The epigenetics and role of Dopa Decarboxylase in heart development

Prickett, Adam

Awarding institution:
King's College London

Title: The epigenetics and role of Dopa Decarboxylase in heart development

Author: Adam Prickett

The epigenetics and role of Dopa Decarboxylase in heart development

Submitted by

Adam Richard Prickett

December 2012

To King's College London for the degree of

Doctor of Philosophy

Department of Medical and Molecular Genetics, King's College London, School of Medicine, London, SE1 9RT, U.K.

All work submitted in this thesis is my own

Adam Prickett

Abstract

Genomic imprinting in mammals subjects a handful of genes to silencing on one allele depending on the parent-of-origin of that allele. The *Ddc_exon1a* gene transcript is under the control of genomic imprinting in the developing and neonatal mouse heart with transcription occurring solely on the paternally inherited allele, with the maternally inherited allele epigenetically silenced. In all other tissues where *Ddc_exon1a* is expressed transcription occurs from both parentally inherited alleles.

CTCF plays a central role in controlling gene expression by regulating chromatin organisation, and has been shown to be fundamental for imprinting of the *Igf2/H19* locus through its binding to the germline differentially methylated region (gDMR). This thesis explores the occurrence of CTCF binding both genome-wide and at imprinting gDMRs in mouse brain in order to assess its relevance to the control of transcription *in vivo*. 49,358 significant binding sites are detected across the genome and binding is enriched at gene coding regions but depleted at distal intergenic regions. 12/20 (60%) imprinted gDMRs are bound by CTCF, of these five bind in a parent-of-origin specific manner implicating CTCF in the control of a subset of imprinted genes including *Ddc_exon1a*. Comparative analysis of CTCF binding in multiple tissues shows a high degree of overlap, and Motif analysis reveals CTCF binds the same canonical motif sequence in each tissue. CTCF binding in the absence of the canonical motif is more tissue-specific.

***Ddc_exon1a* expression is imprinted in the developing heart but is bi-allelic in brain, and the mechanism of imprinting is not known. To explore a model for imprinting control the epigenetic profile of the *Ddc* imprinted locus was examined in detail. Methylation analysis reveals several**

regions that are differentially methylated between heart and brain. One region constitutes a CpG rich region at the promoter of *Grb10,* another imprinted gene located adjacent to *Ddc_exon1a.* The second region is at the promoter of *AK0066690*, a non-coding antisense transcript which initiates in intron four of *Ddc_exon1a*. CTCF binding at the *Ddc/Grb10* locus is assayed in heart and brain, and binding is invariant between tissue types. The *AK0066690* transcript is expressed in the neonatal mouse heart but not in the neonatal brain, consistent with a model of silencing *Ddc_exon1a* on the maternal allele via transcriptional interference.

Ddc_exon1a **codes for the Dopa Decarboxylase (Ddc) protein, which is predominately expressed in the developing myocardium, this points to a role in fetal heart development. The role of Ddc in cardiogenesis is explored using knockout mice lacking *Ddc_exon1a* expression in heart. Expression microarrays were used to detect changes in gene expression, and morphometric analysis using 3D imaging was performed to look for gross morphological changes. Results suggest that Ddc plays a role in regulating cellular proliferation and cardiogenesis of the developing myocardium as mice lacking Ddc show a significant thinning of the apical portion of the right ventricle, a region that shows abundant Ddc expression.**

In this thesis the significance of CTCF binding to imprinting control is examined and the observations applied in a locus specific manner to explore both the mechanistic control, and the functional role of Ddc in the developing mouse heart.

Acknowledgements

First and foremost I would like to thank my supervisor Professor Rebecca Oakey for her guidance, enthusiasm and support, and for encouragement when tasks needed doing. I thank Professor Scott Baldwin, who generously allowed me to visit his Laboratory at Vanderbilt, and his wonderful family who could not have been kinder to me whilst I was there. I am very grateful to all the members of the Oakey laboratory, both past and present particularly Reiner Schulz, Ruth McCole and Mike Cowley, who supported me at the beginning. Finally I thank my parents and grandparents for their moral and financial help and Helen Tracey for being extraordinarily supportive throughout.

Table of Contents

List of Figures

List of Tables

List of Appendices

Abbreviations

DAB	3,3'-Diaminobenzidine
5hmC	5-hydroxymethylcytosine
5-HTP	5-Hydroxytryptophan
m5C	5-methylcytosine
AS	Angelmann syndrome
BAC	bacterial artificial chromosome
bp	base pairs
B	C57BL/6
ChIP	chromatin immunoprecipitation
3C	Chromosome Conformation Capture
CGI	CpG island
CpA	cytosine phosphate adenine (CA dinucleotide)
CpG	cytosine phosphate guanine (CG dinucleotide)
DMR	differentially methylated region
DNMT	DNA methyltransferase
Ddc	Dopa Decarboxylase
DRG	dorsal root ganglion
e	embryonic day
ES cells	embryonic stem cells
EFIC	episcopic fluorescence image capture
EMT	epithelial to mesenchymal transformation
etOH	ethanol
FDR	false discovery rate
F1	filial 1 (first filial generation)
GO	gene ontology
gDMR	germline differentially methylated region
ICR	imprinting control region
IVS	Interventricular septum
kbp	kilo base pairs
LA	left atria
LV	left ventricle
MeDIP	methylated DNA immunoprecipitation
MV	mitral valve
C	*Mus mus castaneus*
ncRNA	non-coding RNA
piRNA	piwi-interacting RNA
poly(A)	polyadenylation
PCR	polymerase chain reaction
PWS	Prader-Willi syndrome
POA	pre-optic area (of the brain)
PGC	primordial germ cell

PCA	principle component analysis
qPCR	quantitative polymerase chain reaction
RT-PCR	reverse transcriptase polymerase chain reaction
RA	right atria
RV	right ventricle
SNP	single nucleotide polymorphism
TpG	thymine phosphate guanine (TG dinucleotide)
TI	transcriptional interference
UPD	uniparental disomy
UTR	untranslated region
WAMIDEX	web atlas of murine genomic imprinting and differential expression

Chapter 1

Introduction

1.1 Genomic Imprinting

In mammalian cells, two complete sets of autosomal genes are present in each somatic cell, one inherited from the mother and one from the father. When expressed, autosomal genes are usually transcribed more or less equally from both parental alleles, however imprinted genes form an exception to this rule. Genomic imprinting results in transcription of some genes from the allele that has been inherited from one parent, while the other allele is epigenetically silenced. This transcription and silencing of imprinted gene alleles is usually consistent between cells of the same type (Figure 1.1).

1.2 The discovery of genomic imprinting

1.2.1 Parental genomes are non-equivalent

Non-equivalence of the parental genomes was first demonstrated by two groups in the mid-1980s, both used elegant nuclear transfer experiments to demonstrate the requirement for genetic contributions from both oocytes and sperm in order for normal embryonic development to proceed ((McGrath, et al., 1984, Surani, et al., 1984)). By transplanting male and female pronuclei into one-cell stage embryos to produce gynogenetic (two female pronuclei) and androgenetic (two male pronuclei) embryos, they showed that embryos were not viable post-implantation, despite the fact that both are diploid. This established that maternal and paternal contributions to the embryo, despite being genetically identical, are not functionally identical and suggested that the parental genomes must be differentially marked by epigenetic mechanisms as a consequence of their origin (Figure 1.2) (Surani, et al., 1986).

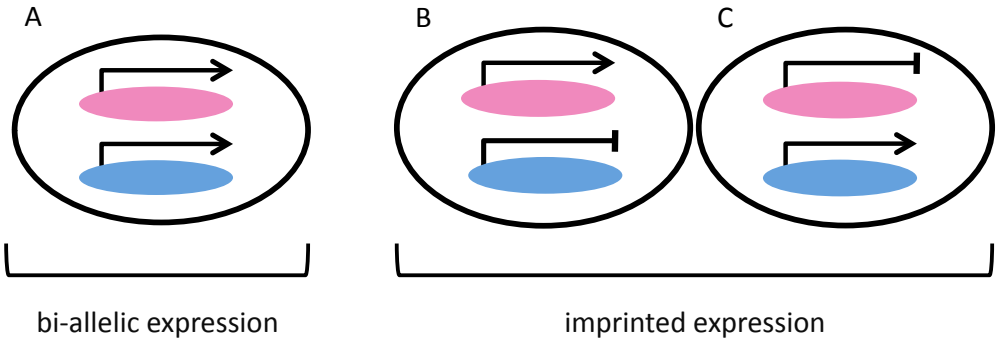


Figure 1.1 Imprinted gene expression. Cartoon showing two gene alleles, one inherited from the father (blue) and one from the mother (pink). (A) Demonstrates bi-allelic expression with the gene being transcribed from both the maternally and paternally inherited copy. (B), (C) Demonstrates imprinted expression. (B) The maternally expressed imprinted gene is being transcribed from the maternal copy only, with the paternal copy silenced. (C) The paternally expressed imprinted gene is being transcribed from the paternal copy only with the maternal copy silenced.

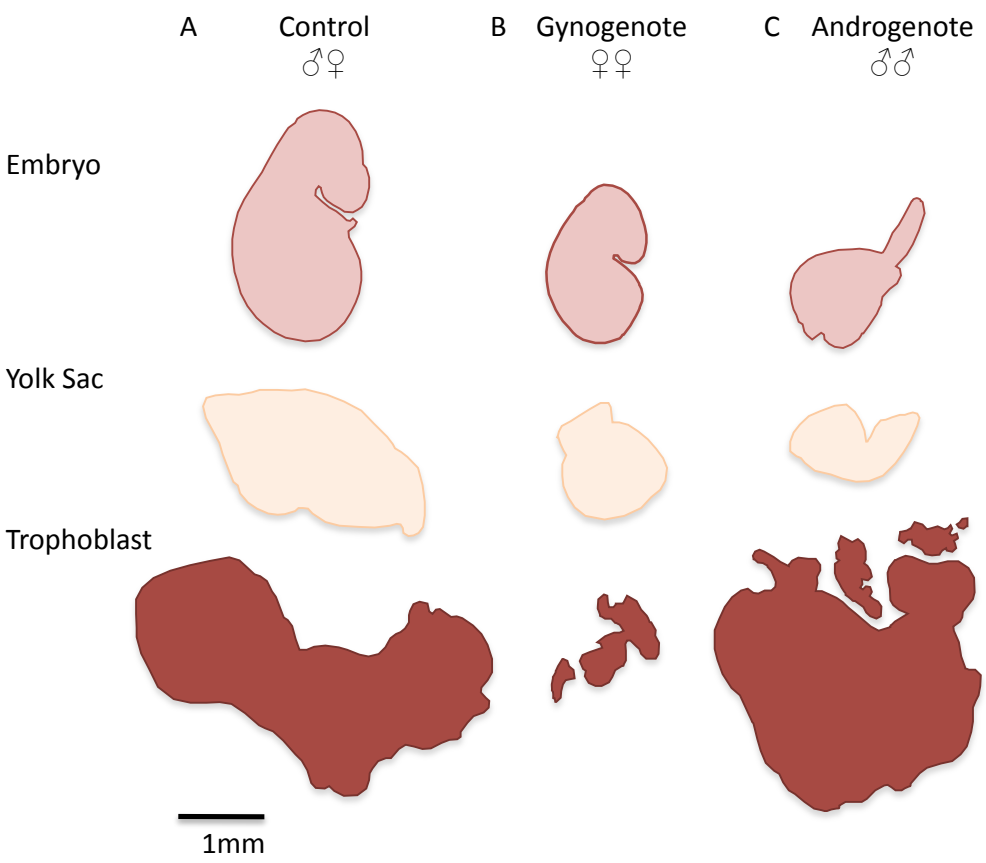


Figure 1.2 Embryonic development requires male and female genomes. Embryonic development from control zygote (A), zygotes containing maternal genomes only (gynogenote - B), and paternal genomes only (androgenote – C). Gynogenote is small and develops to a maximum of 25 somite stage. Androgenote develops to a maximum of 8 somite stage with severe overgrowth of trophoblast. Adapted from Surani 1986

Further evidence supporting these nuclear transfer experiments has come from genetic studies using translocation chromosomes in the mouse. Despite the fact that both maternal and paternal genomic contributions are necessary for embryonic development, some chromosomes appear more tolerant to disruption of these parental effects, for example zygotes with uniparental disomy (UPD) of chromosomes 1, 4, 5, 9, 13, 14 and 15 will survive normally (Cattanach, et al., 1985). By utilizing mice disomic or partially disomic for individual chromosomes Cattanach and Kirk were able to identify which autosomal regions are associated with gross phenotypic abnormalities when inherited from only one parent (Searle, et al., 1978). For example, mice partially disomic for chromosome 11 and 13 produced by intercrossing mice heterozygous for the Robertsonian translocation Rb(11.13)4Bnr, provided evidence that proximal chromosome 11 is responsible for a growth phenotype. When chromosome 11 is of maternal origin only, a growth-retarded phenotype is observed and the reverse phenotype is observed in the paternal chromosome 11 UPD (Cattanach, et al., 1985). Using uniparental duplication covering different genomic regions an "imprinting map" for the mouse genome has been generated (Williamson CM, 2012).

1.2.2 The first identification of imprinted genes in mammals

Genomic imprinting of an individual gene was first described for the *Igf2* gene located on mouse distal chromosome 7, *Igf2* is expressed throughout embryonic development and in the early neonatal period. Initially DeChiara et al. disrupted the *Igf2* gene in cultured embryonic stem cells using gene targeting and noticed that, when transmitted via the male germ line, heterozygous progeny exhibited a growth deficiency when compared to wild-type (DeChiara, et al., 1990). Further experimental analysis revealed that when the mutated allele is transmitted maternally there is no phenotypic difference from the wildtype and that this difference is a result of embryonic and neonatal expression of *Igf2* coming only from the paternal allele in most tissues (DeChiara, et al., 1991). Shortly after the

discovery that *Igf2* is genomically imprinted the adjacent *H19* gene was also confirmed as an imprinted gene with expression coming only from the maternal allele (Bartolomei, et al., 1991), demonstrating that imprinted genes may cluster together in the genome.

1.2.3 Genomic imprinting in therian mammals and flowering plants

Evolutionary evidence suggests genomic imprinting in mammals first appeared following the divergence of therian mammals (placental mammals and marsupials) from prototherians (monotremes) approximately 150 million years ago (Killian, et al., 2001). Imprinting has evolved convergently in flowering plants (Feil, et al., 2007), where it was first identified in maize following analysis of the maternal effect of the R gene, responsible for pigmentation (Kermicle, 1970). There is no evidence that genomic imprinting is conserved in other non-mammalian vertebrates, birds or fish (O'Neill, et al., 2000, Yokomine, et al., 2001, Yokomine, 2005).

1.2.4 Imprinted genes cluster in the genome

Since the discovery of *Igf2*, many more genes have been identified to be imprinted in one or more cell type, and the number of known imprinted transcripts currently stands at approximately 150 in mouse (Schulz, et al., 2008, Williamson CM, 2012) and 72 in humans (Morison, et al., 2005), many of these transcripts are non-coding or derived from retro-transposition events (Wood, et al., 2007) and interestingly mono-allelic expression of imprinted genes is not always conserved between mouse and human. In mammals imprinted genes are frequently organised in clusters (Reik, et al., 2001) or pairs (Wood, et al., 2007) and only occasionally as singletons (Hagiwara, et al., 1997), this organisation is quite similar in the mouse and human genomes. The number of imprinted genes can only be viewed as approximations, as the identification of novel instances of genomic imprinting is frequently confounded by difficulties in testing for genomic imprinting in heterogeneous cell populations such as those found in whole tissue and the technical difficulties of using different techniques

for detection of imprinted transcripts (DeVeale, et al., 2012, Kelsey, et al., 2012). Additionally for many genes there is conflicting evidence in the literature as to whether a gene is indeed imprinted, and different interpretations of the term genomic imprinting means it is often used to refer to situations where transcript expression is biased towards parent of origin (Prickett, et al., 2012). Some chromosomes appear to be hotspots for genomic imprinting, with many imprinted genes present, where other chromosomes have very few (Figure 1.3) (Williamson CM, 2012).

1.3 Imprinting in disease

Many genes that are subject to genomic imprinting are also expressed during embryonic growth, extra-embryonic development and in the developing brain, suggesting that imprinting may be important in mediating placental efficiency, prenatal growth and neurological development. This important role for genomic imprinting in development is further demonstrated by the fact that several imprinted gene clusters show developmental specific regulation of imprinted expression (Santoro, et al., 2011) and many imprinted genes are tissue-specifically imprinted only in extra-embryonic tissues of the placenta and yolk sac and the developing and neonatal brain, whilst remaining bi-allelic in other somatic tissues (Prickett, et al., 2012). Experiments using Robertsonian translocations to examine chromosomal uniparental disomy also frequently demonstrate prenatal lethality or embryonic growth perturbations (Cattanach, 1986). Unsurprisingly, disruption of imprinting has been implicated in multiple human growth syndromes and cancers (Murrell, 2006, Piedrahita, 2011). Disruption of imprinting may be caused by uniparental disomy (UPD), a rare genetic event whereby both copies of a chromosome are inherited either from the mother or from the father. Incidence of UPD of any chromosome is around 1 in 3500 live births (Robinson, 2000). Two well characterised genetic disorders associated with aberrant imprinting are Prader-Willi Syndrome (PWS) and

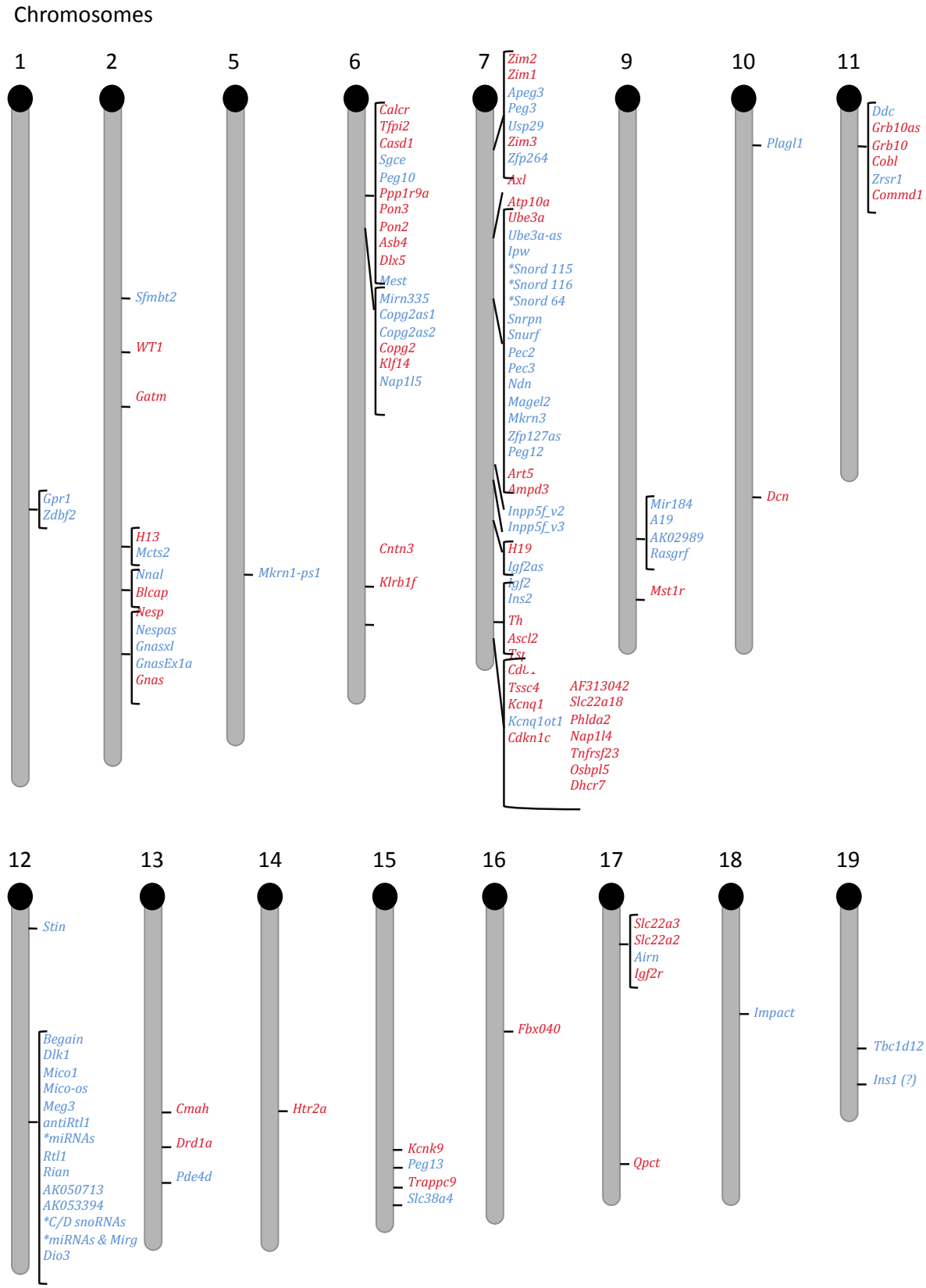


Figure 1.3 Location of imprinted genes. Mouse imprinted genes and gene clusters show that some chromosomes (e.g. Chr7) are imprinting "hot-spots" whereas others have very few or no imprinted genes. Maternally expressed genes are shown in Red, paternally expressed genes in blue. Gene clusters are shown in brackets demonstrating that many imprinted genes fall into clusters. * denotes snRNAs and microRNAs. Data taken from Williamson 2012

Angelmann Syndrome (AS), both of which are linked to abnormalities located on chromosome 15q11-q13 (Donlon, 1988, Magenis, et al., 1987), a region containing the imprinted genes *UBE3A, SNRPN* and *SNURF*. Inheritance of chromosome 5q11-q13 is complex, thus both disorders have a heterogeneous genetic aetiology, but essentially both exhibit genetic abnormities associated with either the paternally (PWS) or maternally (AS) inherited alleles. PWS is associated with a loss of expression of paternally inherited alleles, in 20-30% of cases this is due to complete maternal UPD of chromosome 15 (Bittel, et al., 2005). AS occurs as a result of loss of expression of maternally inherited alleles, in 5% of cases this is from chromosome 15 paternal UPD (Clayton-Smith, et al., 2003). Children with PWS demonstrate hypotonia and hyperphagia, leading to obesity, have short stature, small hands and feet, and exhibit mild mental retardation (Bittel, et al., 2005). The characteristics of AS include severe mental retardation, hyperactivity and seizures (Clayton-Smith, et al., 2003). Another imprinted region on chromosome 11, harbouring the well characterised *IGF2/H19* imprinted locus has been implicated in the imprinting associated disorders Beckwith-Wiedemann Syndrome (BWS) and Silver-Russell Syndrome (SRS). BWS children exhibit intrauterine and postnatal overgrowth, and increased risk of tumours in childhood whereas SRS is characterised by pre and postnatal growth retardation (Eggermann, 2009). In addition to human growth syndromes, the clinical implications of aberrant imprinting have also been observed in the tumourigenesis of multiple cancers (Uribe-Lewis, et al., 2011).

1.4 Evolutionary reasons for genomic imprinting

Since its discovery the evolutionary reasons for genomic imprinting have been energetically sought, with multiple theories proposed. Mono-allelic expression of a gene results in loss of heterozygosity and the functional redundancy that this allows. This renders imprinted genes subject to a complete loss of function if only the expressed allele harbours a deleterious mutation. The advantage of a diploid genome has been well

established (Otto, et al., 1992, Perrot, et al., 1991) therefore this haploid expression of imprinted genes poses an evolutionary paradox. The drawbacks of proposed theories for evolution of imprinting are twofold. Firstly they do not predict, only describe, the existing situation and secondly no single theory adequately allows each imprinted gene to fit the hypothesis.

1.4.1 Parental conflict theory

The first proposed theory was the “parental conflict” or “kinship” theory of evolution for genomic imprinting (Moore, et al., 1991). The salient points of this theory are that the maternal and paternal genomes have different interests in terms of offspring demand on maternal resource allocation, with paternally expressed imprinted genes selected to maximize demand from the mother, and maternally expressed genes to act to balance demand between offspring of current and future litters (Moore, et al., 1991).

1.4.2 Ovarian time bomb theory

The ovarian time bomb hypothesis suggest genomic imprinting has evolved to prevent spontaneous parthenogenetic activation of oocytes that may result in the formation of potentially fatal malignant trophoblastic disease in placental mammals; this is based on the premise that artificially created parthenogenotes have severely underdeveloped trophoblastic tissue (Varmuza, et al., 1994). The hypothesis suggests that maternal silencing of a gene allele by genomic imprinting ensures normal extra-embryonic development can only occur within the uterus, after fertilization from the paternal genome, however it fails to give a solid hypothesis to explain why paternal silencing by genomic imprinting also occurs in mammals.

1.4.3 Co-adaptation theory

More recent theories include the parent-infant co-adaptive process (co-adaptation theory) (Keverne, et al., 2008) and gene dosage hypothesis. The

co-adaptation theory hypothesizes a link between a number of genes imprinted in both the placenta and in the maternal and pup brain, specifically in the hypothalamus. Each of these genes are maternally silenced and paternally expressed in both tissues. Investigation of *Peg3* show mothers carrying a null mutation demonstrate reduced food intake during pregnancy resulting in impairment of weight gain and increase fat reserves normally associated with pregnancy, smaller litter sizes, reduced maternal care particularly with regard to offspring thermostasis and severe impairment in milk letdown, even when the pup is wildtype for these genes. Moreover, when *Peg3* is mutated in the offspring (where it is expressed in both placenta and developing hypothalamus), but remains wildtype in the mother, the phenotype seen is similar to that when mutated in the mother. It is postulated that this effect may be due to hormonally regulated signalling from the placental to the maternal hypothalamus. Thus during pregnancy maternal hypothalamus and fetal placenta and hypothalamus are simultaneously active and selectionary pressure may operate to ensure that offspring extract adequate pre and postnatal resource from the mother and are predisposed towards this same mothering style when adult (Curley, et al., 2004).

1.4.4 Gene dosage theory

It is becoming increasingly clear that genomic imprinting is far more dynamic than first thought. Many “imprinted” genes are known to appear bi-allelically expressed in a tissue specific and temporal specific manner (Prickett, et al., 2012), and some genes exhibit inter-individual heterogeneity with regards to imprinted expression (Riesewijk, et al., 1998). The upshot of this is that genomic imprinting may act to regulate gene expression in a quantitative manner, ensuring correct gene dosage in particular cell types by effectively halving expression. An investigation into the imprinted *Dlk1* gene suggests that a switch from imprinted to bi-allelic expression of *Dlk1* in niche astrocytes and neural stem cells via the specific acquisition of a methylation mark, is critical for dosage-dependent post-

natal neural development (Ferrón, et al., 2011). Mice harbouring a *Dlk1* BAC transgene, recapitulating a re-activation of bi-allelic expression, are born larger than wild-type littermates, but exhibit skeletal abnormities, and show increase postnatal lethality due to a reduced suckling ability (da Rocha, et al., 2009).

1.5 Epigenetics

Epigenetic marks broadly describe reversible modifications to the DNA or DNA environments that are stable and heritable, both through the germline and though cell division. Importantly epigenetic marks can hold and transmit information independently of the genetic code and thus are not subject to traditional Mendelian rules governing genetic inheritance. For the purpose of this thesis the definition of epigenetics is restricted to comprise of direct chemical modification to the DNA itself and to modifications of histone proteins around which DNA is packaged (Figure 1.4).

1.5.1 DNA methylation

DNA methylation is the chemical modification found on the cytosine residue of DNA in eukaryotic cells, and is essential for mammalian embryonic development. The most abundant modification to the cytosine residue is the addition of a methyl group to form 5-methylcytosine (m5C) (Figure 1.5, B) (Holliday, et al., 1975), usually this occurs in the context of a CpG dinucleotide although it has been shown to occur in non-CpG contexts, particularly in embryonic stem cells (Lister, et al., 2009). m5C accounts for around 1% of total DNA bases in a human somatic cell and is present in between 70-80% of genome-wide CpGs. More recently 5-hydroxymethylcytosine has been identified as an alternative covalent modification to the cytosine residue in mammalian genomes (Figure 1.5, C) (Kriaucionis, et al., 2009, Tahiliani, et al., 2009).

1.5.2 CpG Island methylation

CpG dinucleotides are generally depleted from the mammalian genome; where the fraction of genomic (G+C) is 0.4, the expected frequency of a CpG occurring is 0.2 x 0.2 = 0.04, however the observed frequency in the mouse genome is 0.008. This is due to the fact that m5C can be spontaneously converted to uracil (and then thymine), converting CpG to TpG (or CpA on the reverse strand), thus CpGs have been gradually

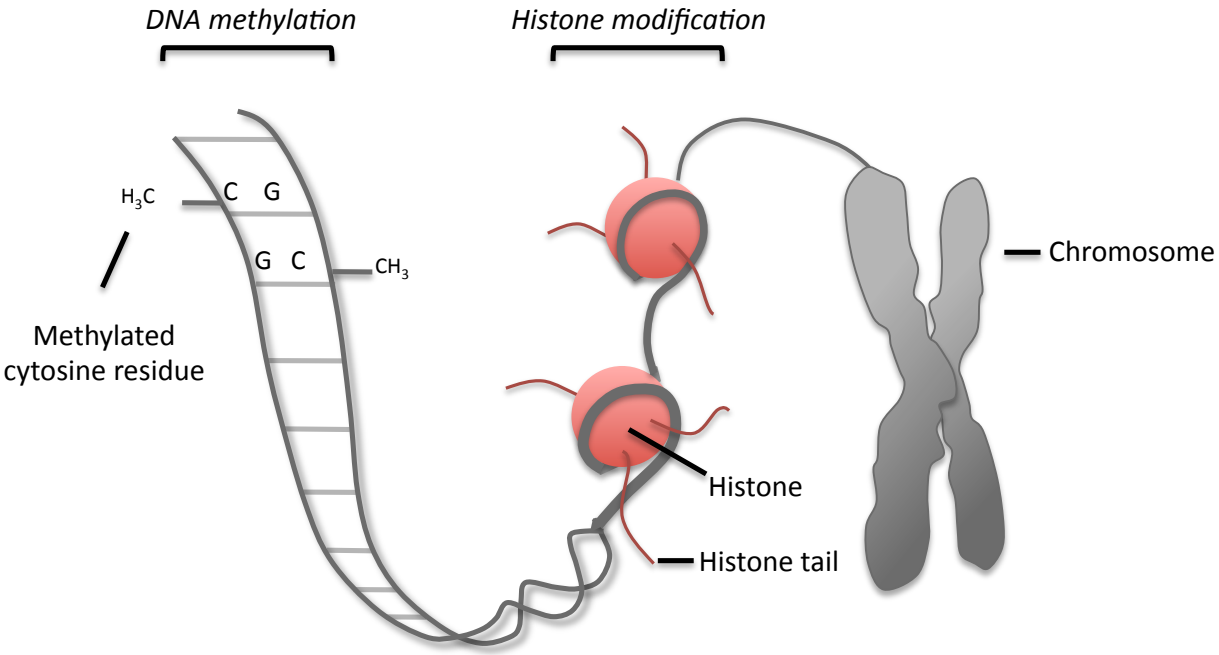


Figure 1.4 Epigenetic marks. Epigenetic variations comprise of methylation of the DNA molecule at the cytosine base usually in the context of a CpG dinucleotide, and Histone modifications, which are covalent changes to amino acid residues on histone tails.

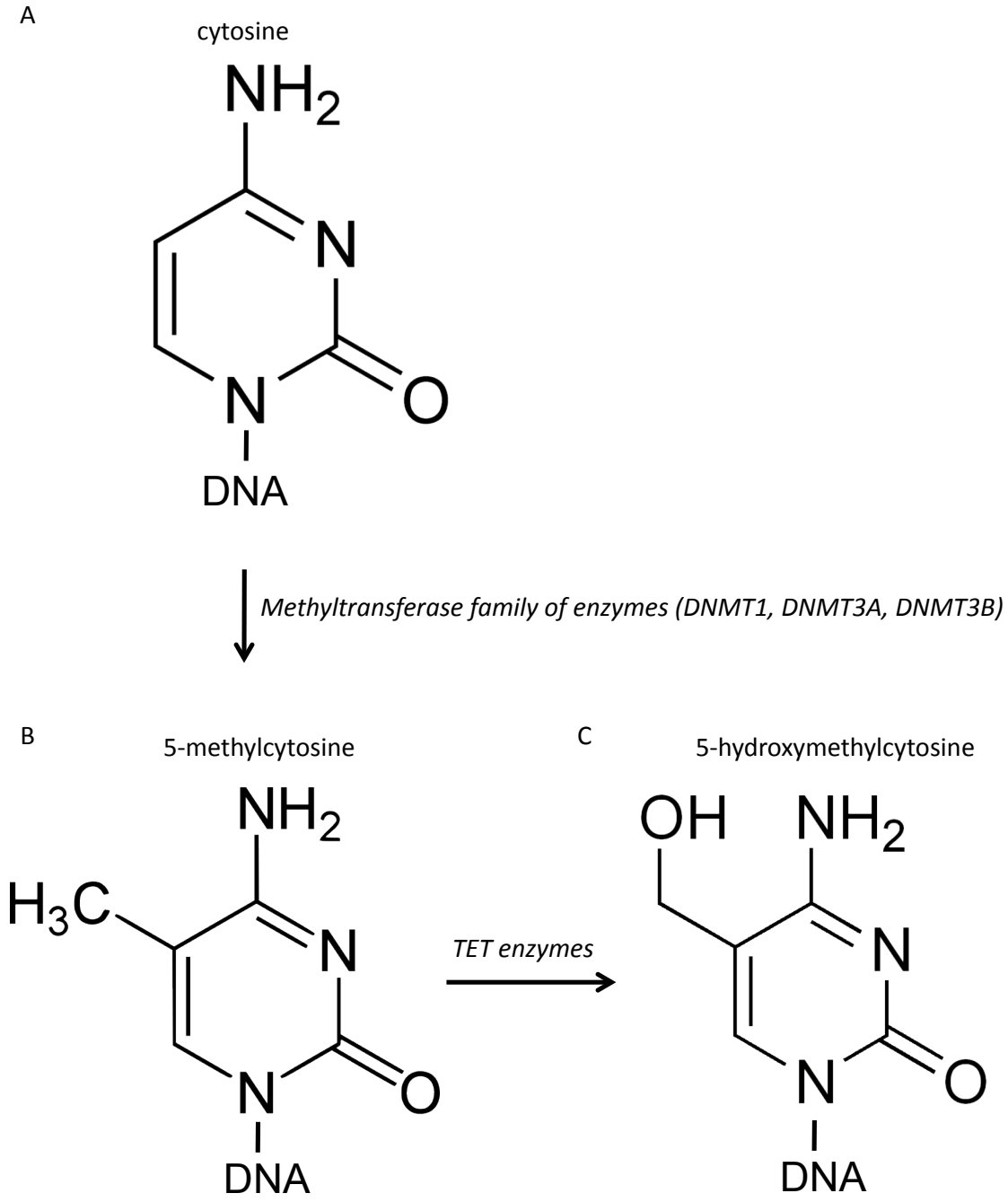


Figure 1.5 DNA methylation at the cytosine base. (A) Unmethylated cytosine residue. (B) Methylated cytosine residue, the methyl group is added at the carbon 5 position, the reaction is catalyzed by the methyltransferase family of enzymes (C) Hydroxymethylated cytosine residue, methylated cytosine is converted to hydroxymethlated cytosine in a reaction catalyzed by the TET family of enzymes.

depleted from vertebrate genomes through evolution (Bird, 1980). In contrast to the overall lack of CpG dinucleotides in the genome there are regions that contain clusters of CpG dinucleotides known as CpG islands (CGIs), by contrast these regions commonly contain unmethylated CpGs and are frequently associated with gene promoters (Bird, 1986). The original and most recognised algorithm used to define a CpG island was developed by Gardiner-Garden and Frommer (Gardiner-Garden, et al., 1987), and was further refined by Takai and Jones to exclude false positives associated with repetitive elements (Takai, et al., 2002). The modified algorithm contains three parameters: 1. GC content >55% 2. Observed CpG / Expected CpG >0.65, and 3. Length >500 bp. Bioinformatic analysis has identified 20,458 CGIs in the mouse genome and 37,531 in the human genome (Han, et al., 2008).

1.5.3 Enzymes responsible for DNA methylation

The DNA methyltransferase (DNMTs) family of enzymes are responsible for the establishment and maintenance of cytosine methylation in the form of 5mC. The mammalian family of DNMTs is comprised of *DNMT1, DNMT3A* and *DNMT3B* along with the accessory protein *DNMT3L* and the tRNA methyltransferase *DNMT2* (Table 1.1) (Goll, 2006). *DNMT1, DNMT3A* and *DNMT3B* function via the transfer of a methyl group from S-adenosylmethionine (SAM) to a cytosine residue of the DNA double helix (Goll, et al., 2005). The currently accepted model of DNA methylation establishment and maintenance is based on one first proposed by Holliday and Pugh (Figure 1.6) (Holliday, et al., 1975). In this model the *de novo* methyltransferase DNMT3A and DNMT3B predominately function to catalyse the establishment of DNA methylation in the germ line and developing embryo (Okano, et al., 1999, Xie, et al., 1999). This process is aided at certain genomic regions by the DNMT3L protein, which lacks the domain associated with SAM binding, but associates with DNMT3A and DNMT3B to modulate their catalytic activity (Bourc'his, et al., 2001, Suetake, 2004). Once methylation has been established, it is maintained

DNA methyltransferase enzyme	Function	Reference
DNMT1	maintenance	Bestor *et al.* (1988)
DNMT3a DNMT3b	establishment	Xie *et al.* (1999) Okano *et al.* (1999)
DNMT3L	no intrinsic enzymatic function, but modulates 3a and 3b activity	Bourc'his D. *et al* (2001) Suetake *et al. (2004)*
DNMT2	tRNA methyltransferase	Goll MG. *et al.* (2006)

Table 1.1 Known DNA methyltransferases with details of their methylation function.

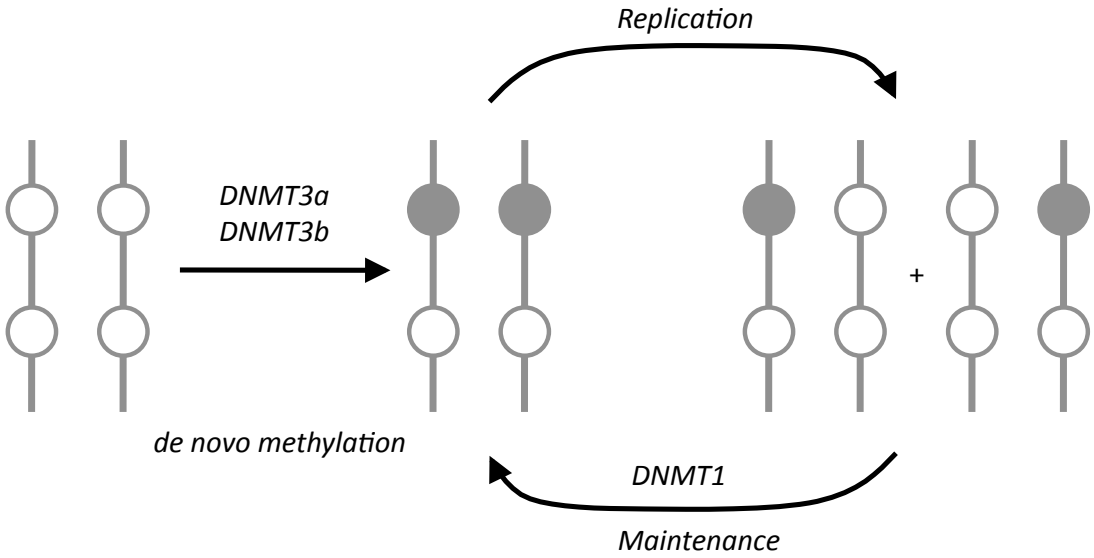


Figure 1.6 DNA establishment and maintenance model as proposed by Holliday and Pugh. Filled circles represent methylated sections of DNA, open circles represent unmethylated regions. *De novo* methyltransferases establish DNA methylation, and patterns of methylation are maintained throughout rounds of cell division and DNA replication via the activity of maintenance methyltransferases.

throughout rounds of cell division by the DNMT1 enzyme (Bestor, et al., 1988), which acts to fully methylated hemimethylated DNA, present after DNA replication. DNMT1 functions through its interaction with the E3 ligase ICBP90 (Np96 in mouse) (Avvakumov, et al., 2008) as ICBP90 demonstrates preferential affinity for hemimethylated DNA over symmetrically methylated DNA through its SET and RING-associated domains (Unoki, et al., 2004). This model for DNA methylation and establishment is supported by the abundance of DNMT3A and DNMT3B in ES cells, during early embryogenesis and in developing germ cells, and their subsequent down-regulation in differentiated somatic cells (Chen, 2002), and by the ubiquitous nature of DNMT1. This model delineates between the actions of DNMT1 and DNMT3A and DNMT3B in the establishment and maintenance mechanisms, however there is a growing body of evidence for overlap of *de novo* and maintenance functions. Knockout of DNMT3A and DNMT3B in ES cell culture reveals progressive loss of methylation over multiple cell divisions and an increase in the amount of hemimethylated sites suggesting that DNMT1 does not act as the sole maintenance enzyme, but that DNMT3A and DNMT3B also play a role in this process (Chen, et al., 2003).

1.5.4 Differential methylation and genomic imprinting

In mammals, Cytosine methylation is a prerequisite for cellular differentiation, development, X-chromosome inactivation and genomic imprinting (Bird, 2002). Most imprinted genes have been shown to associate near regions of DNA that harbour differential methylation patterns determined by the parent of origin from which the regions has been inherited, these are commonly referred to as differentially methylated regions (DMRs). Differential DNA methylation at an imprinted gene locus was first described at the *Igf2* locus where parental methylation differences were observed several thousand base-pairs upstream of the *Igf2* promoter (Sasaki, et al., 1992) and the role of differential methylation in imprinting has been confirmed by analyzing changes in imprinted gene

expression in mice deficient in DNA methyltransferase activity (Li, et al., 1993). Methylation may occur on either the maternal or paternally inherited allele, although maternally methylated DMRs are more common. There are 20 characterised DMRs, associated with imprinted loci that arise from differential methylation of DNA in sperm and oocytes. These are known as germline DMRs (gDMRs) (Table 1.2). A genome-wide screen for asymmetric methylation in adult mouse brain has identified a further 35 regions of differential methylation in the genome, many of which occur near a known imprinted domains (Xie, et al., 2012). A subset of DMRs has been shown, by targeted disruption of the DMR, to control genomic imprinting of the associated transcripts, and thus may be referred to as imprinting control regions (ICRs).

1.5.5 The nucleosome

The nucleosome is comprised of an octamer of histones, two of each H2A, H2B, H3 and H4. DNA is wrapped around each nucleosome in a 1.75 left handed superhelical turn encompassing 147 bp of DNA, which serves to densely package the DNA into the nucleus of the cell (Figure 1.7) (Luger, et al., 1997). In addition to their role as a DNA packaging device, the core histones may harbour post-translational modifications that can function to control the higher order chromatin structure by changing the mobility of the nucleosome, and can thus influence genomic functions by making regions of DNA more or less accessible to DNA binding factors and transcriptional protein complexes (Table 1.3). The post-translational modifications occur predominately on the histone tails (see Figure 1, B, C), which are unstructured regions of the histone proteins that protrude through the DNA, and are not required for assembly of the nucleosome octamer (Luger, et al., 1998). The four core histones are subject to a variety of enzymatically-mediated post-translational modifications to various amino acid residues (Table 1.4). Individual histone modifications do not function in isolation to regulate chromatin structure, making it difficult to define a correlation between a particular modification and specific

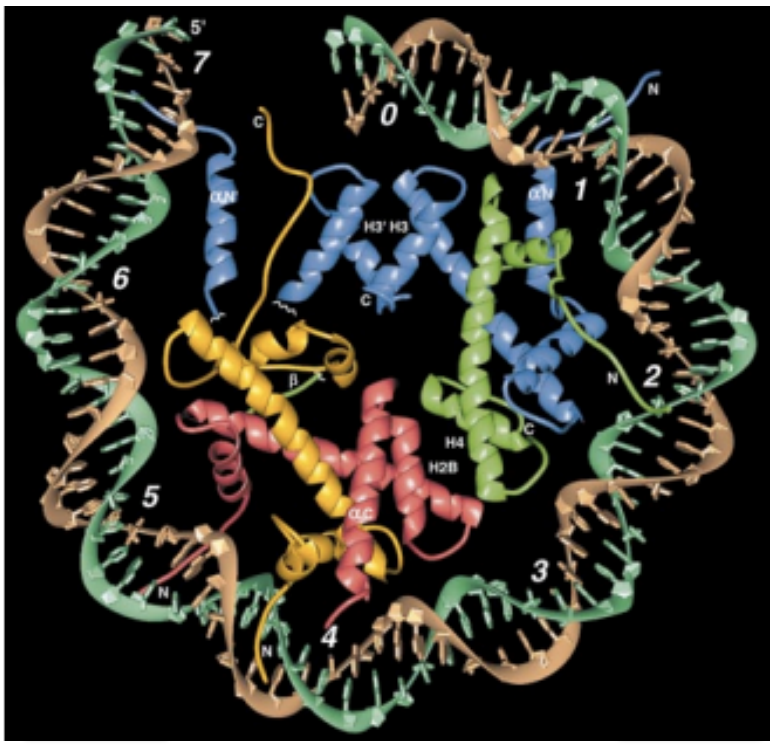
functional significance. However, it is recognised that some histone modifications seems to be involved in particular regions of genomic activity.

Differential histone modifications between active and silent alleles of imprinted genes is described at multiple imprinted loci including *Igf2, Snrpn, Igf2r, U2af-rs1, GNAS, Grb10, ndn, Dlk-gtl2* and *kcnq1* (Carr, et al., 2007, Fournier, et al., 2002, Grandjean, et al., 2001, Hu, et al., 2000, Lau, 2004, Lewis, et al., 2004, Li, 2004, Sakamoto, 2004, Sanz, et al., 2008, Umlauf, et al., 2004, Yang, 2003).

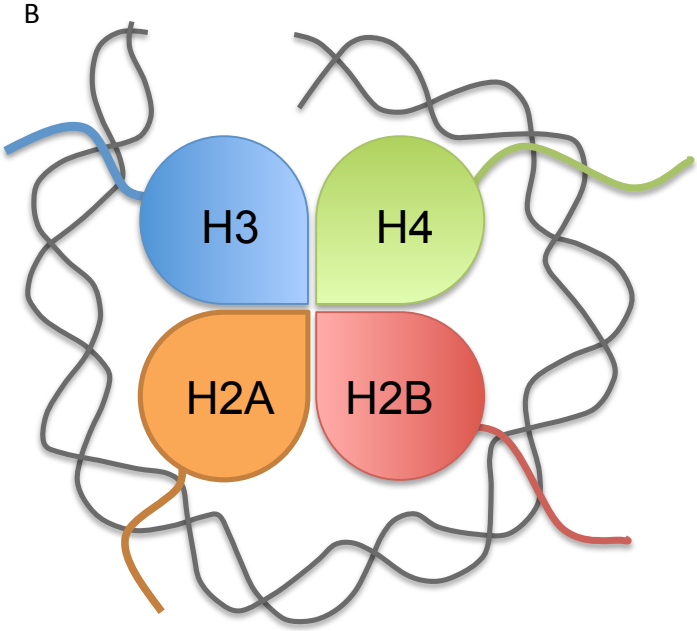
Chromosome	Imprinting control regions	Methylated allele
chr2	*Mcts2*	M
chr2	*Nespas*	M
chr2	*Gnas1a*	M
chr2	*Nnat*	M
chr6	*Peg10*	M
chr6	*Mest*	M
chr6	*Nap1l5*	M
chr7	*Zim2 (Peg3)*	M
chr7	*Snurf/Snrpn*	M
chr7	*Inpp5fv2*	M
chr7	*H19/Igf2*	P
chr7	*Kcnq1ot1*	M
chr9	*Rasgrf1*	P
chr10	*Zac1*	M
chr11	*Grb10*	M
chr11	*U2af1-rs1*	M
chr12	*Gtl/Dlk1*	P
chr15	*Peg13*	M
chr17	*Igf2r/Air*	M
chr18	*Impact*	M

Table 1.2 List of the known germline differentially methylated regions (gDMRs). Maternally methylated alleles are indicated in red and the paternally methylated alleles in blue.

A



B



C

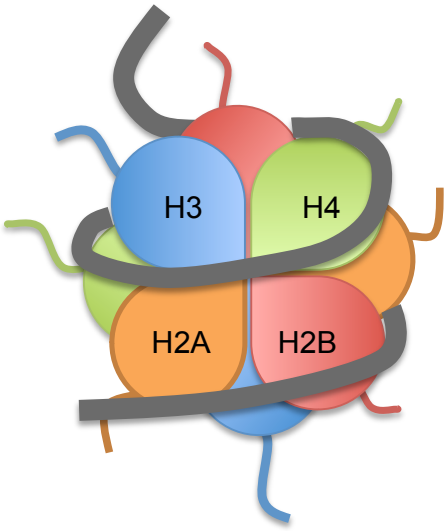


Figure 1.7 The nucleosome. (A) and (B) 73bp half super-helix showing half of the nucleosome octamer. (A) 2.8 Å resolution crystal structure taken directly from (Luger, 1997) shows DNA double helix (orange and green) wrapped around the central histone octamer. (B) Cartoon representation of the crystal structure with H2A shown in orange, H2B in red, H3 in blue and H4 in green. (C) Cartoon showing full 1.73 superhelical turn and full nucleosome comprised of 8 histone proteins.

histone modification	putative functions	reference
H3K36me3	associates with elongation phase of ongoing transcription, in particular at exonic gene regions	Kolasinska-Zwierz, 2009
H4K20me3	associated with constitutive heterochromatin	Kourmouli, 2004
H3K27me3	associated with long-term, polycomb mediated gene silencing via formation of polycomb bodies	Simon, 2009 ; Cheutin, 2012
H3K4me3 and H3K4me3	associated with transcriptionally active and poised genes partiularly at the 5' end and particularly associated with CpG islands	Bernstein, 2002 ; Weber, 2007
H3k9me	associated with heterochromatic regions	Grewal, 2002

Table 1.3 Function of histone modifications. Examples of histone modifications with putative functions.

enzyme family	examples of specific enzymes	example of histone modification
Acetyltransferases	HAT1, CBP/P300, H2A, PCAF/ GCN5, TIP60, H3, HB01	H3K9ac
Deacetylases	SirT2	H4K16
Lysine Methyltransferase	SUV39H1, SUV39H2, G9a, ESET/ SETDB1, EuHMTase/GLP, CLL8, MLL1, MLL2, MLL3, MLL4, MLL5, SET1A, SET1B, ASH1, SET2, NSD1, SYMD2, DOT1, Pr-SET, SUV4, SUV420H2, EZH2, RIZ1	K3K4me, H3k36me
Lysine Demethylases	LSD1/BHC110, JHDM1a, JHDM1b, JHDM2a, JHDM2b, JMJD2A/ JHDM3A, JMJD2B, JMJD2C/ GASC1, JMJD2D	H3K4, H3K36
Arginine Methlytransferases	CARM1, PRMT4, PRMT5	H4R3me

Table 1.4 Histone modifying enzymes. Histone modifying enzyme families listed with details of their methylation function and examples of known genes in the enzyme family.

