Genomic Imprinting in the Brain: the persistent influences from Mom and Dad

Genomic Imprinting in the Brain: the persistent influences from Mom and Dad

A dissertation presented

by

Julio D Perez

to

The Department of Molecular and Cellular Biology

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Biochemistry

Harvard University

Cambridge, Massachusetts

November 2014

Dissertation Advisor: Professor Catherine Dulac By: Julio D Perez

Genomic Imprinting in the Brain: the persistent influences from Mom and Dad

Abstract

Most mammalian genes are equally expressed from the two inherited parental alleles. However, a puzzling subgroup known as imprinted genes are preferentially expressed from either the maternally- or paternally-inherited copy. Interestingly, many imprinted genes identified so far are expressed in the brain and mutations cause striking defects in brain development and function, in some cases leading to mental disorders such as autism-spectrum disorders. To better understand the extent of genomic imprinting in the brain and gain insights into its potential roles, I have investigated genomic imprinting in the Cerebellum. The Cerebellum provides several experimental advantages and has interesting functions, some of them recently associated with autism. Using RNA-Seq I have profiled the maternal and paternal transcriptomes in the developing and adult mouse Cerebellum, and uncovered 124 genes under imprinting regulation, 40 of which had not been described as imprinted before. Interestingly, the parental bias of 50% of detected genes are regulated according to age. Furthermore, parental biases appear to substantially vary across adult brain regions and are often not observed in non-brain tissues. Finally, I observed an overrepresentation of genes involved in programed cell death among imprinted genes, suggesting that the phenomenon of imprinting may target this pathway with interesting functional implications.

Table of Contents

Acknowledgments

My work, as presented here, is largely a reflection of Professor Catherine Dulac's ability to develop the mind of an aspiring scientist. For seven years she organized my ideas, corrected my mistakes, and challenged me to aim for nothing less than the highest quality of science, all while preserving my innate enthusiasm for molecular biology. For this I am forever thankful. I also thank past and present members of the Dulac lab for instructing me and creating a supportive work environment, particularly my closest collaborator Dr. Nimrod Rubinstein, and my frequent sources of advice and technical help Drs. Christopher Gregg, Yoh Isogai, Joseph Bergan, Anne Lanjuin, Leigh Needleman and Stephen Santoro. I also thank John Choi who as an undergraduate joined my research and provided vital assistance for the completion of this work.

I would like to thank members of my thesis committee Professors Alex Schier, Takao Hensch and John Rinn for their guidance, encouragement and support. Furthermore, I feel fortunate for having been a part of the MCB Department and the diverse scientific community it represents, which has expanded my curiosity of biology far beyond the scope of my thesis project. I have specially cherished my experience with other members of the 2008 MCO class who enriched my time with fascinating discussions and frequent laughter.

I am forever grateful to Professor Katarina Ribbeck who in 2007 took me as a naïve summer intern and transformed me into a driven, aspiring scientist by 2008. I also thank Dr. Bodo Stern and Professor Andrew Murray for guidance during that process.

Finally, I thank those friends who always support me, especially Freddyson Martinez, as well as my beloved family including my brother Victor, my sister Ruthzaly and her family, my father Julio Victor who remains the example of excellence I strive for, and my mother Minerva whose never-ending love and faith in me is the biggest reason I am here today.

Chapter One

Neurobiology of Imprinted Genes

Part of this chapter will be incorporated in a review in preparation Dulac, C., Rubinstein, N.D., Perez, J.D, Gregg C *Annu Rev Neurosci*, 2016.

INTRODUCTION

A main advantage of diploidy is the capacity to mask recessive mutations in one of the two copies of a given gene. Diploidy, has permitted highly sophisticated multicellular organisms with high number of cell types as they are more likely to survive the high number of mutations accompanying the numerous DNA replications that occur during development (Otto and Gerstein, 2008). Accordingly most mammalian genes are equally expressed from both of their parental alleles. However, the genomes of eutherian mammals contain a subset of genes that are preferentially expressed from one of the two alleles in a parent-of-origin specific-manner, meaning either from the maternal or the paternal copy of the gene. This phenomenon is known as genomic imprinting and genes under this mode of regulation are called imprinted genes. Genomic imprinting is not to be confused with random monoallelic expression in which each cell or cell-lineage randomly selects to express one of the two alleles in a stochastic manner, such that both alleles are equally expressed at the cell-type or tissue level. By contrast, genomic imprinting implies an epigenetically predetermined decision to preferentially express either the maternal or paternal allele in a particular cell-lineage. The outcome of this form of regulation is that the parental genomes are not functionally equivalent and the functions of the imprinted genes in the individual are mainly or uniquely contributed by one of the parents. Thus, imprinted genes represent an evolutionary paradox as their unique mode of regulation seems to void the advantages of diploidy.

In this chapter I will first provide a historical overview of the experiments that lead to the discovery of genomic imprinting more than 30 years ago, and summarize the major advancements the field has experienced since. I will then describe the biological significance of

this puzzling mode of epigenetic regulation, with an emphasis on the role of genomic imprinting in the brain.

HISTORICAL OVERVIEW OF THE STUDY OF GENOMIC IMPRINTING

Discovery of Genomic Imprinting

The concept of epigenetic marks or imprints that are passed across generations was first postulated in 1921 in the insect *Pseudococcus nipae* (Schrader, 1921). In this species, embryos destined to become males heterochromatinize a haploid set of chromosomes producing a functionally haploid state, while the same is not observed for future females. Thus it was postulated that the genomes are differentially marked according to the sex of the organism (Nur, 1990). Later, in the 1970s parent-of-origin effects were also observed in the inheritance of coloration patterns conferred by the "R" allele in maize (Kermicle, 1970), which later lead to the discovery of imprinted genes in plants (Gehring, 2013).

The phenomenon of genomic imprinting in mammals was independently discovered in the early 1980s in the laboratories of Azim Surani and Davor Solter. With the hope of identifying recessive mutations and improve animal breeding, Surani and colleagues attempted to generate parthenogenetic embryos. The experimental protocol consisted of suppressing second polar body exclusion in fertilized eggs to obtain triploid eggs with two maternal pronuclei and one paternal pronucleus. However, after removing either the paternal pronucleus or one of the maternal pronuclei it became clear that only eggs for which the paternal pronucleus was maintained developed to term, while those in which the paternal pronucleus was removed did not (Surani et

al., 1984a). The authors hypothesized that the maternal and paternal genomes carry different "imprints" established during gametogenesis, resulting in the requirement for both parental genomes for successful embryogenesis. Surani later coined the term "Genomic Imprinting" (Surani, 1984) to describe such effects.

Around the same time, the Solter laboratory had extended nuclear transplantation techniques to remove, add, and replace nuclei across different embryonic cells with high efficiency (McGrath and Solter, 1983). Using this protocol they demonstrated that normal development only occurred when a male and a female pronucleus where transplanted to a cell but not when two female or two male pronuclei were transplanted (McGrath and Solter, 1984a). These publications from the Surani and Solter laboratories not only revealed the existence of genomic imprinting, but also suggested immediately a cycle for its regulation from one generation to the next, which has profound epigenetic implications and will be further discussed in chapter 2.

From these findings several important questions emerged: What are the imprints? When and how are they established? How are they maintained and how do they regulate gene expression? How are they overwritten as the genomes are passed to the next generation? Which genes are imprinted? What is their function and when and where do they act? Below I will provide a timeline of the important advancements in the field of genomic imprinting since the initial discovery of this phenomenon that addresses some of these questions. Remarkably, more than 30 years after its discovery intense research on this topic continues to bring surprises and further questions, while continuing to provide major contributions to the fields of epigenetics, gene expression, development, and more recently to the study of the brain.

Milestones in the field of Genomic Imprinting

1970- Identification of genomic imprinting in flowering plants

Analyses of the inheritance of coloration patterns of maize demonstrated the phenotype of a particular allele named R depended on its parent of origin (Kermicle, 1970). Imprinting in plants was further validated in 1982 (Lin, 1982). Further research has revealed striking correlations between genomic imprinting in mammals and plants, including both homologous and analogous regulatory mechanisms, expression in analogous tissues and execution of analogous functions (Gehring, 2013).

1974-Embryonic lethality caused by the maternal inheritance of the mouse Thp allele

The Thp allele, which is linked to the brachyury marker, showed embryonic lethality when inherited from the mother but not from the father (Johnson, 1974). Additional genetic crosses discarded the possibility that this lethality was a result of interactions with the maternal physiology and it was concluded that the Thp allele is differentially modified when passing through the egg. The Thp allele was later shown to be the Igf2r gene (Lau et al., 1994).

1984-Discovery of genomic imprinting in mammals

Nuclear transfer experiments in mouse zygotes allowing the replacement of parental genomes show both parthenogenetic (carrying only maternal chromosomes) and androgenetic (carrying only one paternal allele) embryos die at early developmental stages (McGrath and Solter, 1984b; Surani et al., 1984b).

1985- First mapping of a genomic region under imprinting regulation

Using a combination of genetic crosses between Robertsonian translocations multiple lines carrying uniparental disomies of specific chromosomal regions were generated, some of

which exhibited different phenotypes depending on the parental origin of the disomy (Cattanach and Kirk, 1985). This technique eventually revealed 15 regions of the mouse genome with detectable parent-of-origin phenotypes (Cattanach and Jones, 1994).

1989-Identification of the first human disorder with parent-of-origin specific inheritance

Although various conditions had been correlated with parent-of-origin inheritance, the realization that Prader-Willi Syndrome was caused either by deletions of the paternal chromosome 15q11-13 region or by maternal uniparental disomies of that region served as a definitive evidence of imprinting regulation of this region, and thus the existence of genomic imprinting in humans (Nicholls et al., 1989). Today seven human disorders caused by abnormalities in imprinted genes are annotated in the Online Mendelian Inheritance in Man database, and it is possible that more are yet to be identified (Peters, 2014).

1989-Conflict theory of genomic imprinting

Before the first imprinted gene was discovered, Haig and Westoby proposed an evolutionary explanation for the existence of genomic imprinting in plants with implications for mammals inspired by associated phenotypes (Haig and Westoby, 1989). The theory (described in more detail below) proposes that imprinting arose due to a "conflict of interests" between the parental genomes over maternal resources during pregnancy. This theory has caused some controversy within the field. Several other theories regarding the evolution of imprinting have also been proposed (Keverne, 2012; Varmuza and Mann, 1994; Wolf and Hager, 2006) but this one remains the most cited and in the eyes of many it possesses the strongest experimental support (Wilkins and Haig, 2003).

1991-Discovery of first imprinted gene

Searching for the allele underlying the Thp maternal phenotype, it was discovered that Igf2r is exclusively expressed from the maternal allele, making it the first imprinted gene identified (Barlow et al., 1991).

1991-Imprinting can be regulated according to tissue type

Igf2, the second discovered imprinted gene, was shown to be expressed exclusively from the paternal allele in the embryo with the exception of the choroid plexus and leptomeninges, indicating imprinting can vary according to tissue (DeChiara et al., 1991; Prickett and Oakey, 2012). Indeed, spatial regulation is now regarded as a common characteristic of imprinted genes.

1991-Identification of the first imprinted lncRNA

H19, the third imprinted gene to be discovered was proposed to be a non-coding RNA based on its nucleotide sequence, which was later confirmed (Bartolomei et al., 1991). Today, over 30% of imprinted genes are non-coding RNAs that conduct important roles in the regulation of imprinted expression (Santoro and Barlow, 2011).

1993-DNA methylation is essential for genomic imprinting

Because DNA methylation represents a stable modification to the genome that can be inherited across cell divisions, it was considered the primary candidate in the mechanistic regulation of genomic imprinting (Li et al., 1993). Furthermore, all imprinted loci known at the time show differential methylation of the parental alleles. However, an essential role of DNA methylation in imprinting was determined after the deletion of the Dnmt1 enzyme caused impaired imprinting expression in the known imprinted loci. Today the great majority of imprinted genes are associated with a least one DMR, which appears to be the primary regulator of genomic imprinting (Bartolomei and Ferguson-Smith, 2011).

Mid 1990s-Clustered organization of imprinted genes

During the mid 1990s the number of identified imprinted genes grew significantly, and it became apparent that many of these genes reside within the same genomic regions. Thus, the concept of imprinted clusters was introduced. It was later determined that imprinting of genes within a cluster often rely on the same DMR among other chromatin regulators. Clustering of imprinted genes is believed to facilitate their epigenetic regulation (Reik and Walter, 2001).

1997-First gene exhibiting parentally biased expression rather than complete allelic silencing

In humans, Angelman syndrome is caused by maternal deletion or uniparental paternal disomies of the chromosome 15q11-13 region. The search for specific imprinted genes affected by these aberrations indicated Ube3a as solely responsible. However, Ube3a appeared to be biallelically expressed in multiple studies. It was later shown that Ube3a was imprinted only in the brain and not in an all-to-none fashion (Albrecht et al., 1997). Rather, only preferential expression of the maternal allele over the paternal allele was observed. Today, an approximately equal number of imprinted genes showing parentally biased expression as those showing monoallelic expression have been identified (Khatib, 2007).

1998-First imprinted gene knock-out exhibiting behavioral abnormalities

Expression patterns of imprinted genes reveal that a substantial number of them are preferentially expressed in the brain. Knockout of the paternally expressed Mest gene reveal impaired maternal behaviors including pup retrieval, nest building and lack of placentophagia (Lefebvre et al., 1998). At present, behavioral impairments have been ascribed to other perturbations of imprinted genes including some showing similar phenotypes as the Mest knockout (Davies et al., 2015).

2000-CTCF binds differentially methylated regions to regulate imprinting

The search for mechanisms regulating allele-specific expression of imprinted genes uncovered a dependence of Igf2 and H19 expression on distant enhancers. Due to its role in regulating long distance enhancer interactions CTCF was a candidate for the regulation of this locus. Indeed, CTCF was found to bind to the Igf2-H19 DMR in a methylation sensitive manner, which is essential for imprinting (Felsenfeld and Bell, 2000). Currently, it is suspected that several additional imprinted loci are regulated by CTCF (Franco et al., 2014).

2002-First Genome-Wide Screen of Imprinted Genes

The advent of cDNA microarrays provided the opportunity to identify new imprinted genes using unbiased screens. The first such screen aimed to identify genes differentially expressed between parthenogenetic and androgenetic embryos and revealed the existence of three novel imprinted genes (Mizuno et al., 2002). Additional microarrays studies discovered a total of 15 imprinted genes. The use of microarrays, however, has now been largely replaced by the use of RNA-seq, which provides more precise and sensitive quantification technology.

2004-Generation of viable animals from parthenogenetic zygotes

The nuclear transfer experiments that lead to the discovery of genomic imprinting demonstrated the requirement of the paternal genome, as parthenogenetic embryos (containing only maternal chromosomes) die by midgestation. However, it was later shown that viable parthenogenesis can be achieved following two major modifications (Kawahara et al., 2007; Kono et al., 2004). The first is the combination of a genome from non-growing oocytes, which have not yet established maternal imprints and thus express genes that are normally expressed from the paternal genome, with fully grown oocytes which have already established the marks that would lead to gene silencing. The second is the deletion of two paternally methylated regions that normally control the paternal-specific expression of Igf2 and Dlk1. In the absence of

these regions both genes can be expressed from the maternal genome. Oocytes reconstructed to carry both of these modifications produce viable and fertile female adults. These results indicated that genomic imprinting is the only barrier for parthenogenesis.

2008-Discovery of Zfp57; a putative imprinting-specific chromatin regulator

In humans, mutations in ZFP57 lead to abnormal methylation and expression of multiple imprinted genes (Mackay et al., 2008). This observation is recapitulated in Zfp57 mouse mutants (Li et al., 2008). While Zfp57 itself is not imprinted it interacts with DMRs during early stages of embryonic development and protects them from global demethylation (Quenneville et al., 2011). Zfp57 may thus facilitate identification of novel imprinted genes.

2008-2010-First Genome-wide attempts to characterize imprinted genes

The advent of RNA-Seq allows the accurate quantification of gene expression in a genome-wide manner. This technology was first used to screen for imprinted genes in 2008 by two independent studies one profiling whole E9.5 embryos and the other P2 brains, both of which only found a handful of previously uncharacterized genes and failed to identify large subsets of known imprinted genes (Babak et al., 2008; Wang et al., 2008). In 2010 our group used RNA-seq to profile imprinted genes across multiple developmental stages and regions of the brain (Gregg et al., 2010). These experiments revealed the existence of many previously uncharacterized imprinted genes, several independently validated, and suggested a remarkable regulation of imprinting according to developmental stages and tissue types, which motivated the work presented in later chapters. This study was later criticized as it was suggested that a substantial number of the newly reported imprinted genes may have been caused by an underestimation of the false discovery rate (DeVeale et al., 2012), and that RNAseq was not a valid approach to investigate genomic imprinting (Kelsey and Bartolomei, 2012). Thus,

additional experiments, including more robust statistical tools and independent validations appear necessary to determine the actual extent of imprinted genes in the brain.

2012-First Genome-Wide Methylation study

Advances in next-generation sequencing technologies have also provided the opportunity to characterize DNA methylation at a single-base resolution. This method has been recently applied to the characterization of imprinted DMRs (Xie et al., 2012). This revealed the existence of over 50 imprinted DMRs in somatic tissues, not all of which are currently associated with imprinted genes. Furthermore, it was later shown that over 20 of these DMRs are not established in the germ lines and are not targeted to the common CpG dinucleotides (Wang et al., 2014). These studies, in combination with profiles of other chromatin regulators, should help elucidate the complex mechanisms underlying the spatio-temporal dynamics of imprinted genes.

2013-Erasure of Genomic Imprints is mediated by Tet1

To maintain normal expression of imprinted genes across generations, erasure of imprints in the germ lines is necessary. The discovery of the ability of Tet enzymes to catalyze DNA demethylation provides an attractive candidate for imprinting erasure. Indeed, it was recently demonstrated that Tet1 plays an essential role in this process (Yamaguchi et al., 2013). Further characterization of the role of Tet-dependent demethylation will improve our understanding of the regulation of imprinted genes and might provide opportunities to perform functional epigenetics.

WHAT IS THE BIOLOGICAL SIGNIFICANCE OF IMPRINTING?

Many theories have been formulated that attempt to explain how genomic imprinting evolved. In general, these theories, postulate cooperative or antagonistic interactions between the parental genomes (Wilkins and Haig, 2003). The conflict theory of genomic imprinting is one of the theories that has received experimental support. This theory argues that genomic imprinting arose in promiscuous mating systems, where an individual father is frequently not related to all the progeny of the female he shares an offspring with. In this scenario it is evolutionary advantageous for the paternal genome to increase the demand on the maternal resources invested in its offspring, while from the maternal side it is evolutionary advantageous to distribute its resources equally among all its progeny (which includes resources for the current litter as well as preservation of potential future pregnancies). Thus, it is predicted that genes silenced in the maternal genome have functions that increase demand on maternal resources, while genes silenced from the paternal genome have functions that reduce demand on maternal resources (Haig and Graham, 1991; Wilkins and Haig, 2003).

A different theory, named the coadaptation theory postulates that imprinting provides phenotypic compatibility between the mother and the offspring during nurturing periods by allowing the expression of the same allele in both entities (Wolf and Hager, 2006). However this theory only predicts the existence of maternally expressed genes and fails to explain paternally expressed genes. A third theory, based on the expression patterns of multiple imprinted genes, proposes that imprinting synchronizes gene expression in two important metabolic tissues, the placenta and the developing hypothalamus, to increase postnatal fitness (Keverne, 2012). Although this and other theories provide explanations for imprinting in the embryonic context,

they are currently vague regarding the postnatal endurance of imprinting. Furthermore, contradictory examples regarding each theory have been found.

From the perspective of experimental biology, however, a general and common function or evolutionary drive for imprinted genes, as a collective group, remains unproven. Here I will review the evidence addressing biological relationships between imprinted genes and their potential synergy.

First, after careful evaluation of imprinting evidence in available literature I have compiled a list of 130 known imprinted genes in the mouse. Associations within this list of imprinted genes will be explored at three biological levels: (1) biochemical and cellular functions of imprinted genes, (2) expression patterns of imprinted genes across tissues and developmental stages, and (3) the roles of imprinted genes in the physiology of developing or mature tissues. I will attempt to take an unbiased approach, independent of current interpretation regarding the biology of imprinted genes. Nevertheless, the reader should be aware that an unavoidable limitation of this review comes from the fact that a large portion of the available data was derived from experiments formulated under a limited biological scope and often interpreted in the context of available evolutionary theories. Finally, I will describe how a systematic characterization of the expression of imprinted genes in a well-defined biological framework will improve our understanding of the overall function of genomic imprinting, particularly in the brain. With this in mind I will define the aims and question addressed by the experiments presented in subsequent chapters.

Biochemical and Cellular Functions of Imprinted Genes

Of the 130 known imprinted genes I was able to compile from data in the mouse, 67% are protein-coding, 16% are lncRNAs, 13 are miRNAs, and 7% are snoRNAs. To determine whether particular amino-acid sequences or domains are enriched among protein-coding imprinted genes I used InterPro web-based predictive models. This analysis did not reveal enrichment in any particular amino-acid sequences or functional domains (data not shown).

The role of imprinted genes in specific cellular processes has only been explored for a small subset of highly studied imprinted genes, leaving the majority substantially less characterized at the biochemical and cellular levels. Thus, at this point it is difficult to draw any conclusion, and our knowledge may clearly be biased by the disproportionate information of a limited few. Nevertheless, analyses of gene ontology (GO) categories may provide a macro-level view of cellular functions commonly performed by imprinted genes. A recent publication conducting GO analyses of imprinted genes (Hamed et al., 2012) did not reveal a discernable picture regarding their role in the cell. Only a few terms were enriched including regulation of catalytic activity, intracellular signaling and ion transport. For instance, the maternally expressed Kcnq1, Kcnk9, Slc22a2, Slc22a3, and Slc22a18 genes, and the paternally expressed Slc38a4 gene are all membrane proteins involved in ion transport (Hamed et al., 2012). Notwithstanding, most other enriched GO terms are represented by a maximum of three genes, out of the 48 analyzed, and the majority of genes are not even assigned to any enriched term. This may point to a limitation in either the resolution of GO annotation and/or the fact that imprinted genes participate in such diverse biological functions which do not provide sufficient power for GO

analyses and alike to detect statistically significant term enrichments. Still, a small number of imprinted genes participate in several common specific cellular functions.

Tissue and Developmental stage-specific expression patterns

Expression patterns

While the specific cellular roles of imprinted genes may be diverse it is possible that at the scale of a cell type, or even a tissue, these roles ultimately integrate into a common function. Hence, imprinted genes may coordinately modify the physiology of particular tissues by regulation of some or all the cell types within that tissue. In this scenario, the expression of imprinted genes would be expected to undergo regulation according to tissue type and/or developmental stage.

A study conducted in 2009 explored the expression of 43 mouse imprinted genes across multiple tissues (Steinhoff et al., 2009a). All non-coding RNAs where excluded from this study, as well as genes whose imprinting has only been described in the placenta, since contamination from the maternal decidua has been shown to led to the identification of false "maternally expressed genes"(Okae et al., 2012). This group of imprinted genes appeared neither expressed in a minority of selected tissues nor ubiquitously expressed across all tissues (such as housekeeping genes). While, individually, each gene showed a unique pattern of expression across tissues and developmental time points, subsets of imprinted genes displayed concerted spatiotemporal common denominators. This included a subset of 15 imprinted genes that appeared expressed at early time points in the whole embryo and in the placenta. These genes included Igf2, Igf2r, H19, Grb10, and Cdkn1c, the imprinting of which were discovered in early

embryos, and which carry out important developmental functions (discussed below). Additionally, enrichment of another subset of 10 genes was observed in brain regions, particularly in the hypothalamus and the pituitary gland. This group included genes such as Peg3, Plagl1, Grb10, Ndn, and Rasgrf1 (Steinhoff et al., 2009b). Some of these genes were discovered

in the brain where they are known to carry out important functions (discussed below). Taken together, this points to a pattern by which specific tissues in specific time points are targeted by subsets of imprinted genes probably due to the impact that these genes have on their development and function.

Monoallelic and Biallelic Expression of Imprinted Genes

Although expression of imprinted genes is often described across multiple tissues, the imprinting status of a gene, meaning biallelic or monoallelic expression has usually be analyzed in only one or a limited number of tissues. This raises the question of whether the imprinting status of a gene is stable or altered across large number of tissues. Unfortunately, available databases of expression across multiple tissues do not provide allele-level resolution and thus do not address the imprinting status of genes. However, allele-specific expression across a handful of tissues has been tested for several imprinted genes. Indeed a systematic characterization of allele-specific expression in the brain revealed substantial regulation in the imprinting status of several genes between different brain regions (Gregg et al., 2010). Among 82 imprinted genes for which expression of the parental alleles has been analyzed in at least two tissues, 28% exhibited tissue-specific imprinting (Prickett and Oakey, 2012). Interestingly, the majority of these genes (10 in total) show imprinting only in the placenta, including Slc22a2, Slc22a3, and Nap1l4. Other genes showed imprinting only in the brain (9 in total) including Ube3a, Copg2,

and Calcr. Finally, a smaller number of genes (5 in total) show imprinting only in the placenta and the yolk sac including Tfpi2, Gatm, and Pon2 (Gregg et al., 2010; Prickett and Oakey, 2012).

These data suggest that a subset of genes are specifically imprinted in the brain and/or placenta, which is correlated with the preferential expression exhibited by other imprinted genes in these tissues (Gregg et al., 2010; Prickett and Oakey, 2012). Interestingly, most of the genes specifically imprinted in the brain or placenta are different from those specifically expressed in these two tissue types, respectively. Together, superimposition of the observations of overall expression and expression differences of the parental alleles suggest that a substantial number of imprinted genes, (approximately 60 genes) are targeted through imprinting or expression to the placenta or the brain (Gregg et al., 2010). Expression profiles of more imprinted genes are clearly warranted for corroborating this idea.

Physiological Roles of Imprinted Genes in the Placenta

The pioneer nuclear transfer studies, which lead to the discovery of genomic imprinting, also highlighted an essential role of imprinting in development of embryonic and extraembryonic tissues. Parthenogenetic embryos initially appear normal but die by E10 due to absence of the placental tissue. On the other hand, androgenetic embryos are grossly developmentally delayed and growth restricted, and die by E8.5, while containing an overabundance of placental tissue (Barton et al., 1984). These experiments are consistent with the enriched expression and role of imprinted genes during embryogenesis, particularly in the placenta. Accordingly, the role of imprinting in the physiology of the placenta has been the focus of many studies.

Before summarizing the role of genomic imprinting in the placenta it is necessary to provide a general overview of its organization and physiology. The placenta is derived from trophoblasts, which is the first discernable cell type arising during embryogenesis. At least eight trophoblast-derived cell types contribute to the placenta, each differing in expression pattern, localization, and physiological roles. The placenta is composed of three broad regions: the parietal trophoblast giant cell (PTGC) layer, the junctional zone, and the labyrinth layer. The PTGC layer lines the implantation site and separates the embryo from maternal tissues. The junctional zone is composed of glycogen cells, which provide an energy source during late gestation for embryonic growth and parturition; and Spongiotrophoblasts, which secrete prolactin hormones that modulate maternal physiology during pregnancies. The labyrinth layer integrates both maternal and fetal blood flows, and regulates nutrient transfer to the embryo (Tunster et al., 2013).

Interestingly, the roles of imprinted genes in the placenta appear to be aimed towards its nutrient transfer functions. For instance, paternally expressed gene 10 (Peg10), which is thought to regulate multiple intracellular signaling cascades, is essential for the formation of both junctional zone and labyrinth layer. This gene is expressed in all trophoblast lineages. Although the paternally transcribed mitogen, Insulin growth factor 2, is widely expressed across the placenta, loss of its expression results in drastic reductions specifically of the labyrinth layer and the glycogen cell population. Maternally expressed Ascl2 produces a transcription factor essential for the lineage maintenance of glycogen cells and spongiotrophoblasts. This gene is specifically expressed in the labyrinth layer and junctional zone during a small time window during midgestation. Loss of function of the maternally expressed Phlda2, which encodes a small cytoplasmic protein of unknown function, results in placental overgrowth accompanied by an

overabundance of glycogen cells. This gene is expressed in cell precursors of the junctional zone and labyrinth cell lineage. Important roles in placental function have been demonstrated for several other imprinted genes (Tunster et al., 2013). These results highlight common roles of imprinted genes in placenta physiology and additionally exemplify how they are brought about by a combination of diverse specific biological functions expressed in the different cell types. By generalization, it is likely that this principle applies in other tissues targeted by imprinting, mainly the brain.

Expression of Imprinted Genes in the Brain

Out of all somatic tissues the brain appears to be the most enriched in the expression of imprinted genes. The brain is also the tissue where the highest number of genes is expressed. The expression of large subsets of imprinted genes in the brain could therefore result from the overall molecular complexity of the brain, in which case the highest number of imprinted genes should be found in the most cellularly heterogeneous brain regions such as the cortex. Alternatively, genomic imprinting may target specific functions of the brain, in which case an enrichment of imprinted genes should be found in specialized brain areas.

The Allen Brain Atlas provides a high-resolution map of gene expression across the brain based on in situ hybridization assays. Using this atlas, our group has previously observed that a large fraction of known imprinted genes (45) are expressed in one or more of the 118 brain regions surveyed (Gregg et al., 2010). Furthermore, while the majority of imprinted genes are ubiquitously expressed in the brain around 15 imprinted genes are expressed in only a subset of brain regions. These genes appear mostly expressed in amygdalar, hypothalamic, and

monoaminergic regions. More specifically, this expression is enriched in the medial and central amygdala, as well as in the medial preoptic area and arcuate nucleus of the hypothalamus. Enriched monoaminergic regions include the dopaminergic ventral tegmental area and substantia nigra, the serotonergic dorsal raphe along with other raphe nuclei of the midbrain, and the noradrenergic locus coeruleus. Additionally, the ventral lateral septum, the nucleus accumbens, the bed nucleus of the stria terminalis, and the periaqueductal gray also show strong representation of imprinted genes (Gregg et al., 2010). These results are reminiscent of those observed across the body in which the expression of few imprinted genes is spatially regulated while most are ubiquitously expressed, and suggest the targeting of imprinted gene functions to specific brain regions, in particular brain regions involved in the regulation of social, feeding and motivated behavior.

These measurements of expression, however, could not assess the actual imprinted status of the expressed genes, nor did they distinguish the expression of individual parental alleles and the strength –or absence of parental allelic expression bias. Since genomic imprinting is not necessarily manifested in an all-to-none fashion, or may even be inexistent for some genes known to be imprinted in non-brain tissue, the true extend of genomic imprinting in a given brain regions is still an open question. Furthermore, the extraordinarily high molecular complexity of the brain begs the question on whether or not additional, yet unidentified neural genes may be subject to genomic imprinting.

Genomic Imprinting and Neural Development

Neurodevelopmental abnormalities in uniparental mouse chimeras

In the late 1980s nuclear transfer experiments were employed to study the functional significance of imprinted genes past the earliest embryonic stages (Fundele and Surani, 1994). Rather than generating uniparental zygotes, which did not survived past early embryonic stages, as was done in the initial pioneering studies of imprinting, in this approach mouse chimeras were generated that contained a mix of normal cells together with uniparental cells. The uniparental cells were either parthenogenetic cells (PG), which carried only maternal chromosomes, or androgenetic cells, which carried only paternal chromosomes. Strikingly, these experiments revealed that PG chimeras exhibit larger brains but smaller bodies than controls while AG chimeras exhibit the opposite pattern: smaller brains and bigger bodies (Keverne et al., 1996). It was later shown that this phenotype could be explained by the expression of a single gene, Grb10 (Charalambous et al., 2003, 2010; Garfield et al., 2011), which is maternally expressed in the body and paternally expressed in the brain. These experiments with chimeras also allowed the tracking of uniparental cells within developing tissues. Intriguingly, data showed that the contributions of the two forms of uniparental cells to brain regions were highly different. PG cells were found to localize mainly to the cortex, striatum, and hippocampus but not hypothalamic regions, while AG cells were found to localize mainly to hypothalamic regions but not cortical areas (Keverne et al., 1996). These results suggested that maternal specific expression is essential to the development of higher brain regions while paternal specific expression targets hypothalamic neurons. To this day, it remained unknown how many and which specific genes contribute to these patterns.

Roles of Imprinted Genes in Neural Development

The experiments with uniparental mouse chimeras clearly highlighted the notion that genomic imprinting plays an important role in brain development. However, these experiments could not resolve the specific biological functions regulated by imprinting and the identity of genes responsible for these functions. Below I will describe several key experiments, which have established the involvement of imprinted genes in several aspects of neural developmental functions including neuronal differentiation, migration, and survival. The majority of these experiments have been performed in the amenable experimental setups provided by corticogenesis and cerebellar development and it is therefore still unclear whether the findings apply to the development of other brain regions.

Neuronal differentiation

Role in the regulation of neuronal differentiation has been observed for a handful of imprinted genes. Plagl1 (Pleiomorphic adenoma gene-like 1) is a paternally expressed transcription factor that regulates the expression of multiple other imprinted genes including Igf2, H19, Cdkn1c, Dlk1, and Rasgrf1 (Hoffmann and Spengler, 2012; Hoffmann et al., 2003; Varrault et al., 2006). During embryonic stages Plagl1 is expressed in actively proliferating brain regions including the arcuate and amygdaloid regions and the cerebellum, and is also expressed post-parturition in the subventricular zone (Valente and Auladell, 2001). In the cerebellum Plagl1 expression is restricted to the cerebellar ventricular zone and external granular layer of lobule IX and in two types of GABAergic interneurons generated from each region. Accordingly, Plagl1 paternal deletions exhibit smaller cerebellar lobule IX and reduce numbers of interneurons (Chung et al., 2011). In the neocortical subventricular zone Plagl1 prevents

astroglial differentiation (Schmidt-Edelkraut et al., 2013) and promotes neural differentiation by the activation of the maternally expressed Cdkn1c. Cdnk1c is a regulatory protein that prevents entry to S phase and cell divisions by inhibiting cyclin D kinases (CDKs). Thus, after Plagl1-depedent activation, Cdkn1c induces cell-cycle arrest of neural progenitors, a necessary step for eventual neural identity (Schmidt-Edelkraut et al., 2014). Indeed, regulation of Cdkn1c levels influences multiple aspects of neuronal differentiation during corticogenesis. The balance between CDKs and Cdkn1c controls the production of neurons in all cortical layers but primarily of layers 5 and 6, where lower Cdkn1c levels result in a higher number of neurons in layers 5 and 6.

Thyroid hormone regulates essential processes during brain development including neural differentiation, migration, myelination, and synaptogenesis (Horn and Heuer, 2010). The paternally expressed Type 3 Deiodinase (Dio3) is the enzyme primarily responsible for thyroid hormone inactivation. Interestingly, Dio3 expression is highly regulated across brain regions and developmental stages (Charalambous and Hernandez, 2013), which were recently shown to coincide with regulations in its imprinted expression (Martinez et al., 2014). In the cerebellum, stimulation of thyroid hormone triggers differentiation of granule cell progenitors from the external layer into internal-layer granule cells. Dio3 regulates the timing of this process and its absence ((Ferrón et al., 2012) in homozygous mutants results in premature disappearance of the external layer, expansion of the overlying molecular layer and reduced cerebellar foliation (Peeters et al., 2013). Furthermore, these animals show abnormalities in locomotor abilities controlled by the cerebellum.

The paternally expressed Dlk1 expresses two types of isoforms. Short isoforms, which encode a membrane protein possessing growth-factor domain similar to the Delta-Notch family,

and long isoforms, which encode a protein that is cleaved in the extracellular side of the membrane resulting in a secreted form of DLK1. Although DLK1 appears to be dispensable for embryonic neurogenesis it plays an essential role in postnatal neurogenesis (Ferrón et al., 2012). Postnatal neurogenesis occurs in specialized microenvironments found in the hippocampus and subventricular zone, which contain niche astrocytes essential for this process. These niche astrocytes excrete the secreted version of DLK1. Strikingly, secreted DLK1 is recognized by a membrane bound DLK1 present in neural stem cells (NSC). This interaction is essential for NSC proliferation. Surprisingly, Dlk1 selectively loses imprinting and is biallelically expressed in both NSCs and astrocytes (Ferrón et al., 2012). Additional imprinted genes, such as Rasgrf1, have also been implicated in the regulation of postnatal neurogenesis (Darcy et al., 2013).

Neuronal Migration

A subset of imprinted genes has also been implicated in the regulation of neuronal migration. The maternally expressed Kcnk9 gene encodes a potassium channel that regulates neuron excitability by mediating membrane resting potential. Neural activity plays an essential role in hebbian connectivity, and has also been implicated in the regulation of neuronal proliferation, differentiation, and migration. In the developing cortex, neuronal migration to layers II and III is largely impaired after Kcnk9 shRNA knockdown indiscriminate of alleles. This phenotype is rescued by reintroducing wild-type Kcnk9 but not by dominant negative mutations. Indeed, this phenotype is associated with decreased permeability of migrating neurons, which raises the resting potential and increases spontaneous calcium transients (Bando et al., 2014). How these abnormal currents regulate neuronal migration remains to be elucidated.

Cdkn1c also regulates cell migration through its mediation of cytoskeleton dynamics. The normally cytoplasmic LIM-domain protein kinase is sequestered to the nucleus upon binding of Cdkn1c. This kinase reduces the actin polymerization by inactivating the actin binding factor cofilin, which in turn reduces cell migration (Tury et al., 2012). Indeed, Cdkn1c mutants have impaired cortical cell migration and reduced populations in higher cortical layers, while overexpression of Cdkn1c results in the opposite effect (Tury et al., 2011).

Neuronal Survival

Programmed cell death or apoptosis, is pervasive throughout the developing brain where it plays essential roles in neural patterning and connectivity (Buss et al., 2006). Accordingly, apoptosis is tightly regulated in the brain according to space and time. Multiple imprinted genes have been associated with neuronal survival including the paternally biased Bcl-xL (Gregg et al., 2010), which represents the primary anti-apoptotic signal of brain cells (Kuan et al., 2000). Bcl-xL primarily prevents apoptosis by counteracting the role of pro-apoptotic proteins, which induce the release of mitochondrial intermembrane proteins that irreversibly engage the apoptosis pathway. In the brain, however, Bcl-xL has been postulated as a general protector of neuronal homeostasis preventing neurotoxicity and neurodegenerative stimuli, besides its anti-apoptotic roles. The neuroprotective mechanisms of Bcl-xL include regulation of mitochondrial bioenergetic efficiency, vesicle recycling, prevention of synaptic long-term depression, neural plasticity and impeding glutamate excitotoxicity (Jonas et al., 2014; Li et al., 2010b).

In addition to Bcl-xL, other imprinted genes have been shown to regulate apoptosis in a region-specific manner. The role of paternally expressed gene 3 (Peg3) in apoptosis appears to be context-dependent. In non-neuronal lineages Peg3 expression is induce by p53, which causes

Bax mitochondrial translocation followed by apoptosis (Deng and Wu, 2000; Relaix et al., 2000). However, absence of Peg3 in the brain results in reduced size correlated with increased apoptosis, suggesting Peg3 serves anti-apoptotic functions in the brain. Furthermore, the penetrance of this phenotype varied according to brain region and the sex of the individual (Broad et al., 2009).

Other imprinted genes expressed in the brain associated with apoptosis and overall neural development include Kcnk9, Ndn, Trappc9, Nnat, Cobl, Adam23, and Grb10 (Gregg et al., 2010; Nantel et al., 1999; Patel and Lazdunski, 2004; Takazaki et al., 2002). These genes will not be described in detail, but they highlight the importance of genomic imprinting in neurodevelopmental processes.

Synaptic Function

Synaptic communication between neurons regulates the wiring of the developing brain and the function of the mature brain. Multiple imprinted genes are associated with synaptic regulation and neural plasticity. Of these the maternally biased Ube3a is the best characterized. Ube3a codes for an E3 ubiquitin ligase localized to postsynaptic compartments. In the cortex, maternal deletion of Ube3a causes reductions in dendritic spine density and length. Ube3a deficient neurons exhibit a higher threshold for long-term potentiation (LTP) induction, which are correlated with Ca^{2+}/calmodulin-dependent protein kinase II (CaMKII) phosphorylation of threonine^{305} (Thr^{305}). Introduction of a CaMKII Thr305Ala mutant rescues LTP and associated phenotypes (Mabb et al., 2011). The targets of Ube3a ligase activity involved in its LTP

regulation are still unclear but one promising candidate is Ephexin5, a RhoA guanine nucleotide exchange factor that prevents synaptic formation (Margolis et al., 2010).

Paternally expressed Rasgrf1 is another imprinted gene, which functions in the synapse. Rasgrf1 directly interacts with the NR2B subunit of NMDA receptors (Krapivinsky et al., 2003). In the hippocampus, Rasgrf1 participates in NMDA receptor-mediated long-term depression (LTD) but not LTP (Li et al., 2006). Homozygous mutants as well as paternal deletions of Rasgrf1 show poor performance in contextual-memory tasks that are dependent on the hippocampus (Brambilla et al., 1997; d'Isa et al., 2011; Giese et al., 2001; Li et al., 2006). In contrast, Rasgrf1 in the amygdala regulates LTP involved in fear conditioning (Brambilla et al., 1997).

Other imprinted genes exerting regulatory functions on the synapse include Xlr3b, Xlr4b, Commd1, Adam23, Ago2, Cdh15, Igf2, Igf2r, Snx14, and Impact (Cubelos et al., 2010; Gregg et al., 2010; Roffe et al., 2013). These genes will not be described in detail, but they highlight the importance of genomic imprinting in synaptic regulation.

Imprinted Genes and Functional Circuits

In earlier parts, I have described work showing the preferential expression of imprinted genes in the brain, the dynamic regulation of their expression levels and imprinting status, plus important functions in the development and physiology of neurons. In turn, these findings raise questions about their role in brain activity and behavior: are imprinted genes involved in regulating a wide range of brain functions and behaviors or do they converge on affecting specific aspects of brain activity? An increasing body of information suggests that the latter is

true as the function of imprinted genes appears to be associated with the regulation of specific neural circuits within the brain.

Hypothalamic neurons

The hypothalamus is a highly conserved, ancient brain region which functions as a master regulator of homeostasis by sensing internal states and coordinating appropriate physiological and behavioral responses. Accordingly, the hypothalamus contains a diverse number of cell types that form complex interactions with each other and with other brain regions (Sternson, 2013). As discussed above, the expression of imprinted genes is highly enriched in the hypothalamus. However, a substantial diversity in the expression of imprinted genes exists within hypothalamic cell types. Strikingly, cell types involved in the regulation of feeding and energy consumption as well as social behaviors appear particularly enriched in the expression and function of a subset of imprinted genes.

Orexin neurons

Orexin neurons of the lateral hypothalamus regulate sleep and wakefulness, energy balance and energy consuming behaviors such as locomotor activity (Burt et al., 2011). The maternally expressed gene Kcnk9 is preferential found in orexin positive neurons of the hypothalamus. Glucose levels are detected in the membrane of orexin neurons and induce an intracellular signaling cascade that modulates ion transportation by Kcnk9 and alter the resting membrane potential of orexin neurons. By this mechanism orexin neurons can encode variations in glucose levels physiologically occurring between meals (Burdakov et al., 2006). Deletion of Kcnk9 leads to exaggerated nocturnal activity, which may also increase feeding (Linden et al., 2007).

POMC neurons

Pro-opiomelanocortin (POMC) neurons of the arcuate nucleus are important regulators of food intake, energy expenditure and glucose metabolism (Sternson, 2013). The maternally biased Asb4 is down-regulated in POMC neurons during fasting periods and increase after injection of leptin but not insulin. Interestingly, insertion of an Asb4 transgene with POMC specific expression results in increased food intake but reduced body weight. Furthermore, these animals show increase rates of energy consumption and locomotor activity (Li et al., 2010a).

Oxytocin neurons

Oxytocin (OT) is a neuropeptide with diverse roles in physiological responses and social, reproductive and parental behaviors. OT neurons of the hypothalamic paraventricular nucleus have been particularly implicated in feeding behavior where they enhance response to satiety signals. Disruption of synaptic release of OT neurons results in overeating (Zhang et al., 2011). Deletions of the paternally expressed genes Peg3 and Ndn result in reductions in the number of oxytocin neurons in the PVN (Li et al., 1999; Muscatelli et al., 2000). Peg3 mutants in particular exhibit impair suckling behavior and often die in the first week of life (Curley et al., 2004). Deletion of another paternally expressed gene, Magel2, does not affect the number of OT neurons (Kozlov et al., 2007) but reduces concentration of mature OT peptides (Schaller et al., 2010). This is associated with poor suckling resulting in newborn death. Interestingly, injections of OT immediately after birth can rescue suckling and neonatal lethality (Schaller et al., 2010).

Other imprinted genes exhibiting preferential expression in these and other feeding-associated neurons are Dlk1, Nnat, Calcr, Gnas, $GnasX_L$ and Dio3. Enrichment of imprinted genes in the regulation of feeding and energy consumption is reminiscent of the role of imprinted

genes in the placenta and embryo. Thus, it appears that resource allocation is a general function of genomic imprinting in mammals.

Monoaminergic Regions

In the brain the monoamines serotonin (5HT), dopamine (DA) and norepinephrine (NE) function as neuromodulators of a wide range of neuronal processes controlling motivated behaviors, feeding and stress responses among others. Monoamine expression is restricted to specific brain nuclei of the midbrain and hindbrain including the dopaminergic ventral tegmental area and substantia nigra, the serotonergic dorsal raphe and the norepinergic locus coeruleus. Several imprinted genes are preferentially expressed in these brain regions. In general, expression of imprinted genes in these regions seems to regulate the concentrations of monoamines resulting in diverse behavioral outcomes.

