Identification of Novel Imprinted Domains in the Therian Lineage

by

Radhika Das

University Program in Genetics and Genomics
Duke University

Date:_______________________
Approved:

____
Randy L. Jirtle, Supervisor

____
Gregory E. Crawford, Chair

____
Susan K. Murphy

____
Paul M. Magwene

____
John M. Greally

Dissertation submitted in partial fulfillment of
the requirements for the degree of Doctor
of Philosophy in the University Program in Genetics and
Genomics in the Graduate School of Duke University

2010

ABSTRACT

Identification of Novel Imprinted Domains in the Therian Lineage

by

Radhika Das

University Program in Genetics and Genomics
Duke University

Date:_______________________
Approved:

____
Randy L. Jirtle, Supervisor

____
Gregory E. Crawford, Chair

____
Susan K. Murphy

____
Paul M. Magwene

____
John M. Greally

An abstract of a dissertation submitted in partial fulfillment of
the requirements for the degree of Doctor
of Philosophy in the University Program in Genetics and
Genomics in the Graduate School of Duke University

2010

Copyright by
Radhika Das
2010

Abstract

Genomic imprinting is the parent-of-origin dependent monoallelic expression of select developmentally important genes that are regulated by epigenetic mechanisms. It is believed to have co-evolved with placentation in the Therian lineage, but it is unclear whether this phenomenon arose in a convergent or divergent manner in the Metatherians (those with a rudimentary placenta) and Eutherians (true placental mammals). Moreover, the epigenetic mechanisms involved in establishing genomic imprinting (DNA methylation or histone modifications) are still poorly defined. To address this question, I studied Metatherian orthologues of Eutherian imprinted loci using the gray short-tailed opossum *Monodelphis domestica* as a model organism. Study of six genes known to be imprinted in Eutherian mammals showed variable imprinted status in *M. domestica*. *L3MBTL* and *HTR2A* were monoallelically expressed; *MEST* had one imprinted and one non-imprinted transcript, while *IMPACT, COPG2* and *PLAGL1* were not imprinted, thus revealing that this phenomenon is conserved at some, but not all loci between Eutherian and Metatherian mammals. Moreover, differential methylation patterns and the associated regulatory non-coding RNAs that characterize imprinted genes in Eutherians are not conserved in Metatherian mammals, exemplified by a novel DMR identified within *IGF2R* which, unlike *Igf2r* in mice, has no associated anti-sense transcript. In contrast, histone modifications, specifically promoter activating H3 Lysine 4 dimethylation, are important in both lineages. Although the evidence does not resolve the question of whether or not imprinted genes evolved through convergent or divergent

evolution, it raises the intriguing possibility that both forms of evolution occurred during establishment of imprinting in these mammals.

The imposition of functional haploidy in the genome by such epigenetic mechanisms necessarily makes imprinted genes more susceptible to deleterious mutations and regulatory perturbations. Thus, imprinting is implicated in a number of developmental disorders, but its role in the etiology of complex human diseases and neurological disorders remains to be determined. I chose to investigate the imprint status of the duplicated locus *DGCR6/DGCR6L* lying within the 22q11.2 microdeletion causative of DiGeorge Syndrome (DGS), because our lab previously predicted genes at this genomic location to be imprinted. My studies revealed that both genes *DGCR6* and *DGCR6L* are monoallelically expressed in the primate lineage, but not in a parent-of-origin dependent manner. Interestingly, *DGCR6L* is not present in the mouse, and *Dgcr6* is expressed from both parental alleles.

Although DGS primarily manifests as facial, limb and heart abnormalities in children, a number of these patients also ultimately present with variable neurocognitive defects. *DGCR6* and *DGCR6L* potentially play a role in neural crest cell migration, thus I focused my studies on determining the effect of the microdeletion at this chromosomal region on the expression of these two genes and the associated neuropathological symptoms. Contrary to the expectation of reduced expression, the study revealed that DGS subjects have a highly dysregulated pattern of *DGCR6* and *DGCR6L* expression as compared to that in controls. Moreover, increased expression of these genes correlated significantly with decreased performance in sustained-attention tests. These data provide

the first evidence that disruption of the normal monoallelic expression pattern of *DGCR6* and *DGCR6L* by hemizygous deletion is involved in the variability in neurocognitive symptoms associated with DiGeorge Syndrome. These also highlight the importance of identifying novel epigenetically regulated and/or imprinted domains to better understand not only their evolution, but also how the inherent epigenetic lability of these regions contributes to modulating complex human diseases and neurological disorders.

Dedication

To my mother and father. Without whom, imprinting or otherwise, none of this was possible. And to JD, for keeping all his promises, but one.

R.D.

Contents

Abstract....................iv

Contents.................... viii

List of Tables....................xi

List of Figures.................... xii

List of Abbreviations.................... xiii

Acknowledgements....................xv

1. Introduction....................1

    1.1 Discovery of Genomic Imprinting....................2

    1.2 Theories for the Evolution of Genomic Imprinting....................6

    1.3 Evolution of Imprinting Mechanisms....................9

    1.4 Genomic Imprinting and Disease Susceptibility....................17

2. Comparative Phylogenetic Analysis of Genomic Imprinting in Marsupials and Eutherians....................24

    2.1 Introduction....................24

    2.2 Materials and Methods....................28

        2.2.1 Sample collection....................28

        2.2.2 Analysis of imprint status....................28

        2.2.3 Analysis of methylation status....................30

        2.2.4 Chromatin Immunoprecipitation Assay....................31

        2.2.5 Real-time PCR....................33

    2.3 Results....................34

        2.3.1 Analysis of imprint status....................34

2.3.2 Analysis of differential methylation 39

2.3.3 Analysis of histone modifications at the promoter region 44

2.4 Discussion 47

3. Investigation of Imprinting in *DGCR6/DGCR6L* within the Primate Lineage 52

3.1 Introduction 52

3.2 Methods and Materials 56

3.2.1 Investigation of expression status in humans 56

3.2.2 Expression analysis in chimpanzees 58

3.2.3 Generation of phylogenetic tree 59

3.3 Results 59

3.3.1 Expression analysis in humans 59

3.3.2 Expression analysis chimpanzee 61

3.3.3 Phylogenetic tree for *DGCR6/DGCR6L* 63

3.4 Discussion 63

4. Determining the Role of Imprinting and Epigenetic Dysregulation in DiGeorge Syndrome 70

4.1 Introduction 70

4.2 Materials and Methods 75

4.2.1 Sample collection 75

4.2.2 Parent-of-origin analysis of the deletion 76

4.2.3 Expression analysis 76

4.2.4 Correlation with neuropsychological data 77

4.2.5 Analysis of methylation status 78

4.3 Results 80
4.3.1 Parent-of-origin and expression analysis 80
4.3.2 Correlation between expression and neuropsychological symptoms 81
4.3.3 Analysis of methylation status 86
4.4 Discussion 89
5. Conclusions 95
Appendix 99
Bibliography 112
Biography 127

List of Tables

Table 1.1: Imprint status of Eutherian imprinted genes in Metatherians 4

Table 2.1: Genes analyzed for imprint status in *M. domestica*..38

Table 2.2: Methylation status of CpG-rich regions in *M. domestica* in the vicinity of orthologues of Eutherian imprinted loci 41

Table 2.3: Allele-specific enrichment of histone modification H3 Lysine 4 in promoter regions of orthologues of Eutherian imprinted loci in *M. domestica* 46

Table 4.1: Number of people exhibiting differential expression levels of *DGCR6* and *DGCR6L* amongst patients and controls 83

Table 4.2: Summary of correlations between *DGCR6* and *DGCR6L* gene expression and neuropsychological symptoms in DGS patient 85

List of Figures

Figure 1.1: Evolution of genomic imprinting in vertebrates 7

Figure 1.2: Imprinting throughout development 11

Figure 1.3: Imprinting regulation of the murine *Igf2/H19* reciprocally imprinted domain (Chromatin Boundary Model) and *Igf2r* (Non-coding RNA Model)...............................13

Figure 2.1: Analysis of imprinting in Eutherian orthologues of Metatherian loci in the opossum *M. domestica* 36

Figure 2.2: Analysis of imprinting at the *MEST* locus in *M. domestica*...........................37

Figure 2.3: Methylation profile of CpG islands in the vicinity of orthologues of Eutherian imprinted loci in *M. domestica* 40

Figure 2.4: Methylation profile and anti-sense transcript analysis around the novel DMR identified in the *IGF2R* gene of *M. domestica* 43

Figure 2.5: Analysis of histone modifications at the promoter regions of orthologues of Eutherian imprinted loci in *M. domestica* 45

Figure 3.1: Duplication at the *DGCR6/DGCR6L* locus 54

Figure 3.2: Analysis of imprinting at the *DGCR6/DGCR6L* locus in human samples 60

Figure 3.3: Analysis of allelic expression profiles for *DGCR6* and *DGCR6L* in human peripheral leucocytes and in chimpanzee liver tissues 62

Figure 3.4: Phylogenetic tree for the *DGCR6/DGCR6L* locus...64

Figure 4.1: Genomic rearrangements in DiGeorge Syndrome and the *DGCR6/DGCR6L* locus...73

Figure 4.2: Gene Expression at the *DGCR6* and *DGCR6L* loci.......................................82

Figure 4.3: Effect of parent-of-origin of the deletion on expression of *DGCR6* and *DGCR6L* 84

Figure 4.4: Methylation analysis adjacent to *DGCR6* and *DGCR6L* 88

Figure 4.5: Hypothetical model for regulation at the *DGCR6* and *DGCR6L* loci.............93

List of Abbreviations

Air	Antisense *Igf2r* RNA
BLAST	Basic Local Alignment Search Tool
bp	Base pairs
cDNA	Complementary DNA
CDS	Coding DNA Sequence
ChIP	Chromatin Immunoprecipitation
CTCF	CCCTC Binding Factor
DGS	DiGeorge Syndrome
DGCR	DiGeorge Critical Region
DMR	Differentially Methylated Region
DNA	Deoxyribonucleic Acid
FISH	Fluorescent *in situ* Hybridization
HDAC	Histone deacytylase
ICR	Imprint Control Region
Igf2	Insulin-like growth factor 2
Igf2r	Insulin-like growth factor 2 receptor
IQ	Intelligence Quotient
IACUC	Institutional Animal Care and Use Committee
IRB	Institutional Review Board

Kb	Kilobase
LINE	Long interspersed element
LTR	Long terminal repeat
Mb	Megabase
MEG	Maternally expressed gene
Mest/Peg1	Mesodermally expressed transcript/Paternally expressed gene 1
M. domestica	*Monodelphis domestica*
mRNA	Messenger RNA
NCBI	National Center for Biotechnology Information
ng	Nanogram
PCR	Polymerase Chain Reaction
PEG	Paternally expressed gene
PIC	Protease inhibitor cocktail
qPCR	Quantitative PCR
RACE	Rapid amplification of cDNA ends
RNA	Ribonucleic Acid
RT	Reverse Transcription
SINE	Short interspersed element
SNP	Single nucleotide polymorphism
XCI	X chromosome inactivation
Xist	X inactivation specific transcript

Acknowledgements

I am indebted to Dr. Randy Jirtle for providing me a laboratory "home" and being my proverbial parent. I am also extremely grateful to all the other members of the Jirtle lab family, especially Jennifer, Kristin, Dale, Dana, David and Autumn for their constant care and help. A special thanks to Dr. Susan Murphy for her support and ideas at every stage of this work and to my committee members, Dr. Greg Crawford, Dr. Paul Magwene and Dr. John Greally, for their invaluable advice. I must also thank Dr. Vandana Shashi, without whom the last portion of this work would not have been possible.

Many others have contributed to each part of this thesis, and I must mention MaryEllen Koran, Zhiqing Huang, Carol Grenier, Nathan Anderson, Michael Kamal, Dr. Kerstin Linblad-Toh, Tarjei Mikkelsen, Dr. Anna Keyte, Arielle Linderman and Dr. Kathleen-Smith, for their contribution in the opossum study. Dr. Priya Kishnani, Kelly Schoch, Dr. David Swofford, Dr. Thomas Kwapil and Dr. Timothy Howard have been instrumental in completing the study on DiGeorge Syndrome. I am also indebted to Dr. Julie Horvath and Dr. Brian Chadwick for their experimental advice.

I would like to end by thanking everyone around me for their overwhelming encouragement and support. A special thanks to the UPGG program, my friends at Duke and in India, especially Nora, Sathya, Christina and Sabarni, who have helped me in every respect possible. Finally, I have to thank my parents, my sister, Munmundi, Prashantda and Joydeep for inspiring me to start and finish this pursuit.

1. Introduction

The field of epigenetics encompasses intracellular events that do not change the underlying DNA sequence, but result in heritable, yet reversible changes in gene expression (Barton, Surani et al. 1984). Several wide-ranging phenomena such as X-chromosome inactivation, position effect variegation in *Drosophila*, and genomic imprinting result from forms of epigenetic regulation. Most of these epigenetically-induced changes in gene expression are mediated by DNA methylation of cytosine at CpG dinucleotide sites, coupled with chemical modifications of histone tails within the chromatin (Kim, Samaranayake et al. 2009). These epigenetic marks can be set or erased both pre- and postnatally, thus an individual's 'epigenome' is malleable and responsive to environmental influences.

Genomic loci that are most vulnerable to such environmental epigenetic changes are imprinted genes and repetitive elements. Repetitive transposable elements are usually heavily methylated to prevent replication; however, some transposable elements have variable levels of CpG methylation. This can result in the formation of 'metastable epialleles', defined as gene alleles that are variably expressed in genetically identical individuals due to epigenetic modifications (Rakyan, Blewitt et al. 2002). The Agouti gene in the viable yellow agouti (A^{vy}) mouse is one example of a metastable epiallele and is frequently used as a biosensor for environmental factors that affect the epigenome (Jirtle and Skinner 2007). Whether genes with metastable epialleles also exist in humans is presently unknown (Morgan and Whitelaw 2009).

The expression of imprinted genes is also epigenetically regulated. They are expressed from only one parental allele due to gamete-specific epigenetic marks established during the early development of the egg and sperm. Genomic imprinting plays a vital role in embryonic growth, and these genes often exhibit tissue or developmental stage-specific monoallelic expression patterns. The imposition of a functional haploid state at imprinted loci greatly increases their prominence in disease formation since a single genetic mutation or lone epigenetic event can alter their function (Murphy and Jirtle 2003). These imprinted loci are involved in several syndromes involving uniparental disomy (UPD), e.g. Beckwith-Wiedemann, Prader-Willi and Angelman syndromes (Glenn, Porter et al. 1993; Kishino, Lalande et al. 1997; Murrell, Heeson et al. 2004), as well as in a variety of cancers (Feinberg 2007). Epigenetic dysregulation may likewise explain the variability and parental effects observed in the pathogenesis of conditions like autism and schizophrenia, but the role of epigenetics in these disorders still remains largely unexplored.

1.1 Discovery of Genomic Imprinting

Animal breeding three millennia ago showed that crossing a female horse with a male donkey gave a mule. Contrary to Mendelian predictions, the reciprocal cross yields a phenotypically different offspring called a hinny (Hunter 2007). These observations show that maternal and paternal genomes do not contribute equally to the progeny.

Mouse nuclear transplantation experiments in 1984 demonstrated that in contrast to the embryonic tissues formed from fusion of sperm and oocyte, tissues formed from two male pronuclei (androgenotes) were very different from those formed from two female pronuclei (gynogenotes) pointing out that the parental genomes are not functionally equivalent (Barton, Surani et al. 1984; Surani, Barton et al. 1984; McGrath and Solter 1986). The first imprinted genes identified were the maternally expressed *Igf2r* (Barlow, Stoger et al. 1991) and the paternally expressed *Igf2* in mice (DeChiara, Robertson et al. 1991). To date, approximately 100 genes have been experimentally confirmed to be imprinted in Eutherians, true placental mammals such as mice and humans (Morison, Ramsay et al. 2005) (http://www.geneimprint.com).

Six out of fourteen tested Eutherian imprinted genes are also imprinted in mammals with a rudimentary placenta, the Metatherians, including the tammar wallaby (*Macropus eugenii*), South American opossum (*Monodelphis domestica*), and the Virginia opossum (*Didelphis virginiana*) (**Table 1.1)** (Killian, Nolan et al. 2001; Weidman, Murphy et al. 2004; Suzuki, Renfree et al. 2005; Weidman, Maloney et al. 2006; Ager, Suzuki et al. 2007; Suzuki, Ono et al. 2007; Smits, Mungall et al. 2008). Interestingly, the orthologs of the six genes imprinted in Metatherians are either biallelically expressed or absent in Prototherians (egg-laying mammals including the platypus and echidna) and Aves (chicken) (Killian, Nolan et al. 2001; Nolan, Killian et al. 2001; Renfree, Ager et al. 2008).

Table 1.1: Imprint status of Eutherian imprinted genes in Metatherians [1]

	Eutherians		Metatherians		
Gene	***Homo sapiens***	***Mus musculus***	***Monodelphis domestica***	***Didelphis virginiana***	***Macropus eugenii***
CDKN1C	Maternal	Maternal	Unknown	Unknown	Biallelic
DIO3	Paternal	Paternal	Unknown	Biallelic	Biallelic
DLK1	Paternal	Paternal	Biallelic	Unknown	Unknown
H19	Maternal	Maternal	Maternal	Unknown	Maternal
IGF2	Paternal	Paternal	Paternal	Paternal	Paternal
IGF2R	Biallelic	Maternal	Maternal	Maternal	Unknown
INS	Paternal	Paternal	Unknown	Unknown	Paternal
MEG3	Maternal	Maternal	No Homolog	Unknown	Unknown
MEST	Paternal	Paternal	Unknown	Unknown	Paternal
NNAT	Paternal	Paternal	No Homolog	No Homolog	No Homolog
PEG10	Paternal	Paternal	Unknown	Unknown	Paternal
SGCE	Paternal	Paternal	Unknown	Unknown	Biallelic
SNRPN	Paternal	Paternal	Unknown	Unknown	Biallelic
UBE3A	Maternal	Maternal	Unknown	Unknown	Biallelic

Note: In this table, paternally expressed genes are listed as "paternal" and maternally expressed genes are listed as "maternal"

Thus, genomic imprinting in vertebrates is restricted to the Metatherian and Eutherian mammals; however, this unique gene regulatory system also evolved independently in some insects and flowering plants (Khosla, Mendiratta et al. 2006; Scott and Spielman 2006).

Most efforts to identify imprinted genes have been experimental, focusing on genes that potentially regulate growth in the mouse embryo (Peters and Beechey 2004). We recently used machine-learning algorithms, utilizing DNA sequence characteristics, to predict imprinted genes at a genome-wide level in both the mouse (Luedi, Hartemink et al. 2005) and human (Luedi, Dietrich et al. 2007). Of the approximately 20,000 autosomal genes annotated in the human genome, we identified 156 (0.75%) novel candidate imprinted genes; significantly fewer than the 600 predicted in the mouse. The 30 percent overlap in the repertoires of mouse and human imprinted genes also emphasizes a marked species difference in genomic imprinting (Luedi, Dietrich et al. 2007). Interestingly, many of the known and candidate human imprinted genes map to chromosomal locations previously linked to complex diseases, indicating that genomic imprinting may play a mechanistic role in their pathogenesis.

1.2 Theories for the Evolution of Genomic Imprinting

In mammals, imprinting originated approximately 150 million years ago during the Jurassic period in the least common ancestor of the two Therian infra-classes, Metatherians and Eutherians **(Figure 1.1)** (Killian, Byrd et al. 2000; Killian, Buckley et al. 2001; Nolan, Killian et al. 2001). The functional haploidy caused by imprinting enables a single genetic mutation or epigenetic modification to inactivate the function of an imprinted gene.

Given the increased disease risk associated with genomic imprinting, it is frequently asked whether or not there is any selective advantage for this unique form of gene regulation. Several theories propose explanations for the adaptive advantage of imprinting evolution in placental mammals. It is postulated that imprinting evolved as a defense against parthenogenesis, and the associated developmental disorders that would emanate from the resulting increased homozygosity of deleterious recessive mutations (Kono 2006). Alternatively, imprinting may have evolved to save the oocytes from spontaneously transforming into lethal teratomas – the ovarian time bomb hypothesis (Weisstein, Feldman et al. 2002). Yet others believe that it serves as a mechanism to modulate complementary expression profiles of maternally expressed genes (MEGs) and paternally expressed genes (PEGs) from different parental chromosomes in order to regulate development – the complementation hypothesis (Kaneko-Ishino, Kohda et al. 2006).

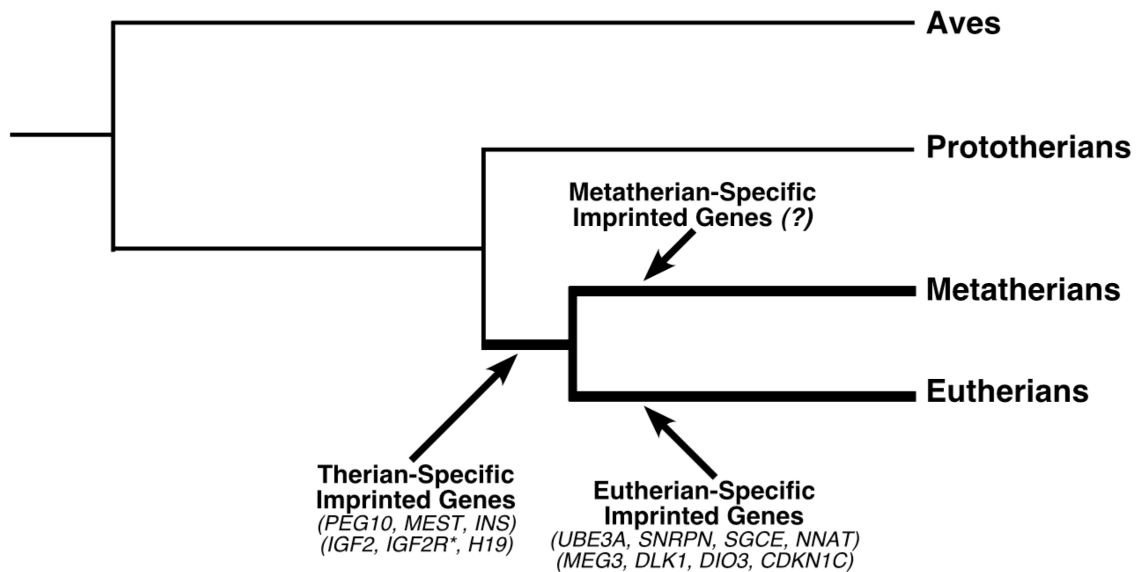


Figure 1.1: Evolution of genomic imprinting in vertebrates

The presence of imprinted genes in Metatherians and Eutherians (heavy lines), but not in Aves and Prototherians (light lines) points to an ancestral origin of imprinting 150–200 million years ago in the Therian mammalian lineage. There is evidence of Eutherian-specific genes but Metatherian-specific imprinted genes have thus far not been identified. *Recent imprint regulatory findings (Weidman, Dolinoy et al. 2006) indicate convergent evolution of *IGF2R* imprinting in Metatherians and Eutherians after lineage divergence rather than divergent evolution in Therians, as shown.

Another hypothesis proposes that imprinted genes were silenced as bystanders when the host genome methylated foreign DNA or retrotransposons in order to repress them – the host defense hypothesis. This has been suggested in the case of imprinted genes like *PEG10*, *Peg12* and *Rtl1*, which all seem to have originated from retrotransposition events. This theory, however, cannot explain why imprinting was limited to placental mammals.

The most accepted evolutionary model of imprinting is the parental conflict hypothesis (Moore and Haig 1991; Haig 2004). This hypothesis is based on the idea that genomic imprinting developed in response to viviparity and polygamy, and speculates that fitness effects during placental development were the principal factors that shaped its evolution. For the first time in evolutionary history, the placenta exists as an interface for both the paternal and maternal genomes to exert their influence on resource allocation within the intra-uterine environment. Moreover, postnatal nurturing behavior (such as lactation) further extends this interaction between the mother and offspring (Murphy, Wylie et al. 2001).

According to the parental conflict hypothesis, in the case of multiple paternities, genes expressed from a paternal allele favor increased maternal investment to enhance his offspring's own fitness at a cost to those fathered by other males. In contrast, genes expressed from the maternal allele maximize reproductive fitness of the mother ensuring availability of resources for all her current and future progeny. This posit successfully explains paternal expression of growth enhancers (e.g. *IGF2* and *KCNQ1OT1*) and

maternal expression of growth inhibitors (e.g. *IGF2R* and *GRB10*) (Haig 2004). Supporting evidence for this model also comes from the discovery of imprinted genes in true placental mammals and marsupials while none have been identified in the egg-laying monotremes or aves **(Table 1.1 and Figure 1.1)** (Killian, Byrd et al. 2000; O'Neill, Ingram et al. 2000; Killian, Nolan et al. 2001; Nolan, Killian et al. 2001). However, this hypothesis fails to clarify the role of imprinting at loci such as the *GNAS* cluster which is mainly involved in energy metabolism.

There is also growing evidence that imprinting may have evolved in response to a need for dosage compensation upon duplication of a gene (Walter and Paulsen 2003; Wood and Oakey 2006). Regional duplication and the danger of over-expression may have initially resulted in random monoallelic silencing of paralogues, utilizing epigenetic modifications such as DNA methylation. Subsequently, the advantage conferred by expression of only one parental allele may have led to gene repression in a germ-line specific manner. This could occur only in placental mammals that utilize epigenetic marking systems to establish parent-specific imprints in the developing gametes.

1.3 Evolution of Imprinting Mechanisms

Genomic imprinting has undoubtedly evolved independently at least three times; in insects (Khosla, Mendiratta et al. 2006), plants (Garnier, Laoueille-Duprat et al. 2008), and the placental mammals (Renfree, Hore et al. 2009). Although there are few conserved genomic characteristics between these three groups of imprinted organisms, the epigenetic mechanisms adopted to mediate parental allele-specific silencing have some

similar characteristics. In certain invertebrate coccids, genomic imprinting is manifested as heterochromatization of the entire paternal genome, selectively in males, and thus serves as the mode of sex-determination as well (Khosla, Mendiratta et al. 2006). In plants, genes such as *MEA* and *FWA* in *Arabidopsis* are imprinted selectively in the nutrient-providing endosperm, and the mechanism for allele-specific expression involves activation by demethylation (Scott and Spielman 2006). This is mediated by the maintenance methyltransferase MET1, and does not appear to require *de novo* methylases.

The situation in vertebrates is more complex. The imprint regulatory marks necessarily undergo a cycle involving their establishment in the primordial germ cells (PGCs) of one generation, maintenance during somatic cell divisions throughout life in the resulting individual, and erasure and re-establishment in the germ cells during embryogenesis to reflect the sex of the individual in which they reside **(Figure 1.2)** (Murphy and Jirtle 2003; Jirtle and Skinner 2007). DNA methylation is an efficient regulatory mechanism for this inherited marking system, since it can be modulated with the help of *de novo* methyltransferases *DNMT3A, DNMT3B* and *DNMT3L* in PGCs, and then maintained throughout life with the aid of *DNMT1* (Hermann, Gowher et al. 2004; Turek-Plewa and Jagodzinski 2005; Jia, Jurkowska et al. 2007).