1.6 Epigenetic reprogramming in the germline

The importance of correct parent-of-origin specific methylation patterns for controlling genomic imprinting was first proposed at the *Igf2r* imprinted locus by Stoger *et al.* (Stöger, et al., 1993) and this mechanism for passing parent-of-origin specific methylation through the germline means that methylation imprints must be erased and re-established during gametogenesis, dependant on the sex of the animal, in order to allow correct transmission of methylation marks to the next generation. These methylation imprints must also be protected from the global DNA demethylation which occurs post-fertilization.

1.6.1 Methylation in the germline

Primordial germ cells (PGC) are derived from the epiblast at around e7.5, and migrate to the genital ridge (Figure 1.8) where they undergo extensive DNA methylation and histone remodelling conducive to their adoption of germ cell fate (Hajkova, et al., 2008, Hajkova, 2010). During this migration, methylation is erased from a level of greater than 70% to less than 10% CpG methylation (Figure 1.9, stage 1), with only a few select genomic elements remaining methylated. This demethylation event provides the environment for the methylation in the germ cell to be reset dependent on the gender of the embryo.

Gender specification in mouse occurs at around e12.5, and PGC methylation is re-established during spermatogenesis and oogenesis during different stages of embryonic and adult development (Figure 1.9, stage 2). In male embryos DNA methylation begins early in prospermatagonia that have been arrested in the G1 phase of mitosis, and is completed before birth. After birth, multiple cell divisions will then occur between the establishment of methylation and the maturation of sperm.

In female embryos, oocyte DNA methylation is established after birth during the follicular growth phase (Hajkova, et al., 2002). Primary oocytes begin meiosis and *de novo* methylation occurs during the diplotene phase

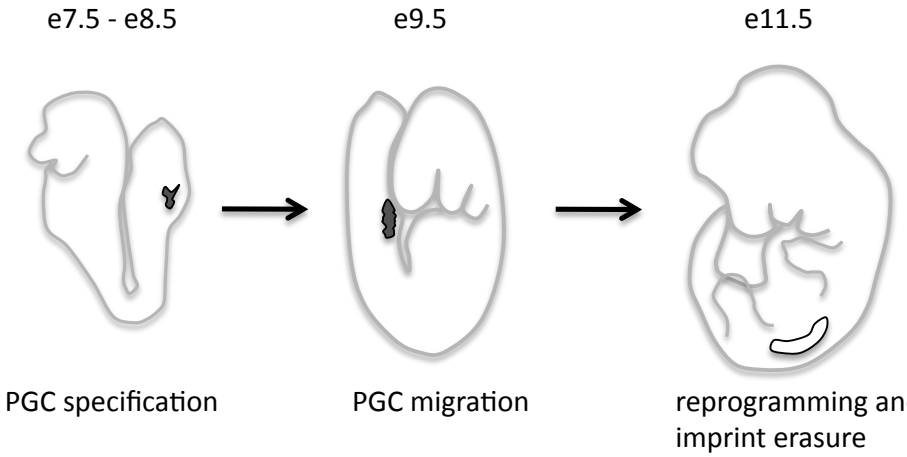


Figure 1.8 Epigenetic reprogramming in the germline. Embryo outline is shown in grey and the primordial germ cell regions are outlined in black. Primordial germ cells (PGCs) are derived from the epiblast at around e7.5 and migrate to the genital ridge by e11.5, where they undergo DNA methylation and histone remodeling. Methylation is erased during migration.

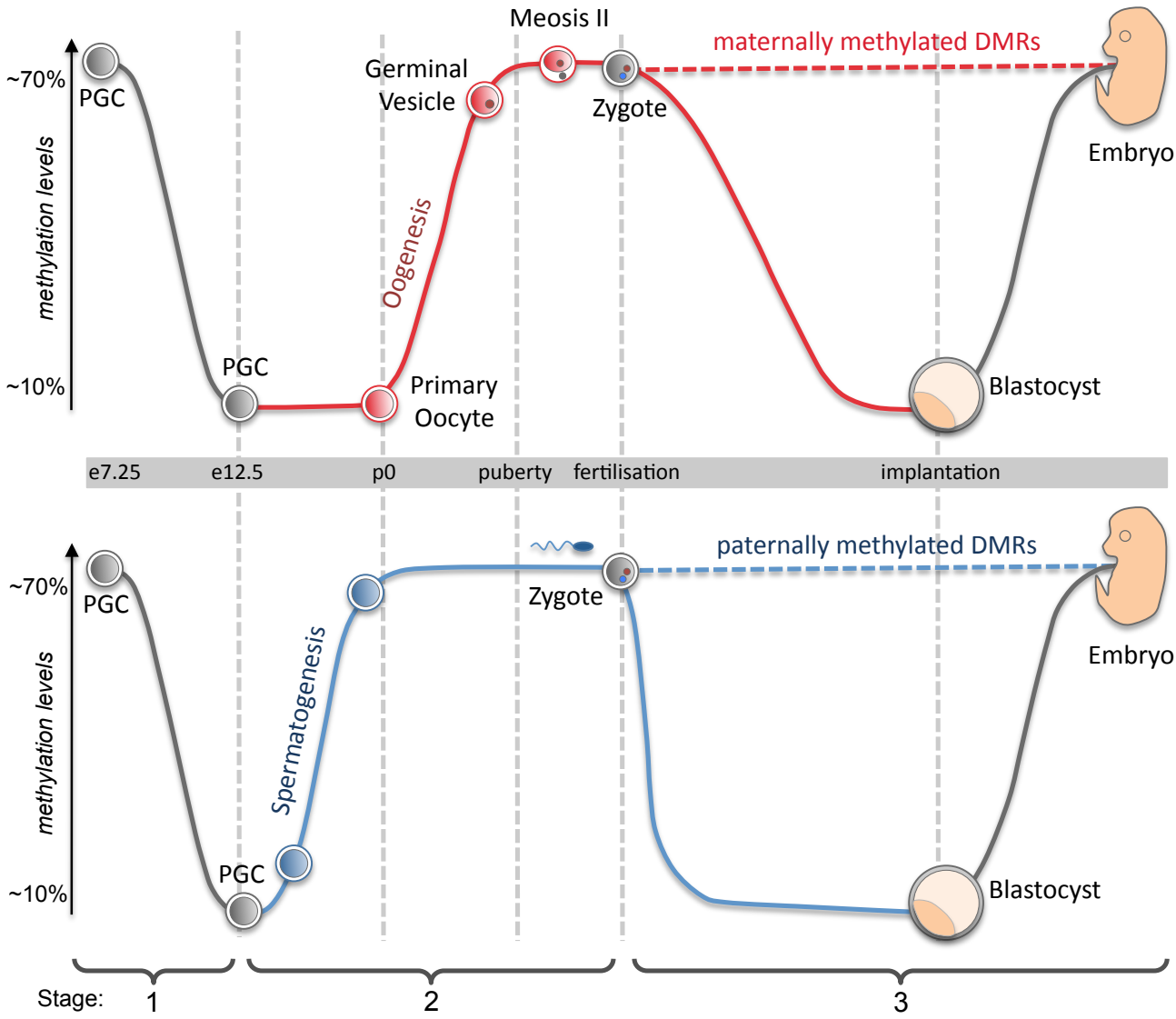


Figure 1.9 Methylation erasure and establishment during germ cell development. Red line indicates methylation establishment and erasure in females and the blue line in males. Stage 1, methylation is erased during germ cell migration, which occurs between e7.5 and e12.5. Stage 2, gender specification occurs at e12.5 and the re-establishment of methylation occurs at different stages of embryonic development in spermatogenesis and oogenesis. Stage 3, after fertilization both paternally and maternally inherited genomes are de-methylated with exception of several genomic regions including methylation at differentially methylated regions (DMRs). Methylation levels are approximate.

of prophase I within the non-dividing oocyte (Smallwood, et al., 2011). At this stage each cell is 4N because as DNA replication has occurred but the developing oocyte has not yet divided. This means that methylation must occur on all 4 chromatids to ensure each final haploid oocyte has the correct methylation pattern.

1.6.2 Establishing imprinted methylation patterns in the germline

The mechanisms for establishing DNA methylation at maternally methylated imprinting control regions in postnatal ovary during the oocyte growth phase have been well studied, with *Dnmt3l* the first methyltransferase family enzyme to be specifically implicated in this process. Oocytes of *Dnmt3l*$^{-/-}$ females exhibit reduced methylation at the imprinted DMR and progeny of the *Dnmt3l* knockout animals lack of methylation on both alleles at maternally methylated gDMRs compared to the wildtype controls, whereas examination of the paternally methylated *Igf2/H19* locus reveals no difference from the wildtype (Bourc'his, et al., 2001). Research using mouse with conditional mutation of *Dnmt3a* and *Dnmt3b* in germ cells has identified co-ordinate regulation of Dnmt3l with Dnmt3a as being essential for the establishment of maternal methylation imprints, with Dnmt3b not implicated in this process (Kaneda, et al., 2004, Smallwood, et al., 2011). The non-catalytic Dnmt3l protein stimulates DNA methylation activity via a direct interaction with Dnmt3a (Suetake, 2004), and there is growing evidence that this process may also be regulated via interaction with histones proteins (Jia, et al., 2007, Ooi, et al., 2007). Evidence from study of mouse embryonic stem cells demonstrates that Dnmt3l methyltransferase family enzyme binds directly to the N terminus of the H3 histone, associating with the first 7 N-terminal amino acids, and dependant on the methylation of the lysine four residue (Ooi, et al., 2007). This provides evidence that histone-mark mediated interactions between the nucleosome and the methyltransferase enzymes could be involved in the establishment of parent-of-origin specific methylation marks at differentially methylated regions (Figure 1.10).

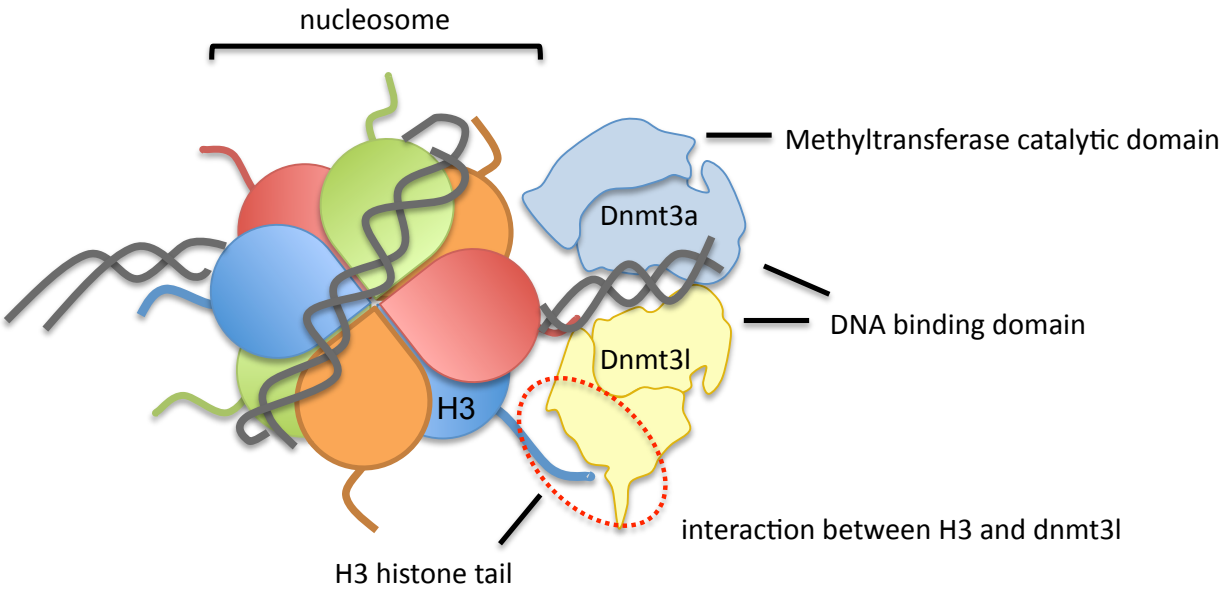


Figure 1.10 Interaction of Histone H3 and Dnmt3l. Evidence from the study of mouse embryonic stem cells suggests Dnmt3l may interact with histone H3 subunit to mediate methyltransferase activity. This model is proposed by Ooi *et al.* (2009), interaction is indicated in red.

Establishment of methylation imprints in prospermatagonia occurs earlier during development and is fully established by p0 (Kato, et al., 2007). Of the three known paternally methylated gDMRs (*H19, Dlk1/Gtl2* and *Rasgrf1*) independent study of each locus using *Dnmt3a* and *Dnmt3b* germline conditional knockout mice has shown that *H19* and *Dlk1/Gtl2* DMRs require Dnmt3a in order to establish methylation during spermatogenesis, with Dnmt3b not being required. Germline methylation at the *Rasgrf* gDMR requires both Dnmt3a and Dnmt3b, with Dnmt3b playing the predominant role in methylation at this locus (Kato, et al., 2007).

The establishment of methylation marks at imprinted regions during gametogenesis is the first part the process to establish and transmit a methylation imprint through the germline as gDMRs must also resist a second wave of global demethylation that occurs after fertilization and allows cell differentiation to begin (Figure 1.9, stage 3). Post-fertilization the maternally and paternally inherited genomes behave differently within the zygote, DNA methylation in the paternal pronucleus is actively erased by a mechanism involving the Tet3 enzyme. Tet3 expressed from the maternal genome within the oocyte converts 5-methylcytosine (5mC) into 5-hydroxymethylcytosine (5hmC) that serves as an intermediate between 5mC and unmethylated cytosine (Gu, et al., 2011, Iqbal, et al., 2011, Wossidlo, et al., 2011). The maternal pronucleus is not subject to the same Tet3 conversion, and methylation is instead lost passively during development due to the fact that Dnmt enzymes do not maintain DNA methylation during replication. Methylated cytosines in gDMRs are resistant to this second wave of demethylation during early development, and the mechanisms responsible for this are not fully understood. Zfp57 in combination with its co-factor, Kap1 has been shown to be involved in methylation maintenance at several, but not all imprinted loci (Li, et al., 2008, Quenneville, et al., 2011, Zuo, et al., 2012), which points towards a role for the Zfp57 and other KRAB zinc finger protein family members as

candidates, indeed mutation in ZFP57 in humans is associated with hypomethylation at six imprinted loci (Mackay, et al., 2008).

piRNAs may be involved in *de novo* methylation at the paternally methylated *Rasgrf* DMR (Watanabe, et al., 2011). During spermatogenesis, piRNAs bind to a non-coding RNA transcribed from the *Rasgrf* locus by recognising the sequence of an LTR-type retrotransposon called *RMER4B* and result in recruitment of the *de novo* methylation machinery to the DMR. Disruption of the piRNA pathway, or of expression of the non-coding RNA, prevents methylation and imprinting of *Rasgrf1.*

1.7 Mechanisms of imprinting

1.7.1 Insulator model of imprinting

This section has been adapted with minor changes from Prickett & Oakey, 2012 (Prickett, et al., 2012).

The most well studied example of a cluster that utilizes insulator binding in order to control genomic imprinting is the *Igf2/H19* cluster (Bartolomei, et al., 1991, DeChiara, et al., 1991). In most tissues *H19* is maternally expressed with *Igf2* being read exclusively from the paternal allele in mouse and human. A germline DMR, methylated on the paternal allele and unmethylated on the maternal allele, which lies in between the *H19* and *Igf2* genes controls imprinting of both genes (Thorvaldsen, et al., 1998, Tremblay, et al., 1995) and this DMR sequence contains a region that binds the zinc finger protein CTCF (Bell, et al., 2000, Hark, et al., 2000). Binding of CTCF at other non-imprinted loci had been previously shown to affect chromatin looping and can mediate chromosomal interactions *in cis* between enhancer elements that may lie many kbps away from the gene promoter (Bell, et al., 1999). Imprinting at the *Igf2/H19* locus is dependent on the binding of CTCF in an allele-specific manner at the DMR (Engel, 2006), binding is observed exclusively on the unmethylated maternal allele as CTCF binds preferentially to unmethylated DNA (Mukhopadhyay, 2004). Chromosome conformation capture (3C) experiments support the idea that

CTCF binding mediates allele-specific chromosomal environments that influence gene expression (Engel, et al., 2008) by controlling interaction of an enhancer distal to *H19* that controls expression of both *H19* and *Igf2* (Leighton, et al., 1995). On the paternally inherited chromosome, methylation prevents CTCF binding at the DMR, thus the enhancer can interact with the promoter *Igf2* allowing for paternal transcription of *Igf2*. Conversely, CTCF binding on the maternal allele blocks this interaction and instead causes the enhancer to promote transcription of *H19* (Figure 1.11, A).

The *Igf2/H19* locus has provided an excellent model for examining the differential binding of a protein between the two alleles of a gene where the transcriptional consequences could be compared directly in the same cell or tissue. CTCF has also been studied at a number of other imprinted loci and it binds the unmethylated allele at the DMRs of *Rasgrf1, Peg1* and the *KvDMR (Kcnq1ot1)* (Fitzpatrick, et al., 2007, Singh, et al., 2011, Yoon, et al., 2005) (although CTCF binding at the *KvDMR* may not be allele-specific (Lin, et al., 2011)). CTCF binding has been shown at the *Grb10* DMR using *in vitro* assays (Hikichi, 2003) and *in vivo* CTCF binds the *Grb10* gene body in mouse fetal liver (Mukhopadhyay, 2004). It has been postulated that CTCF-mediated regulation may be one of the two major control mechanisms at all ICRs (Kim, et al., 2009, Lewis, et al., 2006). Cohesin has also been linked to imprinting mechanisms through its association with CTCF at the *Igf2/H19* ICR (Stedman, et al., 2008) and its role in the allele-specific organization of higher-order chromatin at the *Igf2/H19* locus has been proposed (Nativio, et al., 2009).

1.7.2 Genome-wide CTCF binding and imprinting

Genome-wide assays using ChIP-Seq in mouse ES cells (Chen, et al., 2008, Kagey, et al., 2010) have shown that CTCF and Cohesin, another DNA binding protein, are frequently bound together across the genome. Binding sites coincide with the gDMRs of *Peg13, Zim2 (Peg3), Peg10, Grb10* and

Mest. However, these studies do not observe any binding of CTCF or Cohesin to the *Igf2/H19* ICR suggesting that cell type specificity of binding events and/or differences in the sensitivity of the experimental approaches exists in locus-specific versus genome-wide assays for CTCF binding. These genome-wide data are limited to ES cells (Chen, et al., 2008, Cuddapah, et al., 2009, Martin, et al., 2011) where imprinting is dispensable without affecting viability in culture (Frost, et al., 2011, Kim, et al., 2007, Rugg-Gunn, et al., 2007), CTCF and Cohesin have not been investigated genome-wide in differentiated tissue in a *parent-of-origin* specific manner.

CTCF has a large binding motif (~20bp) with high specificity, which is evolutionarily conserved (Martin, et al., 2011).

1.7.3 Non-coding RNA model of imprinting

A second model of imprinting control is the macro ncRNA mechanism. Macro ncRNAs are large, lack sequence conservation across species and are unspliced or inefficiently spliced. The imprinted locus of *Igf2* and its antisense transcript *Air*, and the *Kcnq1/Kcnq1ot1* locus provide evidence for this model of imprinting.

The *Igf2r/Air* locus is controlled by a maternally methylated DMR that lies within the *Igf2r* gene body. On the unmethylated paternal allele the *Air* non-coding RNA is transcribed in the opposite direction to *Igf2*, and this functions *in cis* to silence *Igf2r*, *Slc22a2* and *Slc22a3* on the paternal allele. On the maternal allele hypermethylation of the DMR prevents transcription of *Air* and *Igf2r, Slc22a2* and *Slc22a3* remain transcriptionally active (Figure 1.11, B) (Sleutels, et al., 2002).

The *Kcnq1* locus contains one paternally expressed transcript, *Kcnq1ot1*, which encodes a long ncRNA, and several maternally expressed protein-coding genes, including *Cdkn1c*, *Mash2*, *Phlda2*. Imprinting is controlled by the maternally methylated DMR known as the KvDMR1, which is located in

an intronic region of *Kcnq1* at the promoter of the *Kcnq1ot1* (Mancini-DiNardo, 2003). On the hypermethylated maternal allele, *Kcnq1ot1* is not transcribed and the surrounding genes remain active, conversely on the hypomethylated maternal allele of KvDMR1, *Kcnq1ot1* is active and the surrounding gene cluster is silenced *in cis* (Mancini-DiNardo, 2006) (Figure 1.11, C).

Transcriptional interference (TI) as a mechanism for gene silencing *in cis*, has been shown in non-mammalian eukaryotic cells of yeast and drosophila, and is caused by the process of transcription of one gene interferes with the transcription of another independent of RNA. In the *Igf2r/Air* cluster, the antisense *Air* transcript overlaps the promoter of the *Igf2r* gene. If the *Air* transcript is truncated so that it no longer overlaps the *Igf2r* promoter this leads to loss of silencing, suggesting TI may provide the mechanism for *in cis* imprinted gene silencing (Pauler, et al., 2012).

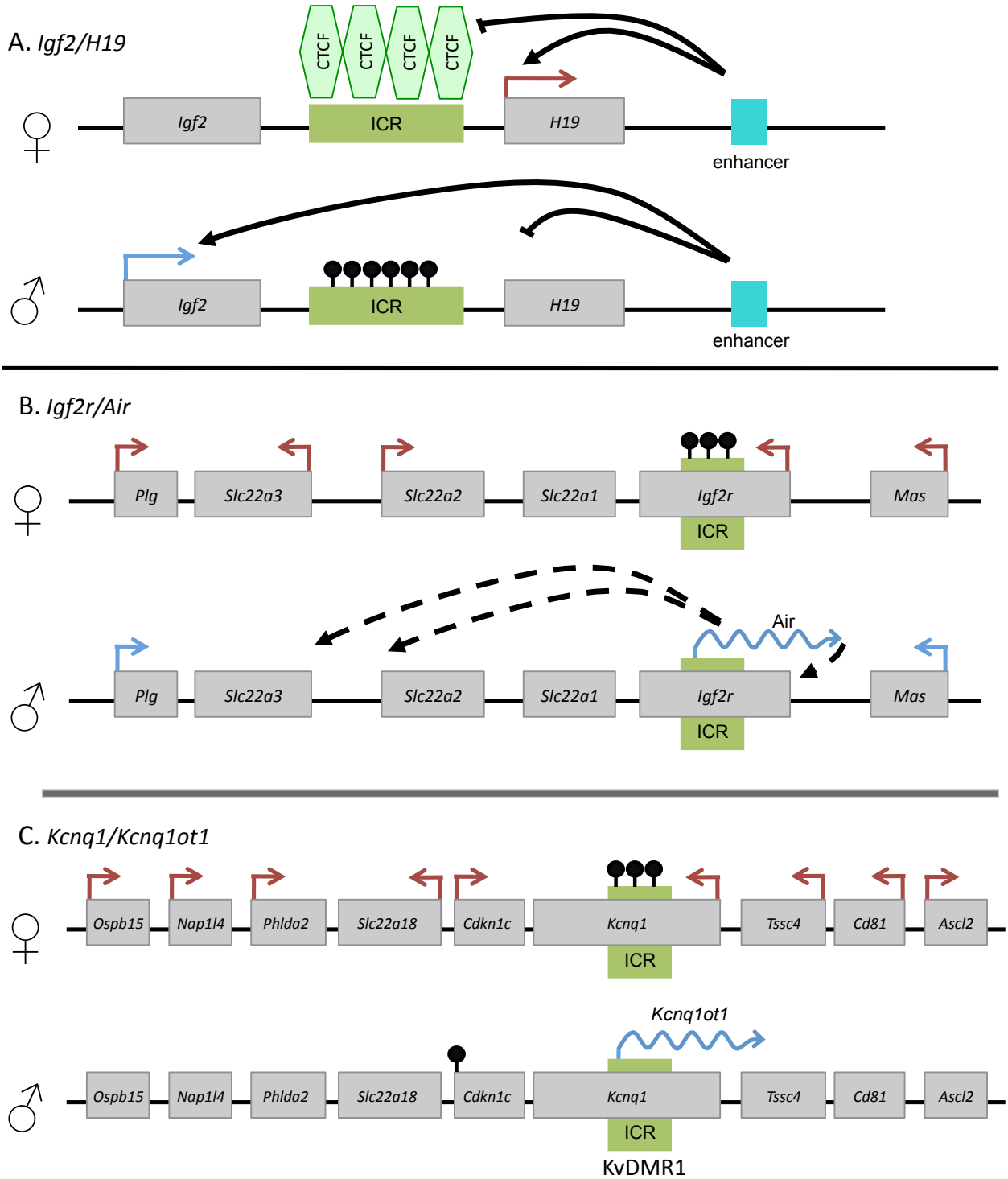


Figure 1.11 Mechanisms of genomic imprinting. Grey boxes denote genes, and arrows indicate transcription with maternal transcription shown in red and paternal in blue. Lollipops indicate regions of DNA methylation at differentially methylated regions. Green boxes denote the imprinting control region. Dashed arrows demonstrate repression in response to transcription of *Air.* Black arrow demonstrates long-range chromosomal interaction. (A) shows the insulator model of imprinting at the *H19/Igf2* imprinted locus, and (B) and (C) illustrate the non-coding RNA model of imprinting at *Igf2r/Air* (B) and *Kcnq1/Kcnq1ot* (C). Loci are not drawn to scale.

1.8 Dopa decarboxylase

1.8.1 Function of Dopa Decarboxylase

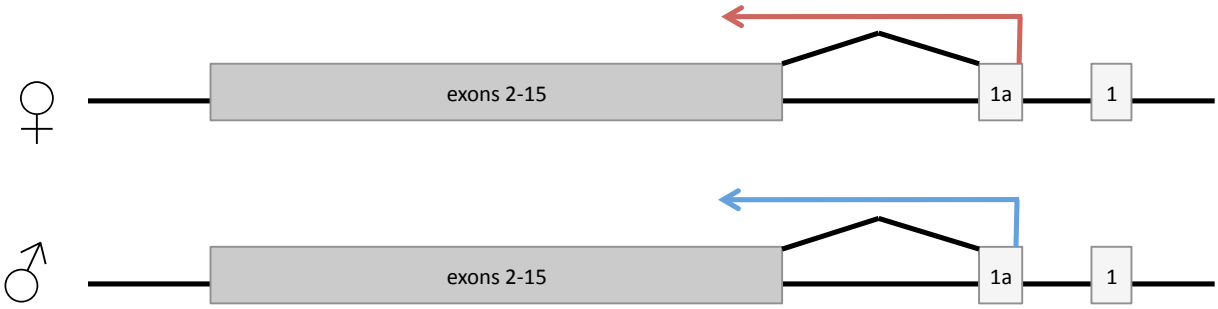
Dopa decarboxylase (Ddc) functions to catalyze the biosynthesis of catecholamine neurotransmitters and serotonin, and as such is expressed at high levels in the dopaminergic and seretonergic neurons of the central and peripheral nervous systems (Christenson, et al., 1972). Polymorphisms in the human *DDC* gene have been correlated with multiple psychiatric disorders include bi-polar disorder (Borglum, et al., 1999) and attention deficit disorder (Kirley, et al., 2002), furthermore reduced *DDC* expression is observed in the neurodegenerative disorder Parkinson's disease (Ichinose, et al., 1994).

Substantial Ddc expression is also present in the liver, pancreas, kidney, intestine and heart (Aguanno, et al., 1996, Hayashi, et al., 1990, Menheniott, et al., 2008) suggesting additional functional roles for Ddc in non-neuronal tissues. Non-neuronal Ddc functions in the kidney, where it is expressed in the proximal tubule (Hayashi, et al., 1990), to regulate solute and water homeostasis by controlling tubule transporter activity in the nephron via interrenal dopamine concentrations (Zhang, et al., 2011). Recent evidence has shown a putative role for Ddc in regulating serotonin concentration in the heart of adult mice (Rouzaud-Laborde, et al., 2012). In the developing heart Menhenniot *et al* report elevated levels of *Ddc* expression in the trabecular cardiomyocytes, suggesting a novel functional role for Ddc in ventricular development (Menheniott, et al., 2008).

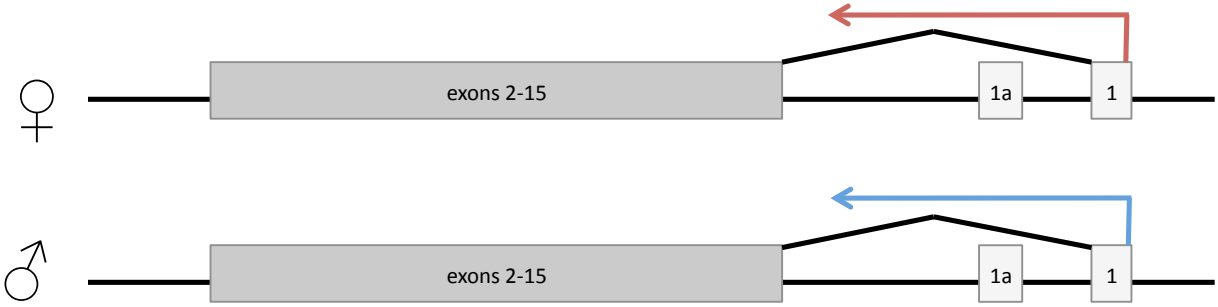
1.8.2 Imprinting of *Ddc* in mouse

In mice, *Ddc* maps to proximal chromosome 11, approximately 25 kbps from a neighbouring gene *Grb10* that is imprinted in mouse and human, and these genes are located in a region of conserved linkage on human chromosome 7. There are two long transcripts of *Ddc*, termed *Ddc_canonical* and *Ddc_exon1a*, both of which contain identical coding sequence and which differ only in the 5' untranslated region of the gene sequence (Figure 1.12) (Menheniott, et al., 2008). *Ddc_canonical* is

Ddc expression in Brain



Ddc expression in Liver



Ddc expression in Heart

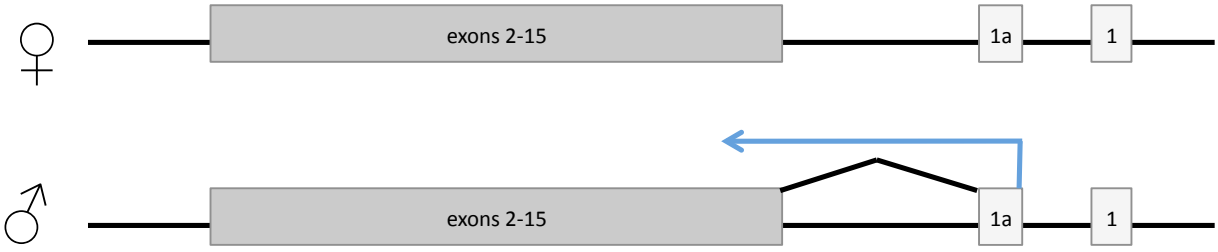


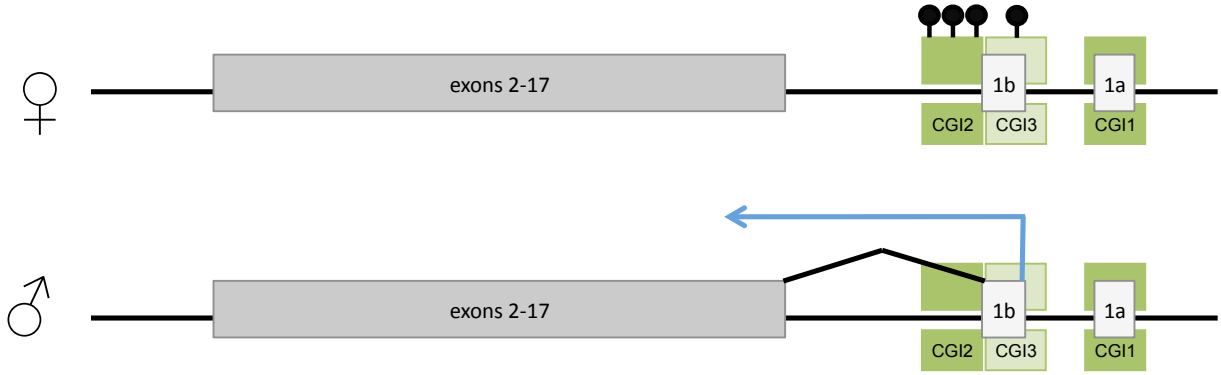
Figure 1.12 *Ddc* gene expression patterns in the heart, brain and liver. Exons 1 and 1a are shown individually (light grey), exons 2 to 15 represents the gene body and are shown together, ignoring introns (dark grey). Transcription is indicated by arrows with red arrows indicating maternal expression and blue arrows paternal expression. *Ddc_exon1a* initiates from exon 1a and is expressed from both parental allele in brain, but only the paternally inherited allele in the developing heart, with the maternal allele epigenetically silenced. *Ddc_canonical* is expressed in peripheral tissues such as liver and kidney, this transcript initiates from exon 1 and is expressed from both parental alleles.

expressed in a bi-allelic fashion in Liver and Kidney. *Ddc_exon1a* is expressed predominately in the Brain and developing heart and in the heart *Ddc_exon1a* is imprinted, being expressed only from the paternal allele (Menheniott, et al., 2008). There is contradictory evidence for imprinting of *Ddc_exon1a* expression in brain, Menheniott *et al.* report that *Ddc_exon1a* is expressed in a bi-allelic fashion in brain, this was assayed in microarray screens on tissue from mice with uniparental duplication of the proximal region of chromosome 11, and by using locus specific allele-specific sequencing assays on neonatal tissues from hybrid subspecies of mice (Menheniott, et al., 2008). Gregg *et al.* report, using an allele-specific RNA-seq screen of gene expression in 3 week old whole mouse brain from hybrid subspecies, that there is a maternal bias of *Ddc_exon1a* (Gregg, et al., 2010), however a more recent re-analysis of these data suggest that this may be a false positive result (DeVeale, et al., 2012). Thus, *Ddc_exon1a* is the first example of a gene imprinted tissue-specifically only in heart. *Ddc_exon1a* expression is present in the developing heart at e13.5 and peaks at the newborn stage before falling to low levels by three weeks (Menheniott, et al., 2008).

1.8.3 *Ddc* imprinting control

A germline differentially methylated region that resides in the 5' UTR of the neighbouring gene *Grb10* controls *Ddc_exon1a* expression in the developing mouse heart. Deletion of the DMR on the paternal allele results in absence of expression of *Ddc_exon1a* transcript in heart, and perturbs the imprinted expression of *Grb10* in heart and brain (Shiura, et al., 2009). It is unclear, however, how deletion of the paternal DMR affects expression of *Ddc_exon1a* in brain. *Grb10* belongs to a small family of cytoplasmic adapter proteins that mediate interactions between cell surface receptor tyrosine kinases and various signalling molecules. It is a rare example of an imprinted gene that is expressed exclusively from the paternal allele in some tissues and from the maternal allele in other tissues (Figure 1.13). *Grb10* expression has been examined in embryogenesis where expression

Grb10 expression in Brain



Grb10 expression as seen in liver and peripheral tissue

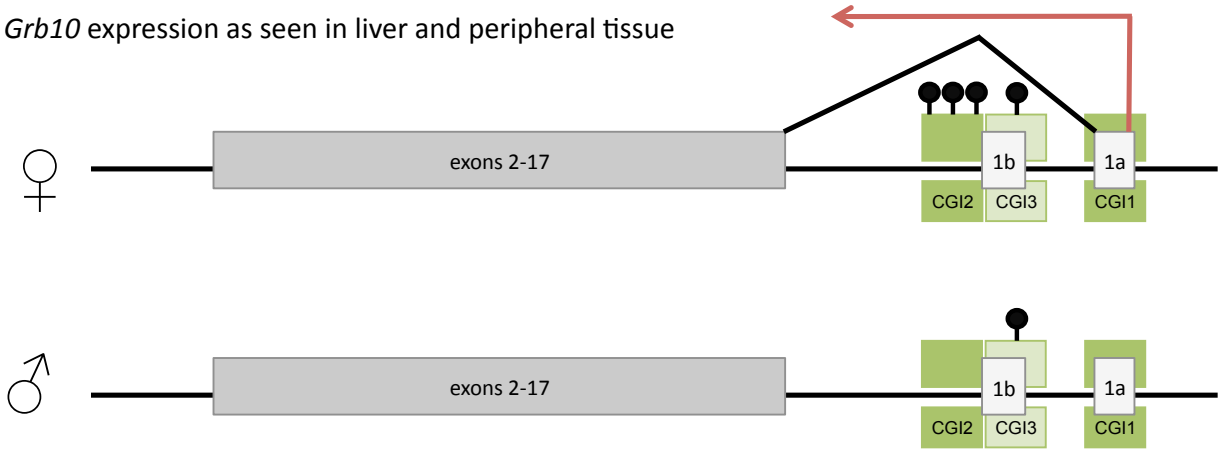


Figure 1.13 Expression of the *Grb10* gene. Exons 1a and 1b are shown individually (light grey), exons 2 to 15 represents the gene body and are shown together, ignoring introns (dark grey). Transcription is indicated by arrows with red arrows indicating expression from the maternally inherited allele and blue arrows expression from the paternally inherited allele. Dark green boxes represent CpG islands, and the light green box represents a regions rich in CpG dinucleotides but that does not meet the criteria for a CpG island due to its length. Filled lollipops represent regions of DNA methylation. CGI2 and CGI3 form part of the imprinting control region. *Grb10* is expressed from the paternal allele in brain, and the maternal allele in peripheral tissues such as liver and muscle.

from the maternal allele occurs in a variety of muscle tissues including facial, intercostal, limb and cardiac muscle, the diaphragm and the tongue, maternal *Grb10* expression is also seen in the choroid plexus and meninges of the brain, the liver, kidneys, lungs, adrenal gland and pancreas. Paternally expressed *Grb10* is seen in the brain, limb cartilage, ribs, head, long bones, lungs, gut and punctuated expression in the developing heart (Charalambous, et al., 2003). Maternal deletion of Grb10 results in fetal overgrowth (Charalambous, et al., 2003) and mice with a paternal deletion of Grb10 exhibit increased social dominance (Garfield, et al., 2011), thus *Grb10* is able to influence different physiological processes dependant on tissue-specific imprinted expression from different parental alleles.

1.8.4 Tissue-specific methylation at the Grb10 DMR

A DMR located at the promoter of *Grb10* controls imprinted expression of *Grb10.* The *Grb10* DMR exhibits tissue-specific methylation differences, which may account for tissue-specific differences in imprinted expression of *Grb10*. There are three distinct CpG rich regions of DNA at the 5' end of the *Grb10* gene, these have been termed CGI1, CGI2 and CGI3. CGI1 is located at the 1A exon of *Grb10* and remains unmethylated on both the maternal and paternal copy in all tissues. The *Grb10* imprinting control region comprises both CGI2 and CGI3. CGI2 is located just over 10 kbps upstream of CGI1 near an alternate promoter 1B exon for *Grb10*. CGI2 constitutes a germline DMR conserved in humans and mouse and is methylated on the maternal allele. CGI3 lies between CGI1 and CGI2, and is specific to mouse. CGI3 is differentially methylated in brain with methylation on the maternal allele and here Grb10 is paternally expressed. In liver and kidney, where *Grb10* expression comes from the maternal allele, CGI3 is fully methylated on both alleles (Figure 1.13) (Arnaud, et al., 2003). *Grb10* transcripts also appear to initiate from different leader exons dependent on whether they are expressed from the maternal or the paternal allele (Figure 1.13). Maternally expressed *Grb10* initiated from exon 1A in liver and kidney, whereas paternally expressed *Grb10* initiates

from exon 1B in the brain (Arnaud, et al., 2003), suggesting tissue-specific changes in methylation may controls a switch in promoter use and direction of imprinting is dependent on which promoter of *Grb10* is active.

1.8.5 Tissue specific imprinting

Of the imprinted genes listed on the Web Atlas of Murine genomic Imprinting and Differential EXpression (WAMIDEX) (Schulz, et al., 2008), at least 23 genes display tissue-specific imprinting, where a gene in expressed in a bi-allelic fashion in multiple tissues yet appears imprinted only in one specific tissue or cell type. The majority of these exhibit imprinted expression only in extra-embryonic tissues, specifically placenta (48%) and yolk sac (9%), or only in the brain, including specific subsets of brain regions (39%). *Ddc* is the only published example of a tissue-specific imprinted gene where imprinting is not localized to the brain or extra-embryonic tissue (Figure 1.14) (Menheniott, et al., 2008). *Ddc* demonstrates a mechanism of tightly controlled epigenetic regulation, which varies between heart and other tissue types. The *Ddc_exon1a* variant is expressed in brain and heart, and is bi-allelic in brain, but expressed from the paternal allele only in the developing heart. This organization of transcripts and expression patterns indicate that epigenetic differences between brain and heart at the *Grb10* DMR are key to controlling imprinted expression, and it is possible that the same epigenetic differences which governs imprinting differences of *Grb10* also control imprinting differences of *Ddc_exon1a* (Prickett, et al., 2012).

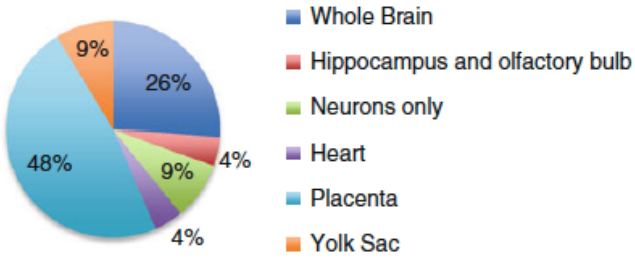


Figure 1.14 Location of tissue-specific imprinting. A large proportion of tissue-specific imprinted transcripts are imprinted in extra-embryonic tissues of the placenta (48 %), yolk sac (9 %) and whole brain (26 %) including subsets of brain regions including the hippocampus (4 %) and neurons (9 %). Only *Ddc* is tissue-specifically imprinted only in heart. Figure taken from Prickett & Oakey, 2012.

1.9 Fetal and neo-natal heart development

1.9.1 Formation of the heart tube e7.0 - e8.5

Cardiomyocyte precursor cells arise from the mesoderm of the primitive streak of the embryo to form paired cardiogenic plates at embryonic day 7.5, adjacent to these plates, endothelial cells differentiate forming the left and right endocardial heart tube. These tubes then fuse across the midline to form the heart tube, which itself has an inner endothelial cell layer and an outer myocardial cell layer with cardiac jelly in between. By e8.0, the heart tube begins to beat, initially in an irregular manner settling to a regular contraction by e9.0 with the venous pole acting as pacemaker (Figure 1) (Savolainen, et al., 2009).

1.9.2 Looping e8.5 – e10.5

The heart tube loops towards the right axis between e8.5 and e10.5 with the atrial chambers locating dorsally to the primitive right and left ventricles, and all the future heart chambers undergo growth. By E9.5 the segmentation of the heart can be seen and the atrioventricular and outflow tract cushions and ridges, which will ultimately form the septal and valvular structures of the heart are visible (Kaufman, 1999).

1.9.3 Cell migration

There are three waves of cell migration of extracardiac cells into the developing heart that occurs during organogenesis. Mesenchyme derived pro-epicardium cells invade the entire external surface of the heart (Komiyama, et al., 1987), dorsal mesocardium cells are incorporated into the primodial atria, and cardiac neural crest cells that migrate to the outflow tract. Dorsal mesocardium functions to correctly orientate the atrial chamber and form the atrial septum. Cardiac neural crest cells contribute to the developing septation and valves of the outflow tract (Gitler, et al., 2003). The epicardium cells subsequently undergos epithelial to mesenchymal transformation (EMT) and form the endothelial and

smooth muscle cells of the coronary vasculature (Perez-Pomares, et al., 1997).

1.9.4 Ventricular morphogenesis e11.0 – e15.0

Myocardial cells, distinguished by distinctive sarcomeric proteins and myofibril assembly are present in the heart by e8.5 as the chambers are delineated, fully differentiated cardiomyocytes are found in the developing atrial and ventricular walls, and it is proliferation of these cardiomyocytes which increases the thickness of the chamber walls and allows the heart to pump under the increasing hemodynamic load associated with embryonic growth (Kirby, 2007). The major growth and development of the ventricular mass occurs between e11.5 and e14, with growth being fed from the compact epicardial layer of cells (Chen, et al., 2002). During this time a subset of cardiomyocytes form specializes structures called trabeculae, which project in the ventricular cavity forming a mesh-like structure that can compose as much as 80% of the myocardial mass at its peak. Trabeculae function to increase myocardial oxygenation, at a period before coronary circulation has been properly established (Wessels, et al., 2003). Between e13 and e14 the trabecular layer compacts rapidly, coinciding with ventricular septation and the establishment of coronary circulation, increasing the mass and thickness of the compact layer that forms the outer ventricle wall. Non-compaction of the trabecular layer presents serious functional consequences for the heart. Several mouse null mutants, for example *RXR-α*$^{-/-}$ animals demonstrate a complete lack of compaction, resulting in death by e14.0 due to a failure of the heart to deal with the increased load placed on the heart during later embryonic stages (Chen, et al., 2002). Partial or incomplete compaction of the myocardial layer is associates with heart failure and sudden cardiac death in humans (Jenni, et al., 2001, Oechslin, et al., 2011).

1.9.5 Myocardial remodelling e15– e18.5

Compared to earlier stages of heart development, changes that occur at around e15.0 are associated mainly with growth and cellular proliferation in the myocardium, and configuration of the right and left atria, valvular development and a change in orientation of the heart, as by this time the fetal heart has achieved its final prenatal configuration (Savolainen, et al., 2009). Cell proliferation and growth increases the mass of the ventricular myocardium and results in the trabeculae becoming compressed within the ventricular wall. This results in a significant increase of proportion and thickness of the compact myocardium (Sedmera, et al., 2000).

1.10 Aims

1.10.1 Parent-of-origin specific binding of CTCF

CTCF binding occurs at the imprinted control region of several imprinted loci including the *H19/Igf2* imprinted locus, however a systematic screen of parent-of-origin specific binding of CTCF at all known imprinted gDMRs in a single tissue type has yet to be performed. Using chromatin immunoprecipitation followed by next generation sequencing on p21 mouse brain, the incidence of CTCF binding at all known gDMRs will be explored, and CTCF binding assessed in a parent-of-origin specific manner. Further analysis of the ChIP-seq data will be performed to look at the genomic location of CTCF, motifs that bind CTCF, the overlap of CTCF binding between different cell types and the incidence of CTCF and cohesin binding in concert in the brain.

1.10.2 Epigenetic control of *Ddc_exon1a* expression in heart

Imprinting of *Ddc_exon1a* is controlled via an epigenetic difference between the two parental alleles, CpG rich regions at the promoter of *Grb10,* are known to control *Ddc_exon1a* expression however the mechanism by which this region controls imprinting has not yet been elucidated. Allele-specific and tissue-specific epigenetic features will be examined at the *Ddc/Grb10* locus to look for differences in methylation

and CTCF binding. Regions of putative differential methylation will be identified using data from MeDIP-seq experiments, and these will be validated using bisulphite PCR and sequencing. Putative CTCF binding regions will be assessed from ChIP-seq in both ES cells and brain, and these regions will be assessed for CTCF binding in heart and brain using ChIP followed by quantitative qPCR.

1.10.3 Role of Ddc in cardiac development

Ddc is expressed at relatively high levels in the embryonic and neonatal heart, but expression is reduced during postnatal life, suggesting it plays a specific role in cardiac development. In order to elucidate the role of Ddc in cardiac development knockout mice lacking Ddc expression in the heart will be used to assay the role of Ddc in controlling heart development during embryogenesis. Cells expressing Ddc in the developing heart will be identified using immunostaining and microscopy and gross morphological changes will be assayed using 3D imaging techniques. In addition expression of downstream gene targets will be examined using expression microarray.

Chapter 2

Methods

All solutions were made with purified and deionised water, prepared with the MilliQ water filtration system (Millipore Corporation, MA, USA), unless otherwise stated.

2.1 Nucleic Acid extraction from tissue

Organs were extracted and snap frozen in liquid nitrogen and stored at -80 °C until required.

2.1.1 DNA extraction

DNA extraction was performed using the DNeasy blood and tissue kit (Qiagen) unless otherwise stated. The protocol used was that for animal tissues using the spin-column. Once isolated DNA was stored at -20 °C.

2.1.2 RNA extraction

RNA was extracted from tissues using the RNeasy mini Kit (Qiagen) as per manufacturers instructions, using the protocol for animal tissues using the spin-column. Tissues were disrupted using a rotary homogenizer. Once isolated, RNA was stored at -80 °C.

2.1.3 RNA and DNA co-extraction

RNA and DNA were extracted concurrently from individual human heart samples using TRIzol® Reagent (Invitrogen). 0.5 mg of tissue was homogenized in 1ml of trizol using the rotary homogenizer and the homogenate transferred to 1.5 ml microcentrifuge tube. 200 µL Chloroform was added and the sample vortexed thoroughly before being centrifuged at 16 000 xg for 15 minutes at 4 °C. The aqueous phase, which contains RNA, was removed to a fresh 1.5ml microcentrifuge tube. 300 µL

of 100% ethanol (etOH) was added to remaining organic phase and the solution thoroughly vortexed and stored on ice for subsequent DNA extraction. 320 µL of ice cold isopropanol was added to the aqueous solution (RNA) and inverted gently to mix. The sample was incubated at room temperature for 10 minutes and centrifuged at 16 000 xg for for 15 minutes at 4 °C. The supernatant was removed leaving white RNA pellet, which was washed using 1 ml of 70% MCB grade ethanol and allowed to dry at room temperature for 10 minutes. To remove DNA contamination, the RNA pellet was resuspended in 180 µL nuclease free H_2O and 18 µL TURBO™ DNase Buffer (Ambion) with 2 µL of TURBO™ DNase solution (Ambion) and incubated at 37 °C for 30 minutes. The sample was placed in a 2 ml phase lock gel microcentrifuge tube (5 PRIME) and 200 µL phenol added, and inverted several times to mix. The solution was incubated at room temperature for 5 minutes and 200 µL Chloroform (Sigma) was added and the sample subjected to centrifugation at 16 000 xg for 5 minutes at room temperature. The aqueous phase was removed to a new 1.5 ml microcentrifuge tube. 0.8 volumes of isopropanol were added and inverted gently to mix and the sample incubated for 10 minutes prior to centrifugation at 16 000 xg for 15 minutes at 4 °C. The supernatant was removed carefully leaving an RNA pellet, which was washed using 1 ml of 70 % MCB grade ethanol and allowed to dry at room temperature for 10 minutes before being resuspended in an appropriate amount of TE pH8 buffer.