Tyrosine hydroxylase (Th) is the rate-limiting enzyme in the synthesis of DA and NE, while Tryptophan hydroxylase is the rate-limiting enzyme of serotonin synthesis. The products of both enzymes are then processed by DOPA-decarboxylase (Ddc) before the eventual generation of all three monoamines. Thus, these enzymes control monoamine levels in the brain. Interestingly both Th and Ddc show preferential maternal expression in the medial preoptic area (Gregg et al., 2010) but it is currently unknown whether they are also imprinted in monoaminergic nuclei.

In the brain, the paternally expressed Sgce is preferentially found in midbrain dopaminergic neurons and the dorsal raphe (Chan et al., 2004). Deletion of Sgce in the brain results in myoclonus, motor deficits, hyperactivity and anxiety- and depression-like behaviors (Yokoi et al., 2006). In dopaminergic nucleus the levels of DA are substantially increase and

correlated with a decrease of Dopamine D2 autoreceptor (Yokoi et al., 2012). These results suggest that imprinting of Sgce regulates dopamine levels in the brain.

The expression of a set of genes essential for DA metabolism in midbrain neurons is controlled by the transcription factors Nurr1 and Pitx3. Nurr1 targets the promoter of Dlk1 and controls its expression in the early embryo. Analyses in the Dlk1 knockouts reveal this gene prevents early expression of the Dopamine transporter (DAT) in the early embryo and thus premature dopaminergic differentiation (Jacobs et al., 2009). However, according to the Allen brain atlas Dlk1 expression is maintained in adult dopaminergic neurons thus is possible that a regulatory role of DAT levels. Since DAT regulates the synaptic reuptake of dopamine, Dlk1 modulation of DAT might also regulate dopamine levels in the brain.

The Gnas locus of distal chromosome 2 shows a complex pattern of imprinting where both maternally expressed and paternally expressed isoforms are produced from the same transcriptional unit. Two protein coding isoforms generated from these loci are particularly interesting: the maternally expressed G-protein subunit alpha (Gnas), the paternally expressed large G-protein subunit alpha (GnasXL). The only difference between the protein products of Gnas and GnasXL is a large amino-terminal domain of GnasXL. Both of these proteins can stimulate adenylyl cyclase in response to receptor activation and in some circumstances they can replace each other in terms of receptor binding (Liu et al., 2011). Although at the cellular level these proteins appear to carry similar process, at physiological level they confer opposing phenotypes. While Gnas appears to be ubiquitously expressed in the brain, GnasXL is restricted to monoaminergic regions and specific hypothalamic nuclei. The absence of Gnas causes tremors, hyperactivity and a decrease metabolic rate leading obesity in adults. Strikingly, the absence of GnasXL results in the opposite phenotype, where animals are inactive,

hypermetabolic and lean as adults. Furthermore, GnasXL mutants reduce suckling behavior as pups (Plagge et al., 2004). Together these results highlight important interactions between the functions of imprinted genes and suggest these effects mainly take place in hypothalamic and monoaminergic regions.

Imprinted Genes in Memory and Cognition

In addition to roles in instinctive behaviors and drives performed by imprinted genes, as described above, a different subset of imprinted genes appears to regulate higher brain functions. In humans multiple neurodevelopmental disorders are associated with mutations in imprinted genes. Maternally inherited mutations of UBE3A cause Angelman syndrome (AS), which is characterized by mental retardation, communications deficits, ataxia and inappropriate laughter. Maternal deletions of Ube3a in mice recapitulate some of the features observed in AS including cognitive deficits and impair motor coordination. For instance, $Ube3a^{m-/p+}$ show substantially slower learning in a contextual fear conditioning paradigm (Jiang et al., 1998). Additionally, $Ube3a^{m-/p+}$ exhibit diminish response to monocular depravation during the visual cortex critical period indicating defects in neuronal plasticity (Sato and Stryker, 2010). In humans, mutations in the maternally expressed TRAPPC9 and KCNK9 also caused mental retardation (Barel et al., 2008; Mochida et al., 2009).

Multiple imprinted genes have also been implicated in the mechanisms of memory formation. As mentioned above deletion of Rasgrf1 impairs memory consolidation in the amygdala and hippocampus (Miller et al., 2013). Recently, it was shown that upregulation of Igf2 in the hippocampus enhances memory consolidation and prevents forgetting. Injections of

recombinant Igf2 after training or memory retrieval recapitulate this effect. Igf2-mediated cognitive enhancement requires protein synthesis, regulator of cytoskeleton associated proteins and glycogen-synthase kinase 3 which regulates receptor trafficking (Chen et al., 2011). Interestingly, Igf2 switches from paternally expressed in the embryo to maternally expressed in the adult brain (Gregg et al., 2010). Determining the significance of this allelic switch in memory consolidation will be particularly important.

CONCLUSION

Are the biological functions of imprinted genes interconnected? This question can be addressed at multiple biological levels. The simplest possible relationship between imprinted genes is that the molecules they produce share biochemical properties indicative of similar cellular functions. However, a close examination shows that there is no apparent relationship between the various molecules produced or cellular function performed by imprinted genes. A second potential characteristic that could link imprinted genes with each other is the existence of correlated expression patterns. Most imprinted genes do not share common expression patterns across the development of body tissues. However, the expression of a substantial subset of imprinted genes is particularly enriched in tow tissues: the placenta and the brain. Alternatively, patterns of differential expression between the two alleles can be assessed for imprinted genes across tissues. Although this only has been analyzed for a selected number of genes in a handful of tissues, a similar enrichment of differential expression between parental alleles can be observed in the brain and the placenta. Importantly, imprinted genes exhibiting spatial regulation of overall expression are by and large different than those in which differential expression between the two alleles is spatially regulated. Finally, associations between imprinted genes can be made at the phenotypic level, if their functions regulate common physiological or behavioral processes. This indeed appears to be the case for multiple genes. In the placenta, imprinted genes regulate levels of energetic supplies stored by this tissue. In the brain, hypothalamic expression of imprinted genes regulates feeding behavior and levels of physical activity by regulation of multiple neuronal types, again highlighting a role of imprinting in resource allocation in different tissues and developmental stages. Furthermore regulation of dopamine levels as well as other monoamines can control multiple motivated behaviors and imprinted genes may influence higher

brain functions by their regulation of neuronal plasticity and memory. Together, these observations strongly indicate biological interactions and common roles of imprinted genes at level of tissue physiology of the brain and placenta.

THESIS AIMS

Many gaps still exist in the understanding of the biology of genomic imprinting. First, the total number of imprinted genes remains unknown. Indeed, genome-wide profiles of the factors controlling genomic imprinting highlight multiple genomic regions not currently associated with preferential parent-of-origin expression (discussed in chapter 2). Moreover, dynamic regulation of known imprinted genes in the brain suggests that different regions may exhibit unique patterns of imprinting. The **first aim** of my thesis work is to establish a robust approach to accurately identify imprinted genes and explore the repertoire of imprinted genes in the cerebellum. This brain area provides unique experimental advantages for RNA-seq experiments and potential follow up functional studies on imprinted genes. The results of these experiments are described in chapter 3.

Imprinted genes have been shown to play important roles in the maturation of the brain. Another experimental advantage of the cerebellum is its unique accessibility during periods of neuronal development. Thus, accessing imprinting during key stages of cerebellar maturation could yield important insights in the role of imprinted genes in these processes. The **second aim** of my thesis work investigates the regulation of imprinted genes during an important milestone of cerebellar development. The results of these experiments are described in chapter 4.

As discussed above the differential expression between the two alleles, which I refer to as parental biases, can be regulated according to tissue types and even across brain regions. However, this aspect of imprinted genes remains largely unexplored and to date only our group has conducted systematic profiling of imprinting across multiple brain regions (Gregg et al., 2010). In my **third aim** I compared the parental biases of a selected number of imprinted genes

in the brain and other body tissues, and also within different brain regions. The results of these experiments are described in chapter 4.

REFERENCES

Albrecht, U., Sutcliffe, J.S., Cattanach, B.M., Beechey, C. V, Armstrong, D., Eichele, G., and Beaudet, A.L. (1997). Imprinted expression of the murine Angelman syndrome gene, Ube3a, in hippocampal and Purkinje neurons. Nat. Genet. *17*, 75–78.

Babak, T., Deveale, B., Armour, C., Raymond, C., Cleary, M.A., van der Kooy, D., Johnson, J.M., and Lim, L.P. (2008). Global survey of genomic imprinting by transcriptome sequencing. Curr. Biol. *18*, 1735–1741.

Bando, Y., Hirano, T., and Tagawa, Y. (2014). Dysfunction of KCNK Potassium Channels Impairs Neuronal Migration in the Developing Mouse Cerebral Cortex. Cereb. Cortex *24*, 1017–1029.

Barel, O., Shalev, S.A., Ofir, R., Cohen, A., Zlotogora, J., Shorer, Z., Mazor, G., Finer, G., Khateeb, S., Zilberberg, N., et al. (2008). Maternally Inherited Birk Barel Mental Retardation Dysmorphism Syndrome Caused by a Mutation in the Genomically Imprinted Potassium Channel KCNK9. Am. J. Hum. Genet. *83*, 193–199.

Barlow, D.P., Stöger, R., Herrmann, B.G., Saito, K., and Schweifer, N. (1991). The mouse insulin-like growth factor type-2 receptor is imprinted and closely linked to the Tme locus. Nature *349*, 84–87.

Bartolomei, M.S., and Ferguson-Smith, A.C. (2011). Mammalian genomic imprinting. Cold Spring Harb. Perspect. Biol. *3*, 1–17.

Bartolomei, M.S., Zemel, S., and Tilghman, S.M. (1991). Parental imprinting of the mouse H19 gene. Nature *351*, 153–155.

Barton, S.C., Surani, M.A., and Norris, M.L. (1984). Role of paternal and maternal genomes in mouse development. Nature *311*, 374–376.

Brambilla, R., Gnesutta, N., Minichiello, L., White, G., Roylance, A.J., Herron, C.E., Ramsey, M., Wolfer, D.P., Cestari, V., Rossi-Arnaud, C., et al. (1997). A role for the Ras signalling pathway in synaptic transmission and long-term memory. Nature *390*, 281–286.

Broad, K.D., Curley, J.P., and Keverne, E.B. (2009). Increased apoptosis during neonatal brain development underlies the adult behavioral deficits seen in mice lacking a functional paternally expressed gene 3(Peg3). Dev. Neurobiol. *69*, 314–325.

Burdakov, D., Jensen, L.T., Alexopoulos, H., Williams, R.H., Fearon, I.M., O'Kelly, I., Gerasimenko, O., Fugger, L., and Verkhratsky, A. (2006). Tandem-pore K+ channels mediate inhibition of orexin neurons by glucose. Neuron *50*, 711–722.

Burt, J., Alberto, C.O., Parsons, M.P., and Hirasawa, M. (2011). Local network regulation of orexin neurons in the lateral hypothalamus. AJP Regul. Integr. Comp. Physiol. *301*, R572–R580.

Buss, R.R., Sun, W., and Oppenheim, R.W. (2006). Adaptive roles of programmed cell death during nervous system development. Annu. Rev. Neurosci. *29*, 1–35.

Cattanach, B.M., and Jones, J. (1994). Genetic imprinting in the mouse: implications for gene regulation. J. Inherit. Metab. Dis. *17*, 403–420.

Cattanach, B.M., and Kirk, M. (1985). Differential activity of maternally and paternally derived chromosome regions in mice. Nature *315*, 496–498.

Chan, P., González-Maeso, J., Ruf, F. d rique, Bishop, D.F., Hof, P.R., and Sealfon, S.C. (2004). ?-Sarcoglycan immunoreactivity and mRNA expression in mouse brain. J. Comp. Neurol. *482*, 50–73.

Charalambous, M., and Hernandez, A. (2013). Genomic imprinting of the type 3 thyroid hormone deiodinase gene: regulation and developmental implications. Biochim. Biophys. Acta (BBA)-General Subj. *1830*, 3946–3955.

Charalambous, M., Smith, F.M., Bennett, W.R., Crew, T.E., Mackenzie, F., and Ward, A. (2003). Disruption of the imprinted Grb10 gene leads to disproportionate overgrowth by an Igf2-independent mechanism. Proc. Natl. Acad. Sci. U. S. A. *100*, 8292–8297.

Charalambous, M., Cowley, M., Geoghegan, F., Smith, F.M., Radford, E.J., Marlow, B.P., Graham, C.F., Hurst, L.D., and Ward, A. (2010). Maternally-inherited Grb10 reduces placental size and efficiency. Dev. Biol. *337*, 1–8.

Chen, D.Y., Stern, S.A., Garcia-Osta, A., Saunier-Rebori, B., Pollonini, G., Bambah-Mukku, D., Blitzer, R.D., and Alberini, C.M. (2011). A critical role for IGF-II in memory consolidation and enhancement. Nature *469*, 491–497.

Chung, S.-H., Marzban, H., Aldinger, K., Dixit, R., Millen, K., Schuurmans, C., and Hawkes, R. (2011). Zac1 plays a key role in the development of specific neuronal subsets in the mouse cerebellum. Neural Dev. *6*, 25.

Cubelos, B., Sebastián-Serrano, A., Beccari, L., Calcagnotto, M.E., Cisneros, E., Kim, S., Dopazo, A., Alvarez-Dolado, M., Redondo, J.M., and Bovolenta, P. (2010). Cux1 and Cux2 Regulate Dendritic Branching, Spine Morphology, and Synapses of the Upper Layer Neurons of the Cortex. Neuron *66*, 523–535.

Curley, J.P., Barton, S., Surani, A., and Keverne, E.B. (2004). Coadaptation in mother and infant regulated by a paternally expressed imprinted gene. Proc. Biol. Sci. *271*, 1303–1309.

d'Isa, R., Clapcote, S.J., Voikar, V., Wolfer, D.P., Giese, P., Brambilla, R., and Fasano, S. (2011). Mice Lacking Ras-GRF1 Show Contextual Fear Conditioning but not Spatial Memory Impairments: Convergent Evidence from Two Independently Generated Mouse Mutant Lines. Front. Behav. Neurosci. *5*.

Darcy, M.J., Trouche, S., Jin, S.-X., and Feig, L.A. (2013). Age-dependent role for Ras-GRF1 in the late stages of adult neurogenesis in the dentate gyrus. Hippocampus *24*, 315–325.

Davies, J.R., Dent, C.L., McNamara, G.I., and Isles, A.R. (2015). Behavioural effects of imprinted genes. Curr. Opin. Behav. Sci. *2*, 28–33.

DeChiara, T.M., Robertson, E.J., and Efstratiadis, A. (1991). Parental imprinting of the mouse insulin-like growth factor II gene. Cell *64*, 849–859.

Deng, Y., and Wu, X. (2000). Peg3/Pw1 promotes p53-mediated apoptosis by inducing Bax translocation from cytosol to mitochondria. Proc. Natl. Acad. Sci. U. S. A. *97*, 12050–12055.

DeVeale, B., van der Kooy, D., and Babak, T. (2012). Critical evaluation of imprinted gene expression by RNA-seq: A new perspective. PLoS Genet. *8*.

Felsenfeld, G., and Bell, A.C. (2000). Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. Nature *405*, 482–485.

Ferrón, S.R., Charalambous, M., Radford, E., Mcewen, K., Wildner, H., Hind, E., Morante-Redolat, J.M., Laborda, J., Guillemot, F., Bauer, S.R., et al. (2012). Postnatal loss of Dlk1 imprinting in stem cells and niche astrocytes regulates neurogenesis. Nature *475*, 381–385.

Franco, M.M., Prickett, A.R., and Oakey, R.J. (2014). The Role of CCCTC-Binding Factor (CTCF) in Genomic Imprinting, Development, and Reproduction. Biol. Reprod.

Fundele, R.H., and Surani, M.A. (1994). Experimental embryological analysis of genetic imprinting in mouse development. Dev. Genet. *15*, 515–522.

Garfield, A.S., Cowley, M., Smith, F.M., Moorwood, K., Stewart-Cox, J.E., Gilroy, K., Baker, S., Xia, J., Dalley, J.W., Hurst, L.D., et al. (2011). Distinct physiological and behavioural functions for parental alleles of imprinted Grb10. Nature *469*, 534–538.

Gehring, M. (2013). Genomic imprinting: insights from plants. Genetics *47*, 187.

Giese, K.P., Friedman, E., Telliez, J.B., Fedorov, N.B., Wines, M., Feig, L.A., and Silva, A.J. (2001). Hippocampus-dependent learning and memory is impaired in mice lacking the Ras-guanine-nucleotide releasing factor 1 (Ras-GRF1). Neuropharmacology *41*, 791–800.

Gregg, C., Zhang, J., Weissbourd, B., Luo, S., Schroth, G.P., Haig, D., and Dulac, C. (2010). High-resolution analysis of parent-of-origin allelic expression in the mouse brain. Sci. (New York, NY) *329*, 643–648.

Haig, D., and Graham, C. (1991). Genomic imprinting and the strange case of the insulin-like growth factor II receptor. Cell *64*, 1045–1046.

Haig, D., and Westoby, M. (1989). Parent-specific gene expression and the triploid endosperm. Am. Nat. 147–155.

Hamed, M., Ismael, S., Paulsen, M., and Helms, V. (2012). Cellular Functions of Genetically Imprinted Genes in Human and Mouse as Annotated in the Gene Ontology. PLoS One *7*, e50285.

Hoffmann, A., and Spengler, D. (2012). Transient neonatal diabetes mellitus gene Zac1 impairs insulin secretion in mice through Rasgrf1. Mol. Cell. Biol. *32*, 2549–2560.

Hoffmann, A., Ciani, E., Boeckardt, J., Holsboer, F., Journot, L., and Spengler, D. (2003). Transcriptional activities of the zinc finger protein Zac are differentially controlled by DNA binding. Mol. Cell. Biol. *23*, 988–1003.

Horn, S., and Heuer, H. (2010). Thyroid hormone action during brain development: more questions than answers. Mol. Cell. Endocrinol. *315*, 19–26.

Jacobs, F.M.J., van der Linden, A.J.A., Wang, Y., von Oerthel, L., Sul, H.S., Burbach, J.P.H., and Smidt, M.P. (2009). Identification of Dlk1, Ptpru and Klhl1 as novel Nurr1 target genes in meso-diencephalic dopamine neurons. Development *136*, 2363–2373.

Jiang, Y.H., Armstrong, D., Albrecht, U., Atkins, C.M., Noebels, J.L., Eichele, G., Sweatt, J.D., and Beaudet, A.L. (1998). Mutation of the Angelman ubiquitin ligase in mice causes increased cytoplasmic p53 and deficits of contextual learning and long-term potentiation. Neuron *21*, 799–811.

Johnson, D.R. (1974). Further observations on the hairpin-tail (Thp) mutation in the mouse. Genet. Res. *24*, 207.

Jonas, E.A., Porter, G.A., and Alavian, K.N. (2014). Bcl-xL in neuroprotection and plasticity. Front. Physiol. *5*, 355.

Kawahara, M., Wu, Q., Takahashi, N., Morita, S., Yamada, K., Ito, M., Ferguson-Smith, A.C., and Kono, T. (2007). High-frequency generation of viable mice from engineered bi-maternal embryos. Nat. Biotechnol. *25*, 1045–1050.

Kelsey, G., and Bartolomei, M.S. (2012). Imprinted Genes … and the Number Is? PLoS Genet. *8*, e1002601.

Kermicle, J.L. (1970). Dependence of the R-mottled aleurone phenotype in maize on mode of sexual transmission. Genetics *66*, 69–85.

Keverne, E.B. (2012). Importance of the matriline for genomic imprinting, brain development and behaviour. Philos. Trans. R. Soc. London Ser. B, Biol. Sci. *368*, 20110327.

Keverne, E.B., Fundele, R., Narasimha, M., Barton, S.C., and Surani, M.A. (1996). Genomic imprinting and the differential roles of parental genomes in brain development. Dev. Brain Res. *92*, 91–100.

Khatib, H. (2007). Is it genomic imprinting or preferential expression? BioEssays *29*, 1022–1028.

Kono, T., Obata, Y., Wu, Q., Niwa, K., Ono, Y., Yamamoto, Y., Park, E.S., Seo, J.-S., and Ogawa, H. (2004). Birth of parthenogenetic mice that can develop to adulthood. Nature *428*, 860–864.

Kozlov, S. V, Bogenpohl, J.W., Howell, M.P., Wevrick, R., Panda, S., Hogenesch, J.B., Muglia, L.J., Van Gelder, R.N., Herzog, E.D., and Stewart, C.L. (2007). The imprinted gene Magel2 regulates normal circadian output. Nat. Genet. *39*, 1266–1272.

Krapivinsky, G., Krapivinsky, L., Manasian, Y., Ivanov, A., Tyzio, R., Pellegrino, C., Ben-Ari, Y., Clapham, D.E., and Medina, I. (2003). The NMDA receptor is coupled to the ERK pathway by a direct interaction between NR2B and RasGRF1. Neuron *40*, 775–784.

Kuan, C.Y., Roth, K.A., Flavell, R.A., and Rakic, P. (2000). Mechanisms of programmed cell death in the developing brain. Trends Neurosci. *23*, 291–297.

Lau, M.M.H., Stewart, C.E.H., Liu, Z., Bhatt, H., Rotwein, P., and Stewart, C.L. (1994). Loss of the imprinted IGF2/cation-independent mannose 6-phosphate receptor results in fetal overgrowth and perinatal lethality. Genes Dev. *8*, 2953–2963.

Lefebvre, L., Viville, S., Barton, S., and Ishino, F. (1998). Abnormal maternal behaviour and growth retardation associated with loss of the imprinted gene Mest. Nat. ….

Li, E., Beard, C., and Jaenisch, R. (1993). Role for DNA methylation in genomic imprinting. Nature *366*, 362–365.

Li, J.-Y., Chai, B.-X., Zhang, W., Wang, H., and Mulholland, M.W. (2010a). Expression of Ankyrin Repeat and Suppressor of Cytokine Signaling Box Protein 4 (Asb-4) in Proopiomelanocortin Neurons of the Arcuate Nucleus of Mice Produces a Hyperphagic, Lean Phenotype. Endocrinology *151*, 134–142.

Li, L., Keverne, E.B., Aparicio, S.A., Ishino, F., Barton, S.C., and Surani, M.A. (1999). Regulation of maternal behavior and offspring growth by paternally expressed Peg3. Science *284*, 330–333.

Li, S., Tian, X., Hartley, D.M., and Feig, L.A. (2006). Distinct roles for Ras-guanine nucleotide-releasing factor 1 (Ras-GRF1) and Ras-GRF2 in the induction of long-term potentiation and long-term depression. J. Neurosci. *26*, 1721–1729.

Li, X., Ito, M., Zhou, F., Youngson, N., Zuo, X., Leder, P., and Ferguson-Smith, A.C. (2008). A Maternal-Zygotic Effect Gene, Zfp57, Maintains Both Maternal and Paternal Imprints. Dev. Cell *15*, 547–557.

Li, Z., Jo, J., Jia, J.M., Lo, S.C., Whitcomb, D.J., Jiao, S., Cho, K., and Sheng, M. (2010b). Caspase-3 activation via mitochondria is required for long-term depression and AMPA receptor internalization. Cell *141*, 859–871.

Lin, B.Y. (1982). Association of endosperm reduction with parental imprinting in maize. Genetics *100*, 475–486.

Linden, A.M., Sandu, C., Aller, M.I., Vekovischeva, O.Y., Rosenberg, P.H., Wisden, W., and Korpi, E.R. (2007). TASK-3 Knockout Mice Exhibit Exaggerated Nocturnal Activity, Impairments in Cognitive Functions, and Reduced Sensitivity to Inhalation Anesthetics. J. Pharmacol. Exp. Ther. *323*, 924–934.

Liu, X.-Y., Liu, Z.-C., Sun, Y.-G., Ross, M., Kim, S., Tsai, F.-F., Li, Q.-F., Jeffry, J., Kim, J.-Y., Loh, H.H., et al. (2011). Unidirectional Cross-Activation of GRPR by MOR1D Uncouples Itch and Analgesia Induced by Opioids. Cell *147*, 447–458.

Mabb, A.M., Judson, M.C., Zylka, M.J., and Philpot, B.D. (2011). Angelman syndrome: insights into genomic imprinting and neurodevelopmental phenotypes. Trends Neurosci. *34*, 293–303.

Mackay, D.J.G., Callaway, J.L.A., Marks, S.M., White, H.E., Acerini, C.L., Boonen, S.E., Dayanikli, P., Firth, H. V, Goodship, J.A., Haemers, A.P., et al. (2008). Hypomethylation of multiple imprinted loci in individuals with transient neonatal diabetes is associated with mutations in ZFP57. Nat. Genet. *40*, 949–951.

Margolis, S.S., Salogiannis, J., Lipton, D.M., Mandel-Brehm, C., Wills, Z.P., Mardinly, A.R., Hu, L., Greer, P.L., Bikoff, J.B., Ho, H.-Y.H., et al. (2010). EphB-mediated degradation of the RhoA GEF Ephexin5 relieves a developmental brake on excitatory synapse formation. Cell *143*, 442–455.

Martinez, M.E., Charalambous, M., Saferali, A., Fiering, S., Naumova, A.K., Germain, D.S., Ferguson-Smith, A.C., and Hernandez, A. (2014). Genomic Imprinting Variations in the Mouse Type 3 Deiodinase Gene Between Tissues and Brain Regions. Mol. Endocrinol.

McGrath, J., and Solter, D. (1983). Nuclear transplantation in mouse embryos. J. Exp. Zool. *228*, 355–362.

McGrath, J., and Solter, D. (1984a). Completion of mouse embryogenesis requires both the maternal and paternal genomes. Cell *37*, 179–183.

McGrath, J., and Solter, D. (1984b). Completion of mouse embryogenesis requires both the maternal and paternal genomes. Cell *37*, 179–183.

Miller, M.B., Yan, Y., Eipper, B.A., and Mains, R.E. (2013). Neuronal Rho GEFs in Synaptic Physiology and Behavior. Neurosci. *19*, 255–273.

Mizuno, Y., Sotomaru, Y., Katsuzawa, Y., Kono, T., Meguro, M., Oshimura, M., Kawai, J., Tomaru, Y., Kiyosawa, H., Nikaido, I., et al. (2002). Asb4, Ata3, and Dcn Are Novel Imprinted Genes Identified by High-Throughput Screening Using RIKEN cDNA Microarray. Biochem. Biophys. Res. Commun. *290*, 1499–1505.

Mochida, G.H., Mahajnah, M., Hill, A.D., Basel-Vanagaite, L., Gleason, D., Hill, R.S., Bodell, A., Crosier, M., Straussberg, R., and Walsh, C.A. (2009). A truncating mutation of TRAPPC9 is associated with autosomal-recessive intellectual disability and postnatal microcephaly. Am. J. Hum. Genet. *85*, 897–902.

Muscatelli, F., Abrous, D.N., Massacrier, A., Boccaccio, I., Le Moal, M., Cau, P., and Cremer, H. (2000). Disruption of the mouse Necdin gene results in hypothalamic and behavioral alterations reminiscent of the human Prader-Willi syndrome. Hum. Mol. Genet. *9*, 3101–3110.

Nantel, A., Huber, M., and Thomas, D.Y. (1999). Localization of Endogenous Grb10 to the Mitochondria and Its Interaction with the Mitochondrial-associated Raf-1 Pool. J. Biol. Chem. *274*, 35719–35724.

Nicholls, R.D., Knoll, J.H.M., Butler, M.G., Karam, S., and Lalande, M. (1989). Genetic imprinting suggested by maternal heterodisomy in non-deletion Prader-Willi syndrome. Nature *342*, 281–285.

Nur, U. (1990). Heterochromatization and euchromatization of whole genomes in scale insects (Coccoidea: Homoptera). Dev. Suppl. 29–34.

Okae, H., Hiura, H., Nishida, Y., Funayama, R., Tanaka, S., Chiba, H., Yaegashi, N., Nakayama, K., Sasaki, H., and Arima, T. (2012). Re-investigation and RNA sequencing-based identification of genes with placenta-specific imprinted expression. Hum. Mol. Genet. *21*, 548–558.

Otto, S.P., and Gerstein, A.C. (2008). The evolution of haploidy and diploidy. Curr. Biol. *18*.

Patel, A.J., and Lazdunski, M. (2004). The 2P-domain K + channels: role in apoptosis and tumorigenesis. Pflügers Arch. - Eur. J. Physiol. *448*, 261–273.

Peeters, R.P., Hernandez, A., Ng, L., Ma, M., Sharlin, D.S., Pandey, M., Simonds, W.F., Germain, D.L.S., and Forrest, D. (2013). Cerebellar Abnormalities in Mice Lacking Type 3 Deiodinase and Partial Reversal of Phenotype by Deletion of Thyroid Hormone Receptor α1. Dx.doi.org.

Peters, J. (2014). The role of genomic imprinting in biology and disease: an expanding view. Nat. Rev. Genet.

Plagge, A., Gordon, E., Dean, W., Boiani, R., Cinti, S., Peters, J., and Kelsey, G. (2004). The imprinted signaling protein XLαs is required for postnatal adaptation to feeding. Nat. Genet. *36*, 818–826.

Prickett, A.R., and Oakey, R.J. (2012). A survey of tissue-specific genomic imprinting in mammals. Mol. Genet. Genomics *287*, 621–630.

Quenneville, S., Verde, G., Corsinotti, A., Kapopoulou, A., Jakobsson, J., Offner, S., Baglivo, I., Pedone, P. V, Grimaldi, G., Riccio, A., et al. (2011). In Embryonic Stem Cells, ZFP57/KAP1 Recognize a Methylated Hexanucleotide to Affect Chromatin and DNA Methylation of Imprinting Control Regions. Mol. Cell *44*, 361–372.

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

Relaix, F., Wei, X. j, Li, W., Pan, J., Lin, Y., Bowtell, D.D., Sassoon, D.A., and Wu, X. (2000). Pw1/Peg3 is a potential cell death mediator and cooperates with Siah1a in p53-mediated apoptosis. Proc. Natl. Acad. Sci. U. S. A. *97*, 2105–2110.

Roffe, M., Hajj, G.N.M., Azevedo, H.F., Alves, V.S., and Castilho, B.A. (2013). IMPACT Is a Developmentally Regulated Protein in Neurons That Opposes the Eukaryotic Initiation Factor 2 Kinase GCN2 in the modulation of Neurite Outgrowth. J. Biol. Chem. *288*, 10860–10869.

Santoro, F., and Barlow, D.P. (2011). Developmental control of imprinted expression by macro non-coding RNAs. Semin. Cell Dev. Biol. *22*, 328–335.

Sato, M., and Stryker, M.P. (2010). Genomic imprinting of experience-dependent cortical plasticity by the ubiquitin ligase gene Ube3a. Proc. Natl. Acad. Sci. U. S. A. *107*, 5611–5616.

Schaller, F., Watrin, F., Sturny, R., Massacrier, A., Szepetowski, P., and Muscatelli, F. (2010). A single postnatal injection of oxytocin rescues the lethal feeding behaviour in mouse newborns deficient for the imprinted Magel2 gene. Hum. Mol. Genet. *19*, 4895–4905.

Schmidt-Edelkraut, U., Hoffmann, A., Daniel, G., and Spengler, D. (2013). Zac1 Regulates Astroglial Differentiation of Neural Stem Cells Through Socs3. Stem Cells *31*, 1621–1632.

Schmidt-Edelkraut, U., Daniel, G., Hoffmann, A., and Spengler, D. (2014). Zac1 Regulates Cell Cycle Arrest in Neuronal Progenitors via Tcf4. Mcb.asm.org.ezp-prod1.hul.harvard.edu.

Schrader, F. (1921). THE CHROMOSOMES OF PSEUDOCOCCUS NIPAe. Biol. Bull. *40*, 259–269.

Steinhoff, C., Paulsen, M., Kielbasa, S., Walter, J., and Vingron, M. (2009a). Expression profile and transcription factor binding site exploration of imprinted genes in human and mouse. BMC Genomics *10*, 144.

Steinhoff, C., Paulsen, M., Kielbasa, S., Walter, J., and Vingron, M. (2009b). Expression profile and transcription factor binding site exploration of imprinted genes in human and mouse. BMC Genomics *10*, 144.

Sternson, S.M. (2013). Hypothalamic Survival Circuits: Blueprints for Purposive Behaviors. Neuron *77*, 810–824.

Surani, M.A.H. (1984). What the papers say: Differential roles of paternal and maternal genomes during embryogenesis in the mouse. BioEssays *1*, 224–227.

Surani, M.A.H., Barton, S.C., and Norris, M.L. (1984a). Development of reconstituted mouse eggs suggests imprinting of the genome during gametogenesis. Nature *308*, 548–550.

Surani, M.A.H., Barton, S.C., and Norris, M.L. (1984b). Development of reconstituted mouse eggs suggests imprinting of the genome during gametogenesis. Nature *308*, 548–550.

Takazaki, R., Nishimura, I., and Yoshikawa, K. (2002). Necdin is required for terminal differentiation and survival of primary dorsal root ganglion neurons. Exp. Cell Res. *277*, 220–232.

Tunster, S.J., Jensen, A.B., and John, R.M. (2013). Imprinted genes in mouse placental development and the regulation of fetal energy stores. Reproduction *145*, R117–R137.

Tury, A., Mairet-Coello, G., and DiCicco-Bloom, E. (2011). The Cyclin-Dependent Kinase Inhibitor p57Kip2 Regulates Cell Cycle Exit, Differentiation, and Migration of Embryonic Cerebral Cortical Precursors. Cereb. Cortex *21*, 1840–1856.

Tury, A., Mairet-Coello, G., and DiCicco-Bloom, E. (2012). The multiple roles of the cyclin-dependent kinase inhibitory protein p57KIP2 in cerebral cortical neurogenesis. Dev. Neurobiol. *72*, 821–842.

Valente, T., and Auladell, C. (2001). Expression pattern of Zac1 mouse gene, a new zinc-finger protein that regulates apoptosis and cellular cycle arrest, in both adult brain and along development. Mech. Dev. *108*, 207–211.

Varmuza, S., and Mann, M. (1994). Genomic imprinting—defusing the ovarian time bomb. Trends Genet. *10*, 118–123.

Varrault, A., Gueydan, C., Delalbre, A., Bellmann, A., Houssami, S., Aknin, C., Severac, D., Chotard, L., Kahli, M., Le Digarcher, A., et al. (2006). Zac1 Regulates an Imprinted Gene Network Critically Involved in the Control of Embryonic Growth. Dev. Cell *11*, 711–722.

Wang, L., Zhang, J., Duan, J., Gao, X., Zhu, W., Lu, X., Yang, L., Zhang, J., Li, G., Ci, W., et al. (2014). Programming and Inheritance of Parental DNA Methylomes in Mammals. Cell *157*, 979–991.

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

Wilkins, J.F., and Haig, D. (2003). What good is genomic imprinting: the function of parent-specific gene expression. Nat. Rev. Genet. *4*, 359–368.

Wolf, J.B., and Hager, R. (2006). A maternal-offspring coadaptation theory for the evolution of genomic imprinting. PLoS Biol. *4*, 2238–2243.

Xie, W., Barr, C., Kim, A., Yue, F., and Lee, A. (2012). Base-resolution analyses of sequence and parent-of-origin dependent DNA methylation in the mouse genome. Cell.

Yamaguchi, S., Shen, L., Liu, Y., Sendler, D., and Zhang, Y. (2013). Role of Tet1 in erasure of genomic imprinting. Nature *504*, 460–464.

Yokoi, F., Dang, M.T., Li, J., and Li, Y. (2006). Myoclonus, motor deficits, alterations in emotional responses and monoamine metabolism in epsilon-sarcoglycan deficient mice. J. Biochem. *140*, 141–146.

Yokoi, F., Dang, M.T., Zhou, T., and Li, Y. (2012). Abnormal nuclear envelopes in the striatum and motor deficits in DYT11 myoclonus-dystonia mouse models. Hum. Mol. Genet. *21*, 916–925.

Zhang, G., Bai, H., Zhang, H., Dean, C., Wu, Q., Li, J., Guariglia, S., Meng, Q., and Cai, D. (2011). Neuropeptide exocytosis involving synaptotagmin-4 and oxytocin in hypothalamic programming of body weight and energy balance. Neuron *69*, 523–535.

Chapter Two

Epigenetic Mechanisms Regulating Imprinted Genes

Part of this chapter will be incorporated in a review in preparation Dulac, C., Rubinstein, N.D., Perez, J.D, Gregg C *Annu Rev Neurosci*, 2016.

INTRODUCTION

Ever since it was discovered, the phenomenon of genomic imprinting has provided a fascinating biological conundrum, raising still largely unsolved questions about its functional significance, and specific molecular mechanisms underlying this unique form of gene regulation. Based on the unique biological features of this unusual form of gene regulation, and before the first imprinted genes were discovered and specific epigenetic mechanisms were associated with the establishment of genomic imprinting, a first hypothetical model of the life cycle of imprinted genes was postulated (Figure 1). First, one should expect imprints to be established at a developmental time point between the birth and the maturation of the parental gametes, as germ cells transmit the genetic material to the offspring and therefore define the identity of its constituent maternal and paternal genomes. Thus, one can postulate that female and male germ lines establish their own unique imprints. Second, the unique markings of each genome associated with imprints must be maintained throughout cell divisions and survive reprograming events occurring in the early embryo, such that they can persist in somatic cell lineages of the developing and mature organism. Third, imprints must be capable of regulating gene expression either directly or indirectly within somatic tissues, in order for them to have a phenotypic effect. Finally, imprints must be erased in the new germ line, otherwise a female individual will pass paternal-specific imprints to it offspring and similarly a male individual will pass maternal-specific imprints to its offspring. Implied within this cycle are unique and complex challenges for the epigenetic and transcriptional machinery underlying genomic imprinting.

In this chapter I will describe our current understanding of the molecular mechanisms that underlie the life cycle of genomic imprinting. Most studies cited will relate to genomic

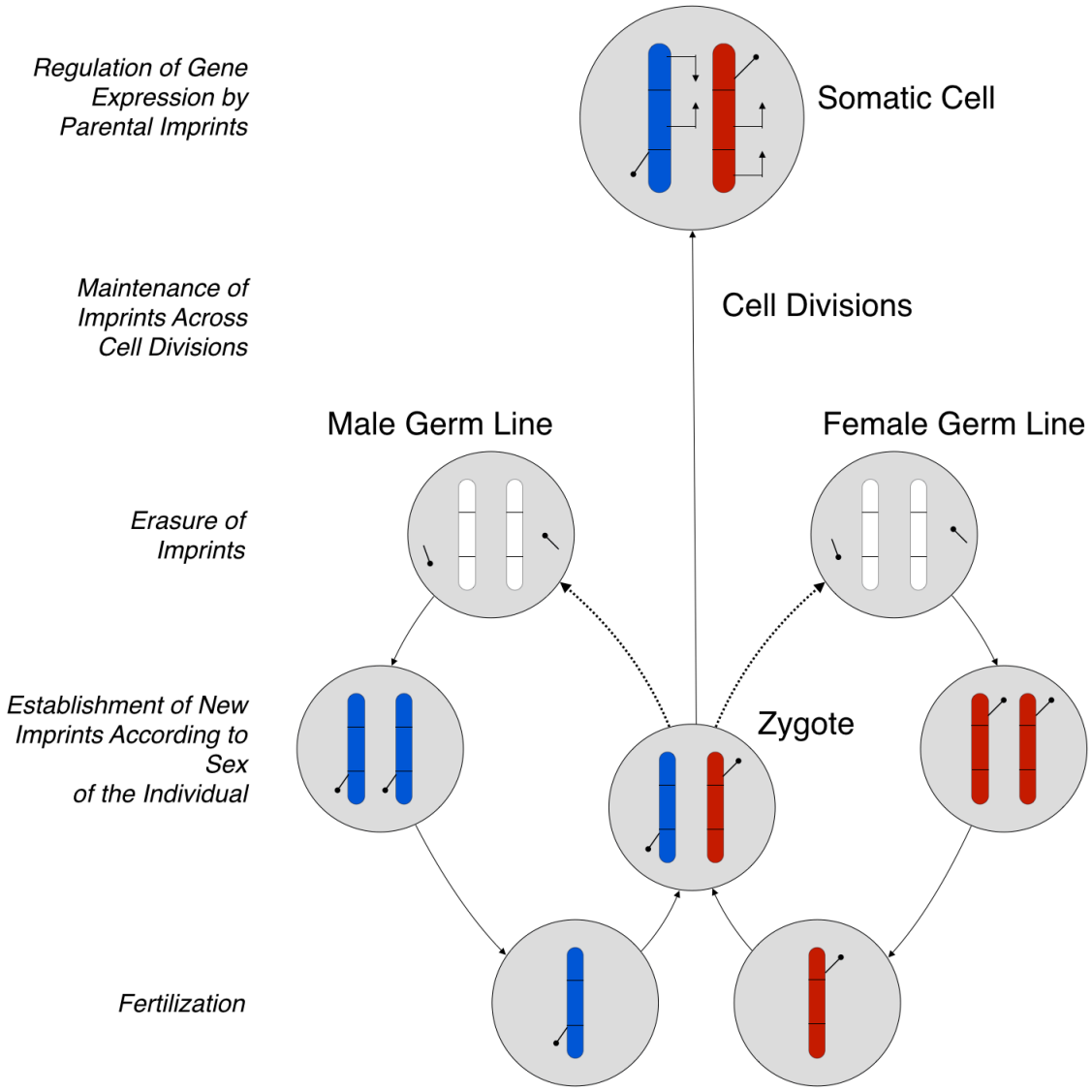

Figure 2.1: Life Cycle of Genomic Imprinting

imprinting in the mouse, unless otherwise specified. This information should aid in the interpretation of our results and in the formulation of future experiments inspired by our data.

GENOMIC ORGANIZATION OF IMPRINTED GENES

Clustered Organization

A striking majority of imprinted genes (90%) are not randomly located in the genome, but are rather colocalized into large gene clusters where they lie side by side. Up to date, 20 imprinted clusters have been identified in the mouse genome (Table 1). These clusters vary significantly from each other both in size, which ranges from 15Kb to 5Mb, and in the number of imprinted genes per cluster, which ranges from 2 to 29. Interestingly, most imprinted clusters contain both maternally and paternally expressed transcripts. Why are imprinted genes clustered is not entirely clear, although a largely accepted hypothesis is that this organization facilitates the regulation of imprinting of multiple genes by coordinated mechanisms (Reik and Walter, 2001).

At first glance, genes within an imprinted cluster do not seem to be functionally related, as is the case for other clustered genes such as Hox genes, olfactory receptors, and β-hemoglobins. This classification, however, depends on the definition of shared function. Indeed, if shared function is broadly defined, such as a common contribution to growth regulation or brain function (Chapter I), genes within imprinted clusters can often be seen as functionally related. For instance, the Prader-Willi cluster on chromosome 7 contains multiple protein coding and non-coding transcripts of diverse molecular and cellular functions. Still deletions of individual genes result in related phenotypes associated with brain function, eating behavior and hormonal regulation.

