2016

Yy1 Gene Dosage Effect and Allele-Specific Expression Analysis of Peg3

Bambarendage Pinithi Upekka Perera
Louisiana State University and Agricultural and Mechanical College, bperer4@gmail.com

Follow this and additional works at: https://digitalcommons.lsu.edu/gradschool_dissertations

Recommended Citation
https://digitalcommons.lsu.edu/gradschool_dissertations/3968

This Dissertation is brought to you for free and open access by the Graduate School at LSU Digital Commons. It has been accepted for inclusion in LSU Doctoral Dissertations by an authorized graduate school editor of LSU Digital Commons. For more information, please contact gradetd@lsu.edu.
YY1 GENE DOSAGE EFFECT AND ALLELE-SPECIFIC
EXPRESSION ANALYSIS OF PEG3

A Dissertation
Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

in

The Department of Biological Sciences

by
Bambarendage Pinithi Upekka Perera
B.S., University of New Orleans, 2011
August 2016
ACKNOWLEDGEMENTS

I must admit that my accomplishments in science are a result of everyone who has supported me from my childhood to graduate school, and I would like to express my heartfelt gratitude to them. First, I would like to thank my mentor and role model, Dr. Joomyeong Kim, for his scholarly advice and guidance in my career path. His hard work and dedication has taught me that science is a fine art of patience and perseverance in search for new observations.

I would like to sincerely thank my committee members Dr. David Donze and Dr. Mark Batzer for constantly encouraging me and providing me with much needed advice throughout my graduate career. I also thank my dean’s representative, Dr. Samithamby Jeyaseelan (Jey) for reviewing and evaluating my research. I would like to recognize Dr. Ryoichi Teruyama and Dr. Scott Herke for their thoughtful discussions about my research, and Dr. John Battista, Dr. Huangen Ding, and Dr. Patrick DiMario for their continuous support. I also thank Dr. William Wischusen for honing my teaching skills and inspiring me to make a difference in the community through teaching.

My research in the laboratory would not have been possible without the support and expertise from the past and present lab members: Dr. Hana Kim, Dr. Hongzhi He, Dr. Suman Lee, Dr. Muhammad Ekram, Dr. Wesley Frey, An Ye, Mohammad Faisal, Corey Bretz, Arundhati Bakshi, Austin Wong, Christopher Peck, Carter Pesson, and Terri Cain. I would like to acknowledge my fellow lab mates as well as my friends Ananya Mukherjee, Kuei-Ting Yang, Bliss Lambert, Jing Yang, Sujeet Kumar, Kyle Kern, Cody Steely, Jasmine Brown, Vallmer Jordan, Himanshu Raje, Qing Wang, Zishuo Cheng, Anuja Pande, Sonu Baral, and all of my fellow biology graduate students from the life sciences building for their encouragement in both my personal and professional life. They all have made my graduate school experience a
memorable and an enjoyable one. I thank the members from the LSU animal facility for their patience and help. I would like to recognize Ms. Priscilla Milligan, and Ms. Chimene Boyd for their help with my transition to graduate school.

Lastly I would like to thank my friends and family for their endless support through many of my obstacles in life. I especially thank my mother, M. H. Prashanthi Sobani Perera, for her hard work and persistence, into molding me to become the woman I am today. I also thank my little sister, B. Isuri Avindika Perera, for constantly reminding me to live by example. I appreciate my father, B. Arunandika Perera, for his encouragement and suggestions to study abroad. I thank my aunt, Usha DeSilva, uncle, Galappaththy DeSilva for taking care of me during my stay in the United States and for becoming my parents away from home. I thank my cousin and my best friend, Ranmali DeSilva, for her continuous encouragement, and Gayan DeSilva for his support. I also thank Ms. Amy Bruins, Mr. Randall Tanaka, Ms. Bette Franken, and Mrs. Krysten McLarty for their interest and help towards my higher education. I owe my utmost gratitude to Dr. Aaron Landry, who constantly inspire me to set higher standards for myself.
TABLE OF CONTENTS

ACKNOWLEDGEMENTS................................................................................................................................. ii

LIST OF TABLES ........................................................................................................................................ vi

LIST OF FIGURES ....................................................................................................................................... vii

LIST OF ABBREVIATIONS ............................................................................................................................ ix

ABSTRACT ...................................................................................................................................................... xi

CHAPTER 1. INTRODUCTION ........................................................................................................................ 1
  1.1 Transcriptional regulation .................................................................................................................... 1
  1.2 Genomic imprinting ............................................................................................................................... 2
  1.3 Peg3 imprinted domain ........................................................................................................................ 5
  1.4 YY1 transcription factor ...................................................................................................................... 10
  1.5 Alternative promoters and detection ................................................................................................. 12
  1.6 Statement of research objectives ...................................................................................................... 16
  1.7 References ......................................................................................................................................... 17

CHAPTER 2. YY1 DOSAGE EFFECT AND BI-ALLELIC EXPRESSION OF PEG3 ............................................ 24
  2.1 Introduction ......................................................................................................................................... 24
  2.2 Materials and Methods ....................................................................................................................... 26
  2.3 Results ................................................................................................................................................ 32
  2.4 Discussion .......................................................................................................................................... 45
  2.5 References ......................................................................................................................................... 49

CHAPTER 3. ALTERNATIVE PROMOTERS OF PEG3 WITH MATERNAL SPECIFICITY ............................. 52
  3.1 Introduction ......................................................................................................................................... 52
  3.2 Materials and Methods ....................................................................................................................... 54
  3.3 Results ................................................................................................................................................ 59
  3.4 Discussion .......................................................................................................................................... 76
  3.5 References ......................................................................................................................................... 81

CHAPTER 4. NGS-BASED 5’RACE FOR ALTERNATIVE PROMOTERS ........................................................... 84
  4.1 Introduction ......................................................................................................................................... 84
  4.2 Materials and Methods ....................................................................................................................... 86
  4.3 Results ................................................................................................................................................ 95
  4.4 Discussion .......................................................................................................................................... 97
  4.5 References ......................................................................................................................................... 98

CHAPTER 5. CONCLUSIONS AND FUTURE DIRECTIONS ............................................................................ 100
  5.1 References ......................................................................................................................................... 103
LIST OF TABLES

2.1 Primer sets used for genotyping experiments .................................................................27
2.2 Primer sets used for RT-PCR and qRT-PCR experiments ..............................................30
2.3 Summary of breeding results ..........................................................................................35
3.1 Primer sets used for genotyping, 5’RACE, RT-PCR, and DNA methylation analyses .......55
3.2 Summary of NGS-based sequencing of 5’RACE libraries from mouse Peg3 ...............61
3.3 Summary of the alternative promoter positions for DNA methylation analysis ...........71
# LIST OF FIGURES

1.1 Mechanisms of genomic imprinting .................................................................4

1.2 Genomic structure of the Peg3 imprinted domain ..............................................7

1.3 YY1 binding sites in the mouse Peg3 locus ..........................................................9

1.4 Transcriptional regulation using alternative promoters ......................................14

2.1 Breeding schemes used to characterize Yy1 roles in the Peg3 imprinted domain ......28

2.2 Genotype and weight profiles of the breeding schemes .......................................29

2.3 Genomic structures of the wild-type and mutant alleles of Peg3 and Yy1 ............32

2.4 Effects of Yy1 gene dosage on the Peg3 imprinted domain ...............................36

2.5 Effects of Yy1 gene dosage on the maternal expression of Peg3 .........................39

2.6 Maternal allele expression of Peg3 using a mouse hybrid cross ........................40

2.7 Maternal expression of Peg3 in mouse brain ......................................................41

2.8 DNA methylation analysis of Peg3 .......................................................................42

2.9 PEG3 immunoreactivity in adult mouse hypothalamus ......................................44

3.1 Genomic structures of the wild-type and mutant alleles of Peg3 .......................60

3.2 Alternative 1st exons and promoters of Peg3 ....................................................62

3.3 RT-PCR analysis of alternative exons in the mouse brain ..................................64

3.4 Alternative transcripts determined by 5’ RACE .................................................66

3.5 Allele and tissue specificity of the alternative transcripts of Peg3 ......................67

3.6 RT-PCR analysis of alternative exons in wild-type mouse tissues ......................69

3.7 DNA methylation analysis of alternative promoters of Peg3 .............................72

3.8 Upstream alternative promoter of human PEG3 ...............................................75

3.9 Evolutionary conservation of mammalian Peg3 domains ....................................77
4.1 Overall scheme for the NGS-based 5’RACE protocol ..................................................85

4.2 Alternative transcripts and promoters identified for mouse and human PEG3 .................96
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>5’RACE</td>
<td>5’ rapid amplification of cDNA ends</td>
</tr>
<tr>
<td>APeg3</td>
<td>antisense Peg3</td>
</tr>
<tr>
<td>cDNA</td>
<td>complementary DNA</td>
</tr>
<tr>
<td>CpG</td>
<td>5’ cytosine – phosphate – guanine 3’</td>
</tr>
<tr>
<td>DMR</td>
<td>differentially methylated region</td>
</tr>
<tr>
<td>ECR</td>
<td>evolutionarily conserved region</td>
</tr>
<tr>
<td>Gnas</td>
<td>gene for guanine nucleotide binding protein</td>
</tr>
<tr>
<td>H19</td>
<td>H19 imprinted maternally expressed transcript (non-protein coding)</td>
</tr>
<tr>
<td>H3K3me3</td>
<td>histone 3 lysine 4 tri-methylation</td>
</tr>
<tr>
<td>H3K9ac</td>
<td>histone 3 lysine 9 acetylation</td>
</tr>
<tr>
<td>H3K27ac</td>
<td>histone 3 lysine 27 acetylation</td>
</tr>
<tr>
<td>ICR</td>
<td>imprinting control region</td>
</tr>
<tr>
<td>Igf2</td>
<td>insulin-like growth factor 2</td>
</tr>
<tr>
<td>Igf2r</td>
<td>insulin-like growth factor 2 receptor</td>
</tr>
<tr>
<td>KAP1</td>
<td>KRAB A box-associated protein 1</td>
</tr>
<tr>
<td>IncRNA</td>
<td>long non-coding RNA</td>
</tr>
<tr>
<td>MIMT1</td>
<td>MER1 repeat containing imprinted transcript 1</td>
</tr>
<tr>
<td>NGS</td>
<td>next-generation sequencing</td>
</tr>
<tr>
<td>ORF</td>
<td>open reading frame</td>
</tr>
<tr>
<td>PCR</td>
<td>polymerase chain reaction</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>Peg3</td>
<td>paternally expressed gene 3</td>
</tr>
<tr>
<td>PHO</td>
<td>Pleiohomeotic</td>
</tr>
<tr>
<td>PRC1</td>
<td>Polycomb Repressive Complex 1</td>
</tr>
<tr>
<td>PVN</td>
<td>paraventricular nucleus</td>
</tr>
<tr>
<td>qRT-PCR</td>
<td>quantitative real-time polymerase chain reaction</td>
</tr>
<tr>
<td>RNAi</td>
<td>RNA interference</td>
</tr>
<tr>
<td>RT-PCR</td>
<td>real-time polymerase chain reaction</td>
</tr>
<tr>
<td>SETDB1</td>
<td>histone-lysine N-methyltransferase SETDB1</td>
</tr>
<tr>
<td>Snrpn</td>
<td>small nuclear ribonucleoprotein polypeptide N</td>
</tr>
<tr>
<td>SON</td>
<td>supraoptic nucleus</td>
</tr>
<tr>
<td>TdT</td>
<td>terminal deoxynucleotidyltransferase</td>
</tr>
<tr>
<td>Usp29</td>
<td>Ubiquitin-specific hydrolase 29</td>
</tr>
<tr>
<td>UTR</td>
<td>un-translated region</td>
</tr>
<tr>
<td>YY1</td>
<td>ying-yang 1</td>
</tr>
<tr>
<td>Zac1/Plag1</td>
<td>pleiomorphic adenoma gene-like 1</td>
</tr>
<tr>
<td>Zfp264</td>
<td>zinc finger protein 264</td>
</tr>
<tr>
<td>Zim1</td>
<td>zinc finger gene, imprinted 1</td>
</tr>
<tr>
<td>Zim2</td>
<td>zinc finger gene, imprinted 2</td>
</tr>
<tr>
<td>Zim3</td>
<td>zinc finger gene, imprinted 3</td>
</tr>
</tbody>
</table>
ABSTRACT

Genomic imprinting is a mechanism that targets epigenetic modifications to regulate gene transcription to express a gene from only one of its two parental alleles. Imprinted genes are typically clustered together and are involved in developmental regulation of the fetus. The paternally expressed gene 3 (Peg3) domain represents one such imprinted gene cluster involved in fetal growth regulation and maternal caring behavior. The transcription and imprinting control of the Peg3 domain requires the transcription factor Yin-Yang 1 (YY1), a protein that plays important roles throughout development.

The first part of this work explores evidence for the hypothesis that half a dosage of YY1 may be involved in controlling the transcription and imprinting of Peg3 in vivo. The results reveal that Yy1 most likely functions as a transcriptional repressor in this domain. The results also provide new evidence for bi-allelic expression of Peg3 in the mouse brain. Altogether, this indicates that the maternal allele of Peg3 is expressed and functional in specific areas of the brain, including the choroid plexus, paraventricular nucleus (PVN), and the supraoptic nucleus (SON).

The observed bi-allelic expression pattern indicates either de-repression of the maternal allele of the known promoter or the presence of alternative promoters for the Peg3 locus. Therefore, the second part of this work demonstrates that several alternative promoters exist for Peg3. The results reveal that these alternative promoters display allele-, tissue-, and developmental stage-specific expression patterns. This suggests that the activity of these alternative promoters have been functionally selected features for the Peg3 imprinted domain during mammalian evolution. The third part of this work develops a novel methodology that detects alternative promoters for Peg3 by incorporating both 5’ rapid amplification of cDNA
ends (5’RACE) and next-generation sequencing (NGS) techniques. The results indicate that this NGS-based 5’RACE protocol is a sensitive and reliable method for detecting low-abundant transcripts and promoters.

Overall, the research presented in this dissertation advances our understanding of how the YY1 transcription factor is involved in controlling the Peg3 imprinted domain and how alternative promoters may contribute to the allele-, tissue- and developmental stage-specific, Peg3 expression patterns observed in the mouse.
CHAPTER 1
INTRODUCTION

1.1 Transcriptional regulation

The central dogma of molecular biology governs the mechanisms underlying each living organism, from prokaryotes to eukaryotes [1, 2]. This includes the transcription of DNA molecules to RNA and translation of the RNA molecules to protein. In eukaryotes, transcriptional regulation is achieved through interactions between the regulatory elements that recruit proper transcription factors required for RNA polymerase II assembly and its activity. This depends on the accessibility of the RNA polymerase II to DNA, in order to generate messenger RNA molecules (mRNA) [3].

In eukaryotic gene transcription, a promoter sequence is required for RNA polymerase to initiate transcription. These regions recruit transcription factors, activators, repressors and RNA polymerase II to enable proper interactions to enable transcription. However, eukaryotic gene transcription is a complex process, which utilizes long-range DNA interactions between enhancers and promoters to initiate transcription. Gene transcription in mammals is a tightly regulated process that incorporates multiple cell-type, tissue-, and developmental stage-specific gene expression patterns with the help of alternative promoters, splicing variations and long non-coding RNA (lncRNA) [3-5]. The current study intends to gain a better understanding of this process.

Chromatin structures affecting epigenetic regulation can be observed through active and repressive histone modifications that result from histone methylation, acetylation, phosphorylation and ubiquitination [3]. For instance, H3 lysine 4 trimethylation (H3K4me3), and H3 lysine 9 acetylation (H3K9ac) encode histone marks responsible for actively transcribed
promoter regions. On the other hand, H3 lysine 9 trimethylation (H3K9me3), and H3 lysine 27 trimethylation (H3K27me3) marks are associated with constitutively and facultatively repressed genes, respectively [6, 7]. It is still unknown whether these modifications are the cause or an effect of the associated chromosomal arrangements.

Another epigenetic modification that controls gene expression is DNA methylation, where the covalent addition of a methyl group to a CG dinucleotide (CpG) occurs via de novo methyltransferases [8]. In many cases, this can either inhibit or facilitate the binding of required transcription factors to the promoter region of a given gene to regulate transcription. Moreover, DNA methylation is tightly linked to histone modifications in forming compact, heterochromatic regions. Thus, it is involved in a multitude of transcriptional processes such as genomic imprinting, X-chromosome inactivation, cancer, aging, and repression of repeat elements [8, 9].

1.2 Genomic imprinting

In sexually reproducing organisms, offspring typically inherit an equal representation of genetic material from both parents. Most genes are equally expressed from two copies, or alleles, with one allele derived from each parent. However, a subset of these autosomal genes shows expression from only one of these alleles due to a process termed genomic imprinting [9]. This process exists in higher order organisms such as plants, fungi, and animals; and in mammals it is found within two subclasses [9-11]. These include the marsupials and placental mammals, which implement genomic imprinting to regulate their complex reproductive systems to control intricate interactions between the mother and the developing fetus [12-14]. According to the parental conflict hypothesis, this embryonic regulation is a result of paternally expressed genes promoting growth and maternally expressed genes suppressing the growth of the developing
embryo [15]. Mammalian genomic imprinting plays critical roles in regulating fetal growth, development, and neurological processes such as maternal behavior [9, 12]. As a consequence, dysfunctional genomic imprinting is implicated in several human diseases, including Prader-Will, Angelman, Beckwith-Wiedemann, Silver-Russel syndromes, and autism spectrum disorders [16, 17].

The existence of genomic imprinting in mammals was first demonstrated over three decades ago by nuclear transfer technology when it failed to generate viable androgenetic embryos (having two paternal genomes) or gynogenetic embryos (having two maternal genomes) in mice [18, 19]. A typical imprinted domain contains a 3-12 imprinted genes spanning roughly 100-1700kb of genomic regions, including at least one IncRNA [9]. They are sensitive to environmental cues, which in turn result in altered cis-acting regulatory mechanisms using epigenetic factors such as DNA methylation, histone modifications, and non-coding RNA. Based on known imprinted gene clusters, a typical imprinted domain is controlled by an imprinting control region (ICR). For instance, paternally expressed genes acquire their maternally imprinted DMR in gametes during oogenesis, while the maternally expressed genes obtain their paternally imprinted DMR in gametes during spermatogenesis [20]. These DNA methylation imprints are effectively erased by demethylation of CpGs during the primordial germ cell stage (gamete production) of development [21].

This phenomenon was confirmed by the identification of two mouse gene clusters including three imprinted genes, Insulin-like growth factor 2 receptor (Igf2r), Igf2, and H19, which showed parental-specific expression patterns [22-24]. The imprinting control of these gene clusters is explained by two current models (Figure 1.1). One is the IncRNA-mediated silencing model, and the other is the insulator model. The Igf2r locus is regulated by the IncRNA-mediated
silencing model, which contains three maternally expressed genes (Slc22a2, Slc22a3, and Igf2r), and a paternally expressed Airn IncRNA (anti-sense to Igf2r) that serves as the ICR for this region [25].

![Diagram](image)

Figure 1.1. Mechanisms of genomic imprinting. The black arrows indicate the direction of transcription. The open and closed circles represent unmethylated and methylated DNA, respectively. (A) The IncRNA model illustrated by the Igf2r cluster. The methylated ICR on the maternal allele silences the Airn non-coding RNA expression, allowing RNA polymerase II to transcribe maternally expressed Igf2r, Slc22a2, and Slc22a3. When unmethylated, the ICR enables the Airn non-coding RNA to silence the maternally expressed genes. (B) The Insulator model is illustrated by the Igf2 cluster. The enhancer (E) transcribes the H19 non-coding RNA when the CTCF transcription factor occupies the unmethylated ICR on the maternal allele. The paternal allele is free of CTCF, thereby allowing E to interact with Ins and Igf2.

The methylated ICR of the maternal chromosome directly silences the Airn promoter, enabling the transcription of Slc22a2, Slc22a3, and Igf2r by recruiting RNA polymerase II [26]. On the contrary, the unmethylated ICR of the paternal chromosome activates the Airn IncRNA promoter, silencing Slc22a2, Slc22a3, and Igf2r through cis regulatory mechanisms. In the
endoderm, the \textit{Igf2} locus is regulated by the insulator model, which contains maternally expressed lncRNA \textit{H19}, paternally expressed genes \textit{Igf2} and \textit{Ins2}, and a gametic DMR serving as an ICR \cite{27}. The unmethylated ICR of the maternal chromosome interacts with a transcription factor, CTCF, which acts as an insulator to prevent the enhancer interaction with the \textit{Igf2} and \textit{Ins} genes. This enables the transcription of the \textit{H19} lncRNA. On the other hand, the methylated ICR of the paternal chromosome activates the \textit{Igf2} and \textit{Ins} genes, inactivating \textit{H19} transcription due to the lack of CTCF binding interactions \cite{28, 29}. Tight regulation of these gene clusters is required as they play essential roles in growth and development before birth \cite{9}.