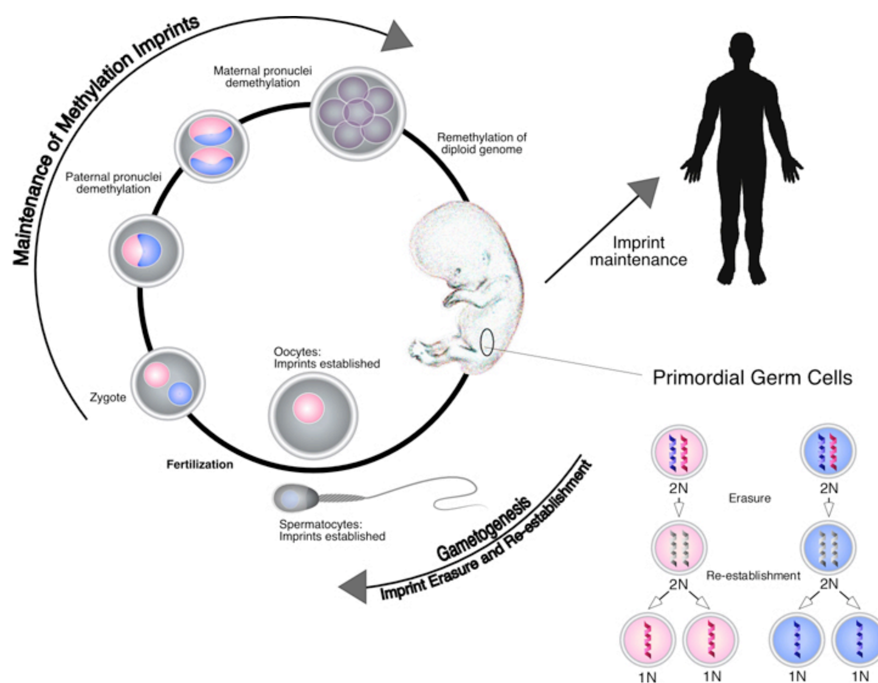


Figure 1.2: Imprinting throughout development

Methylation marks associated with imprinted genes are established at two distinct phases of development. During gametogenesis, the imprint marks present on the maternal (pink) and paternal (blue) chromosomes are erased (grey), followed by establishment of primary imprints which then reflect only the sex of the individual in which they reside. Just after fertilization, a global demethylation event occurs in the zygote, first in the paternal pronucleus (blue), followed by the maternal pronucleus (pink). Imprint marks that were established in the gametes are protected from this demethylation process. Remethylation of the diploid genome occurs around implantation, and includes setting of the secondary imprints, which are maintained throughout the lifespan of the individual.

Proteins like CTCF, coupled with its testis-specific counterpart CTCFL, non-coding RNAs, and methyl-CpG binding proteins (MBDs) which recruit HDACs, are all involved in imprinting regulation, emphasizing the complexity of the imprinting process (Reik and Walter 2001; Klenova, Morse et al. 2002).

A feature of imprinted genes in Eutherians is that they are commonly arranged in clusters (Wan and Bartolomei 2008). This indicates that shared regulatory elements are involved in the epigenetic control of this novel form of gene expression. Imprinting control regions (ICRs) are frequently identified in imprinted gene clusters as discrete *cis*-acting DNA elements that carry a heritable epigenetic mark which distinguishes the two parental alleles (Hoffman, Vu et al. 2000).

The primary imprint mark in Eutherians appears to be a germ-line derived methylation of the DNA at specific CpG sites, wherein these same regions in somatic cells are either methylated or unmethylated (Reik and Walter 2001). The methylation can be paternally derived from the sperm, as in the differentially methylated region (DMR) located upstream of *H19* in the *Igf2/H19* cluster **(Figure 1.3A)** or maternal in origin like DMR2 of murine *Igf2r* **(Figure 1.3B)** (Reik and Walter 2001). Surprisingly, among the six genes known to be imprinted in marsupials, DMRs have only been identified in *PEG10* (Suzuki, Ono et al. 2007) and *H19* (Lawton, Carone et al. 2008; Smits, Mungall et al. 2008).

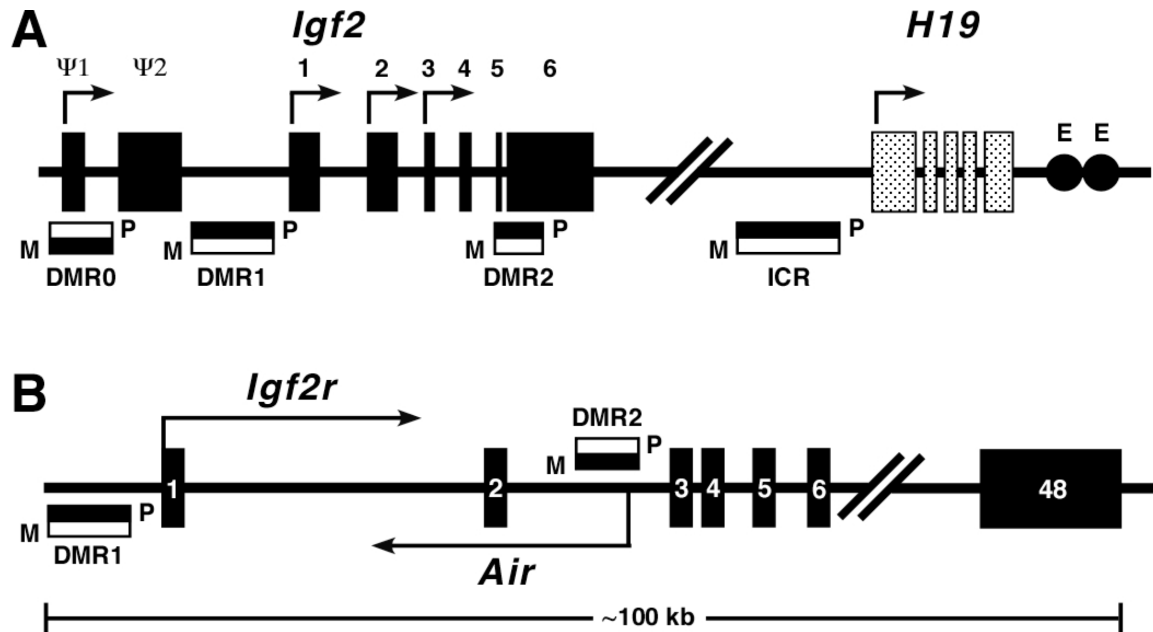


Figure 1.3: Imprinting regulation of the murine *Igf2/H19* reciprocally imprinted domain (Chromatin Boundary Model) and *Igf2r* (Non-coding RNA Model)

(A) The paternal expression of *Igf2* and the maternal expression of *H19* are regulated by allele-specific methylation at differentially methylated regions DMR0, DMR1, DMR2 and the imprint control region (ICR). (B) Maternal expression of *Igf2r* involves two differentially methylated regions, DMR1 and DMR2, and the paternal expression of the *Air* antisense transcript. The black and white boxes below the gene indicate genomic regions of preferential maternal (M) or paternal (P) CpG methylation; black identifies the parental allele that is methylated. Ψ1 and Ψ2 designate placenta specific exons. Significant expression from the Ψ1 promoter is also found in the kidney.

Even if DMRs are ultimately shown to be associated with all imprinted genes in Therian mammals, their presence is not necessarily sufficient to confer parental-dependent, monoallelic gene expression. Recent investigations into the chromatin state of imprint regulatory DMRs reveal that the parental alleles differ both in their hypersensitivity to DNaseI, as well as in covalent modifications of their histone tails (Shibata, Yoshino et al. 1996; Ooi and Henikoff 2007). Generally, repressive modifications such as methylation of lysines 9 and 27 of histone H3 are associated with the silenced allele, whereas activating marks like acetylation of lysines of histones H3 and H4 and methylation of histone H3 lysine 4 accompany the expressed allele (Reinhart and Chaillet 2005).

Human *IGF2R* is biallelically expressed (Killian, Nolan et al. 2001), despite having an intronic DMR2 (Smrzka, Fae et al. 1995). This expression pattern is better predicted by the lack of characteristic histone modifications in the promoter region of the gene, indicating that histone modifications are crucial for imprint regulation at this locus. Histone modifications have been postulated to be the 'primordial imprint mark,' with DNA methylation evolving later to stabilize imprinting (Reik and Lewis 2005). Given the apparent lack of DMRs in the Metatherian lineage, further investigation into the role of histones in imprint regulation in Metatherians is required to help shed light on the ancestral mechanism of imprinting regulation.

Irrespective of the identity of the inherited imprint marks, the parent-of-origin dependent expression of imprinted genes is mediated by at least two different imprinting

mechanisms. The “chromatin boundary model” holds that allele-specific modifications at the ICR affect binding of insulator proteins (e.g. Zinc-finger DNA binding protein CTCF), thereby mediating gene silencing. Such is the case for the *Igf2/H19* imprinted domain (**Figure 1.3A**) (Schoenherr, Levorse et al. 2003). CTCF binds to the maternal unmethylated ICR, shielding *Igf2* from downstream *Igf2/H19* shared enhancers. This results in maternal expression of *H19*. In contrast, CTCF is unable to bind to the `methylated ICR of the paternal allele, allowing enhancer interaction with the *Igf2* promoter and paternal *Igf2* expression. This *H19* promoter methylation also contributes to paternal *H19* silencing.

In the “non-coding RNA model”, production of antisense RNA is critical for the maintenance of imprinting. In the imprinted murine *Igf2r* (**Figure 1.3B**), production of antisense *Air* from DMR2 of the paternal allele after fertilization silences *Igf2r* as well as two other genes, *Slc22a2* and *Slc22a3*, by promoting paternal allele-specific methylation at DMR1 (Sleutels, Zwart et al. 2002). Methylated DMR2 in intron 2 of the maternal allele inhibits the production of *Air*. Thus, DMR1 remains unmethylated, permitting expression of the three genes only from the maternal allele. Interestingly, *Air* is not produced in humans despite the presence of DMR2, and *IGF2R* is biallelically expressed. Even more surprising is that *Air* has not been found in other mammals that show imprinted *IGF2R* expression, such as the dog (O'Sullivan, Murphy et al. 2007) and opossum (Weidman, Dolinoy et al. 2006).

The incomplete conservation of imprinting mechanisms between Eutherians and Metatherians has sparked a debate regarding the timing of the evolution of this phenomenon. Previous results, as well as recent findings in our laboratory, indicate that only a subset of genes imprinted in Eutherians (e.g. mouse and human) are concomitantly imprinted in Metatherians (i.e. opossum and tammar wallaby) (Table **1.1**). *UBE3A* and *SNRPN,* which are in an imprinted gene cluster on human chromosome 15q11.2, lie on different chromosomes in the tammar wallaby, and neither gene is imprinted in this marsupial (Rapkins, Hore et al. 2006). *PEG10* and *SGCE* (human 7q21.3) are imprinted in mice and humans*,* and although both genes are present in the tammar wallaby, only *PEG10* is imprinted (Suzuki, Ono et al. 2007). Interestingly, the DMR of *PEG10* in the tammar wallaby is restricted to its own promoter; however, the maternal allele-specific methylation ultimately spread over the course of evolution to the promoter of the adjacent *SGCE* gene in the mouse and human, leading to Eutherian-specific paternal expression of this gene (Suzuki, Ono et al. 2007). *DLK1* and *MEG3* are present in an imprinted domain at human chromosomal location 14q32.2 (Wylie, Murphy et al. 2000). Of these two genes, only *DLK1* is present in the opossum, and it is not imprinted (Weidman, Maloney et al. 2006). Paternally expressed *NNAT* (human 20q11.23), like *MEG3*, is also only found in Eutherians (Evans, Weidman et al. 2005). Finally, *CDKN1C* and *IGF2* are syntenic in the Metatherian and Eutherian lineages, but only *IGF2* exhibits imprinted expression in marsupials (Ager, Pask et al. 2008). Thus, there are genes that are clearly

imprinted in Eutherians but not in Metatherians; whether the converse is true remains to be determined.

Of those genes that are imprinted in both Metatherian and Eutherian mammals an important question remains. Did imprinting of these orthologous genes evolve convergently after the two Therian lineages separated or did it occur in a common ancestor to Therian mammals (i.e. divergent evolution)? Further investigation into the existence of other imprinted genes and their modes of regulation will undoubtedly shed some light on this question. Also, since placentae have evolved a number of times in lizards and snakes (Thompson and Speake 2006), it is possible that genomic imprinting may have also co-evolved with placentation and viviparity in other vertebrates as well. This intriguing possibility remains to be tested.

1.4 Genomic Imprinting and Disease Susceptibility

Genomic imprinting may be evolutionarily adaptive because of its apparent involvement in brain development (Keverne 2001; Badcock and Crespi 2006; Leung, Vallero et al. 2009) and potential ability to accelerate speciation (Vrana, Guan et al. 1998; Hunter 2007). Nevertheless, the presence of functionally haploid imprinted genes in the human genome has serious consequences for individual health.

Developmental disorders in humans such as Angelman, Prader-Willi, and Beckwith-Wiedemann syndromes result not only from genomic mutations, but also from epigenetic deregulation at *cis*-acting DMRs (Glenn, Porter et al. 1993; Kishino, Lalande et al. 1997; Jiang, Bressler et al. 2004; Murrell, Heeson et al. 2004). Numerous linkage

studies of complex disorders such as autism, bipolar disorder, schizophrenia, Tourette's syndrome, and Silver-Russell syndrome also show a parent-of-origin inheritance preference, suggesting the involvement of imprinted genes in their etiology (Morison, Ramsay et al. 2005).

Silver-Russell syndrome, a congenital disease characterized by growth retardation, is the first human disorder shown to result from epigenetically-mediated imprinting defects affecting two different chromosomes (Eggermann, Eggermann et al. 2008). Ten percent of Silver Russell syndrome patients present with maternal UPD of chromosome 7, and approximately 40% show hypomethylation in the ICR upstream of *H19* at 11p15.5 **(Figure 1.3A**). Interestingly, hypermethylation of the same ICR is associated with the overgrowth disease, Beckwith-Wiedemann syndrome. Consequently, these two developmental syndromes are clinically and epigenetically opposite diseases.

We predicted and confirmed that *KCNK9,* which encodes for the potassium channel TASK3, is maternally expressed in humans (Luedi, Hartemink et al. 2005); it is likewise imprinted in mice (Ruf, Bahring et al. 2007). *KCNK9* is highly expressed in the brain, and maps to 8q24.3, a region previously shown to be linked to bipolar disorder and epilepsy (Kananura, Sander et al. 2002; McInnis, Lan et al. 2003). Maternal inheritance of a receptor inactivating mutation results in the developmental disorder, Birk-Barel mental retardation dysmorphism syndrome (Barel, Shalev et al. 2008). In contrast, genomic amplification and overexpression of *KCNK9* in somatic cells is associated with breast cancer (Mu, Chen et al. 2003); this gene is also overexpressed in colorectal cancers

(Kim, Cho et al. 2004) and glioblastomas (Meuth, Herrmann et al. 2008). The role of LOI at the *KCNK9* locus in the etiology of these cancers has not been investigated.

Myoclonus-dystonia or dystonia 11 is a movement disorder caused by mutations in the father's copy of the paternally expressed epsilon-sarcoglycan (*SGCE)* gene or by maternal UPD at the gene locus (Muller 2009). Psychiatric findings are also common in patients with dystonia 11, including depression, obsessive–compulsive behavior, panic attacks and attention deficit hyperactivity disorder (Kinugawa, Vidailhet et al. 2009).

Transient neonatal diabetes mellitus (TNDM) is a type of diabetes that presents with growth retardation and diabetes in the first week of life. Although there is early remission of the disease, these children tend to also develop diabetes in later life. TNDM is caused by three mechanisms that result in overexpression of the paternally expressed imprinted gene, *HYMA1* (Temple and Shield 2002). These include paternal UPD of chromosome 6, paternally inherited duplication of 6q24, and a methylation defect in a DMR upstream of *HYMA1*.

Abnormal epigenetic reprogramming of imprinted loci is also associated with failure to thrive, conditions found in cloned embryos and stem cells undergoing somatic cell nuclear transfer, as well as in humans born using assisted reproductive technologies (Arnaud and Feil 2005). More recently, imprinting deregulation has been postulated to be involved in the etiology of mental disorders such as autism and schizophrenia because of the potential role imprinted genes play in the development of the human brain (Badcock and Crespi 2008).

According to this theory, a balance between the function of paternally expressed genes in the limbic system, and maternally expressed genes in the cerebral cortex is required for normal human brain development. A skewing of expression from the paternally expressed genes is suggested to promote the development of an 'extreme male brain', characterized by heightened mechanistic cognition and decreased social aptitude, as exemplified by autistic individuals. In contrast, a relative increase in maternally expressed genes is predicted to promote hyper-mentalism, involving increased social and psychological abilities, albeit ultimately at the cost of delusional behavior observed in schizophrenia. The imprinted brain theory suggests that autism is the antithesis of schizophrenia.

Exposure to famine conditions while *in utero* increases the risk of developing cardiovascular disease, obesity, and diabetes (McMillen and Robinson 2005). It also doubles the incidence of schizophrenia (Susser, Neugebauer et al. 1996; St Clair, Xu et al. 2005). Interestingly, Dutch famine victims were recently reported to have epigenetic changes in a number of imprinted gene loci decades after birth (Heijmans, Tobi et al. 2008; Tobi, Lumey et al. 2009). Thus, an important prediction of the imprinted brain theory is that a calorie rich diet could result in an increased prevalence of autism with a concomitant decreased incidence of schizophrenia (Badcock 2009) -- an intriguing situation now observed in Western countries (Der, Gupta et al. 1990; Woogh 2001; Blaxill 2004). A better understanding of the mechanisms by which the environment affects the genesis of neurological disorders will clearly necessitate elucidating the role of

imprinting in normal human brain development.

Because imprinted genes are so integrally involved in regulating early growth and development, mutation or their epigenetic deregulation in somatic cells also increases cancer risk (Feinberg 2005; Feinberg 2007). For imprinted tumor suppressor genes, only a single mutational or epigenetic event is required for complete inactivation because one allele is already nonfunctional due to imprinting. Consequently, the silenced allele of an imprinted gene has been equated to the *first hit*, as defined by Knudson (Knudson 2001) in his two-step model for carcinogenesis. Imprinted oncogenes can also be inappropriately overexpressed in somatic cells through loss of imprinting (LOI) in which the normally silent allele becomes transcriptionally active (Feinberg 2005; Feinberg, Ohlsson et al. 2006). Thus, imprinted genes are also potential human cancer susceptibility loci, such as the imprinted tumor suppressor gene *Igf2r*.

DiGeorge Syndrome is another complex disorder in which parent-of-origin effects have been observed, but direct evidence for the involvement of imprinting has not yet been demonstrated. This congenital disease affects one in 4000 live births, and leads to developmental defects (such as craniofacial abnormalities, limb and digit anomalies, velopharyngeal insufficiency and conotruncal heart defects) as well as psychiatric symptoms that develop later in life. One out of four DGS patients become schizophrenic, while one in six suffer from serious depressive disorders (Maynard, Haskell et al. 2002). The genetic cause has been attributed to a hemizygous deletion at chromosomal location 22q11.2, where rearrangements are mediated by faulty recombination between four low

copy repeats known as LCR22s. Most patients (approximately 85 percent) harbor a 3 Mb deletion, while a subset have a smaller 1.5-2 Mb deletion, but still show all the characteristics of the disease, thus delimiting this as the DiGeorge minimal Critical Region (DGCR). The parental origin of the deletion also plays a role in determining the phenotype since patients carrying a maternal deletion are known to exhibit poorer cognitive ability. Many of the symptoms involve defects that arise during embryonic growth, particularly in formation of the aortic arches (resulting in abnormalities of the heart and appendages) and in migration of neural crest cells (causative of the cognitive and psychiatric anomalies seen). Thus, investigations of genes functionally responsible for this disorder have focused on growth regulatory transcription factors and cell adhesion molecules.

Certain candidates identified by transgenic studies in mice seem to recapitulate the developmental defects observed in humans (for example the transcription factors *Hira* and *Tbx1*), but none of them has the potential to explain all the symptoms simultaneously, thus leading to its classification as a "contiguous gene syndrome." (Maynard, Haskell et al. 2002) Also, since the symptoms are of varying severity in the different afflicted individuals, the involvement of epigenetic factors has been suggested. Interestingly, patients who inherit the deletion from their mothers have more severe learning defects due to much reduced gray matter in the brain, as compared to their paternally-deleted counterparts (Eliez, Antonarakis et al. 2001); however, imprinting analysis of 25 mouse homologs of genes within the critical region did not identify any monoallelically

expressed genes (Maynard, Meechan et al. 2006). Nevertheless, it is possible that the gene(s) actually responsible for this disorder (imprinted or otherwise) have diverged between the two species, leading to human-specific symptoms upon their disruption. Of relevance here, the gene *DGCR6* lying within the 22q11 microdeletion responsible for the disorder, has recently been predicted to be imprinted by computational approaches undertaken in our laboratory (Luedi, Dietrich et al. 2007).

To further our understanding of the role and consequences of genomic imprinting in humans, it is crucial to elucidate mechanisms that first imposed this type of gene regulation. Hence, in this study I have focused on evaluating known Eutherian imprinted loci in the Metatherian lineage. After searching for additional imprinted genes in the South American opossum, *M. domestica*, DNA methylation and histone modifications were examined to elucidate their potential role in imprinting regulation in the marsupials. In this thesis, I provide evidence that histone modifications serve as the primordial imprint regulatory mark. In addition, imprint regulatory marks do not appear to be conserved between Metatherians and Eutherians at the loci investigated, indicating convergent evolution of these imprinted genes. I also investigated whether epigenetic regulation of *DGCR6* (and its duplicate *DGCR6L*) could explain the parent-of-origin inheritance of neurological symptoms seen in DiGeorge Syndrome. I show that dysregulated expression of both genes is causative of specific neurocognitive defects in the DGS patients. This work lays the foundation and justifies future investigation of epigenetics in many such complex disorders.

2. Comparative Phylogenetic Analysis of Genomic Imprinting in Marsupials and Eutherians

2.1 Introduction

Genomic imprinting is an epigenetic phenomenon characterized by the monoallelic expression of a gene in a parent-of-origin dependent manner (Morison, Ramsay et al. 2005). Its discovery in mammals showed that parental genomes are functionally non-equivalent, and that the sex of the parent, not of the resulting progeny, is critical in determining expression of a particular allele (Barton, Surani et al. 1984; Surani, Barton et al. 1984; McGrath and Solter 1986).

In vertebrates, imprinting is believed to have originated approximately 150 million years ago, in the common ancestor of the two Therian infra-classes: Eutherians (true placental mammals like humans and mice), and Metatherians (marsupials like the opossum and tammar wallaby) which have a more rudimentary placenta (Killian, Buckley et al. 2001). The evidence comes from approximately 83 imprinted genes that have been discovered to date in Eutherian mammals, and the six known to be imprinted in the marsupials (Killian, Nolan et al. 2001; Weidman, Murphy et al. 2004; Suzuki, Renfree et al. 2005; Weidman, Maloney et al. 2006; Ager, Suzuki et al. 2007; Suzuki, Ono et al. 2007; Edwards, Mungall et al. 2008). So far their Prototherian and avian homologous counterparts analyzed have been found to be biallelically expressed (Killian, Nolan et al. 2001; Killian, Nolan et al. 2001). To explain this unique phylogenetic distribution of imprinted genes, the “Conflict Hypothesis” proposes that genomic

imprinting evolved in placental mammals in response to polygamy, viviparity and multiple live births (Moore and Haig 1991; Haig 2004). This theory posits that the paternal genome attempts to maximize offspring extraction of nutrients from the mother whereas the maternal genome attempts to reduce this extraction, predicting that paternally and maternally expressed genes enhance and inhibit growth, respectively.

An unusual feature of imprinted genes in Eutherians is that they tend to be clustered in the genome. This suggests the role of shared regulatory elements in the epigenetic control of this phenomenon. In many imprinted gene clusters, "Imprint Control Regions (ICRs)" have been identified as discrete *cis*-acting DNA elements that carry a heritable epigenetic mark which distinguishes the two parental alleles (Hoffman, Vu et al. 2000). These simultaneously and often reciprocally regulate two or more imprinted genes, and can be present up to many hundred Kb away from the genes.

For Eutherians, the imprint mark is usually a germ-line derived cytosine methylation of the DNA at specific CG containing site(s), wherein the two alleles exhibit opposite methylation states (Reik and Walter 2001). The methylation mark can be established in the sperm, as observed for the differentially methylated ICR located upstream of *H19* in the *Igf2/H19* cluster, or in the oocyte, like the differentially methylated region (DMR) in the second intron of *Igf2r*. Surprisingly, of the six genes presently identified to be imprinted in marsupials, a DNA regulatory methylation mark is only present in *PEG10* (Suzuki, Ono et al. 2007), and *IGF2* (Lawton, Carone et al. 2008). Moreover, a DMR is not sufficient to confer imprinting in Eutherians, since *IGF2R* is

biallelically expressed in humans despite having a DMR in intron 2 (Smrzka, Fae et al. 1995).

Differential chromatin states of the parental alleles due to distinct covalent modifications of their histone tails also play a role in regulating expression and imprinting patterns. Repressive modifications like methylation of lysines 9 and 27 of histone H3 function in silencing of an allele whereas, activating marks such as methylation of H3 lysine 4 and acetylation of histone H3 and H4 lysines are associated with the expressed allele. These alternate histone modifications at the promoter region of imprinted genes correlate better with the expression status than even the methylation marks (Grandjean, O'Neill et al. 2001; Yang, Li et al. 2003; Vu, Li et al. 2004). Although the human *IGF2R* locus has an intronic DMR in peripheral tissues, it lacks the promoter-restricted histone code present in the mouse, and is biallelically expressed. Thus it is proposed (Reik and Lewis 2005) that a histone code may be the "primordial imprint mark," and DNA methylation developed later to assist in stabilizing the expression status.

Understanding the basic mechanisms by which imprinting is established in ancestral mammals is critical in furthering our knowledge about imprint gene regulation in humans. Although differential methylation of the two parental alleles is involved in controlling genomic imprinting in Eutherians, there is no clear evidence regarding its role in marsupials. The DNA methyltransferases necessary for conferring imprinting are present in the marsupials (Ding, Patel et al. 2003; Yokomine, Hata et al. 2006) yet differential methylation has only been identified in two of the six imprinted genes

presently identified. This may be due to lack of thorough investigation at and around the imprinted loci in marsupials because of incomplete sequence availability, or due to the actual lack of DMRs in these regions; imprinting being controlled by histone marks in these Therians. The sequencing of the genome of the South American opossum, *M. domestica* allows the investigation of such imprint regulatory regions in a more comprehensive manner.

Interestingly, it has not been determined if chromatin modifications exist at imprinted loci in marsupials even though histone modifications rather than DNA methylation are involved in paternal X chromosome silencing in marsupials (Kaslow and Migeon 1987; Loebel and Johnston 1996). Since imprinted X-inactivation appears to have co-evolved with genomic imprinting in marsupials, it is possible that the same molecular mechanism is used for marking the active and inactive regions in both of these epigenetically regulated phenomena (Lee 2003; Reik and Lewis 2005).