Cluster-ID	Coordinates	Size (MB)	Associated gDMR(s)	Location of gDMR(s)	# of Genes	Protein Coding Genes	Long non-Coding RNAs	Insulator Binding
chr1.Zdbf2	chr1:63182691-63596515	0.41	chr1:63255163-63258892, chr1:63262826-63263859, chr1:63268754-63271944	Intronic, Intronic, Intronic	3	Gpr1, Zdbf2, Adam23	-	-
chr2.Mcts2	chr2:152669461-152831728	0.16	chr2:152686274-152686927 chr2:152707764-152708164	Intronic H13 3UTR	3	Mcts2, H13, Id1, Cox4i2, Bcl2l1	-	CTCF/Cohesin
chr2.Nnat	chr2:157556362-157571274	0.01	chr2:157560051-157560581	Intronic	2	Nnat, Blcap	-	CTCF/Cohesin
chr2.Gnas	chr2:174281237-174346744	0.07	chr2:174292903-174299786 chr2:174327110-174328002	Intronic, Intronic	4	Nesp, Gnas, GnasXL	Nespas	CTCF/Cohesin Cohesin
chr6.Calcr	chr6:3685677-3968375	0.28	-	-	2	Calcr, Tfpi2	-	-
chr6.Sgce	chr6:4600840-5433022	0.83	chr6:4746303-4749370	Intronic	7	Casd1, Sgce, Peg10, Ppp1r9a, Pon1, Pon2, Pon3, Asb4	-	CTCF
chr6.Mest	chr6:30401868-30958990	0.56	chr6:30735840-30739965	Intronic	6	Klhdc10, Tmem209, Ssmem1, Cpa2, Cpa4, Cpa5, Cpa1, Cep41, Mest, Copg2,Tsga13, Klf14	Mit/Lb9	CTCF/Cohesin
chr6.Nap1l5	chr6:58905234-58907126	0.0019	chr6:58907401-58907794	Intronic	2	Herc3, Nap1l5	-	No binding
chr7.Peg3	chr7:6651747-6991084	0.34	chr7:6727356-6733209	Intronic	7	Zim2, Zim1, Peg3, Usp29	Peg3as, Zim3, Zfp264	CTCF/Cohesin
chr7.Snrpn	chr7:57590518-62464510	4.87	chr7:60003140-60005283	Intronic	22	Gabrb3, Atp10a, Ube3a, Snrpn, Snurf, Ndn, Magel2, Mkrn3, Peg12	Ube3a-AS, Ipw, Pec2, Pec3	Cohesin
chr7.Igf2	chr7:142575529-142699510	0.12	chr7:142579886-142582026	Intergenic	4	Igf2, Ins2	H19, Igf2as	CTCF/Cohesin
chr7.Ascl2	chr7:142892752-143756985	0.86	chr7:143294831-143296101	Intronic	15	Th, Ascl2, Tspan32, Cd81, Tssc4, Trpm5, Kcnq1, Cdkn1c, Slc22a18, Phlda2, Nap1l4, Cars, Tnfrsf26, Tnfrsf22, Tnsf23, Osbpl5	Kcnq1ot1	Cohesin
chr9.Rasgrf1	chr9:89826518-90026976	0.2	chr9:89869688-89878981	Intergenic	2	Rasgrf1	A19	No binding
chr10.Plagl1	chr10:13090691-13474396	0.38	chr10:13090122-13092339	Intronic	2	Plagl1, Phactr2	HYMAI	CTCF/Cohesin
chr11.Grb10	chr11:11814101-12464960	0.65	chr11:12023322-12026797 chr11:12035415-12035541	Intronic, Intronic	3	Ddc, Grb10, Cobl	-	CTCF/Cohesin
chr11.Zrsr1	chr11:22834744-22982382	0.15	chr11:22971545-22974145	Intronic	2	Commd1, Zrsr1	-	CTCF/Cohesin
chr12.Meg3	chr12:108860030-110281097	1.42	chr12:109525755-109530399	Intergenic	29	Wars, Wdr25, Begain, Dlk1, Rtl1, Dio3	Meg3, Rian, Rtl1as, Mirg	CTCF
chr15.Peg13	chr15:72512119-73184840	0.67	chr15:72808788-72811583	Intronic	4	Kcnk9, Trappc9, Chrac1, Ago2	Peg13	CTCF/Cohesin
chr17.Igf2r	chr17:12419972-12860122	0.44	chr17:12741760-12742949	Intronic	4	Slc22a3, Slc22a2, Slc22a1, Igf2r, Mas1	Airn	Cohesin
chrX.Xlr3b	chrX:73192207-73243130	0.05	-	-	3	Xlr3b, Xlr4b, F8a, Xlr4c	-	-

Table 2.1: Imprinted Gene Clusters in the Mouse. In associated gDMR column, red denotes maternally imprinted and blue denotes paternally imprinted. In protein coding genes and long non-coding RNA columns, red denotes maternally expressed, blue denotes paternally expressed, magenta denotes genes containing both maternal and paternal isoforms, and black denotes paternally biased expression has not been observed for that gene. In insulator binding column red denotes insulator binding to the maternal allele, blue denotes insulator binding to the paternal allele, and black denotes that the allelic-specific insulator binding is currently unknown.

Singletons

A small number of isolated imprinted genes, or singletons, have also been observed (Table 2). It remains possible that genes flanking such singletons are imprinted as well, yet the particular time points or tissues and cell types in which their expression is parentally biased have not yet been discovered. The *Grb10* imprinted gene serves as an example for this argument, as it was initially believed to be isolated. However it was later shown that its two flanking genes, *Ddc* and *Cobl*, show mild parental biases in the hearth and yolk sac, respectively (Menheniott et al., 2008; Shiura et al., 2009). The molecular mechanisms regulating imprinted expression of these genes are similar to those controlling expression within imprinted clusters.

DNA Sequences Enriched within Imprinted Regions

Substantial efforts have been directed towards the identification of specific genomic sequences that may be predictive of the presence of imprinted genes (Luedi et al., 2005, 2007). Such predictive sequences would help accelerate the discovery of additional imprinted genes and elucidate mechanisms by which genomic imprinting is established in the germ line. Based on limited data, Neumann et al. (Neumann et al., 1995) hypothesized that imprinting arose by the appropriation of the genome defense mechanism for the regulation of expression of transposable elements. Such genomic elements are enriched in repetitive sequences, which are recognized by the host defense mechanism and thus undergo methylation, leading to their heterochromatization as part of a silencing process. Indeed, direct tandem repeats are present in the vicinity of many imprinted regions, and various repetitive elements appear enriched in differentially methylated

regions of imprinted genes (Dindot et al., 2009; Neumann et al., 1995). However, a regulatory role for these elements has been described for only one imprinted gene (Yoon et al., 2002) while for several others they appear to be dispensable (Lewis et al., 2004; Reed et al., 2001; Sunahara et al., 2000).

CpG islands are 200-2,000 bp long DNA sequences enriched for CpG dinucleotides that are found in the promoter regions of most mammalian genes. Methylation of the cytosines in these dinucleotides is associated with transcriptional repression, while an unmethylated state is associated with active transcription. Preliminary observations suggested that, on average, imprinted regions contain more CpG islands than biallelically expressed genes (Paulsen et al., 2000). However, as the number of identified imprinted genes grew, allowing more powerful sequence analyses, it was realized that, while the overall per locus number of CpG islands was similar to that of biallelically expressed genes, imprinted genes contain more intragenic CpG islands (Hutter et al., 2006). As it will be discussed below, in most cases intragenic CpG islands control the expression of overlapping non-coding RNAs, which are often essential for the establishment and regulation of parental-specific expression.

Gene-ID	Coordinates	Associated gDMR(s)	Location of gDMR(s)	Insulator Binding
Sfmbt2	chr2:10370510-10595253	chr2:10372484-10375267	Intronic	No binding
Zfp64	chr2:168893331-168955587	-	-	-
Phf17	chr3:41555734-41616862	-	-	-
Htra3	chr5:35652033-35679782	-	-	-
Axl	chr7:25757273-25788705	-	-	-
Inpp5f	chr7:128611328-128696425	chr7:128688124-128688233	Intronic	CTCF/Cohesin
Cdh15	chr8:122848374-122867397	chr8:122864938-122865178	Intronic	CTCF/Cohesin
Snx14	chr9:88376747-88438951	-	-	-
Pde4d	chr13:109954109-109955963	-	-	-
Htr2a	chr14:74640840-74706859	-	-	-
Slc38a4	chr15:96994823-97055956	chr15:97054000-97055741	Intronic	No binding
Pde10a	chr17:8525372-8986648	-	-	-
Impact	chr18:12972252-12992948	chr18:12972847-12974748	Intronic	CTCF
Tbc1d12	chr19:38836579-38919923	-	-	-

Table 2.2: Isolated imprinted genes in the Mouse. In Gene-ID column red denotes maternally expressed, blue denotes paternally expressed. In associated gDMR column, red denotes maternally imprinted and blue denotes paternally imprinted. In insulator binding column red denotes insulator binding to the maternal allele, blue denotes insulator binding to the paternal allele, and black denotes that the allelic-specific insulator binding is currently unknown.

CHROMATIN REGULATION OF GENOMIC IMPRINTING: DNA METHYLATION

DNA methylation is a biochemical reaction in which a methyl group is added to a cytosine or adenosine base. This process is primarily restricted to Cs in CpG dinucleotides in most cell types. Nevertheless, ES cells also contain significant CH (CA, CC or CT) methylation (Ramsahoye et al., 2000). Moreover, in the brain, neuron but not glial cells accumulate high levels of CH (CA, CC or CT) methylation as the animal matures and reaches adulthood (Lister et al., 2013). The great majority of CpG dinucleotides in the genome are constitutively methylated with the notable exception of the aforementioned CpG islands (CGIs). Besides the defining preferential expression according to parent-of-origin, differential DNA methylation of the parental alleles within regulatory regions (usually CGIs) is the most common feature of imprinted genes. Indeed, nearly all imprinted genes have been associated with at least one differentially methylated region (DMR) that is essential for their imprinted expression, and it has often been stated that methylation is the epigenetic signature that best defines the "imprint" of imprinted genes (Barlow and Bartolomei, 2014).

DNA methylation is believed to be the optimal epigenetic mark for the life cycle of genomic imprinting due to four properties: (1) it can be covalently attached to the individual parental chromosomes in their haploid state in the germ line, (2) it can be inherited during DNA-replication and therefore passed through daughter cells after cell division, (3) it can influence gene expression, and (4) it is a reversible and thus can be erased when it is time to establish a new mark (Bartolomei and Ferguson-Smith, 2011).

Erasure of Imprints: Demethylation

As part of the imprinting life cycle each separate germ line must erase its parent-specific methylation marks to achieve a genomic clean slate where new imprints can be established according to the sex of the individual (Figure 1). This in turn allows the individual to transfer an appropriate program of imprinted expression to its progeny. Failure to erase the marks of the parental generation will result in either absence or overabundance in the expression of imprinted genes in the subsequent generation. Here I will first discuss the general mechanism by which DNA is demethylated followed by a description of how imprinting marks are specifically erased.

DNA demethylation

For many years it has been appreciated that the DNA can be demethylated according to the developmental stage or in a cell-type specific manner. In fact, genome-wide demethylation occurs at two developmental stages: (1) in both sets of primordial germ cells from E8.5 to approximately E15, (2) in both maternal and paternal genomes between the zygote and blastocyst stages. However, the mechanistic nature of these demethylation events is beginning to be understood only now.

Passive Demethylation

Removal of DNA methylation can, in principle, be accomplished by two broad mechanisms: passive demethylation and active demethylation. Passive demethylation occurs during DNA replication by inactivating the molecular machinery responsible of reproducing the methylation marks of the template strand into the daughter strand. After multiple cell divisions

this process will entirely dilute out, and thus erase the methylation marks at a certain genomic locus. Passive demethylation of both parental genomes, occurring between the zygote and blastocyst stage, appears to result from lack Dnmt1 activity, the enzyme responsible for maintenance of DNA methylation, which is preferentially localized to the cytoplasm during this stage (Hirasawa et al., 2008).

Active Demethylation

Active demethylation refers to the enzymatic conversion of 5-methylcytosine back to cytosine (Li and Zhang, 2014). The mechanisms of active demethylation are only now beginning to be understood, and only quite superficially. A major breakthrough has been the discovery that Ten-eleven translocation (Tet) enzymes can oxidize 5-methylcytosine to produce 5-hydroxymethylcytosine (5hmC). These enzymes are able to further oxidize 5hmC into to 5-formylcytosine (5fC) and to 5-carboxylcytosine (5caC), which are proposed intermediate modifications in the process of converting 5-methylcytosine back to cytosine.

Two mechanisms have been suggested to achieve cytosine restoration: direct removal of the oxidized group at position 5 and complete base substitution through the DNA-repair pathway. Today only demethylation through DNA-repair has been confirmed. This occurs through the base excision repair pathway (BER), in which thymine DNA glycosylase (TDG) replaces 5fC and 5caC with unmethylated cytosines. Indeed, TDG null embryos exhibit altered epigenetic regulation, including hypermethylation in regulatory regions of key developmental genes, and die before term (Cortázar et al., 2011; Cortellino et al., 2011).

Erasure of Imprinted Differentially Methylated Regions in Primordial Germ Cells

The process of erasing imprinted DMRs takes place in primordial germ cells (PGCs) as they develop into the gonads. Two major events of demethylation occur during PGC development. First, genome-wide demethylation is observed between E6.5 and E9.5 as PGCs proliferate and migrate. This is mostly due to passive demethylation as the machineries of *de novo* methylation and its maintenance are both significantly down regulated and absent from the nucleus of PGCs (Kurimoto et al., 2008; Seisenberger et al., 2012). DMRs of imprinted genes, however, maintain their methylation state stably during this first wave of demethylation and are independently erased during a second demethylation event between E10.5 to E13.5 as PGCs enter the genital ridge (Seisenberger et al., 2012). During this stage PGCs increase their proliferation rate suggesting passive demethylation (Kagiwada et al., 2013) is the dominant operating mechanism.

However, recent work indicates that while Tet1 is not essential for demethylation of most genomic regions in PGCs it is important for that of imprinted DMRs (Yamaguchi et al., 2012, 2013). This was recognized thanks to the phenotype of the progeny of Tet1 deficient males.

Studies of Tet1-deficient fathers show that a substantial proportion of their progeny die during embryonic stages. Offspring that survive and reach term show a large distribution of sizes skewed towards smaller values. Some of these animals die soon after birth and a similar distribution of sizes is observed among those that survive. This pattern is highly correlated with the state of the placenta, where underdeveloped newborns are associated with smaller placentas with anatomical abnormalities. Interestingly, placental and growth abnormalities observed in affected embryos are reminiscent of the phenotype observed in paternal deletions of Paternally Expressed Gene 10 (Peg10) (Ono et al., 2005). Indeed, the DMR associated with Peg10 is overly

methylated in many of these embryos, together with most other maternally imprinted regions. Notwithstanding, it is important to emphasize that approximately half of the progeny of Tet1-deficient males is normal in size and displays normal physiology. Progeny of Tet1-deficient mothers shows 25% embryonic lethality but does not have a significant effect in the size of surviving embryos.

This study leads to the conclusion that the absence of Tet1 in both germ lines compromises the ability to erase imprints in the parental genomes during PGC development. In the male germ line, the failure to erase imprints in maternal chromosomes will result in a haplotype containing both maternal and paternal specific imprints. Passage of such genome to the next generation will result in abnormal silencing or overexpression of imprinted genes leading to the observed phenotypes in affected embryos. On the other hand, failure to erase imprints of paternal chromosomes still results in an appropriate haplotype containing only paternal-specific imprints and thus normal expression of imprinted genes (**Figure 2.2**) leading to unaffected embryos. An equivalent situation occurs in the female germ line and the substantial difference in maternal and paternal deficiencies of Tet1 are probably due to the fewer imprinting centers that are established in the male germ line (4 up to date) compared to the female germ line (28 imprinting centers up to date) (Wang et al., 2014).

Nevertheless, it is important to point out that abnormalities of imprinted DMRs were highly variable across the affected embryos, and an atypical methylation of a given DMR was

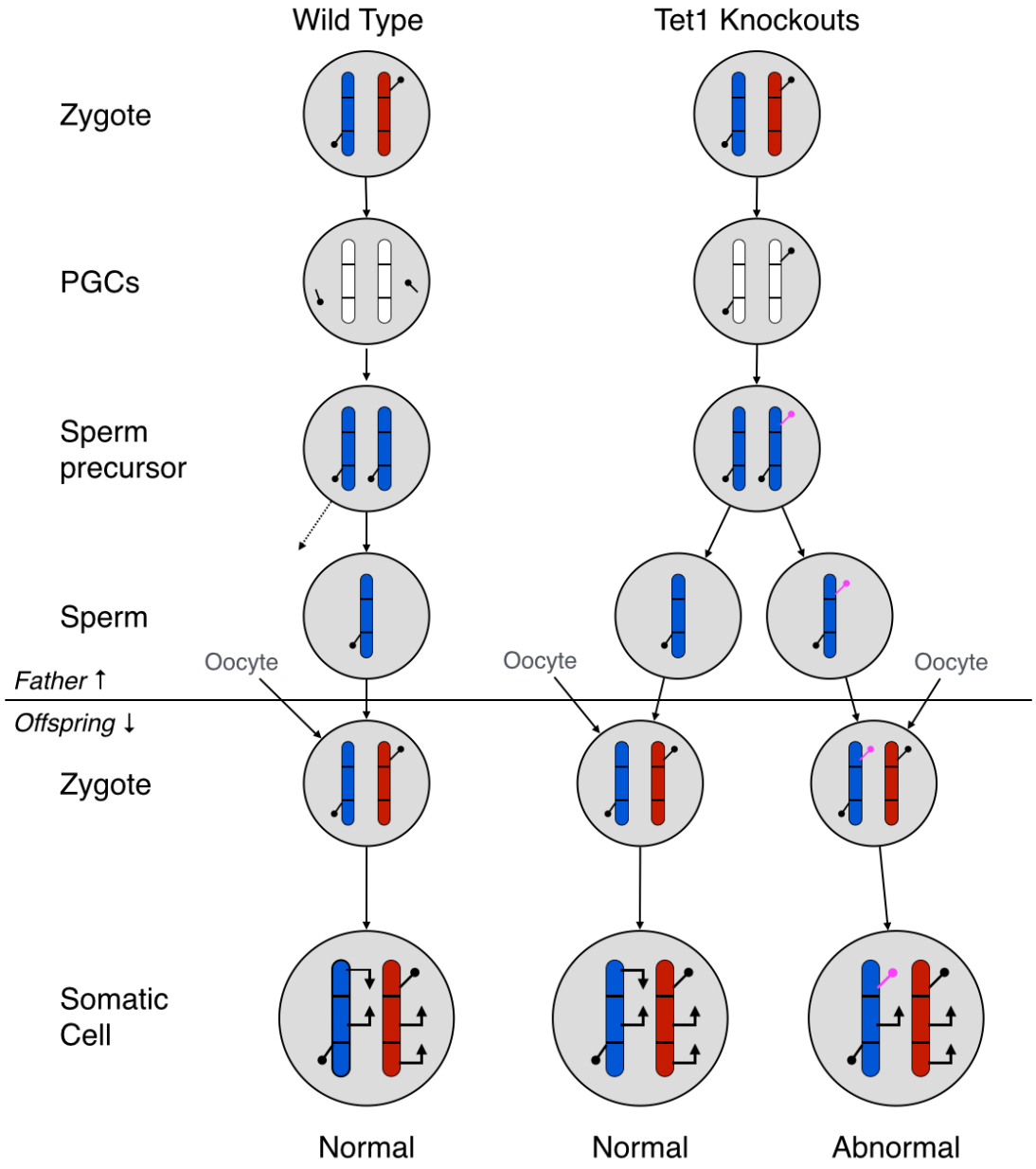

Figure 2.2: Effect of Tet1 paternal knockout on a maternally methylated DMR

not predictive of the state of other imprinted DMRs. This suggests that erasure of imprints can also occur through Tet1 independent mechanisms, perhaps by passive demethylation, and that compensation between Tet1-dependent and -independent mechanisms may vary from one germ cell to the next (Yamaguchi et al., 2013). Accordingly, it has been observed that DMRs of imprinted genes exhibit different onsets and completions of demethylation, and there is substantial variability among-cells in the timing of this process (Lee et al., 2002).

Despite this remarkable progress, several important questions regarding the erasure of imprints in PGCs still remain. How is methylation at DMRs protected during the first genome-wide wave of passive demethylation? Although Dnmt1, the enzyme responsible for maintenance of methylation, remains present in the nucleus during this stage (Hajkova et al., 2002) other components essential for the targeting of Dnmt1 to methylated regions are absent from the nucleus at that time. Thus, it is possible that alternative mechanisms for the targeting Dnmt1 to imprinted regions are active during this stage. Related to this question is whether DMRs are actively or passively demethylated. Tet enzymes are not synonymous with active demethylation. Indeed, in addition to induction of the base excision repair (BER) pathway, Tet enzymes can facilitate passive demethylation since Dnmt1 has drastically lower affinity towards 5hmC compared to 5mC (Guo et al., 2014; Hashimoto et al., 2012a). Consistent with this possibility is the fact that PGCs increase their replication rate during the time window of DMR erasure (Kagiwada et al., 2013). On the other hand, previous work has shown activity of the BER pathway during this time window (Hajkova et al., 2010) and a recent report indicates BER-dependent active demethylation of the Igf2-H19 DMR (Kawasaki et al., 2014). In the future it will be important to assess the genome-wide state of parental imprints in PGCs deficient in BER activity. Finally, it is still unclear what conditions induce the second wave of DNA

demethylation in PGCs, which result with erasure of DMRs. Which factors target the demethylation machinery to these regions and why do different imprinted genes have different timings of demethylation are also open question. Characterizing the functions of key components of the demethylation machinery coupled with chromatin profiles of DMRs during the erasure time window will be necessary to answer these questions.

Establishment of Imprints: *De Novo* DNA Methylation

De Novo DNA Methylation

Methylation marks of the parental genomes are erased between E6.5 and E9.5 in the mouse embryo, before primordial germ cells begin differentiation (Seisenberger et al., 2012). Thus, as differentiation occurs, the opportunity to establish new imprints according to the sex of the individual becomes available. At this time what will constitute the maternal and paternal genomes of the next generation are isolated in the female and male germ lines, respectively. Therefore regulation of imprinted regions are no longer required to occur *in cis*, and genome-wide mechanisms can be employed in the establishment of imprints.

De novo methylation is regulated by Dnmt3 enzymes. Dnmt3a and Dnmt3b are methyltransferases that attach methyl groups to previously unmethylated cytosines. A related protein, Dnmt3L, does not contain the methyltransferase activity but performs an essential modulatory role to the activity of Dnmt3a and Dnmt3b. These enzymes introduce methylation to 60-80% of CpGs in the genome suggesting methylation is the default state for most genomic regions (Smith and Meissner, 2013a). However, a notable exception from this pattern is the common unmethylated state of CpG islands, which will be discussed below. Depending on the

nature of the genomic region, the methylation machinery will function in concert with different chromatin regulators (Smith and Meissner, 2013b). For instance at repressed promoters, the Dnmt3 complex functions in concert with histone deacetylases (HDACs), methyltransferases of the histone 3 lysine 9 (H3K9), and nucleosome remodelers to induce heterochromatin assembly (Myant et al., 2011). At transposable elements, de novo methyltransferase recruitment depends on pi-RNA expression (Aravin et al., 2008).

As it appears that the default state of most genomic cytosines is to be methylated, targeted protection is essential for proper epigenetic configuration of the naïve genome (Hackett and Surani, 2013; Smith and Meissner, 2013b). Multiple mechanisms are thought to mediate regional inhibition of *de novo* methylation. These include: proximal sequence elements, DNA-binding factors including cell-type specific transcription factors and CpG island binding proteins as Cfp1, histone modifications, overlapping transcription, and localized RNA interruptions of dsDNA that form RNA-DNA hybrids and display a single DNA strand known as an R-loop (Chotalia et al., 2009a; Ginno et al., 2012; Lienert et al., 2011; Ooi et al., 2007; Stadler et al., 2011; Thomson et al., 2010). These mechanisms can synergistically act to control the activity of the *de novo* methylation complex at specific genomic regions. For instance, interactions between Dnmt3l and the amino terminus of histone H3 normally recruit Dnmt3a and Dnmt3b to genomic regions. However, trimethylation of lysine 4 of H3 (H3K4me3) abolishes this interaction. This mechanism allows *de novo* methylation to occur only at regions containing unmethylated H3K4 (Ooi et al., 2007). Interestingly, Cfp1 is essential for the presence of H3K4me3 in CGIs (Thomson et al., 2010), providing an interesting example of the integration of multiple mechanisms in localized protection from *de novo* activity.

Regulatory mechanisms of the *de novo* methylation complex such as these, allow differential treatment of homologous loci in the germ lines depending on the sex of the individual.

Establishment of Genomic Imprints

Once the genome has been stripped of its methylation and other epigenetic marks in the germ line, the epigenetic landscape is re-introduced. It is at this stage that imprinted regions become differentially regulated according to the sex of the individual. Re-establishment of the methylation landscape occurs at different developmental time points and cellular contexts in the male and female gonads. Although establishment of germ line DMRs (gDMRs) relies on similar mechanisms in the male and female germ line there are important differences between the two, with substantially more DMRs methylated in oocytes than sperm (Hanna and Kelsey, 2014).

Maternal Imprints

In mice, PGCs undergoing meiosis in the female are arrested at Prophase I from E13.5 until parturition when maturation and ovulation are set to ensue. *De novo* methylation of the female germ line occurs in arrested cells in the later portion of oocyte maturation and is completed when oocytes finish growing and enter a transcriptionally silent stage. Thus, no DNA replication occurs from the time of imprint establishment to the formation of the mature egg in the female germ line. Interestingly, maternal DMRs are established asynchronously but in a stereotypical order according to the diameter of the oocyte rather than its age (Kobayashi et al., 2009; Lucifero, 2004). The oocyte genome is largely unmethylated except for gene-body CGIs

(Shirane et al., 2013) where the Dnmt3a-Dnmt3L complex provides *de novo* methylation activity (Kaneda et al., 2004, 2010).

Currently, 28 maternally imprinted germ line DMRs (gDMRs) have been reported in mice, 20 of which are associated with parent-of-origin expression of at least one gene (Wang et al., 2014). These maternally imprinted DMRs consist of CpG islands at the promoters of coding and non-coding genes (Hanna and Kelsey, 2014). Intensive investigation searched for factors uniquely localized to maternally imprinted regions and that could account for their methylation. Recent work, however, suggests maternal imprints are established as part of a genome-wide mechanism, which in turn is dependent on active transcription over the regions that will become differentially methylated. Thus, it is suspected that transcription itself serves as the signal that helps targets methylation. This is supported by the observation that contrary to sperm, DNA methylation in oocytes is mostly confined to gene bodies of actively transcribed genes. Accordingly, 24 of 28 maternal gDMRs are localized in intronic regions (Wang et al., 2014) and it was demonstrated that transcription over future DMRs of both Gnas and Snrpn associated in the germ line is essential for establishment of maternal methylation (Chotalia et al., 2009b; Smith et al., 2011).

However, as mentioned above, transcription over future DMRs is not restricted to imprinted regions in oocytes. Instead, thousands of additional intragenic CGIs are methylated in the oocyte but not the sperm (Smallwood et al., 2011). Contrary to imprinted DMRs, however, additional oocyte-specific methylated sites are mostly lost after the oocyte is fertilized and before embryo implantation in a period of both passive and active demethylation (Smallwood et al., 2011; Wang et al., 2014). These observations suggest that rather than a targeted establishment of imprinting methylation marks in the oocyte, a directed protection of imprinted DMRs from

demethylation in the embryo is what determines the final set of maternally imprinted genes in a new organism. Indeed, factors that specifically protect imprinted regions in the embryo have been identified and will be discussed in later section.

Paternal Imprints

In the mouse male germ line, reestablishment of methylation begins before meiosis at E13.5 during mitotic arrest of spermatogonia precursor cells and is generally completed by E17.5. Thus, contrary to the female germ line, multiple cycles of DNA replication and cell divisions, where DNA methylation must be maintained, take place before mature sperm are generated. Both Dnmt3a and Dnmt3b are present in sperm and provide *de novo* activity together with their obligated cofactor Dnmt3L (Kaneda et al., 2004). In contrast to oocytes, the sperm genome is universally hypermethylated with the exception of CpG islands, which are, by and large, devoid of methylation (Erkek et al., 2013; Smallwood et al., 2011; Wang et al., 2014).

Only 4 paternally methylated DMRs have been reported in mouse sperm all of which are associated with the parental-specific expression of at least one gene. Unlike the maternally methylated promoter CGIs, paternal DMRs resemble GC-rich DNA elements located in intergenic regions (Hanna and Kelsey, 2014). Methylation of paternal DMRs is believed to occur as part of the global genome methylation observed in sperm precursors.

Strikingly, one paternal DMR associated with the Rasgrf1 gene, undergoes a unique form of establishment of imprint. Out of all four paternal DMRs, the Rasgrf1 DMR of chromosome 9 is the only one that requires the function of Dnmt3b, which is the principal *de novo* methyltransferase of various repetitive elements in the genome. Dnmt3b function can be targeted by PIWI RNAs (piRNAs), which recognize actively transcribing RNAs through sequence

complementary and induce chromatin rearrangements at these genomic loci. Residing within the Rasgrf1 associated DMR is the RMER4B retrotransposon, which is present in a few thousands copies throughout the mouse genome and contains sequences recognized by a piRNA cluster in chromosome 7. An RNA transcribed over the Rasgrf1 element is specifically expressed in prospermatogonia and recognized by piRNAs. This leads to the recruitment of the Dnmt3b complex, and therefore the male germ-line specific methylation of the Rasgrf1 DMR (Watanabe et al., 2011).

A Working Model underlying the Establishment of Parental Imprints

The current knowledge of *de novo* methylation in the gametes suggest that rather than being part of a targeted mechanism, the establishment of parental imprints is part of a default methylation processes in both germ lines (although there is less direct evidence in the male germ line) as summarized in Figure 3. In this model maternal imprints are set as part of a general process in which the oocyte genome undergoes methylation in CGIs associated with transcriptional activity. Paternal imprints are set as part of the global hypermethylation observed in intergenic regions in male gametes. Regions methylated in the male germ line will elude methylation in oocytes due to global hypomethylation. On the other hand, regions that are methylated in oocytes escape methylation in the male germ line due to the global protection of CpG islands. This model predicts the existence of many DMRs in the zygote, in a much higher number than actually associated with imprinted genes. Indeed, it is believed that the global demethylation of both parental genomes that occurs between the zygote and the blastocyst stage removes the majority of DMRs except from those associated with imprinted genes, which are

protected by specific factors that will be discussed below. Thus, the final set of DMRs appears to result from selective maintenance of marks established in the germ lines.

Inheritance of DNA methylation throughout cell divisions

Maintenance of DNA Methylation During DNA-Replication

An essential feature of DNA methylation, which allows it to be the primary mark for parent-of-origin allelic expression, is its ability to be maintained throughout DNA-replication so it is inherited despite cell divisions. Failure of this maintenance will result in passive loss of methylation after multiple cell divisions. The palindromic nature of CpG dinucleotides provides a simple and reliable mechanism to achieve reproduction and maintenance of methylation patterns. Indeed, after DNA replication, methylated CpGs on the template strand is paired with unmethylated CpGs of the new strand forming hemimethylated DNA. Hemimethylated CpGs are recognized and methylated in the new strand by another methyltransferase, Dnmt1. This enzyme shows high affinity for hemimethylated DNA (Jeltsch, 2006), is most abundant during S-phase (Kishikawa et al., 2003), and localizes to the replication fork (Leonhardt et al., 1992). Embryos homozygous for a null allele of Dnmt1 have severe losses of DNA methylation and only develop to the 8-somite stage (Lei et al., 1996).

Recently, a factor essential for Dnmt1 maintenance activity has been identified. Uhrf1 directs Dnmt1 to hemimethylated sites and directly presents hemimethylated base pairs to Dnmt1(Arita et al., 2008). Furthermore, Dnmt1 activity can be inhibited *in cis* by the expression of locally transcribed non-coding RNAs forming stem-loops that directly interact with the enzyme and sequester it away from hemimethylated DNA, preventing transcription refractory

chromatin states (Di Ruscio et al., 2013). This mode of regulation may play a role in paths of cell differentiation by inducing a passive loss of DNA methylation in particular loci. Considering that many imprinted genes are dynamically regulated according to cell type it will be important to investigate whether these types of RNAs exist in imprinted regions and play a role in allelic-specific regulation. Another interesting finding is that the binding and enzymatic activity of Uhrf1 and Dnmt1, respectively, is several orders of magnitude weaker towards hemi-hydroxymethylated (5hm) DNA *in vitro* (Hashimoto et al., 2012b). This raises the possibility that 5hmC facilitates passive methylation loss.

In summary, DNA methylation can be faithfully maintained after DNA replication allowing the inheritance of imprinting marks. However, this step also provides an important node of epigenetic regulation whereby DNA methylation can be modified in a locus and context-dependent manner.

DNA Demethylation in the Early Embryo

In the early embryo, mechanisms insuring methylation maintenance versus erasure play an essential role in the life cycle of genomic imprinting. Once both parental genomes are joined in the zygote, genome-wide reprograming of DNA methylation occurs, such that marks associated with imprinted genes are maintained while unnecessary germ-line specific methylation is removed (Seisenberger et al., 2012). Demethylation occurs in the paternal pronucleus immediately after fertilization by the Tet3-mediated conversion of 5mC to 5hmC. Early studies analyzing 5hmC in the zygote did not detect the modified base in the maternal pronucleus (Gu et al., 2011). More recent work, however, indicates that both Tet3 and 5hmC

appear in the maternal pronucleus at later stages and at lower concentrations than in the paternal pronucleus (Shen et al., 2014).

After this initial period it is observed that both parental genomes undergo widespread demethylation during the cell divisions required to reach the blastocyst stage. Two recent papers demonstrate that there are actually three different mechanisms operating in global demethylation of the early embryo. Beside the above-described passive and active demethylation processes, passive demethylation of 5hmC-converted regions occurs in both genomes (Guo et al., 2014; Shen et al., 2014). This indicates that oxidation of 5mC is not synonymous with active demethylation (Gkountela and Clark, 2014). It is estimated that demethylation of 75% and 87% of the paternal and maternal genomes, respectively, occurs through either of the passive mechanisms (Guo et al., 2014).

Protection of Parental Imprints

The observation of passive and active coordinated systems to demethylate both genomes raises the question of how parental imprints are maintained through this process. Maintenance of a subset of methylation marks in the early embryo, including parental imprints, requires Dnmt1 and Uhrf1 as describe above. However, recent work has identified several additional proteins involved in the protection of differentially methylated regions in the early embryo associated with imprinted gene expression. Some of these proteins are specific to the regulation of parental imprints while others have a broad role in the maintenance of methylation. Moreover, the functions they provide are required for maintenance of methylation in the imprinted allele or protection from methylation of the non-imprinted allele (Kelsey and Feil, 2013).

Zfp57 & Kap1

In humans, mutations in the Zinc finger protein 57 (Zfp57) gene lead to transient neonatal diabetes, a disorder commonly associated with the loss of imprinting of *Plagl1* (Pleiomorphic adenoma gene-like 1). Analysis of the Plagl1 DMR revealed that indeed this region is hypomethylated in individuals with Zfp57 mutations alongside hypomethylation of DMRs associated with several other imprinted genes (Mackay et al., 2008). Another study showed that deletion of Zfp57 in mice results in partial neonatal lethality, and is accompanied by abnormal expression of imprinted genes and loss of methylation in DMRs. Furthermore, it was observed that Zfp57 is expressed in the preimplantation embryo where it was found to bind to an imprinted DMR which was tested (Li et al., 2008).

Subsequently, Zfp57 has been shown to interact with Kap1, a scaffold protein that binds the H3K9 methyltransferase Setdb1 and the heterochromatin protein 1 (HP1). This complex is preferentially located in the imprinted allele of DMRs. Accordingly, H3K9me3 was observed in imprinted regions in a Zfp57 dependent manner. This complex also interacts with Dnmt1 and Uhrf1, and recruits methylation maintenance activity to imprinted regions. Accordingly, loss of Zfp57 results in loss of methylation in these regions. Finally, Zfp57 has been shown to recognize the hexanucleotide motif, TGCCGC, in a methylation-sensitive manner. Remarkably, this motif is present in all known imprinted regions in an average of two copies per locus. Thus, it appears Zfp57 is required for protection of methylation of most, if not all, imprinted DMRs. Several additional regions, not currently associated with imprinted genes, also contain multiple copies of the hexanucleotide motif and exhibit Zfp57 binding. Interestingly, Zfp57 also protects DNA methylation in these loci, which suggests that yet to be unidentified imprinted genes may reside in these regions (Quenneville et al., 2011).

Other Factors Associated with Maintenance of Methylation in the Early Embryo

Other factors are known to participate in the protection of DMRs in the early embryo, however, unlike Zfp57 they do not appear to be specific to imprinted regions. One example is that of Pgc7 (also known as Stella or Dppa3). This protein is essential for the maintenance of methylation at multiple genomic loci including several but not all maternally and paternally methylated regions (Nakamura et al., 2006). Pgc7 binds to H3K9me2, a chromatin mark, which co-localizes with DNA methylation in specific regions, and inhibits Tet3 mediated 5mC oxidation (Nakamura et al., 2012).

In summary, these results indicate that the final set of imprinted DMRs is determined by the specific binding of proteins, particularly Zfp57, which prevents DNA demethylation of these regions in the early embryo. It will be important to determine the identity of other factors associated with these proteins and their interaction with other chromatin regulators. Characterization of the genome-wide localization patterns of such factors could allow the identification of novel imprinted genes and provides a target to specifically perturb imprinted regions without interfering with non-imprinted genes.

Regulation of Gene Expression Through DNA Methylation

As DNA methylation is a mechanism for suppressing gene expression differential methylation of the parental alleles provides a mechanism to regulate allele-specific expression of genes. It is important to note that the methylated allele is not always the one that is actively silenced or expressed. This indicates that methylation does not directly control expression of imprinted genes but rather interacts with additional chromatin regulators and the transcription

machinery in a genomic-context dependent manner to achieve allele-specific expression. Below I describe the various effects that DNA methylation has on gene expression followed by a description of its association with other chromatin regulators.

DNA methylation is conventionally associated with a repressive state of transcription throughout the genome for both genes and transposable elements. For genes this pattern revolves around the methylation state of CpG islands (CGIs) where methylation usually results in transcriptional suppression of the gene. The genomic distribution of CGIs can be organized in three categories: (1) CGIs within promoters that comprise approximately 50% of all CGIs, (2) intragenic CGIs that comprise 25% of all CGIs, and (3) intergenic CGIs accounting for the remaining 25%. The absolute numbers and ratios of these categories are almost the same in mouse and humans (Illingworth et al., 2010).

Promoter CGIs

CGI promoters are, by and large, unmethylated, a state associated with genes poised for transcriptional activity (Deaton and Bird, 2011). Multiple components of the chromatin and transcriptional machinery are associated with this transcriptional state, including H3K4 trimethylation and RNA Polymerase II itself (Guenther et al., 2007; Hargreaves et al., 2009). Nevertheless, a smaller percentage of methylated CGI promoters are observed in somatic cells in specific developmental stages and/or tissue types where they control the repression of genes that play key differentiation roles for the corresponding cells (Meissner et al., 2008; Mohn et al., 2008; Schilling and Rehli, 2007; Shen et al., 2007).

Inter- and Intragenic CGIs

Understanding the function of inter- and intragenic CGIs has been a great challenge for several years since they are not associated with annotated promoters. New evidence suggests that both classes of these CGIs may represent cryptic promoters that also regulate the transcription initiation of RNAs, most of which are non-coding. However, contrary to promoter CGIs, the majority of inter- and intragenic CGIs exhibit substantially denser methylation and are therefore believed to indicate repression (Illingworth et al., 2010; Maunakea et al., 2010). Interestingly, the methylation patterns for intragenic CGIs, in particular, can significantly vary depending on tissue type (Illingworth et al., 2008, 2010; Maunakea et al., 2010). It has been suggested that intragenic CGIs control the expression of long non-coding RNAs (lncRNAs) that in turn regulate the expression of the host gene in a context dependent manner (Deaton and Bird, 2011). As discussed above, maternally imprinted DMRs are mostly intragenic while paternally imprinted DMRs are mostly intergenic. As it will be discussed below both kinds of parental DMRs usually control the expression of lncRNAs. Before describing how differential methylation regulates allele-specific expression of imprinted genes I will introduce two important chromatin regulators involved in imprinting.

OTHER CHROMATIN REGULATORS OF GENOMIC IMPRINTING:

Long Non-coding RNAS

Advancement in high-throughput sequencing technologies has allowed the realization that while a small fraction of the genome has protein-coding potential, most of the genome is transcribed (Amaral et al., 2008). Most of these transcripts are long (usually >200nt) non-coding RNAs (lncRNAs), with no apparent open reading frame and they are almost exclusively located in the nucleus. This type of RNAs has been increasingly implicated in highly specialized and complex functions of the genome (Rinn and Chang, 2012).

LncRNAs are thought to influence chromatin and regulate gene expression in diverse ways. They are thought to exert their effects in either *cis* or *trans*. *Cis*-acting RNAs are constrained to their transcription site and directly act on the regulation of one or several linked genes on the same chromosome. On the other hand, *trans*-acting RNAs diffuse from their transcription site and can regulate multiple genes including those in other chromosomes. Thus, *trans*-acting RNAs, similar to transcription factors and small RNAs, are more likely to converge on, and act within, large gene networks. Nevertheless, both mechanisms involve direct action on targeted genes rather than secondary downstream effects.

Cis-acting lncRNAs (*cis*-lncRNAs) are particularly relevant for genomic imprinting due to their potential to achieve allele-specific effects. The localized function of these transcripts can be due to tethering via Pol-II to their transcriptional origin (Jeon and Lee, 2011; Wang et al., 2008). Alternatively, rapid degradation after transcription can limit their half-life only allowing diffusion to nearby targets but not ectopic sites (Clark et al., 2012; Tani et al., 2012). The

repressive role of *cis*-lncRNAs on expression can be achieved by recruiting chromatin regulators to target regions. In addition and not mutually exclusive, *cis*-lncRNAs could serve as essential scaffolds for the enzymatic activity of protein complexes. Such effect has been observed for the *trans*-lncRNA HotAir, which recruits the H3K27 methyltransferase Polycomb repressive complex 2 and the H3K4 demethylase LSD1 (Rinn et al., 2007; Tsai et al., 2010). The specific roles of lncRNAs in the regulation of imprinted genes will be discussed below.

Insulators

Insulators are DNA-binding proteins that can block interactions between enhancers and promoters when recruited to specific loci. The CTCF insulator is an essential insulator of interactions between regulatory elements, which are particularly important for tissue-specific gene expression. By its targeted binding of several specific DNA sequences, CTCF can restrict promoter-enhancer interactions in tridimensional space. This process requires the stable regulation of chromatin structure and conformation. In addition, CTCF is a major regulator of nuclear architecture since it can form boundaries between euchromatin and heterochromatin domains and determine interactions with nuclear lamina (Merkenschlager and Odom, 2013). Cohesin, a protein complex that regulates the separation of sister chromatids during cell divisions, interacts with CTCF and is essential for its insulator functions (Parelho et al., 2008; Wendt et al., 2008). Finally, CTCF binding is sensitive to DNA methylation, such that CTCF binds to hypomethylated DNA but not to hypermethylated DNA (Merkenschlager and Odom, 2013). The CTCF/Cohesin complex has been shown to regulate parent-of-origin specific expression in a subset of imprinted regions (discussed below).

REGULATION OF ALLELE-SPECIFIC EXPRESSION IN IMPRINTED CLUSTERS: MODELS

Two models of the regulation of allele-specific expression in imprinted clusters have been proposed, the non-coding RNA model and the Insulator model, where each utilizes a different set of molecules that are themselves regulated by DMRs. These two models assign the mode of regulation of parental allele specific expression in imprinted clusters into two broad categories. However, it should be kept in mind that each of the clusters assigned to these categories have their own nuances when analyzed at a more detailed level. Furthermore, the models are based on the study of only a handful of clusters so far, and the mechanisms of allele-specific expression have yet to be described for the majority of imprinted regions . It is thus possible that other models may emerge in the future.

lncRNA model

The regulation of allele-specific expression by a non-coding RNAs has been described for the Igf2r, Kcnq1, and Snrpn clusters. Here I will use the Igf2r cluster as an example of this model (**Figure 2.3**). Igf2r is maternally expressed in most tissues with the brain being a notable exception (Igf2r is thought to biallelically transcribed in the brain). The only DMR present in this cluster, which is established in the germ line, corresponds to a maternally imprinted CGI localized to an intronic region between exons 2 and 3 of the Igf2r gene. This DMR controls the expression of an lncRNA transcribed in the opposite direction of Igf2r. This lncRNAs has been named Airn (Antisense of Igf2r RNA) and is exclusively expressed from the paternal allele. Methylation of the ICR in the maternal allele inhibits the expression of Airn, while the

unmethylated ICR in the paternal allele actively transcribes Airn underlying its paternal specific expression. Several experiments have demonstrated that transcriptional interference of Airn over the Igf2r promoter is responsible for transcriptional silencing of Igf2r in the paternal allele. First, two types of truncations of the Airn transcript where generated by insertion of a polyA signal. The first type caused premature termination of Airn transcription thus preventing its overlap with the Ig2r promoter while the second maintained Airn transcription beyond the Igf2r promoter. The first truncation resulted with biallelic expression of Igf2r while the second had no effect. Furthermore, reversing the Airn promoter thus forcing Airn transcription from the same strand as Igf2r, also resulted with paternal transcription of Igf2r yielding biallelic expression. Finally, repositioning the Airn promoter 700bp before the Igf2r promoter, with no alteration of its orientation, repressed transcription of Igf2r even though marks of active chromatin such as H3K4me3 presence and H3K9me3 absence were still observed in the silenced Igf2r promoter, indicating Airn transcriptional interference is sufficient to silence Ig2r (Latos et al., 2012).

Flanking one side of the Igf2r gene are three solute carrier transporter genes: Slc22a1, Slc22a2, and Slc22a3. Interestingly, Slc22a2 and Slc22a3 are maternally expressed in the placenta while Slc22a1 is biallelically expressed. None of these genes contain a DMR. Instead, the imprinting expression of Slc22a3 is controlled by the *in cis* binding of Airn to the Slc22a3 promoter chromatin, which recruits G9, a histone modifier responsible for the methylation of H3K9. Accordingly, loss of either Airn or G9 results in biallelic expression of this gene (Nagano et al., 2008). Thus, Airn regulation of the Igf2r cluster exemplifies how a lncRNA, whose transcription is controlled by a DMR, can regulate the parental-specific expression of an entire imprinted cluster.