\textbf{1.3 \textit{Peg3} imprinted domain}

Along with fetal growth and development, imprinted domains in the mouse also play critical roles in maternal caring behavior. This is seen from a gene cluster found in the paternally expressed gene 3 (\textit{Peg3}) imprinted domain in mouse chromosome 7. \textit{Peg3} was first discovered over two decades ago during the study of hybrid cDNA libraries from parthenogenetic embryos \cite{30}. Shortly afterwards, \textit{Peg3} was identified for its critical roles during the development and differentiation of the central nervous system and skeletal muscles \cite{31}. The human homolog of \textit{Peg3} was then discovered during gene mapping and mRNA expression analyses spanning the human chromosome 19q13.4 \cite{32}. \textit{Peg3} encodes a Krüppel-type, 2 cysteine 2 histidine (\textit{C2H2}) zinc finger protein that functions as a transcription factor \cite{30-32}. In many cases, an increase in DNA methylation, or hypermethylation, leads to a reduction in gene transcription. \textit{Peg3} is hypermethylated in breast and ovarian cancer patients, implicating that it may function as a tumor suppressor \cite{33, 34}. In addition, \textit{Peg3} also plays important roles in regulating fetal growth rates and maternal caring behavior in mice \cite{35-37}. The \textit{Peg3} gene resides within a cluster of
imprinted genes within a 500-kb domain, and is evolutionarily well conserved among all placental mammals [38]. Although Peg3 has been studied extensively for years, many of the underlying mechanisms of its gene and protein regulation are still unclear.

The imprinted domain of Peg3 includes four paternally expressed genes Peg3, Ubiquitin-specific hydrolase 29 (Usp29), Antisense Peg3 (Apeg3), Zinc finger protein 264 (Zfp264), and three maternally expressed genes Zinc finger gene, imprinted 1, 2, and 3 (Zim1, Zim2, and Zim3), as shown in Figure 1.2 [39-43]. These genes are conserved among placental mammals, with the exception of Zim1, which is found only in mammals with large litter sizes [38]. The Peg3 domain was evolutionarily dynamic in nature, as there are subtle changes between the mouse and human domains that led to significant changes in functionality. For instance, the full-length mouse Usp29 homolog in human, is separated into two genes, MER1 repeat containing imprinted transcript 1 (MIMT1) Inc-RNA, and protein coding USP29 in humans [44]. Moreover, Peg3 and Zim2 genes are fused in humans, sharing the same promoter and multiple alternative splicing variants [45]. All of these imprinted genes are clustered unevenly in the ends of the 500-kb domain, leaving a 250-kb conserved region that lacks any open reading frames (ORFs) in the center [38]. This region, however, contains multiple evolutionarily conserved regions (ECRs) that are most likely involved in controlling the Peg3 domain via cis-regulatory mechanisms [46]. For example, a series of chromosome conformation capture, and in vitro promoter assays indicated that ECR18 interacts with the Peg3 promoter, and is also associated with histone modifiers that correlate with enhancer activity. These include H3 lysine 4 mono-methylation (H3K4me1), and H3 lysine 27 acetylation (H3K27ac). The enhancer activity of this particular ECR has also been reported to act in a tissue specific manner [46]. These ECRs require further investigation for their functional significance, as some of the genomic architectures of enhancers
are shared with promoters as well [47].

(A) *Peg3* domain: mouse chromosome 7

(B) *Peg3* domain: human chromosome 19q13.4

Figure 1.2. Genomic structure of the *Peg3* imprinted domain. The pink and blue boxes denote maternally and paternally expressed genes, respectively. The grey boxes represent the differentially methylated regions (DMR) and the imprinting control region (ICR) of the domain. The two parallel lines indicate that the region is shortened for presentation. The 250-kb region within the mouse *Usp29* is conserved between mammals. The DNA lengths are not drawn to scale. The closed and open circles represent methylated and unmethylated DNA, respectively.

(A) Genomic structure of the mouse *Peg3* locus found in chromosome 7. (B) Genomic structure of the human *PEG3* locus found in chromosome 19q13.4. From mouse to human, *Usp29* diverged into *MIMT1* and *USP29*, while the *Zim2* fused to *PEG3*. The maternally expressed *ZIM2as* is an additional non-coding RNA that is added to the human *PEG3* locus through evolution.

The seven imprinted genes of the *Peg3* imprinted domain are transcribed using five promoters, two of which are bidirectional promoters shared by *Peg3/Usp29* and *Zim3/Zfp264* in mice [38]. DNA methylation analyses have identified *Peg3/Usp29*, *Zim2*, and *Zim3* promoter
regions as DMRs of this domain. However, Peg3-DMR acquires its gametic DNA methylation during oogenesis, and functions as the ICR for this domain [48-51]. This DMR spans the 4.0-kb genomic region that harbors the bidirectional Peg3/Usp29 promoter and their respective first exons (which occupies around 1.5-kb of the Peg3-DMR), as well as an unusual tandem array of repeat sequences that spans the first intron of Peg3 (which occupies the remaining 2.5-kb of Peg3-DMR) [52]. This array contains a well-conserved GGCGCCATCTTT motif, which serves as a DNA binding site for the transcription factor Ying-Yang 1 (YY1), further discussed in section 1.4. This motif is repeated 7 times in the mouse Peg3-DMR and 10 times in the human Peg3-DMR (Figure 1.3) [53, 54]. Researchers extensively use gene deletion based models to study the effect of certain genes or proteins by inactivating or disrupting it with the use of genetic and molecular mechanisms. A 2.5-kb deletion of this ICR region leads to a down-regulation of the Peg3 expression, biallelic expression of Zim2, and a switch in the dominant allele of Zfp264 from paternal to maternal allele expression [50]. This switch hints to an unknown function for the maternal allele of the Peg3-DMR, thus indicating the significance of the ICR in imprinting and transcription of the Peg3 domain [50].

Similar to many of the imprinted genes, Peg3 and Zim1 are highly expressed in the placenta, developing embryos and the hypothalamus region of the brain involved in controlling many of the maternal caring behaviors [12, 30-32, 55]. Such is the case for many of the remaining imprinted genes within the domain, where Zim2 and APEG3 are also expressed in the hypothalamus region [35, 56]. Recent gene deletion experiments involved in the Peg3 locus have indicated that offspring produced from the paternal transmission of the mutation are more susceptible to perinatal lethality. This is due to their difficulties with milk intake or suckling, compared to their wildtype littermates. Once the affected survive to adulthood, these mutant
mice still maintain a stunted growth phenotype compared to their littermates. However, these differences were not as pronounced in the maternal transmission of the mutation [35, 36]. Adult mice lacking the PEG3 protein have also indicated similar phenotypes where the paternally transmitted mutant females lack proper nest building capabilities and problems with milk letdown [37]. The in vivo functional studies on Peg3 to date are based on either a paternal or a maternal deletion of the mutant allele, assuming that the Peg3 maternal allele is methylated, and is thus functionally silent. However, the maternal allele of Peg3 may have previously undetected low levels of expression based on studies described in the following chapters.

Figure 1.3. YY1 binding sites in the mouse Peg3 locus. The maternal and paternal genomic structure of Peg3 represents the transcriptional directions for each gene. The closed and open circles indicate the methylated and unmethylated DNA in found in the ICR, respectively. There are seven YY1 binding sites spanning the ICR of the Peg3 locus. This occurs within the first intron of Peg3, and is shown by the orange triangles that interact with the YY1 transcription factor.
1.4 YY1 transcription factor

The Peg3 imprinted domain includes a conserved repeat sequence that corresponds to several Yin-Yang 1 (YY1) binding sites found near the first intron of Peg3. Thus, the transcription factor YY1 may be involved in controlling the transcription and imprinting status of the Peg3 imprinted domain. YY1 is a multifunctional protein that possesses the dynamic ability of both transcriptional activation and repression [57]. In addition, YY1 can regulate its own transcription by binding to its conserved internal DNA-binding sites found in the first intron of the Yy1 gene during transcription [58]. YY1 is an evolutionarily conserved C2H2 GLI-Krüppel-type zinc finger protein, ubiquitously expressed in vertebrates. YY1 is also an ancient protein with homologs in invertebrates, including Pleiohomeotic (PHO) protein in insects, and PHO-like (PHOL) factor in Drosophila [59-61]. The PHO protein is an important component of the polycomb-group (PcG), which is responsible for silencing of developmental genes through chromatin modulation [61]. On the other hand, Yin-Yang 2 (YY2) and reduced expression 1 (REX1) proteins show sequence homology to YY1. YY2 is similar in function to YY1, while REX1 has structural similarities as well as a few functional similarities to YY1 at its four zinc finger domains. Although YY2 and REX1 have undergone similar retro-transposition events, they have co-evolved to have unique functions in eutherian mammals when compared to YY1 [62, 63].

In 1991, YY1 was originally discovered as a transcriptional repressor that binds the promoter region (P5) of an adeno-associated virus in the absence of an adenovirus inducible element (E1A) [64]. Proteins NF-E1 (YY1), and NF-δ (the mouse ortholog of YY1) were found around the same time period due to their regulatory roles in immunoglobulin promoters and enhancers [57, 65]. Thus, three independent groups simultaneously discovered YY1 and
described similar functionality [57, 64, 65]. YY1 is also involved in the tight regulation of protein-protein interactions in cancer progression. For instance, YY1 plays an important role in controlling the p53 tumor suppressor through multiple pathways, including its interactions with the p53 antagonist, mouse double minute 2 protein (MDM2) [66]. YY1 involvement with histone acetyltransferases (P300, CBP) and histone deacetylases (HDAC1, HDAC2, HDAC3) demonstrates its connection to histone modification as well [66, 67]. Additionally, YY1 shows sensitivity to developmental and environmental indicators as it interacts with proteins involved with epigenetic modifiers such as polycomb repressive complex 1 (PRC1), K\textbf{RAB}_\textbf{A} box-associated protein 1 (KAP1), and histone-lysine N-methyltransferase SETDB1 [38, 68]. Moreover, it has been shown that YY1 may also bind to other genes using specific binding sites in a DNA methylation-sensitive manner [52]. Thus, YY1 may be involved in controlling the genes involved with genomic imprinting.

Some YY1 binding sites have been shown to cluster near ICR regions in mammals [69, 70]. This includes imprinted loci such as Peg3, Xist, Gnas, and Snrpn [69, 71]. For instance, reducing levels of YY1 protein by the use of in vitro and in vivo RNA interference (RNAi) mediated YY1 knockdown methods lead to up-regulation of Peg3 expression and altered DNA methylation levels throughout the Peg3 domain, indicating that it may be a major trans factor that regulates transcription and imprinting [54, 71, 72]. YY1 likely plays a role in the \textit{de novo} DNA methylation of the Peg3-DMR during oogenesis, as in vivo YY1 knockdown in oocytes and blastocysts indicated a loss of DNA methylation in this domain [54]. However, it is still unknown how YY1 regulates the DNA methylation of these ICR regions in these imprinted domains. Therefore, YY1 roles in these imprinted domains require further characterization, using additional in vivo knockout models with lower YY1 protein levels to confirm the data obtained.
from the RNAi-mediated knockdown experiments. Targeted disruption of the \( Yy1 \) gene in mice shows that homozygous mutants of \( Yy1 \) (\( Yy1^{-/-} \)) embryos do not survive beyond early embryogenesis during implantation, which suggests that YY1 has a strong connection to cell proliferation and differentiation [73]. The role of YY1 in the Peg3 imprinted domain may be further examined by utilizing a heterozygous mouse model with a mutation at the \( Yy1 \) locus.

1.5 Alternative promoters and detection

In eukaryotic organisms, gene transcription is a complex process that requires careful assembly of multiple \( trans \)-acting factors as well as \( cis \)-acting DNA elements such as promoters and enhancers. For instance, the Peg3 imprinted domain includes \( trans \)-acting factors such as YY1, and \( cis \)-acting factors such as the ICR, the ECRs, and the five promoter regions. The \( cis \)-regulatory regions spanning the 200-kb evolutionarily conserved region in the Peg3 domain are hypothesized to be involved in tissue-, allele-, and developmental stage-specific expression patterns [38, 46]. In fact, these differences in expression may be responsible for the tight regulatory functions of its domain.

According to multiple high throughput and next-generation sequencing (NGS) methods, the current human genome consists of over 25,000 genes that are responsible for the production of over 100,000 protein-coding and non-coding transcripts [74]. This suggests that majority of the known genes encode for more than one transcript due to variations in transcription initiation, splicing, and polyadenylation. The initial control of gene expression is regulated by transcriptional initiation through the use of promoters [75]. A typical promoter region consists of DNA sequences that enable the recruitment and proper assembly of transcription factors and RNA polymerase II to begin transcription of the respective gene at its first exon. On the other
hand, alternative promoters form transcripts that differ in their first exon or the length of the 5’ untranslated region (5’-UTR), utilizing upstream genomic regions instead of the canonical promoter [74]. This can either generate different ORFs that result in alternative protein isoforms, or the same ORFs with a different 5’-UTR, which can influence both transcriptional and translational regulation (Figure 1.4). Many of the transcripts that originate from alternative promoters also include exon-skipping events, as these are interdependent mechanisms [4]. Further investigation of alternative promoters and splicing variants suggest their involvement in differential gene expression patterns observed in various mammalian tissues, development stages, and cell types [4]. This indicates that the misregulation of alternative transcription may result in various developmental disorders, neuropsychiatric disorders, and cancerous cell growth [4, 76, 77].

In recent studies, the spatial and temporal regulation of alternative transcription results from multiple genetic and epigenetic regulatory mechanisms that occur near the promoter regions. These include histone modifications, changes in DNA methylation, and long-range interactions between the promoter and distal elements such as enhancers that determine the accessibility to RNA polymerase II [47, 74]. For instance, strong alternative promoters often acquire H3K4me3, while developmentally regulated alternative promoters include H3K4me3 as well as H3K27me3 [78]. Moreover, a combination of H3K4me1/H3K4me2, H3K27ac are responsible for enhancer-like chromatin signatures that are involved in recruiting the transcriptional machinery and RNA polymerase II [47, 79]. Thus, it is important to note that the histone architecture where promoters, alternative promoters, and enhancers are localized, share similar modifications, suggesting that some enhancers may also promote a very low level of transcription [47].
Figure 1.4. Transcriptional regulation using alternative promoters. The blue rectangles and circles represent exons and protein of a given gene. The green rectangles and pentagons represent alternative exons and protein isoforms of the same gene. H3K4me3 histone modification is typically found at the +1 region of the DNA. (A) A typical promoter region consists of DNA sequences that recruit RNA polymerase II to transcribe mRNA, which is then spliced and processed before translating into protein. (B) An alternative promoter is usually found upstream of a typical promoter, generating the same protein isoform as the typical promoter. (C) An alternative promoter could also generate a different protein isoform based on its start codon or the ORF.

These alternative transcriptional events are also present in imprinted domains, which are particularly susceptible to epigenetic changes. For example, the small nuclear ribonucleoprotein polypeptide N (SNRPN) gene locus is an imprinted domain associated with Prader-Willi syndrome and Angelman syndrome. The mouse and human SNRPN genomic region span a 500-kb domain, a length that dwarfs its 20-kb protein-coding region. The remaining genomic region
lacks any ORFs, similar to the 200-kb conserved region found in the Peg3 imprinted domain. This genomic region harbors multiple alternative promoters for SNRPN and antisense transcript of ubiquitin protein ligase E3A (UBE3A-ATS), some of which are responsible for setting germline-specific DNA methylation [80-82]. The gene for guanine nucleotide binding protein (GNAS) has a complex imprinted domain that includes maternally, paternally, and biallelically expressed transcripts that allow proper regulation in the signal transduction pathway. Among the multiple tissue and allele-specific alternative promoters found in the GNAS domain, one is responsible for establishing DNA methylation for the downstream canonical promoter in oocytes [83, 84]. Moreover, deep RNA sequencing (RNA-seq) and de novo transcriptome assembly of stage-specific mouse oocytes have indicated that many imprinted genes including pleiomorphic adenoma gene-like 1 (Plagl1/Zac1) contain alternative promoters that may be sensitive to DNA methylation [85]. In addition, the study also suggests that Peg3 may also contain a similar alternative promoter in the Usp29 region of the domain [85]. Thus, it would be of interest to investigate the existence of possible alternative promoters for Peg3, as this may uncover new functions for PEG3 that are tissue-, allele-, and developmental stage-specific. Thus, the prospect of alternative promoters for Peg3 is discussed in chapter three.

As alternative transcription is involved in critical functional roles during development, the study of specific transcriptional variants are imperative to its understanding. Currently, RNA-seq is the most common method of detection for alternative transcription, since it has the ability to reveal the number, structure and abundance of a given alternative transcript. However, it is difficult to predict the strength of alternative promoters and their proper exon structures including exon-intron junctions using this method [74]. Another method that is able to overcome these caveats is 5’ rapid amplification of cDNA ends (5’ RACE) [86]. This technique detects
full-length cDNA from either the 5’ or the 3’ end of a given gene, incorporating a cloning strategy to identify low-abundance transcripts [86]. In comparison to RNA-seq, this method is not sensitive enough to quantify a multitude of transcripts. A novel method that merges both 5’RACE and NGS techniques may solve this problem and allow detection of alternative promoters, and is discussed in the following chapters.

1.6 Statement of research objectives

The process of genomic imprinting has been extensively studied for over three decades, as imprinted genes play critical roles during embryonic development. There are still ongoing investigations to understand how these genes acquire their DNA methylation imprint during gametogenesis. Since few imprinted genes are well studied, many of them have unknown regulatory mechanisms. Peg3 is one such gene, as many aspects of its gene and protein regulation is still unclear. As suggested by previous in vitro and in vivo mouse knockdown experiments, this domain harbors several YY1 binding sites indicating that this transcription factor is likely involved in regulating Peg3 transcription. The first aim of this work is to understand how YY1 is involved in controlling the Peg3 domain in a dosage dependent manner in vivo.

Over the past two decades, Peg3 was thought to be a gene exclusively expressed from the paternal allele. Thus, many of the in vivo functional studies of Peg3 are based on either a paternal or a maternal deletion of the mutant allele, assuming that the Peg3 maternal allele is functionally silent. However, this may not be the case according to the unexpected observations made while addressing the first aim. Peg3 is bi-allelically expressed in specific areas of the mouse brain including the choroid plexus, supraoptic nucleus, and the paraventricular nucleus of
the hypothalamus. In fact, this suggests that the maternal allele of Peg3 is functional and may harbor an allele-specific alternative promoter for its transcription. The second aim of this work is to identify such alternative promoters that are responsible for maternal allele expression of Peg3 in the mouse hypothalamus.

Although there are several methods to detect alternative promoters, they are not sensitive enough to determine transcripts that are of low abundance. For instance, Peg3 is bi-allelically expressed in a small population of cells in the hypothalamus. This requires a method that is more efficient than 5’RACE and more sensitive than RNA-seq. Therefore, the third aim of this work is to introduce a novel method that incorporates an NGS-based 5’RACE method that is specifically designed to identify low-abundant transcripts initiating from alternative promoters.

1.7 References


CHAPTER 2
YY1 DOSAGE EFFECT AND BIALLELIC EXPRESSION OF PEG3

2.1 Introduction

Peg3 (Paternally Expressed Gene 3) was the first imprinted gene identified from the evolutionarily conserved 500-kb domain located in proximal mouse chromosome 7/ human chromosome 19q13.4 [1-3]. Since then, 6 additional imprinted genes have been identified from this domain, including paternally expressed Usp29, Zfp264, APeg3 and maternally expressed Zim1, Zim2, Zim3 [4, 5]. As seen in other imprinted domains, the imprinting and transcription of this imprinted domain is likely regulated through small genomic regions, termed ICRs (Imprinting Control Regions) [6-8]. One genomic region surrounding the promoters of Peg3 and Usp29, termed the Peg3-DMR (Differentially Methylated Region), has been hypothesized to be an ICR for this imprinted domain due to the following features. First, this genomic region has an unusual tandem array of YY1 binding sites [9-11]. Second, the allele-specific DNA methylation on the Peg3-DMR is set up during oogenesis and maintained throughout the lifetime of mammalian species [12-14]. These features are often associated with other ICRs, such as the ICR of H19/Igf2 [11]. A series of subsequent analyses indeed confirmed ICR roles for the Peg3-DMR and also the involvement of Yy1 in the transcription control and DNA methylation of the Peg3 domain [15-18]. In particular, the reduced levels of YY1 protein have been shown to up-regulate the expression levels of the Peg3 domain and also to change the DNA methylation levels of the Peg3-DMR [15-17]. Thus, it has been hypothesized that Yy1 is a major trans factor regulating the transcription and imprinting of the Peg3 domain [11].

