressed in all nucleated cell types and has roles in general cellular processes such as glyceraldehyde-3-phosphate dehydrogenase (GAPDH) in cellular respiration or Beta-actin (β-actin) for cytoskeleton architecture. However, unlike previous assumptions, housekeeping genes differ in their expression levels across various biological samples in both the absence and presence of intrinsic and extrinsic stimuli (Dheda et al., 2004; Schmittgen & Zakrajsek, 2000; Radonic et al., 2003). Therefore, selection of a suitable reference gene(s) requires certain assumptions, mainly that the expression of the reference gene should be uniform across samples, and if applicable, independent of experimental treatment and developmental time points.

A number of mathematical models exist for analyzing data for relative expression across different samples. One such method is the comparative cycle time $\Delta \Delta C_T$ ($2^{\Delta \Delta CT}$) or Livak method (Schmittgen & Livak, 2001), which requires the experimenter to make a few assumptions. Most notably, the PCR efficiency must be close to 1 and uniform across all targeted genes within 5% of each other (including the reference gene). Moreover, this method requires an equivalent amount of starting sample for all reactions. Normalization against known reference gene(s) with stable expression levels, in addition to good RNA quality, can circumvent most of these limitations.
2.7.2 Assay design and validation

Genomic positions were determined by UCSC genome browser (http://genome.ucsc.edu; hg19), from which primer pairs (Integrated DNA Technologies) were designed according to strict parameters using Primer3Plus (http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi):

- Amplicon length: 80-150 bp long
- Primer length: 20/22/30 (minimum/optimum/maximum)
- Primer GC content: 40%/50% (minimum/maximum)
- The GC-content was greater at the 5’ end and less toward the 3’ end (max 3’ stability: 2)
- Each primer had a calculated melting temperature (T\text{m}) of \sim 60°C and maximum T\text{m} difference of 0.1°C between primer sets
- To minimize primer dimer formation, primer pairs had a maximum self-complimentarity of 5 nucleotides and a max 3’ self complimentarity of 2 nucleotides
- Homopolymers of more than 3 nucleotides were avoided

All primer sets were validated in silico for sequence specificity using The BLAST program from the NCBI, as well as empirically by a temperature gradient and separation of PCR products by gel electrophoresis. A no template control (NTC) reaction was included for each primer pair.

2.7.3 Experimental design

The $\Delta\Delta C_T$ method was employed for all assays using SYBR® Green reagents and ROX™ reporter dye (Life Technologies). PCR efficiency (the rate at which polymerase produces amplicons) for each primer pair was determined by the shape of the logarithmic amplification plot and by melt-curve analysis. Assays were designed using a single 384-well optical plate to prevent inter-plate variability and each sample was run in quadruplicate. Internal controls were selected based on the assay type (i.e. gene expression vs. CNV analysis) and control samples were incorporated in all runs. To test for genomic DNA contamination, a NTC control was included for each primer set.
2.7.4 Procedure

For this study, each RT-qPCR assay used singleplex PCR. All wells contained a 16 μl final reaction volume containing 2 μl template (5ng/μl), 8 μl Fast SYBR® Green Master Mix, 0.5 μl each of the forward and reverse primer (10 μM), and 5.0 μl of dH₂O. PCR amplification was achieved using the Applied Biosystems viiA7™ (Life Technologies), with an initial denaturation at 95°C for 2.5 min, followed by 40 cycles at 95°C for 3.0 s and 60°C for 20 s.

2.8 Pyrosequencing

2.8.1 Principle of pyrosequencing

Pyrosequencing is a real-time, sequence-by-synthesis technique based on the detection of light following the release of inorganic pyrophosphate (PPI) to measure allelic frequencies (SNP) and DNA methylation patterns. The theory and methodology behind pyrosequencing is summarized in detail by the following papers summarize the theory and methodology behind pyrosequencing (Dupont et al., 2004; Tost & Gut, 2007) and reviews (Ronaghi, 2001; Fakhrai-Rad et al., 2002); the biochemical basis behind the reaction will be briefly outlined below.

Following the annealing of a sequencing primer, DNA polymerase incorporates a nucleotide complementary to the 3’ end of the immobilized DNA template. Only one nucleotide at any given time is present in the reaction system, which contains a proportional mixture of enzymes, deoxyribonucleotide triphosphates (dNTPs) and substrates dispensed by a cartridge according to a predetermined program. The addition of nucleotides occurs in a stepwise manner as a consequence of four enzymatic reactions. Once DNA polymerase I (*Escherichia coli*) adds a dNTP to single-stranded (ss) DNA, PPI is released and serves as a substrate for ATP sulfurylase (*Saccharomyces cerevisiae*), which converts it to ATP in the presence of adenosine 5’phosphosulfate (Tost & Gut, 2007). Luciferase (*Photinus pyralis*) utilizes energy from ATP to oxidize luciferin to oxyluciferin and
releases light as a by-product. Because dATP is the natural substrate for luciferase, sequencing uses alpha-thio triphosphate (dATPαS) instead, and is recognized and efficiently used by DNA polymerase. A charge-coupled device (CDD) camera simultaneously captures the total amount of light produced in each well by converting photons to a quantifiable charge and are visualized as a series of peaks in a Pyrogram® (Appendix 2). The peak heights represent the quantity of nucleotides added to the 3’ end as the complementary DNA strand is built-up. Apyrase (Solanum tuberosum) degrades all unincorporated dNTPs and extra ATP molecules before the addition of the next nucleotide, after which the cartridge dispenses a new nucleotide according to the dispensation program. Nucleotides are dispensed in alternating wells to ensure that the light measured is actually produced by that specific well (Tost & Gut, 2007). If a dispensed nucleotide is not complementary to the sequence, it will not release PPi and will subsequently be degraded. As such, the bioluminometric signal is directly proportional to the amount of ATP produced, which in turn, is equimolar to the number of PPi released or nucleotides incorporated in the sequence. For instance, if two adjacent nucleotides in a sequence are identical, such as two thymines, the peak height will be twice as high as compared to a peak generated by a single thymine in the sequence (Tost & Gut, 2007). Furthermore, the ratio of nucleotides incorporated at variable positions as seen in SNPs (e.g. A/G) and methylation (i.e. methylated vs. unmethylated) is expressed as the degree of expression or methylation respectively. Consequently, the intrinsic kinetic properties - $K_M$ (μM) and $K_{cat}$ (S⁻¹) – of each enzyme, in addition to the reagent concentrations, dictate the pyrogram profile (Ronaghi, 2001).

2.8.2 Experimental design

Primer and amplicon design

Three primers are required for pyrosequencing: a primer pair for PCR amplification and an internal sequencing primer. All oligonucleotides were purchased from IDT and listed in their respective study chapters. Online softwares OligoCalc (http://www.basic.northwestern.edu/biotools/oligocalc.html)
and Multiple primer analyzer by Thermo Scientific Multiprimer Tool (http://www.thermoscientificbio.com/webtools/multipleprimer/) were used to determine melting temperatures, possible hairpin formation and self- and cross-primer dimers. All primers had a calculated melting temperature ($T_M$) of ~60°C. NCBI BLAST or UCSC BLAT validated that primer pairs were unique to one position within the PCR product.

Amplicons between 100-300 base pairs in length were designed to facilitate amplification and sequencing. PCR primer pairs differed in melting temperature ($T_M$) by less than 1°C and included the 5’ biotinylation of one amplification primer. Usually the reverse primer was biotinylated for a forward sequencing reaction and was opposite to the direction of the sequencing primer. The PCR primers were designed in regions containing no variable positions and were preferably GC-rich near the 5’ end and less toward the 3’ end to improve primer specificity. Similarly, no variable positions were included in the design for sequencing primers, which were 18 to 25 bases (optimally 20bp) in length and had a calculated $T_M$ between 45°C to 50°C.

**PCR Amplification**

PCR amplification of either DNA or cDNA (depending on the application) was used to generate a sufficient amount of template and for sequence specificity. Optimization of PCR conditions (temperature and concentration gradients) was performed using the T100™ Thermal Cycler 186-1096 (BioRad). Each reaction system contained 2.0µl of template (> 10 ng), 1.0µl of each forward and reverse primer (10 uM) and 12.5µl of KAPA2G Fast HotStart ReadyMix in a 25µl reaction volume. The negative control contained no template and a positive control included already optimized primers. PCR conditions followed conventional protocol, with an initial denaturation at 95°C for 1 min, followed by 40 cycles at 95°C for 10s, annealing temperature between 50°C -65°C

---

1Biotin, a B-vitamin, has high affinity for Sepharose. Streptavidin-coated sepharose beads capture biotinylated ssDNA templates for sequencing by in the pyrosequencing assay.
for 10s, and extension at 72°C for 1.0s. The high number of cycles exhausted all primers and minimized sequence interference from potential artefacts. To ensure primers were sequence specific and sufficient template was yielded at the appropriate size, 5.0 μl of PCR product was visualized on a 2.0% agarose gel under trans-UV light.

### 2.8.3 Pyrosequencing procedure

Pyrosequencing was performed using the PyroMark™ Q24 System using Q96 Gold (5x96) reagents (Qiagen) according to manufacturer’s instructions. Each biotinylated PCR product (20 μl) was immobilized to 2.0μl Strepavidin-coated Sepharose (GE Healthcare Life Sciences Division, Piscataway, NJ) in 40μl 2X binding buffer and 18 μl distilled H₂O. The mixture was agitated at 1400 rpm (Thermoscientific) for 10 minutes. The Vacuum Tool (Pyromark™ Q24) captured the beads and rinsed them in 70% ethanol (70 ml), denatured in NaOH (50 ml), and neutralized in wash buffer (70 ml). The beads, now bound to ssDNA template, were released from the vacuum into a 24-well plate and annealed to 0.3 μM of sequencing primer at 80°C for 2.0 min. The plate was cooled to room temperature for 5 minutes and placed in the PyroMark™ Q24 instrument with the appropriate quantity of dNTPs, enzyme and substrate mixtures.

Methylation patterns and SNPs were quantified using the CpG and AQ softwares, respectively (PyroMark Q24 Software). The peak heights were compared to the expected pyrogram heights. Pyrosequencing runs for each primer pair and sample type was optimized and included three negative controls (no template, no sequencing primer, and no template or primer).
Chapter 3: CNV Analysis in ID and Autism Probands

3.1 TRAPPC9

3.1.1 TRAPPC9 function in the CNS

Several lines of evidence demonstrate that TRAPPC9 (trafficking particle protein complex 9; chr8q24.3) is critical for neurodevelopment and brain function. TRAPPC9 has restricted expression in cortical neurons (Mochida et al., 2009) and encodes a multi-functional protein involved in NF-κB signaling (Hu et al., 2005) and vesicular trafficking (Zong et al., 2011).

In mammals, TRAPPC9 positively modulates downstream responses by NF-κB, a ubiquitously expressed transcription factor composed of homo- or heterodimeric complexes of the Rel protein family (p50, p52, p65 (RelA), RelB and c-Rel (Blank et al., 1992; Tak & Firestein 2001). In unstimulated cells, Iκβ proteins bind to NF-κB and prevent its translocation to the nucleus. However following activation by extracellular stimuli (e.g. inflammatory cytokines or brain-specific activators), serine residues of Iκβ are phosphorylated by the signalsome complex (Traenckner et al., 1994; 1995) which consists of protein kinases, including the Iκβ proteins Iκβα and Iκκβ, as well as a non-catalytic regulatory subunit NEMO/IKKγ (Brown et al., 1993; Chen et al., 1995; Koulich et al., 2001). As the main catalytic subunit, phosphorylation by Iκκβ targets Iκβ for ubiquitylation and its subsequent degradation by the 26S proteasome complex in the cytoplasm (Magnani et al., 2000). Degradation of Iκβ frees the NF-κB transcription factor and reveals the nuclear localization signal on RelA (p50/p65), allowing NF-kB dimers to translocate rapidly across the nucleus to regulate expression for genes such as cytokines, adhesion molecules and immunoreceptors in the brain (O’Neill & Kaltschmidt, 1997). NF-κB transcription factors are functionally present in all neuronal and glial cells (O’Neill & Kaltschmidt, 1997), and have critical roles in physiological and pathological responses in the CNS, specifically for neural cell function, synaptic plasticity and cellular immunity (Meffert & Baltimore, 2005; Mattson & Camoandola, 2001). A study by Hu et al.,
revealed that TRAPPC9 enhances NF-κβ activity by increasing the phosphorylation of IKKβ and NIK, and was aptly given the alternative name, NIK (NFκβ-inducing kinase) and IKKβ binding protein (NIBP); conversely, knockdown of TRAPPC9 by RNA interference reduced activation of the NF-κβ pathway. In vitro, TRAPPC9 also increased neurite outgrowth following nerve growth factor-induced neuronal differentiation, thereby suggesting a role for TRAPPC9 in neurogenesis (Hu et al., 2005). Corroborating this finding, increased expression of the anti-apoptotic gene, Bcl-Lx, following TRAPPC9 over expression provides evidence for TRAPPC9 as an essential protein in neuronal survival (Hu et al., 2005). This is particularly relevant in early postnatal life when the majority of synapses are trimmed.

TRAPPC9 also mediates vesicular trafficking of COPII vesicles in early secretory and endocytic pathways as a component of the multi-subunit transport protein particle (TRAPP) (Zong et al., 2011). The TRAPP complex, with the exception of one subunit, is a highly conserved guanine nucleotide-exchange factor between yeast and mammals, serving as a tether for vesicle docking at the target membrane for both endoplasmic reticulum to Golgi and intra-Golgi trafficking (Zong et al., 2012). Yeast S. cerevisiae possess three different TRAPP isoforms sharing a common six-subunit core: TRAPPI, TRAPPII and TRAPPIII (Sacher et al., 2001). For mammals, the TRAPPII complex is the predominant isoform and primarily mediates intra-Golgi trafficking via two identified mechanisms (Yip et al., 2010; Yamasaki et al., 2009). Within the TRAPPII complex, TRAPPC9 is a homolog to an essential gene in yeast, Trs120 (Sacher et al., 2001). Co-immunoprecipitation experiments reveal that TRAPPC9 interacts with TRAPPC2, a key adaptor subunit within the mammalian TRAPPII complex, and TRAPPC10 (Zong et al., 2012). Moreover, TRAPPC9 interacts with p150Glued, a subunit of dynactin that is required for the movement of COPII vesicles toward the ER-Golgi intermediate compartment along microtubules (Zong et al., 2012). Individuals harbouring mutations at TRAPPC9 illustrate the functional implications of TRAPPC9 in neurodevelopment.
3.1.2 TRAPPC9 and ID

To date, five homozygous frameshift or nonsense mutations in TRAPPC9 (MIM 611966) were identified as a cause of NS-ARID in six apparently unrelated consanguineous families segregating ID from South Asia, the Middle East and North Africa (Appendix 2). In these families, parents of affected individuals were carriers for mutations in TRAPPC9 with an exception to a Syrian family described by Jamra and colleagues (2011) which identified a de novo mutation at this gene locus. Similarly, individual cases have emerged in a consanguineous family of East Asian descent and in a non-consanguineous European family, both with no apparent history of ID (Marangi et al., 2012; Koifman et al., 2010). The homozygous mutations comprise two submicroscopic deletions (Koifman et al., 2010; Najmabad et al., 2009), a splice site mutation (Kakar et al., 2012), and four nonsense mutations, three of which are located at the same genomic region (Marangi et al., 2012; Jamra et al., 2011; Mochida et al., 2009; Mir et al., 2009). Nonsense-mediated mRNA decay and decreased protein levels were detected in these patients via RT-qPCR and western blotting (Mir et al., 2009; Mochida et al., 2009; Phillipe et al., 2009), and indicated loss-of-function mutations. Subsequent functional studies by Zong and colleagues (2012) determined that truncating mutations at TRAPPC9 reduce the binding affinity of TRAPPC9 to other proteins in the TRAPPII complex.

Affected individuals carrying homozygous mutations at TRAPPC9 display a range of clinical symptoms synonymous with an idiopathic disorder, including speech impairment, developmental delays of varying severity (including moderate to severe ID), stereotypic behaviours and hand flapping (Appendix 1). Other non-specific features include truncular obesity, growth retardation, hypotonia and dysmorphic facial features. Interestingly, most individuals exhibit postnatal microcephaly or reduced head circumference not concomitant to seizures or epilepsy. Additional shared features include diminished cerebral white matter volume, hypoplasia of the corpus callosum and reduced cerebellar volume.
3.1.3 Heterozygous CNVs at the *TRAPPC9* locus

In addition to the homozygous recessive mutations, we have recently identified a subset (n=20) of ID and ASD individuals carrying rare CNVs that are heterozygous, some of which are from unaffected parents while others are *de novo* in origin (Figure 3.1). Probands with CNVs overlapping the *TRAPPC9* locus were identified and obtained from several groups and databases, including the Autism Genome Project (AGP) database (http://davinci.tcag.ca/agp/), a private database which we have access to via our collaborations with the AGP, the DECIPHER database (GRCh37) (http://decipher.sanger.ac.uk/), and the International Standards for Cytogenomic Arrays (ISCA) consortium (https://www.iscaconsortium.org/). In addition, six independent groups contacted us with reports of probands with *TRAPPC9* CNVs. Different microarray platforms, including Roche NimbleGen 135K array, Illumina Human 1 M array and Agilent 105 K CGH array, screened for structural variations and estimated the breakpoints for CNVs. The relationship of these unique deletions and duplications in autistic and ID etiology require investigation, however there is evidence suggesting the involvement of a conserved imprinting mechanism between mice and humans (genes summarized in Table 3.1).