DNA was extracted from the organic phase by centrifugation at 15 000 xg for 5 minutes at room temperature, the supernatant removed and the pellet washed in 0.1M Sodium Citrate. The sample was centrifuged again and washed in 70 % ethanol, the DNA pellet dried and resuspended in TE pH8.

2.2 Chromatin Immunoprecipitation

2.2.1 Chromatin isolation from tissue samples

Organs were extracted and snap frozen in liquid nitrogen and stored at -80 °C until required. 300-500 mg of tissue was used to prepare chromatin with organs pooled if necessary. Tissues were cut up using a scalpel in 1 ml of cold PBS pH8 (Sigma Aldrich) containing 1x EDTA free complete protease inhibitor (Roche). The tissue and buffer were then transferred into a dounce homogenizer (Fisher) using a pipette and homogenized thoroughly. Samples were centrifuged in 1.6 ml sterile microcentrifuge tubes at 9000 xg for 5 minutes at 4 °C and the supernatant removed and discarded. Cells were washed three more times in 1 ml PBS pH8 plus protease inhibitor, with cells resuspended and centrifuged at 9000 xg for 3 minutes, and supernatant removed to waste after each wash. The cells were then subject to cross-linking using 1 ml of ice-cold 5mM dimethyl 3,3'-dithiobispropionimidate (Sigma Aldrich) in PBS pH8 plus protease inhibitor, at 4°C for 30 minutes. Samples were pelleted as before and resuspended in 1 ml ice-cold quenching buffer (1 mM Tris-HCl pH8, 150mM Sodium chloride (both Fisher Scientific) prepared using Diethylpyrocarbonate (DEPC) treated deionised water (DEPC from Sigma Aldrich)). Samples were washed 3 more times in 1 ml PBS pH8 plus protease inhibitor as previously described. A second cross-linking reaction was performed by re-suspending cells in 1 ml of 1% formalderhyde (Sigma Aldrich) made up in PBS pH8 plus protease inhibitor and incubating at 37°C for 10 minutes in a water bath, samples were washed 3 more times in PBS plus protease inhibitor. Cells were resuspended in 1 ml room temperature lysis buffer (5 µl of 0.1 M phenylmethanesulfonylfluoride (Sigma Aldrich) prepared fresh in ethanol (Fisher Scientific), 40 µl of 25x EDTA free complete protease inhibitor and 950 µl 0.5 mM Tris-HCl pH8, 1% w/v sodium dodecyl sulphate (SDS; Fisher Scientific), 10mM EDTA (Sigma Aldrich) prepared in DEPC-treated water. The chromatin was sonicated using a probe sonicator, samples were kept on ice at all times during the sonication procedure and subjected to 6 cycles of sonication for 1 minute at 40 Amp, with a 1 minute break in between cycles. Samples were approximately quantified using the

NanoDrop 1000 Spectrophotometer (Thermo Scientific) as per manufacturers instructions and stored at -80°C.

2.2.2 Agarose bead Chromatin Immunoprecipitation (ChIP)

All steps were performed using 1.5 ml sterile DNA LoBind Tubes (Eppendorf), unless otherwise described. Chromatin samples were pre-cleared to remove endogenous, non-specific IgG proteins, the amount of pre-cleared chromatin solution was calculated dependent on how many different immunoprecipitations (IP) were to be performed, 500 µl was required for Input and IgG control and 250 µl for each test IP. The pre-clear chromatin solution was prepared on ice using chromatin to a final DNA concentration of 80 ng/µl, 1x EDTA free complete protease inhibitor (Roche), 80 µl Protein A Agarose beads (Millipore) and the remaining volume made up with dilution buffer (70 mM Tris-HCl pH8, 165 mM Sodium chloride, 1mM EDTA, 1% v/v Triton X-100 (Sigma Aldrich) made with DEPC treated water). Samples were then rotated at 4°C for 2 hours at 20 rpm. In batches of 250 µl, the samples were filtered through Costar Spin-X centrifuge tubes with a 0.45 µm membrane pore size (Sigma Aldrich) at 4°C for 10 minutes at 7000 xg to remove the agarose beads, and samples were split into 200 µl in separate 1.5 ml microcentrifuge tubes for each IP treatment and the Input and IgG controls. 400 µl of dilution buffer and treatment antibodies (for details and dilutions see Appendix 2.1) were added to each tube except the Input and all samples rotated at 4°C overnight at 20 rpm.

20 µg tRNA was then added to the antibody treated samples as a carrier and mixed briefly by vortex, a further 60 µl Protein A agarose beads were added to these samples and all samples were then rotated at 4°C for 2 hours at 20 rpm. Antibody treated samples were then washed, by filtering samples through Costar Spin-X centrifuge tubes at 4°C for 1 minute at 5000 xg and the flow through discarded, 700 µl wash buffer 1 (20mM Tris HCl pH8, 150 mM sodium chloride, 0.1% w/v SDS, 1% v/v Triton X 100, 2 mM EDTA, 1x protease inhibitor, 40mM PSMF (made up as 0.1M in EtOH)) was

carefully added to the beads in the filter column, and the samples rotated at 4°C for 10 minutes at 20 rpm before centrifugation at 5000 xg for 1 minute at 4°C at which point the wash buffer is removed to waste. The same wash process was repeated using wash buffer 2 (20mM Tris HCl pH8, 500 mM sodium chloride, 0.1% w/v SDS, 1% v/v Triton X 100, 2 mM EDTA, 1x protease inhibitor, 40mM) and wash buffer 3 (10mM Tris HCl pH8, 250 mM lithium chloride (Sigma Aldrich), 1% v/v IpeGal ca-630 (Sigma Aldrich), 1% w/v sodium deoxycholate (Sigma Aldrich), 1 mM EDTA, 1x protease inhibitor, 40mM PSMF). Samples were then centrifuged at 4°C for 2 minutes at 5000 xg to ensure all wash buffer 3 was removed, 400 µl nuclease-free water (Ambion) was added to the filter column and the beads removed by pipetting to new 1.5 ml microcentrifuge tubes. 200 µl nuclease-free water was added to the input at this step. DNA was then decrosslinked from protein by incubation in 200mM sodium chloride at 65 °C overnight and the treated with 1ul 100 µg/µl RNAase A (Qiagen) at room temperature for 15minutes, then with 2 µl Proteinase-K (Roche) at room temperature for 15 minutes. DNA was purified using a using phenol:chloroform extraction. The samples were placed in a 2 ml phase lock gel microcentrifuge tube (5 PRIME) and 400 µl UltraPure phenol:chloroform:isoamyl alcohol (25:24:1, Invitrogen) solution was added and the tube inverted several times. Samples were spun at 17000 xg for 10 minutes and the aqueous phase removed to a sterile microcentrifuge tube. 40 µl 3M sodium acetate, 1200 µl molecular biology grade ethanol and 2 µl of 2 µg/µl UltraPure glycogen (Invitrogen) were added. Samples incubated on ice for 30 minutes, then centrifuged at 17000 xg for 25 minutes to pellet the DNA, supernatant was removed to waste and the pellet washed in 70% ethanol. A final centrifugation was performed and the ethanol wash removed, the pellets were allowed to dry at room temperature for 20 minutes then resuspended in 36 µl TE pH8 (Sigma).

2.2.3 DNA Analysis – Qubit Quantification™

Concentration of DNA was measured using the Qubit™ dsDNA HS and BR assay (Invitrogen). The HS assay was used for DNA samples where the concentration was estimated to be in the range of 10 pg/µL to 100 ng/µL, and the BR assay for concentration in the range 100 ng /µL to 1000 ng/µL. The Qubit® 2.0 Fluorometer was calibrated prior to each use, with the appropriate standards supplied. All reactions were prepared and analyzed at room temperature as per the manufacturers instructions. Two technical replicates of 1 µL were measured for each DNA sample and the average reading was taken to be the final concentration.

2.2.4 DNA Analysis - Agilent 2100 Bioanalyzer™

DNA fragment length was assayed using the Agilent High Sensitivity DNA Kit (Agilent). The electrodes of the bioanalyzer were cleaned using nuclease-free water before each use. Assays were run according to the manufacturers instructions using 1 µL of DNA sample.

2.3 ChIP-seq

2.3.1 Library preparation for Illumina™ GAIIx sequencing

All steps were performed using 1.5 ml sterile DNA LoBind Tubes (Eppendorf). For each CTCF or Rad21 library preparation, Three identical chromatin immunoprecipitations were pooled to provide enough DNA to proceed with successful library preparation. 4 µl was removed for validation of the ChIP using Taqman™ quantitative PCR for regions "Upstream of *H19* DMR" (DMDup) and "H19 *DMR*" (H19 DMR) (Applied Biosystems, see quantitative PCR method, primers and probe sequences listed in Appendix 2.2), 2 µl was removed for Qubit fluorometer quantification (Invitrogen) using the HS assay and 1 µl for analysis using the HS-DNA bioanalyzer (Agilent). Pooled samples were judged adequate if the quantitative PCR at known regions of enrichment and non-enrichment appeared normal and the DNA concentration as measured by Qubit quantification exceeded 0.8 ng/µl. Due to the large fragment size of DNA,

identified by the HS-DNA bioanalyzer, DNA was sheered using the SS220 sonicator (Covaris) as per manufacturers instructions with the following modifications of settings to achieve a final fragment length of 500 base pairs. 5% Duty Cycle, 3 Intensity, 200 cycles per burst, 65 seconds and the sonication was repeated 2 times for each sample. Samples were then removed to a sterile DNA LoBind tube, and ethanol precipitated by adding 10 µl sodium acetate, 300 µl molecular biology grade ethanol and 1 µl glycogen and incubating on ice for 30 minutes. Samples were centrifuged at 17000 xg for 25 minutes to pellet the DNA, supernatant was removed to waste and the pellet washed in 70% ethanol. A final centrifugation was performed and the ethanol wash removed, the pellets were allowed to dry at room temperature for 20 minutes then resuspended in 43 µl nuclease-free water. 2 µl were subsequently used to quantify DNA concentration using the high sensitivity (HS) Qubit fluorometer quantification system and 1 µl was used to check fragments length using the HS-DNA Bioanalyser as previously described, 40 ng DNA was diluted in 40 µl nuclease-free water to take forward into the library preparation.

ChIP-seq libraries were made using the NEBNext ChIP-Seq Sample Prep Master Mix Set 1 (New England Biolabs) using the manufacturers protocol with the following clarifications. 40ng ChIP-seq DNA was put into the initial reaction not 10 ng as is suggested in the current protocol. For the adapter ligation step Illumina™ adapters were used (Illumina) in a dilution of 1:10 in nuclease-free water. Size selection of the adaptor ligated DNA was performed using gel extraction using 2% UltraPure Low Melting Point Agarose (LMP) agarose gel (Invitrogen), and separation was achieved by running the gel at 4 °C for 3 hours at 60mV, a size of 300-600 bp was extracted. DNA was extracted from the Gel using the Qiagen gel extraction kit (Qiagen) PCR enrichment of adapter ligated DNA was performed using the same mastermix as described in the Illumina protocol using a modified PCR cycle as follows:

Cycle Step	Temp	Time	Cycles
Initial denaturation	98 °C	30 seconds	1
Denaturation	98 °C	10 seconds	
Annealing	65 °C	30 seconds	13
Extension	72 °C	30 seconds	
Final Extension	72 °C	5 minutes	1
	4 °C	Hold	

PCR reactions were cleaned up using the MinElute PCR purification kit (Qiagen) as per Qiagen instructions and samples were eluted in the final step using 50 µl EB buffer.

2.3.2 Library quantification

Quantification of libraries was performed using Taqman™ quantitative PCR (Applied Biosystems, primers and probe sequences listed in Appendix 2.2), using relative quantification against a library previously known to give adequate number of reads when run on the GAIIx platform (previous library courtesy of Dr Jennifer Frost at King's College London). Serial dilutions of the library preparation were made 1:10, 1:1000, 1:10000, 1:100000 and compared to the previous library concentration. (The method for quantitative PCR is described in section 2.7.3.)

2.3.3 Illumina™ GAIIx sequencing

Next generation sequencing libraries were analyzed using the Illumina™ GAIIx sequencing platform, with each sample assigned to a separate lane of an eight lane flow cell. Sequencing was performed at the Biomedical Research Centre (BRC) genomics facility, Guy's Hospital, London by Dr. Efterpe Papouli and Ms Muddassar Mirza.

2.4 Bioinformatic analysis of ChIP-seq data

2.4.1 Alignment and binding peak calling

Sequence reads were aligned to the mouse reference genome (mm9) using Novoalign (v. 2.01.13, http://www.novocraft.com/). USeq (Nix et al. 2008) was used to identify mean peak shift separately for CTCF and Cohesin reads using only the first of each pair end matched reads. Peaks were called using peak shifts and window sizes of 138bp and 144bp for CTCF and Cohesin respectively and an FDR threshold of 95%. A subset of the peaks was obtained to an FDR of 5%, expanded by 500bp upstream and downstream and overlapping peaks merged prior to further analysis. Track comparisons were made to ChIP-seq data from Chen et al, 2008, (Chen, et al., 2008) and the peaks are in good agreement between studies providing confidence in the accuracy of the CTCF binding peaks identified in this study, with the subtle differences likely reflecting tissue specificity.

2.4.2 Assessment of allele-specific binding

Parent-of-origin specific binding was assessed by means of binomial testing, using a custom bioinformatic pipeline. Initially the subset of reads mapping to previously identified peak regions with at least one SNP between the parental mouse strains was obtained using BEDTools (Quinlan, et al., 2010). Only the first read of each paired end was used; the analysis was replicated for the second pair of reads for cross-validation and visual inspection using the UCSC genome browser revealed identical mappings for all peaks that exceeded the genome-wide threshold in both analyses.

Individual reads were assigned to one of the parental alleles using a custom Perl script, utilising the samtools Perl library. Each read was mapped as either derived from the reference sequence (C57BL/6) or from the *Mus mus castaneus* allele on the basis of a SNP between the parental strains. If more than one SNP was present the SNP with the best quality read was used. The assignments were subsequently converted to as

maternally- or paternally-derived and data from the two reciprocal crosses were merged for the CTCF and Cohesin datasets independently.

Counts of maternal and paternal reads were obtained on a per-region basis using MySQL. Reads were only considered if the phred-scaled alignment mapping quality exceeded 50 and the base call quality at the SNP used for mapping of that read exceeded 20. Furthermore, regions were only considered if at least two reads covered the region.

Parent-of-origin specific binding was assessed using a two-sided binomial test (implemented in R) of the maternal versus paternal allelic read counts. Regions were sorted by p-value score using MySQL. Genome-wide significance of p-values was assessed by means of Bonferroni correction. UCSC Bed tracks were prepared at different cutoffs with maternal/paternal annotation.

2.4.3 Binding overlap between different tissues

Peak overlap between binding of CTCF in different tissues were obtained using the bedtools intersectBed command.

Bioinformatic analysis was performed in collaboration with Dr Reiner Schulz (King's College London) and Nikolas Barkas (King's College London). Dr Schulz performed initial analysis of CTCF and Rad21 binding using Useq, and Nikolas Barkas refined this analysis and also developed the custom pipelines used to assess allelic bias and tissue-overlap, including writing all the custom scripts. Analysis was performed using the BRC funded high performance computing cluster at King's College London.

2.5 Mouse stocks and breeding

2.5.1 Housing conditions

C57BL/6 and *Mus mus castaneus* animals were housed at the biological services unit at King's College in accordance with United Kingdom Home

Office breeding licence; colonies were managed with help from Siobhan Hughes and Dr Michael Cowley (King's College London). Conditions were managed with a 12 hour dark/light cycle with 30 minute dawn/dusk lighting, ambient temperature of 21 °C ± 2 °C and humidity of 50 % ± 10 %. Mice were all fed on expanded breeder pellet R&M No3 (Special Diets Services, Essex, UK).

2.5.2 Transgenic mouse lines

Grb10Δ2-4 and *Grb10Δ7* knockouts animal tissues were obtained from the laboratory of Dr Andrew Ward (University of Bath, UK) and dissections were performed with help from Dr Marta Madon (University of Bath). *Ddc* knockout animals were bought commercially from Jackson Laboratories (B6;129S5-*Ddc*$^{Gt(neo)420Lex}$), and the breeding programme and tissue collections were managed by the UC Davis Mouse Biology Program at the University of California, Davis, CA, USA. The breeding programme was designed as part of this thesis. Genotyping of the embedded tissues was performed at UC Davis using the following protocol. DNA was extracted from yolk sac and the following PCR mastermix was used 5 µl of 10x Sigma buffer, 3.5 µl of 25mM $MgCl_2$, 2 µl 10mM dNTPs, 1.5 µl of 20 µM forward primer, 1.5 µM reverse primer, 5 µl Taq Polymerase (Sigma), 31 µl of nuclease free H_2O, 5 µl of yolk sac DNA. Primer sequences are found in Appendix 2.3. PCR conditions are as follows:

Cycle Step	Temp	Time	Cycles
Denaturation	96 °C	27 seconds	
Annealing	61 °C	25 seconds	36
Extension	72 °C	25 seconds	

Genotyping of the frozen samples for microarray was performed at King's on DNA extracted from the carcass of dissected embryos using the DNAeasy Kit (Qiagen) (protocol as described in section 2.1.1). PCR was

performed using the following conditions: 18ul of 1.1x ReddyMix PCR Master Mix (Thermo), 0.5 µl of 20 µM forward primer, 20 µM reverse primer, 1 µl of tail DNA. PCR primers are detailed in Appendix 2.3. Cycling conditions are as follows:

Cycle Step	Temp	Time	Cycles
Initial denaturation	98 °C	30 seconds	1
Denaturation	98 °C	10 seconds	
Annealing	59 °C	30 seconds	35
Extension	72 °C	30 seconds	
Final Extension	72 °C	5 minutes	1

2.6 Bisulphite analysis of DNA

Genomic DNA was extracted from organs that were dissected and snap frozen in liquid nitrogen and stored at -80°C using the DNeasy Blood and Tissue kit (Qiagen). DNA was quantified using the broad range Qubit fluorometer quantification system (Invitrogen).

Two methods for bisulphite treatment were used, these were the manual and kit based protocols.

2.6.1 Manual method

DNA was diluted to a final concentration of 2.5 µg in 30 µl of AE buffer (Qiagen) in a thin walled 0.6 ml microcentrifuge tube, and overlayed with mineral oil (Invitrogen) to prevent evaporation. 3.66 µl of freshly prepared 3M sodium hydroxide (Sigma) was added to each sample and incubated in an oven at 65°C for 20 minutes. Separately, in the absence of light, 500 µl of 200mM hydroquinone (Invitrogen) was added to 8 ml of 3.75M sodium metabisulphite (Invitrogen) the solution equilibrated to pH 5 using 10M Sodium Hydroxide, and left to stand for 5 minutes. All subsequent steps were performed in the absence of light. 400 µl of the

hydroquinone/sodium metabisulphite solution was added to each sample and mixed by inversion. Samples were then incubates at 97°C for 10 minutes then 55°C for 4 hours. Samples were purified using the Qiaquick PCR column (Qiagen) using the manufacturers protocol, and eluted in 200 µl EB buffer (Qiagen). DNA was fully denatured by the addition of 22 µl of 3M sodium hydroxide. Samples were then ethanol precipitated by adding 20 µl sodium acetate, 600 µl molecular biology grade ethanol (Invitrogen) and 1 µl glycogen (Roche) and incubating on ice for 30 minutes. Samples were centrifuged at 17000 xg for 25 minutes to pellet the DNA, supernatant was removed to waste and the pellet washed in 70% ethanol. A final centrifugation was performed and the ethanol wash removed, the pellets were allowed to dry at room temperature for 20 minutes then resuspended in 30 µl TE ph8. The following PCR conditions were used to amplify regions of interest. 5 µl bisulphite converted DNA, 2.5 µl 10x sequencing buffer (Qiagen) , 13.5 µl nuclease free H_2O, 1.5 µl 25 mM $MgCl_2$ (Qiagen), 1 µl 10mM dNTP mix (Invitrogen), 0.5 µl of 20 µM forward primer, 0.5 µl 20 µM reverse primer, 0.5 µl HotStart Taq DNA polymerase (Qiagen). Primers are detailed in Appendix 2.4. Cycling conditions are as follows:

Cycle Step	**Temp**	**Time**	**Cycles**
Initial denaturation	98 °C	30 seconds	1
Denaturation	98 °C	10 seconds	
Annealing	$<T_m$ °C	30 seconds	45
Extension	72 °C	30 seconds	
Final Extension	72 °C	5 minutes	1

2.6.2 Kit based method

Samples were diluted in nuclease-free water to a final dilution of 500 ng in 45 µl. DNA was subjected to bisulphite conversion using the EZ DNA Methylation Kit (Zymo Research) as per the manufacturers instructions

with the following clarifications. Sample incubations during the bisulphite conversion step used 16 cycles of 95°C for 30 seconds followed by 50°C for 60 minutes. The sample PCR conditions were used as for the manual method of bisulphite conversion (section 2.6.1) with a modified number of cycles:

Cycle Step	Temp	Time	Cycles
Initial denaturation	98 °C	30 seconds	1
Denaturation	98 °C	10 seconds	
Annealing	$<T_m$ °C	30 seconds	35
Extension	72 °C	30 seconds	
Final Extension	72 °C	5 minutes	1

2.6.3 Cloning using P-Gem T easy vector system

PCR products were subjected to electrophoresis on a 2% L.M.P. agarose (Invitrogen) gel for 2hours at 70V in TAE running buffer (40 mM Tris, 20 mM acetic acid (Sigma), and 1 mM EDTA). PCR products were extracted using the QIAquick® gel extraction spin protocol (Qiagen) following the manufacturers instructions. DNA was quantified using the nanodrop and 25 ng of DNA was ligated into the P-Gem T easy vector (Promega) as per the manufacturers instructions, with the sample incubated overnight at 4 °C to maximise ligation efficiency. 2 µl of the ligated vector was transformed into 50 µl competent DH5α cells (made by Siobhan Hughes, Dr Michael Cowley and Dr Sabrina Bohm, King's College London) and the cells grown in 150 µl of Super Optimal Broth (SOC) medium (Invitrogen) in a shaking incubator at 37 °C for 1 hour. These cultures were subsequently plated on Lysogeny broth (LB) agar plates which contain 0.5mM IPTG (Sigma), 80µg/ml X-Gal (Invitrogen) and 100µg/ml ampicillin (Sigma), and were grown overnight at 37 °C.

2.6.4 Colony PCR and sequencing

Unique white colonies were picked from the plate into 50 µl of LB (Sigma) containing 100µg/ml ampicillin (Sigma) in a 96-well PCR plate, using a sterile pipette tip, the samples were covered using a PCR adhesive plate sealer (Thermo) and incubated overnight at room temperature on a rotating platform. PCR was performed directly from the LB cultures using 18 µl of 1.1x ReddyMix PCR Mater Mix (Thermo), 0.5 µl of 20 µM T7 primer, 20 µM SP6 primer and 1 µl of LB culture. The cycling conditions are as follows:

Cycle Step	Temp	Time	Cycles
Initial lysis and denaturation	98 °C	5 minutes	1
Denaturation	98 °C	10 seconds	
Annealing	56 °C	30 seconds	35
Extension	72 °C	30 seconds	
Final Extension	72 °C	5 minutes	1

Samples were then sequenced using the sequencing protocol in section 2.6.5.

2.6.5 Sanger sequencing of PCR products

PCR products were purified using the ExoSAP-IT PCR purification kit (GE Healthcare). 11.5 µl nuclease free water and 1 µl ExoSAP-IT was added to 2.5 µl of PCR product, and the solution treated at 37 °C for 15 minutes followed by 80 °C for 15 minutes to deactivate the enzyme. The sequencing reaction was prepared using 2.5 µl of the ExoSAP-IT purified PCR product, 2 µl 5x sequencing buffer, 4 µl 20µM forward or reverse primer, 4.6 µl nuclease free water and 0.5 µl Big Dye terminator. The sequencing reaction was run on the thermocycler using the following conditions:

Cycle Step	Temp	Time	Cycles

Initial denaturation	98 °C	1 minute	1
Denaturation	98 °C	30 seconds	
Annealing	<T_m °C	15 seconds	29
Extension	62 °C	1 minute	

DNA was purified by adding 30 µl 100% etOH and 1 µl 3M Sodium Acetate, and left to precipitate fro 20 minutes at 4 °C before being centrifuged at 3060 xg at 4 °C for 20 minutes. EtOH and Sodium Acetate was removed and the pellet washed with 70 % etOH before being left to dry. DNA was resuspended in 10 µl Hi-Di™ Formamide (Applied Biosystems) by vortexing and heating to 94 °C for 2 minutes. Samples were sequenced on the Applied Biosystems 3730xl according to the manufacturers instructions.

2.7 Gene expression analysis

2.7.1 cDNA synthesis

RNA was converted to cDNA using the Superscript II® first strand synthesis system (Invitrogen) following the manufacturers protocol with the following adaption: cDNA synthesis was primed using the mastermix detailed as follows:

Reagent	**Amount**
Total RNA	2 µg
Random hexamers (50ng/ µl)	1 µl
10mM dNTP mix	1 µl
Nuclease free water	To 10 µl

2.7.2 RT-PCR

PCR on cDNA was performed using the following protocol. 18 µl of 1.1x ReddyMix PCR Mater Mix (Thermo), 0.5 µl of 20 µM F primer, 20 µM R primer and 1 µl cDNA. Primer sequences for all RT-PCR reactions can be found in Appendices 2.5-2.7, and cycling conditions are as follows:

Cycle Step	Temp	Time	Cycles
Initial denaturation	96 °C	2 minutes	1
Denaturation	96 °C	30 seconds	29, unless otherwise specified
Annealing	<T_m °C	30 seconds	
Extension	67 °C	1 minute	
Final Extension	67 °C	5 minutes	1

2.7.3 Quantitative PCR (qPCR)

Absolute quantification qPCR was used to analyse enrichment of chromatin immunoprecipitated DNA for the analysis of CTCF binding in tissues. Primers were designed to regions of interest using the ABI Taqman™ design tool. Primer and FAM probes are detailed in Appendix 2.8. Each qPCR was performed in technical triplicate using the following reaction conditions: 0.5 µl Primers and probe, 5 µl TaqMan Universal PCR Master Mix and 1.5 µl water. For *Ddc* expression analysis qPCR was used to measure the quantity of *Ddc* transcript relative to *Actin* control using the same reaction conditions as for relative quantification, assay details are in Appendix 2.9. The qPCR was analysed on the 7900HT thermocycler using standard conditions.

2.8 Histological analysis using optical microscopy

2.8.1 *Grb10* and Ddc Co-expression

Tissues were collected and fixed for 30 minutes at room temperature in 4 % (w/v) PFA, 0.2 % (v/v) glutaraldehyde in 0.1 % PBS and were then placed in 30% sucrose solution overnight at 4 °C before being embedded in Optimal Cutting Temperature compound (OCT) (VWR international) and

stored at -80 °C until required. Samples were sectioned using a cryostat to a thickness of 10 µm and serial sections were treated to stain for *Grb10* (using the *lacZ* reporter) or Ddc (using immunohistochemistry).

2.8.2 Paternal *Grb10* expression

Samples stained for paternal *Grb10* were incubated in 1 mg/ml X-gal (Sigma Aldrich) diluted in stain base (30 mM $K_4Fe(CN)_6$ (Sigma Aldrich), 30 mM $K_3Fe(CN)_6.3H_2O$ (Sigma Aldrich), 2 mM $MgCl_2$, 0.01 % (w/v) sodium deoxycholate, 0.02 % (v/v) Igepal CA-630 in 0.1 % PBS) for 18 hours at 28°C by pipetting onto sections in a dark, humidified chamber. To permit counter-staining, slides were washed gently in 0.1 % PBS to remove excess X-gal, and cleared in two 2 minute incubations with Histoclear (National Diagnostics). Sections were hydrated in a descending ethanol series, incubating for 2 minutes each in 100 %, 100 %, 95 %, 90 %, 70 %, 50 % ethanol and MilliQ water. Sections were stained in nuclear fast red for 90 seconds and dehydrated in 70 % ethanol (30 seconds), 95 % ethanol (30 seconds) and two 100 % ethanol incubations (1 minute each) followed by Histoclear for 2 minutes prior to mounting with DePex (VWE international).

2.8.3 Ddc Expression

Samples to be stained for Ddc were air dried for 10 minutes and washed twice in PBS with 0.01% v/v Tween-20 (Sigma) for 2 minutes and were fixed with 4% PFA for 30 min. Slide were washed twice in PBS with tween (2 minutes) and placed in H_2O_2 Quenching Buffer (0.3% H_2O_2 in 0.1% PBS) for 30 minutes to quench endogenous peroxidases. Slide were washed twice in PBS with tween (2 minutes) and incubated in 1.5 % blocking serum supplied with the VECTASTAIN® ABC kit (Vector Labs) and the protocol supplied with the VECTASTAIN kit was followed: avidin solution (30 minutes), PBS with tween (3 x 2 minutes), 5 % skim milk in PBS (20 minutes), PBS with tween (3 x 2 minutes), 1:500 Ddc antibody in 1.5 % blocking serum (1 hour), PBS with tween (3 x 2 minutes), biotinylated secondary antibody (30 minutes), PBS with tween (3 x 2 minutes),

VECTASTAIN ABC Reagent (30 minutes), PBS with tween (3 x 2 minutes) 0.05% Diaminobenzidine (DAB) in PBS with tween (10 minutes), PBS with tween (3 x 2 minutes). Slides were washed in water before being counterstained with Harris' haematoxylin (30 seconds), and rinsed in water and 70 % ethanol briefly before incubation in Scott's tap water (Fisher) for 1 minute. Slides were dehydrated in 70 % ethanol (30 seconds), 95 % ethanol (30 seconds) and two 100 % ethanol incubations (1 minute each) followed by Histoclear (2 minutes) prior to mounting with DePex.

2.8.4 Bisected embryo staining for *Grb10*

Embryos were frozen quickly on dry-ice and sagittally bisected with a razor blade, before being stained in X-gal solution, diluted in stain base as described previously, with incubation at 28°C overnight.

2.9 Histological analysis using fluorescence microscopy

2.9.1 Immunostaining for Ddc and ANF

Pregnant female mice were sacrificed and uterine horns containing embryos were dissected. Embryos were left in cold PBS for 10 minutes before being fixed in 4% PFA for 1 hour at 4 °C and embedded in paraffin wax. This was performed at UC Davis, California, USA. Samples were secioned at a thickness of 8 µm and hydrated using the following wash series for 3 minutes each, three times in Xylene (Fisher), 100 %, 100 %, 95 %, 90 %, 70 %, 50 %, three times in PBS. Slides were placed in boiling High pH antigen unmasking solution (Vector Labs) and left for 30 minutes as the solution cooled and wash three times in PBS (5 minutes). Slides were washed in 0.05% Tween-20 in PBS and blocked with 4 % v/v donkey serum (Abcam) in PBS with Tween-20 for 1.5 hours. Primary antibodies were prepared in the following dilutions rabbit-α-mouse Ddc (1:500), goat-α-mouse ANF (1:100) in 0.01% Tween-20, 2 % v/v donkey serum in PBS. Antibody was dropped onto slides and incubated in a humidified chamber at 4 °C for 16 hours. Slides were washed three times in 0.01% Tween-20 in PBS. Secondary antibodies Alexa Fluor 555 donkey-α-goat (Invitrogen) and

Alexa Fluor 647 donkey-α-rabbit (Invitrogen) were diluted 1:300 in 0.01% Tween-20 PBS, dropped onto the slides and incubated for 2 hours at room temperature in the dark. Slides were wased three times in PBS and mounted using ProLong® Gold Antifade Reagent with DAPI (Invitrogen). Details of all antibodies can be found in Appendix 2.10.

2.9.2 Immunostaining for Ddc and MF-20

The same staining protocol as previously descibed was used with the following substitutions; solutions remain the same concentrations unless otherwise specified. Donkey serum was replaced with Goat serum (Abcam), primary antibody for MF-20 was used instead of ANF, this was monoclonal mouse-α-mouse MF-20 at a dilution of 1:50 (from Prof. Scott Baldwin). Secondary antibodies used were Alexa fluor 555 goat-α-rabbit and Alexa fluor 488 goat-α-mouse (Invitrogen) diluted 1:400. Details of all antibodies can be found in Appendix 2.10.

2.10 Western blot analysis

Whole carcass of e15.5 embryo was frozen in liquid nitrogen and smashed using a stainless steel hammer, which had been pre-cooled in liquid nitrogen. 1 mL of RIPA buffer (50 mM Tris-HCl (pH 7.5), 150 mM NaCl, 1 mM EDTA, 1% (w/v) sodium deoxycholate, 0.1% SDS, 1mM PMSF, 1x protease inhibitor (Roche)) was pipette onto the fragmented sample transferred to a 2 mL microcentrifuge tube. Samples were further homogenized using a rotary homogenizer and centrifuged at 16 000 xg for 20 minutes at 4 °C, the supernatant was removed to a fresh 1.6 mL microcentrifuge tube. Total protein concentration was determined using the BCA protein assay kit (Pierce, IL, USA), and samples stored at -20°C. The same protocol was used to extract protein from cultured cells, without the hammer step. 20 µg in 25 µL of each sample was mixed 1:1 with 2x reducing sample buffer (62.5 mM Tris HCl pH 6.8, 2% (w/v) SDS, 6 M Urea, 2 % (v/v) Igepal CA-630, 5 % (v/v) β-mercaptoethanol (SigmaAldrich), 0.02% (w/v) bromophenol blue (Sigma Aldrich), 4 % glycerol (Sigma Aldrich) and

heated to 95°C for 5 minutes. Samples were loaded, with a Spectra™ Multicolor broad range protein ladder (Thermo) onto a 12 % polyacrylamide resolving gel: 12 % polyacrylamide (National Diagnostics), 0.37 M Tris HCl pH 8.8, 0.1 % SDS, 0.05 % AMPS (Sigma Aldrich), 0.05% TEMED (Sigma Aldrich) with a stacking gel (5 % polyacrylamide, 0.12 M Tris HCl pH 6.8, 0.05 % AMPS, 0.1 % TEMED). Samples were electrophoresed at 100 V for 3 hours in running buffer (0.1 % SDS, 25 mM Tris, 208 mM glycine (Sigma Aldrich)). Protein was transferred at 90V for 2 hours to a PVDF membrane (Bio-Rad Laboratories) using Western blot wet transfer buffer (25 mM Tris, 192 mM glycine, 20 % (v/v) methanol (Fisher Scientific). All subsequent incubations took place on an orbital shaker. The membrane was blocked for 90 minutes in 5 % powdered skimmed milk (Marvel, UK) in 0.1 % Tween-20 with PBS. Primary antibodies were diluted rabbit-α-mouse Ddc (1:1000), in 5% milk in 0.1 % Tween-20 with PBS and applied to the membrane and incubated overnight 4 °C. Membranes were washed three times in 0.1 % Tween-20 with PBS for 15 minutes per wash, and incubated in peroxidase-conjugated goat anti-rabbit secondary antibody (DAKO) diluted 1/2000 in 5 % milk in 0.1 % Tween-20 with PBS for 1 hour at room temperature. Three further washes were performed as before and protein detection was performed using ECL system (Amersham Biosciences). Proteins were visualized by exposure on Fuji film and were developed on a Laser 45 developer. For loading control, membranes were stripped by heating at 50 °C for 30 minutes in stripping buffer (100mM 2-Mecaptoethanol, 2 % SDS, 62.5mM Tris-HCl pH6.7), washed three times in 0.1 % Tween-20 with PBS for 15 minutes per wash, and re-probed using mouse-α-mouse Tubulin (1:5000) and rabbit-α-mouse Histone H3 (1:5000), detection was performed as for Ddc. Details of antibodies can be found in Appendix 2.11.

2.11 Analysis of *Ddc* knockout heart

2.11.1 Expression microarray library preparation

Heart samples were dissected from e15.5 embryo and snap frozen in liquid nitrogen. RNA was extracted using the RNeasy Kit as describer previously. Microarray libraries were made using the Illumina TotalPrep RNA amplification kit (Ambion) as per the manufacturers instructions.

2.11.2 Expression microarray analysis

Raw probe signal and raw control signal were pre-processed using NEQC, and the linear modelling was used to make comparisons between genotypes using LIMMA. All bioinformatics were performed with the help of Heba Saadeh (King's College London).

2.11.3 Fixing and embedding for EFIC, EFIC and Analysis

Embryos and dissected neonatal hearts were fixed and embedded in paraffin using the same method as for immunostaining. In order to conduct EFIC analysis, the samples were re-embedded in red aniline dyed wax. Samples were heated to 60 °C to removed the existing wax and washed using xylene 2 times for 30 minutes each on a orbital shaker before being placed in red aniline dyed wax (73% Vybar (Fisher), 24% Paraffin wax (Fisher) and 3% stearic acid containing 0.1% red aniline wax dye) and left for 14 hours at 65 °C to infiltrate the sample. The red aniline wax was changed and the samples left for a further 6 hours at 65 °C before being embedded in fresh red aniline wax. Samples were sectioned and photographed using the EFIC system at Vanderbilt University, with a sectioning size of 5 μm. Samples were photographed at a magnification of 20x using a mCherry and GFP wavelength and the images stored in Volocity 3D image analysis software (Perkin Elmer). Images were checked for quality by visual inspection and rebuilt in 3D using Velocity. Samples were virtually re-sectioned in a plane that bisected the mitral and aortic valve, and the measurements were taken on this plane at the base of the papillary muscle to ensure samples were measured equally. All measurements were made blind and the identity of the samples only revealed just prior to analysis.

Comparisons between samples groups were made using an unpaired Student's T-test.

2.12 Acknowledgement of equipment funding

Sequencing and microarray analysis required the use of BRC Core Facilities provided by the financial support from the Department of Health via the National Institute for Health Research (NIHR) comprehensive Biomedical Research Centre award to Guy's & St Thomas' NHS Foundation Trust in partnership with King's College London and King's College Hospital NHS Foundation Trust. Episcopic fluorescent image capture, performed at Vanderbilt University, TN, USA, was funded by the National Institute of Health (S10-RR27661)

Chapter 3

CTCF and Cohesin binding in mouse brain

3.1 Introduction

3.1.1 CTCF and Cohesin binding in the context of genomic imprinting

As discussed in chapter one, the zinc finger DNA binding protein CTCF has been shown to control imprinting at the *Igf2/H19* imprinted locus in both mice and humans. Furthermore the Cohesin complex has been shown to bind with CTCF at the *H19* gDMR as well as several other imprinted DMRs in mouse tissues and has been implicated in control of gene expression in mouse (Hadjur, et al., 2009) and transcriptional termination in other eukaryotic cells (Gullerova, et al., 2008). It has also been shown that Cohesin is required for cell-type-specific long-range chromosomal interactions *in cis* during cellular differentiation (Hadjur, et al., 2009). Many published manuscripts allude to the importance of CTCF as a mechanism for governing genomic imprinting. The capacity of CTCF to bind DNA varies depending on whether DNA is methylated or unmethylated, and as such it is not an unreasonable to postulate that CTCF may act as a mechanism to 'interpret' the differential methylation marks seen at imprinting control regions and then illicit differential control of gene expression on the two parental alleles. Despite this assumption being widespread, as yet, no systematic screen of allele-specific binding of CTCF has been made in order to investigate the occurrence of CTCF and Cohesin binding at multiple imprinted regions associated DMRs. In order to explore the role of CTCF and Cohesin at all known imprinted loci, allele-specific chromatin immunoprecipitation followed by next generation sequencing (allele-specific ChIP-seq) was performed on three-week-old mouse brain. This thesis sets out to explore the epigenetic control of imprinting of *Ddc_exon1a* in heart, however brain tissue was chosen as a model tissue in

which to investigate CTCF and Cohesin binding at imprinted loci for the following reasons:

1) Imprinting plays a particularly important role is the brain (Davies, et al., 2007). The human imprinting disorders Prader-Willi and Angelman syndromes present with severe behavioural and neurodevelopmental phenotypes (Cassidy, et al., 2000, Lossie, et al., 2001, Williams, et al., 2006). In addition, of the hundred or so imprinted transcripts known in mouse, more than fifty are expressed in the brain (Wilkins, 2008) and thirdly, there are mouse imprinted genes that have been shown to exhibit behavioural phenotypes when disrupted, including *Peg3* (Li, et al., 1999), *Mest* (Lefebvre, et al., 1998), *Nesp55* (Plagge, et al., 2005) and *Grb10* (Garfield, et al., 2011).
2) Extraction of chromatin from whole tissue is a novel experimental procedure and had not been undertaken before these experiments began. When optimising chromatin extraction, the quality and yield of chromatin was far higher in brain than in heart, and this allowed the ChIP-seq experiment to be performed optimally.
3) An additional unforeseen advantage is that a recent genome-wide, allele-specific DNA methylation profile for brain tissue has been compiled (Xie, et al., 2012) allowing for comparison of DNA methylation and CTCF and Cohesin binding to be made.

3.1.2 CTCF binding and tissue differentiation

In order to investigate the role that CTCF plays in cell differentiation and cell lineage specification, genome-wide binding of CTCF in brain was examined in comparison to existing CTCF binding profiles from ES cells (Chen, et al., 2008) and liver (Schmidt, et al., 2010). CTCF binding has reported to be largely invariant in different cell types in experiments using human cell lines (Kim, et al., 2007), however this has not yet been investigated in fully differentiated mouse tissues.

3.2 Results

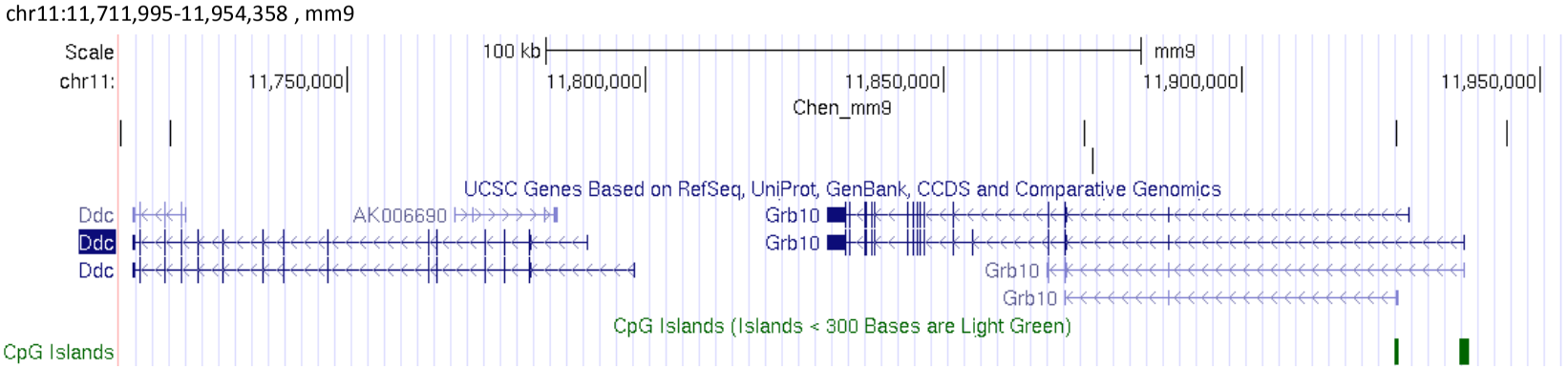
3.2.1 Putative epigenetic mechanisms controlling imprinting of *Ddc*

In order to justify further investigation of CTCF mechanisms in imprinting and in particular a potential role for CTCF binding in controlling expression of *Ddc_exon1a*, an initial in silico screen of CTCF binding at the *Ddc/Grb10* locus, was performed. Binding of CTCF was examined using a ChIP-seq dataset from ES cells (Chen, et al., 2008) which revealed CTCF binding at the *Grb10* DMR (Figure 3.1, B), further CTCF binding sites were observed 5' to the *Grb10* gene body, and two CTCF binding sites were observed 1.3 kbps apart in a intron between exons 2-3 of *Grb10*. Two additional CTCF binding sites are located at the 3' end of the *Ddc* gene body, one downstream of the 3' UTR and one between exons 12 and 13 (Figure 3.1 , A). This ChIP-seq analysis did not distinguish between binding on the maternal or paternally inherited allele, however as binding was observed at the *Grb10* DMR, and other binding sites are seen across the locus, it is pertinent to explore allele-specific binding as a mechanism for controlling imprinting further, as it may relate to control of gene expression at the *Ddc/Grb10* locus. Additional observations of CTCF binding showed that, when analyzed across the entire chromosome 11, multiple regions appear to be rich in CTCF binding separated by regions with sparse CTCF binding. Binding also appears concomitant with gene bodies, with distal intergenic regions displaying little CTCF binding (Figure 3.2), this pattern is further observed across other chromosomes (data not shown).

3.2.2 Chromatin Immunoprecipitation followed by next generation sequencing

Chromatin immunoprecipitation (ChIP) uses antibodies to a protein of interest to precipitate fragments of DNA that are bound to the protein. ChIP for CTCF and Rad21 (a Cohesin subunit) were performed to enrich for sequence fragments bound by CTCF and Cohesin. These DNA fragments were then sequenced using the Illumina™ GAIIx platform. Quality of chromatin preparation and the size of the chromatin fragments that are

A



B

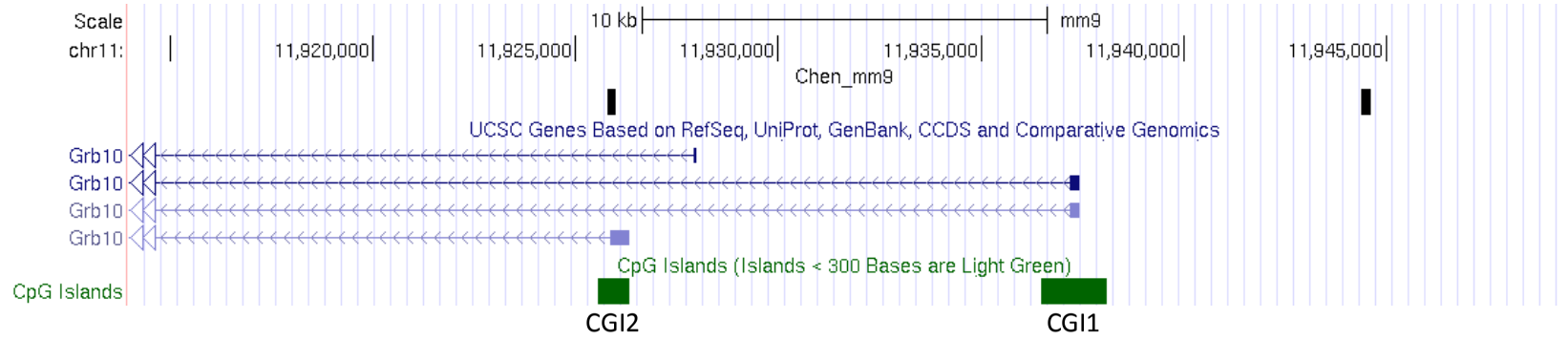


Figure 3.1 CTCF binding in ES cells examined across the *Ddc/Grb10* locus. Black bars in the Chen_mm9 track represent a regions called for CTCF binding, vertical blue bars represent exons and horizontal blue lines represent introns, arrows represent the direction of transciption. (A) shows the entire *Ddc/Grb10* locus and (B) shows just the *Grb10* promoter regions. Regions CGI1 and CGI2 are indicated. CGI2 forms part of the *Grb10* DMR.

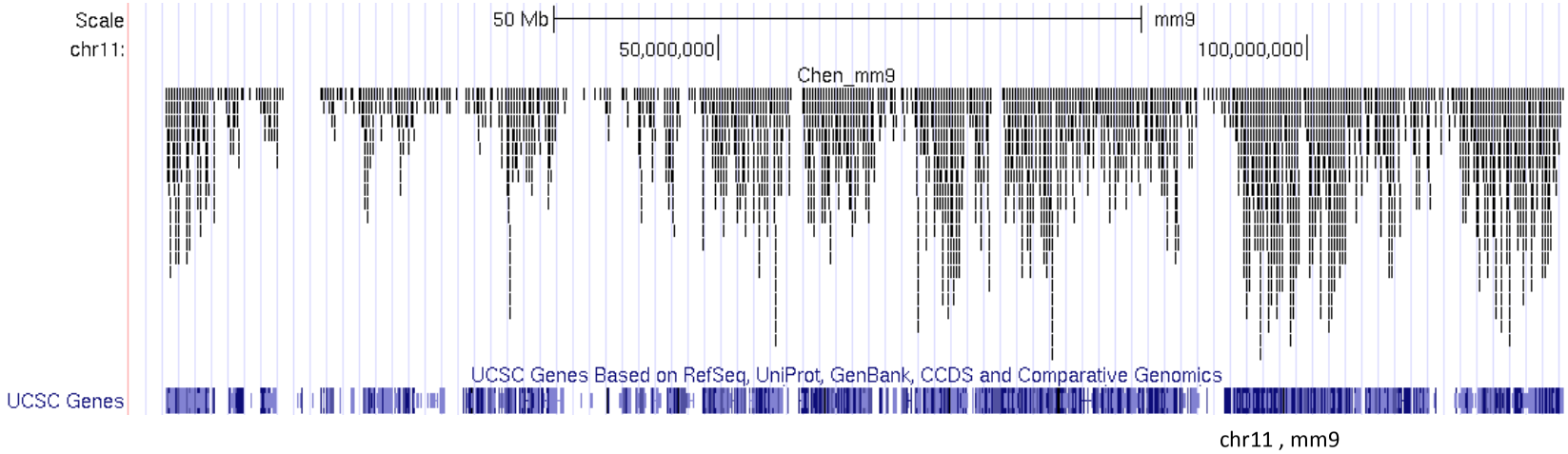


Figure 3.2 CTCF binding regions in embryonic stem cells on chromosome 11. Black bars in the Chen_mm9 track represent a unique regions which binds CTCF. Blue bars represent genes. These data were generated using ChIP-seq from Chen *et al.* 2008. Binding peaks were generated using Useq, and the BED file was created by Dr. Reiner Schulz (King's College London)

used in the immunoprecipitation are integral to successful enrichment; therefore careful optimization of the chromatin extraction protocol was required. DNA fragment length was also a critical issue of the library preparation step of the sequencing protocol and required careful optimization. After sequencing, parent-of-origin specific binding of CTCF and Cohesin was determined utilising SNPs between F1 hybrid mice from C57BL/6 mothers and *Mus mus castaneus* fathers (BxC), together with the reciprocal cross of mice with *Mus mus castaneus* mothers and C57BL/6 fathers (CxB) to discriminate strain-specific from truly parent-of-origin specific binding using some of the ~20 million SNPs between these strains (Yalcin, et al., 2012). Individual reads that aligned to regions identified as CTCF and/or Cohesin binding sites, and that also covered at least one SNP between the two parental strains, were assigned to either the maternal or paternal allele and counted. For each region, the maternal versus paternal read counts were tested for significant deviation from the expected 1:1 ratio, assuming a binomial distribution using calculated p-value cut-offs, and the values were corrected for multiple testing using the Bonferroni method. This method of using hybrid mice bred from inbred lines has been proposed previously to assess allele-specific binding of histone marks from ChIP-seq experiments (Mikkelsen, et al., 2007) however, the authors do not address issues of strain bias or multiple testing.

