GENOMIC IMPRINTING AND X-CHROMOSOME INACTIVATION IN THE GRAY, SHORT-TAILED OPOSSUM, MONODELPHIS DOMESTICA

A Dissertation

by

KORY CHARLES DOUGLAS

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Paul B Samollow
Committee Members,	Scott Dindot
	Charles Long
	Keith Maggert
Head of Department,	Craig Coates

December 2013

Major Subject: Genetics

ABSTRACT

Imprinted genes have been extensively documented in eutherian mammals and exhibit significant interspecific variation, both in the suites of genes that are imprinted and in their regulation between tissues and developmental stages. Much less is known about imprinted loci in metatherian (marsupial) mammals, wherein studies have been limited to a small number of genes imprinted in eutherians. In this dissertation, I used ChIP-seq and RNA-seq approaches to conduct the first *ab initio* search for imprinted autosomal genes in fibroblasts, fetal brain, and placenta of a metatherian mammal, the gray short-tailed opossum, *Monodelphis domestica,* and the first chromosome-wide study of paternally imprinted metatherian X chromosome inactivation.

Evidence from a few genes in diverse species suggests that metatherian X-chromosome inactivation is characterized by exclusive, but incomplete (leaky), repression of genes on the paternally derived X chromosome. Herein I show that the majority of opossum X-linked genes exhibit paternally imprinted expression with 100% maternal-allele expression, whereas ~14% of genes escape inactivation, exhibiting varying levels of biallelic expression. In addition, I have shown that transcriptionally opposing histone modifications correlate strongly with opossum XCI. However, the opossum did not show an association between X-linked gene expression and promoter DNA methylation.

In generating the first genome-wide profile of histone modification states for a metatherian mammal, and coupling it with in-depth gene expression analyses, I

identified the first set of genes imprinted in a metatherian that are not imprinted in eutherian mammals and described transcriptionally opposing histone modifications and differential DNA methylation at the promoters of a subset of these genes. My findings suggest that metatherians use multiple epigenetic mechanisms to mark imprinted genes and support the concept that lineage-specific selective forces can produce sets of imprinted genes that differ between metatherian and eutherian lines. Overall, these studies furnish a comprehensive catalog of parent-of-origin expression status for both autosomal and X-linked genes in a metatherian, *Monodelphis domestica*, and open new avenues for illuminating the mechanisms and evolution of imprinted gene regulation in mammals generally.

ACKNOWLEDGEMENTS

I would like to thank my committee chair, Dr. Paul Samollow, and my committee members, Dr. Scott Dindot, Dr. Charles Long, and Dr. Keith Maggert for their guidance and support throughout the course of this research. I would also like to thank my many collaborators for their contributions to my work especially Madhuri Jasti as well as Dr. Andrew Clark and Dr. Xu Wang at Cornell University. Without their contributions, this work would not have been possible.

I cannot thank my many colleagues at Texas A&M University enough for their willingness to celebrate the successes and help me to survive the struggles. Our many conversations and discussions have shaped me both personally and professionally, and I will always cherish our time together.

Finally, I would like to thank my wife, Lauren, and son, William, for their patience, steadfastness, understanding, and love. I would also like to thank my family and friends for their unending support.

TABLE OF CONTENTS

Page

Page

LIST OF FIGURES

Page

Page

LIST OF TABLES

CHAPTER I

INTRODUCTION

1.1 *Monodelphis domestica* as a Model Organism

Modern mammals comprise three major clades. The subclass Prototheria are egg-laying mammals, of which the platypus and a few species (the number is debated) of echidnas are the only living representatives. The Methatheria ("marsupials") and Eutheria (so-called "placental" mammals) together form the subclass Theria, which are live-bearing mammals. The term placental mammal is somewhat misleading because all therian mammals form a placenta during fetal development, although the degree of placental growth and elaboration is greater in eutherian species.

The gray, short-tailed opossum, *Monodelphis domestica,* is a small, rapidly breeding metatherian species that has been developed as a laboratory animal for more than 30 years. Live specimens have been collected from five geographically distant areas in Brazil and Bolivia and been used to develop five laboratory strains, each of which have been maintained with detailed pedigree information. The history of the laboratory opossum, the maintenance and locations of breeding colonies, and the overall structure of the genetic diversity present in laboratory stocks have been well documented (VandeBerg & Williams-Blangero 2010). *M. domestica* is widely utilized as a model organism for a variety of research fields including but not limited to genetics, neurobiology, comparative immunogenetics, evolutionary biology, physiology, reproductive endocrinology, developmental biology, and environmental carcinogenesis

(Samollow 2006; Kammerer *et al.* 2010; Moustakas *et al.* 2011; Grant *et al.* 2012; Noor *et al.* 2013). The initial draft sequence and assembly of the *M. domestica* genome was released in October 2004, with the most current assembly (Mondom5) constructed and released in October of 2006. Through comparative genomic analysis, Mondom5 has produced insights into therian genome evolution and function, X-chromosome inactivation, and the evolution of non-coding sequences (Mikkelsen *et al.* 2007b). General information about the genome assembly and annotation is presented in Table 1.

Table 1. *M. domestica* genome (Mondom5) characteristics (derived from Mikkelsen *et al.* 2007b).

Chromosome	Sequence Size (Mb)	Estimated Total Size (Mb)	% G+C	Annotated CpG Islands	Annotated Ensembl Genes
1	745	748	37.8	4,067	4,951
2	538	542	38.0	3,779	4,125
3	524	528	37.3	2,505	3,014
4	432	435	37.7	2,356	3,221
5	301	305	37.2	1,285	1,817
6	289	292	38.1	2,117	1,853
7	257	261	36.7	1,097	1,168
8	309	313	37.8	2,161	1,946
X	76	79	40.9	478	561
Total	3475	3503	---	19,845	22,656

The metatherian and eutherian lineages diverged from their most recent common ancestor ~173 – 193 million years ago (Kumar & Hedges 1998; van Rheede *et al.* 2006; Meredith *et al.* 2011). Due to this sister-group relationship metatherians and eutherians share basic genetic structures and processes that reflect elemental or ancestral mammalian characteristics; however, each group has evolved its own distinctive variants

of these characteristics creating opportunities for comparative examination of ancestrally shared but divergent genomic characteristics (Samollow 2008). In this light, the opossum genome has proven extremely useful for investigating evolutionary processes that have shaped the structure and function of mammalian genomes generally, such as those that drive the maintenance and diversification of conserved non-coding elements; relationships between recombination rate, nucleotide composition, and the distribution of repetitive element families; and the birth, proliferation, and death of transposable element families. The same comparative paradigm should prove fruitful for investigating the origins, structures, and evolutionary development of imprinted genes as well.

1.2 Genomic Imprinting

Genomic imprinting, generally, is a phenomenon whereby certain genes, chromosomal regions, or whole-chromosomes show parent-of-origin-specific differential allelic expression rendering the organism functionally hemizygous at the imprinted locus or region. In addition, maternal- or paternal-allele exclusive expression varies between imprinted loci, as well as between developmental stages and tissues in many cases (Dindot *et al.* 2008) and can exhibit "leaky" expression of the silenced allele (discussed below). Imprinted loci have been found in eutherian and metatherian mammals, but not in prototherians, birds, or other vertebrates (Ferguson-Smith 2011).

The failure of proper imprinting can result in mis-segregation of chromosomes during meiosis, abnormal gene expression, developmental abnormalities, and disease states. Nine human imprinting syndromes have been identified: Beckwith-Weidemann,

Angelman, Prader-Willi, Russel-Silver, maternal and paternal derived uniparental disomy of chromosome 14, pseudohypoparathyroidism type 1b, transient neonatal disease, and maternal hypomethylation syndromes (Lawson *et al.* 2013). Considerably increased frequencies of these and other imprinting-related diseases have been observed in offspring resulting from cloning of cattle (Surani *et al.* 1984), swine (Shen *et al.* 2012), and horses (Hinrichs *et al.* 2006), as well as in children conceived by means of assisted reproduction techniques (Thompson & Williams 2005). These increases have been linked to improper regulation and/or resetting of epigenetic modifications during embryogenesis (Amor & Halliday 2008). Furthermore, the failure of imprinting that results in Beckwith-Wiedmann syndrome correlates positively with the formation of Wilm's tumor, a cancer affecting the kidneys (Rivera & Haber 2005).

Normally, the expression of imprinted genes is maintained throughout the life of an individual and imprints are erased and reset according to the sex of the respective parent during gametogenesis, allowing transmission of the sex-specific imprinting pattern to the next generation. How these genes are targeted for imprinting and how the imprint is maintained through early zygotic and embryonic development is unknown. The lack of knowledge concerning the spectrum of imprinted genes in mammals and the molecular mechanisms that target genes for proper establishment and maintenance of imprinting represents a wide gap in our understanding of the scope of imprinting and how imprinting irregularities lead to physiologic dysfunction.

1.3 Occurrence of Imprinted Genes in Metatherian and Eutherian Mammals

Estimates of the number of imprinted genes vary between humans and mice. Simulation studies, based on the molecular and genomic characteristics of known imprinted genes, have predicted ~1% of human and ~2.5% of murine genes are either maternally or paternally imprinted (Luedi *et al.* 2005; Luedi *et al.* 2007). Notwithstanding this high estimate, only 79 and 123 imprinted genes have been characterized in human and mouse, respectively, and of those, only 72 have corresponding homologs in both species, with 44 (~61%) sharing imprinted status (Morison *et al.* 2005). In addition, Hamed *et al.* (2012) examined 25 imprinted genes present in mouse and human for which strong experimental data exists and found that the vast majority (23 of 25) show the same expression pattern in both species whether maternally or paternally imprinted. The strong conservation of the expression states of these shared imprinted genes indicates that their allele-specific, imprinted expression is under selection and is biologically important for proper development in both species. Taken together, the diversity of the suites and expression states of imprinted genes not only illustrates the difficulty in finding and describing imprinted genes but also reveals the magnitude of variation present among the suites of imprinted genes found in different species.

In metatherians, only 19 of the genes that are known to be imprinted in mouse and/or human have been studied with regard to parent-of-origin differential expression (Table 2) (Renfree *et al.* 2008). Of these 19, only 11 have been examined in *M. domestica*, with five showing an imprinted state. Importantly, it has been shown that not

all imprinted genes in metatherians exhibit complete imprinting. "Leaky" expression from the imprinted (repressed) allele has been observed for *Igf2*, *Ins*, *Peg1/Mest*, and *Peg10* (reviewed in Renfree *et al.* 2008). Notably, leaky expression has also been

Table 2. Imprinted genes of three metatherian species: tammar wallaby (*Macropus eugenii*), gray, short-tailed opossum (*M. domestica*), and Virginia opossum (*Didelphis virginiana*).

Gene	Species	Imprinted Allele	Evidence of Leaky Expression	DMR	Reference
PEG1/MEST	*M. eugenii*	Maternal	Yes	No	(Suzuki *et al.* 2005)
PEG10	*M. eugenii*	Maternal	Yes	Yes	(Suzuki *et al.* 2007)
H19	*M. eugenii*	Paternal	No	Yes	(Smits *et al.* 2008)
IGF2	*M. domestica* *M. eugenii*	Maternal	Yes No	Yes Yes	(Suzuki *et al.* 2005; Ager *et al.* 2008b; Lawton *et al.* 2008)
IGF2R	*D. virginiana* *M. domestica* *M. eugenii*	Paternal	No	No	(Killian *et al.* 2000)
INS	*M. eugenii*	Maternal	Yes	No	(Ager *et al.* 2007)
HTR2A	*M. domestica*	Paternal	No	No	(Das *et al.* 2012)
L3MBTL	*M. domestica*	Maternal	No	No	(Das *et al.* 2012)
MEG3	*M. domestica*	No metatherian homolog	-	-	(Weidman *et al.* 2006a; Weidman *et al.* 2006b)
NNAT	*M. domestica*	No metatherian homolog	-	-	(Evans *et al.* 2005)
SNRPN	*M. eugenii*	No	-	-	(Rapkins *et al.* 2006)
UBE3A	*M. eugenii*	No	-	-	(Rapkins *et al.* 2006)
DIO3	M. eugenii	No	-	-	(Edwards *et al.* 2008)
CDKN1C	*M. eugenii*	No	-	-	(Suzuki *et al.* 2005; Ager *et al.* 2008a)
DLK1	*M. domestica* *M. eugenii*	No	-	-	(Weidman *et al.* 2006a; Weidman *et al.* 2006b; Edwards *et al.* 2008)
PLAGL1	*M. domestica*	No	-	-	(Das *et al.* 2012)
IMPACT	*M. domestica*	No	-	-	(Das *et al.* 2012)
COPG2	*M. domestica*	No	-	-	(Das *et al.* 2012)
GRB10	*M. eugenii*	No	-	-	(Stringer *et al.* 2012)

observed from genes on the inactive paternally inherited X chromosome in metatherian females, wherein X-chromosome inactivation occurs as a paternally imprinted phenomenon (Cooper *et al.* 1993; Samollow *et al.* 1995; Hornecker *et al.* 2007) (see below and Chapter 2 for further discussion of X-chromosome inactivation).Additionally, imprinted genes in humans and mice are often found in clusters and are regulated by localized, differentially methylated regions (DMRs) (Hore *et al.* 2007b). Genes in these clusters may be under the control of a single regulatory sequence (imprinting control region: ICR) or be independently regulated (Lopes *et al.* 2003). To date, only one imprinted gene cluster has been found in metatherians, *IgfR*/*H19*, and only three, *Igf2/H19* and *Peg10*, of the eight known imprinted genes in metatherians show evidence of a DMR (Suzuki *et al.* 2007; Lawton *et al.* 2008; Smits *et al.* 2008).

These characteristics not only illustrate the difficulty of finding and describing imprinted genes in different species, but also reveal the magnitude of variation present among the suites of imprinted genes and the mechanisms employed to establish and maintain the imprinted state. *M. domestica* provides us with an excellent opportunity for comparative analysis to expand our knowledge of the variety of imprinted genes and breadth of imprinting mechanisms found in mammals and, thereby, to gather new information with which to fill these gaps in our knowledge of the evolutionary origins, adaptive benefits, and biologic importance of the imprinting phenomenon.

1.4 Genomic Imprinting in Other Species

In addition to its presence in mammals, the phenomenon of genomic imprinting has also been described in insects (Khosla *et al.* 2006; Anaka *et al.* 2009) and to some

extent in other vertebrates and invertebrates (Martin & McGowan 1995; Bean *et al.* 2004) and plants (Tourte *et al.* 1980; Alleman & Doctor 2000). Findings from these and other studies have indicated that disparate organisms utilize similar epigenetic processes to label and/or control the expression of imprinted genes, namely DNA methylation, histone modifications, and non-coding RNAs (further discussed below) (Lippman & Martienssen 2004; Grewal & Elgin 2007), provides evidence that the molecular mechanisms underlying genomic imprinting are derived from ancient regulatory systems that share common ancestry. In further support of this hypothesis, it has been shown that ICRs present in both mouse and human can also silence adjacent genes in transgenic *Drosophila,* although not in a parent-of-origin-specific manner (Lyko *et al.* 1997; Lyko *et al.* 1998). These findings indicate that the mechanisms utilized to control the expression of imprinted genes are present in disparate species, highlight the need to investigate the regulation of imprinted genes in many species which will result in a better understanding of imprinted and monoallelic expression in all organism, and support the concept of genomic imprinting as a model for the study of the mechanisms of gene regulation in general.

1.5 DNA Methylation and Imprinting Control

DNA methylation, usually in the form of methylated cytosine bases, is characteristic of most imprinted genes in eutherians and believed to be integral to the proper function and maintenance of the imprinted state. Methylated cytosines are commonly found at CpG dinucleotides, which themselves are often located within CpG islands, areas of the genome that are strongly enriched for CpG dinucleotides. CpG

islands often occur within or are proximate to the promoters of genes and ICRs (Fatemi *et al.* 2005; Saxonov *et al.* 2006; Feil & Berger 2007). The establishment and maintenance of cytosine methylation is controlled by DNA methyltransferases (DNMTs).

DNMTs comprise a class of enzyme that catalyzes the transfer of a methyl group to cytosine residues, especially at CpG dinucleotides, of DNA. The three major subclasses of DNMTs are DNMT1, DNMT2, and DNMT3. DNMT1 is a maintenance methylase, as it has a strong preference for hemi-methylated DNA, at which it methylates the unmethlylated cytosine residue of the DNA double helix following replication of the methylated parent DNA strand (Pradhan *et al.* 1999). DNMT2 shows only weak methyltransferase activity, and its absence is not known to be associated with methylation or developmental defects in mammals (reviewed in Bestor 2000; Xu *et al.* 2010). DNMT3 has two further subclasses, A and B, which are recognized as *de novo* methylases. These are the responsible for establishing DNA methylation at previously unmethylated cytosine residues (Okano *et al.* 1999). Loss-of-function mutations in these methylases or chemical mutagenesis that causes changes in DNA methylation patterns adversely affect imprinted genes, leading to bi-allelic expression or absence of expression of either allele (Li *et al.* 1992).

One of the best examples of the complexity of genomic imprinting, as well as one of the best-studied imprinting clusters, is the *IGF2/H19* imprinted cluster. *IGF2* codes for Insulin-like Growth Factor 2 and is vital for proper fetal growth and development, and *H19* is a non-coding RNA with unknown function. They are

reciprocally imprinted with *IGF2* expressed solely from the paternal strand and *H19* expressed solely from the maternal strand (Bartolomei *et al.* 1991; DeChiara *et al.* 1991). The imprinted cluster is located on chromosome 11 in humans and consists of *IGF2*, *H19*, a CTCF binding site at a DMR located between the two genes, and a downstream enhancer element.

The proposed model of the transcriptional regulation of this imprinted cluster involves the parent-of-origin specific methylation of the DMR affecting the binding of patterns of CTCF which in turn affects the interaction of the downstream enhancer with the promoters of *IGF2* and *H19* (Bell & Felsenfeld 2000; Hark *et al.* 2000; reviewed in Sha 2008). More specifically, methylation of the DMR at the imprinted cluster on the paternal chromosome prevents binding of the CTCF protein. However, CTCF binds the unmethylated DMR on the maternal chromosome, acting as an insulator that prevents the downstream enhancer from interacting with the promoter of *IGF2* on the maternal chromosome. This causes the enhancer to interact with the promoter of *H19* instead, directing its transcription in a maternally specific manner. On the paternal chromosome, the absence of CTCF at the DMR allows the enhancer to interact with the promoter of *IGF2*, rather than the promoter of *H19*, resulting in the transcription of *IGF2* in a paternally specific manner. It has been shown that the silencing of paternal *H19* is dependent on the upstream DMR on the paternal chromosome, indicating that the absence of CTCF binding alone is insufficient to silence the paternal *H19* gene (Srivastava *et al.* 2000). Interestingly, there is another DMR (DMR1) located at the

promoter of the *IGF2* gene; however, its role in the transcription or imprinting status of *IGF2* has yet to be determined.

Although there is much evidence to support the model described above, more recent work has shown that it is likely an oversimplification of the complex transcriptional regulation of the *IGF2*/*H19* imprinted cluster. Arney (2003) highlights the variation in expression of *IGF2*/*H19* in different tissues as well as the complex array of *cis*-acting elements including insulators, silencers, and enhancers that are often utilized in a dermal-layer-specific, tissue-specific, or developmental-stage-specific manner indicating a more complex mechanism for the control of transcriptional regulation of this imprinted cluster. In addition, Zampieri *et al.* (2012) have shown that post-translational modifications of chromatin-associated proteins (i.e. *PARP1*) can form complexes with *CTCF* and/or *DNMT1* affecting their function and thus the DNA methylation patterns at their binding and/or target sites. Taken together, these studies demonstrate that there is still much to learn about even the most well studied imprinted genes/clusters and the complex mechanisms that are utilized to control their expression states.

1.6 Histone Modifications and Imprinting Control

Along with DNA methylation, histone modifications are correlated with promoters of imprinted genes and ICRs (Fournier *et al.* 2002; Barlow 2011). DNA is packaged into nucleosomes consisting of ~147 base pairs (bp) of DNA wrapped around a histone octamer that comprises 2 copies each of histone proteins 2A, 2B, 3, and 4. The N-terminal domains of each histone, especially H3 and H4, are potential sites for post-

translational modifications including methylation, acetylation, ubiquitination, phosphorylation, and sumoylation (reviewed in Kouzarides 2007). More than 40 histone modifications and their correlations to transcriptional states have been examined in detail (Ernst & Kellis 2010). The changes in chromatin structure resulting from these histone modifications, individually or in combination, are a major source of interest, because they affect access of components of the translational machinery to promoter regions, thereby enhancing or inhibiting transcription rates (Strahl & Allis 2000). Modified histones show varying levels of positive and negative correlation with several kinds of genomic elements and structures, especially promoters, and both coding and non-coding sequences of the gene bodies themselves (Ernst & Kellis 2010; Kharchenko *et al.* 2011; Encode Project Consortium *et al.* 2012; Gifford *et al.* 2013).

At imprinted loci, tri-methylation of the fourth lysine (K) residue on histone 3 (H3K4me3) is associated with promoter regions on chromosomes containing the actively transcribed allele, independent of maternal or paternal origin. Similarly, H3K9me3 is correlated with the transcriptionally silent allele. In addition, these two modified states have been shown to be mutually exclusive at identical sites at the promoters of active versus inactive alleles (Mikkelsen *et al.* 2007a). Another example of imprinting, X-chromosome inactivation in metatherians and trophoblast cells of eutherians, is also associated with histone modifications, especially histone acetylation and H3K27me3 suggesting that these marks might also be important for establishing imprinting expression patterns at metatherian autosomal loci (Wakefield *et al.* 1997; Monk *et al.* 2006; Bernstein *et al.* 2007).

1.7 X-Chromosome Inactivation

X-chromosome inactivation (XCI) is a process in therian mammals that results in the inactivation of one of the two X chromosomes in each cell of the female embryo early in development. In eutherians this event is random with regard to parental origin in the embryo proper, so that about half the cells of the adult female possess an active maternal X chromosome (Xm) only, while the other half have only an active paternal X chromosome (Xp). However, in extra-embryonic tissues, i.e. trophoblast derivatives of mice and cattle, evidence suggests that the Xp is exclusively inactivated (Xue *et al.* 2002; Okamoto & Heard 2006). Unlike the situation in eutherians, XCI is decidedly non-random in metatherians, resulting in all cells of the adult female possessing an active Xm and inactive Xp. The inactive Xp chromosome is enriched for hypoacetylated H4 and H3K27me3 relative to those of the active Xm, which is enriched for activating marks such as H3K4me3 and H3 acetylation (although this latter modification remains poorly characterized) (Keohane *et al.* 1998). These data indicate that histone modifications are correlated with XCI in metatherians, and strongly suggest that histone modifications can be utilized to identify candidate-imprinted regions in the opossum model (Delaval *et al.* 2007; Feil & Berger 2007; Mikkelsen *et al.* 2007a).

1.8 Specific Aims and Structure

The aim of my doctoral research was to search for imprinted loci in *M. domestica* and use the findings to make observations concerning genomic imprinting as a paradigm for gene expression in therians. In order to conduct this search, I utilized certain epigenetic marks (i.e. histone modifications and DNA methylation) that have been

shown to be associated with imprinted genes in eutherians, and genetic crosses designed to enrich for stock-specific single nucleotide polymorphisms (SNPs) that allowed me to track parent-of-origin-specific allelic expression at the mRNA level. I chose to utilize next-generation sequencing technologies to conduct genome-wide analyzes of chromatin states and gene expression. By means of these approaches, both individually and in combination, I have provided the first genome-wide analysis of epigenetic states in any marsupial species to date and described the correlation of these epigenetic states with both X-linked and autosomal gene expression. I also discovered the first marsupial-specific imprinted genes; conducted the first in-depth, chromosome-wide analysis of gene expression of the opossum X chromosome; and correlated monoallelic and parent-of-origin-specific gene expression with activating and repressive histone modifications and DNA methylation states. These data not only add to the suite of imprinted genes in marsupials, but also gives insights into how genes are targeted for imprinting, and how the imprinted state is maintained. Taken together, these observations lay the foundations on which we can begin to address broader questions of why the phenomenon of genomic imprinting exists and the evolutionary implications of such a phenomenon.

To give the remaining chapters context, I would like to briefly explain the overall structure of the dissertation. Chapters II and III are modified versions of submitted manuscripts on the subjects of X-chromosome inactivation and genomic imprinting in fibroblast cell lines, respectively, and are organized in manuscript form. Chapter IV covers the topic of genomic imprinting in tissues, for which the vast majority of data are collected with several verification experiments pending and a manuscript is in

preparation. Chapter V contains extended methods for the preceding chapters and Chapter VI gives concluding remarks and discussion.

CHAPTER II

PATERNALLY IMPRINTED X-CHROMOSOME INACTIVATION (XCI) AND ESCAPERS OF XCI

2.1 Introduction

X-chromosome inactivation (XCI) is a chromosome-wide phenomenon whereby most genes on one of the two of X chromosomes of Therian females are rendered transcriptionally silent early in development (Straub & Becker 2007; Payer & Lee 2008). There are two basic forms of XCI: random XCI (rXCI) and paternally imprinted XCI (pXCI). In rXCI, the choice of X chromosome to be inactivated in any given cell is more-or-less random with regard to the parental source (Heard *et al.* 1997). Once achieved, the inactive state is developmentally stable and clonally inherited throughout subsequent somatic cell divisions, so that the somatic cells of the adult female are approximately equally distributed between those bearing an inactive paternally-derived X (Xp) and those bearing an inactive maternally-derived X (Xm). In mouse, rXCI occurs in epiblast cells, which develop from the inner cell mass of the early embryo (Latham 2005; Okamoto & Heard 2006). However, cells of the mouse trophectoderm layer, which ultimately give rise to extra-embryonic structures including the placenta, display pXCI, in which the Xp is inactivated in all cells, while the Xm always remains active (Huynh & Lee 2001, 2005; Heard & Disteche 2006). pXCI has also been observed in placental tissues of other eutherian mammals, including rat (Wake *et al.* 1976) and cow (Xue *et al.* 2002; Dindot *et al.* 2004). However, this pattern is not

ubiquitous in eutherians, as rXCI has been observed in horse and mule placental tissues (Wang *et al.* 2012), and, notwithstanding a series of conflicting past reports, appears to be the characteristic pattern in human placenta (Moreira de Mello *et al.* 2010). The varying patterns of XCI in trophectoderm-derived tissues of eutherians may have adaptive consequences, although the relatively sparse data show no clear phylogenetic clustering. Thus, while XCI may have important consequences in the development and maintenance of fetal support structures, the molecular mechanisms underlying its inter-specific divergence appear to be evolutionarily labile.

In contrast to the eutherian pattern, data from several metatherian (marsupial) species show that females from this branch of mammals exhibit pXCI in late fetal and adult stage somatic cells (Payer & Lee 2008; Deakin *et al.* 2009). However, there are varying reports that some genes on the inactive Xp exhibit 'leaky' or 'partial' expression (incomplete repression). The particular genes that exhibit leaky expression can differ across species, while within a species the level of paternal-allele expression varies across tissue types, developmental stages, and *in vivo* vs. cultured cells (VandeBerg *et al.* 1987; Cooper *et al.* 1990; Cooper *et al.* 1993; Samollow *et al.* 1995; Koina *et al.* 2005; Hornecker *et al.* 2007). Unfortunately, only five marsupial X-linked genes have been examined with regard to parent-of-origin specific allelic expression, and the limited data derived from different subsets of these five loci in disparate species and interspecific hybrids have not enabled many locus-by-locus comparisons among species. More importantly, the data do not allow an extrapolation of the expression patterns of these few genes to the full X chromosome for any individual species. Thus, it remains unclear

whether pXCI in marsupials is a concerted, chromosome-wide phenomenon, or a piecemeal process that occurs on a region-by-region basis (Cooper *et al.* 1990; Riggs 1990).

Most, but not all genes on the inactive eutherian X chromosome are strongly transcriptionally repressed. Nevertheless, in humans ~15% of X-linked genes located outside the pseudo-autosomal region (PAR) are expressed from both alleles (Disteche *et al.* 2002; Carrel & Willard 2005), whereas in mice only ~3% of X-linked genes located outside of the PAR are biallelically expressed (Yang *et al.* 2010). Non-pseudoautosomal genes expressed from both X chromosomes are designated as rXCI **escapers,** as opposed to monoallelically expressed **non-escapers**, which are subject to full repression through the rXCI mechanism. Based on the a small number of loci examined in marsupials (obtained by non-quantitative isozyme polymorphism, RT-PCR, and RNA-FISH approaches), it appears that genes that escape XCI are common in marsupials, and that homologous genes can show substantially different paternal-allele expression patterns in different species. However, the spotty overlap of gene sets examined in different marsupial species precludes estimation of the proportion of all Xp genes that escape inactivation. Inasmuch as Xp expression for these genes was assayed in various tissues and developmental stages, these differences could reflect both tissue- and species-specific variability.

Mirroring our limited knowledge regarding patterns of Xp-allelic expression in marsupials, information concerning molecular mechanisms of pXCI is also rudimentary. The recently discovered *Rsx* locus, which is expressed exclusively from the inactive Xp

of the gray, short-tailed opossum, *Monodelphis domestica*, appears to be the functional analog of the eutherian *XIST* gene, insofar that its non-coding transcript is suggested to act *in cis* to repress activity of genes on the Xp (Grant *et al.* 2012). Remarkably, despite their analogous functions, *Rsx* and *XIST* show no obvious sequence homology, indicating that they arose independently in metatherians and eutherians. Much less is known about the regulation of expression at the level of individual X-linked genes in marsupials. To date, single-gene bisulfite sequencing of CpG islands around selected X-linked genes has identified no differentially methylated regions (DMRs) (Kaslow & Migeon 1987; Loebel & Johnston 1996; Hornecker *et al.* 2007), suggesting that DNA methylation plays little or no role in marsupial XCI. Two recent chromosome-level immunofluorescence staining studies have shown distinctive epigenetic profiles for the active X (Xa) and inactive X (Xi) in two distantly related marsupials. In the brushtail possum, *Trichosurus vulpecula* (Australasian family Phalangeridae), differential methylation signals on the Xa and Xi correlated well with the locations of active and repressive histone modification marks (Rens *et al.* 2010). For *M. domestica* (American family Didelphidae), histone modifications on the X chromosomes showed suggestive evidence that the Xi is associated with the repressive histone mark H3K27me3 (Mahadevaiah *et al.* 2009; Chaumeil *et al.* 2011). However, the relatively low resolution of these studies makes it impossible to discriminate the locations of modified histones beyond the chromosome-band level, so does not permit the recognition of correlations among individual X-linked gene-expression levels and histone modification status.

To investigate rigorously the many unknowns in marsupial XCI requires a systematic, quantitative, chromosome-wide survey of allelic expression patterns and epigenetic states in a defined set of tissues and developmental stages from one or more marsupial species. To initiate this effort we performed RNA-seq analysis in offspring from two parental stocks and their reciprocal crosses in fetal brain and extra-embryonic membranes (EEM) of the opossum; assayed levels of DNA methylation at promoters in fetal brain and EEM; and examined H3K4me3 and H3K27me3 histone modifications in female fetal brain. The results of this study comprise the first comprehensive catalog of X-linked gene expression from the active Xm and inactive Xp in a marsupial species and the first chromosome-wide assessment of potential epigenetic mechanisms of pXCI in a marsupial at the individual gene level.

2.2 Results

2.2.1 Transcriptome-Wide Quantification of Parent-of-Origin Allele-Specific Expression in Opossum Fetal Brain and EEM

To determine the relative expression levels of genes from the Xm and Xp of female opossums, we performed Illumina RNA-seq on embryonic day 13 (between day 12 – 13 post-coitus) fetal brain and paired extra-embryonic membranes (EEM, counterpart of eutherian placenta) (Figure 1A,B) from reciprocal F_1 hybrid and parental crosses of LL1 and LL2 stocks (Figure A1 and Table B1). Fetuses are born on embryonic day 14 and approximate 11-12 day mouse embryos in overall development (Smith 2001). For a detailed description of opossum fetal development see Mate *et al.* (1994). In total, we obtained 76.5 billion bp of sequence, >80% of which were uniquely mapped to the

opossum reference genome (MonDom5) (see the table on p. 98 in Chapter V). The mapped reads covered more than 10,000 expressed genes with FPKM ≥ 1 in both tissues (Fragments Per Kilobase of transcript per Million mapped reads), and 68,000 exonic SNPs were called from all samples combined (Figure A2). Parent-of-origin allele expression ratios were quantified in each sample from the relative numbers of reads containing the reference and alternative alleles at high-quality SNP positions.

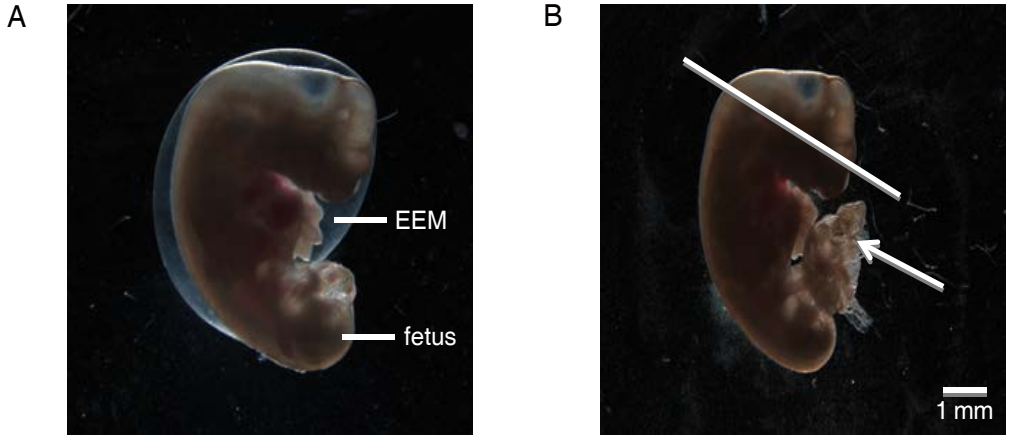


Figure 1. Locations and allelic expression profiles of genes on the opossum X chromosome in fetal brain and extra-embryonic membranes (EEM). (A) Opossum embryonic day 13 (E13) fetus and EEM. (B) E13 fetus with EEM removed. Solid white line is the approximate incision line used to isolate fetal brain. The arrow shows the tissue used for the EEM analysis. (C) The allelic expression percentages for opossum X-linked genes in E13 female fetal brain (*left panel*) and EEM (*right panel*). On the *x*-axis are allelic expression percentages: 0~100% maternal expression to the left and 0~100% paternal expression to the right. The *y*-axis is the physical location of each gene along opossum X chromosome (centromere on top). The red bar is drawn according to the maternal-allele expression percentage and the blue bar is the paternal expression percentage. Among the 176 and 134 expressed X-linked genes with informative SNPs in fetal brain and EEM, respectively, 152 in fetal brain and 110 in EEM had 100% (or close to 100%) maternal expression, and therefore reflect pXCI (nonescaper) genes The names of 24 genes that escape pXCI with > 10% paternal expression (escaper genes) are labelled in the middle panel. The non-coding *Rsx* transcript shows 100% expression from the paternal allele in both tissues. (Figure A2 and Table B1).

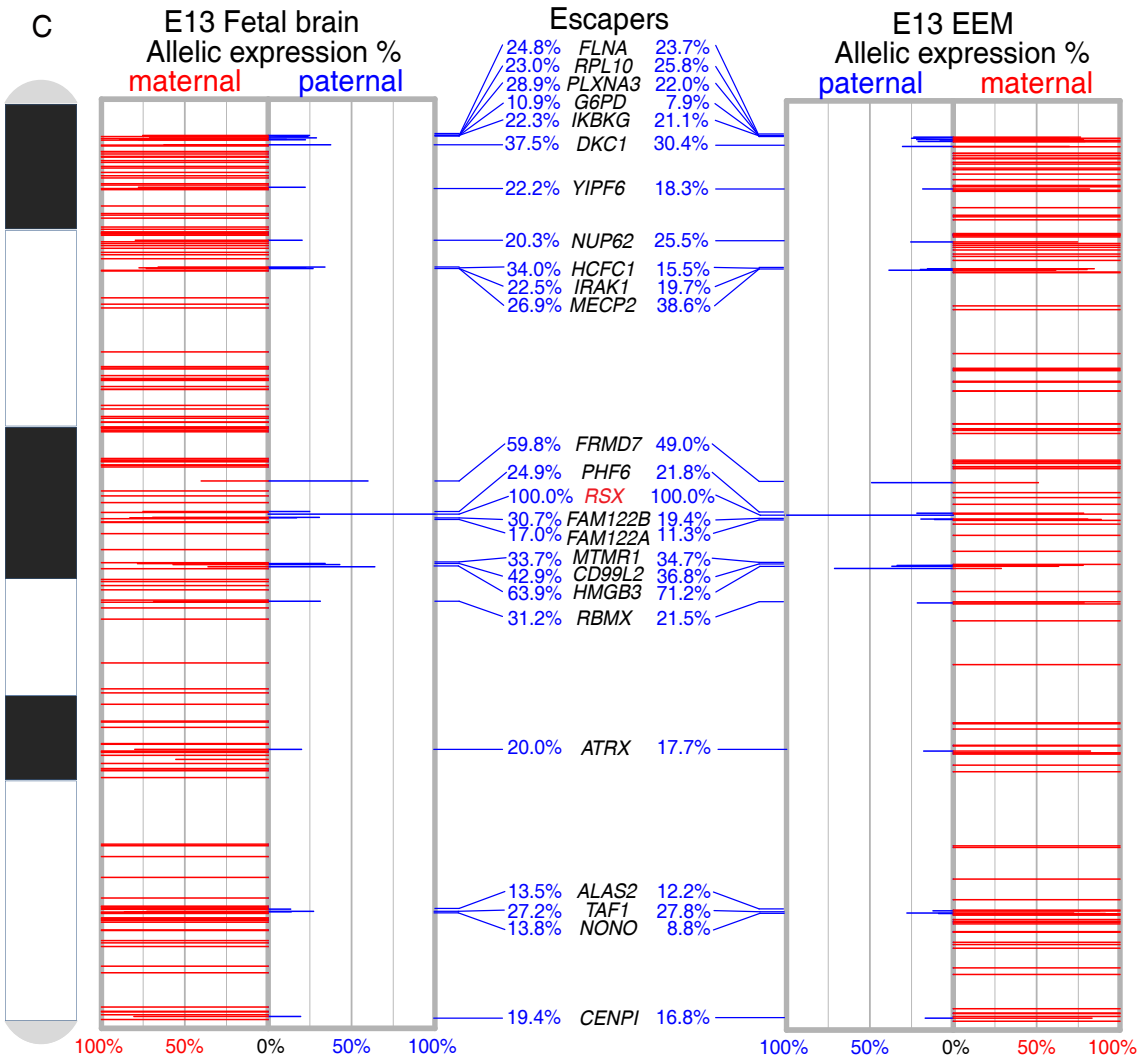
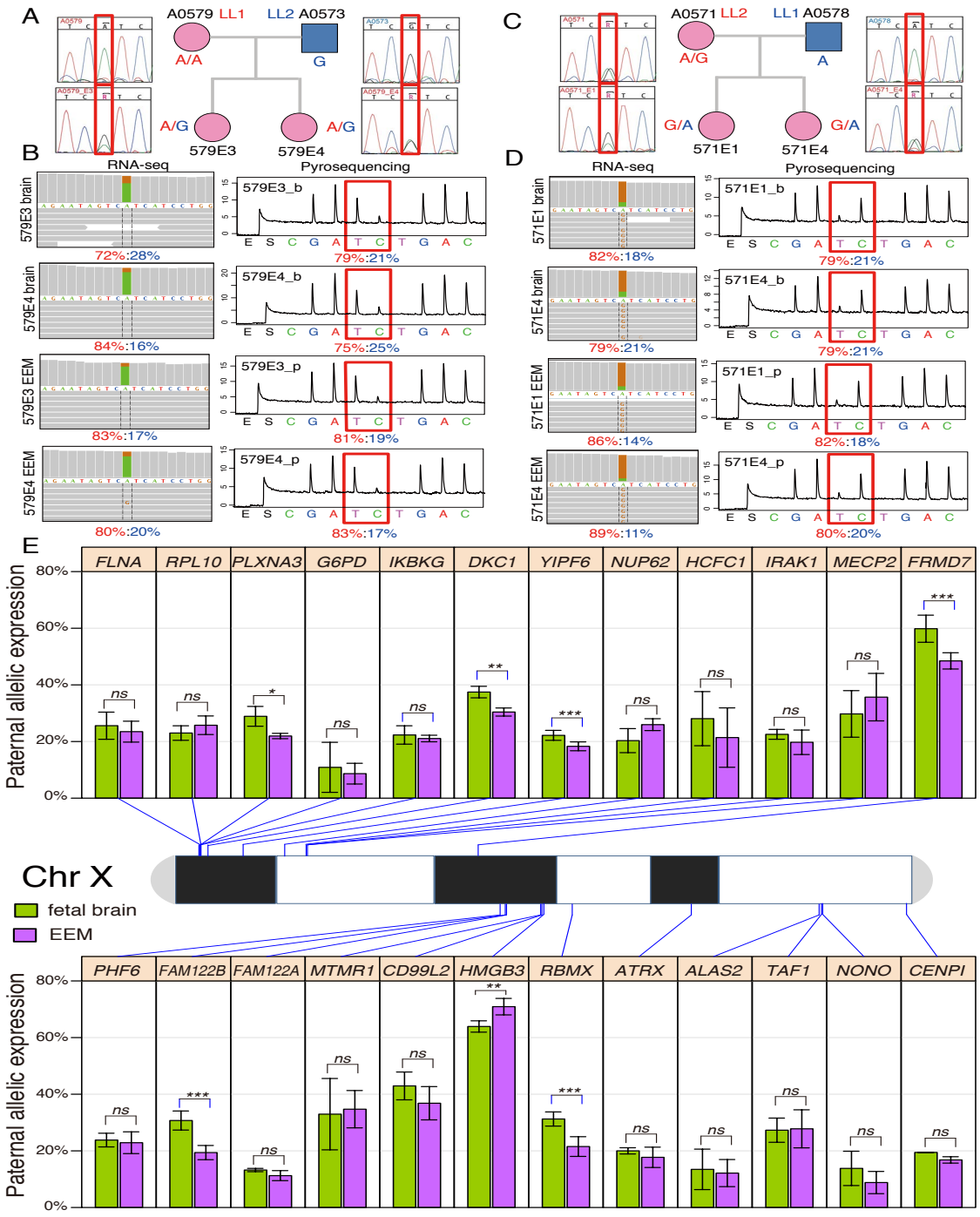


Figure 1 Continued.

2.2.2 Fourteen Percent (24/176) of Opossum X-Linked Genes Escape pXCI

Three-hundred-twelve X-linked gene models (including 19 X-linked genes on chrUn, see tables on pp. 101 and 102) were covered in female fetal brain samples with sufficient expression levels to call *de novo* SNPs (Table B1). Of these, 176 genes had a total of 512 informative heterozygous SNPs that could be used to quantify expression of the Xm and Xp alleles. Among the 176 informative genes, 24 escaped pXCI with more than

Figure 2. Pyrosequencing analysis of maternally vs. paternally derived allele expression ratios for 24 imprinted X-inactivation escaper genes from in opossum female fetal brain and EEM. (A-D) SNP genotyping and pyrosequencing verification for escaper gene *YIPF6* in opossum fetal brain and EEM. Sanger sequencing genotyping confirmed that exonic SNP OMSNP0155110 was informative in all four F_1 embryos: A0579E3 and A0579E4 in LL1 (dam) x LL2 (sire) cross (A); A0571E1 and A0571E4 in the reciprocal F_1 cross LL2 (dam) x LL1 (sire) (B). Biallelic expression was verified by both RNA-seq (*left*) and allele-specific pyrosequencing strategies (*right*) (C, D). Therefore, *YIPF6* is an escaper of pXCI in both tissues. The results for all 24 escaper genes, a non-escaper gene and an autosomal control gene can be found in Figure A2 and Table B2. (E) The opossum X chromosome is drawn in the middle panel with names and physical locations labelled for the 24 escaper-genes identified from RNA-seq data. For each gene, the mean and standard deviation of paternal-allele expression as a percentage of total expression were plotted for fetal brain (green) and EEM (purple). All 24 escaper-genes escape pXCI in both tissues with six genes showing significantly higher paternal leakage in fetal brain as compared to EEM (*PLXNA3*, *DKC1*, *YIPF6*, *FRMD7*, *FAM122B* and *RBMX*). Only one gene showed significantly higher paternal leakage in EEM (*HMGB3*). Statistical significance was assessed by Mann-Whitney U-test (ns: *P*-value > 0.05; *: *P*-value < 0.05; **: *P*-value < 0.01; *: *P*-value < 0.001).**


A
A0579 LL1
LL2 A0573
A/A
G
A/G
A/G
579E3
579E4
C
A0571 LL2
LL1 A0578
A/G
A
G/A
G/A
571E1
571E4
B
RNA-seq
Pyrosequencing
579E3 brain
579E4 brain
579E3 EEM
579E4 EEM
579E3_b
579E4_b
579E3_p
579E4_p
72%:28%
79%:21%
84%:16%
75%:25%
83%:17%
81%:19%
80%:20%
83%:17%
D
RNA-seq
Pyrosequencing
571E1 brain
571E4 brain
571E1 EEM
571E4 EEM
571E1_b
571E4_b
571E1_p
571E4_p
82%:18%
79%:21%
79%:21%
79%:21%
86%:14%
82%:18%
89%:11%
80%:20%
E
Paternal allelic expression
FLNA
RPL10
PLXNA3
G6PD
IKBKG
DKC1
YIPF6
NUP62
HCFC1
IRAK1
MECP2
FRMD7
Chr X
fetal brain
EEM
PHF6
FAM122B
FAM122A
MTMR1
CD99L2
HMGB3
RBMX
ATRX
ALAS2
TAF1
NONO
CENPI

10% paternal-allele expression (**escaper genes**). These genes and the percentage of total transcript that is derived from the Xp allele are listed in Figure 1C.

The remaining 152 genes (**non-escaper genes**) were subject to pXCI with 100% or nearly 100% maternal-allele expression. Only 2% (3/152) of non-escaper genes showed detectable weak paternal expression (>3% paternal expression with >2 paternal allele counts in at least two informative individuals): *MSN* (5.2% paternal expression), *BCAP31* (6.2%) and *PDZD11* (3.3%). In the EEM samples, 134 of 242 covered genes had informative SNPs, and the same 24 genes escaped pXCI with >10% paternal leakage (Figure 1C). Only 3% (4/120) of the non-escaper genes in EEM samples displayed detectable weak paternal expression: *TAZ* (5.0%), *FAM3A* (4.5%), *BCAP31* (5.7%), and *TIMM8A* (5.4%) (Table B1). To confirm the informative X-linked SNPs and XCI status independently, SNPs in escaper genes were validated in multiple samples using Sanger sequencing and allele-specific pyrosequencing (see table on p. 103 and Table B2). This resulted in a 100% validation rate of the RNA-seq expression data (Figure 2A-D; Figure A2). Thus, the majority of X-linked genes show 100% monoallelic maternal expression in both tissues, and the escaper genes are conserved between fetal brain and EEM.

2.2.3 pXCI Profile and Comparison Between Fetal Brain and EEM

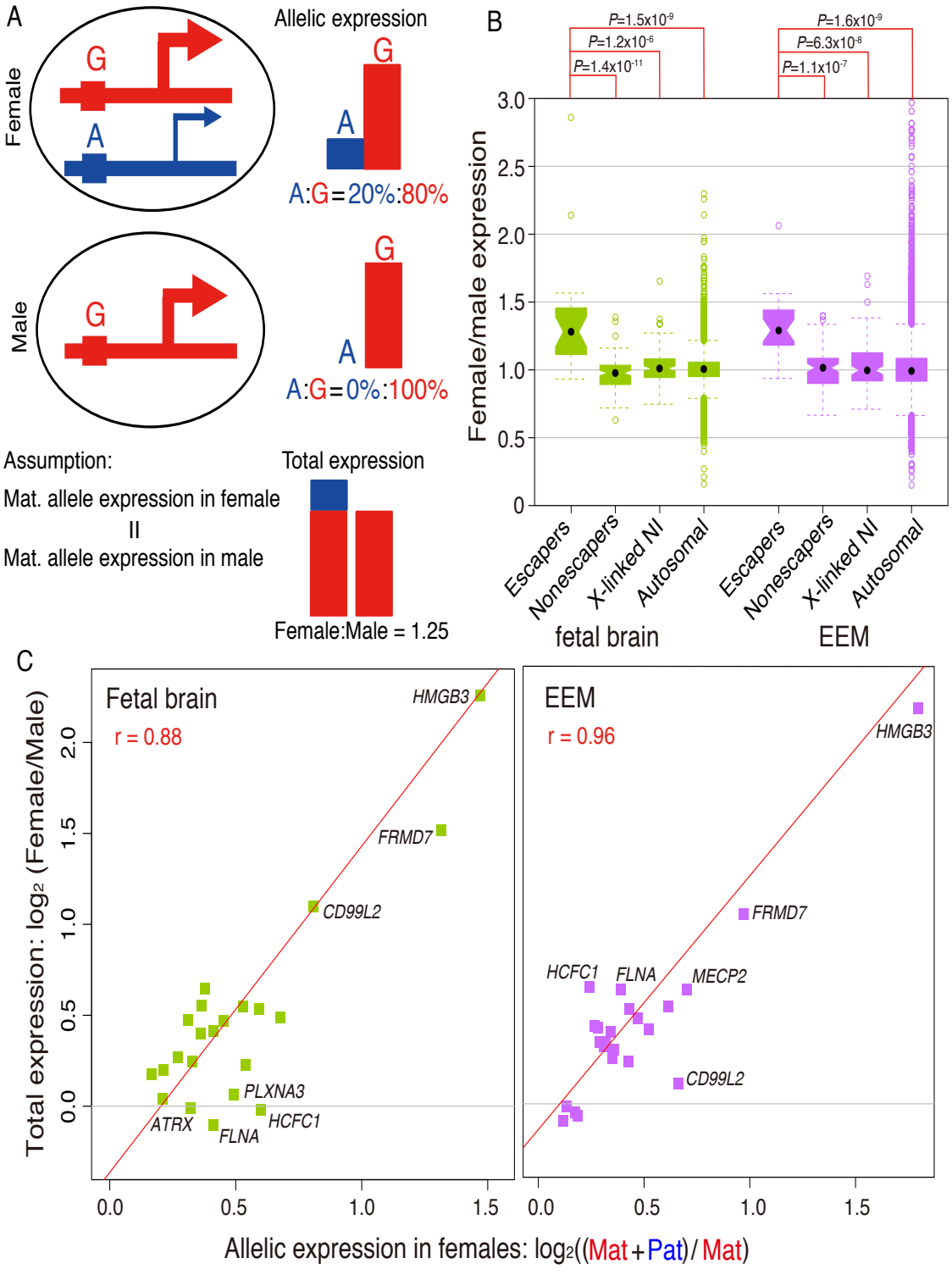
In mouse, X-inactivation patterns are dramatically different between fetal brain (rXCI) and EEM (pXCI), but this is not the case in opossum. Approximately 85% of opossum X-linked genes are subject to complete paternal allele repression through pXCI in both fetal brain and EEM, and the same set of 24 escaper genes is shared between the two tissues (Figure 1C). In terms of the level of paternal leakage, there was no

significant difference between the two tissues for 17 of the 24 escaper genes (Figure 2E). Of the remaining seven genes, six (*PLXNA3*, *DKC1*, *YIPF6*, *FRMD7*, *FAM122B* and *RBMX*) had significantly greater paternal expression in fetal brain than in EEM, and one (*HMGB3*) showed significantly higher paternal expression in EEM than in brain (Mann-Whitney U-test, *P*-value < 0.05). In addition, the maternal/paternal allele-specific expression (M/P) ratio for escapers was considerably greater than 1.0 for most (22/24) of these genes, indicating that the maternal allele was the one preferentially expressed. However, for two genes, *FRMD7* and *HMGB3,* paternal-allele expression exceeded that of the maternal allele (M/P < 1.0) (Figure 2E).