Consequently, in this study we identified additional imprinted loci in the gray short-tailed opossum, *M. domestica*. We distinguished CpG rich regions adjacent to the loci chosen for investigation, and searched for differential methylation. We also analyzed these regions for the activating H3 Lysine 4 dimethylation and the repressive H3 Lysine 9 trimethylation histone tail modifications. Even though differentially methylated regions have been discovered in some marsupial imprinted genes, our findings are consistent with histone modifications playing a significant role in regulating parent-of-origin dependent monoallelic expression in *M. domestica*. The incomplete conservation of Eutherian

imprint regulatory features in marsupials also suggests convergent evolution of these imprinted genes in the two Therian lineages.

2.2 Materials and Methods

2.2.1 Sample collection

M. domestica (gray short-tailed opossum) samples were collected from the laboratory of Dr. Kathleen Smith, Duke University, under protocols approved by the Institutional Animal Care and Use Committee (IACUC). Liver, brain and kidney tissues from 10 individuals and 10 embryos (belonging to three of the adult females) were dissected and stored at -80°C. DNA was extracted using the Qiagen genomic tip protocol 100/G (Qiagen Sciences, Valencia, CA). Total RNA was isolated by homogenization in RNA-Stat60 (Tel-Test, Gainsville, FL) followed by procedures recommended by the manufacturer.

2.2.2 Analysis of imprint status

Bioinformatic searches for orthologues of loci known to be imprinted in Eutherians using the Ensembl genome browser (http://www.ensembl.org/index.html) revealed six candidate imprinted genes, namely *PEG1*/*MEST*, *COPG2*, *HTR2A*, *L3MBTL*, *IMPACT* and *PLAGL1*. These loci were screened for polymorphisms in the adult animals and embryos obtained. Primers were designed to amplify parts of the exonic region of the longest annotated transcript of each gene. PCR was performed using 1.5U of Platinum Taq DNA polymerase (Invitrogen, Carlsband, CA), 15 pmol of primers, 1.5 mM $MgCl_2$ and 10 mM dNTPs in a 25 µl PCR reaction volume. The products were

purified by gel-extraction and elution in spin columns (Sigma-Aldrich, St. Louis, MO), and then sequenced on an ABI 3130 sequencer (Applied Biosystems, Carlsband, CA). For heterozygotes, the liver, brain and kidney RNA were treated with DNase I (Invitrogen, Carlsband, CA), converted to cDNA using Superscript II (Invitrogen, Carlsband, CA) in oligo-dT primed reactions, amplified (using primers detailed in supplementary table 2.1, and conditions as before) and sequenced.

In the case of *PEG1/MEST*, 3′ RACE was performed to identify transcripts (since two transcripts were reported earlier in the marsupial, the tammar wallaby) using the 5′ /3′ RACE kit (Roche Applied Science, Indianapolis, IN). Briefly, first strand synthesis was performed using an oligodT anchor primer, and the cDNA was further amplified using a forward primer placed in Exon 11 (5′ CAGTGAATCCTCATCCAGA 3′), since the differences were seen between exon 11 and the 3′ end in the wallaby. The products obtained were purified by gel extraction and elution using HiPure columns (Roche Applied Science, Indianapolis, IN). Following ligation into the T-easy vector (Promega, Madison, WI) colonies were transformed into JM109 competent cells. These were plated on LB-Agar-Xgal plates, and after overnight incubation at 37° C, white colonies were picked and amplified by whole cell PCR. T7 and SP6 primers were used for amplification with conditions as described above, and the inserts were sequenced to characterize the polymorphisms.

2.2.3 Analysis of methylation status

CpG-rich regions were identified in the vicinity of the six loci selected for investigation in *M. domestica* using custom-designed algorithms from Albert Einstein College of Medicine, New York (Glass, Thompson et al. 2007). Genomic DNA was modified using sodium bisulfite to selectively convert all unmethylated cytosines to uracils, using the Qiagen Epitect Bisulfite Kit (Qiagen Sciences, Inc., Valencia, CA). Fourteen CpG-rich regions in the proximity of the chosen genes were then analyzed for their methylation status. Standard and semi-nested PCR assays were designed for bisulfite treated DNA using Primer 3 (http://frodo.wi.mit.edu/). Converted DNAs were amplified using PCR reactions and products were purified using a QIAquick Gel Extraction Kit (Qiagen Sciences, Inc., Valencia, CA). Purified PCR products were cloned using TOPO TA Cloning Kits (Invitrogen, Carlsband, CA) and plasmid DNA was purified using Montage Plasmid MiniprepHTS 96 Kit (Millipore, Billerica, MA). Plasmid DNA was sequenced and the methylation status of the clones was visualized using CpGviewer (http://dna.leeds.ac.uk/cpgviewer/).

A differentially methylated region was found in *IGF2R* of *M. domestica*. To determine allele-specific differential methylation, an embryo heterozygous for an A/T SNP in this region with a homozygous mother was identified. The bisulfite treated DNA was amplified, ligated, transformed, amplified with T7 and SP6 primers and sequenced, as previously described. Anti-sense transcription was analyzed using strand-specific RT-

PCR followed by PCR; appropriate primers were positioned immediately upstream and downstream of the DMR.

2.2.4 Chromatin Immunoprecipitation Assay

Prior to performing chromatin immunoprecipitation (ChIP) reactions, chromatin was sheared using the Active Motif ChIP-IT Protocol (Active Motif, Carlsbad, CA), modified for the use of tissues instead of cells. Reagents provided by the manufacturer were used. Briefly, 50 mg of tissue was crushed using a pestle and mortar, which was submerged in liquid nitrogen. Then samples were immediately incubated for 10 minutes in 10 ml of Fixation solution (1% Formaldehyde in 1X PBS) to cross-link the DNA and associated proteins. The samples were then washed successively in Glycine-stop-fix solution (10X glycine buffer, 10X PBS and dH_2O) and Cell-scraping solution (10X PBS, dH_2O, and 100 mM PMSF). The samples were incubated on ice for 30 minutes in ice-cold lysis buffer (supplemented with 7.5 μl Protease Inhibitor Cocktail, PIC, and 7.5 μl PMSF), and then dounced gently ten times to aid nuclei release. Finally, the samples were resuspended in 330 μl shearing buffer and transferred to an ice bath in which sonication was performed. To shear the chromatin into 200 to 1000 bp fragments, nuclei were subjected to sonication five times at 35% amplitude in 10 second pulses, with a rest of 50 seconds between each pulse. Cellular debris was pelleted by centrifugation at 13,000 rpm for 12 minutes at 4°C, and the supernatant containing sheared DNA was removed and flash-frozen until further use. In order to check the shearing efficiency, 25 μl of the supernatant was incubated with 5M NaCl and RNase A for 4 hours at 65°C.

After further treatment with Proteinase K, phenol-chloroform extraction and ethanol precipitation, the sheared chromatin was visualized on a 2% agarose gel.

The Active Motif ChIP-IT Express Kit (Active Motif, Carlsbad, CA) was utilized for immunoprecipitation. ChIP grade antibodies specific for H3 Lysine 4 dimethylation and H3 Lysine 9 trimethylation, from Upstate Cell Signaling Solutions (Millipore, Billerica, MA) were used. Rabbit Serum Albumin was also used as a negative control, to judge the level of non-specific binding. Ten µl of sheared chromatin was put aside to serve as the control "input DNA". The remaining sheared chromatin (6.3µg) was incubated overnight at 4°C with Protein G magnetic beads, ChIP Buffer 1, PIC, dH_2O, and the antibody of interest. The beads were washed once with 800 µl ChIP Buffer 1, and twice with 800 µl ChIP Buffer 2. Finally, the beads were re-suspended in 50 µl Elution Buffer AM2 for 15 minutes. Reverse cross-link buffer (50 µl) was then added, and using a magnetic stand, the beads were separated from the immunoprecipitated DNA. Next, all samples including the input (that had been supplemented with 88 µl ChIP Buffer 2 and 2 µl 5M NaCl) were incubated at 94°C for 15 min. Proteins were degraded by incubating with 2 µl Proteinase K for one hour at 37°C, and the reaction was stopped by adding 2 µl Proteinase K Stop solution. Only the input DNA was further purified with RNase A, phenol-chloroform extraction and ethanol precipitation.

All samples were used as template in PCR reactions to determine loci enriched for each histone modification. At least three biological replicates were tested for each locus.

To further assess allele-specific enrichment, input and immunoprecipitated products were purified, ligated, transformed, amplified using T7 and SP6 primers and sequenced, as previously described. The ratio of the two alleles in the input versus the H3K4 dimethyl antibody immunoprecipitated samples was compared. Chi square tests (using Yates correction for small sample size) were performed to determine whether the input or the ChIP samples were enriched for a particular allele.

2.2.5 Real-time PCR

Relative quantification of the immunoprecipitated DNA using the different antibodies at the various loci was performed with the Applied Biosystems 7900HT Fast Real-Time PCR System (Applied Biosystems, Carlsband, CA). ChIP DNA (2 µl) was utilized in 20 µl reactions, conditions being set as per manufacturer's instructions (Active Motif, Carlsbad, CA). Primers were designed to amplify 150-300 bp products, and were checked before use for efficiency and amplification specificity under standard PCR reactions. SybrGreen Assays were performed to judge the relative enrichment of the different histone modifications relative to the input DNA. The cycle threshold (Ct) value for sample input and immunoprecipitated DNA for each experiment was normalized by subtracting the Ct value for control Albumin input and immunoprecipitated DNA, respectively, for each experiment. The enrichment was calculated by subtracting the immunoprecipitated sample's normalized Ct from the normalized input Ct.

2.3 Results

2.3.1 Analysis of imprint status

Availability of the genome sequence for *M. domestica* facilitated the bioinformatic search for orthologues of loci known to be imprinted in Eutherians, using the Ensembl genome browser. We focused on six candidate genes because of their high level of genomic conservation with mouse and humans, namely *PEG1/MEST* (Kobayashi, Kohda et al. 1997; Suzuki, Renfree et al. 2005), *COPG2* (Blagitko, Schulz et al. 1999), *HTR2A* (De Luca, Likhodi et al. 2007), *L3MBTL* (Li, Bench et al. 2004), *IMPACT* (Okamura, Hagiwara-Takeuchi et al. 2000; Okamura and Ito 2006) and *PLAGL1* (Valleley, Cordery et al. 2007). These loci were screened for polymorphisms in the DNA, and subsequently in the cDNA of the informative individuals. The polymorphism is visibly present in the cDNA if the gene is biallelically expressed, whereas visualization of only one allele demonstrates monoallelic expression. Expression was analyzed in three tissues; liver, brain and kidney, representing the three germ layers. Parental origin of the expressed allele was determined by analyzing heterozygous embryos belonging to mothers homozygous at the corresponding loci.

L3MBTL and *HTR2A* were expressed from only the paternal and maternal allele, respectively. In contrast, *IMPACT, COPG2* and *PLAGL1* showed biallelic (non-imprinted) expression in *M. domestica* (**Table 2.1, Figure 2.1, Supplementary Table 2.1**). Interestingly, a SNP in exon 1 of *PEG1/MEST* was present in the cDNA, while a SNP in the 3' UTR was not **(Figure 2.2)**. We hypothesized that this could be due to the

presence of one imprinted and one non-imprinted transcript (or from the presence of an anti-sense transcript originating in this region). Two *PEG1*/*MEST* transcripts were recently discovered in the Eutherians and the marsupial tammar wallaby (Suzuki, Renfree et al. 2005). 3' RACE (Rapid Amplification of cDNA ends) revealed two transcripts, one of which was monoallelicaly expressed. Moreover, in the three genes determined to be imprinted in *M. domestica*, monoallelic expression was observed in all tissues examined.

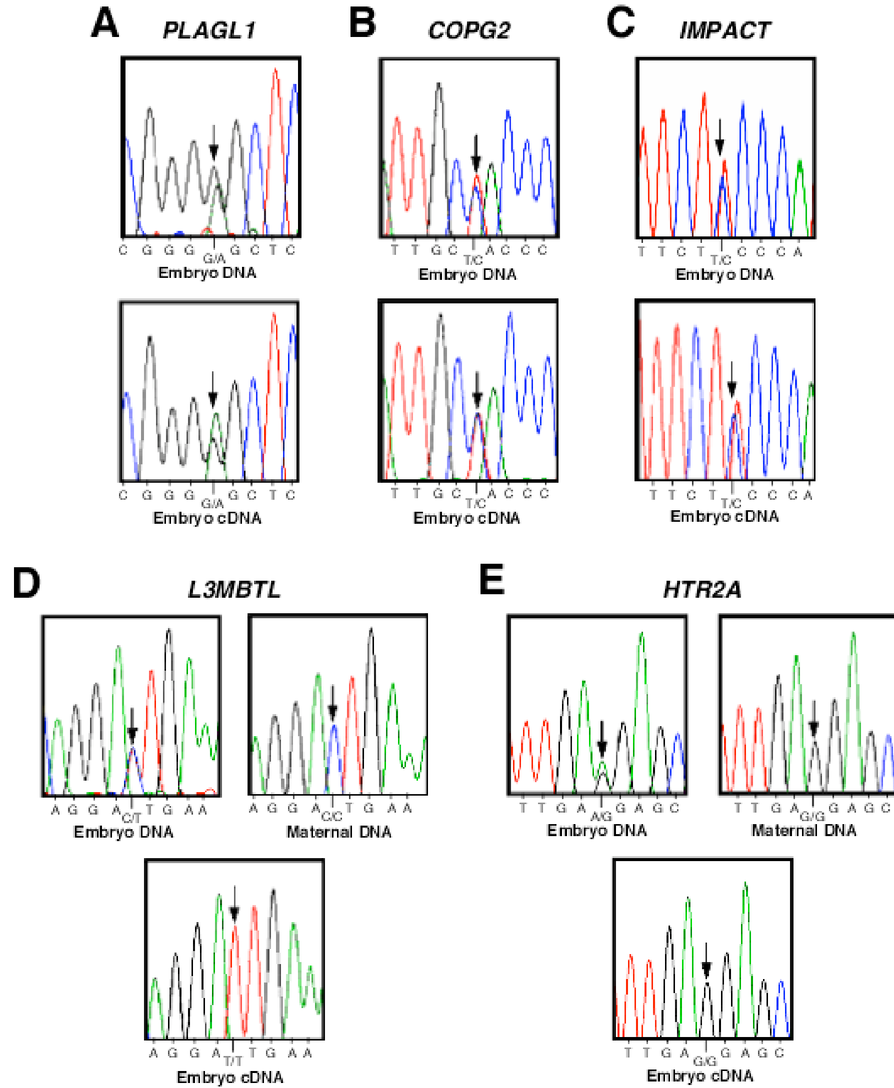


Figure 2.1: Analysis of imprinting in Eutherian orthologues of Metatherian loci in the opossum *M. domestica*

Checking for the persistence of an embryonic genomic polymorphism in the cDNA; arrow denotes polymorphic site. Homozygous maternal genomic DNA was investigated in the event of monoallelic expression. The five marsupial genes investigated were (A) *PLAGL1*, (B) *COPG2*, (C) *IMPACT*, (D) *L3MBTL* and (E) *HTR2A*.

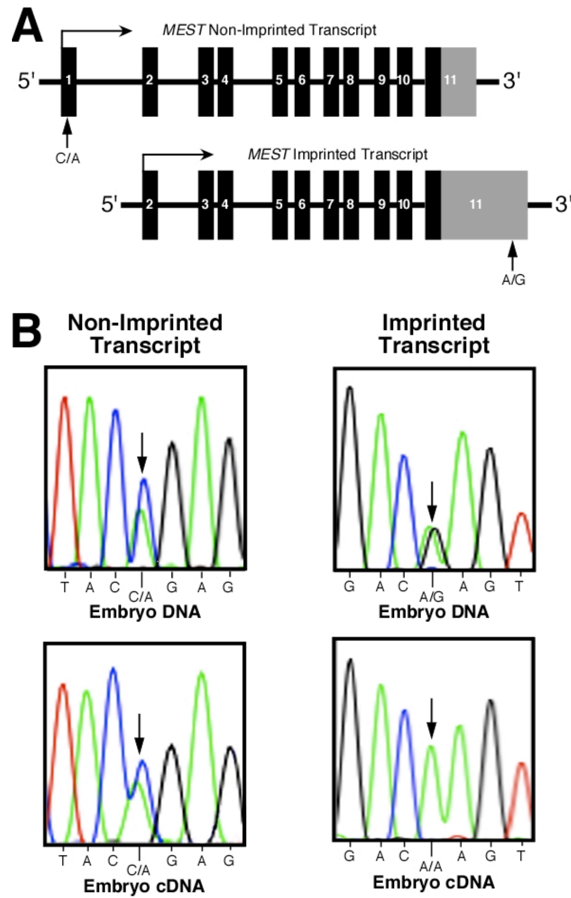


Figure 2.2: Analysis of imprinting at the *MEST* locus in *M. domestica*

(A) The putative *MEST* transcripts are represented graphically where the vertical arrows indicate polymorphic sites used to analyze imprinting. In transcript 1, the arrow marks the site of a C/A polymorphism in the DNA that would be observed in the cDNA if this is a non-imprinted transcript. In transcript 2, the arrow denotes the site of an A/G DNA polymorphism that should not persist in the cDNA, if this is truly a monoallelically expressed transcript. (B) 3′ RACE followed by imprint analysis yielded one imprinted and one non-imprinted transcript.

Table 2.1: Genes analyzed for imprint status in *M. domestica*

Gene	Function	Imprint Status		
		Human	**Mouse**	**Opossum**
COPG2	Binds to CDC42	Paternal	Maternal	**Biallelic**
L3MBTL	Polycomb protein	Paternal	Unknown	**Paternal**
IMPACT	Needed for growth	Biallelic	Paternal	**Biallelic**
HTR2A	Serotonin Receptor	Maternal	Maternal	**Maternal**
PLAGL1	Transcription Factor	Paternal	Paternal	**Biallelic**
MEST	Growth related, mesoderm specific	One paternally expressed, one non-imprinted transcript	One paternally expressed, one non-imprinted transcript	**One imprinted, one non-imprinted transcript**

2.3.2 Analysis of differential methylation

Twelve CpG-rich regions were identified in the vicinity of the six loci selected for imprinting investigation in *M. domestica*, and were further analyzed at the Broad Institute (Boston, MA). Bisulfite treatment and cloning of the chosen regions were performed in order to deduce the methylation status. We would expect an equal proportion of methylated and unmethylated clones if a region exhibits true differential methylation at the two alleles (indicative of imprinting), whereas the clones would exhibit similar methylation profiles if there was no differential methylation.

Results of this analysis revealed that there were no differentially methylated regions in the CpG-rich regions close to the biallelically expressed genes, as expected (**Table 2.2, Supplementary Table 2.2 and 2.3, Figure 2.3**). Importantly, the CpG-rich regions adjacent to the novel imprinted genes identified in *M. domestica* were also not differentially methylated.

IGF2R is the only imprinted gene, besides *IGF2*, that is imprinted in marsupials but not in Prototherians. Previous reports suggest that imprinting occurs in spite of the absence of DMRs at the promoter or intron 2 (sites of the two mouse DMRs that regulate imprinting), or even the non-coding RNA that is also required for imprinting in Eutherian mammals (Weidman, Dolinoy et al. 2006). Thus, additional CpG-rich regions in the vicinity of this gene were screened for differential methylation.

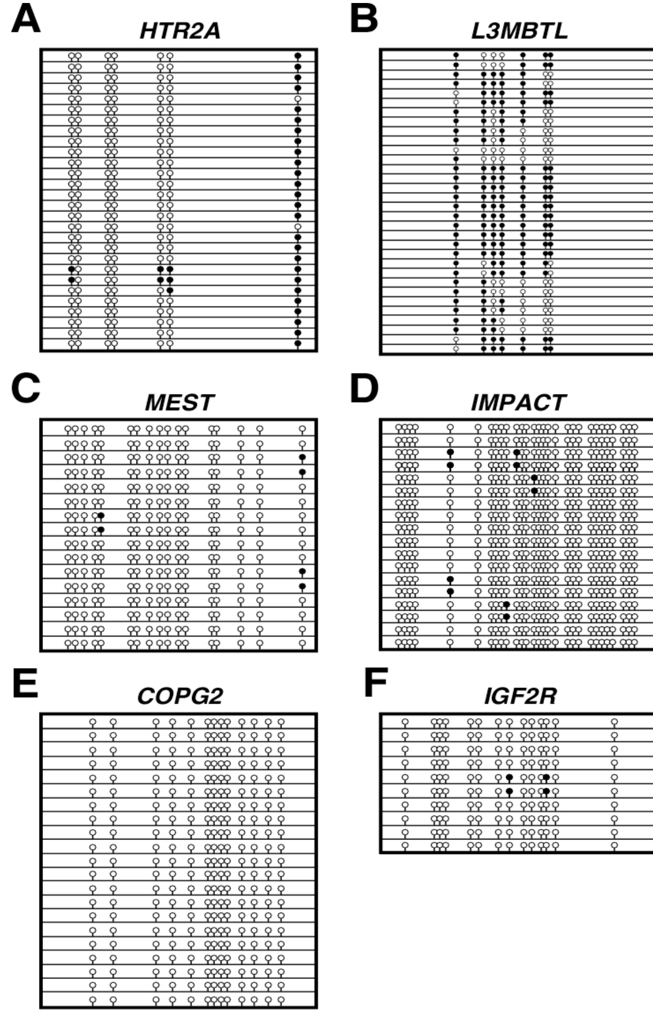


Figure 2.3: Methylation profile of CpG islands in the vicinity of orthologues of Eutherian imprinted loci in *M. domestica*

Representative examples are shown from the 5′ upstream region analyzed in (A) *HTR2A*, (B) *L3MBTL*, (C) *MEST*, (D) *IMPACT*, (E) *COPG2*, and (F) *IGF2R*. Unfilled circles depict unmethylated cytosines at CpG sites, while filled circles depict methylated cytosines.

Table 2.2: Methylation status of CpG-rich regions in *M. domestica* in the vicinity of orthologues of Eutherian imprinted loci

Gene	CpG-rich Region Location	Conservation in Mouse	DMR
PLAGL1	Putative promoter	Yes	**No**
L3MBTL	Putative promoter, Intron 2, Intron 4	Yes Yes Yes	**No** **No** **No**
IMPACT	20 kb upstream of promoter, Putative promoter	Yes Yes	**No** **No**
HTR2A	10 Mb upstream of promoter, Putative promoter	Yes Yes	**No** **No**
COPG2	4 Mb upstream of promoter	Yes	**No**
MEST	4 Mb upstream of promoter	Yes	**No**
IGF2R	38 kb upstream of promoter, Intron 11	Yes No	**No** **Yes**

Two regions were identified: one, 38 kb upstream of the putative transcription start site that could be of regulatory importance since ICRs have been identified several hundreds of kb upstream of the genes they regulate, and the other within intron 11. The CpG island located upstream of *IGF2R* was largely unmethylated (**Figure 2.3**); however, the one lying within intron 11 was differentially methylated (**Figure 2.4**). This DMR appears to be unique to *M. domestica*, since the mouse, dog or human *IGF2R* do not contain a homologous region. Single nucleotide polymorphisms were used to establish the parental origin of the expressed allele. A heterozygous embryo with a homozygous mother was identified, and the maternal allele was expressed in the embryo, as has previously been reported for this gene in *Didelphis virginiana* (Killian, Byrd et al. 2000). Bisulfite treatment and cloning of the DMR region revealed that the maternal (expressed) allele was almost completely methylated, whereas the paternal (non-expressed) allele was nearly completely unmethylated. This is similar to the methylation pattern at DMR2 **(Figure 1.3)** in the mouse, which leads to production of the antisense transcript *Air* (Barlow, Stoger et al. 1991). Search for an anti-sense transcript arising from this region using strand- and anti-sense strand- specific RT primers around the DMR region failed to detect a transcript for any anti-sense strand (**Figure 2.4**, **Supplementary Table 2.4**).

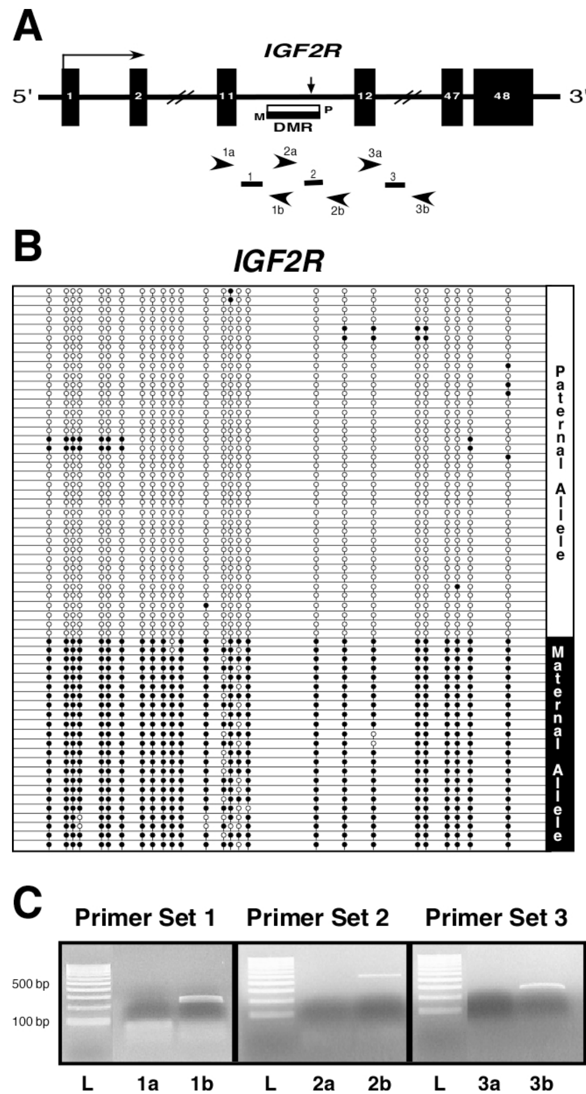


Figure 2.4: Methylation profile and anti-sense transcript analysis around the novel DMR identified in the *IGF2R* gene of *M. domestica*

(A) PCR primers used are labeled as 1, 2 or 3, while the RT primers for anti-sense strand detection are labeled as "a" and that for sense-strand detection as "b". (B) Methylation analysis of a heterozygote. (C) Strand-specific RT-PCR analysis shows a PCR product only by reverse transcription utilizing the "b" primers. L is the 100 bp DNA ladder.

2.3.3 Analysis of histone modifications at the promoter region

Chromatin immunoprecipitation was performed on adult opossum tissues using antibodies specific for H3 Lysine 4 dimethylation and H3 Lysine 9 trimethylation, as well as with Rabbit Serum Albumin to judge the level of non-specific immunoprecipitation. As a control for the antibodies, amplification of the promoter region of the highly expressed albumin gene (*ALB*) in *M. domestica* liver tissue, compared to its own introns and exons, showed that it was enriched for the activating mark H3 Lysine 4 dimethylation at the promoter region, but not for the repressive mark H3 Lysine 9 trimethylation. Similarly, in brain tissues, the promoter of the non-expressed MHC gene was enriched for the repressive H3 Lysine 9 trimethylation mark.