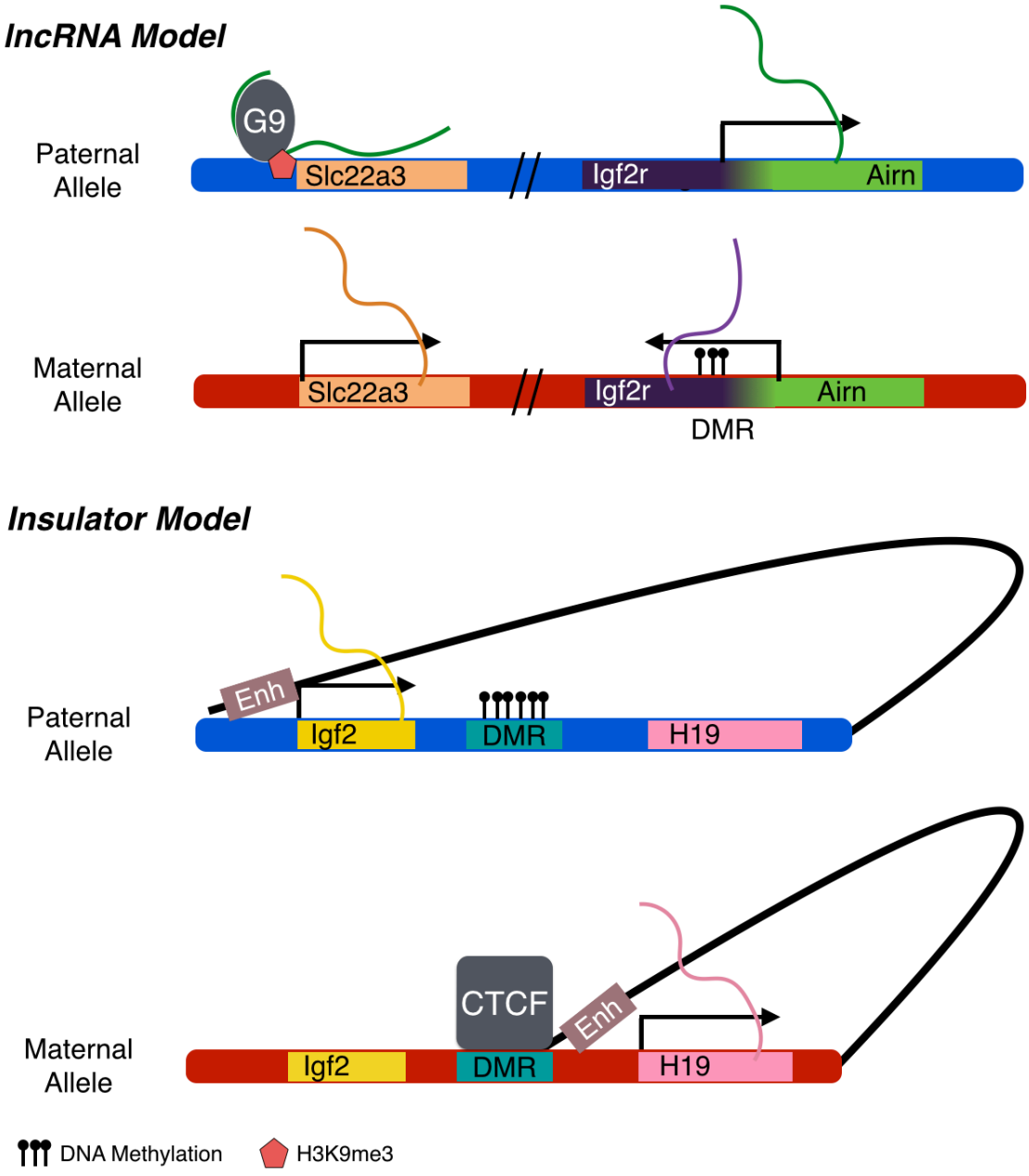

Figure 2.3: Cis-acting mechanism at imprinted gene clusters

For the majority of imprinted clusters at least one lncRNA has been identified but the mechanistic role of only four of them has been described (Barlow and Bartolomei, 2014). Thus, it is possible that imprinted expression of multiple clusters is orchestrated by lncRNAs. However, as it will be discussed below RNA independent allele-specific silencing has also been observed in at least one cluster.

Insulator model

The Igf2 and H19 genes are located next to each other in one of the imprinted cluster regions in distal chromosome 7. Igf2 is expressed from the paternal allele while H19 is expressed from the maternal allele. Early studies indicated that deletions of a nearby endoderm-specific enhancer result in loss of H19 expression if maternally inherited, while loss of Igf2 expression if paternally inherited, suggesting that both genes are regulated by a common enhancer in a parent-of-origin manner (Leighton et al., 1995). Indeed, these genes share multiple enhancers including those specific for skeletal and cardiac muscle development (Ainscough et al., 2000; Kaffer et al., 2000). The only DMR present in this cluster corresponds to a paternally imprinted CGI located in the intergenic region between Igf2 and H19, which was shown to control the imprinting of both genes (Thorvaldsen et al., 1998). Interestingly, the CTCF/Cohesin complex is able to bind the unmethylated DMR allele but not the methylated one. Using chromatin conformation capture, it was shown that long range interactions occur between enhancers and the Igf2 gene in the absence of CTCF, and thus this gene is expressed from the paternal allele (**Figure 2.3**). On the other hand, the unmethylated maternal allele is able to bind CTCF, which blocks the long range interactions with Igf2 and redirects the enhancer to the H19 promoter (Kurukuti et al.,

2006). Thus, in this region the methylation sensitive binding of CTCF to the DMR determines the paternal and maternal specific expression of Igf2 and H19, respectively.

Mouse Rasgrf1, and human Peg13 clusters are reported to follow a similar mode of allele-specific and methylation-sensitive CTCF binding, which is believed to regulate their imprinted expression (Court et al., 2014; Yoon et al., 2005). Interestingly, CTCF also binds in the unmethylated allele in the ICR of Meg3, however, loss of CTCF does not result in loss of imprinted expression but rather increases the expression of the active allele (Lin et al., 2011).

In a recent study, genome-wide CTCF and Cohesin binding was analyzed with allele-specific resolution in a subset of brain regions in the mouse (Prickett et al., 2013). CTCF-Cohesin binding was observed in the unmethylated allele in the majority of DMRs, including H19/Igf2, Peg13, and Mest. However, two genes, Plagl1 and Inpp5f_v2, showed equal CTCF-Cohesin binding in both the methylated and unmethylated alleles. Biallelic CTCF-Cohesin binding of the Plagl1 DMR was also observed in an independent study of mouse and human (Iglesias-Platas et al., 2013). Additionally, several other genes show only CTCF or Cohesin binding but not both, while other genes do not bind either. Finally, Magel2, a gene within the Prader-Willi cluster, has seven different CTCF binding sites near or within its gene body. Specific binding of CTCF-Cohesin to the unmethylated allele of Magel2 was observed at all binding sites (Prickett et al., 2013). Taken together, these results suggest substantial diversity within the genes associated with CTCF/Cohesin and support the idea that a specialized mechanism operates in every imprinted region.

CONCLUSIONS

Remarkable progress has been made recently regarding the mechanisms underlying the life cycle of genomic imprinting and the specific molecular players involved. These findings open new opportunities to further investigate and deepen our understanding of genomic imprinting. The stereotypical genomic locations of maternal and paternal imprinted DMRs, combined with identification of Zfp57 binding sites, may be used to detect novel imprinted genes. Furthermore, experimental manipulations of novel imprinting regulators, such as Zfp57 and Tet1, could facilitate functional epigenetic studies on imprinted genes. The majority of these experiments have been performed in the context of the embryo. However, substantial variation in the expression patterns of imprinted genes after parturition is increasingly appreciated yet the understanding of the mechanisms underlying spatiotemporal regulation of imprinted genes is still lacking. Finally, a selected number of recent publications have proposed communication between different imprinted regions with important consequences in gene expression (Boucher et al., 2014; Stelzer et al., 2014; Varrault et al., 2006). Therefore, future investigation of potential crosstalks between imprinted genes may reveal a new layer of imprinting regulation and suggest coordination between the biological functions of large subsets of imprinted genes.

REFERENCES

Ainscough, J.F., John, R.M., Barton, S.C., and Surani, M.A. (2000). A skeletal muscle-specific mouse Igf2 repressor lies 40 kb downstream of the gene. Development *127*, 3923–3930.

Amaral, P.P., Dinger, M.E., Mercer, T.R., and Mattick, J.S. (2008). The eukaryotic genome as an RNA machine. Science *319*, 1787–1789.

Aravin, A.A., Sachidanandam, R., Bourc'his, D., Schaefer, C., Pezic, D., Toth, K.F., Bestor, T., and Hannon, G.J. (2008). A piRNA Pathway Primed by Individual Transposons Is Linked to De Novo DNA Methylation in Mice. Mol. Cell *31*, 785–799.

Arita, K., Ariyoshi, M., Tochio, H., Nakamura, Y., and Shirakawa, M. (2008). Recognition of hemi-methylated DNA by the SRA protein UHRF1 by a base-flipping mechanism. Nature *455*, 818–821.

Barlow, D.P., and Bartolomei, M.S. (2014). Genomic Imprinting in Mammals. Cold Spring Harb. Perspect. Biol. *6*, a018382–a018382.

Bartolomei, M.S., and Ferguson-Smith, A.C. (2011). Mammalian genomic imprinting. Cold Spring Harb. Perspect. Biol. *3*.

Boucher, J., Charalambous, M., Zarse, K., Mori, M.A., Kleinridders, A., Ristow, M., Ferguson-Smith, A.C., and Kahn, C.R. (2014). Insulin and insulin-like growth factor 1 receptors are required for normal expression of imprinted genes. Proc. Natl. Acad. Sci. U. S. A.

Chotalia, M., Smallwood, S.A., Ruf, N., Dawson, C., Lucifero, D., Frontera, M., James, K., Dean, W., and Kelsey, G. (2009a). Transcription is required for establishment of germline methylation marks at imprinted genes. Genes Dev. *23*, 105–117.

Chotalia, M., Smallwood, S.A., Ruf, N., Dawson, C., Lucifero, D., Frontera, M., James, K., Dean, W., and Kelsey, G. (2009b). Transcription is required for establishment of germline methylation marks at imprinted genes. Genes Dev. *23*, 105–117.

Clark, M.B., Johnston, R.L., Inostroza-Ponta, M., Fox, A.H., Fortini, E., Moscato, P., Dinger, M.E., and Mattick, J.S. (2012). Genome-wide analysis of long noncoding RNA stability. Genome Res. *22*, 885–898.

Cortázar, D., Kunz, C., Selfridge, J., Lettieri, T., Saito, Y., MacDougall, E., Wirz, A., Schuermann, D., Jacobs, A.L., Siegrist, F., et al. (2011). Embryonic lethal phenotype reveals a function of TDG in maintaining epigenetic stability. Nature *470*, 419–423.

Cortellino, S., Xu, J., Sannai, M., Moore, R., Caretti, E., Cigliano, A., Le Coz, M., Devarajan, K., Wessels, A., Soprano, D., et al. (2011). Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair. Cell *146*, 67–79.

Court, F., Camprubi, C., Garcia, C.V., Guillaumet-Adkins, A., Sparago, A., Seruggia, D., Sandoval, J., Esteller, M., Martin-Trujillo, A., Riccio, A., et al. (2014). The PEG13-DMR and brain-specific enhancers dictate imprinted expression within the 8q24 intellectual disability risk locus. Epigenetics Chromatin *7*, 1–13.

Deaton, A.M., and Bird, A. (2011). CpG islands and the regulation of transcription. Genes Dev. *25*, 1010–1022.

Dindot, S. V., Person, R., Strivens, M., Garcia, R., and Beaudet, A.L. (2009). Epigenetic profiling at mouse imprinted gene clusters reveals novel epigenetic and genetic features at differentially methylated regions. Genome Res. *19*, 1374–1383.

Erkek, S., Hisano, M., Liang, C.-Y., Gill, M., Murr, R., Dieker, J., Schübeler, D., Vlag, J. van der, Stadler, M.B., and Peters, A.H.F.M. (2013). Molecular determinants of nucleosome retention at CpG-rich sequences in mouse spermatozoa. Nat. Struct. & Mol. Biol. *20*, 868–875.

Ginno, P.A., Lott, P.L., Christensen, H.C., Korf, I., and Chédin, F. (2012). R-Loop Formation Is a Distinctive Characteristic of Unmethylated Human CpG Island Promoters. Mol. Cell *45*, 814–825.

Gkountela, S., and Clark, A.T. (2014). A Big Surprise in the Little Zygote: The Curious Business of Losing Methylated Cytosines. Cell Stem Cell *15*, 393–394.

Gu, T.-P., Guo, F., Yang, H., Wu, H.-P., Xu, G.-F., Liu, W., Xie, Z.-G., Shi, L., He, X., Jin, S., et al. (2011). The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. Nature *477*, 606–610.

Guenther, M.G., Levine, S.S., Boyer, L.A., Jaenisch, R., and Young, R.A. (2007). A Chromatin Landmark and Transcription Initiation at Most Promoters in Human Cells. Cell *130*, 77–88.

Guo, F., Li, X., Liang, D., Li, T., Zhu, P., Guo, H., Wu, X., Wen, L., Gu, T.-P., Hu, B., et al. (2014). Active and passive demethylation of male and female pronuclear DNA in the Mammalian zygote. Cell Stem Cell *15*, 447–458.

Hackett, J. a, and Surani, M.A. (2013). DNA methylation dynamics during the mammalian life cycle. Philos. Trans. R. Soc. Lond. B. Biol. Sci. *368*, 20110328.

Hajkova, P., Erhardt, S., Lane, N., Haaf, T., El-Maarri, O., Reik, W., Walter, J., and Surani, M.A. (2002). Epigenetic reprogramming in mouse primordial germ cells. Mech. Dev. *117*, 15–23.

Hajkova, P., Jeffries, S.J., Lee, C., Miller, N., Jackson, S.P., and Surani, M.A. (2010). Genome-Wide Reprogramming in the Mouse Germ Line Entails the Base Excision Repair Pathway. Sciencemag.org.ezp-prod1.hul.harvard.edu.

Hanna, C.W., and Kelsey, G. (2014). The specification of imprints in mammals. Heredity (Edinb). *113*, 176–183.

Hargreaves, D.C., Horng, T., and Medzhitov, R. (2009). Control of Inducible Gene Expression by Signal-Dependent Transcriptional Elongation. Cell *138*, 129–145.

Hashimoto, H., Liu, Y., Upadhyay, A.K., Chang, Y., Howerton, S.B., Vertino, P.M., Zhang, X., and Cheng, X. (2012a). Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation. Nucleic Acids Res. *40*, 4841–4849.

Hashimoto, H., Liu, Y., Upadhyay, A.K., Chang, Y., Howerton, S.B., Vertino, P.M., Zhang, X., and Cheng, X. (2012b). Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation. Nucleic Acids Res. *40*, 4841–4849.

Hirasawa, R., Chiba, H., Kaneda, M., Tajima, S., Li, E., Jaenisch, R., and Sasaki, H. (2008). Maternal and zygotic Dnmt1 are necessary and sufficient for the maintenance of DNA methylation imprints during preimplantation development. Genes Dev. *22*, 1607–1616.

Hutter, B., Helms, V., and Paulsen, M. (2006). Tandem repeats in the CpG islands of imprinted genes. Genomics *88*, 323–332.

Iglesias-Platas, I., Court, F., Camprubi, C., Sparago, A., Guillaumet-Adkins, A., Martin-Trujillo, A., Riccio, A., Moore, G.E., and Monk, D. (2013). Imprinting at the PLAGL1 domain is contained within a 70-kb CTCF/cohesin-mediated non-allelic chromatin loop. Nucleic Acids Res. *41*, 2171–2179.

Illingworth, R., Kerr, A., DeSousa, D., Jørgensen, H., Ellis, P., Stalker, J., Jackson, D., Clee, C., Plumb, R., Rogers, J., et al. (2008). A novel CpG island set identifies tissue-specific methylation at developmental gene loci. PLoS Biol. *6*, 0037–0051.

Illingworth, R.S., Gruenewald-Schneider, U., Webb, S., Kerr, A.R.W., James, K.D., Turner, D.J., Smith, C., Harrison, D.J., Andrews, R., and Bird, A.P. (2010). Orphan CpG Islands Identify numerous conserved promoters in the mammalian genome. PLoS Genet. *6*.

Jeltsch, A. (2006). On the enzymatic properties of dnmt1: Specificity, processivity, mechanism of linear diffusion and allosteric regulation of the enzyme. Epigenetics *1*, 63–66.

Jeon, Y., and Lee, J.T. (2011). YY1 Tethers Xist RNA to the inactive X nucleation center. Cell *146*, 119–133.

Kaffer, C.R., Srivastava, M., Park, K., Ives, E., Hsieh, S., Batlle, J., Grinberg, A., Huang, S., and Pfeifer, K. (2000). A transcriptional insulator at the imprinted H19 / Igf2 locus A transcriptional insulator at the imprinted H19 / Igf2 locus. Genes Dev. 1908–1919.

Kagiwada, S., Kurimoto, K., Hirota, T., Yamaji, M., and Saitou, M. (2013). Replication-coupled passive DNA demethylation for the erasure of genome imprints in mice. EMBO J. *32*, 340–353.

Kaneda, M., Okano, M., Hata, K., Sado, T., Tsujimoto, N., Li, E., and Sasaki, H. (2004). Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting. Nature *429*, 900–903.

Kaneda, M., Hirasawa, R., Chiba, H., Okano, M., Li, E., and Sasaki, H. (2010). Genetic evidence for Dnmt3a-dependent imprinting during oocyte growth obtained by conditional knockout with Zp3-Cre and complete exclusion of Dnmt3b by chimera formation. Genes to Cells *15*, 169–179.

Kawasaki, Y., Lee, J., Matsuzawa, A., Kohda, T., Kaneko-Ishino, T., and Ishino, F. (2014). Active DNA demethylation is required for complete imprint erasure in primordial germ cells. Sci. Rep. *4*.

Kelsey, G., and Feil, R. (2013). New insights into establishment and maintenance of DNA methylation imprints in mammals. Philos. Trans. R. Soc. London Ser. B, Biol. Sci. *368*, 20110336.

Kishikawa, S., Murata, T., Ugai, H., Yamazaki, T., and Yokoyama, K.K. (2003). Control elements of Dnmt1 gene are regulated in cell-cycle dependent manner. Nucleic Acids Res. Suppl. 307–308.

Kobayashi, H., Yamada, K., Morita, S., Hiura, H., Fukuda, A., Kagami, M., Ogata, T., Hata, K., Sotomaru, Y., and Kono, T. (2009). Identification of the mouse paternally expressed imprinted gene Zdbf2 on chromosome 1 and its imprinted human homolog ZDBF2 on chromosome 2. Genomics *93*, 461–472.

Kurimoto, K., Yabuta, Y., Ohinata, Y., Shigeta, M., Yamanaka, K., and Saitou, M. (2008). Complex genome-wide transcription dynamics orchestrated by Blimp1 for the specification of the germ cell lineage in mice. Genes Dev. *22*, 1617–1635.

Kurukuti, S., Tiwari, V.K., Tavoosidana, G., Pugacheva, E., Murrell, A., Zhao, Z., Lobanenkov, V., Reik, W., and Ohlsson, R. (2006). CTCF binding at the H19 imprinting control region mediates maternally inherited higher-order chromatin conformation to restrict enhancer access to Igf2. Proc. Natl. Acad. Sci. U. S. A. *103*, 10684–10689.

Latos, P.A., Pauler, F.M., Koerner, M. V, Senergin, H.B., Hudson, Q.J., Stocsits, R.R., Allhoff, W., Stricker, S.H., Klement, R.M., Warczok, K.E., et al. (2012). Airn Transcriptional Overlap, But Not Its lncRNA Products, Induces Imprinted Igf2r Silencing. Sci. (New York, NY) *338*, 1469–1472.

Lee, J., Inoue, K., Ono, R., Ogonuki, N., Kohda, T., Kaneko-Ishino, T., Ogura, A., and Ishino, F. (2002). Erasing genomic imprinting memory in mouse clone embryos produced from day 11.5 primordial germ cells. Development.

Lei, H., Oh, S.P., Okano, M., Jüttermann, R., Goss, K.A., Jaenisch, R., and Li, E. (1996). De novo DNA cytosine methyltransferase activities in mouse embryonic stem cells. Development *122*, 3195–3205.

Leighton, P.A., Ingram, R.S., Eggenschwiler, J., Efstratiadis, A., and Tilghman, S.M. (1995). Disruption of imprinting caused by deletion of the H19 gene region in mice. Nature *375*, 34–39.

Leonhardt, H., Page, A.W., Weier, H.U., and Bestor, T.H. (1992). A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei. Cell *71*, 865–873.

Lewis, A., Mitsuya, K., Constancia, M., and Reik, W. (2004). Tandem repeat hypothesis in imprinting: deletion of a conserved direct repeat element upstream of H19 has no effect on imprinting in the Igf2-H19 region. Mol. Cell. Biol. *24*, 5650–5656.

Li, E., and Zhang, Y. (2014). DNA methylation in mammals. Cold Spring Harb. Perspect. Biol. *6*.

Li, X., Ito, M., Zhou, F., Youngson, N., Zuo, X., Leder, P., and Ferguson-Smith, A.C. (2008). A Maternal-Zygotic Effect Gene, Zfp57, Maintains Both Maternal and Paternal Imprints. Dev. Cell *15*, 547–557.

Lienert, F., Wirbelauer, C., Som, I., Dean, A., Mohn, F., and Schübeler, D. (2011). Identification of genetic elements that autonomously determine DNA methylation states. Nat. Genet. *43*, 1091–1097.

Lin, S., Ferguson-Smith, A.C., Schultz, R.M., and Bartolomei, M.S. (2011). Nonallelic transcriptional roles of CTCF and cohesins at imprinted loci. Mol. Cell. Biol. *31*, 3094–3104.

Lister, R., Mukamel, E. a, Nery, J.R., Urich, M., Puddifoot, C. a, Johnson, N.D., Lucero, J., Huang, Y., Dwork, A.J., Schultz, M.D., et al. (2013). Global epigenomic reconfiguration during mammalian brain development. Science *341*, 1237905.

Lucifero, D. (2004). Gene-specific timing and epigenetic memory in oocyte imprinting. Hum. Mol. Genet. *13*, 839–849.

Luedi, P.P., Hartemink, A.J., and Jirtle, R.L. (2005). Genome-wide prediction of imprinted murine genes. Genome Res. *15*, 875–884.

Luedi, P.P., Dietrich, F.S., Weidman, J.R., Bosko, J.M., Jirtle, R.L., and Hartemink, A.J. (2007). Computational and experimental identification of novel human imprinted genes. Genome Res. *17*, 1723–1730.

Mackay, D.J.G., Callaway, J.L.A., Marks, S.M., White, H.E., Acerini, C.L., Boonen, S.E., Dayanikli, P., Firth, H. V, Goodship, J.A., Haemers, A.P., et al. (2008). Hypomethylation of multiple imprinted loci in individuals with transient neonatal diabetes is associated with mutations in ZFP57. Nat. Genet. *40*, 949–951.

Maunakea, A.K., Nagarajan, R.P., Bilenky, M., Ballinger, T.J., D'Souza, C., Fouse, S.D., Johnson, B.E., Hong, C., Nielsen, C., Zhao, Y., et al. (2010). Conserved role of intragenic DNA methylation in regulating alternative promoters. Nature *466*, 253–257.

Meissner, A., Mikkelsen, T.S., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., Zhang, X., Bernstein, B.E., Nusbaum, C., Jaffe, D.B., et al. (2008). Genome-scale DNA methylation maps of pluripotent and differentiated cells. Nature *454*, 766–770.

Menheniott, T.R., Woodfine, K., Schulz, R., Wood, A.J., Monk, D., Giraud, A.S., Baldwin, H.S., Moore, G.E., and Oakey, R.J. (2008). Genomic imprinting of Dopa decarboxylase in heart and reciprocal allelic expression with neighboring Grb10. Mol. Cell. Biol. *28*, 386–396.

Merkenschlager, M., and Odom, D.T. (2013). CTCF and Cohesin: Linking Gene Regulatory Elements with Their Targets. Cell *152*, 1285–1297.

Mohn, F., Weber, M., Rebhan, M., Roloff, T.C., Richter, J., Stadler, M.B., Bibel, M., and Schübeler, D. (2008). Lineage-Specific Polycomb Targets and De Novo DNA Methylation Define Restriction and Potential of Neuronal Progenitors. Mol. Cell *30*, 755–766.

Myant, K., Termanis, A., Sundaram, A.Y.M., Boe, T., Li, C., Merusi, C., Burrage, J., De Las Heras, J.I., and Stancheva, I. (2011). LSH and G9a/GLP complex are required for developmentally programmed DNA methylation. Genome Res. *21*, 83–94.

Nagano, T., Mitchell, J.A., Sanz, L.A., Pauler, F.M., Ferguson-Smith, A.C., Feil, R., and Fraser, P. (2008). The Air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. Science *322*, 1717–1720.

Nakamura, T., Arai, Y., Umehara, H., Masuhara, M., Kimura, T., Taniguchi, H., Sekimoto, T., Ikawa, M., Yoneda, Y., Okabe, M., et al. (2006). PGC7/Stella protects against DNA demethylation in early embryogenesis. Nature *9*, 64–71.

Nakamura, T., Liu, Y.-J., Nakashima, H., Umehara, H., Inoue, K., Matoba, S., Tachibana, M., Ogura, A., Shinkai, Y., and Nakano, T. (2012). PGC7 binds histone H3K9me2 to protect against conversion of 5mC to 5hmC in early embryos. Nature.

Neumann, B., Kubicka, P., and Barlow, D.P. (1995). Characteristics of imprinted genes. Nat. Genet. *9*, 12–13.

Ono, R., Nakamura, K., Inoue, K., Naruse, M., Usami, T., Wakisaka-Saito, N., Hino, T., Suzuki-Migishima, R., Ogonuki, N., Miki, H., et al. (2005). Deletion of Peg10, an imprinted gene acquired from a retrotransposon, causes early embryonic lethality. Nat. Genet. *38*, 101–106.

Ooi, S.K.T., Qiu, C., Bernstein, E., Li, K., Jia, D., Yang, Z., Erdjument-Bromage, H., Tempst, P., Lin, S.-P., Allis, C.D., et al. (2007). DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. Nature *448*, 714–717.

Parelho, V., Hadjur, S., Spivakov, M., Leleu, M., Sauer, S., Gregson, H.C., Jarmuz, A., Canzonetta, C., Webster, Z., Nesterova, T., et al. (2008). Cohesins Functionally Associate with CTCF on Mammalian Chromosome Arms. Cell *132*, 422–433.

Paulsen, M., El-Maarri, O., Engemann, S., Strödicke, M., Franck, O., Davies, K., Reinhardt, R., Reik, W., and Walter, J. (2000). Sequence conservation and variability of imprinting in the Beckwith-Wiedemann syndrome gene cluster in human and mouse. Hum. Mol. Genet. *9*, 1829–1841.

Prickett, A.R., Barkas, N., McCole, R.B., Hughes, S., Amante, S.M., Schulz, R., and Oakey, R.J. (2013). Genome-wide and parental allele-specific analysis of CTCF and cohesin DNA binding in mouse brain reveals a tissue-specific binding pattern and an association with imprinted differentially methylated regions. Genome Res. *23*, 1624–1635.

Quenneville, S., Verde, G., Corsinotti, A., Kapopoulou, A., Jakobsson, J., Offner, S., Baglivo, I., Pedone, P. V, Grimaldi, G., Riccio, A., et al. (2011). In Embryonic Stem Cells, ZFP57/KAP1 Recognize a Methylated Hexanucleotide to Affect Chromatin and DNA Methylation of Imprinting Control Regions. Mol. Cell *44*, 361–372.

Ramsahoye, B.H., Biniszkiewicz, D., Lyko, F., Clark, V., Bird, A.P., and Jaenisch, R. (2000). Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a. Proc. Natl. Acad. Sci. U. S. A. *97*, 5237–5242.

Reed, M.R., Riggs, A.D., and Mann, J.R. (2001). Deletion of a direct repeat element has no effect on Igf2 and H19 imprinting. Mamm. Genome *12*, 873–876.

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

Rinn, J.L., and Chang, H.Y. (2012). Genome Regulation by Long Noncoding RNAs. Annu. Rev. Biochem.

Rinn, J.L., Kertesz, M., Wang, J.K., Squazzo, S.L., Xu, X., Brugmann, S.A., Goodnough, L.H., Helms, J.A., Farnham, P.J., Segal, E., et al. (2007). Functional Demarcation of Active and Silent Chromatin Domains in Human HOX Loci by Noncoding RNAs. Cell *129*, 1311–1323.

Di Ruscio, A., Ebralidze, A.K., Benoukraf, T., Amabile, G., Goff, L. a, Terragni, J., Figueroa, M.E., De Figueiredo Pontes, L.L., Alberich-Jorda, M., Zhang, P., et al. (2013). DNMT1-interacting RNAs block gene-specific DNA methylation. Nature *503*, 371–376.

Schilling, E., and Rehli, M. (2007). Global, comparative analysis of tissue-specific promoter CpG methylation. Genomics *90*, 314–323.

Seisenberger, S., Andrews, S., Krueger, F., Arand, J., Walter, J., Santos, F., Popp, C., Thienpont, B., Dean, W., and Reik, W. (2012). The Dynamics of Genome-wide DNA Methylation Reprogramming in Mouse Primordial Germ Cells. Mol. Cell *48*, 849–862.

Shen, L., Kondo, Y., Guo, Y., Zhang, J., Zhang, L., Ahmed, S., Shu, J., Chen, X., Waterland, R.A., and Issa, J.P.J. (2007). Genome-wide profiling of DNA methylation reveals a class of normally methylated CpG island promoters. PLoS Genet. *3*, 2023–2026.

Shen, L., Inoue, A., He, J., Liu, Y., Lu, F., and Zhang, Y. (2014). Tet3 and DNA Replication Mediate Demethylation of Both the Maternal and Paternal Genomes in Mouse Zygotes. Cell Stem Cell *15*, 459–470.

Shirane, K., Toh, H., Kobayashi, H., Miura, F., Chiba, H., Ito, T., Kono, T., and Sasaki, H. (2013). Mouse Oocyte Methylomes at Base Resolution Reveal Genome-Wide Accumulation of Non-CpG Methylation and Role of DNA Methyltransferases. PLoS Genet. *9*, e1003439.

Shiura, H., Nakamura, K., Hikichi, T., Hino, T., Oda, K., Suzuki-Migishima, R., Kohda, T., Kaneko-Ishino, T., and Ishino, F. (2009). Paternal deletion of Meg1/Grb10 DMR causes maternalization of the Meg1/ Grb10 cluster in mouse proximal Chromosome 11 leading to severe pre- and postnatal growth retardation. Hum. Mol. Genet. *18*, 1424–1438.

Smallwood, S.A., Tomizawa, S., Krueger, F., Ruf, N., Carli, N., Segonds-Pichon, A., Sato, S., Hata, K., Andrews, S.R., and Kelsey, G. (2011). Dynamic CpG island methylation landscape in oocytes and preimplantation embryos. Nat. Publ. Gr. *43*, 811–814.

Smith, Z.D., and Meissner, A. (2013a). DNA methylation: roles in mammalian development. Nat. Rev. Genet. *14*, 204–220.

Smith, Z.D., and Meissner, A. (2013b). DNA methylation: roles in mammalian development. Nat. Rev. Genet. *14*, 204–220.

Smith, E.Y., Futtner, C.R., Chamberlain, S.J., Johnstone, K.A., and Resnick, J.L. (2011). Transcription is required to establish maternal imprinting at the Prader-Willi syndrome and Angelman syndrome locus. PLoS Genet. *7*.

Stadler, M.B., Murr, R., Burger, L., Ivanek, R., Lienert, F., Schöler, A., van Nimwegen, E., Wirbelauer, C., Oakeley, E.J., Gaidatzis, D., et al. (2011). DNA-binding factors shape the mouse methylome at distal regulatory regions. Nature *480*, 490–495.

Stelzer, Y., Sagi, I., Yanuka, O., Eiges, R., and Benvenisty, N. (2014). The noncoding RNA IPW regulates the imprinted DLK1-DIO3 locus in an induced pluripotent stem cell model of Prader-Willi syndrome. Nat. Publ. Gr. *46*, 551–557.

Sunahara, S., Nakamura, K., Nakao, K., Gondo, Y., Nagata, Y., and Katsuki, M. (2000). The oocyte-specific methylated region of the U2afbp-rs/U2af1-rs1 gene is dispensable for its imprinted methylation. Biochem. Biophys. Res. Commun. *268*, 590–595.

Tani, H., Mizutani, R., Salam, K.A., Tano, K., Ijiri, K., Wakamatsu, A., Isogai, T., Suzuki, Y., and Akimitsu, N. (2012). Genome-wide determination of RNA stability reveals hundreds of short-lived noncoding transcripts in mammals. Genome Res. *22*, 947–956.

Thomson, J.P., Skene, P.J., Selfridge, J., Clouaire, T., Guy, J., Webb, S., Kerr, A.R.W., Deaton, A., Andrews, R., James, K.D., et al. (2010). CpG islands influence chromatin structure via the CpG-binding protein Cfp1. Nature *464*, 1082–1086.

Thorvaldsen, J.L., Duran, K.L., and Bartolomei, M.S. (1998). Deletion of the H19 differentially methylated domain results in loss of imprinted expression of H19 and Igf2. Genes Dev. *12*, 3693–3702.

Tsai, M.-C., Manor, O., Wan, Y., Mosammaparast, N., Wang, J.K., Lan, F., Shi, Y., Segal, E., and Chang, H.Y. (2010). Long noncoding RNA as modular scaffold of histone modification complexes. Sci. (New York, NY) *329*, 689–693.

Varrault, A., Gueydan, C., Delalbre, A., Bellmann, A., Houssami, S., Aknin, C., Severac, D., Chotard, L., Kahli, M., Le Digarcher, A., et al. (2006). Zac1 Regulates an Imprinted Gene Network Critically Involved in the Control of Embryonic Growth. Dev. Cell *11*, 711–722.

Wang, L., Zhang, J., Duan, J., Gao, X., Zhu, W., Lu, X., Yang, L., Zhang, J., Li, G., Ci, W., et al. (2014). Programming and Inheritance of Parental DNA Methylomes in Mammals. Cell *157*, 979–991.

Wang, X., Arai, S., Song, X., Reichart, D., Du, K., Pascual, G., Tempst, P., Rosenfeld, M.G., Glass, C.K., and Kurokawa, R. (2008). Induced ncRNAs allosterically modify RNA-binding proteins in cis to inhibit transcription. Nature *454*, 126–130.

Watanabe, T., Tomizawa, S. i, Mitsuya, K., Totoki, Y., Yamamoto, Y., Kuramochi-Miyagawa, S., Iida, N., Hoki, Y., Murphy, P.J., Toyoda, A., et al. (2011). Role for piRNAs and Noncoding RNA in de Novo DNA Methylation of the Imprinted Mouse Rasgrf1 Locus. Sci. (New York, NY) *332*, 848–852.

Wendt, K.S., Yoshida, K., Itoh, T., Bando, M., Koch, B., Schirghuber, E., Tsutsumi, S., Nagae, G., Ishihara, K., Mishiro, T., et al. (2008). Cohesin mediates transcriptional insulation by CCCTC-binding factor. Nature *451*, 796–801.

Yamaguchi, S., Hong, K., Liu, R., Shen, L., Inoue, A., Diep, D., Zhang, K., and Zhang, Y. (2012). Tet1 controls meiosis by regulating meiotic gene expression. Nature *492*, 443–447.

Yamaguchi, S., Shen, L., Liu, Y., Sendler, D., and Zhang, Y. (2013). Role of Tet1 in erasure of genomic imprinting. Nature *504*, 460–464.

Yoon, B., Herman, H., Hu, B., Park, Y.J., Lindroth, A., Bell, A., West, A.G., Chang, Y., Stablewski, A., Piel, J.C., et al. (2005). Rasgrf1 Imprinting Is Regulated by a CTCF-Dependent Methylation-Sensitive Enhancer Blocker. Mol. Cell. Biol. *25*, 11184–11190.

Yoon, B.J., Herman, H., Sikora, A., Smith, L.T., Plass, C., and Soloway, P.D. (2002). Regulation of DNA methylation of Rasgrf1. Nat Genet *30*, 92–96.

Chapter Three

Transcriptome-wide Analysis of Genomic Imprinting in the Cerebellum

The content of this chapter is part of a manuscript in preparation Perez, J.D., N.D. Rubinstein, D.E. Fernandez, L.A. Needleman, S.W. Santoro, J.J Choi, J.S. Liu, C. Dulac, 2015. New Insights Into Parent-of Origin Expression in the Adult and Developing Brain.

Contributions: Perez, J.D. performed the RNA-seq experiments, pyrosequencing validations, helped calibrate the statistical model to analyze the data presented on this chapter, and interpreted the data. N.D. Rubinstein processed the RNA-seq data, interpreted the data and together with D.E. Fernandez developed the statistical models used in the analysis of the RNA-seq data presented on this chapter. L.A. Needleman, and S.W. Santoro conducted experiments that are not part of this thesis but are part of the manuscript. J.J. Choi contributed to experiments that will be presented in chapter four. J.S. Liu provided guidance and supervision for the statistical models used in the analysis of the RNA-seq data presented on this chapter. C. Dulac, conceived the project, interpreted the data, and provided guidance and supervision.

INTRODUCTION

In contrast to most mammalian genes, which are equally expressed from both alleles, imprinted genes are preferentially expressed from one of the two alleles, in a parent-of-origin-specific manner. Consequently, the functions of genes undergoing imprinting regulation are mainly contributed by either the maternal or paternal genomes, by contrast to the functions of biallelic genes, which are equally contributed by both parental genomes. The regulation of genomic imprinting relies on epigenetic modifications or "imprints" established in the parental germ lines and stably maintained post-fertilization (Bartolomei and Ferguson-Smith, 2011). These imprints may be further modified, or novel imprints may be added, according to tissue types and developmental stages. Hence the preferential parental allelic expression of imprinted genes can be relieved, strengthened, or even switched (Kulinski et al., 2013). Moreover, imprints have been shown to often regulate the imprinted expression of several genes co-localized in genomic regions termed imprinted clusters (Reik and Walter, 2001). Importantly, many imprinted genes have been demonstrated to function in key biological pathways that regulate embryonic development, metabolism, and behavior (Bartolomei and Ferguson-Smith, 2011). Accordingly, impairment of the function or disruption of the epigenetic control of imprinted genes may result in abnormal development, physiology and behavior, and in humans is known to be manifested in a large range of clinical conditions (Peters, 2014).

To gain insights into the unique roles conducted by each parental genome it is necessary to identify all imprinted expressed loci. This in turn will enable a comprehensive investigation of the biological pathways and physiological processes affected by genomic imprinting. It will also help identify regions where mutations or polymorphism could lead to parent of origin specific phenotypes. Furthermore, it will provide additional opportunities to understand the molecular

mechanisms and processes regulating parent of origin expression. Finally, because variations in the imprinted gene repertoire have been observed between species, a definitive list of imprinted genes in multiple mammals will allow comparative studies to determine functions that have been commonly or distinctively targeted throughout mammalian evolution.

Before the first imprinted genes were identified, a series of studies revealed genomic regions under imprinting regulation (reviewed in Cattanach and Jones, 1994). Based on this preliminary map it was speculated that if randomly distributed across the genome, the total number of imprinted genes should not exceed 100 (Solter, 1988). Hayashizaki et al later confirmed this initial estimate after a screen for differentially methylated regions (DMRs) which led to the identification of ~100 DMRs in the mouse genome (Hayashizaki et al., 1994). This calculation assumed a 1 to 1 ratio between imprinted genes and DMRs. Even though the assumptions behind both calculations were quickly proven to be largely incorrect, the "~100 imprinted genes" estimate has been frequently cited since then (Kelsey and Bartolomei, 2012; Morison et al., 2005) Additionally, a bioinformatics study, now largely discredited, attempted to identify genes under imprinting regulation by training statistical models to distinguish DNA features characteristic of known imprinted regions (Luedi et al., 2005). This approach predicted over 600 candidates, among which only 1 of 600 were later confirmed as genuinely imprinted (Ruf et al., 2007).

Most of the imprinted genes known today have been identified either by virtue of their co-localization with know imprinted clusters, or by unbiased screens. From mid 1990s to the early 2000s screens tailored for the detection of genes that were differentially expressed between parthenogenetic (containing only maternally derived chromosomes) or androgenetic (containing only paternally derived chromosomes) embryos led to the discovery of 10 imprinted genes

(Hagiwara et al., 1997; Hatada et al., 2001; Kagitani et al., 1997; Kaneko-Ishino et al., 1995; Miyoshi et al., 1998; Piras et al., 2000; Sandell et al., 2003; Schmidt et al., 2000). These methods were labor intensive and provided very low throughput identification. The advent of cDNA microarrays provided an opportunity to increase the throughput and accelerate the completion of imprinted gene screens. A total of 15 additional genes were subsequently identified using cDNA microarrays to detect differentially expressed genes in parthenogenetic or androgenetic embyros (Choi et al., 2005; Davies et al., 2005; Kobayashi et al., 2009; Kuzmin et al., 2008; Menheniott et al., 2008; Mizuno et al., 2002; Nikaido et al., 2003; Raefski and O'Neill, 2005; Ruf et al., 2006; Schulz et al., 2006).

Presently, New Generation RNA Sequencing (RNAseq) strategies represent the state of the art method to screen for imprinted genes. RNA-Seq provides a highly sensitive digital quantification platform that is able to determine gene expression of a range spanning five orders of magnitude (Mortazavi et al., 2008). Contrary to hybridization technologies, RNA-Seq can potentially detect any transcripts present in a sample instead of an a priori defined set as in microarray platforms. More importantly for imprinted gene screens, the single base resolution of RNA-Seq allows the accurate identification of SNPs in tissues generated by reciprocal crosses. Since its introduction, multiple studies have use RNA-seq to investigate genomic imprinting in various tissues producing a large number of imprinted candidates, 20 of which were validated with independent techniques (Babak et al., 2008a; DeVeale et al., 2012a; Gregg et al., 2010a; Wang et al., 2008a, 2011a).

In 2008, Babak and colleagues published the first screen of imprinted genes using RNA-Seq (Babak et al., 2008a). The transcriptome of eight E9.5 embryos derived from reciprocal crosses between C57Bl/6J and Cast/EiJ mouse strains (4 embryos for each cross), were

independently sequenced using the Illumina platform. Allele-specific expression was quantified first at each SNP position and then SNPs within the same gene were combined to estimate the allele-specific expression of the gene. A binomial distribution was used to calculate the probability to detect differences in expression divergent from the proportions expected by chance. Agreement between both directions of the reciprocal cross was required for statistical significance, but corrections for multiple comparisons were not reported. Novel candidates where corroborated by restriction fragment length polymorphism analysis or Sanger sequencing. The coverage of this study was very shallow (~9 fold transcriptome coverage), significantly weakening its conclusions. Only 32 imprinted genes were detected by this study, among which 26 were previously identified and six were novel. (Babak et al., 2008a).

Shortly after, a second study also used Illumina sequencing to independently profile postnatal day 2 whole brain transcriptomes of postnatal day 2 F1 hybrids derived from reciprocal crosses between PWD/PhJ and AKR/J mouse strains. After estimating the expression of the two alleles at each SNP position, a Storer-Kim test, which is an exact significance test also based on a binomial distribution, was conducted independently for each SNP. Because of multiple comparisons a q-value was also calculated for each SNP. This approach identified a total of 26 candidate imprinted genes, 16 of which where previously known as imprinted, 1 new gene for which imprinting was corroborated using pyrosequencing, and 9 false positives (Wang et al., 2008a).

Using the Allen Brain Atlas database Chris Gregg and collaborators in our group systematically scanned the expression of imprinted genes across the brain, and observed a high representation of imprinted genes in the brain, as well as the existence of specific repertoires of imprinted genes in distinct brain regions (Gregg et al., 2010a). This motivated a genome-wide

investigation of allele-specific expression in some of the enriched brain regions using RNA-seq. Reciprocal crosses between C57Bl/6J and Cast/EiJ were use to assess imprinting in three different tissues: E15 embryonic brain, adult medial Prefrontal Cortex (mPFC) and adult Preoptic Area (POA). Each of these tissues was represented by 4 replicates. After quantifying allelic expression at each SNP, the chance probability of a real parental effect was calculated using a chi-square test. Using a 95% confidence cutoff significance SNPs representing a total of 1,308 imprinted candidates were identified across all three tissues that were common to both sexes. These included 129 known imprinted loci and a small subset of novel genes that were corroborated by Sequenom. Substantial variation in the imprinted status of several genes according to brain region or developmental stage was observed for several genes (Gregg et al., 2010a). In addition, a large fraction of candidate genes exhibited a complex pattern of imprinting expression were a fraction of the SNPs within a gene showed parentally biased expression, which was interpreted as potential isoform-specific forms of imprinting (Gregg et al., 2010a).

Two years later, DeVeale and colleagues published a criticism of the findings by Gregg et_al. Analysis of the relationship of observed parental biases in independent SNPs within the same gene showed substantially more disagreements than would be predicted by simulations assuming random sampling of RNA-Seq, suggesting previously unappreciated technical variation. Finally, a mock cross (analyses that treat one direction of the reciprocal cross as if it was the opposite direction thus nullifying any real imprinting) reported nearly as many candidates as the actual analysis. To improve the true positive: false positive ratio the authors reported two modifications. First, they suggested that summing allele specific quantitations of independent SNPs within a gene could significantly improve the detection of known imprinted genes. Second, use of the false discovery rate estimated by the mock cross helped define a cutoff

for candidate imprinted genes. This, however, led to only ~50% confirmation by pyrosequencing (DeVeale et al., 2012a). Thus, both the use of strong statistical tools as well as thorough independent validation seem imperative (Kelsey and Bartolomei, 2012).

Traditionally, based in large part on data obtained from the first set of known imprinted genes, genomic imprinting has been thought to involve a complete silencing of one of the two parental alleles (Bartolomei and Ferguson-Smith, 2011). For example, the Snrpn gene of the PWS cluster is exclusively expressed from the paternal allele in all tissues, whereas the non-coding RNA H19 is exclusively expressed from the maternal allele during embryogenesis. While examples of imprinted genes showing parental expression biases lower than all-or-none have been reported (Ono et al., 2003; Wang et al., 2011b), the extent of such cases is unknown. On the one hand, biases may be due to the technical inability to detect imprinting at the single-cell type resolution in a heterogeneous tissue, and hence reflects the average of monoallelic expression in certain cell types and biallelic expression in others. On the other hand, it is possible that alternative imprinting mechanisms have evolved to establish parentally biased allelic expression in a continuous range from biallelic to monoallelic expression. These possibilities are not mutually exclusive and regardless of the explanation, genes that exhibit mild to strong parental biases are imprinted.