3.1.4 Imprinting at the Chr15qD3 domain

Mouse distal chromosome 15 (chr15qD3), the syntenic region to chr8q24.3, is a well-characterized imprinting cluster, comprised of three paternally imprinted protein-coding genes including *Trappc9*, and a maternally imprinted intronic long ncRNA (Figure 3.2).

Our research group examined genomic imprinting at chr15qD3 based on the observation that this region contains the known paternally expressed gene, *Peg13*, located within intron 16 of *Trappc9* (Smith et al., 2003). Allelic expression of *Trappc9* was assessed by our collaborator Dr. Kazuhiko Nakabayashi in offspring derived from reciprocally crossed C57BL/6 x JF1 mice in the embryonic
period (15.5 dpc; days post coitum), at postnatal day (PD) 0, and at 8 and 24 weeks postnatal (Appendix 3). Exclusive or preferential expression of \textit{Trappc9} from the maternal allele (paternally imprinted) was tissue specific and restricted to development. In the CNS, \textit{Trappc9} exhibited maternal-specific expression in fetal whole brain in addition to the cerebrum, cerebellum, olfactory bulb, neurons and neural progenitor cells (NPCs); however, \textit{Trappc9} showed equal expression from both alleles in glia, fibroblasts and leukocytes. Allelic imbalance of \textit{Trappc9} persisted until 8 weeks postnatal in mice and subsequently disappeared, at which point the gene became biallelically expressed in all brain tissues analyzed (Appendix 3). Imprinting of \textit{Trappc9} appeared to be restricted to the CNS, as other somatic tissues such as limb, skin, placenta, spinal cord, muscle and heart displayed biallelic expression (Appendix 3). Furthermore, RNA-seq in mouse brain found that paternal imprinting of \textit{Trappc9} in the brain depends on the gene isoform (Gregg et al., 2010). Loss of \textit{Trappc9} imprinting in adult life highlights its evolutionary role in mammalian neurodevelopment.

\textit{Trappc9} resides between two paternally imprinted transcripts. Lying downstream and proximal to \textit{Trappc9} is a \textit{K_{2p}} potassium channel, \textit{Kcnk9}, highly expressed in the brain (Ruf et al., 2007). \textit{Eif2c2} (\textit{argonaute 2}), which lies upstream to \textit{Trappc9} and is also expressed from the maternal allele (Gregg et al., 2010), encodes a protein involved in mRNA processing as part of the RNA-induced silencing complex (RISC) (O’Carroll et al., 2007; Rand et al., 2004). Moreover, the chromatin-remodeling factor 1, \textit{Chrac1}, which is biallelically expressed, is located in-between \textit{Trappc9} and \textit{Eif2c2} (Kelsey, 2011). A paternally expressed long ncRNA, \textit{Peg13}, on the negative strand lies in the 16\textsuperscript{th} intron of \textit{Trappc9}, is presumed to regulate imprinting of these transcripts (Smith et al., 2003; Davies et al., 2004) as the CGI associated with \textit{Peg13} was identified as a maternal DMR in the germ line and in the brain. The promoters of neither \textit{Kcnk9} nor \textit{Eif2c2} appear to regulate their imprinting status (Ruf et al., 2007). The CGI-promoter region of \textit{Kcnk9} is in fact hypomethylated in the germ line for both
sperm and oocytes. It is also worth mentioning that the transcripts exhibit restricted expression to both neonatal and adult mouse brain (Wang et al., 2008; Davies et al., 2004; Smith et al., 2003). \textit{Kcnk9} is highly expressed in neurons and the cerebellum (Smith et al., 2003). Similarly, expression patterns for \textit{Peg13} in mouse brain tissues is primarily neuronal and is highly expressed in hypothalamic regions, the hippocampus and the cerebral cortex (Davies et al., 2004), which parallels to the expression patterns of \textit{Trappc9} in mouse brain (Mochida et al., 2009).

3.1.5 Imprinting at the Chr8q24.3 domain

The extent to which imprinting is conserved at the human syntenic region (chr8q24.3) requires further investigation. It is currently known that \textit{EIF2C2} is not imprinted in humans, while the imprinted status of \textit{TRAPPC9} needs to be determined. \textit{KCNK9} (MIM605874), however, remains paternally imprinted and a mutation to the active copy (maternal allele) has been associated with Birk-Barel syndrome (MIM612292), characterized by ID, hyperactivity, hypotonia and facial dysmorphisms (Barel et al., 2007). Imprinting of \textit{KCNK9}, in parallel to mice, is not regulated by its promoter, which is also hypomethylated in the brain as well as in blood (Ruf et al., 2007). The mechanism governing \textit{KCNK9} imprinting is currently unknown.

Intriguingly, intron 17 of \textit{TRAPPC9} contains numerous overlapping ESTs on the negative strand, specifically AX307073 and AK748923, and a large, proximal CGI (Figure 3.1). ChIP-seq analysis reveals that the CGI contains chromatin signatures indicative of an insulator element, namely enrichment for trimethylation of H3K4 and CTCF binding. The significance of these regulatory marks have yet to be determined in the \textit{PEG13} context. Despite the lack of sequence homology with mouse \textit{Peg13}, these ESTs appear to represent a continuous transcript, and SNPs assessed within this region demonstrate biased expression from the paternal allele in human fibroblasts (Morcos et al., 2011), thereby suggesting that this long ncRNA, aptly named \textit{PEG13}, is the \textit{Peg13} equivalent in
humans. Taken together, the human PEG13 equivalent requires further investigation, along with its relevance to the heterozygous CNVs in our ID and ASD (Figure 3.2).

<table>
<thead>
<tr>
<th>Mouse gene/Human gene</th>
<th>Gene type</th>
<th>Classification</th>
<th>Function</th>
<th>Allelic expression in mice</th>
<th>Allelic expression in humans</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eif2c2/AGO</td>
<td>P</td>
<td>Argonaute RISC catalytic component 2</td>
<td>RNA interference: may interact with DICER and play a role in short-interfering-RNA-mediated gene silencing</td>
<td>Maternal</td>
<td>Biallelic</td>
</tr>
<tr>
<td>Chrac1/CHRAC1</td>
<td>P</td>
<td>Chromatin accessibility complex protein 1</td>
<td>Histone fold protein; interacts with other histone fold proteins to bind DNA in a sequence-independent manner; these dimmers combine with larger enzymatic complexes for DNA transcription, replication and packaging</td>
<td>Biallelic</td>
<td>Biallelic</td>
</tr>
<tr>
<td>Kcnk9/KCNK9</td>
<td>P*</td>
<td>Potassium channel, subfamily K+, member 9</td>
<td>Two-pore potassium channel</td>
<td>Maternal</td>
<td>Maternal</td>
</tr>
<tr>
<td>Peg13/PEG13</td>
<td>RNA</td>
<td>Long ncRNA</td>
<td>Might mediate imprinting of neighbouring genes</td>
<td>Paternal</td>
<td>Unknown</td>
</tr>
<tr>
<td>Trappc9/TRAPPC9</td>
<td>P</td>
<td>Trafficking particle protein complex 9</td>
<td>Protein involved in NF-κB signaling and vesicular trafficking</td>
<td>Maternal</td>
<td>Biallelic</td>
</tr>
</tbody>
</table>

Table 3.1 Summary of genes imprinted at human chr8q24.3 and the syntenic mouse region, chr15qD3. In mice, Trappc9, Kcnk9 and Eif2c2 are paternally-imprinted genes. Conversely in humans, the imprinting status of KCNK9 remains conserved. It is hypothesized in mice and humans that the long ncRNA, Peg13/PEG13, regulates imprinting. P = protein
Figure 3.1 Schematic illustrating the CNV breakpoints spanning the TRAPPC9 locus and relative to neighbouring genes (UCSC/hg19). Blue = duplications and Red = deletions. The CNV breakpoints in ID and autism probands converge near the putative PEG13 locus, which lies on the negative strand of the 17th intron of TRAPPC9. The CGI associated with PEG13 shows enrichment for CTCF and trimethylation and monomethylation of H3K4, which is indicative of an insulator element and active expression respectively. As such, we hypothesize that PEG13 (AK307073 and AK748239) is a novel imprinted gene whose expression may be disrupted by the heterozygous CNVs depending on the parent of origin in which the CNV resides. The maternally-expressed gene, KCNK9, lies downstream to TRAPPC9. CHRAC1, AGO2 and TRAPPC9 show biallelic expression in humans.
Figure 3.2. Schematic representation comparing the mouse Trappc9/Kcnk9/Ago2/Peg13 imprinting cluster with the human syntenic region (Figure not drawn to scale). Allelic expression is indicated by the transcript colours: Blue = paternally-expressed gene; Red = maternally expressed gene; Black = biallelically-expressed gene; and green rectangle = CpG island. (A) Mouse imprinting cluster at chr15D. The ncRNA PEG13 is reciprocally imprinted to the protein-coding genes Kcnk9, Trappc9 and Eif2c2, and lies within intron 16 of Trappc9. Peg13 is presumed to mediate imprinting. (B) Human syntenic region at chr8q24.3. According to current findings, imprinting at chr8q24.3 only remains conserved for KCNK9, a two-pore pore potassium channel highly expressed in the CNS. A maternally-inherited missense mutation in KCNK9 has been associated with Birk-Barel Syndrome. The CpG-promoter of KCNK9 is not differentially methylated in germ cells and brain, suggesting that imprinting of KCNK9 may occur by an alternative mechanism. While surrounding genes are biallelically-expressed and lack DMRs, a non-coding RNA from the minus strand lies in the 17th intron of TRAPPC9, and may be the Peg13 equivalent in humans. Further investigation is required to determine if this ncRNA is imprinted and has a regulatory role in KCNK9 imprinting.
3.2 Materials and Methods

I completed all work in this chapter except the following: Validation of the CNV and sequencing of TRAPPC9 for Proband 1 was performed previously by Kaufman et al., (2011).

3.2.1 Sample Collection

Blood samples from four affected individuals were provided by Suzanne Lewis, Daniel Doherty, Edwin Kirk, and Jennifer Roggenbuck. These samples were collected in Vacutainer ACD Solution B tubes (BD, New Jersey) and Tempus™ Blood RNA tubes (Life Technologies), and shipped to Toronto. All genetic tests, including microarray analysis, karyotyping, Fragile X testing; medical/blood tests (i.e. esophagram, IGF-1 and IGF-BP3; sweat chloride, thyroid levels, and Ca\(^{2+}\) and P levels); and behavioural assessments were performed by a genetic counselor and/or an acting physician. Extraction of nuclei acids was performed as described in Chapter 3, using the high salt method for DNA isolation and a commercial kit for RNA isolation (RNAeasy; Qiagen).

3.2.2 Whole gene sequencing

To confirm that no secondary mutations (i.e. frameshift, missense or nonsense mutations) were present in our probands, all exons of the TRAPPC9 gene were sequenced. PCR primers (IDT) were designed by Liana Kaufman (2011) using UCSC genome browser (build GRCh37/hg19) and Primer3 software. Primers flanked a minimum of 70 bp from each exon boundary to ensure complete amplification of each exon with the exception of exon 2 where two overlapping primer sets were designed to accommodate the ~1 kb size. PCR reactions comprised 5.0 μl 2G Kapa HotStart (2X); 2 μl H₂O; 1.0 μl DNA template (25 ng/μl); and 1.0 μl (10 μM) for each forward and reverse primer. The PCR followed standard conditions with initiation at 95°C for 1 min; and 30 cycles of denaturation at 95°C for 10 s; annealing at 59°C for 1.0 s; and elongation at 72°C for 1.0 s. Amplification of exon 1 used hot star taq polymerase and performed according to the manufacturer’s
protocol (Qiagen). PCR product size was confirmed by gel electrophoresis and visualized using ethidium bromide or Redsafe™ Nucleic acid staining solution (20, 000X) (Frogga Bio) and the Bio Rad Gel Doc™ XR System. PCR products were Sanger sequenced (TCAG, Toronto) and sequences were analyzed by BLAST and FinchTV.

3.2.3 CNV analysis
Validation of the CNVs was performed by RT-qPCR in test subjects using the ∆∆C_{T} method as previously described in Chapter 3. An internal primer pair was used as an endogenous control for all runs and leukocyte DNA from healthy individuals was used as positive controls. Each sample was run in quadruplicate and data was normalized against the internal control. The experimental data was derived from the ViiA™ Software v1.1. The ∆∆C_{T} values and statistical analysis was calculated using the GraphPad Prism Software, with the exception of proband 1 which was derived from the ViiA™ 7 Software.
Table 3.2. List of primers pairs used to sequence TRAPPC9 exons. All primer pairs had a calculated melting temperature of ~60°C and designed to amplify the entire exon with the exception of exon 2, which was > 1kb in size.

<table>
<thead>
<tr>
<th>EXON</th>
<th>Primer Sequence (F &amp; R)</th>
<th>Size (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>CGCCTTGGCAGCTGAACATAC</td>
<td>450</td>
</tr>
<tr>
<td></td>
<td>GAGGAAAGCTTTCTGCCCT</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>GGGTACGCTCTACGCACT</td>
<td>421</td>
</tr>
<tr>
<td></td>
<td>CGGGAGTCACAGATGGTGA</td>
<td></td>
</tr>
<tr>
<td>2-2</td>
<td>CATCACAGACTGCTCTCG</td>
<td>421</td>
</tr>
<tr>
<td></td>
<td>TGCTGCAATAAGAAATCTCC</td>
<td></td>
</tr>
<tr>
<td>2-3</td>
<td>CAAAATGAGGCTCCCTGACT</td>
<td>515</td>
</tr>
<tr>
<td></td>
<td>TCTCTGCTGCTGCTGCTCA</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>GAGAATTGAATATGCCTAGCAG</td>
<td>300</td>
</tr>
<tr>
<td></td>
<td>GCACCCCCAGGCTCTCAAATTTTA</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>GCTAAATGAAGGTCAGCTGACTG</td>
<td>303</td>
</tr>
<tr>
<td></td>
<td>TCCTAAGTTAAGGAAGGAATAACAAAAA</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>CAAGTCATTTTTCTCTAGGCTATG</td>
<td>304</td>
</tr>
<tr>
<td></td>
<td>GGACCACCAAATCATACTC</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>CATGCTCTCTTTAAACAAAGAAA</td>
<td>330</td>
</tr>
<tr>
<td></td>
<td>TGATTTGAGACAGCAACAGCATGAA</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>ATGGGGGAAATGCTGTTGTA</td>
<td>344</td>
</tr>
<tr>
<td></td>
<td>TTATCTGAGCTGCTGAAACC</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>CAGGTCTTCTTTTTAGAAGAGACAG</td>
<td>447</td>
</tr>
<tr>
<td></td>
<td>CAGGGCAAGGCTCTAGTCTCAT</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>GTTTGGTGTTTGGGCATCT</td>
<td>344</td>
</tr>
<tr>
<td></td>
<td>CAAGCCCAAGGTTCATAGTA</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>AATGAAAGCCAGCACCCTTGG</td>
<td>307</td>
</tr>
<tr>
<td></td>
<td>CCGATAACAGGCAACAAAAAA</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>ATCTGAAAACCCAGCTAACC</td>
<td>304</td>
</tr>
<tr>
<td></td>
<td>ACATGCATGAACAGCTGCTCAA</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>TGCCAGAGGATAATTTGGA</td>
<td>265</td>
</tr>
<tr>
<td></td>
<td>TGCTAAGATGCTGCTGCCTCA</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>TGTCGCAAGGCAATgtCTCT</td>
<td>325</td>
</tr>
<tr>
<td></td>
<td>AACCTGTTAGGACTGGCAGATGA</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>GAAGAGGAGCCTGTACTCT</td>
<td>321</td>
</tr>
<tr>
<td></td>
<td>TTACAGCTCTTTTGGCCATT</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>TGCTGATCTTCTTTTGGGGAAG</td>
<td>374</td>
</tr>
<tr>
<td></td>
<td>CTGACTTCAACTGGAATCCCAAAA</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>GGGGATTTGGTAAGTCCCATC</td>
<td>367</td>
</tr>
<tr>
<td></td>
<td>ACAAAAGTTGCCCCAAGAATGTT</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>CAAAGGTGCGAGCAGCAAGA</td>
<td>458</td>
</tr>
<tr>
<td></td>
<td>ACCAAAACAATGCACCTCAACGC</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>GCCTCCTGGTTTGATCAAACT</td>
<td>310</td>
</tr>
<tr>
<td></td>
<td>ACGGGATGCATGACAAAACAA</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>AAGTGGTGTTGGGATTTTGG</td>
<td>300</td>
</tr>
<tr>
<td></td>
<td>TGAGGACTTGGGGTGAGATTAA</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>TGAAGTGCGAGGCACTTTTAG</td>
<td>517</td>
</tr>
<tr>
<td></td>
<td>GTGGGGCCTTTTGAAGTGTAA</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>CCCATCGAGGGTCTGTCCTGTC</td>
<td>309</td>
</tr>
<tr>
<td></td>
<td>TCCCGTGTAGACTCTTCAGT</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>CACACTCTGTGTCCCGAGTG</td>
<td>623</td>
</tr>
<tr>
<td></td>
<td>CACATTTCTCTGCTCCACCA</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>CCTATGGTCCAGGAGGATTCTC</td>
<td>348</td>
</tr>
<tr>
<td></td>
<td>CAGTGAGGCTTTGCTCATT</td>
<td></td>
</tr>
</tbody>
</table>
Table 3.3 qPCR assays for CNV validations. Primer pairs were designed to confirm the direction of the CNV (i.e. deletion or duplication) in the autism and ID probands according to UCSC genome build hg19.