The protein YY1 is a well-known DNA-binding protein with various functions [19, 20]. YY1 can function as a repressor and an activator for the transcriptional regulation of the associated genes [19, 20]. YY1 is also known to interact with many protein complexes that are involved in histone modifications [21, 22]. According to the recent studies, YY1 interacts with two major epigenetic modifiers, including PRC1 (Polycomb Repressive Complex 1) and KAP1 (KRAB A box-Associated Protein 1)/SETDB1 (histone-lysine N-methyltransferase SETDB1) [23, 24], which may provide clues regarding potential roles for YY1 in genomic imprinting. In the case of PRC1, YY1 interacts with PRC1 through YAF2, which might provide a mechanism for permanent and stable repression for the imprinted genes [23]. On the other hand, the interaction between YY1 and KAP1 has been shown to be very specific in ES cells [24]. This cell-type specific interaction might explain the prevalence of YY1 binding sites within the sequences of all the retrotransposons and endogenous retroviruses as well as potential mechanisms for the repression of these DNA elements during early embryogenesis [25, 26]. This protein complex, YY1/KAP1/SETDB1, is particularly relevant to genomic imprinting since DNA methylation on several ICRs with YY1 binding sites all occur during early embryogenesis and gametogenesis [11, 17]. Nevertheless, it is currently unknown how YY1 is involved in establishing DNA methylation on ICRs and other retrotransposons in mammalian genomes.

In the current study, therefore, we sought to characterize the in vivo roles of \( \text{Yy1} \) in the \( \text{Peg3} \) domain using various breeding schemes with a set of newly established mutant alleles of \( \text{Peg3} \). According to the results, \( \text{Yy1} \) indeed functions as a repressor for the \( \text{Peg3} \) domain. During the course of this study, we have also discovered that \( \text{Peg3} \) is expressed bi-allelically in a small subset of cells in mouse brain.
2.2 Materials and Methods

Generating the mutant strains for \textit{Yy1} and \textit{Peg3}

The current study used the following 7 mouse strains. The strain carrying a floxed allele for \textit{Yy1} was obtained from the Jackson Lab (Stock No. 014649, B6.129S4-\textit{Yy1}tm2Yshi/J; [27]). The strain for the CoKO allele of \textit{Peg3} was made using a targeted ES cell from the EUCOMM (European Conditional Mouse Mutagenesis program), and this strain has been maintained in the lab [28]. These two strains were crossed with the Zp3-Cre line from the Jackson Lab (Stock No. 003651, C57BL/6-Tg (Zp3-cre) 93Knw/J) and the Rosa26-FLP line from Jackson Lab (Stock No. 009086, B6.129S4-\textit{Gt} (ROSA)26Sor\textit{^{m1(FLP1)Dym}}/RainJ). The mutagenesis through these breeding derived the \textit{Yy1} mutant strain, and also the FlipKO and DelKO strains for the \textit{Peg3} locus. The following primer sets were used for genotyping of these strains: the deletion of exon 1 for \textit{Yy1}, YY1-CoKO-F (5’-ACCTGGTCTATCGAAAGGAAGCAC-3’) and YY1-genotype-R (5’-TCATCCAAAGTTGAAACCTGCTTTCC-3’); the presence of the expression cassette for the CoKO allele, Peg3-5arm (5’-CCCTCAGCAGAGCTGTTTCCTGCC-3’) and LAR3 (5’-CAACGGGTTCTTCTGTTAGTCC-3’); the deletion and detection of the expression cassette for FlipKO and DelKO, respectively, Peg3-5arm (5’-CCCTCAGCAGAGCTGTTTCCTGCC-3’) and LoxR (5’-TGAACCTGATGGCGAGCTCAGACC-3’); the presence of Zp3-Cre, Zp3-cre-F (5’-TAGGAATCAGGTGAGTGTACT-3’) and oIMR1085 (5’-GTGAAAACAGCATTGCTGTCCTTT-3’); the presence of Rosa26-FLP, oIMR0853 (5’-GCGAAGAGTTGTGCTCCTAACC-3’) and oIMR0852 (5’-AAAGTCGCTCTGAGTTGTTAT-3’). DNA was isolated from ear or tail snips through incubating the tissues at 65°C with the tail lysis buffer (50 mM Tris-Cl at pH 8.0, 100 mM EDTA at pH 8.0, 250 mM NaCl, 1% SDS, along with 20 µg/mL Proteinase K). PCR premix kit (Intron Biotech) was used for genotyping at the
following conditions (step 1, 95°C-30 sec; step 2, 95°C-30 sec, 60°C-30 sec, 72°C-60 sec for 33 cycles; step 3, 72°C-7 min). The information regarding individual primer sequences is also available (Table 2.1).

<table>
<thead>
<tr>
<th>Locus</th>
<th>Name</th>
<th>Sequence (5’→3’)</th>
<th>Primer set</th>
<th>Size (bp)</th>
<th>*Position (mm9, NCBI Build 37)</th>
</tr>
</thead>
<tbody>
<tr>
<td>YY1(Exon1_del)</td>
<td>YY1 CoKO F YY1 genotype R</td>
<td>ACCTGGGTATGCTGAAGAGGAAAGCAGCATCAGAAGGAAAGTCTGCTTCC</td>
<td>1st primer</td>
<td>290</td>
<td>Chr12:108,791-900-108,795,566</td>
</tr>
<tr>
<td>Zp3-cre</td>
<td>Zp3 F eIMR1085</td>
<td>TAGGAATCAGTGGAGTGCTTGAGGAACAGCATTGCTGCTACCTT</td>
<td>1st primer</td>
<td>500</td>
<td>Chr5:115,979-920-unknown</td>
</tr>
<tr>
<td>CoKO</td>
<td>Peg3-5 arm Peg3-LoxR</td>
<td>CCCCCAGAGCTGTTTCTGCCCAACGGGTCTCTGTATGCTCC</td>
<td>1st primer</td>
<td>510</td>
<td>chr7:6,630,678-6,634,857</td>
</tr>
<tr>
<td>DelKO</td>
<td>Peg3-5 arm Peg3-LoxR</td>
<td>CCCCCAGAGCTGTTTCTGCCCAACGGGTCTCTGTATGCTCC</td>
<td>1st primer</td>
<td>505</td>
<td>chr7:6,904,937-6,912,805</td>
</tr>
<tr>
<td>Y chromosome</td>
<td>mSry-F mSry-R</td>
<td>GTCCCGTGGTAGAGGCAACAAGGCAGTCTACTCCAGTCTGCC</td>
<td>1st primer</td>
<td>318</td>
<td>chrY:1,919,267-1,919,524</td>
</tr>
<tr>
<td>Flippase</td>
<td>eIMR0853-eIMR0852-R</td>
<td>GCGAAGAGTTTCTCCTCAACCAAGTCTCTAGTTGTTAT</td>
<td>1st primer</td>
<td>130</td>
<td>Chr8:113,076,214-unknown</td>
</tr>
</tbody>
</table>

**Breeding experiments**

The current study used the following three breeding schemes: Breeding I, female heterozygotes for *Yy1* × male heterozygotes for CoKO of Peg3; Breeding II, female heterozygotes for CoKO × male heterozygotes for *Yy1*; Breeding III, female heterozygotes for *Yy1* × male heterozygotes for DelKO (Figure 2.1). The health status of the pups from these breeding was monitored through measuring their birth weight (Figure 2.2 C). The gender of these pups was also determined through PCR with the following primer set: mSry-F (5’-GTCCCGTGGTAGAGGCAACAAG-3’) and mSry-R (5’-GCAGTCTACTCCAGTCTTGC-3’). All animals were kept in a temperature-controlled environment at 22°C, with 4-5 mice per cage over a 12-hour period of light/dark cycles. Litter size, genotype, birth weight and gender
were all recorded for each mating pair, which were later used to generate a graphical representation of the gender and genotype distribution for each cross (Figure 2.2 A, B).

![Breeding schemes used to characterize Yy1 roles in the Peg3 imprinted domain.](image)

**Figure 2.1.** Breeding schemes used to characterize Yy1 roles in the Peg3 imprinted domain. This study used three breeding schemes: Breeding I, female heterozygotes for Yy1 X male heterozygotes for CoKO allele of Peg3; Breeding II, female heterozygote for CoKO X male heterozygotes for Yy1; and Breeding III, female heterozygotes for Yy1 X male heterozygotes for DelKO allele of Peg3. All of the Peg3 heterozygotes used for these breeding schemes had inherited the mutant alleles maternally, CoKO (-m/+), and DelKO (-m/+). The average litter sizes of the one-day-old progeny are presented in each breeding setup.

**RNA isolation and quantitative RT-PCR analysis**

Total RNA was isolated from the brains of one-day-old neonates using a commercial kit (Trizol, Invitrogen) according to the manufacturer’s protocol. The total RNA was then reverse-transcribed using the M-MLV kit (Invitrogen), and the subsequent cDNA was used as a template for quantitative PCR. This analysis was performed with SYBR Select Master Mix (Applied Biosystems, Life Technologies) using the iCycler iQTM multicolor real-time detection system (Bio-Rad). All qRT-PCR reactions were carried out for 40 cycles under standard PCR conditions with internal controls (28S and β-actin). The results derived from qRT-PCR were further analyzed using the threshold (Ct) value.
Figure 2.2. Genotype and weight profiles of the breeding schemes. (A) Yy1 strain breeding results. Representation of the male and female Yy1 hets and WT, observed in 6 litters of adult mice. There was no significant difference observed between Yy1 hets versus WT offspring ($X^2$ test: $X^2=0.091; df=1; p=0.7630$), although a significant difference was observed between males versus females in Yy1 hets ($X^2$ test: $X^2=10.714; df=1; p=0.0011$). (B) Graphical representation of the genotype distribution (double het, Peg3 het, Yy1 het, and WT) for breeding I, II, and III corresponding to CoKO paternal transmission, CoKO maternal transmission and DelKO paternal transmission with Yy1 het, respectively. A total of 4 litters were used for this analysis consisting of approximately 31 individuals for CoKO paternal transmission, 33 individuals for CoKO maternal transmission, and 23 individuals for DelKO paternal transmission. (C) A graphical representation of the weight distribution for all four genotypes observed from the breeding schemes representing CoKO paternal transmission (blue), CoKO maternal transmission (pink), and DelKO paternal transmission (purple). The percentage of birth weight for neonate mice was calculated by comparing the individual weight at birth to the average weight of each litter for a total of 4 litters. The error bars indicate the standard deviation observed between the birth weight percentages among each genotype. CoKO paternal transmission weight comparison between double heterozygous and wild-type neonates indicate a significant difference $p=0.0121$ using the student t-test. CoKO maternal transmission weight comparison between double heterozygous and wild-type neonates indicate a significant difference $p=0.0094$ using the student t-test. DelKO paternal transmission neonate weight comparison between double heterozygous and wild-type indicates no significant difference showing $p=0.2595$ using the student t-test. All two tailed p-values have been calculated using the paired t-test.
The experiments were performed in triplicates for each imprinted gene (Peg3, Zim1, Usp29). The ΔCt value was initially calculated by subtracting Ct value of a testing replicate of a given gene from the average Ct value of the internal control (28S and β-actin). The fold difference for each replicate was then calculated by raising the ΔΔCt value as a power of 2 [29]. The relative expression levels of all samples were then calculated by dividing the calculated expression level of each sample by the expression level of the wild-type sample. The average and standard deviation for each sample were then calculated by compiling the normalized values. The information regarding individual primer sequences are also available (Table 2.2).

Table 2.2. Primer sets used for RT-PCR and qRT-PCR experiments.

<table>
<thead>
<tr>
<th>Locus</th>
<th>Name</th>
<th>Sequence (5' → 3')</th>
<th>Primer set</th>
<th>Size (bp)</th>
<th>*Position (mm9, NCBI Build 37)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peg3 1-6</td>
<td>Peg3-RT-exon1-F1 Peg3-RT-1262</td>
<td>ATCCCTGAAACGCTCAAGCCCTAGATCCCCGGTTAGGCCAGC</td>
<td>1st primer</td>
<td>291(289)</td>
<td>chr7:6,668,849-6,671,234</td>
</tr>
<tr>
<td>Peg3 1-4</td>
<td>Peg3-RT-14 Peg3-RT-124</td>
<td>GGTTCAGTGGTGCTAGACTGTCACCTGTGGGTGTCGTC</td>
<td>1st primer</td>
<td>304(301)</td>
<td>chr7:6,670,582-6,683,038</td>
</tr>
<tr>
<td>Beta gal Pseudo Exon</td>
<td>Peg3-RT-exon3-F2 Peg3-RT-β gal pseudo-exon</td>
<td>ATCCCTGAAACGCTCAAGCCCTGGAACGGAGGGCCTTG</td>
<td>1st primer</td>
<td>*110</td>
<td>chr7:6,668,849-6,671,234</td>
</tr>
<tr>
<td>Zim1</td>
<td>Zim1-RT-c Zim1-RT-d</td>
<td>GATCACCAGGTGTGGAGCAAAGGAGTCAGGCTCTCTCTGTTGTTGTAGTG</td>
<td>1st primer</td>
<td>405</td>
<td>chr7:6,630,678-6,634,871</td>
</tr>
<tr>
<td>Usp29</td>
<td>Usp29-RT-Pol1 Usp29-RT-PolB</td>
<td>GAGCCCTCGACGCCAGCGCTGCTCCGCGGCTGCTCCT</td>
<td>1st primer</td>
<td>233</td>
<td>chr7:6,964,937-6,912,835</td>
</tr>
<tr>
<td>Actin (beta)</td>
<td>bActin-1a bActin-1b</td>
<td>GACGACCCCTGTCGCTACCCGAACCTGTTAGCTACGAGACATTTC</td>
<td>1st primer</td>
<td>344</td>
<td>chr5:143,666,183-143,666,981</td>
</tr>
</tbody>
</table>

**Immunohistochemistry**

Mice were deeply anesthetized with Nembutal (50mg/kg; i.p) and intracardially perfused with 0.1 M sodium phosphate-buffered saline (PBS: pH 7.2-7.4) followed by a fixative of 4% paraformaldehyde in 0.1 M phosphate buffer (PB; pH 7.2-7.4). Brains were dissected and post-fixed in the same fixative overnight, then transferred to 30% sucrose in 0.1 M PB overnight. Coronal sections (40-µm) containing the SON and PVN were obtained by a sliding microtome (Leica SM2010R, Leica, Mannheim, Germany). Polyclonal antibody against PEG3 was raised in
rabbit and used in free-floating brain slices overnight at 4°C at a concentration of 1:2,000 [16] in PBS containing 0.5% Triton X-100 (PBST). For single staining of PEG3, the brain slices were subsequently incubated with biotinylated goat anti-rabbit antibody (Vector Labs, Burlingame, CA) at 1:200 in PBST. Antigen-antibody interaction was visualized by the ABC-diaminobenzidine method according to the protocol provided by Vector Labs (Burlingame, CA). The brain sections were mounted on gelatin-coated slides, dehydrated, cleared and cover slipped with Permount. The brain sections were rinsed 3 times for 5 min with PBST between each step.

For double immunofluorescence, the brain sections were incubated with PEG3 antibody, followed by incubation with goat anti-rabbit antibody conjugated with DyLight 488 (Jackson ImmunoResearch, West Grove, PA) at 1:400 in PBST overnight. Subsequently, sections were incubated with either oxytocin-neurophysin (NP) or vasopressin-NP mouse monoclonal antibodies (PS38 and PS41, respectively: provided by H. Gainer, NIH) at 1:500 in PBST overnight, followed by incubation with goat anti-mouse antibody conjugated with DyLight 649 (Jackson ImmunoResearch, West Grove, PA) at 1:400 in PBST overnight. In some cases, brain sections were counter stained with DNA labeling dye, 1,5-bis{[(2-[di-methylamino)ethyl]amino]-4, 8-dihydroxyanthracene-9,10-dione (DEAQ5), according to manufacture’s instructions (BioStatus limited, Thermo Scientific). The sections were mounted in polyvinyl alcohol (PVA) with anti-fading agent 1,4-diazabicyclo[2.2.2]octane (DABCO) that consists of 4.8g PVA, 12g glycerol, 12mL dH2O, 24 mL0.2M Tris-HCl, and 1.25g DABCO. Images were acquired with a confocal microscope Leica TCS SP2 spectral confocal microscope, Mannheim, Germany). Optical section thickness was 1 µm. These were viewed in stacks of 5 sections using ImageJ software (NIH).
2.3 Results

Generation of mutant alleles for Peg3 and Yy1

The in vivo roles of Yy1 in the Peg3 domain were investigated using the following mutant alleles, Peg3<sup>tm1aEUCOMMhmgu</sup> and Yy1<sup>tm2Yshi</sup> (Figure 2.3). First, the Peg3 locus was initially targeted through inserting an expression cassette carrying a promoterless LacZ (β-galactosidase) and NeoR (neomycin resistance gene) into its 5<sup>th</sup> intron [30].

![Genomic structures of the wild-type and mutant alleles of Peg3 and Yy1.](image)

Figure 2.3. Genomic structures of the wild-type and mutant alleles of Peg3 and Yy1. (A) Schematic representation of the wild-type and mutant alleles of the mouse Peg3 locus. The 9 exons of Peg3 are indicated by closed boxes in the wild-type (WT) allele. The conditional knockout (CoKO) allele has a 7.1-kb insertion containing a promoterless β-galactosidase (β-Gal) and human β-actin promoter-driven neomycin resistant gene (NeoR). The insertion cassette in the CoKO allele has been removed through FLP-mediated recombination, producing the FlipKO allele. In the FlipKO allele, two LoxP sites flank the exon 6 of Peg3. The Cre recombinase has been used for deleting the exon 6, deriving the DelKO allele for the Peg3 locus. (B) Schematic representation of the wild-type and mutant alleles of the mouse Yy1 locus. The 1<sup>st</sup> exon of Yy1 has been deleted through Cre-mediated recombination, generating the mutant strain for Yy1.
In this knock-in/knock-out scheme, the 3’-side homologous hook contains two LoxP sites flanking the exon 6, deriving a mutant allele that can be ready for conditional knockout experiments. Thus, this mutant allele was named a conditional knockout-ready (CoKO) allele. This CoKO allele was also designed to have immediate mutational effects through truncating the transcription of Peg3 through two poly(A) signals that had been included as part of the inserted cassette (Figure 2.3 A). The predicted mutational effects have been recently confirmed through a study revealing the complete truncation and subsequent growth-related phenotypes among the mutant mice carrying the CoKO allele [28]. The inserted cassette is also flanked by two FRT sites, and thus the mutational effects by the CoKO allele can be rescued by FLP-mediated recombination, deriving a reverted allele (FlipKO). Finally, the FlipKO allele can be mutated again through the Cre-mediated recombination, resulting in the deletion of the exon 6 (DelKO). Both the FlipKO and DelKO alleles have been successfully generated through two consecutive but separate recombination events, and the mutant strains carrying these two alleles indeed displayed the expected outcomes, the absence and presence of growth-related phenotypes, respectively.

The mutant allele of the Yy1 locus has been derived from the floxed allele of Yy1 through Cre-mediated recombination (Figure 2.3 B). This recombination deleted the 3.4-kb genomic region encompassing the promoter and first exon of Yy1, abolishing the transcription and translation of Yy1. According to the results from initial breeding experiments, the homozygotes carrying the mutant allele were embryonic lethal, while the heterozygotes tend to exhibit smaller body size than their wild-type littermates. This is consistent with the observations derived from previous studies [27]. Interestingly, we have also observed a statistically significant gender ratio among the heterozygotes (male : female = 18 : 3) ($\chi^2$ test: $\chi^2=10.714; df=1; p=0.0011$) although
a small number of litters were tested (Figure 2.2 A). The females are less represented in the heterozygous pool of neonates, which might be caused by Yy1 effects on the Xist locus. Overall, 3 different mutant alleles (CoKO, FlipKO, DelKO) for Peg3 and one mutant allele (Yy1 mutant) for Yy1 were successfully generated for a series of breeding experiments as described below.

**Breeding of CoKO and DelKO with Yy1 mutant stains**

We used the following strategy to test the gene dosage effects of Yy1 on the Peg3 domain *in vivo* (Figure 2.1). This strategy involves the crossing of the mutant alleles of two genetic loci, Peg3 and Yy1, wherein the mutant alleles of Peg3 serve as a reporter to monitor the gene dosage effects of Yy1. The CoKO allele expresses β-galactosidase (β-Gal) under the control of the endogenous promoter of Peg3 so that potential Yy1 dosage effects on Peg3 can be inferred through the activity of β-Gal or RT-PCR utilizing the sequence of β-Gal. Two parental alleles of Peg3 are also functionally different due to genomic imprinting by the active paternal versus repressed maternal alleles. Thus, Yy1 gene dosage effects on the paternal and maternal alleles of Peg3 were analyzed separately through a set of reciprocal breeding schemes (Breeding I and II). In Breeding I and II, the female and male heterozygotes (hets) for the mutant allele of Yy1 were crossed with the male and female heterozygotes (hets) for the CoKO allele of Peg3, respectively. We also used another mutant allele of Peg3, DelKO, as an independent reporter allele for this experiment to rule out any artifacts that could originate from the inserted sequence elements within the CoKO allele, such as β-Gal itself and human β-actin promoter-driven NeoR. Thus, the female Yy1 hets were crossed with the male DelKO hets (Breeding III).