In order to detect genes with such parental effects, a powerful genome-wide approach with independent validation should be applied. We identified several avenues to improve on previous RNA-seq based searches of imprinted genes. First, RNA expression may be highly variable across individuals. This variability necessitates the use of a large number of biological samples in order to achieve sufficient power for the accurate detection of various ranges of parental bias. Second, accurate estimation of expression levels from RNA sequencing data

requires estimating the expression levels of transcripts rather than DNA loci. Third, inference of genomic imprinting across all samples should explicitly account for all sources of variation including the effects of all experimental factors (e.g., sex and age) among which are allelic imbalanced expression due to cis eQTLs between the mouse strains used. Finally, because of the high complexity of RNA-seq datasets, results should be subjected to independent experimental validations to avoid false positives. These considerations were not adequately addressed by previous studies (Babak et al., 2008b; DeVeale et al., 2012b; Gregg et al., 2010b; Wang et al., 2008b) likely contributing to their inconsistent findings.

To better characterize genomic imprinting in the brain we chose the cerebellum as our target brain region. Several natural features of the cerebellum make it an attractive entry point to study this phenomenon. First, due to its large size and posterior-most location, the cerebellum allows reproducible pure dissections and enough RNA material to profile independent cerebellums from multiple individuals. Second, contrary to most brain regions, cerebellar development occurs mostly postnatally. This allows access to neurodevelopmental processes such as cell proliferation, migration, and circuit establishment that in other brain regions can be highly inaccessible. As such processes are believed to be under genomic imprinting regulation, investigating the cerebellum during developmental milestones could capture imprinting at one of its most active states. Thirdly, the cerebellum has a well-characterized cell circuit with genetically defined cell types. Moreover, these cell types have a well-established role in regulation of motor behavior. Thus, the cerebellum provides the ability to target manipulations of imprinted genes of interest to specific cell types and a way to screen for potential phenotypes. Finally, the cerebellum is increasingly implicated in Autism Spectrum Disorders (ASD) (Wang et al., 2014). Cellular and anatomical abnormalities in the cerebellum are among the most

common observations in ASD patients (Amaral et al., 2008) and in mouse a Purkinje cell-specific knockout of the ASD associated TSC1 homolog, recapitulates most of the social deficits of the spectrum (Tsai et al., 2012). Considering the strong association between genomic imprinting and social behavior, the cerebellum may harbor imprinted genes implicated in normal and abnormal social conducts.

In this chapter I present the results of a genome-wide screen for imprinted genes in the cerebellum. The experimental design is comprised of 48 individual cerebellums capturing two different developmental stages and representing both sexes. A statistical model that infers genomic imprinting based on transcript-level estimates of allele-specific expression, which explicitly accounts for all factors in our experimental design was employed to analyze this data. This approach reports 124 genes under imprinting regulation in the cerebellum, 50 of which had not been previously reported. Using pyrosequencing, imprinting in 41 of these genes was confirmed. Most novel genes exhibit subtle parental expression biases rather than all-to-none silencing. A substantial portion of novel genes is located in the vicinity of known imprinted cluster. Finally, an overrepresentation of genes involved in programed cell death was observed in the dataset suggesting genomic imprinting particularly targets this pathway.

RESULTS

Transcriptome-wide profile of allele-specific expression in the cerebellum

To study genomic imprinting in the brain we sequenced RNA from 48 individual cerebellums derived from F1 hybrids of Cast/EiJ and C57Bl/6J reciprocal crosses (**Figure 3.1**). Half of these cerebellums were collected at postnatal day 8 (P8), a period in which granule cells are highly multiplied and are migrating to the inner granule layer. The other half of cerebellums were collected from adult animals (P60). Each age group had equal numbers of males and females. Therefore, our experimental design includes six replicates for each cross, sex, and age combination, which allows inferring genomic imprinting across all individuals and the effects that age and sex have on it, while controlling for cross effects.

RNA-seq libraries generated from tissues with high cellular heterogeneity and high transcriptome complexity, such as brain, are highly likely to contain multiple transcripts (isoforms) with different abundances. For this reason it is important to estimate the expression level of each transcript. Accurate transcript-level expression estimates are also necessary even if one only wants to obtain estimates of expression at the gene level, as the expression level of a gene is merely the sum of expression levels of all its isoforms. However, since many reads do not map uniquely to a single transcript (e.g., reads originating from constitutive exons) or even to single genomic locations, transcript expression levels (and hence gene expression levels) can only be estimated with some degree of certainty (i.e., measurement error) (Jiang and Wong, 2009; Trapnell et al., 2013; Turro et al., 2011). This uncertainty is even more pronounced when estimating allele-specific expression levels due to the high sequence similarity between the two alleles. In addition, mapping RNA-seq data to a haploid reference genome can bias the alignments towards one of the alleles and may therefore lead to inaccurate estimates of allele-

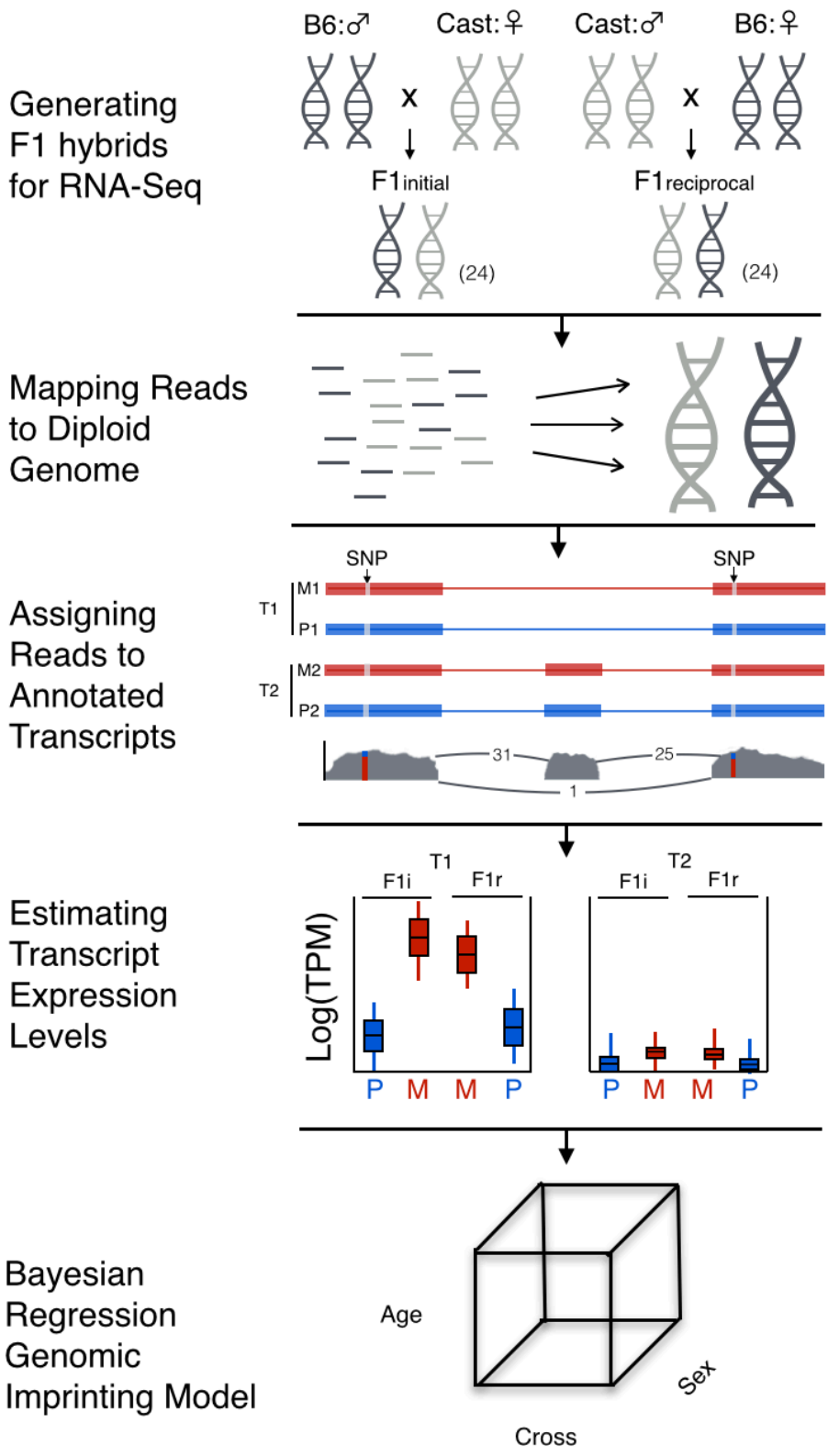

Figure 3.1: Experimental setup and major steps of the data analysis

specific expression levels (Vijaya Satya et al., 2012). Finally, in order to infer genomic imprinting, the expression-level uncertainties at each sample need to be combined with the across-samples-variability (biological and technical) and in addition, the effects of all factors in the experimental design (e.g., sex, age, and in the casewere collected from adult animals (P60). Each age group had equal numbers of males and females. Therefore, our experimental design includes six replicates for each cross, sex, and age combination, which allows inferring genomic imprinting across all individuals and the effects that age and sex have on it, while controlling for cross effects.

RNA-seq libraries generated from tissues with high cellular heterogeneity and high transcriptome complexity, such as brain, are highly likely to contain multiple transcripts (isoforms) with different abundances. For this reason it is important to estimate the expression level of each transcript. Accurate transcript-level expression estimates are also necessary even if one only wants to obtain estimates of expression at the gene level, as the expression level of a gene is merely the sum of expression levels of all its isoforms. However, since many reads do not map uniquely to a single transcript (e.g., reads originating from constitutive exons) or even to single genomic locations, transcript expression levels (and hence gene expression levels) can only be estimated with some degree of certainty (i.e., measurement error) (Jiang and Wong, 2009; Trapnell et al., 2013; Turro et al., 2011). This uncertainty is even more pronounced when estimating allele-specific expression levels due to the high sequence similarity between the two alleles. In addition, mapping RNA-seq data to a haploid reference genome can bias the alignments towards one of the alleles and may therefore lead to inaccurate estimates of allele-specific expression levels (Vijaya Satya et al., 2012). Finally, in order to infer genomic imprinting, the expression-level uncertainties at each sample need to be combined with the

across-samples-variability (biological and technical) and in addition, the effects of all factors in the experimental design (e.g., sex, age, and in the case of mouse hybrids the cross of each sequenced subject) must also be taken into account.

To address these challenges for the purpose of inferring transcriptome-wide genomic imprinting, we applied the following steps in the analysis of our RNA-seq data (**Figure 3.1** and described in more details in the Methods). We first generated C57, CAST diploid genomes and transcriptomes by incorporating C57 and CAST single nucleotide and short insertion and deletion polymorphisms (SNPs and indels, obtained from the Mouse Genome Project: ftp://ftp-mouse.sanger.ac.uk/REL-1303-SNPs_Indels-GRCm38) into the Mus musculus GRCm38 reference genome sequence. We then mapped each of our C57×CAST hybrid RNA-seq libraries to this diploid genome using STAR RNA-seq aligner (Dobin et al., 2013). For each mapped RNA-seq library, we then estimated the expression levels with their respective uncertainties of each transcript from each allele in the C57, CAST diploid transcriptome using MMSEQ (Turro et al., 2011). Finally, we developed a statistical model testing for genomic imprinting at each transcript as well as the effect that any factor (namely, cross, sex, and age in this study) has on it across all samples in our experimental design. Specifically, our model is a Bayesian variable selection regression model, extending Chipman et al. (1997) by accounting for the measurement error in the response. In our settings, we define the response as the difference between the paternal and the maternal expression levels for a given transcript in each sample in the experiment, i.e., the parental bias. The model therefore computes a posterior distribution of the magnitude of the parental bias and of the effects that all factors have on it, along with their posterior probabilities (PPs) of being significantly different from zero.

We fitted our model to 47,676 heterozygous transcripts from 26,651 genes, which were expressed at sufficient levels (Methods). The distribution of the PPs of the parental effect shows that most transcripts are not inferred to be under imprinting regulation (**Figure 3.2.A**). It also clearly shows a group of transcripts with PP > 0.95 of the parental effect, which we set as our cutoff for calling an effect significant (Methods), i.e., calling the parental bias imprinting. Notably, the distribution of the PPs of the cross effect (**Figure 3.2.A**) clearly shows how widely prevalent it is in these hybrids indicating that both crosses are necessary for inferring genomic imprinting or monoallelic expression (as opposed to Deng et al., 2014). In addition, none of the transcripts imprinted are reported to have a sex-effect with PP above the 0.95 cutoff, indicating that genomic imprinting is sex invariant in the mouse cerebellum. Conversely, a small group of imprinted transcripts are found to have age-effect PPs above the 0.95 cutoff indicating age-regulated imprinting (describe in chapter four).

Among the 124 genes inferred to be imprinted (represented by 169 transcripts), 74 were previously reported as such (either identified before 2010 or independently validated by Gregg et al., 2010 or DeVeale et al., 2012). The remaining 50 genes had not been described as imprinted before (**Figure 3.2.B**). To independently evaluate imprinting in all these candidates we used pyrosequencing, a real-time sequence-by-synthesis approach relying on light emissions after nucleotide incorporation (Wang and Elbein, 2007). As positive and negative controls, we respectively tested 11 known imprinted genes and 11 randomly selected genes with no significant parental effects according to our RNA-seq analysis. We tested an average of two SNPs per gene (Methods) across 12 P60 and/or 12 P8 individual cerebellums, different from those used in our RNA-seq experiment. We estimated parental effects in these data using the model we developed for the same task for RNA-seq data, where there is no measurement

uncertainty. This confirmed significant parental effects for 41 of the candidate novel imprinted genes and the expected significant and non-significant parental effects for all positive and negative controls, respectively (**Figure 3.3.A**). Among the novel and known imprinted genes we observed a slightly higher number of genes with a paternal effect (**Figure 3.2.C**). Interestingly, five genes preferentially express different isoforms from the maternal and paternal alleles (**Figure 3.2.C** and see below).

The distribution of the parental biases of the novel and known imprinted genes span the entire range of slightly above 50:50 % to 100:0 % and shows a bimodal shape at the two extremes (**Figure 3.3.B**). Notably, the mild parental bias mode is enriched with novel imprinted genes (**Figure 3.3.B**) and these biases show strong correspondence between the RNA-seq and the pyrosequencing analyses (**Figure 3.3.A** and see specific examples in **Figure 3.4**), indicating they are accurately quantified. This emphasizes that genomic imprinting cannot be assumed to be manifested as an all-or-none parental bias in RNA measurements obtained from tissue samples. Accordingly, using the term monoallelic expression to describe genomic imprinting in such context may be inadequate. In summary, considering known imprinted genes detected in this analysis as true positives, these results indicate that our approach has a true positive rate of ~93% and increases the number of mouse imprinted genes by ~26%.

We also considered the hypothesis that several true imprinted genes may not meet our parental bias PP cutoff (i.e., false negatives). We therefore subjectively selected for pyrosequencing evaluation 18 genes with PPs of their parental effect below 0.95, and successfully confirmed a significant parental effect in 10 of these 18 genes (not shown). As a positive control for this step, we chose the Casd1 gene, which has been reported to be imprinted in other tissues (Ono et al., 2003) and obtained a parental effect PP of 0.89 (not shown).

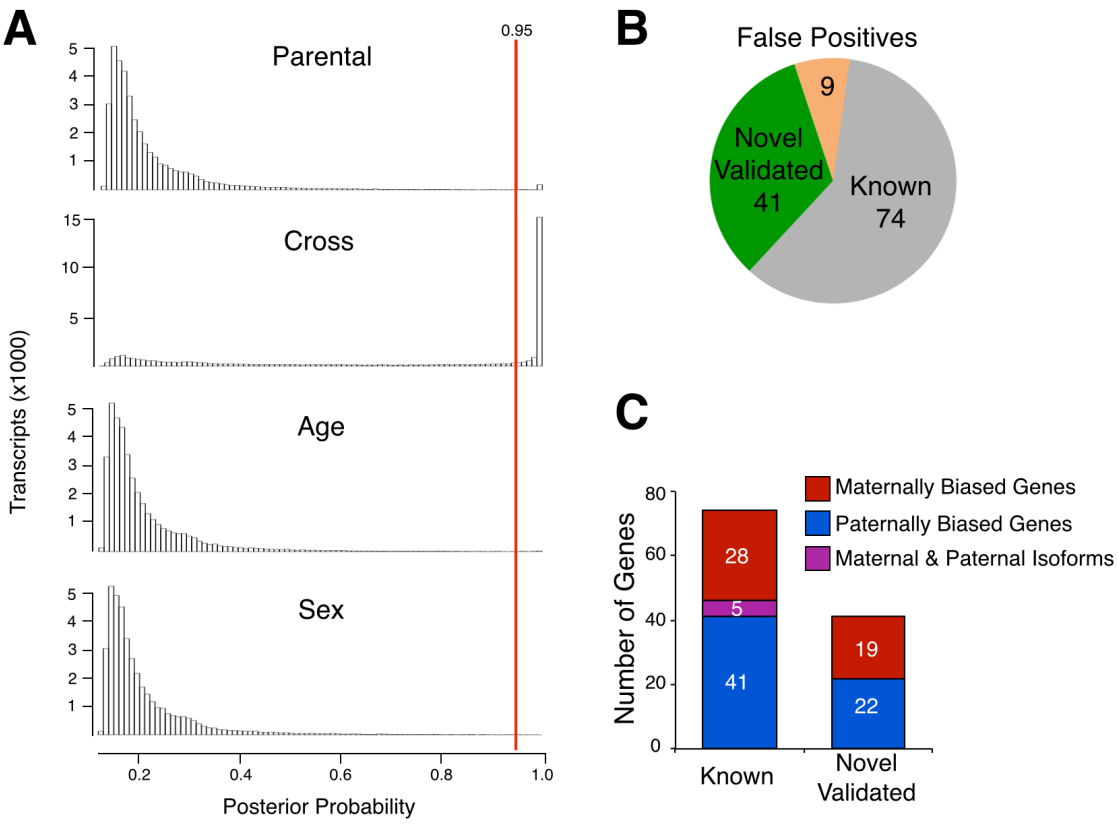

Figure 3.2: Summary of allele specific expression in the cerebellum

(A) Distributions of posterior probabilities of all four factors analyzed. Significance cutoff were selected at 0.95. **(B)** Breakdown of previously identified and novel genes represented by at least one transcript above the 0.95 cutoff of parental effects. The proportion of novel genes independently validated by pyrosequencing is also shown. **(C)** Breakdown parentally biased expression.

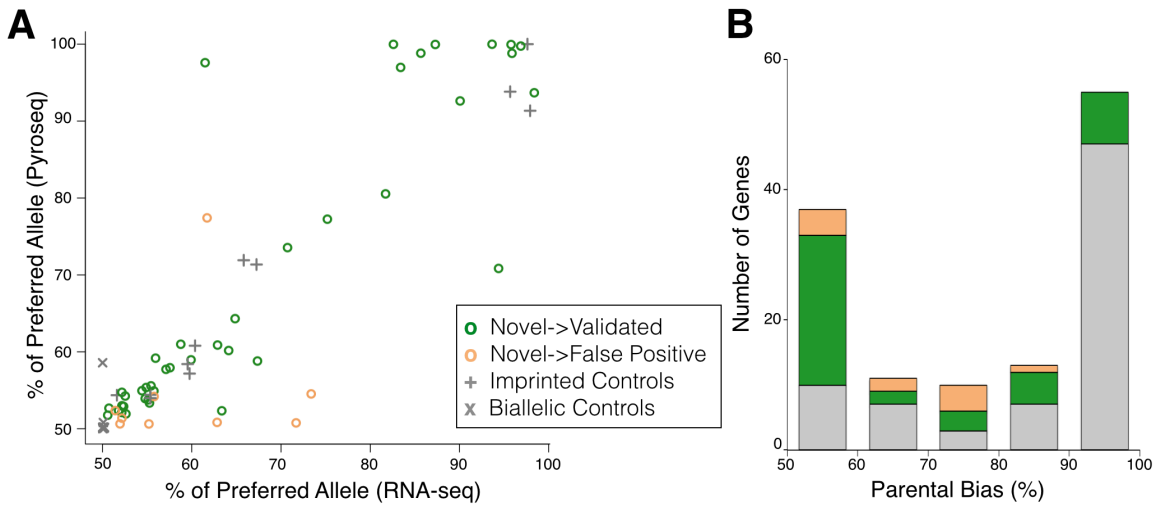

Figure 3.3: Differential expression between parental alleles occur in spectrum ranging from all-to-none to minor parental biases

(A) Comparison of the percentage of total expression accounted by the preferentially expressed allele as estimated in RNA-seq vs. Pyrosequencing. **(B)** Distribution of parentally biased expression in the Cb.

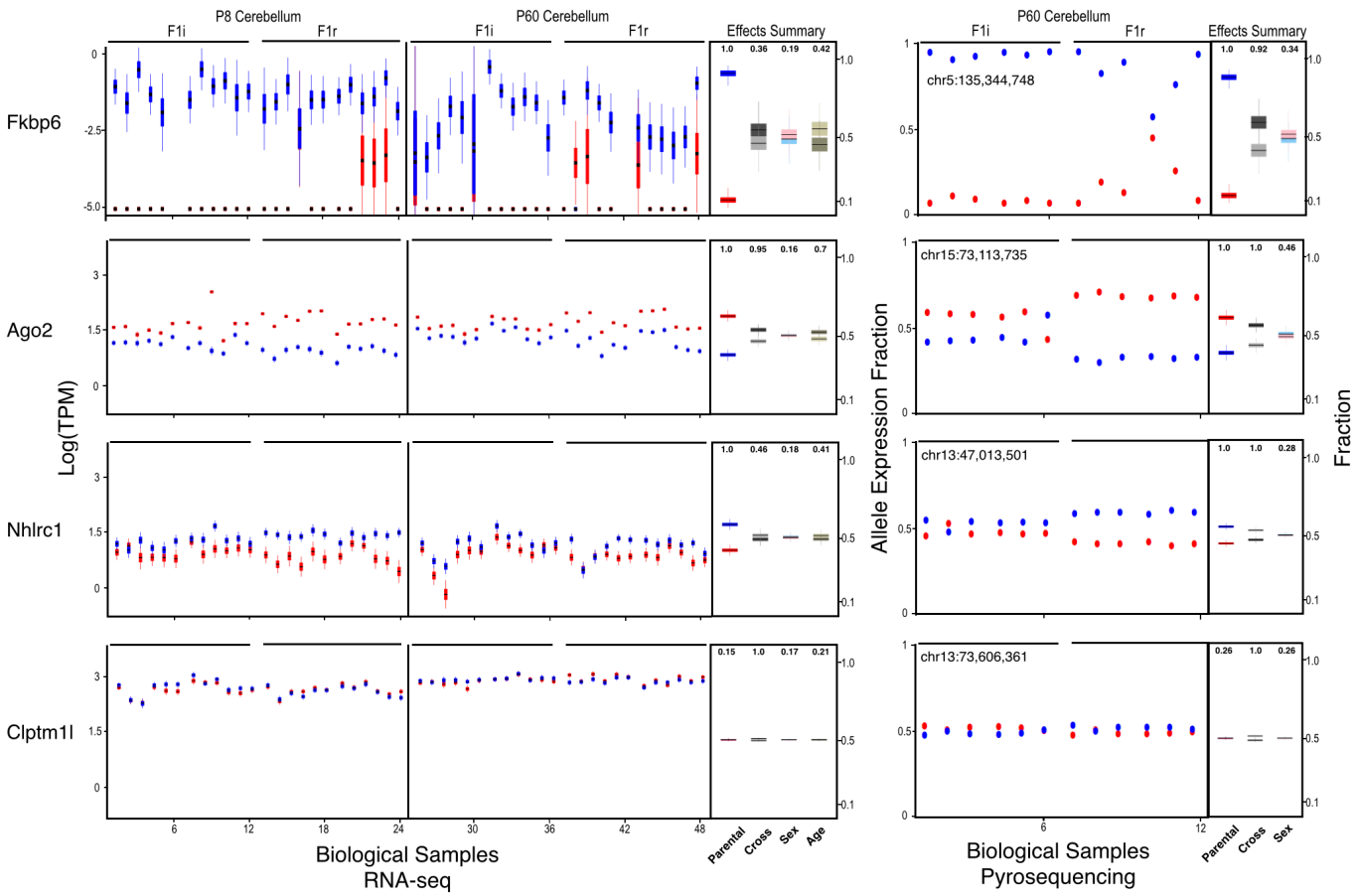

Figure 3.4: Examples of genes imprinted in the cerebellum and a biallelic control

Fkbp6, Ago2 and Nhlrc1 show parentally biased expression in the Cb as observed with RNA-seq and confirmed with pyrosequencing. Clptm1l shows biallelic expression in the Cb in both RNA-seq and pyrosequencing experiments. For each replicate red indicates maternal expression while blue indicates paternal expression. Y-axis of RNA-seq result is in the log of transcript per million units. Effects summary, shows the average effect on all four factors. For the cross effect gray represents F1rs and black represents F1is. For sex pink represent female and cyan represents male. For age bright olive green represents P8 and dark olive green represents P60.

Isoform-Specific Imprinting

Our approach estimates the parental bias at the transcript level and indeed, in most cases, multiple isoforms of the same gene show a consistent imprinting pattern (e.g., all the expressed isoforms of the Rian gene, **Figure 3.5**). For such cases, a gene-level analysis would most likely also infer a significant parental bias. This however, is not guaranteed for genes with parental effects specific to certain isoforms. Indeed, isoform-specific imprinting is known to occur in genes harboring a paternally expressed gene within an intron, resulting in the generation of different isoforms from the maternal and paternal alleles (Gregg et al., 2010a; McCole and Oakey, 2008a). Although it is unclear exactly how such regulation arises, transcriptional interference by the inner paternally expressed gene likely plays an important role (McCole and Oakey, 2008b). Our analysis detected almost all of the previously reported cases of genes with isoforms subjected to such regulation (**Figure 3.5**) and further detected additional imprinted transcripts in some of these loci. This includes either or both of two short transcripts of the Herc3 gene (indistinguishable by our sequence data) from a promoter upstream to the large 25 exons-long transcript (Gencode transcript IDs: ENSMUST00000141600.1 and ENSMUST00000122981.1), which are preferentially expressed by the maternal allele (**Figure 3.5**). Other known cases of isoform-specific imprinting are due to differential methylation of alternative promoters, as in the case of the Gnas (Peters and Williamson, 2007) and Grb10 (Arnaud et al., 2003) genes. Surprisingly, at the paternally expressed Mest gene we detect a novel maternally expressed short isoform (Gencode transcript ID: ENSMUST00000149496.1) whose transcription starts at exon 9 (**Figure 3.5**). This transcript is presumably a non-coding

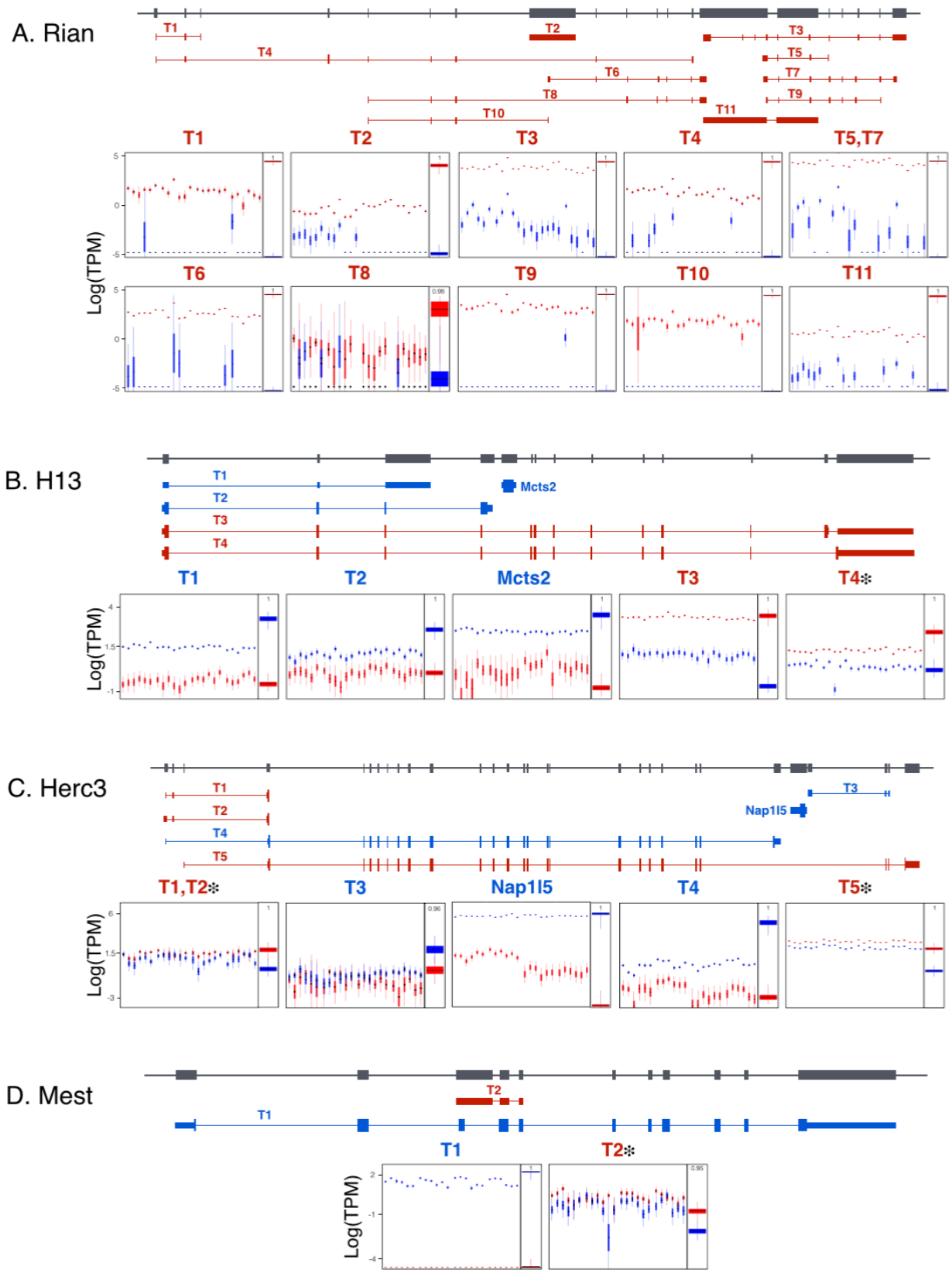

Figure 3.5: Assessment of imprinting at the isoform level

RNA since no ORF could be found in it. These results therefore emphasize the importance of analyzing allelic-biased expression at the transcript level.

Genomic Locations of Imprinted Genes

As described above our method detects imprinting where the parental expression bias ranges from weak expression biases to absolute silencing of an allele (**Figure 3.3.B & 3.6.A**). Interestingly, a substantial proportion of these novel imprinted genes localize to the vicinity of imprinted genes exhibiting stronger parental biases, thereby expanding imprinted clusters (**Figure 3.6.B**). For example, Ankrd34c and Ctsh, two genes that exhibit subtle paternal biases (the parental bias of Ankrd34c is significant only according to the pyrosequencing data), are located up- and down-stream to Rasgrf1, a gene exclusively expressed from the paternal allele (**Figure 3.6.C**). If this phenomenon is common we would expect to observe a strong and consistent decay of the parental bias as a function of the distance from strongly biased genes. To test this hypothesis we defined genes for which the parental bias is at least 85% to 15% as imprinting cluster centers and assigned any other imprinted genes with an intergenic distance of up to 1Mbp from them as members of their clusters (Methods). This analysis indeed confirmed a statistically significant negative effect of the distance from the cluster center on the magnitude of the parental bias (P-value = 0.01, Methods), supporting our hypothesis (**Figure 3.6.D**). It is possible that genes with weak parental biases are affected by the silencing taking place at the location of their neighboring strongly biased genes, perhaps through the chromatin environment, and that this effect has a weak fitness cost. However, it is also possible that either these weak

biases are selectively advantageous or that such genes are strongly imprinted in a small fraction of cerebellum cell types or in other tissues or time points.

The clustered organization of imprinted genes is regarded as one of their hallmark characteristics and is thought to reflect common regional control (Reik and Walter, 2001). If imprinting regulation of genes in which we detect weak parental biases and which expand imprinted clusters is functionally important we would expect natural selection to operate against disruption of their clustered organization. That is, we would expect the clustered organization, or micro-synteny, of such genes to be conserved during mammalian evolution. On the other hand, if imprinting regulation of these genes is not functionally important, we would not expect the micro-synteny of these genes to be conserved across mammalian genomes. To test these hypotheses, we derived all pairs of adjacent genes in the mouse genome (Methods) and estimated the propensity of their mammalian orthologs to be adjacent as well (**Figure 3.6.E**, Methods), using a probabilistic phylogenetic model analyzing phyletic patterns of presence and absence (Cohen and Pupko, 2011). This revealed that the mean propensity of adjacency of orthologs of mouse adjacent imprinted gene pairs, with parental biases lower than 85% to 15%, is significantly higher than the mean propensity of adjacency of orthologs of all mouse adjacent gene pairs (P-value = 4.8×10-5, **Figure 3.6.E**). This finding therefore supports the hypothesis that imprinting regulation of mouse genes, for which we detect weak parental biases, is functionally important. It may also suggest that the orthologs of many of these genes could be imprinted in the analyzed comparative mammalian species. In addition to newly identified imprinted genes associated with known imprinted clusters, several novel imprinted genes appear isolated from any other known imprinted gene (>2Mbp away). Interestingly, previous studies have reported a differential methylation between the two parental alleles for some of these genes. For example, differential

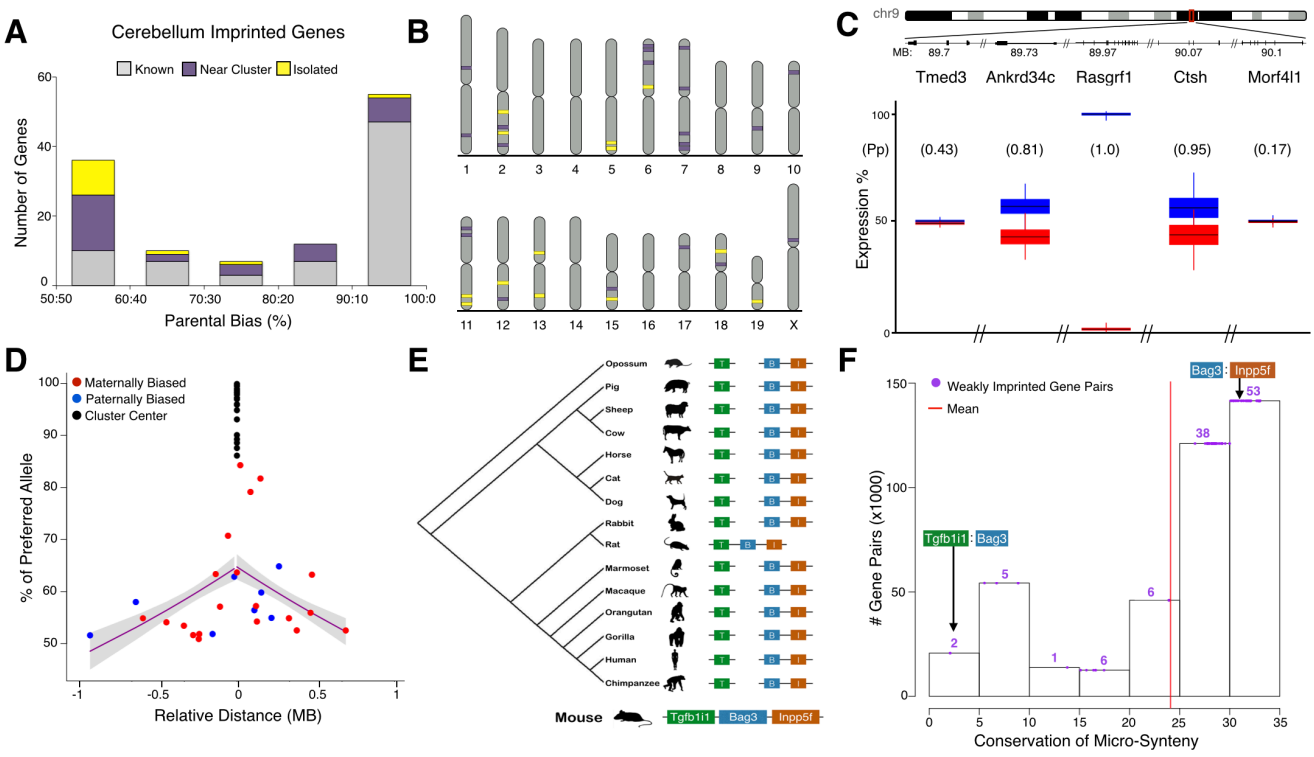

Figure 3.6: Genomic patterns of weakly imprinted genes

(A) Distribution of parentally biased according to genomic location relative to known imprinted regions. **(B)** Chromosomal location of imprinted genes. **(C)** Example of an imprinted region where an all-to-none imprinted gene is flanked by weakly imprinted genes. **(D)** Stereotypical decay in the degree of preferential expression of one parental allele over the other. **(E)** Example of mammalian conservation of neighborship between an all-to-none imprinted gene and a flanking weakly imprinted gene. **(F)** Distribution of the mammalian conservation of microsynteny for neighbor genes in the mouse.

methylation was observed in a region of chromosome 13 immediately downstream of Nhlrc1 (Xie et al., 2012), a novel paternally biased gene in our results (**Figure 3.4**). In humans, mutations in Nhlrc1 cause Lafora progressive myoclonic epilepsy (Romá-Mateo et al., 2012), a fatal neurological disorder characterized by the presence of massive intracellular inclusions observed in several neuronal cell types across the brain including the cerebellar granule cells. Differential methylation but not parentally biased expression in the brain was also reported within the Actinin alpha 1 (Actn1) gene at chromosome 12 (Calaway and Domínguez, 2012), which codes for a protein that regulates cytoskeleton interactions with the membrane. Our results show that this gene is indeed preferentially expressed from the paternal allele in the cerebellum.

Imprinting of the Apoptosis Pathway

We investigated whether specific functional categories were enriched among the list of cerebellum imprinted genes. One pathway which our data clearly suggest to be highly targeted by imprinting is programmed cell death or apoptosis, which includes seven genes previously exhibiting parent-of-origin monoallelic expression as well as five genes exhibiting parentally biased expression (**Figure 3.7.A**). Noticeably, most of the maternal genes promote apoptosis while most of the paternal genes inhibit apoptosis. Two genes, the paternally biased BclXL (Gregg et al., 2010a) and maternally biased Bag3, are of particular interest (**Figure 3.7.B**). Both genes have multiple functions independent from each other, not all of which are related to apoptosis. However, interactions between the two prevent mitochondrial release of factors controlling activation of caspases and thus the irreversible commitment to undergo apoptosis (Jacobs and Marnett, 2009).

In addition to genes directly associated with the apoptosis machinery, we also observe parental bias in genes encoding various proteins that regulate apoptosis in a context-dependent manner. One function that appears particularly well represented is the regulation of apoptosis by cell-to-cell interactions. These include the paternally expressed Dlk1, the paternally biased Adam23 and Actn1, and the maternally biased Bmf. Dlk1 encodes an atypical Notch ligand that has been implicated in apoptosis induction after loss of cellular contact (Kawakami et al., 2006). Adam23 is a member of the disintegrin protein family that regulates cell adhesion via interactions with Integrin and is propose as a negative regulator of tumor growth (Cal et al., 2000; Takada et al., 2005). Actn1 mediates interactions between the cytoskeleton and membrane by its interactions with Integrin and loss of these interactions promotes apoptosis (Triplett and Pavalko, 2006). Finally, Bmf is released from its normal microtubule localization to the cytoplasm in response to loss of cell membrane attachments where its interactions with BclXL induces apoptosis (Puthalakath, 2001). Together, these data suggest the specific targeting and regulation of the apoptosis pathway by genomic imprinting.

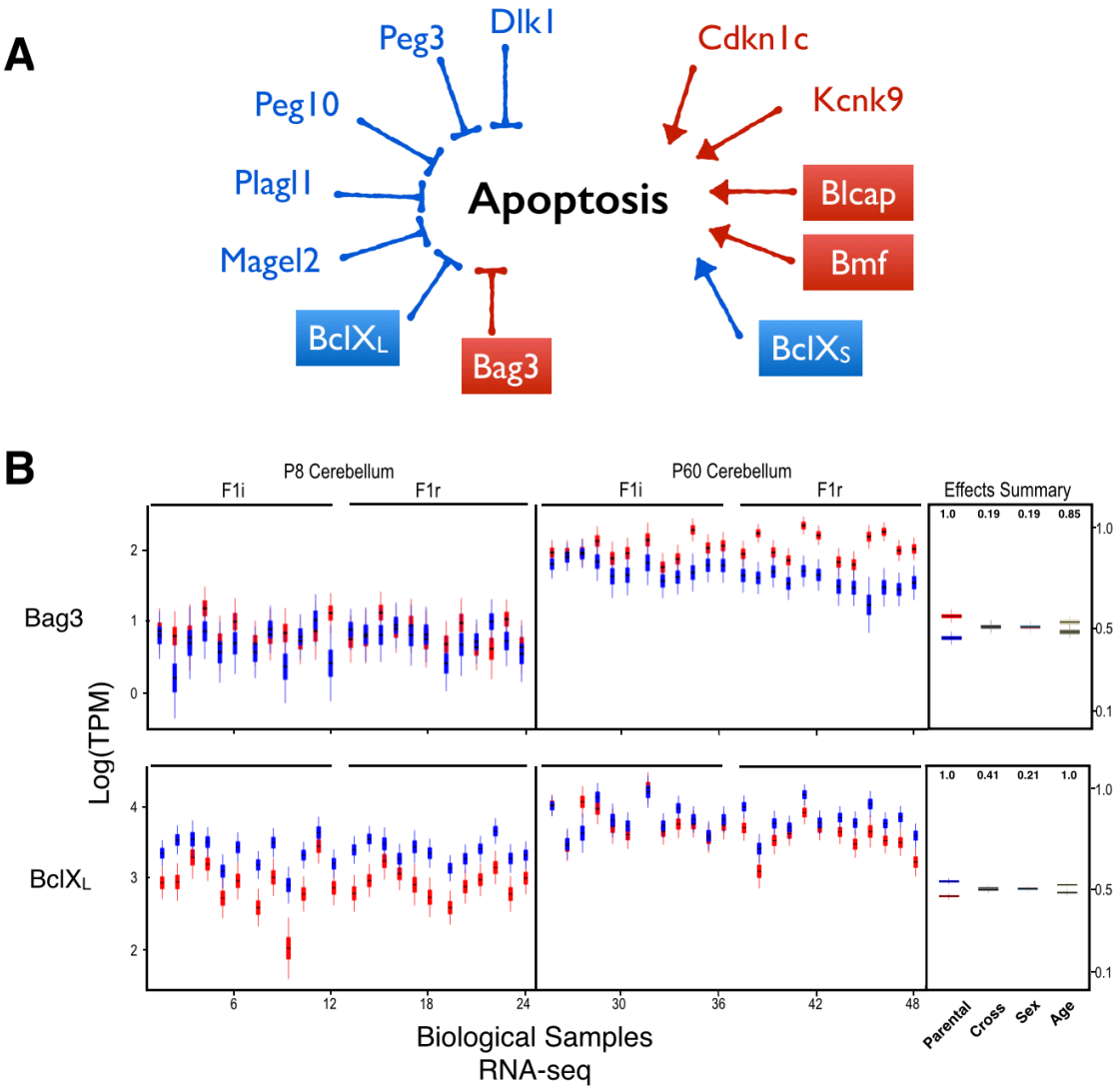

Figure 3.7: Abundance of anti- and pro-apoptotic under genomic imprinting regulation

(A) Imprinted genes involved in the regulation of apoptosis. Color background signifies parentally biased expression. **(B)** Bag3 and $BclX_L$ are weakly imprinted in the cerebellum.

DISCUSSION

Genome-Wide Analysis of Genomic Imprinting in the Cerebellum

In this study we have profiled genomic imprinting in the mouse cerebellum using a newly developed statistical approach that enables the analysis of RNA-seq data and detects parental allelic bias at the individual transcript level, while simultaneously estimating the effects of known factors in the experimental design such as age, sex, and the mouse cross. At the statistical significance level chosen, our approach detected 74 previously validated imprinted genes as well as 50 novel imprinted genes, 41 of which were successfully confirmed by pyrosequencing. This therefore indicates that our approach has a ~93% true positive rate. If we also take into account 10 additional genes, which had a statistical significance of parental bias below the chosen cutoff but were confirmed as imprinted by pyrosequencing, our study increases the number of known imprinted genes in the mouse by ~26%.

Importantly, in our data, most of the newly discovered imprinted genes and a considerable number of the known imprinted genes show parental expression biases rather than the traditionally described all-to-none silencing of a given parental allele. This observation, already made in Gregg et al., 2010 a, b, is made possible by the powerful and sensitive genome-wide experimental design. Currently, our experiments cannot distinguish whether the weak parental biases are uniform in all cells of the tissue analyzed, or result from averaging variable magnitudes of parental biases specific to individual cell types, as was reported for the nephron-specific cell type imprinting of the Gnas gene (Weinstein et al., 2000) and the neuron-specific imprinting of the Snx14 gene (Huang et al., 2014). A more sensitive approach of allelic expression analysis at cellular resolution will be needed to resolve this issue.