3.2.3 Optimization of chromatin fragmentation

Chromatin was extracted from frozen 3 week BxC and CxB mouse brain. Chromatin was fragmented via sonication so that the majority of fragment sizes were <10 kbps, to allow for efficient antibody precipitation. Two types of sonication were tested:

1) Probe Sonicator
2) BioRupter

BxC and CxB chromatin was split four ways and subjected to independent treatments, two were subjected to probe sonication, and two were subjected to BioRupter sonication. Different durations of each sonication

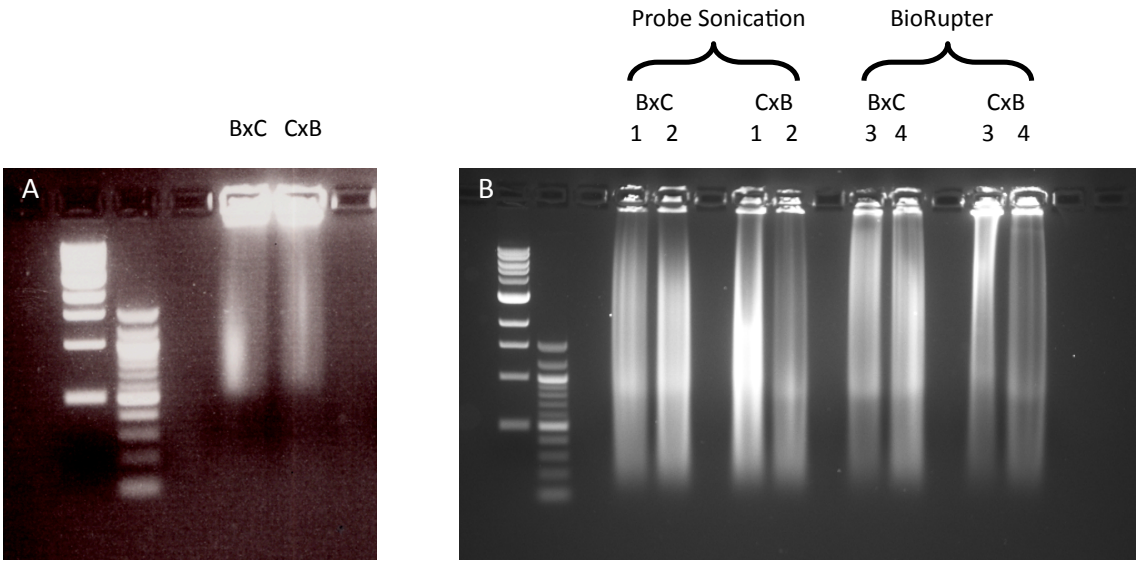
method were tested. After sonication, each treatment was decrosslinked in 1.2 M NaCl at 65 degrees overnight and visualised using a 1.5 % agarose gel and the effectiveness of sonication was judged visually (Figure 3.3). This optimization demonstrated that probe sonication marginally outperformed BioRupter sonication, with little difference between duration of sonication for either treatment. Chromatin was therefore fragmented using probe sonication for 6x. 1 minute On/Off cycle at 40 amplitude prior to the immunoprecipitation step.

3.2.4 DNA fragmentation post immunoprecipitation

Previous attempts GAIIx sequencing of CTCF and Rad21 ChIP on 3 week mouse brain conducted by Ruth McCole (PhD Student, Oakey laboratory) yielded high clonal duplication of reads and therefore a low number of unique reads, thus diminishing the statistical power of the experiment. Analysis of the mean fragment size after immunoprecipitation revealed that this is most likely due to the fact that enrichment is far higher for fragments of DNA > 2 kbps in length (Figure 3.4). This means that during the size selection step of the library preparation step prior to sequencing, where fragment sizes of 300-500 bp are selected, most of the immunoprecipitation will be lost, and only relatively few fragments will be amplified during the amplification step of the library preparation, resulting in high clonal duplication. Therefore an additional DNA sonication step was added to in order to fragment the enriched DNA into sizes optimal for size selection during library preparation.

3.2.5 Covaris™ sonication of Immunoprecipitated DNA

Sonication of the immunoprecipitated DNA was performed using the Covaris™ S-series ultrasonication (Covaris Inc.). No protocol for fragmentation of ChIP DNA is available, so three separate optimizations of low concentration DNA protocols were performed using control ChIP maternal from mouse brain (Figure 3.5). After optimization, the following



C

Treatment	
1	6x. 1 min On/Off cycles 40 amplitude
2	9x. 1 min On/Off cycles 40 amplitude
3	30x. 30s On/Off cycles
4	50x. 30s On/Off cycles

Figure 3.3 Optimization of chromatin sonication. (A) Chromatin before sonication. (B) After sonication using both probe sonicator and BioRupter, using 4 different treatments (1-4) shown in (C). Samples were de-crosslinked after sonication as per method and electrophoresed on a 1.5% agarose gel.

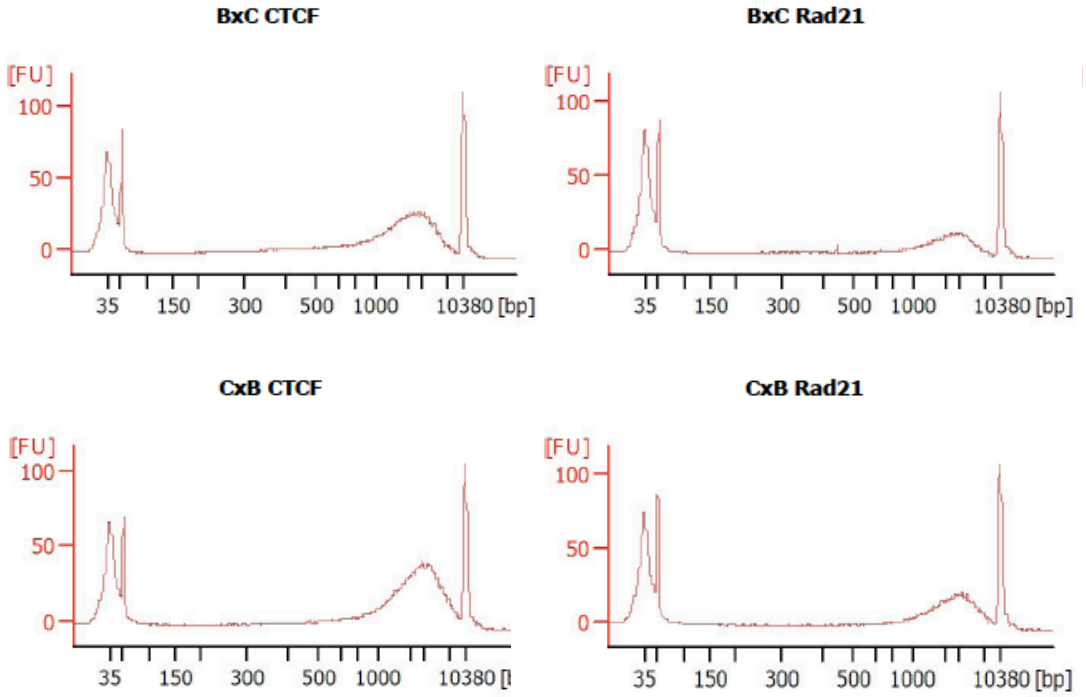
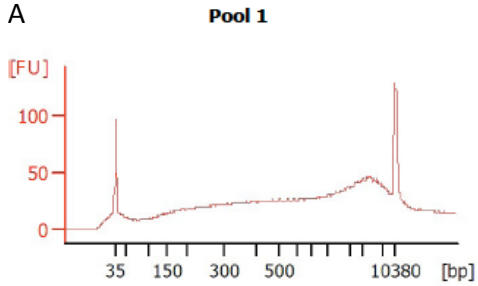


Figure 3.4 CTCF and Rad21 ChIP Fragment length analysis. DNA immunoprecipitated using CTCF and Rad21 antibody was anayzed using the Agilent™ HS DNA bioanalyzer. The average size of fragments immunoprecipitated is greater than 2kb for both CTCF and Rad21 antibodies.

A



B

	1x.
Duty Cycle	5 %
Intensity	3
Cycles per burst	200
Times (s)	80

	2x.
Duty Cycle	5 %
Intensity	3
Cycles per burst	200
Times (s)	80

	3x.
Duty Cycle	5 %
Intensity	3
Cycles per burst	200
Times (s)	80

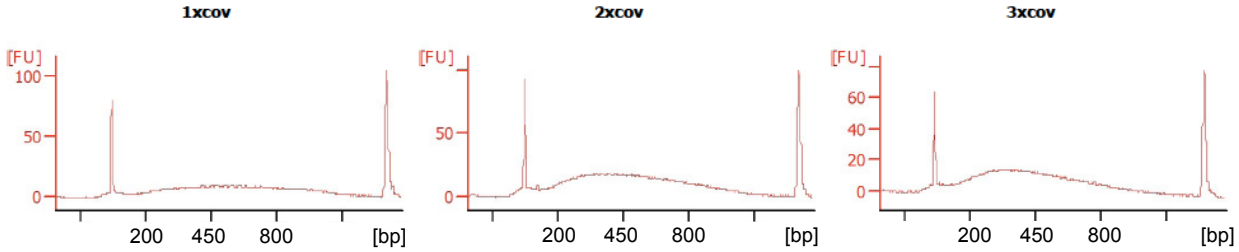


Figure 3.5 Optimization of Covaris™ ultrasonication. (A) Fragment length before sonication. (B) Fragment length after sonication, samples were subjected to 1x, 2x or 3x cycles of stated operating conditions. 2x. cycles produced optimum yield of DNA fragment sized between 300-500bp in length.

protocol was adopted for shearing of DNA fragments from the immunoprecipitation:

Target Size 500bp	2 Cycles of:
Duty Cycle	5 %
Intensity	3
Cycles per burst	200
Times (s)	80

CTCF and Rad21 immunoprecipitated DNA were subjected to Covaris ultrasonication, as described above; this yielded an optimum concentration of DNA fragments at 300-500 bp (Figure 3.6). Fragmented DNA was used to make sequencing libraries, and analysis of fragment length after library preparation showed an average fragment length for all four libraries of between 350 and 430 bp (Figure 3.7).

3.2.6 Read Statistics

Library preparation was completed using the NEBnext sequencing kit and the libraries were sequenced using the Illumina™ GAIIx sequencing platform (Section 2.3.3). The loading of the libraries onto the GAIIx was carried out by the Biomedical Research Centre genomics facility at King's College London, specifically Dr Efterpi Papouli. Bioinformatic analyses were performed in collaboration with Nikolas Barkas (PhD Student, Oakey Lab), who provided programming expertise required to develop the bioinformatic pipeline. After sequencing, reads were mapped to the mouse genome, version mm9, using novoalign (v. 2.01.13, http://www.novocraft.com/). The percentage of identical reads, which represent clonal duplications, was below 6% for all samples (Figure 3.8, B), this represented a significant reduction from that observed in ChIP-seq experiments conducted prior to addition of the fragmentation step to the ChIP-seq protocol. Duplicates of reads were removed so that only uniquely mapped reads were used in further analyses. After duplicate removal this

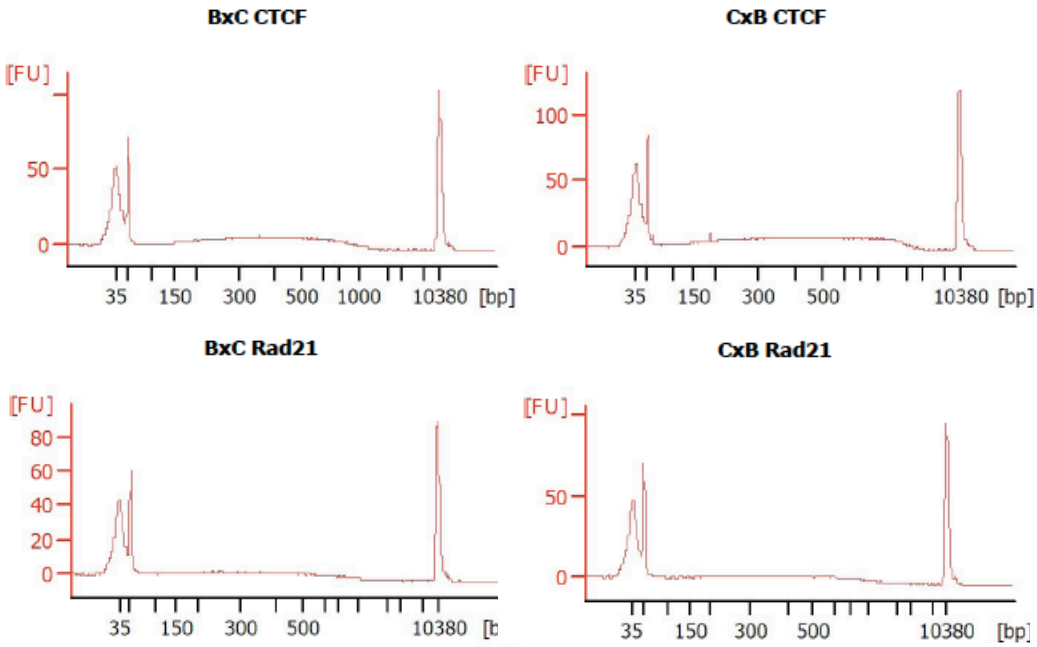


Figure 3.6 Fragmented ChIP DNA. Immunoprecipitated DNA from both reciprocal crosses using both CTCF and Rad21 were fragmented using the optimized conditions. Each of the four treatments was analyzed using the Agilent ™ HS DNA assay, and demonstrate fragment sizes between 150 and 800 bp with the peak between 300 and 500 bp.

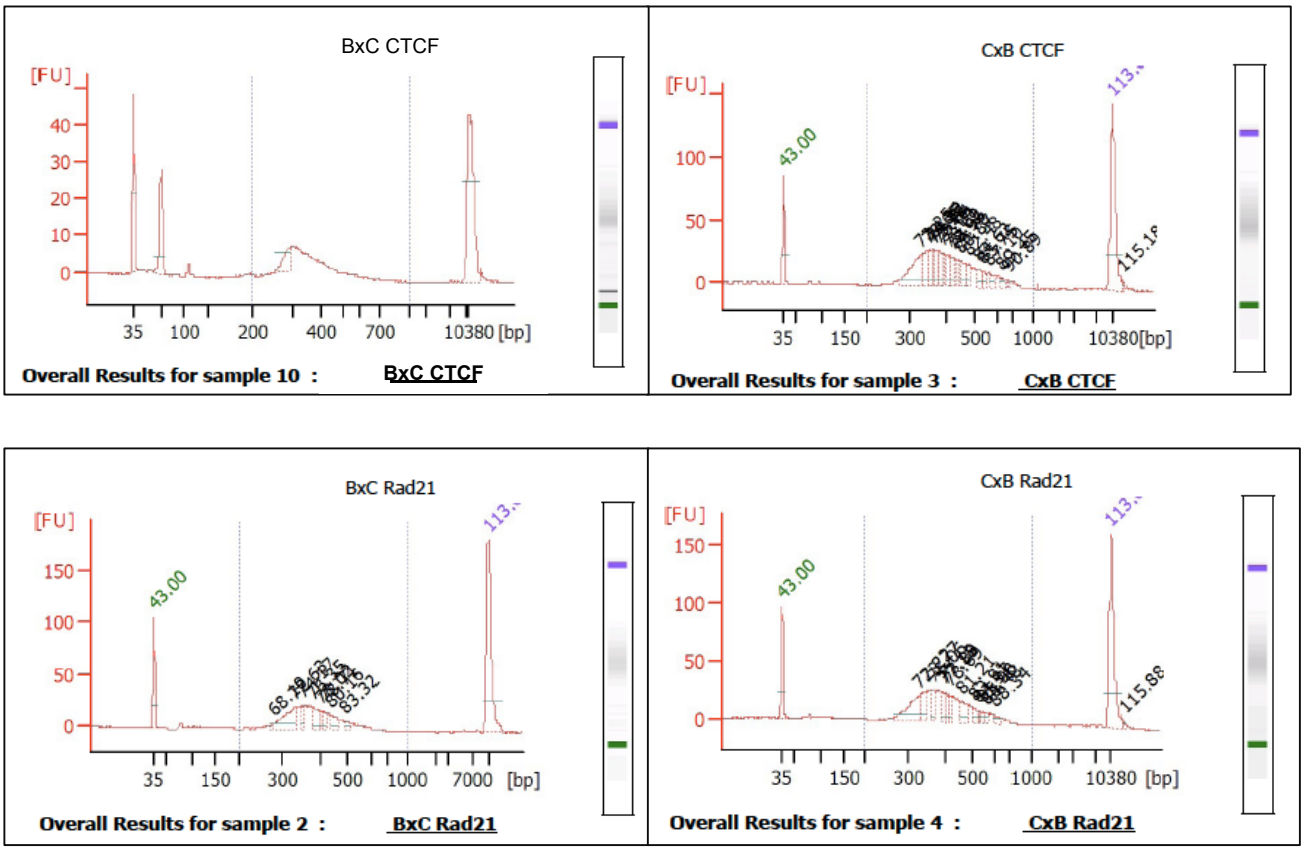


Figure 3.7 ChIP-seq Libraries. Detailed analysis of fragment length of CTCF and Rad21 immunoprecipitated analyzed using the Agilent ™ HS DNA assay. Analysis demonstrates fragment lengths between 300-500bp were captured with the average length of fragment for each falling between 350 and 430 bp.

resulted in 114,443,577 BXC and 120,127,056 CXB CTCF reads and 123,240,573 BXC and 107,956,976 CXB Cohesin reads (Figure 3.8, A and G).

3.2.7 CTCF and Cohesin binding in the genome

USeq (Nix, et al., 2008) was used to identify regions of binding separately for CTCF and Cohesin. The mean peak shift was calculated separately for CTCF and Cohesin using only the first of each paired end matched reads. Peaks were then "called" using peak shifts and window sizes of 138bp and 144bp for CTCF and Cohesin. Peaks were considered representative of a true binding signal if the statistical threshold of a 5 % False Discovery Rate (FDR) was reached. Using these criteria 49,358 CTCF binding sites and 52,938 Cohesin binding sites were detected across the genome.

3.2.8 Validation of CTCF and Cohesin ChIP-seq

Quantitative-PCR assays of *Igf2/H19, Peg10, Nap1l5* and *Nnat* were utilised to validate the ChIP-Seq findings at imprinted loci, *Grb10* is also included and supports previous studies suggestive of *Grb10* binding in brain (Hikichi, 2003). The qPCR validation (Figure 3.9) is in good agreement with ChIP-Seq data (Table 3.1 and Table 3.2, A) with the exception of CTCF binding at *Nnat* and Cohesin binding at *Peg10*. CTCF binding at the *Nnat* DMR has a p=0.08 assayed using qPCR, this is just above the cutoff for binding, therefore the discrepancy between qPCR and ChIP-Seq is attributed to experimental differences between the two methods. At *Peg10*, Cohesin binding is detected by qPCR but no binding peak was identified in the ChIP-Seq analysis. When the stringency of the ChIP-Seq peak detection is relaxed, two binding regions roughly 1 kbps either side of the qPCR regions assayed are detected (Figure 3.9, C), which may account for the signal seen using the qPCR assay and again can be attributed to experimental differences between the two methods.

3.2.9 Allele-specific binding of CTCF

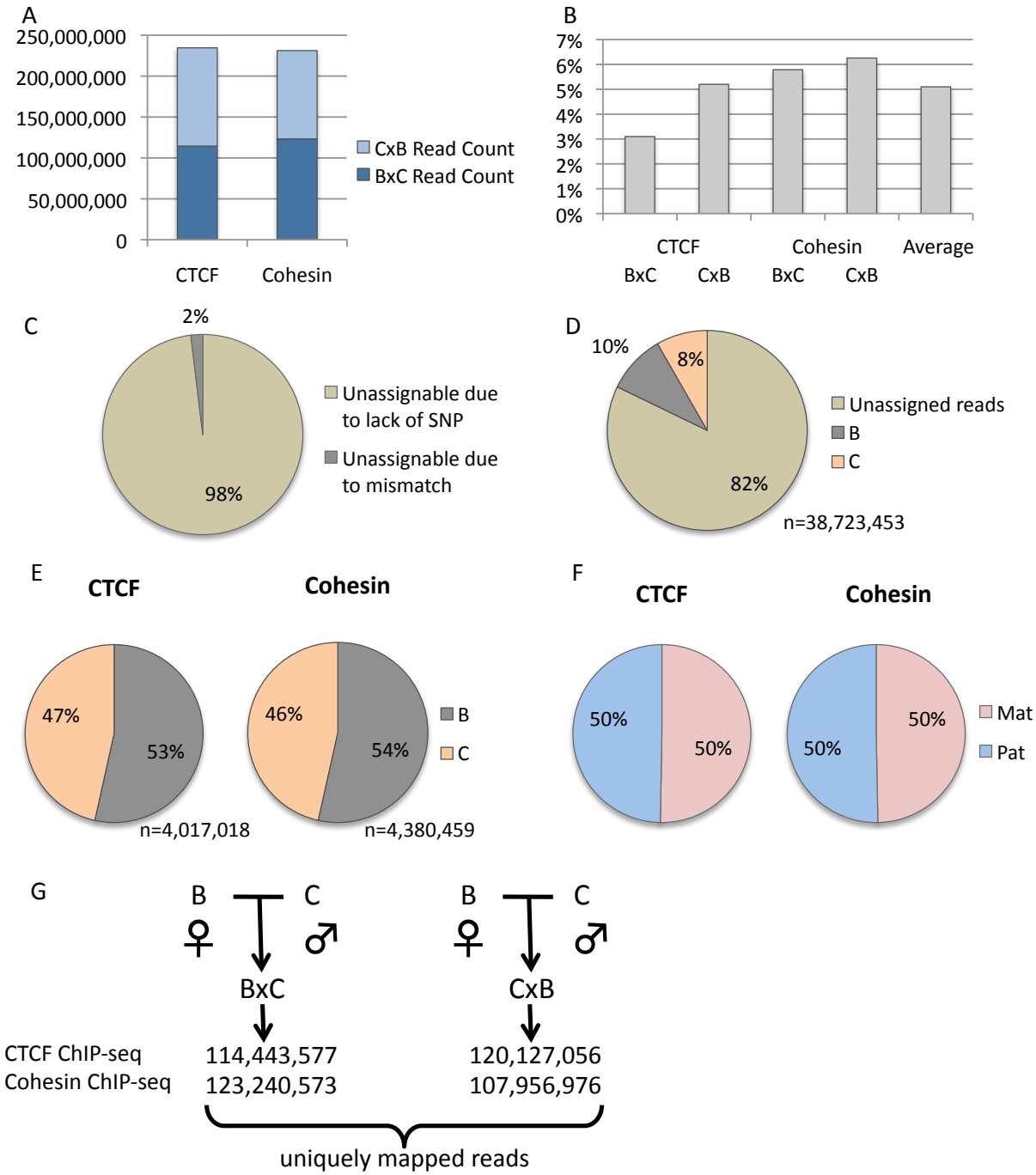


Figure 3.8 Read statistics from ChIP-seq. (A) The number of uniquely mapped reads for CTCF and cohesin ChIP-seq, reads are split by reciprocal cross category; CxB and BxC. (B) The percentage of reads that demonstrate clonal duplication for each lane are indicated. (C,D) Reads were mapped to C57Bl/6 (referred to as B) or *Mus musculus castaneus* (referred to as C) where possible. Of the reads which we not mapped 98% due to an absence of a SNP between B and C mouse strains, and 2% because of a mismatch (ie. the SNP did not map to either B or C). Of the reads which mapped within a binding peak, 10% mapped to B, 8% to C and 82% remained unassigned to either B or castaneus. (E,F, G) A bias was seen towards B for both CTCF and cohesin, however this was resolved after the concatination of the BxC and CxB cross reads.

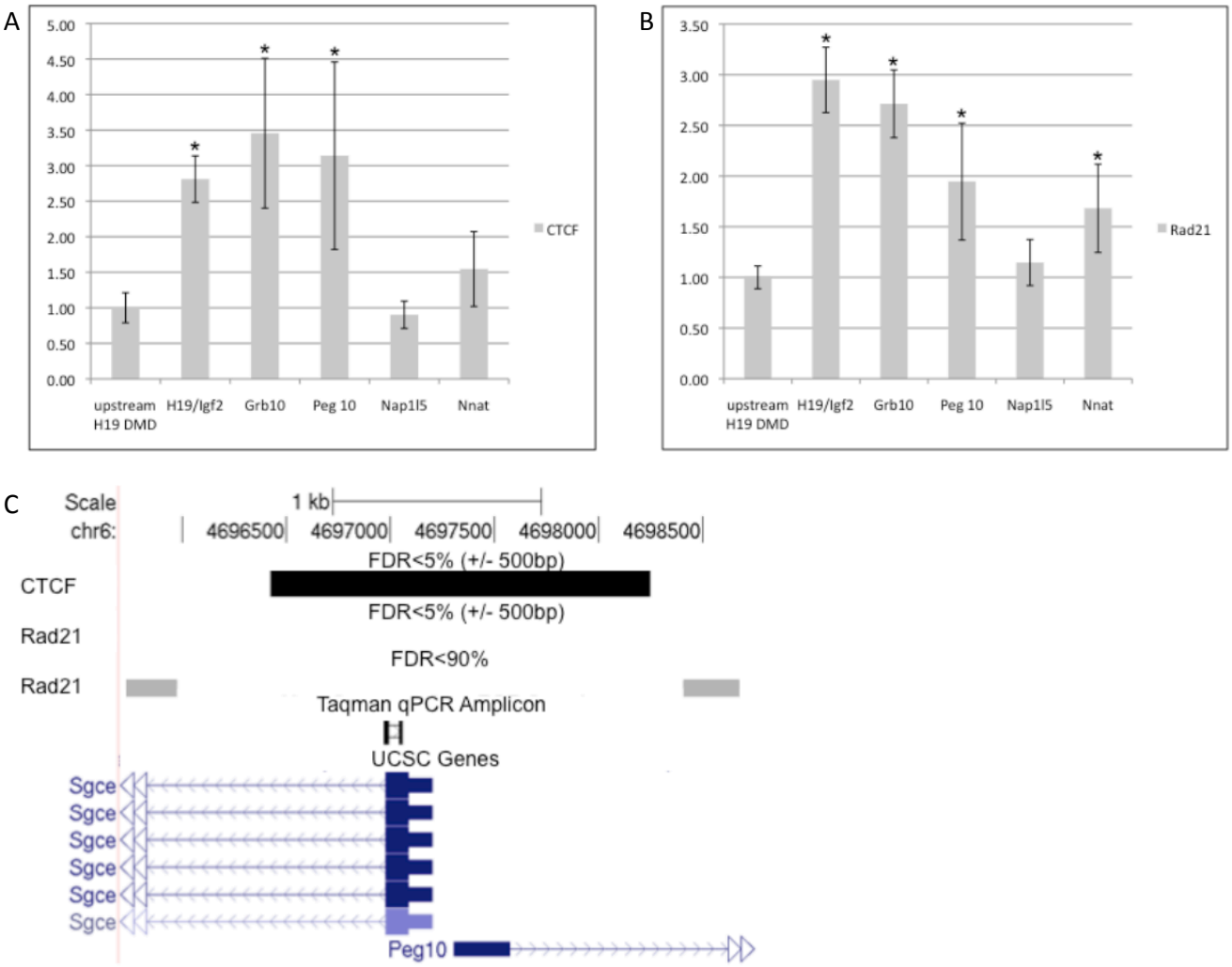


Figure 3.9 Validation of ChIP-seq binding. (A & B). Quantitative PCR analysis showing average CTCF (A) and Cohesin (Rad21) (B) enrichment with standard deviations at five gDMRs. Regions that differ from the negative control upstream H19 gDMR (DMD) with a $p<0.05$ are denoted with an asterisk indicating that these regions show statistically significant levels of CTCF or Cohesin binding above the level of the negative control. C. UCSC screenshot showing CTCF binding peak called at <5% FDR (black) and Cohesin binding peaks called at <90% FDR (grey). The Peg10 TaqMan™ qPCR location is indicated.

In order to assess whether a peak was representative of binding on both parentally inherited alleles, or was representative of parent-of-origin specific binding, a novel bioinformatic pipeline was created. Binding peaks at an FDR of 5 % were expanded by 500 bp upstream and downstream and overlapping peaks merged prior to further analysis, which allowed full length of all reads to be used to map the reads to either the maternal or paternal allele. Sequencing reads that fell within the expanded peaks were then assigned to the maternal or paternal allele using known SNPs between C57BL/6 (B) and *Mus mus castaneus* (C). Considering both CTCF and Cohesin experiments together, of the sequencing reads that mapped within a binding peak 82% could not be assigned, 10% mapped to B and 8% to C (Figure 3.8, D) indicating that a bias is present between the strains of mice with more B reads mapping to the reference genome. This bias was evident towards B for both CTCF and Cohesin and is approximately equivalent in both experiments (Figure 3.8, E). This reference bias is compensated for after the process of concatenation of the reads from the forward and reverse reciprocal crosses (Figure 3.8, E & F). Of the reads that could not be assigned most were not assigned due to the lack of a SNP between the two strains of mice used in the study (~98%) and a tiny proportion (around 2%) was due to a mismatch at the SNP (ie. the base did not map to either B or C genomes) (Figure 3.8, C).

3.2.10 Allele-specific binding of CTCF and Cohesin at imprinting gDMRs

In order to assess the importance of CTCF and Cohesin binding at imprinted loci, binding was examined in detail at 20 known imprinted gDMRs (Table 3.1 and Table 3.2). Of these, 50 %, including the *Grb10* DMR, demonstrated co-ordinate binding of both CTCF and Cohesin at, or within, 1 kbps of the DMR. Of the remainder, 10 % bind CTCF alone, 25% bind Cohesin alone and 15 % bind neither CTCF nor Cohesin at the DMR (Figure 3.10).

The 12 gDMRs that bound CTCF were examined further for allele-specific binding, utilizing the bioinformatic pipeline described in previously in this

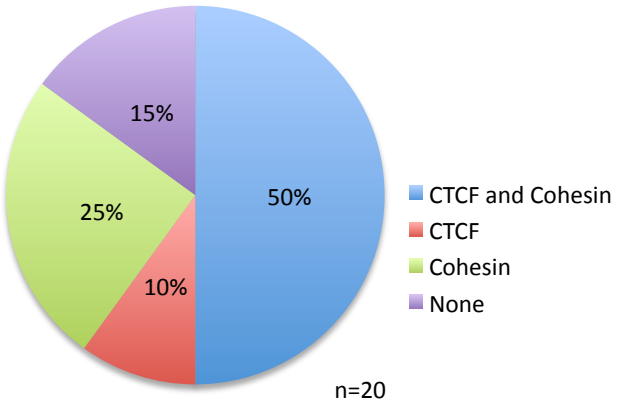


Figure 3.10 CTCF and Cohesin binding at gDMRs. 20 well characterized gDMRs were assessed for CTCF and Cohesin binding. 50% demonstrate binding of both CTCF and Cohesin, 10% bind CTCF alone, 25% bind Cohesin alone and 15% bind neither CTCF or Cohesin.

chapter and in detailed in chapter 2 (section 2.4.1). These analyses demonstrated that five gDMRs bound CTCF in a parent of origin specific manner specifically the *Igf2/H19, Mest, Peg13, Zim2* and *Gtl2/Dlk1* DMRs. *Mest and Zim2* were not previously known to bind CTCF in this manner. Of the seven remaining CTCF bound DMRs, which returned a non-significant p-value for parent-of-origin specific binding, this was mostly due to insufficient power of the binomial test. Parent-of-origin specific CTCF and Cohesin binding at gDMRs could only be tested where one or more SNP(s) were located within the binding peak at the region and a further limitation for the detection of parent-of-origin specific binding occurs when a SNP is located at the periphery of the binding region where there are fewer aligned reads and the statistical power of the binomial test is diminished. This limited the assessment of parent-of-origin specific binding of CTCF at the *Grb10, Mcts2, Nespas, Nnat, U2af1-rs1, Impact* and *Peg10* loci. To look for trends of allele-specific binding at these seven loci, 95 % confidence intervals of the proportion of maternal to paternal reads were calculated to demonstrate allele-specific binding (Figure 3.11). Confidence intervals show the allele-specific binding of CTCF at *Igf2/H19, Mest, Peg13, Zim2 and Gtl2/Dlk1* and are suggestive of parent-of-origin specific binding at *Grb10, Nespas, U2af-rs1* and *Peg10*. In each instance parent-of-origin specific binding is on the unmethylated allele (Table 3.1 and Table 3.2). The only gDMR that decisively showed bi-allelic binding was the *Inpp5f_v2* gDMR, the *Nnat* gDMR could not be assessed as there was no SNP within or near the gDMR.

Allele-specific Cohesin binding was also observed in co-ordination with CTCF at *Igf2/H19* and the *Peg13* gDMRs, with 13 further gDMRs bound by Cohesin, but not in an parent-of-origin specific manner. Scrutiny of the confidence intervals for allele-specific binding of Cohesin revealed much less of a bias towards maternal or paternal binding at all gDMRs compared to the allele-specific effects seen where CTCF binds at imprinted gDMRs. This suggests that Cohesin does not respond to epigenetic differences in

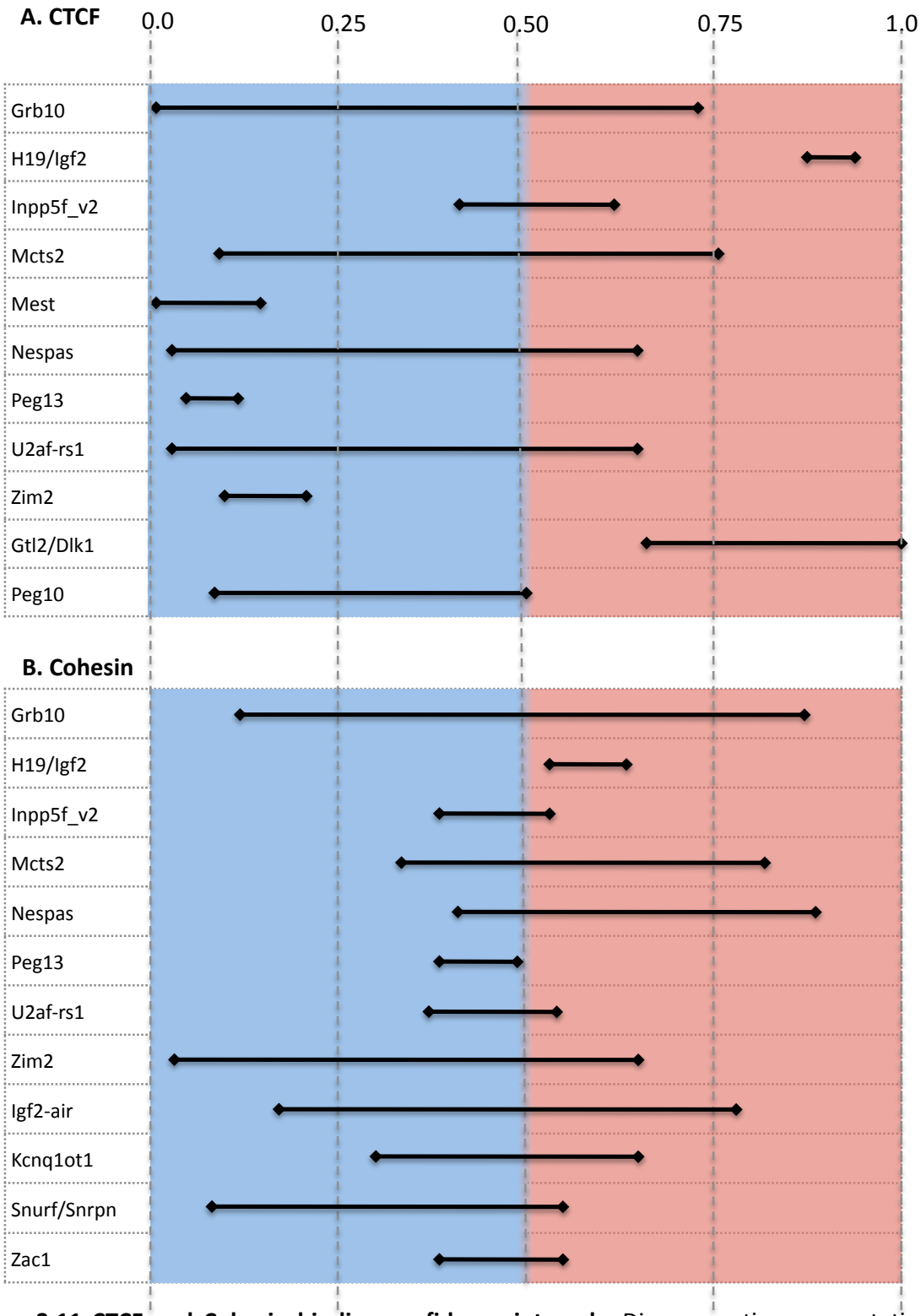


Figure 3.11 CTCF and Cohesin binding confidence intervals. Diagrammatic representation of the 95 % confidence intervals for (A) CTCF and (B) Cohesin (Rad21) binding. Blue represents binding on the paternal allele and red binding on the maternal allele. Binomial tests illustrate that allele-specific CTCF binding favours the unmethylated paternal allele for *Grb10, Nespas, U2af-rs1* and *Peg10* even when the binomial test is not significant.

			CTCF Binding				Cohesin Binding			
gDMR (Wamidex)	**Chr. Position**	**Methylated Allele**	**Distance from peak to gDMR (kb)**	**Useq FDR**	**Binding Allelle**	**P-value for Allele-specific binding**	**Distance from peak to gDMR (kb)**	**Useq FDR**	**Binding Allelle**	**P-value for Allele-specific binding**
Grb10	chr11:11,925,485-11,925,790	Maternal	0	<0.001	Not Sig	N/A	0	0.966	Not Sig	N/A
H19/Igf2	chr7:149,766,168-149,768,424	Paternal	0	0.003	Maternal	8.11E-48	0	<0.001	Maternal	7.60E-04
Inpp5f_v2	chr7:135,831,788-135,832,156	Maternal	0.076	0.001	Biallelic	N/A	0	0.872	Not Sig	N/A
Mcts2	chr2:152,512,491-152,513,011	Maternal	0.937	2.141	Not Sig	N/A	0.955	0.877	Not Sig	N/A
Mest	chr6:30,686,709-30,687,273	Maternal	0	0.018	Paternal	1.54E-12	0	<0.001	No Data	N/A
Nespas	chr2:174,121,208-174,126,482	Maternal	0	1.984	Not Sig	N/A	0	1.609	Not Sig	N/A
Nnat	chr2:157,385,786-157,387,398	Maternal	0.588	0.133	No Assigned reads	N/A	0	0.058	No Data	N/A
Peg13	chr15:72,636,765-72,642,079	Maternal	0	0.002	Paternal	3.18E-39	0	<0.001	Paternal	1.31E-02
U2af1-rs1	chr11:22,871,842-22,872,319	Maternal	0.107	4.684	Not Sig	N/A	0.05	0.023	Not Sig	N/A
Zim2 (Peg3)	chr7:6,680,287-6,684,827	Maternal	0	<0.001	Paternal	3.00E-20	0	2.483	Not Sig	N/A

Table 3.1 CTCF and Cohesin binding at gDMRs. Imprinted regions that demonstrate both CTCF and cohesin binding are listed with their corresponding gDMR positions. The methylated allele is shown and CTCF binding indicated according to the criteria used in this ChIP-seq study. The distance of the binding peak to the DMR is included to illustrate whether the binding is coincident with the DMR (0) or up to 1.5kb away. Most CTCF binding is at or very close to the DMR with the exception of *Mcts2* and *Nnat*. The false discovery rate (FDR) for binding peaks, an indication of confidence in the signal for each peak, is listed and was required to be <5%. The parental allele specificity is shown but for some regions these data did not reach significance or, "no data" was registered which was due either to the absence of a SNP or insufficient sequence reads over the SNP. p-values are given for the allele specific binding. Most Cohesin binding is coincident or very close with the exception of *Mcts2* again. The FDR is given and required to be <5% for binding. Parental specific binding is indicated with corresponding p-values. References to previous binding are indicated. Allele specific binding is graphically indicated blue indicates binding on the paternally inherited allele and red binding on the maternally inherited allele.

A. CTCF Binding							
gDMR (Wamidex)	**Chr. Position**	**Meylated Allele**	**Distance from peak to gDMR (kb)**	**Useq FDR**	**Binding Allelle**	**P-value for Allele-specific binding**	**Previous evidence for binding**
Gtl2/Dlk1	chr12:110,761,563-110,768,989	Paternal	0	<0.001	Maternal	3.91E-03	Lin et al. (2011)
Peg10	chr6:4,697,209-4,697,507	Maternal	0	0.550	Not Sig	N/A	None

B. Cohesin Binding							
gDMR (Wamidex)	**Chr. Position**	**Meylated Allele**	**Distance from peak to gDMR (kb)**	**Useq FDR**	**Binding Allelle**	**P-value for Allele-specific binding**	**Previous evidence for binding**
Gnas-exon1A	chr2:174,153,279-174,153,502	Maternal	0.848	0.027	No Data	N/A	
Igf2r-air	chr17:12,934,163-12,935,573	Maternal	0	3.876	Not Sig	N/A	
Kcnq1ot1	chr7:150,481,060-150,481,397	Maternal	0	0.177	Not Sig	N/A	Fitzpatrick et al. (2007)
Snurf/ Snrpn	chr7:67,149,878-67,150,301	Maternal	0.427	2.869	Not Sig	N/A	
Zac1	chr10:12,810,276-12,810,604	Maternal	0	<0.001	Not Sig	N/A	

C. No CTCF or Cohesin binding			
gDMR (Wamidex)	**Chr. Position**	**Meylated Allele**	**Previous evidence for binding**
Nap1l5	chr6:58,856,690-58,857,056	Maternal	
Rasgrf1	chr9:89,774,406-89,774,691	Paternal	Yoon et al. (2005)
IMPACT	chr18:13,130,706-13,132,250	Maternal	

Table 3.2 gDMRs that bind CTCF or Cohesin or neither. Listed are imprinted regions that demonstrate either (A) CTCF binding, (B) Cohesin binding or (C) neither, shown with their corresponding gDMR positions. For each DMR the methylated allele is shown and CTCF binding indicated according to the criteria used in this ChIP-seq study. The distance of the binding peak to the DMR is included to illustrate whether the binding is coincident with the DMR (0) or up to 1.5kb away. The false discovery rate (FDR) for binding peaks, is listed and was required to be <5%. The parental allele specificity is shown where significant with p-values given for the allele specific binding. Most Cohesin binding is coincident or very close with the exception of *GNAS-exon1a*. The FDR is given and required to be <5% for binding. No regions showed allele-specific binding with statistical confidence

methylation at gDMRs in the same way that CTCF does, but may play a role in transcriptional activation at some loci, and therefore binds preferentially on the active allele. This 'independent regulation' model excludes Cohesin from actively interpreting epigenetic marks to regulate gene expression and limits it parent-of-origin specific binding role to a bystander effect. An alternative 'co-ordinate regulation' model is that Cohesin does not bind in an allele-specific manner, unless it binds co-ordinately with CTCF, this may be suggestive of a mechanism whereby binding of CTCF directs the Cohesin complex to the unmethylated allele of the chromosome, and CTCF/Cohesin together then act to regulate epigenetic control of gene expression. Binding at several loci contradict this model, however, for example the confidence intervals at the *Nespas* gDMR are suggestive of paternal binding of CTCF and maternal binding of Cohesin, which would be incompatible with the 'co-ordinate regulation' model.

3.2.11 Overlap of CTCF and Cohesin binding

To investigate co-ordinated roles for CTCF and Cohesin more broadly across the genome, a further genome-wide analysis of overlap of CTCF and Cohesin binding was conducted, which demonstrates that in brain tissue of the 49,358 CTCF binding sites and 52,938 Cohesin binding sites, just over half of the CTCF (55 %) and Cohesin (51 %) sites overlapped (Figure 3.12). These analyses suggest that there may be both independent and combined roles for these two proteins in the mouse genome, and this may also apply at imprinted loci.

From these results and previous evidence at other imprinted loci it is proposed that multiple mechanisms are responsible for the epigenetic control of gene silencing associated with genomic imprinting. CTCF may play a major role in controlling imprinting at multiple gDMRs by binding to the unmethylated allele, and thus may provide a one model of imprinting control. These results also suggest that Cohesin may play a role in imprinting control when co-ordinately bound with CTCF. These analyses

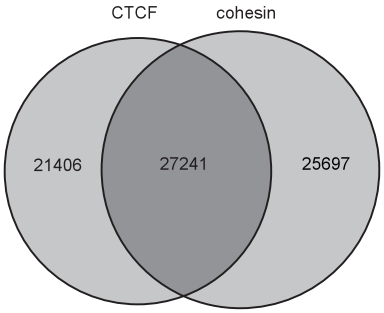


Figure 3.12 Overlap of CTCF and Cohesin binding in mouse brain. Just over half of CTCF (55%) and Cohesin (51%) binding sites are shared suggesting both independent and combinatorial functions for CTCF and Cohesin.

both at gDMRs and genome-wide illustrate both an independent and co-ordinated roles for these factors, and this supports previous studies into CTCF and Cohesin binding (Lin, et al., 2011).

3.2.12 CTCF binds the consensus motif in brain

CTCF has been shown to bind to a specific DNA sequence motif in ES cells (Chen, et al., 2008) and liver (Schmidt, et al., 2010) (Figure 3.13, A). In order to consider CTCF binding motifs in brain, CTCF binding peaks in brain were analysed for overrepresented sequence motifs using the MEME motif finding tool (Bailey, et al., 2006). The consensus binding motif was identified in brain with a high degree of certainty ($p = 2.8 \times 10^{-70}$) and visual inspection of the motif compared to those previously identified in ES cells and Liver show that the motif is extremely similar to those previously identified (Figure 3.13, A) (Chen, et al., 2008, Schmidt, et al., 2010).

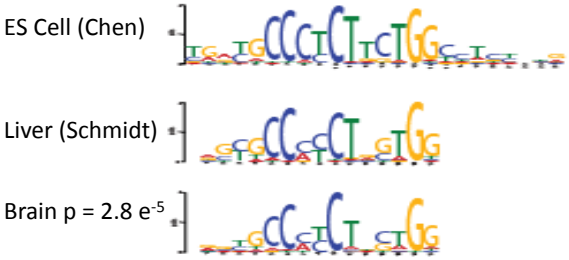
3.2.13 CTCF binds preferentially to unmethylated DNA in brain

During the course of these experiments, genome-wide base pair resolution data from bisulphite-seq experiments on 8 week mouse brain were published (Xie, et al., 2012) facilitating comparisons between the level of DNA methylation and binding of CTCF in mouse brain. This analysis confirmed the previously reported finding that CTCF does not bind effectively to methylated DNA. Genome-wide CpG methylation levels are approximately 60.8 % in mouse brain, however methylation at empirically observed CTCF binding locations is 51.9 %. Cytosine methylation, when not in a CpG context, is also reduced in CTCF binding sites despite the paucity of 5-methylcytosine outside of CpGs (2.5 % genome wide compared to 2.1 % within CTCF binding sites), both these observation are highly significant ($p < 1 \times 10^{-6}$) (Figure 3.13, B).

3.2.14 Genomic location of CTCF binding in mouse brain

Analysis of the genomic distribution of CTCF shows that CTCF binds most frequently in regions of DNA that contain gene-coding regions. CTCF

A



B

% methylation

Cytosine context	Genome Wide	Within CTCF binding site	P-value
All CpG Residues	60.8%	51.9%	<1x10-6
All Non CpGs	2.5%	2.1%	<1x10-6

C

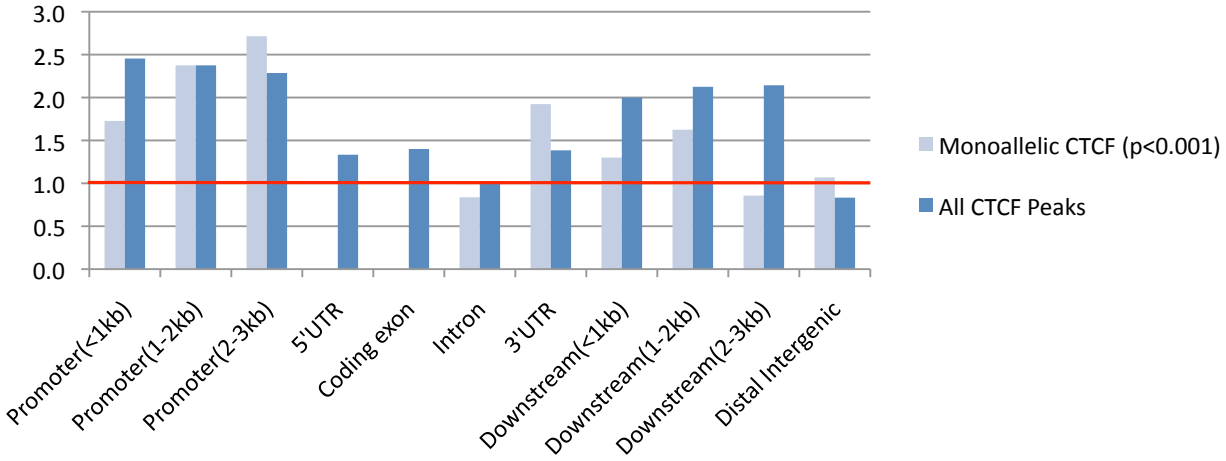


Figure 3.13 CTCF binding analyses (A) Meme motif finder was executed on the CTCF binding locations identified by ChIP-seq in Brain tissue, and this was compared to motifs identified previously in ES cells and Liver. All three datasets found a similar motif, although there was slight variation in the Liver motif. (B) The level of Cytosine methylation was examined, using methylation data from Xie *et al.* (2012) to look at cytosine methylation at CTCF binding sites in brain compared to genome wide. This confirmed that CTCF prefers to bind DNA in the absence of cytosine methylation (C) Genomic locations of CTCF binding are represented normalized to the proportion of the genome that constitutes each location (represented by the red line). This was considered for all CTCF peaks, and separately for CTCF that bind on only one allele. CTCF binding appears enriched upstream and downstream of gene bodies and depleted in introns and intergenic regions. A similar patterns is seen for mono-allelic CTCF binding sites, however these sites are also completely depleted in coding exons and 5' UTRs.

binding is over-represented at promoter regions, 5' UTRs, coding exons, 3' UTRs and regions immediately downstream of gene bodies, and is only under-represented at distal intergenic regions of DNA (Figure 3.13, C). This analysis substantiates previously published evidence that CTCF plays a significant role in controlling gene expression (Bell, et al., 1999), and suggests that it plays no role in non-transcriptional cell processes, for example DNA replication or DNA damage repair.