The regions upstream of *L3MBTL*, *HTR2A* and *IGF2R* were enriched for the activating mark H3 Lys4 dimethylation but not for the repressive H3 Lys9 trimethylation mark (**Table 2.3**, **Figure 2.5, Supplementary Table 2.5**). To assess whether the activating modification was associated with only one allele, ChIP was performed on individuals polymorphic in these regions. Clones of the immunoprecipitated samples as well as input DNA were analyzed, revealing that in the case of *IGF2R*, *L3MBTL* and *HTR2A*, H3 Lys4 dimethylation was significantly enriched on one allele (**Table 2.3**), while the input contained equal representation of both alleles. In the case of *L3MBTL*, the enriched allele was established as the paternal (expressed) allele since the polymorphic individual's father was homozygous at this locus.

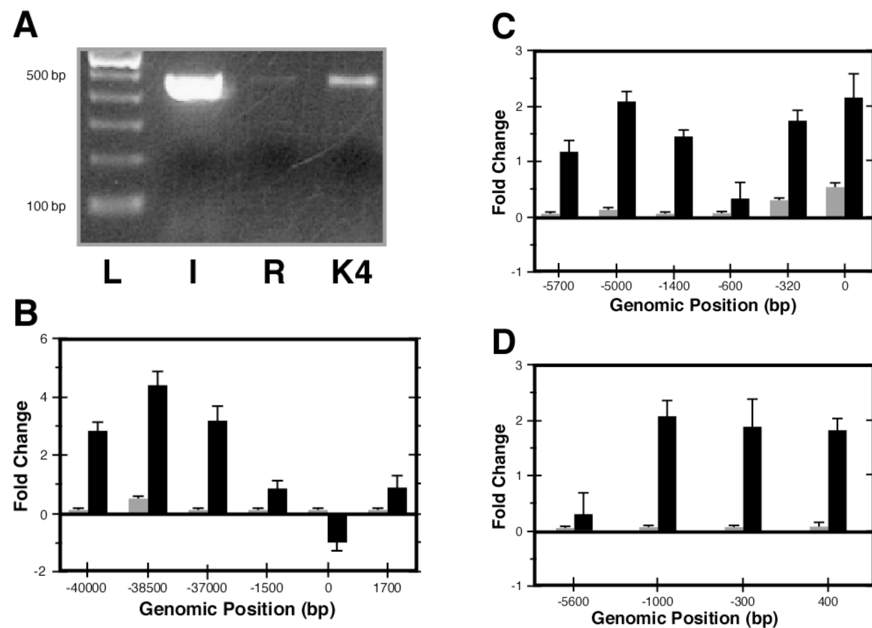


Figure 2.5: Analysis of histone modifications at the promoter regions of orthologues of Eutherian imprinted loci in *M. domestica*

(A) Analysis at the *IGF2R* putative promoter region by chromatin immunoprecipitation with H3K4 dimethyl antibody and subsequent PCR is shown, where, L denotes the 100 bp ladder, I the input, R the sample immunoprecipitated with Rabbit Serum Albumin (non-specific pull-down) and K4 the sample immunoprecipitated with H3K4 dimethyl antibody. (B), (C), (D) Analysis of enrichment by real-time PCR at the promoter regions of *IGF2R*, *HTR2A* and *L3MBTL*, respectively. The fold difference was plotted on the Y-axis and the distance from the transcription start site on the X- axis, with 0 representing the putative transcription start site. The black bars represent H3K4 dimethylation while the grey bars represent H3K9 trimethylation.

Table 2.3: Allele-specific enrichment of histone modification H3 Lysine 4 in promoter regions of orthologues of Eutherian imprinted loci in *M. domestica*

Gene	Poly-morphism	Genomic Location	Input: No. of Clones		H3K4 dimethyl ChIP sample: No. of clones		p-value for allele-specific enrichment	
			Allele 1	Allele 2	Allele 1	Allele 2	Input	H3K4 ChIP sample
L3MBTL	C/A	1: 315,862,133	7	11	18	9	0.346	0.083
HTR2A	G/C	4: 322,101,216	8	7	4	11	0.715	0.018
IGF2R	T/C	2: 445,407,861	5	6	1	10	0.6698	0.0063

2.4 Discussion

Genomic imprinting is thought to have evolved in placental mammals in order to tightly regulate allocation of resources to progeny. Selective silencing of one of the parental alleles, at least in Eutherians, is achieved in part by *cis*-acting imprint control regions (ICRs) that are selectively methylated on one of the two alleles. These ICRs also appear to be associated with allele-specific activating or repressive histone modifications on the underlying chromatin. The mechanisms leading to the establishment of imprinting in marsupials, however, is unclear since differential methylation has not been observed at four of the six imprinted loci previously investigated. The role of allele-specific histone modifications has never been investigated.

In this study, the imprint status of *M. domestica* orthologues of genes imprinted in Eutherians (i.e., mice and humans) was determined. Those found to be imprinted were screened for epigenetic marks such as differential methylation and histone modifications known to play a critical role in imprinting establishment in Eutherians. Three of the six genes analyzed (*L3MBTL*, *HTR2A* and *MEST*) exhibited monoallelic expression in this marsupial, showing substantial conservation with Eutherian-specific patterns. Interestingly, *MEST* had one monoallelic and one biallelic transcript, which is consistent with reports in mice, humans and the tammar wallaby (Kobayashi, Kohda et al. 1997; Suzuki, Renfree et al. 2005). Furthermore, parent-of-origin expression analysis showed that *L3MBTL* is expressed from the paternal allele, while *HTR2A* is expressed from the maternal allele, also consistent with mouse and human expression patterns **(Table 2.1)**.

MEST parent-of-origin expression remains undetermined since no informative polymorphisms in parent-offspring duos could be found; however, due to high conservation of imprinting at this locus, it is likely that the gene is paternally expressed, just as it is in another marsupial, the tammar wallaby (Suzuki, Renfree et al. 2005).

Although the Metatherian marsupials share many homologous features with Eutherians, differences are also apparent. In contrast to Eutherians, the imprint status of all three imprinted genes identified is maintained in tissues derived from the three germ layers (liver, brain and kidney), whereas polymorphic imprinting is common in Eutherians (Morison, Ramsay et al. 2005). Additionally, only a subset of genes imprinted in mice and humans appear to be imprinted in marsupials, even though the genes themselves are conserved. Interestingly, genes lying within conserved Eutherian "imprinted domains" appear to lack coordinate imprinting in the opossum. For example, in *M. domestica, MEST* has at least one imprinted transcript, while its adjacent gene *COPG2* is expressed from both alleles. In mice, however, both genes are imprinted. Similarly, in the tammar wallaby, *UBE3A* and *SNRPN* (the only two genes of the PWS-AS domain that could be identified) lie on different chromosomes and are not imprinted (Rapkins, Hore et al. 2006); while imprinted expression is restricted only to the retrotransposed gene *PEG10* within its imprinted domain (Suzuki, Ono et al. 2007).

Moreover, the imprint regulatory features also seem to differ between the two lineages. Out of the 14 CpG rich regions analyzed in *M. domestica*, a number of which are known to serve as ICRs for the adjacent genes in Eutherians, only one DMR was

identified – a novel DMR within *IGF2R* intron 11. Identification of this DMR demonstrates the possibility that such regions may have been missed in previous analyses due to incomplete sequence availability. Future studies need to focus on a regulatory role and germ cell of origin for this DMR in order to elucidate its function in governing imprinting at this locus. Given that the maternal allele is methylated in somatic tissues, it is possible that this mark is a maternally-derived gametic imprint. However, study of gametes will be required to determine if this methylation represents a primary imprint mark (established in gametes) or a secondary mark (established post-fertilization). Nonetheless, the lack of homology between the *IGF2R* intron 11 DMR and this region with human, dog or mouse sequence suggests that DMRs may not be conserved between Eutherians and Metatherians. Other DMRs, such as the *PEG10* DMR, also vary between Metatherians and Eutherians. In the tammar wallaby, this DMR is limited to the promoter region, while in Eutherians, the maternal allele-specific methylation spreads to the promoter of its adjacent gene, *SGCE*, leading to its paternal expression as well (Suzuki, Ono et al. 2007). Interestingly, the newly reported DMR in *M. domestica IGF2* bears 50% homology to humans (Lawton, Carone et al. 2008), yet its demethylation leads to activation of the maternal allele, instead of silencing the paternal allele as seen in mice.

Histone modifications at the promoters of imprinted genes also vary between Eutherians and *M. domestica*. The presence of the allele-specific activating histone mark H3 Lysine 4 dimethylation at the putative promoters of *IGF2R*, *HTR2A* and *L3MBTL* is consistent with a significant role in determining the parental-specific expression status in

Eutherians. However, unlike Eutherians, the repressive mark H3 Lysine 9 trimethylation is absent from the promoters of imprinted genes in *M. domestica*. Nevertheless, this is the first study in marsupials that demonstrates the presence of histone modifications at imprinted loci and supports that they are strong candidates for the "primordial imprint mark" in the absence of differential methylation. The presence/absence of other histone modifications as well as the germ-line derivation of these marks remains to be elucidated. Moreover, experiments showing abolishment of imprinting from disruption of these histone tail modifications will be necessary to prove their definitive role in the imprinting process.

Based upon the differences of imprinted genes and regulatory mechanisms between Metatherians and Eutherians, it appears that the features of genomic imprinting observed in Eutherians are not derived from the Metatherian lineage. Imprinting seems to have evolved independently in the Metatherians and Eutherians, probably with different selective forces acting on the growth-regulatory genes (Edwards, Rens et al. 2007; Hore, Rapkins et al. 2007). Recent findings add complexity to this notion. The *H19* non-coding RNA was found in the tammar wallaby, and has been shown to be maternally expressed and associated with a differentially methylated region in the *IGF2* imprinted domain (Smits, Mungall et al. 2008). This is the first and only locus where all the regulatory features of imprinting seem to be conserved between Eutherians and Metatherians. Thus, Smits et al. suggest sequential evolution of imprinting between the two lineages. However, they also suggest that imprinting of "singleton imprinted genes" such as *MEST*,

IGF2R and *PEG10*, which are controlled by simple regulatory mechanisms, evolved at separate times in marsupials and true placental mammals. It would be informative to identify genes that are imprinted in the marsupials, and not in the true placental mammals, to further corroborate this view. Our results show that while there are similarities in imprinted genes between Metatherians and Eutherians, stark differences also exist. It is imperative to extend the evolutionary analyses of imprint control mechanisms to many more loci and species before we can conclusively determine whether the evolution of imprinting occurred in a convergent or divergent manner, or possibly both.

3. Investigation of Imprinting in *DGCR6/DGCR6L* within the Primate Lineage

3.1 Introduction

Monoallelic gene expression evolved in mammals in three forms: 1) random expression, 2) random inactivation of one X chromosome to mediate dosage compensation in females and 3) genomic imprinting. Genomic imprinting is a heritable epigenetic phenomenon characterized by parent-of-origin dependent monoallelic expression. Imprinting plays an important role in regulating embryonic growth, and is highly tissue and developmental stage specific. Due to the functional haploidy of imprinted genes, they are particularly vulnerable to deleterious mutations, and have been implicated in several disorders involving uniparental disomies as well as in cancers (Murphy and Jirtle 2003). However, their role in modulating complex phenotypes like autism and schizophrenia remains largely unexplored, even though epigenetic dysregulation can potentially explain the variability and parental effects observed in such abnormalities (Petronis 2001). DiGeorge Syndrome (DGS) is another complex disorder characterized by a wide range of physical and mental anomalies, and a majority of the patients develop depressive disorders or schizophrenia later in life (Maynard, Haskell et al. 2002). Parent-of-origin effects have also been seen in DGS, suggestive of a role for genomic imprinting (Eliez, Antonarakis et al. 2001). DGS patients carry a deletion on chromosome 22 considered the DiGeorge Critical Region. *DGCR6* sits at chromosome location 22q11.2, within this DiGeorge Critical Region. If *DGCR6* is imprinted, this

potentially would explain the parental effects and pathology associated with the deletion of this locus.

DGCR6 was predicted to be imprinted using a genome wide bioinformatics approach designed by our laboratory (Luedi, Dietrich et al. 2007). DNA sequence characteristics were used with a combination of multiple classifier algorithms to specifically predict human imprinted genes. In contrast, studies searching for imprinted genes in an orthologous region in mice showed this gene to be biallelically expressed (Maynard, Meechan et al. 2006).

DGCR6 function is not well defined, but since it bears homology to the *Drosophila melanogaster* gonadal protein (involved in development) and to laminin, it has been implicated in cell and tissue re-organization. Hence, its disruption could lead to neural crest cell migration defects and the pathological conditions observed in DiGeorge Syndrome. A chicken model also highlights the role of this gene as a master regulator of development by influencing expression of several downstream genes like *Tbx1*, which happens to also be implicated in DGS (Hierck, Molin et al. 2004). The most intriguing observation is the association of this region to schizophrenia susceptibility in completely independent linkage studies (Liu, Heath et al. 2002).

Interestingly, the *DGCR6* locus underwent duplication specifically in the primate lineage (**Figure 3.1**). Both copies, *DGCR6* and *DGCR6L*, lay within the repeat regions sc11.1a and sc11.b, respectively (Edelmann, Stankiewicz et al. 2001).

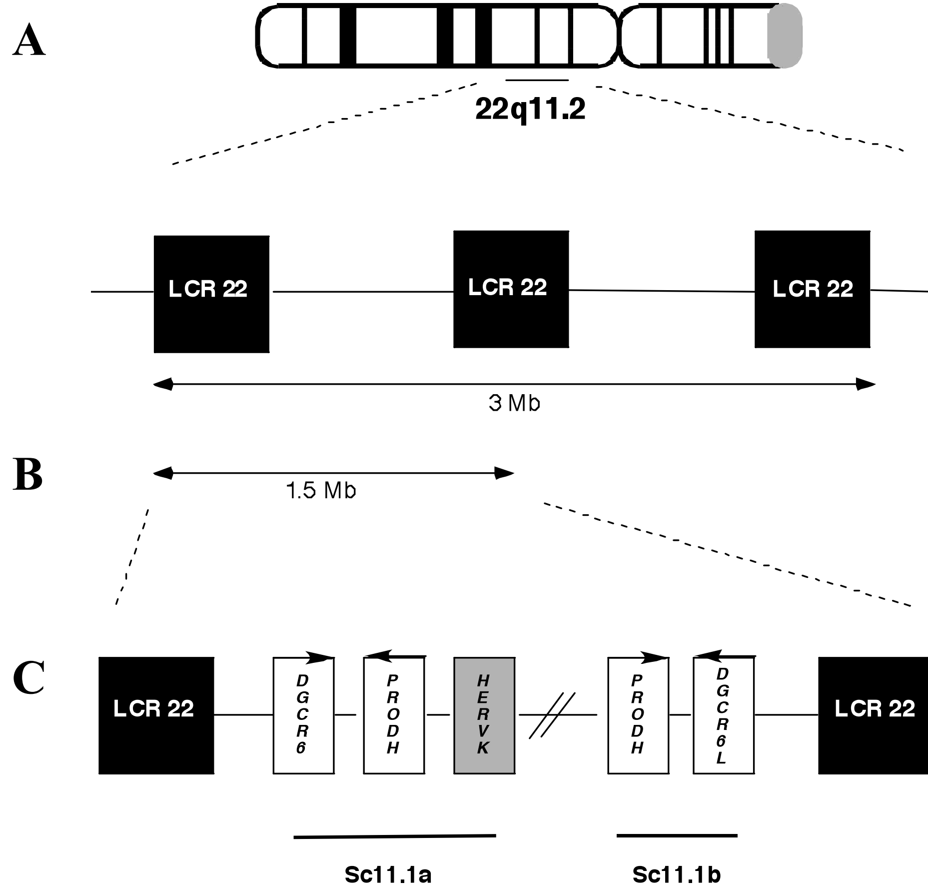


Figure 3.1: Duplication at the *DGCR6/DGCR6L* locus

(A) DiGeorge Syndrome patients bear a well-defined 3Mb microdeletion at 22q11.2, mediated by low copy repeats (LCR22s) flanking the breakpoints. (B) Patients with a nested 1.5 Mb deletion exhibit all the characteristics of the syndrome, thus this is termed the DiGeorge Critical Region. (C) The critical region in primates bears a duplication containing the genes *DGCR6/DGCR6L* and *PRODH*/pseudogenized *PRODH* on the duplicated regions (named sc11.1a and sc11.1b), respectively. An endogenous retroviral insert HERV-K is located at the distal end of sc11.1a.

The duplication has been confirmed in numerous primate species, supporting the view that it originated at least 12 million years ago, and may even have predated the divergence of the Catarrhines (old world monkeys) and Platyrrhines (new world monkeys). What is unique to humans, however, is the insertion of retrovirus HERV-K downstream of *DGCR6* (Barbulescu, Turner et al. 1999). An even more intriguing fact about the gene is that both copies are still functional, despite being 97% identical in structure, while the duplicate copy of neighboring gene *PRODH* lying within sc11.b has become pseudogenized. This suggests that *DGCR6* may have been dosage-compensated upon duplication by a mechanism other than silencing of one copy. Imposition of monoallelic expression of each copy may have occurred followed by the acquisition of silencing in a parent-of-origin specific manner.

This mode of imprinting evolution has been suggested for other genes, including members of the Makorin gene family, wherein the ancestral gene *Mkrn1* is not imprinted, while its retrotransposed copy *Mkrn3* is paternally expressed in mice and humans (Gray, Hernandez et al. 2000). Similarly, the small ribonucleoprotein *SNRPPB'/SNRPB* has a brain-specific paternally expressed paralogue *SNRPN* (Gray, Smithwick et al. 1999). Hence, a model has been proposed wherein duplication of biallelically expressed genes is believed to cause monoallelic expression of the paralogues by epigenetic mechanisms like methylation (Walter and Paulsen 2003). Subsequently, divergence of the two paralogues might make monoallelic expression obsolete. Alternatively, the benefits of

parent-specific expression (as explained by the conflict hypothesis) may lead to the establishment and persistence of imprinting in either one or both copies of the gene.

Thus, despite mouse studies suggesting biallelic expression for *DGCR6* (Maynard, Meechan et al. 2006), the above facts support further investigation of its imprint status in primates. We analyzed *DGCR6* and *DGCR6L* expression first in humans and then in chimpanzees, which also possess a duplicated locus. We created a phylogenetic tree to determine when this duplication evolved, and to correlate this information with expression patterns. Our studies indicate that this duplication is likely to have occurred in a common ancestor of the great apes and led to random rather than parent-of-origin expression of these genes.

3.2 Methods and Materials

3.2.1 Investigation of expression status in humans

Eighty human fetal tissues were chosen for imprint analysis since *DGCR6/ DGCR6L* are highly expressed in brain, liver and kidney during early development (ascertained by Affymetrix microarray in the Susan Murphy laboratory, Duke University Medical Center). Forty fetal DNA samples were obtained from the Susan Murphy laboratory, Duke University Medical Center, and another 40 from the Jirtle laboratory fetal frozen tissue databank, Duke University Medical Center (obtained from NIH-supported Laboratory of Human Embryology at the University of Washington) with approval from the Duke University Institutional Review Board.

Total DNA was isolated from liver tissues using lysis buffer ATL (Qiagen Sciences, Inc., Valencia, CA) and Proteinase K digestion, followed by standard phenol:chloroform:isoamyl alcohol purification, and ethanol precipitation.

PCR was performed using 30ng DNA in 25 µl reactions, using 15 pmol primers, 1.5 mM $MgCl_2$, 10 mM dNTPs and 1.5 units Platinum *Taq* DNA Polymerase (Invitrogen, Carlsband, CA). PCR primers were designed to uniquely amplify each copy of the gene (**Supplementary Table 3.1**). Correctly sized products were extracted using spin-columns (Sigma-Aldrich, St. Louis, MO) from 1.5% agarose gels, and sequenced on an ABI 3130 sequencer (Applied Biosystems, Carlsband, CA) to examine the sequence for polymorphisms.

RNA was extracted from fetal liver, brain and kidney tissues of the informative individuals using RNA Stat-60 (Tel-Test, Gainsville, FL) according to the manufacturer's protocol. Residual DNA was removed from 1 µg RNA using DNaseI and then reverse transcribed using copy-specific primers and Superscript II (Invitrogen, Carlsband, CA). Negative controls containing no reverse transcriptase were included to rule out the possibility of products being produced from residual DNA contamination. The cDNA obtained was amplified in the same region, and the products sequenced by techniques previously described. The persistence of the polymorphism in the cDNA indicated biallelic expression whereas the presence of only one allele in the cDNA indicated monoallelic expression. In order to identify the existence of a parent-of-origin effect, decidua DNA was isolated from the mothers of heterozygous informants. Laser Capture

Microdissection was used to separate deciduas (maternal in origin) from placenta (fetal in origin). The decidua samples were amplified and sequenced to identify those homozygous at the corresponding SNP location. Homozygous maternal DNA was matched with the fetal cDNA to see whether the maternal or paternal allele was expressed in the monoallelic samples.

The study was also extended to determine whether imprinting was present in adult peripheral blood leucocytes. Exonic polymorphisms were specifically sought due to the low level of expression of this gene in peripheral blood, as determined by real time PCR (described in next chapter). One hundred DNA and corresponding RNA samples were obtained from the Susan Murphy laboratory, Duke University Medical Center for this purpose under a protocol approved by the Duke University Institutional Review Board, and the expression analysis was performed as previously described.

3.2.2 Expression analysis in chimpanzees

Fifteen liver tissue samples of the chimpanzee *Pan troglodytes* were collected from the Southwest National Primate Research Center, San Antonio, Texas, with IACUC approval.

The UCSC and Ensembl Genome Browsers (http://genome.ucsc.edu/, http://www.ensemblgenomes.org/) were utilized to look for orthologues of *DGCR6* as well as *DGCR6L*, using the complete genome sequence. Copy-specific primers were designed to uniquely amplify each copy of the gene.

DNA and RNA were isolated utilizing the same protocols used for human fetal tissues. PCR was also performed as previously described, and the expression status was determined utilizing the same strategy outlined earlier.

3.2.3 Generation of phylogenetic tree

The coding DNA sequences for human *DGCR6*/*DGCR6L* orthologues were obtained for several species (chimpanzee, gorilla, tree shrew, bushbaby, opossum, rat, mouse, dog) from the Ensembl genome browser. The opossum was utilized as the outgroup. Amino acid sequences for the putative proteins were derived using MacClade software (http://macclade.org/macclade.html), and then aligned. The alignment was used as input data in the PAUP software (http://paup.csit.fsu.edu/) designed by Dr. David Swofford, Duke University, to generate a phylogenetic tree.

3.3 Results

3.3.1 Expression analysis in humans

SNPs were identified within intron 2 and exon 3 of nine samples for *DGCR6* and in intron two of five samples for *DGCR6L*. Both genes were monoallelically expressed in the brain, liver and kidney of heterozygous samples (**Figure 3.2**). In the parent-of-origin expression analysis, decidua DNA from the mothers of heterozygotes was sequenced to look for those exhibiting homozygosity at the corresponding SNP locations.

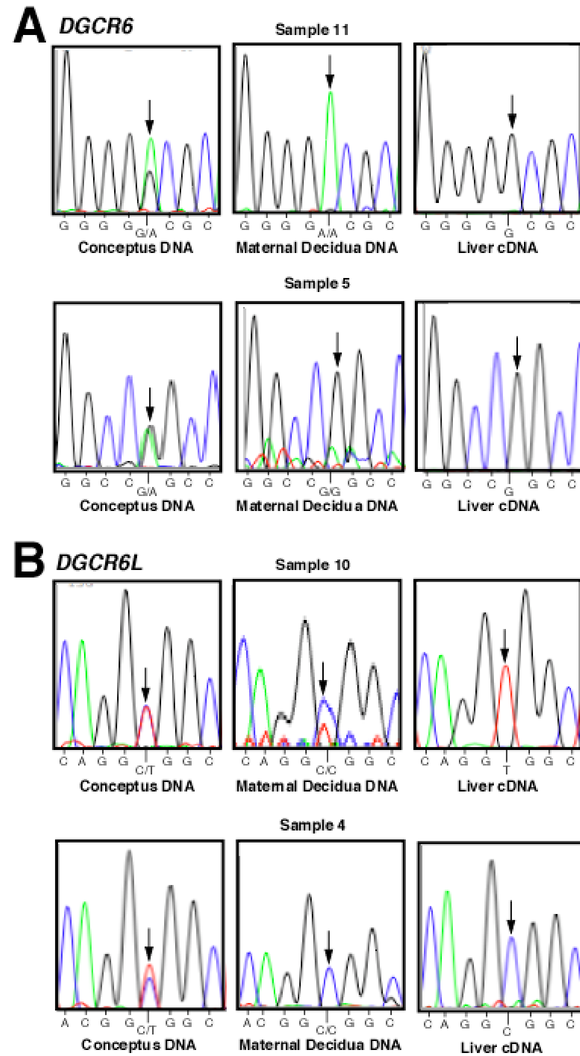


Figure 3.2: Analysis of imprinting at the *DGCR6/DGCR6L* locus in human samples

(A) Investigation of the *DGCR6* locus revealed monoallelic expression, due to non-persistence of a genomic polymorphism in the cDNA. However, the paternal allele was expressed in the case of sample 11, whereas the maternal allele was expressed in sample 5. (B) The *DGCR6L* locus exhibited a similar random monoallelic expression pattern, with the paternal allele expressed in sample 10 and the maternal allele in sample 4.

In the case of *DGCR6*, two conceptuses expressed the same allele as their homozygous mothers, whereas another two were found to express the paternally-derived allele (**Figure 3.2A**). These results indicate random monoallelic expression of the gene rather than imprinted expression. To ensure that the detected monoallelic expression was not due to amplification bias caused by low transcript abundance, we performed six experimental replicates following reverse transcription to check for consistent amplification of the same allele. Moreover, the fetal tissue expression database indicated that the genes were highly expressed in brain, liver and kidney further reducing the likelihood of amplification bias. A similar situation was observed for *DGCR6L*, wherein one conceptus consistently expressed the same allele as the homozygous mother, while a second conceptus always expressed the paternal allele (**Figure 3.2B**).

Peripheral leucocyte samples were also examined to determine if monoallelic expression was maintained in adult tissues. This investigation yielded three samples with polymorphisms in exon 3 for *DGCR6*, but none for *DGCR6L*. Subsequent RT-PCR and cDNA analysis confirmed monoallelic expression even in adult tissues (**Figure 3.3A**).

3.3.2 Expression analysis chimpanzee

Both *DGCR6* and *DGCR6L* in the chimpanzee were annotated in Ensembl and their sequences were used for expression analysis. Sequencing yielded two heterozygotes for *DGCR6* and one heterozygote for *DGCR6L*. cDNA analysis revealed that both genes were monoallelically expressed (**Figure 3.3B**).