The imprinting regulation of several newly identified genes in the cerebellum is corroborated by several lines evidence. First, we find that ~65% of these genes are located by, or at known imprinted clusters, and this localization is observed to be evolutionarily conserved suggesting it is under purifying selection. In addition, three of the isolated biased genes contain regions differentially methylated between the parental alleles. Second, many of these genes are implicated in the same biological pathways, including cell survival and apoptosis, as monoallelically expressed genes.

Apoptosis

A number of imprinted genes are known to play important roles in the apoptosis pathway. In this study we find additional apoptosis-related genes imprinted with parental biases (**Figure 3.7.A**). Apoptosis plays an important role during brain development where it regulates cell population size and adjusts neuronal circuits by removing poorly connecting cells (Buss et al., 2006). Accordingly, most of the imprinted apoptosis-related genes observed in our study exhibit both stronger overall expression and parental biases in the developing cerebellum. We observe a significant number of pro- and anti-apoptotic genes differentially expressed according to brain region. Considering that most neurons in the adult brain are post-mitotic, the functional significance of such spatial regulation is still unclear. Nevertheless, some imprinted genes are known to be highly pleiotropic, hence alternative functions of these imprinted apoptotic genes in the adult brain cannot be ruled out (Jiao and Li, 2011; Li et al., 2010).

MATERIALS AND METHODS

Animal Subjects

F1 hybrids were generated by reciprocally crossing C57Bl/6J and Cast/EiJ mouse strains, where we denote by F1i an F1 hybrid derived from a C57Bl/6J father and a Cast/EiJ mother, and by F1r an F1 hybrid derived from a Cast/EiJ father and a C57Bl/6J mother. For the RNA-seq data, we used 48 animal subjects covering both crosses, both sexes, and two age groups: developing animals sacrificed at postnatal day 8, called P8, and adult animals sacrificed in the range of postnatal days 56-64, called P60. Our experimental design is balanced thus having three factors: cross, sex, and age, with six animal replicates in each factor block.

RNA-Seq Sample Preparation

RNA was isolated from dissected Cerebella using Trizol (Life Technologies) according to the manufacturer instructions and further purified using DNase I digestion and the RNeasy kit (Qiagen). We required samples to have RNA integrity score of above 9, according to the Agilent 2100 Bioanalyzer, to be used for RNA-seq library preparation. For each sample we used 3ug of total RNA to prepare libraries according to the Illumina Tru-Seq RNA sample preparation protocol. Sample purity and integrity was confirmed using the Agilent 2100 Bioanalyzer. We selected an average fragment size of 250 bp. Each animal subject was used to prepare a single library and was sequenced on an individual lane generating 59 bp single-end reads. The average read depth across our samples was 168,991,714.5.

Reference Genome and Transcriptome Data

We created a C57, CAST diploid genome by incorporating C57 and CAST SNPs and indels (obtained from the Mouse Genome Project: ftp://ftp-mouse.sanger.ac.uk/REL-1303-SNPs_Indels-GRCm38) into the Mus musculus GRCm38 reference genome sequence. We additionally created a transcriptome annotation set as follows. We first downloaded the Gencode (Engström et al., 2013; Steijger et al., 2013) M2 mouse main gene annotation (gencode.vM2.annotation) General Transfer Format (GTF) file and removed from it the following RNA types: Mt_rRNA, Mt_tRNA, miRNA, rRNA, snRNA, snoRNA, Mt_tRNA_pseudogene, tRNA_pseudogene, snoRNA_pseudogene, snRNA_pseudogene, scRNA_pseudogene, rRNA_pseudogene, miRNA_pseudogene, as they are not supposed to be represented in our RNA libraries. In order to include as an extensive set of transcripts as possible and to specifically cover retroposed genes due to their known involvement in genomic imprinting (McCole and Oakey, 2008b), we followed these steps to add additional annotated transcripts to gencode.vM2.annotation. First, we downloaded the Gencode Retrotransposed (gencode.vM2.2wayconspseudos) GTF file and the ucscRetroInfo2 mm10 mouse genome assembly retrogenes annotation file from the University of California Santa Cruz (UCSC) genome browser (Karolchik et al., 2014). We eliminated any redundancy between the two transcript sets by selecting the longest transcript between any two transcripts represented in both files. Following that, we eliminated any redundancy between that retrogene set and gencode.vM2.annotation transcript set by selecting the longest transcript between any two transcripts represented in both sets. In order to remove any redundancy between our retrogene set and single-exon protein coding transcripts (which is the structural hallmark of retrogenes) in gencode.vM2.annotation we kept the longest of any intersecting protein-coding single-exon

transcript in gencode.vM2.annotation and retrogene in our retrogenes set. Subsequently, we added all transcripts from the UCSC knownGene mm10 mouse genome assembly annotation file, which are not indicated to be represented in the gencode.vM2.annotation set. Finally, we added all functional RNAs from the functional RNA database (fRNAdb, Mituyama et al., 2009), which did not intersect with any of the transcripts in the set we generated in the previous steps and is longer than our 59 bp read length. Altogether, this amounted in 117,643 transcripts. We then used the UCSC liftOver utility to generate a C57, CAST diploid transcriptome set from our generated transcriptome set.

Processing of RNA-seq Data

Each RNA-seq library was first subjected to quality and adapter trimming using the Trim Galore utility (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore) with stringency level 3. We then mapped each of our C57×CAST hybrid RNA-seq sequenced libraries to the C57, CAST diploid genome and transcriptome splice junctions using STAR RNA-seq aligner (Dobin et al., 2013) allowing a maximum of three mismatches. Specifically, we mapped the data twice where after the first mapping step we incorporated valid splice junctions which were reported by STAR to exist in our RNA data. Subsequent to the second STAR mapping step we filtered any alignments that did not map to our transcript set using custom code. We note that by doing so we allow the reads to unbiasedly align to their best locations in the splice-junction aware genome and subsequently keep our alignments of interest as opposed to aligning the read data directly to the transcriptome (e.g., Roberts and Pachter, 2013). Following that, we estimated the expression levels with their respective uncertainties of each transcript in our C57, CAST diploid transcriptome using MMSEQ (Turro et al., 2011). MMSEQ uses a Bayesian model for

estimating expression levels and therefore computes a posterior distribution of the expression levels of each transcript in Fragment Per Kilobase per Million (FPKM) units. We first transformed these posterior FPKM samples to Transcript per Million units (TPM) as TPM units were shown to be unbiased and more interpretable (Wagner et al., 2012). Following that, for each transcript which had a posterior median TPM below 0.01, which we set as our minimal expression level cutoff, we set all of its TPM posterior samples to zero.

Using an extensive transcriptome annotation set on the one hand has the advantage of estimating the expression levels (and therefore testing for parentally-biased expression) of as many known transcripts that are expressed in the tissue from which RNA was purified, as possible. One the other hand, it is very likely that highly similar transcripts (e.g., NAGNAG alternative splice forms, Bradley et al., 2012) will not be distinguished by the read data. This, in turn, would increase the expression level uncertainties of such lowly identifiable transcripts and would therefore reduce power when testing whether their expression is parentally biased. Ideally, one would detect such lowly identifiable transcripts and combine their expression level estimates into that of single merged transcript, indicating that either of them or both of them are expressed. Such combined transcripts would therefore have lower expression level uncertainty than their constituents, which would therefore increase power when testing whether their expression is parentally biased. To achieve this, we adopted the approach of Turro et al. (2014) for collapsing lowly identifiable transcripts based on the posterior correlation of their expression level estimates, tailored for a diploid transcriptome case. That is, we require that the across-samples mean posterior correlation of expression level estimates of two homologous pairs of transcripts from each of the two alleles each be lower than a defined cutoff (which we empirically set to -0.25), in order to combine each of these pairs into a single combined transcript, yielding a

homologous combined transcripts. This consistency between the alleles in the collapsing process ensures that the resulting combined transcripts are identical for the two alleles and can therefore be tested for parentally biased expression.

A Statistical Model for Inferring Genomic Imprinting from an RNA-seq Experimental Design

For inferring whether a given transcript is imprinted, we define our estimand of interest as the difference in the expression level between its paternal and maternal alleles, i.e., the parental bias. Intuitively, if the parental bias is approximately zero across all samples we would conclude that the transcript is not imprinted. The reality of RNA-seq data, however, is more complicated than that. First, as mentioned above, we do not obtain accurate estimates of expression levels but rather estimates with uncertainty. Next, our experimental design may include inherent factors that can affect allele-specific expression levels to various extents, such as the mouse cross, sex, and age. Whether of interest or not, the effects of these factors need to be accounted for. In addition, even though our nearly genetically identical animal subjects are treated with similar conditions, thereby minimizing effects of any additional factors to the ones specified above (such as environment), we still expect biological variability in RNA expression across our subjects (e.g., due to litter effects). Finally, we would expect technical variability across experiments to add to the biological variability, yet unless addressed explicitly (e.g., sequencing the same RNA library as technical replicates), the two are indistinguishable.

To address all of these issues we developed a statistical model for inferring genomic imprinting from our experimental design for every transcript in our annotation set. Specifically, we have chosen to extend the Bayesian variable selection regression model of Chipman et al. (1997) by accounting for the measurement error in the response, as uncertainties are naturally

propagated in a Bayesian framework. In our model, the response of sample j ($\hat{y}_j$, where $j = 1,...,n$ samples) for a certain transcript is the mean posterior difference between the paternal and maternal log posterior TPM samples, i.e.,

$$\hat{y}_j = \frac{\sum_{s \in S} \log(TPM_s^{p_j} + c) - \log(TPM_s^{m_j} + c)}{S} \quad (1)$$

We denote by s a specific posterior sample of S posterior samples, and by c a constant that we add to the TPM expression in order to avoid taking the log of zero (we use $c = 0.01$). Since regression parameters are sensitive to the scale of inputs (Gelman, 2008) yet we wish to have a common interpretation for all transcripts we fit our model to, we divide the response $\hat{y}_j$ by the standard deviation of $\hat{y}_j$'s across all $j \in n$ samples and denote this scaled response as $\hat{y}'_j$.

We define the uncertainty (or measurement error) of the response as:

$$\hat{\varepsilon}_j = \frac{\sum_{s \in S} (\log(TPM_s^{p_j} + c) - \log(TPM_s^{m_j} + c) - \hat{y}_j)^2}{S - 1} \quad (2)$$

We thus model:

$$(\hat{y}'_1,...,\hat{y}'_j,...,\hat{y}'_n) = y' \mid z \sim MVN(z, E) \quad (3)$$

We denote $z = (z_1,...,z_j,...,z_n)$ as the unobserved (latent) true value of the response, and $E = diag(\hat{\varepsilon})$ as the covariance matrix of the response errors $\hat{\varepsilon} = (\hat{\varepsilon}_1,...,\hat{\varepsilon}_j,...,\hat{\varepsilon}_n)$. We denote by n the total number of samples. In our experimental design, each of our k factors, namely, cross, sex, and age has two levels, and each factor is represented by n_{rep} replicates, therefore n

corresponds to $2^k \times n_{rep}$ observations. We model the, unobserved, true value of the response, z, as:

$$z \mid \beta, \sigma^2 \sim MVN(X, \beta, \Sigma) \tag{4}$$

We denote the experimental design matrix by $X = (X_1, ..., X_p)$, the regression factor parameters by $\beta = (\beta_1, ..., \beta_p)$, and across-samples errors as $E = diag(\sigma^2)$. In our model we define the parental bias as the mean response across all samples, which is the intercept of the regression. Therefore, X_1 is a column vector of 1's, β_1 is the parameter that quantifies the effect of the parental bias, and $p = k + 1$ is the number of factors. Since we are interested in testing whether each effect i is significant (i.e., whether β_i is significantly different from zero), we model the β's as a mixture of two normals, such that the first normal is centered around zero with a small variance, representing a non-significant effect, and the second normal is centered around the estimated β_i, and considers the effect as significant:

$$\beta_i \mid \sigma^2, \tau_i, c_i, \delta_i \sim \begin{cases} N(0, \sigma^2 \tau_i^2), & \delta_i = 0 \\ N(0, \sigma^2 (c_i \tau_i)^2), & \delta_i = 1 \end{cases} \tag{5}$$

δ_i can be interpreted as a random variable that indicates whether factor i has an effect on the response. $\delta's$ are therefore modeled as Bernoulli i.i.d., with probability p_i for each δ_i:

$$\pi(\delta) \propto \prod_{i \in p} p_i \tag{6}$$

The τ's and c's are hyper-parameters for scaling the two normals and p_i is the prior probability that factor i has a significant effect on the response. Finally, we put a conjugate prior on σ^2:

$$\sigma^2 \sim IG(\nu / 2, \nu\lambda / 2) \tag{7}$$

Where ν and λ are hyper-parameters for the location and scale of σ^2, respectively.

The joint distribution of the observed data, $y^{'}$, the covariance matrix of the response errors, $E = diag(\hat{\varepsilon})$, and the parameters and latent variables, $\theta = (z, \beta, \sigma^2, \delta)$ is:

$$f(\hat{y}^{'}, E; \theta) = f(\hat{y}^{'} \mid z, E)\pi(z \mid \beta, \sigma^2)\pi(\beta \mid \delta, \sigma^2)\pi(\delta)\pi(\sigma^2) \tag{8}$$

Where,

$$f(\hat{y}^{'} \mid z, E) \propto \mid E \mid^{-1/2} \exp\left\{-\frac{1}{2}(\hat{y}^{'} - z)^T E(\hat{y}^{'} - z)\right\} \tag{9}$$

$$\pi(z \mid \beta, \sigma^2) \propto \mid \Sigma \mid^{-1/2} \exp\left\{-\frac{1}{2}(z - X\beta)^T \Sigma^{-1}(z - X\beta)\right\} \tag{10}$$

$$\pi(\beta \mid \delta, \sigma^2) \propto (\sigma)^{-p} \exp\left\{-\frac{1}{2}\beta^T \Sigma_{\delta}^{-1} \beta\right\} \tag{11}$$

Where,

$$\Sigma_{\delta} = diag(\sigma c_i^{\delta_i} \tau_i) \tag{12}$$

$$\pi(\sigma^2) \propto (\sigma^2)^{-\frac{\nu}{2}+1} \exp\left\{-\frac{\nu\lambda}{2\sigma^2}\right\} \tag{13}$$

And,

$$\pi(\delta) \sim multinom(p) \tag{14}$$

The full conditional distribution of z is:

$$\begin{aligned} &f(z \mid \sigma^2, \beta, \hat{y}^{'}) \propto f(\hat{y}^{'} \mid z, E)\pi(z \mid \beta, \sigma^2) \propto \\ &\mid E \mid^{-1/2} \exp\left\{-\frac{1}{2}(\hat{y}^{'} - z)^T E(\hat{y}^{'} - z)\right\} \times \\ &\mid \Sigma \mid^{-1/2} \exp\left\{-\frac{1}{2}(z - X\beta)^T \Sigma^{-1}(z - X\beta)\right\} \end{aligned} \tag{15}$$

Where by dropping the terms not involving z we get:

$$f(z \mid \sigma^2, \beta, \hat{y}^{'}) \propto \exp\left\{-\frac{1}{2}\left[(\hat{y}^{'} - z)^T E(\hat{y}^{'} - z) + (z - X\beta)^T \Sigma^{-1}(z - X\beta)\right]\right\} \tag{16}$$

We use the proposition that if:

$$f(z \mid \mu, y, S, \Sigma) \propto \exp\left\{-\frac{1}{2}\left[(z - \mu)^T S^{-1}(z - \mu) + (y - z)^T \Sigma^{-1}(y - z)\right]\right\} \tag{17}$$

Then (see proof in Gelman et al., 2013),

$$z \sim MVN(\mu_z, \Lambda_z) \tag{18}$$

Where,

$$\mu_z = \Lambda_z(S^{-1}\mu + \Sigma^{-1}y) \tag{19}$$

And,

$$\Lambda_z = (S^{-1} + \Sigma^{-1})^{-1} \tag{20}$$

We thus obtain,

$$f(z \mid \sigma^2, \beta, \hat{y}^{'}) \propto \exp\left\{-\frac{1}{2}\left[(z - \mu_z)^T \Lambda_z^{-1}(z - \mu_z)\right]\right\} \tag{21}$$

Where,

$$\mu_z = \Lambda_z(\Sigma^{-1}X\beta + E^{-1}\hat{y}^{'}) \tag{22}$$

And,

$$\Lambda_z = (E^{-1} + \Sigma^{-1})^{-1} \tag{23}$$

And therefore,

$$z \mid \sigma^2, \beta, \hat{y}^{'} \sim MVN(\mu_z, \Lambda_z) \tag{24}$$

The full conditional distribution of β is:

$$f(\beta \mid z,\sigma^2,\delta) \propto \pi(z \mid \beta,\sigma^2)\pi(\beta \mid \delta,\sigma^2) \propto$$
$$|\Sigma|^{-1/2} \exp\left\{-\frac{1}{2}(z-X\beta)^T\Sigma^{-1}(z-X\beta)\right\} \times \sigma^{-p} \exp\left\{-\frac{1}{2}\beta^T\Sigma_\delta^{-1}\beta\right\} \quad (25)$$

Where by dropping the terms not involving β we get:

$$f(\beta \mid z,\sigma^2,\delta) \propto \exp\left\{-\frac{1}{2}\left[(z-X\beta)^T\Sigma^{-1}(z-X\beta)+\beta^T\Sigma_\delta^{-1}\beta\right]\right\} \quad (26)$$

We use the proposition that if:

$$y \mid \beta,\Sigma \sim MVN(X\beta,\Sigma) \quad (27)$$

And,

$$\beta \mid D \sim MVN(0,D) \quad (28)$$

Then (see proof in Lindley and Smith, 1972),

$$\beta \mid y,\Sigma,D \sim MVN(\mu_\beta,\Lambda_\beta) \quad (29)$$

Where,

$$\mu_\beta = \Lambda_\beta X^T\Sigma^{-1}y \quad (30)$$

And,

$$\Lambda_\beta = (X^T\Sigma^{-1}X + D^{-1})^{-1} \quad (31)$$

We thus obtain,

$$f(\beta \mid z,\sigma^2,\delta) \propto \exp\left\{-\frac{1}{2}\left[(\beta-\mu_\beta)^T\Lambda_\beta^{-1}(\beta-\mu_\beta)\right]\right\} \quad (32)$$

Where,

$$\mu_\beta = \Lambda_\beta(X^T\Sigma^{-1}z) \quad (33)$$

And,

$$\Lambda_\beta = (X^T \Sigma^{-1} X + \Sigma_\delta^{-1})^{-1} \tag{34}$$

And therefore,

$$\beta \mid z, \sigma^2, \delta \sim MVN(\mu_\beta, \Lambda_\beta) \tag{35}$$

The full conditional distribution of σ^2 is:

$$\begin{aligned} & f(\sigma^2 \mid z, \beta, \delta) \propto \pi(z \mid \beta, \sigma^2)\pi(\beta \mid \delta, \sigma^2)\pi(\sigma^2) \propto \\ & [\sigma^2]^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2}(z - X\beta)^T (z - X\beta)\right\} \times [\sigma]^{-p} \exp\left\{-\frac{1}{2\sigma^2}\beta^T \Sigma_\delta^{-1} \beta\right\} \times \\ & [\sigma^2]^{-\frac{\nu}{2}} \exp\left\{-\frac{\nu\lambda}{2\sigma^2}\right\} \end{aligned} \tag{36}$$

Where by dropping the terms not involving σ^2 we get:

$$f(\sigma^2 \mid z, \beta, \delta) \propto [\sigma^2]^{-\frac{(n+p+\nu)}{2}-1} \exp\left\{-\frac{1}{2\sigma^2}\left[\nu\lambda + (z - X\beta)^T (z - X\beta) + \beta^T \Sigma_\delta^{-1} \beta\right]\right\} \tag{37}$$

Therefore,

$$\sigma^2 \mid z, \beta, \delta \sim IG\left(\frac{1}{2}(n + p + \nu), \frac{1}{2}\left[\nu\lambda + (z - X\beta)^T (z - X\beta) + \beta^T \Sigma_\delta^{-1} \beta\right]\right) \tag{39}$$

The full conditional distribution of δ after dropping the terms not involving δ is:

$$f(\delta \mid z, \beta, \sigma^2) \propto \pi(\beta \mid \delta, \sigma^2)\pi(\delta) \tag{40}$$

The joint conditional distribution of δ, however, is unknown, and therefore it is more suitable to sample each δ_i independently, given the set $\delta_{[-i]} = \{\delta_1, .., \delta_{i-1}, \delta_{i+1}, ..., \delta_p\}$. Using equation 40 we get:

$$f(\delta_i \mid \delta_{[-i]}, z, \beta, \sigma^2) \propto \pi(\delta_i \mid \delta_{[-i]}\beta, \sigma^2)\pi(\delta_i, \delta_{[-i]}) \tag{41}$$

Therefore,

$$p(\delta_i = 1 \mid \delta_{[-i]}, z, \beta, \sigma^2) =$$

$$\frac{\pi(\beta \mid \delta_i = 1, \delta_{[-i]}, \sigma^2)\pi(\delta_i, \delta_{[-i]})}{\pi(\beta \mid \delta_i = 1, \delta_{[-i]}, \sigma^2)\pi(\delta_i = 1, \delta_{[-i]}) + \pi(\beta \mid \delta_i = 0, \delta_{[-i]}, \sigma^2)\pi(\delta_i = 0, \delta_{[-i]})} =$$

$$\frac{\pi(\delta_i, \delta_{[-i]})}{\pi(\delta_i = 1, \delta_{[-i]}) + \frac{\pi(\beta \mid \delta_i = 0, \delta_{[-i]}, \sigma^2)}{\pi(\beta \mid \delta_i = 1, \delta_{[-i]}, \sigma^2)}\pi(\delta_i = 0, \delta_{[-i]})} \tag{42}$$

Where $\frac{\pi(\beta \mid \delta_i = 0, \delta_{[-i]}, \sigma^2)}{\pi(\beta \mid \delta_i = 1, \delta_{[-i]}, \sigma^2)}$ is the ratio of the normal mixture for β from equation 11.

Our Gibbs sampler therefore follows these steps:

Initialize the parameters.

Iterate until convergence on sampling each parameter from its full conditional distribution:

Sample $z \mid \sigma^2, \beta, \hat{y}^{'} \sim MVN(\mu_z, \Lambda_z)$ (equations 18, 22, and 23).

Sample $\beta \mid y, \Sigma, D \sim MVN(\mu_\beta, \Lambda_\beta)$ (equations 29, 33, and 34).

Sample $\sigma^2 \mid z, \beta, \delta \sim IG\left(\frac{1}{2}(n + p + \nu), \frac{1}{2}\left[\nu\lambda + (z - X\beta)^T(z - X\beta) + \beta^T\Sigma_\delta^{-1}\beta\right]\right)$

For $i = 1, \ldots, p$, sample δ_i according to equation 42.

In all our analyses we ran our Gibbs sampler for 10,000 iterations discarding the first 1,000 as burn-in.

In our experimental design all factors have binary levels: parental effect: paternal and maternal; cross effect: F1i and F1r; sex effect: male and female; and, age effect: P8 and P60. Since in this study we are interested in contrasting the two levels in each effect we set the X matrix cross, sex, and age columns to 1 for F1i's, males, and P8's, and to -1 for F1r's, females , and P60's. Each of the β parameters, which quantifies the effect of the corresponding factor,

can therefore be split into the respective levels of the factor. For example, the β which quantifies the parental effect corresponds to the mean effect across samples. Therefore:

$$\hat{\beta} = \frac{\sum_{j \in n} y_j}{n} = \frac{\sum_{j \in n} \left[\log(TPM_{p_j}) - \log(TPM_{m_j}) \right]}{n} = \frac{\log\left[\prod_{j \in n} \left(TPM_{p_j} \middle/ TPM_{m_j} \right) \right]}{n} \tag{43}$$

Therefore:

$$\exp(\hat{\beta}) = \frac{\sqrt[n]{\prod_{j \in n} TPM_{p_j}}}{\sqrt[n]{\prod_{j \in n} TPM_{m_j}}} \tag{44}$$

The geometric means $\sqrt[n]{\prod_{j \in n} TPM_{p_j}}$ and $\sqrt[n]{\prod_{j \in n} TPM_{m_j}}$ correspond to the median paternal and maternal expression across samples. Denoting the biallelic expression as:

$$\sqrt[n]{\prod_{j \in n} TPM_{p_j}} + \sqrt[n]{\prod_{j \in n} TPM_{m_j}} = 1 \tag{45}$$

we can therefore represent the median paternal and maternal fractions in terms of β:

$$\sqrt[n]{\prod_{j \in n} TPM_{p_j}} = \frac{\exp(\beta)}{1 - \exp(\beta)} \tag{46}$$

And

$$\sqrt[n]{\prod_{j \in n} TPM_{m_j}} = \frac{1}{1 - \exp(\beta)} \tag{47}$$

Our model computes samples from the posterior distributions of each of the parameters. For inferring an effect as statistically significant, we require that the mean posterior value of its corresponding δ parameter be higher than 0.95.

As described above, our model has several hyper-parameters that need to be specified. Namely, the normal mixture prior distribution of each β_i is scaled by respective τ_i and c_i. The c parameter acts as a multiplying constant that determines how much higher an important effect has to be relative to a negligible effect in order to be considered significant. The τ parameter will therefore determine the posterior probability of an important effect being significant. The shape and scale of the inverse gamma prior distribution of σ^2 is determined by ν and λ, respectively. The ν parameter should be set to a relative uninformative value, yet larger than zero in order to avoid obtaining low values of σ^2 which will result in overinflated posteriors for selecting effects as significant. Finally, each δ_i has prior probability p_i of being selected.

We follow Chipman et al. (1997) in setting ν to a value near two (specifically 2.5) and setting $\lambda = \frac{2\sqrt{\text{var}(\hat{y}')}}{5}\frac{1}{\nu}\left(\frac{\nu}{2}-1\right)$. Since $\text{var}(\hat{y}') = 1$ for all transcripts, for $\nu = 2.5$ we get $\lambda = 0.04$. We chose τ and c of all transcripts to have empirical values of 0.1 and 4.25, respectively, in order to provide a good separation between imprinted and non-imprinted transcripts at the selected PP = 0.95 cutoff. In addition, we set to p of all effects to 0.1, reflecting our prior belief of the effects being significant. Although the choice of these hyper-parameter values is arbitrary to some extent, if the model is robust this choice should not affect the ranking of the δ estimates but merely their values. Therefore, in order to evaluate how the inference of genomic imprinting by our model is affected by our choice of empirical hyper-parameter values we performed the following sensitivity analysis. The empirical value that we chose for each of the five hyper-parameters: τ, c, ν, λ, and p, was perturbed by selecting four other values. Specifically, we perturbed the empirical $\tau = 0.1$ value with 0.005, 0.01, 1, and 2 thus both lowering and elevating

the posterior probability of an effect begin significant; we perturbed the empirical $c = 4.25$ value only with the higher values: 1.7, 3.4, 5.3125, and 10.625 as 4.25 was found to be around the minimal value for detecting significant effects; we perturbed the empirical $\nu = 2.5$ value with 5, 12.5, 25, and 50 as ν cannot assume values lower than 2; we perturbed the empirical $\lambda = 0.04$ value with 0.002, 0.004, 0.4, and 0.8 shifting the σ^2 prior distribution from informative to uninformative; and finally, we perturbed the empirical $p = 0.1$ value with higher prior probabilities of 0.2, 0.3, 0.4, and 0.5. In each such perturbation we re-fitted our model to the data where all other hyper-parameter values are held fixed at their original empirical unperturbed values (thereby achieving a one-at-a-time sensitivity analysis). We assessed the perturbed results using a receiver operating characteristic (ROC) analysis in which the imprinted transcripts obtained by the unperturbed inferences were used as the ground-truth positives, and all other non-imprinted transcripts were used as ground-truth negatives. For practical computational considerations, for all perturbations we used a random sample of 10% of the 38,112 transcripts in our data set. For this sample to reflect the transcripts proportionally with respect to the PP of their parental effect, we binned the 38,112 parental effect PP distribution to a 100 bin histogram and randomly sampled 10% from each bin. In all perturbations the area under the ROC curve (AUC) was found to be 1, except for the $\tau = 1$ and $\tau = 2$ perturbations which obtained AUCs of 0.99 and 0.95, respectively. These results therefore indicate that the ranking of transcripts according to the PP of the parental effect, obtained by our model, is robust to the choice of hyper-parameter empirical values. Therefore, the practice employed in this study of selecting empirical values, setting a parental effect PP cutoff, and experimentally confirming all transcripts with a PP of parental effect above that cutoff, is a reasonable choice for obtaining reliable inference of genomic imprinting from RNA-seq data.

In order to detect all imprinted transcripts in our data we fitted our model to all 48 P8 and P60 samples, thus estimating cross, sex, and age effects on the parental bias. This, however, requires that a given transcript is expressed in both age groups. In addition, a significant age effect may mean that a transcript has a strong parental bias in one age group but none in the other. In that case, the PP of the overall parental bias may not be significant. For these reasons we additionally fitted our model to the 24 P8 and P60 samples, independently, thus estimating only cross and sex effects on the parental bias for each of the age groups. We thus report all transcripts with a parental effect PP > 0.95 in the combined P8, P60 dataset, or exclusively expressed in one of the age group datasets with a parental effect PP > 0.95 as imprinted. Transcripts with an age effect PP > 0.95 in the P8, P60 dataset and with a parental effect PP > 0.95 in either of the individual P8 and P60 datasets were additionally reported as imprinted.

Analysis of the Decay of Parentally-Biased Expression as a Function of Distance from Imprinted Cluster Centers

To quantify the effect that physical chromosomal distance of imprinted genes from their corresponding imprinted cluster center has on the magnitude of their parentally-biased expression we performed the following analysis. We started off with the list of transcripts which were either validated to be imprinted in this study or their corresponding genes were validated to be imprinted in previous studies. Retrogenes were excluded from this list as the context of their genomic location is not strongly relevant for the question addressed by this analysis. For each transcript in this list we computed the parental bias as:

$$\max\left(\frac{\exp(\beta)}{1-\exp(\beta)}, \frac{1}{1-\exp(\beta)}\right) \tag{48}$$

In words, the median parental expression fraction across all samples (see equations 43-47). In order to represent a gene by a single representative parental bias, for each gene we chose the isoform with the most significant parental bias (i.e., the isoform with the maximal posterior mean δ), derived from the model fitted to both age groups. In cases of genes encoded within a host gene the gene with the maximal parental bias was chosen as the representative of that genomic location. In order to assign genes to clusters we linearly scanned each chromosome from its 5' to 3' end. The most upstream gene on a chromosome (where we represent the start and end sites of genes encoded on the antisense strand as their end and start sites on the sense strand, respectively), was assigned to the first imprinted cluster of that chromosome. Then, if the start site of the next downstream gene is within 1Mbp of the end site of the previous upstream gene it was assigned to the same cluster, else it was assigned to a new cluster. Thus, the maximal distance between any two clusters can be 1Mbp. Practically, this inter-cluster distance is much larger. We then defined the genes with a parental bias larger than 85% to 15% as candidate centers of their respective clusters. Since in some imprinted clusters more than a single gene met this condition, we grouped all physically consecutive candidate cluster center genes to a single gene which start site was defined as the start site of the most upstream candidate center gene and end site was defined as the end site of the most downstream candidate center gene. If an imprinted cluster resulted with more than one such group of candidate center genes flanked by genes with parental biases lower than 85% to 15%, it was broken down to sub-clusters centered on each of these candidate centers. We defined the boundaries of each sub-cluster as half the distance between the end site of its center gene and the start site of the center gene of the adjacent sub-cluster (or up to 1M bp in the cases of the most up- and down-stream sub-clusters) and thus flanking genes were assigned to the sub-cluster which boundary was downstream to

their start site location. We finally removed any gene which start site was more than 1Mbp away from its respective cluster center. We then fitted a general linear fixed effects model to non-cluster center genes, where we defined the response as the parental bias and the fixed effect as the intergenic distance from the cluster center.

Analysis of the Evolutionary Propensity for Clustering of Imprinted Genes

To test whether the clustered organization of mouse imprinted genes with weak parental-expression biases (lower than 85% to 15%) is conserved in mammalian evolution, which therefore suggests functional importance, we carried out the following analysis. We defined a pair of adjacent genes in the mouse reference genome as any two genes which intergenic distance was below 1Mbp. Using this set of adjacent pairs of genes as reference we then constructed a phyletic pattern of orthologous gene pairs as follows. We downloaded the gene orthology assignment between the mouse reference genome and all mammalian reference genomes available in the Ensembl genome browser (Flicek et al., 2014), which are assembled at the chromosome level. This included the following genomes: opossum, pig, sheep, cow, horse, cat, dog, rat, marmoset, rhesus, orangutan, gorilla, chimpanzee, and human (Figure 3E). For each comparative genome, if both orthologs of a mouse adjacent gene pair exist in that genome, we labeled this orthologous pair by "1" if their intergenic distance is lower than 1Mbp (i.e., they are also adjacent in the comparative genome) and "0" otherwise (i.e., they are not adjacent in the comparative genome). If either or both orthologs do not exist in the comparative genome we labeled this pair as "?" (i.e., unknown). Ensembl tables of orthology assignment provide two types of homology between a query and search genomes: a one-to-one and a one-to-many homology. In addition, these tables also provide a binary confidence score for the orthology

assignment (0 for low and 1 for high). As a conservative approach, we filtered out all one-to-many and low confidence orthologs unless the genes orthologous to the mouse pair were found to be within 1Mbp of each other. As a result we obtained a phyletic pattern, which is a type of a multiple sequence alignment where each site is a gene pair orthologous to an adjacent gene pair in the mouse genome with the characters of "1", "0", and "?". All phyletic pattern columns in which all sites were "?", were removed thus obtaining a phyletic patterns of 477,279 sites. We then fitted to this phyletic pattern and corresponding phylogenetic species tree (Figure 3E) a probabilistic phylogenetic model estimating the substitution rates between presence and absence ("1" and "0", respectively) and vice-versa (Cohen and Pupko, 2011). We specifically allowed the presence/absence ratio to vary across sites and the presence and absence probabilities at the root of the phylogenetic tree to be independently estimated. As a result, we obtained a posterior expectation of the absence-to-presence substitution rate (i.e., rate of gene-pair gain: "0" to "1") and a posterior expectation of the presence-to-absence substitution rate (i.e., rate of gene-pair loss: "1" to "0") for each site in the phyletic patters. The propensity of each orthologous gene pair to be adjacent is therefore the posterior expectation of its rate of gene-pair gain divided by the posterior expectation of its rate of gene-pair loss. In order to test whether the propensity of adjacency of orthologs of mouse weakly parentally-biased imprinted gene pairs (with a bias lower than 85% to 15%), is significantly higher than what would be expected by chance, we compared the mean propensity of adjacency of these gene pairs with that of all gene pairs in the phyletic pattern using a one-sided z-test.

Pyrosequencing Validations

Validations were performed in cerebellums derived from a different batch of young (P8) and adult (P60) F1 hybrids from that use in RNA-Seq experiments. RNA purification and quality control followed the same procedures described above. An average of three SNPs suitable for pyrosequencing analyses were identified for each candidate gene. Previously described three-primer strategy (Royo et al., 2007), including a 3'-biotinylated primer common to all reactions, was employed for the amplification of each targeted SNP. All primers were design using Pyromark Assay Design 2.0. Pyromark One Step RT-PCR kit (Qiagen) was used for the amplification of each targeted region, followed by purification of single stranded biotinylated DNA according to the manufacturer instructions. Pyrosequencing was performed on the Pyromark Q96 MD (Qiagen). For each SNP at least 12 replicates were separately analyzed. Statistical significance of results obtained for each SNP were analyzed using a version of our Bayesian Regression Genomic Imprinting Model customized for pyrosequencing data.

REFERENCES

Amaral, D.G., Schumann, C.M., and Nordahl, C.W. (2008). Neuroanatomy of autism. Trends Neurosci. 31, 137–145.

Arnaud, P., Monk, D., and Hitchins, M. (2003). Conserved methylation imprints in the human and mouse GRB10 genes with divergent allelic expression suggests differential reading of the same mark. Hum. Mol. ….

Babak, T., Deveale, B., Armour, C., Raymond, C., Cleary, M.A., van der Kooy, D., Johnson, J.M., and Lim, L.P. (2008a). Global survey of genomic imprinting by transcriptome sequencing. Curr. Biol. 18, 1735–1741.

Babak, T., Deveale, B., Armour, C., Raymond, C., Cleary, M.A., van der Kooy, D., Johnson, J.M., and Lim, L.P. (2008b). Global survey of genomic imprinting by transcriptome sequencing. Curr. Biol. 18, 1735–1741.

Bartolomei, M.S., and Ferguson-Smith, A.C. (2011). Mammalian genomic imprinting. Cold Spring Harb. Perspect. Biol. 3.

Bradley, R.K., Merkin, J., Lambert, N.J., and Burge, C.B. (2012). Alternative splicing of RNA triplets is often regulated and accelerates proteome evolution. PLoS Biol. 10.

Buss, R.R., Sun, W., and Oppenheim, R.W. (2006). Adaptive roles of programmed cell death during nervous system development. Annu. Rev. Neurosci. 29, 1–35.

Cal, S., Freije, J.M., López, J.M., Takada, Y., and López-Otín, C. (2000). ADAM 23/MDC3, a human disintegrin that promotes cell adhesion via interaction with the alphavbeta3 integrin through an RGD-independent mechanism. Mol. Biol. Cell 11, 1457–1469.

Calaway, J., and Domínguez, J. (2012). Intronic Parent-of-Origin Dependent Differential Methylation at the Actn1 Gene Is Conserved in Rodents but Is Not Associated with Imprinted Expression. PLoS One.

Cattanach, B.M., and Jones, J. (1994). Genetic imprinting in the mouse: implications for gene regulation. J. Inherit. Metab. Dis. 17, 403–420.

Chipman, H., Hamada, M., and Wu, C. (1997). A Bayesian variable-selection approach for analyzing designed experiments with complex aliasing. Technometrics.

Choi, J.D., Underkoffler, L.A., Wood, A.J., Collins, J.N., Williams, P.T., Golden, J.A., Schuster, E.F., Loomes, K.M., and Oakey, R.J. (2005). A novel variant of Inpp5f is imprinted in brain, and

its expression is correlated with differential methylation of an internal CpG island. Mol. Cell. Biol. 25, 5514–5522.

Cohen, O., and Pupko, T. (2011). Inference of gain and loss events from phyletic patterns using stochastic mapping and maximum parsimony-A simulation study. Genome Biol. Evol. 3, 1265–1275.

Davies, W., Isles, A., Smith, R., Karunadasa, D., Burrmann, D., Humby, T., Ojarikre, O., Biggin, C., Skuse, D., Burgoyne, P., et al. (2005). Xlr3b is a new imprinted candidate for X-linked parent-of-origin effects on cognitive function in mice. Nat. Genet. 37, 625–629.

Deng, Q., Ramskold, D., Reinius, B., and Sandberg, R. (2014). Single-cell RNA-seq reveals dynamic, random monoallelic gene expression in mammalian cells. Science (80-.). 343, 193–196.

DeVeale, B., van der Kooy, D., and Babak, T. (2012a). Critical evaluation of imprinted gene expression by RNA-seq: A new perspective. PLoS Genet. 8.

DeVeale, B., van der Kooy, D., and Babak, T. (2012b). Critical evaluation of imprinted gene expression by RNA-seq: A new perspective. PLoS Genet. 8.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: Ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21.

Engström, P.G., Steijger, T., Sipos, B., Grant, G.R., Kahles, A., Rätsch, G., Goldman, N., Hubbard, T.J., Harrow, J., Guigó, R., et al. (2013). Systematic evaluation of spliced alignment programs for RNA-seq data. Nat. Methods 10, 1185–1191.

Flicek, P., Amode, M.R., Barrell, D., Beal, K., Billis, K., Brent, S., Carvalho-Silva, D., Clapham, P., Coates, G., Fitzgerald, S., et al. (2014). Ensembl 2014. Nucleic Acids Res. 42.

Gelman, A. (2008). Scaling regression inputs by dividing by two standard deviations. Stat. Med. 27, 2865–2873.

Gelman, A., Carlin, J., Stern, H., and Dunson, D. (2013). Bayesian data analysis.

Gregg, C., Zhang, J., Weissbourd, B., Luo, S., Schroth, G.P., Haig, D., and Dulac, C. (2010a). High-resolution analysis of parent-of-origin allelic expression in the mouse brain. Sci. (New York, NY) 329, 643–648.

Gregg, C., Zhang, J., and Weissbourd, B. (2010b). High-resolution analysis of parent-of-origin allelic expression in the mouse brain. Science (80-.). 329, 643–648.

Hagiwara, Y., Hirai, M., Nishiyama, K., Kanazawa, I., Ueda, T., Sakaki, Y., and Ito, T. (1997). Screening for imprinted genes by allelic message display: identification of a paternally expressed gene Impact on mouse chromosome 18. Proc. Natl. Acad. Sci. U. S. A. 94, 9249–9254.
Hatada, I., Morita, S., Obata, Y., Sotomaru, Y., Shimoda, M., and Kono, T. (2001). Identification of a new imprinted gene, Rian, on mouse chromosome 12 by fluorescent differential display screening. J. Biochem. 130, 187–190.
Hayashizaki, Y., Shibata, H., Hirotsune, S., Sugino, H., Okazaki, Y., Sasaki, N., Hirose, K., Imoto, H., Okuizumi, H., Muramatsu, M., et al. (1994). Identification of an imprinted U2af binding protein related sequence on mouse chromosome 11 using the RLGS method. Nat. Genet. 6, 33–40.
Huang, H.S., Yoon, B.J., Brooks, S., Bakal, R., Berrios, J., Larsen, R.S., Wallace, M.L., Han, J.E., Chung, E.H., Zylka, M.J., et al. (2014). Snx14 regulates neuronal excitability, promotes synaptic transmission, and is imprinted in the brain of mice. PLoS One 9.
Jacobs, A.T., and Marnett, L.J. (2009). HSF1-mediated BAG3 expression attenuates apoptosis in 4-hydroxynonenal-treated colon cancer cells via stabilization of anti-apoptotic Bcl-2 proteins. J. Biol. Chem. 284, 9176–9183.
Jiang, H., and Wong, W.H. (2009). Statistical inferences for isoform expression in RNA-Seq. Bioinformatics 25, 1026–1032.
Jiao, S., and Li, Z. (2011). Nonapoptotic Function of BAD and BAX in Long-Term Depression of Synaptic Transmission. Neuron 70, 758–772.
Kagitani, F., Kuroiwa, Y., Wakana, S., Shiroishi, T., Miyoshi, N., Kobayashi, S., Nishida, M., Kohda, T., Kaneko-Ishino, T., and Ishino, F. (1997). Peg5/Neuronatin is an imprinted gene located on sub-distal chromosome 2 in the mouse. Nucleic Acids Res. 25, 3428–3432.
Kaneko-Ishino, T., Kuroiwa, Y., Miyoshi, N., Kohda, T., Suzuki, R., Yokoyama, M., Viville, S., Barton, S.C., Ishino, F., and Surani, M.A. (1995). Peg1/Mest imprinted gene on chromosome 6 identified by cDNA subtraction hybridization. Nat. Genet. 11, 52–59.
Karolchik, D., Barber, G.P., Casper, J., Clawson, H., Cline, M.S., Diekhans, M., Dreszer, T.R., Fujita, P.A., Guruvadoo, L., Haeussler, M., et al. (2014). The UCSC Genome Browser database: 2014 update. Nucleic Acids Res. 42.

Kawakami, T., Chano, T., Minami, K., Okabe, H., Okada, Y., and Okamoto, K. (2006). Imprinted DLK1 is a putative tumor suppressor gene and inactivated by epimutation at the region upstream of GTL2 in human renal cell carcinoma. Hum. Mol. Genet. 15, 821–830.

Kelsey, G., and Bartolomei, M.S. (2012). Imprinted Genes … and the Number Is? PLoS Genet. 8, e1002601.

Kobayashi, H., Yamada, K., Morita, S., Hiura, H., Fukuda, A., Kagami, M., Ogata, T., Hata, K., Sotomaru, Y., and Kono, T. (2009). Identification of the mouse paternally expressed imprinted gene Zdbf2 on chromosome 1 and its imprinted human homolog ZDBF2 on chromosome 2. Genomics 93, 461–472.

Kulinski, T.M., Barlow, D.P., and Hudson, Q.J. (2013). Imprinted silencing is extended over broad chromosomal domains in mouse extra-embryonic lineages. Curr. Opin. Cell Biol. 25, 297–304.

Kuzmin, A., Han, Z., Golding, M.C., Mann, M.R.W., Latham, K.E., and Varmuza, S. (2008). The PcG gene Sfmbt2 is paternally expressed in extraembryonic tissues. Gene Expr. Patterns 8, 107–116.

Li, Z., Jo, J., Jia, J.M., Lo, S.C., Whitcomb, D.J., Jiao, S., Cho, K., and Sheng, M. (2010). Caspase-3 activation via mitochondria is required for long-term depression and AMPA receptor internalization. Cell 141, 859–871.

Lindley, D., and Smith, A. (1972). Bayes estimates for the linear model. J. R. Stat. Soc. Ser. B (… 5.

Luedi, P.P., Hartemink, A.J., and Jirtle, R.L. (2005). Genome-wide prediction of imprinted murine genes. Genome Res. 15, 875–884.