<table>
<thead>
<tr>
<th>Source (Proband ID)</th>
<th>Coordinates/CNV type</th>
<th>Primer sequence (F/R)</th>
<th>Size of amplicon (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proband 1</td>
<td>chr8:141211615-141387894 Deletion</td>
<td>AGGCCTTAATACCTGGGTCACAA AGTTCAGGGCTACAAGTGAAGG</td>
<td>125</td>
</tr>
<tr>
<td>Proband 3</td>
<td>chr8:141210317-141298087 Deletion</td>
<td>AGAGCACACACGATAAATGG TGGAGGATAAAGCATCTCAAGT</td>
<td>97</td>
</tr>
<tr>
<td>Proband 8</td>
<td>chr8:141211815-141268818 Deletion</td>
<td>CACTAGCAGAAGGAGCATGAAA GGCTGAGTATGAGGAAACCCTA</td>
<td>118</td>
</tr>
<tr>
<td>Proband 20</td>
<td>chr8:140821366-141268818 Duplication</td>
<td>AGAGCACACACGATAAATGG TGGAGGATAAAGCATCTCAAGT</td>
<td>97</td>
</tr>
<tr>
<td>Internal control for CNVs (Region on chromosome 7 not known to contain CNVs in autism populations)</td>
<td></td>
<td>GAAGCAGGACTCTAAGTCCAGA TGCTAGAGGAGTGGGACAAGTA</td>
<td>140</td>
</tr>
</tbody>
</table>
3.3 Results

3.3.1 TRAPPC9 CNVs confer phenotypic variation

Twenty probands carrying rare heterozygous CNVs overlapping TRAPPC9 exons show varying severities of developmental delay, ID or ASD, accompanied by a range of other clinical phenotypes (Table 3.3). All probands were from unrelated families from various geographical regions with no known history of ID or ASD.

An autistic individual (Proband 1) carrying a 176 kb deletion was identified by a study conducted in Vancouver (Riendeau, 2009). The CNV was confirmed by fluorescence in situ hybridization (FISH) (Riendeau, 2009) and by RT-PCR (Kaufman et al., 2011) in both the mother and proband (Figure 3.3a). Whole gene sequencing of TRAPPC9 in the mother and proband revealed no coding mutations (Kaufman et al., 2011), indicating that the deletion is the only genetic variant at this gene locus.

Other autistic cases with developmental delay were found in the ISCA Database, as well as by Sanders et al., (2011) and individual collaborators (Table 3.3). Some deletions/duplications overlap with TRAPPC9 in addition to genes distal or proximal to TRAPPC9, including EIF2C2, KCNK9 and CHRAC1 (Probands 4, 6 9 and 10). Mutation testing could not be performed in these probands because DNA and RNA have not yet been obtained.

We also identified individuals harbouring deletions or duplications overlapping TRAPPC9 exons without autism, but displaying abnormal behavioural patterns and developmental delay. An individual (Proband 3) inherited a 129 kb deletion overlapping TRAPPC9 at exons 13-17 from a healthy, normal father. The individual has normal intelligence and exhibits mild behavioral problems and stereotypical behaviors. RT-qPCR confirmed the deletion in both the father and the proband and coding sequences of TRAPPC9 did not contain any additional genetic aberrations (Figure 4.3b). A girl 9 years at assessment (Proband 20) with ID has no history of epileptic seizures, as identified and
evaluated by Dr. Edwin Kirk. She has a paternally inherited duplication overlapping exons 17 to 21 in TRAPPC9; however, the father is unaffected and healthy (Figure 3.3d). The duplication overlaps with AX748239 and it may affect TRAPPC9 function. Subsequent sequencing of TRAPPC9 in the proband and both parents also did not reveal any additional mutations. MRI scans of the brain revealed spaces in the cerebral spinal fluid (CSF) within the posterior cranial fossa (which contains the brain and cerebellum), in addition to the cisterna magna. She is also microcephalic (head circumference < 2nd percentile). There were no abnormalities in the white matter or the corpus callosum. She has a vocabulary of 20-30 words, and although she can walk unassisted, she is ataxic. The subject has occasional hand-flapping and limited fine-motor skills – she cannot hold a pen and nor is she toilet trained. She also has dysmorphic facial features. Methylation testing for Angelman’s syndrome was negative. We also confirmed the patient to be negative for MECP2 mutation.

A de novo 57 kb deletion was identified in an individual with ID (Proband 8); however, the allele on which the deletion is located is currently unknown (Figure 3.3c). This individual is short in stature and has a distinctive facial appearance. The individual is also noted to have borderline hypothyroidism. Her hair is described as sparse and brittle. Fragile X testing and screening for Noonan syndrome were both negative. We also confirmed the patient to be negative for MECP2 mutation. Furthermore, TRAPPC9 sequencing in this proband did not identify other coding mutations.

3.3.2 Validation of CNVs in human leukocytes

As described above, whole genome sequencing of TRAPPC9 did not reveal any coding mutations in our probands and RT-qPCR confirmed the heterozygous nature of the CNVs. The fact that the CNVs are heterozygous and no other coding mutation were found indicates that these individuals are not compound heterozygotes for TRAPPC9 mutations, which is clearly a recessive disease gene.
Therefore, the disease mechanism, if related to the CNVs, must be independent of TRAPPC9 if imprinting of TRAPPC9 is ruled out.

3.3.3 Relevance for characterizing PEG13 in humans

The CNV deletion in Proband 1 was previously shown to have no effect on TRAPPC9 expression in blood as compared to the healthy parents, including the mother who carries the deletion (Kaufman et al., 2011). Moreover, our collaborator (Dr. Miho Ishida, UCL, UK) recently showed that TRAPPC9 is not imprinted in humans using the SNPs rs3735801 and rs3735802 within coding regions of TRAPPC9. Of the 58 fetal DNA samples that were sequenced, 22 cases were heterozygous for both SNPs, 17 of which were also heterozygous for rs3735803 and 8 additional cases were heterozygous for SNP rs3735803. As such, five heterozygous cases with brain and other tissues were available for assessment of TRAPPC9 allelic expression. In all tissues tested, including brain, TRAPPC9 exhibited biallelic expression, indicating that TRAPPC9 is not imprinted in human foetus.

Further evidence for the involvement of disrupted imprinting in disorder etiology is the analysis of a maternally inherited truncating mutation at TRAPPC9 in unaffected carriers from a consanguineous family segregating ID (Mir et al., 2009). Kaufman et al. (2011) found that unaffected carriers with the maternally inherited mutations were phenotypically normal, while paternally inherited mutations displayed delayed cognitive development and poor performance at school. All family members who were heterozygous for the mutations coming from the maternal allele also showed very high performance at school. The cognitive performance of individuals carrying the mutation on the paternal allele indicates parent-of-origin effects (POE) at chr8q4.3. However, it should be noted that evidence for POE is weak because of the small sample size. Also, there may be a culturally related gender bias, whereby greater efforts would be made to educate males in this family. Taken together, the absence of additional coding mutations at TRAPPC9 supports the hypothesis that the proband
CNVs may exert their effects via an alternative mechanism, such as imprinting of another proximal gene.
### Table 3.3 CNVs identified in TRAPPC9 and summary of clinical phenotypes.

Different CHIP platforms identified the estimated breakpoints of CNVs overlapping TRAPPC9 and, in some cases, neighboring genes including KCNK9 and EIF2C2. N/O = no overlap

<table>
<thead>
<tr>
<th>PROBAND ID</th>
<th>DATA SOURCE</th>
<th>MICROARRAY USED FOR ANALYSIS</th>
<th>CNV SIZE AND TYPE</th>
<th>CNV COORDINATES (HG19)</th>
<th>AFFECTED EXONS</th>
<th>OTHER GENES OVERLAPPED</th>
<th>INHERITANCE</th>
<th>PHENOTYPE SEQUENCING OF TRAPPC9</th>
<th>CNV VALIDATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Suzanne Lewis Lab, UBC</td>
<td>Agilent 105K CGH array</td>
<td>176 kb deletion</td>
<td>chr8:141211615-141387894</td>
<td>8-17</td>
<td>None</td>
<td>Maternal</td>
<td>Autism</td>
<td>Yes</td>
</tr>
<tr>
<td>2</td>
<td>Marvin Natowicz, Cleveland Clinic</td>
<td>Unknown</td>
<td>129 kb deletion</td>
<td>chr8:141142433-141318712</td>
<td>18-19</td>
<td>None</td>
<td>Unknown/De Novo</td>
<td>Profound global developmental delays and epilepsy</td>
<td>No</td>
</tr>
<tr>
<td>3</td>
<td>Dan Doherty, University of Washington</td>
<td>Roche NimbleGen 135K array (SignatureChipOS version 2, hg18 assembly)</td>
<td>88 kb deletion</td>
<td>chr8:141210317-141298087</td>
<td>13-17</td>
<td>None</td>
<td>Paternal</td>
<td>Behavioral problems and stereotypical behaviors</td>
<td>Yes</td>
</tr>
<tr>
<td>4</td>
<td>ISCA Database</td>
<td>Unknown</td>
<td>660.5 kb deletion</td>
<td>chr8:141070844-141731423</td>
<td>1-17</td>
<td>CHRAC1, EIF2C2, PTK2, AK130220, AX748239</td>
<td>Maternal</td>
<td>ASD/DD</td>
<td>No</td>
</tr>
<tr>
<td>5</td>
<td>ISCA Database</td>
<td>Unknown</td>
<td>106 kb deletion</td>
<td>chr8:141297551-141403714</td>
<td>8-13</td>
<td>None</td>
<td>Unknown</td>
<td>ASD/DD</td>
<td>No</td>
</tr>
<tr>
<td>6</td>
<td>Bridget Fernandez, Memorial</td>
<td>Illumina Human 1M array</td>
<td>1 Mb duplication</td>
<td>chr8:140016461-141168970</td>
<td>18-23</td>
<td>KCNK9, C8orf17, AX748239</td>
<td>Maternal</td>
<td>Autism and speech delays</td>
<td>Yes</td>
</tr>
<tr>
<td>7</td>
<td>TCAG Autism Chromosome Rearrangement Database</td>
<td>Unknown</td>
<td>281 kb duplication</td>
<td>chr8:140789805-141070993</td>
<td>18-21</td>
<td>None</td>
<td>Maternal</td>
<td>Autism/PD</td>
<td>No</td>
</tr>
<tr>
<td>8</td>
<td>Jennifer Roggenbuck</td>
<td>Unknown</td>
<td>57 kb deletion</td>
<td>chr8:141211815-141268818</td>
<td>16-17</td>
<td>None</td>
<td>De Novo</td>
<td>ID/short stature, facies</td>
<td>Yes</td>
</tr>
<tr>
<td>Proband 9=Proband 6?</td>
<td>DECIPHER #250624</td>
<td>Unknown</td>
<td>1.15Mb</td>
<td>chr8:140016461-141168970</td>
<td>18-23</td>
<td>KCNK9, C8orf17, AX748239</td>
<td>Unknown</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>Database</td>
<td>Sample Type</td>
<td>Description</td>
<td>Chromosome</td>
<td>Genes</td>
<td>Cytoband</td>
<td>Disease</td>
<td>Source</td>
<td>Sex</td>
</tr>
<tr>
<td>---</td>
<td>-------------------------</td>
<td>-------------</td>
<td>-------------</td>
<td>------------</td>
<td>-------</td>
<td>----------</td>
<td>---------</td>
<td>--------</td>
<td>-----</td>
</tr>
<tr>
<td>10</td>
<td>ISCA Database nssv576297</td>
<td>Unknown</td>
<td>1.165 Dup</td>
<td>chr8:140876636-142041711</td>
<td>1-19</td>
<td>CHRAC1, EIF2C2, PTK2, AX748239</td>
<td>Unknown</td>
<td>Autism; ID</td>
<td>No</td>
</tr>
<tr>
<td>11</td>
<td>ISCA Database nssv582488</td>
<td>Unknown</td>
<td>246Kb del</td>
<td>chr8:140980052-141226564</td>
<td>15-17</td>
<td>AX748239</td>
<td>Unknown</td>
<td>Developmental delay</td>
<td>No</td>
</tr>
<tr>
<td>12</td>
<td>ISCA Database nssv579107</td>
<td>Unknown</td>
<td>5.7Mb dup</td>
<td>chr8:140459470-146280161</td>
<td>1-21</td>
<td>KCNK9, CHRAC1, EIF2C2, PTK2, AX748239, Many genes</td>
<td>Unknown</td>
<td>Not reported</td>
<td>No</td>
</tr>
<tr>
<td>13</td>
<td>Sanders et al, 2011 #11194</td>
<td>Unknown</td>
<td>1.291Kb Del</td>
<td>chr8:141120454-141121745</td>
<td>N/O</td>
<td>AK130220</td>
<td>Unknown</td>
<td>mother</td>
<td>No</td>
</tr>
<tr>
<td>14</td>
<td>Sanders et al, 2011 #11290</td>
<td>Unknown</td>
<td>21.143Kb Del</td>
<td>chr8:141203709-141224852</td>
<td>N/O</td>
<td>AK130220</td>
<td>Unknown</td>
<td>father</td>
<td>No</td>
</tr>
<tr>
<td>15</td>
<td>Sanders et al, 2011 #11290</td>
<td>Unknown</td>
<td>21.143Kb Del</td>
<td>chr8:141203709-141224852</td>
<td>N/O</td>
<td>AK130220</td>
<td>Unknown</td>
<td>Autism proband</td>
<td>No</td>
</tr>
<tr>
<td>16</td>
<td>Sanders et al, 2011 #11644</td>
<td>Unknown</td>
<td>2.881Kb dup</td>
<td>chr8:141117789-141120670</td>
<td>N/O</td>
<td>none</td>
<td>Unknown</td>
<td>mother</td>
<td>No</td>
</tr>
<tr>
<td>17</td>
<td>Sanders et al, 2011 #11825</td>
<td>Unknown</td>
<td>216bp Dup</td>
<td>chr8:141120454-141120670</td>
<td>N/O</td>
<td>none</td>
<td>Unknown</td>
<td>Autism proband</td>
<td>No</td>
</tr>
<tr>
<td>18</td>
<td>Sanders et al, 2011 #11846</td>
<td>Unknown</td>
<td>216bp Dup</td>
<td>chr8:141120454-141120670</td>
<td>N/O</td>
<td>none</td>
<td>Unknown</td>
<td>Autism proband</td>
<td>No</td>
</tr>
<tr>
<td>19</td>
<td>Sanders et al, 2011 #12224</td>
<td>Unknown</td>
<td>216bp Dup</td>
<td>chr8:141120454-141120670</td>
<td>N/O</td>
<td>none</td>
<td>Unknown</td>
<td>father</td>
<td>No</td>
</tr>
<tr>
<td>20</td>
<td>Edwin Kirk</td>
<td>CGH array</td>
<td>~500Kb Dup</td>
<td>chr8:140821366-141268818</td>
<td>17-21</td>
<td>AX748239</td>
<td>Paternal</td>
<td>intellectual disability; microcephalic (HC &lt; 2nd centile)</td>
<td>Yes</td>
</tr>
</tbody>
</table>
Figure 3.3 Validating heterozygous CNVs in autism and ID probands by RT-qPCR. Gene copy number within the TRAPPC9 locus was measured in probands relative to three normal, healthy controls. (A) Confirmation of a maternally-inherited deletion in Proband 1. (B) Confirmation of a paternally-inherited deletion from an unaffected father in Proband 3. (C) Confirmation of a de novo deletion in Proband 8. (D) Confirmation of a parentally inherited duplication in Proband 20.
Chapter 4: Expression analysis of *PEG13*

All work in this chapter was completed by me except the following: Acquisition and processing of human fetal tissues and downstream applications, specifically nucleic acid extraction, RT-PCR and genotyping was conducted by Dr. Miho Ishida from University College London (UK).

Current data supports the evolutionary conservation of mouse *Peg13* and its exclusive expression from the paternal allele in humans. Given that the deletions/duplications of the ID and autism probands lie proximal to or overlap with the putative *PEG13* locus and no other mutations were found at *TRAPPC9*, indicates that the heterozygous CNVs in our probands may affect the expression of *PEG13* in an allele-specific manner. Based on the hypothesis that *PEG13* is maternally imprinted in humans, deletions/duplications proximal to the *PEG13* locus on the paternal allele may affect *PEG13* expression and have downstream effects on *KCNK9* expression. In contrast, disruptions on the maternal allele should have no effect on *PEG13* function or expression, or on *KCNK9* expression.