We have obtained four litters from each of the three breeding schemes, and examined their litter sizes as well as individual health status by measuring their weights (Table 2.3 and Figure 2.2 C). The litter sizes of one-day-old mice derived from breeding I, II, and III (8.25,
7.75, and 6.5 respectively) were close to the normal litter size (8) of the same genetic background (129/B6), indicating no embryonic lethality associated with these breeding schemes. These three breeding schemes produced four different genotypes of the progeny: double hets \([\text{Peg3}^{+/-}\text{Yy1}^{+/-}]\), Peg3 hets \([\text{Peg3}^{+/-}]\), Yy1 hets \([\text{Yy1}^{+/-}]\), and wild-type [WT]. According to the results from genotyping, all four different genotypes were represented at the predicted Mendelian ratio (1: 1: 1: 1) among the progeny from the three breeding experiments, confirming that the progeny with each genotype is viable until birth (Figure 2.2 B). This is again consistent with the previous observation, that no embryonic lethality is associated with these breeding schemes. The weight profiles, however, indicated that the double het progeny tend to be smaller and weaker than their littermates (Figure 2.2 C). Furthermore, none of the double hets from breeding III survived past their weaning age, indicating the severity of the combined mutational effects of Yy1 and Peg3 on the postnatal survival of individual mice. In summary, the three breeding schemes successfully produced the progeny with all possible combinations of genotypes.

Table 2.3. Summary of breeding results

<table>
<thead>
<tr>
<th>Breeding setup</th>
<th>Peg3 (+/-p)</th>
<th>Peg3 (-m/+p)</th>
<th>Peg3 (+/-)</th>
<th>Peg3 (-m/)</th>
<th>Yy1 (+/-)</th>
<th>WT</th>
<th># of litters</th>
<th>Average litter size (adult)</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>F</td>
<td>M</td>
<td>F</td>
<td>M</td>
<td>F</td>
<td>M</td>
<td>F</td>
<td>M</td>
</tr>
<tr>
<td>I</td>
<td>5</td>
<td>3</td>
<td>-</td>
<td>-</td>
<td>6</td>
<td>1</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>II</td>
<td>-</td>
<td>-</td>
<td>2</td>
<td>7</td>
<td>-</td>
<td>-</td>
<td>3</td>
<td>7</td>
</tr>
<tr>
<td>III</td>
<td>2</td>
<td>2</td>
<td>-</td>
<td>-</td>
<td>3</td>
<td>5</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

**Yy1 dosage effects on paternal allele of Peg3**

The gene dosage effects of Yy1 on the Peg3 domain was analyzed mainly with the progeny derived from Breeding III crossing female Yy1 hets and male Peg3 DelKO hets. A set of one-day-old pups with four genotypes (double hets, Peg3 DelKO hets, Yy1 hets, WT) was used for preparing total RNA, cDNA and subsequent qRT-PCR analyses (Figure 2.4).
Figure 2.4. Effects of Yy1 gene dosage on the Peg3 imprinted domain. (A) Genomic structure of the Peg3 imprinted domain: maternally expressed Zim1 and paternally expressed Peg3 and Usp29. (B) The current study used two sets of RT-PCR primers for Peg3: the first set amplifying exons 1-4 and the second set amplifying exons 3-6. (C) RT-PCR analyses of the progeny derived from Breeding I (female Yy1 heterozygotes X male Peg3 DelKO heterozygotes). RT-PCR amplifying exons 1-4 and exons 3-6 were performed using the total RNA isolated from the neonatal brains with 4 genotypes (lanes 1-4). In the case of RT-PCR amplifying exons 1-4, the PCR products from the pups with four genotypes represent the expression from the paternal allele since Peg3 is paternally expressed. In the case of RT-PCR amplifying exons 3-6, the PCR products from the pups with two genotypes (lane 3, Yy1-/+; lane 4, WT) still represent the expression from the paternal allele of Peg3, but the products from the pups with the other genotypes (lane 1, Yy1-/+ Peg3+/-; lane 2, Peg3+/-) represent the expression from the maternal allele. The mRNA from the paternal allele of Peg3, DelKO, cannot be detected by the RT-PCR amplifying exons 3-6 since the exon 6 is deleted in the DelKO allele. (D) Quantitative RT-PCR analyses using the total RNA isolated from the pups with four genotypes: Yy1-/+ Peg3+/- (1), Peg3+/- (2), Yy1-/+ (3), WT (4). The expression values of each gene were normalized first with an internal control (28S) and later with the values from the WT pup (lane 4). The expression levels of Peg3 were analyzed using the primer set amplifying exons 1-4, thus representing the expression levels of the paternal allele. This series of qRT-PCR analyses were repeated three independent times from cDNA synthesis to qRT-PCR. Error bars indicate standard deviations for observed triplicates.
Actual dosage effects of Yy1 were tested by comparing the expression levels of a given gene between double hets vs Peg3 DelKO hets (lane 1 vs 2 in Figure 2.4 C) and Yy1 hets vs WT (lane 3 vs 4 in Figure 2.4 C). For the Peg3 locus, two sets of primers were also used to measure the expression levels, which included the primer set amplifying exon 1 through 4 for the paternal allele and the primer set amplifying exon 3 through 6 for the maternal allele expression (Figure 2.4 B). Since the DelKO allele lacks exon 6, the primer set for exon 3-6 will amplify its corresponding product only from the normal maternal allele that has been inherited from female Yy1 hets. On the other hand, the primer set for exon 1-4 will amplify its product mainly from the paternal allele due to the paternal expression of Peg3. This series of expression analyses also included the two adjacent genes of Peg3, maternally expressed Zim1 and paternally expressed Usp29 (Figure 2.4 A). Other imprinted genes, such as Zim2, Zim3 and Zfp264, were not included due to their very low expression levels in neonatal brain [31, 32].

According to the results from qRT-PCR, the expression levels of the paternal allele of Peg3 were 1.5-fold higher in double hets compared to Peg3 DelKO hets (lanes 1 and 2, Figure 2.4 D), and Yy1 hets compared to WT (lanes 3 and 4, Figure 2.4 D). In both sets, the half dosage of Yy1 coincides with the up-regulation of Peg3, suggesting a repressor role for Yy1 in the paternal allele of Peg3. Interestingly, the expression level of Peg3 is 3-folds lower in double hets compared to Yy1 hets (lanes 1 and 3, Figure 2.4 D), and in Peg3 DelKO hets compared to WT (lanes 2 and 4, Figure 2.4 D). This phenomenon is likely associated with exon 6 deletion in both samples; thus, the observed down-regulation might be caused by the degradation of the Peg3 mRNA lacking exon 6 and thus the ORF (Open Reading Frame) through the NMD (Non-sense mRNA Decay) pathway [33]. In the case of Zim1, the half dosage of Yy1 also correlates with the up-regulation of Zim1, yet the levels of this up-regulation (16 fold) were much higher than those
observed from Peg3 (1.5 fold). By contrast, the half dosage of Yy1 did not result in any major change in the expression levels of Usp29. Overall, this series of expression analyses concluded that the half dosage of Yy1 coincides with the up-regulation of both Peg3 and Zim1, suggesting a repressor role for Yy1 for both genes in the Peg3 imprinted domain. This series of analyses was repeated with 3 technical replicates and 2 biological replicates, and the overall conclusion was reproducible with these independent trials.

**Yy1 dosage effect on maternal allele of Peg3**

Yy1 dosage effects on the maternal allele of Peg3 were initially analyzed by detecting the expression of β-Gal in the whole mount and sectioned samples prepared from the progeny of Breeding II inheriting the CoKO allele with β-Gal maternally (Figure 2.1). Although this series of experiments was not fruitful due to the low sensitivity of the β-Gal staining, we were able to detect low levels of the maternal expression of Peg3 through RT-PCR (Figure 2.5 D). This suggests that the paternal allele of Peg3 is intact and functional, and yet the maternal allele, CoKO, is still expressed (Figure 2.5 D). To further investigate the observed maternal expression of the Peg3 locus, we decided to use the progeny of Breeding III inheriting the DelKO allele paternally, based on the following reason. The maternal allele in this progeny contains the normal, unmodified Peg3 locus, yet it can be differentiated from the paternal DelKO allele (lacking exon 6) with the primer set amplifying Peg3 exon 3-6 (Figure 2.5 A). According to the initial survey (Figure 2.5 B), the low levels of Peg3 expression from the maternal allele were indeed observed from the neonate brains among all progeny with the inherited DelKO allele (Figure 2.5 C). This confirmed the maternal, and thus bi-allelic, expression of the Peg3 locus in the brain. Subsequent qRT-PCR analyses further revealed that the relative expression level of the maternal allele of Peg3 was about 0.5% of the paternal allele (Figure 2.5 F, G, Figure 2.6).
Figure 2.5. Effects of Yyl gene dosage on the maternal expression of Peg3. (A) Schematic representation of the Peg3 locus showing the positions of the two sets of primers used for RT-PCR. (B) RT-PCR products with the primer set amplifying exons 3-6. Comparison of the expression levels between the neonate brains of pups with the following two genotypes, Yy1-/+; Peg3+-/ (lane 1) and Peg3+-/ (lane 2), indicate up-regulation of the maternal allele of Peg3 by the half dosage of Yy1. (C) Additional RT-PCR analyses further confirming the maternal expression of Peg3 as well as up-regulation of the maternal expression of Peg3 by half dosage of Yy1. Neonate brains of pups with the genotypes Yy1-/+ Peg3+-/ (1), Peg3+-/ (4), and WT (7) are the same as the samples shown in Figure 2.5 B (2), (3), and (4), respectively. An additional set of pups were obtained from Breeding III (marked with an asterisk *), and subsequently used for the RT-PCR analyses. (D) Yy1 dosage effect on the maternal allele of Peg3. Additional RT-PCR analyses confirming the maternal expression of Peg3 with an additional set of pups (1-7) obtained from breeding II (CoKO maternal transmission with a half dosage of Yy1). The primer set amplifying exon 1-4 was used to illustrate the paternal allele expression of Peg3, while the primer set amplifying the β-Gal insertion cassette (exon 3 - β-Gal pseudo-exon) was used to show the maternal allele expression of Peg3. (E) qRT-PCR analyses measuring the levels of the paternal allele of Peg3 affected by the half dosage of Yy1. This analysis used the primer set amplifying exons 1-4. (F) qRT-PCR analysis showing relative expression levels of the maternal to paternal allele of Peg3. This analysis used the primer set amplifying exons 3-6. (G) qRT-PCR analyses measuring the expression levels of the maternal allele of Peg3 affected by the half dosage of Yy1. This analysis used the primer set amplifying exons 3-6.
Figure 2.6. Maternal allele expression of Peg3 using a mouse hybrid cross. (A) Schematic representation of the Peg3 locus. Positions are indicated for two sets of primers used for qRT-PCR to distinguish the maternal and the paternal alleles of a PWD/B6 hybrid mouse strain. A female PWD mouse was mated with a B6 male to give rise to hybrid progeny. Using two single nucleotide polymorphisms (SNPs), two primers were designed to distinguish Peg3 alleles from PWD (maternal) and the B6 (paternal). RNA was isolated and subsequent cDNA was generated from the hypothalamus and the rest of the brain from PWD/B6 hybrid progeny to detect allele specific Peg3 expression levels. (B) qRT-PCR analyses measuring the levels of Peg3 maternal allele expression in PWD/B6 hypothalamus and brain compared to their parental strains. Allele specific reverse primers were combined with a forward primer corresponding to Peg3 exon 6 to amplify mRNA from Peg3 exon 6-9 to determine the relative expression levels of Peg3 in hybrid tissues compared to their parental strains. The average expression levels of Peg3 was normalized to β-actin and subsequently compared to B6 and PWD respectively. The percentage of maternal Peg3 expression was calculated using the maternal to paternal expression ratio of the PWD/B6 hybrid tissues.

The half dosage of Yy1 also coincided with the 1.5-fold up-regulation of Peg3 (Figure 2.5 G), which is similar to the up-regulation level seen in the paternal allele (Figure 2.5 E). Given the similar changes of Peg3 expression levels observed between the paternal and maternal alleles by a half dosage of Yy1, this is considered to be a transcriptional up-regulation of the already active maternal allele, rather than de-repression of the repressed maternal allele by genomic imprinting.
In summary, this series of analyses concluded that the maternal allele of Peg3 is normally expressed at very low levels in the brain, and that the half dosage of Yy1 also causes an up-regulation of Peg3 on the maternal allele.

**Bi-allelic expression patterns in the specific areas of mouse brain**

The observed low levels of expression from the maternal allele of Peg3 were further investigated with RT-PCR (Figure 2.7) and immunohistochemistry. We first surveyed the maternal expression of Peg3 using the total RNA isolated from the neonate mouse heads of the Peg3 het mice inheriting the DelKO allele paternally (Figure 2.5 B, C). We repeated an RT-PCR assay on a set of the total RNA isolated from the different parts of the adult mouse brain (Figure 2.7 A).

![Figure 2.7](image)

Figure 2.7. Maternal expression of Peg3 in mouse brain. (A) RT-PCR testing the maternal expression of Peg3. The total RNA was isolated from the different parts of a 4-month-old male mouse with the paternally transmitted DelKO allele (midbrain, cerebellum, olfactory, hypothalamus, and pituitary). These RNA were analyzed with the two sets of primers amplifying exon 1-4 and exon 3-6, confirming the maternal expression of Peg3 in the midbrain and hypothalamus. (B) qRT-PCR analyses were also performed to measure the relative expression levels of the paternal and maternal alleles of Peg3 between the RNA samples isolated from the different parts of the adult brain, including the midbrain (M), cerebellum (C), olfactory (O), hypothalamus (H), and pituitary (P). Parts of the adult mouse brain showing Peg3 maternal allele expression have been marked with an asterisk (*).
The maternal expression was detected mainly in the hypothalamus and mid brain sections with the expression levels being slightly higher in the hypothalamus than in the mid brain (Figure 2.7 B). This was somewhat different from the expression pattern observed from the paternal allele, which showed global expression throughout the entire brain. This suggests that the observed maternal expression is specific to certain areas of the brain including the midbrain and the hypothalamus regions. The relative expression levels of maternal to paternal alleles in these cell types are much lower based on qRT-PCR data (about 1% of the paternal level, Figure 2.7 B, Figure 2.6). This suggests that Peg3 expression is most likely bi-allelic in a small population of cells in the midbrain and hypothalamus regions. Moreover, DNA methylation analyses using DNA derived from tissues pertaining to bi-allelic expression shows no major methylation differences when compared to tissues derived from other areas of the brain (Figure 2.8).

Figure 2.8. DNA Methylation analysis of Peg3. Methylation levels of the Peg3-DMR were determined using COBRA. A set of genomic DNA isolated from the cortex and choroid plexus of two mice (WT and Yy1+/+) was treated with bisulfite conversion. The amplified PCR products from the Peg3-DMR were digested with HphI and TaqIa enzymes. The digestion pattern revealed half methylation in both CP and CTRL without any major difference, indicating no obvious methylation difference in the choroid plexus with Peg3 biallelic expression. This suggests that small populations of cells are likely biallelic and/or an unknown alternative promoter may derive the maternal expression. The observed pattern is also true between WT and YY1−/−, indicating no major effect on the biallelic expression of Peg3 by Yy1.
To elucidate the location of maternal Peg3 expression in the brain, we performed immunostaining of PEG3 using a series of whole coronal sections containing the hypothalamus (10 sections collected every 200 µm) from an adult Peg3 hets inheriting the CoKO and DelKO alleles paternally (Figure 2.9). Brain slices from the wild-type littermates were also included as a positive control. There are numerous PEG3 immunoreactive cells ubiquitously located in the brain sections from WT mice. The intensity of immunoreactivity appeared especially high in the hypothalamic region (Figure 2.9 Aa) and in the choroid plexus in the lateral ventricles (Figure 2.9 Ba). In contrast, there was no detectable PEG3 immunoreactivity in the brain slices from the paternally transmitted CoKO or DelKO mice, except in the hypothalamic paraventricular (PVN) and supraoptic (SON) nuclei (Figure 2.9 Ab) and in the choroid plexus (Figure 2.9 Ba), which was expected given the very low (0.5%) maternal expression of Peg3. Prominent PEG3-immunoreactive cells observed in the PVN and SON in the CoKO and of DelKO mice (Figure 2.9 Ab) is also consistent with the results from RT-PCR, which showed Peg3 expression in the hypothalamus (Figure 2.7 A, B). The choroid plexus is a tuft of capillaries with an overlying epithelial covering. PEG3 immunofluorescence labeling was counterstained with a DNA dye, DRAQ5, to investigate whether PEG3 immunoreactivity is located in the capillary lining endothelial cells or the choroid plexus epithelial cells. PEG3 immunoreactivity was located exclusively in the cuboidal epithelia typically found in the choroid plexus epithelium, but was absent from the endothelia of the capillaries (Figure 2.9 Bb). Because the PVN and SON contain neurons synthesizing neurohypophysial hormones, oxytocin and vasopressin, double immunofluorescence detection of PEG3 and oxytocin/vasopressin was employed to determine whether the maternal expression of Peg3 occurs specifically in oxytocin-and/or vasopressin-synthesizing neurons.
Figure 2.9. PEG3 immunoreactivity in adult mouse hypothalamus. (A) PEG3 immunoreactivity in the adult mouse hypothalamus of two littermates. Image a represents global PEG3 immunoreactivity in a 4-month-old, wild-type female mouse brain. Image b represents a 4-month-old female mouse brain including the paternally transmitted DelKO allele with arrows indicating PEG3 immunoreactive neurons located in the paraventricular nucleus (PVN) and the supraoptic nucleus (SON). (B) PEG3 immunoreactivity of the maternal allele in the choroid plexus from the paternally transmitted DelKO allele. Image a represents PEG3 immunoreactivity of the choroid plexus located in the lateral ventricle. Image b represents immunoreactivity of PEG3 (green) and DRAQ5 nuclear staining (blue), with a smaller arrow indicating PEG3 immunoreactive cuboidal choroid plexus epithelia, and a larger arrow indicating non-PEG3 immunoreactive endothelia of capillaries. (C) Double immunostaining of PEG3 (green) and Oxytocin-OT/Vasopressin-VP (blue) immunoreactive neurons in the PVN from the paternally transmitted DelKO allele. Overlay of PEG3/OT and PEG3/VP indicates PEG3 immunoreactivity of the maternal allele coinciding predominantly with OT/VP immunoreactive neurons in the PVN. (D) Double immunostaining of PEG3 (green) and Oxytocin-OT/Vasopressin-VP (blue) immunoreactive neurons in the SON from the paternally transmitted DelKO allele. Overlay of PEG3/OT and PEG3/VP indicates PEG3 immunoreactivity of the maternal allele coinciding predominantly with OT/VP immunoreactive neurons in the SON.
The double labeling revealed that PEG3-immunoreactivity was found in oxytocin- and vasopressin- immunoreactive neurons in both the PVN (Figure 2.9 C) and SON (Figure 2.9 D). In summary, this series of qRT-PCR and immunostaining experiments strongly suggest the maternal allele expression of Peg3 in specific cell types of the mouse brain, thus confirming its bi-allelic expression in these cell types.

2.4 Discussion

In the current study, we tested the in vivo effects of Yy1 gene dosage on the Peg3 imprinted domain with various breeding schemes utilizing the mutant alleles. The results indicated that the half dosage of Yy1 coincides with the up-regulation of Peg3 and Zim1, suggesting a repressor role of Yy1 in the imprinted domain. The results also posed an unexpected observation that the maternal allele of Peg3 is normally expressed, and thus the expression of Peg3 is bi-allelic in the specific areas of the brain, such as the choroid plexus and the SON and PVN of the hypothalamus. Overall, these results confirm the in vivo repressor role of Yy1 that had been previously observed from in vitro studies, and also provide a new insight regarding the bi-allelic expression of Peg3 in the mouse brain.

According to the present study results (Figure 2.4), the half dosage of Yy1 appears to coincide with the up-regulation of Peg3 and Zim1, suggesting a repressor role for Yy1 in the Peg3 domain. A similar observation has been previously made multiple times through a series of in vitro and in vivo experiments, demonstrating the global up-regulation of the Peg3 domain in a response to the low levels of the YY1 protein [15-17]. This domain-wide response along with multiple YY1 binding sites within the 1st intron of Peg3 have been the two major observations suggesting the possibility that Yy1 is a major trans factor controlling the transcription of this 500-
kb domain [11]. This prediction is overall well supported by the current study utilizing much more controlled in vivo systems than the previous in vitro systems [15-17]. Nevertheless, the current study was unable to replicate another previous observation that the low levels of YY1 protein may be responsible for DNA hypomethylation on the Peg3-DMR (Figure 2.8). According to the results (Figure 2.5), the half dosage of Yy1 does not appear to affect the epigenetic imprint of the maternal allele of Peg3 although this is still somewhat inconclusive due to the technical limitations associated with the sensitivity of the β-Gal staining. Instead, the boosted expression levels observed from the maternal allele of Peg3 in Yy1 hets compared to those of WT mice is thought to be caused by the up-regulation of the already active maternal allele, rather than by the de-repression of the repressed maternal allele via DNA hypomethylation. Although we need to further investigate this aspect in the near future, the inability to detect the predicted hypomethylation in the current study could be due to the following reasons. First, half dosage of Yy1 might not be sufficient enough to derive a similar observation made from in vitro studies. Second, the pups severely affected by the DNA hypomethylation might not be viable so that the breeding schemes used for the current study could not produce the pups with predicted epigenetic imprints. Overall, the current study utilizing in vivo systems again confirms that Yy1 functions as a transcriptional repressor for the Peg3 imprinted domain.