2.2.4 Genes that Escape pXCI Have Higher Total Expression in Females

In therian mammals, the disparity in X-linked gene dosage between XX females and XY males is 'compensated' by X-chromosome inactivation. For X-linked opossum genes that are subject to complete pXCI (non-escaper genes), only the maternal copy is expressed in both sexes, leading to similar total expression levels in males and females. Assuming that the expression level of the maternal allele is the same in both sexes without any sex differences or feedback compensation mechanism, higher total expression in females is expected for escaper genes with paternal leakage, and the female/male expression ratio should correlate with the degree of paternal leakage (Figure 3A). To determine if this is true for opossum escaper genes, we compared the distribution of female/male expression ratios for escaper genes, non-escaper genes, non-informative X-linked genes (a small minority of which could be unidentified escaper

Figure 3. Comparison of paternal X-linked gene expression percentage and female/male expression ratios. (A) Diagram showing individual alleles and total expression in female and male samples for a hypothetical escaper gene. In females (top panel), escaper genes have a maternal allele (G, in red) and a paternal allele (A, in blue). The maternal allele is fully active and the paternal allele accounts for 20% of the total expression. In males (middle panel), there is one maternal allele (G, in red). Assuming maternally derived allele expression is the same in females and males, the estimated female/male total expression ratio would be (80% + 20%)/80% = 1.25 (bottom panel). For non-escaper genes, the paternal allele is repressed in females so the female/male expression ratio is 1. (B) Boxplot of female/male total expression ratios for imprinted X inactivation escaper genes (Esc), non-escaper genes (Nesc), X-linked genes with no informative SNPs (X-linked NI) and autosomal genes (Auto) in fetal brain (green) and EEM (purple). The median female/male expression ratio is 1.3 for escaper genes and 1.0 for the other three groups of genes in both tissues. (C) Scatterplot of allele-specific expression ratio and female/male total expression ratios for escaper genes in fetal brain (left) and EEM (right) with normalized FPKM > 5. The y-axis is the female/male total expression ratio (log2 scale); the x-axis is (maternal + paternal)/maternal-allele expression in females (log2 scale). A linear relationship is observed for escaper genes in both fetal brain (green squares) and EEM (purple squares).


A
Female
G
A
Allelic expression
G
A
A:G=20%:80%
Male
G
G
A
A:G=0%:100%
Assumption:
Mat. allele expression in female
II
Mat. allele expression in male
Total expression
Female:Male = 1.25
B
P=1.5x10-9
P=1.2x10-6
P=1.4x10-11
P=1.6x10-9
P=6.3x10-8
P=1.1x10-7
Female/male expression
3.0
2.5
2.0
1.5
1.0
0.5
0
Escapers
Nonescapers
X-linked NI
Autosomal
Escapers
Nonescapers
X-linked NI
Autosomal
fetal brain
EEM
C
Fetal brain
r = 0.88
HMGB3
FRMD7
CD99L2
PLXNA3
ATRX
FLNA
HCFC1
EEM
r = 0.96
HMGB3
FRMD7
HCFC1
FLNA
MECP2
CD99L2
Total expression: log2 (Female/Male)
0.0
0.5
1.0
1.5
2.0
0.0
0.5
1.0
1.5
Allelic expression in females: log2((Mat+Pat)/Mat)

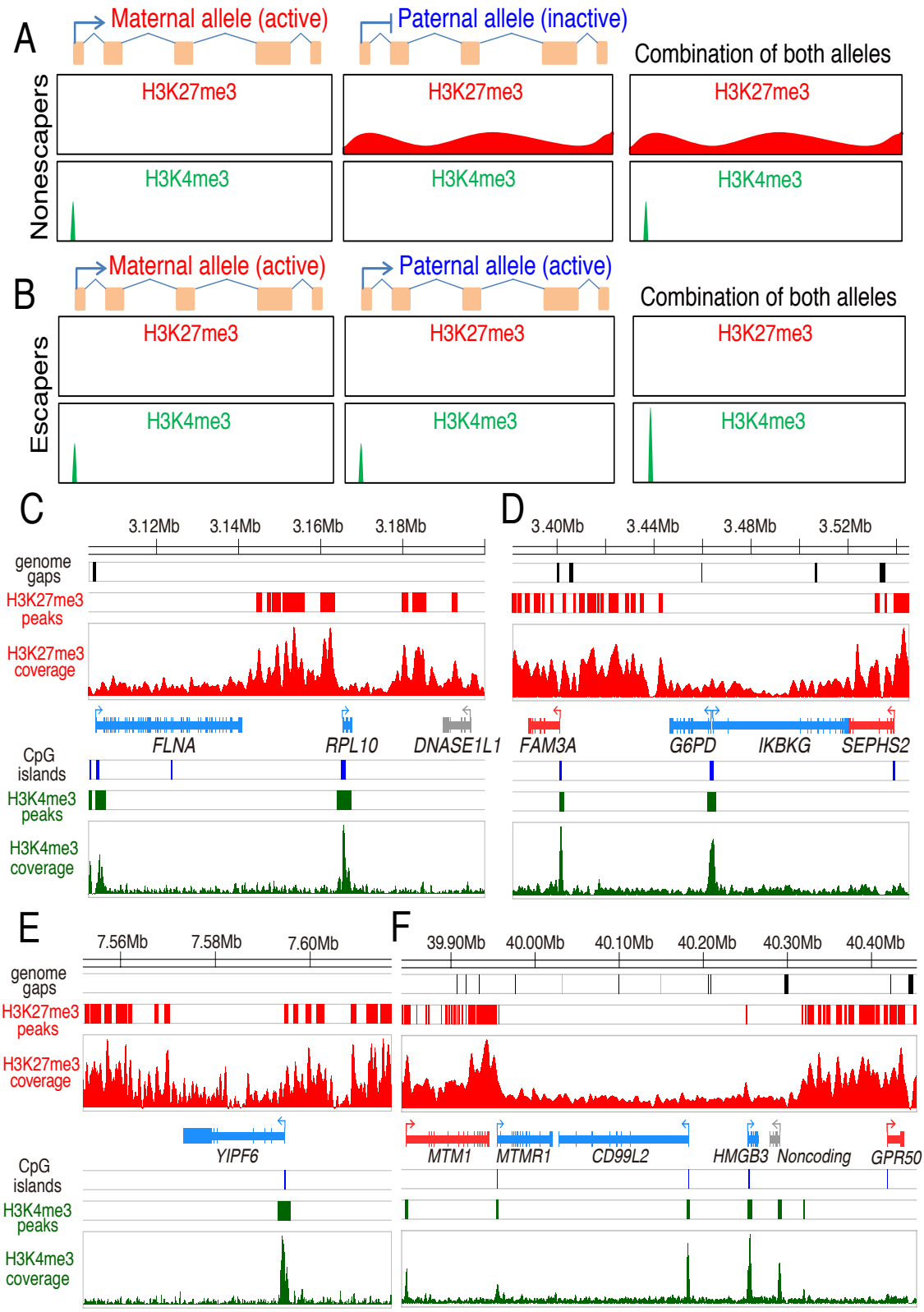
genes that could slightly inflate the female/male expression ratio of this class), and autosomal genes in both fetal brain and EEM samples (Figure 3B).

As expected, the median female/male expression ratio is much higher (1.3) for escaper genes (~1.0 for the other three groups of genes in both tissues), and the distribution for escaper genes is significantly different from the other three groups (*P*-value < 0.001, Kolmogorov-Smirnov Test). These results indicate that XCI escaper genes have significantly higher expression levels in females than in males (Figure 3B). To assess the correlation between increased female expression and the degree of paternal leakage, we plotted female/male total expression ratios against (maternal + paternal)/maternal-allele expression percentage in females. A linear relationship was observed in both fetal brain (Pearson's r = 0.88, *P*-value = 4.4×10^{-8}, Figure 3C) and EEM (Pearson's r = 0.96, *P*-value = 1.5×10^{-13}, Figure 3C) suggesting that the higher expression of escaper genes in females is due to leaky expression of alleles from the 'inactive' Xp.

2.2.5 Histone Modification Profiling by ChIP-Seq

To study histone state profiles of pXCI in opossum, we performed native ChIP-seq experiments in fetal brain for two critical epigenetic marks that correlate with X-linked gene expression in eutherian mammals: the "on-mark" (H3K4me3; tri-methylation of lysine 4 on histone H3) and the "off-mark" (H3K27me3; tri-methylation of lysine 27 on histone H3). H3K4me3 is found at promoter regions, CpG islands, and is often associated with active transcription in all eukaryotes. H3K27me3 is located throughout the gene body and associated with repressed transcription. If histone modifications are

Figure 4. Depletion of H3K27me3 marks at opossum XCI escaper genes. (A) Expected histone modification profile for non-escaper genes assuming pXCI is regulated by H3K4me3 (mark-of-activation, or on-mark) and H3K27me3 (mark-of-repression, or off-mark). For non-escaper genes, the maternal allele is active, therefore the on-mark is present at the promoter and the off-mark is absent (left panel). In contrast, the paternal allele is repressed, with the off-mark covering the gene region and a lack of on-mark at the promoter region (middle panel). In females, with two parental alleles, both on- and off-marks will be observed for non-escaper genes (right panel). Males will have only an active, maternally derived allele. (B) Expected histone modification profile for escaper genes. Since both parental alleles are active, the on-mark is present at the promoter and the off-mark is absent (left and middle panels). In females, with two parental alleles, the on-mark is present and the off-mark is depleted for escaper genes (right). (C-F) H3K4me3 and H3K27me3 histone modification profile in four X-chromosome regions in female fetal brain samples from ChIP-seq experiments. The top three diagrams in each panel are plots of genome gap locations (black bars), significant H3K27me3 peaks, and H3K27me3 coverage (red bars), respectively. Gene models are shown in the middle panel, color-coded according to pXCI status (blue: escaper; red: non-escaper; gray: unknown due to lack of an informative SNP). The bottom three diagrams of each panel show CpG island locations (blue bars), significant H3K4me3 peaks, and H3K4me3 coverage profile (green bars). For non-escaper genes (*FAM3A*, *SEPHS2*, *MTM1* and *GPR50*), the H3K4me3 mark is present at the promoter CpG island, suggesting active transcription of the maternal allele; the H3K27me3 peaks cover almost the entire gene body, consistent with repression of the paternal allele. For the escaper genes (*FLNA*, *RPL10*, *G6PD*, *IKBKG*, *YIPF6*, *MTMR1*, *CD99L2* and *HMGB3*), H3K4me3 mark is present at the promoter CpG island, suggesting active transcription; H3K27me3 marks are depleted across the gene body, consistent with the biallelic expression from both parental allele. This pattern was seen for all 23 escaper genes with an H3K4me3 peak at the promoter region (Figure A4 and Table B3).


A
Maternal allele (active)
Paternal allele (inactive)
Combination of both alleles
Nonescapers
H3K27me3
H3K4me3
B
Maternal allele (active)
Paternal allele (active)
Combination of both alleles
Escapers
H3K27me3
H3K4me3
C
3.12Mb
3.14Mb
3.16Mb
3.18Mb
genome gaps
H3K27me3 peaks
H3K27me3 coverage
FLNA
RPL10
DNASE1L1
CpG islands
H3K4me3 peaks
H3K4me3 coverage
D
3.40Mb
3.44Mb
3.48Mb
3.52Mb
FAM3A
G6PD
IKBKG
SEPHS2
E
7.56Mb
7.58Mb
7.60Mb
YIPF6
F
39.90Mb
40.00Mb
40.10Mb
40.20Mb
40.30Mb
40.40Mb
MTM1
MTMR1
CD99L2
HMGB3
Noncoding
GPR50

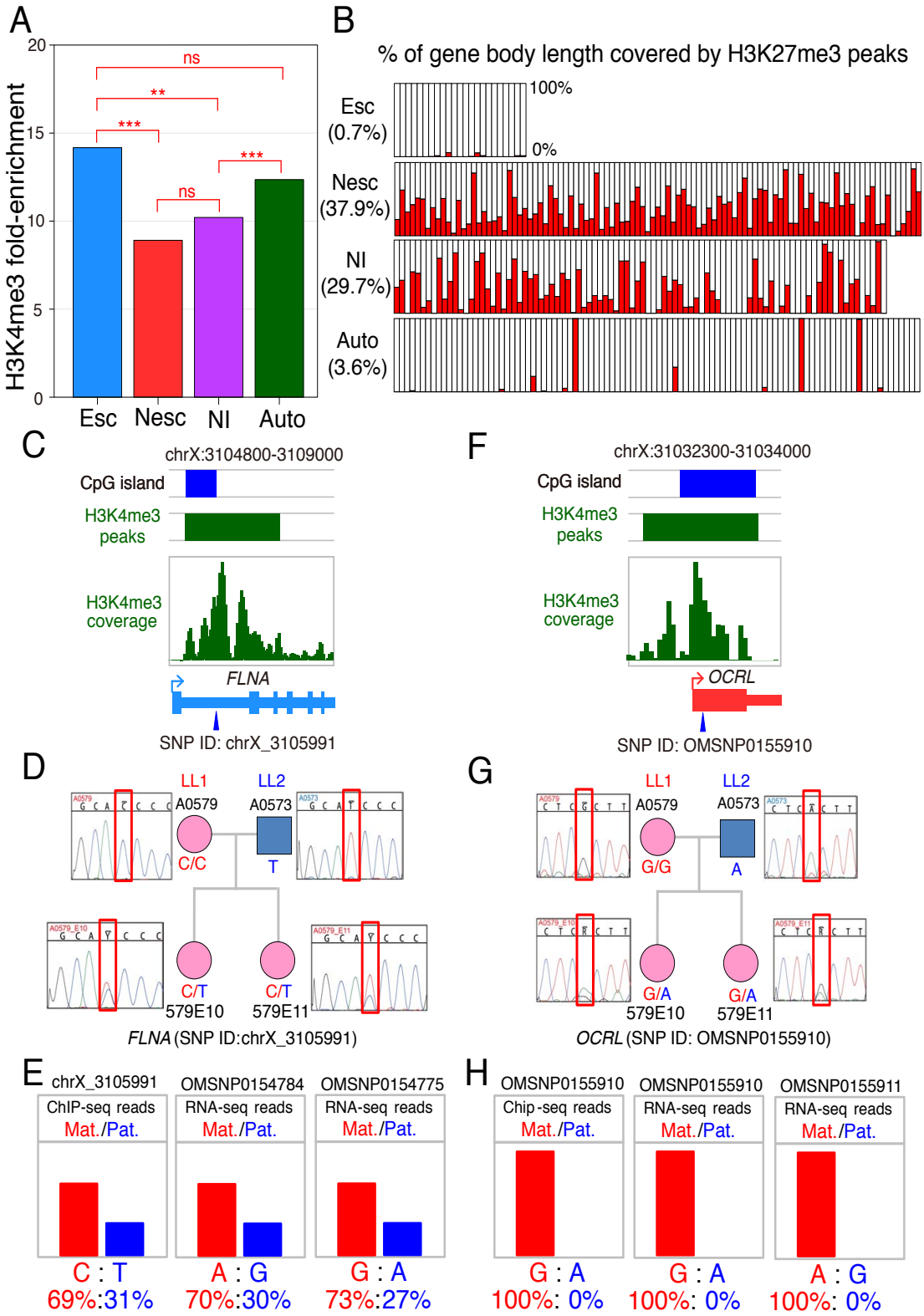
correlated with transcriptional states and possibly playing important roles in pXCI regulation for non-escaper genes in opossum, the promoter of the active maternal allele would be expected to be enriched for H3K4me3 and the repressed paternal allele enriched for H3K27me3 across the promoter and gene body. Overall, we should observe both the on-mark and off-mark simultaneously at X-linked non-escaper genes in females (Figure 4A). For escaper genes with biallelic expression, the on-mark should be present and the off-mark absent, resulting in depletion of H3K27me3 marks compared to the non-escaper genes (Figure 4B). Biallelically expressed autosomal genes are expected to have the same profile as the X-linked escaper genes.

In ChIP-seq data from female fetal brain, 23 of the 24 escaper genes showed significant H3K4me3 peaks at the promoter region, overlapping the promoter CpG islands. As expected, all 23 genes with an on-mark peak were depleted of H3K27me3 peaks beginning at the promoter and spanning the gene body, whereas the both H3K4me3 and H3K27me3 marks were present at flanking non-escaper genes (Figure 4C-F and Figure A4). In control male fibroblast cell lines, the H3K4me3 mark was present for expressed X-linked genes, and the H3K27me3 marks were absent for all expressed X-linked genes (Figure A4 and Table B3), which is consistent with the fact that males only have one X chromosome and it is active. The remarkable association between escape of pXCI and depletion of H3K27me3 marks suggests that the H3K27me3 modification is critical for the repression of the paternal allele of opossum X-linked genes, and depletion of H3K27me3 is a hallmark of escaper genes. To evaluate enrichment of on-marks (H3K4me3) we quantified peak intensity as 'fold-

enrichment' relative to background level (Table S3). The fold-enrichment was significantly higher for escaper genes than for non-escaper genes (P-value = 4.6×10^{-5}, Kolmogorov-Smirnov test) and non-informative X-linked genes (P-value = 0.004, Figure 5A), but there was no significant difference between escaper genes and autosomal genes (P-value > 0.05). This finding is consistent with the hypothesis that biallelically expressed genes have stronger on-mark peaks due to the presence of the on-mark on both parental chromosomes (Figure 4A). For the off-mark (H3K27me3), the average percentages of gene body length covered by significant peaks were calculated for escaper genes, non-escaper genes, non-informative X-linked genes, and 100 randomly selected autosomal genes (Figure 5B). The off-mark peaks were completely absent on escaper genes, and most autosomal genes lacked off-mark peaks, with a small subset of non-expressed genes having 100% off-mark coverage. In contrast, most non-escaper genes and non-informative X-linked genes were virtually covered with off-mark peaks (Figure 5B).

To further validate whether the parent-of-origin allele-specific histone modification profile was consistent with our hypothesis, we examined three escaper genes (*FLNA*, *YIPF6* and *FAM122B*) and five non-escaper genes (*OCRL*, *PNCK*, *GPC4*, *ITM2A* and *PDZD11*) that had informative SNPs overlapping the H3K4me3 peaks. For all three escaper genes, we observed ChIP-seq reads containing both SNP alleles, indicating that the on-mark was present on both parental X chromosomes (Figure 5C-E and Figure A5). For the five non-escaper genes, only maternal-allele-containing reads were found at the

Figure 5. Histone modification is correlated with maternal vs. paternal-allele expression of X-linked escaper and non-escaper genes. (A) H3K4me3 peak fold-enrichment for pXCI escaper genes (Esc), non-escaper genes (Nesc), X-linked genes with no informative SNPs (X-linked NI) and autosomal genes (Auto) in female fetal brain samples. Escaper genes show significantly higher enrichment than non-escapers or NI genes, because both parental alleles are expressed, similar to autosomal genes. Statistical significance was assessed by Kolmogorov-Smirnov test (ns: *P*-value > 0.05; **: *P*-value < 0.01; *: *P*-value < 0.001). (B) Barplot of percentage of H3K27me3 peak coverage in the gene region for pXCI escaper genes with on-mark peaks (Esc), non-escaper genes (Nesc), X-linked genes with no informative SNPs (X-linked NI) and 100 randomly selected autosomal genes (Auto) in female fetal brain samples. Average coverage is 0.7% for escaper genes, indicating the depletion of the off-marks. Among the X-linked NI genes, 11 have the similar histone modification profiles as the escaper genes, which are potential escapers of pXCI in opossum (Table B4). (C-E) Allele-specific H3K4me3 modification for escaper gene *FLNA* in female fetal brain ChIP-seq data from LL1 x LL2 cross. (C) Top to bottom: 5'-end gene model, CpG island location (blue bar), and H3K4me3 peak/coverage profile (green bars). There is a SNP (X_3105991) under the H3K4me3 peak with sufficient coverage to infer individual parent-of-origin allele-specific histone modification. (D) Sanger sequencing genotyping results for the fetuses A0579E10 and A0579E11 (used for ChIP-seq experiments) and their parents confirmed that this SNP is informative in both F_1 offspring. (E) Highly skewed H3K4me3 reads from the maternal allele (64%) and paternal allele (34%) at X_3105991, suggest both parental alleles are active. This is consistent with the allele-specific expression profile at SNP OMSNP0154784 in the RNA-seq data and SNP OMSNP0154775 from the allele-specific pyrosequencing results. (F-H) Allele-specific H3K4me3 modification for nonescaper gene *OCRL* in female fetal brain ChIP-seq data from LL1 x LL2 cross. (F) Top to bottom: 5'-end gene model, CpG island location (blue bar), and the H3K4me3 peak/coverage profile (green bars). There is a SNP (OMSNP0155910) under the H3K4me3 peak with sufficient coverage to infer individual allele histone modifications. (G) Sanger sequencing genotyping results in the two embryos (A0579E10 and A0579E11) and their parents confirmed that this SNP is informative in both embryos. (H) From the ChIP-seq data, we observed 100% of the H3K4me3 reads from the maternal allele at OMSNP0155910, suggesting the on-mark is only present at the maternal allele. This is consistent with the maternal-specific expression at OMSNP0155910 and OMSNP0155911 in the RNA-seq data. Allele-specific ChIP-seq quantification for two additional escaper genes (*YIPF6* and *FAM122B*) and four additional non-escaper genes (*PNCK*, *GPC4*, *ITM2A* and *PDZD11*) in female fetal brain can be found in Figure A5.**

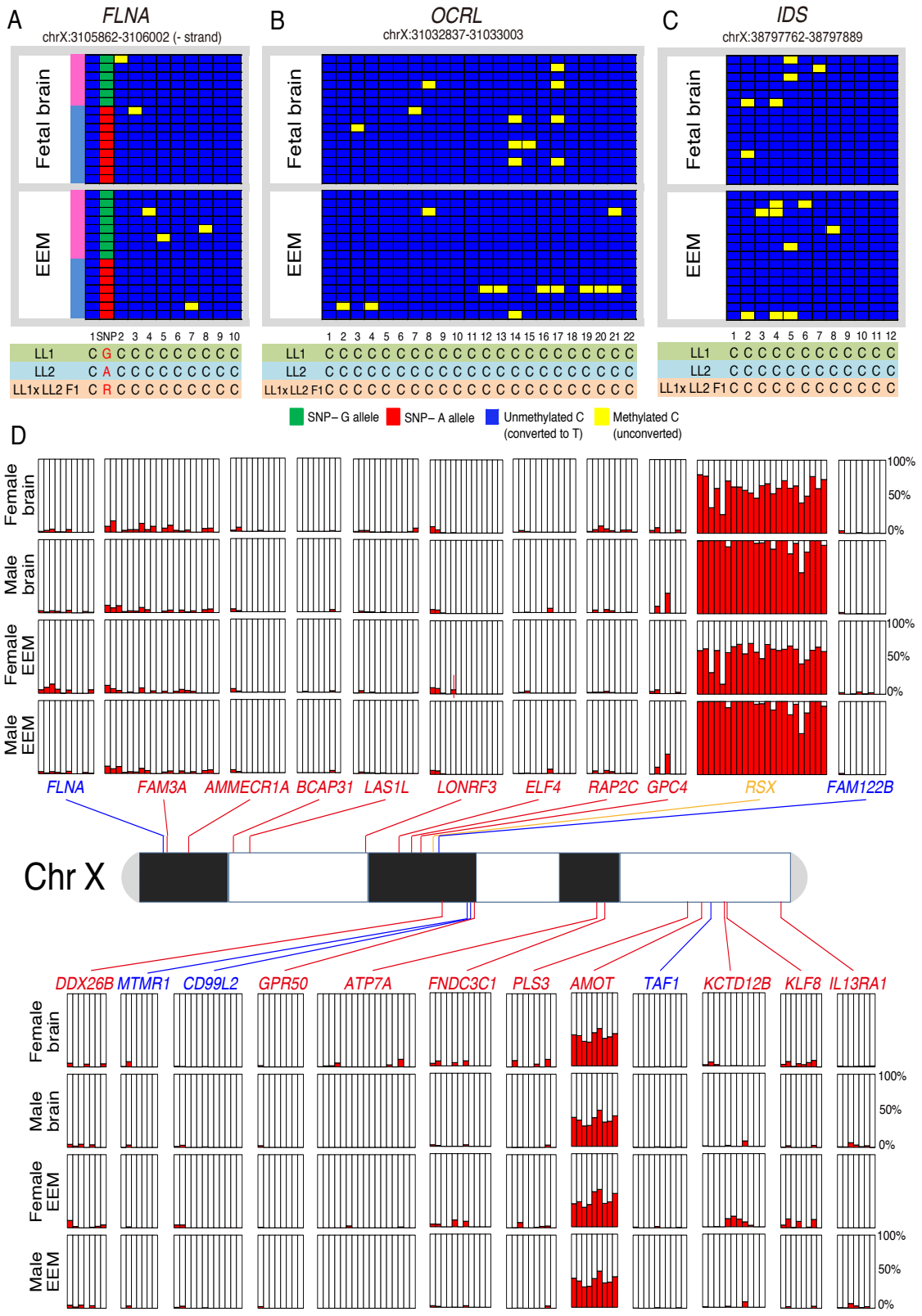

A
H3K4me3 fold-enrichment
20
15
10
5
0
ns
**
***
***
ns
Esc
Nesc
NI
Auto
B
% of gene body length covered by H3K27me3 peaks
Esc (0.7%)
Nesc (37.9%)
NI (29.7%)
Auto (3.6%)
100%
0%
C
chrX:3104800-3109000
CpG island
H3K4me3 peaks
H3K4me3 coverage
FLNA
SNP ID: chrX_3105991
F
chrX:31032300-31034000
CpG island
H3K4me3 peaks
H3K4me3 coverage
OCRL
SNP ID: OMSNP0155910
D
LL1
LL2
A0579
A0573
C/C
T
C/T
C/T
579E10
579E11
FLNA (SNP ID:chrX_3105991)
G
LL1
LL2
A0579
A0573
G/G
A
G/A
G/A
579E10
579E11
OCRL (SNP ID: OMSNP0155910)
E
chrX_3105991
ChIP-seq reads Mat./Pat.
C : T
69%:31%
OMSNP0154784
RNA-seq reads Mat./Pat.
A : G
70%:30%
OMSNP0154775
RNA-seq reads Mat./Pat.
G : A
73%:27%
H
OMSNP0155910
Chip-seq reads Mat./Pat.
G : A
100%: 0%
OMSNP0155910
RNA-seq reads Mat./Pat.
G : A
100%: 0%
OMSNP0155911
RNA-seq reads Mat./Pat.
A : G
100%: 0%

SNP positions, consistent with allele-specific epigenetic modification resulting in monoallelic-maternal expression (Figure 5F-H and Figure A5). From the overall and allele-specific histone modification profiles of escaper and non-escaper genes in fetal brain, we conclude that histone modifications are an important epigenetic feature of pXCI in opossum and could possibly be involved directly in pXCI regulation.

2.2.6 Absence of Promoter CpG Island DNA Methylation in Both Escaper and Non-Escaper Genes

In addition to histone modifications, in eutherian mammals differential DNA methylation of promoters is a key regulatory mechanism for XCI. Promoter CpG islands are methylated exclusively on the inactive allele of X-linked genes (Lock *et al.* 1986).If DNA methylation plays the same role in opossum XCI, differential methylation (on the inactive paternal allele) would be present at non-escaper genes, while little or no methylation would be expected for escaper genes, consistent with their biallelic expression. Lack of methylation is also expected at promoters of all expressed X-linked genes in males, because males have only one (active) X chromosome. To study whether promoter DNA methylation is correlated with XCI status in opossum, we quantified DNA methylation percentages at promoter CpG islands for 24 X-linked genes using bisulfite sequencing and PyroMark assays. Promoters of all five escaper genes assayed (*FLNA*, *FAM122B*, *MTMR1*, *CD99L2* and *TAF1*) were unmethylated in female samples, showing less than 1% average methylation across all CpG sites in the promoter regions, which is consistent with biallelic expression (Figure 6A, Figure A6 and Table B5). Surprisingly, non-escaper genes also lacked differential methylation in females.

Figure 6. Promoter CpG islands are unmethylated at both escaper and non-escaper genes. (A-C) Bisulfite sequencing of the promoter CpG islands for an escaper gene *FLNA* (A) and two non-escaper gene, *OCRL* (B) and *IDS* (C) in opossum female fetal brain and EEM. Each panel shows multiple CpG sites at the promoter for the corresponding gene/tissue. Yellow boxes depict methylated CpGs and blue boxes are unmethylated CpGs. A SNP was found in the *FLNA* promoter and the transmission direction could be inferred, enabling quantification of methylation on each allele. The maternal allele is G (green boxes) and the paternal allele is A (red boxes). The promoter CpG islands were not methylated in these tissues. (D) Profile of promoter CpG island methylation status in 5 escaper genes (labeled in blue), 17 non-escaper genes (in red) and *Rsx* (in orange) using the PyroMark assay in female and male fetal brain and EEM samples. Each red bar represents the methylation percentage at one CpG site, from 0% to 100%. The raw pyrograms for selected genes are shown in Figure A6 and Figure A7. All five escaper genes (*FLNA*, *FAM122B*, *MTMR1*, *CD99L2* and *TAF1*) lacked methylation, with 1.0% average methylation across all genes/tissues. 16 of the 17 non-escaper genes (*FAM3A, AMMECR1, BCAP31, LAS1L, LONRF3, ELF4, RAP2C, GPC4, DDX26B, GPR50, ATP7A, FNDC3C1, PLS3, KCTD12B, KLF8* and *IL13RA1*) also lacked methylation, with 1.6% average methylation across all genes/tissues (Table B5). *AMOT* has ~ 40% methylation across promoter CpG sites, but there is no difference between female and male samples.

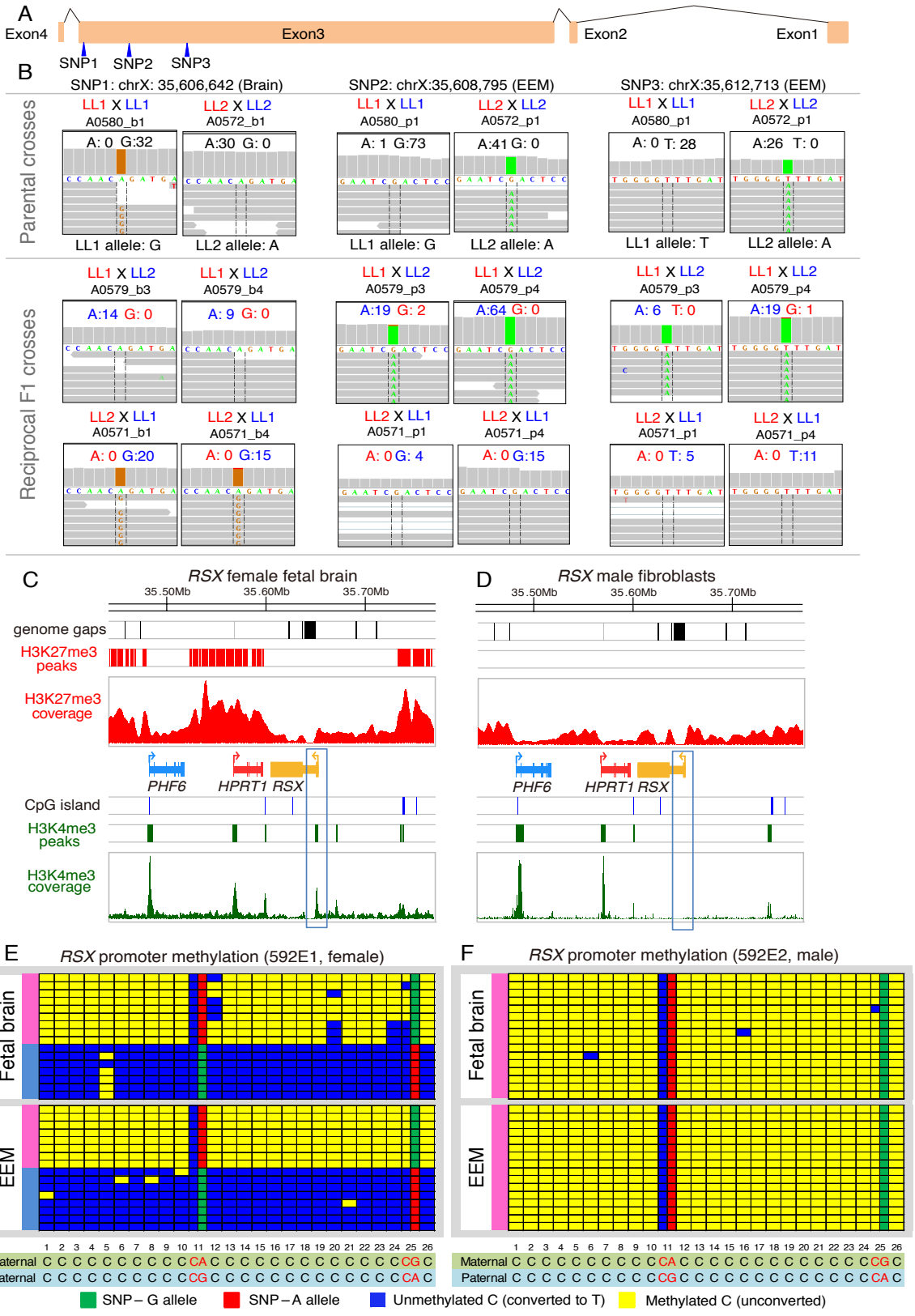

A
FLNA
chrX:3105862-3106002 (- strand)
B
OCRL
chrX:31032837-31033003
C
IDS
chrX:38797762-38797889
Fetal brain
EEM
SNP
LL1
LL2
LL1x LL2 F1
SNP– G allele
SNP– A allele
Unmethylated C (converted to T)
Methylated C (unconverted)
D
Female brain
Male brain
Female EEM
Male EEM
100%
50%
0%
FLNA
FAM3A
AMMECR1A
BCAP31
LAS1L
LONRF3
ELF4
RAP2C
GPC4
RSX
FAM122B
Chr X
DDX26B
MTMR1
CD99L2
GPR50
ATP7A
FNDC3C1
PLS3
AMOT
TAF1
KCTD12B
KLF8
IL13RA1

For 18 of the 19 assayed non-escaper genes, promoter CpG islands were unmethylated in both fetal brain and EEM tissues of both sexes (*OCRL*, *IDS*, *FAM3A, AMMECR1, BCAP31, LAS1L, LONRF3, ELF4, RAP2C, GPC4, DDX26B, GPR50, ATP7A, FNDC3C1, PLS3, KCTD12B, KLF8* and *IL13RA1*), with an overall average methylation percentage of 1.6% (Figure 6B-D, Figure A6 and Table B5). The remaining gene (*AMOT*) exhibited about 40% methylation across promoter CpG sites, but there was no difference between the sexes (Figure 6D). Thus, promoter CpG islands are unmethylated for both escaper and non-escaper genes, and there is no correlation between promoter CpG DNA methylation and X-inactivation status in opossum.

2.2.7 Allele-Specific Expression and Epigenetic Profile for the Non-Coding *Rsx* Gene

In eutherians, an X-linked non-coding transcript (*Xist*) is transcribed exclusively from the inactive X chromosome (Xi) in females and coats the Xi in *cis* to repress the expression of genes from this chromosome. Recently, a female-specific non-coding transcript (*Rsx*) with *Xist*-like functional properties was discovered in a number of adult tissues in opossum. DNA-RNA duplex FISH results showed indirect evidence that *Rsx* transcripts coat the Xi in a manner similar to those from *Xist* (Grant *et al.* 2012). Here, we independently confirmed, in different tissues and developmental stages, that *Rsx* is indeed expressed exclusively in females. Three informative SNPs between LL1 and LL2 stocks were found in the non-repetitive region in exon 3, and we were able to directly quantify the parent-of-origin allelic expression level for *Rsx* in female E13 fetal brain and EEM (Figure 7A). Only paternal allele reads were detected in both tissues,

Figure 7. Allele-specific expression, allele-specific methylation and histone modification profile for the non-coding RNA *Rsx*. (A) Exon models for the *Rsx* gene, with three SNPs between LL1 and LL2 lines indicated (blue triangles). (B) Allele-specific expression analysis of RNA-seq data for SNP1 in fetal brain and SNP2, 3 in EEM in female individuals. Monoallelic-paternal expression was found at all three SNP positions in both tissues. (C-D) H3K4me3 and H3K27me3 histone modification profile at the *Rsx*-containing region in female fetal brain (C) and control male fibroblast samples (D) from ChIP-seq experiments. In each panel, plotted from the top to the bottom, are the genome gap locations (black bars), significant H3K27me3 peaks (red bars), H3K27me3 coverage (red), gene models, CpG island locations (blue bars), significant H3K4me3 peaks (green bars) and H3K4me3 coverage (green). The escaper gene (*PHF6*) in this region is labeled in blue, the non-escaper gene (*HPRT1*) is labeled in red. and *Rsx* is labeled in orange. In females, the on-mark (H3K4me3) was present at the *Rsx* promoter (in blue box), which is consistent with the active transcription of the paternal allele, whereas in males the *Rsx* on-mark peak is missing, consistent with absence of transcription. However, H3K27me3 marks were depleted in the gene body in both sexes suggesting that, unlike non-escaper genes, expression of *Rsx* is not regulated by the H3K27me3 modification. (E, F) Bisulfite sequencing of *Rsx* promoter region in female (E) and male (F) fetal brain and EEM samples. Each panel shows multiple CpG sites at the promoter for the corresponding sex/tissue. Yellow boxes depict methylated CpGs and blue boxes depict unmethylated CpGs. Two SNPs found in this region (position 11 and 25) were genotyped in both parents by Sanger sequencing to infer the transmission direction and quantify methylation of the two alleles. In males, the promoter CpG island was 100% methylated across all 25 CpG sites in both tissues, consistent with silencing of *Rsx* in males; in females, the maternal allele is methylated and the paternal allele is unmethylated in both tissues, consistent with monoallelic-paternal expression of *Rsx*. These findings suggest that the *Rsx*-allele expression is regulated by differential allelic methylation, but not influenced by histone modification states.


A
Exon4
Exon3
Exon2
Exon1
SNP1
SNP2
SNP3
B
SNP1: chrX: 35,606,642 (Brain)
SNP2: chrX:35,608,795 (EEM)
SNP3: chrX:35,612,713 (EEM)
Parental crosses
LL1 X LL1
A0580_b1
A: 0 G:32
LL2 X LL2
A0572_b1
A:30 G: 0
LL1 allele: G
LL2 allele: A
LL1 X LL1
A0580_p1
A: 1 G:73
LL2 X LL2
A0572_p1
A:41 G: 0
LL1 allele: G
LL2 allele: A
LL1 X LL1
A0580_p1
A: 0 T: 28
LL2 X LL2
A0572_p1
A:26 T: 0
LL1 allele: T
LL2 allele: A
Reciprocal F1 crosses
LL1 X LL2
A0579_b3
A:14 G: 0
LL1 X LL2
A0579_b4
A: 9 G: 0
LL1 X LL2
A0579_p3
A:19 G: 2
LL1 X LL2
A0579_p4
A:64 G: 0
LL1 X LL2
A0579_p3
A: 6 T: 0
LL1 X LL2
A0579_p4
A:19 G: 1
LL2 X LL1
A0571_b1
A: 0 G:20
LL2 X LL1
A0571_b4
A: 0 G:15
LL2 X LL1
A0571_p1
A: 0 G: 4
LL2 X LL1
A0571_p4
A: 0 G:15
LL2 X LL1
A0571_p1
A: 0 T: 5
LL2 X LL1
A0571_p4
A: 0 T:11
C
RSX female fetal brain
35.50Mb
35.60Mb
35.70Mb
genome gaps
H3K27me3 peaks
H3K27me3 coverage
PHF6
HPRT1
RSX
CpG island
H3K4me3 peaks
H3K4me3 coverage
D
RSX male fibroblasts
35.50Mb
35.60Mb
35.70Mb
PHF6
HPRT1
RSX
E
RSX promoter methylation (592E1, female)
Fetal brain
EEM
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26
Maternal C C C C C C C C C CA C C C C C C C C C C C C CG C
Paternal C C C C C C C C C CG C C C C C C C C C C C C CA C
F
RSX promoter methylation (592E2, male)
Fetal brain
EEM
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26
Maternal C C C C C C C C C CA C C C C C C C C C C C C CG C
Paternal C C C C C C C C C CG C C C C C C C C C C C C CA C
SNP – G allele
SNP – A allele
Unmethylated C (converted to T)
Methylated C (unconverted)

indicating that the expression of *Rsx* is exclusively from the Xp (Figure 7B). The histone modification profile for *Rsx* shows the presence of the on-mark (H3K4me3) in females but not in males, which is consistent with female-specific expression (Figure 7C,D). If repression of the maternal *Rsx* allele is correlated with the off-mark (H3K27me3), similar to non-escaper genes, then the off-mark peak should be present in both females and males. However, we did not observe any off-mark peaks across the entire *Rsx* gene body in either sex (Figure 7C,D).

To assess the *Rsx* DNA methylation profile, we searched for and annotated a novel CpG island at the proposed *Rsx* promoter, which is associated with an H3K4me3 peak. We quantified the DNA methylation percentage at all 23 CpG sites within the *Rsx* promoter CpG island in fetal brain and EEM of both sexes. Unlike other X-linked genes that are subject to pXCI, the *Rsx* CpG island was unmethylated in males but differentially methylated in females in both tissues; maternal alleles showed 100% methylation and paternal alleles showed 0% methylation (Figure 7E-F, Figure A6 and Table B5), which is consistent with exclusive paternal-allele expression. Therefore, we propose that *Rsx* parent-of-origin allelic expression is regulated by differential promoter DNA methylation, but does not involve the H3K27me3 histone modification mechanism that is associated with X-linked escaper and non-escaper gene expression.

2.3 Discussion

2.3.1 XCI Completely Silences Most Paternal X-Linked Alleles in Opossum, but a Substantial Fraction Escape Inactivation in Fetal Brain and EEM

Unlike mouse and rat females which display random X-chromosome inactivation (rXCI) in both embryonic and adult tissues and paternally-imprinted X inactivation (pXCI) in extra-embryonic tissues and placenta, pXCI is characteristic of both somatic tissues and extra-embryonic membranes (EEM) in marsupials. Results from five previously assayed opossum X-linked genes suggest that pXCI is incomplete with locus-, species-, and tissue-specific levels of paternal allele leakage (incomplete repression). However, it is unclear whether this incomplete pXCI status is the general pattern for all X-linked genes or is a biased reflection of the small number of genes examined. To address this question, we performed RNA-seq analyses of fetal brain and EEM samples from animals from two opossum laboratory lines and their reciprocal F_1 offspring. Parent-of-origin allele-specific expression was quantified chromosome-wide to determine relative expression of SNP alleles unambiguously attributable to each of the parents.

Of the X-linked genes examined, 86% (152/176) showed ~100% monoallelic maternal expression, and 14% (24/176) exhibited strong maternal-allele expression accompanied by leaky expression from the paternally derived allele. Most of the latter (22/24) showed a paternal-allele contribution of less than 50% to the total mRNA for each respective gene. These results demonstrate conclusively for the first time that the

large majority of genes on a marsupial X-chromosome is subject to complete pXCI with only a small minority (14%) of partial escapers in E13 fetal brain and EEM.

2.3.2 The pXCI Pattern is Remarkably Similar Between Opossum Fetal Brain and EEM

In some eutherian mammals, such as mouse and rat, rXCI takes place in somatic cells, whereas pXCI occurs in trophectoderm, resulting in distinctly different patterns of XCI gene silencing and escape in the embryo proper and its extra-embryonic membranes. In contrast, the X-inactivation status is remarkably similar between opossum E13 fetal brain and EEM with exactly the same set of escaper genes in both tissues and similar paternal-allele expression for each gene as well (conserved escape). Similarity of the XCI profiles of fetal brain and EEM and the lack of tissue-specific escapers could be consequences of both tissues being derived from adjacent regions of the unilaminar blastocyst in early marsupial development, as opposed to their origins from discrete inner cell mass and trophectoderm structures in eutherians (Reviewed by (Zeller & Freyer 2001) (Selwood & Johnson 2006); see also (Mate *et al.* 1994).

2.3.3 DNA Methylation and Histone Modification Profiles Reveal Potential Regulation Mechanisms of pXCI in Opossum

Previous epigenetic studies of pXCI at the cytogenetic level in marsupials have suggested strongly that histone modifications are critical in the regulation of pXCI (Koina *et al.* 2009; Mahadevaiah *et al.* 2009; Rens *et al.* 2010; Chaumeil *et al.* 2011); however, there are inconsistent results concerning the role of DNA methylation in regulation (Loebel & Johnston 1993, 1996; Rens *et al.* 2010). To acquire base-pair

resolution to infer histone states at the individual-gene level, we performed the first genome-wide ChIP-seq analyses in opossum and found both H3K4me3 (on-marks) and H3K27me3 (off-marks) to be strongly correlated with the inactivation status of X-linked genes. Depletion of H3K27me3 was observed only for escaper genes, consistent with expression from both alleles of these genes. In addition, promoter DNA methylation was absent for all but one gene examined, and methylation level was not correlated with parent-of-origin allele-expression for any X-linked genes.

We also characterized transcription of *Rsx*, the recently discovered *Xist*-like gene that is expressed exclusively from the Xp in adult opossum female cells. As previously reported in adult tissues, *Rsx* expression was found to be extremely female-biased in fetal brain and EEM, and we directly demonstrated monoallelic expression of the paternally derived *Rsx* allele in both tissues. Beyond this, we profiled histone modifications across the gene as well as DNA methylation at the promoter. H3K27me3 peaks were not found on the *Rsx* gene body, and, in contrast to escaper genes, the CpG island at the *Rsx* promoter is differentially methylated in females, with 100% methylation of the maternal allele and virtually no methylation of the paternal allele. These findings suggest the functional convergence of the epigenetic regulatory mechanisms of *Rsx* and *Xist* despite the absence of sequence homology between these two non-coding transcripts. It remains unresolved whether *Rsx* and *Xist* have a single common origin in a distant ancestor, or whether this is a case of true evolutionary convergence.

2.3.4 Predicting XCI Status of Genes Using Escaper/Non-Escaper Characteristics

The histone data also allow us to predict escaper genes that do not have informative SNPs. In this study, about 60% of the expressed X-linked loci examined, or 176 genes, had informative SNPs in the reciprocal crosses from which we discovered 24 pXCI escaper genes in E13 fetal brain and EEM. These escaper genes share four major characteristics: 1. By definition, they display biallelic expression in allele-specific RNA-seq and pyrosequencing results; 2. The H3K27me3 marks are depleted across the entire gene body, consistent with biallelic expression; 3. Fold-enrichment of H3K4me3 peaks is significantly higher for escaper genes than for non-escapers; and 4. Almost all escaper genes show significantly higher expression in females relative to males. Among 136 expressed X-linked genes without informative SNPs, we searched for candidate escaper genes using to the second, third, and fourth characteristics (H3K27me3 peak coverage < 5%, H3K4me3 fold-enrichment > 8, and female/male expression ratio > 1.05) and found 11 additional candidate-escaper genes (Table B4). Three of these candidate escapers were confirmed as genuine escapers in fibroblast cell lines with informative SNPs (data not shown). Overall, these results comprise the first comprehensive catalog of pXCI status in opossum fetal brain and EEM, and show that between 10 and 15% of X-linked genes escape pXCI.

2.3.5 A Model for pXCI Regulation in Opossum

Based on our data, we propose the following model for the epigenetic regulation of pXCI in opossum. At some stage of germ line development or gametogenesis, the *Rsx* promoter is methylated in female opossums, but not in males, resulting in an

imprinted (silenced) *Rsx* on maternally derived X chromosomes. After fertilization, this imprint results in paternal-specific expression of *Rsx* in the developing embryo. The *Rsx* transcript coats the paternal X chromosome in *cis* (similar to the *Xist* transcript on eutherian X chromosomes), and recruits the polycomb group (PcG) proteins and other factors to target H3K27 for methylation on the paternally derived X chromosome. For reasons yet to be determined, escaper genes are avoided during this chromatin remodeling process and remain active on both parental chromosomes. In eutherian mammals, the promoter CpG islands of X-linked genes that are subject to imprinted expression in the placenta are methylated on the paternal allele only, presumably to strengthen/stabilize their repression (Lock *et al.* 1987). In opossum, we did not observe promoter DNA methylation of X-linked genes (except *Rsx*), suggesting a de-coupling of histone modification and promoter DNA methylation. In summary, was propose that pXCI in opossum arises from a *Xist*-like regulating mechanism via the non-coding *Rsx* gene and H3K27me3 the histone modifications, but promoter DNA methylation does not play a direct role.