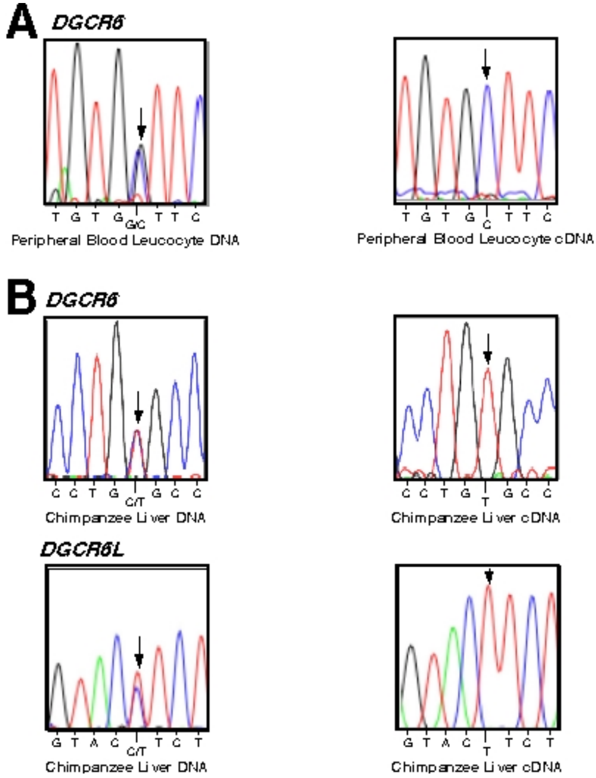


Figure 3.3: Analysis of allelic expression profiles for *DGCR6* and *DGCR6L* in human peripheral leucocytes and in chimpanzee liver tissues

(A) *DGCR6* is monoallelically expressed in human peripheral leucocytes due to non-persistence of the genomic polymorphism in the cDNA. (B) Both *DGCR6* and *DGCR6L* are expressed from only one allele in chimpanzees.

3.3.3 Phylogenetic tree for *DGCR6/DGCR6L*

The phylogenetic tree generated by maximum parsimony (**Figure 3.4**) indicates that the great apes (human, gorilla, chimpanzee) have a duplicated *DGCR6/DGCR6L* locus, while other mammals possess only one copy of the gene. Thus, the duplication probably arose in a common ancestor to the great apes. Interestingly, the paralogs in different species are more related to each other than the orthologs in the same species.

3.4 Discussion

Mammals exhibit genomic haploidy in a parent-of-origin dependent manner or in a random pattern and are able to use this to their evolutionary advantage (Keverne 2009). Although haploidy increases genomic susceptibility to deleterious mutations, it has proven benefits such as fine-tuning of fetal growth (genomic imprinting), compensating for X-chromosome gene dosage in female mammals (X-inactivation), and increasing the specificity and diversity of receptors in the immune and olfactory systems (random monoallelic expression). Nevertheless, disruption of monoallelic expression patterns also undoubtedly leads to disorders and diseases (Raslova, Komura et al. 2004). In fact, epigenetic lability may explain the phenotypic variability observed in many complex human disorders.

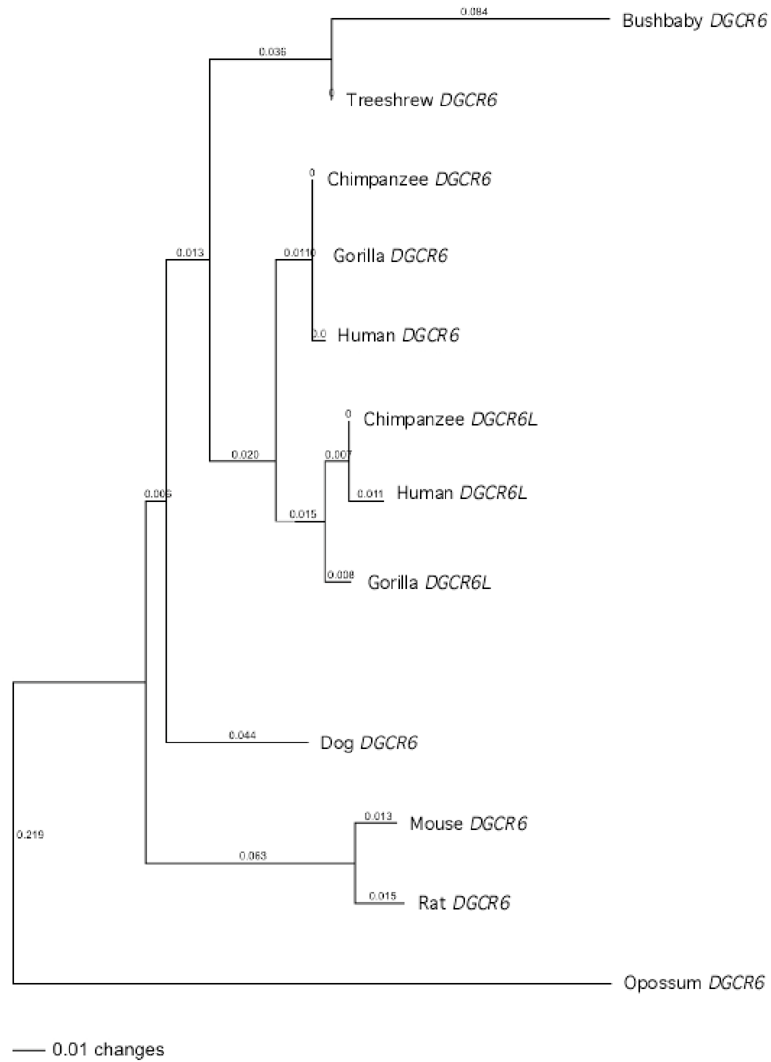


Figure 3.4: Phylogenetic tree for the *DGCR6/DGCR6L* locus

The tree was generated by maximum parsimony using the PAUP software. Coding DNA sequences from common Eutherians were utilized, and the Metatherian (Opossum) was used as an outgroup. *DGCR6* seems to have duplicated specifically in the primate lineage, and the paralogs are more related to each other than the orthologs.

In order to identify monoallelic expression associated with a complex disorder the gene *DGCR6* and its duplicate copy *DGCR6L*, that reside in a region deleted in DiGeorge Syndrome patients, were investigated. *DGCR6* is predicted to be imprinted specifically in humans by machine-learning algorithms developed in our laboratory (Luedi, Dietrich et al. 2007). Moreover, the presence of the duplication enhances the likelihood of monoallelic expression since this is a potential mechanism for acquiring dosage compensation.

Our experimental results demonstrate that *DGCR6* and *DGCR6L* are both monoallelically expressed in human fetal tissues. Moreover, this expression pattern appears to be established early in development, since it is maintained in fetal liver, brain and kidney tissues arising from the three germinal layers. *DGCR6* is also monoallelically expressed in adult peripheral blood leucocytes. However, contrary to our prediction, the gene is not imprinted, but rather, appears to show random monoallelic expression.

To determine whether the monoallelic expression correlates with the duplication of this locus, we investigated the expression status in our nearest neighbor, the chimpanzee. Chimpanzees are reported to have a duplicated locus, but not a retroviral insertion distal to the sc11.1a region. Since both these features could potentially lead to silencing, analysis of chimpanzees was anticipated to help explain if either of these phenomenon is responsible for this expression pattern. We found that both genes were monoallelically expressed in chimpanzees. We were unable to distinguish the parent-of-origin for the observed monoallelic expression due to the lack of parental samples.

However, based on our studies in humans, it is reasonable to expect that chimpanzees also display random monoallelic expression. Our results support that duplication at this locus is responsible for the observed monoallelic expression pattern, rather than the retroviral insertion that is specific to the humans.

To further explore the association between gene duplication and monoallelic expression, other genes lying within the sc11 duplication, such as *PRODH,* were also examined for random monoallelic expression. Unfortunately, we were unable to decipher *PRODH* expression status by the strategy utilized above due to extremely low levels of gene expression. As a negative control, we checked the status of another gene lying within the deletion involved in DiGeorge Syndrome, *UFD1L,* which is not duplicated, and this was shown to be biallelically expressed. Hence, our studies show the acquisition of random monoallelic expression at a duplicated locus in humans.

To better understand the exact timing of the *DGCR6* duplication that appears to result in monoallelic expression, we generated a phylogenetic tree utilizing orthologous gene sequences for several different mammalian genomes annotated to date. We observed a distinct branch that represented the great apes (humans, gorilla, chimpanzee), which have two copies of the gene. Thus, it appears likely that duplication of the gene occurred in a common ancestor to this group. Monoallelic expression probably evolved due to dosage compensation after duplication of this locus.

Random monoallelic expression is now believed to be more common in autosomal genes than was once previously thought (Gimelbrant, Hutchinson et al. 2007; Wang, Valo et al. 2007). The evolutionary advantage or functional significance of random monoallelic expression in autosomal genes is not yet completely understood. The phenomenon may be a bystander effect of nearby imprinted genes or could be maintained to compensate for gene dosage in the genome (Sano, Shimada et al. 2001). Random monoallelic expression could also result from low levels of a transcription factor that requires fine regulation (Sano, Shimada et al. 2001). Nonetheless, it is beneficial in increasing the specificity and diversity of immune and olfactory receptors (Keverne 2009), as well as in specifying cell identity (Esumi, Kakazu et al. 2005). Moreover, it has been argued that monoallelically expressed genes are easier to turn off when making critical developmental decisions (Sano, Shimada et al. 2001). Another possible advantage is the fact that random allelic expression ensures that at least 50 percent of the descendent cells are alive even if a deleterious mutation affects one allele of a cell (Watanabe and Barlow 1996).

Random monoallelic expression is regulated by alternate active or inactive chromatin states at each parental chromosome, with the polycomb protein Eed playing a major role in modifying histones during development to ensure only one allele is expressed (Krueger and Morison 2008). Asymmetric DNA methylation, non-coding RNAs and localization within the nucleus are also instrumental in controlling this process (Keverne 2009). Asynchronous DNA replication is another hallmark of monoallelic

expression (Gimelbrant, Ensminger et al. 2005), and this was recently shown to be closely associated with regions of tandem duplication in mice (Gimelbrant and Chess 2006). For example, *Nubp2*, *Igfals* and *Jsap1* are three mouse genes lying within a complex genomic region that were recently identified to be expressed randomly from only one allele (Sano, Shimada et al. 2001).

The genes *DGCR6* and *DGCR6L* also reside in a complex genomic region and exhibit monoallelic expression in humans and chimpanzees. This is the first time that such random monoallelic expression has been demonstrated by analysis of whole tissues rather than in single cells, indicating that it is established extremely early in development. This must be confirmed with another strategy like FISH (for determination of asynchronous replication), since the active allele is known to replicate earlier than the inactive allele in a non parent-of-origin dependent manner. However, if a parental bias is seen during replication, this could also indicate that there are imprinted transcripts present in this region. Due to low level transcription of each of these, PCR amplification might selectively amplify one or the other in different reactions, giving the impression of random monoallelic expression. Multiple imprinted transcripts have been observed at the *GNAS* locus, where both paternally and maternally expressed transcripts arise from the same region, and this was not deciphered by simple PCR based strategies (Peters, Wroe et al. 1999). Moreover, the question regarding whether or not *DGCR6* and *DGCR6L* are coordinately expressed from the same parental allele for an individual still remains to be answered. This needs to be addressed using high-throughput techniques which can track

polymorphisms over long genomic regions, since the two genes are approximately 1.5 Mb apart on the same chromosome.

Genes that are expressed in a random monoallelic manner are more sensitive to dosage disruption, and this might have important consequences in understanding their role in different disorders. For example, *FLI1* is monoallelically expressed at an early stage of megakaryopoiesis, and its loss due to a hemizygous deletion in Paris-Trousseau/Jacobsen syndrome leads to thrombocytopenia (Raslova, Komura et al. 2004). It is possible that hemizygous loss of *DGCR6/DGCR6L* could lead to dosage imbalance and the variable developmental defects observed in DiGeorge Syndrome. Thus, it is imperative to test the expression and functional significance of these genes in patient samples.

Using expression and phylogenetic analysis, our investigation demonstrates random monoallelic expression of *DGCR6* and *DGCR6L* that most likely developed because of *DGCR6* duplication in a common ancestor to the great apes. These human results contrast with reports that indicate biallelic expression of *Dgcr6* in the mouse. The epigenetic differences between mice and humans show that the human epigenome, rather than that in the mouse, should be studied to enhance understanding of the human complex disorder, DiGeorge Syndrome, that results from hemizygous deletion at this chromosomal location.

4. Determining the Role of Imprinting and Epigenetic Dysregulation in DiGeorge Syndrome

4.1 Introduction

Genomic imprinting is the parent-of-origin dependent, monoallelic silencing of genes that renders loci functionally haploid and increases their vulnerability to disorders resulting from gene mutation. Several clearly defined syndromes are attributed to disrupted imprinting patterns (Das, Hampton et al. 2009), but imprinted genes are also implicated in a number of complex disorders such as bipolar disorder , schizophrenia and autism (Petronis 2001). In these disorders, epigenetic modulation can explain variable features such as age of onset, severity, and parental effects (Petronis 2001); however, no single genes have been identified yet that cause these neurological disorders.

DiGeorge Syndrome is a complex disorder in which parent-of-origin effects have been observed, but direct evidence for the involvement of imprinting has yet to be demonstrated. This congenital disease affects one in 4000 live births, and leads to developmental defects (e.g. craniofacial abnormalities, limb and digit anomalies, velopharyngeal insufficiency and conotruncal heart defects) as well as psychiatric symptoms that develop later in life (Maynard, Haskell et al. 2002). Children with the 22q11.2 deletion are reported to have impaired cognition including a low IQ (mean 75), impaired sustained attention, executive function and verbal memory (Shashi, Kwapil et al. 2010). Major psychotic disorders develop in 25-40% of affected individuals, and tend to follow the same courses in the general population, especially for DGS associated

schizophrenia (Bassett, Chow et al. 2005). The genetic cause has been attributed to a hemizygous deletion at chromosomal location 22q11.2, where rearrangements are mediated by faulty recombination between four low copy repeats known as LCR22s. Approximately 85 percent of patients carry a 3 Mb deletion. A minority with a smaller 1.5-2 Mb deletion still show all the characteristics of the disease, delimiting this as the DiGeorge minimal Critical Region (DGCR). Many of the symptoms involve defects arising during embryonic growth, particularly in formation of the aortic arches, resulting in abnormalities of the heart and appendages and in the migration of neural crest cells, causing the formation of cognitive and psychiatric anomalies. Thus, investigations of genes functionally responsible for this disorder have focused on growth regulatory transcription factors and cell adhesion molecules. Genes studied by creating transgenic knockouts in mice, such as the transcription factors *HIRA* and *Tbx1*, recapitulate many of the developmental defects observed in humans. Nevertheless, none of the candidate gene knockouts explains all the symptoms, leading to the classification of DGS as a "contiguous gene syndrome" (Maynard, Haskell et al. 2002). Furthermore, symptoms vary in their severity and onset, suggesting the involvement of epigenetic factors in their etiology. To quantify the brain-specific parental effects of the 22q11 deletion, magnetic resonance imaging has been employed (Eliez, Antonarakis et al. 2001). These studies showed that white matter was reduced in all DGS patients, but that gray matter showed a larger reduction in volume when the deletion was of maternal origin, indicating the involvement of at least one imprinted gene in the cognitive/psychiatric manifestations of

DGS. Such parental-specific cognitive effects have been seen in the context of deletions of other imprinted domains such as 15q11, which contribute to Angelman and Prader Willi Syndromes. For these reasons the mouse orthologues of genes lying within the DiGeorge Critical Region were investigated for their imprint status. Surprisingly, all genes in the region orthologous to DGCR were found to be biallelically expressed (Maynard, Meechan et al. 2006). Since the repertoires of imprinted genes are species-dependent, it is possible that imprinted genes exist in humans that are absent in mice. Computational approaches in our laboratory predict that the gene *DGCR6* lying in the DGS region (**Figure 4.1**) is imprinted in humans but not in mice (Luedi, Dietrich et al. 2007) Thus, we decided to investigate the potential role of *DGCR6* in DiGeorge Syndrome.

The exact function of *DGCR6* is presently unknown, but its homology to the *Drosophila melanogaster* Gonadal (*gdl*) gene (instrumental for gonadal and germ cell development), as well as to the laminin gamma chain, implicates a role in neural crest cell migration (Edelmann, Stankiewicz et al. 2001). Moreover, chicken models of DGS, created by *DGCR6* antisense construct injection, highlight its effect on pharyngeal arch development as well as in the regulation of other genes implicated in DGS e.g. *TBX-1* and *UFD1L* (Hierck, Molin et al. 2004). Also intriguing are completely independent genetic linkage studies implicating this region in schizophrenia susceptibility (Karayiorgou, Morris et al. 1995; Liu, Heath et al. 2002).

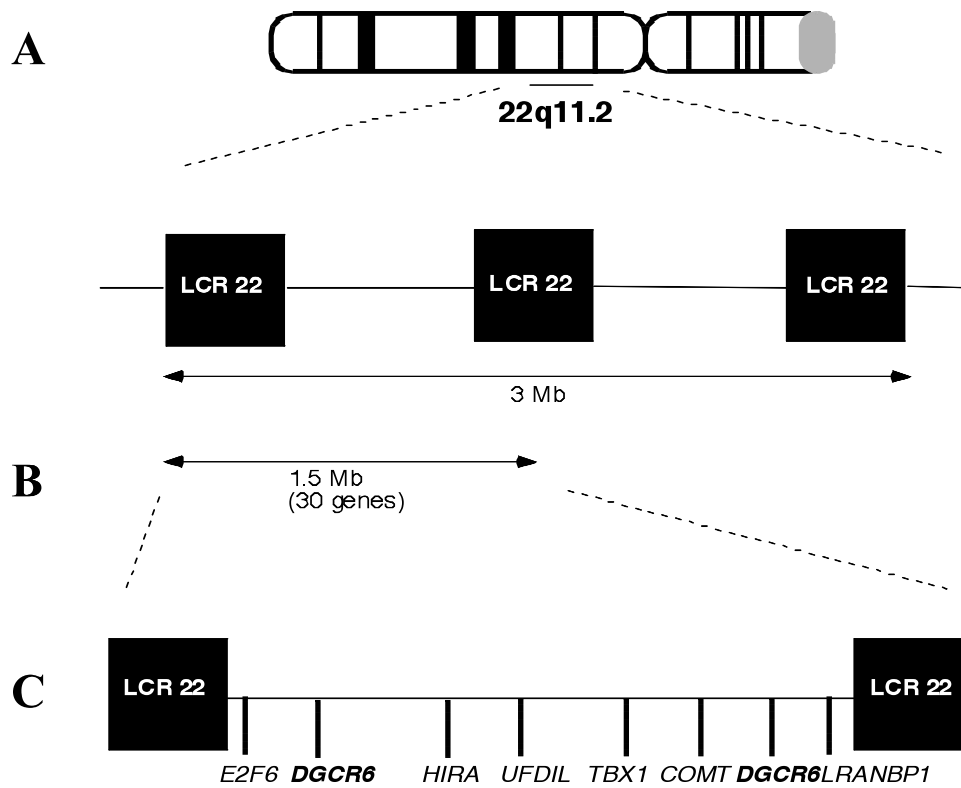


Figure 4.1: Genomic rearrangements in DiGeorge Syndrome and the *DGCR6/DGCR6L* locus

(A) DiGeorge Syndrome patients bear a well-defined 3 Mb microdeletion at 22q11.2, mediated by low copy repeats (LCR22s) flanking the breakpoints. (B) The DiGeorge Critical Region is a nested 1.5 Mb deletion that exhibits all the characteristic psychological and psychiatric symptoms of the syndrome. (C) Approximately 30 genes lie within the critical region, and those shown are implicated in the disease phenotype, including *DGCR6* and its duplicate copy *DGCR6L.*

An intriguing fact about this gene is that it has been duplicated in the primate lineage (the extra copy named as *DGCR6*-like, i.e., *DGCR6L*), with both genes residing in the 1.5 Mb deletion defined as the DiGeorge Critical Region. Consistent with the 97% identity of the cDNA sequences and only seven amino acid differences between them, both *DGCR6* and its duplicate copy *DGCR6L* have similar expression profiles (Edelmann, Stankiewicz et al. 2001) and are suggested to be redundant in function. Since duplication can lead to dosage compensation by random silencing of one allele of each paralogue and later evolve into parental-allele specific silencing (Walter and Paulsen 2003), the imprint status of *DGCR6* was assessed in humans.

Given the function of *DGCR6/DGCR6L* in neural crest cell migration and the possibility of their being imprinted, we hypothesized that dysregulation of expression caused by the deletion of one allele might be at the root of the variability in neurodevelopmental abnormalities observed in DGS. In our analysis, we observed that the DGS deletion led to deregulated expression of *DGCR6* and *DGCR6L* in patients, rather than reduced expression, when compared to control individuals. Moreover, those bearing a maternal deletion exhibited much higher transcript levels for both genes. This overexpression correlated with decreased performance in cognitive tests, particularly the sustained attention test, a hallmark of schizophrenia. Thus, *DGCR6/DGCR6L* levels can potentially help predict the onset of schizophrenia in DGS patients. We were also able to identify a differentially methylated region (DMR) in fetal tissues, which may be important in regulation of expression during development.

4.2 Materials and Methods

4.2.1 Sample collection

Blood was collected from 36 DiGeorge Syndrome patients (mean age = 12.2 years/ SD = 3.2 years/ males = 22, females =14) carrying a 22q11.2 microdeletion confirmed by flourescent *in situ* hybridization (FISH) and 13 controls (age and gender matched to the subjects, mean age = 12.7 years/ SD = 2.5 years/ males = 8, females =5) from Duke University Medical Center and Wake Forest University Health Sciences, under protocols approved by the Institutional Review Board of these institutions. The subjects were all non-psychotic. The ethnicity of the patients and controls was 90% Caucasian and 10% Afro-American.

A three-generation pedigree was drawn to ascertain developmental or genetic disorders, and mental illness, learning disabilities or other cognitive defects in the families of the DGS patients. Personal or family histories of cognitive defects, psychotic illness or congenital anomalies were used as exclusion criteria for the control subjects. DGS patients with an IQ of <50 were excluded from the study, as were control subjects with an IQ>115. This minimized the intellectual disparities between the two groups and in the case of the DGS children, ensured the performance of the neurocognitive battery.

Blood for DNA extraction (5 ml) was collected in K-EDTA tubes, and 2.5 ml of blood was drawn for RNA extraction in PAXgene tubes that contain a proprietary reagent to stabilize total RNAs (Qiagen sciences, Inc., Valencia, CA).

4.2.2 Parent-of-origin analysis of the deletion

DNA was extracted from the blood samples using the Gentra Puregene Kit (Qiagen sciences, Inc., Valencia, CA) at Wake Forest University Health Sciences. Briefly, RBCs were lysed with RBC Lysis solution, and the remaining white cell pellet was treated successively with Cell Lysis Solution, RNase A and Protein Precipitation solution to obtain pure DNA in solution. This was followed by isopropanol precipitation and resuspension of the DNA pellet in DNA Hydration Solution. These patients were then genotyped for 450 SNPs corresponding to the 1.5 Mb deleted region using the MassARRAY system and the iPLEX assay from Sequenom (Sequenom, Inc., San Diego, CA). SNPs were selected using the Tagger program built into the Haploview software for the HapMap data, which utilizes linkage disequilibrium to identify a minimal set of SNPs for this region. Blood from both parents was also collected for extraction of DNA and parent-of-origin analysis of the deletion which was performed by genotyping the same SNPs and correlating their inheritance patterns.

4.2.3 Expression analysis

RNA was extracted from control and patient samples using the PAXgene Blood RNA Kit (Qiagen sciences, Inc., Valencia, CA) for blood collected in the PAXgene tubes. Briefly, whole blood was pelleted, washed and treated with Buffers BR1, BR2 (Lysis buffer) and Proteinase K. RNA was precipitated utilizing Ethanol and passed through a spin column to specifically collect RNA. The RNA was reverse transcribed using oligo dT primers and Superscript II (Invitrogen, Carlsband, CA). Differences in

expression levels were quantified using Taqman real-time PCR on the Applied Biosystems (ABI) 7900HT Fast Real-Time PCR System (Applied Biosystems, Carlsband, CA). Custom-designed ABI gene expression assays for *DGCR6* (Hs00606390_mH) and *DGCR6L* (Hs_00819920_mH) were utilized, and each reaction was performed in triplicate. Beta-Actin was used as an internal control for these reactions. The Delta Ct value was calculated by subtracting the Ct value of the gene from the Ct value of B-Actin for each sample. The Delta Delta Ct value was calculated by subtracting the control mean Delta Ct value from the Delta Ct value for each sample, and the fold difference was plotted as 2^{DDCt}.

To analyze the effect of the parent-of-origin on expression, samples were grouped into three classes: those showing underexpression, average expression or overexpression. Samples that showed two SD times lower/higher fold expression than the control mean were considered as samples with under/overexpression respectively, and the rest were retained in the average expression group. Chi-square tests were performed to evaluate significance in expression difference between the maternally and paternally deleted patients versus the controls within the three classes.

4.2.4 Correlation with neuropsychological data

Neuropsychological data were collected from the patients and controls at Duke University Medical Center and Wake Forest University Health Sciences. The Wechsler Intelligence Scale for children (WISC) was used for intelligence testing. Based upon the NIMH Measurement and Treatment Research to Improve Cognition in Schizophrenia

(MATRICS) (Kern, Green et al. 2004) task force recommendation assessments were made for vigilance/attention, verbal learning and reasoning, as well as executive function since these are typically impaired in individuals with schizophrenia. Executive function was measured with the Wisconsin Card Sort Test (WCST), which involves matching of cue cards containing different shapes and colors (Chelune and Baer 1986). Sustained attention was assessed with the identical pairs (IP) and AX conditions of the Continuous Performance Test (CPT), which are based on recognition of identical numbers or a pattern of numbers, respectively (Cornblatt and Erlenmeyer-Kimling 1985; Erlenmeyer-Kimling, Rock et al. 2000). In addition, the California Verbal Learning Test (CVLT-C), that entails recalling words read from a list (Otto, Bruder et al. 1994) was used to assess verbal learning and memory.

We used independent sample t-tests to assess the correlation between performance in the various cognitive tests and the expression levels of *DGCR6* and *DGCR6L* separately.

4.2.5 Analysis of methylation status

Genomic DNA from human fetal liver samples was bisulfite treated to selectively convert the unmethylated cytosines to uracils, using the Qiagen Epitect Bisulfite Kit (Qiagen Sciences, Inc., Valencia, CA). Primers were designed to specifically amplify bisulfite-treated DNA in the genomic region containing *DGCR6* and *DGCR6L*, as well as 2 Kb upstream and 3 Kb downstream of each gene using Epidesigner-beta (http://www.epidesigner.com) from Sequenom (Sequenom, Inc., San Diego, CA). The

specificity of the primers (**Supplementary Table 4.1**) was checked by BLAST against the entire bisulfite-treated genome (http://bisearch.enzim.hu). PCR was performed using 1.5U of Platinum Taq DNA polymerase (Invitrogen, Carlsband, CA), 15 pmol of primers, 1.5 mM $MgCl_2$ and 10 mM dNTPs in a 50 µl PCR reaction volume. A differentially methylated region was identified using the Sequenom MassARRAY system (Sequenom, Inc., San Diego, CA) (Hartmer, Storm et al. 2003; Ehrich, Field et al. 2006). The results were analyzed using the Epityper software provided along with the Sequenom MassARRAY system.

To confirm the differential methylation status, bisulfite treated DNAs were amplified by PCR (**Supplementary Table 4.1**) and products were purified using a High Pure PCR Product Purification Kit (Roche Applied Science, Indianapolis, IN). Purified PCR products were cloned using TOPO TA Cloning Kits (Invitrogen, Carlsband, CA). The ligated plasmids were plated on LB-Agar-Xgal plates, and after overnight incubation at 37° C, white colonies were selected and amplified by whole cell PCR. T7 and SP6 primers were used for amplification with conditions as described above, and the inserts sequenced on the ABI 3130 sequencer (Applied Biosystems, Carlsband, CA). DNA from six human sperm samples, obtained from ABS (Analytical Biological Services, Inc., Wilmington, DE) were also bisulfite treated and analyzed in a similar manner.