Given the fact that one allele of Peg3 is already repressed by genomic imprinting, it is interesting to speculate why the remaining active allele requires further repression by another transcription factor, Yy1. This may be related to the potential functions of Peg3. According to the recent studies, Peg3 is predicted to be a major regulator controlling autophagy in endothelial cells [34, 35]. Many stimuli from environment, such as starvation, can stimulate autophagy along with Peg3, resulting in a temporal up-regulation of Peg3. Restoring it back to normal levels of
Peg3 is likely required for the proper function of cells, which may use other unknown regulatory mechanisms. In that regard, it is relevant to note that the mTOR (mechanistic target of rapamycin) signaling pathway is known to repress autophagy, in which Yy1 acts as a major contributing factor [36]. Thus, it is reasonable to predict that Yy1 may be involved in controlling the dynamically fluctuating levels of Peg3, which may be triggered by environmental and developmental cues. Recent studies on histone modification profiles have also indicated that promoters of several imprinted genes in the Peg3 domain interacts with one evolutionarily conserved region, ECR18, suggesting its key roles played in the transcription and imprinting control of Peg3 domain as a distant regulatory element [37]. Thus, it is conceivable that Yy1 may affect histone modification profiles contributing to the up-regulation of Peg3 expression instead of DNA methylation changes, using such distant regulatory elements. Although speculative, this may be a reason why Yy1 is needed for the repression of Peg3, which requires further investigation in near future.

The expression of Peg3 appears to be bi-allelic in the specific areas of the brain (Figures 2.7, 2.9). The detection of the maternal expression of Peg3 from the both mutant models, CoKO and DelKO, rules out the possibility that this detection is due to unknown artifacts associated with mutagenesis. Also, this rules out the possibility that the observed maternal expression of Peg3 is caused by some functional compensation between two parental alleles. For instance, the loss-of-function type mutation on the paternal allele (DelKO) might render the cells to de-repress the repressed maternal allele. In the case of the progeny inheriting the CoKO allele maternally (Figure 2.1), however, the paternal allele of Peg3 is intact and functional, and yet the maternal allele, CoKO, is still expressed (Figure 2.5 D). Therefore, this strongly supports the idea that the observed maternal expression is reflecting the genuine bi-allelic expression of Peg3 in normal
mice, which has been previously unnoticed. Nevertheless, this new observation is intriguing given the following reasons. First, there is another imprinted gene, Igf2, which is known to be bi-allelic in the choroid plexus [38, 39]. Given a very small number of imprinted genes in mammalian genomes, the bi-allelic expression of two imprinted genes (Peg3 and Igf2) in the same small area of mouse brain seems to be a very rare coincidence. At the same time, the choroid plexus is known to play a major role in controlling the concentration of ions such as Na\(^+\), Cl\(^-\), HCO\(_3\)^-, and K\(^+\) in the cerebrospinal fluid, and thus should be very critical for the normal function of neurons in the brain. Thus, this rare coincidence may be an indication that some functional constraints derive the bi-allelic expression of these two imprinted genes in the choroid plexus. Second, several previous reports have predicted that the maternal allele of Peg3 may be functional at some unknown stages and/or in specific cell types since the homozygous animals for several mutant alleles targeting the Peg3 locus are not viable although the paternal heterozygotes are still viable [18, 28]. This prediction is further supported by the observed bi-allelic expression of Peg3 in that the maternal allele of Peg3 is indeed expressed and functional in the specific areas of brains. The lack of both the paternal and maternal expression of Peg3 in these brain areas might contribute to the observed lethality of the homozygous mutant animals. Although this is likely, we need to first investigate the functional contribution of the maternal allele of Peg3 to the choroid plexus and other areas, such as the PVN and SON of hypothalamus. Overall, the current study reports, for the first time, the bi-allelic expression of Peg3 in specific areas of mouse brain, and thus it would be of great interest to follow up the functional significance of the observed bi-allelic expression in the near future.
2.5 References


CHAPTER 3
ALTERNATIVE PROMOTERS OF PEG3 WITH MATERNAL SPECIFICITY

3.1 Introduction

_Peg3_ (paternally expressed gene 3) is the first imprinted gene identified from an evolutionarily conserved 500-kb domain localized in human chromosome 19q13.4 and mouse proximal chromosome 7 [1-3]. This domain harbors 6 additional imprinted genes: the paternally expressed _Usp29_, _Zfp264_, and _APeg3_, and the maternally expressed _Zim1_, _Zim2_, and _Zim3_ [4-7]. _Peg3_ is structurally comprised of 9 exons that are spread over a 25-kb distance in both human and mouse. Interestingly, the 200-kb upstream region of both human and mouse _PEG3_ lacks any obvious ORF (Open Reading Frame), but this region has been well preserved during mammalian evolution. According to recent studies, this 200-kb region is filled with small ECRs (Evolutionarily Conserved Regions), which are putative _cis_-regulatory elements for _Peg3_ and other imprinted genes [8, 9]. _Peg3_ has been shown to be involved in controlling maternal-caring behaviors and fetal growth rates [10]. Accordingly, _Peg3_ is highly expressed in neuronal cells as well as embryos and placentas [3, 11]. The mechanistic basis for these expression patterns and also the paternal-specific expression of _Peg3_ is still under investigation. Nevertheless, the potential _cis_-regulatory elements found within the 200-kb region are hypothesized to be involved in regulating the tissue- and allele-specific expression patterns of _Peg3_, given their unusual evolutionary conservation.

As seen in the _Peg3_ locus, other imprinted domains also have similar unusual genomic layouts. For instance, the upstream region of human and mouse _SNRPN_ is more than 500 kb in

---

2 This chapter previously appeared as Perera, B. P. U., Teruyama, R., and Kim, J. (2016) Alternative promoters of _Peg3_ with maternal specificity under Scientific Reports and reprinted with permission of Scientific Reports.
length, which is relatively large compared to a 20-kb transcribed area from its coding region. This 500-kb region also lacks any obvious ORFs, but this region harbors multiple alternative promoters for *Snrpn* and *Ube3a-ATS*, an antisense transcript gene to *Ube3a* [12]. In fact, some of these alternative promoters are critical for establishing germline-specific DNA methylation marks [13, 14]. A similar case has been observed from the *Gnas* domain, where the transcription of the locus is driven by multiple alternative promoters of various tissue and allele specificity [15, 16]. Yet, the transcription by one particular promoter during oogenesis is believed to be responsible for establishing oocyte-specific DNA methylation marks for the downstream promoter [17]. There are additional cases of imprinted domains that harbor multiple alternative promoters in their large upstream regions include *Zac1* and *Grb10*. In the *Zac1* locus, the multiple promoters are responsible for its various tissue-specific expression patterns [17-19]. On the other hand, the two alternative promoters of *Grb10* are known to display both allele and tissue specificity [20, 21]. Given these examples, it is reasonable to predict that the 200-kb upstream region of the *Peg3* locus may also have similar alternative promoters, although the *Peg3* locus has long been known to have only one promoter in both human and mouse.

A recent study, however, demonstrated that the expression of *Peg3* appears to be bi-allelic in certain areas of the adult mouse brain, such as the hypothalamus and choroid plexus [22]. This may be an indicator for either de-repression of the maternal allele of the known promoter or the presence of potential alternative promoters for the *Peg3* locus. In the current study, therefore, we sought to characterize the observed bi-allelic expression pattern with a strategy involving 5’RACE and NGS-based (Next Generation Sequencing) deep sequencing experiments. With this approach, we were able to identify several alternative promoters for both human and mouse *PEG3*. A series of expression analyses further revealed that these alternative
promoters display allele-, tissue-, and stage-specific expression patterns. More detailed results are described in the following section.

3.2 Materials and Methods

Generating the mutant strains for Peg3

The current study used the KO2 and DelKO mutant mouse strains derived from KO2-Neo and CoKO strains, respectively. The strain for the CoKO allele of Peg3 was made using a targeted ES cell from the EUCOMM (European Conditional Mouse Mutagenesis program), and this strain has been maintained in the lab [23]. The strain for the KO2-Neo allele of Peg3 was made using the targeted ES cells from Darwin Transgenic Mouse Core Facility of Baylor College of Medicine, and this strain has been maintained in the lab (H. He et al, unpublished). The CoKO mouse strain was crossed with the Rosa26-FLP line from Jackson Lab (Stock No. 009086, B6.129S4-Gt (ROSA)26Sor<sup>tm1(FLP1)Dym</sup>/RainJ) to allow for FLP-mediated recombination, generating the FlipKO allele. Both the FlipKO and KO2-Neo strains were crossed with the Zp3-Cre line from the Jackson Lab (Stock No. 003651, C57BL/6-Tg (Zp3-cre) 93Knw/J). The mutagenesis through these breeding derived the DelKO and KO2 strains for the Peg3 locus [22, 24]. The following primer sets were used for genotyping of these strains: the deletion and detection of the expression cassette for FlipKO, Peg3-CoKO-F (5’-ATGAGTCTCGATCCAGGTATGCC-3’) and LoxR (5’-TGAACTGATGGCGAGCTCAGACC-3’), Peg3-5arm (5’-CCCTCAGCAGAGCTTCTGCTGCC-3’); the deletion and detection of the expression cassette for DelKO, Peg3-5arm (5’-CCCTCAGCAGAGCTTCTGCTGCC-3’) and LoxR (5’-
TGAACTGATGGCGAGCTCAGACC-3’, Peg3-5arm (5’-CCCTCAGCAGAGCTTCTCTGACC-3’), and Peg3-rev (5’-ACCCATTCTCATGCAGTCCAGACG-3’), the deletion of exon 1 for Peg3 and Usp29, bac2082-F (5’-ACAACCCGGAGTTTAGAGCAGAC-3’) and bac6710-R (5’-GGATGTAAGATGGAGGACTGT-3’); the presence of Zp3-Cre, Zp3-cre-F (5’-TAGGAATCACCGTGAGGTCTCT-3’) and oMR1085 (5’-GTGAAACACGATTGCTGCTACTT-3’). DNA was isolated from ear or tail snips after incubating the tissues at 65°C with the tail lysis buffer (50 mM Tris-Cl at pH 8.0, 100 mM EDTA at pH 8.0, 250 mM NaCl, 1% SDS, 20 µg/mL Proteinase K). PCR premix kit (Intron Biotech) was used for genotyping at the following conditions (step 1, 95°C-2 min; step 2, 95°C-30 sec, 60°C-30 sec, 72°C-60 sec for 33 cycles; step 3, 72°C-7 min). The information regarding individual primer sequences is also available (Table 3.1). All animal studies were approved by the Louisiana State University Institutional Animal Care and Use Committee (IACUC) and were performed in accordance with approved guidelines and regulations of the LSU Division of Laboratory Animal Medicine, Baton Rouge, Louisiana, U.S.A.

Table 3.1 Primer sets used for genotyping, 5’RACE, RT-PCR, and DNA methylation analyses

<table>
<thead>
<tr>
<th>Locus</th>
<th>Name</th>
<th>Sequence (5’ &gt; 3’)</th>
<th>Primer set</th>
<th>Size (bp)</th>
<th>Position (mm9, NCBI Build 37)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peg3-ko KO (genotyping) Peg3-ko KO</td>
<td>Peg3-ko KO F</td>
<td>ATGACTTCAGACGCTAGCCCAGCTAGGGTCAGCAGAACC</td>
<td>1st primer</td>
<td>m290</td>
<td>mChr7: 6,660,691–9,668,412</td>
</tr>
<tr>
<td>Peg3-ko KO (genotyping) Peg3-ko KO</td>
<td>Peg3-ko KO R</td>
<td>TGAAGCTAAGGGCACTAGGGTCAGCAGAACC</td>
<td>2nd primer</td>
<td>n700</td>
<td>mChr7: 6,660,691–9,668,412</td>
</tr>
<tr>
<td>Zp3-Cre (genotyping) Zp3-Cre</td>
<td>Zp3-Cre F</td>
<td>TAGGAATCACCGTGAGGTCTCTGAGCAAGGATGTAAGATGGGAGGACTGTGCTTGTACTT</td>
<td>1st primer</td>
<td>n500</td>
<td>mChr5: 136,455,787–unknown</td>
</tr>
</tbody>
</table>
5’ RACE and sequencing

Total RNAs were isolated from adult mouse hypothalamus and one-day-old mouse neonate heads using the Trizol RNA isolation kit (Invitrogen). Total RNAs for normal adult and neonate human brains were purchased from BioChain (cat: R1234035-50; R1244035-50). The
total RNA (2.5-5µg) was mixed with gene-specific primers substituting random primers: Ex2-R1 (5’-AGTCTTCCTCTTGGCCGTTGTC-3’) for KO2 mice, Ex6-R2 (5’-CCAAAATGTGGTCTTGCATACACAG-3’) for DelKO mice, and Ex2-R1 (5’-TCCCTTCCTCTCCTCGCCAGTCG-3’) for human, and reverse-transcribed using the M-MuLV reverse transcriptase (New England Biolabs, cat: M0253S). The cDNA products were purified using phenol chloroform extraction and ethanol precipitation. The 5′-ends of the purified cDNA were further modified by the tailing reaction using dGTPs and terminal deoxynucleotidyl transferase according to the manufacturer’s protocol (New England Biolabs, cat: M0315S). The tailed cDNA was amplified using two primers: the tail long primer (5’-GGTTGTAGCTCTTCTAGATCCCCCCCCCCCCCC-3’) and internal gene-specific primers: Ex2-R2 (5’-TCCTCTTGCCAGTTGTCTCCAA-3’) for KO2 mice, Ex6-R3 (5’-ATGTGGTCTTGCATACAGGAAAGA-3’) for DelKO mice, and Ex2-R2 (5’-CTTCTTCTCGCCAGTCGCTTCC-3’) for human. The amplified DNA was re-amplified with a set of nested primers: the tail out primer (5’-GGTTGTAGCTCTTCTAGA-3’) and additional internal gene-specific primers as anchors (Table 3.1). The PCR products were further purified and sequenced using next generation sequencing (NGS) for analysis of 5′ RACE products [25].

RNA isolation and quantitative RT-PCR analysis

Total RNA was isolated from the olfactory bulb, midbrain, hypothalamus, cerebellum, thymus, heart, lung, liver, kidney, fat, testis, 14.5-dpc placenta and ovary from adult male and female mice, heads of one-day-old neonates, and 14.5-dpc embryos using a commercial kit (Trizol reagent, Life technologies, cat: 15596018) according to the manufacturer’s protocol. Total RNAs for normal human adult brain and the neonate brain were purchased from BioChain (cat: R1234035-50; R1244035-50). The total RNA was then reverse-transcribed using the M-
MuLV reverse transcriptase (NEB, cat: M0253S). A set of normalized cDNA for human heart, liver, kidney, and placenta was obtained from BioChain as well. The cDNA fragments were used as a template for PCR amplification (Maxime PCR Premix Kit, Intron Biotech) to check for mRNA transcripts at the following conditions (step 1, 95°C-2 min; step 2, 95°C-30 sec, 60°C-30 sec, 72°C-60 sec for 36 cycles; step 3, 72°C-7 min). The information regarding individual primer sequences is also available (Table 3.1).

**DNA methylation analysis**

DNA was first isolated from the hypothalamus, olfactory bulb, midbrain, cerebellum, and kidney of WT adult mice, and subsequently treated with the bisulfite conversion protocol using the EZ DNA methylation kit (Zymo Research, cat: D5002). The converted DNA was used for PCR amplification of the upstream exons regions E1, U1, U2, and U3 (Maxime PCR Premix Kit, Intron Biotech). The following primer combination was used to amplify the upstream U1 promoter: Bis-Peg3-RACEF1-F1 (5’- GTTGGGAATGGAAAATATTAAAGATAA -3’) and Bis-Peg3-RACEF1-R1 (5’- AAAATCAAAAACATACACAAAACATACAA -3’). Upstream U2 promoter was amplified by using the following primer combination: ECR4-Bis-a (5’- ATTGGTTTATAGTTAGGGAAGGAAGTAGT -3’) and ECR4-Bis-b (5’- AAATCTCTCTAAACATAATACTATTCTAT -3’). The following primer combination was used to amplify the upstream U3 promoter: Bis-Peg3-RACEF5-F (5’- GTAGGTAAGAATTTATTGGATAAGAGT -3’) and Bis-Peg3-RACEF5-R (5’- CTTTCTTTCTCTTTTCTTTATACATATAT -3’). The upstream E1 promoter was amplified by using the following primer combination: mPeg3-pro-bis-a.1 (5’- GTTTTTGTAAGGATTTTGATAAGGAG -3’) and mPeg3-pro-bis-b (5’- CACCCCCAACACCATCTAAACTCTACAAAC -3’). Each PCR product was further analyzed
by restriction enzyme digestion-based COBRA (COBined Bisulfite Restriction Assay) and sequencing [26, 27]. For the COBRA analysis, PCR products were digested with various restriction enzymes (NEB). Each PCR product was also used for next-generation sequencing (NGS). Detailed information regarding oligonucleotide sequences, and COBRA is also available (Table 3.1).

3.3 Results

**Mutant alleles of Peg3**

In the current study, two mutant alleles of mouse Peg3 were used to identify the alternative 1st exons and corresponding promoters that may be localized upstream of the known 1st exon and promoter of Peg3. The known 1st exon of Peg3 is localized next to the 1st exon of the adjacent gene Usp29, controlling the transcription of both Peg3 and Usp29 bidirectionally (Figure 3.1 A). The 4-kb genomic region surrounding this bidirectional promoter is also differentially methylated between two alleles: unmethylated paternal versus methylated maternal alleles [8]. Thus, this promoter is functional only from the paternal allele. This 4-kb DMR of Peg3 was deleted using a floxed allele recently generated in the lab named KO2-Neo (H. He et al, unpublished). The Zp3-cre strain involves a cre-recombinase enzyme that is expressed in the female germ line. When bred with the KO2-Neo strain, the floxed sequence is recognized and deleted in the growing oocytes. Thus, generating the KO2 allele with a deletion of the known Peg3 promoter region including its first exon (Figure 3.1 B, C). The current study also used another strain, CoKO, which is designed to truncate the transcription of Peg3 and also to delete the exon 6 with two flanking LoxP sites [23]. This CoKO strain was first bred with the Flippase strain, and later with the Zp3-cre strain to delete exon 6, generating the second mutant allele.
DelKO for the current study (Figure 3.1 C). Since Peg3 is expressed mainly from the paternal allele, the two mutant alleles were transmitted paternally to remove the main transcript of Peg3 for the KO2 strain and also to have the main transcript without exon 6 for the DelKO strain [24].

Figure 3.1. Genomic structures of the wild-type and mutant alleles of Peg3. (A) Schematic representation of the wild-type alleles of mouse Usp29 and Peg3 loci. Gray and black boxes indicate the exons of Usp29 and Peg3, respectively. Transcriptional direction for each gene is represented by arrows with corresponding colors. (B) Schematic representation of the KO2 mutant allele. The KO2-Neo mutant has a 6.0-kb insertion containing a PGK promoter-driven neomycin resistance gene (NeoR) as selection marker, shown by the red box. NeoR, along with the 4.0-kb promoter region of Peg3 containing first exons of Usp29 and Peg3 are flanked by two LoxP sites as indicated with triangles. Cre-recombinase was used to delete the first exons of Usp29 and Peg3, producing the KO2 mutant. (C) Schematic representation of the DelKO mutant allele. The conditional-ready knockout (CoKO) allele has a 7.1-kb insertion containing a promoterless β-galactosidase gene (β-Gal) as shown by the blue box and human β-actin promoter-driven neomycin resistance gene (NeoR) as shown by the red box. The insertion cassette of the CoKO allele is removed by FLP-mediated recombination, producing the FlipKO allele which has the 6\(^{\text{th}}\) exon of Peg3 flanked by two LoxP sites as indicated with triangles. Cre recombinase was used to delete exon 6, producing the DelKO allele for the Peg3 locus.
Alternative 1st exons and promoters of Peg3

To identify potential alternative 1st exons of Peg3, a series of 5’RACE (Rapid Amplification cDNA Ends) experiments were performed using total RNA isolated from the tissues of the two mutant strains and their wild-type littermates (Figure 3.2 A). Total RNA was isolated from the brains of neonates given that high levels of Peg3 expression were found in neuronal cells. The hypothalamus of adult mice has already shown maternal expression of Peg3, suggesting the presence of potential alternative promoters for Peg3 [22]. Thus, this portion of brain was also included for the current study. The total RNA from the neonatal brains and the adult hypothalamus were first reverse-transcribed with gene-specific primers: Ex2-R1 for the total RNA from KO2(+/p) and Ex6-R1 for the total RNA from DelKO(+/p). These initial cDNA fragments generated by Ex2-R1 and Ex6-R1 primers were further amplified using two different nested PCR schemes (RACE I and II, Figure 3.2 A). The amplified libraries were finally sequenced using a NGS (Next Generation Sequencing) platform [25], and the results are summarized in Table 3.2.