2.3.6 Presence of Y-Linked Homologs

No extensive sequence data from the opossum Y chromosome have been published, but two genes, *ATRX* and *RSP4,* have been shown to have X- and Y-linked homologs in *M. domestica* (Jegalian & Page 1998; Carvalho-Silva *et al.* 2004). Thus, the possibility that some X-linked escaper genes found in our study could have undetected Y-linked homologs must be considered. In another marsupial, the tammar wallaby (family Macropodidae, Australia), ten homologs of X-linked genes have been detected and

mapped to the Y chromosome (Murtagh *et al.* 2012). In our study, six of these genes were found to have informative SNPs, and all six escaped pXCI in both fetal brain and EEM. Interestingly, four opossum escaper genes (*ATRX*, *PHF6*, *HCFC1* and *RBMX*) have very similar levels of expression in males and females (female/male ratio < 1.2) in fetal brain or EEM, unlike the other escaper genes, which have higher expression in females. If these genes have retained their active Y-linked homologs in opossum, their expression from the Y chromosome could equalize dosage between the sexes, and at least partially explain the lack of inactivation at these loci in females.

2.3.7 Evolution of pXCI in Therian Mammals

The idea that the eutherian and metatherian patterns of XCI are descended from an ancient form of XCI in the common therian ancestor has long been attractive for its parsimony (Cooper *et al.* 1971; Riggs 1990; Wakefield *et al.* 1997). The occurrence of pXCI in extra-embryonic tissues of rodents suggests that the marsupial pXCI pattern might represent the primitive state from which rXCI later arose in the cells of the inner cell mass, the uniquely eutherian developmental structure that gives rise to the embryo proper. However, the discovery that marsupials lack the *Xist* locus (Duret *et al.* 2006; Davidow *et al.* 2007; Hore *et al.* 2007a; Shevchenko *et al.* 2007) casts doubt on the evolutionary relationship between pXCI and rXCI, suggesting they might have evolved independently. In this paper, we describe the occurrence of paternal-specific *Rsx* expression and maternal-specific promoter methylation, as well as histone modification profiles for X-linked genes in female opossums. The similarity of the *Rsx* expression pattern and transcript function to that of the eutherian *Xist* locus, together with the

fundamental association of specific epigenetic marks with X-linked gene activity states in both opossum and eutherian mammals is comfortably consistent with a common origin for paternal and random XCI. However, independent origins and convergence of *Rsx* and *Xist* functions is also a credible possibility, and epigenetic processes are common to gene regulatory systems across the biotic world. Indeed, widespread differential promoter DNA methylation for X-linked genes has not been found in opossum, platypus (a prototherian mammal) or non-mammalian vertebrates, suggesting that DNA methylation-dependent regulation at promoter CpG islands is the acquired state in eutherian mammals. Overall then, the molecular similarities alone do not address the issue of the ancestral vs. derived relationship between the two XCI patterns. Lacking corresponding data from an appropriate outgroup species to establish the evolutionary polarity to these molecular characteristics, the proposal that rXCI arose from an ancestral pXCI mechanism remains attractive, but speculative.

2.3.8 Effective X-Linked Hemizygosity in Female Marsupials: Implications and Solutions

In this study, we found that opossum *Rsx* is a paternally expressed (maternally imprinted) gene not only in EEM but also in fetal brain. For the annotated opossum X-linked genes examined, all non-escaper genes are 100% imprinted with monoallelic, maternal expression, while escaper genes exhibit biallelic, but unequal expression of both alleles, with all but two showing preferential expression from the maternal allele. The opossum X thus acts as a chromosome-wide paternal imprinting cluster with a small minority of genes escaping imprinting and one maternally imprinted gene (*Rsx*) in both

EEM and somatic tissues. As a result, the female is effectively hemizygous for the majority of X-linked genes, as is the male. This condition will manifest the presence of deleterious recessive mutations on the Xm in all female cells, greatly reducing fitness in females, just as in males. By contrast, rXCI in the somatic tissues of eutherian mammals confers some of the advantages of heterozygosity because females express both parental alleles at X-linked loci, albeit individually in different cells. Through rXCI, dosage compensation between the two sexes is achieved without sacrificing many of the advantages of diploidy. The theoretical population genetics of this problem has been investigated, and conditions for invasion of random X-inactivation into a population with imprinted X-inactivation is clearly favored under conditions that depend on dominance and the degree of sex-specific selection (Connallon & Clark 2013).

However, not all opossum X-linked genes are subject to pXCI. It could be argued that the escaping status of some genes is the derived state because ~85% of X-linked genes have 100% monoallelic maternal expression and most escaper genes show preferential maternal expression. Such leaky expression might have been established during evolution by positive selection, although we cannot rule out other forms of selection acting on these loci. By expressing both alleles, the hemizygosity problem is solved; and, by selectively escaping X-inactivation, escaper genes could be up-regulated in total expression levels in females. Furthermore, there is almost no overlap between the opossum pXCI escapers and human/mouse rXCI escapers (Table B6) (Carrel & Willard 2005; Yang *et al.* 2010), which could be due to pXCI escapers of marsupials

facing different selection pressures from those that impinge upon rXCI escapers of eutherian mammals.

2.4 Methods

2.4.1 Opossum Tissue Dissection and RNA Extractions

Embryonic day 13 (E13) fetal brain and extra-embryonic membranes (EEM) were dissected from reciprocal F_1 crosses (Supplemental Figure S1 and Table S1) of two random-bred stocks of *Monodelphis domestica*, LL1 and LL2 (VandeBerg & Williams-Blangero 2010). Fetal sex was determined as described in Chapter V. Total RNAs were extracted using BCP (1-bromo-2 chloropropane), precipitated with isopropanol, and resuspended in RNase-free water. Potential DNA contamination was removed by both DNase I treatment and Qiagen RNeasy Plus Mini kit (Qiagen, CA). See Chapter V for further details.

2.4.2 mRNA-Seq and Data Analysis

mRNA-seq libraries were made using Illumina TruSeq RNA Sample Prep Kit and sequenced on a HiSeq 2000 platform (Illumina Inc., CA). Image analysis and base calling were performed using Illumina software. RNA-seq reads were aligned to the reference genome assembly (MonDom5) using TopHat v1.4.1 (Trapnell *et al.* 2009) with three mismatches allowed. The metric for total expression level, FPKM (Fragments Per Kilobase-pair of exon Model), was calculated for all samples using Cufflinks v1.3.0 (Trapnell *et al.* 2010) based on all mapped reads. The RNA-seq data were deposited in the *Gene Expression Omnibus* (*GEO*) *database* under accession number GSE45211.

2.4.3 Quantification of Parent-of-Origin Allelic Expression from RNA-Seq Data

68,000 SNPs were called in combined RNA-seq data from uniquely mapped reads using SAMtools software (Li *et al.* 2009). Allele-specific expression was quantified as the ratio of the number of reference allele-containing reads divided by the total coverage at each high-coverage SNP position (≥8 X) in brain and EEM tissues (Wang *et al.* 2008; Wang *et al.* 2011). Allele transmission direction was inferred from the parental crosses and supplemented using additional information from other LL1 individuals for which RNA-seq data was available (data not shown). See Chapter V.

2.4.4 Validation of X-Linked Escaper and Non-Escaper Genes by Allele-Specific Pyrosequencing

To verify the escaping status, allele-specific pyrosequencing were performed on all 24 escaper genes, one non-escaper gene (*HPRT1*), and one autosomal control gene (*GPM6B*) (Figure 2 and Figure A2) at informative SNP positions confirmed by Sanger sequencing. For non-escaper genes, 20 SNPs (one per gene) were randomly selected and genotyped by Sanger sequencing. All 20 were verified as heterozygous in at least two of the four female samples (Appendix B: Table B2). See Chapter V for further details.

2.4.5 ChIP-Seq and Data Analysis

Native-ChIP was conducted on a primary fibroblast cell line (derived from adult male ear pinna), fetal brain, and EEM using a method modified from (Dindot *et al.* 2009). Illumina libraries were constructed at Global Biologics, LLC, and sequenced at the University of Missouri – Columbia DNA Core Facility and Genomics Resources Core Facility at Weill Cornell Medical College. Raw reads were quality filtered,

trimmed, and aligned using Bowtie in the Galaxy suite (Giardine *et al.* 2005; Blankenberg *et al.* 2010; Goecks *et al.* 2010). Aligned reads were visualized on the UCSC genome browser (Kent *et al.* 2002) and IGV (Robinson *et al.* 2011; Thorvaldsdottir *et al.* 2012) and significant peaks were called using the MACS algorithm (Zhang *et al.* 2008). See Chapter V for further details. The ChIP-seq data *were deposited* in the *GEO database* under accession number GSE45186.

2.4.6 Bisulfite-Sequencing and PyroMark Assays on Promoter DNA

Two μg of genomic DNA were treated with EpiTech Bisulfite Kit from Qiagen, Inc. Treated DNA was amplified using primers designed by Methyl Primer Express v 1.0 (Applied Biosystems). PCR products were cloned using the TopoTA Cloning® Kit (Life Technologies) and sequenced by Sanger sequencing. Methylation percentages at promoter CpG sites were also quantified using the Pyromark assay on a PSQ 96MA Pyrosequencer (Qiagen, CA). See Chapter V for further details.

CHAPTER III

CHIP-SEQ IDENTIFIES THE FIRST MARSUPIAL-SPECIFIC IMRPINTED GENE

3.1 Introduction

Genomic imprinting, generally, is an epigenetic phenomenon whereby certain genes show parent-of-origin-specific allelic expression. It affects a minority of genes in the genomes of therian mammals (eutherians and metatherians) but has not been detected in prototherians, birds, or other vertebrates. In human and mouse, 79 and 123 imprinted genes have been characterized, respectively, with only ~60% of these genes sharing imprinted status (Morison *et al*. 2005). In addition to interspecific differences, imprinted expression of some loci has been shown to vary between cell types, tissues, developmental stages, and gene isoforms; and in some cases, 'leaky' expression of the repressed allele has been observed especially in placenta (Kosaki *et al*. 2000; Umlauf *et al*. 2004; Dindot *et al*. 2008; Wang *et al*. 2011; Bebbere *et al*. 2013; Wang *et al*. 2013). These variable characteristics compound the difficulty of finding and describing imprinted genes, reveal the magnitude of variation present among suites of imprinted genes between species, and underscore the dynamic expression patterns of imprinted genes within an individual.

In metatherian (a.k.a. marsupial) mammals, genomic imprinting has been examined primarily in the tammar wallaby (*Macropus eugenii*: Australian family Macropodidae), gray short-tailed opossum, (*Monodelphis domestica*: American family Didelphidae), and Virginia opossum (*Didelphis virginiana*: Didelphidae), wherein only

19 genes, each already known to be imprinted in human and/or mouse, have been scrutinized in one or another of these species with regard to parent-of-origin specific allele expression. Eight of these 19 loci have been shown to be imprinted in at least one of these marsupial species; nine show biallelic expression; and two have no marsupial homolog (Renfree *et al*. 2008; Samollow 2008; Das *et al*. 2012; Stringer *et al*. 2012). Of the eight marsupial imprinted genes, only *IGF2* and *H19* are located in an imprinted cluster and associated with an imprinting control region (ICR), both of which are hallmarks of imprinted loci in eutherian mammals (Smits *et al*. 2008; Barlow 2011). The remaining six marsupial imprinted genes are individually imprinted, associated with no known clusters, and mechanisms that regulate their expression remain unknown.

Beyond interspecific comparative analyses to infer the evolutionary origins and adaptive significance of imprinted genes, the process of genomic imprinting is, *per se*, an invaluable model system for studying the epigenetic regulation of genes generally. For example, interspecific comparisons of imprinted and non-imprinted orthologs have led to the identification of certain structural features, such as SINEs and LINEs and their *cis*-acting epigenetic elements, that can affect the imprintability of a gene (Greally 2002; Murphy & Jirtle 2003; Samollow 2006). Further, the identification of differential DNA methylation between the two parental alleles at imprinted loci in eutherians has not only provided insight concerning the epigenetic regulation of these loci, but has also led to the development of a paradigm for studying *cis*-acting mechanisms of gene regulation at non-imprinted loci as well (Ferguson-Smith 2011). Finally, the interaction of genomic elements and epigenetic modifications at imprinted loci has revealed links between

epigenetic states, chromatin structure, and transcriptional activity. A comprehensive catalogue of imprinted loci across a broader range of therians, including eutherian and marsupial species alike, with descriptions of the molecular mechanisms that establish and maintain the imprinted state, can illuminate the evolutionary history and mechanisms of genomic imprinting generally and perhaps reveal heretofore unrecognized selective pressures that act on a gene to target it for imprinted expression.

Various epigenetic marks have been associated with imprinted genes and ICRs in eutherians, most notably cytosine methylation and histone modifications. Differential methylation of cytosine residues at CpG dinucleotides within CpG islands has been found at both ICRs and promoter regions of imprinted genes and occurs in a parent-of-origin-allele-specific manner (Reik & Walter 2001). Some of these parent-of-origin-specific differentially methylated regions (DMRs) are established in the germ-line and maintained throughout all developmental stages and tissues, whereas other DMRs arise after fertilization and occur in tissue-specific or developmental-stage-specific patterns (Feil *et al*. 1994; Shemer *et al*. 1997). Furthermore, the loss of DNA methylation at the promoter region or ICR of an imprinted gene or imprinted gene cluster leads to the loss of the imprinted state, resulting in biallelic expression (Li *et al*. 1992; Wutz & Barlow 1998; Anwar *et al*. 2012).

Differential histone modification states have also been associated with ICRs and promoter regions of imprinted genes. Transcriptionally repressive modifications such as trimethylation of lysine 9 of histone 3 (H3K9me3) and trimethylation of lysine 27 of histone 3 (H3K27me3) are present at the ICRs and/or promoters of the repressed allele,

whereas transcriptionally active marks such as H3K4me3 and H3K9Ac are present at the ICRs and promoters of the actively expressed allele (Yang *et al*. 2003; Mikkelsen *et al*. 2007a; Dindot *et al*. 2009). Along with DNA methylation, these histone modifications create open or closed chromatin states, which can alter the accessibility of DNA to transcriptional machinery, thereby affecting transcription rates. In addition, certain of these transcriptionally opposing histone modifications have been shown to be mutually exclusive at identical sites in the promoter regions of active versus repressed alleles at imprinted loci (e.g. H3K4me3 vs. H3K9me3; H3K9Ac vs. H3K9me3) suggesting a potentially powerful approach for seeking candidate-imprinted loci, independent of expression-based SNPs (Regha *et al*. 2007; Samollow 2008; Dindot *et al*. 2009).

Previous searches for imprinted genes in marsupials have focused on a small number of loci that are already known to be imprinted in eutherians, and only a few of these have sought to describe DNA methylation and histone modification profiles of CpG islands at putative promoter regions (Killian *et al*. 2000; Evans *et al*. 2005; Suzuki *et al*. 2005; Rapkins *et al*. 2006; Weidman *et al*. 2006b; Ager *et al*. 2007; Suzuki *et al*. 2007; Lawton *et al*. 2008; Smits *et al*. 2008; Das *et al*. 2012; Stringer *et al*. 2012). Clear evidence of differential DNA methylation at marsupial imprinted genes has been found at only two loci, and only the DMR at the *IGF2-H19* imprinting cluster has been shown to regulate transcription of a marsupial imprinted gene (Lawton *et al*. 2008; Smits *et al*. 2008). In addition, the marsupial X chromosome, which exhibits paternal imprinting for most loci in females, is strongly deficient in CpG island methylation (Loebel & Johnston 1996; Rens *et al*. 2010).

Data addressing histone modifications at promoters and CpG islands of marsupial imprinted genes are extremely limited, with only two histone modifications, H3K3me2 and H3K9me3, examined for opossum *Igf2r*, *Htr2A*, and *L3mbtl*. These genes exhibit enrichment of H3K4me2 but not H3K9me3 at promoters (Das *et al*. 2012). Fluorescence *in-situ* hybridization analyses of wallaby and brush-tailed possum (*Trichosurus vulpecula*: Australasian family Phalageridae) X chromosomes using antibodies to specific histone modifications showed a whole-chromosome level correlation between repressive and activating histone marks on the inactive and active X chromosomes, respectively, consistent with a possible role for histone modification states in the transcriptional regulation of genes on the marsupial X (Wakefield *et al*. 1997; Koina *et al*. 2009; Rens *et al*. 2010). Overall, however, understanding of the role of epigenetic processes in the regulation of gene activity in marsupial species has been hindered by a lack of genome-wide assessments of marsupial genomes by allele-specific expression and histone modification profiling.

Taking advantage of continuously improving next-generation sequencing technologies and the high-quality draft assembly of the *M. domestica* genome, we are now able to search for marsupial-specific imprinted genes and analyze fundamental signals of imprinting on a genome-wide basis. To accomplish this, we conducted reciprocal crosses of animals from two *M. domestica* stocks and used ChIP-seq to perform the first *ab initio* search for putative gene promoters that are concurrently marked by mutually-exclusive, transcriptionally opposing histone modifications as a means to identify candidate-imprinted genes.

3.2 Methods

3.2.1 Animals and Tissue Collection

For the ChIP-seq experiments, animals from two laboratory stocks (LL1 and LL2) of the opossum, *M. domestica* were utilized (VandeBerg & Williams-Blangero 2010). For initial ChIP-seq profiling, primary fibroblasts were cultured from ear pinna of a male F_1 (ID# A0514) from an LL1 X LL2 mating and collected using standard methods (Figure A8). For further experiments, reciprocal crosses were conducted between LL1 and LL2 stocks, and primary fibroblast lines were established from ear pinnae collected from the parents in each cross, and from four F_1 (LL1 X LL2) and four F_1 (LL2 X LL1) individuals using standard methods (Figure A8B). All procedures involving opossums were approved by the Texas A&M University, College Station, Institutional Animal Care and Use Committee (TAMU Animal Use Protocols 2011-141 and 2011-191).

3.2.2 Chromatin Immunoprecipitation (ChIP) and ChIP-Seq

Native-ChIP (N-ChIP) was conducted on low (< 5) passage, primary fibroblasts from male A0514 using a method modified from Dindot *et al*. (2009). Harvested fibroblast cells were washed in PBS and homogenized in 500 μL of Buffer I (0.3 M sucrose, 60 mM KCl, 15 mM NaCl, 5 mM MgCl2, 0.1 mM EDTA, 15 mM Tris, 0.5 mM DTT, 0.1 mM PMSF). The sample was centrifuged for 5 min. at 3000 X g, the supernatant was removed, and the pellet was re-suspended in 200 μL of Buffer I. Cells were lysed on ice for 5 minutes by adding 200 μL of Buffer II (Buffer I + 4 μL of NP40), and nuclei were isolated by centrifugation for 20 minutes at 10,000 X g through

1.5 mL of Buffer III (1.2 M sucrose, 60 mM KCl, 15 mM NaCl, 5 mM $MgCl_2$, 0.1 mM EDTA, 15 mM Tris, 0.5 mM DTT, 0.1 mM PMSF). The nuclei-enriched pellet was washed with Buffer I, centrifuged, and re-suspended in 350 μL of micrococcal nuclease digestion buffer (0.32 M sucrose, 4 mM $MgCl_2$, 50 mM Tris, 0.1 mM PMSF). Chromatin was digested using 10 units of micrococcal nuclease (Sigma, N5386) for 10 minutes at 37°C. The reaction was stopped using 50 μL of 0.5 M EDTA.

For an input control, 100 μL of digested chromatin was removed before treatment with antibodies and the DNA fraction was extracted. For ChIP, 4.0 μg of digested chromatin was incubated at 4°C overnight with one of the following antibodies: anti-H3K4me3 (Millipore #07-473), anti-H3K9Ac (Millipore #CS200583), anti-H3K9me3 (Millipore #07-442), anti-H3K27me3 (Millipore #07-449), or non-specific, rabbit IgG (Millipore #12-370). Antibody-bound chromatin was isolated using Dynabeads® Protein A (Invitrogen), washed, and eluted according to manufacturer's specifications. N-ChIP and input DNA were purified using Qiagen MiniElute Spin Columns (Qiagen, CA) and enrichment was verified using real-time PCR (data not shown). Non-indexed Illumina libraries were constructed at Global Biologics, LLC (Columbia, MO) and sequenced on an Illumina GAIIx at the University of Missouri – Columbia DNA Core Facility using 51- or 101-base chemistry. Image analysis and base calling were performed using Illumina software.

3.2.3 ChIP-Seq Analysis

Raw sequence reads were filtered for quality and mapped to the MonDom5 genome assembly using Bowtie in the Galaxy suite (Giardine *et al*. 2005; Blankenberg *et*

al. 2010; Goecks *et al*. 2010). A seed length of 28 bases was used with a maximum of 2 mismatches permitted between the seed and reference genome, and only the best alignment reported for each read. Significant peaks of enrichment were identified for each histone modification using Model-based Analysis for ChIP-seq (MACS) using the input control option (Zhang *et al*. 2008). The ChIP-seq data were deposited in the GEO database under accession number GSE47723. Ensembl gene models (release 64) were used and annotated CpG island coordinates were obtained from the UCSC genome browser (Kent *et al*. 2002). Putative promoters were defined as regions 5,000 bases upstream to 500 bases downstream of annotated transcription start sites (TSSs). Overlaps (minimum of one base pair) between features were assessed using scripts in the BEDTools package (Quinlan & Hall 2010). In order to be considered a candidate-imprinted gene, the putative promoter of a gene had to be concurrently marked by significant H3K4me3, H3K9Ac, and H3K9me3 peaks, and contain an annotated CpG island.

3.2.4 SNP Discovery in Candidate-Imprinted Genes

PCR primers were designed using Primer3 (Rozen & Skaletsky 2000) to amplify 600-700 bases of the putative 3'-untranslated region (UTR) of each candidate-imprinted gene as well as *Igf2r* (Table B9). Genomic DNA (gDNA) was extracted from livers of the eight individuals comprising the P generations of each cross using standard protocols and was PCR amplified (20 μL reaction volume) for each primer set using AmpliTaq Gold polymerase (Invitrogen). After an initial denaturation of 5 minutes at 95°C, 38 PCR cycles were conducted at 95°C for 30 seconds, 54°C for 30 seconds, 72°C for 30

seconds followed by a final extension for 7 minutes at 72°C. PCR optimization was conducted where necessary. To confirm PCR amplification, 3 μL of PCR product was run and visualized on a 1% agarose gel (data not shown). All PCR products for each of the eight parents were pooled, eight indexed Illumina libraries were created from each pool, and 101 bases were sequenced on an Illumina GAIIx at the University of Missouri – Columbia DNA Core Facility. Raw reads were filtered for quality, mapped to the MonDom5 genome assembly, and SNPs were called using MPileup in the SAMTools package (Li *et al*. 2009). Variant regions were required to have a minimum of 20X coverage to be considered as candidate SNPs.

3.2.5 Verification of Imprinting Status

Total RNA and gDNA were extracted from six of the eight fibroblast cell lines from the F_1 generation using standard protocols (two F_1 animals, A0703 and A0716, are absent from the SNP analysis due to the lack of success in establishing fibroblast lines from these animals). Total RNA was treated with DNase I and converted to cDNA using the SMARTer cDNA Synthesis Kit (Clontech). PCR reactions were conducted as previously described, and gDNA and cDNA PCR products were sequenced on an ABI 3730XL at Beckman-Coulter Genomics, Inc. (Danvers, MA). Sequences were viewed in Sequencher4.10™ .

To quantify maternal/paternal allele expression ratios, pyrosequencing PCR was conducted on cDNA from one F_1 male and one F_1 female from each of the LL1 X LL2 and LL2 X LL1 crosses. Pyrosequencing PCR and sequencing primers were designed using the PyroMark Assay Design Software Version 2.0.1.15 (Qiagen, CA).

Pyrosequencing PCR amplification was carried out in a 40 μl system using Ampli-Taq Gold polymerase under the following cycling conditions: 1 cycle of 95° C for 5 min; 45 cycles of 95° C for 45 sec, 57° C for 30 sec, and 72° C for 20 sec; followed by 1 cycle of 72° C for 10 min. PCR products were prepared according to the manufacturer's protocol and loaded on the PSQ 96MA Pyrosequencer with PyroMark Gold Reagents (Qiagen, CA) using the Allele Quantification method (AQ). Two technical replicates were done for each gene in each sample. Overall, variation between replicates was negligible (≤ 3%), and the final expression percentages were determined by averaging the results from each run.

3.2.6 Analysis of CpG Island Methylation

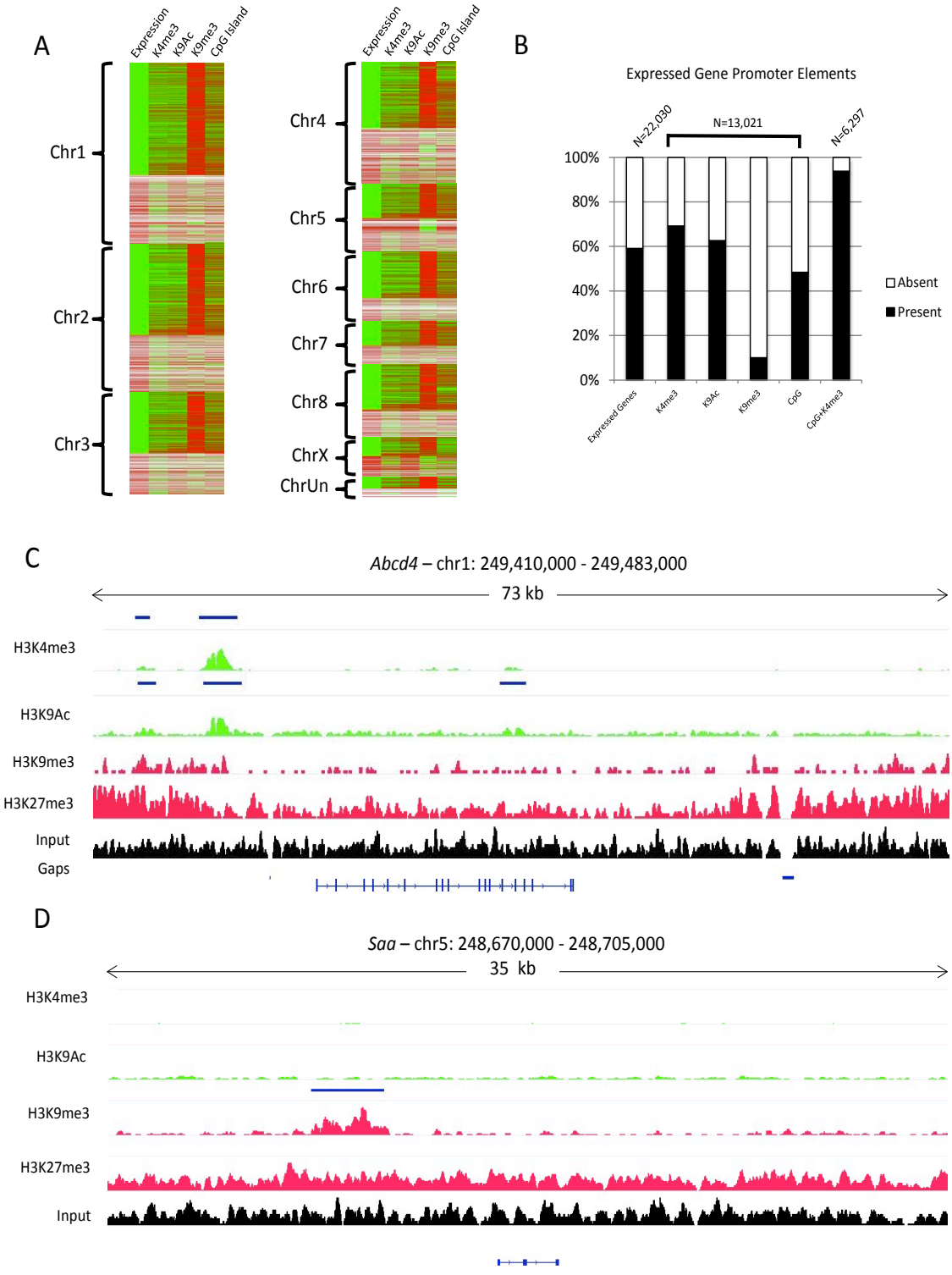
To assess the methylation status of promoter CpG islands, gDNA was isolated from fibroblasts from two F_1 animals from each reciprocal cross (4 total animals) and treated with sodium bisulfite to convert unmethlyated cytosines to uracils using the Qiagen EpiTect Bisulfite Kit (Qiagen, Inc). PCR primers were designed to amplify bisulfite converted DNA using Methyl Primer Express Software (Applied Biosystems, Inc). BS-PCR products were gel purified, sub-cloned using the TOPO TA Cloning Kit (Invitrogen), and blue/white screened using XGal (40 mg/mL). For each cloned PCR product, plasmids were purified from at least 16 positive white colonies and were sequenced at Beckman Coulter Genomics by the Sanger dideoxy-chain termination method using the M13 forward primer. Sequences were inspected and analyzed using Sequencher4.10™ .

3.3 Results

3.3.1 ChIP-Seq Results

More than 436 million Illumina reads from male A0514 fibroblasts were uniquely mapped to the current *M. domestica* genome assembly (MonDom5). The two marks of activation (MOAs) examined, H3K4me3 and H3K9Ac, gave 79,412 and 52,511 unique peaks of enrichment, respectively (MACS, $p \leq 10^{-5}$). The two marks of repression (MORs) examined, H3K9me3 and H3K27me3, gave 56,719 and 16,592 unique peaks of enrichment, respectively (MACS, $p \leq 10^{-5}$) (Table B7). We next analyzed the overlap of each histone modification with promoters of annotated genes and CpG islands. Of the 22,030 annotated genes in MonDom5, 13,021 showed expression in at least one of four male-fibroblast cell lines as determined by RNA-seq (data not shown), and 9,012 (69%) were marked by H3K4me3 (Figure 8A,B). About half of these expressed genes have an annotated CpG island at the promoter and 93% of these CpG islands were marked with H3K4me3 regardless of transcriptional state (Figure 8B). Thus, the promoters of the transcribed genes (e.g. *Abcd4*, Figure 8C) showed enrichment for both MOAs and were deficient for MORs, whereas the promoters of repressed genes (e.g. *Saa* Figure 8D) showed a deficiency in MOAs and, in some cases, an enrichment of H3K9me3. The distribution of H3K27me3 was diffuse across the genome with most significant peaks occurring in intergenic regions. Promoters and gene bodies of biallelically expressed genes and known imprinted genes showed a depletion of H3K27me3. In addition, H3K27me3 has not been shown in other

Figure 8. Summary of fibroblast ChIP-seq results. A) Heatmap of RNA expression, H3K4me3, H3K9Ac, H3K9me3, and CpG Islands. Expression was determined using RNA-seq data from 4 male fibroblast cell lines. Green = expressed/presence of element, Red = not expressed/absence of element. White = not expressed and no elements. B) Graph of percentage of genes that are expressed, percentage of promoters of expressed genes with indicated histone modifications or CpG island, and percentage of CpG islands marked by H3K4me3. K4 = H3K4me3; K9Ac = H3K9Ac; K9me3 = H3K9me3, C) and D) Histone profiles across an expressed gene (*Abcd4*) and a repressed gene (*Saa*). Blue bars = significant peak of enrichment; green panels = marks of activation; red panels = marks of repression; black panel = input. Gaps (if present) and gene annotations are shown in the bottom panels.


A
Expression
K4me3
K9Ac
K9me3
CpG Island
Chr1
Chr2
Chr3
Chr4
Chr5
Chr6
Chr7
Chr8
ChrX
ChrUn
B
Expressed Gene Promoter Elements
N=22,030
N=13,021
N=6,297
100%
80%
60%
40%
20%
0%
Absent
Present
Expressed Genes
K4me3
K9Ac
K9me3
CpG
CpG+K4me3
C
Abcd4 – chr1: 249,410,000 - 249,483,000
73 kb
H3K4me3
H3K9Ac
H3K9me3
H3K27me3
Input
Gaps
D
Saa – chr5: 248,670,000 - 248,705,000
35 kb
H3K4me3
H3K9Ac
H3K9me3
H3K27me3
Input

species to be mutually exclusive with the MOAs used in this study. For these reasons H3K27me3 was excluded from further examination.

In addition to the promoters discussed above, we examined overlap of the various histone modifications with each other and all annotated-putative promoters (35,105) in the MonDom5 assembly. Of the H3K9Ac peaks, 47,275 (90%) overlapped with an H3K4me3 peak by at least one base pair, and 6,410 (11%) H3K9me3 peaks overlapped with an H3K4me3 peak (Figure 9A, Table B8). Additionally, 11,580 (52%) promoter-associated CpG islands were marked by a significant H3K4me3 peak.

Of the 35,105 putative promoters, 16,620 (~46%) were marked with H3K4me3, 7,871 also have an annotated CpG island, and 178 of them were also concurrently marked with H3K9Ac and H3K9me3 (Figure 9B). No X-linked genes met these criteria. This is noteworthy because animal A0514 was a male and possessed only a single X chromosome; patterns reflective of imprinted expression of X-linked genes would have signaled false positive outcomes. That none was observed provides an internal control indicating the accuracy of ChIP-seq procedures in this study. These 178 autosomal

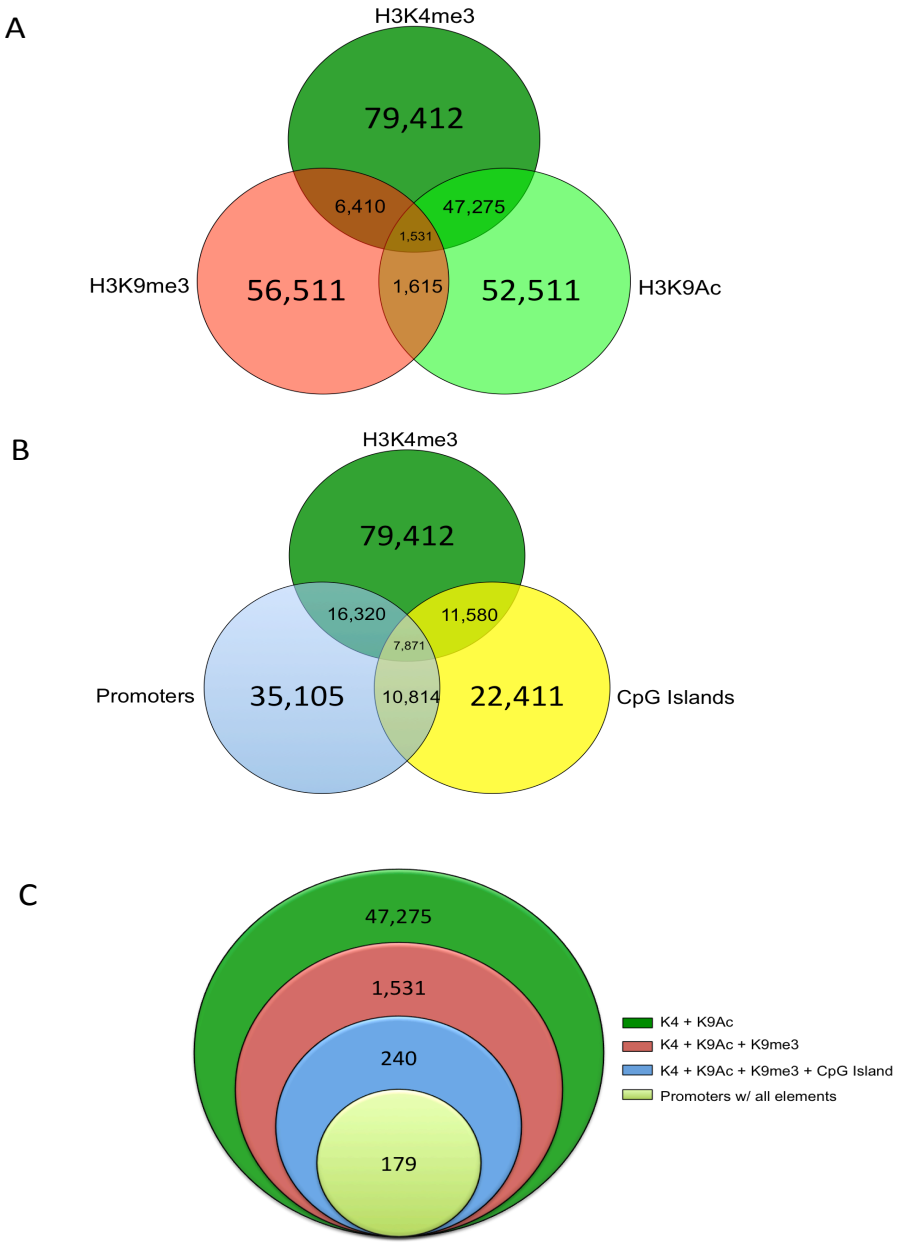


Figure 9. Venn diagrams representing overlaps of significant histone peaks, annotated CpG islands, and putative promoters. A) Significant peaks for H3K4me3 (dark green), H3K9Ac (light green) and H3K9me3 (red). B) Overlaps of H3K4me3 (green), annotated CpG islands (yellow), and putative promoters (blue). C) Overlaps of H3K4me3+H3K9Ac peaks (green), all three histone modification peaks (red), all three histone modification peaks and CpG islands (blue), and all elements and putative promoters (light green).

genes with putative promoters marked by two MOAs, one MOR (H3K9me3), and a CpG island were considered candidate-imprinted genes and targeted for SNP discovery along with *Igf2r* (Figure 9C). *Igf2r* is known to be imprinted in *M. domestica*, and has a promoter CpG island, but it did not show overlapping enrichment of MOAs and H3K9me3. The histone modification states of the remaining annotated opossum imprinted genes, *Htr2A*, *L3mbtl*, and *Mest,* were also examined and showed the presence of MOAs at their promoters but lacked overlapping MOAs and H3K9me3 (Addition File 1: Figure S5). However, informative SNPs for these genes were not present in our crosses, precluding our assessment of their imprinted/non-imprinted states. Furthermore, the *Igf2-H19* imprinted cluster is not present in the current MonDom5 assembly and consequently was not accessible for this study.

3.3.2 SNP Search for Candidate-Imprinted Genes

Primers were designed to target the 3'-UTR for each of the 178 candidate-imprinted genes and *Igf2r*, allowing for amplification of both gDNA and cDNA with the same primer sets (Table B9) The primer panel was run on liver DNA from the eight animals in the P generation to search for 'trackable' parent-specific SNPs between the reciprocal crosses. Of the 179 genes tested, 38 - 49 genes, depending on the cross, showed at least one such SNP in individual crosses (Tables B10 - B13). We selected 30 genes that had a trackable SNP in at least one family in each reciprocal cross, and 21 of these showed specific 3'-UTR amplification in cDNA of the F_1 generation. The PCR products from gDNA and cDNA of these genes were Sanger sequenced to qualitatively

assess monoallelic vs. biallelic gene expression, and 17 of them gave high quality sequences from both templates (Tables B10 - B13).

3.3.3 *Meis1* is Paternally Imprinted in *M. domestica* Fibroblasts

Three annotated genes with promoters concurrently marked by the two MOAs and H3K9me3 (MOR) were clearly heterozygous for SNP variants in gDNA and showed strong allele-biased expression of alleles in cDNA: *Meis1* (ENSMODG00000003396), *Cstb* (ENSMODG00000021035), and *Rpl17* (ENSMODG00000011184) (Figure 10A-C; Table 1, Figure A9). However, only *Meis1* showed parent-of-origin specific expression (Figure A10). Quantification of allele-specific expression by pyrosequencing showed 92% and 77% expression of the maternal allele for *Meis1* for one animal from each reciprocal cross (A0695 and A0727 respectively) (Figure 10C). Four additional F_1 animals were examined for monoallelic expression, and informative animals showed strong maternal-allele biased expression with an average of 82% of transcripts originating from the maternal allele. Both *Cstb* and *Rpl17* also exhibited mono-allelic expression; however, the pattern was not parent-of-origin-specific, but rather appeared to be allele-specific (Figure A10; Table B14). *Igf2r* also showed monoallelic expression in the one F_1 animal (A0695) that was heterozygous for a trackable SNP (Figure A10).

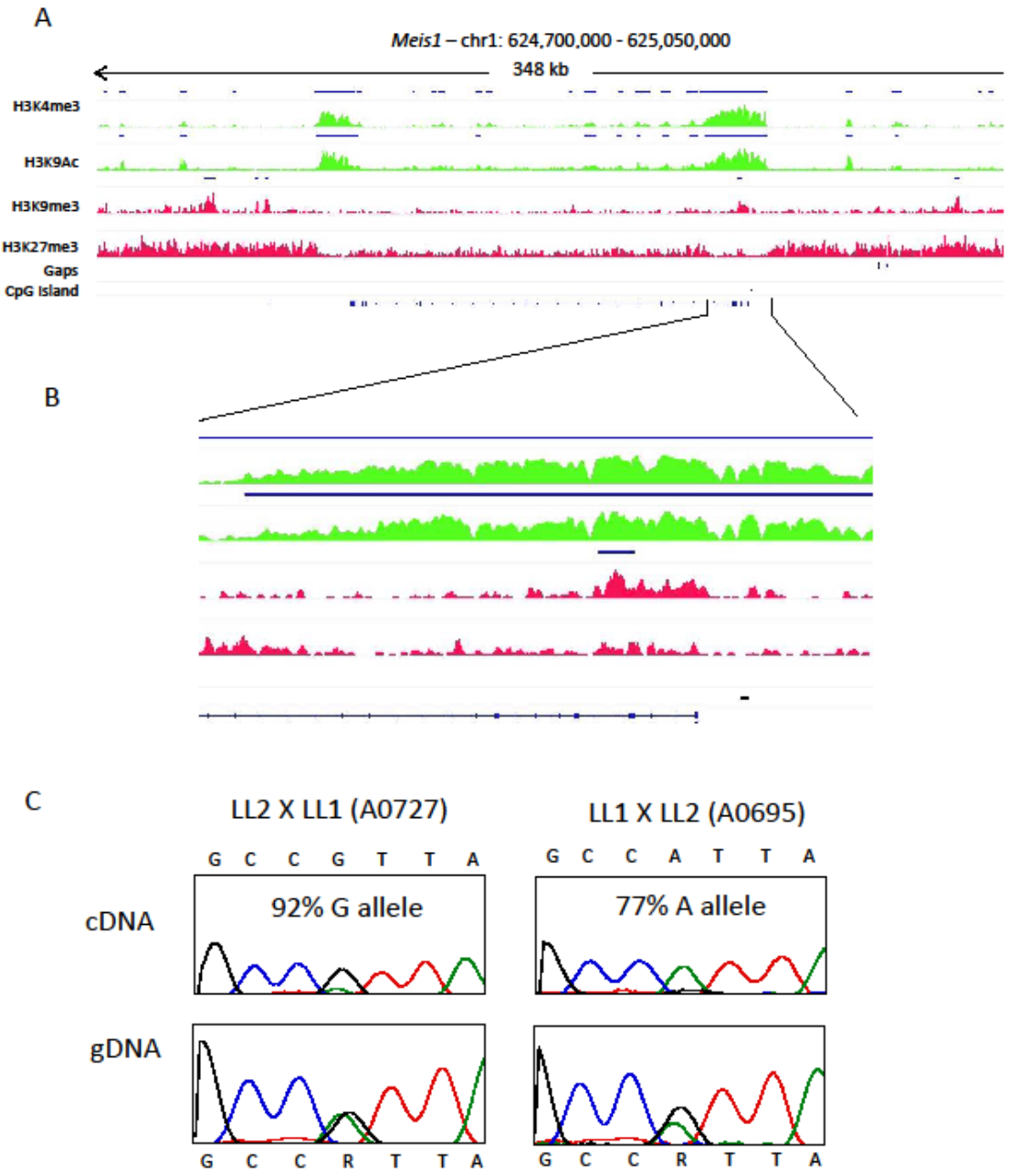


Figure 10. ***Meis1*** **is maternally imprinted in** ***M. domestica*** **fibroblasts. A) Histone profiles of** ***Meis1*****. Two MOAs (two green panels), two MORs (two red panels), assembly gaps, annotated promoter CpG island, and the gene annotation. Significant histone peaks (MACS, $p \leq 10^{-5}$) are indicated by blue bars above the peaks. B) Enlargement of the promoter region of** ***Meis1.*** **C) Genomic DNA and cDNA genotypes of one animal from each reciprocal cross. Percent of maternal allele contribution as determined by pyrosequencing is shown in each box.**

3.3.4 Methylation States of Promoters

We next assayed cystosine methylation at promoter CpG islands of the four monoallelically expressed genes. For *Meis1*, we assayed 16 CpG dinucleotides across the promoter and found a hypomethylated state with no evidence of differential methylation (Figure 11A). The promoters of *Cstb, Rpl17,* and *Igf2r* were also hypomethylated with no differential methylation (for primers see Table B15). Recently, Das *et al*. (2012) discovered a differentially methylated CpG island in intron 11 of *Igf2r* in the liver, brain, and kidney of *M. domestica*. We assayed 18 CpG dinucleotides across this same CpG island and found this DMR in fibroblasts as well (Figure A11). However, we were unable to assess allele-specific methylation patterns, as a parent-of-origin specific SNP was not present in this region in our animals. The hypomethylated states of the promoters of *Meis1*, *Cstb, Rpl17*, and *Igf2r,* as well as the DMR in intron 11 of *Igf2r,* were also verified in three other F_1 animals: A0690 (female), A0716 (male), and A0727 (male).

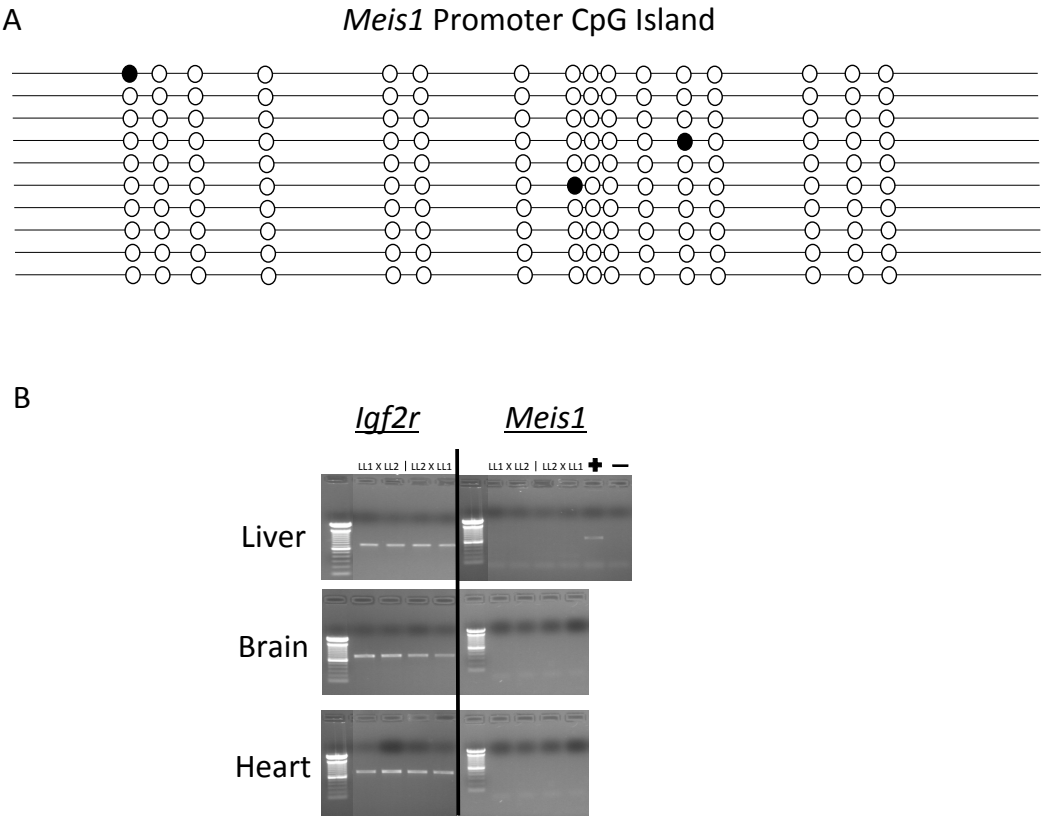


Figure 11. DNA methylation profiles of *Meis1* and tissue-specific expression pattern. A) Bisulfite converted DNA was cloned and sequenced. Each line represents an individual PCR product. The annotated promoter CpG island was assayed for DNA methylation. Unfilled circles represent unmethylated cytosines at CpG dinucleotides and filled circles represent methylated cytosines. B) 1% agarose gel of PCR amplicons (cDNA) generated from liver, brain, and heart mRNA. *Igf2R* is expressed in all three tissues, but *Meis1* is not expressed. Ladder = 100bp DNA Ladder; Positive control = fibroblast cDNA.

3.4 Discussion

Of the 35,105 putative promoters assayed in our analysis of *M. domestica* fibroblasts, only ~46% (16,320) were marked by H3K4me3. This fraction is considerably smaller than the 74% and 71% of the promoters marked by this expression-associated modification in cultured human and mouse cells (Guenther *et al*. 2007;

Mikkelsen *et al*. 2007a), respectively, and is most likely an artifact of inaccuracy in the annotation of the *M. domestica* gene set. The initial set of predicted protein-coding and non-coding genes was produced by analyzing similarity with well-annotated eutherian gene sets, a practice that is expected to underrepresent or overlook diverged orthologs, paralogs, and marsupial-specific genes (Goodstadt *et al*. 2007; Mikkelsen *et al*. 2007b). Further annotation has relied on individual sequencing of genes-of-interest, as well as a small number of RNA-seq data sets that are enriched for the 3' ends of genes, leaving the 5' annotation of many genes incomplete or inaccurate. This issue is underscored by a recent, comprehensive RNA-seq study of the *M. domestica* X chromosome that reveals that the 5' ends of nearly half of the genes on the X chromosome are incorrectly annotated in the MonDom5 assembly, with ~30% having a transcription start site more than 5 kb upstream from the first annotated 5' exon (X Wang, KC Douglas, AG Clark, PB Samollow, unpublished data). Annotation issues of this kind, especially at the 5' ends of genes, pose a significant challenge for correlating promoter histone modification states with transcriptional states. In light of this limitation, our results likely underestimate the number of opossum promoters marked, either independently or concurrently, by MOAs and/or H3K9me3.