Allelic expression of *PEG13* (AK307073; chr8:141105709-141107269; hg19) (Figure 4.1) was assessed using human fibroblasts, whole blood and fetal tissues. DNA was genotyped at SNP loci within the transcripts of interest to identify heterozygous individuals and the relative abundance of each allele was quantified by pyrosequencing. Expression profiles of *PEG13* and *KCNK9* were also studied in various somatic and germ line tissues to determine if these genes share similar patterns of expression. Informative SNPs of cDNA sequence suggests that *PEG13* is preferentially expressed from the paternal allele in human fetal brain and kidneys, in addition to fibroblasts, but not in skin, placenta, and whole blood. Moreover, both *PEG13* and *KCNK9* exhibit high expression in human cerebellum and whole brain.
4.1 Materials and Methods: Allelic expression of *PEG13*

4.1.1 Human fetal tissues

Sample Collection

Human fetal tissues, brain, skin, placenta and kidney, were obtained from the Moore Tissue bank at the UCL Institute of Child Health (UK). The Hammersmith, Queen Charlotte’s and Chelsea and Acton Hospital Research Ethics Committee approved the samples (Project Registration 2001 /6029 and 2001/6028).

RNA Isolation and cDNA Synthesis

Tissue samples were homogenized in 1ml TRIzol (Invitrogen) per 100 mg of tissue. TRIzol® Reagent (Invitrogen) extraction of RNA was performed on tissues according to the manufacturer’s instructions. RNA samples were treated with TURBO™ DNase (Ambion®) for residual DNA. First strand cDNA synthesis was performed using 1μg of RNA volume with Promega M-MLV RT in a 25μl reaction according to manufacturer’s protocol. Reverse Transcriptase was omitted in a separate reaction for each sample to screen for DNA contamination. The cDNA samples were shipped to Toronto on dry ice and stored at -20°C.

<table>
<thead>
<tr>
<th>Tissue Type</th>
<th># of heterozygotes</th>
<th># of DNA/cDNA samples available for genotyping</th>
<th># of cDNA samples available for pyrosequencing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brain</td>
<td>10</td>
<td>10</td>
<td>8</td>
</tr>
<tr>
<td>Kidney</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Liver</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Placenta</td>
<td>3</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Skin</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Blood</td>
<td>10</td>
<td>10</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>26</td>
<td>26</td>
<td>14</td>
</tr>
</tbody>
</table>

*Table 4.1. List of human fetal tissues samples heterozygous for SNP rs4289794.* A total of 10 individuals had brain samples in addition to other somatic tissues that were informative for rs4289794. Maternal DNA was available for four individuals. Fourteen samples, including eight brain samples, were available for pyrosequencing.
PCR amplification and genotyping

PCR was performed using BIOTAQ DNA polymerase (Bioline). Cycling conditions began with an incubation at 94°C for 5 min, followed by 35 cycles of 94°C for 30 sec, 58°C for 30 sec and 72°C for 30 sec, and final extension at 72°C for 2 min. DNA and cDNA was amplified by primer sequences in 5’ to 3’ direction for SNP genotyping (Table 4.1). Prior to sequencing, PCR products were purified with microCLEAN (Microzone) following manufacture’s instruction. BigDye® Terminator v1.1 Cycle Sequencing Kit (Life Technologies) carried out the sequencing reaction. Allelic expression was analyzed by comparing the peak heights of the parental alleles in the electropherograms.

4.1.2 Human Fibroblasts and leukocytes

RNA Isolation and cDNA Synthesis

Total RNA was extracted from human fibroblasts and leukocytes using the NucleoSpin® RNA kit (Machery-Nagel) following the manufacturer’s guidelines. First-strand cDNA was synthesized from 1μg of RNA using Superscript III™ Reverse Transcriptase. Reverse Transcriptase was omitted in a parallel reaction for each sample to ensure the absence of DNA.

Genotyping: Identification of SNPs

PCR primers (IDT) are listed in Table 4.2. PCR was carried out using KAPA2G™ Fast HotStart ReadyMx (2X) (Kapa Biosystems) in a 10μl reaction volume containing 1.0μl (30 ng/μl) template. Q-Solution (Qiagen) in place of H2O facilitated the amplification of GC-rich sequences. Each PCR reaction followed an initial denaturation at 95°C and 35 cycles of 95°C for 10 sec, 60°C for 10 sec and 72°C for 1 sec. PCR products were checked for DNA contamination and integrity by 2% gel electrophoresis and Sanger sequenced by TCAG (Toronto).
SNPs Genotyped | PCR Primers (annealing temperature) | Amplicon Size
---|---|---
rs4289794 | GCTTGGTTTGTGCATTCTT TCATTTGGTCTGCTTTCC (58˚C) | 203
rs4289794 & rs2270409 | TTCTGCAGTGCTCCCTA CCAGACAGGACCCAAAGAAA (61˚C) | 794
rs3802217 | CGTCTGGTCCGTAGTGTG GAGTGTTCTCCGTGTTGGAT (61˚C) | 677
rs4455807 | CCTCCACAGGGTAGAACA AGATTGCAGGCTCGTTC (61˚C) | 323

Table 4.2. PCR assays for SNP loci at the putative PEG13 locus. This table outlines the primer sets used to sequence SNPs within the PEG13 region and includes the amplicon size and annealing temperatures used for each PCR program.

4.1.3 Pyrosequencing validation

PCR primer pairs flanked the SNP loci within PEG13 and featured 5’ biotin modification of the reverse primer. Table 4.3 lists all the oligonucleotides used, including sequencing primers. To optimize PCR conditions, cDNA was diluted and aliquoted into tubes for 10 μl reaction as previously described in Chapter 2. PCR for pyrosequencing was carried out in 25 μl reaction volumes comprising 2.0 μl of cDNA template, 12.5 μl of KAPA 2G and 1.0μl (10 μM) of each forward and reverse primer. Amplification conditions were as described in Chapter 2. PCR products were pyrosequenced using the PyroMark™ Q24 System (Qiagen) according to the manufacturer’s instructions and results were analyzed using the AQ software included with the instrument (Chapter 3). Pyrosequencing assays for human fetal tissues were performed in triplicate and we calculated the mean and S.D. for each sample.

4.2 Materials and methods: Tissue expression of PEG13

The functional role of PEG13 in humans is currently unknown. An important aspect in characterizing PEG13 is a description of its tissue expression profile. As previously mentioned, intronic long ncRNAs with regulatory functions tend to share similar expression patterns as their protein-coding counterparts. Therefore, the comparative C_T method for qPCR was used to show the relative expression of PEG13 and KCNK9 across various human tissues. Here, we show that concomitant to KCNK9, PEG13 expression is restricted to the brain and cerebellum.
4.2.1 Sample collection

Human RNA was obtained from the Human Total RNA Master Panel II (BD Biosciences Clontech). Fibroblast and leukocyte RNA were from normal, healthy individuals.

4.2.2 cDNA synthesis

First-strand cDNA was synthesized according to standard procedures from 1.0 μg RNA using Superscript Reverse® Transcriptase III (Life Technologies) (Chapter 3). Prior to use, RNA was diluted 1:1 with nuclease-free H₂O (Ultrapure™ Invitrogen™). Reverse Transcriptase was omitted in parallel reactions for each sample to ensure the absence of DNA contamination.

4.2.3 RT-qPCR

All oligonucleotides were purchased from IDT and designed using the primer express software (Applied Biosystems Inc., Foster city, CA). With the exception of PEG13, primers spanned intron junctions between exons to limit amplification to the coding sequence (CDS), as well as to prevent amplification of genomic DNA. All cDNA samples and primers were optimized by standard PCR and visualized under UV light following gel electrophoresis. Because the \( \Delta \Delta C_T \) method assumes equal amplification efficiency for all samples, a serial dilution for each primer set was created. qPCR was carried out using SYBR green reagents and performed in quadruplicate in a 384-well plate. A NTC was included for each primer set. Cerebellum was used as a reference gene and Human B-Actin was used as an internal control. Data was normalized against B-actin prior to statistical analysis using GraphPad Prism Software and values were represented as a fold-change in expression compared to cerebellar tissue. An unpaired two-tailed Student’s t-test was used to assess differences in gene expression across tissue samples.
<table>
<thead>
<tr>
<th>SNP and region</th>
<th>Amplification Primers (annealing temperature °C)</th>
<th>Amplicon Size</th>
<th>Sequencing Primer</th>
<th>Sequence length</th>
<th>Nucleotide dispensation</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs4289794 chr8:141,106,145-141,106,258</td>
<td>Forward GGCGCACATTTCCACCTAGT Reverse (B)-GGCTTGGTTTGTGCATTCT (62°C)</td>
<td>114</td>
<td>Primer 1 GGGACTCATAAATATGACT Primer 2 GGACCTCATAAATATGACTG</td>
<td>36</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>27</td>
<td>4</td>
</tr>
<tr>
<td>rs2270409 chr8:141,105,656-141,105,838</td>
<td>Forward CCAGACAGGACCCAAAAGAAA Reverse (B) – CTCCTCTCCCAAGACCTCT (52°C)</td>
<td>183</td>
<td>Primer 1 GAGGGGGACTCATAAATAT Primer 2 CACTCATAAATATGACTG</td>
<td>16</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>28</td>
<td>4</td>
</tr>
<tr>
<td>rs3802217 chr8:141,107,907-141,108,043</td>
<td>Forward CCCAGGGTCATGACGCCTA Reverse (B) - GGTGGGCATCTTTATCTTCA (58°C)</td>
<td>139</td>
<td>Primer 1 CAGTCTTACCGGGAAG</td>
<td>23</td>
<td>6</td>
</tr>
</tbody>
</table>

Table 4.3 Oigonucleotides used for quantifying allelic expression of SNPs by pyrosequencing. This table outlines the PCR primer sets used and the annealing temperatures for the PCR program, in addition to the sequencing primer used for pyrosequencing. *(B) denotes 5’ nucleotide biotin modification.

**Figure 4.1 Schematic of the PEG13 locus within intron 17 of TRAPPC9** (not drawn to scale). The blue rectangular boxes represent the PEG13 transcript, which encompasses AK307073 (chr8:141,105,709-141,107,269) and AX748239. (chr8:141,111,257-141,113,775). The SNPS rs2270409, rs4289794, rs3802217, rs4455807 and rs7828256 were used to assess allelic expression of PEG13. The green rectangular box is the CpG island embedded within PEG13 and the black rectangular boxes show the centromere (Cen) and telomere (Tel).
4.3 Results

4.3.1 *PEG13* is a novel paternally expressed gene in human fetal brain

Fifty-five DNA samples from human fetal tissues were genotyped for SNP rs4289794 to identify heterozygous individuals (Figure 4.1). Among the twenty-nine samples that were informative, ten had brain samples available in addition to other somatic tissues for RNA extraction (Appendix 4).

*PEG13* showed predominant or preferential expression from one allele in seven brain samples, and had biased expression in fetal kidney. Three brain samples, however, expressed both gene copies although this may be due to the heterogeneous nature of the brain tissues such that imprinting of *PEG13* may be restricted to a specific brain region such as cerebellum. Biallelic expression was observed in all other tissues tested i.e. placenta and skin (Table 4.4; Appendix 4). The parental genotype was available for one case and confirmed that *PEG13* is paternally-expressed, as the mother was homozygous (C/C, homozygous) for the non-expressed allele (Figure 4.2). Relative allelic expression was quantified for the seven brain samples by pyrosequencing and is presented in Figure 4.3 as a ratio of allelic expression: 0:100 represents imprinted expression (silencing of one allele) and 50:50 represents biallelic expression. Since these exact ratios are not always represented biologically, transcripts were considered to be preferentially or exclusively expressed (imprinted) if the difference in allelic bias between the alleles was > 60% for one allele, whereas expression between 40-60% for an allele indicated biallelic expression. Out of the seven heterozygous samples, only six samples confirmed biased expression in fetal brain (including the individual with the maternal genotype) (Figure 4.3; Appendix 4). The discrepancy between the Sanger sequencing data and pyrosequencing data may be attributed to bias in the PCR reaction or sequence heterogeneity.
Figure 4.2 Electropherograms showing the imprinting status of PEG13. SNP rs4289794 was used to assess the parental origin of PEG13 expression in human fetal brain. A) Genomic DNA sequence of case 1 (C/T heterozygous). B) Brain cDNA sequence of case 1 (Monoallelic, T expression). C) DNA sequence of case 1’s mother, who is homozygous for the non-expressed allele (C/C homozygous).

<table>
<thead>
<tr>
<th>Case</th>
<th>Tissue</th>
<th>Active allele</th>
<th>Active allele expression (%)</th>
<th>Maternal genotype</th>
<th>GA (weeks)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Brain</td>
<td>T</td>
<td>87%</td>
<td>C/C</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>Kidney</td>
<td>Pref T</td>
<td>76%</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Liver</td>
<td>C/T</td>
<td>N/A</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Skin</td>
<td>C/T</td>
<td>58%</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Placenta</td>
<td>C/T</td>
<td>N/A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Brain</td>
<td>C</td>
<td>77%</td>
<td>C/T</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>Skin</td>
<td>C/T</td>
<td>54%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Brain</td>
<td>C</td>
<td>45%</td>
<td>C/T</td>
<td>10</td>
</tr>
<tr>
<td>4</td>
<td>Brain</td>
<td>C</td>
<td>62%</td>
<td>C/T</td>
<td>10</td>
</tr>
<tr>
<td>5</td>
<td>Brain</td>
<td>T</td>
<td>89%</td>
<td>NA</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>Placenta</td>
<td>C/T</td>
<td>N/A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Brain</td>
<td>T</td>
<td>90%</td>
<td>NA</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>Placenta</td>
<td>C/T</td>
<td>67%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Brain</td>
<td>T</td>
<td>90%</td>
<td>NA</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>Placenta</td>
<td>C/T</td>
<td>53%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Brain</td>
<td>C/T</td>
<td>N/A</td>
<td>NA</td>
<td>14</td>
</tr>
<tr>
<td>9</td>
<td>Brain</td>
<td>C/T</td>
<td>57%</td>
<td>NA</td>
<td>15</td>
</tr>
<tr>
<td>10</td>
<td>Brain</td>
<td>C/T</td>
<td>67%</td>
<td>NA</td>
<td>10</td>
</tr>
</tbody>
</table>

Table 4.4 Number of informative heterozygous samples for SNP rs4289794 at PEG13. Seven out of ten brain samples showed predominant expression from an allele and was confirmed in six brain samples by pyrosequencing. The paternal origin of PEG13 expression was confirmed by case 1’s maternal genotype who was homozygous (C/C) for the non-expressed allele. For sample 3 (red), the pyrosequencing assay did not validate the Sanger Sequencing data. N/A = sample cDNA was not available; Pref = preferential expression from one allele (both alleles were expressed; however the peak height for one allele was greater than the other allele in the electropherogram).
4.3.2 Allelic expression of *PEG13* is conserved in human fibroblasts

Fibroblasts were analyzed in parallel to fetal tissues to confirm imprinting of *PEG13* in humans. Allelic expression was quantified by pyrosequencing in ten unrelated fibroblast cell lines heterozygous for the SNPs rs2270409; rs4289794; and rs3802217 (refer to Table 3.2 for list and SNP array data). Based on the assumption that the SNPs tested belong to the same non-coding transcript, it would be expected that all SNPs have preferential expression. All individuals, even those that did not exhibit preferential expression from the data provided by Dr. Tomi Pastinen, had predominant to exclusive allelic expression at all SNPs (Figure 4.4; Appendix 5). The discrepancy between the findings may be due to experimental error, either with the inaccuracy of reads from the Illumina platform or PCR amplification bias for the pyrosequencing reaction. The results, however, are consistent with predictions for a novel imprinted locus. Indeed, a distinguishing feature for imprinted genes is concordant allelic or biased expression at neighbouring SNPs within the region of interest (DeVeale et al., 2012). Furthermore, the degree of allele-specific expression (ASE) appeared to occur independent of age (5 months to 54 years), and no correlation was observed between age and the degree of biased expression. To determine the parental transmission of the expressed allele, four fibroblast trios (fibroblast RNA from proband and leukocyte DNA from parents) were assessed for the abovementioned SNPs, in addition to rs4455807. The probands were screened for and did not have any developmental or neuropsychiatric disorders. Among the probands who were informative, manual inspection of electropherogram peaks showed preferential or exclusive expression at each SNP with the exception of two probands (Proband 1 and 3) at rs2270409. The parental genotypes that were informative were available (Figures 4.5, 5.7, 4.9 and 4.11) and validated the paternal expression of *PEG13* where the maternal genotype was homozygous for the non-expressed allele. Tables 4.5-4.8 summarize the relative expression at each SNP. Pyrosequencing data was in agreement with the Sanger sequencing data (Figures 4.6, 4.8, and 4.10). Therefore, we confirmed that PEG13 is preferentially-expressed from the paternal allele in humans.
Figure 4.3 Relative expression of PEG13 alleles in human fetal tissues. The coloured bars represent the proportion of expression (%) at PEG13 rs4189794. PEG13 shows biased expression in human fetal brain and kidney, and biallelic expression in skin and placenta. Pooled commercial RNA was run in parallel, indicating that PEG13 shows biased expression in human cerebellum and kidney. C = Cytosine; T = Thymine
Figure 4.4 Relative allelic expression of *PEG13* in human fibroblasts. The coloured bars represent the ratio of allelic expression (%) at *PEG13* rs4189794 and rs3802217. PEG13 shows preferential expression in most individuals, however the parent-of-origin could not be determined as parental DNA was not available. A = adenine; C = cytosine; G = guanine; T = thymine
**Figure 4.5** Electropherograms showing the allelic bias of *PEG13* at rs2270409. SNP rs2270409 was used to assess the parental origin of *PEG13* allelic expression in fibroblast trios. A) Genomic DNA sequence of proband 2 (G/A heterozygous). B) Fibroblast cDNA sequence of proband 2 (preferential expression from the A allele). C) DNA sequence of proband 2’s mother, who is homozygous for the non-expressed allele (G/G). D) Genomic DNA sequence of proband 2’s father who carries the preferentially expressed allele (G/A, heterozygous).