Table 3.2. Summary of NGS-based sequencing of 5’RACE libraries from mouse Peg3

<table>
<thead>
<tr>
<th>Tissues (genotype)</th>
<th>Neonatal brain (KO2)</th>
<th>Neonatal brain (WT)</th>
<th>Neonatal brain (WT)</th>
<th>Adult hypoth (DelKO)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RACE scheme (# of total read)</td>
<td>I (575)</td>
<td>I (1843)</td>
<td>II (5473)</td>
<td>II (1197)</td>
</tr>
<tr>
<td>E1-E2-9</td>
<td>6.60</td>
<td>93.54</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>U1-U0-E2-9</td>
<td>0.00</td>
<td>0.00</td>
<td>8.17</td>
<td>78.96</td>
</tr>
<tr>
<td>U2-U0-E2-9</td>
<td>67.83</td>
<td>0.33</td>
<td>77.96</td>
<td>0.00</td>
</tr>
<tr>
<td>U3-U0-E2-9</td>
<td>0.00</td>
<td>0.00</td>
<td>0.09</td>
<td>7.60</td>
</tr>
</tbody>
</table>
Figure 3.2. Alternative 1st exons and promoters of Peg3. (A) Map of the mouse Peg3 locus. Gray and black boxes indicate the exons of Usp29 and Peg3, respectively. Transcriptional direction for each gene is represented by arrows with corresponding colors. The solid boxes indicate the position of the exons of Peg3, labeled E1 through E9, followed by ovals to indicate the position of upstream alternative exons. The closed oval represents the shared upstream exon U0, and open ovals represent alternative 1st exons U1, U2, and U3, respectively. The deleted Peg3 exons corresponding to KO2(+/-p) and DelKO(+/-p) mutant alleles are shown using parentheses and dashed lines. The two extended arrows show the anchoring primers used for nested PCR after 5’ RACE: Ex2-R2 and Ex2-R3 for KO2(+/-p) mutant and its WT counterpart, as shown by scheme I. The two extended arrows underneath it shows the anchoring primers used for nested PCR after 5’ RACE: Ex6-R3 and U0 for DelKO(+/-p) adult hypothalamus and a WT neonate brain as indicated by the scheme II. A summary of the sequence reads is shown in Table 3.2. (B) The exon structures of Peg3 alternative transcripts identified from the mouse brain. The E1 - E2 exon structure indicates the transcription by the main promoter of Peg3. The exon U1 is positioned 20-kb upstream of E1, the 1st exon of Peg3. The alternative transcript starting from the U1 exon connects to U0 (16-kb upstream of E1) and E2 exons while skipping E1. The exon U2 is positioned 26-kb upstream of E1. This alternative transcript starting from the U2 exon connects to U0 and E2 exons while skipping U1 and E1 exons. The exon U3 is positioned 163-kb upstream of E1. The alternative transcript starting from the U3 exon connects to U0 and E2 exons while skipping U2, U1 and E1 exons. Genomic distances are not drawn to scale.
Inspection of the sequence reads from four individual libraries provided the following conclusions (Table 3.2). First, detailed analyses confirmed the presence of four new upstream exons, thus named U0 through U3. The genomic positions of these exons relative to that of the known 1st exon (E1) are as follows: U3 (163 kb), U2 (26 kb), U1 (20 kb) and U0 (16 kb upstream of E1). These upstream exons are spread throughout the middle 200-kb genomic region of the Peg3 imprinted domain, which is quite unexpected and remarkable, as the previously identified exons of Peg3, E1-E9, spans only a 25-kb genomic distance (Figure 3.2 B). Second, the three exons U1, U2, and U3 have a clear exon-intron border at their 3’-ends, but not at the 5’-ends, suggesting that the 5’-end of these exons likely contain transcription start sites. This further suggests that the immediate upstream regions of these three exons should serve as alternative promoters for the Peg3 locus. This is also the case for E1 exon, which has long been known as the 1st exon for the Peg3 locus. On the other hand, U0 exon has a clear exon-intron border in either end, suggesting that this exon is likely connected downstream to the transcripts that start from its upstream regions starting from U1, U2, and U3. None of the transcripts starting from its upstream regions, U1, U2 and U3, are connected to E1. Instead, they all skip the E1 exon, and are directly connected to the E2 exon. This agrees with the fact that the 5’-end sequence of E1 contains a promoter and transcription start site, but not a splicing acceptor site. The exon structures of these newly identified transcripts are summarized in Figure 3.2 B.

Third, the newly identified transcripts U1, U2, and U3 all appear to contain a proper combination of the previously identified downstream exons of Peg3, including E2 through E6, based on a subset of the long sequence reads that had been derived from the libraries. Thus, we performed an independent series of RT-PCR experiments to determine the extent of the transcripts starting from U1, U2 and U3, particularly focusing on the 3’-ends of these transcripts.
These analyses confirmed that all the transcripts starting from U1, U2 and U3 indeed contain properly spliced downstream exons, including E2 through E9 (Figure 3.3).

Figure 3.3. RT-PCR analysis of alternative exons in the mouse brain. (A) A schematic of the mouse Peg3 locus with RT-PCR primer combinations. Gray and black boxes indicate the exons of Usp29 and Peg3, respectively. The red arrows indicate the directionality of primers: E9-R coupled with E1, U0, U1, U2, and U3 specific primers to amplify the respective exons. The deleted Peg3 exons corresponding to KO2(+/−p) and DelKO(+/−p) mutant alleles are shown using parentheses and dashed lines. The RT-PCR panel indicates the expression patterns of Peg3 using total RNA isolated from DelKO(+/−p) adult hypothalamus and KO2(+/−p) neonate brain tissues. The U3, U2, and U1 represent primer combinations targeting the upstream 1st exons of Peg3 with E9-R primer combination; U0 targets a shared upstream exon of Peg3 with the E9-R primer, while E1 targets the main promoter of Peg3 with the E9-R primer, to show the expression profile preferred by each exon. (B) Genotyping of the paternally transmitted KO2 allele. The schematic represents the mouse locus for the paternal and maternal alleles of Peg3. The arrows indicate the primer combinations used for PCR amplification. The dotted box represents the deleted region corresponding to the bidirectional promoter. (C) Structural map of the upstream alternative exons of Peg3. The blue rectangles represent Peg3, Usp29, and APeg3 exon structures within the 6658776-6703943 genomic region of mouse chromosomes 7, with arrows indicating the respective transcriptional directions. The U0-E9 exon structure represents the genomic region transcribed by the shared upstream alternative exon of Peg3, with arrows indicating its transcriptional direction. The UCSC genome browser was used to visualize the exon structure of U0-E9.
Since the known ORF (Open Reading Frame) of Peg3 spans from E3 through E9, this further suggests that these transcripts may serve as mRNA templates that produce the PEG3 protein. Fourth, each of the identified transcripts was represented differently compared to the other transcripts in the four libraries of 5’RACE products based on the number of its sequence reads (Table 3.2). The U2 represented the highest amount of transcripts (77%, 4267/5473), followed by U1 (8.17%, 447/5473) and U3 (0.09%, 5/5473), according to the results from the third library that had been derived from the WT neonatal brain. The remaining 14.7% (754/5473) products were non-specific sequence reads. This trend was also true for two other libraries derived from neonates, with U2 transcript showing the highest percentage among the newly identified transcripts. Interestingly, this was not the case in the library derived from the hypothalamus of the adult DelKO mice (Table 3.2, Figure 3.4). Instead, U2 transcript was not detected at all while the remaining U1 and U3 transcripts were well represented. It is relevant to note that the tissue source of this library was the adult hypothalamus, rather than the neonate brain, which was used for the other three libraries. Thus, U2 transcript might be neonatal-specific, which was further tested in the following section (Figure 3.5). On a different note, the relative representation of U2 transcript to the known main transcript (E1-E9) was 0.33 to 93%, indicating that the three identified transcripts are overall very minor compared to the primary transcript starting from E1, as shown by the results from the second library. Since these alternative transcripts were also detected from the wild-type animals, they likely represent the genuine transcripts of Peg3 rather than potential artifacts that could be generated as an outcome of knockout mutations. Scheme I and scheme II (Table 3.2) are different in terms of the gene-specific primers used for the respective nested PCR scheme as shown by Figure 2. Even in the case of Scheme II, two different nested primers were also used between two different libraries: a shared upstream exon
(U0) as an anchor for the DelKO and WT tissues (Table 3.2) versus the exon 4 as an anchor for the hypothalamus of adult DelKO (Figure 3.4 B). Thus, comparison of the results from these two schemes needs a caution. Taken together, this series of 5’RACE and NGS sequencing confirmed that the Peg3 locus harbors three alternative 1\textsuperscript{st} exons and promoters that have never been previously described. The three alternative 1\textsuperscript{st} exons and promoters are distributed throughout a relatively large genomic interval (the middle 200-kb region of the Peg3 imprinted domain), which as of yet has not been well characterized.

Figure 3.4. Alternative transcripts determined by 5’ RACE. (A) Map of the mouse Peg3 locus. Gray and black boxes indicate the exons of \textit{Usp29} and Peg3, respectively. Transcriptional direction for each gene is represented using arrows with corresponding colors. A solid red arrow indicates the position of the first exon of Peg3, labeled E1, followed by dotted arrows to indicate the position of upstream alternative exons. The ovals represent upstream exons U0, U1, U2, and U3, respectively. The extended arrow shows the gene-specific primer Ex6-R2 used for DelKO(+/p) mouse hypothalamus cDNA synthesis. (B) Percentage of Peg3 alternative transcripts identified from adult mouse hypothalamus. The Ex12,4 indicates the anchoring primer used for nested PCR. The percentage of transcripts preferred by DelKO(+/p) adult mouse hypothalamus was calculated by counting the sequences specific for E1, U0, U1, U2, and U3.
Allele and tissue specificity of the alternative promoters of *Peg3*

The three alternative transcripts of *Peg3* were further characterized in terms of their allele and tissue specificity using RT-PCR experiments (Figure 3.5).

![Figure 3.5](image-url). Allele and tissue specificity of the alternative transcripts of *Peg3*. (A) A schematic of the mouse *Peg3* locus with RT-PCR primer combinations. Gray and black boxes indicate the exons of *Usp29* and *Peg3*, respectively. The transcriptional direction for each gene is represented with arrows and corresponding colors above the map. The arrows below the map indicate the directionality of primers: *Peg3*-RT-Ex6-R3 was coupled with U0, U1, U2, and U3 specific primers to amplify the respective alternative exons. The DelKO\(^{(+/p)}\) mutant has a deletion on exon 6 of the *Peg3* paternal allele, enabling the detection of its maternal allele expression. Similarly, the DelKO\(^{(-m/+)}\) mutant has a deletion on exon 6 of the *Peg3* maternal allele, enabling the detection of *Peg3* paternal allele expression. (B) Allele, tissue, and stage specificity of upstream alternative promoters of *Peg3*. The left RT-PCR panel shows the maternal expression patterns of the identified alternative promoters using total RNA isolated from tissues of DelKO\(^{(+/p)}\) mice: OB (olfactory bulb), MB (midbrain), HT (hypothalamus), CB (cerebellum), TM (thymus), HR (heart), LG (lung), LV (liver), KD (kidney), FT (fat), TT (testis), OV (ovary), and NB (neonate head). The right RT-PCR panel shows the paternal expression patterns using total RNA isolated from tissues of DelKO\(^{(-m/+)}\) mice: OB (olfactory bulb), MB (midbrain), HT (hypothalamus), CB (cerebellum), and NB (neonate head). The U3-Ex6R3, U2-Ex6R3, and U1-Ex6R3 primer combinations amplified the upstream alternative 1\(^{st}\) exons of *Peg3*, whereas U0-Ex6R3 primer combination amplified a shared upstream exon of *Peg3* to show the expression
profile preferred by each upstream exon. The combination of exon 1 and exon 2 primers was used to explore the expression pattern of Usp29. Equal amounts of total RNA were used for RT-PCR, which were further normalized and visualized by β-actin expression levels. We repeated this analysis using three independently derived replicates.

For allele specificity, an RT-PCR scheme targeting exon 6 was employed to differentiate the two alleles using the DelKO mutant, which contains a deletion in its exon 6 of Peg3 (Figure 3.5 A). The tissues isolated from DelKO(+/p) and DelKO(-/+m) were used to test the maternal and paternal expression of the alternative transcripts, respectively (Figure 3.5 B). For tissue specificity, total RNA was isolated from twelve individual tissues of adult mice, including the olfactory bulb (OB), midbrain (MB), hypothalamus (HT), cerebellum (CB), thymus (TM), heart (HR), lung (LG), liver (LV), kidney (KD), fat (FT), testis (TT), and ovary (OV). For stage specificity, total RNA was also isolated from one-day-old neonatal brains (NB), and the results from these neonatal brains were compared with those from the four different parts of the adult brains (OB, MB, HT, CB). Two sets of total RNA from DelKO(+/p) and DelKO(-/+m) were first reverse-transcribed, normalized with the expression level of an internal control (β-actin), and finally used for the expression analyses of the alternative transcripts.

The results from this series of expression analyses are summarized as follows. First, U3 transcript was detected at very low levels only in the hypothalamus (HT) and cerebellum (CB) of the adult mice, but not in the neonatal brains (NB), indicating its specific expression in HT and CB in the adult stage. This expression pattern was also independently confirmed through the samples from wild-type mice, showing exclusive expression of U3 transcript in the mid-brain and cerebellum (Figure 3.6 B). More importantly, U3 transcript was detected only in the tissues of DelKO(+/p), indicating maternal allele-specific expression. Second, U2 transcript appeared to
be neonatal-specific, as expression was detected only in the two neonatal brain samples, DelKO\(^{(+/-p)}\) and DelKO\(^{(-m/+)}\).

**Figure 3.6.** RT-PCR analysis of alternative exons in wild-type mouse tissues. (A) A schematic of the mouse Peg3 locus with RT-PCR primer combinations. Gray and black boxes indicate the exons of Usp29 and Peg3, respectively. Transcriptional direction for each gene is represented using arrows with corresponding colors. The red arrows indicate the directionality of primers: Ex6-R3 coupled with E1, U0, U1, U2, and U3 specific primers to amplify the respective exons. (B) Allele, tissue, and stage-specificity of upstream alternative exons of Peg3. The RT-PCR panel shows the expression patterns of Peg3 using total RNA isolated from tissues of wild-type mice: OB (olfactory bulb), MB (midbrain), HT (hypothalamus), CB (cerebellum), TM (thymus), HR (heart), LG (lung), LV (liver), KD (kidney), FT (fat), TT (testis), OV (ovary), NB (neonate brain), MEF (mouse embryonic fibroblasts), E14.5 (Embryo14.5-dpc.), and PC14.5 (Placenta 14.5-dpc.). U3-Ex6R3, U2-Ex6R3, and U1-Ex6R3 primer combinations target the upstream 1st exons of Peg3; U0-Ex6R3 primer combination targets a shared upstream exon of Peg3, whereas E1-Ex6R3 primer combination targets the main promoter of Peg3 to show the expression profile preferred by each exon. The combination of exon 1 and exon 2 primers was used for the expression pattern of Usp29. The β-actin expression profile serves as a control to visualize the relative mRNA levels.
This also indicated bi-allelic expression of U2 transcript although the expression levels from the paternal allele were much higher than those from the maternal allele. Thus, the expression of U2 transcript is bi-allelic but with a bias toward the paternal allele. Independent analyses also indicated that U2 transcript is not detectable in 14.5-dpc embryos and placentas, further confirming its neonatal-specific expression (Figure 3.6 B). Third, the expression of U1 transcript was detected in the hypothalamus (HT) and also in the ovary (OV). Interestingly, U1 transcript was present in the hypothalamus (HT) of DelKO^{(+/-p)} and absent from that of DelKO^{(-m/+)}, thus indicating a maternal allele-specific expression pattern for U1, similar to U3 transcript. The expression of U1 transcript in ovary (OV) was further investigated by performing an independent RT-PCR and also 5’ RACE experiments using mature oocytes. U1 transcript was not detected at all in mature oocytes (data not shown), indicating that the observed expression of U1 transcript likely originated from either the early-stage oocytes or the somatic cells of ovary. Fourth, a series of RT-PCR surveying U0 exon detected the expression from the hypothalamus (HT), cerebellum (CB), and ovary (OV) of DelKO^{(+/-p)} and also from the neonatal brains (NB) of both DelKO^{(+/-p)} and DelKO^{(-m/+)}. This is reasonable since the three alternative transcripts, U1, U2, and U3, which contain the U0 exon, were detected from the same tissues that show the U0 transcript. Interestingly, the same RT-PCR also detected the expression from the testis (TT), which was not detected by the previous RT-PCR surveying U1, U2, U3 transcripts. This suggests that additional unknown transcripts may exist in testis. Multiple minor PCR products observed throughout this expression analysis (U3, U2, U1 and U0) were confirmed to result from splicing variations (Figure 3.5 B). Taken together, this series of expression analyses concluded that U1 and U3 transcripts are maternal allele-specific whereas U2 is bi-allelic with a bias toward the paternal
allele. In addition, the main expression sites of these alternative transcripts include the hypothalamus for U1 and U3 and the neonatal brain for U2.

**DNA methylation analysis of the alternative promoters of Peg3**

DNA methylation status of the three alternative promoters was further analyzed given their allele and tissue-specific expression patterns. Genomic DNA was first isolated from four different parts of the brain: olfactory bulb (OB), hypothalamus (HT), midbrain (MB), cerebellum (CB), as well as kidney (KD) of a wild-type female adult mouse (Figure 3.7). The isolated DNA was treated with the bisulfite conversion protocol [26], and the converted DNA was used as a template for PCR amplification. Four different regions were targeted for DNA methylation analyses: E1, U1, U2 and U3 promoter regions. The amplified PCR products were analyzed first by COBRA (Combined Bisulfite Restriction Analysis [27]), and later by individual sequencing (Table 3.3, Figure 3.7).

**Table 3.3. Summary of the alternative promoter positions for DNA methylation analysis.**

<table>
<thead>
<tr>
<th>Upstream promoter/exon (Chr: 7)</th>
<th>Position (NCBI37/mm9) direction towards Peg3</th>
<th>Position used for DNA methylation analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>U3</td>
<td>5’ (~157 bp) – 6,846,590 3’ 5’ 6,846,749 – 6,846,565 3’</td>
<td></td>
</tr>
<tr>
<td>U2</td>
<td>5’ (~ 79 bp) – 6,709,554 3’ 5’ 6,760,018 – 6,759,695 3’</td>
<td></td>
</tr>
<tr>
<td>U1</td>
<td>5’ (~178 bp) – 6,703,769 3’ 5’ 6,704,151 – 6,703,900 3’</td>
<td></td>
</tr>
<tr>
<td>U0</td>
<td>5’ 6,699,550 – 6,699,412 3’ N/A</td>
<td></td>
</tr>
<tr>
<td>E1</td>
<td>5’ (~164 bp) – 6,682,968 3’ 5’ 6,846,749 – 6,846,566 3’</td>
<td></td>
</tr>
</tbody>
</table>

The promoter region of an imprinted gene usually derives 50% methylation level due to its allele-specific methylation status: one allele is methylated while the other allele is unmethylated. The allele-specific methylation pattern is very uniform and complete among individual CpG sites in a given promoter and also among individual tissues. This is the case for
E1 promoter, which is part of the 4-kb Peg3-DMR, showing 50% methylation level among all the tissues tested.

Figure 3.7. DNA methylation analysis of alternative promoters of Peg3. (A) DNA methylation analysis of the E1, U1, U2, and U3 promoter regions. Genomic DNA was purified from the different parts of the brain (olfactory bulb, hypothalamus, midbrain, cerebellum) and kidney of a two-month-old female mouse, and then used for bisulfite conversion. The bisulfite-converted DNAs were subsequently amplified with PCR using specific primer sets designed for each promoter region (Table S1). The amplified PCR products were analyzed by COBRA. The restriction enzymes used for each digestion is shown on the left side of the panel, while the right side of the panel indicates the promoter regions under investigation. The red lines indicate methylated DNA whereas the blue lines indicate unmethylated DNA. The methylation levels of U2 and E1 promoter regions were calculated using the ImageJ software for three independent trials. (B) Quantitative methylation analysis of U1 and U3 promoter regions. The bisulfite-converted DNAs were amplified with PCR and used for NGS-based deep sequencing to obtain the methylation level for each adult mouse tissue. A single red box and a blue box indicate a single read for methylated and an unmethylated CpG site, respectively. Each column in the methylation array represents a single CpG site for the respective tissue sample. The U3 and U1 promoter regions represent DNA methylation changes observed from four and six CpG sites, respectively. The overall percentage of methylation is indicated at the bottom of each tissue for comparison. The number of reads are represented by the (n) underneath each tissue sample.
On the other hand, the methylation levels and patterns of the three alternative promoters are quite different from those from E1. First, U1 promoter also showed around 50% methylation levels, but the levels were variable among the tissues, ranging from 43% in the hypothalamus (HT) to 61% in the cerebellum (CB). Detailed inspection further indicated that the methylation level and pattern of each CpG site were also variable compared to that of the remaining CpG sites. This is quite different from the uniform and complete pattern observed from E1 promoter, thus suggesting that the DNA methylation on U1 promoter may not be of allelic origin. The same may be the case for U3 promoter, as its individual CpG sites displayed fluctuating levels of DNA methylation. Compared to the U1 promoter, the overall methylation levels of U3 promoter were greater, ranging from 68% in the hypothalamus (HT) to 82% in the olfactory bulb (OB) and kidney (KD). This is somewhat consistent with the higher expression levels observed from U1 transcript than those from U3 transcript (Figure 3.5 B). Also, the methylation levels of both U1 and U3 promoters in the hypothalamus were the lowest among the tissues tested, which also agrees with the fact that the expression of U1 and U3 transcripts were observed only from the hypothalamus. Finally, U2 promoter displayed the greatest DNA methylation levels among the alternative promoters, ranging from 85.4% in the cerebellum (CB) to 92.3% in the olfactory bulb (OB). This is again consistent with the fact that the expression of U2 transcript was not detectable at all in the adult tissues (Figure 3.5 B). Overall, this series of analyses concluded that the methylation levels and patterns of the alternative promoters are variable among individual tissues and also among individual CpG sites. Thus, the DNA methylation on the alternative promoters is somewhat different from the typical pattern of allelic origin, although the expression of U1, U2, and U3 transcripts appears to be biased in either the maternal or the
paternal allele. Nevertheless, this aspect of the results is inconclusive at the moment, which requires further investigation in the near future.