We were, however, able to identify 178 genes that were concurrently marked by MOAs and H3K9me3 within 5 kb of an annotated 5' exon. Twenty-one of these were expressed in fibroblasts and had an informative SNP in each reciprocal cross. Importantly, only six of them showed 100% overlap of significant peaks of H3K4me3, H3K9Ac, and H3K9me3, and half of these exhibited strongly biased allele expression, of

which one, *Meis1*, was expressed in a parent-of-origin specific manner; i.e., is imprinted. The high frequency of monoallelic expression among genes with 100% overlap of transcriptionally opposing histone marks suggests that complete peak overlap be adopted as an essential criterion in future *ab initio* searches for imprinted genes in non-eutherian species.

Expression of the MEIS family of genes in eutherians is often strongly developmental-stage and cell-type specific and, accordingly, we were unable to detect *Meis1* transcripts in several adult opossum somatic tissues examined (Figure 4B). The remaining two monoallelically expressed genes, *Cstb* and *Rpl17*, showed allele-biased expression that was independent of parent-of-origin, emphasizing the importance of conducting reciprocal crosses to detect genuine parent-of-origin specific expression patterns, a practice that has been absent from many past studies of marsupial imprinted genes.

Assessment of the transcriptional state of these three monoallelically expressed genes reveals the first case of an imprinted gene in a marsupial that is not known to be imprinted in any other organism, and suggests a role for histone modification states in the occurrence of monoallelic-expression of genes in the opossum and perhaps other marsupial genomes. Contrastingly, methylation analysis of DNA from these fibroblasts failed to find evidence of DMRs at annotated CpG islands in the promoter regions of this novel imprinted gene or either of the two non-imprinted monoallelically expressed genes. This is consistent with past reports that DMRs are rare or absent from marsupial orthologs of eutherian imprinted genes.

Examination of the four previously known annotated opossum imprinted genes, *Igf2r*, *Htr2A*, *L3mbtl*, and *Mest* failed to detect transcriptionally opposing histone modifications at their respective promoters or their gene bodies. *Igf2r* is not imprinted in humans but is imprinted in mouse, sheep, dog, and marsupials (wallaby and opossums). In mouse, the transcriptional regulation of *Igf2r* is controlled by a DMR in intron 2 and by an antisense transcript (*Air*) (Wutz *et al*. 1997; Sleutels *et al*. 2002). Interestingly, the DMR at intron 2 is present in human, mouse, and sheep, but absent in dog and marsupials (Wutz *et al*. 1997; Sleutels *et al*. 2003). Transcriptionally opposing histone states have been associated with the imprinted state, or lack thereof, in human and mouse; but the full-length *Air* antisense transcript has only been described in mouse (Vu *et al*. 2006; O'Sullivan *et al*. 2007). *Htr2A* and *L3mbtl* show variation of imprinted status in human organs sampled and are associated with certain disease states that correlate with aberrant DMRs, but no studies of associated histone states have been reported for these loci (Kobayashi *et al*. 1997; Li *et al*. 2004; De Luca *et al*. 2007).

We were able to assess the imprinting status at the *Igf2r* locus, but a lack of suitable SNP variants in our animals prevented us from analyzing expression patterns of *Htr2A*, *L3mbtl*, and *Mest* in the present study. It is possible that these genes are not imprinted in opossum fibroblasts, in which case the absence of transcriptionally opposing histone modifications would be expected. Alternatively, any or all of these three genes could be imprinted in opossum fibroblasts but not marked or regulated by the specific histone modifications we examined, or DMRs, but rather by some yet-to-be identified genomic elements or regulatory mechanisms such as non-coding RNA

transcripts. If so, there could be additional imprinted loci in fibroblasts that went undetected by our strategy relying on only four histone marks.

Although *Meis1* showed parent-of-origin-specific allele expression in three individual fibroblast cell lines, there was 'leaky' expression of the paternal allele in some samples. Leaky expression of the repressed allele has been observed for some imprinted genes in eutherians and for some paternally imprinted X-linked genes in marsupials (Samollow *et al*. 1987; Umlauf *et al*. 2004; Wang *et al*. 2008; Suzuki *et al*. 2009; Wang *et al*. 2011). At the *G6pd* locus, the degree of paternal allele leakiness is age-dependent, with adults showing greater levels of paternal leakage than embryos (Samollow *et al*. 1995). Similarly, studies in eutherians have demonstrated a loss of allele-specific gene regulation for X-linked genes in a passage-number-dependent manner in primary cell lines (Migeon *et al*. 1985). Although we used low passage fibroblast cell lines, the cells were originally grown from adult tissue, and the combination of adult source and increasing passage could have resulted in higher levels of leakiness. Alternatively, it is possible that the epigenetic regulation of imprinted loci in marsupials is not as stable as in eutherians due to the apparent lack of differential DNA methylation at these loci. Furthermore, most studies of marsupial imprinted gene expression have not utilized highly sensitive assays, such as pyrosequencing, to measure allele-specific expression of imprinted genes; so leaky expression of the repressed allele could more prevalent than previously believed.

The MEIS gene family comprises three homeobox genes in humans (*MEIS1, MEIS2, MEIS3*), which have been implicated as cofactors of Hox proteins and are

heavily involved in development, cell proliferation, and disease states. They have been shown to be essential for organ development in mouse, zebrafish, chicken, and *Drosophila* (Moens & Selleri 2006; Choe *et al*. 2009; Sanchez-Guardado *et al*. 2011; Carbe *et al*. 2012; Mahmoud *et al*. 2013). In the absence of protein functional data, we are unable to determine definitively which *MEIS* gene ortholog is represented by the imprinted opossum *Meis* locus. Nevertheless, this locus was matched by the Illumina reads to the exclusion of other *Meis* paralogs and comparative synteny analysis shows gene content and order of the genomic region flanking this locus is strongly conserved with that containing the human *MEIS1* and mouse *Meis1* genes. Finally, a reciprocal Blast search strategy using the opossum predicted mRNA and amino acid sequences from the Ensembl annotation indicates that the opossum-imprinted *Meis* gene shares the greatest sequence similarity with the *MEIS1* gene of human, mouse, and rat. Hence, we feel confident that this is the ortholog of *MEIS1*. This gene has been shown to regulate cell proliferation, growth, and differentiation in embryos and fetuses, and in adult animals is active in highly proliferative tissues (Unnisa *et al*. 2012; Mahmoud *et al*. 2013). These functional characterisitcs and the paternally imprinted state of opossum *Meis1* are consistent with the conflict model for the evolution of genomic imprinting, which is based on competing fitness interests of the paternal and maternal genomes with regard to maternal resource allocation to the developing offspring (Moore & Haig 1991; Haig 2004).

3.5 Conclusion

In this first comprehensive report on histone profile states in any marsupial species, we have described the genomic landscapes for four canonical histone modifications, H3K4me3, H3K9Ac, H3K9me3, and H3K27me3 and successfully identified a novel imprinted gene in opossum as well as two monoallelically expressed genes. These results demonstrate the practicality of an *ab initio* strategy for discovering imprinted genes in non-eutherian mammals and, potentially, non-mammalian species as well. Overall, the findings support the conclusion that specific histone modifications are conserved features that mark the promoters of some imprinted genes in all therians, but also suggest that marsupials use multiple epigenetic mechanisms for imprinting, some of which are distinct from those known in eutherians; e.g., DNA methylation appears to play little, if any, role in regulating the imprinted state in marsupial mammals. Furthermore, while the imprinting status of some genes is conserved across therians, identification of a marsupial-specific imprinted locus, *Meis1*, which is not known to be imprinted in any eutherian species examined, bolsters the concept that lineage-specific differences in selective pressures may have led to phylogenetically distinct variants of the imprinting phenomenon.

CHAPTER IV

GENOMIC IMPRPINTING IN FETAL BRAIN AND EXTRA-EMBRYONIC MEMBRANES

4.1 Introduction

In eutherian mammals, imprinted genes have been shown to be enriched in fetal and placental tissues as compared to adult tissues (Kosaki *et al.* 2000; Umlauf *et al.* 2004; Dindot *et al.* 2008; Wang *et al.* 2011; Bebbere *et al.* 2013; Wang *et al.* 2013). Using this knowledge and what we have learned from our studies of XCI and fibroblast imprinting, we applied both RNA-seq and ChIP-seq approaches to detect novel imprinted genes in fetal brain and EEM of *M. domestica*. However, the examination of autosomal loci in tissues from these *M. domestica* crosses poses a greater challenge than the X-chromosome inactivation study discussed in Chapter II.

First, the two stocks, LL1 and LL2, share a considerable amount of genetic material due to historical matings of LL1 and LL2 stocks to achieve sustainable numbers of animals with the LL2 genetic background. Although it has been estimated that LL2 animals share 22% or less of their genetic material with LL1 animals (VandeBerg & Williams-Blangero 2010), the actual percentages remain unmeasured. Our method for determining monoallelic expression at any locus using RNA-seq data is based on the assumption that certain trackable SNPs are stock-specific, homozygous, and fixed in a particular stock. This allows us to assume that the F_1 generation is heterozygous at the gDNA level at loci-of-interest given the genotypes of the expressed sequences of the P

generation. When alleles are shared between stocks, this assumption is violated, thus generating the possibility of concluding that a gene is imprinted when in fact it is not (i.e. a false positive). Furthermore, the risk of a false positive is greater for autosomal loci, with a maximum of 4 alleles/ per gene per mating, than for X-linked loci, with a maximum of 3 alleles per gene per mating.

Secondly, it has been shown in mouse and human that histone state profiles are highly dynamic throughout early development and can vary at individual loci, with undifferentiated cells having many loci that are marked by both active and repressive histone modifications (i.e. bivalent loci). As development proceeds and cells differentiate, bivalent loci resolve to a univalent state, becoming marked by either active or repressive histone modifications that are, in many instances, cell-type specific (Mikkelsen *et al.* 2007a). The early stage of development of the fetal brain, and our limited knowledge of the cell-type specific developmental patterns of *M. domestica*, make the interpretation of correlations of histone modifications with gene expression more challenging than is the case with uniform fibroblast cultures or even fully differentiated adult tissues.

Lastly, assessing the transcriptional and histone states of genes within organs, which by definition comprise multiple cell types, is problematic. We are unable to assess cell-type specific imprinting or histone modifications leaving only genes that are imprinted in all or the vast majority of cells in the tissue. Despite these challenges and limitations, we were able to utilize the fetal brain and EEM RNA-seq data sets and the fetal brain ChIP-seq data sets generated in Chapter II to identify additional candidate-

imprinted genes in opossum fetal brain and EEM. All analyses and methods used can be found in the methods sections of Chapter II and III or are presented in Chapter V.

4.2 Results

4.2.1 Imprinted Genes in Fetal Brain and EEM

Employing a conservative search strategy, we identified 22 candidate-imprinted genes and a previously known imprinted gene, *Ig2r*, all of which exhibited monoallelic expression in fetal brain and/or EEM according to the RNA-seq data (Table 3). This set of candidate-imprinted genes contains known protein coding genes, non-coding RNAs, and genes of unknown biotype. To confirm heterozygosity in the F_1 generation and the parent-of-origin of the alleles, candidate-imprinted genes identified were genotyped by Sanger sequencing of PCR products generated from gDNA of the P and F_1 generations. Sanger sequencing results showed that 14 genes were heterozygous in the F_1 generation and showed trackable parent-of-origin-specific alleles in at least one of the two tissues, enabling us to postulate their imprinted status, paternally (Pat.) or maternally (Mat.) imprinted (Table 3 and Table B16). Two genes, *Praak1* and *Parp4*, were not heterozygous in the F_1 generation and were labeled not assessable (NA), as we were unable to assess their imprinting status. The remaining seven genes were not testable (NT) due to the difficulty in obtaining high quality Sanger data. The 14 genes that did give high quality Sanger data and were heterozygous with a trackable, parent-of-origin specific SNP in the F_1 generation were selected for verification of imprinted status via pyrosequencing analysis. Additionally, we conducted a search for SNPs in the other previously known, annotated imprinted genes in *M. domestica* but did not find a SNP in

Htr2a, *L3mbtl, Mest*, or *Meis1*; thus we are unable to assess the imprinting status of these genes in fetal brain and EEM.

Table 3. Candidate-imprinted genes in fetal brain and EEM as determined by RNA-seq. Refseq gene names, Ensemble gene IDs, location of the informative SNP, and predicted imprinted status (IS) in fetal brain and EEM are shown. NA = Not Annotated; Pat. = Paternally imprinted; Mat. = Maternally imprinted; ND = Not Detectable/Low coverage; NA = Not Assessable; NT = Not Testable via Sanger.

Refseq Gene Name	Ensembl Gene ID	SNP Location		Fetal Brain IS	EEM IS
		Chr	Pos.		
Unknown_gene_1	NA	1	432003410	Pat.	ND
Pou5fl	NA	1	469390121	Mat.	NI
Npdc1	NA	1	469395728	Mat.	Mat.
Prkaa1	ENSMODG00000003591	2	35625423	NA	NA
Rwdd2	ENSMODG00000018383	2	338813088	Mat.	Mat.
Zfp68	ENSMODG00000008516	2	522422185	Mat.	Mat.
Fam169a	ENSMODG00000001914	3	49751281	Pat.	Pat.
Matn2	ENSMODG00000006840	3	362553338	NT	NT
Unknown_gene_5	NA	3	509558241	Mat.	Mat.
Unknown_gene_2	NA	3	525556810	NT	NT
Rpl36al	ENSMODG00000024968	4	105027808	Pat.	Pat.
Parp4	ENSMODG00000007894	4	288455210	NA	NA
Lig1	ENSMODG00000013344	4	412759243	NT	NT
Unknown_gene_6	NA	6	291639563	Mat.	ND
Nkrf	ENSMODG00000024032	6	291750260	Mat.	Mat.
Fkbp4	ENSMODG00000018354	8	116416577	NT	NT
Smc6	ENSMODG00000011685	Un	22460755	Pat.	Pat.
Xm_001366097.1	ENSMODG00000011693	Un	22534496	Pat.	Pat.
Unknown_gene_7	NA	Un	23482575	NT	NT
Csnk1a1	ENSMODG00000018038	Un	28330592	NT	NT
Unknown_gene_3	NA	Un	28488850	Pat.	ND
Unknown_gene_4	NA	Un	73948586	NT	NT
Igf2r	ENSMODG00000007100	2	442547043	Mat.	Mat.

To independently confirm the expression patterns of the 13 candidate-imprinted genes and *Igf2r*, we designed primers for allele-specific pyrosequencing of these loci. Ten of the 14 genes gave specific pyrosequencing PCR products in fetal brain and EEM (Table 4), and the pyrosequencing assays confirmed the imprinted state for all ten. It is noteworthy that all of the novel-imprinted genes showed some degree of leaky expression from the repressed allele (Table 4). Figure 12 shows the Sanger, RNA-seq, and pyrosequencing results for the maternally imprinted (paternally expressed) gene, *Nkrf*.

Table 4. Summary of pyrosequencing results to confirm imprinted expression. The gene name, tissues, biotype, imprinting status, and expression percentages in fetal brain and EEM of the expressed allele averaged over two runs is presented. FB = Fetal Brain; EEM = Extra-Embryonic Membranes; Mat. = Maternally Imprinted; Pat. = Paternally Imprinted; NM = Not Measured; TBD = To Be Determined.

Gene Name	Tissue	Gene Biotype	Imprinted Allele	% Expression of Active Allele (FB)	% Expression of Active Allele (EEM)
Igf2r	FB & EEM	Coding	Mat.	89.60%	91.95%
Unk_gene_1	FB & EEM	Non-coding	Pat.	87.50%	NM
Smc6	FB & EEM	Coding	Pat.	97.09%	96.89%
Unk_gene_5	FB & EEM	TBD	Mat.	88.67%	84.15%
Nkrf	FB & EEM	Coding	Mat.	97.43%	97.30%
Pou5fl (Oct3/4)	FB only	Coding	Mat.	TBD	TBD
Npdc1	FB & EEM	Coding	Mat.	87.75	84.28
Rwdd2a	FB & EEM	Coding	Mat.	77.43%	82.64%
Zfp68	FB & EEM	Coding	Mat.	TBD	TBD
Unk_gene_6	FB only	TBD	Mat.	92.25	64.18%

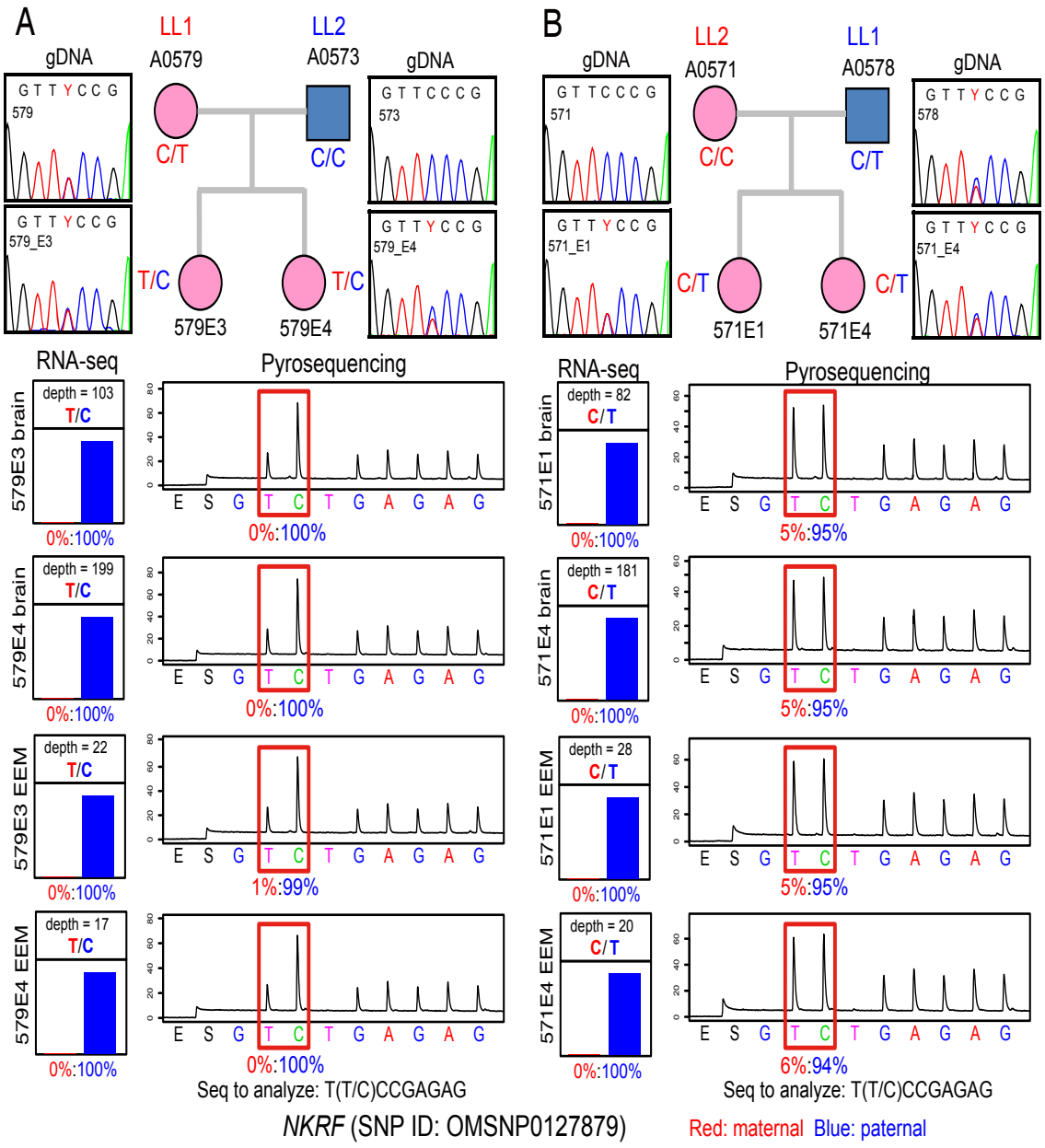


Figure 12. ***Nkrf*** **is maternally imprinted in both fetal brain and EEM. (A-B) SNP genotyping, RNA-seq allele counts, and pyrosequencing verification for maternally imprinted gene** ***Nkrf*** **in opossum fetal brain and EEM. Sanger sequencing genotyping (A, B top panels) confirmed that exonic SNP OMSNP0127879 was informative in all four F_1 embryos: A0579E3 and A0579E4 in LL1 x LL2 cross (A); A0571E1 and A0571E4 in the LL2 x LL1 cross (B). Imprinted expression was verified by both RNA-seq (left) and allele-specific pyrosequencing strategies (right) (A, B bottom panels).**

4.2.2 Histone Modification at Imprinted Genes in Fetal Brain

We assessed the presence of two marks of activation (MOAs), H3K4me3 and H3K9Ac, and one mark of repression (MOR), H3K27me3, at the putative promoters of the candidate-imprinted genes in fetal brain (Table5 and Figure 13). Of the 23 candidate-imprinted genes, the promoters of all but two, *Pou5fl* and *Unknown_Gene_7*, were marked with H3K4me3, and 18 contained an annotated promoter CpG island (Table 5). However, only 10 genes had a significant H3K27me3 peak at the promoter, and of these, seven were concurrently marked with H3K4me3 and contained an annotated promoter CpG island. Our findings suggest that H3K4me3 correlates with promoters of imprinted genes in opossum whereas H3K27me3 is not as prevalent at these loci.

Table 5. The presence of at least one MOA, MOR, and annotated CpG islands at putative promoters of candidate-imprinted genes is shown. Genes in bold were chosen for pyrosequencing verification. Pat. = Paternally imprinted; Mat. = Maternally Imprinted; ND = Not Detectable/Low coverage; NI = Not Imprinted; NT = Not Testable via Sanger or pyrosequencing; Y = Present; N = Absent.

Gene Name	Imprinted Status Fetal Brain	Imprinted Status EEM	MOAs	MORs	CpG
Unknown_gene_1	Pat.	ND	Y	Y	N
Pou5f1	Mat.	NI	N	Y	N
Npdc1	Mat.	Mat.	Y	N	Y
Prkaa1	NI	NI	Y	Y	Y
Rwdd2	Mat.	Mat.	Y	N	Y
Zfp68	Mat.	Mat.	Y	N	Y
Fam169a	Pat.	Pat.	Y	Y	Y
Matn2	NT	NT	Y	Y	Y
Unknown_gene_5	Mat.	Mat.	Y	N	N
Unknown_gene_2	NT	NT	Y	N	Y
Rpl36al	Pat.	Pat.	Y	N	Y
Parp4	NI	NI	Y	N	N
Lig1	NT	NT	Y	N	Y
Unknown_gene_6	Mat.	ND	Y	Y	Y
Nkrf	Mat.	Mat.	Y	N	Y
Fkbp4	NT	NT	Y	N	Y
Smc6	Pat.	Pat.	Y	Y	Y
Xm_001366097.1	Pat.	Pat.	Y	Y	Y
Unknown_gene_7	NT	NT	N	N	Y
Csnk1a1	NT	NT	Y	Y	Y
Unknown_gene_3	Pat.	ND	Y	-	Y
Unknown_gene_4	NT	NT	Y	Y	N
Igf2r	Mat.	Mat.	Y	N	Y

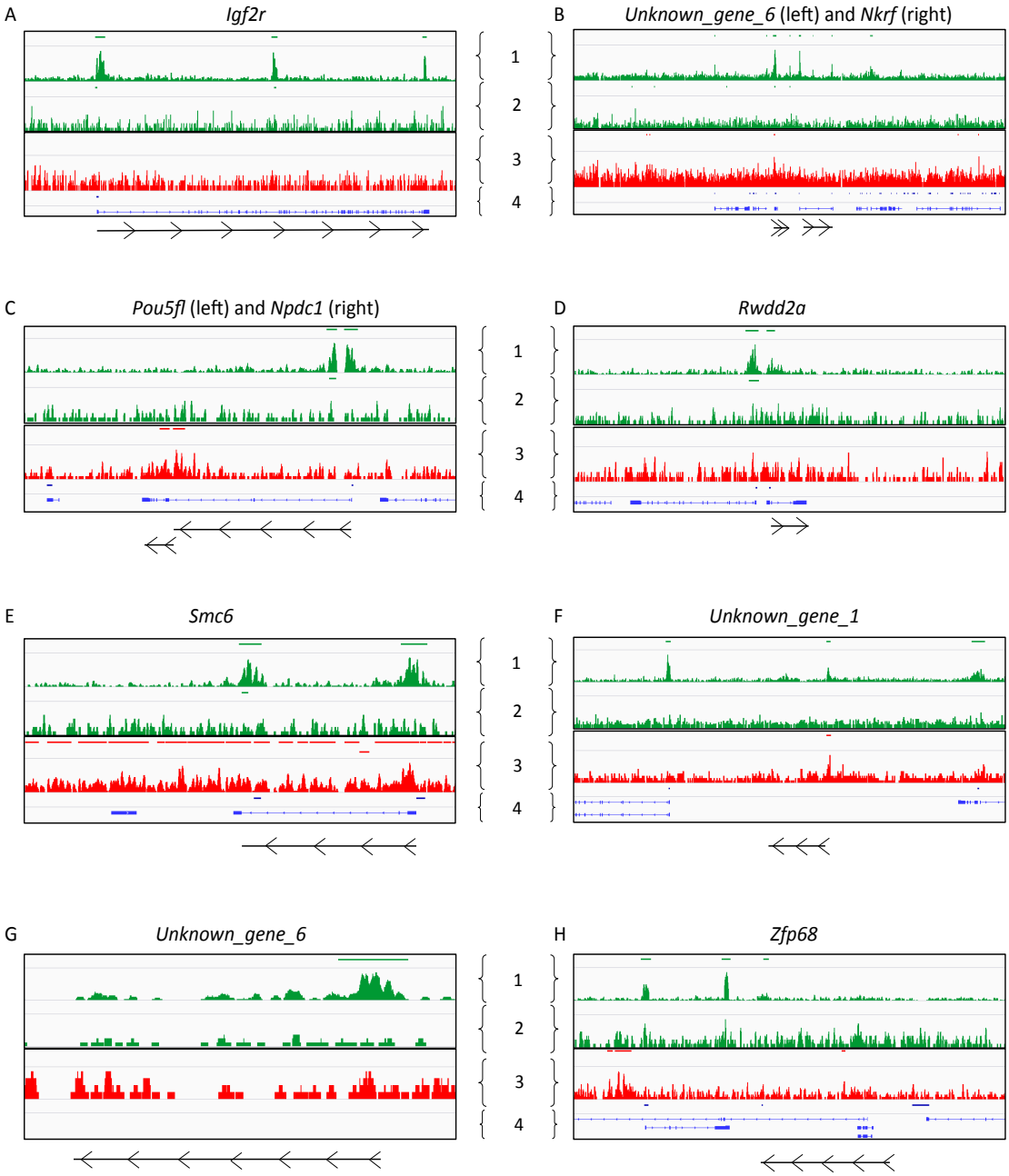


Figure 13. Histone modifications across selected candidate-imprinted genes. A) *Igf2r*; B) *Unknown_gene_6* and *Nkrf*; C) *Pou5fl* and *Npdc1*; D) *Rwdd2a*; E) *Smc6*; F) *Unknown_gene_1*; G) *Unknown_gene_5*; H) *Zfp68*. 1 = H3K4me3; 2 = H3K9Ac; 3 = H3K27me3; 4 = Annotated CpG islands (top) and gene annotation (bottom), if present. Revised gene annotation according to our RNA-seq data is presented below each box. Arrows indicated direction of transcription.

4.2.3 Comparison of Histone State Profiles between Fibroblasts and Fetal Brain

We found a noticeable difference in peak morphology for H3K9me3 and H3K27me3 between fetal brain and fibroblast cell lines. In fibroblast cells, the profile of H3K9me3 exhibits many significant intra- and inter-genic peaks that are associated with a subset of promoters; however, the H3K9me3 signal in fetal brain is diffuse with only a few significant peaks (Figure 14A). The H3K27me3 signal in fibroblasts was generally diffuse across the autosomes with only a small number of significant peaks at promoters, and in most cases, shows a depletion of signal, or trough, at promoters and across gene bodies of transcribed genes. In fetal brain, the distribution of the H3K27me3 signal is concentrated at the promoters of genes, is not depleted across gene bodies, and shows, in many cases, a significant peak of enrichment at promoters (Figure 14B). In contrast to the two MORs, the distribution and morphology of H3K4me3 peaks show a similar distribution between the two tissues (Figure 14D).

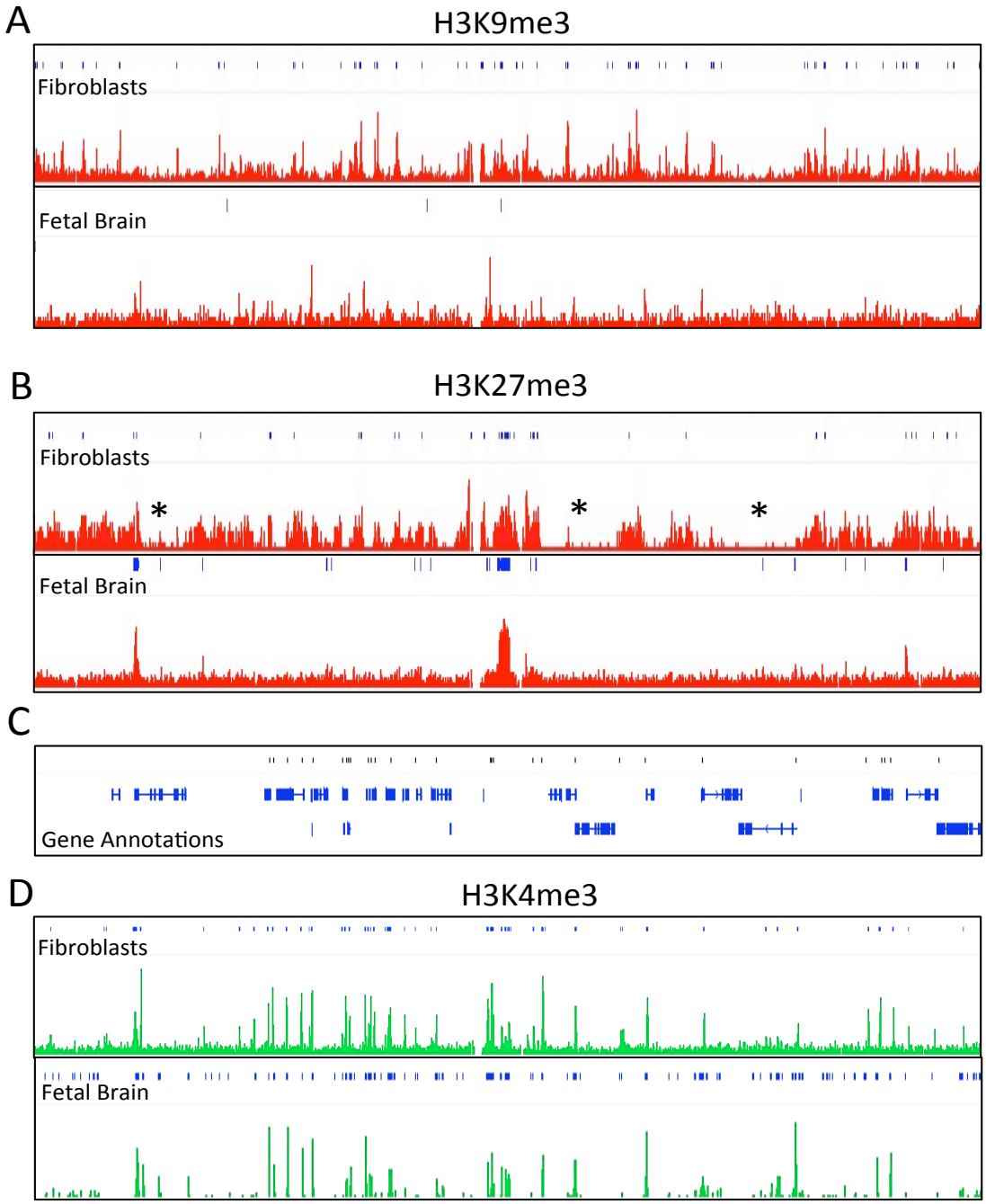


Figure 14. Examples of histone peak morphology in ChIP-seq samples from fibroblast and fetal brain. Comparison of histone peak morphologies of the region of chromosome 2 (456,062,440 - 458,903,191) for H3K9me3 (A), H327me3 (B), and H3K4me3 (D). Significant MACS peaks are indicated by blue bars above histone profile plots. Areas of H3K27me3 depletion across promoters and gene bodies are indicated by asterisks (B). (C) Annotated Ensembl genes (blue bars) and CpG islands (black bars) in this region.

4.2.4 *Smc6* is X-Linked

The histone modification profile of *Smc6* led us to postulate that it was X-linked (see Chapter II). To test this, we conducted DNA-FISH using the BAC clone that contains *Smc6* (VMRC18:415P26), and it mapped to the opossum X chromosome (Figure 15). This result not only demonstrates our ability to find imprinted genes but also illustrates the value of pairing expression data with histone state profiles to correctly identify imprinted genes.

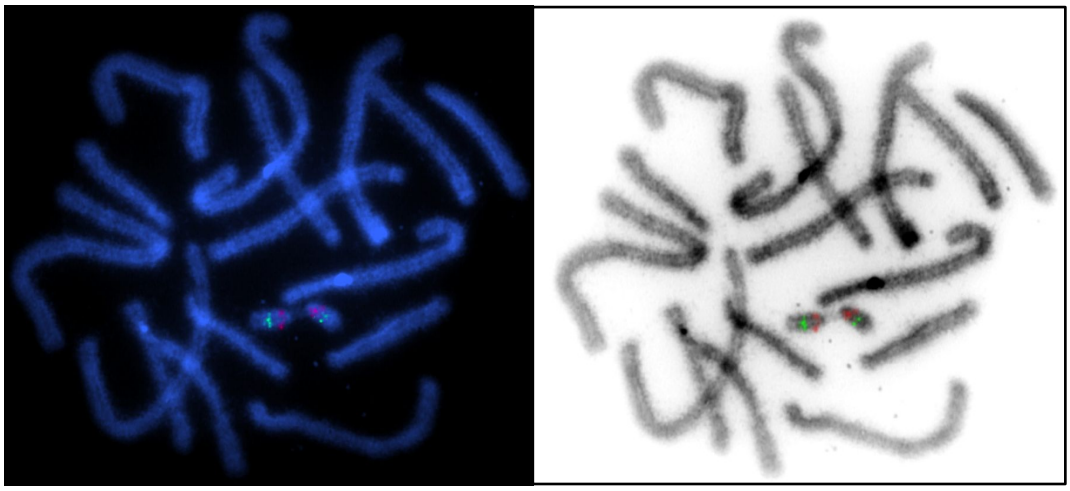


Figure 15. *Smc6* BAC clone maps to the opossum X chromosome using DNA-FISH. Red = *Smc6* (VMRC16:415P26); Green = X-linked BAC (VMRC18:608C5).

4.2.5 DNA Methylation of Promoter CpG Islands for Novel Imprinted Genes

To determine if DNA methylation was present at the promoters of novel imprinted genes, we quantified DNA methylation percentages at the annotated promoter CpG islands for *Rwdd2a* (13 CpGs assayed) and *Unknown_gene_1* (6 CpGs assayed) using bisulfite sequencing (Figure 16 and Table 6) and *Rwdd2a* and *Npdc1* using the

PyroMark assay (Table 6). We also included the X-linked gene *Smc6* (12 CpGs assayed), which is expected to lack promoter CpG island methylation, in the bisulfite sequencing analysis.

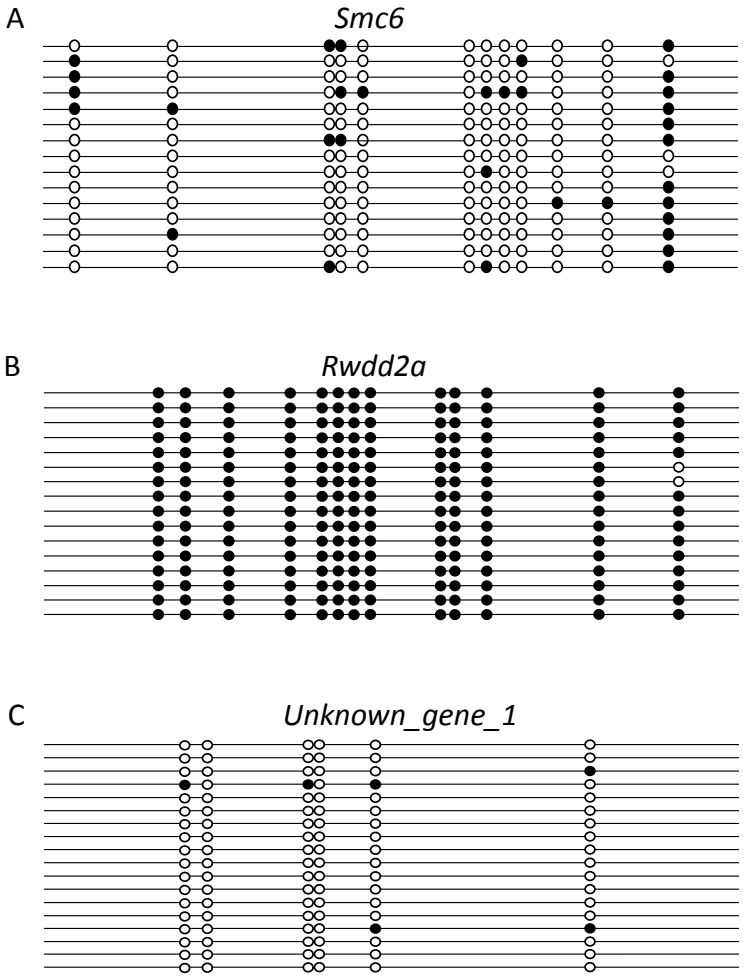


Figure 16. Summary of bisulfite sequencing results of promoter regions of A) *Smc6*, B) *Rwdd2a*, and C) *Unknown_gene_1* in fetal brain of A0592E1. Each line represents a single sequenced clone. Open Circles = Unmethylated Cytosine; Filled Circles = Methylated Cytosine.

Table 6. Target sequence coordinates for analyses of DNA methylation at promoters of novel imprinted genes. Four genes were assayed. BS = Bisulfite Sequenced; PM = Pyromark Assay.

Gene	Assay	Target Location	Size	Methylated State
Smc6	BS	chrUn:22,489,504-22,489,642	139 bp	No Meth.
Unknown_gene_1	BS	chr1:432,036,966-432,037,112	147 bp	No Meth.
Rwwd2a	BS	chr2:338,812,440-338,812,574	135 bp	Full Meth.
Rwwd2a	PM	chr2:338,812,606-338,812,815	209 bp	DMR
Npdc1	PM	chr1:469,421,901-469,422,117	216 bp	DMR

Our bisulfite sequencing data show that *Smc6* and *Unknown_gene_1* lack promoter DNA methylation as well as a DMR; however, *Rwdd2a* showed 100% methylation at all cytosines examined. In addition, the Pyromark assay confirmed the 100% methylated state found at the *Rwdd2a* region indicated by bisulfite sequencing, but also revealed a DMR within the same CpG island, approximately 100 bp downstream of the sequence analyzed via bisulfite sequencing. We also discovered a DMR at the promoter CpG island at *Npdc1* which is the first report of promoter CpG island methylation at any autosomal imprinted gene in opossum.

4.3 Conclusions and Future Work

We have conducted the first *ab intio* search for imprinted genes in the fetal brain and EEM of a marsupial and identified 22 novel, candidate-imprinted genes and a previously described imprinted gene, *Igf2r*. Sanger and pyrosequencing analyses have confirmed the imprinted state of 10 of these genes; however, DNA-FISH showed that the paternally imprinted gene *Smc6*, which is annotated on Chromosome Un, is X-linked bringing the final count of confirmed, autosomal imprinted genes to nine. Eight of the

nine novel, confirmed-imprinted genes have not been shown to be imprinted in any marsupial or eutherian tissues examined making this the first set of marsupial-specific imprinted genes discovered in any marsupial tissue. Interestingly, all of the nine novel-imprinted genes and *Igf2r* exhibit some 'leaky' expression from the imprinted (repressed) allele. This phenomenon has been well documented for imprinted genes on the opossum X chromosome (see Chapter II) and for the opossum, fibroblast-specific imprinted gene, *Meis1* (see Chapter III). Leaky expression of imprinted genes, whether autosomal or X-linked, appears to be common in opossum; however, the mechanism and biological or evolutionary significance of this expression pattern has yet to be determined. One possible explanation for such leaky expression in marsupials could be a more relaxed epigenetic regulation of imprinted loci due to a general lack of strong differential DNA methylation concomitant with the occurrence of transcriptionally opposing histone modifications at the promoters of imprinted loci.

H3K4me3, a mark-of-activation, was found at the promoters of all but one of the confirmed-imprinted genes, *Pou5f1*; and two genes, *Unknown_gene_1* and *Unknown_gene_6*, were simultaneously marked with H3K4me3 and H3K27me3 (a mark-of-repression). Neither of these genes have been previously annotated in the opossum genome, and their sequence alignments and characteristics suggest that they are novel, non-coding RNAs, although this speculation has not been verified. In addition, the promoter CpG islands of three genes, *Rwdd2a*, *Unknown_gene_1,* and *Npdc1*, have been examined for DNA methylation using bisulfite sequencing and Pyromark analyses and have given interesting results inasmuch as adjacent sections of CpG islands show

different methylation patterns. We also discovered two novel DMRs at the promoters of *Rwdd2a* and *Npdc1*, and a fully methylated section less than 100 bp upstream of the DMR at *Rwdd2a*. Confirmation withstanding, our preliminary data suggests that promoter CpG island methylation does correlate with promoters of some marsupial imprinted genes and possibly plays a role in regulating imprinting expression in fetal brain and EEM. This differs drastically from X-linked genes (Chapter II) and the imprinted and monoallelically expressed genes in fibroblasts (Chapter III) which lacked promoter CpG island methylation in all cases (except for *Rsx* on the X chromosome) and represents the first evidence of promoter CpG island methylation for imprinted, autosomal genes in opossum.

To better understand this set of imprinted genes in opossum fetal brain and EEM, more work is needed. An in-depth annotation of the imprinted genes, both protein-coding and non-coding RNAs, is ongoing, but the identification of possible targets for the non-coding RNAs is also needed. The newly identified DMRs, as well as promoters that appear to lack them, need to be confirmed by both bisulfite sequencing and Pyromark assay, and the promoters of more imprinted genes need to be examined to fully assess the extent and biological significance of promoter CpG island methylation for imprinted loci in opossum. Furthermore, a search for parent-specific SNPs, which will allow the tracking of allele-specific methylation and transcriptionally opposing histone modifications, needs to be conducted to assess how well these highly suggestive epigenetic features correlate with the transcriptional states of the active and repressed (imprinted) alleles generally. Completion of these experiments is critical to elucidate the

possible mechanisms of imprinting in opossum and complete the catalogue of imprinted genes in opossum fetal brain and EEM.

CHAPTER V

EXTENDED METHODS

5.1 Chapter II Extended Methods

5.1.1 Opossum Lines and Crosses

Two random-bred stocks of *Monodelphis domestica*, designated LL1 and LL2 derived from the Population 1 and Population 2 stocks described by VandeBerg and Williams-Blangero (2010) were used to generate reciprocal F_1 embryos (Table 7). The LL1 stock was derived as a subgroup from Population 1 ancestors, which is the same ancestral stock that furnished DNA for the MonDom5 opossum genome sequence (Mikkelsen *et al.* 2007b); LL2 was derived by admixture of Population 1 and Population 2 animals and comprises an approximate 1:7 mixture of the Population 1 and Population 2 genetic backgrounds, respectively (John L. VandeBerg, personal communication). Embryos from the parental crosses of LL1 x LL1 and LL2 x LL2 were also collected to assist in determining the direction of allelic transmission. To control for shared alleles segregating in the two stocks, the same males were used for the F_1 and parental crosses, and the females in F_1 and parental crosses were full siblings (Figure A1, A-D). To collect prenatal stage animals of known gestational ages, the time of copulation was determined by videotaping paired animals as described by Mate *et al*. (1994) with minor modifications. All procedures involving opossums were approved by the Institutional Animal Care and Use Committee of Texas A&M University, College Station (TAMU Animal Use Protocols 2011-141 and 2011-191).

Table 7. List of animals, developmental stages, and tissues used for RNA-seq experiments.

No.	Sample ID*	Sex	Dam ID	Sire ID	Dam Stock	Sire Stock	Develop. Stage	Tissue
1	A0571_p1	female	A0571	A0578	LL2	LL1	E13.0**	EEM***
2	A0571_p4	female	A0571	A0578	LL2	LL1	E13.0	EEM
3	A0579_p3	female	A0579	A0573	LL1	LL2	E13.0	EEM
4	A0579_p4	female	A0579	A0573	LL1	LL2	E13.0	EEM
5	A0571_b1	female	A0571	A0578	LL2	LL1	E13.0	Fetal head
6	A0571_b4	female	A0571	A0578	LL2	LL1	E13.0	Fetal brain
7	A0579_b3	female	A0579	A0573	LL1	LL2	E13.0	Fetal head
8	A0579_b4	female	A0579	A0573	LL1	LL2	E13.0	Fetal brain
9	A0580_p1	female	A0580	A0578	LL1	LL1	E12.5	EEM
10	A0580_p5	male	A0580	A0578	LL1	LL1	E12.5	EEM
11	A0572_p1	female	A0572	A0573	LL2	LL2	E12.5	EEM
12	A0572_p3	male	A0572	A0573	LL2	LL2	E12.5	EEM
13	A0580_b1	female	A0580	A0578	LL1	LL1	E12.5	Fetal head
14	A0580_b5	male	A0580	A0578	LL1	LL1	E12.5	Fetal head
15	A0572_b1	female	A0572	A0573	LL2	LL2	E12.5	Fetal head
16	A0572_b3	male	A0572	A0573	LL2	LL2	E12.5	Fetal head

*: "p" = placenta/EEM; "b" = fetal brain/fetal head.**: embryonic day 13. ***: extra-embryonic membranes.

5.1.2 Tissue Selection, Dissection, and Sex-Typing

Based on video evidence of mating, females were euthanized at 13 days post-copulation (d.p.c.) and the fetuses and respective extra-embryonic membranes (EEM) were collected by dissection and placed in either RNA*later* (Ambion) or phosphate buffered saline (PBS) and stored at -80 ^{o}C. Head/brain and EEM tissues were dissected from each fetus (Figure 1A, B). At this early stage of development, brain size was very small; and despite identical gestation times, fetuses within a litter varied in size. For smaller fetuses, the head was further dissected to remove jaws, muzzle, and other non-cranial structures. The remaining cranial region and the corresponding EEM tissues were used for RNA extraction, whereas the limbs were used for genomic DNA

extraction. For the two largest fetuses (sample_ID: A0571_b4 and A0579_b4), we were able to identify and isolate the frontal cortices and other brain structures.

To avoid maternal cell contamination during the collection of EEM, we followed the umbilical cord to the uterine/placenta interface and retreated the scissors ~1-2 mm before cutting. The fetuses were sex-typed using Y-chromosome-specific primers (Figure A1.E) developed in collaboration with the Opossum Y chromosome Mapping Project (Page Laboratory, Whitehead Institute for Biomedical Research, Cambridge, MA) using a Y-specific BAC sequence containing the opossum sex-determining region on the Y (*Sry*) gene (AC239615.3, JN086997.1). These primers amplify a fragment of the Y-linked *Sry* gene in male cells/tissues in a highly replicable manner but show no amplification in female cells/tissues. Two female fetuses from each reciprocal cross, and one male and one female fetus from each stock-specific cross (total of eight individuals) were used for RNA extraction.

5.1.3 Total RNA and Genomic DNA Extractions and QC

After collection, the tissues were homogenized in TriReagent (Invitrogen) and total RNA was extracted using BCP (1-bromo-2 chloropropane), precipitated with isopropanol, and resuspended in RNase-free water. Potential DNA contamination was removed by both DNase I treatment and Qiagen RNeasy Plus Mini kit (Qiagen, CA). RNA concentrations and A260nm/A280nm ratios were checked with a NanoDrop ND-1000 Spectrophotometer. The RNA quality was validated on both 1% agarose gels and an Agilent 2100 Bioanalyzer. To extract genomic DNA, tissue was minced and digested

with proteinase K overnight followed by phenol:chloroform extraction, ethanol precipitation overnight, and elution in TE.

5.1.4 Illumina mRNA-Seq and Sequence Alignment

mRNA-seq libraries were made from brain and EEM RNA samples (1-3 µg total RNA input) of the four individuals in reciprocal F_1 crosses and four individuals in the parental crosses described above, using the Illumina TruSeq RNA Sample Prep Kit (Illumina Inc., CA). The 16 libraries (eight from the F_1 crosses and eight from the parental crosses) were multiplexed and run on four 51 bp single-end lanes each, on an Illumina HiSeq 2000 instrument (Illumina Inc., CA). Image analysis and base calling were performed using Illumina software. In total, 1.5 billion short reads (76.5 billion bps) were generated from the 16 samples. The q-score and nucleotide distribution QC indicated good RNA-seq data quality. The reads were aligned to the opossum reference genome assembly (MonDom5, http://genome.ucsc.edu/) using TopHat v1.4.1 (Trapnell *et al.* 2009) with three mismatches allowed. Eighty-two percent of the reads were uniquely mapped to the opossum reference genome (Table 8).

5.1.5 Quantification of Total and Parent-of-Origin Allele-Specific Expression from X-Linked Gene RNA-Seq Data

The metric for total gene expression level is derived from counts of transcript reads that match specific loci in the reference genome. This metric, FPKM (Fragments Per Kilobase-pair of exon Model), was calculated for all samples using Cufflinks v1.3.0 (Trapnell *et al.* 2010) based on all mapped reads from the TopHat alignments. The multiple mapped reads were weighted using the "-u" parameter in Cufflinks. The

expression level was normalized across brain and EEM samples using quartile normalization.

Table 8. Summary of Illumina raw and mapped read data for RNA-seq experiments. All experiments were run on Illumina HiSeq2000.