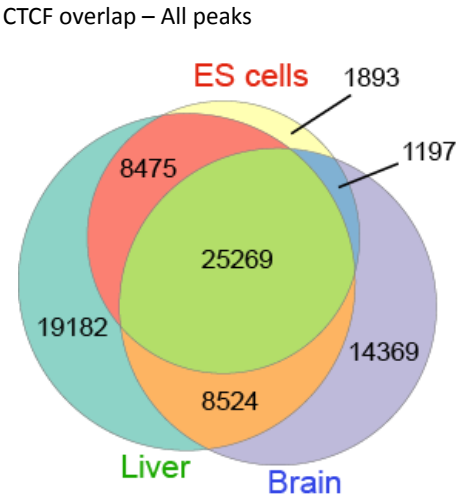
Analysis of the genomic distribution of parent-of-origin specific CTCF binding sites reveals a similar pattern to that seen at all CTCF binding sites genome-wide, with one obvious difference. Parent-of-origin CTCF binding is under-represented at 5' UTRs and coding exons. This may be due to it playing a role analogous to the *Igf2/H19* DMR where CTCF functions to insulate a promoter from a distal enhancer element (Figure 3.13, C).

3.2.15 Tissue specific differences in genome-wide CTCF binding

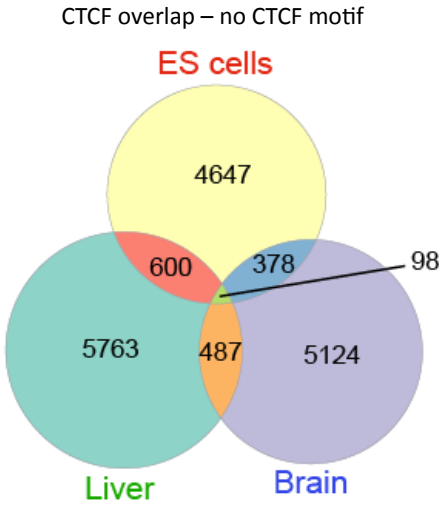
The distribution of binding sites genome-wide in 3 week brain were compared to ES and liver cells (Chen, et al., 2008, Schmidt, et al., 2010) and significant overlap was observed (Figure 3.14, A). To identify the correct size of peak to make appropriate comparisons between the datasets, iterative comparisons with expanding peak sizes, of overlap of binding between brain vs. liver (Schmidt, et al., 2010) and ES-cells (Chen, et al., 2008) were made, and compared with overlap between brain vs. a randomly generated set of peaks. Above a peak size of 1 kbps, any increase in overlap between the dataset is due to noise (Figure 3.14, C & D).

A small number of CTCF binding sites are present exclusively in ES cells, presumably related to pluripotency, and around a third of binding sites in each differentiated tissue (brain and liver) are unique to these tissue only (Figure 3.14, A). These findings show that CTCF binding changes globally during lineage commitment, and that these changes are dependent on the cell type, they also demonstrate that there are a substantial proportion of

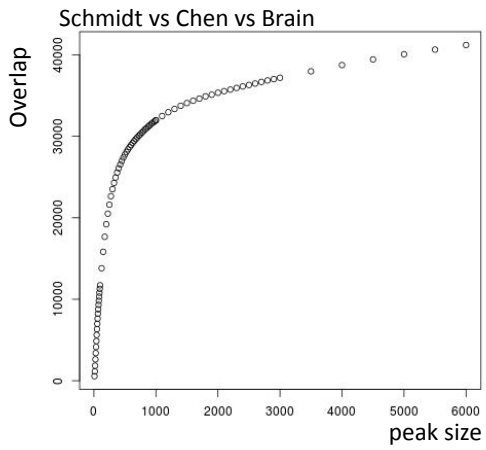
A



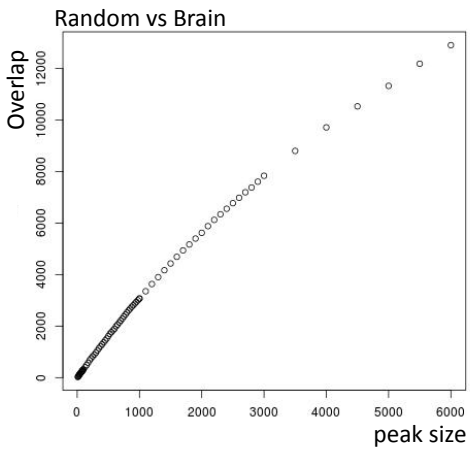
B



C



D



E

% Unique Peaks

	ES Cells	Liver	Brain
All peaks	5.1%	31.2%	29.1%
No motif	81.2%	82.9%	84.2%

Figure 3.14 Overlap of CTCF binding between tissues. (A) CTCF binding sites, as determined by ChIP-seq, compared between ES cells, liver and brain tissue demonstrating significant overlap of CTCF binding in different tissues. (B) Binding sites determined by ChIP-seq, but that do not contain the CTCF motif, were compared between ES-cells, liver and brain tissue. Overlap of non-motif binding was much lower than when all binding sites were considered (C & D) To identify the correct size of peak to make appropriate comparisons between the datasets, the optimum size for the binding peak was determined by iterative comparison of overlaps in brain, liver (Schmidt) and ES-cell (Chen) datasets. From the graph in (C), above a peak size of 1kb, any increase in overlap between the dataset is due to noise. The graph in (D) is the control for (C) whereby a randomly created dataset is plotted against brain CTCF ChIP-Seq data. (E) The percentages of shared peaks for each tissue type for all peaks and for non-motif peaks.

CTCF sites that are invariant across tissue types consistent with previously published data (Kim, et al., 2007).

3.2.16 CTCF binding in the absence of the consensus motif

Interestingly, the same comparison between tissues for CTCF binding in the absence of the consensus binding motif shows a much-reduced overlap, suggesting tissue-specific binding of CTCF occurs in the absence of the primary identified motif (Figure 3.14, B). The number of unique CTCF peaks in brain is 29.1 % with a motif (84.2 % without), ES cells 5.1 % (81.2 % without) and liver 31.2 % (82.9 % without) (Figure 3.14, E).

Comparing the overlap of empirical binding between brain, liver and ES cells in absence of the consensus binding motif suggests other sequences may direct tissue-specific CTCF binding during differentiation. Of the binding peaks found in brain, 14 % lacked the canonical motif, which is lower than the proportion identified in liver where 25 % of sites lack their consensus sequence (Schmidt, et al., 2010). As these data illustrate that in the absence of a canonical motif CTCF binds selectively in different tissues; it is pertinent to investigate the sequence characteristics of tissue-specific CTCF binding sites in different tissue.

3.2.17 Motif finding at tissue-specific binding sites

Binding sites identified as unique to a specific tissue were analyzed individually for motifs using MEME. Two similar CTCF motifs are found in binding sites unique to ES cells and liver, but not in sites unique to brain (Figure 3.15). A unique brain-specific motif was identified, suggesting a potential sequence target for brain specific CTCF binding. Another unique binding motif is also observed in ES cells. These observations suggest that sequence-specific mechanism exists to ensure CTCF binding profiles can vary between different tissue types (Figure 3.15).

3.3 Discussion

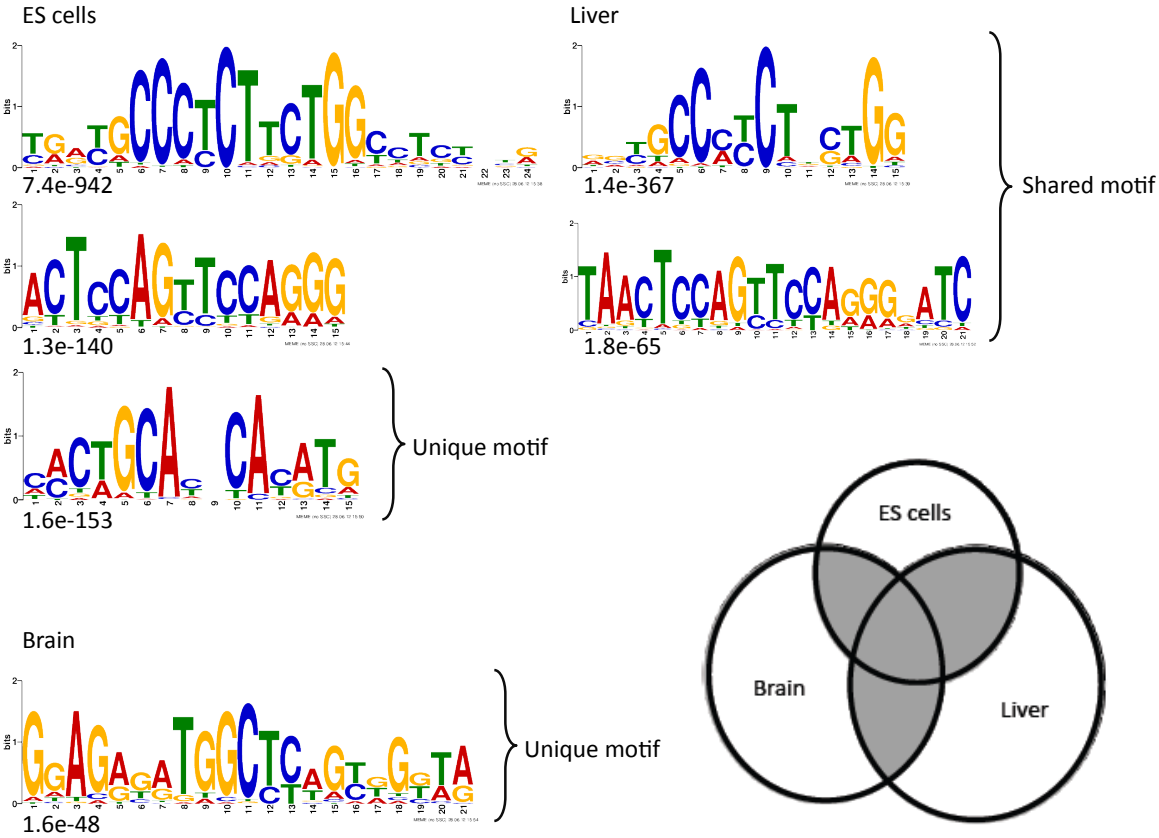


Figure 3.15 Tissue-specific CTCF binding motif discovery. Binding sites identified as unique to a specific tissue were analyzed individually for consensus motifs using Meme, all shared peaks were excluded from this analysis. ES cells and liver but not brain identified the known CTCF motif (not shown). Two similar motifs were identified in ES cells and liver, and further unique motifs were identified for ES cells and for brain. White sections of Venn diagram illustrate the non-shared binding peaks used for motif discovery in Meme.

These analyses provide the first systematic assessment of CTCF and Cohesin binding at all known gDMRs. The results demonstrate that allele-specific CTCF binding occurs at multiple imprinted loci, supporting the assumption that CTCF binding plays mechanistic role in controlling imprinting at some loci. Eight imprinted gDMRs exhibit CTCF binding which is likely to be allele-specific, in the presence of Cohesin. Two further gDMRs exhibit allele-specific binding of CTCF independently of Cohesin.

Co-ordinate and separate biding of CTCF and Cohesin, both at gDMRs and genome-wide, is also suggestive of the multiple functions of the two proteins when bound separately or in combination. However Cohesin, when bound alone at gDMRs may not be involved in imprinting control, as it is not parent-of-origin specific in the absence of CTCF. In addition to illustrating a role for CTCF at multiple imprinted loci, these results also show that other imprinting mechanisms must play a role in gene silencing, as several imprinted loci do not bind CTCF at the gDMR.

Investigation of CTCF binding genome-wide reveals the importance of CTCF in controlling genomic regions rich in coding genes, and that CTCF does not bind to regions of methylated CpGs, concordant with previously published data. Comparison of CTCF binding across different tissues suggests a role for CTCF in determining differential gene expression observed in different cell types and this is potentially directed by the ability of CTCF to bind different seqence motifs in different tissues. CTCF is known to bind different sequences via its multiple zinc-finger domains (Filippova, et al., 1996) however these analysis provide the first evidence that it uses particular sequence motifs in different tissues to direct tissue-specific binding.

Chapter 4

Epigenetic characterisation of the *Ddc/Grb10* locus

4.1 Introduction

The *Ddc_exon1a* transcript is expressed predominately in the developing heart and in the brain. Examination of *Ddc_exon1a* expression using microarray and locus specific sequencing approaches have shown this transcript to be imprinted in heart, with bi-allelic expression in brain (Menheniott, et al., 2008). Subsequent RNA-seq experiments suggest that *Ddc_exon1a* may be imprinted in the pre-optic area of the brain (Gregg, et al., 2010), however these data have been subject of hotly contested debate (DeVeale, et al., 2012, Kelsey, et al., 2012). In this chapter a model of imprinting at the locus is explored combining expression analysis in mouse and human with epigenetic analysis based on common mechanisms of imprinting control seen at other loci and the work covered in chapter 3. These included methylation differences, CTCF binding and analysis of antisense transcription.

4.2 Co-ordinate regulation of *Grb10* and *Ddc*

Ddc and *Grb10* genes are located adjacent to one another on chromosome 11 approximately 32.5 kbps apart and form an imprinted cluster (Figure 4.1); both have expression patterns that are controlled by a region of DNA located at the 5' end of the *Grb10* gene body, which also constitutes the gDMR. Expression of *Grb10* is well characterised in brain and peripheral tissue such as liver and analysis of mice harbouring maternal or paternal duplication of chromosome 11 reveals transcripts of *Grb10* initiates from different leader exons when expressed from the paternal allele compared to the maternal allele (Arnaud, et al., 2003). In brain *Grb10* is imprinted and expressed from the paternal allele, expression initiates from a brain specific promoter at exon 1B (Figure 4.2, A). In the liver expression comes

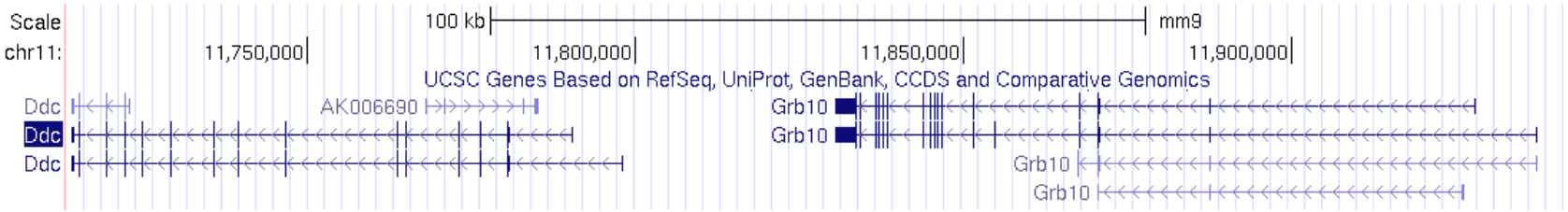


Figure 4.1 *Ddc/Grb10* locus. Screenshot from the UCSC browser (genome build mm9) showing the *Ddc/Grb10* locus. Genes are shown in blue with vertical lines representing exons and horizontal lines introns, chevrons indicate the direction of transcription.

from the maternal allele, and here *Grb10* expression comes from the exon 1A (Figure 4.2, B) (Arnaud, et al., 2003), thus it has been postulated that a switch from paternal *Grb10* expression to maternal *Grb10* expression may be dependent on which leader exon is being used.

Imprinted expression of *Grb10* has not been precisely defined in heart. Existing data show *Grb10* is widely expressed from the maternal allele in the heart during development with paternal *Grb10* expression far more spatially restricted. Maternal and paternal imprinted expression of *Grb10* expression has previously been examined using transgenic animals that carry a *lacZ* reporter under the control of *Grb10* expression (Figure 4.3, A) (Charalambous, et al., 2004, Garfield, et al., 2011). Using the same strategy, histological analysis of whole mount bisected embryos of transgenic mice that harbour a β-*geo* cassette (*Grb10KO*) inherited either via the maternal or the paternal line was performed (Figure 4.3, B & C) (wholemount X-gal staining and imaging, were performed by Dr Michael Cowley, King's College London). X-gal staining indicated extensive maternal expression throughout the heart and a small subset of cells displaying paternal expression (Figure 4.3, C).

Previously published data utilizing RNA-seq to identify imprinting, has suggested that *Ddc_exon1a* might also be imprinted in specific regions of the neo-natal brain (poster presentation by Christian Gregg, Harvard, MA, USA, Gordon Epigenetics Meeting, Boston 2009) which is contrary to published data (Menheniott, et al., 2008). Choroid plexus and lepto meninges elicit polymorphic imprinting of several imprinted genes including *Grb10, Igf2* and *H19* (Charalambous, et al., 2004, Garfield, et al., 2011), thus it was postulated that bi-alleleic expression of *Ddc_exon1a* in one part of the brain might be masking imprinted transcripts in other brain regions. Building on the hypothesis that both *Ddc_exon1a* and *Grb10* are imprinted and expressed from the paternal allele in regions of the brain and in a subset of cardiomyocytes, a model of co-ordinate imprinting

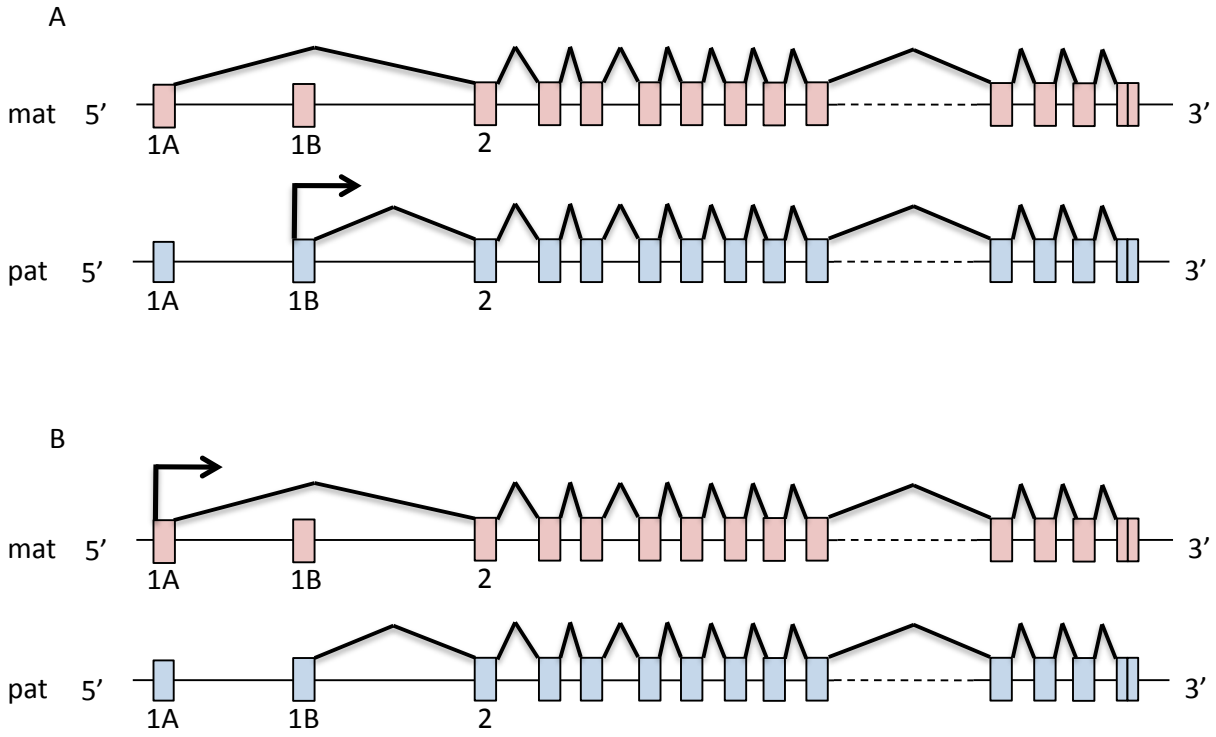


Figure 4.2 *Grb10* expression in mouse. Allele specific expression and promoter use observed in Arnaud *et al.* (2003) using mice with uniparental duplication of *Grb10* on chromosome 11. Boxes represent exons and arrows indicate direction of transcription. Red boxes represent the maternal allele and blue the paternal allele. (A) shows imprinted expression of *Grb10* in brain and (B) shows imprinted expression of *Grb10* in liver.

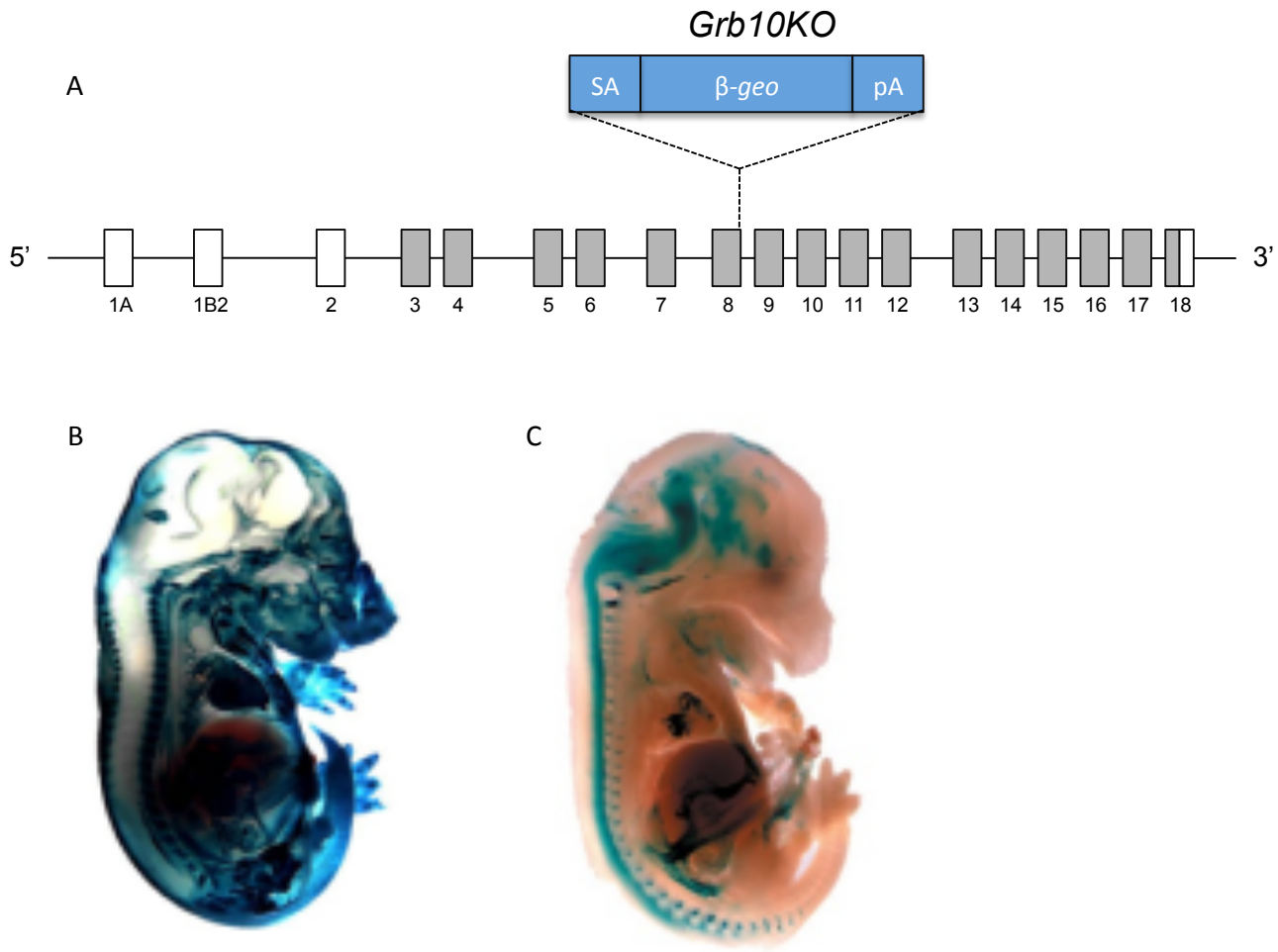


Figure 4.3 Parental origin of *Grb10* expression. Grb10 reporter expression when transmitted through the maternal or paternal line at e15.5. (A) The *Grb10* gene was truncated by the addition of a *β-geo* cassette in exon 8 creating a truncated *Grb10* gene with a *LacZ* reporter (*Grb10KO)*. (B) When the *Grb10KO* gene is transmitted from the mother extensive expression is seen in peripheral tissues including muscle, liver and heart. (C) When the *Grb10KO* allele is transmitted through the paternal line expression is mainly in the brain and CNS and a subset of cells in the developing heart.

regulation was developed predicting these two cell types behave similarly in terms of *Grb10* and *Ddc* expression(Figure 4.4, A). This model proposed that an epigenetic switch induces a co-ordinated change in *Ddc* and *Grb10* expression between peripheral tissue (such as liver) compared to a subset cells found in the brain and heart (Figure 4.4, A & B). The mechanism would involve epigenetic factors in the brain and subset of heart cells being permissive for paternal expression of *Ddc_exon1a* and *Grb10* from the 1B allele, with expression from the maternal allele epigenetically silenced. In order to test this model, histological analysis of paternal *Grb10* expression was tested in heart to assess for co-expression with Ddc. *Ddc_exon1a* imprinting was analysed in specific brain regions and transcript analysis of paternal *Grb10* was performed.

4.2.1 Expression of paternal *Grb10* in the developing heart

To assess co-expression of *Grb10* from the paternal allele with *Ddc_exon1a* in heart, histological analysis was performed on *Grb10KO* mice harbouring a *lacZ* reporter construct within the *Grb10* gene body. The *Grb10KO* alleles were generated using a commercial gene trap method, where a β-*geo* cassette was integrated randomly into the genome of embryonic stem cells. The cassette consists of a 5' splice accepter site, a *lacZ* reporter gene, a neo resistance gene for embryonic stem cell selection and a poly(A) site to terminate transcription. The β-*geo* cassette is located at the 3' end of exon 8 (Figure 4.3, A). When the reporter transcript is transmitted through the paternal line *Grb10* expression is perturbed and expression of the lacZ protein occurs. By using mice where the *Grb10KO* allele has been inherited from the father ($Grb10^{+/KO}$), specific tissue localization of Grb10 expression can be observed as blue staining when stained with X-gal. Mice harbouring the *Grb10KO* allele inherited on the paternal allele have been characterised phenotypically and expression of *Grb10* modulates social behaviour, presumably via expression in the brain, however the effect of paternal expression in heart has yet to be elucidated, and is complicated by ubiquitous maternal expression of *Grb10* (Garfield, et al., 2011). Maternal

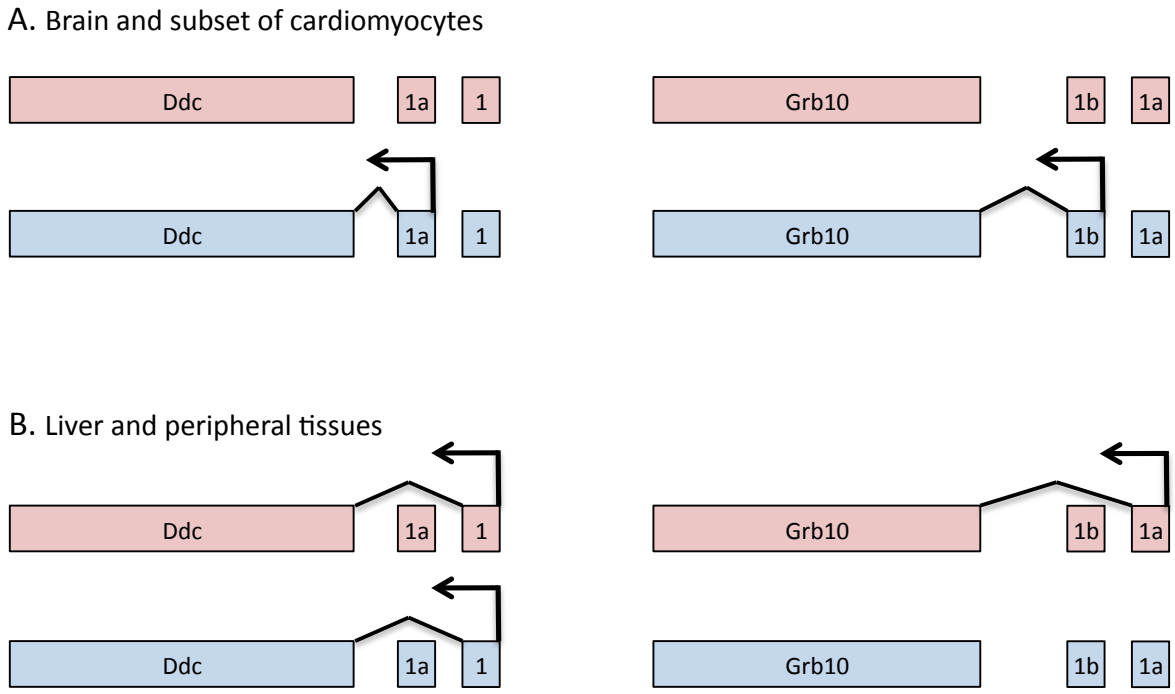


Figure 4.4 Proposed model of coordinated imprinting of *Ddc* and *Grb10*. (A) Epigenetic environment environment in brain and subset of heart cells is permissive of paternal expression of *Grb10* initiating at exon 1b and paternal expression of *Ddc* initiation at exon 1a. (B) In Liver and peripheral tissues expression of *Grb10* is maternal and initiates from exon 1a, whilst *Ddc* expression is from exon 1 and is from both maternal and paternal alleles.

deletion of Grb10 results in fetal overgrowth in liver and overall bodyweight, however no difference has been observed in the size of neonatal heart, thus the role of *Grb10* in heart is still unknown (Charalambous, et al., 2003, Garfield, et al., 2011). In order to analyze overlap of expression between paternally expressed *Ddc_exon1a* and paternally expressed *Grb10* in the developing mouse heart, tissue from e18.5 day $Grb10^{+/KO}$ embryos (from Dr Andrew Ward, Unviersity of Bath, UK) were dissected and cryoembedded in OCT. Serial sections from the cryoembedded embryos were mounted and stained alternately for either Ddc using antibody staining or for paternal *Grb10* expression using X-gal. Immunological staining of the Ddc protein in heart is used as a proxy for paternal *Ddc_exon1a* expression since there is no *Ddc_canonical* and negligible maternal *Ddc_exon1a* expression in the developing heart ((Menheniott, et al., 2008) and discussed in section 5.3.2). Consecutive saggital sections were analyzed for paternal *Ddc_exon1a* and paternal *Grb10* expression (Figure 4.5, A & C). Ddc expression, is shown using Dab staining and is predominately localized to the right ventricular myocardium and the right portion of the interventricular septum (IVS) (Figure 4.5, B & D). Expression of paternal *Grb10* appears punctate and localized to the IVS and the atrio-ventricular septum (Figure 4.5, A & C) and to a small region of the right ventricular myocardium. Importantly expression of paternal *Grb10* and paternal *Ddc_exon1a* in heart does not overlap. *Ddc_exon1a* is far more extensively expressed throughout the myocardium than paternal *Grb10*. Additional analysis for paternal *Grb10* staining in specific regions of the right ventricle is shown in Figure 4.6, again indicating that it does not overlap specifically with Ddc expression in the myocardium. Dense clusters of paternal *Grb10* expressing cells are found in the right myocardial wall (Figure 4.6). These clusters of cells may represent components of the developing conduction system, however this was not explored as part of this thesis.

4.2.2 Transcript Analysis of *Grb10* in heart

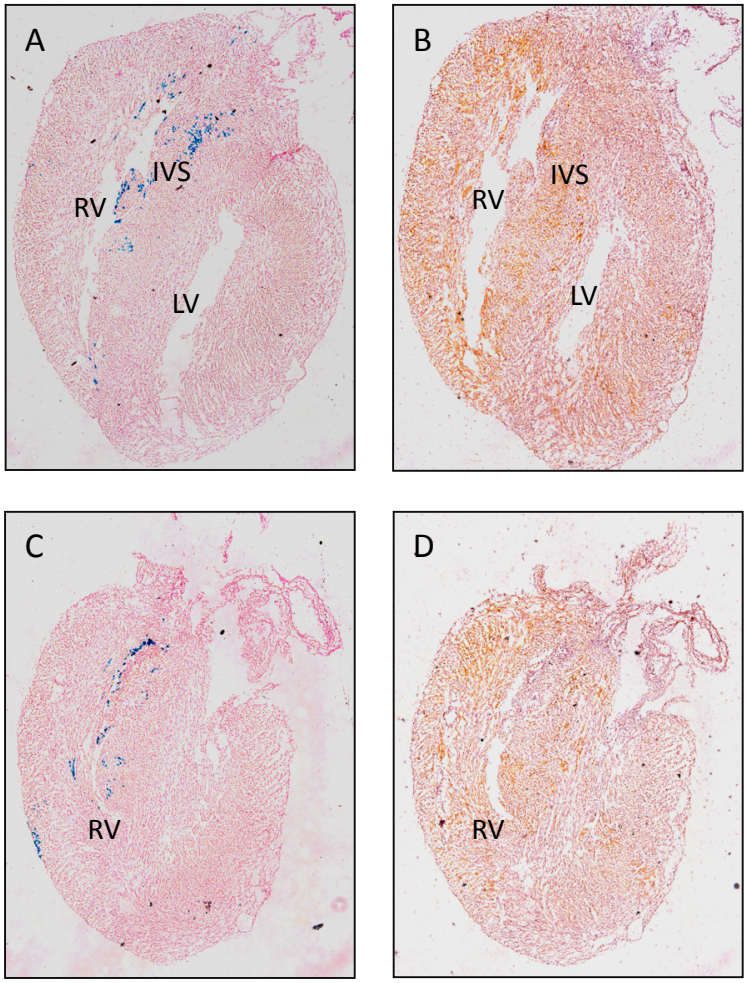


Figure 4.5 Co-expression analysis of paternal *Grb10* expression and Ddc expression in e18.5 heart. (A) and (B) are serial sections from the same heart sample in the sagittal plane showing the Right Ventricle (RV) and Left Ventricle (LV). (C) and (D) are taken from the same heart in the same plane and are also serial sections of each other, this time showing only the Right Ventricle, where Ddc staining appears most abundant. (A) and (C) are stained using X-gal and show paternal *Grb10* expression in blue. (B) and (D) are stained for Ddc using DAB (Brown). Expression of Ddc is far more extensive with paternal *Grb10* restricted mainly to the inter ventricular septum (IVS).

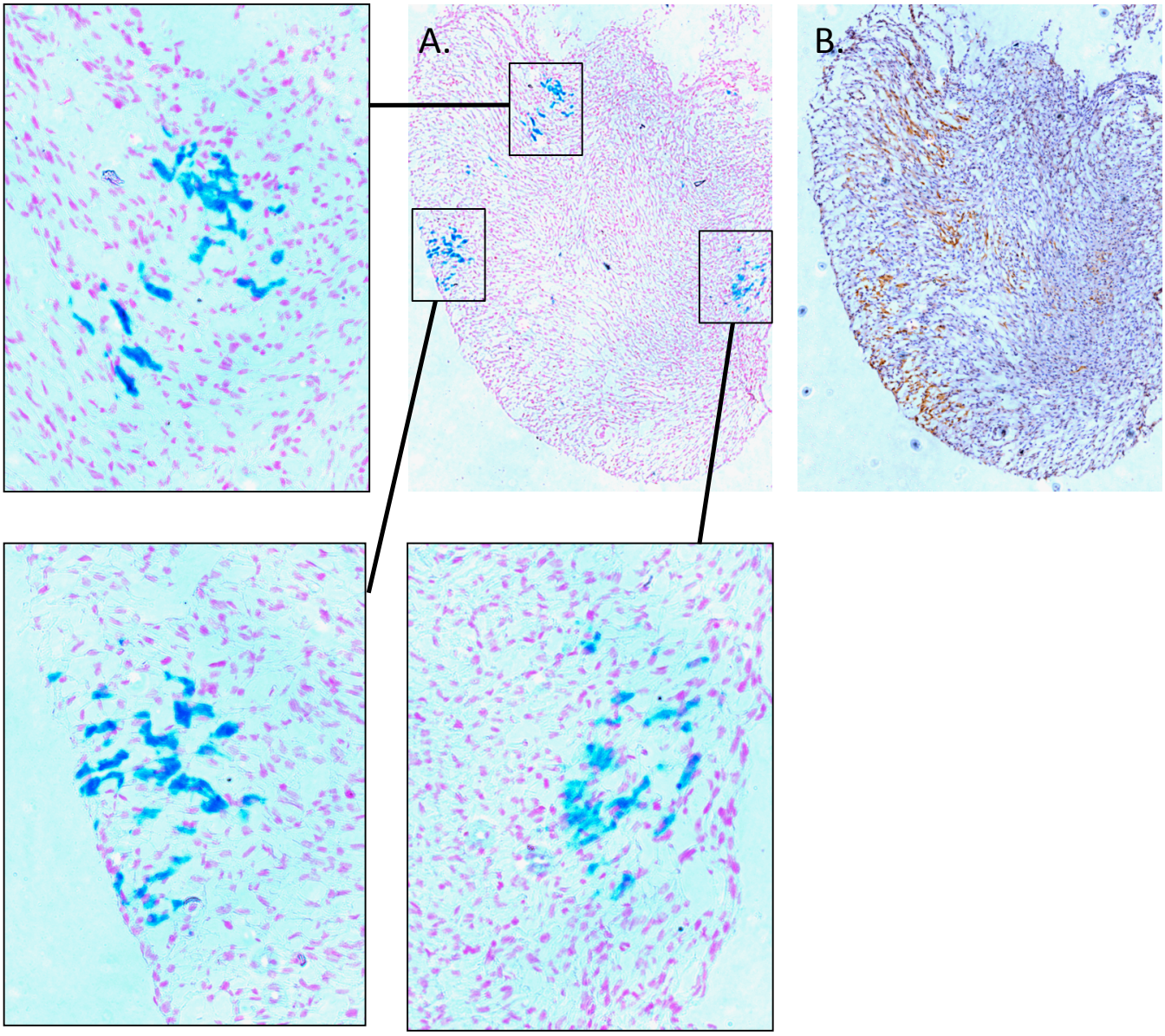


Figure 4.6 Detailed co-expression analysis of paternal *Grb10* and Ddc in e18.5 heart. Co-expression analysis of (A) paternal *Grb10* expression (Blue) and (B) Ddc expression (Brown) in e18.5 heart in the right ventricular myocardium in serial sections. Paternal *Grb10* expression is spatially restricted to distinct regions. Ddc is more widely expressed, but is not specifically co-localised with paternal Gr*b10.*

In order to assess which promoters are being used to initiate transcription of *Grb10* in the developing heart, a separate *Grb10* knockout mouse (*Grb10Δ2-4*) in which exons two to four of *Grb10* has been replace with a *lacZ* reported construct, was used to determine which exon is being used to initiate expression of both the paternal and maternal alleles. The *Grb10Δ2-4* knockout allele was generated using the same strategy as for the *Grb10KO* alleles previously described. Forward primers were designed to be complimentary to exons 1A and 1B, and reverse primers were designed to be complimentary either to exons 2 or the β-*geo* cassette (Figure 4.7, A). RNA was isolated from heart, brain and liver of mice where the *lacZ* reporter had been transmitted through the paternal line (samples from mice used in these experiments were supplied by Dr Andrew Ward, University of Bath, UK) (Figure 4.7, A). PCR assayed *Grb10* transcript expression from both parental alleles independently in each of the three tissues (Figure 4.7, B,C,D). These data revealed that maternal *Grb10* expression in heart, brain and liver initiates at exon 1B, confirming previously published results (Arnaud, et al., 2003). However analysis of paternal *Grb10* expression reveals a more complex picture of expression. Firstly, in agreement with published data, there is no paternal expression of *Grb10* observed in liver and there is paternal expression present in heart and brain (Figure 4.7, B). In contrast to results from Arnaud *et al.* paternal expression of *Grb10* in the brain appears to initiate from both exon 1A and exon 1B, and *de novo* analysis of paternal expression in heart reveals that the transcript initiates exclusively from exon 1A, with no expression from the 1B promoter as hypothesised (Figure 4.7, B & C).

4.2.3 Imprinting of *Ddc_exon1a* in brain

In order to explore the imprinted expression of *Ddc_exon1a* in the brain in more detail, allele-specific expression analysis was performed on sub-regions of the brain taken from six week old adult mice. Intercross F1 animals from C57BL/6 mothers and *Mus mus castaneus* fathers were used (BxC mice) along with the reciprocal cross using *Mus mus castaneus*

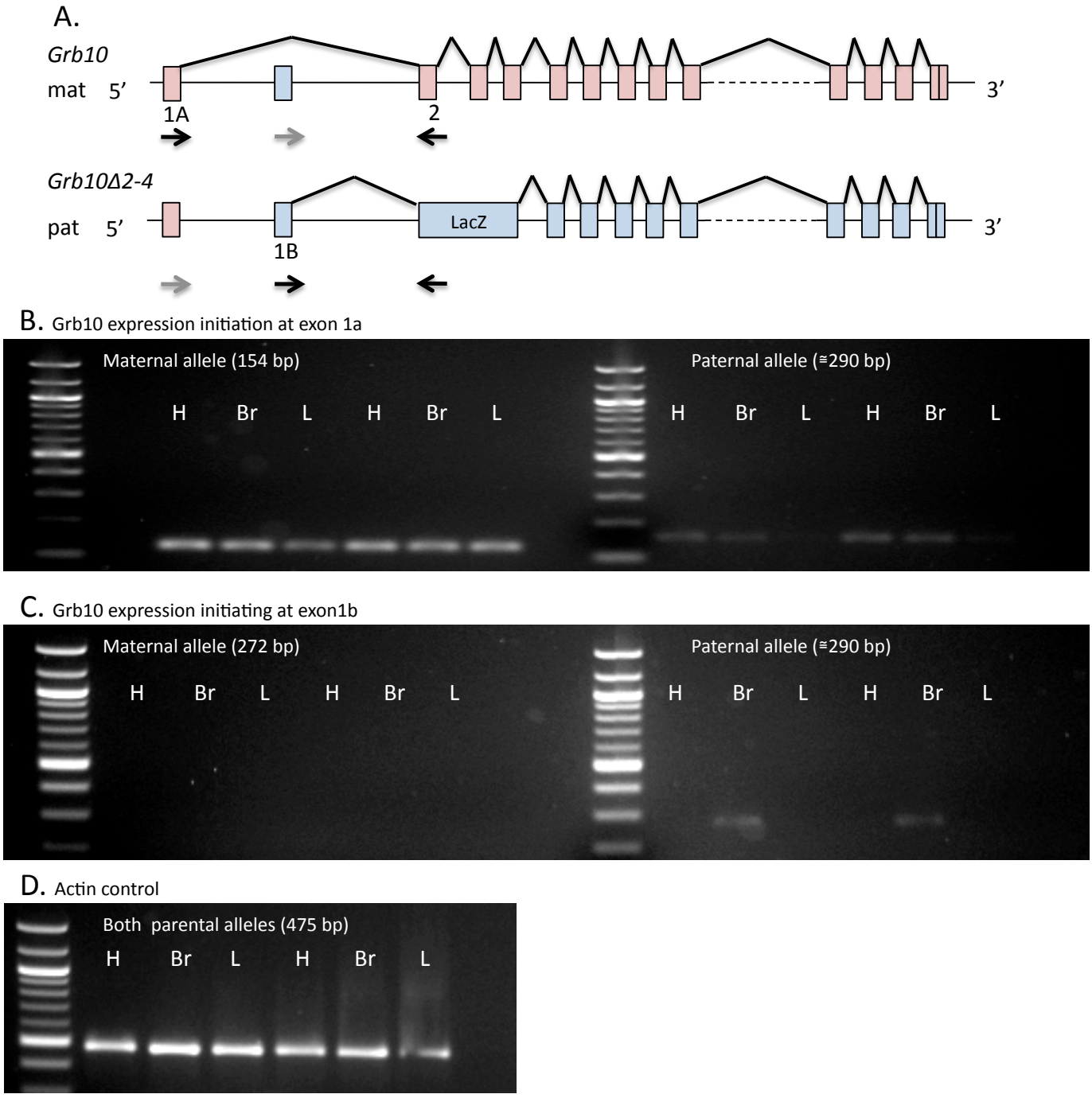


Figure 4.7 Transcript analysis of *Grb10.* (A) The parental origin of *Grb10* expression was differentiated using *LacZ* insertion transmitted through the paternal line. Promoter use was determined using RT-PCR with forward primers in exon 1a and exon 1b and reverse primers in exon 2 or the LacZ reporter (primers shown as black arrows) (B) RT-PCR for maternal and paternal *Grb10* expression initiating from exon 1a in Heart (H), Brain (Br) and Liver (L) reveals maternal expression in each tissue tested, patenal expression from exon 1a is seen in heart and brain only. (C) RT-PCR for maternal and paternal *Grb10* expression initiating from exon 1b reveals no maternal expression initiating from 1b in any tissues tested, paternal expression from exon 1b is only seen in brain tissue. (D) Control RT-PCR to show approximately equal loading of cDNA for each sample.

mothers and C57BL/6 fathers (CxB mice). RNA isolated from the heart of BxC and CxB animals was converted to cDNA and *Ddc_exon1a* transcripts were amplified using PCR and sequenced. Utilizing a G/A SNP between the two subspecies of mice located in exon 4 of the Ddc gene (chr11: 11,776,278; mm9) imprinted expression of *Ddc_exon1a* was assessed, with the presence of the SNP in the sequence trace indicative of bi-allelic expression. The results of these analyses (Figure 4.8) show that in whole brain, diencephalon (which gives rise to the pre-optic area), cerebellum and brain stem *Ddc_exon1a* expression is bi-allelic.

4.2.4 Summary of co-ordinate regulation

These analyses reveal that although *Ddc* and *Grb10* form part of a single imprinted cluster and the gDMR at the promoter of *Grb10* controls expression of *Ddc_exon1a*, expression and imprinting of both genes are not coupled and tissue specific imprinted expression of *Ddc_exon1a* in heart can be analyzed independently of *Grb10*. In addition *Grb10* promoter use appears less important in the switch from paternal to maternal *Grb10* expression than previously hypothesised. These data also show that *Ddc_exon1a* is bi-allelic throughout the brain, therefore epigenetic comparisons can now be made between heart and brain in order to search for the epigenetic basis for the change in from bi-allelic to imprinted expression.

RNA-seq analysis of *Ddc_exon1a* in brain using reciprocal crosses has since suggested that expression of *Ddc_exon1a* may be biased towards the maternal allele in the pre-optic area of the brain (Gregg, et al., 2010), however a subsequent validation of this experimental approach has called into question its validity and has suggested that this result is likely to be a false positive (DeVeale, et al., 2012).

4.3 Epigenetic analysis of the *Ddc/Grb10* locus.

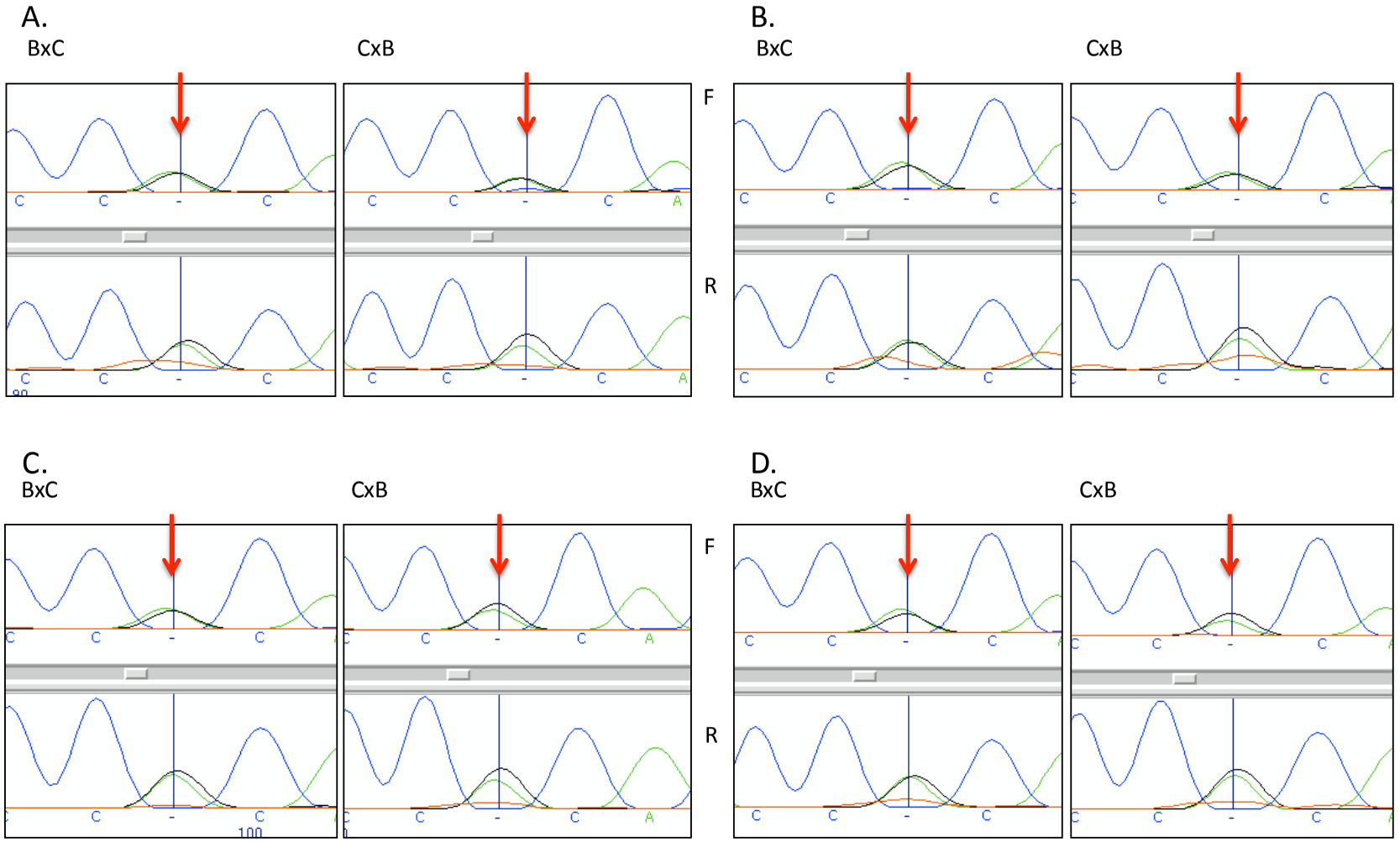


Figure 4.8 Bi-allelic expression of *Ddc_exon1a* in brain regions. Sequence traces of *Ddc_exon1a* transcripts show bi-allelic expression in dissected brain sub regions. RNA was extracted whole brain (A) and from Pre-optic area (B), cerebellum (C) and brain stem (D) of six week old BxC and CxB mice and *Ddc_exon1a* was sequenced using sanger sequencing over a regions with a known polymorphism between C57BL/6 and *Mus mus castaneus*. Each regions demonstrated expression from both parental copies of *Ddc_exon1a* which is indicated by red arrows.