**Identification of the alternative 1st exon and promoter of human PEG3**

The genomic structure of the Peg3 imprinted domain is well conserved among individual mammals, including humans [2, 8]. Thus, we tested the presence of potential alternative 1st exons and promoters of human PEG3. This test also used a similar strategy involving 5’RACE and NGS sequencing as described above (Figure 3.8 A). The total RNA isolated from adult and neonate brain tissues was first reverse-transcribed with a gene-specific primer targeting exon 2 of human PEG3 (Ex2-R1), and later these cDNA fragments were further amplified with a set of nested primers (Ex2-R2 and Ex2-R3). The two amplified libraries were sequenced using a NGS platform. The results are summarized in Figure 3.8 B. According to the results, about 74% of the sequence reads from the adult brain (255,571/342,369) belong to the known 1st exon of human PEG3 (E1). On the other hand, 0.13% of the sequence reads (433/342,369) derived from a 120-bp region that is located 2.4-kb upstream region of E1 exon. Yet, these sequences were not connected to the E1 exon, but directly connected to the 2nd exon of PEG3 (E2). Also, this set of cDNA sequences shows a similar pattern as seen from the three alternative promoters of mouse Peg3, as a clear exon-intron border exists in their 3’-end sequences but not in their 5’-end sequences. Therefore, this exon is predicted to be an alternative 1st exon of human PEG3, thus named U1. Interestingly, this U1 exon was not detected in the library derived from the neonate brain, thus indicating an adult-specific expression (Figure 3.8 B). To further confirm this possibility, we also performed a set of RT-PCR testing the expression pattern of U1 exon (Figure 3.8 C). As expected, the expression levels of the human PEG3 transcript starting from E1 exon
were very high in both adult and neonatal brains (AB and NB), and also in the other adult tissues, including heart (HR), liver (LV), kidney (KD) and placenta (PC).

Figure 3.8. Upstream alternative promoter of human PEG3. (A) A schematic of the human PEG3 locus with RT-PCR primer combinations. Gray and black boxes indicate the exons of MIMT1 and PEG3, respectively. Transcriptional direction for each gene is represented using arrows with corresponding colors. The black arrow indicates the directionality of Ex2-R1, the primer used for cDNA synthesis. The open oval represents the alternative 1st exon U1. (B) 5’ RACE analysis for upstream exons of PEG3. The Ex2-R2 indicates the anchoring primer used for nested PCR after 5’ RACE. This primer was coupled with U1 and E1 specific primers to amplify the respective exons. The percentage of transcripts preferred by the adult human brain was calculated by counting the sequences specific for U1 and E1. (C) RT-PCR analysis of upstream alternative exons in human tissues. The RT-PCR panel shows the expression patterns of PEG3 using cDNA from AB (adult brain), NB (neonate brain), HT (adult heart), LV (adult liver), KD (adult kidney) and PC (adult placenta). U1-Ex2R2 and E1-Ex2R2 primer combinations amplified the upstream exons of PEG3 to show the expression profile preferred by each 1st exon. The Ex1-Ex2 primer combination amplified the expression pattern of MIMT1 in the AB and NB. Equal amounts of total RNA were used for RT-PCR, which were normalized by β-actin expression levels.
On the other hand, the expression of the human $PEG3$ transcript starting from U1 was detected at very low levels only from the adult brain (AB) and placenta (PC). This low level of expression in adult brain agrees with the initial detection of this exon from the library that had been derived from adult brain. Also, this appears to be somewhat similar to the mouse U1 and U3 transcripts since they are also expressed only in the adult stage. Overall, this series of 5’RACE and expression analyses confirmed that human $PEG3$ also harbors an alternative 1$^{\text{st}}$ exon and promoter. Thus, alternative 1$^{\text{st}}$ exons and promoters may be evolutionarily conserved features that are associated with the mammalian Peg3 domain.

### 3.4 Discussion

In the current study, we have identified three alternative promoters for mouse $Peg3$ and one alternative promoter for human $PEG3$. These alternative promoters are localized within the 200-kb upstream regions of mouse and human $PEG3$, which are well conserved among mammals in terms of genomic structure. In mouse, two of these promoters, U1 and U3, derive maternal-specific expression whereas the remaining promoter, U2, derives bi-allelic expression of $Peg3$ with a paternal allele bias. These promoters are also tissue and stage-specific: U1 and U3 transcripts are detected in the hypothalamus of adult mice whereas U2 transcript only in neonatal-stage brains. In humans, U1 transcript is detected at relatively very low levels in adult brain and placenta. Overall, the identification of alternative promoters in both mouse and human suggest that these alternative promoters may be functionally selected features for the $Peg3$ imprinted domain during mammalian evolution.

The current study has identified alternative promoters for human and mouse $PEG3$ with a strategy involving 5’RACE and NGS-based deep sequencing experiments (Figure 3.2, Figure
The identified alternative promoters are all located within the 200-kb genomic interval that is upstream of the Peg3 locus (Figure 3.9).

This 200-kb genomic interval has been well conserved during mammalian evolution although this region lacks an obvious ORF [8]. According to recent studies, this genomic interval contains 18 evolutionarily conserved regions (ECRs), and these ECRs are all associated with H3K4me1 and/or H3K27ac [9]. Thus, these ECRs are thought to be potential enhancers. Some of the alternative promoters are in fact closely associated with these ECRs. For instance, U2 is localized between ECR3 and ECR4, whereas U3 is located nearby ECR18. This may be an
indication that some of the ECRs are enhancers for these alternative promoters. On the other hand, a similar series of inspection was unable to confirm the association of any histone marks, such as H3K4me3, with these alternative promoters, which might be caused by their very low abundance. Recent studies from the bovine genome also revealed that a deletion of a 20-kb genomic interval encompassing U2 promoter might be responsible for the stillbirths observed among the calves that had been derived through artificial insemination [28]. This observation is intriguing in that U2 promoter is very specific in neonatal-stage brains (Figure 3.5). This further suggests that these alternative promoters may play very unique, but critical roles for the transcription control of Peg3. Overall, the identification of alternative promoters in mouse and human PEG3 appears to provide an answer for a long-standing, puzzling question: why is the middle 200-kb genomic region of the Peg3 domain preserved so well during mammalian evolution? This region most likely harbors several cis-regulatory elements critical for the Peg3 domain, and one set of these should be the newly identified alternative promoters for human and mouse PEG3.

The three alternative promoters of mouse Peg3 are quite interesting in terms of their expression patterns (Figure 3.5). First, one of the main expression sites shared among the three promoters is the hypothalamus. It is well known that Peg3 is highly expressed in the hypothalamus [1, 3, 10, 22], and the in vivo functions of Peg3 are closely associated with several roles played by this part of brain, such as milk provision and maternal-caring behavior [10]. The similar expression pattern shared between the new alternative promoters (U1, U3) and the known major promoter (E1) suggests that U1 and U3-driven transcripts likely play similar roles as the main transcript by E1 promoter. This further suggests that unknown cis-regulatory elements may exist for this hypothalamus-specific expression of these promoters. Second, U2 promoter shows
a highly sensitive, stage-specific expression, showing its expression exclusively in one-day-old neonates, but not in 14.5-dpc embryos or adult tissues (Figure 3.6). Thus, this promoter might be functional only during a particular stage when neonates’ brain is known to undergo a series of major organization processes, especially in the hypothalamus region [29, 30]. Therefore, the main roles of this promoter might be associated with many changes occurring during this transition period of brain development. Third, two of the alternative promoters, U1 and U3, show maternal allele-specific expression patterns. It is puzzling at the moment why these alternative promoters derive maternal-specific expression for the paternally expressed Peg3 in the hypothalamus. One possible explanation would be that some of cell populations within the hypothalamus might require additional dosage of Peg3. As a consequence, Peg3 might be bi-allelic in those cell populations: maternal by U1 and U3 and paternal by E1 promoter. Since the alternative promoters reside within the paternally expressed Usp29 transcriptional region, transcriptional interference by Usp29 could be another possible explanation for the allele-specific expression patterns observed by U1, U2 and U3 [31]. In spite of the observed allele-specific expression, however, DNA methylation pattern suggests that it is unlikely of allelic origin (Figure 3.7). Nevertheless, the absence of sequence polymorphisms in the alternative promoter regions between different subspecies of the mouse hampered the inquiry of allele-specific DNA methylation. Overall, the newly identified alternative promoters exhibit very interesting expression profiles, which have not been observed before. Thus, characterizing these promoters should provide additional insights regarding the in vivo function of Peg3.

Alternative promoters have also been identified from other imprinted domains, such as Grb10, Snrpn/Ube3a, Gnas and Zac1 domains. Interestingly, the two alternative promoters for a given locus tend to be opposite in terms of their allele-specific expression. The examples include
maternal and paternal-specific promoters in *Grb10* and *Gnas* [15-17, 20]. Because of the current study, this list includes the maternal and paternal-specific promoters for the *Peg3* locus as well. According to the observations from other domains, several *in vivo* roles are possible for the alternative promoters of the *Peg3* locus. First, the two promoters for *Grb10* have two complementary expression patterns. The paternal and maternal promoters of *Grb10* are responsible for its expression in neuronal and non-neuronal cells, respectively [32]. Thus, this might be feasible for the alternative promoters for *Peg3*: the paternal and maternal promoters could function in different population of cells within the hypothalamus. Second, one upstream alternative promoter for the *Snrpn* locus is also involved in regulating the allele-specific expression of the nearby gene, *Ube3a*, which is located 500-kb downstream of *Snrpn*. This particular promoter derives the paternal-specific transcription of a long non-coding *Ube3a-ATS* [12]. Furthermore, this antisense transcript is thought to be responsible for maintaining the maternal-specific expression of *Ube3a* [13, 14]. The alternative promoters of *Peg3* might also play similar roles. In particular, U3 promoter is located 160-kb upstream of *Peg3*, yet the transcription initiated from this promoter extends all the way downstream to the *Peg3* locus. It is possible that the transcription starting from U3, particularly from the maternal allele, might be related to the expression of two maternally expressed downstream genes of *Peg3*, *Zim1* and *Zim2*. Besides these two possible roles, the alternative promoters might be also involved in the DNA methylation setting for the Peg3-DMR during oogenesis. A similar situation is known to occur in the *Gnas*, *Snrpn*, and *Zac1* loci [13, 14, 17, 19]. According to a deep transcriptome sequencing study of the mouse oocytes, several alternative promoters were identified for individual imprinted genes including *Peg3* [19]. However, none of the three alternative promoters of the *Peg3* locus seem to be functional and thus detectable in mature oocytes, based
on our preliminary results obtained so far (data not shown). Thus, these promoters might not play a similar role in DNA methylation setting as seen in these loci. Overall, the information from the other imprinted domains appears to provide several hints regarding the potential roles that might be played by the alternative promoters of Peg3. Thus, it should be very interesting to pursue these aspects of the identified alternative promoters in the near future.

3.5 References


CHAPTER 4
NGS-BASED 5’RACE FOR ALTERNATIVE PROMOTERS

4.1 Introduction

In eukaryotic cells, gene transcription is regulated through several cis-regulatory DNA elements, such as promoters, enhancers, silencers and insulators [1]. Identifying promoters is one of the initial steps for characterizing transcriptional regulation for any given gene. In model organisms, the promoters of the majority of genes have been identified through large-scale sequencing efforts [2]. According to recent results, however, transcripts are quite often detected from the upstream regions of many genes that are outside of the previously defined transcribed regions. This indicates that additional transcription start sites and alternative promoters may exist for many genes [3]. In the current study, thus, we have modified and adapted the existing 5’RACE (Rapid Amplification of cDNA Ends, [4]) approach into an NGS (Next Generation Sequencing)-based new protocol that can identify the alternative promoters for these transcripts.

The main strength of this new protocol is its sensitivity: this can detect the 5’-ends of mRNA that are of very low abundance. The main ideas and detailed steps for this protocol are described below.

Two main ideas have been incorporated into this new protocol. First, the initial cDNA for 5’RACE needs to be prepared with the reverse transcription reaction using gene-specific oligonucleotide primers (Figure 4.1 B). These gene-specific primers also need to be derived from the known 2nd exons, but not from the known 1st exons for genes. The known 1st exons are not a good choice since these 1st exons do not have splicing acceptor sites at their 5’ ends. Instead,
they have transcription start sites, thus they cannot be connected to any upstream exons.

Figure 4.1. Overall scheme for the NGS-based 5’RACE protocol. (A) A flow chart shows the entire procedure of the NGS-based 5’RACE protocol. (B) A schematic illustration represents the genomic structure of a given locus with its promoters and exons being indicated by triangles and rectangles, respectively. The bent arrow indicates the transcriptional direction of the gene. The previously known transcript and its alternative transcript are shown in black and light blue, respectively. The total RNA isolated from the desired tissue is first reverse-transcribed using gene-specific primers, as indicated by the black arrow underneath the exon structure. The resulting cDNA product is then purified and modified with the tailing reaction using dGTP and TdT, as indicated by the extended blue and black arrows with a ‘G’ mark. The cDNA is then used as a template for a nested PCR scheme I with the following two primers: an internal gene-specific primer (gray arrow) and a primer made of a unique sequence plus a C-tail (indicated by XC). The product from PCR I is subsequently used for PCR II with another internal gene-specific primer (gray arrow) and the second primer specific for the unique sequence (arrow indicated by X). The resulting products are end-repaired and ligated to barcoded adapters as shown by the green boxes. A set of several 5’RACE libraries can be pooled together for NGS-based sequencing. The sequences from NGS runs are sorted based on the barcoded primer portion of the cDNA products. These sorted sequences are further divided into individual groups based on their different combinations of exon joining. These groups of the sorted sequences are analyzed to examine their exon structures using the UCSC genome browser. Each group with a different exon joining can also be analyzed for the frequency (or representation) in a given library.

By contrast, the known 2nd exons have proper splicing acceptor sites, thus should be ideal regions anchoring the upstream alternative exons that have not been identified so far. Second,
potential unknown alternative 1st exons most likely represent the sequences that belong to very low abundant transcripts, given all the sequencing efforts by the genomic community. Therefore, this new protocol utilizes NGS-based approaches to identify the sequences derived from these very minor transcripts. NGS-based sequencing should allow the identification of these low abundant transcripts given their sequencing capacity, million reads per library. This new protocol, therefore, includes several steps that can easily convert PCR products into NGS libraries.

4.2 Materials and Methods

The detailed steps for the new protocol are as follows (Figure 4.1 A). In Step 1, the total RNA isolated from tissues samples is reverse-transcribed with gene-specific primers (dark arrow in Figure 4.1 B). The subsequent cDNA needs to be purified through phenol/chloroform extraction followed by ethanol precipitation. In Step 2, the cDNA is modified through the tailing reaction using dGTP and terminal deoxynucleotidyl transferase. In Step 3, a pool of the G-tailed cDNA is targeted and enriched for a given gene through a nested PCR scheme involving two sets of primers: two gene-specific primers (gray arrows) and two primers targeting the G-tailed portion of cDNA (XC and X; blue arrows). A small amount of PCR products from each PCR needs to be separated on a 2% agarose gel to monitor the proper amplification of cDNA. In Step 4, the amplified product from the 2nd PCR is modified with the end-repair reaction, and later with the ligation reaction to be analyzed with NGS sequencing. The current protocol has been tested multiple times using a NGS platform (PGM2, Ion Torrent), which requires two specific sequences at either end of all the DNA fragments to be analyzed (A and P adaptors). Thus, the two adaptors containing these sequences need to be ligated onto either end of the cDNA. An
NGS platform usually allows the simultaneous sequencing of multiple libraries, thus several cDNA libraries derived from multiple tissues are also ligated individually with a set of adaptors with different barcodes. In Step 5, the pool of individual barcoded libraries is amplified with PCR, and subsequently sequenced using a NGS platform. In Step 6, the raw sequence reads from NGS runs are sorted first by their barcodes, and later by the sequences of the two primers: X and the gene-specific primer used for the 2nd PCR of the nested scheme (Figure 4.1 B). This sorting process can be executed through several Unix command lines, which has been included below. These filtered reads are further processed to identify alternative exons using a PERL script, which can detect and count the number of the raw reads that have different combinations of exon joining with the initial exon (E2 in Figure 4.1 B). In Step 7, each group of the raw reads displaying different exon joining can be mapped to the genome sequence of a given species using the UCSC genome browser, which will then identify new alternative exons.

**Reagents**

A Adapter with barcode:

5’-CCATCTCATCCCTGCGTGTCCTCGAGTCXXGAT-3’ (X, 10mer barcode)

A adaptor primer: 5’-CCATCTCATCCCTGCGTGTC-3’

Bst2.0 WarmStart DNA polymerase (includes 10X Isothermal Amplification Buffer; New England Biolabs. Cat. No. M0538S)

DNA Clean & Concentrator™-25 (includes spin columns, DNA Binding Buffer, Wash Buffer, Elution Buffer, Zymo, Cat. No. D4005)

Gene-specific primers for the gene of interest
NEBNext End Repair Module (includes NEBNext End Repair Enzyme Mix, NEBNext End Repair Reaction Buffer; New England Biolabs, Cat. No. E6050S)

P1 Adaptor:

5’-CCACTACGCCTCCGCTTTTCTCTCTATGGGCGAGTCGGTGAT-3’

P1 Adapter Primer: 5’-CCACTACG CCTCCGCTTTCC-3’

Phenol/Chloroform from Sigma (cat: P2069-200ML)

Reverse transcription reagents from New England Biolabs (NEB)

T4 DNA Ligase (includes 10X T4 DNA Ligase Reaction Buffer; New England Biolabs, Cat. No. M0202S)

XC primer: 5’-GGTTGTGAGCTCTTCTAGATCCCCCCCCCCCCNN-3’

X primer: 5’-GGTTGTGAGCTCTTCTAGA-3’

TE buffer (1X, pH 8.0)

Terminal Transferase from NEB (cat: M0315S)

Trizol reagent from Life Technologies (cat: 15596-018)

Zymoclean Gel DNA Recovery Kit (includes spin columns, ADB Buffer, Wash Buffer, Elution Buffer; Zymo, Cat. No. D4001)

**Equipment**

2100 Agilent Bioanalyzer System

Agarose gel (2%) and equipment for electrophoresis

Centrifuge (Benchtop centrifuge for 1.5ml microcentrifuge tubes)

E-Gel Precast Agarose Electrophoresis System

Ion torrent PGM NGS machine

Thermocycler
cDNA synthesis and purification

1. Isolate RNA from the tissue of interest using Trizol reagent according to the manufacturer’s protocol. Use equal amounts of RNA to synthesize cDNA using gene specific primers.

2. Prepare the cDNA synthesis reaction as shown below.

<table>
<thead>
<tr>
<th>Component</th>
<th>Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total RNA (5µg)</td>
<td>Xµl</td>
</tr>
<tr>
<td>dNTP (2.5mM)</td>
<td>8µl</td>
</tr>
<tr>
<td>Gene-specific primer (10µM)</td>
<td>2µl</td>
</tr>
<tr>
<td>H₂O</td>
<td>Yµl</td>
</tr>
<tr>
<td>Total</td>
<td>32µl</td>
</tr>
</tbody>
</table>

3. Incubate at 65°C for 5 mins in the thermocycler.

4. Prepare the following items to complete the reverse transcription.

<table>
<thead>
<tr>
<th>Component</th>
<th>Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>Template from (3)</td>
<td>32µl</td>
</tr>
<tr>
<td>5X RT buffer</td>
<td>8µl</td>
</tr>
<tr>
<td>RNase Inhibitor (10U/µL)</td>
<td>2µl</td>
</tr>
<tr>
<td>DTT (0.1M)</td>
<td>4µl</td>
</tr>
<tr>
<td>M-MuLV RT enzyme (200u/µL)</td>
<td>2µl</td>
</tr>
<tr>
<td>Total</td>
<td>48µl</td>
</tr>
</tbody>
</table>
5. Incubate at 42°C for 1 hour for reverse transcription and 90°C for 10 mins to inactivate the enzyme with the thermocycler.