No.	Sample_ID*	read length	multiplexing	# of lanes	total number of reads	uniquely mapped reads	mapping percentage
1	A0571_p1	51bp	1-8 8-plex	4	122,984,871	96,491,161	78.46%
2	A0571_p4	51bp	1-8 8-plex	4	102,676,227	80,265,550	78.17%
3	A0579_p3	51bp	1-8 8-plex	4	97,126,305	75,901,042	78.15%
4	A0579_p4	51bp	1-8 8-plex	4	93,176,519	70,214,242	75.36%
5	A0571_b1	51bp	1-8 8-plex	4	111,441,595	92,237,141	82.77%
6	A0571_b4	51bp	1-8 8-plex	4	104,401,730	87,413,805	83.73%
7	A0579_b3	51bp	1-8 8-plex	4	123,279,342	102,502,500	83.15%
8	A0579_b4	51bp	1-8 8-plex	4	107,001,858	88,789,694	82.98%
9	A0580_p1	51bp	9-16 8-plex	4	89,170,950	73,057,075	81.93%
10	A0580_p5	51bp	9-16 8-plex	4	80,905,285	66,088,617	81.69%
11	A0572_p1	51bp	9-16 8-plex	4	57,906,830	47,408,285	81.87%
12	A0572_p3	51bp	9-16 8-plex	4	76,082,440	62,107,048	81.63%
13	A0580_b1	51bp	9-16 8-plex	4	78,348,683	66,295,368	84.62%
14	A0580_b5	51bp	9-16 8-plex	4	111,042,946	93,760,306	84.44%
15	A0572_b1	51bp	9-16 8-plex	4	75,672,460	63,848,790	84.38%
16	A0572_b3	51bp	9-16 8-plex	4	78,107,923	65,288,060	83.59%

*: "p" = placenta/EEM; "b" = fetal brain/fetal head

The RNA-seq read coverage and mapping percentages were homogeneous across all the samples, rendering the before-normalization and after-normalization M-A plots equally informative (Figure A3). We covered 11,465 Ensembl opossum gene models with FPKM≥1 in all eight brain samples and 10,518 gene models in the eight EEM samples. The RNA-seq data were deposited in the *Gene Expression Omnibus* (*GEO*) *database* under accession number GSE45211.

SNP positions were called in combined RNA-seq data only from reads that mapped uniquely to the opossum reference sequence with ≥40 matching reads, using SAMtools software (Li *et al.* 2009). Problematic SNPs, such as those with a third allele, near an indel position, or at the exon-intron junctions, were removed. In total, 68,000 SNPs were called. The reference and alternative allele counts were summarized at high quality X-linked SNP positions. We further selected high-coverage SNPs with ≥8 X coverage in at least one of two female F_1 individuals in both reciprocal crosses. Nineteen additional X-linked genes assigned to MonDom5 ChrUn were also included (Table 9). Retrotransposed X-linked genes were excluded from the analysis (Table 10).

Table 9. X-linked genes on the unmapped scaffold (ChrUN) covered in RNA-seq data.

No.	Scaffold ID	Coordinates on ChrUn	Gene name
1	scaffold_217	5892577-6515230	*ABCB7*
2	scaffold_217	5892577-6515230	*KIAA2022*
3	scaffold_217	5892577-6515230	*RLIM*
4	scaffold_217	5892577-6515230	*SLC16A2*
5	scaffold_251	23882535-24256940	*RNF170-like*
6	scaffold_251	23882535-24256940	*Rab-1B-like*
7	scaffold_251	23882535-24256940	*KLHL4*
8	scaffold_251	23882535-24256940	*ENSMODG00000022796*
9	scaffold_261	27708438-28006343	*MTMR8*
10	scaffold_261	27708438-28006343	*XM_001378411.1*
11	scaffold_298	37707632-37883514	*FUNDC2*
12	scaffold_298	37707632-37883514	*MTCP1*
13	scaffold_298	37707632-37883514	*BRCC3*
14	scaffold_298	37707632-37883514	*PICALM-like*
15	scaffold_298	37707632-37883514	*VBP1*
16	scaffold_298	37707632-37883514	*RAB39B*
17	scaffold_352	45973239-46059762	*SLC9A6*
18	scaffold_352	45973239-46059762	*MMGT1-like*
19	scaffold_1524	75514428-75525886	*EMD*

Table 10. Retrotransposed X-linked genes excluded from the analysis.

No.	Ensembl trans ID	Ensembl gene ID	Gene name
1	ENSMODT00000009535	ENSMODG00000007539	*HAUS7*
2	ENSMODT00000039334	ENSMODG00000025336	*HSP90AB1-transposed*
3	ENSMODT00000039297	ENSMODG00000025324	*LUC7L2-transposed*
4	ENSMODT00000008835	ENSMODG00000006990	*ENSMODG00000006990*
5	ENSMODT00000021722	ENSMODG00000017105	*PABPC1L2B*
6	-	-	*X.288*
7	ENSMODT00000009094	ENSMODG00000007195	*RNASET2-transposed*
8	ENSMODT00000026898	ENSMODG00000021142	*TAX1BP1-transposed*
11	ENSMODT00000017947	ENSMODG00000014093	*SSU72-transposed*
13	ENSMODT00000022195	ENSMODG00000017486	*SET-transposed*
14	ENSMODT00000011525	ENSMODG00000009064	*KDM2A-transposed*

To quantify allele-specific expression in brain and EEM tissues from the reciprocal crosses, we calculated the ratio of reference allele-containing reads divided by the total coverage at each identified SNP position (Wang *et al.* 2008) (Table B1). The transmission directions were inferred from the parental crosses and SNP variant data from other LL1 individuals for which RNA-seq data was available (data not shown).

5.1.6 X-Linked SNP Genotyping by Sanger Sequencing

To confirm the parental origins of the two alleles at escaper loci, we genotyped the F_1 individuals and their parents at informative SNP positions using Sanger sequencing (Table B2). Primers targeting informative SNPs were designed using Primer3 (http://frodo.wi.mit.edu/). DNA was PCR amplified using TaqGold® polymerase, purified, and Sanger sequenced at Beckman Coulter Genomics (Danvers, MA). Gel purification was necessary for some samples due to the presence of non-specific PCR products. Sequences were viewed, aligned, and analyzed using

Sequencher 4.10®. For non-escaper genes, in which expression of only a single allele was observed, the possibility that any particular informative SNP might not be heterozygous in all F_1 individuals had to be considered. This was necessary because the LL1 and LL2 lines are not 100% inbred and some alleles are shared between them at segregating loci. To check whether the SNPs were heterozygous in the female F_1 individuals, we classified the informative SNPs into six classes (Table 11), and randomly selected 20 SNPs (one SNP per gene) for genotyping by Sanger sequencing. All 20 of these were verified as heterozygous in at least two of the four female samples (Table B2).

Table 11. Informative SNP classes of non-escaper genes.

Class *	Count	With parental crosses support	Ref. allele	Segregating within other LL1 individuals?	Class description
1	75	YES	LL1	NO	best "fixed" SNPs
2	9	YES	LL2	NO	best "fixed" SNPs
3	49	YES	LL1	unknown	2nd best "fixed" SNPs
4	2	YES	LL2	unknown	2nd best "fixed" SNPs
5	173	NO	LL1	NO	2nd best "fixed" SNPs, with less support from the parental cross
6	24	NO	LL2	NO	2nd best "fixed" SNPs, with less support from the parental cross

*: SNPs outside these six classes belong to class 0.

5.1.7 Validation of X-Linked Escaper and Non-Escaper Gene Expression by Allele-Specific Pyrosequencing

Fifteen of twenty-four escaper genes possessed more than one informative SNP in the RNA-seq dataset. Judged from the abundances of RNA-seq reads containing these linked SNPs, the paternal allelic expression levels are consistent for SNP sites within a gene (Table B1). This agreement between multiple SNPs within the same gene provided an internal validation for the allele-specific expression quantification. To verify the escaping status of these genes and obtain an estimate of paternal allelic expression levels using an independent method, we performed allele-specific pyrosequencing on all 24 escaper genes, one non-escaper gene (*HPRT1*), and one autosomal control gene (*GPM6B*) (Figure 2A-D, Figure A2, and Table B2). Pyrosequencing PCR and sequencing primers were designed to target informative exonic SNP positions within selected genes using PyroMark Assay Design Software Version 2.0.1.15 (Qiagen, CA). To eliminate potential amplification bias, all primers were checked to guarantee that they did not individually overlap base positions that differed between the LL1 and LL2 parents.

Pyrosequencing PCR amplification was carried out in 40 µl system using Ampli-Taq Gold polymerase (Life Technologies) under the following cycling conditions: 1 cycle of 95^{o} C for 5 min, 45 cycles of 95^{o} C for 45 sec, 57^{o} C for 30 sec, 72^{o} C for 20 sec, followed by 1 cycle of 72^{o} C for 10 min. PCR products were prepared according to the manufacturer's protocol and then loaded on the PSQ 96MA Pyrosequencer (Qiagen,

CA) with the PyroMark Gold Reagents (Qiagen, CA) using the Allele Quantification method (AQ). Two technical replicates were done for each gene in each sample.

5.1.8 Native ChIP-Seq and Data Analysis

Native-ChIP (N-ChIP) was conducted on a primary fibroblast cell line (derived from adult ear pinna), fetal brain, and EEM using a method modified from Dindot *et al.* (2009). Briefly, total tissue samples of fetal brain and EEM were washed in PBS and homogenized in 500 µl of Buffer I (0.3 M sucrose, 60 mM KCl, 15 mM NaCl, 5 mM $MgCl_2$, 0.1 mM EDTA, 15mM Tris, 0.5 mM DTT, 0.1 mM PMSF). The sample was centrifuged for 5 min. at 3000g, the supernatant was removed, and the pellet was re-suspended in 200 µl of Buffer I. Cells were lysed for 5 minutes on ice by adding 200 µl of Buffer II (Buffer I + 4 µl of NP40), and nuclei were isolated by centrifugation of lysed cells for 20 min at 10,000Xg through 1.5 ml of Buffer III (1.2 M sucrose, 60 mM KCl, 15 mM NaCl, 5 mM $MgCl_2$, 0.1 mM EDTA, 15 mM Tris, 0.5 mM DTT, 0.1 mM PMSF). The pellet was washed with Buffer I, centrifuged, and re-suspended in 350 µl of MNase (micrococcal nuclease) Digestion Buffer (0.32 sucrose, 4 mM $MgCl_2$, 50 mM Tris, 0.1 mM PMSF). Chromatin was digested using 10 units of MNase (Sigma, N5386) for 10 min at 37°C. The reaction was stopped using 50 µl of 0.5 M EDTA. For input control, 100 µl of digested chromatin were removed and stored at -20°C.

Equal aliquots of the remaining digested chromatin (EEM = 2.0 µg/rxn; fetal brain = 11 µg/rxn) were incubated at 4° C overnight with anti-H3K4me3 (Millipore #07-473), anti-H3K9Ac (Millipore #CS200583), anti-H3K9me3 (Millipore #07-442), anti-H3K27me3 (Millipore #07-449), or non-specific, rabbit IgG (Millipore #12-370).

Antibody-bound chromatin was isolated using Dynabeads® Protein A (Invitrogen), washed, and eluted. N-ChIP and input DNA were purified using the Qiagen MiniElute Spin Columns (Qiagen) and enrichment was verified using real-time PCR. Illumina libraries were constructed at Global Biologics, LLC, and sequenced at the University of Missouri – Columbia DNA Core Facility and Genomics Resources Core Facility at Weill Cornell Medical College (New York, NY). Raw reads were quality filtered, trimmed, and aligned using Bowtie in the Galaxy suite (Giardine *et al.* 2005; Blankenberg *et al.* 2010; Goecks *et al.* 2010). Aligned reads were visualized on the UCSC genome browser (Kent *et al.* 2002) and Integrative Genomics Vizualizer (IGV) (Robinson *et al.* 2011; Thorvaldsdottir *et al.* 2012) and significant peaks were called using the MACS algorithm (Zhang *et al.* 2008) (Table B3). The ChIP-seq data were deposited in the *Gene Expression Omnibus* (*GEO) database* under accession number GSE45186.

5.1.9 Re-Annotation of X-Linked Promoters

Due to the incompleteness of the annotation of the *Monodelphis* genome, it was necessary to re-annotate the X-linked genes to ensure that the 5' exons, UTRs and putative promoters were located correctly in the corresponding gene model. We used predicted RNA structure from TopHat, the presence of CpG islands (both currently and newly annotated), and the presence of H3K4me3 peak to annotate new 5' exons and putative promoters for 312 X-linked genes (results not shown).

5.1.10 Bisulfite-Sequencing of Promoter DNA

Two μg of genomic DNA were treated with bisulfite using the EpiTech Bisulfite Kit from Qiagen, Inc. Treated DNA was PCR amplified using primers designed by Methyl Primer Express v 1.0 (Applied Biosystems). One μl of bisulfite-treated gDNA was used in PCR amplification in 50 μl reaction using Ampli-Taq Gold polymerase (Life Technologies) under the following cycling conditions: 1 cycle of 95^{o} C for 5 min, 35 cycles of 95^{o} C for 15 sec, 50° or 55^{o} C for 30 sec, 72^{o} C for 20 sec, followed by 1 cycle of 72^{o} C for 10 min. PCR products were cloned using the TopoTA Cloning® Kit (Life Technologies). For each cloned PCR product, plasmids were purified from at least 16 transformed colonies and Sanger sequenced at Beckman Coulter Genomics (Danvers, MA) using the M13 forward primer. Sequences were viewed, aligned, and analyzed using Sequencher 4.10®.

5.1.11 Quantification of DNA Methylation Percentage Using PyroMark Assays

Quantification of methylation percentages in individual consecutive CpG sites was achieved with high reproducibility by pyrosequencing of bisulfite-treated DNA using the Pyromark Assay method. Bisulfite conversion was carried out on 500 ng genomic DNA of fetal brain and EEM samples in both sexes with the Qiagen EpiTect Bisulfite Kit (Qiagen, CA). PyroMark primers were designed to target the CpG islands with PyroMark Assay Design Software Version 2.0.1.15 (Qiagen, CA). PCR products were prepared, run, and analyzed on the PSQ 96MA Pyrosequencer (Qiagen, CA) with PyroMark CpG software 1.0.11. Background subtraction was done using the “control

peak heights" feature. Each sample was repeated twice and the average of the runs was used for the analysis.

5.2 Chapter III Extended Methods

5.2.1 Western Blot Using Histone Antibodies Against *M. domestica* Proteins

All antibodies used in this study were created in mouse or rabbit. To validate that the antibodies bind Histone 3 in *M. domestica*, we extracted protein from female liver tissue and ran a Western Blot Analysis. To enrich for histone proteins, nuclei were isolated using a sucrose gradient as described in section 5.1.8 above and placed in 1 mL of 0.2 N HCl (acid extraction) overnight at 4^{o} C. The samples were centrifuged at 6,500 X g for 10 minutes at 4^{o} C and the supernatant, which contains basic proteins, was removed and stored at -20^{o} C. Protein concentrations were determined using the Bradford assay. Purified preparations (1 μg of total basic protein) and controls (1 μg of total protein extracted from HeLa cells, not acid extracted) were loaded on 1% agaorse gels and blotted on nylon membranes and visualized using antibodies against specific histone modification states by standard Western Blot techniques. The expected 17 kDa bands revealing the presence of histone proteins were present in all samples and a more intense signal was seen in the acid extracted, *M. domestica* samples (Figure 17).

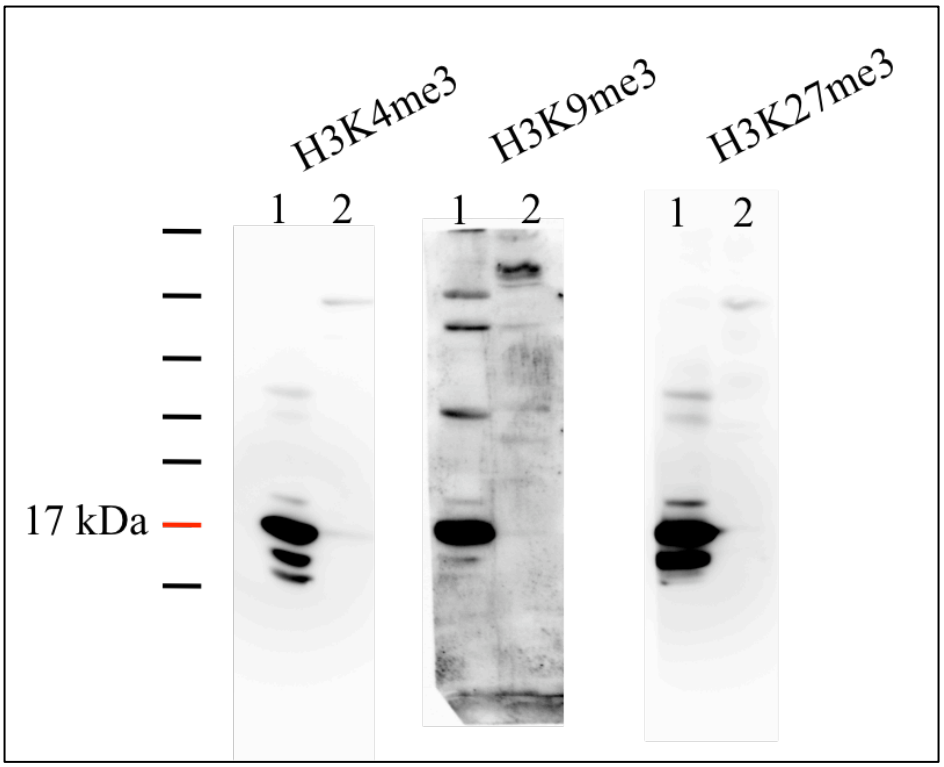


Figure 17. Western blot analysis using antibodies specific to a posttranslational modification of Histone 3. H3K4me3 (Millipore #07-473), H3K9me3 (Millipore #07-442), and H3K27me3 (Millipore #07-449). 1 – basic proteins extracted using HCl from *M. domestica* liver tissue. 2 – protein extracted from human HeLa cells, not acid extracted. The Histone 3 band can be seen at 17 kD in all samples and is stronger in the acid extracted, *M. domestica* samples.

CHAPTER VI

CONCLUSIONS

The aim of my dissertation research was to identify, on a genome-wide scale, previously unknown imprinted loci in *M. domestica* and use my findings to make observations concerning the epigenetic and functional characteristics of these genes in comparison to what is known about imprinted gene expression in therian mammals generally. More specifically, I wanted to employ a search strategy that was not constrained by *a priori* knowledge of the imprinted states of genes in eutherian mammals. In order to conduct this search, I utilized specific histone modifications that have been shown to be associated with imprinted genes in eutherians, as well as genetic crosses designed to enrich for stock-specific SNPs, which allowed me to track parent-of-origin-specific-allele expression patterns. Using RNA-seq and ChIP-seq approaches, I have demonstrated the ability to detect imprinted genes, both X-linked and autosomal, based on epigenetic marks and structural characteristics of promoter regions in opossum fetal brain, EEM, and fibroblast cells and to predict with considerable accuracy the transcriptional states of X-linked and autosomal genes using histone modification profiles.

6.1 Detection of Autosomal and X-Linked Candidate-Imprinted Genes in *M. domestica* by Epigenetic Profiling

Applying ChIP-seq to a male fibroblast cell line and female fetal brain showed that transcriptionally active and repressive histone modifications are associated with the

transcriptional states of both autosomal and X-linked genes in *M. domestica*. This enabled me to use the simultaneous overlap of these marks at annotated loci to identify candidate-imprinted and monoallelically expressed genes. I also observed differences in histone modification profiles between fetal brain and fibroblast cells, especially for the repressive histone modifications. By coupling these findings with parent-of-origin-specific SNP data for individual genes (e.g., *Meis1*) and genome-wide RNA-seq data, I was able to detect the imprinted states (maternal or paternal) of 10 novel imprinted genes, the known-imprinted gene *Igf2r*, and 177 imprinted X-linked loci in the *M. domestica* genome.

Additionally, I analyzed DNA methylation patterns at promoter CpG islands for 26 X-linked genes and seven autosomal imprinted genes (including *Igf2r*). Overall, X-linked loci lack DNA methylation with the noteworthy exception of the marsupial-specific *Rsx* locus, which has a promoter CpG island DMR that is methylated on the maternal strand only (parent-of-origin-specific methylation). In contrast, two of the seven autosomal genes, *Rwdd2a* and *Npdc1*, have DNA methylation and/or a DMR at promoter CpG islands and one, *Igf2r*, also has a DMR in a CpG island located in intron 11. These are the first reported cases of promoter CpG island methylation for genes that are expressed in a parent-of-origin-allele-specific manner in *M. domestica* and suggests a previously undescribed role for DNA methylation at imprinted loci in fetal brain and EEM of *M. domestica*. These data encourage a more in-depth analysis of DNA methylation in these tissues using genome-wide techniques such as whole genome

bisulfite sequencing (BS-seq) or methylation dependent immunoprecipitation sequencing (MeDIP-seq).

6.2 Chromosome-Wide Characterization of Paternally Imprinted X-Chromosome Inactivation in *M. domestica*

6.2.1 Cataloging of Imprinting and Non-Imprinted X-Linked Genes

The cataloging of X-linked and autosomal imprinted genes in multiple species and correlating specific epigenetic marks with the expression patterns of these loci enables informed speculation regarding the biological and evolutionary forces that have shaped gene expression throughout mammalian evolution. X-chromosome inactivation is a therian-specific phenomenon that has been postulated to be a dosage compensation mechanism that maintains equal expression of X-linked genes between males and females (Lyon 1961; Ohno 1967; Heard *et al.* 1997); however, prior to the research reported here, no systematic, chromosome-wide assessment of the extent of XCI or its epigenetic regulation had been conducted in any marsupial. I found that most X-linked genes of *M. domestica* are indeed paternally imprinted; however, ~14% escape inactivation. In eutherians, the percentages of genes that escape XCI vary across species, and the suites of genes that escape XCI differ between species. Adding to this complexity, only one gene that escapes pXCI in *M. domestica* has a homologue that escapes XCI in examined tissues of human or mouse (i.e., the overlap in XCI escaper genes between *M. domestica* and the best studied eutherian models is virtually nonexistent). Thus, it would appear that the phenomenon of escape from XCI is

conserved throughout the therian lineage but that species-specific selective pressures have targeted the escape of different genes in different lineages.

6.2.2 Epigenetic Patterns at Paternally Imprinted X-Linked Genes in *M. domestica*

The epigenetic regulation of XCI has been intensively studied in eutherians. Its establishment has been shown to be dependent upon expression of *Xist*, a *cis*-acting non-coding RNA gene which is expressed exclusively from and coats the inactive X chromosome preventing the transcription of most of its genes. This inactive state is reinforced by allele-specific DNA methylation and histone modifications (Silva *et al.* 2003; Hellman & Chess 2007; Ball *et al.* 2009). When I began this study, no *Xist* homologue or functionally analogous gene had been discovered in any metatherian (but see below), and only broad scale (i.e., FISH) experiments examining DNA methylation and histone modifications on the marsupial X chromosome had been conducted. Data from my research, reported herein, indicate that eutherians and metatherians utilize the same repressive histone modification, H3K27me3, to mark individual genes on the inactive X chromosome. Also similar to the eutherian condition, the active histone modification, H3K4me3, marks the promoters of actively transcribed X-linked genes; however, in contrast to eutherians, DNA methylation at promoter CpG islands is not present as a complementary, augmenting repressive mark at inactivated X-linked genes of *M. domestica*. This suggests that the use of histone modifications to mark the inactivated X chromosome was present before the eutherian-metatherian split, but that DNA methylation of X-linked genes arose later, as a eutherian-specific epigenetic mechanism, possibly to help stabilize the inactive state. Alternatively, DNA methylation

could have been present in the common therian ancestor and subsequently maintained in eutherians as an epigenetic regulatory mechanism and lost in metatherians, although this situation seems less parsimonious.

6.2.3 Regulation and Expression of the Marsupial-Specific *Rsx* Locus

Recently, *Rsx*, a functional analog of *Xist*, was discovered in *M. domestica* and implicated as a central participant in the establishment of XCI in marsupials. As mentioned above, during the establishment of XCI in eutherians, *Xist* transcripts coat and silence the inactive X chromosome; these transcripts also recruit factors such as repressive histone modifications and DNA methylation that help maintain the silenced state of the chromatin. In this study, I provide the first documentation of imprinted expression for *Rsx* (expressed from the paternally derived X only), a promoter DMR that is methylated exclusively on the maternal chromosome, and the presence of the activating histone modification H3K4me3, but absence of the repressive histone modification H3K27me3.

The expression of *Rsx* from the paternal allele and the presence of an allele-specific DMR at its promoter parallel the situation in eutherians in that the repressed eutherian X chromosome expresses the *cis*-acting *Xist*, and DNA methylation is used to silence *Xist* on the active X chromosome. In contrast, however, the histone modification profiles differ greatly between the *Rsx* and *Xist* loci. In eutherians, repressive histone marks (i.e., H3K27me3) are associated with the repressed *Xist* locus and act in combination with *Tsix*, a *Xist* antisense transcript, and DNA methylation to repress its expression on the active X chromosome. In *M. domestica,* the repressive histone

modifications are absent. Instead, the histone modification profile of the *Rsx* locus resembles that of an escaper of pXCI inasmuch that its promoter is marked by the active histone modification, H3K4me3, but lacks the repressive histone modification, H3K27me3. This suggests that the repressive histone modifications characteristic of genes on the eutherian X chromosome do not mark the silenced, maternally derived *Rsx* locus in *M. domestica* and leaves open the possibility that DNA methylation is the only epigenetic regulator of *Rsx*.

6.3 Autosomal Imprinted Genes in *M. domestica*

6.3.1 Expression

The expression and epigenetic characteristics of autosomal imprinted loci of *M. domestica* differ in several important ways from those of paternally imprinted X-linked loci. First, the majority (~86%) of paternally imprinted X-linked loci show 100% expression from the maternal allele, whereas, all of the confirmed autosomal imprinted loci show a substantial level (>10%) of leaky (partial) expression from the repressed allele. Leaky expression of imprinted loci is not confined to metatherian-imprinted genes; it has also been documented from some imprinted genes in eutherians (Wang *et al.* 2011; Wang *et al.* 2013). The biologic (physiologic) relevance of leaky expression of the repressed allele at imprinted loci is not well understood and has sometimes been dismissed as an experimental artifact or attributed to a low level of random background transcription at individual loci with little to no biological importance. However, leaky expression was observed for all nine marsupial-specific imprinted genes described in this study and supports to the idea that leaky expression of the repressed allele is common, at

least in *M. domestica*, and the expression of the repressed allele might be biologically relevant. Future work is needed in both eutherian and metatherian model organisms to address this issue.

6.3.2 Histone States

The epigenetic characteristics of paternally imprinted X-linked loci of *M. domestica* show a different pattern than those of the autosomal imprinted loci. For X-linked loci, the presence of the examined repressive histone modification (H3K27me3) correlates positively with the transcriptionally repressed paternal allele, whereas the examined active histone modification (H3K4me3) correlates positively with the transcriptionally active maternal allele. Similarly, for autosomal loci active histone modifications are present at most imprinted gene promoters; however, the correlation of repressive histone modifications at the repressed alleles is not as strong as that found at X-linked loci. Only three of the nine confirmed, imprinted genes (*Meis1*, *Unknown_gene_1* and *Unknown_gene_6*) have histone marks of activation and repression occurring simultaneously at their annotated promoters. In addition, DNA methylation is virtually absent at all but two X-linked promoters but has been found at two of the four promoters of autosomal imprinted loci examined in this study, both of which have a histone mark of activation but lack a mark of repression.

6.3.3 Differential DNA Methylation

Historically, only one metatherian imprinted locus, *Peg10*, has been shown to have differential DNA methylation at promoter CpG islands and only two others, *Igf2* and *Igf2r*, have DMRs in other parts of the gene or region (Suzuki *et al.* 2007; Smits *et*

al. 2008; Das *et al.* 2012). The general paucity of DMRs, coupled with the general lack of histone modification data, led some to propose that transcriptionally opposing histone modifications might be the fundamental (and perhaps only) epigenetic mechanism employed by metatherians to regulate the expression of imprinted loci (Renfree *et al.* 2008; Samollow 2008). The results from my research indicate that active histone modifications do indeed mark the promoters of most, if not all, imprinted genes in *M. domestica*; but that, transcriptionally repressive histone modifications and DNA methylation are not required for imprinted expression. Nevertheless, my data also indicate that promoters of some metatherian-imprinted loci do show transcriptionally opposing histone modifications. These variable occurrences of epigenetic modifications at imprinted loci in *M. domestica* indicate that the diversity of regulatory mechanisms used to establish and maintain the imprinted state can be variable even within a single species and underscore the complexity of imprinting regulation across therian lineages.

6.3.4 Challenges and Opportunities

The discovery of novel marsupial-imprinted genes provides new opportunities for gaining additional insight into the biologic and evolutionary forces that have shaped the phenomenon of genomic imprinting in mammals. Mammalian imprinted genes were first discovered in eutherians and shown to have functions important for embryonic and fetal growth. These early discoveries led to the development of theories about why certain genes were imprinted (while most are not). One of these theories, The "Conflict" or "Kinship" Theory, emerged in the 1980s and 1990's as the prevailing model for selective advantage that could promote imprinting. It proposes that genomic imprinting

evolved in therian mammals (all of which form placental attachments to the mother) in response to polygamy, viviparity, and multiple births (Moore & Haig 1991; Haig 2004). It juxtaposes the reproductive strategies of males vs. females for maximizing reproductive success by noting that offspring of different fathers in multiple-paternity litters compete for the same maternal resources. Maximization of fitness for any one father is achieved by his progeny extracting maternal resources better than the progeny of other fathers, whereas for the mother, the best strategy is to provide resources equitably among all her offspring. This creates a "conflict" between the paternal and maternal genomes for genes that influence resource allocation. This parental conflict is manifested by imprinted genes, which are directly or indirectly involved in fetal growth, and is well exemplified by the interplay of *IGF2* and *IGF2R,* which are involved in fetal vs. placental growth and development and are reciprocally imprinted; *IGF2* being expressed from the paternal allele only (maternally imprinted), whereas *IGF2R* is expressed from the maternal allele only (paternally imprinted).

The Conflict model fits well with known functions of some imprinted genes (e.g., *IGF2* and *IGF2R*) and has also been extended to imprinted genes that affect cognition, feeding behavior, and social development; however, this model does not explain all imprinted genes. In fact, the Conflict model has not provided explanatory power for the parent-of-origin specific expression pattern of most imprinted genes, but in view of the logical power and broad acceptance of the Conflict model, few alternative hypotheses as to why these genes are imprinted have been proposed (the few that have been proposed were quickly dismissed as evolutionarily unstable or logically flawed). It is possible that

some imprinted genes are simply evolutionary "casualties" of their genomic location (close proximity to other selected imprinted genes). While they themselves might not be important for fetal growth and/or development, their physical proximity to other selected imprinted genes or regions, along with an ability to compensate for monoallelic expression by upregulation of the active allele, might be the driving force behind their imprinted state. Alternatively, there could be solid but as yet unidentified biological advantages for the imprinting of genes that are not involved in fetal growth and development. In any case, lacking evidence that Conflict Theory can explain all imprinting, it seems prudent to remain open to and actively seek alternative hypotheses.

Following upon this theme, I have described 10 imprinted genes in *M. domestica* (nine novel genes and *Igf2r*) of which seven are protein coding. An analysis of the functions associated with these protein-coding genes indicates involvement in cell proliferation (*Meis1*), fetal growth (*Igf2r*), neurological development (*Npdc1* and *Rwdd2a*), regulation of immunity (*Nkrf*), and homeobox genes/transcription factors that could potentially control the transcription of many genes at different times throughout development (*Pou5fl* and *Zfp68*). However, only two of these genes, *Meis1* and *Igf2r* have an obvious and direct connection to fetal or cell growth thus can be explained by the Conflict Theory. My data indicate that, as in eutherians, the imprinting phenomenon in marsupials is complex, and although the imprinted state of some genes is conserved between eutherians and metatherians, some genes have also been uniquely selected for imprinted expression in *M. domestica* or possibly across the metatherian lineage in general. In establishing a novel method for identifying imprinted genes using ChIP-seq

and RNA-seq approaches in *M. domestica*, I have provided a new way to discover imprinted genes in non-eutherian mammals that could be applied to non-placental species, including non-mammalian vertebrates as well. The discovery of imprinted genes in a broader range of vertebrate taxa (and perhaps beyond) could provide novel insights into the phenomenon of genomic imprinting and help to generate alternative hypotheses for its origin and evolutionary advantages.

6.4 Conclusion

In conclusion, my dissertation research has produced the first genome-wide analysis of four histone modifications in fibroblast cells and female fetal brain, and their correlation with the transcriptional states of autosomal and X-linked genes. It has also provided new data concerning the imprinted expression of genes in fibroblasts, fetal brain and EEM of *M. domestica* from autosomal genes and the paternally imprinted X chromosome, the imprinted state of which is specific to marsupials. These data have allowed me to describe comprehensively, and for the first time, the phenomena of X-chromosome inactivation and genomic imprinting in *M. domestica*, provided new insights into the possible biological and evolutionary forces that established and maintain these phenomena, and generated new hypotheses for future testing.

REFERENCES

Ager E., Suzuki S., Pask A., Shaw G., Ishino F. & Renfree M.B. (2007) Insulin is imprinted in the placenta of the marsupial, Macropus eugenii. *Dev Biol* **309**, 317-28.

Ager E.I., Pask A.J., Gehring H.M., Shaw G. & Renfree M.B. (2008a) Evolution of the CDKN1C-KCNQ1 imprinted domain. *BMC Evol Biol* **8**, 163.

Ager E.I., Pask A.J., Shaw G. & Renfree M.B. (2008b) Expression and protein localisation of IGF2 in the marsupial placenta. *BMC Dev Biol* **8**, 17.

Alleman M. & Doctor J. (2000) Genomic imprinting in plants: observations and evolutionary implications. *Plant Mol Biol* **43**, 147-61.

Amor D.J. & Halliday J. (2008) A review of known imprinting syndromes and their association with assisted reproduction technologies. *Hum Reprod* **23**, 2826-34.

Anaka M., Lynn A., McGinn P. & Lloyd V.K. (2009) Genomic imprinting in Drosophila has properties of both mammalian and insect imprinting. *Dev Genes Evol* **219**, 59-66.

Anwar S.L., Krech T., Hasemeier B., Schipper E., Schweitzer N., Vogel A., Kreipe H. & Lehmann U. (2012) Loss of imprinting and allelic switching at the DLK1-MEG3 locus in human hepatocellular carcinoma. *PLoS One* **7**, e49462.

Arney K.L. (2003) H19 and Igf2--enhancing the confusion? *Trends Genet* **19**, 17-23.

Ball M.P., Li J.B., Gao Y., Lee J.H., LeProust E.M., Park I.H., Xie B., Daley G.Q. & Church G.M. (2009) Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells. *Nat Biotechnol* **27**, 361-8.

Barlow D.P. (2011) Genomic imprinting: a mammalian epigenetic discovery model. *Annu Rev Genet* **45**, 379-403.

Bartolomei M.S., Zemel S. & Tilghman S.M. (1991) Parental imprinting of the mouse H19 gene. *Nature* **351**, 153-5.

Bean C.J., Schaner C.E. & Kelly W.G. (2004) Meiotic pairing and imprinted X chromatin assembly in Caenorhabditis elegans. *Nat Genet* **36**, 100-5.

Bebbere D., Bauersachs S., Furst R.W., Reichenbach H.D., Reichenbach M., Medugorac I., Ulbrich S.E., Wolf E., Ledda S. & Hiendleder S. (2013) Tissue-specific and minor inter-individual variation in imprinting of IGF2R is a common feature of Bos taurus Concepti and not correlated with fetal weight. *PLoS One* **8**, e59564.

Bell A.C. & Felsenfeld G. (2000) Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. *Nature* **405**, 482-5.

Bernstein B.E., Meissner A. & Lander E.S. (2007) The mammalian epigenome. *Cell* **128**, 669-81.

Bestor T.H. (2000) The DNA methyltransferases of mammals. *Hum Mol Genet* **9**, 2395-402.

Blankenberg D., Von Kuster G., Coraor N., Ananda G., Lazarus R., Mangan M., Nekrutenko A. & Taylor J. (2010) Galaxy: a web-based genome analysis tool for experimentalists. *Curr Protoc Mol Biol* **Chapter 19**, Unit 19 0 1-21.

Carbe C., Hertzler-Schaefer K. & Zhang X. (2012) The functional role of the Meis/Prep-binding elements in Pax6 locus during pancreas and eye development. *Dev Biol* **363**, 320-9.

Carrel L. & Willard H.F. (2005) X-inactivation profile reveals extensive variability in X-linked gene expression in females. *Nature* **434**, 400-4.

Carvalho-Silva D.R., O'Neill R.J., Brown J.D., Huynh K., Waters P.D., Pask A.J., Delbridge M.L. & Graves J.A. (2004) Molecular characterization and evolution of X and Y-borne ATRX homologues in American marsupials. *Chromosome Res* **12**, 795-804.

Chaumeil J., Waters P.D., Koina E., Gilbert C., Robinson T.J. & Graves J.A. (2011) Evolution from XIST-independent to XIST-controlled X-chromosome inactivation: epigenetic modifications in distantly related mammals. *PLoS One* **6**, e19040.

Choe S.K., Lu P., Nakamura M., Lee J. & Sagerstrom C.G. (2009) Meis cofactors control HDAC and CBP accessibility at Hox-regulated promoters during zebrafish embryogenesis. *Dev Cell* **17**, 561-7.

Connallon T. & Clark A.G. (2013) Sex-specific selection and the evolution of X inactivation strategies. *PLoS Genetics* **In Press**.

Consortium E.P., Dunham I., Kundaje A., Aldred S.F., Collins P.J., Davis C.A. & *al. e.* (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57-74.

Cooper D.W., VandeBerg J.L., Sharman G.B. & Poole W.E. (1971) Phosphoglycerate kinase polymorphism in kangaroos provides further evidence for paternal X inactivation. *Nat New Biol* **230**, 155-7.

Cooper D.W., Johnston P.G., VandeBerg J.L. & Robinson E.S. (1990) X-chromosome inactivation in marsupials. In: *Mammals From Pouches and Eggs: Genetics, Breeding and Evolution of Marsupials and Monotremes* (eds. by Graves JAM, Hope R & Cooper DW), pp. 269-75. CSIRO, Melbourne, Australia.

Cooper D.W., Johnston P.G., Watson J.M. & Graves J.A.M. (1993) X-inactivation in marsupials and monotremes. *Semin. Dev. Biol* **4**, 117-28.

Das R., Anderson N., Koran M.I., Weidman J.R., Mikkelsen T.S., Kamal M., Murphy S.K., Linblad-Toh K., Greally J.M. & Jirtle R.L. (2012) Convergent and divergent evolution of genomic imprinting in the marsupial Monodelphis domestica. *BMC Genomics* **13**, 394.

Davidow L.S., Breen M., Duke S.E., Samollow P.B., McCarrey J.R. & Lee J.T. (2007) The search for a marsupial XIC reveals a break with vertebrate synteny. *Chromosome Res* **15**, 137-46.

De Luca V., Likhodi O., Kennedy J.L. & Wong A.H. (2007) Parent-of-origin effect and genomic imprinting of the HTR2A receptor gene T102C polymorphism in psychosis. *Psychiatry Res* **151**, 243-8.

Deakin J.E., Chaumeil J., Hore T.A. & Marshall Graves J.A. (2009) Unravelling the evolutionary origins of X chromosome inactivation in mammals: insights from marsupials and monotremes. *Chromosome Res* **17**, 671-85.

DeChiara T.M., Robertson E.J. & Efstratiadis A. (1991) Parental imprinting of the mouse insulin-like growth factor II gene. *Cell* **64**, 849-59.

Delaval K., Govin J., Cerqueira F., Rousseaux S., Khochbin S. & Feil R. (2007) Differential histone modifications mark mouse imprinting control regions during spermatogenesis. *EMBO J* **26**, 720-9.

Dindot S.V., Kent K.C., Evers B., Loskutoff N., Womack J. & Piedrahita J.A. (2004) Conservation of genomic imprinting at the XIST, IGF2, and GTL2 loci in the bovine. *Mamm Genome* **15**, 966-74.

Dindot S.V., Antalffy B.A., Bhattacharjee M.B. & Beaudet A.L. (2008) The Angelman syndrome ubiquitin ligase localizes to the synapse and nucleus, and maternal deficiency results in abnormal dendritic spine morphology. *Hum Mol Genet* **17**, 111-8.

Dindot S.V., Person R., Strivens M., Garcia R. & Beaudet A.L. (2009) Epigenetic profiling at mouse imprinted gene clusters reveals novel epigenetic and genetic features at differentially methylated regions. *Genome Res* **19**, 1374-83.

Disteche C.M., Filippova G.N. & Tsuchiya K.D. (2002) Escape from X inactivation. *Cytogenet .Genome Res.* **99**, 36-43.

Duret L., Chureau C., Samain S., Weissenbach J. & Avner P. (2006) The Xist RNA gene evolved in eutherians by pseudogenization of a protein-coding gene. *Science* **312**, 1653-5.

Edwards C.A., Mungall A.J., Matthews L., Ryder E., Gray D.J., Pask A.J., Shaw G., Graves J.A., Rogers J., Dunham I., Renfree M.B. & Ferguson-Smith A.C. (2008) The evolution of the DLK1-DIO3 imprinted domain in mammals. *PLoS Biol* **6**, e135.

Ernst J. & Kellis M. (2010) Discovery and characterization of chromatin states for systematic annotation of the human genome. *Nat Biotechnol* **28**, 817-25.

Evans H.K., Weidman J.R., Cowley D.O. & Jirtle R.L. (2005) Comparative phylogenetic analysis of blcap/nnat reveals eutherian-specific imprinted gene. *Mol Biol Evol* **22**, 1740-8.

Fatemi M., Pao M.M., Jeong S., Gal-Yam E.N., Egger G., Weisenberger D.J. & Jones P.A. (2005) Footprinting of mammalian promoters: use of a CpG DNA methyltransferase revealing nucleosome positions at a single molecule level. *Nucleic Acids Res* **33**, e176.

Feil R., Walter J., Allen N.D. & Reik W. (1994) Developmental control of allelic methylation in the imprinted mouse Igf2 and H19 genes. *Development* **120**, 2933-43.

Feil R. & Berger F. (2007) Convergent evolution of genomic imprinting in plants and mammals. *Trends Genet* **23**, 192-9.

Ferguson-Smith A.C. (2011) Genomic imprinting: the emergence of an epigenetic paradigm. *Nat Rev Genet* **12**, 565-75.

Fournier C., Goto Y., Ballestar E., Delaval K., Hever A.M., Esteller M. & Feil R. (2002) Allele-specific histone lysine methylation marks regulatory regions at imprinted mouse genes. *EMBO J* **21**, 6560-70.

Giardine B., Riemer C., Hardison R.C., Burhans R., Elnitski L., Shah P., Zhang Y., Blankenberg D., Albert I., Taylor J., Miller W., Kent W.J. & Nekrutenko A. (2005) Galaxy: a platform for interactive large-scale genome analysis. *Genome Res* **15**, 1451-5.

Gifford C.A., Ziller M.J., Gu H., Trapnell C., Donaghey J., Tsankov A., Shalek A.K., Kelley D.R., Shishkin A.A., Issner R., Zhang X., Coyne M., Fostel J.L., Holmes L., Meldrim J., Guttman M., Epstein C., Park H., Kohlbacher O., Rinn J., Gnirke A., Lander E.S., Bernstein B.E. & Meissner A. (2013) Transcriptional and Epigenetic Dynamics during Specification of Human Embryonic Stem Cells. *Cell* **153**, 1149-63.

Goecks J., Nekrutenko A. & Taylor J. (2010) Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol* **11**, R86.

Goodstadt L., Heger A., Webber C. & Ponting C.P. (2007) An analysis of the gene complement of a marsupial, Monodelphis domestica: evolution of lineage-specific genes and giant chromosomes. *Genome Res* **17**, 969-81.

Grant J., Mahadevaiah S.K., Khil P., Sangrithi M.N., Royo H., Duckworth J., McCarrey J.R., VandeBerg J.L., Renfree M.B., Taylor W., Elgar G., Camerini-Otero R.D., Gilchrist M.J. & Turner J.M. (2012) Rsx is a metatherian RNA with Xist-like properties in X-chromosome inactivation. *Nature* **487**, 254-8.

Greally J.M. (2002) Short interspersed transposable elements (SINEs) are excluded from imprinted regions in the human genome. *Proc Natl Acad Sci U S A* **99**, 327-32.

Grewal S.I. & Elgin S.C. (2007) Transcription and RNA interference in the formation of heterochromatin. *Nature* **447**, 399-406.

Guenther M.G., Levine S.S., Boyer L.A., Jaenisch R. & Young R.A. (2007) A chromatin landmark and transcription initiation at most promoters in human cells. *Cell* **130**, 77-88.

Haig D. (2004) Genomic imprinting and kinship: how good is the evidence? *Annu Rev Genet* **38**, 553-85.

Hamed M., Ismael S., Paulsen M. & Helms V. (2012) Cellular functions of genetically imprinted genes in human and mouse as annotated in the gene ontology. *PLoS One* **7**, e50285.

Hark A.T., Schoenherr C.J., Katz D.J., Ingram R.S., Levorse J.M. & Tilghman S.M. (2000) CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus. *Nature* **405**, 486-9.

Heard E., Clerc P. & Avner P. (1997) X-chromosome inactivation in mammals. *Annu Rev Genet* **31**, 571-610.

Heard E. & Disteche C.M. (2006) Dosage compensation in mammals: fine-tuning the expression of the X chromosome. *Genes Dev* **20**, 1848-67.

Hellman A. & Chess A. (2007) Gene body-specific methylation on the active X chromosome. *Science* **315**, 1141-3.

Hinrichs K., Choi Y.H., Love C.C., Chung Y.G. & Varner D.D. (2006) Production of horse foals via direct injection of roscovitine-treated donor cells and activation by injection of sperm extract. *Reproduction* **131**, 1063-72.

Hore T.A., Koina E., Wakefield M.J. & Marshall Graves J.A. (2007a) The region homologous to the X-chromosome inactivation centre has been disrupted in marsupial and monotreme mammals. *Chromosome Res* **15**, 147-61.

Hore T.A., Rapkins R.W. & Graves J.A. (2007b) Construction and evolution of imprinted loci in mammals. *Trends Genet* **23**, 440-8.

Hornecker J.L., Samollow P.B., Robinson E.S., VandeBerg J.L. & McCarrey J.R. (2007) Meiotic sex chromosome Inactivation in the marsupial *Monodelphis domestica*. *Genesis* **45**, 696-708.

Huynh K.D. & Lee J.T. (2001) Imprinted X inactivation in eutherians: a model of gametic execution and zygotic relaxation. *Curr Opin Cell Biol* **13**, 690-7.

Huynh K.D. & Lee J.T. (2005) X-chromosome inactivation: a hypothesis linking ontogeny and phylogeny. *Nat Rev Genet* **6**, 410-8.

Jegalian K. & Page D.C. (1998) A proposed path by which genes common to mammalian X and Y chromosomes evolve to become X inactivated. *Nature* **394**, 776-80.

Kammerer C.M., Rainwater D.L., Gouin N., Jasti M., Douglas K.C., Dressen A.S., Ganta P., Vandeberg J.L. & Samollow P.B. (2010) Localization of genes for V+LDL plasma cholesterol levels on two diets in the opossum Monodelphis domestica. *J Lipid Res* **51**, 2929-39.

Kaslow D.C. & Migeon B.R. (1987) DNA methylation stabilizes X chromosome inactivation in eutherians but not in marsupials: evidence for multistep maintenance of mammalian X dosage compensation. *Proc Natl Acad Sci U S A* **84**, 6210-4.

Kent W.J., Sugnet C.W., Furey T.S., Roskin K.M., Pringle T.H., Zahler A.M. & Haussler D. (2002) The human genome browser at UCSC. *Genome Res* **12**, 996-1006.

Keohane A.M., Lavender J.S., O'Neill L.P. & Turner B.M. (1998) Histone acetylation and X inactivation. *Dev Genet* **22**, 65-73.

Kharchenko P.V., Alekseyenko A.A., Schwartz Y.B., Minoda A., Riddle N.C., Ernst J., Sabo P.J., Larschan E., Gorchakov A.A., Gu T., Linder-Basso D., Plachetka A., Shanower G., Tolstorukov M.Y., Luquette L.J., Xi R., Jung Y.L., Park R.W., Bishop E.P., Canfield T.K., Sandstrom R., Thurman R.E., MacAlpine D.M., Stamatoyannopoulos J.A., Kellis M., Elgin S.C., Kuroda M.I., Pirrotta V., Karpen G.H. & Park P.J. (2011) Comprehensive analysis of the chromatin landscape in Drosophila melanogaster. *Nature* **471**, 480-5.

Khosla S., Mendiratta G. & Brahmachari V. (2006) Genomic imprinting in the mealybugs. *Cytogenet Genome Res* **113**, 41-52.

Killian J.K., Byrd J.C., Jirtle J.V., Munday B.L., Stoskopf M.K., MacDonald R.G. & Jirtle R.L. (2000) M6P/IGF2R imprinting evolution in mammals. *Mol Cell* **5**, 707-16.

Kobayashi S., Kohda T., Miyoshi N., Kuroiwa Y., Aisaka K., Tsutsumi O., Kaneko-Ishino T. & Ishino F. (1997) Human PEG1/MEST, an imprinted gene on chromosome 7. *Hum Mol Genet* **6**, 781-6.

Koina E., Wakefield M.J., Walcher C., Disteche C.M., Whitehead S., Ross M. & Marshall Graves J.A. (2005) Isolation, X location and activity of the marsupial homologue of SLC16A2, an XIST-flanking gene in eutherian mammals. *Chromosome Res* **13**, 687-98.

Koina E., Chaumeil J., Greaves I.K., Tremethick D.J. & Graves J.A. (2009) Specific patterns of histone marks accompany X chromosome inactivation in a marsupial. *Chromosome Res* **17**, 115-26.

Kosaki K., Kosaki R., Craigen W.J. & Matsuo N. (2000) Isoform-specific imprinting of the human PEG1/MEST gene. *Am J Hum Genet* **66**, 309-12.

Kouzarides T. (2007) Chromatin modifications and their function. *Cell* **128**, 693-705.

Kumar S. & Hedges S.B. (1998) A molecular timescale for vertebrate evolution. *Nature* **392**, 917-20.

Latham K.E. (2005) X chromosome imprinting and inactivation in preimplantation mammalian embryos. *Trends Genet* **21**, 120-7.

Lawson H.A., Cheverud J.M. & Wolf J.B. (2013) Genomic imprinting and parent-of-origin effects on complex traits. *Nat Rev Genet* **14**, 609-17.

Lawton B.R., Carone B.R., Obergfell C.J., Ferreri G.C., Gondolphi C.M., Vandeberg J.L., Imumorin I., O'Neill R.J. & O'Neill M.J. (2008) Genomic imprinting of IGF2 in marsupials is methylation dependent. *BMC Genomics* **9**, 205.