<table>
<thead>
<tr>
<th>Family</th>
<th>Active allele</th>
<th>Active allele expression (%)</th>
<th>Maternal Genotype</th>
<th>Proband Genotype</th>
<th>Age (years)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>G</td>
<td>99%</td>
<td>G/G</td>
<td>G/A</td>
<td>21</td>
</tr>
<tr>
<td>2</td>
<td>Pref. A</td>
<td>60%</td>
<td>G/G</td>
<td>G/A</td>
<td>20</td>
</tr>
<tr>
<td>3</td>
<td>G</td>
<td>77%</td>
<td>G/A</td>
<td>G/A</td>
<td>7</td>
</tr>
<tr>
<td>4</td>
<td>G/A</td>
<td>60%</td>
<td>G/A</td>
<td>G/A</td>
<td>18</td>
</tr>
</tbody>
</table>

Table 4.5 Number of informative heterozygous samples for SNP rs2270409 at *PEG13*. Two out of four probands showed predominant expression from an allele, and was confirmed by pyrosequencing. The paternal origin of *PEG13* expression was confirmed by proband 2’s maternal genotype, who was homozygous (G/G) for the non-expressed allele. Pref.A = preferential expression from adenine (the electropherogram showed expression for both alleles; however A had a higher peak).

**Figure 4.6** Relative expressions of *PEG13* alleles in fibroblasts at rs2270409. The coloured bars represent the proportion of expression (%) at *PEG13* rs2270409. PEG13 shows preferential expression in two probands. DNA sample from a control leukocyte that was heterozygous for the SNP was used as a positive control.
**Figure 4.7** Electropherograms showing the allelic bias of PEG13 at rs4289794. SNP rs4289794 was used to assess the parental origin of PEG13 allelic expression in fibroblast trios. A) Genomic DNA sequence of proband 1 (C/T heterozygous). B) Fibroblast cDNA sequence of proband 1 (Preferential A expression). C) DNA sequence of proband 1’s mother, who is homozygous for the non-expressed allele (C/C). D) Genomic DNA sequence of case 2 father who carries the expressed alleles (C/T, heterozygous).

<table>
<thead>
<tr>
<th>Family</th>
<th>Active allele</th>
<th>Active allele expression (%)</th>
<th>Maternal Genotype</th>
<th>Proband Genotype</th>
<th>Age (years)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>T</td>
<td>86%</td>
<td>C/C</td>
<td>C/T</td>
<td>21</td>
</tr>
<tr>
<td>2</td>
<td>C</td>
<td>81%</td>
<td>C/T</td>
<td>C/T</td>
<td>20</td>
</tr>
<tr>
<td>3</td>
<td>C</td>
<td>71%</td>
<td>C/T</td>
<td>C/T</td>
<td>7</td>
</tr>
<tr>
<td>4</td>
<td>C</td>
<td>94%</td>
<td>C/T</td>
<td>C/C</td>
<td>18</td>
</tr>
</tbody>
</table>

**Table 4.6** Number of informative heterozygous samples for SNP rs4289794 at PEG13. All four probands showed predominant to exclusive allelic expression and was confirmed by pyrosequencing. The paternal origin of PEG13 expression was confirmed by case 1’s maternal genotype, who was homozygous (C/C) for the non-expressed allele.

**Figure 4.8** Relative expressions of PEG13 alleles in fibroblasts. The coloured bars represent the proportion of expression (%) at PEG13 rs4289794. PEG13 shows preferential expression in all probands. DNA sample from a control leukocyte that was heterozygous for the SNP was used as a positive control.
**Figure 4.9 Electropherograms showing the allelic bias of PEG13 at rs380221.** SNP rs380221 was used to assess the parental origin of PEG13 allelic expression in fibroblast trios. A) Genomic DNA sequence of proband 2 (C/T heterozygous). B) Fibroblast cDNA sequence of proband 1 ( Preferential A expression). C) DNA sequence of proband 2’s mother, who is homozygous for the non-expressed allele (C/C). D) Genomic DNA sequence of case 2 father who carries the expressed alleles (C/T, heterozygous).

<table>
<thead>
<tr>
<th>Family</th>
<th>Active allele</th>
<th>Active allele expression (%)</th>
<th>Maternal Genotype</th>
<th>Proband Genotype</th>
<th>Age (years)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>C</td>
<td>99%</td>
<td>C/C</td>
<td>C/T</td>
<td>21</td>
</tr>
<tr>
<td>2</td>
<td>Pref. C</td>
<td>81%</td>
<td>C/T</td>
<td>C/T</td>
<td>20</td>
</tr>
<tr>
<td>3</td>
<td>C</td>
<td>84%</td>
<td>C/T</td>
<td>C/T</td>
<td>7</td>
</tr>
<tr>
<td>4</td>
<td>C</td>
<td>99%</td>
<td>C/C</td>
<td>C/T</td>
<td>18</td>
</tr>
</tbody>
</table>

**Table 4.7 Number of informative heterozygous samples for SNP rs380221 at PEG13.** All probands showed predominant or exclusive allelic expression and was confirmed by pyrosequencing. The paternal origin of PEG13 could not be confirmed in any of the fibroblasts Pref. C = the electropherogram showed expression for both alleles, however C had a higher peak.

**Figure 4.10 Relative expressions of PEG13 alleles in fibroblasts.** The coloured bars represent the proportion of expression (%) at PEG13 rs380221. PEG13 SHOWS exclusive expression in all probands. DNA sample from a control leukocyte who was heterozygous for the SNP was used as a positive control.
**Figure 4.11 Electropherograms showing the allelic bias of PEG13 at rs4455807.** SNP rs4455807 was used to assess the parental origin of PEG13 allelic expression in fibroblast trios. A) Genomic DNA sequence of proband 3 (C/T heterozygous). B) Fibroblast cDNA sequence of proband 3 (Preferential A expression). C) DNA sequence of proband 3’s mother, who is homozygous for the non-expressed allele (G/G). D) Genomic DNA sequence of case 2 father who carries the expressed alleles (G/T, heterozygous).

<table>
<thead>
<tr>
<th>Family</th>
<th>Active allele</th>
<th>Active allele expression (%)</th>
<th>Maternal Genotype</th>
<th>Proband Genotype</th>
<th>Age (years)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>G</td>
<td>N/A</td>
<td>G/G</td>
<td>G/G</td>
<td>21</td>
</tr>
<tr>
<td>2</td>
<td>Pref. T</td>
<td>N/A</td>
<td>G/T</td>
<td>G/T</td>
<td>20</td>
</tr>
<tr>
<td>3</td>
<td>T</td>
<td>N/A</td>
<td>G/G</td>
<td>G/T</td>
<td>7</td>
</tr>
<tr>
<td>4</td>
<td>G</td>
<td>N/A</td>
<td>G/G</td>
<td>G/G</td>
<td>18</td>
</tr>
</tbody>
</table>

**Table 4.8 Number of informative heterozygous samples for SNP rs4455807 at PEG13.** All probands showed predominant expression from an allele; however this could not be confirmed by pyrosequencing. The paternal origin of PEG13 expression was confirmed by proband 3’s maternal genotype, who was homozygous (G/G) for the non-expressed allele Pref. T = the electropherogram showed expression for both alleles, however C had a higher peak; N/A = not available.
4.3.3 PEG13 is highly expressed in human brain

PEG13 and KCNK9 displayed similar expression profiles in both germ and somatic tissues (Figure 5.14 a and b). Both genes were prominently expressed in the brain with highest expression in the cerebellum (p<0.5), indicating that PEG13 expression and its function may also be region-specific in the CNS. PEG13 exhibited minimal expression in skeletal muscle, kidney and placenta, while KCNK9 expression was non-existent in other tissues except for kidney and adrenal gland. The absence of PEG13 and KCNK9 in leukocytes supports earlier findings that PEG13 is biallelically-expressed in this cell. Consequently, we could not address the hypothesis that CNVs in the ID and Autism probands may have downstream effects on PEG13 and KCNK9 expression.

4.3.4 PEG13 in Autism and ID Probands

We postulated that CNVs overlapping TRAPPC9 and adjacent to the putative PEG13 locus may alter the expression of KCNK9 and PEG13 according to the parent of origin. RNA and DNA from leukocytes extracted from whole blood were available for Proband 3 (including father) and Proband 8, while only RNA was available for Proband 1. However, given that leukocytes do not express PEG13 or KCNK9 (Figure 5.14b), RT-qPCR could not accurately measure gene expression in our probands. Instead, DNA was genotyped for informative SNPs within the PEG13 locus to determine if PEG13 is imprinted in human leukocytes. Proband 8 and the father of Proband 3 were heterozygous for PEG13 rs7828256, while Proband 3 did not have any informative SNPs (Figure 4.15). The cDNA of Proband 1 was also genotyped and showed expression for both gene copies at rs4289793, indicating that PEG13 is biallelically expressed in human blood. Similarly, both alleles were expressed in the cDNA sequences in Proband 8 and the father of Proband 3. Although we could not elucidate any changes in PEG13 expression, based on the fact that PEG13 imprinting and expression is restricted to the brain, suggests that the CNVs may manifest specifically in the brain, and the genetic aberrations observed in our probands are not detectable with our current materials.
Figure 4.12 Relative expression of *PEG13* and *KCNK9* in human tissues and cells by RT-qPCR. A) *KCNK9* expression is restricted to the brain, particularly the cerebellum. B) *PEG13* has a similar expression pattern as *KCNK9*, but also shows minimal expression in other somatic tissues, including skeletal muscle, kidney, adrenal gland and placenta. Both genes are not expressed in leukocytes. */ ** = p <0.05
Figure 4.13 Electropherograms in ID and autism probands with deletions proximal to PEG13. Deletions in probands had no effect on PEG13 expression levels in leukocytes. A) Proband 1 showed biallelic expression at rs4289794. b) Leukocyte cDNA sequence of proband 3’s father who is heterozygous for rs7828256 (C/G). c) Father’s cDNA sequence was heterozygous (C/G), indicating that PEG13 is not imprinted in leukocytes. Proband 3 did not have any informative SNPs at the PEG13 locus; therefore the paternal DNA was used instead. d) Proband 8 DNA sequence at rs7828256 (C/G, heterozygous). E) Proband 8 cDNA sequence was heterozygous ((C/G, heterozygous), which also confirmed biallelic expression of PEG13 in human leukocytes.
Chapter 5: Methylation analysis of CpG islands

All work in this chapter was completed by me except the following: Identification of the PEG13-DMR in fetal tissues was performed by Dr. Kazu Nakabayashi from the National Research Institute for Child Health and Development (Tokyo, Japan). Sperm DNA was provided by The Krembil Family Epigenetics Lab (CAMH, Toronto).

Differentially methylated CGIs are predictive features of imprinted genes, and to date, the ICs for a number of long ncRNAs have been identified as maternally-methylated DMRs. Here, we assessed methylation patterns at the PEG13-associated CGI and the KCNK9-associated CGI by bisulfite pyrosequencing in fibroblasts and leukocytes, in conjunction to combined bisulfite restriction analysis (COBRA) in fetal tissues. Methylation was also studied in Probands 3, 8 and 20 to identify whether the CNVs affect methylation at the PEG13-DMR or KCNK9 CGI. As PEG13 is a paternally expressed gene, we hypothesized that methylation would be maternally derived and therefore, the DMR would be hypomethylated in sperm DNA.

We identified the PEG13-DMR in human brain, whole blood and fibroblasts, and maternal-specific methylation in placenta. Low methylation levels were indicative of hypomethylation in sperm and a germ line-derived DMR, corroborating our findings that PEG13 is a paternally expressed gene.

5.1 Materials and Methods

5.1.1 Sample collection

Isolation of DNA from Human Sperm

Sperm DNA from 10 healthy males (Table 5.1) was extracted using a standard phenol-chloroform method. Briefly, semen was washed twice with 150mM NaCl and 10mM EDTA solution (pH 8.0) and centrifuged at 1000 x g for 10min. The pellet was vortexed and resuspended in lysis buffer (6.0
guanidine thiocyanate, 5 M NaCl, 30% Sarkosyl, 1M DTT, 20, 000 mg/ml, 240μl H2O) and incubated at 56°C for 2 h, after which 2.4 ml of isopropanol was added for DNA precipitation. The spooled DNA was removed from solution and transferred to a tube containing 2.0ml 0.1M sodium citrate in 10% EtOH. After 30 min at RT, the buffer was removed and DNA was washed with 70% EtOH and air-dried. The DNA was rehydrated in 10 mM Tris-HCl (pH 8.0), and incubated at 65°C for 1h until DNA was in solution.

Isolation of DNA from whole blood

The sample set (n=10) was taken from 440 unrelated individuals of European Caucasian ancestry recruited at The Center for Addiction and Mental Health (CAMH, Toronto) and when possible, age-matched to the sperm DNA sample set (Table 5.2). Individuals were screened and negative for psychiatric conditions and serious medical conditions as described by Xu et al., (2014). DNA was isolated from whole blood using a high-salt method as described in Chapter 3. Individual fibroblast cell lines were obtained from the Coriell Institute (Camden, New Jersey) (Table 3.2). DNA was extracted from fibroblasts using the Nucleospin® RNA kit following manufacturer’s protocol as previously described in Chapter 3.

<table>
<thead>
<tr>
<th>Sample</th>
<th>Age</th>
<th>Medical problem</th>
<th>Smoker</th>
<th>Rx Meds</th>
<th>Supplements</th>
<th>Ethnicity</th>
</tr>
</thead>
<tbody>
<tr>
<td>C37</td>
<td>21</td>
<td>N/I</td>
<td>N</td>
<td>-</td>
<td>-</td>
<td>Caucasian</td>
</tr>
<tr>
<td>C117</td>
<td>24</td>
<td>CN</td>
<td>N</td>
<td>-</td>
<td>CN-VitD</td>
<td>Caucasian</td>
</tr>
<tr>
<td>C10</td>
<td>25</td>
<td>Eczema</td>
<td>Y</td>
<td>Hydrocortis</td>
<td>-</td>
<td>Caucasian</td>
</tr>
<tr>
<td>C46</td>
<td>29</td>
<td>HBP</td>
<td>Y</td>
<td>-</td>
<td>VitD</td>
<td>Caucasian</td>
</tr>
<tr>
<td>C16</td>
<td>32</td>
<td>N/I</td>
<td>N</td>
<td>-</td>
<td>CN</td>
<td>Caucasian</td>
</tr>
<tr>
<td>C77</td>
<td>33</td>
<td>CN</td>
<td>N</td>
<td>CN</td>
<td>CN</td>
<td>Caucasian</td>
</tr>
<tr>
<td>C28</td>
<td>39</td>
<td>Cleft palate</td>
<td>N</td>
<td>-</td>
<td>-</td>
<td>Caucasian</td>
</tr>
<tr>
<td>C104</td>
<td>40</td>
<td>CN</td>
<td>Y</td>
<td>Cephalexin</td>
<td>CN</td>
<td>Mixed</td>
</tr>
<tr>
<td>C44</td>
<td>41</td>
<td>N/I</td>
<td>N</td>
<td>-</td>
<td>Multivitamin</td>
<td>Asian</td>
</tr>
<tr>
<td>C125</td>
<td>52</td>
<td>CN</td>
<td>N</td>
<td>CN</td>
<td>CN</td>
<td>Caucasian</td>
</tr>
</tbody>
</table>

Table 5.1 List and description of individuals from which sperm DNA was collected. DNA from ten sperm samples was used to assess methylation at PEG13-DMR. CN = ; HBP = high blood pressure; N/I = not identified; N = No Y = Yes
<table>
<thead>
<tr>
<th>Sample</th>
<th>Age</th>
<th>Ethnicity</th>
</tr>
</thead>
<tbody>
<tr>
<td>GBP 7208</td>
<td>25</td>
<td>Caucasian</td>
</tr>
<tr>
<td>GBP 7014</td>
<td>26</td>
<td>Caucasian</td>
</tr>
<tr>
<td>GBP 7148</td>
<td>26</td>
<td>Caucasian</td>
</tr>
<tr>
<td>GBP 7031</td>
<td>29</td>
<td>Caucasian</td>
</tr>
<tr>
<td>GBP 7012</td>
<td>33</td>
<td>Caucasian</td>
</tr>
<tr>
<td>GBP 7113</td>
<td>33</td>
<td>Caucasian</td>
</tr>
<tr>
<td>GBP 7288</td>
<td>39</td>
<td>Caucasian</td>
</tr>
<tr>
<td>GBP 7058</td>
<td>40</td>
<td>Caucasian</td>
</tr>
<tr>
<td>GBP 7047</td>
<td>41</td>
<td>Caucasian</td>
</tr>
<tr>
<td>GBP 7207</td>
<td>52</td>
<td>Caucasian</td>
</tr>
</tbody>
</table>

**Table 5.2 List and description of control leukocyte DNA.** DNA isolated from whole blood for 10 healthy males that were age-matched to DNA from sperm samples.