McCole, R.B., and Oakey, R.J. (2008a). Unwitting hosts fall victim to imprinting. Epigenetics 3, 258–260.

McCole, R.B., and Oakey, R.J. (2008b). Unwitting hosts fall victim to imprinting. Epigenetics 3, 258–260.

Menheniott, T.R., Woodfine, K., Schulz, R., Wood, A.J., Monk, D., Giraud, A.S., Baldwin, H.S., Moore, G.E., and Oakey, R.J. (2008). Genomic imprinting of Dopa decarboxylase in heart and reciprocal allelic expression with neighboring Grb10. Mol. Cell. Biol. 28, 386–396.

Mituyama, T., Yamada, K., Hattori, E., Okida, H., Ono, Y., Terai, G., Yoshizawa, A., Komori, T., and Asai, K. (2009). The functional RNA database 3.0: Databases to support mining and annotation of functional RNAs. Nucleic Acids Res. 37.

Miyoshi, N., Kuroiwa, Y., Kohda, T., Shitara, H., Yonekawa, H., Kawabe, T., Hasegawa, H., Barton, S.C., Surani, M.A., Kaneko-Ishino, T., et al. (1998). Identification of the Meg1/Grb10 imprinted gene on mouse proximal chromosome 11, a candidate for the Silver-Russell syndrome gene. Proc. Natl. Acad. Sci. U. S. A. 95, 1102–1107.

Mizuno, Y., Sotomaru, Y., Katsuzawa, Y., Kono, T., Meguro, M., Oshimura, M., Kawai, J., Tomaru, Y., Kiyosawa, H., Nikaido, I., et al. (2002). Asb4, Ata3, and Dcn Are Novel Imprinted Genes Identified by High-Throughput Screening Using RIKEN cDNA Microarray. Biochem. Biophys. Res. Commun. 290, 1499–1505.

Morison, I.M., Ramsay, J.P., and Spencer, H.G. (2005). A census of mammalian imprinting. Trends Genet. 21, 457–465.

Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat. Methods 5, 621–628.

Nikaido, I., Saito, C., Mizuno, Y., Meguro, M., Bono, H., Kadomura, M., Kono, T., Morris, G.A., Lyons, P.A., Oshimura, M., et al. (2003). Discovery of imprinted transcripts in the mouse transcriptome using large-scale expression profiling. Genome Res. 13, 1402–1409.

Ono, R., Shiura, H., Aburatani, H., Kohda, T., Kaneko-Ishino, T., and Ishino, F. (2003). Identification of a large novel imprinted gene cluster on mouse proximal chromosome 6. Genome Res. 13, 1696–1705.

Peters, J. (2014). The role of genomic imprinting in biology and disease: an expanding view. Nat. Rev. Genet.

Peters, J., and Williamson, C. (2007). Control of imprinting at the Gnas cluster. Epigenetics.

Piras, G., El Kharroubi, A., Kozlov, S., Escalante-Alcalde, D., Hernandez, L., Copeland, N.G., Gilbert, D.J., Jenkins, N.A., and Stewart, C.L. (2000). Zac1 (Lot1), a Potential Tumor Suppressor Gene, and the Gene for varepsilon -Sarcoglycan Are Maternally Imprinted Genes: Identification by a Subtractive Screen of Novel Uniparental Fibroblast Lines. Mol. Cell. Biol. 20, 3308–3315.

Puthalakath, H. (2001). Bmf: A Proapoptotic BH3-Only Protein Regulated by Interaction with the Myosin V Actin Motor Complex, Activated by Anoikis. Sci. (New York, NY) 293, 1829–1832.

Raefski, A.S., and O'Neill, M.J. (2005). Identification of a cluster of X-linked imprinted genes in mice. Nat. Genet. 37, 620–624.

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

Roberts, A., and Pachter, L. (2013). Streaming fragment assignment for real-time analysis of sequencing experiments. Nat. Methods 10, 71–73.

Romá-Mateo, C., Sanz, P., and Gentry, M. (2012). Deciphering the role of malin in the lafora progressive myoclonus epilepsy. IUBMB Life.

Royo, J.L., Hidalgo, M., and Ruiz, A. (2007). Pyrosequencing protocol using a universal biotinylated primer for mutation detection and SNP genotyping. Nat. Protoc. 2, 1734–1739.

Ruf, N., Dünzinger, U., Brinckmann, A., Haaf, T., Nürnberg, P., and Zechner, U. (2006). Expression profiling of uniparental mouse embryos is inefficient in identifying novel imprinted genes. Genomics 87, 509–519.

Ruf, N., Bahring, S., Galetzka, D., Pliushch, G., Luft, F.C., Nurnberg, P., Haaf, T., Kelsey, G., and Zechner, U. (2007). Sequence-based bioinformatic prediction and QUASEP identify genomic imprinting of the KCNK9 potassium channel gene in mouse and human. Hum. Mol. Genet. 16, 2591–2599.

Sandell, L.L., Guan, X.-J., Ingram, R., and Tilghman, S.M. (2003). Gatm, a creatine synthesis enzyme, is imprinted in mouse placenta. Proc. Natl. Acad. Sci. U. S. A. 100, 4622–4627.

Schmidt, J. V, Matteson, P.G., Jones, B.K., Guan, X.-J., and Tilghman, S.M. (2000). The Dlk1 and Gtl2 genes are linked and reciprocally imprinted. Genes Dev. 14, 1997–2002.

Schulz, R., Menheniott, T.R., Woodfine, K., Wood, A.J., Choi, J.D., and Oakey, R.J. (2006). Chromosome-wide identification of novel imprinted genes using microarrays and uniparental disomies. Nucleic Acids Res. 34, e88.

Solter, D. (1988). Differential imprinting and expression of maternal and paternal genomes. Annu. Rev. Genet. 22, 127–146.

Steijger, T., Abril, J.F., Engström, P.G., Kokocinski, F., Akerman, M., Alioto, T., Ambrosini, G., Antonarakis, S.E., Behr, J., and Bertone, P. (2013). Assessment of transcript reconstruction methods for RNA-seq. Nat. Methods 10, 1177–1184.

Takada, H., Imoto, I., Tsuda, H., Nakanishi, Y., Ichikura, T., Mochizuki, H., Mitsufuji, S., Hosoda, F., Hirohashi, S., Ohki, M., et al. (2005). ADAM23, a possible tumor suppressor gene, is frequently silenced in gastric cancers by homozygous deletion or aberrant promoter hypermethylation. Oncogene 24, 8051–8060.

Trapnell, C., Hendrickson, D.G., Sauvageau, M., Goff, L., Rinn, J.L., and Pachter, L. (2013). Differential analysis of gene regulation at transcript resolution with RNA-seq. Nat. Biotechnol. 31, 46–53.

Triplett, J.W., and Pavalko, F.M. (2006). Disruption of -actinin-integrin interactions at focal adhesions renders osteoblasts susceptible to apoptosis. AJP Cell Physiol. 291, C909–C921.

Tsai, P.T., Hull, C., Chu, Y., Greene-Colozzi, E., Sadowski, A.R., Leech, J.M., Steinberg, J., Crawley, J.N., Regehr, W.G., and Sahin, M. (2012). Autistic-like behaviour and cerebellar dysfunction in Purkinje cell Tsc1 mutant mice. Nature 488, 647–651.

Turro, E., Su, S.-Y., Gonçalves, Â., Coin, L.J., Richardson, S., and Lewin, A. (2011). Haplotype and isoform specific expression estimation using multi-mapping RNA-seq reads. Genome Biol. 12, R13.

Turro, E., Astle, W.J., and Tavaré, S. (2014). Flexible analysis of RNA-seq data using mixed effects models. Bioinformatics 30, 180–188.

Vijaya Satya, R., Zavaljevski, N., and Reifman, J. (2012). A new strategy to reduce allelic bias in RNA-Seq readmapping. Nucleic Acids Res. 40.

Wagner, G.P., Kin, K., and Lynch, V.J. (2012). Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. Theory Biosci. 131, 281–285.

Wang, H., and Elbein, S.C. (2007). Detection of allelic imbalance in gene expression using pyrosequencing. Methods Mol. Biol. 373, 157–176.

Wang, S.S.-H., Kloth, A.D., and Badura, A. (2014). The Cerebellum, Sensitive Periods, and Autism. Neuron 83, 518–532.

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

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

Wang, X., Soloway, P.D., and Clark, A.G. (2011a). A Survey for Novel Imprinted Genes in the Mouse Placenta by mRNA-seq. Genetics 189, 109–122.

Wang, X., Soloway, P.D., and Clark, A.G. (2011b). A Survey for Novel Imprinted Genes in the Mouse Placenta by mRNA-seq. Genetics 189, 109–122.

Weinstein, L.S., Yu, S., and Ecelbarger, C.A. (2000). Variable imprinting of the heterotrimeric G protein G(s) alpha-subunit within different segments of the nephron. Am. J. Physiol. Renal Physiol. 278, F507–F514.

Xie, W., Barr, C., Kim, A., Yue, F., and Lee, A. (2012). Base-resolution analyses of sequence and parent-of-origin dependent DNA methylation in the mouse genome. Cell.

Chapter Four

Temporal and Spatial Regulation of Genomic Imprinting in the Brain

The content of this chapter is part of a manuscript in preparation Perez, J.D., N.D. Rubinstein, D.E. Fernandez, L.A. Needleman, S.W. Santoro, J.J Choi, J.S. Liu, C. Dulac, 2015. New Insights Into Parent-of Origin Expression in the Adult and Developing Brain.

Contributions: Perez, J.D. performed the RNA-seq experiments, pyrosequencing experiments, helped calibrate the statistical model to analyze the data presented on this chapter, and interpreted the data. N.D. Rubinstein processed the RNA-seq data, interpreted the data and together with D.E. Fernandez developed the statistical models used in the analysis of the RNA-seq data presented on this chapter. L.A. Needleman, and S.W. Santoro conducted experiments that are not part of this thesis but are part of the manuscript. J.J. Choi contributed to experiments that will be presented in chapter four. J.S. Liu provided guidance and supervision for the statistical models used in the analysis of the RNA-seq data presented on this chapter. C. Dulac, conceived the project, interpreted the data, and provided guidance and supervision.

Introduction

The developing and mature brain express a high number of imprinted genes and brain function is compromised in a number of imprinting-associated conditions (Gregg et al., 2010a; Keverne and Curley, 2008; Wilkinson et al., 2007a).A classical example of imprinted-related disease, the Prader-Willi syndrome, is caused by loss of paternal expression in the q11-13 region of human chromosome 15, resulting in abnormal development, hyperphagia, mental retardation, and volatile behavior (Peters, 2014). Angelman syndrome (AS) is caused by loss of maternal expression at the same genomic region, resulting in mental retardation, impaired speech, and a joyful demeanor (Peters, 2014). Another condition, the Birk-Barel mental retardation syndrome (Barel et al., 2008), is caused by mutations in the human maternally expressed Kcnk9 gene. In mice, experimental work has also highlighted a particular influence of genomic imprinting on behavior. For examples, paternal deletion of the Grb10 gene increases social dominance (Garfield et al., 2011a) and paternal deletions of the Mest and Peg3 loci result in deficient maternal behaviors (Lefebvre et al., 1998; Li et al., 1999). Interestingly, a large portion of conditions or phenotypes associated with imprinting exhibit specific impairments of social behavior (Isles et al., 2006). To date, the neuronal underpinnings of these conditions remain largely obscure. From a neurobiology standpoint, deciphering the specific roles of genomic imprinting in the brain can help better understand regulatory mechansims underlying brain development, physiology, and complex behaviors. From a genetic standpoint, the study of genomic imprinting in the brain is advantageous, since this phenomenon targets many neural genes and associated biological functions.

It is important to consider that, irrespective of their imprinted status, the expression levels of imprinted genes is not uniform across the brain but rather are known to vary significantly among brain regions (Gregg et al., 2010a). In addition, expression levels of imprinted genes has also been shown to vary throughout brain development (Wilkinson et al., 2007a). These expression-level differences may be associated with changes in the magnitude of the parental-expression bias. Indeed, developmental- and brain region-specific imprinting has been reported for some genes. For example, the AS associated Ube3a gene is maternally expressed in the hippocampus and cerebellar Purkinje Cells, while bialleic expression is observed in the anterior commissure and optic chiasm (Albrecht et al., 1997). Furthermore, Ube3a shows significant paternal expression in the cortex immediately after birth that becomes biallelic by postnatal day 14 (Sato and Stryker, 2010). Therefore, it is possible that imprinting regulation of many other genes varies across brain regions and among different CNS cell types. Since it is not known whether all mouse imprinted genes have been discovered, studying imprinting in different brain regions and across different time points is of high interest. Further examples of the regulation and functions of imprinted genes in the brain are described in chapter 1.

As described in chapter 3, we have performed a genome-wide screen for imprinted genes in the cerebellum. Our experimental design is comprised of 48 individual cerebellums capturing two different developmental stages and representing both sexes. We observed 114 genes under imprinting regulation in the cerebellum, 74 of which have been previously described, and 40 of which were as confirmed imprinted using pyrosequencing.

In this chapter, we have investigated the regulation of imprinted genes in both the time and spatial axis. First we asked whether imprinting expression is regulated according to the developmental stage of the cerebellum. Interestingly, when the repertoire and expression of

imprinted genes are compared between P8 and P60 cerebellum, half of the identified genes appear to be regulated according to age, either in their parental biases, in their overall level of gene expression, or both. For some genes, the developmental regulation is restricted to the preferentially expressed allele, while for others it is restricted to the non-preferentially expressed allele. Second, we asked if the parental biases of a selected group of imprinted genes are regulated across the different tissues. Using pyrosequencing for allele specific quantification we find that a substantial number of genes are uniquely imprinted in the brain. We then reveal considerable variation of the magnitude of the parental bias of many of these imprinted genes across different brain regions. Strikingly, genes from same as well as independent imprinted clusters seem to share similar patterns of variation, suggesting coordinated regulation. Thus, we uncover a highly dynamic mode of imprinting regulation according to tissue and/or developmental stage, where a unique program of imprinting is manifested in the brain and can be further modified according to brain region and developmental stage.

RESULTS

As described in chapter 3 we have sequenced RNA from 48 individual cerebellums from the offspring of reciprocal crosses, half of which were collected at postnatal day 8 (P8), a period in which granule cells are highly multiplied and are migrating to the inner granule layer. The other half of cerebellums were collected from adult animals (P60). Each age group had equal numbers of males and females. Besides its ability to quantify significant preferential expression from one of the parental alleles, our newly developed bayesian regression model also allows the identification of significant effects of age on parental biases and on overall expression levels.

Developmental Regulation of Genomic Imprinting in the Cerebellum

The importance of genomic imprinting during development, particularly in the developing brain, has been clearly established (Wilkinson et al., 2007b). Furthermore, age-dependent regulation of the expression and/or imprinted state of some genes has also been described (Gregg et al., 2010b). Here we took advantage of the cerebellar postnatal maturation to investigate the regulation of imprinting during important milestones of neuronal development. We detected 57 imprinted genes (50% of all imprinted genes expressed in the cerebellum) for which either the parental bias and/or the total expression (paternal + maternal) level are regulated according to developmental stage (age effect PP > 0.95; **Figure 4.1A**). This includes 11 genes in which both the parental bias and the total expression level are affected by age, 17 genes in which only the parental bias is affected by age, and 29 genes in which only the total expression level changes with age. Among those, the parental bias in 21 genes remains age invariant and eight are not expressed in the adult. Two striking patterns are apparent among the age-regulated genes. First, in a disproportionally high number of genes, both the parental bias and the total expression level

are stronger in P8 cerebellum (21 genes versus 9 in which both the parental bias and the total expression level are higher in P60, and 11 and 9 genes in which either the parental bias is higher in P8 and total expression level higher in P60 or the opposite, respectively, P-value = 0.048; χ^2 test). The cerebellum undergoes important developmental milestones at P8. It is therefore possible that these changes in expression and parental bias indicate an active involvement of these genes in this process. Secondly, the age effect on the magnitude of the parental bias and on the level of total expression are positively correlated (Pearson correlation coefficient = 0.34; P-value = 3.1×10^{-4} **Figure 4.1B**). Such a pattern may artificially arise if the power to detect parental biases is strongly correlated with expression levels. This option, however, is not supported in our data (data not shown). Alternatively, it is possible that either the preferentially expressed allele and/or the non-preferentially expressed allele experience a significant change in expression levels along development, thereby altering both the magnitude of the parental bias as well as the level of the total expression. To test this hypothesis we fitted our model to the data where we defined the response as either the paternal expression levels or the maternal expression levels. This analysis indeed revealed that age regulated imprinting (age effect PP > 0.95) is achieved either by a significant change in the expression level of the preferred allele (31 out of 57 genes, shown along the X-axis in **Figure 4.1C**), a significant change in the expression level of the non-preferred allele (13 out of 57 genes, shown along the Y-axis in **Figure 4.1C**), or a significant change in the expression levels of both alleles (13 out of 57 genes, shown along the diagonal in **Figure 4.1C**), indicating that altering the expression level of the preferred allele is the common mode through which age regulated imprinting is achieved (P-value = 0.003; χ^2 test).

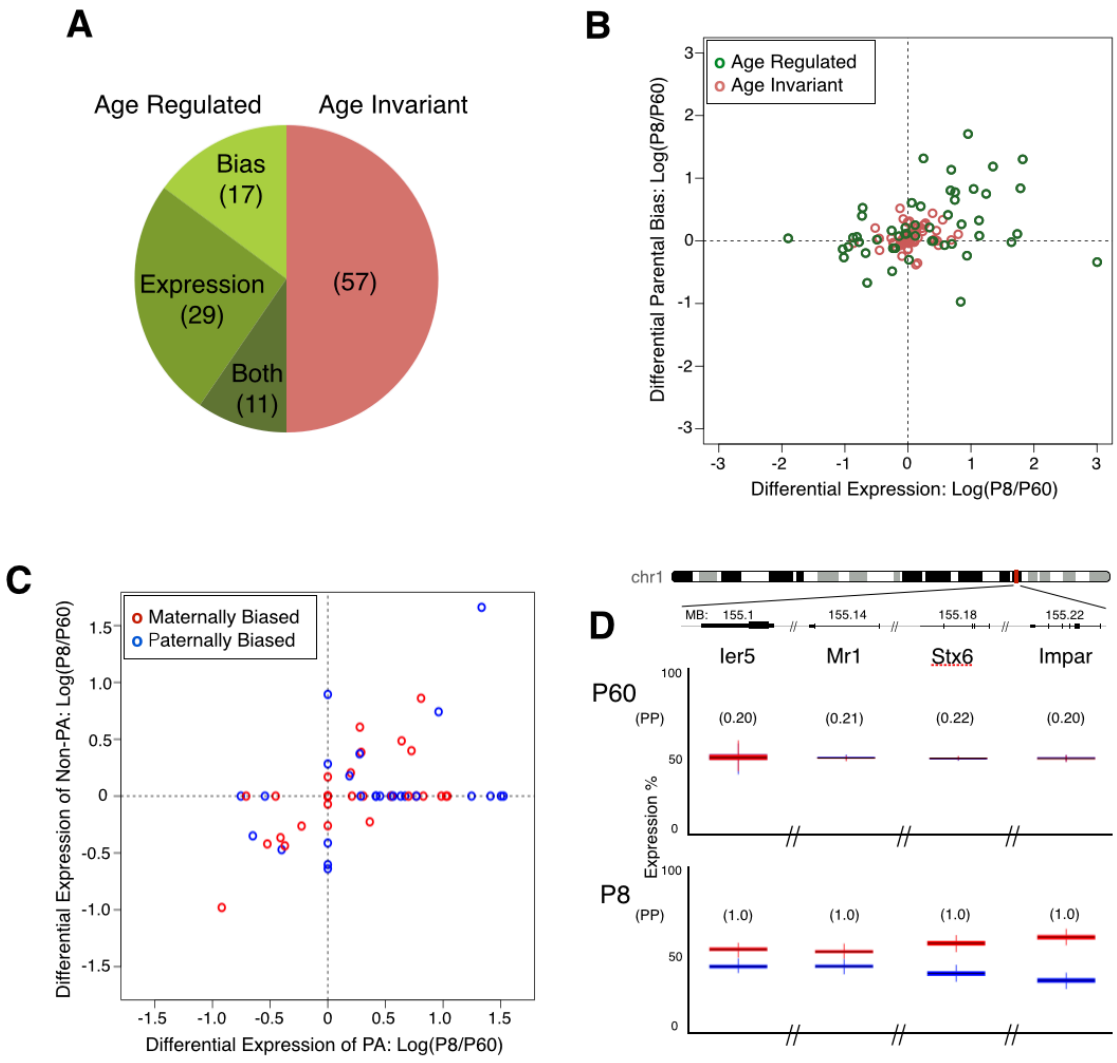

Figure 4.1: Age Regulation of Imprinted Genes in the Cerebellum

(A) Proportions of imprinted genes in the cerebellum regulated by age. **(B)** Comparison of the age regulation in the overall expression of imprinted genes and in their parental biases reveals positive correlation. **(C)** Comparison of the age regulation in the expression of the preferred and the non-preferred allele of imprinted genes. **(D)** Novel imprinted cluster of distal chromosome 1 exhibits is regulated according to age.

Several genes for which the parental bias in the cerebellum is affected by age are associated with developmental processes such as cell proliferation, differentiation, and survival. For instance the Asb4 gene, which regulates embryonic stem-cell differentiation (Townley-Tilson et al., 2014), exhibits a strong maternally-biased expression during cerebellar development but is biallelicaly expressed during adulthood. This change in parental bias is achieved by a significant decrease in maternal expression and an increase in paternal expression, which interestingly, maintains the total level of expression approximately constant (**Figure 4.2**). The growth suppressor Grb10 gene, on the other hand, exhibits biallelic expression at P8 but exclusive paternal expression in the adult, which is achieved by suppression of the maternal allele (**Figure 4.2**). Strikingly, we observe a switch in the parental bias for the transcription factor Zim1, from maternal during development to paternal bias in the adult cerebellum, which is achieved by reduction in the maternal expression level from above the paternal expression level to below it (**Figure 4.2**). Remarkably, we uncovered a novel imprinted locus at the distal end of chromosome 1, which exhibits age dependent regulation. The genes Ier5, Mr1, Stx6, and the putative BC034090 gene, which we name here Impar (for Imprinted and Age Regulated) are located side by side within 136 KB all show a maternal bias during cerebellum development but biallelic expression in the adult (**Figure 4.1D**). For all genes in the locus, this shift in parental bias is achieved by a reduction in the expression level of the maternal allele and to a lesser extent in the paternal allele. It is interesting to note that Stx6 is believed to regulate neuronal migration and formation of processes (Kabayama et al., 2008; Tiwari et al., 2011), two events necessary for the integration of granule cells to the cerebellar circuit occurring at the P8 stage. It would be therefore interesting to determine whether the imprinting regulation of this gene affects this

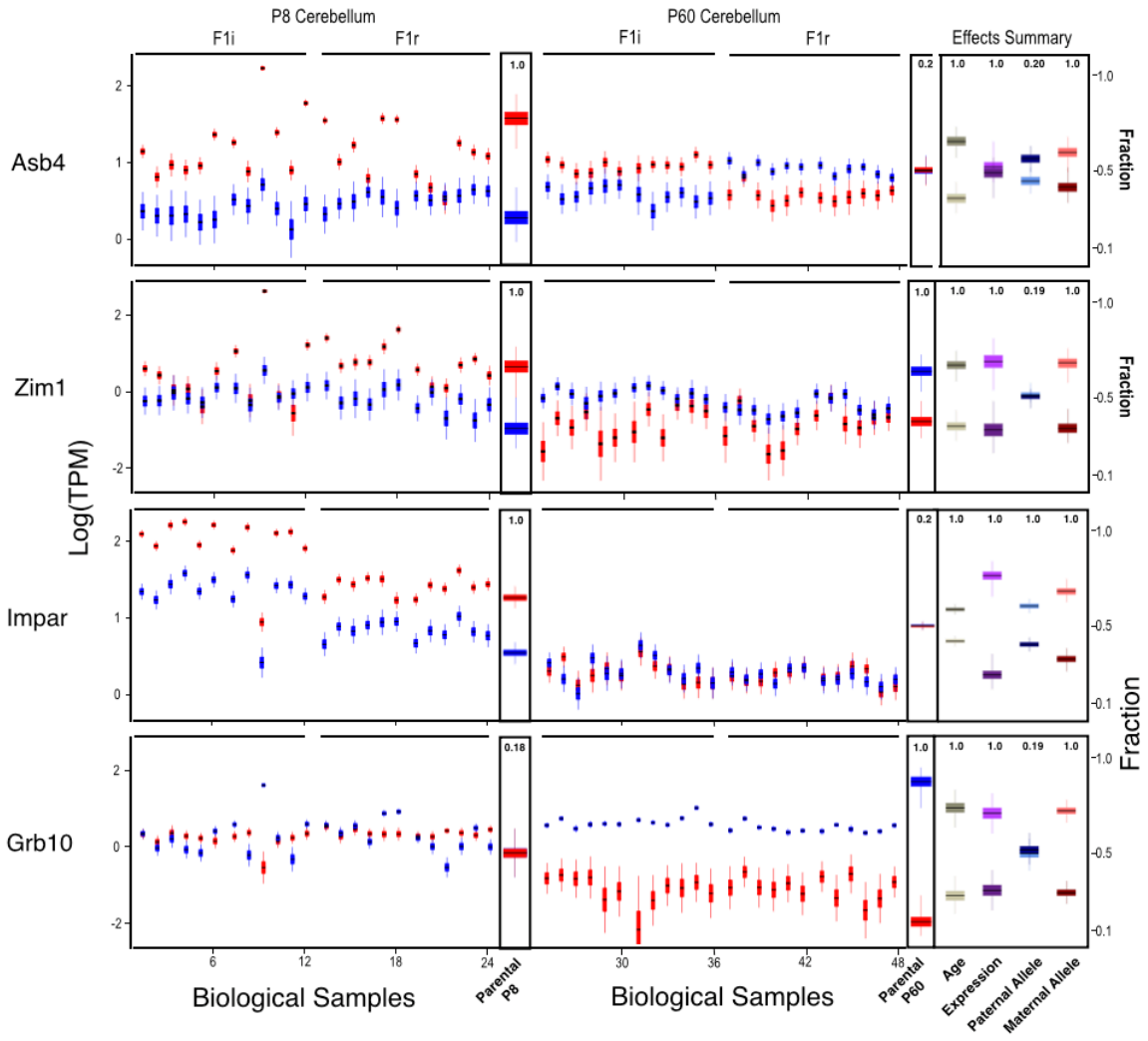

Figure 4.2: Imprinted Genes Regulated According to Age

Asb4 is maternally biased during development but biallelically expressed in the adult. Still the overall expression of the gene is maintained in both ages. Only the maternal allele of Zim1 is regulated according to age, while the paternal allele is maintain at a baseline level. Both alleles of Impar are regulated during development, however the maternal allele is express higher at P8. Decrease expression of Grb10 is mediated by downregulation of the non-preferentially expressed allele (maternal allele) while the paternal allele is expressed at a constant level. For each replicate red indicates maternal expression while blue indicates paternal expression. Y-axis of RNA-seq result is in the log of transcript per million units. Effects summary, shows the average effect on tested factors. For age bright olive green represents P8 and dark olive green represents P60.

process. Finally, we also observe age effects on the parental biases of specific isoforms of Herc3, Mest, and H13, which all show isoform-specific imprinting (data not shown).

Spatial Regulation of Genomic Imprinting

Tissue-dependent regulation has been described for several imprinted genes (Gregg et al., 2010a; Prickett and Oakey, 2012). We decided to further investigate the extent of tissue-specific imprinting in the brain by quantifying the parentally-biased expression of 28 imprinted genes (20 known and 8 novel) by pyrosequencing analysis of 16 brain macro-regions and seven non-brain tissues, all from adult animals (**Figure 4.3A** and Methods). This analysis revealed several striking patterns of parental bias across different genes and tissues (**Figure 4.3B, 4.4B**).

Igf2, which is maternally biased in the brain (Gregg et al., 2010b) but paternally biased outside the brain stands in contrast to the brain paternally biased and body maternally biased Grb10 (Charalambous et al., 2003, 2010; Garfield et al., 2011b). The Igf2r gene, which during developmental periods is maternally expressed and exerts a function that is antagonistic to that of the paternally expressed Igf2 (Filson et al., 1993; Ludwig et al., 1996), maintains a strong maternal bias in the body, which is relaxed in most of the brain (**Figure 4.3B, 4.4B**). The fact that these three genes regulate growth in the tissues in which they are expressed, suggests different contributions from the parental genomes to the brain and body. In addition, the Kcnk9 gene is robustly maternally biased throughout the brain whereas in the body it is either not expressed, biallelicaly expressed, or paternally biased. Finally, the Zim1 gene shows an intriguing pattern of both paternal and maternal biases both in the brain and in the body.

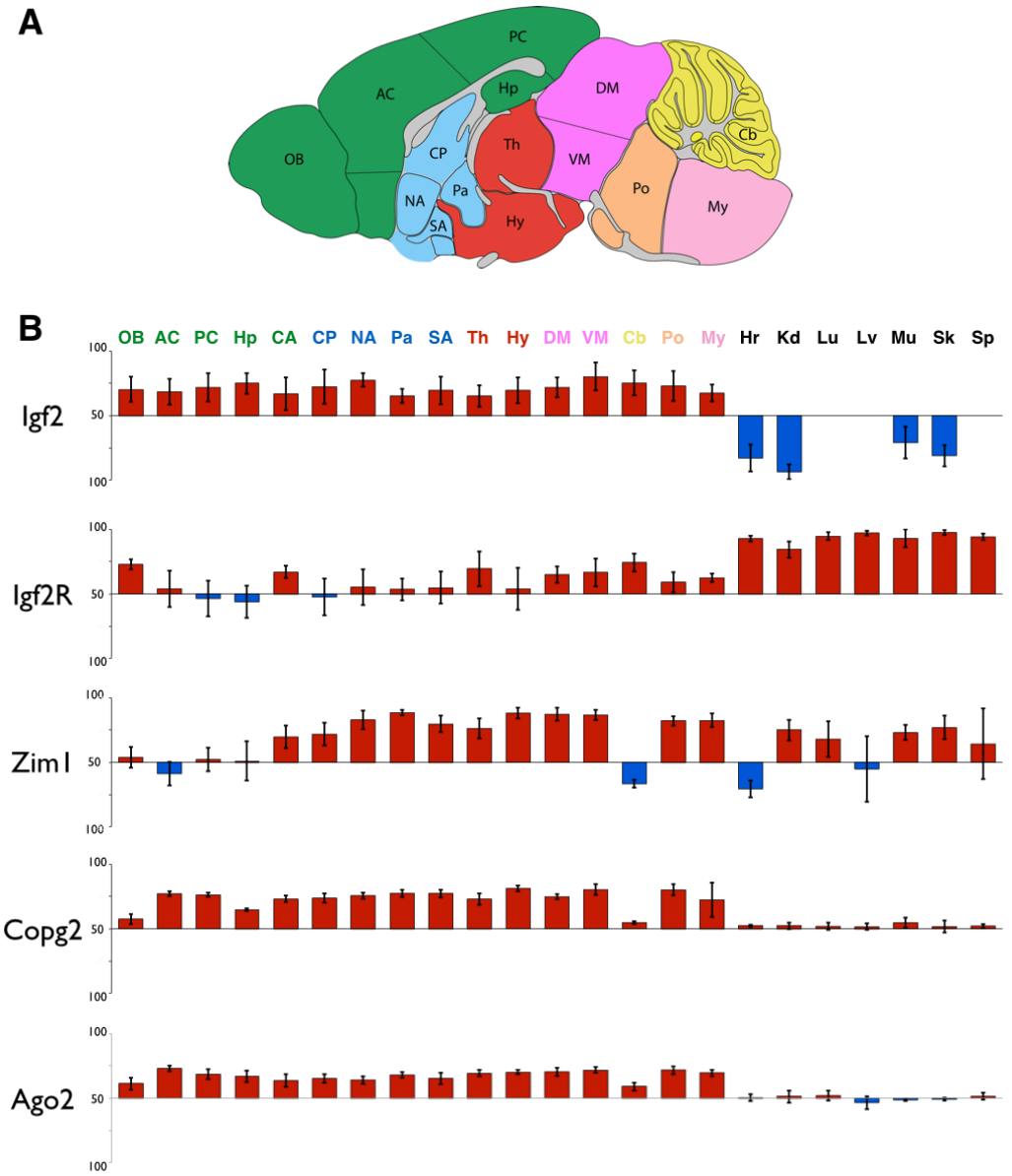

Figure 4.3: Spatial regulation of Parental Biases

(A) Legend of brain regions analyzed colored accordingly to broad developmental relatedness of the regions. OB: Olfactory bulb, AC: Anterior Cortex, PC: Posterior Cortex, Hp: Hippocampus, CP: Caudate Putamen, NA: Nucleus Accumbens, Pa: Pallidum, SA: Striatiu-like Amygdala, Th: Thalamus, Hy: Hypothalamus, DM: Dorsal Midbrain, VM: Ventral Midbrain, Cb: Cerebellum, Po: Pons, My: Medulla. **(B)**Examples of genes whose parental bias is regulated according to organ (brain vs body tissues) and of genes whose parental bias is regulated within the brain. Origin of bar graphs represents biallelic expression positive values represent preferential maternal expression (colored red) while negative values represent preferential expression of the paternal allele (colored blue).

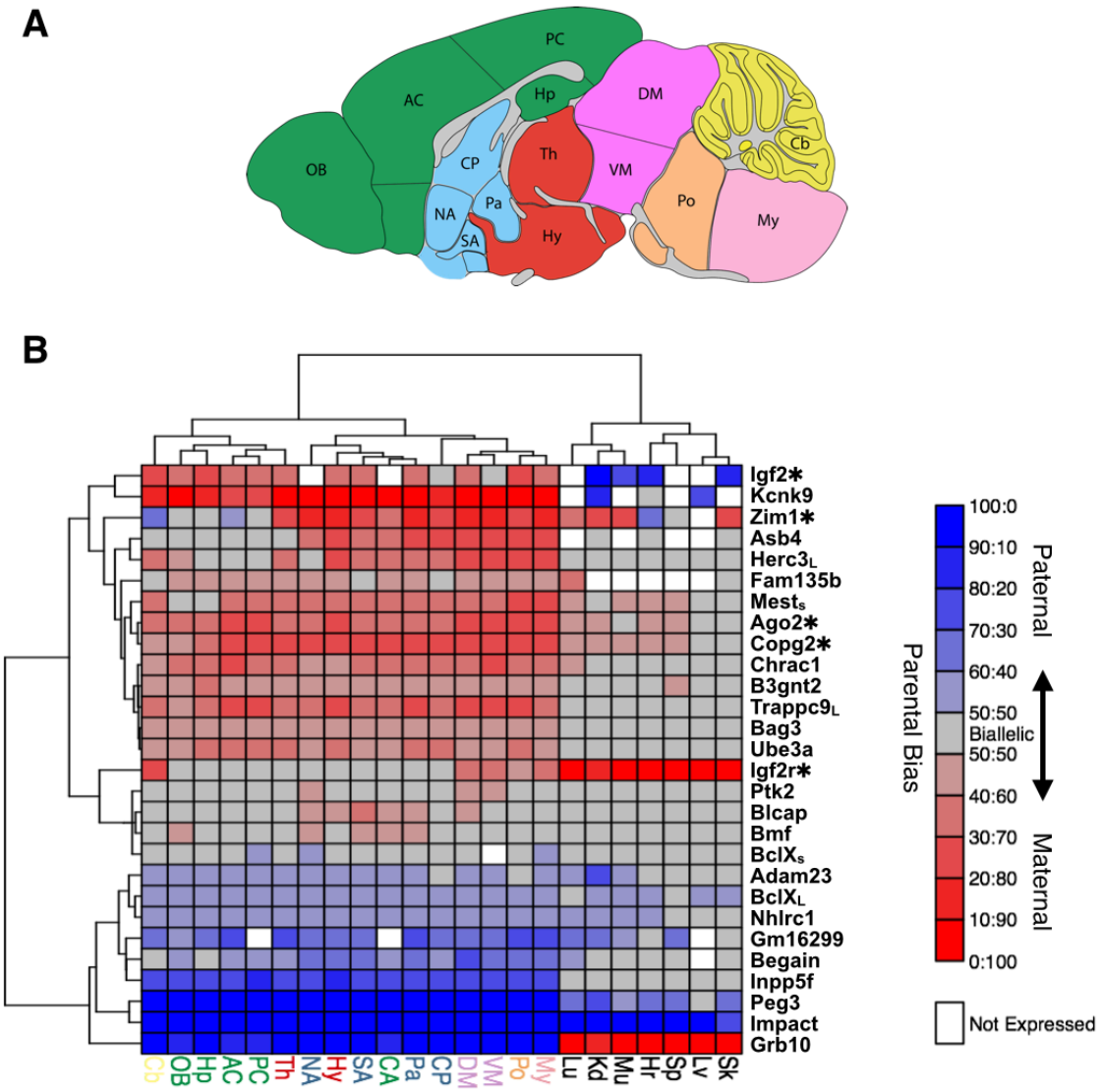

Figure 4.4: Hierarchical clustering based on the parental biases of the Imprinted Genes

(**A)** Legend of brain regions described in figure 4.3 explanations. (**A)** Hierarchical cluster based on deviations from biallelic expression in imprinted genes.

To determine associations in the spatial regulation of imprinted genes we performed a hierarchical clustering analysis based on the parental biases of tested genes. Imprinted genes clearly cluster into three main clades, one that includes maternally biased genes, one that includes paternally biased genes, and one that includes genes that are sporadically biased in the brain, most of which are maternally biased. Several genes exhibit sharp contrasts in their parental bias between the brain and body. This includes genes which are exclusively or nearly exclusively biased in and throughout the brain, such as the maternally biased Ube3a (Rougeulle et al., 1997), Trappc9, Bag3, and B3gnt2 and the paternally biased BclX long isoform (Bcl2l1L), Inpp5f (Choi et al., 2005), and Begain.

Moreover, parental biases of imprinted genes clearly separate tissues into two main clades, non-brain and brain, where in the latter the parental bias is much more consistent and robust. The brain is further subdivided into additional sub-clades, which roughly group developmentally related regions. One sub-clade clusters most of the telencephalon, the thalamus, and cerebellum. The other sub-clade is further split into a clade that groups mesencephalic and rhombencephalic regions and a clade that groups diencephalic and basal ganglia regions.

In addition to the contrasting imprinting patterns between the body and the brain, the brain itself shows considerable imprinting dynamics. We therefore repeated the previous clustering analysis, this time confined to the brain (data not shown). This revealed that the Trappc9$_L$, Chrac1, and Ago2 genes, which co-localize to an imprinted cluster in the distal end of chromosome 15, exhibit a very similar pattern of maternal bias across the brain, which is stronger in the cortex and weaker in the olfactory bulb, hippocampus, and Cerebellum. The Copg2 and Mest genes, which co-localize near the centromeric region of chromosome 6, exhibit a strongly similar

pattern to that of genes on the distal end of chromosome 15 (**Figure 4.5**). The Zim1, Asb4, and Herc3 genes, which are located in the proximal end of chromosome 7, the proximal end chromosome 6, and near the centromeric region of chromosome 6, respectively, also exhibit shared patterns of biallelic expression (or in the case of Zim1 weak paternal biases) in telencephalic regions and cerebellum but strong maternal biases in other brain regions. These results suggest that the brain executes region-specific programs of imprinting. Moreover, the genes utilized in these programs are not necessarily from the same imprinted cluster, suggesting a higher order of regulation.

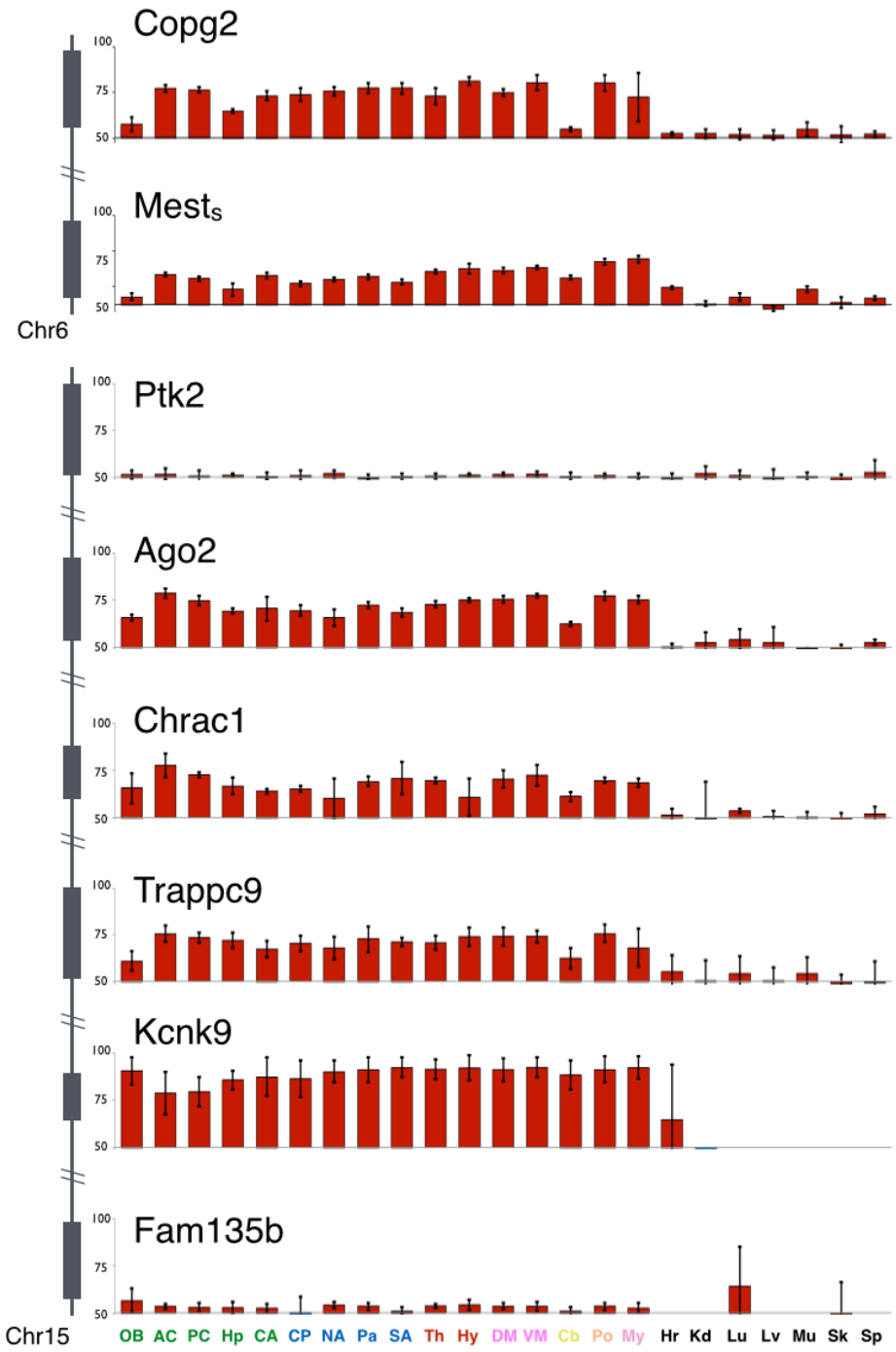

Figure 4.5: Genes within two separate cluster exhibit similar patterns of parental bias regulation across the brain.

Spatiotemporal Regulation of Genomic Imprinting

The significant spatial dynamics of imprinting in the brain begs the question of the developmental stages at which these specificities are established, as it may bare relevance to the function of the genes in the targeted tissues. To address this question we performed a pyrosequencing analysis of the parental biases of 13 genes, which were inferred to be temporally and/or spatially regulated in the previous analyses, at postnatal days 0, 8, 15, and 64 in the cortex, hypothalamus, and cerebellum, which were observed to have contrasting imprinting patterns in the previous analysis. We additionally analyzed the parental bias of these genes in the entire E15 brain. This analysis revealed substantial spatio-temporal dynamics of the parental bias of several genes (**Figure 4.6**). For instance, Blcap experiences a gradual decrease in maternal bias along the age axis, which is consistent across the three analyzed brain regions. In contrast, the switch from maternal to paternal bias of Zim1 during cerebellum development, mentioned above, occurs gradually along development in both the cerebellum and cortex, yet is not mirrored in the hypothalamus where maternal bias is strongly maintained up to adulthood. Moreover, this analysis revealed that the sharp contrast between the parental biases of Igf2 and Grb10, observed across the brain, seem to be co-temporally regulated. The switch in the expressed allele for both genes happens earlier in the cortex and hypothalamus than in the cerebellum, which roughly coincides with the completion of their development (Sillitoe and Joyner, 2007). These results demonstrate that imprinting is a remarkably dynamic process, and suggest that the observed highly coordinated spatio-temporal regulation of parent-of origin expression may in turn orchestrate development across different brain regions.