In order to assess the epigenetic differences that can account for a switch from paternal expression and maternal silencing of *Ddc_exon1a* in heart, to bi-allelic expression observed in the brain, epigenetic control mechanisms, known to control genomic imprinting at other imprinted loci, were explored at the *Ddc/Grb10* locus. Tissue specific epigenetic differences between heart and brain were sought and a new model that could account for the silencing of the maternal allele in heart was proposed using these data.

4.3.1 CTCF binding across the *Ddc/Grb10* locus in heart and brain

As examined in detail in chapter 3, CTCF binds at the controlling DMR of many imprinted clusters including the *Grb10* DMR and binding at the *Grb10* DMR is likely to be on the unmethylated maternal allele in brain. CTCF has been shown to mediate interactions between enhancer and promoter elements at the chick *beta-globin* locus (Bell, et al., 1999) and control imprinting at the *Igf2/H19* imprinted locus where it binds on the unmethylated maternal allele (Bell, et al., 2000, Hark, et al., 2000, Kanduri, et al., 2000). Analysis of CTCF binding in newborn heart, brain and liver was performed in order to profile tissue-specific differences in CTCF binding which may account for silencing of *Ddc_exon1a* on the maternal allele in heart, but not in brain. Binding was assessed using chromatin immunoprecipitation (ChIP) on BxC and CxB tissue to ensure no strain specific differences exist, followed by quantitative PCR analysis. Using data from the genome-wide analysis of CTCF binding performed in chapter 3 and by mining ChIP-seq data from ES-cell line (Chen, et al., 2008) five regions shown to bind CTCF in ES cells and/or brain across the locus were examined for CTCF binding using custom designed Taqman™ assays. CTCF binding was assumed to be binary (ie. CTCF is bound or not bound) and this was assessed if binding was significantly different from the known negative control region DMDup, a region just upstream of the *H19* DMD known not to bind CTCF in multiple tissues including brain and ES cells. All qPCR results were normalized to genomic input DNA that was not treated with antibody

during the ChIP. Results from the qPCR experiments are shown in Figure 4.9. Regions bound by CTCF are denoted and summarised in Figure 4.10. These results demonstrate that CTCF is bound at four of the five regions examined, including the *Grb10-2* probe that examines the *Grb10* DMR. The only region that does not bind CTCF lies just upstream of the *Grb10* promoter. CTCF binding is invariant across the locus in heart and brain, and also in liver, an additional control tissue (Figure 4.9, A, B, C). Thus it is concluded that tissue-specific differences in CTCF binding cannot account for the silencing of the maternal allele in heart, although it cannot be ruled out that there is an additional binding regions in heart that is not present in brain, which has been missed in this analysis.

4.3.2 DNA methylation across the *Ddc/Grb10* locus

Genome-wide methylation data obtained from a study conducted by Fabien Guidez (Ellen Solomon Laboratory, King's College London) was used to scan the *Ddc/Grb10* locus for regions that demonstrate methylation differences between different tissues. MeDIP-seq utilized an antibody to 5-methylcytosine to precipitate regions of DNA rich in cytosine methylation followed by next generation sequencing as a mechanism to quantify the enrichment of DNA across the genome. MeDIP-seq was performed on the brain and heart of mice and comparative methylation was assessed using Useq (Nix, et al., 2008). Methylation differences were assessed with a "sliding window" of both 1000 bp and 500 bp independently (Figure 4.11). Two regions of tissue specific differential methylation (FDR < 1 %) were identified as regions of interest in the context of methylation differences that could play a role in controlling silencing of *Ddc_exon1a* specifically on the maternal allele in the developing and neonatal heart (Figure 4.11, A & B).

One region of differential methylation between heart and brain tissues is located at the *Grb10* DMR (Figure 4.11, A), this region has previously been analysed for tissue-specific methylation differences between brain and

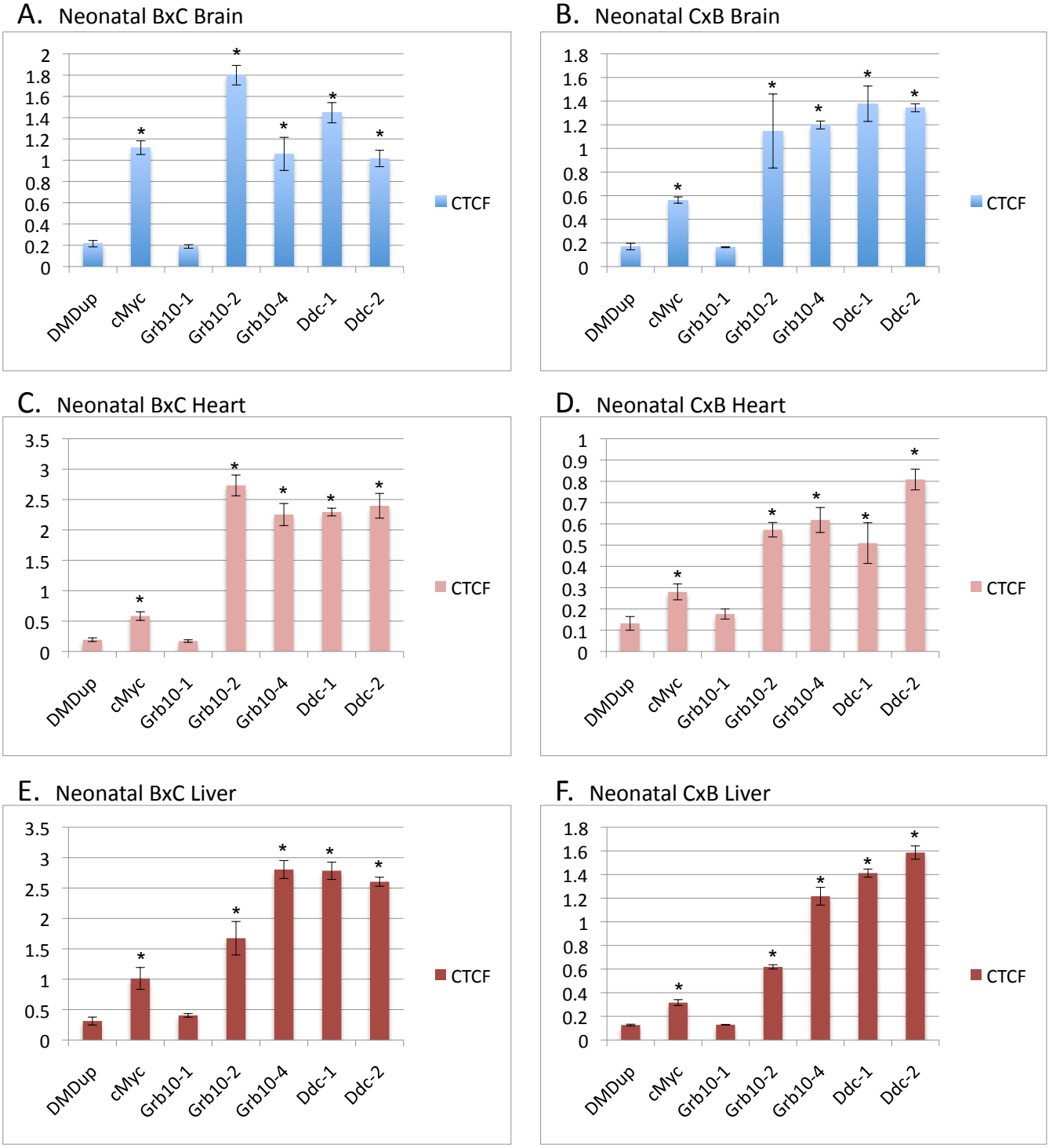
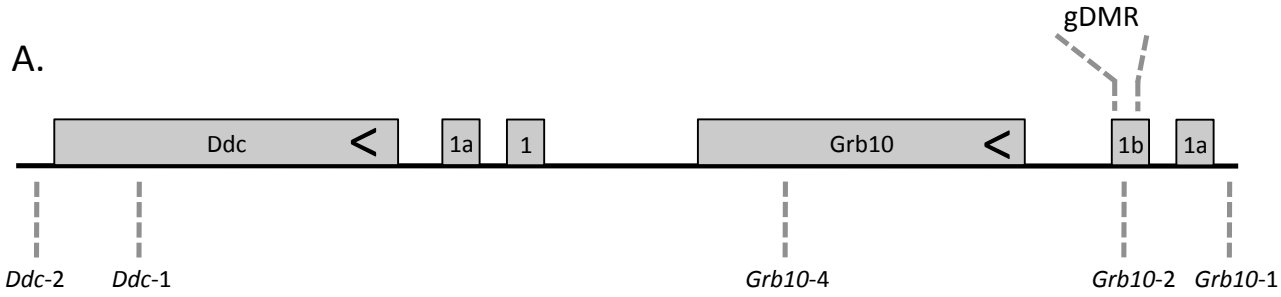


Figure 4.9 CTCF binding at the *Ddc/Grb10* locus. Quantitative PCR (qPCR) analysis showing average CTCF enrichment with standard deviations at five regions across the *Ddc/Grb10* locus in neonatal brain (A & B), heart (C & D) and liver tissue (E & F). Experiment was assessed in both BxC (A, C, E) and CxB animals (B, D, F) ensuring no strain bias was observed. Regions that differ from the negative control upstream *H19* gDMR (DMDup) with a $p < 0.05$ are denoted with an asterisk indicating that these regions show statistically significant levels of CTCF binding, above the level of the negative control. cMyc is a positive control for CTCF binding.

A.



B.

	Neonatal Brain		Neonatal Heart		Neonatal Liver	
	BxC	**CxB**	**BxC**	**CxB**	**BxC**	**CxB**
cMyc (Positive control)	Y	Y	Y	Y	Y	Y
Grb10-1	N	N	N	N	N	N
Grb10-2	Y	Y	Y	Y	Y	Y
Grb10-4	Y	Y	Y	Y	Y	Y
Ddc-1	Y	Y	Y	Y	Y	Y
Ddc-2	Y	Y	Y	Y	Y	Y

Figure 4.10 Summary of CTCF binding at the *Ddc/Grb10* locus. (A) Cartoon showing the *Grb10/Ddc* locus showing the location of the five regions assayed for CTCF binding names Grb10-1, Grb10-2, Grb10-3, Grb10-4, Ddc-1 and Ddc-2. Also shown is the location of the *Grb10* DMR. Genes are shown as grey boxes, with the alternate promoters shown individually direction of transcription is shown with black chevron. (B) Summarized results from the qPCR CTCF binding assay.

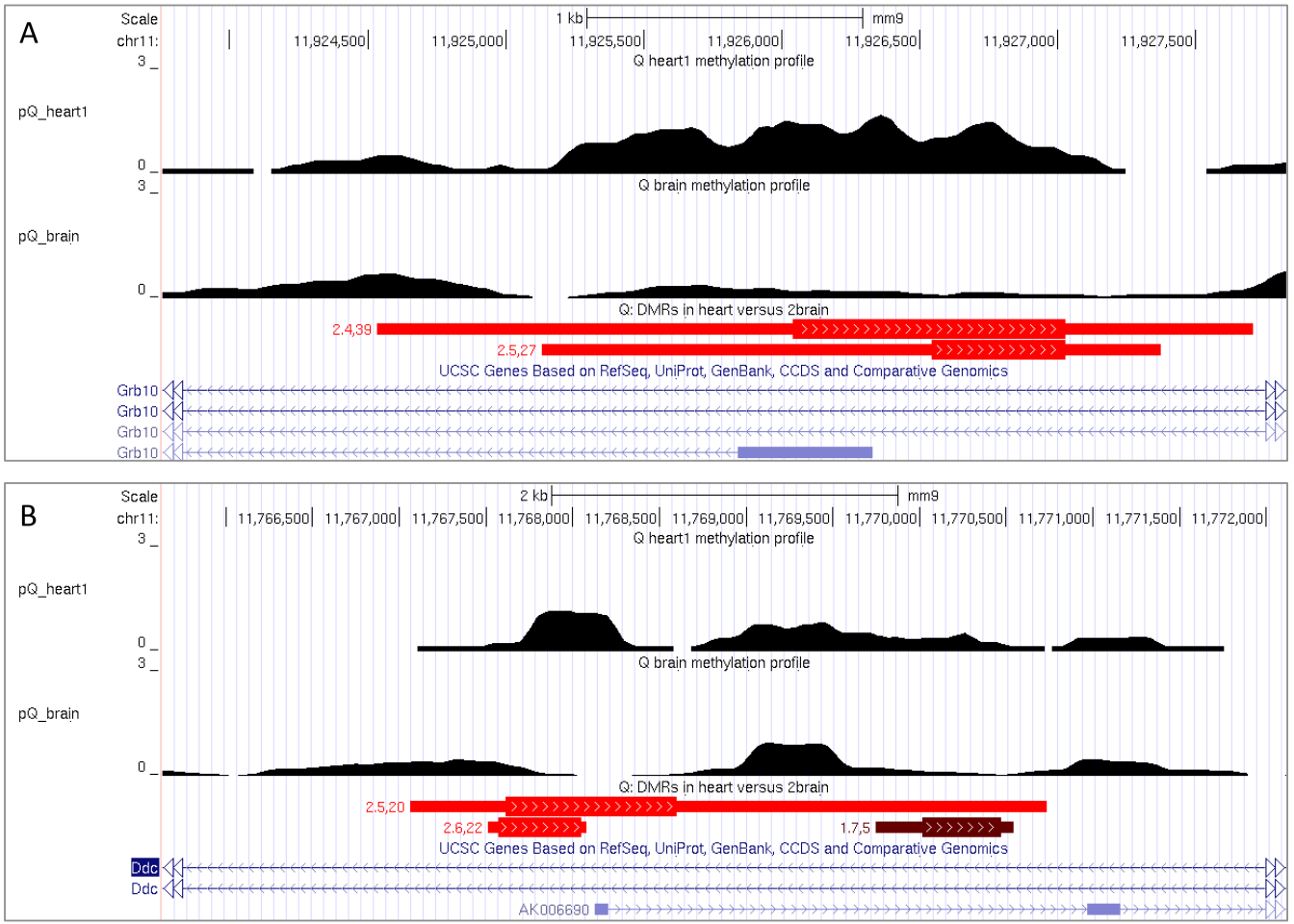


Figure 4.11 MeDIP-seq analysis of *Ddc/Grb10* locus. Density plots of methylation shown using UCSC genome browser at the *Grb10* DMR (A) and at the *AK0066690* promoter (B). The methylation density is shown in black with black representing DNA methylation. The upper most plot for each regions represents methylation levels in heart and the lower plot methylation levels in brain. Methylation differences between heart and brain were assessed statistically using Useq, using two different parameters. Regions demonstrating statistical differences in methylation are shown as red bars, with the 'large window' parameter showing a wider bar than the 'small window' parameter. Genes are indicated in blue.

liver (Arnaud, et al., 2003), MeDIP-seq analysis revealed a enrichment fold change difference of $\log_2(2.5)$ for the 500 bp and $\log_2(2.4)$ for the 1000 bp analysis, which equates to approximately a 25-30% increase in enrichment in heart when compared to brain. Previous analysis of this regions for methylation differences has been made by Arnaud *et al.* using mice with uniparental duplication of the part of chromosome 11 containing *Ddc* and *Grb10* to assess methylation at or near the *Grb10* DMR at two separate regions, which they term CGI2 and CGI3 (Figure 4.12) (Arnaud, et al., 2003). CGI2 constitutes the germline differential methylated region, verified by bisulphite analysis of the region by Arnaud *et al.* in oocytes where the *Grb10* DMR (CGI2) is methylated, and sperm where this region is hypomethylated. The *Grb10* DMR methylation remains invariant between kidney, liver and brain and is maternally methylated and paternally unmethylated in each tissue. CGI3 is a region rich in CpG dinucleotides, but which does not quite meet the criteria for a CpG island due to its size (Gardiner-Garden, et al., 1987). CGI3 lies just downstream of the *Grb10* exon 1B promoter and constitutes a methylation difference between liver and brain. Analysis of CGI3 reveals this region is hypermethylated in both sperm and oocytes, and analysis of UPD tissues show this pattern of methylation on both maternal and paternal inherited alleles is maintained in kidney and liver tissues. In brain, however, methylation is lost on the paternal allele; therefore CGI3 is differentially methylated between the two alleles, with the maternally inherited allele remaining hypermethylated (Arnaud, et al., 2003). In order to compare methylation at the *Grb10* gDMR (CGI2) and at CGI3 in brain, to that in heart, methylation was assessed at both loci using allele-specific bisulphite analysis. As epigenetic regulation of gene expression at imprinted loci works predominately *in cis*, methylation differences on the maternal allele were of particular interest, because expression of *Ddc_exon1a* on the maternal allele constitutes the transcriptional difference between brain and heart. Allele-specific examination of methylation revealed that methylation of CGI2 in heart (Figure 4.13, A) remains consistent to that seen in brain

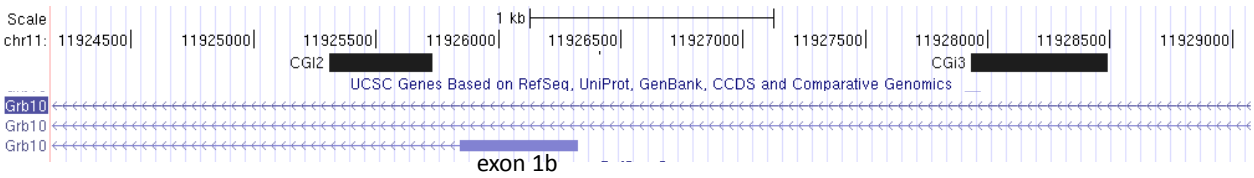


Figure 4.12 Genomic location of CGI2 and CGI3. Screenshot of UCSC genome browser showing the location of regions of CGI2 and CGI3 being assessed using bisulphite analysis (black bars), *Grb10 i*s shown in blue, exon 1b is denoted.

(Figure 4.13, B) and liver (Figure 4.14, A), with the gDMR demonstrating maternal methylation and paternal hypomethylation in all tissues, this result is unremarkable since the gDMR has previously been shown to be invariant between tissues at this and other imprinted loci (Arnaud, et al., 2003, Bartolomei, 2009), however examination of methylation at the *Grb10* DMR was still pertinent because dynamic methylation changes at gDMRs is not unprecedented (Oswald, et al., 2000). Examination of methylation at CGI3 reveals that methylation differences exist here between brain and heart. Bisulphite analysis revealed differential methylation between the parental alleles in brain (Figure 4.15, B) with the maternal allele remaining methylated and the paternal allele unmethylated in both BxC and CxB hybrids. In the heart, by contrast, methylation is maintained on the paternal allele (Figure 4.15, A), at a similar frequency (55 – 59%) to that observed in liver (Figure 4.15, C), thus the region retains the methylation pattern observed in oocytes and sperm is methylated on both parentally inherited alleles. Analysis of methylation at CGI3 reveals tissue-specific methylation differences at or near the *Grb10* DMR region previously shown to regulate expression of *Ddc_exon1a* in heart (Shiura, et al., 2009) and it is biologically plausible that an epigenetic difference between heart and brain here could control expression differences between the two tissues. However, these results show that methylation at CGI3 varies between heart and brain only on the paternal allele, and the expression difference between the two tissues is only observed on the maternal allele. There is currently no evidence for epigenetic marks controlling expression of imprinted genes *in trans*, therefore it is unlikely that methylation differences on the paternal allele are influencing the imprinted expression difference, which occurs on the maternal allele at this locus. There is evidence that CTCF mediates a trans-chromosomal interaction between the *Igf2/H19* DMR locus and the non-imprinted *Wsb1/Nf1* locus (Ling, et al., 2006) and a recent genome-wide analysis of CTCF mediated chromosomal interaction in ES cells reveals 336 CTCF mediated *trans-chromosomal* interacting loci, however this experiment

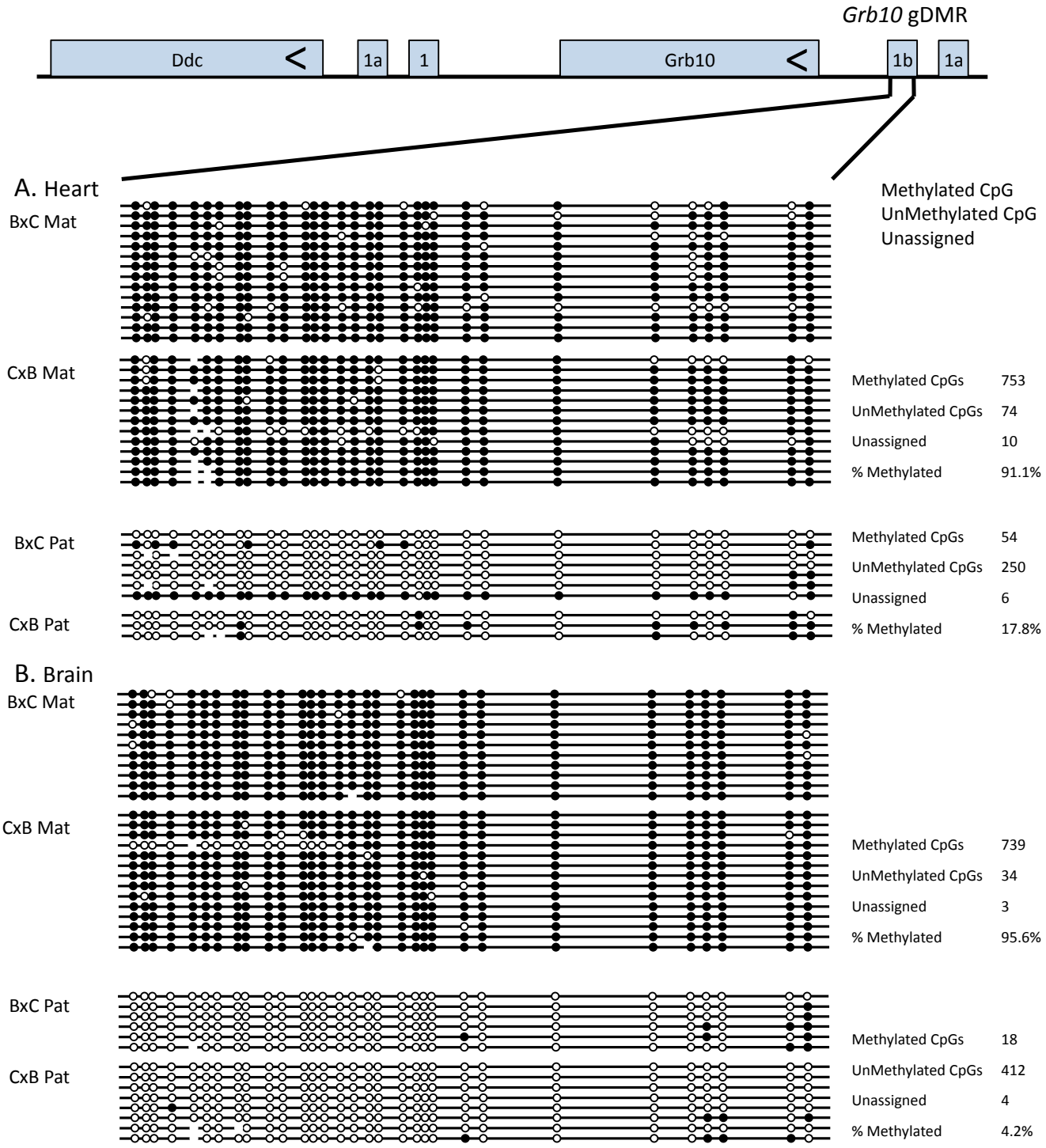


Figure 4.13 Bisulphite analysis of CGI2 in neonatal heart (A) and brain (B). Maternal (Mat) and paternal (Pat) alleles are shown separately and were distinguished using a known SNP between C57BL/6 and *Mus mus castaneus.* Each line represents an individual clone and each circle represents a CpG dinucleotide. Filled circles represent a methylated cytosine and blank circles represent unmethylated cytosines. Where methylation couldn't be assigned due to sequencing quality a gap is left. Results are summarized in the right most column.

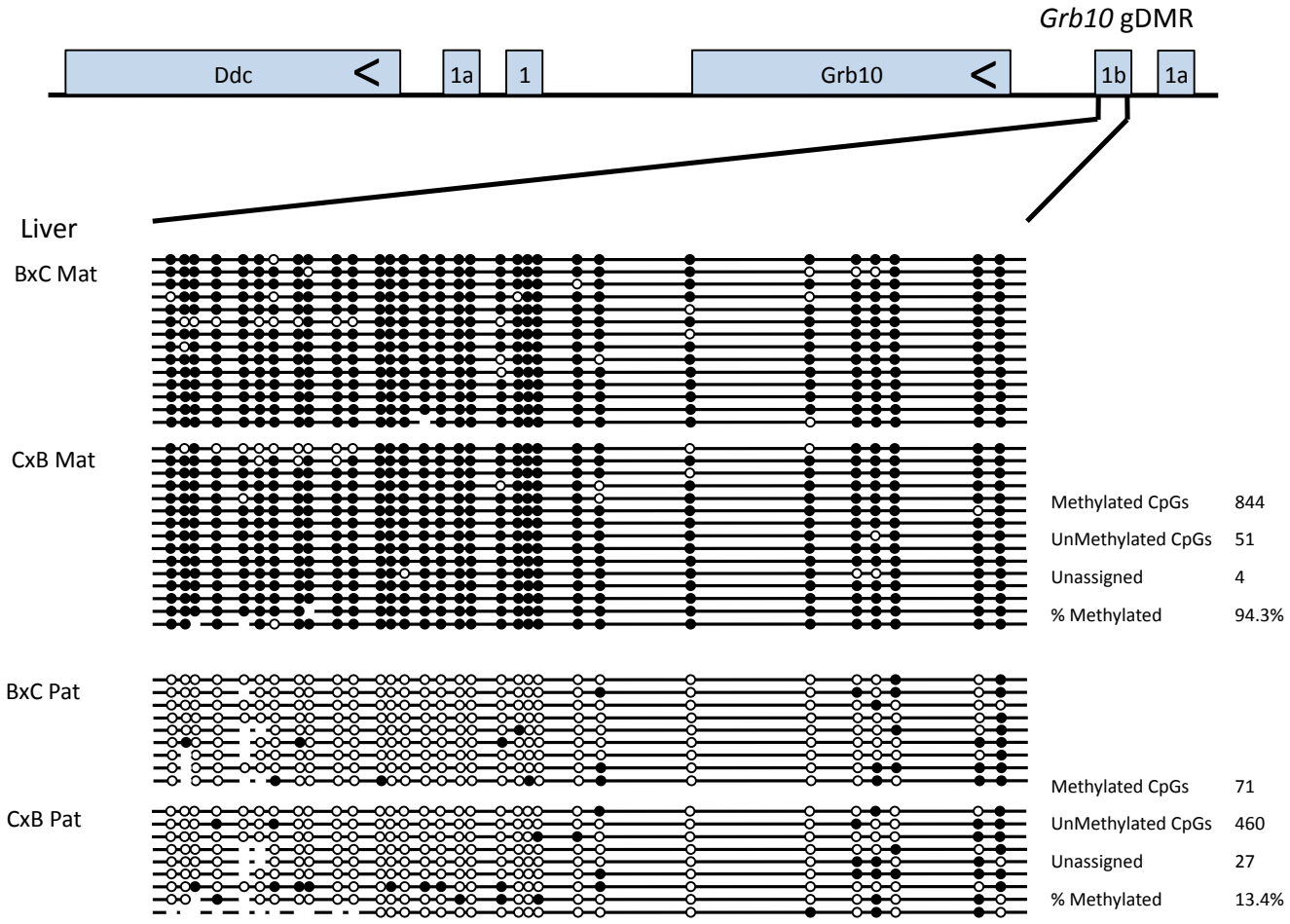


Figure 4.14. Bisulphite analysis of CGI2 in neonatal liver. Maternal (Mat) and paternal (Pat) alleles are shown separately and were distinguished using a known SNP between C57BL/6 and *Mus mus castaneus.* Each line represents an individual clone and each circle represents a CpG dinucleotide. Filled circles represent a methylated cytosine and blank circles represent unmethylated cytosines. Where methylation couldn't be assigned due to sequencing quality a gap is left. Results are summarized in the right most column.

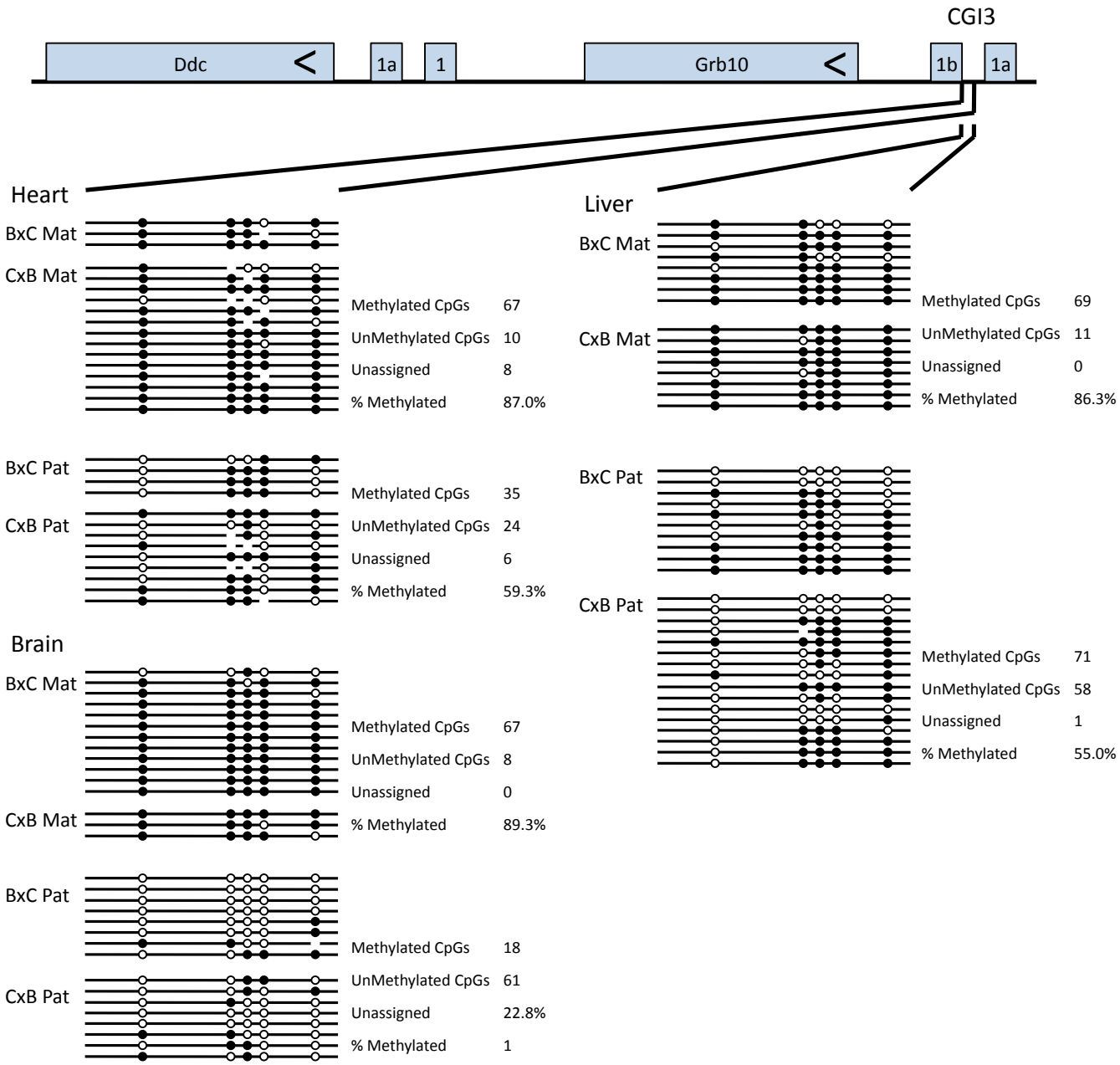


Figure 4.15 Bisulphite analysis of CGI3 in neonatal heart (A), brain (B) and liver (C). Maternal (Mat) and paternal (Pat) alleles are shown separately and were distinguished using a known SNP between C57BL/6 and *Mus mus castaneus.* Each line represents an individual clone and each circle represents a CpG dinucleotide. Filled circles represent a methylated cytosine and blank circles represent unmethylated cytosines. Where methylation couldn't be assigned due to sequencing quality a gap is left. Results are summarized in the right most column.

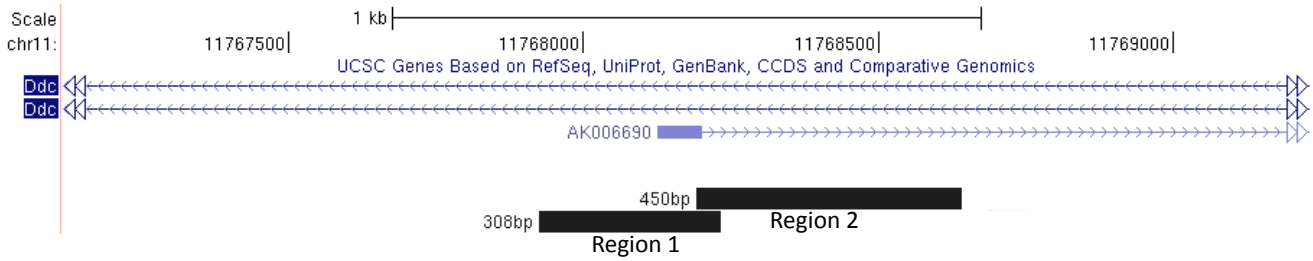


Figure 4.16 Bisulphite amplicons at the *AK0066690* promoter. Screenshot of UCSC genome browser showing the location regions of the *AK0066690* promoter being assessed using bisulphite analysis (black bars), *AK0066690* and *Ddc* genes are shown in blue. Two regions were assayed: Region 1 and Regions 2.

was unable to distinguish binding *in trans* between maternal and paternal copies of the same chromosome (Handoko, et al., 2011). A more likely explanation is that differential methylation between parental alleles combines with another epigenetic mark nearer the *Ddc_exon1a* promoter.

The second region identified from the genome-wide screen of methylation differences between heart and brain coincides with the annotated promoter of a transcript that is transcribed in an antisense direction to *Ddc_exon1a*, and initiates transcription between exons four and five of *Ddc*. Antisense transcripts are implicated in the control of imprinting a several loci and has been well studied as a mechanism of imprinting control at the *Igf2r/Air* locus (Pauler, et al., 2007, Pauler, et al., 2012, Sleutels, et al., 2002). Bisulphite analysis of methylation at the *AK0066690* promoter was performed in neonatal heart, brain and liver at two overlapping regions (region 1 and regions 2, Figure 4.17), parent of origin specific methylation could not be established at region 1 due to the lack of a SNP other than C/T between C57BL/6 and *Mus mus castaneus* at this locus (C/T SNPs cannot be used to determine parental origin when using bisulphite analysis), however maternal and paternal alleles are assigned in regions 2 and there is no indication of allele-specific methylation (Figure 4.17). Results revealed tissue-specific methylation difference with this regions appearing partially methylated in heart and liver and hypomethylated in brain. A model for imprinting control that satisfies all known observations was therefore proposed. It is hypothesised that control of imprinting in heart is exerted by the combinatorial effect of tissue specific methylation at the *AK0066690* promoter and differential methylation at the *Grb10* DMR (Figure 4.18). This model predicts that methylation differences at the *AK0066690* promoter observed between heart and brain result in transcription of *AK0066690* in heart but not in brain. *AK0066690* transcription then functions to silence *Ddc_exon1a* via transcriptional interference. In heart however, *AK0066690* is selectively silenced on the paternal allele though interaction with the *Grb10* DMR region, an

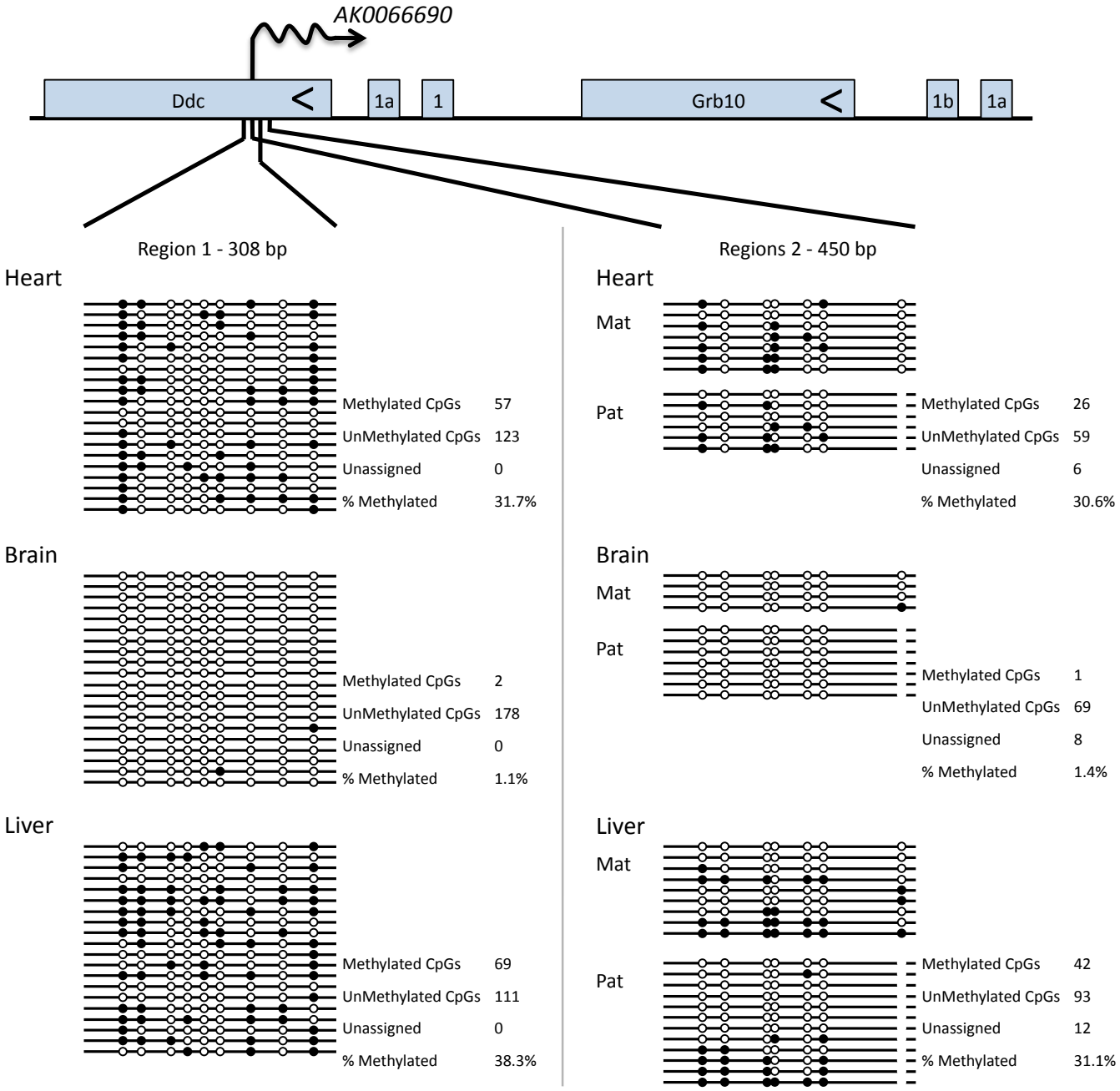


Figure 4.17 Bisulphite analysis of the *AK0066690* promoter in neonatal heart, brain and liver. Two regions we assayed (Region 1 and Region 2), parent of origin could only be assessed to clones in region 2. Each line represents an individual clone and each circle represents a CpG dinucleotide. Filled circles represent a methylated cytosine and blank circles represent unmethylated cytosines. Where methylation couldn't be assigned due to a polymorphism a gap is left. 30-40% methylation is observed in the heart and liver with no obvious difference between the parental alleles, the both regions are hypomethylated in brain. BxC and CxB tissues were used, however the results are combined in these analyses.

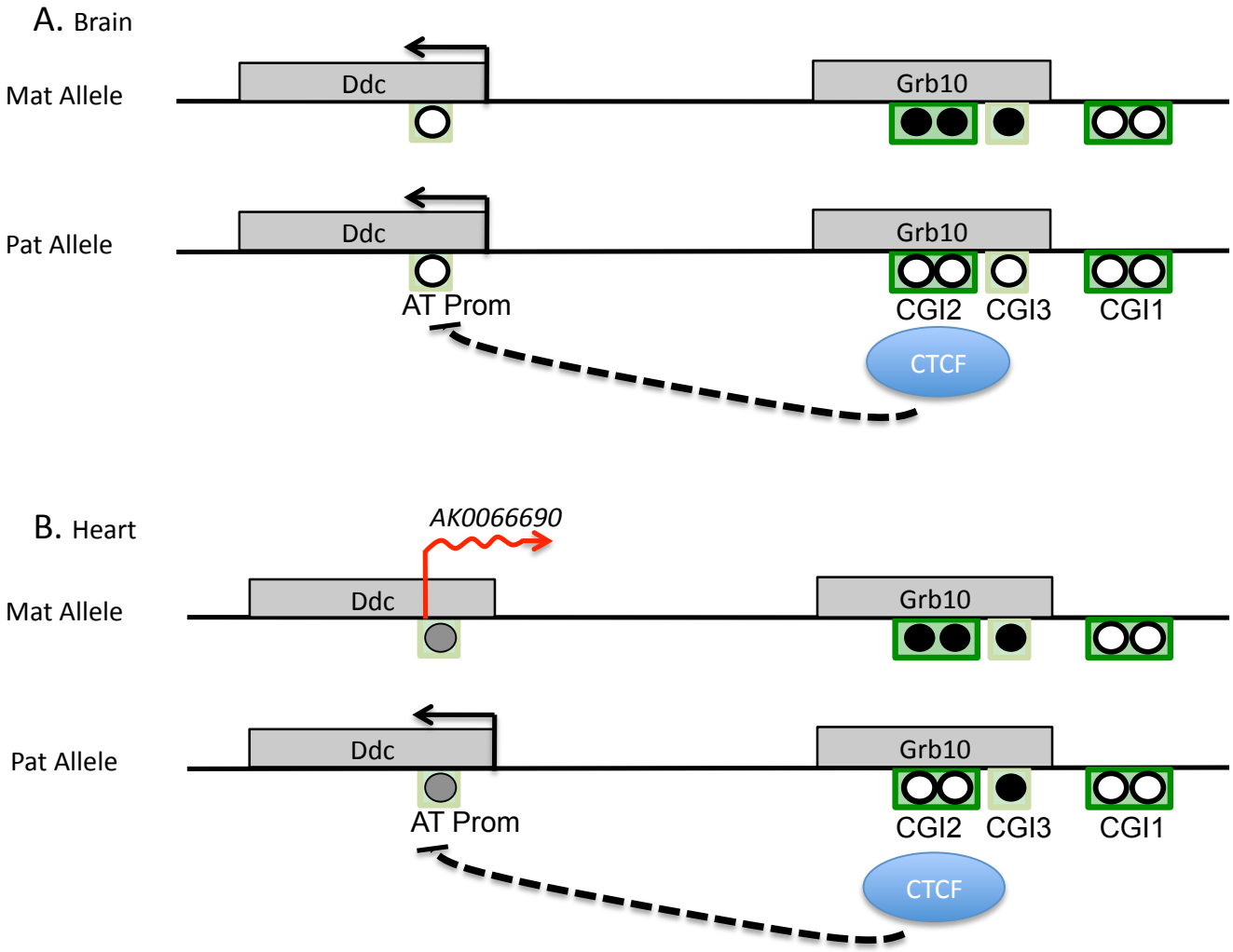


Figure 4.18 Model for the regulation of imprinting of *Ddc_exon1a.* A proposed model of *Ddc* imprinting control in (A) brain and (B) heart. Genes are show as grey boxes. Dark green boxes represent CpG islands and light green boxes represent regions rich in CpG but which do not classify as CpG islands due to their size. Circles represent methylation levels at these regions with filled circles indicating hypermethyaltion, white circles representing hypomethylation and grey circles representing partial methylation. Arrows represent transcription and the red arrows represents transcription of. CTCF binding at the DMR is shown in blue and the proposed mechanistic interaction is indicated by the dotted arrow.

interaction possibly mediated by CTCF binding at the DMR. This model is supported by analysis of data from ChIA-PET experiments looking at chromosomal interactions in ES cells, which reports CTCF mediated looping within the *Ddc/Grb10* locus (Handoko, et al., 2011), and by data from *Grb10* DMR knockout experiments, which shows that in the absence of the paternal copy of the *Grb10* DMR (and therefore its ability to selectively silence *AK006690* on the paternal allele), no paternal expression of *Ddc_exon1a* is observed (Shiura, et al., 2009).

4.3.3 *AK0066690* expression

In order to test the antisense transcription model of silencing, RT-PCR primers were designed to the *AK0066690* transcript and expression was tested in neonatal mouse heart, brain and liver tissues. RT-PCR revealed expression of the antisense transcript only in heart tissue, which fits with the hypothesised model of imprinting (Figure 4.19, A). The RT-PCR product was sequenced and aligned back to the genome to ensure that amplification was truly detecting the antisense transcript (Figure 4.19, C). Multiple RT-PCR primers were designed to test expression of *AK0066690,* however only one primer pair amplified a product that subsequently aligned back to the expected regions of the genome when sequenced, suggesting this transcript is not correctly annotated in genome build mm9. These analyses provide evidence supporting the model proposed, however, due to the lack of a SNP between C57BL/6 and *Mus mus castaneus* in the portion of *AK0066690* amplified by RT-PCR, allele-specific analysis of expression could not be performed. This leaves further scope for investigation in order to test the hypothesised model by fully characterising the antisense transcript using RACE, and assessing whether its expression is imprinted.

4.4 *Ddc_exon1a* expression in human

In order to explore whether *Ddc_exon1a* imprinting is conserved between mouse and humans, mono-allelic expression of *DDC_EXON1A* was assessed

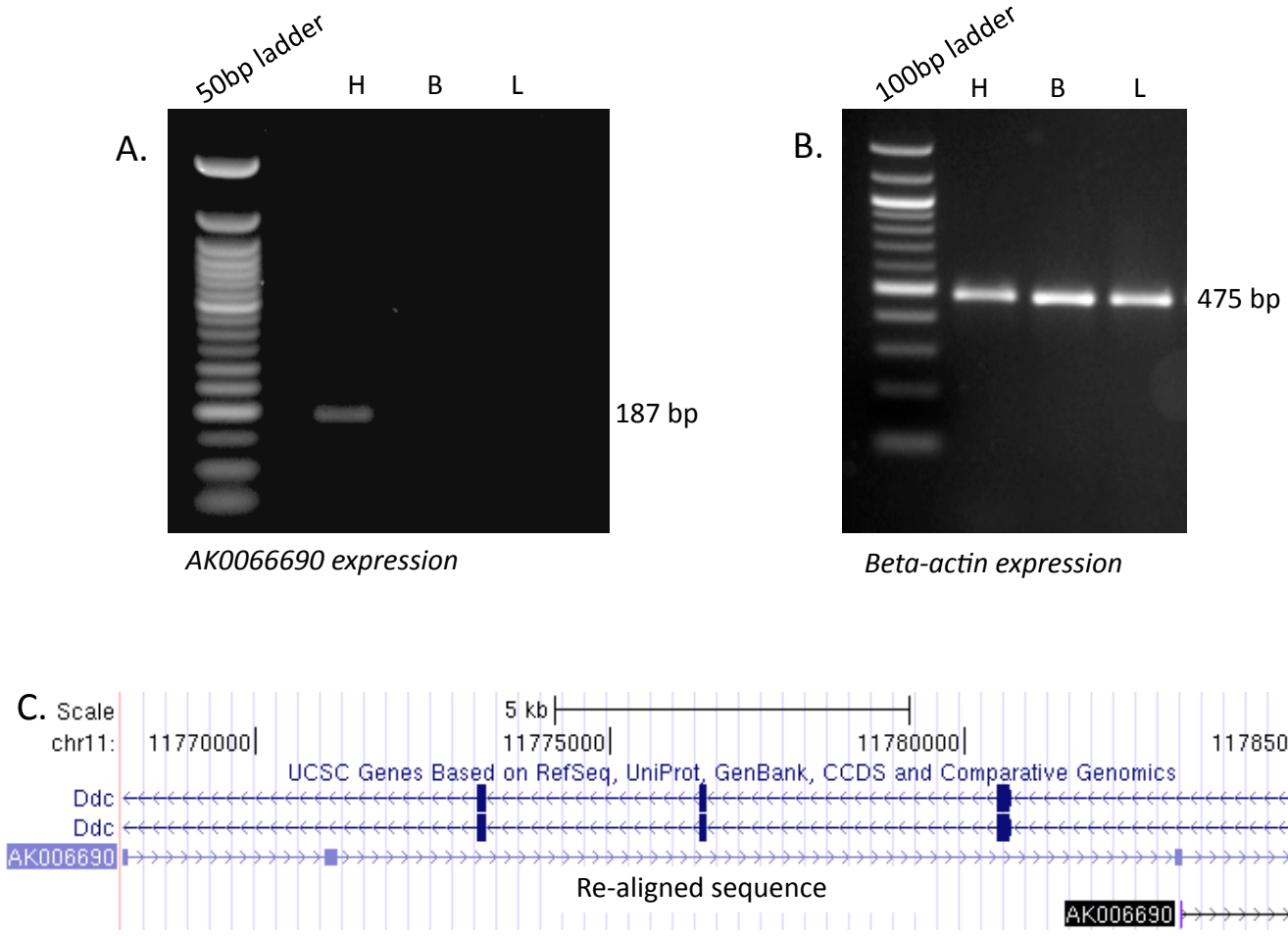


Figure 4.19 *AK0066690* expression analysis. (A) *AK0066690* expression in the neonatal heart, brain and liver shown by RT-PCR. (B) RT-PCR for beta-Actin showing cDNA is approximately equal for each tissue tested. (C) The *AK0066690* PCR product was excised and sequenced, before being realigned to the mouse genome, showing the product is indeed the antisense transcript.

in human fetal heart samples using RT-PCR followed by sequencing. 25 heart samples were obtained from the Human Development Biology Resource tissue bank (http://www.hdbr.org , University of Newcastle and UCL Institute of Child Health), ranging from four weeks gestation to 13 weeks gestation. Using DNA extracted from the samples all exons of *DDC_EXON1A* were sequenced to look for single nucleotide polymorphisms (SNPs) between the maternally and paternally inherited alleles. Three samples were identified that had a SNP in the coding portion of *DDC_EXON1A*, RNA from these samples was then sequenced to see if both, or only one parental copy of the gene was being transcribed. Two of the samples showed clear mono-allelic expression at the RNA level (Figure 4.20, A & B). The other sample (sample 17070) demonstrated a bias of expression coming from one allele, as evidenced on the sequence trace (Figure 4.20, C). Although this bias was not quantitative, it may indicate incomplete silencing of one allele, or polymorphic imprinting. Polymorphic imprinting in human is not without precedent and been shown at the *IGF2R* locus (Oudejans, 2001). In order to assess the parental origin of *DDC_EXON1A* expression in human, human fetal samples with maternal DNA samples were collected from the Clinical Genetics Department (collaboration with Dr Melita Irving, Guy's and St. Thomas' NHS trust), however no informative samples have so far been identified.