Isolation of DNA from Fibroblasts

### 5.1.2 Bisulfite Conversion of DNA

Bisulfite conversion for DNA samples from fibroblasts, leukocytes and sperm was carried out with the EZ DNA Methylation-Lightning Kit™ (Zymo Research) using 1.0μg of DNA. Converted DNA was eluted in 20μl of TE buffer (pH 8.0). A brief summary of the reaction protocol is outlined below in Figure 6.1.

**Figure 5.1** Bisulfite conversion involves three main steps following DNA strand separation (98°C for 8 min). Steps (1) and (2), the actual bisulfite conversion, occurs at 54°C for 1 h, and step (3) occurs at RT for 20 min. (1) Sulphonation: the addition of bisulfite (HSO₃⁻) to the 5-6 bond of cytosine; (2) Hydrolic deamination: the removal of the amine group (NH₂) from the cytosine-bisulfite intermediate, resulting in the loss of ammonia and the creation of a cytosine-uracil derivative; (3) Alkali desulphonation: the removal of sulphonate from the uracil ring by alkali treatment. 5-mC is protected from sodium bisulfite and retains its integrity as a cytosine residue.

Bisulfite conversion involves the deamination of unmethylated cytosine residues to uracil by sodium bisulfite while leaving methylated cytosines intact (Figure 3.3) (Frommer et al., 1992; Clark et al.,
Because DNA polymerases cannot discriminate between uracil and thymine, subsequent PCR amplifications replace uracil with thymine while amplifying 5mC (and 5hmC) residues as cytosines (Warnecke et al., 2002). Downstream applications for bisulfite-treated DNA generally involve assays that quantitatively determine the ratio of thymine “T” versus cytosine “C” at specific CpG sites. Both strands can be studied individually for DNA methylation due to a loss of DNA complementarity following bisulfite conversion.

Bisulfite treatment is advantageous primarily because of its efficiency in converting ~99% of all cytosine residues to uracil (Holmes et al., 2014). However, despite the range of bisulfite-based applications, there remain certain limitations and caveats. First, bisulfite conversion leads to considerable DNA degradation. Earlier bisulfite conversion methods involved the use of harsh conditions (i.e. long incubation periods, strong chemical agents and high temperatures), resulting in fragmented single stranded DNA and very low DNA concentration (Raizis et al., 1995). However, while conditions have improved with the introduction of ready-made kits and decreased incubation times, up to 90% of DNA undergoes degradation following bisulfite conversion. Therefore both DNA quality and quantity prior to bisulfite treatment is important before commencing downstream applications, particularly for PCR, which depends on the number of DNA molecules. Secondly, the formation of artifacts such as incomplete conversion and strand reannealing can lead to artificial results (Warnecke et al., 2002). Furthermore, errors in PCR amplification (e.g. PCR bias and hybrid products) can hinder sequencing results due to background interference (Warnecke et al., 2002). Optimization of bisulfite PCR conditions – annealing temperature, DNA polymerase and Nested PCR –can circumvent some amplification-based problems (Li & Tollefsbol, 2011).

5.1.3 Quantification of DNA methylation

CGI analysis and primer design
CpG islands were identified using EMBOSS CpGplot (EMBL-EBI) and Methyl Primer Express v1.0 software (Life Technologies: http://www.appliedbiosystems.com). Primers for PCR amplification were designed using Microsoft Word and followed recommended guidelines for bisulfite treated DNA (Patterson et al., 2011):

1. Primers were 25-30 bp in length with Primers a melting temperature of ~ 60˚C, and did not differ by more than 1˚C for each pair
2. Primers had ~20% T-content (including both converted and non-converted T)
3. The 3’ end of primers (when possible) were a converted thymine (T) to ensure amplification of bisulfite converted DNA
4. Primers were not designed within repetitive regions, palindromes, self-chain alignments or polymorphic regions
5. Primer sequences were devoid of CpG nucleotides to prevent biased PCR amplification
6. Amplicon length did not exceed 300 bp to ensure maximum yield of the PCR product and to facilitate amplicon capturing during the pyrosequencing reaction

PCR assays were designed to amplify a part of the CGIs associated with *PEG13* and *KCNK9* and encompassed 5 to 7 CpG sites. All reverse primers featured a 5’ biotin modification to create a forward sequencing assay.

5.1.4 PCR amplification

Each PCR reaction contained 2.0-3.0μl of bisulfite-treated DNA in a 25μl reaction volume and was carried out using 12.5μl of KAPA2G™ Fast HotStart ReadyMx (2X) (Kapa Biosystems), 7.5-8.5μl H₂O and 1.0μl of each forward and reverse primer (10μM). DNA or polymerase was omitted in a separate reaction for each sample to check for DNA contamination. Cycling conditions began with an initial incubation at 95˚C for 1 min, followed by 40 cycles of 95˚C for 10 sec, 50˚C to 63˚C for 10sec, and 72˚C for 1 sec. Table 6.3 lists the specific annealing temperatures and oligonucleotides
used for each amplicon (IDT). The quality and quantity of PCR product in addition to DNA contamination was verified by 2% gel electrophoresis.

5.1.5 Pyrosequencing and statistical analysis

PCR products were pyrosequenced using the PyroMark™ Q24 System according to the manufacturer’s instructions (Qiagen) and each CpG site was analyzed using the pyrogram generated by the CpG software provided by the manufacturer where methylation at each CpG site was represented as a percentage (%) ranging from 0% to 100% (Chapter 3). Methylation profiles were visualized as a box and whisker plot showing the median and interquartile ranges at a single CpG site for all samples (n=10), except for methylation analysis in the ID and autism probands. Statistical analysis was performed using GraphPad Software and involved calculations of the median, S.D., and interquartile values.

5.1.6 Combined-bisulfite restriction analysis

Combined-bisulfite restriction analysis (COBRA) was performed to identify the PEG13-DMR and allele-specific methylation (ASM) in DNA samples from human placenta and fetal brain. DNA samples were treated to bisulfite by The EZ DNA Methylation-Gold Kit™ (Zymo Research) using 1.0 μg of DNA. COBRA-primers used to investigate the PEG13-DMR were: forward (KS185) 5’TTATAGTTATTTGTAGGGTAGGGAA-3’; and reverse (KS186) 5’CTTAAAAACCTAAACTAAATATCCTCCC-3’. The primers were complementary to bisulfite treated DNA and did not contain CpG sites. Following PCR amplification, amplicons were digested by BstUI restriction enzyme (New England BioLabs) and fragments were separated by polyacrylamide gel electrophoresis (PAGE). Quantification of methylation was represented as a ratio of all digested fragments to the total number of undigested and digested fragments.
Table 5.3 DNA methylation assays. PCR conditions and oligonucleotides used for bisulfite pyrosequencing and regions analyzed across all biological samples for sperm, leukocytes and fibroblasts. (B) represents 5′ biotin modification of the primer.

<table>
<thead>
<tr>
<th>Region</th>
<th>Amplification Primers and hybridization temperature</th>
<th>Product length (bp)</th>
<th>Sequencing Primers</th>
<th>No. of CpGs</th>
<th>Sequence length</th>
<th>No. of nucleotide dispensation</th>
</tr>
</thead>
</table>
| **PEG13 DMR** chr8:141,110,441-141,110,641 | Forward GAGTYGTAGTGGTATATGGGAAGATT  
                    Reverse (B)- AATACRACACCCAACCAATACTCCAA | 201 | Primer 1 GTGGTATATGGAAGATTTGG | 7           | 50             | 49                          |
| **PEG13, DMR** chr8:141,108,632-1411087555 | Forward GATGTTGAAAGTTATGATTGAGATT  
                    Reverse (B)-ACACCACCTCCACAATATAAA | 122 | Primer 1 GGTGTTTTATTTAGTGGA | 7           | 46             | 40                          |
| **PEG13, DMR** chr8:141,108,731-141,108,8807 | Forward GTTATTATTTTTGTATAGTAGGAATGT  
                    Reverse (B)-AATCTCAATACATAACTTCAACATCC | 140 | Primer 1 TATTGAGATTTAGTGGTG | 7           | 67             | 60                          |
| **PEG13, DMR** chr8:141,107,853-141,108,138 | Forward TAGGATAAAGATGTTAGTTGGAGTTTG  
                    Reverse (B)–ACAACTACCTCATAAAAAACATAAACTAAC (60.5) | 287 | Primer 1 GTTTATTTAGTTGTAGTG | 7           | 61             | 61                          |
| **PEG13, DMR** chr8:141,107,731-141,108,007 | Forward TTTGTTAATAAGGAGTTTTAATGGTG  
                    Reverse (B)–TCAAAATACCTCAATTCACAAAAATACATA | 277 | Primer 1 GTATATGGAAGATTTGG | 5           | 54             | 47                          |
| **KCNK9 CGI** chr8:140,714,342-140,714,644 | Forward TTAGGGGTTTGGGAGTAAAAATTTGG | 333 | Primer 1 GAGTHTAGAGAAGTGG | 4           | 51             | 40                          |
|             | Forward TTAGGGGTTTGGGAGTAAAAATTTGG  
                    Reverse (B)–AAATAATCCCATACAAACACCCCAC | 333 | Primer 2 GTAAGTGTAAGGTTTAGTTG | 5           | 26             | 27                          |
Figure 5.2 Plots of methylation at the *PEG13-CGI* in human leukocyte DNA. Box and whisker plots show median and interquartile ranges at single CpG sites for 10 healthy individuals (Table 6.2). The error bars indicate the S.D across samples. Methylation patterns were assessed at four regions. (A) and (B) Region 1 (chr8:141,107,853-141,108,138) and region 2 (chr8:141,107,731-141,108,007) demonstrate high levels of methylation, synonymous to hypermethylation. (C) and (D) Methylation at region 3 (chr8:141108632-1411087555) and region 4 (chr8:141,108,731-141,108,8807) have differential methylation, indicating a DMR.
Figure 5.3 Plots of methylation at the PEG13-CGI in human fibroblasts DNA. Box and whisker plots show median and interquartile ranges at single CpG sites for 10 healthy individuals (Table 3.1). The error bars indicate the S.D. of methylation in all samples. (A) Methylation profiles at region 1 (chr8:141,107,853-141,108,138), similar to leukocytes, is highly methylated akin to hypermethylation. (B) and (C) Methylation profiles at region 2 (chr8:141108632-141108755) and region 3 (chr8:141,108,731-141,108,8807) showed differential methylation, indicating a DMR.
Figure 5.4 Plots of methylation at the PEG13-CGI in human sperm DNA. Box and whisker plots show median and interquartile ranges at single CpG sites for 10 healthy individuals. The error bars indicate the S.D. across samples. (A) Methylation profiles at Region 1 (chr8:141,108,731-141,108,8807) demonstrate low methylation levels (i.e. hypomethylation) in sperm as compared to (B) differential methylation in leukocytes. (C) and (D) Methylation profiles at Region 2 (chr8:141,110,441-141,110,641) and Region 3 (chr8:141108632-1411087555) illustrate a hypomethylated pattern.
5.2 Results

To determine if DNA methylation levels reflect allelic expression of *PEG13*, individual CpG sites were profiled at various regions within the CpG island overlapping *PEG13* (chr8: 141107838-141110984; GRCh37/hg19) and the *KCNK9* CGI (chr8:140,714,586-140,718,259; GRCh37/hg19) by bisulfite pyrosequencing. For pyrosequencing, methylation levels were measured as a ratio of cytosine:thymine residues (%) at each CpG site and were defined as a range where 0:100 to 30:70 was considered lowly methylated (hypomethylation), 70-30 to 100:0 was highly methylated (hypermethylation) and a ratio between 30:70 to 60:40 was indicative of a DMR. The regions selected spanned the entire *PEG13* CGI for an accurate representation of methylation patterns.

5.2.1 Identification of the *PEG13*-DMR

Methylation was assessed at four regions within the *PEG13*-CGI for human leukocyte DNA (n=10). Figure 5.2 illustrates methylation patterns at single CpG sites for each region. Regions 1 (chr8:141,107,853-141,108,138) comprised seven CpG sites and region 2 (141,107,731-141,108,007) had five CpG sites. Both regions were highly methylated and consistent with a hypermethylated pattern, where region 1 and region 2 had mean methylation of 93.0 ± 4.14% and 90.0 ± 5.55%, respectively. Region 3 (chr8:141108632-1411087555) encompassed seven CpG sites with a mean methylation level of 40.75 ± 8.34%, indicative of a DMR. Similarly, region 4 (chr8:141,108,731-141,108,8807) spanned seven CpG sites and had an average methylation level of 39.69 ± 3.96%. Fibroblasts (n=10) shared similar methylation patterns as leukocytes (Figure 5.3). Region 1 (141,107,853-141,108,138) in fibroblasts was also hypermethylated (89.49 ± 5.54%), and was similar to leukocytes. Region 2 (chr8:141108632-1411087555) had a mean methylation level of 38.06 ± 4.53% and region 3 (chr8:141,108,731-141,108,8807) had an average methylation level of 36.40 ±6.66%). We identified the *PEG13*-DMR in human leukocytes and fibroblasts, and supports imprinting of *PEG13* in fibroblasts.
5.2.2 *PEG13*-CpG island is a maternal DMR

Methylation patterns at three different regions within the *PEG13*-CGI (chr8: 141107838-141110984; GRCh37/hg19) were investigated using human placental and fetal brain DNA by combined bisulfite restriction analysis (COBRA) (Figure 5.5; Appendix 6). Region 1 (chr8:141107809-141108250) and region 3 (chr8:141109944-141110307) exhibited hypermethylated and hypomethylated patterns, respectively. Region 2 (chr8:141108200-141108523) was found to be differentially methylated, however the parental origin could not be determined due to the absence of informative SNPs within this region. Intriguingly, region 1 showed predominant maternal methylation in placental DNA, suggesting that region 2 acquires methylation on the maternal allele.

Accordingly, methylation patterns were assessed at three regions in sperm DNA (Figure 5.4). As expected, the average methylation at each region was lowly methylated and indicative of hypomethylation. Region 1 (chr8:141,108,731-141,108,880) encompassed five CpG sites within the *PEG13*-DMR, had a mean methylation of 1.80% ± 1.8% in sperm as compared to 42.2 ± 1.8% in whole blood. The hypomethylated profile observed in sperm suggests that the *PEG13*-DMR is a maternally-derived DMR.
Figure 5.5. Schematic diagram of the human TRAPPC9/PEG13 locus and identification of the PEG13 DMR in human fetal brain and placenta (Figure not drawn to scale). One of the blue rectangular boxes represents the PEG13 transcript (AK307073) and lies within intron 17 of TRAPPC9. The black arrows indicate the transcriptional direction and the black rectangular boxes depict the centromere (cent) and telomere (tel). Methylation at the PEG13-associated CpG island (green rectangular box) was assessed at three regions. Each circle denotes a single CpG site, where methylated and unmethylated are black and white circles, respectively. Allelic methylation at region 1 (chr8:141107809-141108250) was evaluated using placental DNA at rs3802217, where G was the paternal allele and A was the maternal allele, which was methylated. The overall methylation pattern at region 1 indicates hypermethylation. Region 2 (chr8:141108200-141108523) and region 3 (chr8:141109944-141110307) were assessed using fetal brain DNA. Region 2 exhibited differential methylation, indicating the PEG13-DMR. Conversely, region 3 showed methylation levels reflecting hypomethylation.
5.2.3 KCNK9-associated CpG island is hypomethylated

To support our hypothesis that PEG13 influences KCNK9 expression, we measured DNA methylation levels at the KCNK9 CGI (chr8: 140714586-140718259) (Figure 5.6). The region analyzed (chr8:140,714,585-140,714,615) contained 5 CpG sites and had low levels of methylation in both leukocytes (9.3 ± 3.39%) and fibroblasts (6.65 ± 4.38%), similar to sperm. Methylation was assessed at four CpG sites just outside the promoter region (chr8:140,714,444-140,714,512) as a comparison, where the overall methylation levels varied (26.2± 4.98%). The overall hypomethylated pattern at the KCNK9 CGI in leukocytes and fibroblasts indicates that a mechanism other than promoter methylation regulates KCNK9 imprinting.