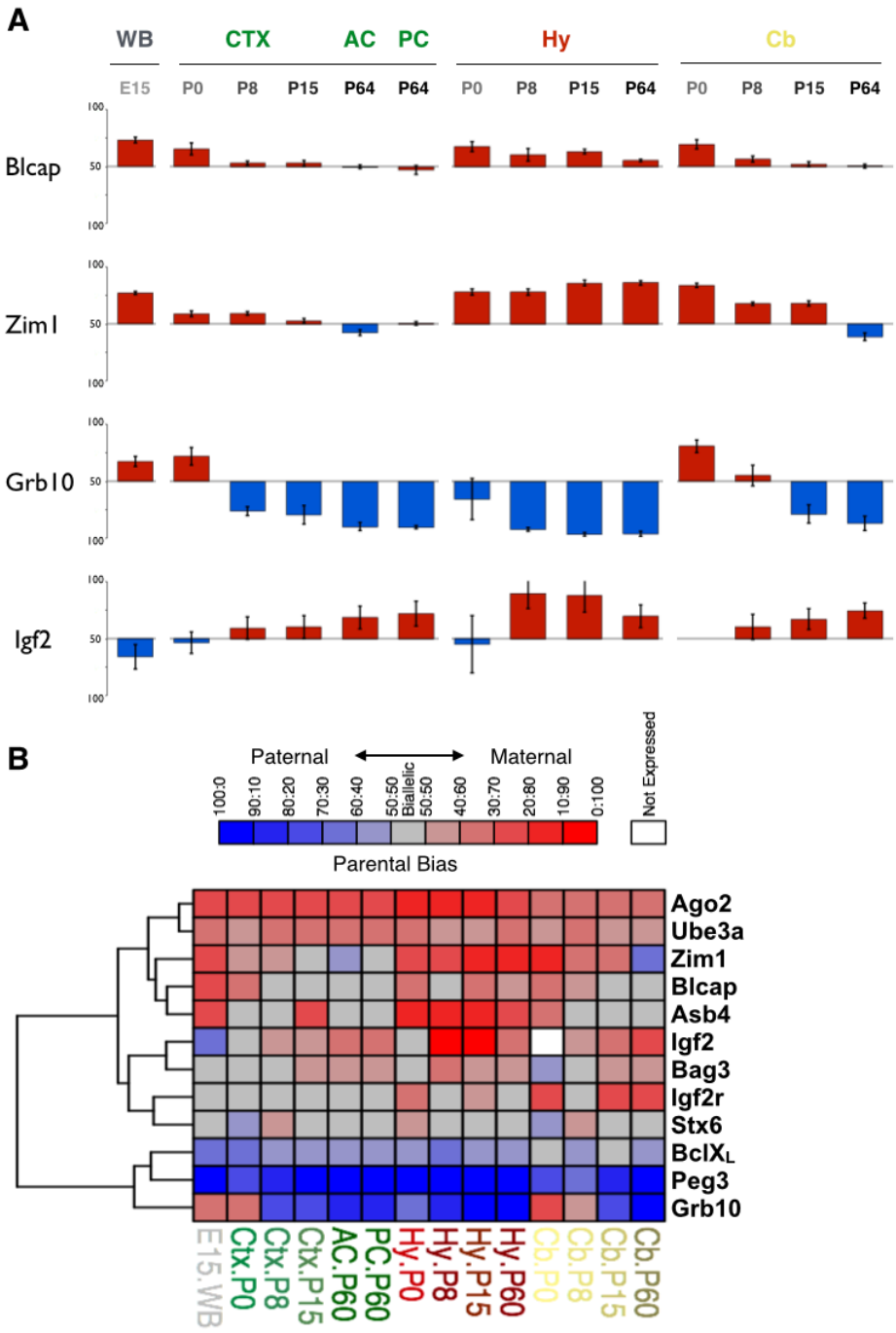

Figure 4.6: Spatiotemporal Regulation of Imprinted Genes

(A) Four genes analyzed at different timepoints in during development of the Cortex, Hypothalamus and Cerebellum. Blcap gradually loses imprinting at the same time for different brain regions. Zim1 shows selective lost and maintenance developmental maternal biases. Grb10 and Igf2 are two genes switching preferentially expressed alelle **(B)** Hierarchical cluster based on deviations from biallelic expression in imprinted genes.

Spatial regulation of Apoptosis related imprinted genes

The RNA-seq analysis revealed that the maternal bias of Bag3 substantially increases from P8 to P60 whereas $BclX_L$ shows an opposite significant decrease in paternal bias between these two time points (see chapter 3, figure 3.7B). This observation motivated us to test whether these two apoptosis genes show opposing imprinting patterns during development. We therefore tested the parental bias of these two genes in the E15 brain and in the cortex, hypothalamus, and cerebellum of postnatal days 0, 8, 15, and 64. This analysis revealed that, with the exception of the hypothalamus, the parental biases of Bag3 and $BclX_L$ show a significant anti-correlation (Pearson correlation coefficient = -0.69; P-value = 0.026, **Figure 4.7A**). A similar analysis across 16 regions of the adult brain (**Figure 4.7B**), further corroborated this finding, such that, with the expectation of the ventral midbrain and pons, the parental biases of Bag3 and $BclX_L$ show a significant anti-correlation (Pearson correlation coefficient = -0.56; P-value = 0.04, **Figure 4.9B**). These results demonstrate that the further away from biallelic expression one of the two genes is, the closer to biallelic expression the other gene is, and vice versa. This pattern is most pronounced in the cortex and cerebellum, as well as in central striatal regions, caudate putamen, and hippocampus.

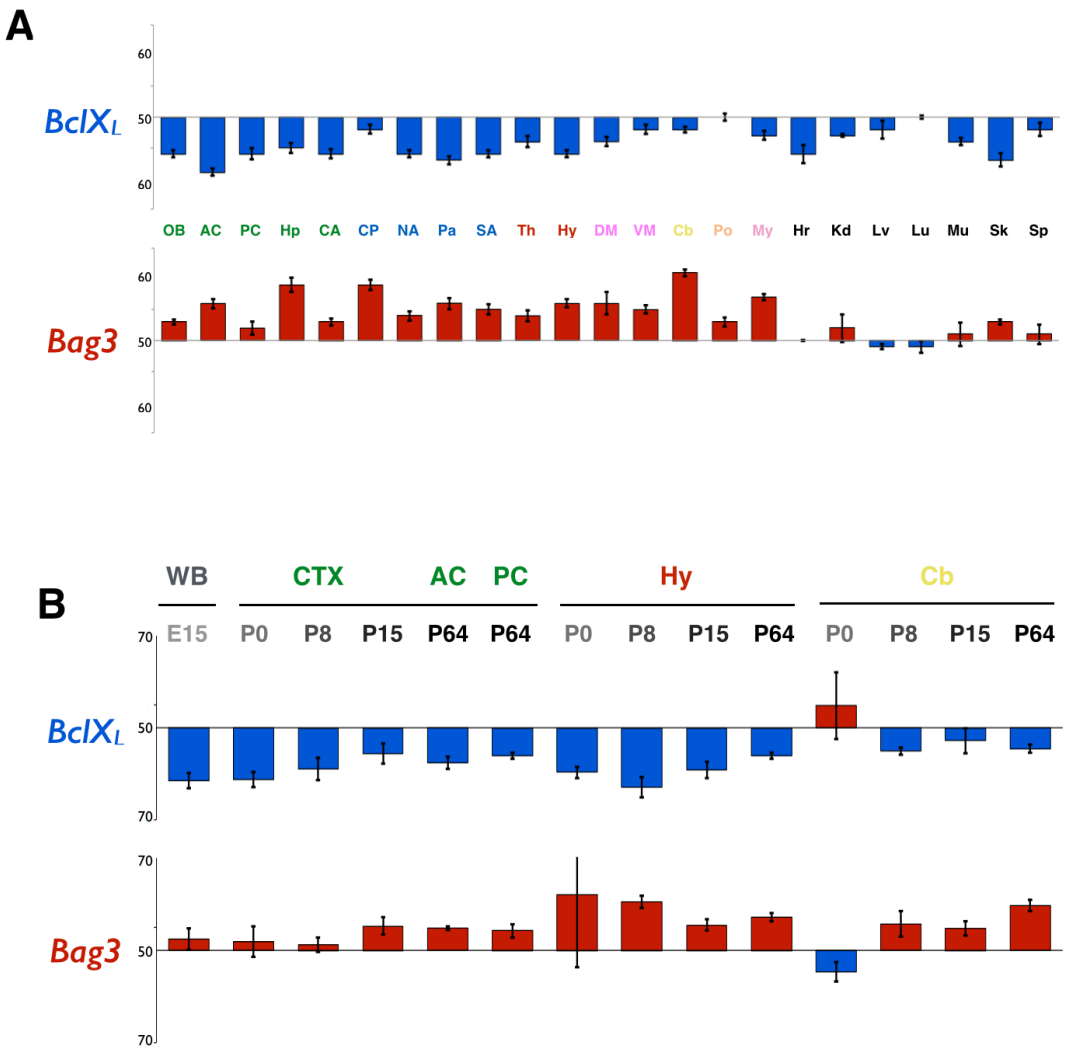

Figure 4.7: Apparent anti-correlation in parental biases of $BclX_L$ and Bag3

(A) Spatial regulation of BclX and Bag3 in the adult. **(B)** Temporal regulation of BclX and Bag3 in during development.

DISCUSSION

Here we have investigated the spatial, temporal and spatiotemporal regulation of imprinted genes in specific brain regions and found striking regulation of parental bias in gene expression. First we show that the parental bias and/or expression levels of half of the identified imprinted genes are regulated according to cerebellar development. Imprinting was investigated in the midst of an important milestone (Postnatal day 8), which constitutes the peak of granule cell precursor proliferation, the active migration of granule cells to the inner granule layer and the establishment of connectivity between mossy fibers, granule cells and Purkinje cells. Most of the regulated genes, for example the E3 ubiquitin ligase Asb4 and the pro-apopotic transcription factor Blcap, exhibit stronger imprinting during this time point than in the mature cerebellum. We also demonstrate that multiple imprinted genes maintain their degree of parental bias but change their expression according to age, including the cationic amino-acid transporter Slc38a4 and Rasgrf1 (described in chapter 1). Together, almost half of the genes imprinted in the cerebellum have their imprinting and/or expression regulated according to age. Indeed, several genes like Plagl1 and Cdkn1c are known to regulate cell survival and differentiation during embryonic development. On the other hand, a smaller subset including Sintaxin 6, and the protocadherins Pcdhb12 and Pcdhb20, is associated with cell motility and synapse formation suggesting imprinting is also involved in the regulation of neuronal wiring.

We have also investigated the spatial regulation of both strongly and mildly imprinted genes and identified two main features. First, imprinting patterns in the brain highly differs from that of other body tissues. We show that a large portion of genes appear only imprinted in the brain, while few genes show weak or null imprinting in the brain and strong imprinting in other tissues. We also show that some genes such as Igf2 and Grb10 (Blagitko et al., 2000) switch their

preferentially expressed parental allele between the brain and the rest of the body. These results clearly indicate cells in the brain execute a very specific program of genomic imprinting that is distinct from other tissues.

This in turn raising a fundamental question: what is the functional significance of these changes in imprinting between the brain and non-brain tissues, and between different brain regions? Since its discovery, studies of genomic imprinting have noticed a key association of this mode of gene regulation with brain function. One study observed contrasting roles of maternal and paternal genomes in the brain and in the body by generating chimeric animals in which a subset of cells contained either two maternal genomes or two paternal genomes. Embryos of the former developed bigger brains and smaller bodies, while embryos of the latter developed smaller brains and bigger bodies, than normal (Keverne et al., 1996). More recently however, it has been shown that this observation can be largely explained by tissue-specific imprinting of a single gene, Grb10 (Charalambous et al., 2003, 2010; Garfield et al., 2011b). The fact that we observed a large number of both maternally and paternally biased genes under brain-specific imprinting regulation may suggest that genomic imprinting plays a more intricate and complex roles in the brain than the rest of the body. A role in the regulation of behaviors is appreciated by the neuropsychiatric features of human imprinting syndromes and phenotypes observed in mouse knockouts of imprinted genes (Davies et al., 2015). Not mutually exclusive however, is the possibility that imprinting in the brain serves to regulate energy demand and consumption. During development, a great majority of imprinted genes are expressed in the placenta, a transient tissue responsible for the nourishment and protection of the embryo. Mutations or alterations in the expression of several imprinted genes results in embryonic growth and metabolic abnormalities related to placental function (Tunster et al., 2013). Considering that the

brain is the most energetically demanding tissue of the body, imprinting could thus perform an analogous role in the brain than it does in the placenta.

Secondly, striking changes in the degree and even sometimes nature of parental bias of specific genes across various brain regions reveal a previously unappreciated dynamics of imprinting regulation inside the brain. Remarkably, multiple genes appear to share similar dynamics of parental bias, in particular genes belonging to the same genomic clusters, but also genes located in distinct clusters and chromosomes.

Our data identified at least two imprinting regulatory patterns. Pattern 1 consists of strong imprinting in the core of the brain (Striatum, Thalamus, Hypothalamus, Midbrain, Pons and Medulla) but weak or no imprinting in Cortical regions, Olfactory Bulb and cerebellum. Pattern 2 also consists of strong imprinting in the brain core and in the Cortex, while the Olfactory Bulb, Cerebellum and Hippocampus show weaker imprinting. These patterns could suggest cross-talk between genes sharing imprinting signatures mediated by common epigenetic regulators. Alternatively, these signatures may reflect the cellular composition of these brain areas, where regions showing strong imprinting contain more cell-types with monoallelic expression for the genes in question and regions exhibiting mild imprinting may contain less. Irrespective, these results are consistent with the idea that imprinted genes are functionally interconnected. Considering some of the behavioral overlaps observed in available imprinted gene mouse knockouts it would be interesting to see whether those sharing imprinting signatures show similar and potentially additive phenotypes. Furthermore, identifying the cell-types with imprinted expression of these genes may help delineate functional circuits underpinning observed phenotypes.

Finally, the anti-apoptotic genes Bag3 and $BclX_L$ shows a significant anti-correlation to $BclX_L$ it would be important to compare the effects of allele-specific deletions of $BclX_L$ and Bag3 in the brain and the phenotype when both genes are altered. Overall, our findings expand the view of genomic imprinting regulation particularly in the brain and motivate further studies of imprinting at the cell-type level.

Materials and Methods

RNA-Seq experiments

Animals subjects use in RNA-seq experiments, preparation of RNA-seq samples as well as processing and analysis of RNA-seq data are described in chapter 3.

Pyrosequencing Analyses

The protocol use in pyrosequencing experiments is described in chapter 3. The coordinates of the Allen Reference Atlas guided dissections of all brain regions interrogated.

REFERENCES

Albrecht, U., Sutcliffe, J.S., Cattanach, B.M., Beechey, C. V, Armstrong, D., Eichele, G., and Beaudet, A.L. (1997). Imprinted expression of the murine Angelman syndrome gene, Ube3a, in hippocampal and Purkinje neurons. Nat. Genet. *17*, 75–78.

Barel, O., Shalev, S.A., Ofir, R., Cohen, A., Zlotogora, J., Shorer, Z., Mazor, G., Finer, G., Khateeb, S., Zilberberg, N., et al. (2008). Maternally Inherited Birk Barel Mental Retardation Dysmorphism Syndrome Caused by a Mutation in the Genomically Imprinted Potassium Channel KCNK9. Am. J. Hum. Genet. *83*, 193–199.

Blagitko, N., Mergenthaler, S., Schulz, U., Wollmann, H.A., Craigen, W., Eggermann, T., Ropers, H.H., and Kalscheuer, V.M. (2000). Human GRB10 is imprinted and expressed from the paternal and maternal allele in a highly tissue- and isoform-specific fashion. Hum. Mol. Genet. *9*, 1587–1595.

Charalambous, M., Smith, F.M., Bennett, W.R., Crew, T.E., Mackenzie, F., and Ward, A. (2003). Disruption of the imprinted Grb10 gene leads to disproportionate overgrowth by an Igf2-independent mechanism. Proc. Natl. Acad. Sci. U. S. A. *100*, 8292–8297.

Charalambous, M., Cowley, M., Geoghegan, F., Smith, F.M., Radford, E.J., Marlow, B.P., Graham, C.F., Hurst, L.D., and Ward, A. (2010). Maternally-inherited Grb10 reduces placental size and efficiency. Dev. Biol. *337*, 1–8.

Choi, J.D., Underkoffler, L.A., Wood, A.J., Collins, J.N., Williams, P.T., Golden, J.A., Schuster, E.F., Loomes, K.M., and Oakey, R.J. (2005). A novel variant of Inpp5f is imprinted in brain, and its expression is correlated with differential methylation of an internal CpG island. Mol. Cell. Biol. *25*, 5514–5522.

Davies, J.R., Dent, C.L., McNamara, G.I., and Isles, A.R. (2015). Behavioural effects of imprinted genes. Curr. Opin. Behav. Sci. *2*, 28–33.

Filson, A.J., Louvi, A., Efstratiadis, A., and Robertson, E.J. (1993). Rescue of the T-associated maternal effect in mice carrying null mutations in Igf-2 and Igf2r, two reciprocally imprinted genes. Development *118*, 731–736.

Garfield, A.S., Cowley, M., Smith, F.M., Moorwood, K., Stewart-Cox, J.E., Gilroy, K., Baker, S., Xia, J., Dalley, J.W., Hurst, L.D., et al. (2011a). Distinct physiological and behavioural functions for parental alleles of imprinted Grb10. Nature *469*, 534–538.

Garfield, A.S., Cowley, M., Smith, F.M., Moorwood, K., Stewart-Cox, J.E., Gilroy, K., Baker, S., Xia, J., Dalley, J.W., Hurst, L.D., et al. (2011b). Distinct physiological and behavioural functions for parental alleles of imprinted Grb10. Nature *469*, 534–538.

Gregg, C., Zhang, J., Weissbourd, B., Luo, S., Schroth, G.P., Haig, D., and Dulac, C. (2010a). High-resolution analysis of parent-of-origin allelic expression in the mouse brain. Sci. (New York, NY) *329*, 643–648.

Gregg, C., Zhang, J., and Weissbourd, B. (2010b). High-resolution analysis of parent-of-origin allelic expression in the mouse brain. Science (80-.). *329*, 643–648.

Isles, A.R., Davies, W., and Wilkinson, L.S. (2006). Genomic imprinting and the social brain. Philos. Trans. R. Soc. Lond. B. Biol. Sci. *361*, 2229–2237.

Kabayama, H., Tokushige, N., Takeuchi, M., and Mikoshiba, K. (2008). Syntaxin 6 regulates nerve growth factor-dependent neurite outgrowth. Neurosci. Lett. *436*, 340–344.

Keverne, E.B., and Curley, J.P. (2008). Epigenetics, brain evolution and behaviour. Front. Neuroendocrinol. *29*, 398–412.

Keverne, E.B., Fundele, R., Narasimha, M., Barton, S.C., and Surani, M.A. (1996). Genomic imprinting and the differential roles of parental genomes in brain development. Dev. Brain Res. *92*, 91–100.

Lefebvre, L., Viville, S., Barton, S., and Ishino, F. (1998). Abnormal maternal behaviour and growth retardation associated with loss of the imprinted gene Mest. Nat. ….

Li, L., Keverne, E.B., Aparicio, S.A., Ishino, F., Barton, S.C., and Surani, M.A. (1999). Regulation of maternal behavior and offspring growth by paternally expressed Peg3. Science *284*, 330–333.

Ludwig, T., Eggenschwiler, J., Fisher, P., D'Ercole, A.J., Davenport, M.L., and Efstratiadis, A. (1996). Mouse mutants lacking the type 2 IGF receptor (IGF2R) are rescued from perinatal lethality in Igf2 and Igf1r null backgrounds. Dev. Biol. *177*, 517–535.

Peters, J. (2014). The role of genomic imprinting in biology and disease: an expanding view. Nat. Rev. Genet.

Prickett, A.R., and Oakey, R.J. (2012). A survey of tissue-specific genomic imprinting in mammals. Mol. Genet. Genomics *287*, 621–630.

Rougeulle, C., Glatt, H., and Lalande, M. (1997). The Angelman syndrome candidate gene, UBE3AIE6-AP, is imprinted in brain. Nat. Genet.

Sato, M., and Stryker, M.P. (2010). Genomic imprinting of experience-dependent cortical plasticity by the ubiquitin ligase gene Ube3a. Proc. Natl. Acad. Sci. U. S. A. *107*, 5611–5616.

Sillitoe, R. V, and Joyner, A.L. (2007). Morphology, Molecular Codes, and Circuitry Produce the Three-Dimensional Complexity of the Cerebellum. Annu. Rev. Cell Dev. Biol. *23*, 549–577.

Tiwari, A., Jung, J.-J., Inamdar, S.M., Brown, C.O., Goel, A., and Choudhury, A. (2011). Endothelial Cell Migration on Fibronectin Is Regulated by Syntaxin 6-mediated 5 1 Integrin Recycling. J. Biol. Chem. *286*, 36749–36761.

Townley-Tilson, W.H.D., Wu, Y., Ferguson, J.E., and Patterson, C. (2014). The ubiquitin ligase ASB4 promotes trophoblast differentiation through the degradation of ID2. PLoS One *9*.

Tunster, S.J., Jensen, A.B., and John, R.M. (2013). Imprinted genes in mouse placental development and the regulation of fetal energy stores. Reproduction *145*, R117–R137.

Wilkinson, L.S., Davies, W., and Isles, A.R. (2007a). Genomic imprinting effects on brain development and function. Nat. Rev. Neurosci. *8*, 832–843.

Wilkinson, L.S., Davies, W., and Isles, A.R. (2007b). Genomic imprinting effects on brain development and function. Nat. Rev. Neurosci. *8*, 832–843.

Chapter Five

Discussion, Future Directions and Concluding Remarks

OVERVIEW

Genomic Imprinting represents a fascinating biological phenomenon in which homologous chromosome regions within the same nucleus behave in drastically different ways according to their parental origin. Within defined chromosomal regions and loci, imprinted genes are bound by epigenetic mechanisms to preferentially express one of the two parental alleles. Thus, in a first level of organization, imprinted genes can be subdivided into two categories: those preferentially expressed by the maternal allele and those preferentially expressed from the paternal allele. Biased parental expression has been found to be primarily controlled by DNA methylation marks established in the parental germ lines. These marks are maintained after fertilization through the multiple cell divisions leading to the developing and ultimately mature organism. In addition to DNA methylation, the mechanisms ultimately controlling imprinting expression are diverse and complex, and include regulation by long non-coding RNAs, genomic insulators and histone modifications.

Imprinted genes generate a diverse group of transcripts, only a portion of which have protein-coding potential. In cells, proteins and RNAs encoded by imprinted genes perform a wide array of functions ranging from transcriptional regulation, to intracellular signaling to membrane signal detection. Imprinted genes are found in all tissues tested so far, but intriguingly, many imprinted genes are preferentially expressed in the placenta, where they play essential roles in nutrient transfer and nourishment of the embryo, and in the brain where their functions are only beginning to be elucidated.

This preferential expression of imprinted genes in the brain makes the phenomenon of genomic imprinting all the more intriguing. The brain is the most complex organ of the body that

is ultimately in charge of orchestrating body homeostasis and animal behavior. Thus, what are the roles of these genes in the physiology of the brain and how do they affect normal and pathological brain functions? Interestingly, the expression of imprinted genes appears to substantially vary across different regions of the brain and seems particularly enriched in hypothalamic nuclei and monoaminergic regions (Gregg et al., 2010). Accordingly, mutations in imprinted genes in both mouse and humans can lead to feeding disorders, hyper- or hypo-activity, social impairments and cognitive deficits. But what are the dynamics of imprinted genes within the brain? When and where do they exert the roles leading to the observed phenotypes? Finally, what is the full repertoire of the genes that are imprinted in the brain, and how does it compare to other tissues?

My thesis work provides information relevant to all of these questions. A critical decision at the onset of my work was to focus our initial investigation on the cerebellum. Indeed, although the cerebellum has so far not been particularly associated with the biology of imprinted genes, as were other regions such as the hypothalamus, this brain structure provides unique advantages in experimental design aimed at genome-wide investigation of genomic imprinting. Briefly, its large size and anatomical location allow accurate and reproducible dissections, providing sufficient material for single-cerebellum RNA-seq data. Additionally it provides a well-defined genetic circuit with testable phenotypic features. Finally its postnatal development provides readily access to developmental milestones that are more concealed in other brain areas. Thus my work began by profiling imprinted gene expression in the cerebellum of young (P8) and adult animals.

TRANSCRIPT LEVEL INFERENCES AND DETECTION OF IMPRINTING BY BRAIM

Besides noting 74 previously known imprinted genes, we detected and validated 40 novel imprinted genes. Overall our method provides a 92% true positive rate, which is significantly higher than a recently reported approach (Babak, 2012; DeVeale et al., 2012). We credit our successful predictions to 3 unique features in our statistical analysis and experimental design. First, we performed an analysis of imprinting at the transcript level, and not at the gene level as was done in all other studies. This is important, as it provides in most cases a more accurate quantification of expression levels of alleles of a given gene (for both maternal and paternal variants of a transcript). This is because transcripts, not genes are the molecules generating the reads measured by RNA-seq. Second, we have developed a bayesian regression allelic imbalance model (BRAIM) that is able to test for parental effects on allele-specific expression, while taking into account other factors of the experimental design including cross, sex and age in this study. Third, our experimental design included a large number of replicates (48 total), which in turn provided sufficient statistical power to detect parental effects of small to high magnitudes. Finally, by performing comprehensive validation of our results by the independent method of pyrosequencing, we were able to exclude false positives from our analyses. In summary, BRAIM is a new and powerful approach that allows a sensitive and rigorous study of genomic imprinting at the transcript level, while taking into account the effects of multiple experimental factors.

ISOFORM-SPECIFIC IMPRINTING

The power of our approach was best demonstrated by the accurate detection of isoform-specific imprinting (genes expressing different sets of transcripts from the maternal and paternal

alleles) in Gnas, Trappc9, H13, Herc3, Commd1 and Inpp5f 6 of 7 loci exhibiting this mode of expression. Blcap was missed because it does not contain SNPs in regions of one alternatively expressed transcript, making the isoform-specific imprinting inaccessible to our analysis. In addition, we discovered a new example of isoform-specific imprinting within the known imprinted gene Mest. In six of the eight cases, isoform-specific imprinting is associated with paternal expression of either a retrogene or a non-coding RNA, within intronic regions of a host gene (Gregg et al., 2010; McCole and Oakey, 2008). Transcriptional interference is thought to induce unique transcripts on the host's paternal allele (McCole and Oakey, 2008). Interestingly we find two novel paternally expressed long non-coding RNAs, Gm16299 and FR0149454, within intronic regions of host genes Rbm20 and Ccdc40, respectively. However, expression of the Rbm20 and Ccdc40 was not detected in the cerebellum. It is predicted that these host genes will exhibit isoform specific imprinting in tissues where the lncRNAs are also expressed. Finally, an important question is whether isoform-specific imprinting results in alternative functions of the host gene. One interesting case is Herc3, which encodes for a Hect domain E3 ubiquitin ligase with multiple catalytic clefts in the maternal allele. On the other hand, the paternal allele also encodes a Hect E3 ubiquitin ligase but with substantially less catalytic domains (Marchler-Bauer et al., 2011). Thus, paternal and paternal alleles encode proteins with potentially distinct functionalities, and future investigations should assess the cellular roles of these two proteins, and in turn imprinting dependent functional variability.

PATERNALLY BIASED IMPRINTING

The majority of novel imprinted genes discovered by our approach exhibit moderate to weak expression differences between the two parental alleles, referred to here as parental biases,

rather than monoallelic expression, which implies the all or none silencing of one of the two parental allleles. Even though preferential parental expression rather than clear cut monoallelic expression has already been reported for multiple genes, this form of imprinting remains largely overlooked (Gregg, 2014; Khatib, 2007). Parental biases observed at the level of a tissue can arise from subtle differences between parental allele expressions within all cells within the tissue. Alternatively, parental biases at the tissue level may result from monoallelic expression within a particular cell type, while other cell types display biallelic expression (Gregg, 2014). To help distinguish between these models, single-cell or cell-type resolution is needed. At this moment single cell RNA-seq platforms are not sensitive and reliable enough for the systematic characterization of genomic imprinting (but see Deng et al., 2014) and other forms of cell-type specific characterization may help address this question. For example, SNP-sensitive single-molecule FISH may be instrumental in addressing this important question (Levesque et al., 2013).

GENOMIC FEATURES OF NOVEL IMPRINTED GENES

Of the 40 imprinted genes discovered by our approach, a large subset exhibits genomic features predicted in canonical genomic imprinting. Fkbp6, exhibits an almost exclusive paternal expression in the cerebellum. Interestingly this gene has been identified as a target of Zfp57, a DNA-binding protein essential for the protection of imprinted DMRs during global demethylation in the early embryo (Quenneville et al., 2011). Because Zfp57 functions as an imprinting-specific protector, its genome localization might serve in the identification of additional imprinted genes, an avenue that we are interested in exploring further. In addition, Nhlrc1 and Actn1 were found to exhibit paternally biased expression in our study. Interestingly,

maternally imprinted DMRs where identified for both genes in earlier studies (Esperón et al., 2012; Xie et al., 2012).

Finally, a common feature of imprinted genes is their clustered organization. Thus, adjacency to known imprinted regions is highly predictive of imprinting regulation. Accordingly, a large proportion of the genes we discovered were found to map to the vicinity of known imprinted genes. Interestingly, however, most of these genes exhibit parental biased expression, which is in contrast to the monoallelic expression observed by the neighboring gene. For example, Ankrd34c and Ctsh, two novel paternally biased genes, flank both sides of the paternally expressed Rasgrf1. Are these features reflective of the local chromatin environment? One interesting possibility is that imprinting genes flanking imprinted clusters are regulated through mechanisms of position-effect variegation. In fruit flies position-effect variegation is when a normally euchromatin region juxtaposed with a heterochromatin region, acquires heterochromatic state silencing genes in the locus. Once acquired this pattern can be inherited through multiple cell division (Elgin and Reuter, 2013). Consistent with this idea, restricted heterochromatin states of the silent allele of imprinted genes have been reported (Regha et al., 2007). This model provides interesting implications for the single cell nature of parentally biased genes, as it would support the idea that flanking genes may be silenced in the inactive allele in only subset of cells. However, inconsistent with this proposition are flanking genes exhibiting preferential expression from the opposite allele to that of the mononallelically expressed neighbor, which are also observed in our data. For example, the paternally expressed Plagl1 and the maternally biased Phactr2 are next to each other. Regardless of the precise mechanism underlying parental bias versus monoallelic expression, a key issue that will need to be addressed is whether parental biases in flanking genes have functional implications. Our phylogenetic

analysis demonstrates an above average microsynteny conservation for flanking imprinted genes and neighbors, which may therefore provide support for such a functional role of parental bias. As I will discuss in a later portion of this chapter, genetic approaches through targeted manipulations of parentally biased genes will help address this issue directly.

DEVELOPMENTAL REGULATION OF GENOMIC IMPRINTING

Multiple imprinted genes have been implicated in brain development, where they have been shown to regulate a wide range of developmental processes (Wilkinson et al., 2007). Consistent with a role of imprinted genes in neurodevelopmental processes, our comparison of imprinting between developing and adult cerebellum revealed that 57 genes, meaning half of all the genes found imprinted in the cerebellum, are regulated according to age. Of these 29 genes show differences in their overall expression but not in parental biases, 17 genes show changes in parental biases but not their overall expression, and 11 genes exhibited changes in both overall expression and parental biases. Interestingly, most genes show higher expression and/or parental biases during development compared to the adult stage, suggesting that the parental bias in gene expression if developmentally regulated, and thus, may play roles in neurodevelopment. Among these were four maternally biased genes, Ier5, Mr1, Stx6 and the here named Impar (BC034090) localized side by side in distal chromosome 1, which represent a novel imprinted cluster discovered in our study. Remarkably these four genes coordinately lose their maternal biases and substantially reduce their overall expression in the adult. One of them Stx6, has been implicated in the regulation of neuronal migration and neurite formation (Kabayama et al., 2008; Tiwari et al., 2011). I look forward to learn more about the neurodevelopmental roles of these genes.

SPATIAL REGULATION OF PARENTAL BIASES

Motivated by previous studies showing regulation in the expression and, in some cases, imprinting status of imprinted genes across body tissues and brain regions (Albrecht et al., 1997; Gregg et al., 2010; Prickett and Oakey, 2012) we investigated the spatial regulation of parental bias of 28 known and novel imprinted genes. A total of 16 brain regions and 7 somatic non-neural tissues were investigated. Three interesting notions emerge from this analysis. First, we uncovered a striking difference between the brain and other body tissues, such that most genes tested were found to be imprinted in the brain, but not in other tissues. Second, a substantial variability in the degree of parental bias was observed between different brain areas. Third, genes within a cluster, as well as occasional genes located in distinct chromosomes display similar patterns of paternal biases across brain regions, raising two important questions for future investigations: What are the mechanism controlling such rich spatial regulation? And what is its biological significance?

An attractive mechanism that could potentially be associated with the regulation of brain-specific imprinting patterns is DNA methylation of non-CpG cytosines. Recent work has uncovered high levels of non-CpG methylation in the brain, particularly in adult individuals (Lister et al., 2013). DNA methylation of non-CpG cytosines is believed to be associated with two properties unique to neurons. First, since most neurons in the brain are post-mitotic, loss of non-CpG methylation through cell division is absent in neurons (see chapter 2). Second, the de-novo methyltransferase Dnmt3a is expressed in adult neurons and in principle can be responsible for accumulating levels of non-CpG methylation in the mature brain. Methylation of non-CpG sites in neurons is enriched within the gene bodies of repressed genes (Lister et al., 2013). It is still unknown whether non-CpG participates in the silencing of repressed genes or if it is simply

a by-product of transcriptional silencing. Interestingly, non-CpG methylation specifically in the paternal allele of Ago2 has been observed. Consistently, our data show that maternally biased expression of Ago2 is only observed in the brain. However, it is unclear what role if any this plays in the imprinting expression and whether this mechanism is capable of explaining the extensive variation in parental biases observed between brain regions. The participation of multiple additional factors is to be expected.

At this point we can only speculate about the functional significance regarding the regulation of parental biases in the brain. Changes in the degree of parental bias of a gene could be due either to expression increments of the preferentially expressed allele or due to expression decreases of non-preferentially expressed allele. Since our data was collected using pyrosequencing, which only provides the ratio of expression between the two alleles, and does not inform on the absolute gene expression level, observed changes in parental biases may, or may not be associates with up- or down regulation of genes, effectively modulating gene dosage, and thus possibly titrate the function of the corresponding genes.

SPATIOTEMPORAL REGULATION OF PARENTAL BIASES

We have analyzed 12 imprinted genes during 5 developmental time points in three brain regions. We observe once again substantial variations in parental biases according to time and space. Two genes display particularly striking variation in parental expression: Igf2 and Grb10. In the embryo Igf2 is paternally expressed while Grb10 is maternally expressed. This pattern is also maintained in body tissues. However, between late embryonic stages and a few days after birth, the parental expression of these genes disappears in the brain and becomes biallelic prior to P8. Strikingly, after this time point, both genes preferentially express the opposite parental allele,

thus Igf2 becomes maternally biased while Grb10 becomes paternally expressed. Interestingly this process occurs in the hypothalamus, followed by the cortex and finally in the cerebellum. This is consistent with the order in which each region matures.

In the embryo the presence of Igf2 promotes growth (DeChiara et al., 1990) while the presence of Grb10 reduces growth (Charalambous et al., 2003). However, in the brain neither gene appears to regulate metabolism or growth. Instead, Igf2 has been shown to enhance memory consolidation in the hippocampus and it has been proposed to function as a cognitive enhancer (Chen et al., 2011). On the other hand, absence of Grb10 in the brain results in an increase of social dominance behavior and barbering of cagemates (Garfield et al., 2011). Together these experiments highlight the intricate regulation of imprinted genes during development and the important functional role of the affected genes.

FUNCTIONAL SIGNIFICANCE OF PARENTAL BIAS AND ITS REGULATION

What, if any, is the functional significance of the widespread and dynamically regulated parental bias identified in our study? An attractive hypothesis may lie on the requirement for different brain regions, and possibly neuronal types, to tightly regulate the dosage of certain genes for proper function. Converging evidence from mental disorders associated with only slight over- or under-regulation of certain genes (Ramocki and Zoghbi, 2008) suggests that proper brain function may indeed require precise gene dosage. Our study reveals that certain genes undergo striking variations in parental bias from one brain region to the next, or from one developmental stage to the other. For example Argonaute 2 shows a very modest 60% maternal bias in the cerebellum and olfactory bulb, but a robust 80% maternal expression in the cortex. Similarly, Zim1 shows biallelic expression in cortical regions, 85% maternal bias in the pallidum

and hypothalamus and 60% paternal expression in the cerebellum. Igf2r, shows biallelic expression throughout the brain except in the cerebellum where 80% of its expression is maternal.

Changes in allelic expression in between developmental stages or brain areas may result in modulation of gene dosage provided that the absolute expression level is modified accordingly. RNAseq analysis of P8 and P60 cerebellum enables us to directly investigate how absolute expression levels and parental bias co-vary. Remarkably, we find that in the majority of cases, age regulated imprinting (age effect PP > 0.95) is achieved by a significant change in the expression level of the preferred allele (31 out of 57 genes), such that increase (or decrease) in gene dosage is associated with both increase (or decrease) in parental bias and absolute expression level.

The importance of this phenomenon, and our hypothesis on the regulation of gene dosage through changes in parental bias can be directly assessed by mouse genetics, in which phenotypes resulting from specific deletions of the maternal and paternal alleles of biased genes are compared, and we are currently pursuing several avenues towards this goal.

The interpretation of the results will depend on the biological origin of the parental bias. As described above the observed imprinting might arise from a parental biased within all cells of a tissue (scenario 1) or from the mix of mononallelically expressing cells with biallelically expressing cells (scenario 2). In scenario 1, the deletion of the preferentially expressed allele will result in a larger dosage reduction than deletions of the non-preferentially expressed allele. If indeed the cell is sensitive to the dosage of the gene a phenotype of higher penetrance should be observed when the preferential allele is deleted than when the non-preferential allele is. Such phenotype would strongly suggest that the observed parental bias is functionally relevant.

In scenario 2, deletion of the preferentially expressed allele results in the complete absence of the gene in mononallelically expressing cells, whereas deletion of the non-preferentially expressed allele does not affect mononallelically expressing cells, and makes biallelically expressing cells equivalent to mononallelically expressing cells. In this case if the gene is essential for the cell only deletions of the preferentially expressed allele is expected to cause a phenotype.

Thus, regardless of the biological origin of the parental bias, observation of a differential phenotype between deletions of the preferentially expressed allele and the non-preferentially expressed allele of certain genes would give credence to the functional importance of its imprinting.

Currently, our group has taken this approach with the paternally biased BclX by generating both maternal and paternal-specific allelic deletions restricted to the brain (Nestin:Cre x flx Bclx). Interestingly, whereas BclX maternal deletions show normal brain size, paternal deletions exhibit a smaller, brain. Analyses of three brain regions shows reduce areas in all regions for the paternal deletion. Finally, less neurons than glial cells, and less excitatory than inhibitory neurons, are present in the brains of paternal deletions, but not maternal deletions. These results highlight the importance of BclX brain functions and corroborates is imprinted. Depending on the biological origin of the paternal bias of BclX these observations could be due to BclX dosage sensitivity (scenario 1) or BclX cell-type specific monoallelic expression (scenario 2). Thus combining mouse genetics with single cell profiling of BclX allele-specific expression should provide important insights about the nature and function of parental biases. Multiple other interesting genes exhibiting parental biases, like Bag3 and Ago2, could be tested by this approach.

CONCLUSION

Altogether this work has expanded our knowledge of genomic imprinting in the brain, and offers new opportunities to investigate the functions of genomic imprinting in brain development and behavior. Moreover, analyses on the developmental and tissue-specific regulation of imprinting have uncovered highly dynamic regulation of parental biases. These findings represent significant progress in our understanding of how functions of imprinted genes are organized within the complex landscape of the brain, and open exciting new avenues of research on the mechanisms underlying parental bias in gene expression, and on the function of this unique form of gene regulation.

REFERENCES

Albrecht, U., Sutcliffe, J.S., Cattanach, B.M., Beechey, C. V, Armstrong, D., Eichele, G., and Beaudet, A.L. (1997). Imprinted expression of the murine Angelman syndrome gene, Ube3a, in hippocampal and Purkinje neurons. Nat. Genet. *17*, 75–78.

Babak, T. (2012). Identification of imprinted loci by transcriptome sequencing. Methods Mol. Biol. *925*, 79–88.

Charalambous, M., Smith, F.M., Bennett, W.R., Crew, T.E., Mackenzie, F., and Ward, A. (2003). Disruption of the imprinted Grb10 gene leads to disproportionate overgrowth by an Igf2-independent mechanism. Proc. Natl. Acad. Sci. U. S. A. *100*, 8292–8297.

Chen, D.Y., Stern, S.A., Garcia-Osta, A., Saunier-Rebori, B., Pollonini, G., Bambah-Mukku, D., Blitzer, R.D., and Alberini, C.M. (2011). A critical role for IGF-II in memory consolidation and enhancement. Nature *469*, 491–497.

DeChiara, T.M., Efstratiadis, A., and Robertsen, E.J. (1990). A growth-deficiency phenotype in heterozygous mice carrying an insulin-like growth factor II gene disrupted by targeting. Nature *345*, 78–80.

Deng, Q., Ramskold, D., Reinius, B., and Sandberg, R. (2014). Single-cell RNA-seq reveals dynamic, random monoallelic gene expression in mammalian cells. Science (80-.). *343*, 193–196.

DeVeale, B., van der Kooy, D., and Babak, T. (2012). Critical evaluation of imprinted gene expression by RNA-seq: A new perspective. PLoS Genet. *8*.

Elgin, S.C.R., and Reuter, G. (2013). Position-effect variegation, heterochromatin formation, and gene silencing in Drosophila. Cold Spring Harb. Perspect. Biol. *5*.

Esperón, E. de la C., Cordier, G., and Engel, N. (2012). A genomic reservoir for Tnfrsf genes is developmentally regulated and imprinted in the mouse. Epigenetics *7*, 626–634.

Garfield, A.S., Cowley, M., Smith, F.M., Moorwood, K., Stewart-Cox, J.E., Gilroy, K., Baker, S., Xia, J., Dalley, J.W., Hurst, L.D., et al. (2011). Distinct physiological and behavioural functions for parental alleles of imprinted Grb10. Nature *469*, 534–538.

Gregg, C. (2014). Known unknowns for allele-specific expression and genomic imprinting effects. F1000Prime Rep. *6*.

Gregg, C., Zhang, J., Weissbourd, B., Luo, S., Schroth, G.P., Haig, D., and Dulac, C. (2010). High-resolution analysis of parent-of-origin allelic expression in the mouse brain. Sci. (New York, NY) *329*, 643–648.

Kabayama, H., Tokushige, N., Takeuchi, M., and Mikoshiba, K. (2008). Syntaxin 6 regulates nerve growth factor-dependent neurite outgrowth. Neurosci. Lett. *436*, 340–344.

Khatib, H. (2007). Is it genomic imprinting or preferential expression? BioEssays *29*, 1022–1028.

Levesque, M.J., Ginart, P., Wei, Y., and Raj, A. (2013). Visualizing SNVs to quantify allele-specific expression in single cells. Nat. Publ. Gr. *10*, 865–867.

Lister, R., Mukamel, E. a, Nery, J.R., Urich, M., Puddifoot, C. a, Johnson, N.D., Lucero, J., Huang, Y., Dwork, A.J., Schultz, M.D., et al. (2013). Global epigenomic reconfiguration during mammalian brain development. Science *341*, 1237905.

Marchler-Bauer, A., Lu, S., Anderson, J.B., Chitsaz, F., Derbyshire, M.K., DeWeese-Scott, C., Fong, J.H., Geer, L.Y., Geer, R.C., Gonzales, N.R., et al. (2011). CDD: A Conserved Domain Database for the functional annotation of proteins. Nucleic Acids Res. *39*.

McCole, R.B., and Oakey, R.J. (2008). Unwitting hosts fall victim to imprinting. Epigenetics *3*, 258–260.

Prickett, A.R., and Oakey, R.J. (2012). A survey of tissue-specific genomic imprinting in mammals. Mol. Genet. Genomics *287*, 621–630.

Quenneville, S., Verde, G., Corsinotti, A., Kapopoulou, A., Jakobsson, J., Offner, S., Baglivo, I., Pedone, P. V, Grimaldi, G., Riccio, A., et al. (2011). In Embryonic Stem Cells, ZFP57/KAP1 Recognize a Methylated Hexanucleotide to Affect Chromatin and DNA Methylation of Imprinting Control Regions. Mol. Cell *44*, 361–372.

Ramocki, M.B., and Zoghbi, H.Y. (2008). Failure of neuronal homeostasis results in common neuropsychiatric phenotypes. Nature *455*, 912–918.

Regha, K., Sloane, M.A., Huang, R., Pauler, F.M., Warczok, K.E., Melikant, B., Radolf, M., Martens, J.H.A., Schotta, G., Jenuwein, T., et al. (2007). Active and Repressive Chromatin Are Interspersed without Spreading in an Imprinted Gene Cluster in the Mammalian Genome. Mol. Cell *27*, 353–366.

Tiwari, A., Jung, J.-J., Inamdar, S.M., Brown, C.O., Goel, A., and Choudhury, A. (2011). Endothelial Cell Migration on Fibronectin Is Regulated by Syntaxin 6-mediated 5 1 Integrin Recycling. J. Biol. Chem. *286*, 36749–36761.

Wilkinson, L.S., Davies, W., and Isles, A.R. (2007). Genomic imprinting effects on brain development and function. Nat. Rev. Neurosci. *8*, 832–843.

Xie, W., Barr, C., Kim, A., Yue, F., and Lee, A. (2012). Base-resolution analyses of sequence and parent-of-origin dependent DNA methylation in the mouse genome. Cell.