4.5 Discussion

The mechanism of *Ddc_exon1a* imprinting is complex and the analysis performed in this chapter serves to increase the knowledge of the epigenetic marks at the locus and allows models of imprinting to be proposed. Analysis of *Ddc_exon1a* imprinting in brain, and the lack of co-expression of *Ddc_exon1a* and paternal *Grb10* in fetal heart suggest that the imprinting of *Ddc* and *Grb10* is not directly coupled, and means that they can be considered independently, despite the fact *Ddc_exon1a* expression is under control of the *Grb10* DMR (Shiura, et al., 2009). Results in this chapter unequivocally show *Ddc_exon1a* expression is bi-allelic in

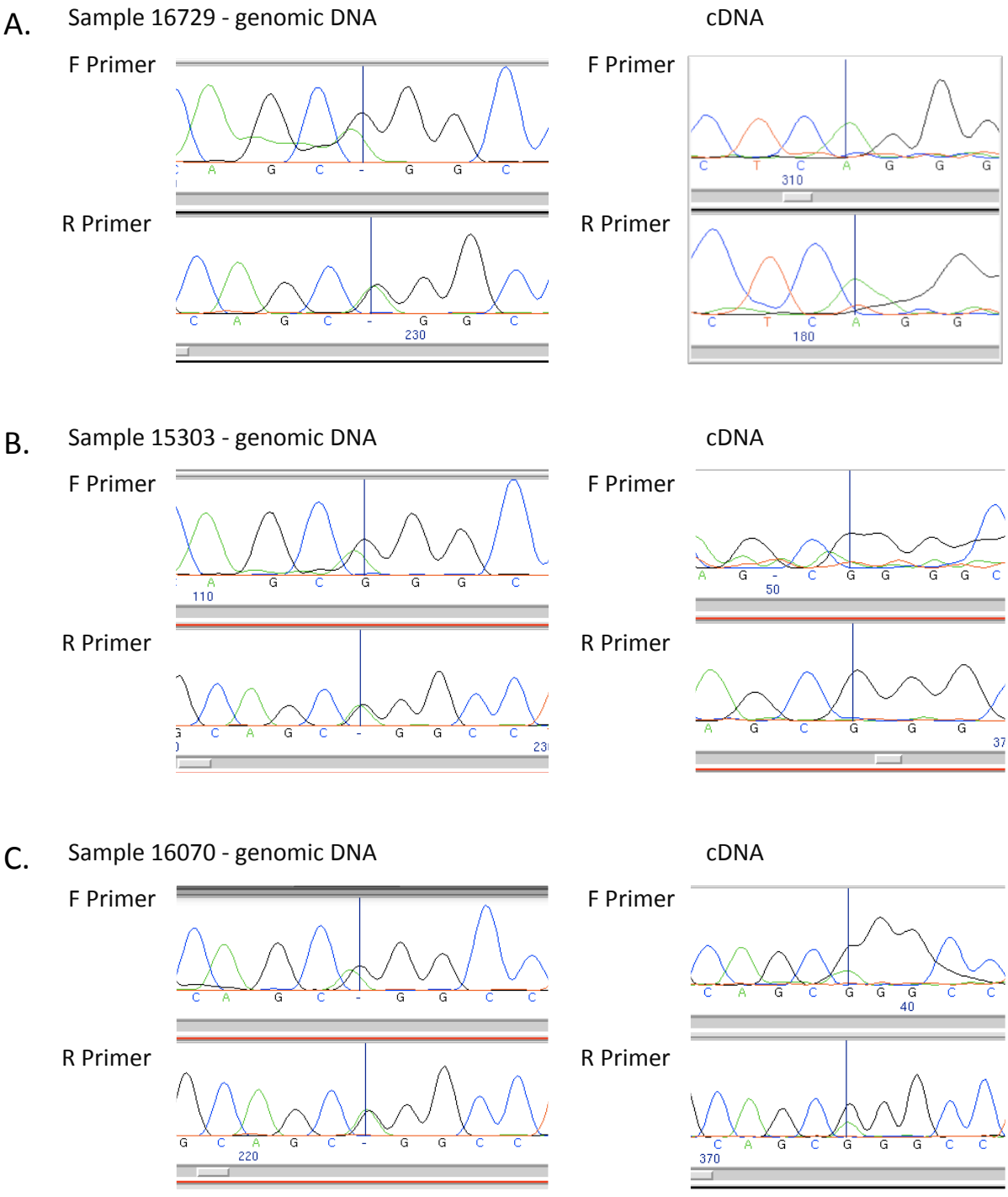


Figure 4.20 Monoallelic expression of *DDC_EXON1A* in human. Three human heart samples obtained from termination of pregnancy (A, B & C) were assayed for monoalleleic expression of *DDC_EXON1A.* Sequence traces of genomic DNA (left) show an A/G SNP in exon 14 in each sample, this corresponds with rs11575542 categorized in dbSNP 129. Sequencing of cDNA (right) reveals mono-allelic expression in two samples (A & B) and an allelic bias in one (C).

brain, whilst being imprinted in heart, this means that direct epigenetic comparison can be made between brain and heart tissue elucidate the mechanism of *Ddc_exon1a* silencing on the maternal allele.

Analysis of CTCF binding in neonatal heart and brain shows that binding is invariant between these tissues across the *Ddc/Grb10* locus at regions identified from the ES cell and whole brain CTCF ChIP-seq screen. However, these data do not rule out binding in heart at another regions and a further limitation ChIP qPCR experiments is that they do not distinguish between imprinted or bi-allelic binding. ChIP-seq analysis performed in chapter 3 shows a signal for CTCF binding on the unmethylated allele at the *Grb10* DMR in 3 week brain and a genome-wide analysis of CTCF binding reveals an affinity for binding unmethylated sections of DNA. This, when coupled with published evidence of the low affinity binding of CTCF to methylated DNA (Mukhopadhyay, 2004), suggests CTCF is most likely bound to the paternal allele in neonatal heart and brain, as the maternal allele is hypermethylated in both tissues.

CGI2, CGI3 and the *AK0066690* promoter region were identified as regions of interest when looking for differential methylation between heart and brain observed using MeDIP-seq comparison of methylation in heart and brain. CGI3 was considered of particular interest due to its implication in the tissue specific imprinting of *Grb10* (Arnaud, et al., 2003), and *AK0066690* promoter because antisense transcription is implication in imprinting control at other loci (Sleutels, et al., 2002). Allele-specific bisulphite analysis of these regions show differences in methylation at CGI3 between heart and brain, but differences are limited to the paternal allele, with the maternal allele remaining methylated in both tissues. There is no precedent for imprinting being controlled by selective silencing *in trans,* thus a mechanism where the *Grb10* DMR interacts with another epigenetic mark to control imprinting of *Ddc_exon1a* is hypothesised. Methylation differences in heart and brain are observed at the *AK0066690* promoter, with the regions appearing hypomethylated in brain, and partially

methylated in heart. The hypothesised model proposes that methylation mediates expression of *AK0066690* in heart, but not in brain and this is supported by RT-PCR expression analysis of the *AK0066690* transcript. There are a number of caveats to these results, firstly *AK0066690* is expressed in heart, where its promoter is partially methylated; the promoter is unmethylated in brain where expression is silenced. DNA Methylation usually correlates with transcriptional silencing; however there is evidence at low CpG content promoters showing methylation does not preclude gene activity (Weber, et al., 2007). Secondly, *AK0066690* expression analysis proved particularly difficult, with only one RT-PCR primer combination resulting in successful amplification of the transcript. This is almost certainly due to *AK0066690* being incorrectly annotated in Ref-seq, a feature not uncommon for non-coding antisense transcripts, therefore this transcript will need to be properly defined, and any parent-of-origin specific expression assessed, in order to further test the hypothesised model. Ultimately, therefore, the work in this chapter forms the basis for future investigation into the mechanism of *Ddc_exon1a* imprinting.

Chapter 5

Ddc_exon1a in the developing mouse heart

5.1 Introduction

Ddc_exon1a expression is present in the developing and neonatal heart (Menheniott, et al., 2008). The function of the Ddc protein in heart is yet to be elucidated, however published data suggests that expression is observed in the trabecular cardiomyocytes which form part of the developing myocardium and which functions to provide capacity for the developing myocardium to increase oxygenation prior to the formation of a coronary circulation. Later in development the trabecular cardiomyocytes increase the mass and thickness of the myocardium, enabling it support the increase in haemodynamic pressure associated with a fully functioning circulation. This chapter examines expression of *Ddc* in the developing heart and considers the role of Ddc in the heart, by utilizing knockout mice lacking *Ddc* expression in the heart.

5.2 Expression analysis of Ddc in developing heart

To assay regions of Ddc expression in the developing heart e15.5 day hearts were analysed using immunostaining, and visualised using fluorescence microscopy. At e15.5 trabeculation is almost complete, with trabecular cardiomyocytes forming a relatively compact layer lining the ventricles.

5.2.1 Ddc expression is localised to the right ventricle

Initially immunostaining was performed to assay only Ddc expression in the heart, sections were stained using rb-anti-Ddc (Abcam) and counterstained with DAPI nuclear stain. Serial sections from paraffin embedded whole embryos were taken through the heart in the saggital plane, from the left lateral side to the right lateral side. Figure 5.1 shows three sections.

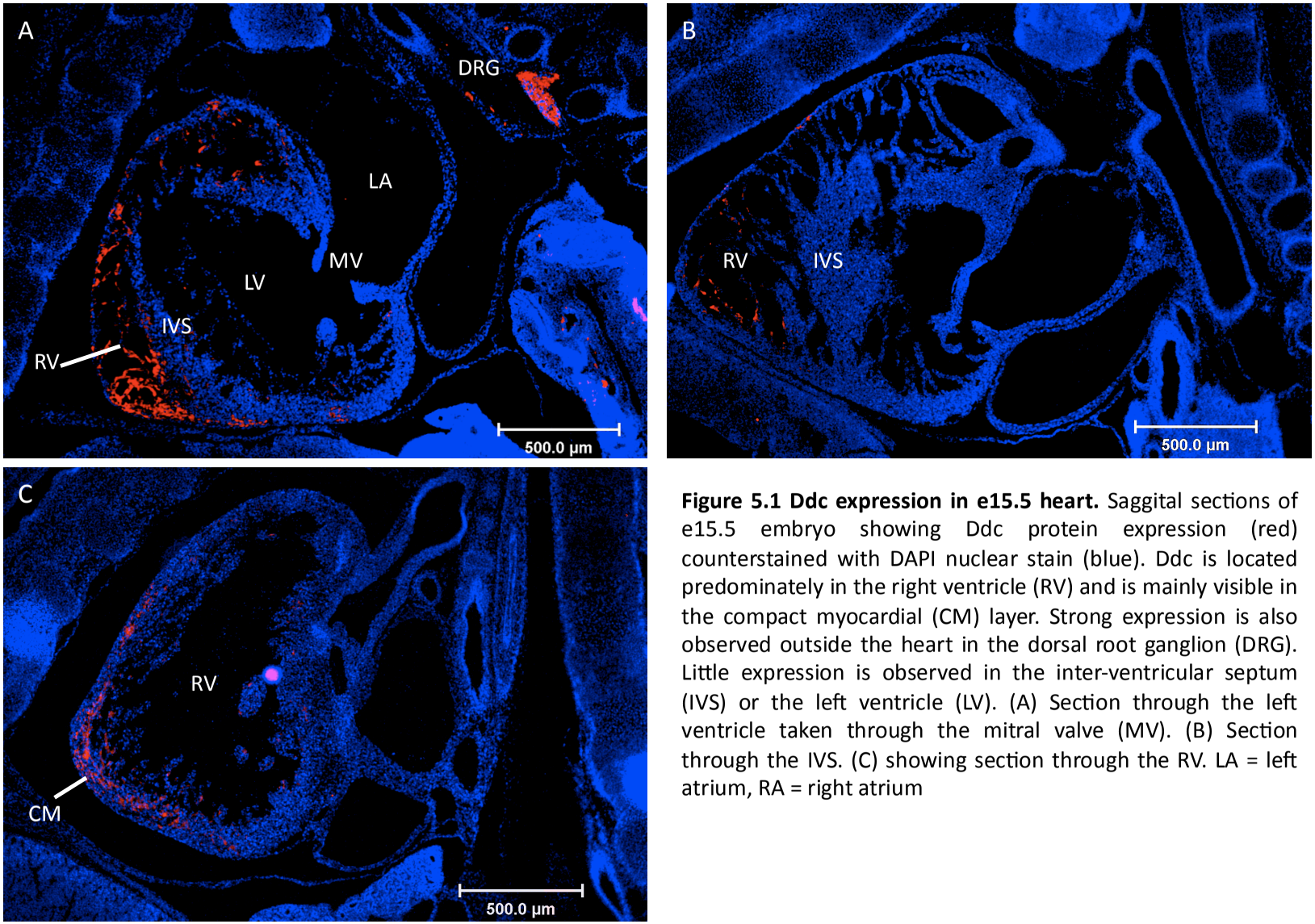


Figure 5.1 Ddc expression in e15.5 heart. Saggital sections of e15.5 embryo showing Ddc protein expression (red) counterstained with DAPI nuclear stain (blue). Ddc is located predominately in the right ventricle (RV) and is mainly visible in the compact myocardial (CM) layer. Strong expression is also observed outside the heart in the dorsal root ganglion (DRG). Little expression is observed in the inter-ventricular septum (IVS) or the left ventricle (LV). (A) Section through the left ventricle taken through the mitral valve (MV). (B) Section through the IVS. (C) showing section through the RV. LA = left atrium, RA = right atrium

Observation through the middle of the left ventricle (LV) shows Ddc staining (Red) does not appear as abundantly expressed in the left ventricular trabecular layer, but instead is concentrated in the anterior portion of the right ventricle (RV) (Figure 5.1, A), additional staining can be seen in the dorsal root ganglion. The saggital view taken through the interventricular septum (IVS), again shows that Ddc expression is located predominately in the right ventricle (Figure 5.1, B). Expression of Ddc in the right ventricle view (Image 5.1, C) shows that Ddc is localized to the compact layer of the right ventricular myocardium, with little expression in the trabecular 'spongy' layer. As this result ran contrary to previously published research, further double immunostaining assays using markers for cardiac muscle (MF-20) and for trabecular cardiomyocytes (ANF) were performed. A four chamber section of the e15.5 heart was made in the coronal plane and stained for Ddc (Red) and MF-20 (Green) using DAPI as a counterstain, again Ddc expression is localised predominately to the apical portion of the right ventricle at the base of the IVS. In addition, weaker expression of Ddc is seen in the compact later of the left and right ventricular myocardium and the IVS, no expression is observed in the trabecular layer or in either the left or right atria (Figure 5.2). Images shown are for one biological replicate, however similar results were observed in two further biological replicates.

5.2.2 Ddc is not restricted to trabecular cardiomyocytes

In order to assess the expression of Ddc in the context of trabecular cardiomyocytes, double staining was performed with antibodies to Ddc and ANF, the latter is a marker of trabecular cardiomyocytes and atria, images shows that there is minimal co-expression of Ddc and ANF (Figure 5.3) and confirmed using a higher resolution (x20) view of the left ventricle, which shows the compact layer of trabecular cardiomyocytes stained for ANF, with negligible co-expression of Ddc in these cells (Figure 5.4). When observed from the coronal section through the right ventricle expression of Ddc is localized predominately to the right side of the IVS and the compact

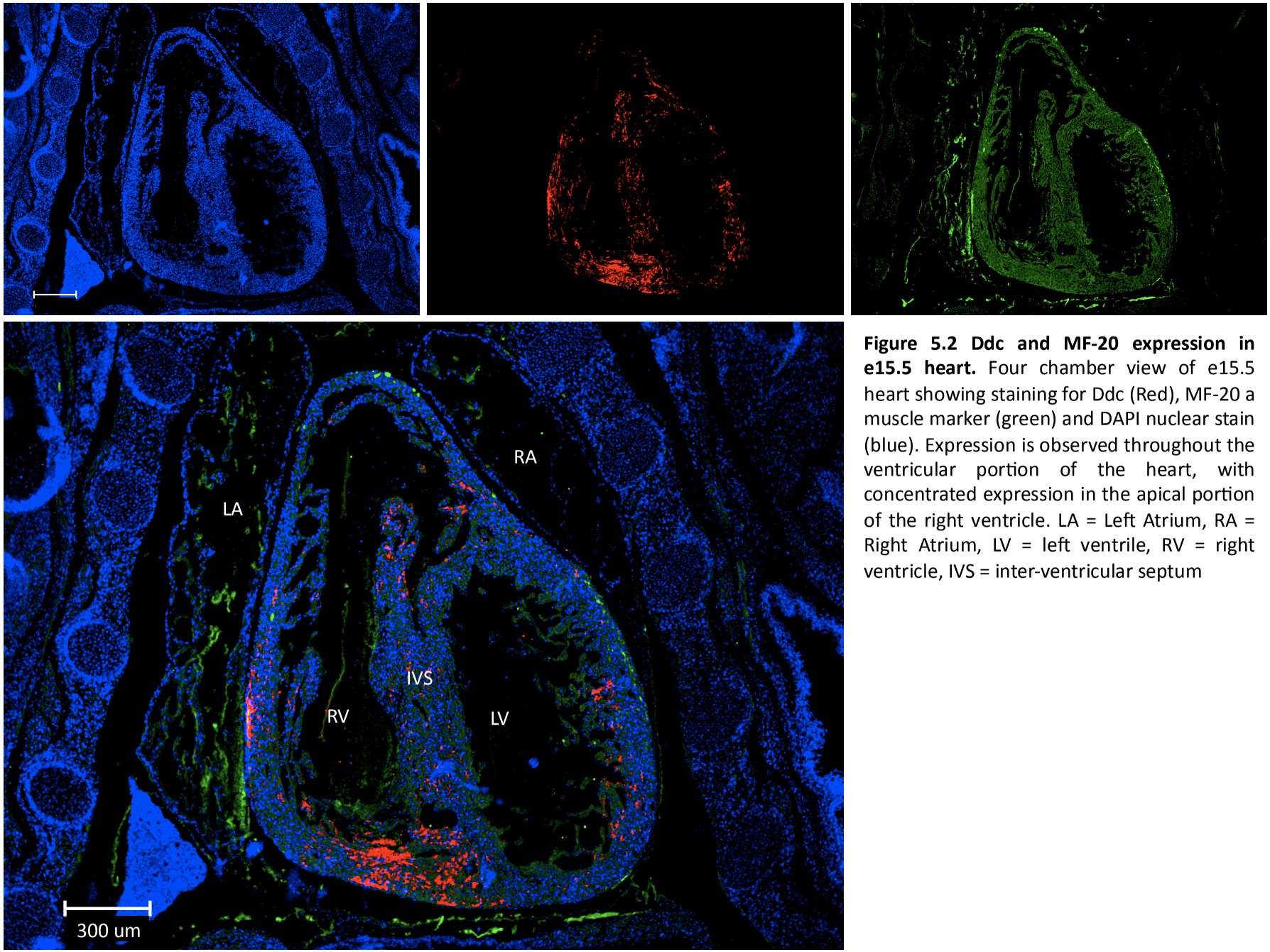


Figure 5.2 Ddc and MF-20 expression in e15.5 heart. Four chamber view of e15.5 heart showing staining for Ddc (Red), MF-20 a muscle marker (green) and DAPI nuclear stain (blue). Expression is observed throughout the ventricular portion of the heart, with concentrated expression in the apical portion of the right ventricle. LA = Left Atrium, RA = Right Atrium, LV = left ventrile, RV = right ventricle, IVS = inter-ventricular septum

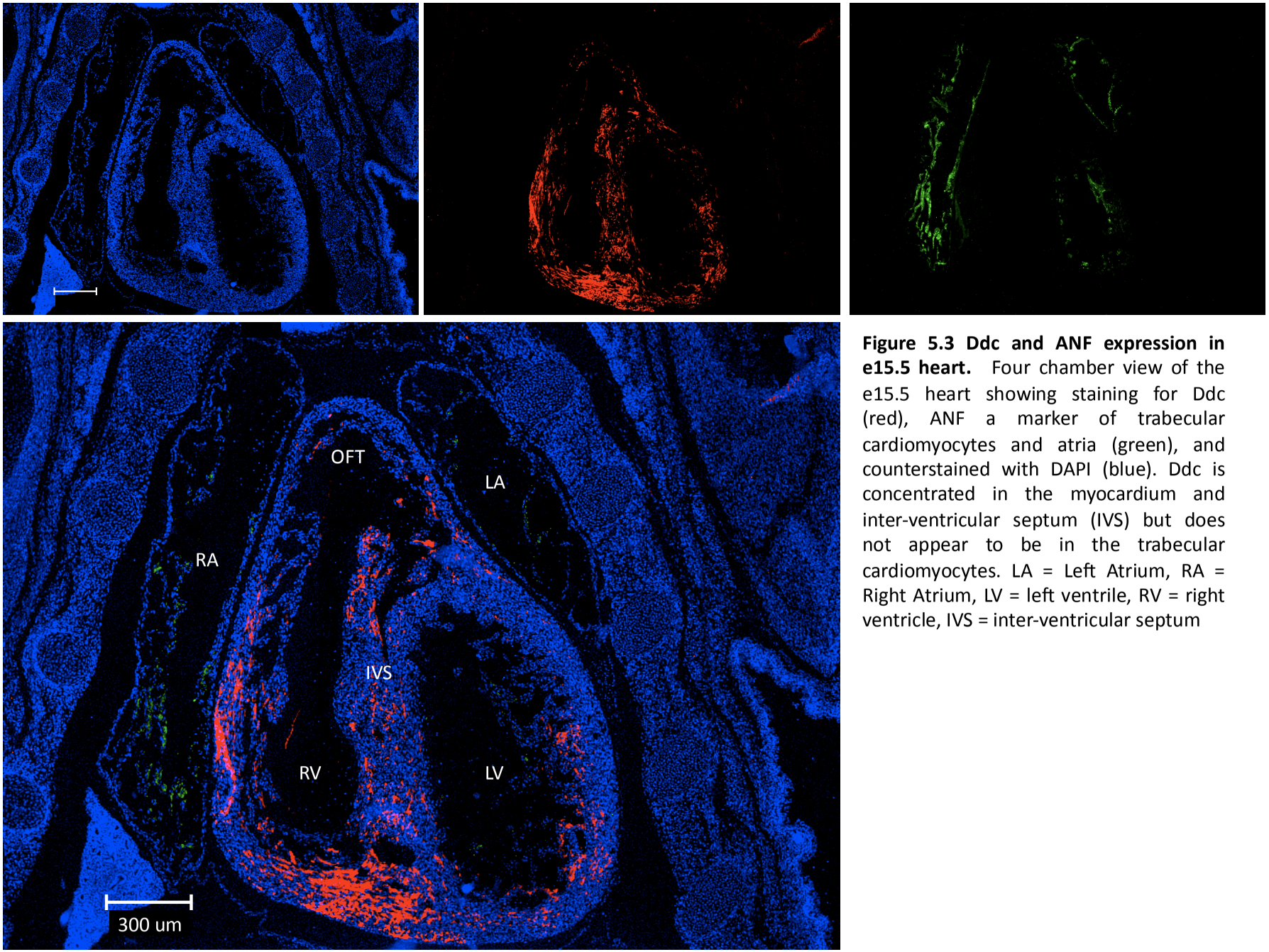


Figure 5.3 Ddc and ANF expression in e15.5 heart. Four chamber view of the e15.5 heart showing staining for Ddc (red), ANF a marker of trabecular cardiomyocytes and atria (green), and counterstained with DAPI (blue). Ddc is concentrated in the myocardium and inter-ventricular septum (IVS) but does not appear to be in the trabecular cardiomyocytes. LA = Left Atrium, RA = Right Atrium, LV = left ventrile, RV = right ventricle, IVS = inter-ventricular septum

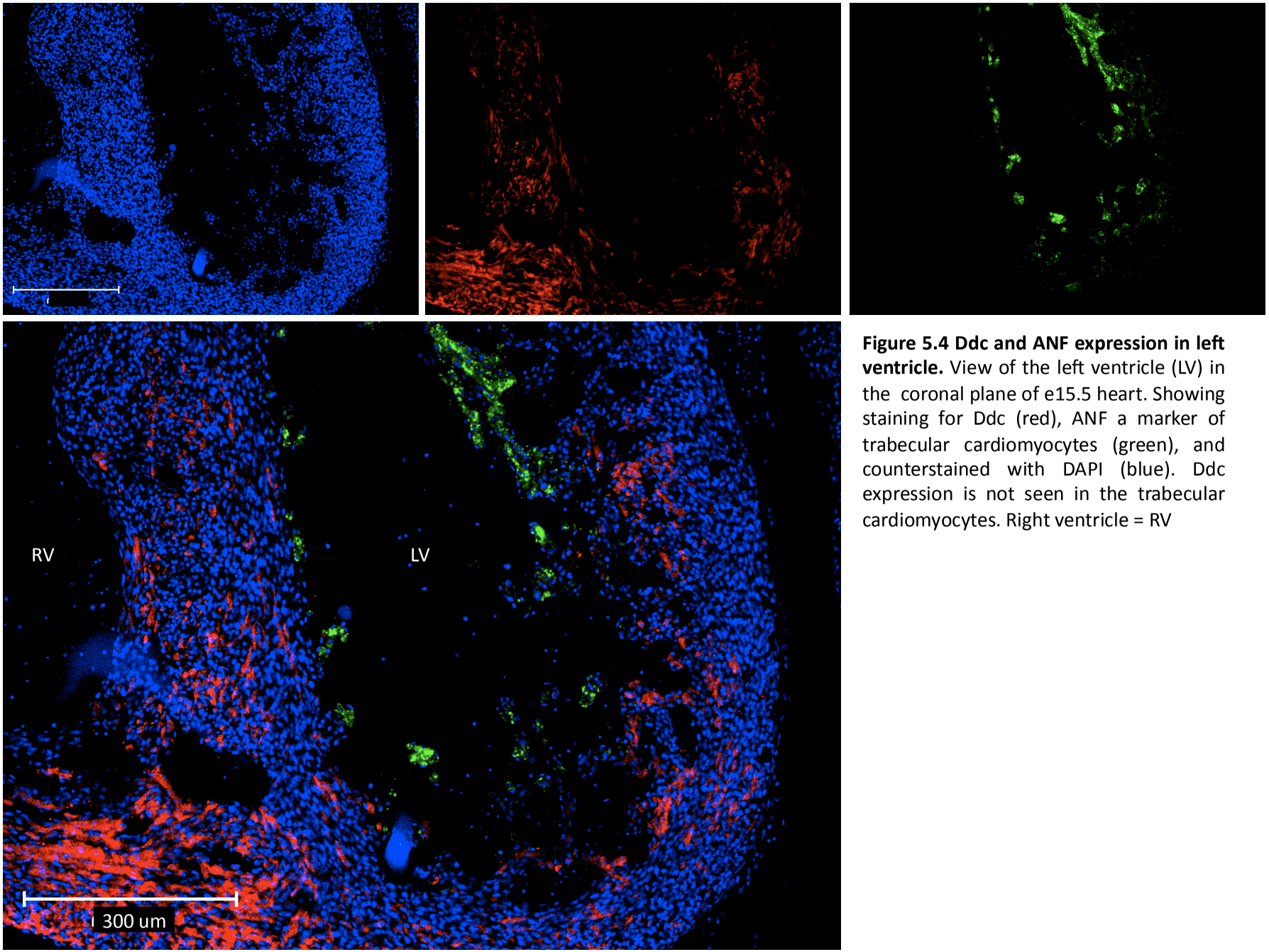


Figure 5.4 Ddc and ANF expression in left ventricle. View of the left ventricle (LV) in the coronal plane of e15.5 heart. Showing staining for Ddc (red), ANF a marker of trabecular cardiomyocytes (green), and counterstained with DAPI (blue). Ddc expression is not seen in the trabecular cardiomyocytes. Right ventricle = RV

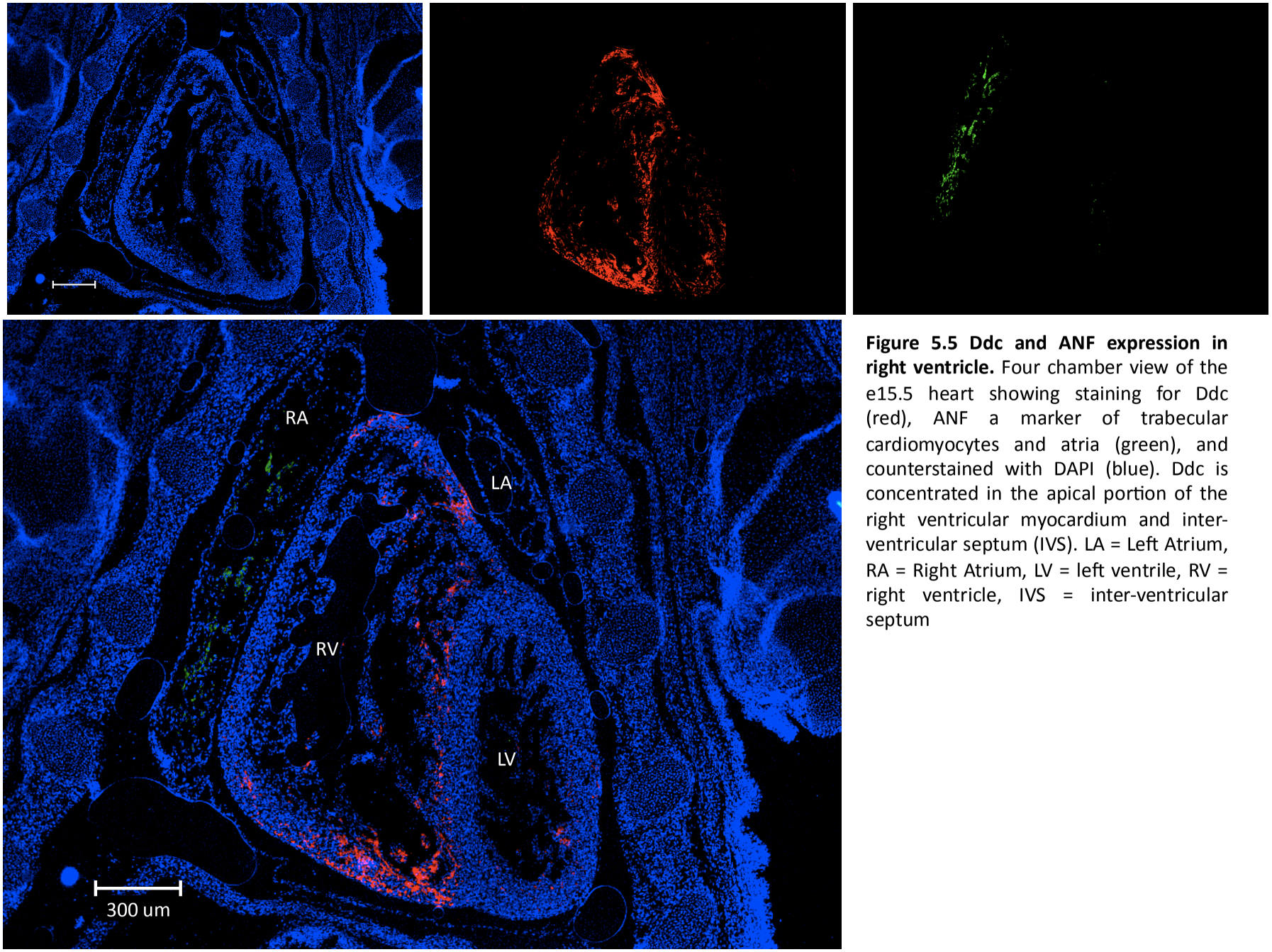


Figure 5.5 Ddc and ANF expression in right ventricle. Four chamber view of the e15.5 heart showing staining for Ddc (red), ANF a marker of trabecular cardiomyocytes and atria (green), and counterstained with DAPI (blue). Ddc is concentrated in the apical portion of the right ventricular myocardium and inter-ventricular septum (IVS). LA = Left Atrium, RA = Right Atrium, LV = left ventrile, RV = right ventricle, IVS = inter-ventricular septum

layer of the apical portion of the right ventricle (Figure 5.5). Although expression is more predominate in the compact layer there is some expression of Ddc at the interface between the compact and trabecular layer, with limnited expression in the trabecular layer.

5.3 Characterisation of *Ddc* knockout animals

Mice harbouring a germline deletion of *Ddc* have been generated by Lexicon Genetics™ ($Ddc^{Gt(neo)420Lex}$) and were explored with genomic and functional studies. Homozygous null mice for *Ddc* ($Ddc^{-/-}$) die late in prenatal development (Eppig, et al., 2012), possibly due to a global lack of neurotransmitter synthesis in the brain and the central nervous system.

However, due to the tissue-specific genomic imprinting of *Ddc* in heart, mice heterozygous for deletion of *Ddc* where the deletion has been passed through the male line, can be used as a effective 'conditional' knockout of *Ddc* in heart.

5.3.1 *Ddc* knockout breeding

A programme was designed to breed animals with a specific paternally or maternally inherited deletion of *Ddc,* this was then carried out by the UC Davis Mouse Biology Program (MBP), CA., USA. In order to establish mice with a paternal germline deletion of *Ddc* ($Ddc^{+/-}$), Female *Ddc* wildtype mouse were bred with males heterozygous for deletion of *Ddc* (Ddc^{HET}) (Figure 5.6). Assuming expected Mendelian ratios, 50 % of the F1 progeny carry a *Ddc* knockout allele and these are paternal *Ddc* knockouts ($Ddc^{+/-}$). The remaining 50 % of the litter do not carry a knockout allele and are used as wildtype controls (Figure 5.6). For control purposes the reverse cross using females heterozygous for deletion of *Ddc* (Ddc^{HET}) and wildtype males was performed giving F1 progeny with either a maternal deletion of *Ddc* ($Ddc^{-/+}$) or wildtype genotype. Finally F1 progeny of heterozygous males and females were bred to produce full maternal and paternal *Ddc* knockout ($Ddc^{-/-}$) animals. All embryos at e15.5 and p0 dissected and either fixed and embedded, or snap frozen in liquid nitrogen.

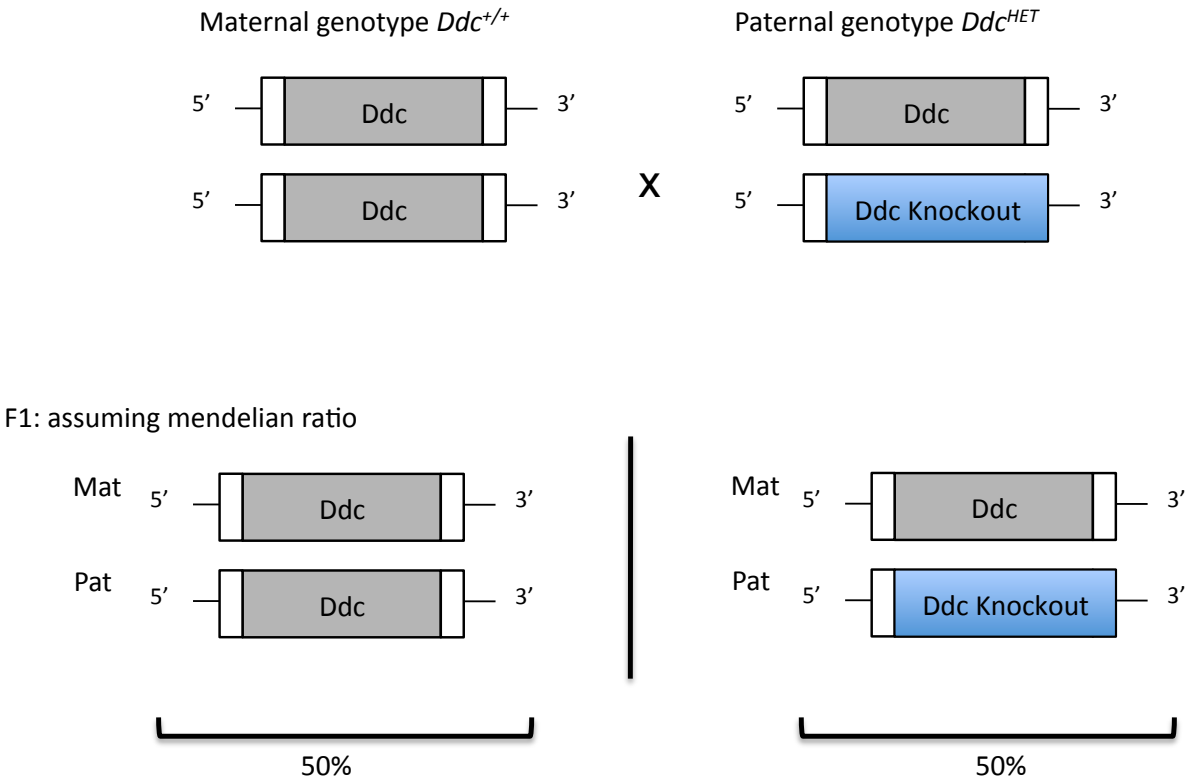


Figure 5.6 Breeding program used to generate *Ddc*$^{+/-}$ knockout mice. Wildtype females (*Ddc*$^{+/+}$) were bred with males heterozygous for the *Ddc* Knockout allele (*Ddc*$^{HET)}$. Assuming mendelian ratios of F1 progeny, 50% will be Wildtype (*Ddc*$^{+/+}$), and 50% will be heterozygous for the *Ddc* knockout allele with the mutated allele inherited on the paternal allele (*Ddc*$^{+/-}$).

Ddc expression was ablated via insertion of the VICTR 48 Omnibank Vector upstream of exon 2, with loss of function is achieved using two premature poly(A) signals attached to a splice accepter sequence (Figure 5.7). In order to confirm ablation of the coding *Ddc* gene and absence of the Ddc protein in knockout animals, gene expression and western blot analyses were performed.

5.3.2 Expression analysis of *Ddc* knockout in heart

RT-PCR analysis was performed on RNA from neonatal heart tissue of $Ddc^{+/+}$, $Ddc^{-/+}$, $Ddc^{+/-}$ and $Ddc^{-/-}$ animals. RNA was converted to cDNA using the protocol described in section 2.7.1. Primers were designed to assay *Ddc_canonical* and *Ddc_exon1a* transcripts, as well as the predicted truncated transcripts and actin control. RT-PCR analyses showed *Ddc_exon1a* expression as expected in the $Ddc^{+/+}$ and the $Ddc^{-/+}$ samples (Figure 5.8, B & C), however the analyses also revealed an unexpected weak band showing the *Ddc_exon1a* transcript present in the $Ddc^{+/-}$ heart (Figure 5.8, D). Expression of *Ddc_exon1a* is controlled by genomic imprinting and is seen only from the paternal allele (Menheniott, et al., 2008) therefore expression of *Ddc_exon1a* is not expected in $Ddc^{+/-}$ heart because the knockout allele is being inherited from the paternally inherited allele. Two hypotheses were developed to account for this result and are these are shown in Figure 5.9. The first is the 'transcriptional read-though' hypothesis whereby expression of the full length *Ddc_exon1a* transcript on the paternal allele is amplified in the RT-PCR assay because splicing occurs over the poly(A) insert (Figure 5.9, A). This was discounted because RT-PCR data from the $Ddc^{-/-}$ animals revealed no *Ddc_exon1a* expression, suggesting that the observed transcript in the $Ddc^{+/-}$ heart is likely to be coming from the maternal allele. The second hypothesis suggests that *Ddc_exon1a* expression observed in the $Ddc^{+/-}$ heart is coming from the maternal allele as a result of incomplete silencing of *Ddc_exon1a* expression (Figure 5.9, B). This may account for the RT-PCR product

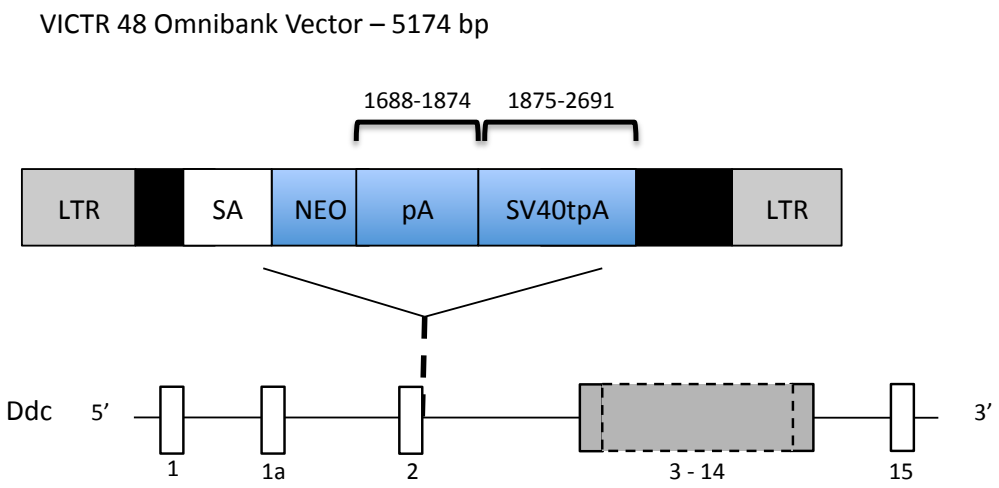


Figure 5.7 VICTR 48 Omnibank vector in *Ddc*$^{GT(neo)420Lex}$. The VICTR 48 vector is inserted downstream of exon 2. The vector contains a splice acceptor site (SA), neomycin resistance gene (NEO), polyadenylation site (pA) and an SV40 triple polyadenylation sequence (SV40tpA) flanked by retroviral long terminal repeats (LTR). White boxes represent non coding exons and grey boxes coding exons.

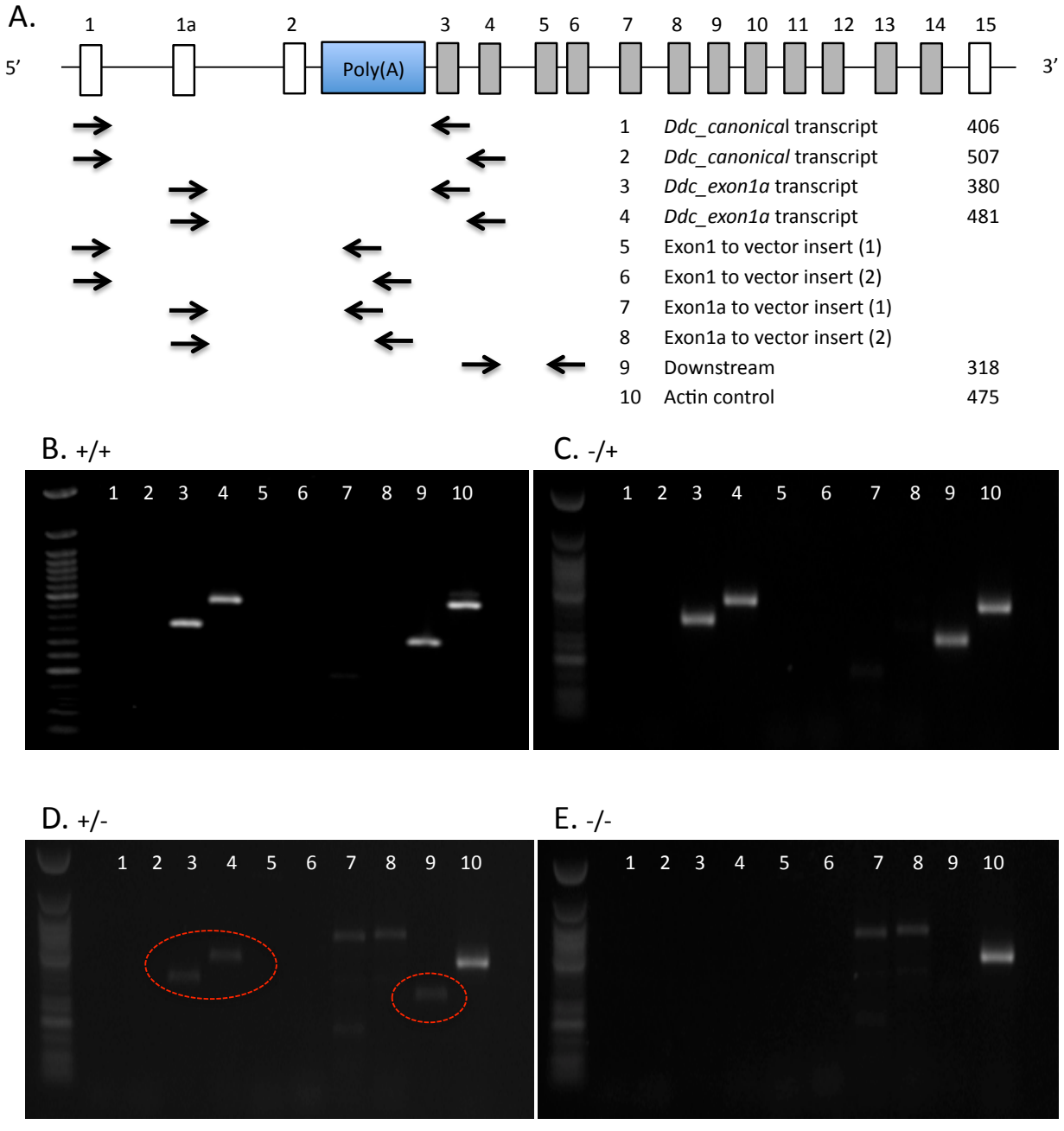


Figure 5.8 RNA transcript analysis in *Ddc* knockout mouse heart at e15.5. (A) Primers locations are indicated by black arrows, forward primers were designed in exons 1, exon 1a and exon 4. Reverse primers we designed in exons 3, exon 4 and exon 5, as well as two in the VICTR48 vector. RT-PCR experiments 1-10 are shown to the right of the primers. Results are shown for (A) $Ddc^{+/+}$, (B) $Ddc^{-/+}$, (C) $Ddc^{+/-}$, (D) $Ddc^{-/-}$. Expression is as expected except in (D), where *Ddc_exon1a* expression is present despite the fact the maternal allele is epigenetically silenced (highlighted in red)

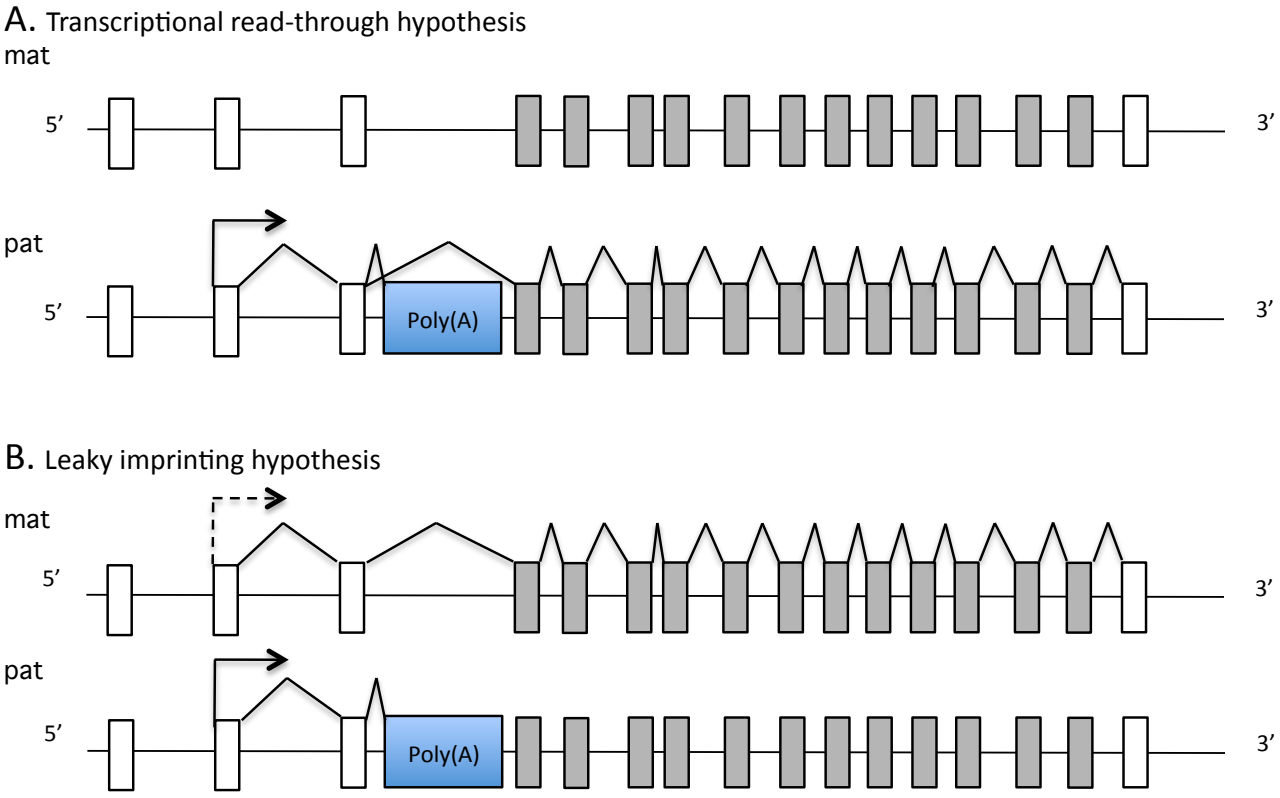


Figure 5.9 Possible mechanisms for *Ddc_exon1a* expression in *Ddc* knockout hearts. (A) Transcriptional read-through hypothesis. *Ddc_exon1a* expression is not completely ablated by the insertion of the poly(A) signal as some splicing occurs over signal. (B) Leaky imprinting hypothesis. *Ddc_exon1a* expression in the heart is only from the paternal allele, with the maternal allele epigenetically silenced, however there is residual expression of the maternal allele as silencing is not 100% efficient.

appearing as a very faint band, because residual expression from the maternal allele is low. To confirm this, quantitative PCR (qPCR) was performed on each heart sample to assess *Ddc* expression in each of the knockout genotypes (*Ddc*$^{+/+}$, *Ddc*$^{-/+}$, *Ddc*$^{+/-}$ *and Ddc*$^{-/-}$) (Figure 5.10, A). This experiment validated the second hypothesis as residual *Ddc* expression in present in the *Ddc*$^{+/-}$ hearts. In addition expression is not as high in *Ddc*$^{-/+}$ animals as in the wildtype, suggesting that the full wildtype *Ddc* expression observed in heart comprises of expression of *Ddc_exon1a* from the paternal allele, and to a lesser extent from the maternal allele. The qPCR analysis of *Ddc_exon1a* indicates a 42-fold decrease in expression in the *Ddc*$^{+/-}$ heart when compared to the *Ddc*$^{-/+}$ heart, with negligible expression from the maternal allele.