Li E., Bestor T.H. & Jaenisch R. (1992) Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* **69**, 915-26.

Li H., Handsaker B., Wysoker A., Fennell T., Ruan J., Homer N., Marth G., Abecasis G. & Durbin R. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-9.

Li J., Bench A.J., Vassiliou G.S., Fourouclas N., Ferguson-Smith A.C. & Green A.R. (2004) Imprinting of the human L3MBTL gene, a polycomb family member located in a region of chromosome 20 deleted in human myeloid malignancies. *Proc Natl Acad Sci U S A* **101**, 7341-6.

Lippman Z. & Martienssen R. (2004) The role of RNA interference in heterochromatic silencing. *Nature* **431**, 364-70.

Lock L.F., Melton D.W., Caskey C.T. & Martin G.R. (1986) Methylation of the mouse hprt gene differs on the active and inactive X chromosomes. *Mol Cell Biol* **6**, 914-24.

Lock L.F., Takagi N. & Martin G.R. (1987) Methylation of the Hprt gene on the inactive X occurs after chromosome inactivation. *Cell* **48**, 39-46.

Loebel D.A. & Johnston P.G. (1993) Analysis of DNase 1 sensitivity and methylation of active and inactive X chromosomes of kangaroos (Macropus robustus) by in situ nick translation. *Chromosoma* **102**, 81-7.

Loebel D.A. & Johnston P.G. (1996) Methylation analysis of a marsupial X-linked CpG island by bisulfite genomic sequencing. *Genome Res* **6**, 114-23.

Lopes S., Lewis A., Hajkova P., Dean W., Oswald J., Forne T., Murrell A., Constancia M., Bartolomei M., Walter J. & Reik W. (2003) Epigenetic modifications in an imprinting cluster are controlled by a hierarchy of DMRs suggesting long-range chromatin interactions. *Hum Mol Genet* **12**, 295-305.

Luedi P.P., Hartemink A.J. & Jirtle R.L. (2005) Genome-wide prediction of imprinted murine genes. *Genome Res* **15**, 875-84.

Luedi P.P., Dietrich F.S., Weidman J.R., Bosko J.M., Jirtle R.L. & Hartemink A.J. (2007) Computational and experimental identification of novel human imprinted genes. *Genome Res* **17**, 1723-30.

Lyko F., Brenton J.D., Surani M.A. & Paro R. (1997) An imprinting element from the mouse H19 locus functions as a silencer in Drosophila. *Nat Genet* **16**, 171-3.

Lyko F., Buiting K., Horsthemke B. & Paro R. (1998) Identification of a silencing element in the human 15q11-q13 imprinting center by using transgenic Drosophila. *Proc Natl Acad Sci U S A* **95**, 1698-702.

Lyon M.F. (1961) Gene action in the X-chromosome of the mouse (Mus musculus L.). *Nature* **190**, 372-3.

Mahadevaiah S.K., Royo H., VandeBerg J.L., McCarrey J.R., Mackay S. & Turner J.M. (2009) Key features of the X inactivation process are conserved between marsupials and eutherians. *Curr Biol* **19**, 1478-84.

Mahmoud A.I., Kocabas F., Muralidhar S.A., Kimura W., Koura A.S., Thet S., Porrello E.R. & Sadek H.A. (2013) Meis1 regulates postnatal cardiomyocyte cell cycle arrest. *Nature* **497**, 249-53.

Martin C.C. & McGowan R. (1995) Genotype-specific modifiers of transgene methylation and expression in the zebrafish, Danio rerio. *Genet Res* **65**, 21-8.

Mate K.E., Robinson E.S., Vandeberg J.L. & Pedersen R.A. (1994) Timetable of in vivo embryonic development in the grey short-tailed opossum (Monodelphis domestica). *Mol Reprod Dev* **39**, 365-74.

Meredith R.W., Janecka J.E., Gatesy J., Ryder O.A., Fisher C.A., Teeling E.C., Goodbla A., Eizirik E., Simao T.L., Stadler T., Rabosky D.L., Honeycutt R.L., Flynn J.J., Ingram C.M., Steiner C., Williams T.L., Robinson T.J., Burk-Herrick A., Westerman M., Ayoub N.A., Springer M.S. & Murphy W.J. (2011) Impacts of the Cretaceous Terrestrial Revolution and KPg extinction on mammal diversification. *Science* **334**, 521-4.

Migeon B.R., Wolf S.F., Axelman J., Kaslow D.C. & Schmidt M. (1985) Incomplete X chromosome dosage compensation in chorionic villi of human placenta. *Proc Natl Acad Sci U S A* **82**, 3390-4.

Mikkelsen T.S., Ku M., Jaffe D.B., Issac B., Lieberman E., Giannoukos G., Alvarez P., Brockman W., Kim T.K., Koche R.P., Lee W., Mendenhall E., O'Donovan A., Presser A., Russ C., Xie X., Meissner A., Wernig M., Jaenisch R., Nusbaum C., Lander E.S. & Bernstein B.E. (2007a) Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448**, 553-60.

Mikkelsen T.S., Wakefield M.J., Aken B., Amemiya C.T., Chang J.L., Duke S., Garber M., Gentles A.J., Goodstadt L., Heger A., Jurka J., Kamal M., Mauceli E., Searle S.M., Sharpe T., Baker M.L., Batzer M.A., Benos P.V., Belov K., Clamp M., Cook A., Cuff J., Das R., Davidow L., Deakin J.E., Fazzari M.J., Glass J.L., Grabherr M., Greally J.M., Gu W., Hore T.A., Huttley G.A., Kleber M., Jirtle R.L., Koina E., Lee J.T., Mahony S., Marra M.A., Miller R.D., Nicholls R.D., Oda M., Papenfuss A.T., Parra Z.E., Pollock D.D., Ray D.A., Schein J.E., Speed T.P., Thompson K., VandeBerg J.L., Wade C.M., Walker J.A., Waters P.D., Webber C., Weidman J.R., Xie X., Zody M.C., Graves J.A., Ponting C.P., Breen M., Samollow P.B., Lander E.S. & Lindblad-Toh K. (2007b) Genome of the marsupial Monodelphis domestica reveals innovation in non-coding sequences. *Nature* **447**, 167-77.

Moens C.B. & Selleri L. (2006) Hox cofactors in vertebrate development. *Dev Biol* **291**, 193-206.

Monk D., Arnaud P., Apostolidou S., Hills F.A., Kelsey G., Stanier P., Feil R. & Moore G.E. (2006) Limited evolutionary conservation of imprinting in the human placenta. *Proc Natl Acad Sci U S A* **103**, 6623-8.

Moore T. & Haig D. (1991) Genomic imprinting in mammalian development: a parental tug-of-war. *Trends Genet* **7**, 45-9.

Moreira de Mello J.C., de Araujo E.S., Stabellini R., Fraga A.M., de Souza J.E., Sumita D.R., Camargo A.A. & Pereira L.V. (2010) Random X inactivation and extensive mosaicism in human placenta revealed by analysis of allele-specific gene expression along the X chromosome. *PLoS One* **5**, e10947.

Morison I.M., Ramsay J.P. & Spencer H.G. (2005) A census of mammalian imprinting. *Trends Genet* **21**, 457-65.

Moustakas J.E., Smith K.K. & Hlusko L.J. (2011) Evolution and development of the mammalian dentition: insights from the marsupial Monodelphis domestica. *Dev Dyn* **240**, 232-9.

Murphy S.K. & Jirtle R.L. (2003) Imprinting evolution and the price of silence. *Bioessays* **25**, 577-88.

Murtagh V.J., O'Meally D., Sankovic N., Delbridge M.L., Kuroki Y., Boore J.L., Toyoda A., Jordan K.S., Pask A.J., Renfree M.B., Fujiyama A., Graves J.A. & Waters P.D. (2012) Evolutionary history of novel genes on the tammar wallaby Y chromosome: Implications for sex chromosome evolution. *Genome Res* **22**, 498-507.

Noor N.M., Mollgard K., Wheaton B.J., Steer D.L., Truettner J.S., Dziegielewska K.M., Dietrich W.D., Smith A.I. & Saunders N.R. (2013) Expression and Cellular Distribution of Ubiquitin in Response to Injury in the Developing Spinal Cord of Monodelphis domestica. *PLoS One* **8**, e62120.

O'Sullivan F.M., Murphy S.K., Simel L.R., McCann A., Callanan J.J. & Nolan C.M. (2007) Imprinted expression of the canine IGF2R, in the absence of an anti-sense transcript or promoter methylation. *Evol Dev* **9**, 579-89.

Ohno S. (1967) *Sex chromosomes and sex-linked genes*. Springer-Verlag, Berlin, New York etc.

Okamoto I. & Heard E. (2006) The dynamics of imprinted X inactivation during preimplantation development in mice. *Cytogenet Genome Res* **113**, 318-24.

Okano M., Bell D.W., Haber D.A. & Li E. (1999) DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* **99**, 247-57.

Payer B. & Lee J.T. (2008) X chromosome dosage compensation: how mammals keep the balance. *Annu Rev Genet* **42**, 733-72.

Pradhan S., Bacolla A., Wells R.D. & Roberts R.J. (1999) Recombinant human DNA (cytosine-5) methyltransferase. I. Expression, purification, and comparison of de novo and maintenance methylation. *J Biol Chem* **274**, 33002-10.

Quinlan A.R. & Hall I.M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-2.

Rapkins R.W., Hore T., Smithwick M., Ager E., Pask A.J., Renfree M.B., Kohn M., Hameister H., Nicholls R.D., Deakin J.E. & Graves J.A. (2006) Recent assembly of an imprinted domain from non-imprinted components. *PLoS Genet* **2**, e182.

Regha K., Sloane M.A., Huang R., Pauler F.M., Warczok K.E., Melikant B., Radolf M., Martens J.H., Schotta G., Jenuwein T. & Barlow D.P. (2007) Active and repressive chromatin are interspersed without spreading in an imprinted gene cluster in the mammalian genome. *Mol Cell* **27**, 353-66.

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

Renfree M.B., Ager E.I., Shaw G. & Pask A.J. (2008) Genomic imprinting in marsupial placentation. *Reproduction* **136**, 523-31.

Rens W., Wallduck M.S., Lovell F.L., Ferguson-Smith M.A. & Ferguson-Smith A.C. (2010) Epigenetic modifications on X chromosomes in marsupial and monotreme mammals and implications for evolution of dosage compensation. *Proc Natl Acad Sci U S A* **107**, 17657-62.

Riggs A.D. (1990) Marsupials and Mechanisms of X-Chromosome Inactivation. *Australian Journal of Zoology* **37**, 419-41.

Rivera M.N. & Haber D.A. (2005) Wilms' tumour: connecting tumorigenesis and organ development in the kidney. *Nat Rev Cancer* **5**, 699-712.

Robinson J.T., Thorvaldsdottir H., Winckler W., Guttman M., Lander E.S., Getz G. & Mesirov J.P. (2011) Integrative genomics viewer. *Nat Biotechnol* **29**, 24-6.

Rozen S. & Skaletsky H. (2000) Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol* **132**, 365-86.

Samollow P.B., Ford A.L. & VandeBerg J.L. (1987) X-linked gene expression in the Virginia opossum: differences between the paternally derived Gpd and Pgk-A loci. *Genetics* **115**, 185-95.

Samollow P.B., Robinson E.S., Ford A.L. & Vandeberg J.L. (1995) Developmental progression of Gpd expression from the inactive X chromosome of the Virginia opossum. *Dev Genet* **16**, 367-78.

Samollow P.B. (2006) Status and applications of genome resources for the gray, short-tailed opossum, *Monodelphis domestica*, an American marsupial model for comparative biology. *Australian Journal of Zoology* **54**, 173-96.

Samollow P.B. (2008) The opossum genome: insights and opportunities from an alternative mammal. *Genome Res* **18**, 1199-215.

Sanchez-Guardado L.O., Irimia M., Sanchez-Arrones L., Burguera D., Rodriguez-Gallardo L., Garcia-Fernandez J., Puelles L., Ferran J.L. & Hidalgo-Sanchez M. (2011) Distinct and redundant expression and transcriptional diversity of MEIS gene paralogs during chicken development. *Dev Dyn* **240**, 1475-92.

Saxonov S., Berg P. & Brutlag D.L. (2006) A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proc Natl Acad Sci U S A* **103**, 1412-7.

Selwood L. & Johnson M.H. (2006) Trophoblast and hypoblast in the monotreme, marsupial and eutherian mammal: evolution and origins. *Bioessays* **28**, 128-45.

Sha K. (2008) A mechanistic view of genomic imprinting. *Annu Rev Genomics Hum Genet* **9**, 197-216.

Shemer R., Birger Y., Riggs A.D. & Razin A. (1997) Structure of the imprinted mouse Snrpn gene and establishment of its parental-specific methylation pattern. *Proc Natl Acad Sci U S A* **94**, 10267-72.

Shen C.J., Cheng W.T., Wu S.C., Chen H.L., Tsai T.C., Yang S.H. & Chen C.M. (2012) Differential differences in methylation status of putative imprinted genes among cloned swine genomes. *PLoS One* **7**, e32812.

Shevchenko A.I., Zakharova I.S., Elisaphenko E.A., Kolesnikov N.N., Whitehead S., Bird C., Ross M., Weidman J.R., Jirtle R.L., Karamysheva T.V., Rubtsov N.B., VandeBerg J.L., Mazurok N.A., Nesterova T.B., Brockdorff N. & Zakian S.M. (2007) Genes flanking Xist in mouse and human are separated on the X chromosome in American marsupials. *Chromosome Res* **15**, 127-36.

Silva J., Mak W., Zvetkova I., Appanah R., Nesterova T.B., Webster Z., Peters A.H., Jenuwein T., Otte A.P. & Brockdorff N. (2003) Establishment of histone h3 methylation on the inactive X chromosome requires transient recruitment of Eed-Enx1 polycomb group complexes. *Dev Cell* **4**, 481-95.

Sleutels F., Zwart R. & Barlow D.P. (2002) The non-coding Air RNA is required for silencing autosomal imprinted genes. *Nature* **415**, 810-3.

Sleutels F., Tjon G., Ludwig T. & Barlow D.P. (2003) Imprinted silencing of Slc22a2 and Slc22a3 does not need transcriptional overlap between Igf2r and Air. *EMBO J* **22**, 3696-704.

Smith K.K. (2001) Early development of the neural plate, neural crest and facial region of marsupials. *J Anat* **199**, 121-31.

Smits G., Mungall A.J., Griffiths-Jones S., Smith P., Beury D., Matthews L., Rogers J., Pask A.J., Shaw G., VandeBerg J.L., McCarrey J.R., Renfree M.B., Reik W. & Dunham I. (2008) Conservation of the H19 noncoding RNA and H19-IGF2 imprinting mechanism in therians. *Nat Genet* **40**, 971-6.

Srivastava M., Hsieh S., Grinberg A., Williams-Simons L., Huang S.P. & Pfeifer K. (2000) H19 and Igf2 monoallelic expression is regulated in two distinct ways by a shared cis acting regulatory region upstream of H19. *Genes Dev* **14**, 1186-95.

Strahl B.D. & Allis C.D. (2000) The language of covalent histone modifications. *Nature* **403**, 41-5.

Straub T. & Becker P.B. (2007) Dosage compensation: the beginning and end of generalization. *Nat Rev Genet* **8**, 47-57.

Stringer J.M., Suzuki S., Pask A.J., Shaw G. & Renfree M.B. (2012) GRB10 imprinting is eutherian mammal specific. *Mol Biol Evol* **29**, 3711-9.

Surani M.A., Barton S.C. & Norris M.L. (1984) Development of reconstituted mouse eggs suggests imprinting of the genome during gametogenesis. *Nature* **308**, 548-50.

Suzuki J., Jr., Therrien J., Filion F., Lefebvre R., Goff A.K. & Smith L.C. (2009) In vitro culture and somatic cell nuclear transfer affect imprinting of SNRPN gene in pre- and post-implantation stages of development in cattle. *BMC Dev Biol* **9**, 9.

Suzuki S., Renfree M.B., Pask A.J., Shaw G., Kobayashi S., Kohda T., Kaneko-Ishino T. & Ishino F. (2005) Genomic imprinting of IGF2, p57(KIP2) and PEG1/MEST in a marsupial, the tammar wallaby. *Mech Dev* **122**, 213-22.

Suzuki S., Ono R., Narita T., Pask A.J., Shaw G., Wang C., Kohda T., Alsop A.E., Marshall Graves J.A., Kohara Y., Ishino F., Renfree M.B. & Kaneko-Ishino T. (2007) Retrotransposon silencing by DNA methylation can drive mammalian genomic imprinting. *PLoS Genet* **3**, e55.

Thompson J.R. & Williams C.J. (2005) Genomic imprinting and assisted reproductive technology: connections and potential risks. *Semin Reprod Med* **23**, 285-95.

Thorvaldsdottir H., Robinson J.T. & Mesirov J.P. (2012) Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform*.

Tourte Y., Kuligowski-Andres J. & Barbier-Ramond C. (1980) [Different behaviour of paternal and maternal genomes during embryogenesis in the fern, Marsilea (author's transl)]. *Eur J Cell Biol* **21**, 28-36.

Trapnell C., Pachter L. & Salzberg S.L. (2009) TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105-11.

Trapnell C., Williams B.A., Pertea G., Mortazavi A., Kwan G., van Baren M.J., Salzberg S.L., Wold B.J. & Pachter L. (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* **28**, 511-5.

Umlauf D., Goto Y., Cao R., Cerqueira F., Wagschal A., Zhang Y. & Feil R. (2004) Imprinting along the Kcnq1 domain on mouse chromosome 7 involves repressive histone methylation and recruitment of Polycomb group complexes. *Nat Genet* **36**, 1296-300.

Unnisa Z., Clark J.P., Roychoudhury J., Thomas E., Tessarollo L., Copeland N.G., Jenkins N.A., Grimes H.L. & Kumar A.R. (2012) Meis1 preserves hematopoietic stem cells in mice by limiting oxidative stress. *Blood* **120**, 4973-81.

van Rheede T., Bastiaans T., Boone D.N., Hedges S.B., de Jong W.W. & Madsen O. (2006) The platypus is in its place: nuclear genes and indels confirm the sister group relation of monotremes and Therians. *Mol Biol Evol* **23**, 587-97.

VandeBerg J.L., Robinson E.S., Samollow P.B. & Johnston P.G. (1987) X-linked gene expression and X-chromosome inactivation: marsupials, mouse, and man compared. *Isozymes Curr. Top. Biol. Med. Res.* **15**, 225-53.

VandeBerg J.L. & Williams-Blangero S. (2010) The laboratory opossum (*Monodelphis domestica*). Chapter 19. pp 246-261. In: *UFAW Handbook on the Care and Management of Laboratory and Other Research Animals. 8th edition.* (eds. by Hubrecht R & Kirkwood J). John Wiley & Sons, Ltd., Chichester, England.

Vu T.H., Jirtle R.L. & Hoffman A.R. (2006) Cross-species clues of an epigenetic imprinting regulatory code for the IGF2R gene. *Cytogenet Genome Res* **113**, 202-8.

Wake N., Takagi N. & Sasaki M. (1976) Non-random inactivation of X chromosome in the rat yolk sac. *Nature* **262**, 580-1.

Wakefield M.J., Keohane A.M., Turner B.M. & Graves J.A. (1997) Histone underacetylation is an ancient component of mammalian X chromosome inactivation. *Proc Natl Acad Sci U S A* **94**, 9665-8.

Wang X., Sun Q., McGrath S.D., Mardis E.R., Soloway P.D. & Clark A.G. (2008) Transcriptome-wide identification of novel imprinted genes in neonatal mouse brain. *PLoS One* **3**, e3839.

Wang X., Soloway P.D. & Clark A.G. (2011) A survey for novel imprinted genes in the mouse placenta by mRNA-seq. *Genetics* **189**, 109-22.

Wang X., Miller D.C., Clark A.G. & Antczak D.F. (2012) Random X inactivation in the mule and horse placenta. *Genome Res* **22**, 1855-63.

Wang X., Miller D.C., Harman R., Antczak D.F. & Clark A.G. (2013) Paternally expressed genes predominate in the placenta. *Proc Natl Acad Sci U S A* **110**, 10705-10.

Weidman J.R., Dolinoy D.C., Maloney K.A., Cheng J.F. & Jirtle R.L. (2006a) Imprinting of opossum Igf2r in the absence of differential methylation and air. *Epigenetics* **1**, 49-54.

Weidman J.R., Maloney K.A. & Jirtle R.L. (2006b) Comparative phylogenetic analysis reveals multiple non-imprinted isoforms of opossum Dlk1. *Mamm Genome* **17**, 157-67.

Wutz A., Smrzka O.W., Schweifer N., Schellander K., Wagner E.F. & Barlow D.P. (1997) Imprinted expression of the Igf2r gene depends on an intronic CpG island. *Nature* **389**, 745-9.

Wutz A. & Barlow D.P. (1998) Imprinting of the mouse Igf2r gene depends on an intronic CpG island. *Mol Cell Endocrinol* **140**, 9-14.

Xue F., Tian X.C., Du F., Kubota C., Taneja M., Dinnyes A., Dai Y., Levine H., Pereira L.V. & Yang X. (2002) Aberrant patterns of X chromosome inactivation in bovine clones. *Nat Genet* **31**, 216-20.

Xu F., Mao C., Ding Y., Rui C., Wu L., Shi A., Zhang H., Zhang L. & Xu Z. (2010) Molecular and enzymatic profiles of mammalian DNA methyltransferases: structures and targets for drugs. *Curr Med Chem* **17**, 4052-71.

Yang F., Babak T., Shendure J. & Disteche C.M. (2010) Global survey of escape from X inactivation by RNA-sequencing in mouse. *Genome Res* **20**, 614-22.

Yang Y., Li T., Vu T.H., Ulaner G.A., Hu J.F. & Hoffman A.R. (2003) The histone code regulating expression of the imprinted mouse Igf2r gene. *Endocrinology* **144**, 5658-70.

Zampieri M., Guastafierro T., Calabrese R., Ciccarone F., Bacalini M.G., Reale A., Perilli M., Passananti C. & Caiafa P. (2012) ADP-ribose polymers localized on Ctcf-Parp1-Dnmt1 complex prevent methylation of Ctcf target sites. *Biochem J* **441**, 645-52.

Zeller U. & Freyer C. (2001) Early ontogeny and placentation of the grey short-ailed opossum, *Monodelphis domestica* (Didelphidae: Marsupialia): contribution to the reconstruction of the marsupial morphotype. *J. Zool. Syst. Evol. Res.* **39**.

Zhang Y., Liu T., Meyer C.A., Eeckhoute J., Johnson D.S., Bernstein B.E., Nusbaum C., Myers R.M., Brown M., Li W. & Liu X.S. (2008) Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137.

APPENDIX A

SUPPLEMENTAL FIGURES

Figure A1. The scheme for the opossum crosses and sex genotyping results for the XCI project.
(A-B) Reciprocal F1 crosses between LL1 and LL2 animals. **(C-D)** Parental crosses of LL1 and LL2 animals. In the four crosses, three LL1 animals and three LL2 animals were used. LL1 individuals: A0579 (female) and A0580 (female) are full sibs; LL2 individuals: A0571 (female), A0572 (female) and A0573 (male) are full sibs. A (C/T) SNP was shown (LL1: T and LL2: C). **(E)** Sex genotyping results for opossum embryos. The samples selected for Illumina RNA-seq are labeled with an asterisk. For expanded figure see attached file XCI Supplemental Figures, Figure S1-S2.

Figure A2. RNA-seq, SNP genotyping and pyrosequencing verification results for *HPRT1*, *GPM6B* and *RBMX* in opossum fetal brain and EEM samples.
(A-D) Non-escaper gene *HPRT1*. **(A)** F1 cross of LL1 (mother) x LL2 (father). **(B)** Reciprocal F1 cross of LL2 (mother) x LL1 (father). **(C)** LL1 parental cross. **(D)** LL2 parental cross. From the Sanger sequencing genotyping results, the SNP (OMSNP0222971) is informative in three embryos (571E1, 571E4 and 572E1). In brain/head and EEM tissues of all three individuals, 100% maternal expression was observed from both RNA-seq and allele-specific pyrosequencing verification. Therefore, HPRT1 is subject to imprinted XCI with zero paternal leakage in both tissues. The target sequence for pyrosequencing is (T/C)TTATCTCC.**(E-H)** Autosomal control gene *GPM6B*. **(E)** F1 cross of LL1 (mother) x LL2 (father). **(F)** Reciprocal F1 cross of LL2 (mother) x LL1 (father). **(G)** LL1 parental cross. **(H)** LL2 parental cross. *GPM6B* is an autosomal gene in opossum on chromosome 7. From the Sanger sequencing genotyping results, the SNP (7_27283330) is informative in three embryos (579E3, 579E4 and 571E4). In brain/head and EEM tissues of all three individuals, biallelic expression was observed from both RNA-seq and allele-specific pyrosequencing verification, which is expected for autosomal genes with two parental alleles. The target sequence for pyrosequencing is (T/C)GAGACT. The Sanger sequencing traces were not shown here because an indel polymorphism in the amplicon shifted the trace, but the genotypes could be determined by the CodonCode Aligner software.**(I-H)** Escaper gene *RBMX*. **(I)** F1 cross of LL1 (mother) x LL2 (father). **(J)** Reciprocal F1 cross of LL2 (mother) x LL1 (father). **(K)** LL1 parental cross. **(L)** LL2 parental cross. From the Sanger sequencing genotyping results, the SNP (OMSNP0156027) is informative in five embryos (579E3, 579E4, 571E1, 571E4 and 580E1). In brain/head and EEM tissues of all five individuals, biallelic expression was observed from both RNA-seq and allele-specific pyrosequencing verification. Therefore, RBMX is an escaper of imprinted XCI in both tissues. The target sequence for pyrosequencing is G(C/G)TATGGTGGT (on the minus strand). For figures of other genes, see attached file XCI Supplemental Figures, Figures S4-S28.

Figure A3. M-A plot for brain and EEM expression levels before and after normalization. For figure, see attached file XCI Supplemental Figures, Figure S3.

Figure A4. Histone modification H3K4me3 and H3K27me3 peaks and coverage profile for X chromosome region containing escaper genes in female brain ChIP-seq data. In each figure, plotted in the top panel are the genome gap locations and the H3K27me3 peaks and coverage. Gene models are shown in the middle panel, color-coded based on their imprinted XCI status (blue: escapers; red: non-escapers; gray: not known due to lack of informative SNPs). In the bottom panel are the CpG island locations and the H3K4me3 peaks and coverage profile. **(A)** For the escaper gene *PLXNA3*, H3K4me3 was present at promoter CpG islands, suggesting active transcription. H3K27me3 marks were depleted across the gene body, consistent with biallelic expression. The downstream gene *UBL4A* does not have informative SNPs, but its histone modification profile suggests it is a candidate escaper (Table B4). The other three non-informative genes (*ATP6AP1*, *GDI1* and *SLC10A3*) were covered with H3K27me3 peaks across the entire gene body, consistent with non-escaper status. **(B)** For the escaper gene *DKC1*, H3K4me3 were present at promoter CpG islands, suggesting active transcription. H3K27me3 marks were depleted across the gene body, consistent with biallelic expression. For the nonescaper gene *MPP1*, the H3K27me3 peaks covered the entire gene body, consistent with monoallelic expression. **(C)** For the escaper gene *NUP62*, H3K4me3 was present at the promoter CpG island, suggesting active transcription. H3K27me3 were depleted across the gene body, consistent with biallelic expression. For the nonescaper gene *RBM41*, the H3K4me3 peak was present and the H3K27me3 peaks covered the entire gene body, consistent with monoallelic expression. **(D)** For the three escaper genes in this region (*HCFC1*, *IRAK1* and *MECP2*), H3K4me3 marks were present at promoter CpG islands, suggesting active transcription. H3K27me3 marks were depleted across the gene body, consistent with biallelic expression. For the two non-escapers (*TKTL1* and *LOC10002972*), the H3K4me3 marks were present and the H3K27me3 peaks covered the entire gene body, consistent with monoallelic expression. For the rest four genes, there is no informative SNP to infer the XCI status from RNA-seq data, but the histone modification profile of *NAA10* is consistent with escaper status (Table B4). **(E)** For the escaper gene *FRMD7*, H3K4me3 was present at promoter CpG islands, suggesting active transcription. H3K27me3 marks were depleted across the gene body, consistent with biallelic expression. For the non-escaper gene *RAP2C*, the H3K4me3 peak was present and the H3K27me3 peak covered the entire gene body, consistent with monoallelic expression. The histone modification profile of the upstream non-informative gene *MST4* is consistent with non-escaping status. **(F)** For the three escaper genes (*PHF6*, *FAM122B* and *FAM122A*), H3K4me3 marks were present at promoter CpG islands, suggesting active transcription. H3K27me3 marks were depleted across the gene body, consistent with biallelic expression. For the two nonescapers (*HPRT1* and *MOSPD1*), the H3K4me3 marks were present and the H3K27me3 peaks covered the entire gene body, consistent with monoallelic expression. **(G)** For the escaper gene *RBMX*, H3K4me3 mark was present at

the promoter CpG island, suggesting active transcription. H3K27me3 marks were depleted across the gene body, consistent with biallelic expression. For the two non-escapers (*ARHGEF6* and *TM9SF2*), the H3K4me3 marks were present and the H3K27me3 peaks covered the entire gene body, consistent with monoallelic expression. **(H)** For the escaper gene *ATRX*, H3K4me3 mark was present at the promoter CpG island, suggesting active transcription. H3K27me3 marks were depleted across the gene body, consistent with biallelic expression. For the two nonescapers (*MAGT1* and *COX7B*), the H3K4me3 marks were present and the H3K27me3 peaks covered the entire gene body, consistent with monoallelic expression. The non-informative upstream gene *FGF16* was covered with H3K27me3 peaks across the entire gene body, consistent with nonescaper status. **(I)** For the two escaper genes (*TAF1* and *NONO*), H3K4me3 marks were present at promoter CpG islands, suggesting active transcription. H3K27me3 marks were depleted across the gene body, consistent with biallelic expression. For the three nonescapers (*APEX2*, *ZMYM3* and *NLGN3*), H3K27me3 peaks covered the entire gene body, consistent with monoallelic expression. Three non-informative genes (*OGT*, *RHOG* and *ITGB1BP2*) in the H3K27me3 depleted region are consistent with escaper status. **(J)** For the escaper gene *CENPI*, H3K4me3 mark was present at promoter CpG islands, suggesting active transcription. H3K27me3 marks were depleted across the gene body, consistent with biallelic expression. For the non-escaper *SYTL4*, the H3K27me3 peaks covered the entire gene body, consistent with monoallelic expression. The downstream gene *CSTF2* does not have informative SNPs, but its histone modification profile suggests it is a candidate escaper. The other two non-informative genes (*TMEM35* and *XKRX*) were covered with H3K27me3 peaks across the entire gene body, consistent with non-escaper status. For expanded figures see attached file XCI Supplemental Figures, Figures S29-S44.

Figure A5. Allele-specific histone modification H3K4me3 for escaper and non-escaper genes in female brain ChIP-seq data from LL1 x LL2 cross. (A) Left: the 5'-end gene model, CpG island location and the H3K4me3 peak/coverage profile for escaper gene *YIPF6*. There is one SNP (X_7594487) under the H3K4me3 peak with enough coverage to infer allele-specific histone modification. **Middle:** Sanger sequencing genotyping results in the two embryos (579E10 and 579E11, used for ChIP-seq experiments) and their parents confirmed that the SNP is informative in both embryos. **Right:** from the ChIP-seq data, we observed 64% of the H3K4me3 reads from the maternal allele and 36% from the paternal allele at X_7594487, suggesting both parental alleles are active. This is consistent with allele-specific expression profile at SNP OMSNP0155108 in the RNA-seq data and SNP OMSNP0155110 from the allele-specific pyrosequencing results. **(B) Left:** the 5'-end gene model, CpG island location and the H3K4me3 peak/coverage profile for escaper gene *FAM122B*. There is one SNP (OMSNP0156061) under the H3K4me3 peak with enough coverage to infer allele-specific histone modification. **Middle:** Sanger sequencing genotyping results in the two embryos (579E10 and 579E11) and their mother confirmed that the SNP is informative in both embryos. **Right:** from the ChIP-seq data, we observed the H3K4me3 reads from both parental alleles at OMSNP0156061, suggesting both parental alleles are active. This

is consistent with biallelic expression from the allele-specific pyrosequencing results. **(C) Left:** The 5'-end gene model, CpG island location and the H3K4me3 peak/coverage profile for non-escaper gene *PNCK*. There is one SNP (OMSNP0155237) under the H3K4me3 peak with enough coverage to infer allele-specific histone modification. **Middle:** Sanger sequencing genotyping results in the two embryos (579E10 and 579E11) and their parents confirmed that the SNP is informative in both embryos. **Right:** from the ChIP-seq data, we observed 100% of the H3K4me3 reads from the maternal allele at OMSNP0155237, suggesting the on-mark is only present at the maternal allele. This is consistent with maternal-specific expression at OMSNP0155237 and OMSNP0155219 in the RNA-seq data. **(D) Left:** The 5'-end gene model, CpG island location and the H3K4me3 peak/coverage profile for non-escaper gene *GPC4*. There is one SNP (OMSNP0156005) under the H3K4me3 peak with enough coverage to infer allele-specific histone modification. **Middle:** Sanger sequencing genotyping results in the two embryos (579E10 and 579E11) and their parents confirmed that the SNP is informative in both embryos. **Right:** from the ChIP-seq data, we observed 100% of the H3K4me3 reads from the maternal allele at OMSNP0156005, suggesting the on-mark is only present at the maternal allele. This is consistent with maternal-specific expression at OMSNP0156005 and OMSNP0156006 in the RNA-seq data. **(E) Left:** The 5'-end gene model, CpG island location and the H3K4me3 peak/coverage profile for non-escaper gene *ITM2A*. There is one SNP (OMSNP0156531) under the H3K4me3 peak with enough coverage to infer allele-specific histone modification. **Middle:** Sanger sequencing genotyping results in the two embryos (579E10 and 579E11) and their parents confirmed that the SNP is informative in both embryos. **Right:** from the ChIP-seq data, we observed 100% of the H3K4me3 reads from the maternal allele at OMSNP0156531, suggesting the on-mark is only present at the maternal allele. This is consistent with maternal-specific expression at OMSNP0156531 in the RNA-seq data. **(F) Left:** The 5'-end gene model, CpG island location and the H3K4me3 peak/coverage profile for two X-linked genes, *PDZD11* (+ strand) and *KIF4A* (- strand). They are organized in head-to-tail orientation, and they share one CpG island and one H3K4me3 peak. *PDZD11* is a nonescaper gene (colored in red) and the escaping status for *KIF4A* (colored in gray) is unknown due to lack of informative exonic SNPs. There is one SNP (OMSNP0223343) under the H3K4me3 peak with enough coverage to infer allele-specific histone modification. **Middle:** Sanger sequencing genotyping results in the two embryos (579E10 and 579E11) and their parents confirmed that the SNP is informative in both embryos. **Right:** From the ChIP-seq data, we observed 100% of the H3K4me3 reads from the maternal allele at OMSNP0223343, suggesting the on-mark is only present at the maternal allele. This is consistent with maternal-specific expression of *PDZD11* at OMSNP0223343 and OMSNP0156925 in the RNA-seq data. For expanded figures, see attached file XCI Supplemental Figures, Figures S45-S50.

Figure A6. Pyrograms for PyroMark analysis of the methylation profile at promoter CpG islands of X-linked genes. Pyrograms for PyroMark analysis of the methylation profile at escaper gene *CD99L2* **(A)** and non-escaper gene *LAS1L* **(B)** promoter CpG islands. The bisulfite converted target sequence to analysis is shown on

the top for each gene. Methylation percentages for the CpG positions are quantified in female fetal brain, male fetal brain, female EEM and male EEM (from top to bottom). For pyrograms of more gene promoters, see XCI Supplemental Figures, Figures S51 A-I and K-Z.

Figure A7. Pyrograms for PyroMark analysis of the methylation profile at *Rsx* promoter CpG island. Four PyroMark primer sets were designed to profile the methylation at *RSX* promoter CpG island **(A-D)**. The bisulfite converted target sequence to analysis is shown on the top for primer. Methylation percentages for the CpG positions are quantified in female fetal brain, male fetal brain, female EEM and male EEM (from top to bottom). For expanded pyrograms of *RSX* promoter see attached file XCI Supplemental Figures, Figure 51J.

Figure A8. Pedigree Information for fibroblast cell lines A) Crosses and animals used for ChIP-seq experiments. Animal IDs and stock source (LL1 or LL2) are indicated. B) Crosses and animals used for DNA and RNA verification experiments. Animal IDs are indicated. Top Panel: LL1 females crossed with LL2 males. Bottom Panel: LL2 females crossed with LL1 males.

Figure A9. **Histone modification profile for *Cstb*and *Rpl17*.** Green panels = ChIP-seq raw read alignments for H3K4me3 (top) and H3K9Ac (bottom). Red panels = ChIP-seq raw read alignments for H3K9me3 (top) and H3K27me3 (bottom). Black panel = Input. Blue bars above ChIP-seq panels are areas of significant enrichment determined by MACS ($p \leq 10^{-5}$). Blue bar in bottom panel represents the gene annotation with the direction of transcription indicated by the black arrows. Annotated CpG islands are indicated by black bars one panel above the annotation.

Figure A10. Pedigrees and Sanger Sequencing Results for A) *Meis1*, B) *Cstb*, C) *Rpl17*, and D) *Igf2r* from reciprocal crosses. Pedigrees and Sanger Sequencing Results for A) *Meis1*, B) *Cstb*, C) *Rpl17*, and D) *Igf2r* from reciprocal crosses. Animal IDs are presented as A0xxx, Sanger sequences generated from cDNA (top) and gDNA (bottom) are shown for informative animals. Black arrows indicate positions of informative SNPs. Only cDNA data is shown for *Rpl17*. Lower case letters for *Meis1* indicate leaky expression of the imprinted allele. Sanger data are not definitive for gene expression and were used exclusively to screen for probable heterozygotes for pyromark analysis of relative allelic expression.

Figure A11. Histone modification profile and DNA methylation at *Igf2r*. Promoter, CpG island hypomethylation, and a differentially methylated region at the CpG island at intron 11 for animal A0694. Top two panels are the histone modifications, H3K4me3 and H3K9Ac and significant peaks as called by MACS ($p \leq 10^{-5}$).

Figure A12. Histone modification profiles at A) *Htr2a*, B) *L3mbtl,* and C) *Mest*. Green panels = ChIP-seq raw read alignments for H3K4me3 (top) and H3K9Ac

(bottom). Red panels = ChIP-seq raw read alignments for H3K9me3 (top) and H3K27me3 (bottom). Black panel = input. Blue bars above ChIP-seq panels are areas of significant enrichment determined by MACS ($p \leq 10^{-5}$). Blue bar in bottom panel represents the gene annotation with the direction of transcription indicated by the black arrow. Annotated CpG islands are indicated by black bars one panel above the annotation. Gaps in the genome assembly are in the panel above the CpG Island panel (only seen in B).

Figure A1.

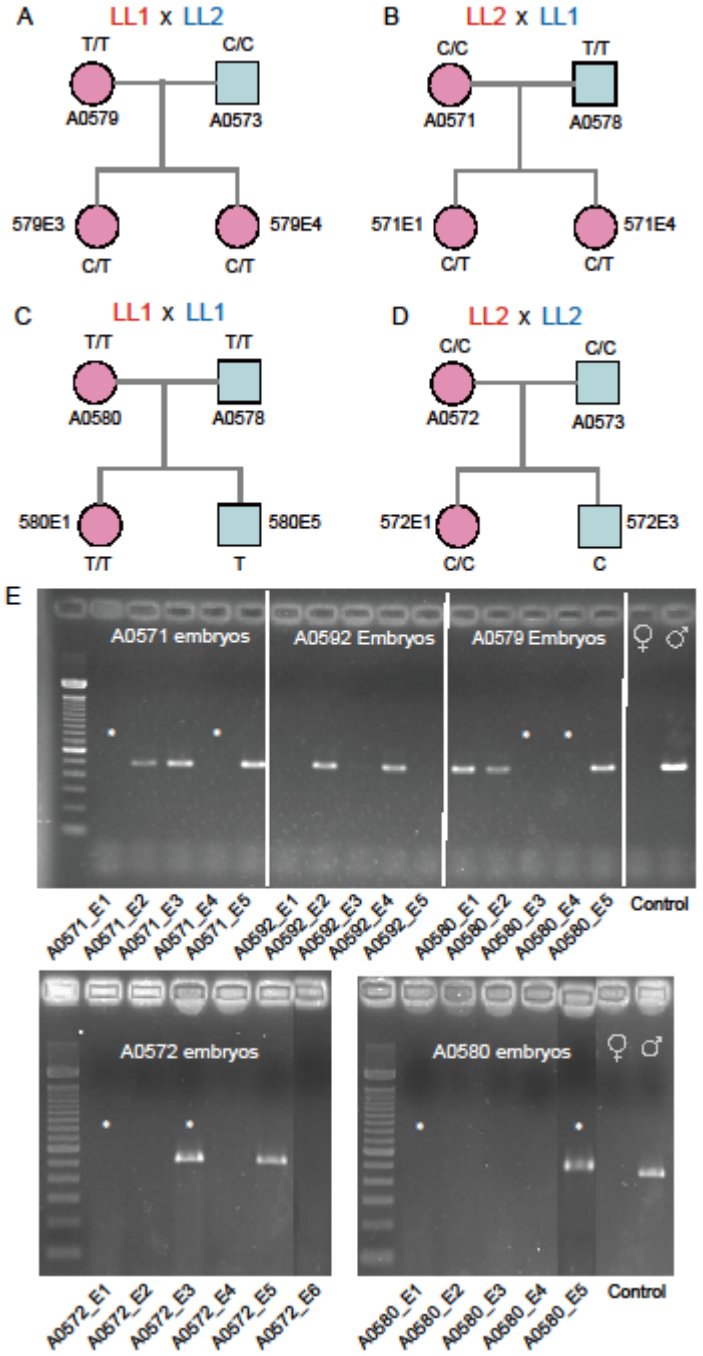

A
LL1 x LL2
T/T
C/C
A0579
A0573
579E3
579E4
C/T
C/T
B
LL2 x LL1
C/C
T/T
A0571
A0578
571E1
571E4
C/T
C/T
C
LL1 x LL1
T/T
T/T
A0580
A0578
580E1
580E5
T/T
T
D
LL2 x LL2
C/C
C/C
A0572
A0573
572E1
572E3
C/C
C
E
A0571 embryos
A0592 Embryos
A0579 Embryos
♀ ♂
A0571_E1
A0571_E2
A0571_E3
A0571_E4
A0571_E5
A0592_E1
A0592_E2
A0592_E3
A0592_E4
A0592_E5
A0580_E1
A0580_E2
A0580_E3
A0580_E4
A0580_E5
Control
A0572 embryos
A0580 embryos
♀ ♂
A0572_E1
A0572_E2
A0572_E3
A0572_E4
A0572_E5
A0572_E6
A0580_E1
A0580_E2
A0580_E3
A0580_E4
A0580_E5
Control

Figure A2.

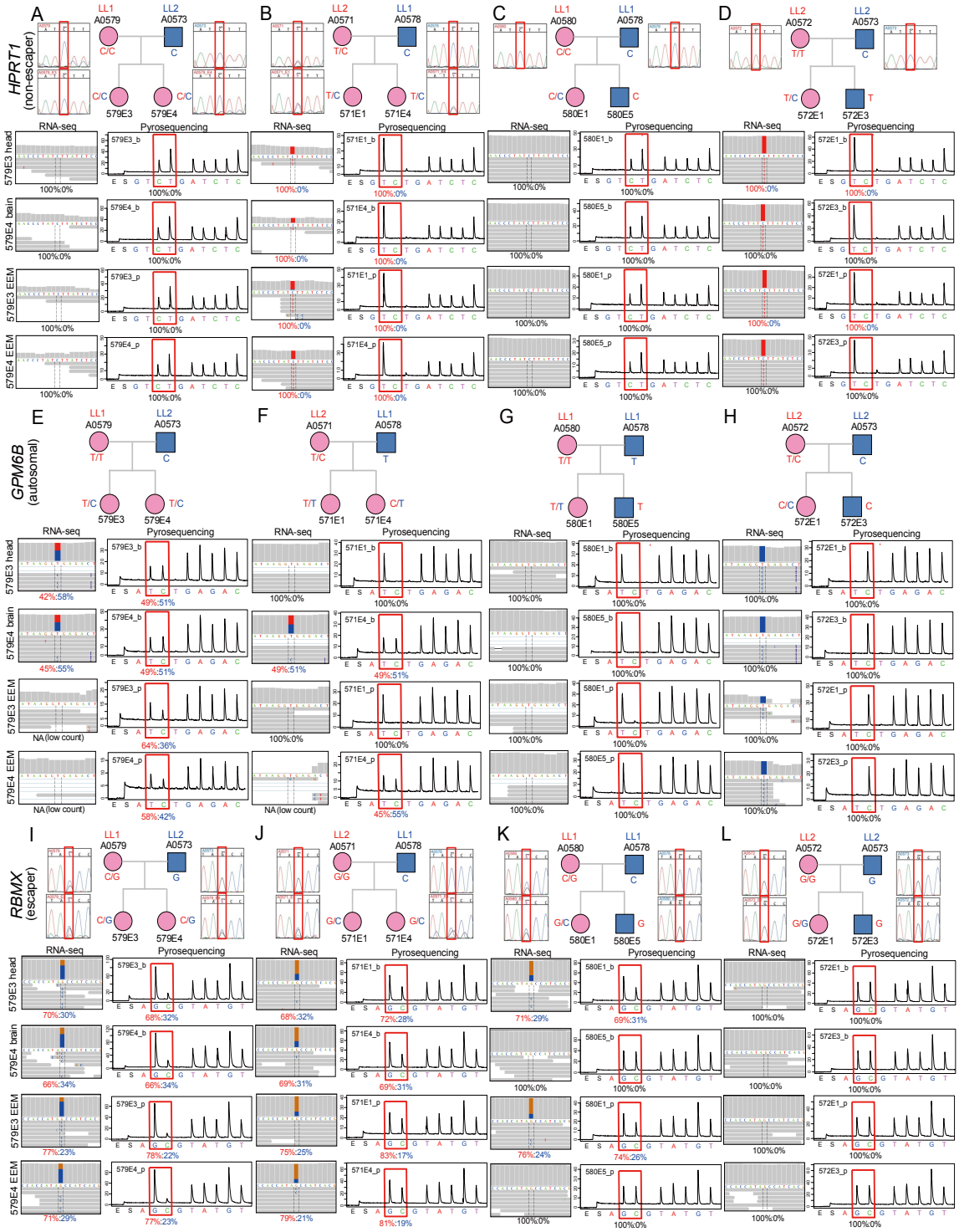

A
B
C
D
HPRT1
(non-escaper)
LL1
LL2
A0579
A0573
A0571
A0578
A0580
A0572
C/C
C
T/C
T
T/T
579E3
579E4
571E1
571E4
580E1
580E5
572E1
572E3
RNA-seq
Pyrosequencing
579E3 head
579E4 brain
579E3 EEM
579E4 EEM
100%:0%
E
F
G
H
GPM6B
(autosomal)
42%:58%
49%:51%
45%:55%
NA (low count)
64%:36%
58%:42%
C/T
I
J
K
L
RBMX
(escaper)
G/G
G/C
C/G
G
70%:30%
68%:32%
66%:34%
72%:28%
69%:31%
71%:29%
77%:23%
78%:22%
75%:25%
83%:17%
76%:24%
74%:26%
79%:21%
81%:19%

Figure A4.