5.2.4 Methylation in ID and autism probands

Methylation patterns at the PEG13-CCGI were analyzed using DNA from leukocytes of ID and autism probands to determine if the CNVs proximal or distal to the PEG13-DMR affect methylation. Similar methylation patterns at the PEG13-DMR and CGI of KCNK9 were observed in all probands as compared to leukocyte DNA from healthy controls (Figure 5.7). Region 1 (chr8: 141,110,441-141,110,641) encompassed seven CpG sites. For the deletion probands, a mean methylation level of 37.9 ± 2.54% and 38.1 ± 2.12% was observed for Proband 3 and 8 respectively. Proband 20, who has a paternally-inherited duplication, had a mean methylation level of 32.6 ± 2.51 %. All probands had low methylation levels at the KCNK9 CGI similar to control leukocytes, where Proband 3, 8 and 20 had mean methylation levels of 6.40 ± 1.52%, 8.40 ± 2.70% and 10.4 ± 3.78%. Taken together, the CNVs do not appear to affect methylation at the PEG13-DMR or KCNK9 CGI, suggesting that either the disorder mechanism is tissue specific (i.e. not in leukocytes), or the CNVs may exert their pathological effect by alternative method.
Figure 5.6 Plots of methylation at the *PEG13*-CGI and *KCNK9* CGI in human leukocytes and fibroblasts. Box and whisker plots show median and interquartile ranges at single CpG sites for 10 healthy individuals. The error bars indicate the S.D. across samples. (A) Methylation profiles at individual CpG sites outside the *KCNK9* CpG island (chr8:140,714,444-140,714,512) were highly variable and differed significantly (p < 0.5). (B) and (C) Methylation at region 1 (chr8:140,714,585-140,714,615) in leukocytes and fibroblasts were lowly methylated, indicative of hypomethylation.
Figure 5.7 Methylation at the PEG13 and KCNK9 CpG islands derived from bisulfite pyrosequencing in Autism and ID proband leukocytes. (A) Methylation profiles at Region 1 were not different between probands and controls, (C) Methylation profiles at region 2 the KCNK9 CpG island was lowly methylated in all probands, similar to control leukocytes. Circles represent the methylation at a single CpG site. Black circles = proband 3; Red circles = Proband 8; White circles = Proband 20
Chapter 6 Summary and conclusions, general discussion and future directions

The data presented here highlights the association between genomic imprinting and neurodevelopment by providing evidence for a novel imprinted gene, *PEG13*, showing preferential expression towards the paternal allele in human fetal brain. Further hypotheses require investigation in brain samples from both normal, healthy individuals and our deletion probands in conjunction with functional *in vitro* experiments to a) determine the specific role of *PEG13* on *KCNK9* imprinting in the brain; and to b) elucidate any neurophysiological implications for ID and ASD.

6.1 General Discussion

*PEG13* is a paternally expressed ncRNA in the brain

We identified an intronic long ncRNA running antisense to the intellectual disability gene *TRAPPC9* and illustrate partial conservation of the imprinting that was identified for mouse *Peg13* at the human syntenic region (chr8q24.3). Human *PEG13* was found to be preferentially expressed from the paternal allele like its mouse equivalent (Wang et al., 2008; Davies et al., 2004; Smith et al., 2003), specifically in fetal brain, as well as fibroblasts and kidney. Interestingly, *PEG13* is biallelically expressed in placenta. There is currently no data indicating the evolutionary loss of imprinting for *PEG13* in the placenta before species divergence between mice and humans. Moreover, whether knockout of *Peg13* or deregulated imprinting confers embryonic lethality or abnormal behavioural and/or growth phenotypes in mice requires investigation. The fact that *PEG13* is biallelically expressed in placenta supports the possibility that *PEG13*, along with its brain-specific expression, may have arisen in a lineage specific to ethuerian mammals, or was recently acquired in mice and maintained in humans. Indeed, unlike most imprinted long ncRNAs, and dissimilar to mouse *Peg13*, human *PEG13* lies within a non-imprinted gene (i.e. *TRAPPC9*), an arrangement analogous to a microimprinted domain. Microimprinted domains are hypothesized to have eutherian-specific
lineages acquired through late retrotransposition events. For example, neuronatin \((NNAT)\), a paternally-expressed gene, resides within the single intron of \(BLCAP\) on chr20q11.2 and is highly expressed in the CNS from mid-gestation to early postnatal life (Evans et al., 2001). A subsequent study revealed that imprinting of \(NNAT\) is restricted to placental mammals, with no traceable lineage in monotremes or marsupials (Evans et al., 2005). As such, phylogenetic analysis across species for \(PEG13\) may help to address the evolutionary impact of this gene and its role in the brain.

Nonetheless, various lines of evidence support the importance of \(PEG13\)’s imprinted expression in brain function and development. First, \(PEG13\) remains imprinted from fetal (Law, 2014; this thesis) to adult life (Court et al., 2014). Secondly, \(PEG13\) is the only gene along with \(KCNK9\) who’s imprinting status remains conserved from mice. The fact that only seven out of ten fetal brain samples displayed preferential expression may be attributed to the heterogeneity of the tissue samples. Lastly, \(PEG13\) has a similar expression profiles as the mouse gene: we showed that \(PEG13\) is highly expressed in the brain (Davies et al., 2004), as well as in adrenal gland, and has minimal expression in kidney, heart and placenta (Smith et al., 2003). Davies et al., (2004) found that mouse Peg13, similar to other paternally-expressed genes in the brain, has a neuron-specific localization in olfactory bulbs, cerebral cortex, hippocampus and thalamus. \(In situ\) hybridization of \(PEG13\) in human brain tissues (or other techniques such as gene expression microarray, SAGE (serial analysis of gene expression) or RNA-seq) are required to determine if this brain-specific expression pattern is conserved. Furthermore, we identified a DMR in the CGI associated with \(PEG13\) in all cells and tissues examined, including fetal brain. Analysis of the \(PEG13\)-DMR showed that the maternal allele was methylated in extra-embryonic tissue, and hypomethylated in sperm, indicative of a germ line-derived maternal DMR.

Previous studies have identified imprinted long ncRNAs in mouse brain, although few have been characterized in humans. A recent survey identified 34 imprinted ncRNAs in mouse brain which
corresponded to the Allen Brain Atlas (ABA) (Nikaido et al., 2003), and included Air (Sleutals et al., 2005), Copg2as (Lee et al., 2000), Gtl2 (Wylie et al., 2000), Rian (Cavaillé et al., 2002), and Mirg (Gregg et al., 2010). Subsequent analysis of ABA genes in mouse brain illustrated that long ncRNAs have regional expression patterns with enrichment in the hippocampus, cerebellum, olfactory bulb and cerebral cortex, in addition to subcellular compartmentalization (Mercer et al., 2008). Some cis-antisense and intronic transcripts show expression patterns corresponding to their protein-coding genes of neurological relevance (Mercer et al., 2008), indicating that unlike previous suppositions; some long ncRNAs are not merely genomic artefacts. A proportion of long ncRNAs been characterized as imprinted genes in mouse brain and impart regulation in brain development and function (Davies et al., 2004). The ncRNA Air is paternally expressed in glial cells and regulates maternal expression of Igf2r, and loss of Air imprinting in neurons results in biallelic expression of Igf2r (Yamasaki et al., 2008). Mercer et al., (2008) also identified a novel antisense transcript, AK045070, which shows parallel expression patterns to the nuclear receptor, Coup-Tfll, in mouse brain. Another imprinted long ncRNA, Kcnq1ot1, which shows region-specific expression in tissues, may also confer silencing in specific brain regions in addition to the placenta (Korostowski et al., 2012; Robins et al., 2012). Based on these findings, further analysis is required to determine how PEG13 plays a role in KCNK9 imprinting and if it acts similar to other antisense transcripts in the brain.

Accordingly, a recent study by Court et al., (2014) confirmed that PEG13 is maternally imprinted in human adult frontal cortex, and demonstrated that the mechanism by which PEG13 regulates KCNK9 expression parallels the reciprocal expression of H19 and Igf2 through higher-order chromatin remodeling (Kurukuti et al., 2003; Murrell et al., 2004) (Figure 8.1). Analogous to our findings, the study authors found that the imprinted expression of PEG13 was associated with methylation at the PEG13-DMR on the maternal allele. Furthermore, the non-methylated state of the PEG13-DMR on
the paternal allele was found to preferentially bind to CTCF-cohesin in the brain (Court et al., 2014). Most notably, the study identified a novel brain-specific enhancer element distal to PEG13 and located in the 17\textsuperscript{th} intron of TRAPPC9. Subsequent chromatin-conformation capture (3C) experiments revealed that in human cerebellum, the distal enhancer interacts with the KCNK9 promoter (which also has CTCF enrichment) and the PEG13-DMR through chromatin looping according to the parental allele (Court et al., 2014) (Figure 7.1). In other words, the enhancer interacts with PEG13-DMR depending on its methylated state: on the paternal allele, the unmethylated PEG13-DMR possesses enhancer-blocker activity (or acts as an insulator) as it can bind to CTCF-cohesin to induce higher order chromatin folding, thereby preventing the enhancer from interacting with the promoter of KCNK9. Moreover, enrichment of the epigenetic marks H3Kme1, H3K27ac and p300 in at the PEG13-DMR was restricted to the brain and not found in leukocytes (Court et al., 2014), thus supporting our conclusion that imprinting of PEG13 is primarily restricted to the brain. The higher-order chromatin silencing model has been previously described in Saccharomyces cerevisiae where chromatin looping has been observed at the processed promoter (5’end) and terminator (3’end) regions of genes transcribed by RNAP II and is required for transcriptional activation (O’Sullivan et al., 2004). As indicated above, chromatin looping also separates promoters and enhancers in the Igf2-H19 imprinting cluster according to the methylation status of the IC (Murrell et al., 2004). This enables Igf2 to reside within open chromatin for gene expression or within repressive chromatin depending on the allele, where binding of CTCF to the Igf2/H19 IC facilitates chromatin packaging on the maternal allele for silencing of Igf2 (Kurukuti et al., 2006). The Kcnqlot1 long ncRNA has also has a role in the imprinted silencing of Kcnql by affecting chromatin conformation, which in turn, influences interactions between a local enhancer and the Kcnql promoter (Korostowski et al., 2012).
We found no significant differences in methylation levels between our probands and normal individuals at the PEG13-DMR or KCNK9 CGI. Our results here do not show any association between the CNVs and disorder aetiology in our probands. However, we cannot disregard the possibility for a mechanism in the imprinting context. First, the epigenetic changes may be far more robust in tissues that are actually affected by PEG13 and KCNK9 activity, namely the brain. Secondly, other epigenetic mechanisms, such as chromatin conformation and histone modifications may constitute the primary epigenetic marks regulating PEG13-KCNK9 imprinting. Downstream effects of deletions/duplications around PEG13 may manifest by modulations to chromatin architecture. Indeed, the juxtaposition of euchromatic and heterochromatic domains by nucleoproteins and epigenetic marks define chromosomal regions and dictate the positioning of regulatory elements and the accessibility of interacting proteins (Kleinjan & van Heyningen, 1998). Structural rearrangements, translocations, inversions, deletions, duplications, can influence gene expression through “position effects” by altering the genomic position or chromatin environment in which a gene or regulatory element reside (Feuk et al., 2006). Position effect refers to the stochastic spreading of heterochromatin across a euchromatic region (which assumes the conformation of heterochromatin) due to chromosomal rearrangements, resulting in the inactivation of expressed genes (Kleinjan & van Heyningen, 1998). This phenomenon was originally observed in Drosophila for eye pigmentation and referred to as position effect variegation (PEV) (Wallrath & Reuter, 1995). Genes encoding chromatin-modifying proteins, either as enhancers or silencers of gene expression including Su (var) proteins, were identified via direct mutagenesis and associated to genes silenced by PEV (Fodor et al., 2010; Tschiersch et al., 1994). Furthermore, yeast also exhibit gene transcriptional silencing by position effect (Gottcschling et al., 1990) and is a natural occurrence in
Figure 6.1 Proposed model for PEG13-mediated imprinting of KCNK9 in the brain through higher order chromatin looping (based on diagram from Court et al., 2014) (Not drawn to scale). A brain-specific enhancer is located in intron 17 of TRAPPC9 and has enrichment for H3K4me1, H3K27ac and p300. (A) On the paternal allele, the unmethylated PEG13-DMR binds to a CTCF-cohesin complex to facilitate chromatin looping, and blocks the enhancer from interacting with the promoter of KCNK9, resulting in silencing of KCNK9. (B) Conversely on the maternal allele, the methylated PEG13-DMR cannot interact with the enhancer element, allowing the enhaser to associate with the KCNK9 promoter region to facilitate KCNK9 transcription on the maternal allele. The specific role of the long ncRNA PEG13 in this mechanism is currently unknown. Grey circles = unmethylated and black circles = methylated; Green rectangular boxes = CGI
mammals (Capel et al., 1993; Koopman et al., 1991). Chromatin-mediated inactivation has been connected to human diseases such as aniridia (Fant et al., 1994), Potocki-Shaffer syndrome (Feuk et al., 2006; Wakui et al., 2008), and α-Thalassemia (Barbour et al., 2000). Position effect has also been inferred in developmental disorders where mutations to gene coding sequences have not be identified in affected individuals (Lee et al., 2006; Velagaleti et al., 2005). Indeed, breakpoints of chromosomal rearrangements are often distal to coding sequences and open reading frames, which are free of mutations but concomitant to gene downregulation through various mechanisms. Position effects manifest by altering chromatin structure around a regulatory element, dissociating transcriptional units and regulatory elements, and/or may perturb methylation at gene promoters (Barbour et al., 2000; Kleinjan & van Heyningen, 1998). Indeed, insulator elements help demarcate chromatin boundaries through chromatin looping and higher order chromatin structure (Labrador & Corces, 2002). It is therefore plausible that CNVs may affect how chromatin is organized within and surrounding the PEG13 locus. Intriguingly, and in support of this hypothesis, the chromatin-remodelling complex CHRAC1 resides proximal and upstream to TRAPPC9. CHRAC1, which comprises five subunits including the ATPase ISWI and topoisomerase II, increases the accessibility to chromatin by facilitating nucleosome assembly and disassembly (Varga-Weisz et al., 1997). Whether the genomic position of CHRAC1 is merely coincidental and if its expression is required for cis-regulation of chr8q24.3 remains unexplored.

PEG13: Potential implications in ID and ASD etiology

Given that PEG13 and KCNK9 are not highly expressed in leukocytes, we could not identify any changes in gene expression that might result from deletions proximal or distal to the PEG13-DMR. Additionally, we could not determine if loss of imprinting at PEG13 contributes to the neurological profiles observed in our ID patients. Based on the assumption that PEG13 regulates imprinting at KCNK9, it is possible to expect some phenotypic similarities between our affected ID and autistic
individuals and those with Birk-Barel Syndrome. Barel et al., (2007) described an imprinting syndrome caused by a missense mutation replacing glycine at position 236 by arginine (G236R) in exon 2 on the maternal copy of KCNK9. The affected individuals displayed mild to severe ID, generalized hypotonia in early life which progressed to muscle weakness in major muscle groups, as well as unique facial dysmorphisms including elongated faces. Subsequent whole cell electrophysiology in oocytes found that the K2p9.1 (G236R) channel produced no measurable currents. Moreover, oocytes expressing the wild-type and mutant channel in a heterologous system elicited a decrease in K+ currents as compared to wild-type cells; the same phenotype was observed in oocytes co-expressing the mutant channel along with TASK-1, a channel which under normal physiological conditions heterodimerizes with TASK-3 in the human and mouse CNS (Barel et al., 2007). Interestingly, the study authors did not report any behavioural abnormalities other than ID in the affected individuals, and since this study, no other disorders associated with KCNK9 have been reported. Intriguingly, one of our study subjects (proband 20) has ID, ataxia, as well as characteristic dysmorphisms (Appendix 4). While these comparisons are not conclusive, the overlap of certain clinical features suggests that it would be worthwhile to obtain more clinical information from our probands for more in-depth analyses.

Role of KCNK9/TASK-3 in the brain

PEG13 expression was mostly restricted to the brain with a tissue-specific pattern similar to other identified brain-specific ncRNAs (Qureshi et al., 2010). Moreover, PEG13’s high expression in the cerebellum suggests region-specific and sub-compartamental localization in the brain. The fact that KCNK9 and PEG13 exhibit overlapping expression profiles suggests some functional importance, specifically with regards to how PEG13 regulates KCNK9 expression. Moreover, both genes are highly expressed in the cerebellum, which influences both motor and non-motor functions, and
ncRNAs have subcellular localization in cerebellar Purkinje cells (i.e. diffused in the soma, nuclei, or foci) (Mercer et al., 2008).

The KCNK9 gene encodes the K_2p9.1 potassium channel (also known as TASK-3) from the two pore domain (K_2p) subfamily. K_2p channels are among the largest class of K^+ channels with over 50 related genes identified in Caenorhabditis elegans (Bargmann, 1998); comparatively in mammals, 14 genes have been identified, of which 11 are functional and 8 are highly expressed in the CNS (Talley et al., 2003). Structurally, K_2p channels are characterized by their four membrane spanning domains and two pore-forming domains (2P/4TM) which dimerize to form a conducting K^+ pore (Lesage & Lazdunski, 2000). These background or “leaky” channels are both voltage and time-independent (Duprat et al., 1997), producing basal outward rectifying potassium currents that not only maintain resting membrane potential (RMP), but also modulate cell excitability by affecting the duration, frequency and amplitude of action potentials during hyperpolarization and depolarization (Talley et al., 2001). This is particularly important for neuronal excitability, heart rate, muscle contractions, as well as hormone secretion. Furthermore, K_2p channels are responsive to a variety of other chemical and physical agents, including volatile anesthetics (Lesage, 2003; Veale et al., 2007), oxygen (Mulkey et al., 2007), neurotransmitters, mechanical stretch, lipids, and G-protein coupled receptors (Duprat et al., 2007). However, despite their structural similarities, K_2p channels share relatively low sequence similarity and consequently, exhibit diverse cellular functions and expression patterns in mammals.