DNA from 15 patients and 5 control samples, as well as two normal adult brain samples, obtained from ABS (Analytical Biological Servies, Inc., Wilmington, DE) were

also bisulfite treated and analyzed in the DMR regions as well as promoter regions for the two genes using the same primers and the Sequenom methodology previously described.

4.3 Results

4.3.1 Parent-of-origin and expression analysis

Genotyping of the 36 DGS patients and their parents showed 13 with deletions that were maternal in origin, and 12 with paternally-inherited deletions.

The expression of *DGCR6* and *DGCR6L* was analyzed in human fetal liver, brain and kidney as well as in peripheral blood leucocytes of unaffected individuals to judge the expression levels during the prenatal and postnatal periods. Both genes were highly expressed in fetal tissues. Expression was almost two-fold lower in blood, but because only blood was available from the DGS patients, we proceeded with our analysis using blood samples. We anticipated that one set of patients would have null expression status (because of deletion of the active allele), whereas the other set would show consistent expression (due to deletion of the non-expressed allele). In contrast to these expectations, *DGCR6* and *DGCR6L* both had dysregulated expression patterns, with no consistent levels observable in the different patient groups (**Figure 4.2**). There was less variation in expression levels among the control specimens.

When we analyzed the effect of parent-of-origin of the deletion, those with a maternally derived deletion exhibited much higher levels of expression of both genes **(Table 4.1, Figure 4.3)**. The p-value for difference in distribution for individuals

showing over-expression between maternally deleted patients and controls was less than 0.0001 (highly significant), while the p-value between paternally deleted patients and controls was 0.0019.

4.3.2 Correlation between expression and neuropsychological symptoms

Neurocognitive tests were performed on controls and patients, and the patients performed worse on all tests, as expected. On correlating expression of the two genes with performance in the various tests, expression of both genes tracked together with variability in neurocognitive ability. Overexpression of these genes had a detrimental effect on performance, particularly for certain neurocognitive tests, as indicated by the significant negative correlation (**Table 4.2, Supplementary Table 4.1**). The strongest correlation was observed for the verbal comprehension test and the sustained attention test (CPT-IP). Since the maternally deleted patient set contained a higher number of individuals exhibiting overexpression, this was also the set that had poorer cognition.

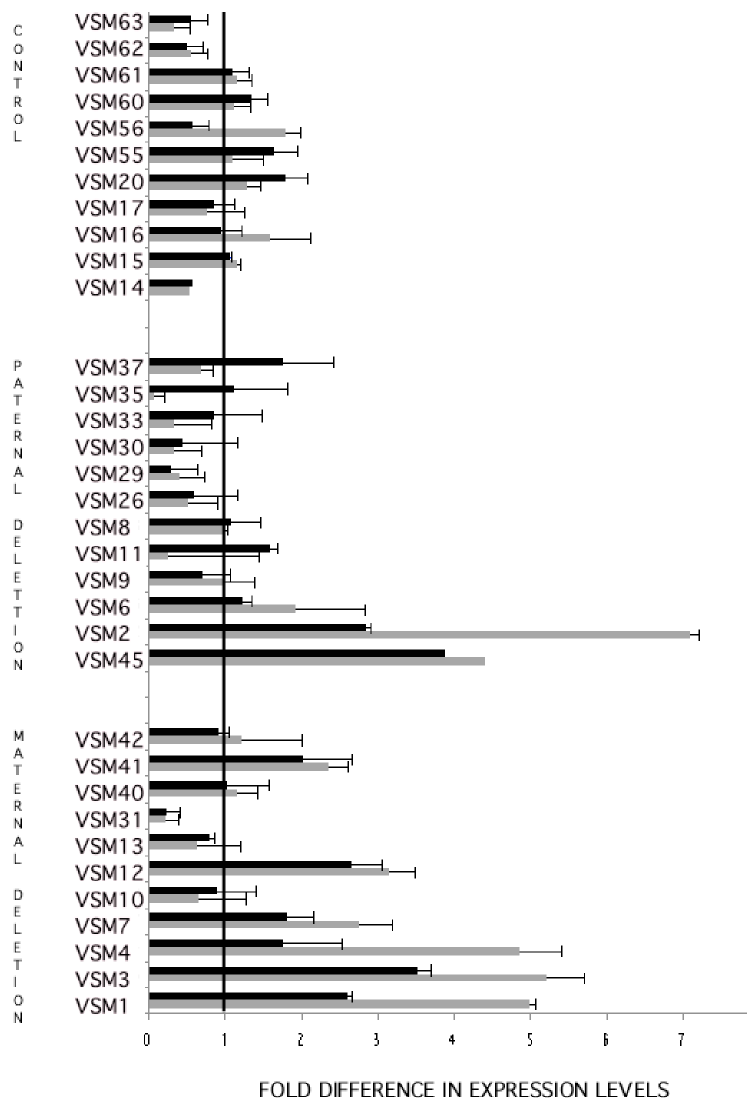


Figure 4.2: Gene Expression at the *DGCR6* and *DGCR6L* loci

Expression levels for *DGCR6* (black bar) and *DGCR6L* (grey bar) in controls and patients with paternal and maternal deletions are shown. The fold difference in expression level between the sample and the mean expression of the controls (black line) is shown on the X-axis.

Table 4.1: Number of people exhibiting differential expression levels of *DGCR6* and *DGCR6L* amongst patients and controls

	DGCR6			*DGCR6L*		
	Maternal Deletion	Paternal Deletion	Control	Maternal Deletion	Paternal Deletion	Control
Low Expression	1	5	1	1	1	1
Average Expression	4	4	9	4	9	9
High Expression	7	3	1	7	2	1

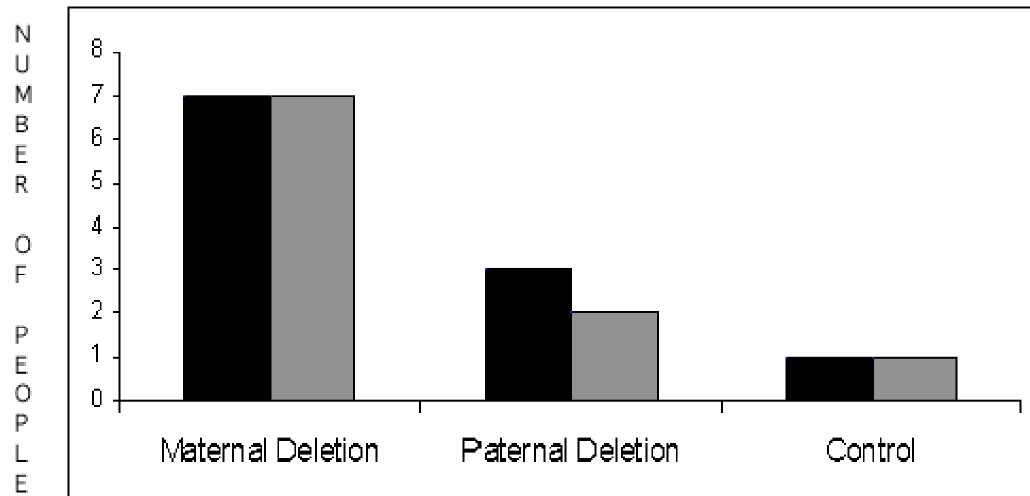


Figure 4.3: Effect of parent-of-origin of the deletion on expression of *DGCR6* and *DGCR6L*

People showing high expression levels of *DGCR6* (black bar) and *DGCR6L* (grey bar) in controls and patients with paternal and maternal deletions (as calculated in the Table 4.1) are shown. The number of individuals is shown on the Y-axis. Chi-square tests showed that the p-value for the difference in distribution between controls and maternally deleted individuals (< 0.0001) was highly significant.

Table 4.2: Summary of correlations between *DGCR6* and *DGCR6L* gene expression and neuropsychological symptoms in DGS patient

Test	*DGCR6*	*DGCR6L*
Global Assessment (GAS)	-0.082	-0.005
Any anxiety disorder (DISC)	-0.344 *	-0.255
IQ (WISC)- Verbal Comprehension	-0.363 *	-0.335 *
IQ- Perceptual Organization	-0.102	-0.183
IQ- Working memory	-0.027	0.174
IQ- Processing speed	-0.044	0.029
Executive Function (WCST)- PE	0.096	0.091
Executive Function- NPE	-0.190	-0.284
Executive Function- PCLR	-0.002	-0.058
Sustained Attention- CPT AX	-0.305	-0.397 *
Sustained Attention- CPT IP	-0.313 *	-0.356 *

* Correlation significant at the 0.05 level (Two-tailed test)

4.3.3 Analysis of methylation status

To find a possible explanation for the higher than normal *DGCR6* and *DGCR6L* expression in DGS patients, we compared the methylation levels at the promoter regions of the two genes in 15 patients, which included representatives from the maternally and paternally deleted sets, and five controls (**Figure 4.4B**). We observed that this region was primarily unmethylated in all samples, regardless of the deletion.

Differentially methylated regions are well-known epigenetic regulatory elements, and thus we next searched *DGCR6* and *DGCR6L* for DMRs covering the coding sequences as well as 3kb upstream and 2 Kb downstream of each gene. We located a small (approximately 200bp) region lying 1 Kb downstream of each gene that exhibited 50% methylation in fetal liver and brain.

By performing analysis of cloned alleles generated using sodium bisulfite modified genomic liver and brain DNA, predominantly methylated or unmethylated clones were observed in an approximate 50:50 distribution, confirming this as a differentially methylated region (DMR) (**Figure 4.4A**). To investigate the germ-line derivation of this mark, we analyzed six human sperm samples in the same region and found these to be fully methylated (**Figure 4.4B**). As this region contains no polymorphisms, no further assignment of parental origin of methylation could be determined, but our results are consistent with these DNA methylation marks being inherited from the father.

Investigation of the DMR in control and patient peripheral blood samples showed that both patient and control samples were fully methylated in this region (**Figure 4.4B**). DNA from normal adult brain also demonstrated complete methylation at these DMRs for *DGCR6* and *DGCR6L*.

Figure 4.4: Methylation analysis adjacent to *DGCR6* and *DGCR6L*

(A) Methylation profile of the cloned PCR products from the differentially methylated region identified in the vicinity of *DGCR6* and *DGCR6L* of human fetal liver samples. Unfilled circles depict unmethylated cytosines at CpG sites, while filled circles depict methylated cytosines, and each line denotes an individual cloned allele.

(B) Methylation profiling utilizing Sequenom methodology. The level of methylation is depicted as circles varying in color from yellow (0% methylation) to blue (100% methylation).

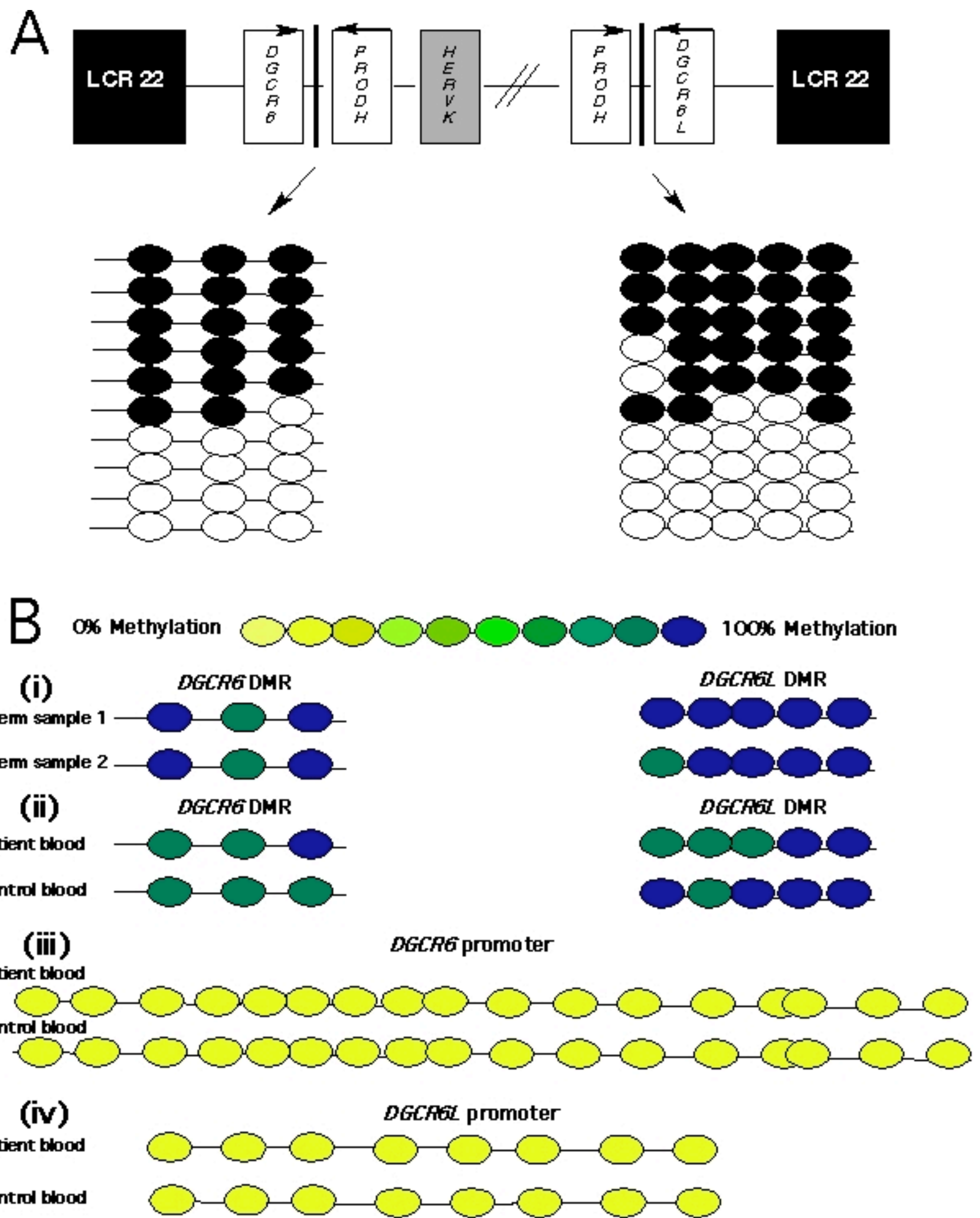
i) Two human sperm samples exhibit complete methylation at the DMR region of *DGCR6* and *DGCR6L.*

ii) The *DGCR6* and *DGCR6L* DMRs in adult blood samples are also completely methylated, as represented in patient VSM1 and control VSM16.

iii) The promoter region of *DGCR6* is completely unmethylated in adult blood samples, shown for patient VSM1 and control VSM16.

iv) The promoter region of *DGCR6L* is also completely unmethylated in the same samples.

Figure 4.4: Methylation analysis adjacent to *DGCR6* and *DGCR6L*



4.4 Discussion

DiGeorge Syndrome (DGS) or 22q11 deletion syndrome is characterized by a multitude of physical deformations involving defects in neural crest cell migration during fetal development, as well as psychopathological symptoms ranging from learning disabilities to schizophrenia (Maynard, Haskell et al. 2002). The effect of the 22q11 deletion on brain development has been quantified using magnetic resonance imaging. These studies found that white matter was reduced in all DGS patients but maternal inheritance of the deletion was correlated with greater reduction in cortical gray matter and increased severity of language-learning disabilities (Eliez, Antonarakis et al. 2001). This indicates a possible role of genomic imprinting in the variable cognitive abilities and psychiatric symptoms associated with the disorder. Though mouse studies ruled out parent-of-origin specific expression of 25 genes lying within the DGS critical region (Maynard, Meechan et al. 2006), computational predictions in our lab predicted *DGCR6*, lying within the deleted locus, to be imprinted in humans but not in mice (Luedi, Dietrich et al. 2007). This gene is believed to be required for neural crest cell migration during development (Hierck, Molin et al. 2004), and has also been linked to schizophrenia susceptibility in completely independent association studies (Liu, Heath et al. 2002). We hypothesized that some patients could be phenotypically null for *DGCR6*, due to either chromosomal deletion or epigenetic dysregulation, resulting in the variability in neurocognitive phenotypes associated with the disease.

Thus, a study was designed to compare the levels of expression of *DGCR6* and its duplicate copy *DGCR6L* between DGS patients with maternal and paternal 22q11.2 deletions, as well as with normal age and sex matched controls. Contrary to expectations based on the hemizygous state of *DGCR6* and *DGCR6L* in DGS subjects, a large number of patients showed dysregulated rather than reduced *DGCR6* and *DGCR6L* expression when compared to control individuals. There were two groups that could be clearly distinguished, one with expression levels relatively similar to the controls, and the other which exhibited gross overexpression. The majority of the maternally deleted set belonged to the latter category, indicating a significant parent-of-origin effect on the expression levels of these genes.

Moreover, the expression of *DGCR6* and *DGCR6L* was highly correlated with performance in neurocognitive tests. Overexpression led to lower scores in sustained attention tests in the DGS patients. Sustained attention in these patients is independent of their IQ, and thus is likely to be an integral part of the neurocognitive phenotype. Additionally, impaired sustained attention is a hallmark of schizophrenia, with the DGS patients showing decreased sustained attention with the onset of schizophrenia (Silver and Feldman 2005). Based on our findings, it appears that overexpression of *DGCR6/DGCR6L* is associated with the impaired sustained attention observed in this syndrome.

In order to explain this dysregulated expression pattern, we searched for differential methylation of the promoters in *DGCR6* and *DGCR6L* of the DGS patients as

well as for other differentially methylated regions (DMRs). Though there were no noticeable differences at the promoter that correlated with variable expression, we were able to identify a stretch of CG sites 1 Kb downstream of *DGCR6L* and *DGCR6* that were differentially methylated in fetal tissues. This was shown to have an allele-specific methylation pattern and was fully methylated in the sperm, suggesting that the maternal allele would be unmethylated. Furthermore, this region was highly conserved in the primates, but was not conserved in the mouse. Also pertinent to these findings is that previous studies identified polymorphisms in this region that are associated with schizophrenia (Liu, Heath et al. 2002; Zinkstok, Schmitz et al. 2008). Hence, it appears likely that this is a vital regulatory region required for modulation of gene expression when the genes *DGCR6/DGCR6L* are required in exacting amounts, probably early in development. However, it is also possible that there is an unidentified region(s) lying far from the genes themselves that directs gene expression, and whose disruption leads to the observed dysregulation.

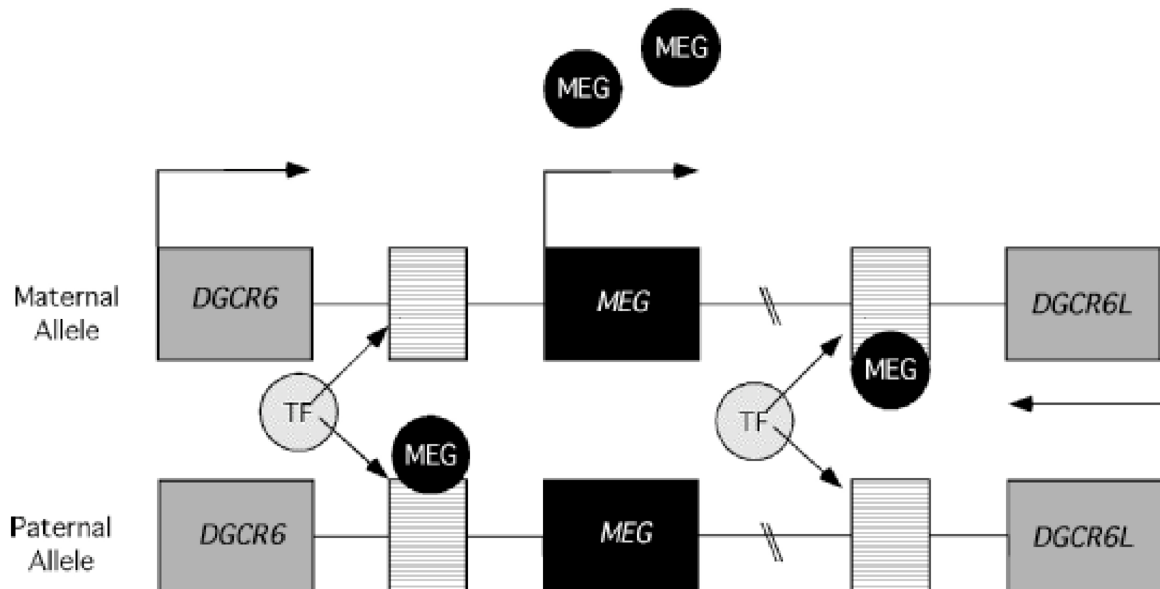
Thus, our investigations reveal that the vital genes *DGCR6* and *DGCR6L* are involved in neurocognitive development, and their deletion leads to some of the observed variability in psychosis in DiGeorge syndrome. Although there is no evidence for their being imprinted in the patient blood samples it is known that they are randomly monoallelically expressed due to the duplication of the locus (data in previous chapter). Also, a fetal DMR 1 Kb downstream of each gene probably serves as a regulatory region that helps to control their expression levels. *Cis*-acting activating transcription factors

may bind the unmethylated allele during development to direct expression, with acquisition of full methylation when the gene products are no longer necessary. These transcription factors may be in competition with repressors that mediate monoallelic gene silencing of the *DGCR6/DGCR6L* locus (**Figure 4.5**). If these repressors were produced from a maternally expressed gene lying in the DiGeorge Critical Region, deletion of specifically the maternal allele would lead to a lack of repressors and associated deregulation, as is observed in the DGS patients.

Another important outcome of our study is the observed difference in epigenetic regulation of *DGCR6* in mice and humans, which demonstrates that epigenetically regulated genes are different between these two species. There might yet be unidentified imprinted genes or other monoallelically expressed genes lying within the DGS critical region, and understanding their dosage and regulation directly in humans will be the key to understanding the role of epigenetics and genomic imprinting in complex diseases such as this one.

It is also evident from our findings that both *DGCR6* and *DGCR6L* are expressed at very low levels in adult blood; the limitation in when and where expression levels can be measured in DGS individuals means that these results may not reflect developmental events. Although it would be difficult to execute, a study of conceptuses carrying the 22q11 deletion would determine the temporal and spatial expression levels of these and other important genes during critical windows of development.

Figure 4.5: Hypothetical model for regulation at the *DGCR6* and *DGCR6L* loci



Graphical representation of the DiGeorge Critical region. *DGCR6* and *DGCR6L* are shown as grey boxes, while a Maternally expressed gene (*MEG*) residing within the DiGeorge Critical Region is depicted as a black box. The shaded box represents a differentially methylated region (DMR) of regulatory importance. Proteins are depicted as circles, the black one being the repressor produced from the MEG locus, and the light gray one representing an activating transcription factor, TF. *DGCR6* and *DGCR6L* are expressed from the maternal or paternal allele depending on the methylation state of the DMR that directs the binding of TF or MEG. In this model, *DGCR6* is shown as maternally expressed and *DGCR6L* as paternally expressed. Deletion of the maternal allele would lead to deregulation due to lack of MEG production.

In conclusion, we have identified the contribution of the *DGCR6*/*DGCR6L* locus in the variability associated with neurosis in DGS patients. The overexpression could potentially be a diagnostic tool to predict the onset of schizophrenia. Moreover, defining the exact function of these genes, as well as understanding their regulation by genetic as well as epigenetic studies would prove to be of great therapeutic value in the future.

5. Conclusions

Genomic imprinting, the epigenetic phenomenon involving parent-of-origin specific monoallelic gene expression is proposed to have evolved in placental mammals to modulate intrauterine resource allocation. In support of this hypothesis, imprinted genes have been reported in Eutherians (e.g. mice and humans). While six of these genes are also imprinted in the Metatherians (e.g. kangaroo and opossum), none of the genes investigated in the egg-laying Prototherians (e.g. platypus and echidna) are imprinted.

Selective inactivation of one of the parental alleles, at least in Eutherians, is achieved in part by parent-of-origin specific cytosine methylation at imprint regulatory regions. These "imprint control centers" are sequences that are selectively methylated on one of the two alleles, thereby affecting transcriptional regulation. The regions also appear to be associated with allele-specific activating or repressive histone modifications on the underlying chromatin. In the Metatherian lineage, however, the method of imprint regulation remains largely unknown, since regulatory features like DNA methylation have been associated with only two imprinted loci thus far. Phylogenetic insights into such mechanisms would undoubtedly help resolve the much-debated question on whether DNA methylation or chromatin modifications are the "primordial imprint mark", and whether such regulatory patterns arose in a convergent or divergent manner.

Herein I studied Metatherian orthologues of Eutherian imprinted loci in *Monodelphis domestica* (gray short-tailed opossum) due to the availability of the genomic sequence. My investigation revealed that *L3MBTL* and *HTR2A* were

monoallelically expressed; *PEG1/MEST* had one imprinted and one non-imprinted transcript, while *IMPACT, COPG2* and *PLAGL1* were not imprinted in this marsupial species. Searching for operational imprint control mechanisms, I showed an absence of differentially methylated regions (DMRs) around all of the loci investigated except for *IGF2R*, wherein the novel DMR in intron 11 had neither conservation with Eutherians – nor an associated anti-sense transcript needed for imprint regulation of this gene in mice. Further investigation, however, revealed that the promoter regions of all imprinted loci were enriched for the activating histone modification H3 Lysine 4 dimethylation, like in Eutherians. Thus, it is likely that histone modifications help establish imprinted gene expression in Metatherians. Nonetheless, experiments to prove that these marks are necessary and sufficient for imprinting and derived in the germ cells are required before a conclusive role in establishment of imprinting can be attributed to them. Hence, genomic imprinting is conserved only in a fraction of loci in the two Therian infra-classes. Additionally, the regulatory marks are not fully conserved between Metatherian and Eutherian mammals. Thus, our results are consistent with imprinted gene evolution being both divergent and convergent in Therian mammals. Some genes (such as *IGF2*) were imprinted in a common ancestor to marsupials and Therian mammals, while others like *IGF2R, HTR2A* and *L3MBTL* probably evolved imprinting separately in the two lineages.

The functional haploidy rendered by imprinting makes these genes more susceptible to the deleterious effects of recessive mutations. Thus, proper elucidation of such loci is crucial to understanding the etiology of complex disorders with parental

effects resulting in epigenetic dysregulation. Machine-learning algorithms developed in the Jirtle laboratory predicted the gene *DGCR6* as being imprinted in humans, and it lies within a deletion at 22q11.2 commonly found in DiGeorge Syndrome patients. DiGeorge Syndrome (DGS) is a congenital anomaly where affected individuals show developmental defects of the face, heart and limbs, as well as psychiatric problems. It has been reported that patients with a maternally inherited deletion have a reduction in cortical gray matter. This indicates a possible role of genomic imprinting in the variable cognitive abilities and psychiatric symptoms associated with the disorder. Interestingly, *DGCR6* is duplicated in the primate lineage. This may have led to dosage compensation in the form of random monoallelic expression or genomic imprinting. Thus, I chose to investigate the expression pattern of the genes *DGCR6* and *DGCR6L* in humans and our nearest evolutionary neighbor, the chimpanzee. My investigation revealed that the genes were monoallelically expressed in humans and chimpanzees, but the monoallelic expression is not parentally dependent. Nevertheless, this unique random monoallelic expression pattern could have important effects in modulating the disease phenotype in the hemizygously deleted DGS patients.