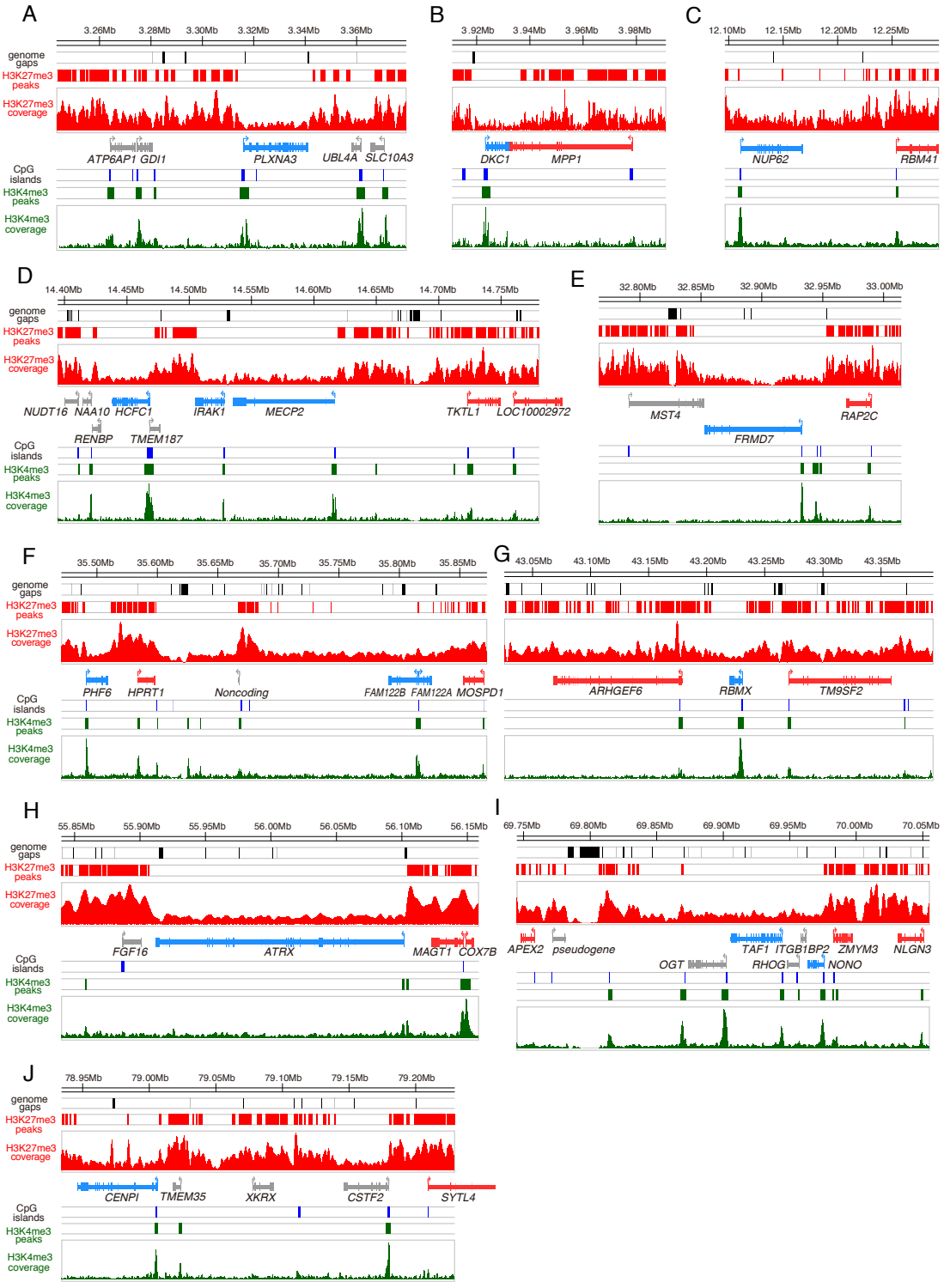

A
3.26Mb 3.28Mb 3.30Mb 3.32Mb 3.34Mb 3.36Mb
genome gaps
H3K27me3 peaks
H3K27me3 coverage
ATP6AP1 GDI1
PLXNA3
UBL4A SLC10A3
CpG islands
H3K4me3 peaks
H3K4me3 coverage
B
3.92Mb 3.94Mb 3.96Mb 3.98Mb
DKC1 MPP1
C
12.10Mb 12.15Mb 12.20Mb 12.25Mb
NUP62 RBM41
D
14.40Mb 14.45Mb 14.50Mb 14.55Mb 14.60Mb 14.65Mb 14.70Mb 14.75Mb
NUDT16 NAA10 HCFC1 IRAK1 MECP2 TKTL1 LOC10002972
RENBP TMEM187
E
32.80Mb 32.85Mb 32.90Mb 32.95Mb 33.00Mb
MST4 RAP2C
FRMD7
F
35.50Mb 35.60Mb 35.65Mb 35.70Mb 35.75Mb 35.80Mb 35.85Mb
PHF6 HPRT1 Noncoding FAM122B FAM122A MOSPD1
G
43.05Mb 43.10Mb 43.15Mb 43.20Mb 43.25Mb 43.30Mb 43.35Mb
ARHGEF6 RBMX TM9SF2
H
55.85Mb 55.90Mb 55.95Mb 56.00Mb 56.05Mb 56.10Mb 56.15Mb
FGF16 ATRX MAGT1 COX7B
I
69.75Mb 69.80Mb 69.85Mb 69.90Mb 69.95Mb 70.00Mb 70.05Mb
APEX2 pseudogene TAF1 ITGB1BP2 ZMYM3 NLGN3
OGT RHOG NONO
J
78.95Mb 79.00Mb 79.05Mb 79.10Mb 79.15Mb 79.20Mb
CENPI TMEM35 XKRX CSTF2 SYTL4

Figure A5.

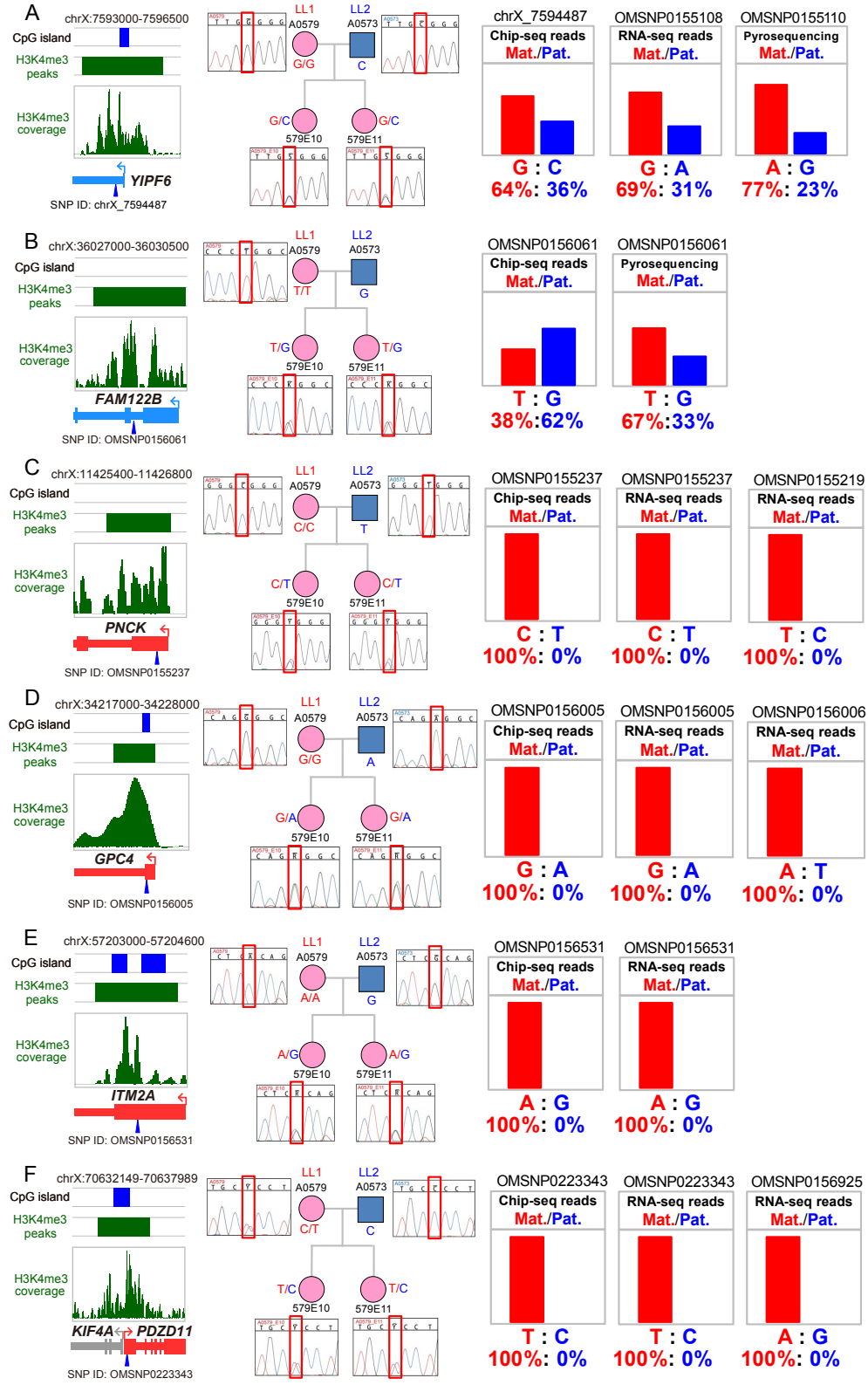

A
chrX:7593000-7596500
CpG island
H3K4me3 peaks
H3K4me3 coverage
YIPF6
SNP ID: chrX_7594487
LL1
A0579
G/G
LL2
A0573
C
G/C
579E10
G/C
579E11
chrX_7594487
Chip-seq reads Mat./Pat.
G : C
64%: 36%
OMSNP0155108
RNA-seq reads Mat./Pat.
G : A
69%: 31%
OMSNP0155110
Pyrosequencing Mat./Pat.
A : G
77%: 23%
B
chrX:36027000-36030500
FAM122B
SNP ID: OMSNP0156061
T/T
G
T/G
579E10
T/G
579E11
OMSNP0156061
Chip-seq reads Mat./Pat.
T : G
38%:62%
OMSNP0156061
Pyrosequencing Mat./Pat.
T : G
67%:33%
C
chrX:11425400-11426800
PNCK
SNP ID: OMSNP0155237
C/C
T
C/T
579E10
C/T
579E11
OMSNP0155237
Chip-seq reads Mat./Pat.
C : T
100%: 0%
OMSNP0155237
RNA-seq reads Mat./Pat.
C : T
100%: 0%
OMSNP0155219
RNA-seq reads Mat./Pat.
T : C
100%: 0%
D
chrX:34217000-34228000
GPC4
SNP ID: OMSNP0156005
G/G
A
G/A
579E10
G/A
579E11
OMSNP0156005
Chip-seq reads Mat./Pat.
G : A
100%: 0%
OMSNP0156005
RNA-seq reads Mat./Pat.
G : A
100%: 0%
OMSNP0156006
RNA-seq reads Mat./Pat.
A : T
100%: 0%
E
chrX:57203000-57204600
ITM2A
SNP ID: OMSNP0156531
A/A
G
A/G
579E10
A/G
579E11
OMSNP0156531
Chip-seq reads Mat./Pat.
A : G
100%: 0%
OMSNP0156531
RNA-seq reads Mat./Pat.
A : G
100%: 0%
F
chrX:70632149-70637989
KIF4A
PDZD11
SNP ID: OMSNP0223343
C/T
C
T/C
579E10
T/C
579E11
OMSNP0223343
Chip-seq reads Mat./Pat.
T : C
100%: 0%
OMSNP0223343
RNA-seq reads Mat./Pat.
T : C
100%: 0%
OMSNP0156925
RNA-seq reads Mat./Pat.
A : G
100%: 0%

Figure A6.

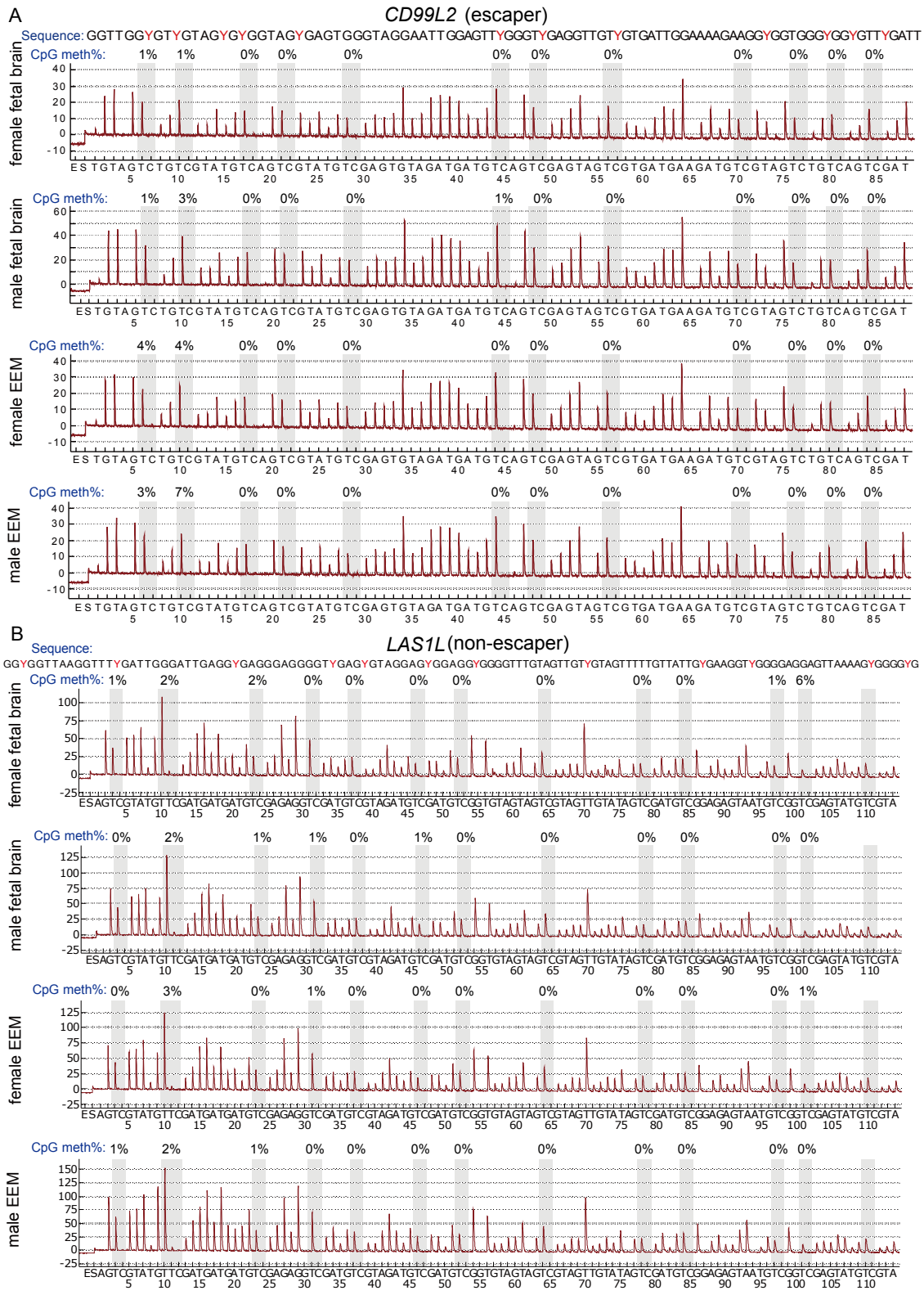

A
CD99L2 (escaper)
Sequence: GGTTGGYGTYGTAGYGYGGTAGYGAGTGGGTAGGAATTGGAGTTYGGGTYGAGGTTGTYGTGATTGGAAAAGAAGGYGGTGGGYGGYGTTYGATT
female fetal brain
male fetal brain
female EEM
male EEM
B
LAS1L (non-escaper)
Sequence: GGYGGTTAAGGTTTYGATTGGGATTGAGGYGAGGGAGGGGTYGAGYGTAGGAGYGGAGGYGGGGTTTGTAGTTGTYGTAGTTTTTGTTATTGYGAAGGTYGGGGAGGAGTTAAAAGYGGGGYG

Figure A7.

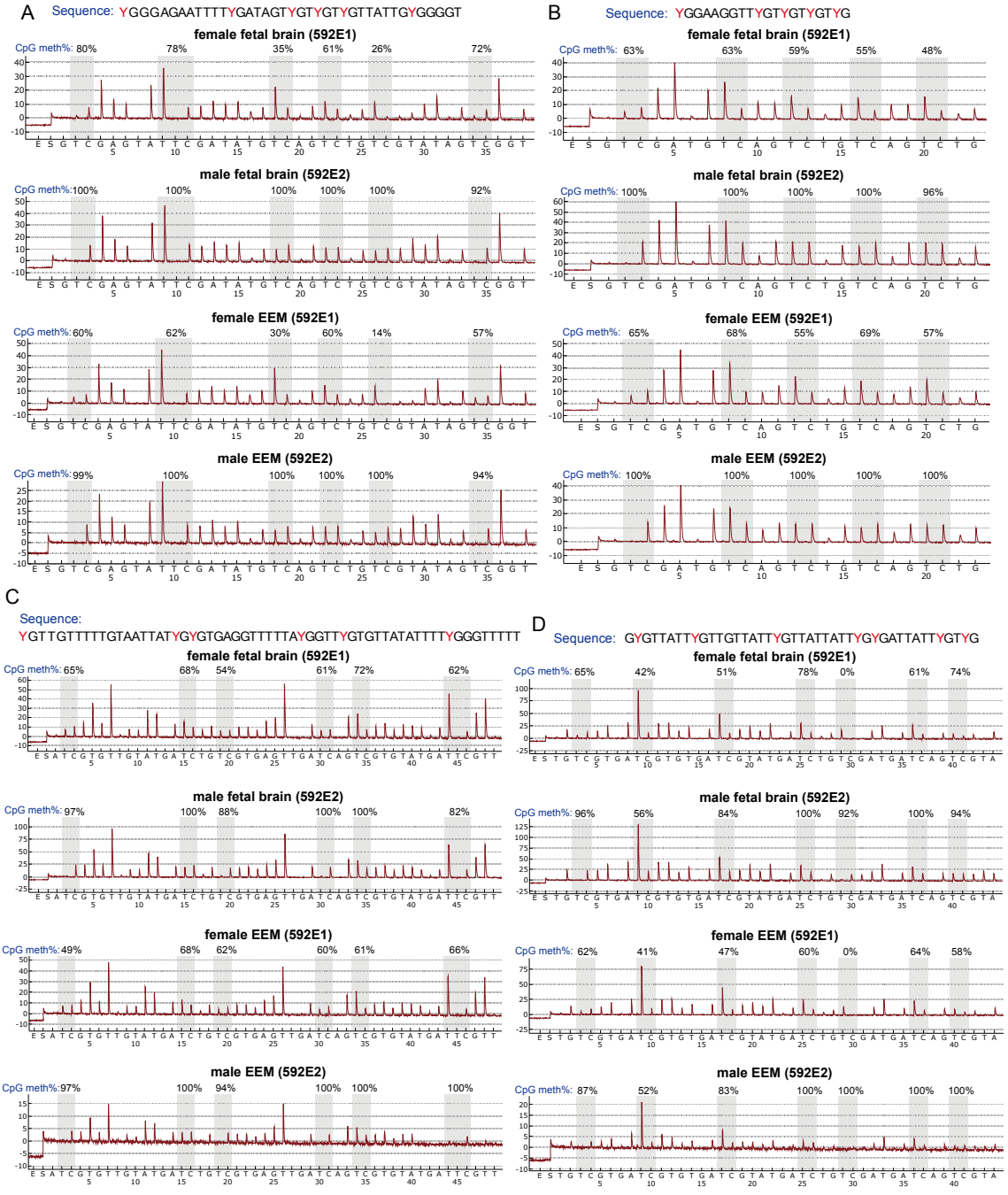
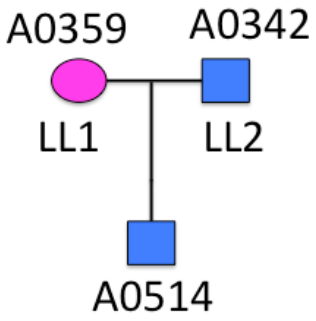

A
Sequence: YGGGAGAATTTTYGATAGTYGTYGTYGTTATTGYGGGGT
female fetal brain (592E1)
CpG meth%: 80% 78% 35% 61% 26% 72%
male fetal brain (592E2)
CpG meth%: 100% 100% 100% 100% 100% 92%
female EEM (592E1)
CpG meth%: 60% 62% 30% 60% 14% 57%
male EEM (592E2)
CpG meth%: 99% 100% 100% 100% 100% 94%
B
Sequence: YGGAAGGTTYGTYGTYGTYG
female fetal brain (592E1)
CpG meth%: 63% 63% 59% 55% 48%
male fetal brain (592E2)
CpG meth%: 100% 100% 100% 100% 96%
female EEM (592E1)
CpG meth%: 65% 68% 55% 69% 57%
male EEM (592E2)
CpG meth%: 100% 100% 100% 100% 100%
C
Sequence:
YGTTGTTTTTGTAATTATYGYGTGAGGTTTTTAYGGTTYGTGTTATATTTTYGGGTTTTT
female fetal brain (592E1)
CpG meth%: 65% 68% 54% 61% 72% 62%
male fetal brain (592E2)
CpG meth%: 97% 100% 88% 100% 100% 82%
female EEM (592E1)
CpG meth%: 49% 68% 62% 60% 61% 66%
male EEM (592E2)
CpG meth%: 97% 100% 94% 100% 100% 100%
D
Sequence: GYGTTATTYGTTGTTATTYGTTATTATTYGYGATTATTYGTYG
female fetal brain (592E1)
CpG meth%: 65% 42% 51% 78% 0% 61% 74%
male fetal brain (592E2)
CpG meth%: 96% 56% 84% 100% 92% 100% 94%
female EEM (592E1)
CpG meth%: 62% 41% 47% 60% 0% 64% 58%
male EEM (592E2)
CpG meth%: 87% 52% 83% 100% 100% 100% 100%

Figure A8.

A)


A0359
A0342
LL1
LL2
A0514

B)

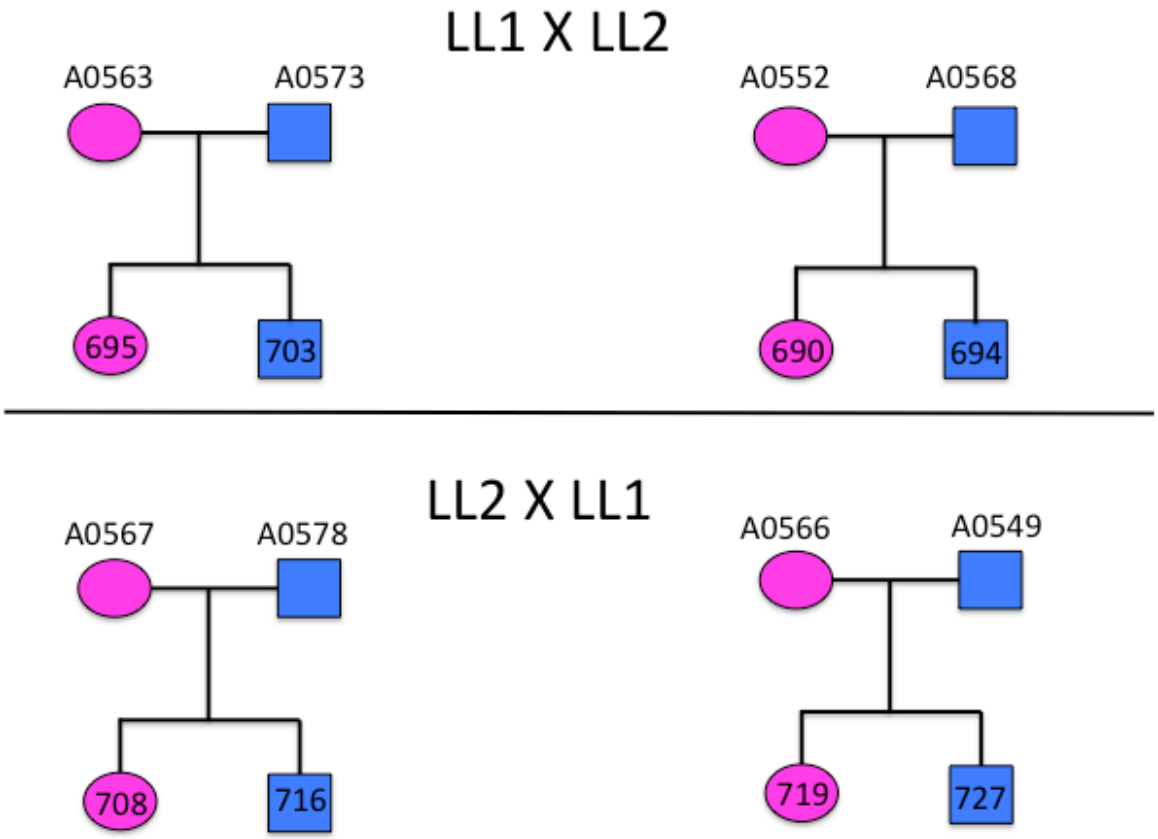
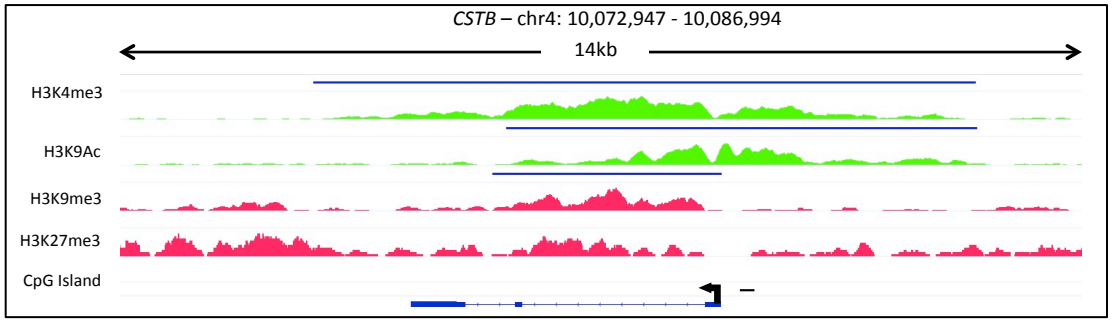

LL1 X LL2
A0563
A0573
695
703
A0552
A0568
690
694
LL2 X LL1
A0567
A0578
708
716
A0566
A0549
719
727

Figure A9.

A)


CSTB – chr4: 10,072,947 - 10,086,994
14kb
H3K4me3
H3K9Ac
H3K9me3
H3K27me3
CpG Island

B)

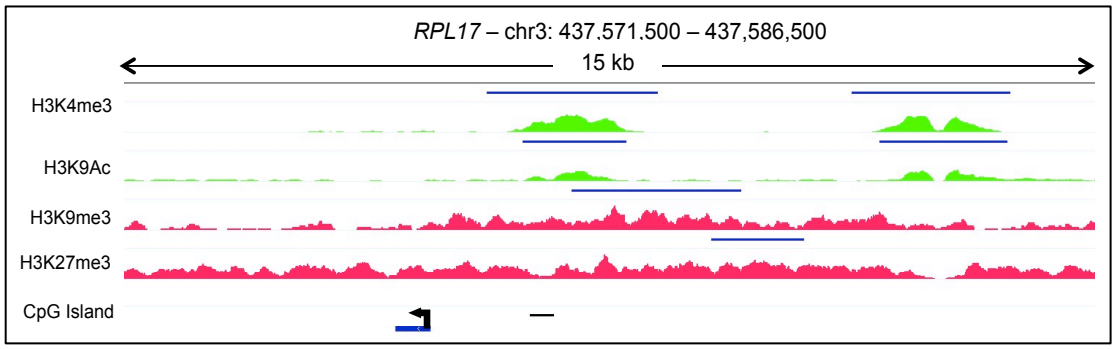
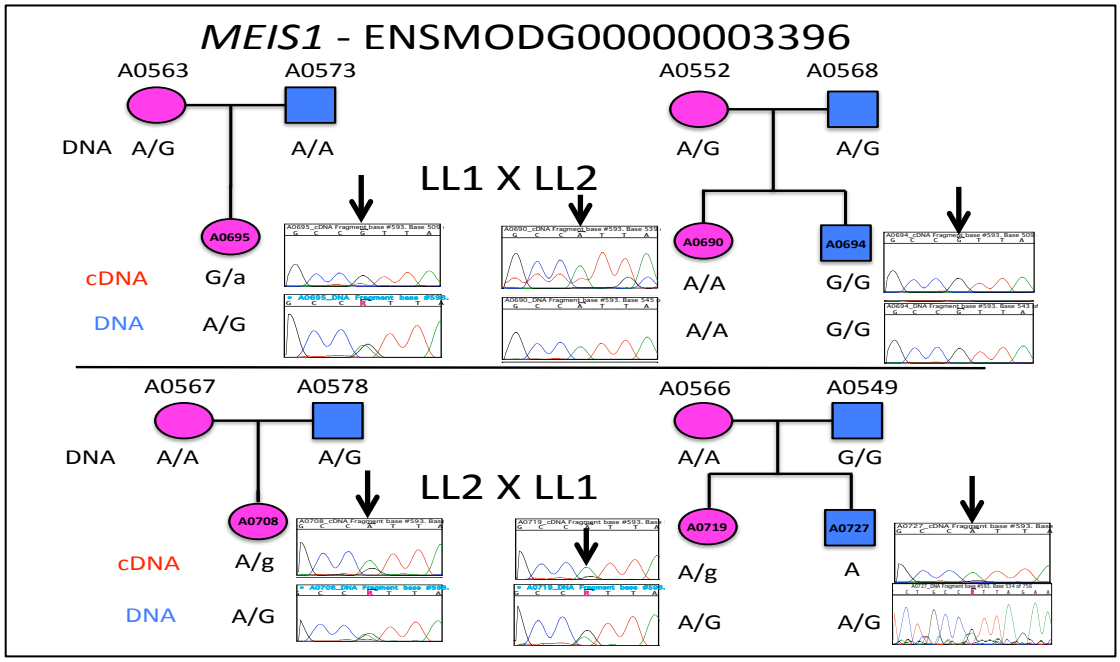
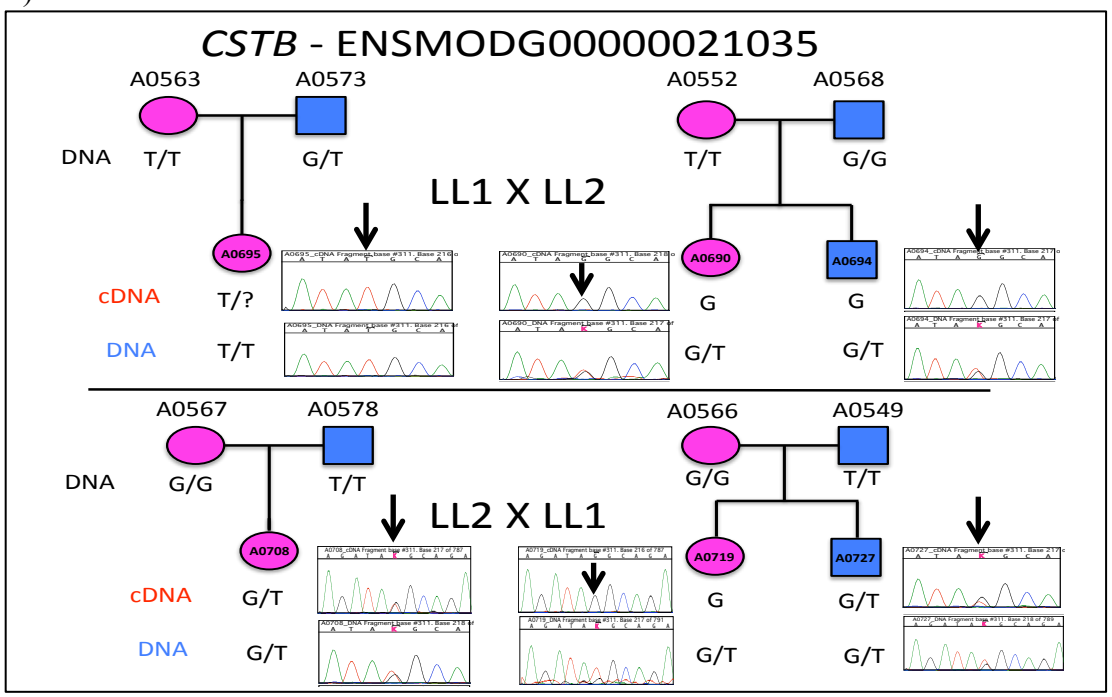

RPL17 – chr3: 437.571.500 – 437,586,500
15 kb
H3K4me3
H3K9Ac
H3K9me3
H3K27me3
CpG Island

Figure A10.

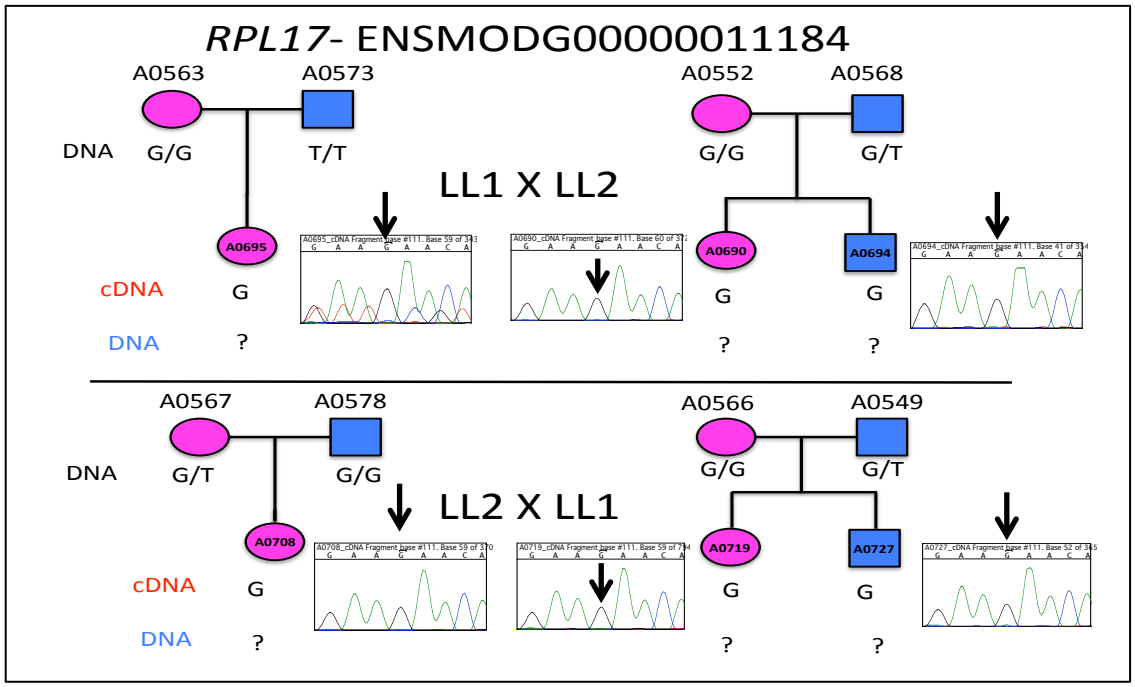
A)



B)



C)


RPL17- ENSMODG00000011184
A0563
A0573
DNA
G/G
T/T
LL1 X LL2
A0552
A0568
G/G
G/T
A0695
A0690
A0694
cDNA
G
G
G
DNA
?
?
?
A0567
A0578
DNA
G/T
G/G
LL2 X LL1
A0566
A0549
G/G
G/T
A0708
A0719
A0727
cDNA
G
G
G
DNA
?
?
?

D)

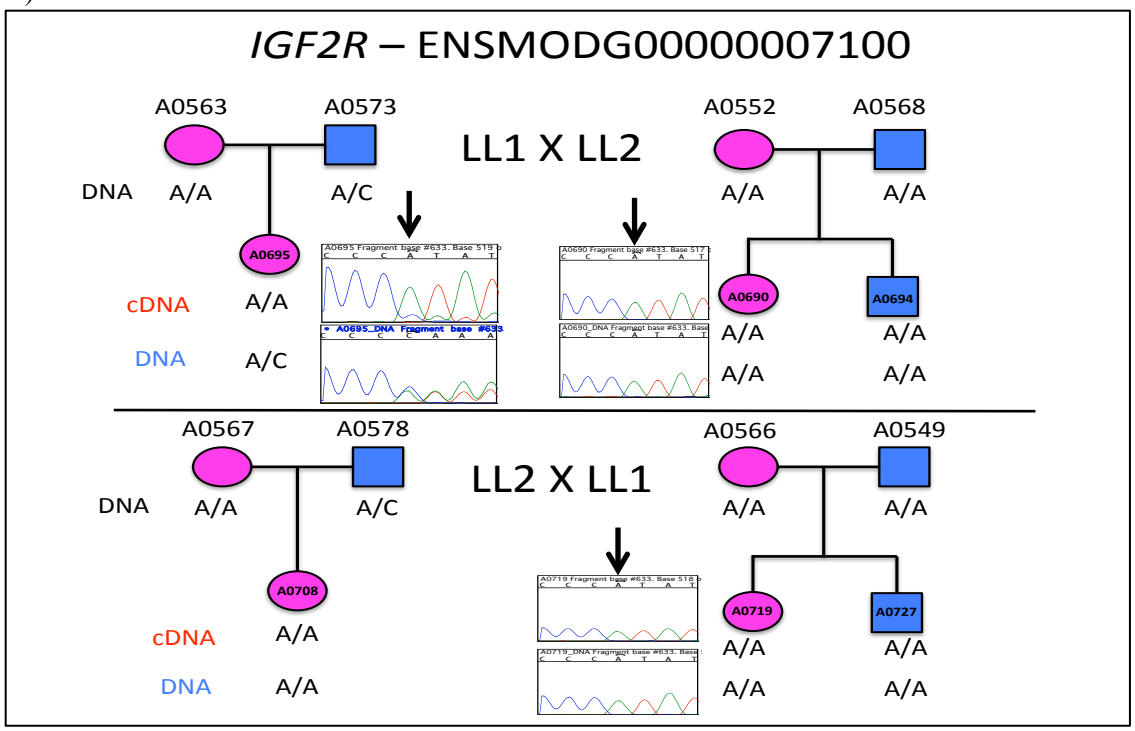

IGF2R – ENSMODG00000007100
A0563
A0573
LL1 X LL2
A0552
A0568
DNA
A/A
A/C
A/A
A/A
A0695
A0690
A0694
cDNA
A/A
A/A
A/A
DNA
A/C
A/A
A/A
A0567
A0578
LL2 X LL1
A0566
A0549
DNA
A/A
A/C
A/A
A/A
A0708
A0719
A0727
cDNA
A/A
A/A
A/A
DNA
A/A
A/A
A/A

Figure A11.

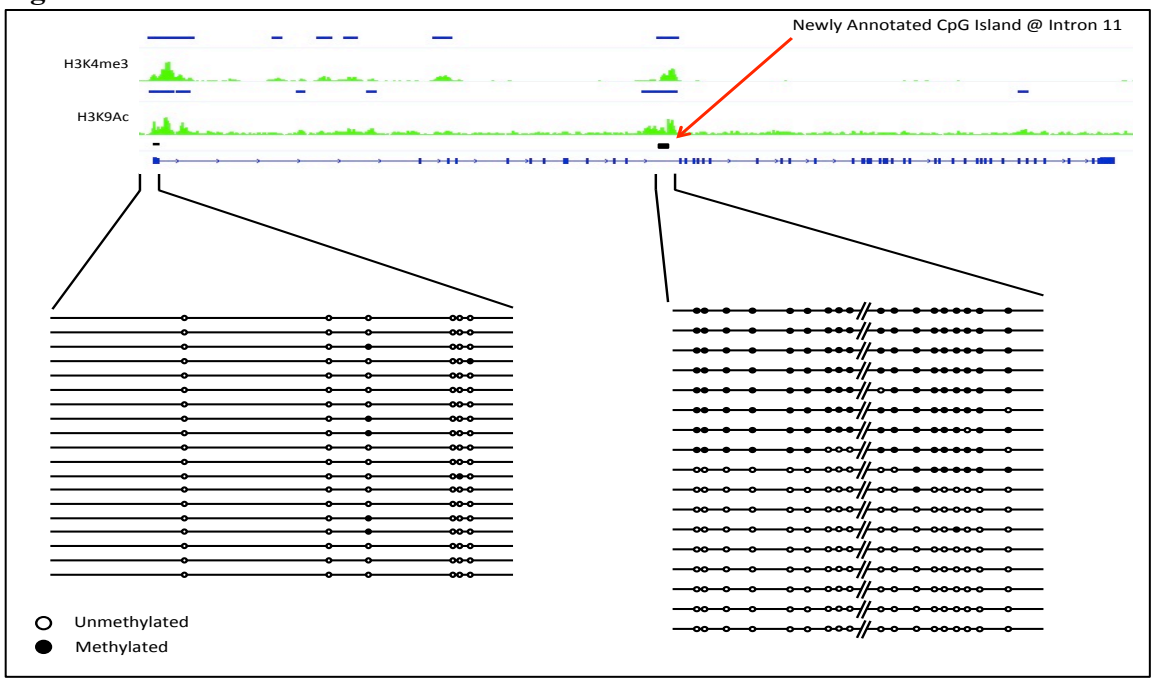
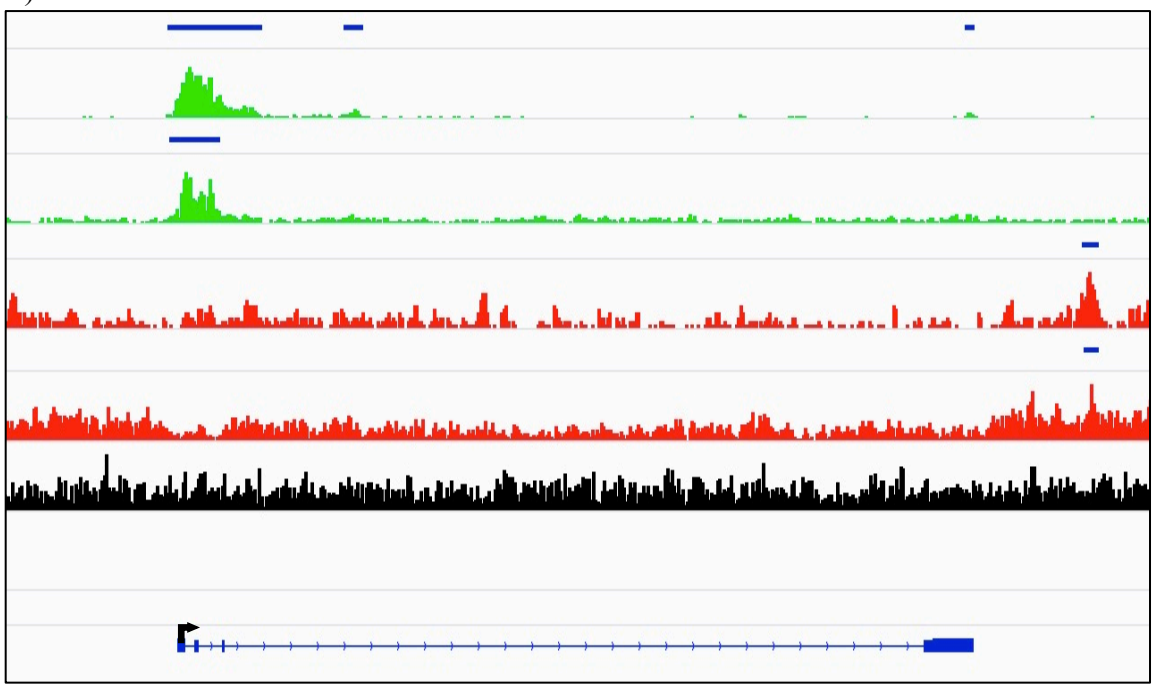
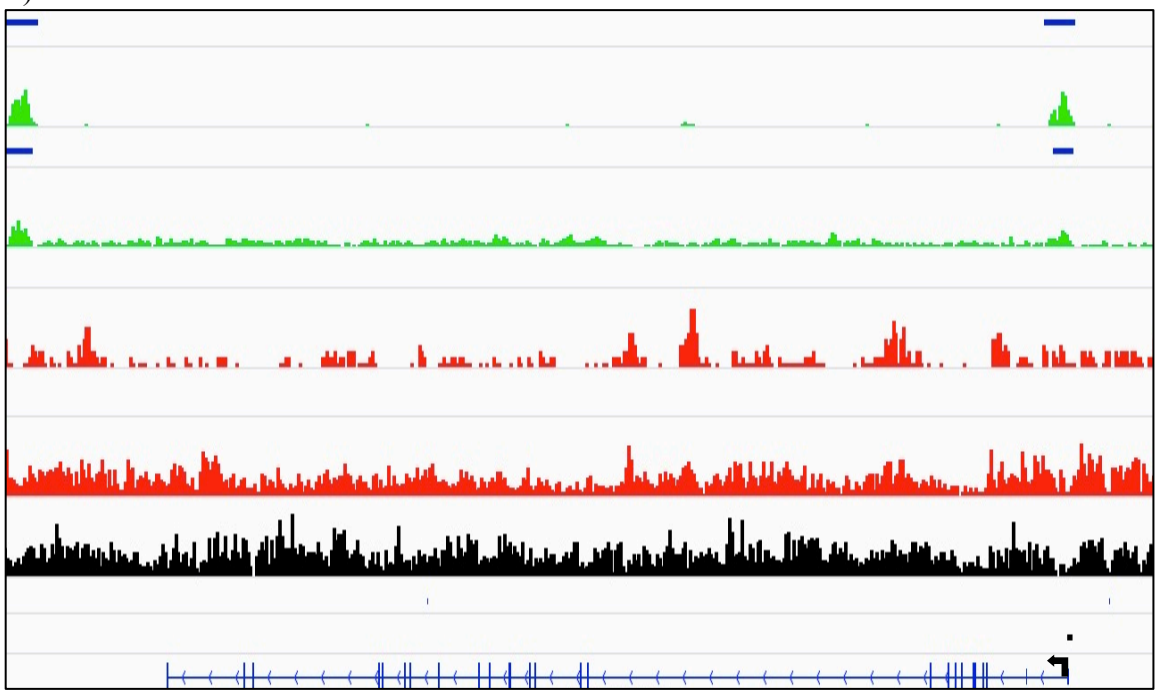

Newly Annotated CpG Island @ Intron 11
H3K4me3
H3K9Ac
Unmethylated
Methylated

Figure A12.

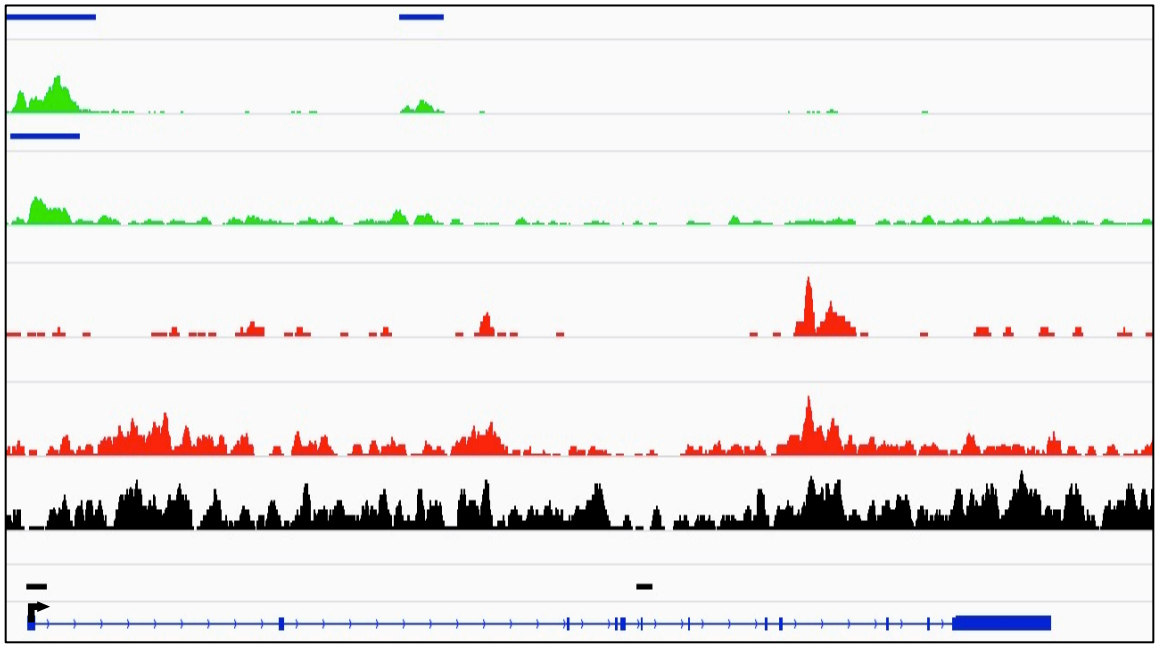
A) *Htr2a*



B) *L3mbtl*



C) *Mest*



APPENDIX B

SUPPLEMENTAL TABLES

Table B1. Chromosome-wide imprinted XCI status profile and Illumina RNA-seq SNP counts in opossum fetal brain and EEM. (A) SNP count summary for escaper genes from fetal brain and EEM RNA-seq data. (B) Table S1B. SNP count summary for nonescaper genes from fetal brain and EEM RNA-seq data. (C) chromosome-wide imprinted XCI status profile in opossum fetal brain. **(D)** Chromosome-wide imprinted XCI status profile in opossum EEM. For table see attached file Table S1.xlsx.

Table B2. Allele-specific pyrosequencing verification results for escaper genes and selected non-escaper genes in brain and EEM. (A) SNP genotyping results in the F1 embryo and their parents by Sanger sequencing for selected escaper genes. (B) SNP genotyping results in the F1 embryos by Sanger sequencing for selected non-escaper genes. (C) Allele-specific pyrosequencing verification results. For table see attached file Table S2.xlsx.

Table B3. H3K4me3 and H3K27me3 significant peaks on X chromosome in female fetal brain and male fibroblasts ChIP-seq data. (A) Significant H3K4me3 peaks on X chromosome in female fetal brain ChIP-seq data. (B) Significant H3K27me3 peaks on X chromosome in female fetal brain ChIP-seq data. (C) Significant H3K4me3 peaks on X chromosome in male fibroblasts ChIP-seq data. **(D)** Significant H3K27me3 peaks on X chromosome in male fibroblasts ChIP-seq data. For table see attached file Table S3.xlsx.

Table B4. Candidate escaper genes without informative SNPs to quantify the allele-specific expression.

No.	gene ID	gene name	H3K4me3 fold-enrichment	% of H3K27me3 coverage	Female/male expression ratio
1	ENSMODG00000008997	*UBL4A*	13.21	0.00%	1.26
2	ENSMODG00000008581	*SSR4*	18.5	0.00%	1.15
3	ENSMODG00000011547	*NAA10*	12.77	4.98%	1.37
4	ENSMODG00000014984	*THOC2*	10.13	0.00%	1.07
5	ENSMODG00000003985	*ZFP347L*	16.3	1.03%	1.23
6	ENSMODG00000004615	*OGT*	17.18	0.00%	1.19
7	ENSMODG00000004864	*KIF4A*	12.33	4.85%	1.09
8	ENSMODG00000009860	*CCDC22*	13.65	0.00%	1.34
9	ENSMODG00000009765	*KDM5C*	13.21	0.00%	1.34
10	ENSMODG00000011070	*RPS4X*	8.37	0.00%	1.11
11	ENSMODG00000011785	*CSTF2*	18.5	0.00%	1.26

Table B5. Average promoter CpG methylation percentages for 5 escaper and 17 non-escaper genes in female and male fetal brain and EEM from PyroMark sequencing results.