Cloning of the K_2p channels in heterologous systems has enabled for their classification into five subfamilies based on their electrophysiological and pharmacological properties: the acid-sensitive K^+ (TASK), the weakly inward rectifying K^+ channel (TWIK), the TWIK-related K^+ channel (TREK), the TWIK-related arachidonic acid stimulated K^+ channel (TRAALK) and tandem-pore domain halothane-
inhibited K+ (THIK) (review by Goldstein et al., 2001; 2005). The mammalian TASK family comprises TASK-1 (Duprat, 1997), TASK-2 (Medhurst et al., 2001; Lesage et al., 2000), TASK-3 (Kim et al., 2000) and the non-functional TASK-5 (Ashmole et al., 2001; Duprat et al., 1997), which are all expressed in rodents and humans. Only TASK-1 and TASK-3 share sequence homology (Rajan et al., 2002), and as their nomenclature implies with the exception for TASK-5, respond to slight variations in extracellular pH (Lesage & Lazdunski, 2000; Rajan et al., 2000; Kim et al., 2000). K$_{2p}$9.1 activity may also be modulated by local volatile anesthetics such as halothane and isoflurane (Czirják & Enyedi, 2003), and serve as a therapeutic target.

Previous expression studies illustrate that TASK channels are widely expressed in mammals and exhibit different tissue expression profiles in the CNS and peripheral tissues (Chapman et al., 2000; Medhurst et al., 2001); our current work is in agreement with these reports, specifically for TASK-3, and poses some interesting questions with regards to the functional implications of PEG13 and its affect on KCNK9 expression. Indeed, TASK-3 is prominently expressed in the human brain where it is restricted to the cerebellum, and has little to no expression in other somatic tissues (Chapman et al., 2000; Medhurst et al., 2001). Interestingly, TASK-3 expression is more regionally distributed in the rodent CNS, showing moderate to high expression in select hypothalamic nuclei, the amygdala, the neocortex and spinal cord, yet its marked expression in the cerebellum, specifically the granule cell layer, remains conserved (Karschin et al., 2001; Talley et al., 2001; Vega-Saenz et al., 2001). TASK-3 expression overlaps with some other KCNK transcripts, particularly with TASK-1, in both neuronal and non-neuronal tissues (Karschin et al., 2001; Talley et al., 2000). The complementary expression patterns in the brain between TASK-1 and TASK-3 are not surprising given that these channels share 54% sequence homology and have been shown to heterodimerize in these regions (Aller et al., 20005; Millar et al., 2000; Talley et al., 2003).
At the physiological level, knockout TASK-3 in cortical granular neurons fire at higher frequencies during depolarization, as expected (Brickley et al., 2007), indicating a role in controlling neuronal output and excitability. Accordingly, loss of TASK-3 expression is associated with impaired memory and altered cognitive function (Linden et al., 2007). Therefore, dysregulation of PEG13 expression may lead to a similar neurophysiological phenotype.

Hydroxymethylation: another epigenetic signature in the brain

Bisulfite sequencing cannot distinguish between 5-methylcytosine (5mC) and 5-hydroxymethylcytosine (5hmC) (Hahn et al., 2013). Consequently, 5mC may not account for all the methylated marks we observed at the PEG13-DMR, namely in brain tissues. Previous studies showed that the mammalian brain is enriched in 5hmC marks (Guo et al., 2011; Szulwach et al., 2011) and accounts for 0.20% of cytosine bases in fetal cortex and 0.87% in adult cortex (Lister et al., 2013). However, although 5hmC does not represent a significant fraction of cytosine methylation, it is widely hypothesized to serve as a functional intermediate in regulating gene expression as the majority of 5hmC marks are found at active genomic sequences i.e. promoters, intergenic regions and gene bodies (Jin et al., 2010; Robertson et al., 2011). Moreover, genes required for neuronal function such as ion channels and scaffold proteins show enrichment for 5hmC marks (Robertson et al., 2011). Recent whole-genome studies illustrate dynamic changes to the brain methylome that occur in development and with aging, and recognize 5hmC as a key regulatory mark (Guo et al., 2014; Lister et al., 2013; Xie et al., 2012). Lister and colleagues (2013) found that while fetal genomes are enriched in 5mC marks, early postnatal genomes were marked by the accrualment of conserved non-CG methylation, including 5hmC (Szulwach et al., 2011); this trend, which continues into adolescence, coincides with the primary period of neurogenesis (Lister et al., 2013). Furthermore, the increase of 5hmC marks from fetal life to adulthood is restricted to neuronal cells and not glial cells (Hahn et al., 2013; Lister et al., 2013). 5hmC profiles are also region-specific, thus adding another
layer of gene regulation. For instance, 5hmC marks in human cerebellum were shown to significantly increase from fetal to adult life, and the change was correlated to changes in gene expression (Wang et al., 2012). Presently, the role of 5hmC in the context of imprinting has yet to be addressed and the purpose of 5hmC versus 5mC methylation is not well-understood, although certain studies postulate that 5hmC may be a stronger silencing mark than 5mC (Robertson et al., 2011). Therefore, measuring hmC marks at the PEG13 DMR, as well as in relation to KCNK9, may provide more insight into the function of this conserved imprinted locus in neurodevelopment and adulthood.

6.2 Conclusions
This study identified a novel imprinted gene PEG13 in human fetal brain and provides evidence that the long ncRNA has a role in the imprinting of KCNK9. The tissue- and temporal-specific imprinting of mouse Trappc9 is not conserved in humans. The role of the heterozygous deletions/duplications in ID and ASD etiology require further investigation, however based on our current knowledge, a disease mechanism including PEG13 cannot be ruled out. There is the possibility that these structural variations may affect chromatin architecture surrounding the PEG13 locus, though future studies will need to address this hypothesis.

6.3 Future Directions
Studies using induced pluripotent stem cell (iPS) generated from our ID and autistic probands and differentiated to neurons will enable us to properly study the effects of the heterozygous deletions in the brain given that the expression of PEG13 and KCNK9 are restricted to the CNS. Moreover, other epigenetic mechanisms, such as chromatin remodeling and 5hmC marks will need to be studied at the PEG13 locus by chromatin immunoprecipitation (ChIP) assays and protein assays such as western blotting in neurons. Additionally, if deletions/duplications affect KCNK9 expression, electrophysiology studies measuring K+ currents will provide more insight into the neurophysiologic effects in ID and ASD etiology for our probands. In fact, electrophysiological studies, as well as
facial dysmorphology, and neurological testing would be helpful in order to compare the clinical presentation of these CNV individuals to reported cases of Birk-Barel syndrome. Knockdown of the long ncRNA by small interfering RNAs (siRNAs) or transfection of $PEG13$ in cells may also help determine the role of $PEG13$ in $KCNK9$ imprinting.
References


Arima, T. et al., (2005). ZAC, LIT1 (KCQ1OT1) and p57KIP2 (CDKN1C) are in an imprinted gene network that may play a role in Beckwith–Wiedemann syndrome. Nucleic acids research, 33(8), 2650-2660.


Francks, C. et al., (2007). LRRTM1 on chromosome 2p12 is a maternally suppressed gene that is associated paternally with handedness and schizophrenia. *Molecular psychiatry, 12*(12), 1129-1139.


Kaufman, L. et al. (2011, October). Evidence of imprinting at the autosomal recessive intellectual disability gene, TRAPPC9, and a possible role for CNVs at this locus in autism. Poster presented at the meeting of The American Society of Human Genetics, Montreal, Canada.


Morita, S. et al. (2007). One argonaute family member, Eif2c2 (Ago2), is essential for development and appears not to be involved in DNA methylation. Genomics. 89(6), 687-696.


Figure A.1 Pyrosequencing traces analyzed to validate allelic expression and methylation. The x-axis shows the order of nucleotide dispensation and the y-axis measures the amount of nucleotide added for each reaction well (the amount of light produced). The ascending slope of a peak is determined by DNA polymerase and sulfurylase activity, while the peak height depends on the oxidative capacity of luciferase, and the rate of nucleotide degradation by apyrase determines the descending slope (Rongahi et al., 2010). (A) Pyrogram quantifying allelic expression at a SNP (B) Pyrogram measuring DNA methylation.
Table A.2 List of studies which have identified TRAPPC9 as a cause of NS-ARID. All studies with the exception of the Koifman et al., (2011) study used homozygosity mapping to determine segregation of disease-causing mutations in TRAPPC9. N/A = Not available; N/I = Not identified; ** This mutations was identified in three families (Jamra et al., 2011, Mochida et al., 2009, Mir et al., 2009)
A.3 Expression analysis of *Trappc9* in mice. Allelic expression of *Trappc9* was assessed in B6 males crossed with JF1 females, and JF1 males crossed with B6 females to account for strain variation. *Trappc9* showed preferential or exclusive expression from the maternal allele in fetal brain, neuroprogenitor cells and neural cells. Allelic imbalance disappeared at 8 weeks postanatal. These findings were confirmed by pyrosequencing.
<table>
<thead>
<tr>
<th>Sample</th>
<th>Description</th>
<th>SNP#1</th>
<th>SNP#2</th>
</tr>
</thead>
<tbody>
<tr>
<td>BJ15.5_whole brain</td>
<td>Brain tissue from 15.5 dpc</td>
<td>C/T (B6/JF1)</td>
<td>T/C (B6/JF1)</td>
</tr>
<tr>
<td>JB15.5_whole brain</td>
<td>Brain tissue from 15.5 dpc</td>
<td>C/T (B6/JF1)</td>
<td>T/C (B6/JF1)</td>
</tr>
<tr>
<td>BJ P0♀_brain1</td>
<td>Brain tissue from postnatal day 0 female</td>
<td></td>
<td></td>
</tr>
<tr>
<td>JB P0♀_brain1</td>
<td>Brain tissue from postnatal day 0 female</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BJ P0♀_brain2</td>
<td>Brain tissue from postnatal day 0 female</td>
<td></td>
<td></td>
</tr>
<tr>
<td>JB P0♀_brain2</td>
<td>Brain tissue from postnatal day 0 female</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BJ P0♀_brain3</td>
<td>Brain tissue from postnatal day 0 female</td>
<td></td>
<td></td>
</tr>
<tr>
<td>JB P0♀_brain3</td>
<td>Brain tissue from postnatal day 0 female</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BJ P0♀_brain5</td>
<td>Brain tissue from postnatal day 0 female</td>
<td></td>
<td></td>
</tr>
<tr>
<td>JB P0♀_brain5</td>
<td>Brain tissue from postnatal day 0 female</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BJ 24W♀_brain1</td>
<td>Brain tissue from 24 weeks female</td>
<td></td>
<td></td>
</tr>
<tr>
<td>JB 24W♀_brain1</td>
<td>Brain tissue from 24 weeks female</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BJ 24W♀_brain3</td>
<td>Brain tissue from 24 weeks female</td>
<td></td>
<td></td>
</tr>
<tr>
<td>JB 24W♀_brain3</td>
<td>Brain tissue from 24 weeks female</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BJ 24W♀_brain5</td>
<td>Brain tissue from 24 weeks female</td>
<td></td>
<td></td>
</tr>
<tr>
<td>JB 24W♀_brain5</td>
<td>Brain tissue from 24 weeks female</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BJ 8W♂_brain1</td>
<td>Brain tissue from 8 weeks male</td>
<td></td>
<td></td>
</tr>
<tr>
<td>JB 8W♂_brain1</td>
<td>Brain tissue from 8 weeks male</td>
<td></td>
<td></td>
</tr>
<tr>
<td>JB 8W♂_brain1</td>
<td>Brain tissue from 8 weeks male</td>
<td></td>
<td></td>
</tr>
<tr>
<td>JB 8W♂_brain1</td>
<td>Brain tissue from 8 weeks male</td>
<td></td>
<td></td>
</tr>
<tr>
<td>JB 5W♂_brain1</td>
<td>Brain tissue from 5 weeks male</td>
<td></td>
<td></td>
</tr>
<tr>
<td>JB 5W♂_brain1</td>
<td>Brain tissue from 5 weeks male</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SNP#1:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SNP#2:</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Allelic expression patterns of Trappc9 measured by pyrosequencing

Summary

<table>
<thead>
<tr>
<th>Sample # (re-numbered)</th>
<th>Sample name</th>
<th>Paternal(%)</th>
<th>Maternal(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 BJ fibroblast</td>
<td>44.3</td>
<td>56.2</td>
<td></td>
</tr>
<tr>
<td>2 BJ fibroblast</td>
<td>48.1</td>
<td>51.9</td>
<td></td>
</tr>
<tr>
<td>3 BJ_glia2</td>
<td>42.4</td>
<td>57.6</td>
<td></td>
</tr>
<tr>
<td>4 BJ_glia1</td>
<td>47.3</td>
<td>52.7</td>
<td></td>
</tr>
<tr>
<td>5 BJ_glia1</td>
<td>44.5</td>
<td>55.5</td>
<td></td>
</tr>
<tr>
<td>6 BJ_NPC6</td>
<td>27.2</td>
<td>72.8</td>
<td></td>
</tr>
<tr>
<td>7 BJ NPC6</td>
<td>26.4</td>
<td>71.6</td>
<td></td>
</tr>
<tr>
<td>8 BJ neuronal1</td>
<td>19.2</td>
<td>80.8</td>
<td></td>
</tr>
<tr>
<td>9 BJ neuronal2</td>
<td>28.4</td>
<td>71.6</td>
<td></td>
</tr>
<tr>
<td>10 BJ neuronal2</td>
<td>29.9</td>
<td>70.1</td>
<td></td>
</tr>
<tr>
<td>11 BJ neuronal1</td>
<td>31.2</td>
<td>68.9</td>
<td></td>
</tr>
<tr>
<td>12 BJ 2AWS_cerebrum</td>
<td>28.9</td>
<td>72.0</td>
<td></td>
</tr>
<tr>
<td>13 BJ 2AWS_cerebrum</td>
<td>26.5</td>
<td>73.4</td>
<td></td>
</tr>
<tr>
<td>14 BJ P6_cerebellum</td>
<td>17.3</td>
<td>82.7</td>
<td></td>
</tr>
<tr>
<td>15 BJ P6_cerebellum</td>
<td>21.5</td>
<td>78.5</td>
<td></td>
</tr>
<tr>
<td>16 BJ P6_cerebellum</td>
<td>25.3</td>
<td>74.7</td>
<td></td>
</tr>
<tr>
<td>17 BJ P6_cerebellum</td>
<td>26.8</td>
<td>73.2</td>
<td></td>
</tr>
<tr>
<td>18 BJ15.5_whole brain</td>
<td>22.9</td>
<td>77.1</td>
<td></td>
</tr>
<tr>
<td>19 BJ15.5_whole brain</td>
<td>20.4</td>
<td>79.6</td>
<td></td>
</tr>
</tbody>
</table>

Mouse_Trappc9_Pyrosequencing_100514
A.4. Allelic expression analysis for **PEG13**. Electropherograms showing allelic expression of **PEG13** in fetal brain, kidney, skin and placenta using SNP rs4289794. Among the ten individuals who were informative, seven showed preferential expression at **PEG13**. Three samples showed biallelic expression for **PEG13** in fetal brain.
3 brain samples showed biallelic expression

E2 Brain cDNA
E4 Brain cDNA
E18 Brain cDNA
Reverse

7 brain samples showed monoallelic expression (newly extracted RNA)

F69 Brain cDNA
Stock RNA
Fresh RNA
Reverse

F69 Kidney
Fresh RNA
NA

F69 Liver
Stock RNA

F69 Genomic
F69 Mother

F77 Brain cDNA
F77 Skin cDNA
F77 Genomic
F77 Mother

F73 Brain cDNA
F73 Genomic
F73 Mother

PEG13 imprinting analysis
A.5. Relative expression of alleles at *PEG13* in ten fibroblasts samples at SNPs rs2270409. G = guanine; A = adenine. All individuals showed preferential or exclusive expression from an allele; however the parent-of-origin could not be determined because we did not have parental DNA.
A.6. Identification of the PEG13-DMR. Three regions of the CpG island investigated for allelic methylation are shown as region 1 (chr8:141107809-141108250), region 2 (chr8:141108200-141108523), and region 3 (chr8:141109944-141110307). Allelic methylation of region 1 was assessed in placental and cord DNA using rs3802217. Methylation in cord blood was highly methylated, while methylation in placenta varied, but was overall highly methylated. Methylation profiles at region 2 in placenta, cord blood and fetal brain shows differential methylation, indicating the PEG13-DMR. Conversely, methylation at region 3 was lowly methylated in all tissues tested.