I determined expression patterns of *DGCR6* and *DGCR6L* in DGS patients, and how this correlated with specific neurological symptoms. Many patients showed dysregulated *DGCR6* and *DGCR6L* expression when compared to that in control individuals. The maternally deleted patient set had a much higher number of patients exhibiting aberrant overexpression. Moreover, increased expression was associated with

lower scores in psychiatric attention-span tests in these patients. This suggests that these genes are critical during brain development, and that increased expression throughout life may play a role in the neurological symptoms associated with this disorder. I searched for differentially methylated regions (DMRs) that could potentially serve to regulate expression patterns at this locus and identified two CG-rich regions that are present 1 Kb downstream of *DGCR6L* and *DGCR6L.* These are differentially methylated in fetal tissues, but that exhibit complete methylation in adult peripheral blood leucocytes.

In conclusion, my study shows that dysregulation at the *DGCR6/DGCR6L* locus contributes to the variability observed in DGS related psychosis. This has important consequences for diagnosis and treatment of schizophrenia not only in DGS patients, but also in the general population. It also draws attention to the fact that epigenetically regulated genes are different between mice and humans, and that human studies are essential to understand the role of epigenetics in modulating complex human diseases like DGS.

Appendix

Supplementary Table 2.1: SNPs identified for Imprint Analysis

Gene	Forward Primer	Reverse Primer	SNP Location- Chromosome: coordinate (Ensembl build 43)
L3MBTL	GAGACTAATCACCTCACCTCCT AGTCCTCAGCCAGCCAG	CCGAGAAGAGGGAGCC GAGAAGAATATAGGGTGAGAGTGA	1: 385,955,543 1: 385,957,311
HTR2A	GTGAGTATCATTTGTGTTCCCT GAGACACCCAGAACTGAACAC	ACTGAGTGACAGAGATAAAGCTG GCATCTGGAGGGCTTC	4: 322,102,343 4: 322,105,260
MEST	GGATCCCCTGTGTGGCTC CTCCACATTCTCTTCCAATAG	TCGTCCCCAGCCTCAC ACAGAGCAAGGAAAACGAC	8: 190,443,889 8: 190,463,734
PLAGL1	CATGAAATGATATTGTGTGGTTC	CAGACACCTAGCGAGAAAGG	2: 422,112,031
COPG2	GTGGCCATGAAACACC	AAAGCAAACAGGACAAGTATTAC	8: 190,486,069
IMPACT	AGAGACCGACTGTGGC	CTGTAAGCAGAGTGTCCAT	3: 260,030,960
IGF2R	TTCTACAGATCATAAATTAAGC	CTTTAATCATTTCCTCCC	2: 442,443,695

Note: Primers are listed in 5′ to 3′ direction

Supplementary Table 2.2: Primers used for Methylation Analysis

Gene Product	Forward Primer 1	Forward Primer 2
COPG2.3	GTTATTAGGAGGGTTAGGGTTTAGT	None
COPG2_As1	ATTAATGGTTTTAAGTTTTGGGTATGG	None
COPG2_As2	GGTTTTAAGTATTAGGGGTTTTAAGAG	None
HTR2A.1	GGTTGGAGAAGTTGTATATTTATGT	None
HTR2A.2	GTAAGGTAGAAGAGAAGAAGTAAGTT	None
IGF2R_A.3	GTGGGAAGTTGAGGGTTTTT	None
IGF2R_A_As1	TGGTTATAGGGATAAGGTTAGGTAAAA	None
IGF2R_A_As2	GGAAGTTTATGGATTAGGGGTTATTAG	None
IGF2R_A_As4	TTTAGGAGTTGTGTAGGTAAGGAAGTG	None
IGF2R_B.1	GTTTATTTGTATATGGAAAGGAGAA	None
IGF2R_B.2	GTTTTTTTGGATTTTGAAGATTTGT	None
IGF2R_B_As1	GGATTATTTTTAGGGAGGAAATGATTA	TGGAAGAATTTTATGGTATTTTTGAAG
IMPACT_A.1	TAAGATTAGATTGGATTTTAAGTTG	None
IMPACT_A.2	AGTTAGAGAGGGTTGATATTTGTTT	None
IMPACT_A.3	ATGTGATTAGTTTTATTTTGGATTT	None
IMPACT_B.2	AAAAGTAGAGGTGGAGGGTTGTATAG	None
IMPACT_B_As1	AGTTGTATTTAAAAGTGTGTTTTAGGATT	TTTATTTGTGGTAGTTGATATAAATTTTG
L3MBTL_A_As1	ATGTAGGTTAGAGAGGGGTTTTGTTAT	None
L3MBTL_B_As1	AATAATTAGGTTTATTTGGAGGGAATG	TTATGGGTTTATTAGTTATGGGTTGAG
L3MBTL_B_As2	GATTGATTCGTAGGTTGTTTTTAGGTA	None
L3MBTL_C.1	TTTTATAGTTTTTGTAATGGGAGTT	None
L3MBTL_C_As1	TGATTAGGTTAGGGATTTTAGGAGTTT	GTAGAGATGGGGAGATTTTAGAGGTAG
L3MBTL_C_As2	ATTTATGGGTTATTGAAGGGTTTTATT	None
MEST_A_As2	TGTAGTTTAATGGTTGTTAGGGGTAAT	None
MESTB_As1	TTTTGTATATAATGAGGTAAAATTTGTGAA	None
MESTB_As2	GGTTGGGGTATGGATAGTTTTGG	None
MethControl.2	TAAAAGGATTGGAAGTTATGTTATTGG	ATATTTATAATTGGAGGGTTGGAGAAA
PLAGL1_As1	AGTAGTGTTGTTAATAAATTGAAGGATTT	GAAAGTATGTTTTTATTGGTTTTTGGA
PLAGL1_As2	TGTTGGGTTGTTGTTTTATTTAGGTAT	None
Unmeth_control_As1	GGGAAGGGAAGTGTTTAGTTTTT	None
Unmeth_control_As2	GGAAGGAGTTTAGAGTAGTTTTTAGTTGG	None

Note: Primers are listed in 5′ to 3′ direction

Supplementary Table 2.2: Primers used for Methylation Analysis (continued)

	Reverse Primer 1	Reverse Primer 2
COPG2.3	CTAAAAATAACCTAATTTCCAAAAA	None
COPG2_As1	CTCTTAAAACCCCTAATACTTAAAACC	None
COPG2_As2	AACTAAAATCCCTCCTCCTCCACTAAT	None
DIO3.1	AAAAATAACTTAACGAAACCCTAA	None
HTR2A.1	AACTTACTTCTTCTCTTCTACCTTAC	None
HTR2A.2	AATTAACCACACTCGAAATACTAAT	None
IGF2R_A.3	ACATCTTCCTTCTCCAAACTTAC	None
IGF2R_A_As1	CTAATAACCCCTAATCCATAAACTTCC	None
IGF2R_A_As2	CTTCTCCTCCCCACTTCTCTATC	None
IGF2R_A_As4	CCCACTCCAAACCAAATAAAAA	None
IGF2R_B.1	ACAAATCTTCAAAATCCAAAAA	None
IGF2R_B.2	CTTCAAAAATACCATAAAATTCTTC	None
IGF2R_B_As1	CAATCTCTACCATTCCCTCTAATTCTA	None
IMPACT_A.1	TCATTCTAAAAAAAATCCTCCATAC	None
IMPACT_A.2	AAATCCAAAATAAAACTAATCACAT	None
IMPACT_A.3	AAACAAAATACCTCTTAATCAAAAA	None
IMPACT_B.2	CAACATTCTCTCCACTTAAAAACTC	None
IMPACT_B_As1	AAAACTATACAACCCTCCACCTCTACT	None
L3MBTL_A_As1	AAATTCCCTTTTCCTAAAATCCTTTAT	ATTCCCCTAAACTTAAATCAACATTTC
L3MBTL_B_As1	TACCTAAAAACAACCTACGAATCAATC	None
L3MBTL_B_As2	AACCCTAACCTAAAATAACCAAACTCC	AAACTCCTAAAATCCCTAACCTAATCA
L3MBTL_C.1	TTCTTACATCCCTAACTCTACCTAA	None
L3MBTL_C_As1	AATAAAACCCTTCAATAACCCATAAAT	None
L3MBTL_C_As2	CAAATCATTTAAACTCCACAATAACCT	CTTTACCTCCTTTCCTAATCCTTAAAC
MEST_A_As2	ACCTTATATTATCCATTTCCCAAAACT	CTAAAAACACTCCCCACTCCATTAC
MESTB_As1	CCAAAACTATCCATACCCCAACC	None
MESTB_As2	CAAATAAACAAAAACAACCAATAAACA	None
MethControl.2	TTACAATAACATTTACAATAACCTACATCA	None
PLAGL1_As1	ATACCTAAATAAAACAACAACCCAACA	None
PLAGL1_As2	TTTCTTTTCTCTCTACAAACAAAATCT	AACTATCAAACACCAACCAAAAATTAC
Unmeth_control_As1	CCCAACTAAAAACTACTCTAAACTCCTT	None
Unmeth_control_As2	AACCTAAAACTTCTCTACCCCTAAAAA	None

Supplementary Table 2.3: Genomic Co-ordinates of Primers used for Methylation Analysis

Summary of Gene Products	Genomic Start	Finish	Size
COPG2.3	186437881	186438252	371
COPG2_As1	186436617	186436874	257
COPG2_As2	186436847	186437315	468
HTR2A.1	312085321	312085795	474
HTR2A.2	312085769	312086146	377
IGF2R_A.3	442406608	442407048	440
IGF2R_A_As1	442405350	442405645	295
IGF2R_A_As2	442405618	442405948	330
IGF2R_A_As4	442406351	442406648	297
IGF2R_B.1	442481958	442482459	501
IGF2R_B.2	442482434	442482929	495
IGF2R_B_As1	442482902	442483403	501
IMPACT_A.1	260054981	260055316	335
IMPACT_A.2	260055322	260055856	534
IMPACT_A.3	260055831	260056390	559
IMPACT_B.2	260031046	260031493	447
IMPACT_B_As1	260030640	260031076	436
L3MBTL_A_As1	385955382	385955901	519
L3MBTL_B_As1	385956549	385956980	431
L3MBTL_B_As2	385956953	385957398	445
L3MBTL_C.1	385957265	385957551	286
L3MBTL_C_As1	385957384	385957831	447
L3MBTL_C_As2	385957804	385958110	306
MEST_A_As2	186358831	186359320	489
MESTB_As1	186360936	186361329	393
MESTB_As2	186361307	186361718	411
MethControl.2	126702427	126702820	393
PLAGL1_As1	422128619	422129116	497
PLAGL1_As2	422129089	422129421	332
Unmeth_control_As1	103840827	103841343	516
Unmeth_control_As2	103841314	103841552	238

<u>Supplementary Table 2.4</u>: Primers used for *IGF2R* Anti-sense Detection

Strand-specific RT-primer	Forward Primer	Reverse Primer	Product Size
1) 1a: CATCAATATCTGTCACAAAG 1b: AACGCTTAGATTTGGCTT	GGTGAGTATAACATCCAGAGAA	GCCTCAGTCTAAGAAGTTCTC	180
2) 2a: GATGTTGTCAGAAGGTTGT 2b: GTCGGTTTCAACAATCC	ATGTGATACACCGTAGTCG	ATGAGTATGTGTTTACATGTACG	476
3) 3a: CTAGGCAAGTTTGTGTCTTC 3b: AACCTTTTATTCATCCTGTT	GTGTCAAGTATTTGAAGGTAGGA	GCTTTAAGACACATCAAAAGGAA	242

Note: Primers are listed in 5′ to 3′ direction

<u>Supplementary Table 2.5</u>: Primers Used for Analysis of Histone Modifications

Gene Product	Forward Primer	Reverse Primer	Distance from Transcription Start Site
HTR2Ap6	GACATTACCTAGTCCACTCTCCT	AGAAGCTGAGGCAAACAGA	-2600
HTR2Ap3	CAAATTACCATCTGATTGTGTG	TGAAAACTTACATGAGGCTTCA	-1000
HTR2Ap1	TGTGGGTGACACTGGATA	ACCTATTTTACTCAGCATGTGT	-300
HTR2Ain1	GTGAGTATCATTTGTGTTCCCT	ACTGAGTGACAGAGATAAAGCTG	+400
L3MBTLp6	CCTTATCTTCCATCCAAGAATC	AGGTCCTGAGTTCAAATTTGAC	-2700
L3MBTLp5	GATGAAGGTCACACAACCA	GCTGCATGACAAGGAAAG	-2000
L3MBTLp4	CCATTTGCCTCCTATGTCTT	ATTTGTACATGCCAGTGGA	-1400
L3MBTLp3	TGTGACAGTCTCTCCAGCTC	CTGATGGTGGTGGTGTTG	-900
L3MBTLp1	CATTTAGACTCCACAATGACCT	CAGTCCTATCCAGCTCAAGG	-350
L3MBTL-1	CTTGTGTGCAGGTGGTTG	TGGCTGGCTGAGGACTT	0
IGF2RAqPCR-1	CCAGAGAGATTAAATGACTTGC	GAAGTTTCTGAAAGCCTACCTT	-40000
IGF2RAqPCR	GCTGCTGTGTGTGTAGTTCCT	TGCCTCTCCTTTCCTTCTT	-38500
IGF2RAqPCR-2	TATGGTATCTCCTCACAAGTGG	CACACAGTAGGTACTAGCAATGG	-37000
IGF2RqPCRp2	GAGAAGCATCATGTAGAAATTG	ACTCTGCAATCCAGTGATACT	-1500
IGF2RqPCREx1	GATGCTGTTGATATTGAAAATC	CGAAGTATTCCGATTAAATGC	0
IGF2RqPCRin1	CCAAATGGAAGCCAGTAACT	ATGCTACAGCTCACAATGC	+1700
Albumin_qPCR	CAGATGCTATTACTGGAAAGTA	GGATTCAGCCAAGTAGAAGA	+50

Note: Primers are listed in 5′ to 3′ direction

<u>Supplementary Table 3.1:</u> Primers used for Analysis of Imprint Status of *DGCR6/DGCR6L*

Gene	Forward Primer	Reverse Primer	SNP Location-Chromosome: coordinate
Human *DGCR6*	GCTCTGCTGGGTTGGTAA	CTGGCACCCTCAGC	22:18897814 22:18897797
Human *DGCR6L*	TCTGCTGGGCTGGCAA	CACACTTGGTATCCGCAAAG	22:20303738
Chimp *DGCR6*	CTGCTGCCAAGTCCCA	CACTCAGGCCCAGGGTAG	22:17349086
Chimp *DGCR6L*	GGACTTGCCATGACAGTG	TTGGTATCCGCAAGGA	22:18806191
Macaque *DGCR6* (long)	ACTACTGAGGGGCAACGCAAT	TCCTCCTCGGGGTCTGC	Chromosome 10 contig
Macaque *DGCR6* (short)	GCAGACCCCGAGGAGGA	GAAGCTCATTGCTGCCGC	Chromosome 10 contig

Note: Primers are listed in 5′ to 3′ direction. UCSC Jan 2010 has been used.

Supplementary Table 4.1: Primers Used for Methylation Analysis of *DGCR6/DGCR6L*

Oligo Name	Forward Primer	Reverse Primer 1	Genomic Start	Size
DGCR6_DMR	TTATTGGAAATTTTTGGGAATTTTT	CAAACCCACCTAAAAAACTTAACCT	18900030	451
DGCR6_pr1	GGGTTAGAGTTTATAGGATTTGGGT	ACCACCAAAAACTCCCCAAC	18893183	269
DGCR6_pr2	TTAGATTTAGGTTTTGTGGGTTTTG	TAAAACTCATACCTTCCTACTCCCC	18893423	216
DGCR6_pr3	GGTGGGATTTTATATGGTTTTG	ACCTAAACAACTCCTTCACCAAACT	18893553	446
DGCR6L_DMR	ATTTAGGTGTGGGTTGATTTTGATA	CACTCCCTCTTTATTACCCAAAACT	20300879	282
6L_2_00	TTTGGGTAATTTTTTTATTAGGTTT	AACAAAACCCTCCTTAAAATCTCC	20307400	348
6L_2_001_	TGGGTAATTTTTTTATTAGGTTTTGTAG	ACCTAAACTAAACAAAACCCTCCTT	20307402	356
6L_2_002	ACCTAAACTAAACAAAACCCTCCTT	GGAGATTTTAAGGAGGGTTTTGTT	20307724	374
6L_2_003	TTAAGGAGGGTTTTGTTTAGTTTAGG	ACCCAAATCCTATAAACTCTAACCC	20307731	233
6L_2_004	AAGGAGGGTTTTGTTTAGTTTAGGT	CCTTAAACCCAAATCCTATAAACTC	20307733	237
6L_2_005	GGGTTAGAGTTTATAGGATTTGGGT	AAAACCTTTACAAAAACCATTCCTC	20307939	319
6L_2_006	GGGTTAGAGTTTATAGGATTTGGGT	TACTCTATAAACAAACCCCCAACAA	20307939	441
6L_2_007	TTGTTAGAGTTAGGTTAGGTGGGTG	TAACACTTCAAAAAAACCCCAATTA	20308256	342
6L_2_008	TTGTTAGAGTTAGGTTAGGTGGGTG	TACTAACACTTCAAAAAAACCCCAA	20308256	345
6L_2_009	TTGTTAGAGTTAGGTTAGGTGGGTG	TTACTAACACTTCAAAAAAACCCCA	20308256	346
6L_2_012	GTGTTTTAGATTTAGGTTTTGTGGG	CTCCAAAAAAAATTTTACTCAACCC	20307729	244
6L_2_013	TTGTTTTGATTTAGAGGTTTTAGGG	AAAAAAATTTTACTCAACCCAAACTC	20307734	337
6L_2_014	TGTTTTTTGTTGTTTTATGGATAAGTT	CAAAACCCACAAAACCTAAATCTAA	20307943	447
6L_2_015	GATTTTGGTAGGGTTTTTGTAGGAG	CACAAAACCTAAATCTAAAACACCA	20307950	317
6L_2_016	ATTTTGGTAGGGTTTTTGTAGGAGT	CCCTAAAACCTCTAAATCAAAACAA	20308046	220
6L_2_017	TTTGTTTGGAAGGGATTATAAGGAT	CCCTACCAAAATCAAATCAAATAAA	20308254	439

Note: Primers are listed in 5′ to 3′ direction

<u>Supplementary Table 4.2</u>: Complete set of Correlations between *DGCR6/DGCR6L* Gene Expression and Performance in Neurocognitive Tests

		DGS *DGCR6*	DGS *DGCR6L*	Control *DGCR6*	Control *DGCR6L*
GAS General Assessment	Pearson Correlation	-.082	.005	-.334	-.447
	Sig. (2-tailed)	.641	.976	.346	.195
	N	35	35	10	10
DISC Any Anxiety Disorder	Pearson Correlation	-.344*	-.255	.[a]	.[a]
	Sig. (2-tailed)	.043	.139	.000	.000
	N	35	35	9	9
DISC any depressive diagnosis	Pearson Correlation	-.003	-.095	.[a]	.[a]
	Sig. (2-tailed)	.988	.587	.000	.000
	N	35	35	9	9
DISC AD/HD	Pearson Correlation	.222	.065	.225	.319
	Sig. (2-tailed)	.201	.709	.560	.403
	N	35	35	9	9
DISC Inattention Symptoms	Pearson Correlation	.139	-.037	.295	.327
	Sig. (2-tailed)	.426	.833	.440	.390
	N	35	35	9	9
DISC Hyperactive Symptoms	Pearson Correlation	.145	.115	.535	.505
	Sig. (2-tailed)	.405	.511	.138	.166
	N	35	35	9	9
DISC ODD	Pearson Correlation	.100	.024	.560	.693*
	Sig. (2-tailed)	.567	.893	.117	.038
	N	35	35	9	9
DISC any diagnosis	Pearson Correlation	-.001	-.002	.225	.319
	Sig. (2-tailed)	.995	.993	.560	.403
	N	35	35	9	9

<u>Supplementary Table 4.2</u>: Complete set of Correlations between *DGCR6/ DGCR6L* Gene Expression and Performance in Neurocognitive Tests (continued)

WISC Full Scale IQ	Pearson Correlation	-.158	-.127	-.022	-.043
	Sig. (2-tailed)	.357	.461	.952	.907
	N	36	36	10	10
WISC Verbal Comprension factor	Pearson Correlation	-.363*	-.335*	.172	.103
	Sig. (2-tailed)	.032	.049	.635	.776
	N	35	35	10	10
WISC Perceptual Organization factor	Pearson Correlation	-.102	-.183	-.015	.003
	Sig. (2-tailed)	.559	.293	.966	.994
	N	35	35	10	10
WISC Working Memory	Pearson Correlation	.027	.174	.215	.127
	Sig. (2-tailed)	.879	.325	.550	.727
	N	34	34	10	10
WISC Processing Speed factor	Pearson Correlation	-.044	-.029	-.487	-.490
	Sig. (2-tailed)	.807	.869	.154	.151
	N	34	34	10	10
WIAT Broad Reading	Pearson Correlation	.009	.040	.394	.401
	Sig. (2-tailed)	.961	.820	.259	.251
	N	35	35	10	10
WIAT Broad Mathematics	Pearson Correlation	-.025	-.008	.096	.195
	Sig. (2-tailed)	.888	.964	.791	.589
	N	35	35	10	10
WIAT Spelling	Pearson Correlation	-.142	-.045	.374	.434
	Sig. (2-tailed)	.409	.793	.287	.210
	N	36	36	10	10
WCST_PE	Pearson Correlation	.096	.091	.106	-.176
	Sig. (2-tailed)	.577	.596	.719	.547
	N	36	36	14	14

<u>Supplementary Table 4.2</u>: Complete set of Correlations between *DGCR6/ DGCR6L* Gene Expression and Performance in Neurocognitive Tests (continued)

WCST_NPE	Pearson Correlation	-.190	-.284	.094	-.146
	Sig. (2-tailed)	.268	.093	.749	.618
	N	36	36	14	14
WCS_PCLR	Pearson Correlation	.002	-.058	.024	-.274
	Sig. (2-tailed)	.993	.737	.935	.344
	N	36	36	14	14
WCST_CC	Pearson Correlation	-.177	-.277	.082	-.291
	Sig. (2-tailed)	.300	.102	.781	.313
	N	36	36	14	14
CPT AX	Pearson Correlation	-.305	-.397*	-.083	-.323
	Sig. (2-tailed)	.063	.014	.808	.333
	N	38	38	11	11
CPT IP	Pearson Correlation	-.313	-.356*	-.177	-.469
	Sig. (2-tailed)	.056	.028	.583	.124
	N	38	38	12	12
cvlta15z	Pearson Correlation	.014	-.101	.086	.075
	Sig. (2-tailed)	.934	.564	.827	.848
	N	35	35	9	9
Parent CBCL Activities t-score	Pearson Correlation	-.267	-.155	.065	.263
	Sig. (2-tailed)	.121	.373	.825	.364
	N	35	35	14	14
Parent CBCL Social t-score	Pearson Correlation	-.159	-.159	-.116	-.006
	Sig. (2-tailed)	.360	.361	.692	.983
	N	35	35	14	14

<u>Supplementary Table 4.2</u>: Complete set of Correlations between *DGCR6/ DGCR6L* Gene Expression and Performance in Neurocognitive Tests (continued)

Parent CBCL Attention (VI) t-score	Pearson Correlation	.108	-.159	.163	.271
	Sig. (2-tailed)	.537	.363	.578	.349
	N	35	35	14	14
Parent CBCL Delinquent (7) t-score	Pearson Correlation	.013	-.251	.534*	.502
	Sig. (2-tailed)	.941	.145	.049	.067
	N	35	35	14	14
Parent CBCL Aggreesive (8) t-score	Pearson Correlation	.129	-.171	.341	.350
	Sig. (2-tailed)	.461	.325	.232	.219
	N	35	35	14	14
Parent CBCL Internalizing t-score	Pearson Correlation	-.116	-.326	.056	-.038
	Sig. (2-tailed)	.506	.056	.850	.896
	N	35	35	14	14
Parent CBCL Externalizing t-score	Pearson Correlation	.228	-.089	.309	.216
	Sig. (2-tailed)	.188	.613	.283	.459
	N	35	35	14	14
Parent CBCL Problem total t-score	Pearson Correlation	.225	-.070	.173	.079
	Sig. (2-tailed)	.194	.690	.554	.788
	N	35	35	14	14
Parent SSRS Skills total standard score	Pearson Correlation	.118	.179	-.282	-.129
	Sig. (2-tailed)	.513	.318	.499	.760
	N	33	33	8	8
Parent SSRS EIH standard score	Pearson Correlation	.101	-.064	.181	.259
	Sig. (2-tailed)	.577	.725	.668	.536
	N	33	33	8	8

Supplementary Table 4.2: Complete set of Correlations between *DGCR6/DGCR6L* Gene Expression and Performance in Neurocognitive Tests (continued)

Parent CBCL School t-score	Pearson Correlation	-.505**	-.098	-.091	-.338
	Sig. (2-tailed)	.002	.583	.769	.259
	N	34	34	13	13
Parent CBCL Competency total t-score	Pearson Correlation	-.314	-.168	.014	.039
	Sig. (2-tailed)	.070	.342	.964	.899
	N	34	34	13	13
Parent CBCL Anxious/Depressed (III) t-score	Pearson Correlation	-.019	-.210	.229	.087
	Sig. (2-tailed)	.915	.226	.431	.768
	N	35	35	14	14
Parent CBCL Withdrawn (I) t-score	Pearson Correlation	-.101	-.149	.005	-.171
	Sig. (2-tailed)	.564	.394	.986	.560
	N	35	35	14	14
Parent CBCL Somatic (II) t-score	Pearson Correlation	.057	-.263	-.094	-.076
	Sig. (2-tailed)	.746	.127	.750	.796
	N	35	35	14	14
Parent CBCL Social (IV) t-score	Pearson Correlation	.196	.091	.439	.193
	Sig. (2-tailed)	.259	.604	.117	.509
	N	35	35	14	14
Parent CBCL Thought Problems (V) t-score	Pearson Correlation	.079	-.109	.287	.221
	Sig. (2-tailed)	.654	.532	.320	.447
	N	35	35	14	14

Bibliography

Ager, E., S. Suzuki, et al. (2007). "Insulin is imprinted in the placenta of the marsupial, Macropus eugenii." Dev Biol **309**(2): 317-328.

Ager, E. I., A. J. Pask, et al. (2008). "Evolution of the CDKN1C-KCNQ1 imprinted domain." BMC Evol Biol **8**: 163.

Arnaud, P. and R. Feil (2005). "Epigenetic deregulation of genomic imprinting in human disorders and following assisted reproduction." Birth Defects Res C Embryo Today **75**(2): 81-97.

Badcock, C. and B. Crespi (2006). "Imbalanced genomic imprinting in brain development: an evolutionary basis for the aetiology of autism." J Evol Biol **19**(4): 1007-1032.

Badcock, C. and B. Crespi (2008). "Battle of the sexes may set the brain." Nature **454**(7208): 1054-1055.

Badcock, C. R. (2009). The imprinted brain : how genes set the balance between autism and psychosis. London ; Philadelphia, Jessica Kingsley Publishers.