Gene name	CGI_start	CGI_end	pXCI status	# of CpGs	Average CpG methylation percentage			
					female brain	male brain	female EEM	male EEM
FLNA	3105210	3105995	Escaper	10	1.43%	1.58%	4.08%	2.46%
FAM3A	3400962	3401678	Nonescaper	20	4.45%	3.23%	2.40%	1.80%
AMMECR1	5988670	5989919	Nonescaper	10	0.85%	0.56%	0.69%	0.98%
BCAP31	11514655	11515266	Nonescaper	8	0.12%	0.37%	0.40%	0.30%
LAS1L	13387573	13387813	Nonescaper	13	1.10%	0.43%	0.49%	0.35%
LONRF3	27458981	27459730	Nonescaper	14	1.03%	0.79%	1.64%	0.91%
ELF4	31571699	31572038	Nonescaper	11	0.29%	0.61%	0.35%	0.03%
RAP2C	32989986	32990510	Nonescaper	9	2.85%	1.20%	0.72%	3.20%
GPC4	34223543	34224297	Nonescaper	7	1.97%	5.25%	1.43%	2.91%
RSX	35651221	35651426	RSX	23	60.57%	95.03%	56.34%	95.92%
FAM122B	36029151	36029318	Escaper	8	0.63%	0.26%	1.07%	1.87%
DDX26B	36362206	36362830	Nonescaper	7	1.89%	1.95%	2.69%	9.17%
MTMR1	39955076	39955459	Escaper	7	1.14%	0.47%	0.19%	1.21%
CD99L2	40182295	40182801	Escaper	12	0.17%	0.41%	0.65%	0.87%
GPR50	40418664	40419488	Nonescaper	8	0.23%	0.29%	0.09%	0.28%
ATP7A	56159948	56160563	Nonescaper	17	1.29%	0.00%	0.22%	0.47%
FNDC3C1	56480924	56481174	Nonescaper	11	1.99%	0.52%	2.20%	0.92%
PLS3	67094850	67095761	Nonescaper	9	2.55%	0.39%	1.41%	3.70%
AMOT	68740642	68741518	Nonescaper	9	40.98%	36.72%	37.15%	58.91%
TAF1	69944348	69945053	Escaper	10	0.14%	0.25%	0.35%	0.84%
KCTD12B	71564312	71565156	Nonescaper	11	0.88%	0.86%	4.13%	1.98%
KLF8	71604414	71607119	Nonescaper	8	4.30%	0.66%	4.71%	1.16%
IL13RA1	78514312	78514793	Nonescaper	7	0.89%	2.08%	0.00%	2.83%

Table B6. Opossum imprinted XCI escapers and the random XCI status for their human/mouse ortholog genes.

Gene name	Opossum escaping status	Human random XCI status	Mouse random XCI status
FLNA	Escaper	Non-escaper	Non-escaper
RPL10	Escaper	Non-escaper	Non-escaper
PLXNA3	Escaper	Non-escaper	Non-escaper
G6PD	Escaper	Non-escaper	Non-escaper
IKBKG	Escaper	Escaper	Non-escaper
DKC1	Escaper	Non-escaper	Non-escaper
YIPF6	Escaper	NA	Non-escaper
NUP62	Escaper	NA	Non-escaper
HCFC1	Escaper	Non-escaper	Non-escaper
IRAK1	Escaper	Non-escaper	Non-escaper
MECP2	Escaper	Non-escaper	Non-escaper
FRMD7	Escaper	NA	Non-escaper
PHF6	Escaper	Non-escaper	Non-escaper
FAM122B	Escaper	NA	Non-escaper
FAM122A	Escaper	NA	Non-escaper
MTMR1	Escaper	Non-escaper	Non-escaper
CD99L2	Escaper	Non-escaper	Non-escaper
HMGB3	Escaper	Non-escaper	Non-escaper
RBMX	Escaper	Non-escaper	Non-escaper
ATRX	Escaper	Non-escaper	Non-escaper
ALAS2	Escaper	Non-escaper	Non-escaper
TAF1	Escaper	Non-escaper	Non-escaper
NONO	Escaper	Non-escaper	Non-escaper
CENPI	Escaper	NA	Non-escaper

Table B7. Summary of ChIP-Seq data for fibroblast cells using 4 histone modifications and an input control.

Histone Modification	Raw Reads ($X10^6$)	Filtered and Aligned Reads ($X10^6$)	Enrichment Peaks ($p \leq 10^{-5}$)	Overlap with Putative Ensembl Gene Promoters[2]
H3K4me3	97.0	71.3	79,412	16320
H3K9Ac	158.4	112.4	56,719	13420
H3K9me3	74.7	56.9	52,511 (159,734)[1]	4514
H3K27me3	157.3	118.0	16, 592	NA
Input	120.2	78.2	NA	NA

[1] - significant peaks determined using MACs ($p \leq 10\text{-}3$)
[2] - 500 bp upstream to 500 bp downstream of first annotated

Table B8. Summary overlaps of significant peaks with each other, annotated CpG islands, and annotated putative promoters.

	H3K4me3	H3K9Ac	H3K9me3	K4me3+K9Ac	K4me3+ K9Ac+K9me3
H3K4me3 (n=79,412)	----	45,331	----	----	----
H3K9Ac (n=52,511)	47,275	----	----	----	----
H3K9me3 (n=56,719)	6,410	1,615	----	1,531	----
Putative Promoters (n=35,301)	16,320	10,959	3,163	13,176	253
CpG Islands (n=22,441)	11,580	9,061	188	9,319	240
Promoters + CpG Islands[1] (n=10,814)	7,871	6,759	773	6,803	136 (178)[2]

[1] – Annotated CpG Islands within the 5.5 kb range of putative annotated promoters
[2] – Number in parenthesis represents overlapping peaks at promoters and CpGs using a lower level of significance for H3K9me3 peak calls ($p \leq 10\text{-}3$)

Table B9. Candidate-imprinted genes Ensembl Gene ID, Associated common gene name if applicable, 3' UTR coordinates as annotated, and Forward and Reverse PCR Primers.

Ensembl Gene ID	Chr	Start	End	Forward Primer	Reverse Primer
ENSMODG00000000633	1	50888333	50889333	GAACCTCATGCAGTCTGTGAAG	TGTTTGGTAAGGGAAAGCCTAC
ENSMODG00000006639	1	78887646	78888646	TTGAACAGCCAGAACTCAAGAA	GGTTTACTTCAAACAGCAGCAA
ENSMODG00000008434	1	86454736	86455736	ACGACAGAAAGGACCAAGTGAT	TATAGTTGCCCAAAAACCCATC
ENSMODG00000010523	1	101428590	101429590	TCCAGAGAAAGGTTCCAGTGTT	GGTTTGAAGGCTCTTGCTACAC
ENSMODG00000011165	1	108942899	108943899	CTGAACCTCTCCCGCATAATAG	TGAACTGACTCTTACGGAGCTG
ENSMODG00000019402	1	117607423	117608423	AGCTATGCTTGGGAGATCAAAG	AAGTTCCAGAACCAGTTTCCAA
ENSMODG00000011135	1	148353086	148354086	CACATGCCAAAGAAGAAAACCT	CCCCCAACACAAATAAGAAACA
ENSMODG00000000464	1	197125347	197126347	TCCGACTTTGTAGAACCCAGAT	TTCATTCCAAGCCCACTATTCT
ENSMODG00000000164	1	198202335	198203335	CCTTCAAGAACAGGGATAAATG	GGGGAAAGAAGATGATTGAGAA
ENSMODG00000000158	1	198271412	198272412	GATTCAGGGTTCCTCTCCTCTT	GTAAGAACCATCCAGCCATCTC
ENSMODG00000003366	1	228646013	228647013	ACTACAGGTCAAAAGGGCAGTC	GCCTGGATGTGAAAGTTATCAA
ENSMODG00000011362	1	275456081	275457081	TTTGATAAGCCTTCTCCTGCTC	CCTCATCTCACCACACTCTTTG
ENSMODG00000011684	1	276674516	276675516	CCCAGCAATTGAATAAAGGAAC	CTACCCTTCCCCATAAACCATT
ENSMODG00000012905	1	285526641	285527641	ACCCCAGTCATCTCTTGTTTGT	CTGGAGGACACATGATTTTTCA
ENSMODG00000012822	1	285630813	285631813	GAACATGGCCTACTGCTTCTTT	GCTTTGCCTTTATGTTCCTAGC
ENSMODG00000010724	1	333227663	333228663	CATGAGGTTTCTCTCACCACTG	CCCCCATAGGAAGATTACCAA
ENSMODG00000023761	1	333244982	333245982	TTCTCTCACCATGGACTCTCAA	TCATCCATTCCCAGAGTCAATA
ENSMODG00000010717	1	333338010	333339010	AGGAAAAGAAGTAGGGTCAGCA	TGAGGGGTGGAGAAGAATAAGA
ENSMODG00000009937	1	342089584	342090584	TGCAGTTGCACTACTAGGCATT	CAGACAAGTGGTGGAGAAATTG
ENSMODG00000009436	1	343914253	343915253	CATGTCCAGTATGAGTGGTGGA	ACCACTTTAACAGGGAATGGTG
ENSMODG00000007182	1	367060313	367061313	TTGCACCAACTAATGGAAAGTG	TACAGGGATACTTCGGAGCAGT
ENSMODG00000006223	1	372593415	372594415	ATACAACCAACCTTGGCAACTC	AACCTCCACAACCACAGTCTTT
ENSMODG00000003806	1	385459213	385460213	GGGGAATGCTTAAGATTTGTGT	CACCCCATCCCTAAATGAGAT
ENSMODG00000001669	1	389035953	389036953	AGAAATGGGGCATCTGAAACTA	CACACGTCTGATGAAGGTTTTT
ENSMODG00000019490	1	401848708	401849708	ACAGGACACCTTTGTGGAACTT	GGCATGGTTTGGAGTTACTAGG
ENSMODG00000019496	1	401900281	401901281	TCCACACATAGATGGCTGTCTT	TGAACTCACCACTTCCCTCTCT
ENSMODG00000012957	1	436892842	436893842	AAACCAGGCTTGGTCTCATAAA	AAGTAGGGTGAGCCAAATCCTT
ENSMODG00000017051	1	472620853	472621853	TTCAGTGTCACCTGCAATTTCT	AGGAATGTCCCAACTCAGGTAA
ENSMODG00000016314	1	495758719	495759719	ATCCCCTTTCTTCACTCGAATA	AAAAGAAGGTGGAGCAGTTGTC
ENSMODG00000016229	1	499725383	499726383	ACTAAGAGGAAGGAAGCGGTCT	GCCTTTAATGGAAGGAGGATTC
ENSMODG00000015575	1	510096298	510097298	CTTGAACTCTCCCACTCTCCTG	GAGCTATGCCTCTCATCTCCTG
ENSMODG00000014618	1	531430233	531431233	AGGAAGCCACAGTTTGAACATT	TTCTGAATTCCCCTCAGAGCTA
ENSMODG00000011515	1	563790769	563791769	TGGGTGCAAATTATGGAGTACA	TGCAATTATCGATCCATCTGAG
ENSMODG00000011328	1	564790600	564791600	GCCCTTTTTCCTTTTACAGCTT	ACTCCCATGTCTGGAAGTCAGT
ENSMODG00000008637	1	589524319	589525319	AGAAGGCAATGCATACAACTCC	TCTGAGCAGGAGACTTCCTTTC
ENSMODG00000006163	1	600377628	600378197	GTGGTCCAGAAATTCTCAGACAG	GATATGGCCCTGGAAGGAATTA
ENSMODG00000005428	1	605428140	605429140	TTTCACGATCACTCCAGAATTG	TATTGCCGAGAAACTAGGAGGA
ENSMODG00000003919	1	621177657	621178657	CCAGGTTACCATAGAGTTTGACC	GTGGCCCAAACATCAAGTTAGA
ENSMODG00000003396	1	624802971	624803971	GAATTGTTTAGGGCGGGTTTAT	TCATGCTGACTCTACTGCTTGA
ENSMODG00000002285	1	631466134	631467134	AAATTGGGTCAAGGGAACTAGG	GTAGCTTTTGGCAATCCTGAAC

Table B9 continued

ENSMODG00000001803	1	639694124	639695124	AAGCAAAGCGGAGTAAAATCAG	CATCCCAAGTGTGTGATGAAAT
ENSMODG00000008750	1	662002860	662003860	TTCGTGCCTACATTGTGAGTTT	ATGAATGGAATTGGAAGACTCG
ENSMODG00000008486	1	663121133	663122133	TGAGCTATGGCAGAATTTGGTA	ATGTGCATCACTTTCAGCTACA
ENSMODG00000003201	1	698983517	698984517	TTATTGCCAGGACCTTTTCAGT	TGCTACCGGAAACTTTTAATCC
ENSMODG00000011200	1	704999526	705000526	AGCTCGTCTTACAGCAGAGGAA	AGAAGATCAACGGCTTTTTCAC
ENSMODG00000009450	1	716933050	716934050	GGTGGTTTCTACCACACCAAGT	ATACTCAAAGCCAACGAGTGTG
ENSMODG00000009386	1	717367383	717368383	CCTCCTACAGCATCGAGCAC	GGGTGGATTCTTTCGTGATTTA
ENSMODG00000015604	1	717916461	171917461	AAGTTAAAAACCCTGGGAGGAG	AGTTGGGCTCTGCTATTCTGTC
ENSMODG00000001321	2	57215440	57216440	GGAGGAACAGACAGAACAATCC	ACTTCCTCTTTCCCCATGATGT
ENSMODG00000007194	2	78706632	78707632	AGCATTTATTCAGGTGCTGTCC	TTAGTCCAGCAAGAAGTGAGCA
ENSMODG00000011721	2	82226320	82227320	AGGAAATGTCCATTAGCCAGAA	TGCATTCAGTGTTCAAGACTCC
ENSMODG00000001558	2	109874517	109875517	GGATTTGCATTGCAGACTCTAA	TGACATTTGAGCACAGACAGAA
ENSMODG00000004622	2	131134701	131135701	GTTCAGTGGGAGATGGAAGTTT	TATCAATTAGCCCAAGGCAAAG
ENSMODG00000006246	2	139634610	139635610	ATCTGTGCTCAACAGCATCTTC	AACTGCTTCAGCATTCAGATCA
ENSMODG00000008286	2	154429949	154430949	GGGTAAGTTTACAGGCCAAATG	CCAGTCCGAACATAAACTCTCC
ENSMODG00000014128	2	169499702	169500702	TGCATGTTGGGTGTATCCTAAA	CAGCATGGATGAAGACTTTGAG
ENSMODG00000016946	2	191002496	191003496	CAGCATCATGTAACCTGGAGAA	GAGCAGAGACTCAAACTCAGCA
ENSMODG00000014859	2	194012782	194013782	AGAAGATTCGAGAGGGAGAGGT	CTACTGCTCCAAATCCCAAACT
ENSMODG00000014651	2	194867160	194868160	TAAAACCAGTTTGGGAGGAATG	TGCTCAGCCTGAAACTCAAATA
ENSMODG00000024021	2	206114036	206115036	GCAACCTGCTTATCAAGTCTGA	CCGAGAGACAGAGGGAGAGATA
ENSMODG00000007640	2	213236964	213237964	AATAGCAGGGGAGAAGGGTAAC	ACTCCCCACAGGTGAATGTAAA
ENSMODG00000004465	2	225162279	225163279	ACTCTGACCAAGCAGGTTTAGC	CTACCTCACAGGGTTGTTGTGA
ENSMODG00000003779	2	229916066	229917066	GAAAATTCTTCAGCCTGGAACT	GACGATGACCTCTTCTTTCGAG
ENSMODG00000003743	2	229961937	229962937	GCATGGCCTTCCAATAAAGATA	TAAAAATGCAGCCCTTCTATGG
ENSMODG00000003701	2	229995182	229996182	ACAATGAAGGAGCTAATGCACA	CCACCATAAATTCCATTCTGAC
ENSMODG00000007792	2	254268634	254269634	AGGGCTGATGAAACTCCAGATA	TGTCGTGCAAATAAGGGTAGAA
ENSMODG00000008266	2	260064845	260065845	TCTTCTGGATAGAAGTGGCAAT	CAATCAACATACATGGGAGAGG
ENSMODG00000013704	2	275269390	275270390	GTTCTTTCCCTTCCTTCTAGGC	CCCCCAAGACATTCTTCTCTCT
ENSMODG00000012785	2	285218809	285219809	TAATCCTTATCTCCGGCCTGTA	GGCTTTGTAGGTCAGAGGAAGA
ENSMODG00000018800	2	302215607	302216607	AGGTGCTGGAAAGCAAAGTTAT	GATGGATGAGGGAGTTTGTTTC
ENSMODG00000018721	2	307679005	307680005	AATTGGATGTGCTCTCGAAGAT	ACTTTTCTATGGCCAGGGATG
ENSMODG00000018444	2	332904143	332905143	AAAGAATGGGAAGGGCAGATAG	TGTAGTTTGATTCCCCAACTGT
ENSMODG00000006175	2	451158631	451159631	CAAAAAGCTTGAAGACCTCACA	CTGACACGGACTCATCAACAGT
ENSMODG00000004472	2	459548415	459549415	CACCGGTTTCAAAGTGTCTGTA	TTTTGTCATCATCATCCTCGTC
ENSMODG00000004326	2	460038121	460039121	ACCTAAAACTTCTGCAGGCTCA	AATACATAACGGAAGCCAAGGA
ENSMODG00000005319	2	491245071	491246071	GTTTCAGGAGGTCTGGTTCACT	ACTCTGTTCATCTCCCTCCAAC
ENSMODG00000018801	2	495754229	495755229	GTTCTAAATCACCTGGGCAAAG	GTAGGAGCTGGTTTGGCATATC
ENSMODG00000019221	2	505025234	505026234	CTCGTCGAGTTTCTTTGCTACC	AAACAATATGGCCGTGAGAAAC
ENSMODG00000000570	2	520269915	520270915	TCCTACCCATGAGGGATTTTTA	GGGTCTTCACAAAGATCTGCAT
ENSMODG00000016394	2	520272577	520273577	GAGCCCAGTGACACTATTGAGA	GTCCTGCCACTGCAAGATTTAT
ENSMODG00000023140	2	522386355	522387355	GAATGTGGAAAAGGATTTGGAC	CCCATAACTTTCTCCATTCCAG
ENSMODG00000019470	3	16801021	16802021	TAGAGGGAGTTTTCCTCACCTG	GCAAAGACCCTGTCGTTTTAAT
ENSMODG00000025244	3	49543210	49543819	CTGGTACCGTCCACCTGTTCT	GAGGACAGACCTTCATGTGTGA
ENSMODG00000015548	3	52330287	52331287	ACCATGCTCCAGCATAAGATTT	TTTCAGGAACTGGTGAAACAGA
ENSMODG00000003564	3	69451400	69452400	CAACTACCCCTGCTACATAGGC	TCAGACCTCAGAATGGTGAAAA

Table B9 continued

ENSMODG00000000399	3	87470085	87471085	ATCTTCCATTTGCTGGAGACAT	GACACAGTGAGTTGGGGTGATA
ENSMODG00000001446	3	94853941	94854941	TGTGGTTCATCTACCAAAAAGC	GTCCACAGGTAGGTGTCACAAT
ENSMODG00000006950	3	162249049	162250049	TTCAGTCATGATTGTCCTGGTC	CACTTCCACAGCAACAGTATCC
ENSMODG00000006961	3	162533985	162534985	TAAGCCAGGATCAGAACAGAGC	CCATCTTATGGGCTATGACAAA
ENSMODG00000010525	3	194975067	194976067	AAGCTGGTGGAGCTCTCTTCT	CAAGTTTCTCCGGAGTATGTCC
ENSMODG00000020132	3	208749463	208750463	TCGACAACATATATCGCACTCA	AGGTAAAACCCCCATTGAAACT
ENSMODG00000020286	3	232703757	232704757	GTTTGGTACACCACAAGCTTTC	TTTTCCTAGGTTCAGGATGTGG
ENSMODG00000020357	3	240202567	240203567	GATGATGAACTTCGAACTGTGC	AGCAGCATGATTTTCTCCAGAT
ENSMODG00000021462	3	269085279	269086279	ATCTGGACCAAGTGCAAATACC	CCCGGTCAATGCTAATAACCTA
ENSMODG00000006222	3	365999412	366000412	TCTACCAAACTGGCTTTTGACA	GCTGAGCAAGGAAACGTAATTT
ENSMODG00000004746	3	377688004	377689004	CCTCCTAAGTGGTTCAGGAAGA	TCAAACCTAGAAGGTTGCCATC
ENSMODG00000001735	3	404001553	404002553	AGTGAGCTGCTGAGGAAAGAAC	TCAAGGAAATCCTTTCAAGCTC
ENSMODG00000000768	3	421671148	421672148	CCCTTAAGGATGGTTTTCCTCT	TGCTGAGGAGTTGGTATGAGAA
ENSMODG00000000764	3	421755076	421756076	CTCACTGCACATGCTTCATTTT	CTGGCCTACACAGCTTTACTGA
ENSMODG00000006967	3	429190670	429191670	TGTAATTAAGACGCCAGGGAGTA	GACTCGAGGACAAGGTGAAGAC
ENSMODG00000011184	3	437574717	437575717	ACCAGAAATGGGAAAAGCTGTA	CAAAAACTTATGGCATGGGAGT
ENSMODG00000011203	3	437738813	437739813	CAAGAGCTGACCAAGCAGTAGA	AATGGAAGAGTCAGGACCTCAA
ENSMODG00000000718	3	440376529	440377529	GCTACAAGTAGCCTGTGCCTCT	AAGTGGCATCAGAGATGAGGTT
ENSMODG00000001178	3	443006006	443007006	GCCAAATTGGATCAAGTACAGG	CTCACTTCACCCGCAGATATTA
ENSMODG00000001236	3	443062526	443063526	GAGAAAGTGATGGAAAGAGAGA GA	GGGGATTTCAAGTCCTTTCACT
ENSMODG00000003841	3	460433663	460434663	ACAAAGGAAGAACTGGCACTTG	GGACAGAAATCCAACAAACTCC
ENSMODG00000003520	3	475399984	475400984	GACCATTACCCTTGAAGTGGAG	CTTACCGGCAAAGATCAGTCTC
ENSMODG00000014519	3	477439465	477440465	AGAACAAAGGGACAAGCAACAT	CAGTTGTTGCTCAACTCTGACC
ENSMODG00000025637	3	499708709	499709709	GGAGAAGGTTCTGGGCTAATCT	GACCTTTTGAATGCAGTGAATG
ENSMODG00000023660	3	501254997	501255997	CAGAGATGTGACCTTTGGCATA	GGGAGAAACCTTATGAATGCAA
ENSMODG00000009557	3	520340337	520341337	ACACGTGTACAGAGGCAGCAC	CCACTTTAGGACATCGACATCA
ENSMODG00000021035	4	10077076	10078076	GGGTGCCGTAAGTTTCTATCAG	AAACAAGCCATTGCAACTATCC
ENSMODG00000021051	4	12639112	12640112	ACCAGAGTGATGTGGTGGAAG	AAAGAAGCAAGCACAGTTTTCC
ENSMODG00000007427	4	41989413	41990413	AAAAACCCCTTTCCACAGTACA	AGATTGTGTCTTCTGCAAGCTG
ENSMODG00000001888	4	62096907	62097907	GCCAATTTCTCCATTCAGTTTC	GCATTGCCTCCTGAAAGAAATC
ENSMODG00000021670	4	62103338	62104338	ACCTGGAAATATGCCTCTCTGA	GTCTGTTGGCCATTCTAGATCC
ENSMODG00000018047	4	67631607	67632607	GAGAGTCCCTCTGTTCTTCCAA	CTGAAGCAAGAAACAGTCCAAA
ENSMODG00000018216	4	83632774	83633774	TTCTTTGCCTTTCCAGATCACT	GTTTCACTCAGCAACACTGGTC
ENSMODG00000000944	4	132780390	132781390	CTAGAAAAACGTGGCGAAGACT	ATCTATGCGCACTGAGGAATCT
ENSMODG00000005628	4	168318450	168319450	TACTCATGGGGAATGTTCACTG	CTTTCCAGTCCTGGGGAACTAT
ENSMODG00000005665	4	168505310	168506310	TACAAAACCCTCACACAGTGGT	CAGCAAAGAAAATCTTGGGTCT
ENSMODG00000009617	4	187738965	187739965	CTCTCAGGTTGGCCTAGAGAAA	GGAAGAGGGATCCATGGTAAAG
ENSMODG00000011664	4	208188190	208189190	CCTTCAGGAATACCAAGTCCAG	CTGGAAGGGTCCAAAGTTACAT
ENSMODG00000012380	4	221510565	221511565	CAGGGATTGCAGTTCATACTCA	GTGACAAGAATGAGGCACAAAA
ENSMODG00000014615	4	245943367	245944367	AGGATATAAGGTGGCCTCAACA	CAGCAATGTAATTTGGTTCCAG
ENSMODG00000000824	4	259665304	259666304	ACATACACACACACGCAGGTTT	AAGTCCGTCAAGTGTCATAGCA
ENSMODG00000004297	4	325137648	325138648	AATCCAGTGGAAGAGAATGGAA	CTTGCAAAATCTAGGCACAGAA
ENSMODG00000004498	4	328701595	328702595	TGCAAGTTTGTCCTGTTGAGTT	GTACACCATCTCCCAGCAAGAC
ENSMODG00000012783	4	359890347	359891347	CTCCTCAACAAAACAAAACACG	GCTGACTTTCTTTGGTGGAATC
ENSMODG00000013020	4	360458330	360459330	TAGCTACAAAACAGCACGCATT	GGCCACTAATGGCATAACCATA

Table B9 continued

ENSMODG00000023715	4	385763442	385764442	CACTGCGGACAAATACCTACAA	CCCTAGTCATTCTAGCCCAGTG
ENSMODG00000009049	4	386060673	386061673	AGCTGGAGCTGGAGATGGAG	CTGGAAAGGAGGGGTGATTT
ENSMODG00000002366	4	387524083	387525083	AACTGCCAAATCATGACTTCCT	GAGAAACAGAAGGGAAAAAGCA
ENSMODG00000006096	4	394682732	394683732	CAAAGACGGTCTGGGATTCTAC	AGCAATATGAGATGCTCAAACG
ENSMODG00000023559	4	401589273	401590273	ATCATGCACCTAACTTGTGCTC	CACATCAGAGAATTCACACTGGA
ENSMODG00000017226	4	429135945	429136945	CCTAACAATAAGCCCAAACTGC	GAAGAAGGAACTGGGGAAAACT
ENSMODG00000012053	5	35181234	35182234	CAGCCTTGACAAGTTGAGTGAT	GTGCTCAAAGTGGTTCTTACCC
ENSMODG00000004762	5	79622982	79623982	TGAAGTTCAGCGTCAACAAGAT	TTGGGGGAACAAAGTACATAGA
ENSMODG00000020706	5	168647869	168648869	AGTAAGGAAGATGCTGGGGAAT	TGCGTAGATGCTAGGGATACAA
ENSMODG00000020677	5	170377569	170378569	GGCAAGGAAAACAAAAGTCCTA	GGGATGAAACCATAGGTGAACA
ENSMODG00000000034	5	195763940	195764940	GTTCCTGCAACCCAACATACTT	ACCACCAGCTAACACTTTTGTG
ENSMODG00000000423	5	203290281	203291281	AGTGATGGTGTCTCTGCAGTTG	ATAAGGAAACTAAGGCCCATGA
ENSMODG00000006522	5	242663942	242664942	TGATCACCCTAGTGCAAATGTT	ACAACAATCCAACCAATTCCTC
ENSMODG00000007335	5	247048855	247049855	TGCAAATAGCTCACTGGCTTTA	ATTGCACCCTTTCACATTCTCT
ENSMODG00000008096	5	250096219	250097219	GAACCTGGCTCACCAGATCTAC	AACACCAAACTTGACAGCAAGA
ENSMODG00000026540	5	250107281	250108281	AAAGAAAGAAAGAAGCCCAACC	CTCCCTCGTCTATTCCCTTTTC
ENSMODG00000009646	5	250270154	250271154	AGGAAACGATATCCGAAAGACA	AGAGAGCTCAAGGACTGGACTG
ENSMODG00000013386	6	2293076	2294076	ATACTGTATTGGGCTCCTCTGG	TTCTGGCTGAATTTGGAGAGAT
ENSMODG00000003495	6	15524842	15525842	CAAAGAACGCACTCATCTGAAC	AGACTCGCAGTACGAGGAGGAT
ENSMODG00000003820	6	36296449	36297449	ACCATTTGGGGGTTTTCTTAGT	CCCAGGATCAGCTTCAATACAT
ENSMODG00000025162	6	36333515	36334515	ATATTCTCCTCCCCTCTGCTTC	GGTTTCCTGGACACTCTGATCT
ENSMODG00000005539	6	54243962	54244962	GGGAAGGAAGGGTTAAGCTACA	GCAGAGGAAATCCTTGACGTAT
ENSMODG00000008046	6	62389852	62390852	ATAGCACTTTGCTGTTTGAGCA	CTTTTCCAAAGCCCTTACCTTT
ENSMODG00000002641	6	94039653	94040653	CCTACTTGTGCCATGTAAGTGC	AAGTGGTTCCATTTTCAAGCTC
ENSMODG00000003635	6	102275090	102276090	GGAATAAGGAAGTCACCACTGC	ACAAGGAAAAGCTCGAATCAAC
ENSMODG00000004516	6	236726356	236727356	CAGAGAAGATCCCTCCTCCTTT	GGGCTACTTTACAGCTTTTGGA
ENSMODG00000004830	6	242565134	242566134	AGGTTTTCAGCACCAACTGAAC	ATCTGGCTCAGCCTTACTTCAC
ENSMODG00000004766	7	82619875	82620875	CGACTTTGAAAGCAGGAGAACT	CTGATTTGCCTAATTGCTGACA
ENSMODG00000015025	7	171107979	171108979	GTCTTGAGGCAACTTGAACCTC	ATTCTCAAGATGCTGACCATGA
ENSMODG00000015605	7	178177078	178178078	TCAGTTGCTGATTGTCCATCTT	ATCACCAACCATCATTACACCA
ENSMODG00000015703	7	244712188	244713188	ACGGCCACAAGAGGAAAAAGT	CCAAGACAGGAGTCAAGATTCG
ENSMODG00000014421	8	25879609	25880609	ATGGTACAGGCTTGGACACTTT	CAGAGAGTTGGGAGCCATTTAG
ENSMODG00000020516	8	62627976	62628976	TGTAATGGAGTGATACGCTTGG	AAATCCAATATCTCCCCCAACT
ENSMODG00000017458	8	93333479	93334479	TATTTCAGGGTGTTGATGATGC	GCCCTTAACTATTTTCCACTGC
ENSMODG00000018312	8	114355198	114356198	GACTTCTCGAGAGCCAAAGAAC	TCTTCTAGGATTGCCTGAAAGC
ENSMODG00000018401	8	117841274	117842274	CACTCAGTGGAGTGGAACAAGT	CACATCAAAACTGATGGCAAAG
ENSMODG00000018521	8	122746963	122747963	CTGGCATAACAGCAAACTTGAA	CAGATTTTAGGGGGTTACAGGA
ENSMODG00000016607	8	126406300	126407300	ACCTTACCAAATACCCAGATGC	TACCACATTGCTTTTCATCACC
ENSMODG00000014499	8	190464437	190465437	ATGAAGGCAAGGTACTCTCAGC	TAGATTGGCCTTAGGAGATGGA
ENSMODG00000013750	8	201464609	201465609	CACTTCCAAGCTGCTACCTACC	CCATGTTCTTCCTAGCTGTGTG
ENSMODG00000013737	8	201570807	201571807	CCTTTTGGATCTGCTCAAGAAC	AGGCTCAAGGCTTTAGAATGTG
ENSMODG00000006521	8	232504676	232505676	ATGGTGGTTGAAACTACGACCT	ACTGAGAGCTCTGGAATCTGCT
ENSMODG00000008021	8	244170301	244171301	CTGGAGAGAGGAAAGAAGGCTA	GGCAAAGTTAGACTTGGAATGC
ENSMODG00000009719	8	256334468	256335468	CGCAGAAAGCCATGATACATT	TGCTGCTACTGACGTCACAATA

Table B9 continued

ENSMODG00000009818	8	256730655	256731655	CTGAAAATTGCAGTTTGACCAC	CAACACCATGATGATCTGGAA
ENSMODG00000023924	8	304139473	304140473	CATCGACTTCCTCTACCAGGTC	GGTGGAATTTGGGAGTTTACAA
ENSMODG00000021336	Un	7775211	7776211	CATCAAGTTCGACCTGAACAAG	ACTAATCACGGTTGGGGAATGT
ENSMODG00000020581	Un	38479834	38480834	ATTAAACCCCATGAGGAAAACC	GCGTTGGTCTGAGTGAATAAAG
ENSMODG00000007100	2	442547176	442548176	GGGAGAGAAGGAGAGAAAAGGA	GAAGCACCCGAGAAACTAAAGA

Table B10. SNP variation detected between individuals A0563 (LL1) X A0573 (LL2).
42 total SNPs. Genes chosen for SNP confirmation are indicated by an asterisk.

Chromosome	Start	End	Ensembl Gene ID	SNP A0563	SNP A0573	POSITION
CHR1	50888330	50889340	ENSMODG00000000633	A	AG	50888587*
CHR1	108942890	108943900	ENSMODG00000011165	G	AG	108943512
CHR1	148353080	148354090	ENSMODG00000011135	A	AG	148353432
CHR1	285526640	285527650	ENSMODG00000012905	T	CT	285527301*
CHR1	285630810	285631820	ENSMODG00000012822	C	CT	285631226
CHR1	333227660	333228670	ENSMODG00000010724	C	CT	333227821*
CHR1	333244980	333245990	ENSMODG00000023761	AG	A	333245463*
CHR1	333338010	333339010	ENSMODG00000010717	G	C	333338396
CHR1	343914250	343915260	ENSMODG00000009436	C	GC	343915130*
CHR1	510096290	510097300	ENSMODG00000015575	A	AT	510096540
CHR1	589524310	589525320	ENSMODG00000008637	C	T	589524693
CHR1	624802970	624803980	ENSMODG00000003396	AG	A	624803564*
CHR1	717916460	717917470	ENSMODG00000015604	C	CT	717917106
CHR2	78706630	78707640	ENSMODG00000007194	G	AG	78707300
CHR2	169499700	169500710	ENSMODG00000014128	AG	A	169499823
CHR2	194012780	194013790	ENSMODG00000014859	A	AG	194013292
CHR2	275269390	275270390	ENSMODG00000013704	AG	G	275269467
CHR2	285218800	285219810	ENSMODG00000012785	C	A	285219179*
CHR2	302215600	302216610	ENSMODG00000018800	A	AC	285219179*
CHR2	307679000	307680010	ENSMODG00000018721	A	G	307679839*
CHR3	16801020	16802030	ENSMODG00000019470	AT	T	16801290*
CHR3	49543210	49543820	ENSMODG00000025244	A	AG	49543500*
CHR3	421671140	421672150	ENSMODG00000000768	G	GT	421671879
CHR3	437574710	437575720	ENSMODG00000011184	G	T	437574828*
CHR3	460433660	460434670	ENSMODG00000003841	T	CT	460434242*
CHR3	501254990	501256000	ENSMODG00000023660	A	AG	501255667
CHR4	12639110	12640120	ENSMODG00000021051	GT	G	12639255
CHR4	41989410	41990420	ENSMODG00000007427	T	A	41989927
CHR4	62096900	62097910	ENSMODG00000001888	G	C	62097383
CHR4	67631600	67632610	ENSMODG00000018047	G	T	67632141
CHR4	132780390	132781390	ENSMODG00000000944	G	AC	132780626
CHR4	168318450	168319450	ENSMODG00000005628	A	G	168319167

Table B10 continued

CHR4	259665300	259666310	ENSMODG00000000824	AG	G	259665471*
CHR5	242663940	242664950	ENSMODG00000006522	A	AT	242664163
CHR6	236726350	236727360	ENSMODG00000004516	G	AG	236726484
CHR6	242565130	242566140	ENSMODG00000004830	G	A	242565229
CHR7	244712180	244713190	ENSMODG00000015703	AC	C	244712322
CHR8	25879600	25880610	ENSMODG00000014421	C	CT	25879766
CHR8	62627970	62628980	ENSMODG00000020516	G	AG	62628747*
CHR8	201570800	201571810	ENSMODG00000013737	T	AG	201570943
CHR8	244170300	244171310	ENSMODG00000008021	C	T	244171254
CHR8	256730650	256731660	ENSMODG00000009818	A	AT	256731126*

Table B11. SNP variation detected between individuals A0552 (LL1) X A0568 (LL2).
42 total SNPs. Genes chosen for SNP confirmation are indicated by an asterisk.

				SNP		
Chromosome	Start	End	Ensembl Gene ID	A0552	A0568	POSITION
CHR1	276674510	276675520	ENSMODG00000011684	AT	AT	276675391
CHR1	285526640	285527650	ENSMODG00000012905	T	CT	285527301*
CHR1	285630810	285631820	ENSMODG00000012822	A	AG	285631526
CHR1	333227660	333228670	ENSMODG00000010724	T	GT	333228053*
CHR1	333244980	333245990	ENSMODG00000023761	A	AG	333245463*
CHR1	333338010	333339010	ENSMODG00000010717	G	C	333338396
CHR1	343914250	343915260	ENSMODG00000009436	C	CG	343915130*
CHR1	663121130	663122140	ENSMODG00000008486	CT	C	663121514*
CHR2	109874510	109875520	ENSMODG00000001558	C	CT	109874653
CHR2	139634610	139635610	ENSMODG00000006246	C	G	139635294*
CHR2	194012780	194013790	ENSMODG00000014859	CT	T	194013302
CHR2	206114030	206115040	ENSMODG00000024021	AG	G	206114857
CHR2	225162270	225163280	ENSMODG00000004465	C	T	225162765
CHR2	275269390	275270390	ENSMODG00000013704	AG	G	275269467
CHR2	285218800	285219810	ENSMODG00000012785	AC	C	285219179*
CHR2	302215600	302216610	ENSMODG00000018800	G	GT	302216483*
CHR2	307679000	307680010	ENSMODG00000018721	A	G	307679357*
CHR2	332904140	332905150	ENSMODG00000018444	C	CT	332904537
CHR2	442547170	442548180	ENSMODG00000007100	A	A	442547809*
CHR2	459548410	459549420	ENSMODG00000004472	G	AG	459548616
CHR2	505025230	505026240	ENSMODG00000019221	G	A	505025646
CHR2	520272570	520273580	ENSMODG00000016394	G	CG	520270108
CHR3	49543210	49543820	ENSMODG00000025244	A	G	49543500*
CHR3	232703750	232704760	ENSMODG00000020286	G	AG	232704431
CHR3	269085270	269086280	ENSMODG00000021462	G	AG	269085368*
CHR3	377688000	377689010	ENSMODG00000004746	C	CT	377688211
CHR3	437574710	437575720	ENSMODG00000011184	G	GT	437574828*
CHR3	437738810	437739820	ENSMODG00000011203	G	AG	437739703

Table B11 continued

CHR3	460433660	460434670	ENSMODG00000003841	T	C	460434242*
CHR4	10077070	10078080	ENSMODG00000021035	T	GT	10077387*
CHR4	12639110	12640120	ENSMODG00000021051	GT	G	12639255
CHR4	41989410	41990420	ENSMODG00000007427	T	AT	41989927
CHR4	62096900	62097910	ENSMODG00000001888	CT	C	62097717
CHR4	67631600	67632610	ENSMODG00000018047	T	C	67631863
CHR4	168318450	168319450	ENSMODG00000005628	AG	A	168319167
CHR4	259665300	259666310	ENSMODG00000000824	AG	G	259665471*
CHR5	250270150	250271160	ENSMODG00000009646	AG	A	250270818*
CHR6	2293070	2294080	ENSMODG00000013386	C	CT	2293331
CHR6	94039650	94040660	ENSMODG00000002641	AG	G	94040117
CHR8	25879600	25880610	ENSMODG00000014421	C	T	25879766
CHR8	62627970	62628980	ENSMODG00000020516	CT	T	62628670*
CHR8	256730650	256731660	ENSMODG00000009818	A	AT	256731126*

Table B12. SNP variation detected between individuals A0567 (LL2) X A0578 (LL1).
49 total SNPs. Genes chosen for SNP confirmation are indicated by an asterisk.

Chromosome	Start	End	Ensembl Gene ID	SNP A0567	SNP A0578	POSITION
CHR1	50888330	50889340	ENSMODG00000000633	G	AG	50888812*
CHR1	108942890	108943900	ENSMODG00000011165	A	AG	108943512
CHR1	148353080	148354090	ENSMODG00000011135	AT	T	148354005
CHR1	276674510	276675520	ENSMODG00000011684	T	AT	276675391
CHR1	285526640	285527650	ENSMODG00000012905	T	CT	285527301*
CHR1	285630810	285631820	ENSMODG00000012822	A	AG	285631526
CHR1	333227660	333228670	ENSMODG00000010724	GT	T	333228053*
CHR1	333244980	333245990	ENSMODG00000023761	AG	A	333245463*
CHR1	333338010	333339010	ENSMODG00000010717	C	G	333338396
CHR1	343914250	343915260	ENSMODG00000009436	C	CG	343915130*
CHR1	401900280	401901290	ENSMODG00000019496	C	T	401900959
CHR1	589524310	589525320	ENSMODG00000008637	C	CT	589524693
CHR1	624802970	624803980	ENSMODG00000003396	A	AG	624803564*
CHR1	663121130	663122140	ENSMODG00000008486	C	CT	663121514*
CHR2	57215440	57216440	ENSMODG00000001321	CT	T	57215808
CHR2	78706630	78707640	ENSMODG00000007194	AG	A	78706823
CHR2	109874510	109875520	ENSMODG00000001558	C	CT	109874653
CHR2	139634610	139635610	ENSMODG00000006246	G	C	139635294*
CHR2	206114030	206115040	ENSMODG00000024021	G	AG	206114857
CHR2	275269390	275270390	ENSMODG00000013704	G	AG	275269467
CHR2	285218800	285219810	ENSMODG00000012785	A	AC	285219179*
CHR2	307679000	307680010	ENSMODG00000018721	G	A	307679839*
CHR2	442547170	442548180	ENSMODG00000007100	A	C	442547809*
CHR2	459548410	459549420	ENSMODG00000004472	A	AG	459549205
CHR2	505025230	505026240	ENSMODG00000019221	A	AG	505025646
CHR2	522386350	522387360	ENSMODG00000023140	CT	C	522386432

Table B12 continued

CHR3	16801020	16802030	ENSMODG00000019470	T	G	16801298*
CHR3	49543210	49543820	ENSMODG00000025244	AG	AG	49543500*
CHR3	269085270	269086280	ENSMODG00000021462	AG	G	269085368*
CHR3	377688000	377689010	ENSMODG00000004746	C	CT	377688068
CHR3	437574710	437575720	ENSMODG00000011184	GT	G	437574828*
CHR3	437738810	437739820	ENSMODG00000011203	AG	G	437739703
CHR3	443006000	443007010	ENSMODG00000001178	T	CT	443006178
CHR3	460433660	460434670	ENSMODG00000003841	C	T	460434242*
CHR3	499708700	499709710	ENSMODG00000025637	AG	A	499709016
CHR4	10077070	10078080	ENSMODG00000021035	GT	T	10077387*
CHR4	12639110	12640120	ENSMODG00000021051	G	GT	12639255
CHR4	41989410	41990420	ENSMODG00000007427	A	T	41989927
CHR4	67631600	67632610	ENSMODG00000018047	CT	C	67632076
CHR4	221510560	221511570	ENSMODG00000012380	AG	A	221511019
CHR4	259665300	259666310	ENSMODG00000000824	G	AG	259665471*
CHR5	250270150	250271160	ENSMODG00000009646	T	CT	250270554*
CHR6	2293070	2294080	ENSMODG00000013386	CT	C	2293331
CHR6	54243960	54244970	ENSMODG00000005539	T	C	54244551*
CHR8	25879600	25880610	ENSMODG00000014421	T	CT	25879766
CHR8	62627970	62628980	ENSMODG00000020516	CT	T	62628207*
CHR8	201570800	201571810	ENSMODG00000013737	G	AG	201571369
CHR8	232504670	232505680	ENSMODG00000006521	A	AC	232505189
CHR8	256730650	256731660	ENSMODG00000009818	T	AT	256731126*

Table B13. SNP variation detected between individuals A0566 (LL2) X A0549 (LL1).
38 total SNPs. Genes chosen for SNP confirmation are indicated by an asterisk.

				SNP		
Chromosome	Start	End	Ensembl Gene ID	A0566	A0549	POSITION
CHR1	50888330	50889340	ENSMODG00000000633	AG	G	50888812*
CHR1	285526640	285527650	ENSMODG00000012905	CT	T	285527301*
CHR1	285630810	285631820	ENSMODG00000012822	CT	C	285631226
CHR1	333227660	333228670	ENSMODG00000010724	GT	GT	333228053*
CHR1	343914250	343915260	ENSMODG00000009436	G	CG	343915130*
CHR1	624802970	624803980	ENSMODG00000003396	A	G	624803564*
CHR1	663121130	663122140	ENSMODG00000008486	C	CT	663121514*
CHR2	139634610	139635610	ENSMODG00000006246	CG	C	139635294*
CHR2	275269390	275270390	ENSMODG00000013704	CT	C	275269633
CHR2	285218800	285219810	ENSMODG00000012785	C	AC	285219179*
CHR2	302215600	302216610	ENSMODG00000018800	T	G	302216483*
CHR2	307679000	307680010	ENSMODG00000018721	G	AG	307679357*
CHR2	332904140	332905150	ENSMODG00000018444	AG	G	332904873
CHR2	505025230	505026240	ENSMODG00000019221	A	AG	505025646
CHR2	522386350	522387360	ENSMODG00000023140	T	C	522386432

Table B13 continued

CHR3	16801020	16802030	ENSMODG00000019470	A	AT	16801605*
CHR3	49543210	49543820	ENSMODG00000025244	G	AG	49543590*
CHR3	377688000	377689010	ENSMODG00000004746	CT	C	377688211
CHR3	421671140	421672150	ENSMODG00000000768	GT	GT	421671879
CHR3	443006000	443007010	ENSMODG00000001178	CT	T	443006178
CHR3	460433660	460434670	ENSMODG00000003841	C	CT	460434242*
CHR3	499708700	499709710	ENSMODG00000025637	G	AG	499709016
CHR4	10077070	10078080	ENSMODG00000021035	G	T	10077387*
CHR4	41989410	41990420	ENSMODG00000007427	A	T	41989927
CHR4	67631600	67632610	ENSMODG00000018047	CT	C	67632076
CHR4	132780390	132781390	ENSMODG00000000944	AC	C	132780626
CHR4	187738960	187739970	ENSMODG00000009617	CT	T	187739279
CHR4	259665300	259666310	ENSMODG00000000824	G	AG	259665471*
CHR4	360458330	360459330	ENSMODG00000013020	A	AT	360458582
CHR5	242663940	242664950	ENSMODG00000006522	AT	A	242664163
CHR5	250270150	250271160	ENSMODG00000009646	A	A	250270818*
CHR6	2293070	2294080	ENSMODG00000013386	C	AC	2293370
CHR6	36333510	36334520	ENSMODG00000025162	AG	A	36333856*
CHR6	54243960	54244970	ENSMODG00000005539	CT	T	54244551*
CHR6	94039650	94040660	ENSMODG00000002641	G	AG	94040117
CHR7	178177070	178178080	ENSMODG00000015605	CT	C	178177875
CHR7	62627970	62628980	ENSMODG00000020516	CT	T	62628207*
CHR7	201570800	201571810	ENSMODG00000013737	AT	T	201570943

Table B14. Summary of Pyrosequencing results for *Meis1*, *Cstb*, and *Rpl17.* Animal ID and cross type are indicated (1 - A0xxx - LL1 X LL2 and 2 - A0xxx- LL2 X LL1). The parental genotypes are shown with maternal allele listed first followed by reference (Ref.) and alternative (Alt.) alleles and their respective expression percentages. SNPs identified in Supplemental Figure F3 were used to assay allele specific expression for *Meis1* and *Cstb*. Genotypes for *Rpl17* were inferred from the PCR-Seq data due to the lack of quality Sanger reads for the gDNA.

Animal ID	Cross Type	Gene	Genotype	Ref. Allele	Alter. Allele	% Ref Allele	% Alt Allele
A0690	1	*Meis1*	A/A	A	G	NM	NM
A0694	1	*Meis1*	G/G	A	G	0.0%	100.0%
A0695	1	*Meis1*	G/A	A	G	23.1%	76.9%
A0719	2	*Meis1*	A/G	A	G	74.8%	25.2%
A0727	2	*Meis1*	A/G	A	G	93.2%	6.8%
A0690	1	*Cstb*	T/G	T	G	0.0%	100.0%
A0694	1	*Cstb*	G/T	T	G	0.0%	100.0%
A0695	1	*Cstb*	T/T	T	G	100.0%	0.0%
A0719	2	*Cstb*	G/T	T	G	1.6%	98.4%
A0727	2	*Cstb*	G/T	T	G	15.6%	84.4%
A0690	1	*Rpl17*	G/T	G	T	0.0%	100.0%
A0694	1	*Rpl17*	G/T	G	T	0.0%	100.0%
A0695	1	*Rpl17*	A/T	G	T	0.0%	100.0%
A0719	2	*Rpl17*	G/T	G	T	0.0%	100.0%
A0727	2	*Rpl17*	G/T	G	T	0.0%	100.0%

Table B15. Bisulfite PCR primers for *Meis*, *Cstb*, *Rpl17*, and *Igf2r*. Primers designed using Methyl Primer Express Software (Applied Biosystems, Inc.) to target the promoter CpG islands in bisulfite treated DNA. Two primers produced amplicons for *Cstb*. For *Igf2r*, we designed two primers each for the promoter CpG island and the CpG island at intron 11.

Gene	CpG Island Location	Forward and Reverse Primers	Amplicon Size (bp)
Meis1	chr1:624957358-624957620	F-GATTTAGGGTTGGAGAAAGTTAG R-CAAAAAAAAAAAAAATCCCTCT	205
Cstb_1	chr4:10081961-10082165	F-ATTTATTGTTTAAAAGTGGGAGG R-AAAAACAAAAAACTCAAATTTCC	274
Cstb_2	chr4:10081961-10082165	F-ATGGAAGGAAGGAGTTTAGTT R-AAATTCTTATCTTAAAAAAATCAACCT	274
Rpl17	chr3:437577629-437578003	F-GGAAAAAGTTTTTGGAAATTGT R-AAAATTAACCAAATAACAACCCC	175
Igf2r Promoter	chr2:442405660-442406525	F-ATATTGGTTATAGGGATAAGGTTAGG R-CATAAACTTCCCAAAATACTTCAC	283
Igf2r Promoter	chr2:442405660-442406525	F-TTTGAGATGAGTGTTAGAAAATT R-AACTAATAACCCCTAATCCATAA	157
Igf2r Intron 11	chr2:442405660-442406525	F-AAGTGGTAAAAGGTTTTTTAATGTT R-AAATCTTTAATCATTTCCTCCC	224
Igf2r Intron 11	chr2:442405660-442406525	F-TTTATTTAGTTAAATTGTTTGGAAGAA R-AAAAAAACCCAATAAAAAAACC	161

Table B16. SNP genotyping results in the F_1 and P generations by Sanger sequencing for candidate-imprinted genes. For table see attached file Table S4.xlsx.