6. Mix the cDNA with an equal amount of 1X TE. Monitor the cDNA quality with PCR using gene specific primers. Check the amplified PCR products through a 2% agarose gel electrophoresis at 100V for 30 mins.

<table>
<thead>
<tr>
<th>Cycle number</th>
<th>Denature</th>
<th>Anneal</th>
<th>Extension</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>95°C, 2min</td>
<td></td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>95°C, 30s</td>
<td>60°C, 30s</td>
<td>72°C, 60s</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td>72°C, 7min</td>
</tr>
</tbody>
</table>

7. Perform phenol/chloroform extraction. First, bring the cDNA volume up to total 200µl with 1X TE. Add 100µl of phenol/chloroform, vortex, and centrifuge at 13,000 rpm for 6 mins. Save the top 150µl aqueous solution for the next step.

8. Precipitate the cDNA with 100% ethanol at -20°C for 1 hour as described below.

<table>
<thead>
<tr>
<th>Component</th>
<th>Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>cDNA template (7)</td>
<td>150µl</td>
</tr>
<tr>
<td>Sodium Acetate</td>
<td>15µl</td>
</tr>
<tr>
<td>Glycogen</td>
<td>3µl</td>
</tr>
<tr>
<td>100% Ethanol</td>
<td>300µl</td>
</tr>
<tr>
<td>Total</td>
<td>468µl</td>
</tr>
</tbody>
</table>

9. Centrifuge the cDNA solution at 13,000 rpm for 15 mins at 4°C, and pipet the supernatant out of the tube.
10. Add 1mL of 70% ethanol, centrifuge the sample(s) at 13,000 rpm for 10 mins at 4°C, and pipet out the supernatant. Air-dry the pellet for 10 mins until the ethanol is evaporated completely.

11. Resuspend the pellet with 37µl of H2O. This can be stored at -20°C for a few days.

**Tailing reaction and nested PCR procedure**

12. Prepare the following items to complete the tailing reaction.

<table>
<thead>
<tr>
<th>Component</th>
<th>Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>Purified cDNA template (11)</td>
<td>37µl</td>
</tr>
<tr>
<td>Cobalt Chloride (2.5mM)</td>
<td>5µl</td>
</tr>
<tr>
<td>dGTP (1mM)</td>
<td>2.5µl</td>
</tr>
<tr>
<td>TdT buffer (10x)</td>
<td>5µl</td>
</tr>
<tr>
<td>TdT enzyme</td>
<td>0.5µl</td>
</tr>
<tr>
<td>Total</td>
<td>50µl</td>
</tr>
</tbody>
</table>

13. Incubate at 37°C for 30 mins and later 70°C for 10 mins to stop the reaction using the thermocycler.

14. For the nested PCR 1, use 1µl of the G-tailed cDNA fragments with the XC primer (5’-GGTTGTGAGCTCTTCTAGATCCCCCCCCCN3’) and an internal gene-specific primer to amplify the desired transcript. Monitor the amplified products with a 2% agarose gel electrophoresis at 100V for 30 mins. This same PCR amplification conditions were applied as step 6 (this reaction employed 36 cycles).
15. For the nested PCR 2, use 1µl of the nested PCR 1 product with the X primer (5’-GGTTGTGAGCTCTTCTAGA-3’) and another internal gene-specific primer to amplify the desired transcript. Monitor the PCR products on a 2% agarose gel electrophoresis.

<table>
<thead>
<tr>
<th>Cycle number</th>
<th>Denature</th>
<th>Anneal</th>
<th>Extension</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>95°C, 2min</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>95°C, 30s</td>
<td>60°C, 30s</td>
<td>72°C, 60s</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td>72°C, 7min</td>
</tr>
</tbody>
</table>

Prepare the libraries for NGS

16. Prepare the end repair reaction as described below.

<table>
<thead>
<tr>
<th>Component</th>
<th>Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCR product (15)</td>
<td>Xµl</td>
</tr>
<tr>
<td>End repair enzyme mix</td>
<td>5µl</td>
</tr>
<tr>
<td>End repair reaction buffer (10X)</td>
<td>10µl</td>
</tr>
<tr>
<td>H₂O</td>
<td>Yµl</td>
</tr>
<tr>
<td>Total</td>
<td>100µl</td>
</tr>
</tbody>
</table>

17. Incubate at 20°C for 30 mins in a thermocycler

18. Prepare a 2% agarose gel for size selection.

19. Load the entire DNA sample into one or two wells along with a 100-bp size marker in a separate well, and run the gel at 130V for 40 mins.
20. Excise a section of the agarose gel containing the PCR products with the desired size range (<400bp length), and extract the DNA from the gel using Zymoclean Gel DNA Recovery Kit. Finally, elute the DNA in 20µl of the elution buffer.

21. Prepare the adapter ligation reaction as described below.

<table>
<thead>
<tr>
<th>Component</th>
<th>Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>Size selection sample</td>
<td>17µl</td>
</tr>
<tr>
<td>Bst2.0 warm polymerase (8000 U/ml)</td>
<td>1µl</td>
</tr>
<tr>
<td>Bst2.0 polymerase buffer (10X)</td>
<td>4µl</td>
</tr>
<tr>
<td>T4 DNA ligase (400,000 U/ml)</td>
<td>4µl</td>
</tr>
<tr>
<td>T4 DNA ligase buffer (10X)</td>
<td>4µl</td>
</tr>
<tr>
<td>P1+A Barcode Adapter (10 pmole/ul)</td>
<td>10µl</td>
</tr>
<tr>
<td>Total</td>
<td>40µl</td>
</tr>
</tbody>
</table>

22. Incubate the sample at 25°C for 25 mins, 65°C for 30 mins, and finally 80°C for 20 mins to stop the reaction.

23. Repeat the size selection step as described above (20-21) to exclude any adapter dimers.

Elute the DNA sample in 6µl of H₂O.

24. Perform PCR to amplify the eluted DNA with a set of the two primers corresponding to P and A1 adapter: P1 adapter primer, 5’-CCACTACGCTCCGTGTTTCTTCT-3’; A adapter primer, 5’-CCATCTCATCCCTGCCTGTCTCC-3’.

25. The detailed program for PCR is shown below. The typical number of PCR cycles ranges from 10 to 14.

<table>
<thead>
<tr>
<th>Cycle number</th>
<th>Denature</th>
<th>Anneal</th>
<th>Extension</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>95°C, 5min</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

93
26. Clean up the final PCR product with DNA Clean & Concentrator Kit.

27. Size-select the amplified library with E-Gel Precast Agarose Electrophoresis system.

28. Clean up the selected DNA library with DNA Clean & Concentrator Kit.

29. Perform NGS-based sequencing with the prepared library.

**Bioinformatics process**

The raw sequences with the different barcodes are first sorted and provided as individual fastq files from NGS platform. These fastq files are processed in the following manner to identify alternative exons and promoters.

1. Count the number of the cDNA sequences that contain the sequences of both the G-tailed portion and the 2\textsuperscript{nd} nested gene-specific primer portion with the following command line (grep –E ‘X part of the G-tailed primer sequence’ ‘the name of fastq file’ | grep –E ‘the sequence of the 2\textsuperscript{nd} nested primer’ | wc -l). Once the number of the cDNA sequences is reasonable (at least greater than 0), then retrieve the sequences with the following command line (grep –E ‘X part of the G-tailed primer sequence’ ‘the name of fastq file’ | grep –E ‘the sequence of the 2\textsuperscript{nd} nested primer’ > ‘the name of the sorted file with the forward orientation’). These processes need to be repeated with a set of complementary sequences to retrieve the sequence reads with the reverse direction.

2. Count and group the sequences in the sorted file in the forward orientation based on their different combinations of exon joining using a Perl script with the following command line (perl test\_RP\_counter.pl ‘the name of the sorted file with the forward orientation’
‘the sequence of the 2nd nested primer’). The execution of this command line generates the number of the sequences containing each of the 16 different combinations of randomized 2 bp sequences (4x4) right after the sequence of the 2nd nested primer, which is designed to be located at the 3’-end of the known 2nd exon. This process also needs to be repeated with the opposite orientation (perl test_RP_counter.pl ‘the name of the sorted file with the reverse orientation’ ‘the complementary sequence of the 2nd nested primer’). This process will provide the information regarding how many sequences contain different exon joining and how many different groups of exon joining are present in a sequenced library.

3. Group the sequences based on the identified different combinations of exon joining. This can be executed with the following command line (grep –E ‘the sequence of the 2nd nested primer plus two additional bases’ ‘the name of the sorted file with the forward orientation’ > ‘the name of the groups with different exon joining’). This needs to be repeated again with the sequences with the opposite orientation.

Map the representative individual sequences from each group to the genome sequence of a given species using the BLAT search in the UCSC genome browser. This mapping process will first compare the exon structure of each group against the known exon structure of a given gene, which will eventually identify its new alternative exons and promoters.

4.3 Results

The feasibility of this protocol has been tested through identifying alternative promoters for an imprinted gene called PEG3 (Paternally expressed gene 3) (Figure 4.2). This gene is
comprised of 9 exons that are spread throughout 25-kb genomic regions in both human and mouse, and is highly expressed in the brain [5-7].

Figure 4.2. Alternative transcripts and promoters identified for mouse and human PEG3. (A) The genomic structure of PEG3. The black and blue boxes represent the 9 previously known exons and the newly identified alternative exons of PEG3, respectively. The bent black arrow indicates the transcriptional direction of PEG3. The E1-E2 transcript contains the known 1st and 2nd exons of PEG3. The mouse U1-E2 transcript harbors the newly identified alternative 1st exon, but skips the known 1st exon of Peg3. The human U1-E2 transcript starts from the newly identified alternative 1st exon, which also skips the known 1st exon of PEG3. (B) The relative frequency of alternative transcripts for mouse and human PEG3. The table summarizes the relative representation (%) of the transcripts with E1 (the known 1st exons) and U1 (alternative 1st exons) of PEG3. The relative representations (%) are calculated through dividing the number of the reads containing the E1, mU1 (mouse), and hU1 (human) exons by the total number of the reads for each library. The total number of reads for each library represents the sorted sequences that still contain both the G-tail and gene-specific primer portions at their 5’ and 3’-ends. (C) RT-PCR analyses for alternative exons. The E1-E2 primer combination targets the transcript containing the known 1st and 2nd exons of PEG3. The U1-E2 primer combination detects the transcript with the newly found alternative 1st exon connected to the 2nd exon of PEG3. The top and bottom panels represent the RT-PCR products from mouse and human brains, respectively.

Interestingly, the 200-kb upstream region of this gene is well conserved among all the mammals without any additional ORFs (Open Reading Frames), thus this region has been suspected to
harbor other unknown cis-regulatory elements [8]. This new protocol, therefore, has been applied to test whether this region contains any unknown alternative promoters. First, the total RNA from mouse and human brains were individually used for generating 5’RACE cDNA libraries, and later these libraries were sequenced using a NGS platform. Each library derived on average several hundred thousand reads, and these raw reads were further analyzed with the bioinformatics strategy described previously. According to the results, the majority of cDNA sequences were indeed derived from the transcripts with the known exon combination (E1-E2): 93% for mouse and 74% for human PEG3 (Figure 4.2 B). However, this protocol also detected some minor transcripts from both mouse and human, and these transcripts were very low in abundance, based on their representations in the libraries: 0.33% for mouse and 0.13% for human PEG3. Detailed inspection further revealed that these minor transcripts skip the known 1st exons but are connected to previously unknown alternative exons (U1 in Figure 4.2 A). It also revealed that the new U1 exons of PEG3 are not conserved between mouse and human genomic regions. Individual RT-PCR analyses indeed confirmed the presence of these minor transcripts for mouse and human PEG3 (Figure 4.2 C). Until now, there has never been any clue suggesting the presence of alternative promoters for the PEG3 locus, although this locus has been intensively studied for more than two decades. Thus, this new protocol appears to be very robust in identifying low abundant transcripts and corresponding alternative promoters.

### 4.4 Discussion

This newly developed protocol, NGS-based 5’RACE, may be used for other similar studies that intend to identify low abundant transcripts for their gene of interest based on the following reasons. First, this protocol is very sensitive: it can detect very low abundant
transcripts and corresponding alternative promoters. As demonstrated above, this protocol successfully detected a minor transcript that makes up only 0.1% of all transcripts from human PEG3. This level of sensitivity cannot be achieved with the existing 5’RACE protocols. This protocol also provides clear exon structures for a given locus, which is not always feasible with other NGS data, such as RNA-seq, due to their relatively short read lengths. Second, the new protocol can estimate the relative expression levels of the alternative transcripts for a given locus among various tissues or stages. The number of raw sequence reads belonging to each transcript can be easily converted into its relative representation in a given cDNA library, which can further indicate the strength of the corresponding alternative promoter. This type of information could be very important since the in vivo function of a given gene often manifests through its tissue and stage-specific alternative promoters. Third, this protocol is reliable and reproducible. We have tested more than 15 rounds of 5’RACE followed by NGS runs, and each round has been consistently successful for identifying new alternative exons and promoters. Despite these strengths, however, the current protocol also has one caveat. We have been selecting DNA fragments less than 400 bp in length for actual NGS runs due to the read length limitation that can be afforded by current NGS platforms. Thus, any 5’RACE cDNA products greater than 400 bp in length cannot be sequenced through the current protocol. Nevertheless, this NGS-based 5’RACE protocol appears to be a very sensitive and reliable method in identifying low abundant alternative transcripts and promoters.

4.5 References


CHAPTER 5
CONCLUSIONS AND FUTURE DIRECTIONS

Genomic imprinting plays critical functional roles in regulating fetal growth and development along with neurological processes such as maternal behavior [1, 2]. This is seen from a gene cluster found in the Peg3 imprinted domain, exclusively expressed in placental mammals [3-5]. These genes are unevenly distributed in the ends of the 500-kb domain, leaving a 250-kb conserved region that contains multiple ECRs that are most likely involved in controlling the Peg3 domain via cis-regulatory mechanisms [6, 7]. The ICR of this region is located at a DMR that corresponds to a bidirectional promoter of Peg3 and Usp29 [8]. This region also contains several YY1 binding sites that suggest the involvement of trans acting factors for imprinting regulation [9-11]. Therefore, it is safe to state that these cis- and trans-acting factors are important for imprinting and transcription of the Peg3 domain. This dissertation presents data indicating how dosage of Yy1 is involved in controlling the Peg3 imprinted domain in vivo, and the discovery of alternative promoters that give rise to the Peg3 expression originating from its maternal allele in the brain.

Recently, reduced levels of YY1 using in vitro and in vivo RNA interference (RNAi) mediated YY1 knockdown models led to an up-regulation of Peg3 expression and altered DNA methylation levels throughout the Peg3 domain, suggesting that it may be a major trans factor that regulates transcription and imprinting [11-14]. However, the in vivo function of YY1 is not well understood. The next step is to understand how YY1 is involved in controlling the Peg3 domain in a dosage dependent manner. Here, we examined the in vivo effects of Yy1 gene dosage on the Peg3 imprinted domain with various breeding schemes utilizing two sets of mouse mutant alleles. The results indicated that half dosage of Yy1 coincides with the up-regulation of Peg3 and Zim1, suggesting that Yy1 acts as a repressor in the Peg3 imprinted domain. This study also
led to an unexpected observation that the maternal allele of *Peg3* is expressed in the adult mouse brain, despite the fact that *Peg3* is maternally imprinted and silent in other tissues. The expression of *Peg3* was found to be bi-allelic in the specific areas of the brain, including the choroid plexus, the PVN and the SON of the adult mouse hypothalamus. This finding is quite significant, as the hypothalamus region of the brain produces the needed hormones to control caring behavior and development [2, 15]. PEG3 protein immunoreactivity was detected in the hypothalamus, along with its enrichment in neurons involved in synthesizing the hormones oxytocin and vasopressin, which play critical roles in labor, milk production and sexual behavior [16]. Although the exact roles of the maternal allele expression of *Peg3* in these cell types are currently unknown, this new finding suggests that the maternal allele may be functional in specific cell types. Moreover, it provides insight regarding the previously unnoticed bi-allelic expression of *Peg3* in the mouse brain. Overall, these results confirm the repressor role of *Yy1* in the *Peg3* domain and also provide a new insight regarding the bi-allelic expression of *Peg3* in mouse brain. However, the mechanism underlying the maternal-allele expression patterns observed in mouse brain remains to be elucidated.

The observed bi-allelic expression pattern may be a result of either de-repression of the maternal allele of the known promoter or the presence of alternative promoters for the *Peg3* locus. To explore this mechanism, we sought to identify the possibility of alternative promoters responsible for its transcription and gene expression patterns using a method that employs NGS-based 5’RACE. As a consequence, in chapter three we identified three alternative promoters for mouse *Peg3* and one alternative promoter for human *PEG3* upstream of their major promoter. In the mouse, two of these alternative promoters drive maternal-specific expression of *Peg3* specifically in the hypothalamus of the adult brain, while the third promoter drives bi-allelic
expression of Peg3 with a paternal bias only in the neonatal-stage brain. In humans, an alternative transcript is also detected at relatively very low levels in the adult brain and placenta. These results reveal that alternative promoters display allele-, tissue-, and developmental stage-specific Peg3 expression patterns. This is a novel discovery, as most functional studies on Peg3 assume that the maternal allele is completely silent, and that a paternal transmission of a knockout mutant is consequently devoid of all Peg3 expression. Therefore, many of the behavioral and functional studies of Peg3 thus far may need to be carefully re-evaluated. These alternative promoters are localized within the 200-kb upstream region of Peg3, in between ECRs, suggesting that some enhancer-like elements have the potential to act as promoters as well [17]. The identification of alternative promoters in both mouse and human models suggests that these alternative promoters may be functionally selected features for the Peg3 imprinted domain during mammalian evolution. Together, these findings contribute to the overall understanding of the imprinting and transcription of Peg3 that was not reported during the past two decades of research.

We modified the existing 5’RACE approach into an NGS-based new protocol in order to identify alternative promoters for Peg3. In this work, we present a new method for detecting alternative promoters that incorporates 5’RACE starting from the known 2nd exons of genes and NGS-based sequencing of the subsequent cDNA products. This method can successfully detect levels of mRNA that comprise less than 1% of total transcripts from a given gene, which to date could not be achieved with existing 5’RACE techniques. Moreover, the relative expression levels of the alternative transcripts can be estimated as an indicator of the alternative promoter strengths. This protocol also provides clear exon structures for a given locus, which is not always feasible with other NGS methods, such as RNA-seq, due to their relatively short read lengths.
Finally, we provide a new method to identify alternative promoters using an NGS-based 5’RACE techniques, which was instrumental in the maternal allele-specific alternative promoters responsible for the bi-allelic expression of Peg3. Overall, these results indicate that this NGS-based 5’RACE protocol is a sensitive and reliable method for detecting low-abundant transcripts and promoters.

Based on the findings presented here, it would be of interest to further explore the mechanistic and functional significance of these alternative promoters in vivo. For instance, exploring the behavior of these alternative promoters in the absence of its major promoter would provide more information about the transcriptional regulation of the alternative promoters. This could be observed by utilizing various breeding schemes for existing mouse mutant models for Peg3. Evaluating the spatial expression of these alternative promoters in the brain may also provide insight to their functional significance. Sequential deletion of these alternative promoters using transgenic and knockout mouse models would be informative tools for loss-of-function studies for Peg3. We speculate that some of these alternative promoters (such as U1) may be involved in functions such as early developmental regulation and DNA methylation setting of the Peg3 ICR [18]. Even though the exact mechanism for alternative promoter transcription is not well explored, this dissertation provides the knowledge and technological advancements essential for future studies in eukaryotic gene transcriptional regulation, and aids in further understanding the process of genomic imprinting in the Peg3 imprinted domain.

5.1. References


APPENDIX: PERMISSION TO INCLUDE PUBLISHED WORK


License agreement and author copyright

Scientific Reports does not require authors to assign copyright of their published original research papers to the journal. Articles are published under a CC BY license (Creative Commons Attribution 4.0 International License). The CC BY license allows for maximum dissemination and re-use of open access materials and is preferred by many research funding bodies. Under this license users are free to share (copy, distribute and transmit) and remix (adapt) the contribution including for commercial purposes, providing they attribute the contribution in the manner specified by the author or licensor (read full legal code).

Alternative Creative Commons licenses are available on request: authors should contact the editorial office (scientificreports@nature.com) on acceptance of an article to discuss these options.

Visit our open research site for more information about Creative Commons licensing.
Bambarendage Pinithi Upekka Perera is the daughter of Mawanana Hewa Prashanthi Sobani Perera and Bambarendage Arunandika Perera. She was born in Colombo, Sri Lanka. She attended high school at Sirimavo Bandaranaike Vidyalaya located in Colombo, Sri Lanka. She attended Delgado Community College in the fall of 2007 and graduated with an Associate degree in general sciences in the spring of 2009 with chancellor honors. She attended the University of New Orleans in the fall of 2009 and graduated with a Bachelor of Science in biological sciences in the fall of 2011. She began her doctoral research afterward under the supervision of Dr. Joomyeong Kim. She will graduate with a Doctor of Philosophy in biological sciences in the summer of 2016. She will begin her post-doctoral training at the University of Michigan School of Public Health in the fall of 2016.