Barbulescu, M., G. Turner, et al. (1999). "Many human endogenous retrovirus K (HERV-K) proviruses are unique to humans." Curr Biol **9**(16): 861-868.

Barel, O., S. A. Shalev, et al. (2008). "Maternally inherited Birk Barel mental retardation dysmorphism syndrome caused by a mutation in the genomically imprinted potassium channel KCNK9." Am J Hum Genet **83**(2): 193-199.

Barlow, D. P., R. Stoger, et al. (1991). "The mouse insulin-like growth factor type-2 receptor is imprinted and closely linked to the Tme locus." Nature **349**(6304): 84-87.

Barton, S. C., M. A. Surani, et al. (1984). "Role of paternal and maternal genomes in mouse development." Nature **311**(5984): 374-376.

Bassett, A. S., E. W. Chow, et al. (2005). "Clinical features of 78 adults with 22q11 Deletion Syndrome." Am J Med Genet A **138**(4): 307-313.

Blagitko, N., U. Schulz, et al. (1999). "gamma2-COP, a novel imprinted gene on chromosome 7q32, defines a new imprinting cluster in the human genome." Hum Mol Genet **8**(13): 2387-2396.

Blaxill, M. F. (2004). "What's going on? The question of time trends in autism." Public Health Rep **119**(6): 536-551.

Chelune, G. J. and R. A. Baer (1986). "Developmental norms for the Wisconsin Card Sorting test." J Clin Exp Neuropsychol **8**(3): 219-228.

Cornblatt, B. A. and L. Erlenmeyer-Kimling (1985). "Global attentional deviance as a marker of risk for schizophrenia: specificity and predictive validity." J Abnorm Psychol **94**(4): 470-486.

Das, R., D. D. Hampton, et al. (2009). "Imprinting evolution and human health." Mamm Genome **20**(9-10): 563-572.

De Luca, V., O. Likhodi, et al. (2007). "Parent-of-origin effect and genomic imprinting of the HTR2A receptor gene T102C polymorphism in psychosis." Psychiatry Res **151**(3): 243-248.

DeChiara, T. M., E. J. Robertson, et al. (1991). "Parental imprinting of the mouse insulin-like growth factor II gene." Cell **64**(4): 849-859.

Der, G., S. Gupta, et al. (1990). "Is schizophrenia disappearing?" Lancet **335**(8688): 513-516.

Ding, F., C. Patel, et al. (2003). "Conservation of Dnmt1o cytosine methyltransferase in the marsupial Monodelphis domestica." Genesis **36**(4): 209-213.

Edelmann, L., P. Stankiewicz, et al. (2001). "Two functional copies of the DGCR6 gene are present on human chromosome 22q11 due to a duplication of an ancestral locus." Genome Res **11**(2): 208-217.

Edwards, C. A., A. J. Mungall, et al. (2008). "The evolution of the DLK1-DIO3 imprinted domain in mammals." PLoS Biol **6**(6): e135.

Edwards, C. A., W. Rens, et al. (2007). "The evolution of imprinting: chromosomal mapping of orthologues of mammalian imprinted domains in monotreme and marsupial mammals." BMC Evol Biol **7**: 157.

Eggermann, T., K. Eggermann, et al. (2008). "Growth retardation versus overgrowth: Silver-Russell syndrome is genetically opposite to Beckwith-Wiedemann syndrome." Trends Genet **24**(4): 195-204.

Ehrich, M., J. K. Field, et al. (2006). "Cytosine methylation profiles as a molecular marker in non-small cell lung cancer." Cancer Res **66**(22): 10911-10918.

Eliez, S., S. E. Antonarakis, et al. (2001). "Parental origin of the deletion 22q11.2 and brain development in velocardiofacial syndrome: a preliminary study." Arch Gen Psychiatry **58**(1): 64-68.

Erlenmeyer-Kimling, L., D. Rock, et al. (2000). "Attention, memory, and motor skills as childhood predictors of schizophrenia-related psychoses: the New York High-Risk Project." Am J Psychiatry **157**(9): 1416-1422.

Esumi, S., N. Kakazu, et al. (2005). "Monoallelic yet combinatorial expression of variable exons of the protocadherin-alpha gene cluster in single neurons." Nat Genet **37**(2): 171-176.

Evans, H. K., J. R. Weidman, et al. (2005). "Comparative phylogenetic analysis of blcap/nnat reveals eutherian-specific imprinted gene." Mol Biol Evol **22**(8): 1740-1748.

Feinberg, A. P. (2005). "A genetic approach to cancer epigenetics." Cold Spring Harb Symp Quant Biol **70**: 335-341.

Feinberg, A. P. (2007). "An epigenetic approach to cancer etiology." Cancer J **13**(1): 70-74.

Feinberg, A. P., R. Ohlsson, et al. (2006). "The epigenetic progenitor origin of human cancer." Nat Rev Genet **7**(1): 21-33.

Garnier, O., S. Laoueille-Duprat, et al. (2008). "Genomic imprinting in plants." Epigenetics **3**(1): 14-20.

Gimelbrant, A., J. N. Hutchinson, et al. (2007). "Widespread monoallelic expression on human autosomes." Science **318**(5853): 1136-1140.

Gimelbrant, A. A. and A. Chess (2006). "An epigenetic state associated with areas of gene duplication." Genome Res **16**(6): 723-729.

Gimelbrant, A. A., A. W. Ensminger, et al. (2005). "Monoallelic expression and asynchronous replication of p120 catenin in mouse and human cells." J Biol Chem **280**(2): 1354-1359.

Glenn, C. C., K. A. Porter, et al. (1993). "Functional imprinting and epigenetic modification of the human SNRPN gene." Hum Mol Genet **2**(12): 2001-2005.

Grandjean, V., L. O'Neill, et al. (2001). "Relationship between DNA methylation, histone H4 acetylation and gene expression in the mouse imprinted Igf2-H19 domain." FEBS Lett **488**(3): 165-169.

Gray, T. A., L. Hernandez, et al. (2000). "The ancient source of a distinct gene family encoding proteins featuring RING and C(3)H zinc-finger motifs with abundant expression in developing brain and nervous system." Genomics **66**(1): 76-86.

Gray, T. A., M. J. Smithwick, et al. (1999). "Concerted regulation and molecular evolution of the duplicated SNRPB'/B and SNRPN loci." Nucleic Acids Res **27**(23): 4577-4584.

Haig, D. (2004). "Genomic imprinting and kinship: how good is the evidence?" Annu Rev Genet **38**: 553-585.

Hartmer, R., N. Storm, et al. (2003). "RNase T1 mediated base-specific cleavage and MALDI-TOF MS for high-throughput comparative sequence analysis." Nucleic Acids Res **31**(9): e47.

Heijmans, B. T., E. W. Tobi, et al. (2008). "Persistent epigenetic differences associated with prenatal exposure to famine in humans." Proc Natl Acad Sci U S A **105**(44): 17046-17049.

Hermann, A., H. Gowher, et al. (2004). "Biochemistry and biology of mammalian DNA methyltransferases." Cell Mol Life Sci **61**(19-20): 2571-2587.

Hierck, B. P., D. G. Molin, et al. (2004). "A chicken model for DGCR6 as a modifier gene in the DiGeorge critical region." Pediatr Res **56**(3): 440-448.

Hoffman, A. R., T. H. Vu, et al. (2000). "Mechanisms of genomic imprinting." Growth Horm IGF Res **10 Suppl A**: S18-19.

Hore, T. A., R. W. Rapkins, et al. (2007). "Construction and evolution of imprinted loci in mammals." Trends Genet **23**(9): 440-448.

Hunter, P. (2007). "The silence of genes. Is genomic imprinting the software of evolution or just a battleground for gender conflict?" EMBO Rep **8**(5): 441-443.

Jia, D., R. Z. Jurkowska, et al. (2007). "Structure of Dnmt3a bound to Dnmt3L suggests a model for de novo DNA methylation." Nature **449**(7159): 248-251.

Jiang, Y. H., J. Bressler, et al. (2004). "Epigenetics and human disease." Annu Rev Genomics Hum Genet **5**: 479-510.

Jirtle, R. L. and M. K. Skinner (2007). "Environmental epigenomics and disease susceptibility." Nat Rev Genet **8**(4): 253-262.

Kananura, C., T. Sander, et al. (2002). "Tandem pore domain K(+)-channel TASK-3 (KCNK9) and idiopathic absence epilepsies." Am J Med Genet **114**(2): 227-229.

Kaneko-Ishino, T., T. Kohda, et al. (2006). "Complementation hypothesis: the necessity of a monoallelic gene expression mechanism in mammalian development." Cytogenet Genome Res **113**(1-4): 24-30.

Karayiorgou, M., M. A. Morris, et al. (1995). "Schizophrenia susceptibility associated with interstitial deletions of chromosome 22q11." Proc Natl Acad Sci U S A **92**(17): 7612-7616.

Kaslow, D. C. and B. R. Migeon (1987). "DNA methylation stabilizes X chromosome inactivation in eutherians but not in marsupials: evidence for multistep maintenance of mammalian X dosage compensation." Proc Natl Acad Sci U S A **84**(17): 6210-6214.

Kern, R. S., M. F. Green, et al. (2004). "NIMH-MATRICS survey on assessment of neurocognition in schizophrenia." Schizophr Res **72**(1): 11-19.

Keverne, B. (2009). "Monoallelic gene expression and mammalian evolution." Bioessays **31**(12): 1318-1326.

Keverne, E. B. (2001). "Genomic imprinting and the maternal brain." Prog Brain Res **133**: 279-285.

Khosla, S., G. Mendiratta, et al. (2006). "Genomic imprinting in the mealybugs." Cytogenet Genome Res **113**(1-4): 41-52.

Killian, J. K., T. R. Buckley, et al. (2001). "Marsupials and Eutherians reunited: genetic evidence for the Theria hypothesis of mammalian evolution." Mamm Genome **12**(7): 513-517.

Killian, J. K., J. C. Byrd, et al. (2000). "M6P/IGF2R imprinting evolution in mammals." Mol Cell **5**(4): 707-716.

Killian, J. K., C. M. Nolan, et al. (2001). "Monotreme IGF2 expression and ancestral origin of genomic imprinting." J Exp Zool **291**(2): 205-212.

Killian, J. K., C. M. Nolan, et al. (2001). "Divergent evolution in M6P/IGF2R imprinting from the Jurassic to the Quaternary." Hum Mol Genet **10**(17): 1721-1728.

Kim, C. J., Y. G. Cho, et al. (2004). "Altered expression of KCNK9 in colorectal cancers." APMIS **112**(9): 588-594.

Kim, J. K., M. Samaranayake, et al. (2009). "Epigenetic mechanisms in mammals." Cell Mol Life Sci **66**(4): 596-612.

Kinugawa, K., M. Vidailhet, et al. (2009). "Myoclonus-dystonia: an update." Mov Disord **24**(4): 479-489.

Kishino, T., M. Lalande, et al. (1997). "UBE3A/E6-AP mutations cause Angelman syndrome." Nat Genet **15**(1): 70-73.

Klenova, E. M., H. C. Morse, 3rd, et al. (2002). "The novel BORIS + CTCF gene family is uniquely involved in the epigenetics of normal biology and cancer." Semin Cancer Biol **12**(5): 399-414.

Knudson, A. G. (2001). "Two genetic hits (more or less) to cancer." Nat Rev Cancer **1**(2): 157-162.

Kobayashi, S., T. Kohda, et al. (1997). "Human PEG1/MEST, an imprinted gene on chromosome 7." Hum Mol Genet **6**(5): 781-786.

Kono, T. (2006). "Genomic imprinting is a barrier to parthenogenesis in mammals." Cytogenet Genome Res **113**(1-4): 31-35.

Krueger, C. and I. M. Morison (2008). "Random monoallelic expression: making a choice." Trends Genet **24**(6): 257-259.

Lawton, B. R., B. R. Carone, et al. (2008). "Genomic imprinting of IGF2 in marsupials is methylation dependent." BMC Genomics **9**: 205.

Lee, J. T. (2003). "Molecular links between X-inactivation and autosomal imprinting: X-inactivation as a driving force for the evolution of imprinting?" Curr Biol **13**(6): R242-254.

Leung, K. N., R. O. Vallero, et al. (2009). "Imprinting regulates mammalian snoRNA-encoding chromatin decondensation and neuronal nucleolar size." Hum Mol Genet.

Li, J., A. J. Bench, et al. (2004). "Imprinting of the human L3MBTL gene, a polycomb family member located in a region of chromosome 20 deleted in human myeloid malignancies." Proc Natl Acad Sci U S A **101**(19): 7341-7346.

Liu, H., S. C. Heath, et al. (2002). "Genetic variation at the 22q11 PRODH2/DGCR6 locus presents an unusual pattern and increases susceptibility to schizophrenia." Proc Natl Acad Sci U S A **99**(6): 3717-3722.

Loebel, D. A. and P. G. Johnston (1996). "Methylation analysis of a marsupial X-linked CpG island by bisulfite genomic sequencing." Genome Res **6**(2): 114-123.

Luedi, P. P., F. S. Dietrich, et al. (2007). "Computational and experimental identification of novel human imprinted genes." Genome Res **17**(12): 1723-1730.

Luedi, P. P., A. J. Hartemink, et al. (2005). "Genome-wide prediction of imprinted murine genes." Genome Res **15**(6): 875-884.

Maynard, T. M., G. T. Haskell, et al. (2002). "22q11 DS: genomic mechanisms and gene function in DiGeorge/velocardiofacial syndrome." Int J Dev Neurosci **20**(3-5): 407-419.

Maynard, T. M., D. W. Meechan, et al. (2006). "No evidence for parental imprinting of mouse 22q11 gene orthologs." Mamm Genome **17**(8): 822-832.

McGrath, J. and D. Solter (1986). "Nucleocytoplasmic interactions in the mouse embryo." J Embryol Exp Morphol **97 Suppl**: 277-289.

McInnis, M. G., T. H. Lan, et al. (2003). "Genome-wide scan of bipolar disorder in 65 pedigrees: supportive evidence for linkage at 8q24, 18q22, 4q32, 2p12, and 13q12." Mol Psychiatry **8**(3): 288-298.

McMillen, I. C. and J. S. Robinson (2005). "Developmental origins of the metabolic syndrome: prediction, plasticity, and programming." Physiol Rev **85**(2): 571-633.

Meuth, S. G., A. M. Herrmann, et al. (2008). "The two-pore domain potassium channel TASK3 functionally impacts glioma cell death." J Neurooncol **87**(3): 263-270.

Moore, T. and D. Haig (1991). "Genomic imprinting in mammalian development: a parental tug-of-war." Trends Genet **7**(2): 45-49.

Morgan, D. K. and E. Whitelaw (2009). "The role of epigenetics in mediating environmental effects on phenotype." Nestle Nutr Workshop Ser Pediatr Program **63**: 109-117; discussion 117-109, 259-168.

Morison, I. M., J. P. Ramsay, et al. (2005). "A census of mammalian imprinting." Trends Genet **21**(8): 457-465.

Mu, D., L. Chen, et al. (2003). "Genomic amplification and oncogenic properties of the KCNK9 potassium channel gene." Cancer Cell **3**(3): 297-302.

Muller, U. (2009). "The monogenic primary dystonias." Brain **132**(Pt 8): 2005-2025.

Murphy, S. K. and R. L. Jirtle (2003). "Imprinting evolution and the price of silence." Bioessays **25**(6): 577-588.

Murphy, S. K., A. A. Wylie, et al. (2001). "Imprinting of PEG3, the human homologue of a mouse gene involved in nurturing behavior." Genomics **71**(1): 110-117.

Murrell, A., S. Heeson, et al. (2004). "An association between variants in the IGF2 gene and Beckwith-Wiedemann syndrome: interaction between genotype and epigenotype." Hum Mol Genet **13**(2): 247-255.

Nolan, C. M., J. K. Killian, et al. (2001). "Imprint status of M6P/IGF2R and IGF2 in chickens." Dev Genes Evol **211**(4): 179-183.

O'Neill, M. J., R. S. Ingram, et al. (2000). "Allelic expression of IGF2 in marsupials and birds." Dev Genes Evol **210**(1): 18-20.

O'Sullivan, F. M., S. K. Murphy, et al. (2007). "Imprinted expression of the canine IGF2R, in the absence of an anti-sense transcript or promoter methylation." Evol Dev **9**(6): 579-589.

Okamura, K., Y. Hagiwara-Takeuchi, et al. (2000). "Comparative genome analysis of the mouse imprinted gene impact and its nonimprinted human homolog IMPACT: toward the structural basis for species-specific imprinting." Genome Res **10**(12): 1878-1889.

Okamura, K. and T. Ito (2006). "Lessons from comparative analysis of species-specific imprinted genes." Cytogenet Genome Res **113**(1-4): 159-164.

Ooi, S. L. and S. Henikoff (2007). "Germline histone dynamics and epigenetics." Curr Opin Cell Biol **19**(3): 257-265.

Otto, M. W., G. E. Bruder, et al. (1994). "Norms for depressed patients for the California verbal learning test: associations with depression severity and self-report of cognitive difficulties." Arch Clin Neuropsychol **9**(1): 81-88.

Peters, J. and C. Beechey (2004). "Identification and characterisation of imprinted genes in the mouse." Brief Funct Genomic Proteomic **2**(4): 320-333.

Peters, J., S. F. Wroe, et al. (1999). "A cluster of oppositely imprinted transcripts at the Gnas locus in the distal imprinting region of mouse chromosome 2." Proc Natl Acad Sci U S A **96**(7): 3830-3835.

Petronis, A. (2001). "Human morbid genetics revisited: relevance of epigenetics." Trends Genet **17**(3): 142-146.

Rakyan, V. K., M. E. Blewitt, et al. (2002). "Metastable epialleles in mammals." Trends Genet **18**(7): 348-351.

Rapkins, R. W., T. Hore, et al. (2006). "Recent assembly of an imprinted domain from non-imprinted components." PLoS Genet **2**(10): e182.

Raslova, H., E. Komura, et al. (2004). "FLI1 monoallelic expression combined with its hemizygous loss underlies Paris-Trousseau/Jacobsen thrombopenia." J Clin Invest **114**(1): 77-84.

Reik, W. and A. Lewis (2005). "Co-evolution of X-chromosome inactivation and imprinting in mammals." Nat Rev Genet **6**(5): 403-410.

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

Reinhart, B. and J. R. Chaillet (2005). "Genomic imprinting: cis-acting sequences and regional control." Int Rev Cytol **243**: 173-213.

Renfree, M. B., E. I. Ager, et al. (2008). "Genomic imprinting in marsupial placentation." Reproduction **136**(5): 523-531.

Renfree, M. B., T. A. Hore, et al. (2009). "Evolution of genomic imprinting: insights from marsupials and monotremes." Annu Rev Genomics Hum Genet **10**: 241-262.

Ruf, N., S. Bahring, et al. (2007). "Sequence-based bioinformatic prediction and QUASEP identify genomic imprinting of the KCNK9 potassium channel gene in mouse and human." Hum Mol Genet **16**(21): 2591-2599.

Sano, Y., T. Shimada, et al. (2001). "Random monoallelic expression of three genes clustered within 60 kb of mouse t complex genomic DNA." Genome Res **11**(11): 1833-1841.

Schoenherr, C. J., J. M. Levorse, et al. (2003). "CTCF maintains differential methylation at the Igf2/H19 locus." Nat Genet **33**(1): 66-69.

Scott, R. J. and M. Spielman (2006). "Genomic imprinting in plants and mammals: how life history constrains convergence." Cytogenet Genome Res **113**(1-4): 53-67.

Shashi, V., T. R. Kwapil, et al. (2010). "Evidence of gray matter reduction and dysfunction in chromosome 22q11.2 deletion syndrome." Psychiatry Res **181**(1): 1-8.

Shibata, H., K. Yoshino, et al. (1996). "Inactive allele-specific methylation and chromatin structure of the imprinted gene U2af1-rs1 on mouse chromosome 11." Genomics **35**(1): 248-252.

Silver, H. and P. Feldman (2005). "Evidence for sustained attention and working memory in schizophrenia sharing a common mechanism." J Neuropsychiatry Clin Neurosci **17**(3): 391-398.

Sleutels, F., R. Zwart, et al. (2002). "The non-coding Air RNA is required for silencing autosomal imprinted genes." Nature **415**(6873): 810-813.

Smits, G., A. J. Mungall, et al. (2008). "Conservation of the H19 noncoding RNA and H19-IGF2 imprinting mechanism in therians." Nat Genet **40**(8): 971-976.

Smrzka, O. W., I. Fae, et al. (1995). "Conservation of a maternal-specific methylation signal at the human IGF2R locus." Hum Mol Genet **4**(10): 1945-1952.

St Clair, D., M. Xu, et al. (2005). "Rates of adult schizophrenia following prenatal exposure to the Chinese famine of 1959-1961." JAMA **294**(5): 557-562.

Surani, M. A., S. C. Barton, et al. (1984). "Development of reconstituted mouse eggs suggests imprinting of the genome during gametogenesis." Nature **308**(5959): 548-550.

Susser, E., R. Neugebauer, et al. (1996). "Schizophrenia after prenatal famine. Further evidence." Arch Gen Psychiatry **53**(1): 25-31.

Suzuki, S., R. Ono, et al. (2007). "Retrotransposon silencing by DNA methylation can drive mammalian genomic imprinting." PLoS Genet **3**(4): e55.

Suzuki, S., M. B. Renfree, et al. (2005). "Genomic imprinting of IGF2, p57(KIP2) and PEG1/MEST in a marsupial, the tammar wallaby." Mech Dev **122**(2): 213-222.

Temple, I. K. and J. P. Shield (2002). "Transient neonatal diabetes, a disorder of imprinting." J Med Genet **39**(12): 872-875.

Thompson, M. B. and B. K. Speake (2006). "A review of the evolution of viviparity in lizards: structure, function and physiology of the placenta." J Comp Physiol B **176**(3): 179-189.

Tobi, E. W., L. H. Lumey, et al. (2009). "DNA methylation differences after exposure to prenatal famine are common and timing- and sex-specific." Hum Mol Genet **18**(21): 4046-4053.

Turek-Plewa, J. and P. P. Jagodzinski (2005). "The role of mammalian DNA methyltransferases in the regulation of gene expression." Cell Mol Biol Lett **10**(4): 631-647.

Valleley, E. M., S. F. Cordery, et al. (2007). "Tissue-specific imprinting of the ZAC/PLAGL1 tumour suppressor gene results from variable utilization of monoallelic and biallelic promoters." Hum Mol Genet **16**(8): 972-981.

Vrana, P. B., X. J. Guan, et al. (1998). "Genomic imprinting is disrupted in interspecific Peromyscus hybrids." Nat Genet **20**(4): 362-365.

Vu, T. H., T. Li, et al. (2004). "Promoter-restricted histone code, not the differentially methylated DNA regions or antisense transcripts, marks the imprinting status of IGF2R in human and mouse." Hum Mol Genet **13**(19): 2233-2245.

Walter, J. and M. Paulsen (2003). "The potential role of gene duplications in the evolution of imprinting mechanisms." Hum Mol Genet **12 Spec No 2**: R215-220.

Wan, L. B. and M. S. Bartolomei (2008). "Regulation of imprinting in clusters: noncoding RNAs versus insulators." Adv Genet **61**: 207-223.

Wang, J., Z. Valo, et al. (2007). "Monoallelic expression of multiple genes in the CNS." PLoS One **2**(12): e1293.

Watanabe, D. and D. P. Barlow (1996). "Random and imprinted monoallelic expression." Genes Cells **1**(9): 795-802.

Weidman, J. R., D. C. Dolinoy, et al. (2006). "Imprinting of opossum Igf2r in the absence of differential methylation and air." Epigenetics **1**(1): 49-54.

Weidman, J. R., K. A. Maloney, et al. (2006). "Comparative phylogenetic analysis reveals multiple non-imprinted isoforms of opossum Dlk1." Mamm Genome **17**(2): 157-167.

Weidman, J. R., S. K. Murphy, et al. (2004). "Phylogenetic footprint analysis of IGF2 in extant mammals." Genome Res **14**(9): 1726-1732.

Weisstein, A. E., M. W. Feldman, et al. (2002). "Evolutionary genetic models of the ovarian time bomb hypothesis for the evolution of genomic imprinting." Genetics **162**(1): 425-439.

Wood, A. J. and R. J. Oakey (2006). "Genomic imprinting in mammals: emerging themes and established theories." PLoS Genet **2**(11): e147.

Woogh, C. (2001). "Is schizophrenia on the decline in Canada?" Can J Psychiatry **46**(1): 61-67.

Wylie, A. A., S. K. Murphy, et al. (2000). "Novel imprinted DLK1/GTL2 domain on human chromosome 14 contains motifs that mimic those implicated in IGF2/H19 regulation." Genome Res **10**(11): 1711-1718.

Yang, Y., T. Li, et al. (2003). "The histone code regulating expression of the imprinted mouse Igf2r gene." Endocrinology **144**(12): 5658-5670.

Yokomine, T., K. Hata, et al. (2006). "Evolution of the vertebrate DNMT3 gene family: a possible link between existence of DNMT3L and genomic imprinting." Cytogenet Genome Res **113**(1-4): 75-80.

Zinkstok, J., N. Schmitz, et al. (2008). "Genetic variation in COMT and PRODH is associated with brain anatomy in patients with schizophrenia." Genes Brain Behav **7**(1): 61-69.

Biography

Place and Date of Birth

Pune, Maharashtra, India

17th January 1982

Education

- 2005 to present: PhD, University Program in Genetics & Genomics, Duke University, USA
- 2005: M. Sc. Biomedical Sciences, Delhi University, India
- 2003: B. Sc. Zoology (Honors), Delhi University, India

Publications

- Dolinoy D.C., **Das R**., Weidman J.R., Jirtle R.L.., Metastable epialleles, imprinting, and the fetal origins of adult diseases. Pediatr Res, 2007. 61(5 Pt 2): p. 30R-37R.
- Mikkelsen T.S., *et al*., Genome of the marsupial Monodelphis domestica reveals innovation in non-coding sequences. Nature, 2007. 447(7141): p. 167-77.
- **Das R**., Hampton D.H., Jirtle R.L. Imprinting evolution and human health (Mammalian Genome, e-pub ahead of print)

Conferences

- The Rank Prize Funds: Mini-Symposium on Unravelling Common Epigenetic Processes Affecting Crop Quality and Human Development

Cumbria, UK; April 2009

Talk: *Role of imprinting in DiGeorge syndrome*

- EMBO Workshop on Genomic Imprinting

  Singapore, September 2008

  Talk: *Phylogenetic analysis of genomic imprinting in the marsupial Monodelphis domestica reveals incomplete conservation of Eutherian imprint regulatory features*

- Gordon Research Conference- Mechanism of Toxicity

  Portland, Maine, USA; July 2008

  Talk: *Epigenetic Gene Regulation: Linking Early Developmental Environment to Adult Disease*

- Dissecting the role of disease and environment in Complex Diseases

  Hyderabad, India; December 2007

  Talk: *Epigenetic Gene Regulation: Linking Early Developmental Environment to Adult Disease*

- Keystone Conference: Epigenetics: Regulation of Chromatin Structure in Disease and Development

  Breckenridge, USA; April 2007

  Poster: *Phylogenetic analysis of Imprinting in Marsupials reveals substantial but incomplete conservation with Eutherian-specific patterns*