5.3.3 Western blot analysis

In order to confirm the absence of the Ddc protein when the gene is truncated, and therefore to rule out the "transcriptional read-through" hypothesis, western blot analysis of Ddc protein expression was examined in *Ddc*$^{+/+}$ and *Ddc*$^{-/-}$ animals. Protein was extracted from whole carcass and the western blot performed using the protocol described in section 2.10. The results revealed that there is no Ddc expression in the *Ddc*$^{-/-}$ animals compared to the wildtype (Figure 5.10, B) therefore it is concluded that there is no transcription read-through of the knockout alleles.

5.4 Gene expression microarray analysis

Gene expression microarray was used to assess the biological consequences of *Ddc* knockout in heart; gene expression in *Ddc*$^{+/-}$ animals was compared to *Ddc*$^{-/+}$ using a gene expression microarray. The Illumina™ WG-6 mouse microarray platform was used to assess gene expression using 45 281 probes, and analyses were performed in biological replicate, using two separate six lane arrays. The microarrays were performed on three *Ddc*$^{+/+}$, four *Ddc*$^{-/+}$, four *Ddc*$^{+/-}$ and one *Ddc*$^{-/-}$ heart. *Ddc*$^{-/-}$ genotype could only be performed as a singleton because of difficulties in obtaining

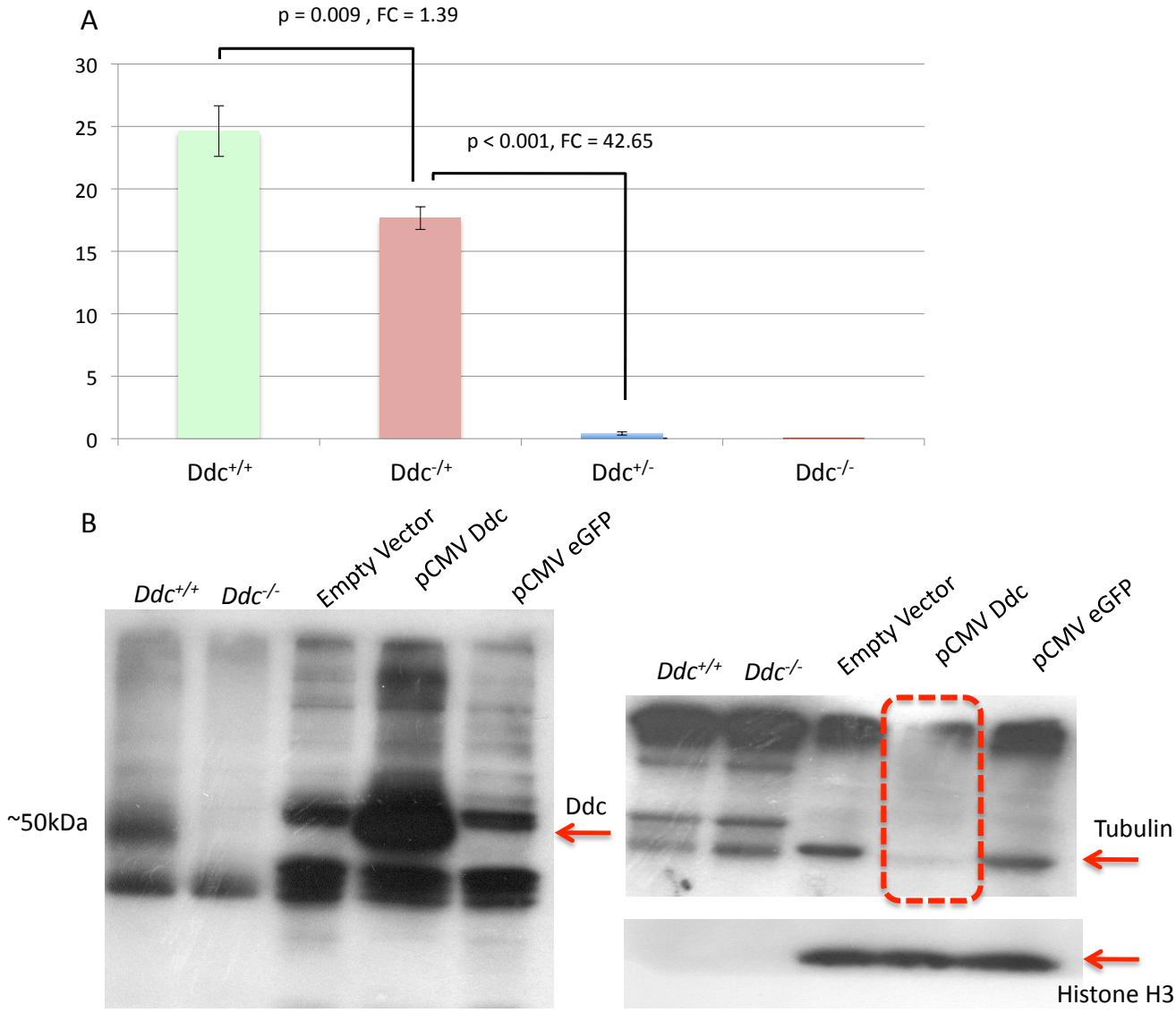


Figure 5.10 Quantitative PCR and western blot analyses of *Ddc* knockout. (A) Quantitative PCR analysis of e15.5 heart for *Ddc* transcript in $Ddc^{+/+}$, $Ddc^{-/+}$, $Ddc^{+/-}$, $Ddc^{-/-}$ genotypes showing statistical comparisons and fold change (FC) difference. (B) Western blot analysis on protein extract from either $Ddc^{+/+}$ *or* $Ddc^{-/-}$ whole carcass and NIH3T3 fibroblast cell lines transfected with either Empty Vector, Ddc under the control of the cauliflower mosaic virus promoter (pCMV Ddc) or eGFP under the control of the cauliflower mosaic virus promoter (pCMV eGFP). The western blot was probed using antibodies raised against Ddc, followed by Tubulin and Histone H3 as loading controls. Tubulin detection in the pCMV Ddc transfected fibroblast did not occur adequately, (highlighted by red box) possibly as a consequence of the detection of high levels of Ddc at the same location, so the blot was re-probed for histone H3, showing the three fibroblast cell line protein extract were equally loaded. Western blot analysis revealed no Ddc expression in the $Ddc^{-/-}$ carcass.

tissues from this genotype at e15.5. The microarray libraries were hybridized to the arrays in two separate batches. This strategy was adopted because of the scarcity of the frozen tissue samples and concerns over the quality of the array therefore scope for repetition of the experiment was planned. Data analysis was performed in collaboration with Heba Saadeh (King's College London) and results and discussion can be found in appendix 5.1 .

5.5 Structural analysis of hearts with reduced *Ddc_exon1a* expression

In order to assess phenotypic differences in the developing hearts that lack *Ddc_exon1a* expression, episcopic fluorescence image capture (EFIC) was performed on e15.5 $Ddc^{+/+}$, $Ddc^{-/+}$ and $Ddc^{+/-}$ hearts to look for gross morphological abnormalities $Ddc^{-/-}$ hearts could not be assessed due to lack of samples available. EFIC analysis utilizes episcopic images, captured during serial sectioning of embedded tissue to model a 3D image stack, which can then be re-sectioned in any plane (Weninger, et al., 2002).

The advantages of using EFIC for the study of heart morphology, compared to traditional histological analysis, are multi-fold. Embryos analysed by EFIC are lightly fixed and embedded in vybar/paraffin dyed with red alanine to suppress exongenous fluorescence during image capture. Individual sections of a specific thickness are removed and discarded before the remaining embryo is then captured *in situ* in the vybar block using a fluorescence microscope. Each image captured is stored and used to rebuild a 3D model in silico (Figure 5.11). The primary advantage of EFIC is that once rebuilt, images can be re-sectioned in identical planes across samples, thus allowing identical comparison to be made, when even slight variations in the sectioning plane of a sample may produce large variation in measurements made. Another advantage of the EFIC system is that the image captured remains embedded in the vybar block and, as such, is not

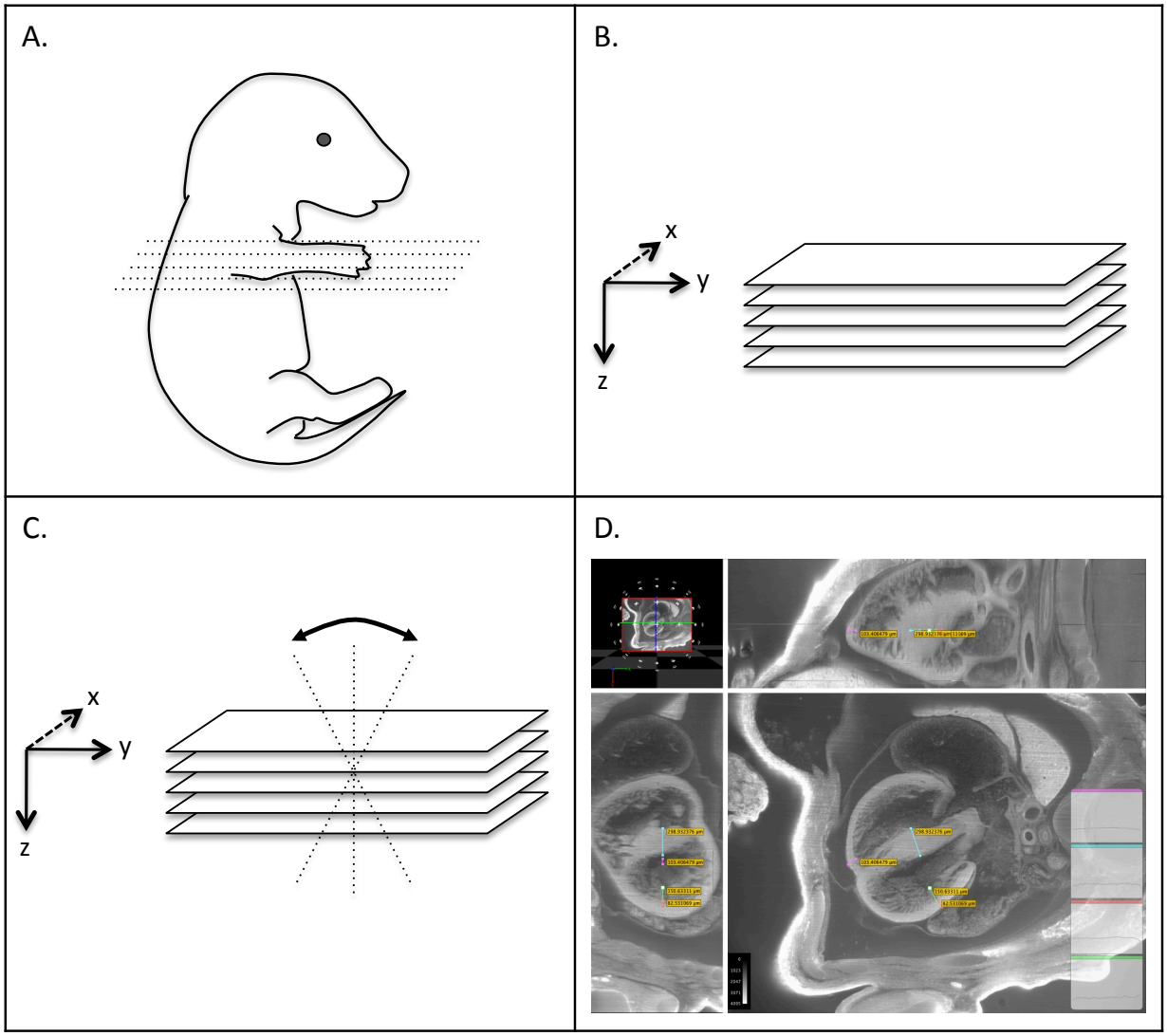


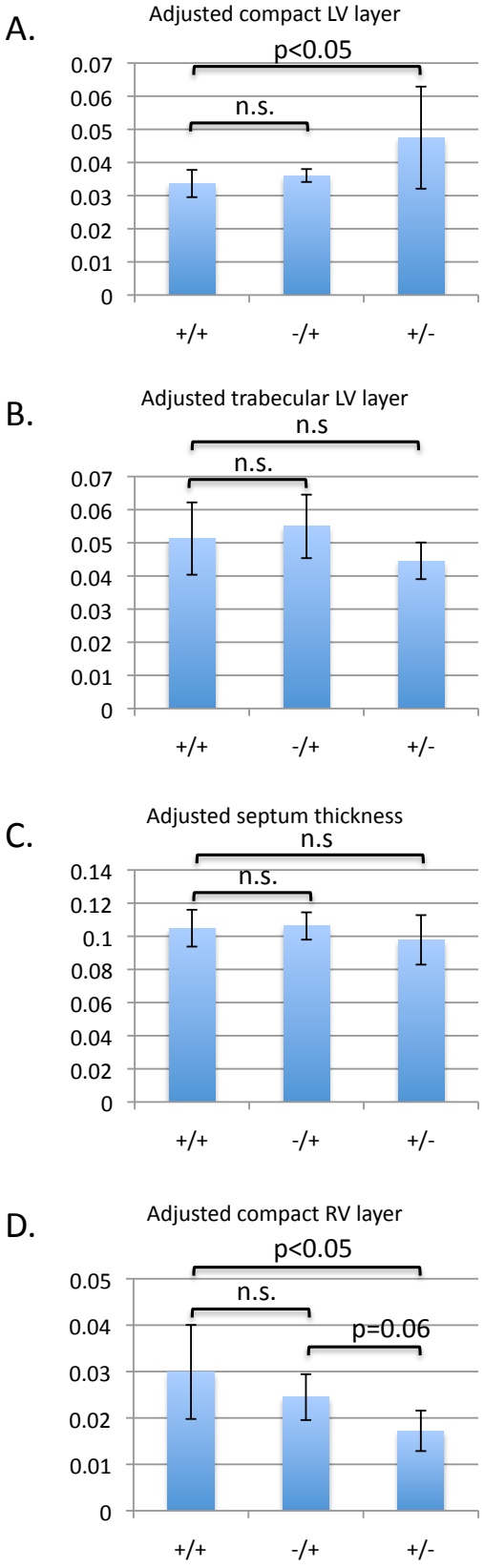
Figure 5.11 Episcopic fluorescence image capture. (A) The embryo is embedded in analine dyed paraffin/vybar and sections are removed sequentially, with the embryo imaged *in situ* using a fluorescence microscope after each section is removed. (B) The images are then stacked to rebuild a 3D model of the embryo using Velocity™ software. (C) The embryo can then be orientated and re-sectioned in any plane. (D) Measurements can them be made using equivalent sections.

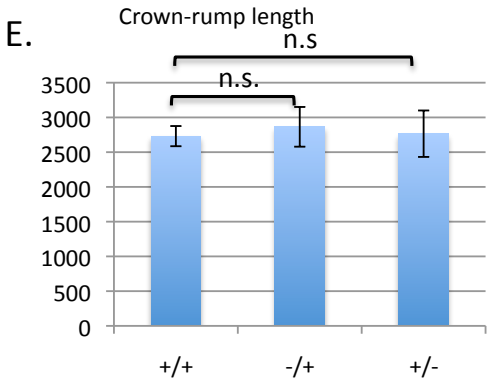
subject to distortions associated with cutting thin sections on a microtome and transferring them onto slides.

5.5.1 EFIC analysis of e15.5 heart

Multiple measurements were made of e15.5 day hearts from the three genotypes being assessed. Firstly the compact, and trabecular layers of the left ventricle were analyzed to determine whether there were any differences between the wildtype and *Ddc*$^{-/+}$ mice, with a view to hypothesising a role for Ddc in trabeculation and compaction. Secondly, in the same plane, the width of the interventricular septum was measured at its widest point. Thirdly, the thickness of the compact layer, at the apical region of the right ventricle was measure in the same direction as the septum, this region constitutes the most dense regions of Ddc protein expression as observed by immunostaining (Figure 5.12, F). No further measurements were made of the right ventricle due to difficulty in distinguishing between the trabecular layer and the compact layer of some samples, and due to a lack of a defined reference point with which to standardise all sample measurements. All measurements were adjusted to crown/rump length of the embryo, to control for variations in embryonic size. Embryonic size itself was also analyzed to check for developmental differences between genotypes (Figure 5.11, E). Due to the complex nature of the breeding programme and difficulty in obtaining multiple samples the analysis was performed on six *Ddc*$^{+/+}$, three *Ddc*$^{-/+}$ and three *Ddc*$^{+/-}$ embryos. Statistical comparisons were made between wildtype *Ddc*$^{+/+}$ and the other two genotypes independently, as the low number of replicates in the *Ddc*$^{-/+}$ vs *Ddc*$^{+/-}$ cohorts meant comparison were limited.

The results revealed that no statistically significant difference for any ventricular or structural abnormalities in the *Ddc*$^{-/+}$ embryos when compared to the wildtype *Ddc*$^{+/+}$ embryos (Figure 5.12, A, B, C, D, E), this can be attributed to the fact that when the *Ddc* knockout allele is being passed down on the epigenetically silenced maternal allele, the absence of a working copy of the gene is negated. A small difference is observed when





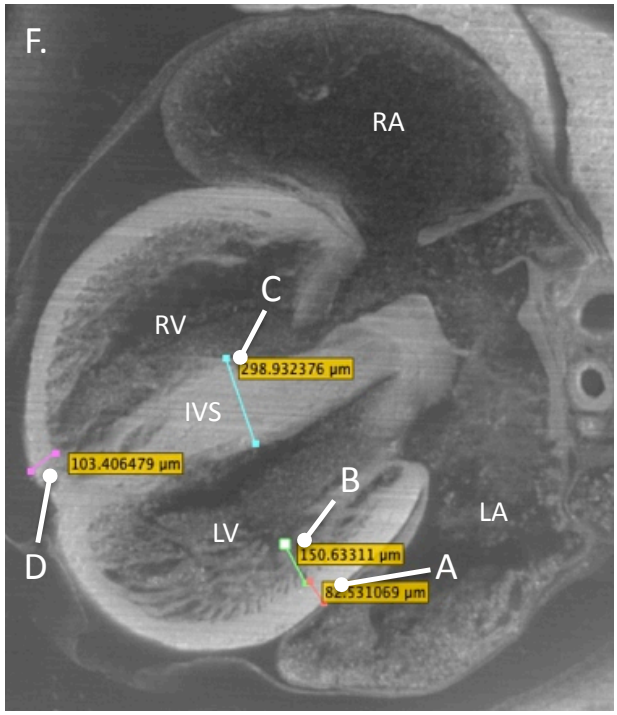


Figure 5.12 Morphological analysis of *Ddc* knockout hearts. Statistical comparisons made between *Ddc*$^{+/+}$, *Ddc*$^{-/+}$ *and Ddc*$^{+/-}$ hearts. A,B,C and D are all normalized to crown rump length. (A) Thickness of the compact left ventricular myocardium at the base of the papillary muscle. (B) Thickness of the trabeculated layer of the left ventricle (LV) at the base of the papillary mucle. (C) Thickness of the inter-ventricular septum (IVS) at the widest point. (D) Thickness of the compact right ventricular (RV) myocardium at the apical point, parallel to the IVS. (E) Crown rump length. (F) Depiction of where measurements are made.

comparing compact RV thickness (Figure 5.12, D), but this is not enough to reach statistical significance and could be attributed to noise, however in the context of previously discussed expression analysis, which indicates a small but statistically significant difference in *Ddc* expression in heart between $Ddc^{+/+}$ *and* $Ddc^{-/+}$ animals, it should not be ruled out that the lack of a working maternal copy may have an effect on RV thickness. Analysis of the $Ddc^{+/-}$ embryos compared to $Ddc^{+/+}$ reveals a significant increase in the thickness of the compact layer of the left ventricle (Figure 5.12, A), and a decrease in thickness of the compact layer in the right ventricle (Figure 5.12, D). No statistically significant differences between the three genotypes were observed for LV trabecular layer, septum thickness, or overall embryo size as measured by crown-rump length (Figure 5.12, B, C, E). A difference in right ventricular thickness between $Ddc^{-/+}$ and $Ddc^{+/-}$ embryos was noted (Figure 5.12, D), as it fell just short of statistical significance (p=0.06), probably due to the low sample number in each of the knockout genotypes (n=3 for both groups).

5.6 Discussion

The majority of *Ddc* expression in the developing heart is seen in the apical portion of the developing right ventricle, specifically in the compact layer of the myocardium. Results from morphological analysis reveal that a decrease in expression of Ddc in this portion of the right ventricle results in reduction in the thickness of the right ventricular myocardium, which can be seen when visually comparing wildtype *Ddc* hearts with $Ddc^{+/-}$ hearts at e15.5 (Figure 5.13). When combined with results from expression microarray analysis it is proposed that this may be due to decreased cellular growth and proliferation in the developing myocardium. More unexpectedly it appears the absence of Ddc in the heart results in an increase in thickness of the compact layer of the left ventricular myocardium, in a region of the e15.5 heart that displays less Ddc expression (Figure 5.2). Microarray analysis shows that disruption of *Ddc_exon1a* expression in the heart does not have a large influence on

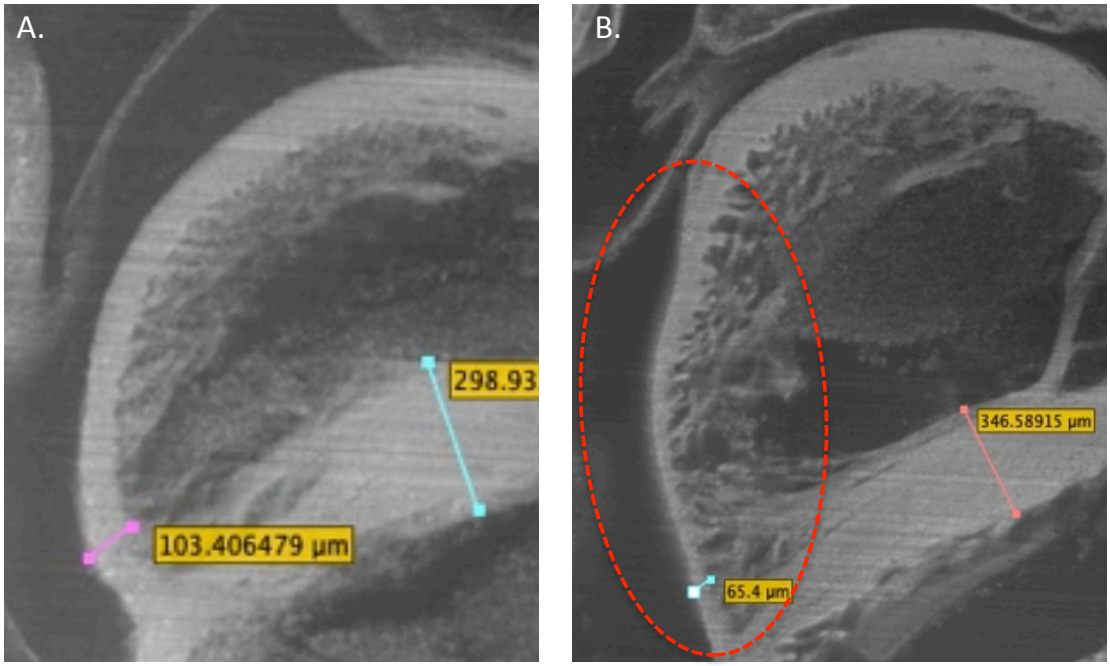


Figure 5.13 Visual comparison of *Ddc*$^{+/+}$ and *Ddc*$^{+/-}$ hearts at e15.5. Comparison of (A) *Ddc*$^{+/+}$ and (B) *Ddc*$^{+/-}$ heart at e15.5 shows hypoplastic myocardium in the right ventricle of *Ddc*$^{+/-}$ heart, highlighted in red.

global gene expression, suggesting Ddc does not function to directly regulate transcription of genes as part of a downstream process. However, there are a significant number of genes that show small, yet statistically significant up or down-regulation in response to paternal *Ddc_exon1a* knockout in heart and this suggests that Ddc is biochemically mediating cellular processes possibly via its enzymatic function. Ddc plays a fundamental role in the biosynthesis of catecholamine neurotransmitter and serotonin (Livingstone, et al., 1983) in the central nervous system (CNS). Outside of the CNS, serotonin regulates many biological processes via multiple receptors (Berger, et al., 2009, Hoyer, et al., 1994). These processes included motility, secretion and visceral sensitivity in the gastrointestinal tract (Cirillo, et al., 2011), vasoconstriction and vasodilation (Kaumann, et al., 2006). Serotonin has also been implicated multiple developmental processes in the developing heart via the serotonin signalling receptor $5HT_{2B}R$, which is expressed in cardiovascular tissues of embryonic and adult mice. Mice harbouring a germline deletion of the *5HT$_{2B}$R* gene exhibit an increased incidence of embryonic and neonatal lethality due to severe cardiac phenotypes, with death seen from e10.5 due to myocardial rupture indicated by leakage of blood into the pericardial cavity. Surviving *5HT$_{2B}$R* $^{-/-}$newborn mice display a severe ventricular hypoplasia caused by impaired proliferative capacity of cardiomyocytes and abnormal sarcomeric differentiation, as 5HT2BR is required for correct cytoskeleton assembly to membrane structures by regulating expression of N-cadherin (Nebigil, et al., 2000, Nebigil, et al., 2001, Nebigil, et al., 2001). Adult mice lacking *5HT$_{2B}$R* expression exhibit significant decrease in cardiomyocyte size and number, resulting in thinning of the ventricular wall and a reduction in ventricular mass. It is proposed that Serotonin signalling via the $5HT_{2B}$ receptor regulates expression of *ErbB-2* (Nebigil, et al., 2000), which is shown to control differentiation in developing heart (Lemke, 1996), however significant changes in *ErbB-2* transcript expression were not observed in the microarray analysis performed on *Ddc* knockout hearts. From the

observation made in this chapter it is proposed that *Ddc_exon1a* may be functioning in specific regions of the fetal heart to drive cell proliferation and cytoskeletal organisation of cardiomyocytes during development of the myocardium possibly via its enzymatic function to increase serotonin levels, which would in turn activate the $5HT_{2B}$ receptor to increase cell proliferation. However, this role is putative and these results forms the basis for future experiments which may confirm this role.

Chapter 6

Final discussion

The discovery of the first imprinted gene in the early 1990s (DeChiara, et al., 1991) prompted a period of energetic study, which culminate in the discovery of around 140 examples of genes subject to genomic imprinting in mouse (Schulz, et al., 2008, Williamson CM, 2012). Novel sequencing technologies that have become widely available since 2000 have allowed high throughput quantitative analysis of the transcriptome (Brenner, et al., 2000, Reinartz, et al., 2002) with deep coverage and base pair resolution. However, rather than opening the floodgates for the discovery of many new imprint transcripts, RNA-seq has been successfully used to identify a relatively small number of novel imprinted transcripts on top of those already known (Babak, et al., 2008, Wang, et al., 2008) suggesting that it is unlikely that the total number of imprinted genes will exceed current estimates of up to 200 (Kelsey, et al., 2012). With this in mind, it is perhaps an ideal time to look at common themes in tissue-specific expression of imprinted genes and control of genomic imprinting. A screen of genes imprinted in a single tissue or cell type, demonstrates that this type of tissue-specific imprinting is most common in neuronal cell types, and in the extra-embryonic tissues of the placenta and yolk sac (Prickett, et al., 2012). *Ddc_exon1a* is the only imprinted gene that has so far been identified that does not follow this pattern, being imprinted specifically in the heart and remaining bi-allelic in brain (Menheniott, et al., 2008). This study focuses on the function of *Ddc_exon1a* in the developing heart, its mechanism of imprinting and how these relate to common features of other imprinted genes and their imprinting control regions.

CTCF binding has long been implicated in genomic imprinting as it binds in an allele-specific manner to the *Igf2/H19* DMR, one of the first imprinting

control regions to be discovered (Thorvaldsen, et al., 1998). Investigating allele-specific CTCF binding patterns in mouse brain shows that CTCF binds at 12 of the 20 characterised imprinting control regions. Of these, five ICRs bind CTCF on the unmethylated allele, and five demonstrate a binding bias towards the unmethylated allele. These results support the assertion that CTCF plays a central role at imprinted loci by epigenetically mediating chromosomal interactions in order to control gene expression (Recillas-Targa, et al., 2006, Ren, et al., 2012), whilst acknowledging there must be other mechanisms to control imprinting at other loci. More generally CTCF has multiple functions, mediating chromosomal looping by acting as a barrier between adjacent chromosomal regions (Bell, et al., 1999), directly mediating long range chromosomal interaction (Hadjur, et al., 2009, Yusufzai, et al., 2004) and enhancing transcription (Kuzmin, et al., 2005). It has been proposed that different functions may be a result of the context of CTCF binding (Gaszner, et al., 2006), and it is possible that co-ordinate binding of Cohesin with CTCF may influence function of CTCF. CTCF has been shown to physically associate with Cohesin via the Stag1 (Scc3/SA1) subunit in human cells (Wendt, et al., 2008). Cohesin is involved in tissue-specific transcriptional control (Faure, et al.) and is associated with the mediator complex that has a role in transcriptional activation (Taatjes). Studies have also shown a link between Cohesin, the mediator complex, transcription and chromatin looping (Kagey, et al., 2010). Of the 20 gDMRs examined we detected co-ordinate binding of CTCF and Cohesin at 10 (50%), compared to the genome-wide incidence of co-ordinate binding which is ~30%. In addition a novel class of gDMR, that binds only Cohesin was identified and this observation supports the assertion that CTCF is not required to position Cohesin onto DNA (Rubio, 2008), however it is unlikely that this mechanism is responsible for interpreting the methylation imprint as binding does not occur in a parent-of-origin specific manner.

CTCF binding at an imprinting control region represents a mechanism by which differential methylation between parental alleles can influence gene

expression by blocking a DNA-protein interaction (Mukhopadhyay, 2004, Recillas-Targa, et al., 2006), however there are no published examples where tissue-specific changes in CTCF binding govern tissue-specific differences in imprinting, and evidence suggests that CTCF binding in mammals is largely invariant between cell types (Chen, et al., 2012, Kim, et al., 2007). In order to explore the effects of tissue-specific binding of CTCF as a potential mechanism for controlling differential gene expression in different tissue types, genome-wide comparisons of CTCF binding were made between ES cells, brain and liver. These analyses provide evidence for tissue-specific variation in CTCF binding. In addition, a link between tissue-specific binding and the preference for CTCF to bind to different DNA motifs is reported for the first time.

Understanding epigenetic features of other imprinted loci will help to inform on the control of genomic imprinting of *Ddc_exon1a* in the heart, however the unique features of *Ddc_exon1a* mean direct parallels cannot easily be drawn. In this thesis epigenetic features of the locus in brain and heart are explored to look for differences that account for the switch from bi-allelic expression to imprinted expression. Analysis of the *Grb10* DMR shows that CTCF binding here is invariant between heart and brain, however existing work by Shiura *et al.* shows that expression of *Ddc_exon1a* in heart disappeared when the DMR is knocked out on the paternal allele (Shiura, et al., 2009). These results suggest a requirement for a second epigenetic mark, working in concert with the DMR to regulate the tissue-specific imprinting of *Ddc_exon1a* in heart.

The most intriguing line of enquiry for this second epigenetic mark is found at the promoter of the antisense transcript *AK0066690,* a region hypomethylated in brain relative to heart. There is precedence for antisense transcripts regulating imprinting at other loci. The *Igf2r* locus is an excellent example of an imprinted gene being regulated by an antisense transcript, known as *Air* (Sleutels, et al., 2002). At the *Igf2* locus, the antisense transcript is silenced on the maternal allele permitting the

transcription of the maternal *Igf2r* gene. Conversely the *Air* is transcribed from the paternal allele in turn silencing *Igf2r* on the paternal allele. This process is regulated by a region of maternal methylation at the promoter of *Air*. The promoter of *AK0066690* differs from *Igf2r* as it is not differentially methylated between the parental alleles, however interplay between the *Grb10* DMR and tissue-specific epigenetic differences at the antisense promoter may function to regulate expression of *AK0066690* from the maternal allele-specifically in heart. Unfortunately allele-specific expression of *AK0066690* could not be examined without further characterization of the transcript that was not possible within the time limits of this thesis, however *AK0066690* expression is observed in heart but not in brain, consistent with this model. These analyses allow a model for imprinting control of *Ddc_exon1a* to be proposed.

Further work is required to support the antisense transcription model. Demonstration that *AK0066690* is maternally expressed and paternally silenced would support this model for *Ddc_exon1a* silencing, however parent-of-origin specific expression could not be assayed using F1 hybrid mice from BxC and CxB crosses as the annotated transcript does not contain any polymorphisms between the two subspecies. There are several methods that could be exploited to test imprinting in the absence of a SNP, firstly pUPD11 and mUPD11 tissues could be used and expression measure using quantitative PCR, however UPD tissues could not be sourced. In addition further characterization of the full length of the *AK0066690* transcript may yield exons that cover polymorphisms between Bl6 and *cast* which would allow allele specific expression to be tested.

In order to build evidence for an analogous mechanism at *AK0066690/Ddc_exon1a* to that seen at *Igf2r/Air,* the length of the *AK0066690* transcript would need to be elucidated thus ensuring transcription runs through the *Ddc_exon1a* promoter. In order to silence *Igf2r, Air* must transcribe through the *Igf2r* promoter (Latos, et al., 2012).

This is because convergently transcribing RNA polymerase II molecules cannot transcribe past each other, resulting in stalling and inefficient transcription (Hobson, et al.,2012). Northern blot analysis can be used to assess the length of the *AK0066690* transcript and its expression in heart compared to brain. In addition 3' RACE would unequivocally show if transcription overlaps the *Ddc_exon1a* promoter, although ncRNAs are often not polyadenylated which poses an additional challenge to this assay.

The model of imprinting control proposed also suggests a long range chromosomal interaction between the imprinting control region and the tissue specifically methylated promoter of *AK0066690* mediated via CTCF. Chromosomal conformation capture experiments can be used to assess long range chromosomal interactions and this would build on evidence of an interaction within the *Ddc/Grb10* locus seen using ChIA-PET analysis in ES cells (Handoko, et al., 2011). These experiments are often confounded by false positives, however, novel methods which build on these technologies will allow these interactions to be assessed with greater certainty (Sexton, et al., 2012).

Finally, a comparison between *Ddc_exon1a* expression in mouse and human reveals that *DDC_EXON1A* is mono-allelically expressed in human, it is therefore probable the control of imprinting exhibits a conserved mechanism. Genetic and epigenetic elements conserved from mouse to human may be indicative of features implicated in the imprinting control of *Ddc_exon1a* in heart, thus comparisons between the two species may be made to look for common genetic features at the locus.

Tissue-specific imprinting of *Ddc_exon1a* in heart is unique. In order to investigate the role of Ddc in heart development, and to consider the evolutionary reasons for imprinting of *Ddc_exon1a*, the role of Ddc in heart was assessed using microarray and morphological analyses. The results suggest a role for the Ddc protein in the cell proliferation in the cardiac

muscle of the developing heart, particularly in the apical portion of the right ventricle where Ddc expression strongest. Ddc has multiple characterised enzymatic functions but the most functionally relevant in this context is its ability to catalyse the conversion 5-Hydroxytryptophan (5-HTP) into serotonin (Livingstone, et al., 1983). Serotonin has multiple signalling functions regulated via various serotonin receptors (Hoyer, et al., 1994), but in the context of heart development serotonin functions to regulate cardiomyocyte proliferation and cytoskeletal organization in the developing heart via the 5-HT_{2B} receptor (5-HT_{2B}R). Expression of Ddc in the developing heart may increase levels of serotonin locally, driving this developmental process. It is clear from the microarray analysis that Ddc does not directly influence downstream gene regulation in heart, therefore it is postulated that the effect seen is more subtly regulated via a biochemical feedback mechanism. These effects can be studied in more detail using biochemical profiling experiments to measure levels of serotonin using HPLC coupled with electrochemical detection.

Analysis of *Ddc_exon1a* function can be further investigated using *Ddc* conditional knockout animals, under the control of a heart specific promoter driven cre recombinase. This would eliminate any confounding factors resulting from heterozygous deletion of *Ddc* in other tissues, or other as yet unidentified tissues that may possibly display imprinted *Ddc* expression. A conditional *Ddc* knockout has already been used to assess the function of Ddc in kidney (Zhang, et al., 2011).

The mechanism of *Ddc_exon1a* imprinting cannot be explained adequately by applying knowledge gained from the study of imprinting control mechanisms at other imprinted loci, despite this the *Ddc* locus exhibits several characteristics in common with them. It is proposed that *Ddc_exon1a* imprinting involves differential CTCF binding and antisense transcription, combining two previously separate models of imprinting control (Ideraabdullah, et al., 2008). *Ddc_exon1a* is the only example of a

heart-specific imprinted gene, and this thesis demonstrates that *Ddc_exon1a* expression plays a role in regulating cell proliferation in heart by increasing cell growth and division in the apical portion of the right ventricle, and regulating division in the compact left ventricular myocardium. One might propose that the correct balance of cell division in the developing myocardium is dependent on very tight dosage regulation of Ddc, and that genomic imprinting of *Ddc_exon1a* has evolved to achieve this in a manner similar to that seen at *Igf2. Ifg2* is a driver of cell division and growth (Beck, et al., 1987) and imprinting is required to finely balance dosage control of this powerful embryonic mitogen in order to correctly regulate fetal growth and development (Searle, et al., 1990). This theory leaves further scope to explore evidence for the evolution of *Ddc_exon1a* imprinting in heart in order to ascertain whether cell proliferation is dose-dependent in response to local levels of Ddc, which may provide further evidence that imprinting is necessary to control gene dosage of Ddc in heart. To test this theory of dosage dependent control of Ddc the *Ddc* gene can be inserted in a BAC construct to assess phenotypes associated with upregulated gene dosage. This method has been used previously to assess upregulated gene dosage of the *Dlk1* in mouse (Charalambous, et al., 2012). The mechanistic control of *Ddc_exon1a* imprinting, and the reason for tissue-specific imprinting in the developing heart will provide excellent scope for investigating both the mechanistic and evolutionary function of genomic imprinting in the future.

References

Aguanno, A., Afar, R. and Albert, V. R. (1996). Tissue-specific expression of the nonneuronal promoter of the aromatic L-amino acid decarboxylase gene is regulated by hepatocyte nuclear factor 1. J Biol Chem. 271, 4528-4538

Arnaud, P., Monk, D., Hitchins, M., Gordon, E., Dean, W., Beechey, C. V., Peters, J., Craigen, W., Preece, M., Stanier, P., Moore, G. E. and Kelsey, G. (2003). Conserved methylation imprints in the human and mouse GRB10 genes with divergent allelic expression suggests differential reading of the same mark. Human Molecular Genetics. 12, 1005-1019

Avvakumov, G. V., Walker, J. R., Xue, S., Li, Y., Duan, S., Bronner, C., Arrowsmith, C. H. and Dhe-Paganon, S. (2008). Structural basis for recognition of hemi-methylated DNA by the SRA domain of human UHRF1. Nature. 455, 822-825

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. Current Biology. 18, 1735-1741

Bailey, T. L., Williams, N., Misleh, C. and Li, W. W. (2006). MEME: discovering and analyzing DNA and protein sequence motifs. Nucleic Acids Res. 34, W369-373

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

Bartolomei, M. S. (2009). Genomic imprinting: employing and avoiding epigenetic processes. Genes & Development. 23, 2124-2133

Beck, F., Samani, N. J., Penschow, J. D., Thorley, B., Tregear, G. W. and Coghlan, J. P. (1987). Histochemical localization of IGF-I and -II mRNA in the developing rat embryo. Development. 101, 175-184

Bell, A. C., West, A. G. and Felsenfeld, G. (1999). The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. Cell. 98, 387-396

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

Berger, M., Gray, J. A. and Roth, B. L. (2009). The expanded biology of serotonin. Annu Rev Med. 60, 355-366

Bestor, T., Laudano, A., Mattaliano, R. and Ingram, V. (1988). Cloning and sequencing of a cDNA encoding DNA methyltransferase of mouse cells. The carboxyl-terminal domain of the mammalian enzymes is related to bacterial restriction methyltransferases. Journal of molecular biology. 203, 971-983

Bird, A. (2002). DNA methylation patterns and epigenetic memory. Genes & Development. 16, 6-21

Bird, A. P. (1980). DNA methylation and the frequency of CpG in animal DNA. Nucleic Acids Research. 8, 1499-1504

Bird, A. P. (1986). CpG-rich islands and the function of DNA methylation. Nature. 321, 209-213

Bittel, D. C. and Butler, M. G. (2005). Prader–Willi syndrome: clinical genetics, cytogenetics and molecular biology. Expert Reviews in Molecular Medicine. 7,

Borglum, A. D., Bruun, T. G., Kjeldsen, T. E., Ewald, H., Mors, O., Kirov, G., Russ, C., Freeman, B., Collier, D. A. and Kruse, T. A. (1999). Two novel variants in the DOPA decarboxylase gene: association with bipolar affective disorder. Mol Psychiatry. 4, 545-551

Bourc'his, D., Xu, G. L., Lin, C. S., Bollman, B. and Bestor, T. H. (2001). Dnmt3L and the establishment of maternal genomic imprints. Science. 294, 2536-2539

Brenner, S., Johnson, M., Bridgham, J., Golda, G., Lloyd, D. H., Johnson, D., Luo, S., McCurdy, S., Foy, M., Ewan, M., Roth, R., George, D., Eletr, S., Albrecht, G., Vermaas, E., Williams, S. R., Moon, K., Burcham, T., Pallas, M., DuBridge, R. B., Kirchner, J., Fearon, K., Mao, J. and Corcoran, K. (2000). Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. Nature biotechnology. 18, 630-634

Carr, M. S., Yevtodiyenko, A., Schmidt, C. L. and Schmidt, J. V. (2007). Allele-specific histone modifications regulate expression of the Dlk1-Gtl2 imprinted domain. Genomics. 89, 280-290

Cassidy, S. B., Dykens, E. and Williams, C. A. (2000). Prader-Willi and Angelman syndromes: sister imprinted disorders. Am J Med Genet. 97, 136-146

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

Cattanach, B. M. (1986). Parental origin effects in mice. Journal of embryology and experimental morphology. 97 Suppl, 137-150

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. Proceedings of the National Academy of Sciences of the United States of America. 100, 8292-8297

Charalambous, M., Menheniott, T. R., Bennett, W. R., Kelly, S. M., Dell, G., Dandolo, L. and Ward, A. (2004). An enhancer element at the Igf2/H19 locus drives gene expression in both imprinted and non-imprinted tissues. Developmental biology. 271, 488-497

Charalambous, M., Ferron, S. R., da Rocha, S. T., Murray, A. J., Rowland, T., Ito, M., Schuster-Gossler, K., Hernandez, A. and Ferguson-Smith, A. C. (2012). Imprinted gene dosage is critical for the transition to independent life. Cell Metab. 15, 209-221

Chen, H., Tian, Y., Shu, W., Bo, X. and Wang, S. (2012). Comprehensive identification and annotation of cell type-specific and ubiquitous CTCF-binding sites in the human genome. PLoS ONE. 7, e41374

Chen, T. (2002). A Novel Dnmt3a Isoform Produced from an Alternative Promoter Localizes to Euchromatin and Its Expression Correlates with Active de Novo Methylation. Journal of Biological Chemistry. 277, 38746-38754

Chen, T., Ueda, Y., Dodge, J. E., Wang, Z. and Li, E. (2003). Establishment and Maintenance of Genomic Methylation Patterns in Mouse Embryonic Stem Cells by Dnmt3a and Dnmt3b. Molecular and Cellular Biology. 23, 5594-5605

Chen, T. H. P., Chang, T.-C., Kang, J.-O., Choudhary, B., Makita, T., Tran, C. M., Burch, J. B. E., Eid, H. and Sucov, H. M. (2002). Epicardial induction of fetal cardiomyocyte proliferation via a retinoic acid-inducible trophic factor. Developmental biology. 250, 198-207

Chen, X., Xu, H., Yuan, P., Fang, F., Huss, M., Vega, V. B., Wong, E., Orlov, Y. L., Zhang, W., Jiang, J., Loh, Y.-H., Yeo, H. C., Yeo, Z. X., Narang, V., Govindarajan, K. R., Leong, B., Shahab, A., Ruan, Y., Bourque, G., Sung, W.-K., Clarke, N. D., Wei, C.-L. and Ng, H.-H. (2008). Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. Cell. 133, 1106-1117

Christenson, J. G., Dairman, W. and Udenfriend, S. (1972). On the identity of DOPA decarboxylase and 5-hydroxytryptophan decarboxylase

(immunological titration-aromatic L-amino acid decarboxylase-serotonin-dopamine-norepinephrine). Proc Natl Acad Sci U S A. 69, 343-347

Cirillo, C., Vanden Berghe, P. and Tack, J. (2011). Role of serotonin in gastrointestinal physiology and pathology. Minerva Endocrinol. 36, 311-324

Clayton-Smith, J. and Laan, L. (2003). Angelman syndrome: a review of the clinical and genetic aspects. Journal of medical genetics. 40, 87-95

Cuddapah, S., Jothi, R., Schones, D. E., Roh, T.-Y., Cui, K. and Zhao, K. (2009). Global analysis of the insulator binding protein CTCF in chromatin barrier regions reveals demarcation of active and repressive domains. Genome Research. 19, 24-32

Curley, J. P., Barton, S., Surani, A. and Keverne, E. B. (2004). Coadaptation in mother and infant regulated by a paternally expressed imprinted gene. Proceedings of the Royal Society B: Biological Sciences. 271, 1303-1309

da Rocha, S. T., Charalambous, M., Lin, S. P., Gutteridge, I., Ito, Y., Gray, D., Dean, W. and Ferguson-Smith, A. C. (2009). Gene dosage effects of the imprinted delta-like homologue 1 (dlk1/pref1) in development: implications for the evolution of imprinting. PLoS Genet. 5, e1000392

Davies, W., Isles, A. R., Humby, T. and Wilkinson, L. S. (2007). What are imprinted genes doing in the brain? Epigenetics. 2, 201-206

DeChiara, T. M., Efstratiadis, A. and Robertson, 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

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

DeVeale, B., van der Kooy, D. and Babak, T. (2012). Critical Evaluation of Imprinted Gene Expression by RNA–Seq: A New Perspective. PLoS Genetics. 8, e1002600

Donlon, T. A. (1988). Similar molecular deletions on chromosome 15q11.2 are encountered in both the Prader-Willi and Angelman syndromes. Human genetics. 80, 322-328

Eggermann, T. (2009). Silver-Russell and Beckwith-Wiedemann Syndromes: Opposite (Epi)Mutations in 11p15 Result in Opposite Clinical Pictures. Hormone Research. 71, 30-35

Engel, N. (2006). CTCF binding sites promote transcription initiation and prevent DNA methylation on the maternal allele at the imprinted H19/Igf2 locus. Human Molecular Genetics. 15, 2945-2954

Engel, N., Raval, A. K., Thorvaldsen, J. L. and Bartolomei, S. M. (2008). Three-dimensional conformation at the H19/Igf2 locus supports a model of enhancer tracking. Human Molecular Genetics. 17, 3021-3029

Eppig, J. T., Blake, J. A., Bult, C. J., Kadin, J. A. and Richardson, J. E. (2012). The Mouse Genome Database (MGD): comprehensive resource for genetics and genomics of the laboratory mouse. Nucleic Acids Res. 40, D881-886

Faure, A. J., Schmidt, D., Watt, S., Schwalie, P. C., Wilson, M. D., Xu, H., Ramsay, R. G., Odom, D. T. and Flicek, P. (2012). Cohesin regulates tissue-specific expression by stabilizing highly occupied cis-regulatory modules. Genome Res. 22, 2163-2175

Feil, R. and Berger, F. (2007). Convergent evolution of genomic imprinting in plants and mammals. Trends in genetics : TIG. 23, 192-199

Ferrón, S. R., Charalambous, M., Radford, E., McEwen, K., Wildner, H., Hind, E., Morante-Redolat, J. M., Laborda, J., Guillemot, F., Bauer, S. R., Fariñas, I. and Ferguson-Smith, A. C. (2011). Postnatal loss of Dlk1 imprinting in stem cells and niche astrocytes regulates neurogenesis. Nature. 475, 381-385

Filippova, G. N., Fagerlie, S., Klenova, E. M., Myers, C., Dehner, Y., Goodwin, G., Neiman, P. E., Collins, S. J. and Lobanenkov, V. V. (1996). An exceptionally conserved transcriptional repressor, CTCF, employs different combinations of zinc fingers to bind diverged promoter sequences of avian and mammalian c-myc oncogenes. Mol Cell Biol. 16, 2802-2813

Fitzpatrick, G. V., Pugacheva, E. M., Shin, J. Y., Abdullaev, Z., Yang, Y., Khatod, K., Lobanenkov, V. V. and Higgins, M. J. (2007). Allele-specific binding of CTCF to the multipartite imprinting control region KvDMR1. Mol Cell Biol. 27, 2636-2647

Fournier, C., Goto, Y., Ballestar, E., Delaval, K., Hever, A. M., Esteller, M. and Feil, R. (2002). Allele-specific histone lysine methylation marks regulatory regions at imprinted mouse genes. The EMBO Journal. 21, 6560-6570

Frost, J., Monk, D., Moschidou, D., Guillot, P. V., Stanier, P., Minger, S. L., Fisk, N. M., Moore, H. D. and Moore, G. E. (2011). The effects of culture on genomic imprinting profiles in human embryonic and fetal mesenchymal

stem cells. Epigenetics : official journal of the DNA Methylation Society. 6, 52-62

Gardiner-Garden, M. and Frommer, M. (1987). CpG islands in vertebrate genomes. J Mol Biol. 196, 261-282

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., Wilkinson, L. S., Isles, A. R. and Ward, A. (2011). Distinct physiological and behavioural functions for parental alleles of imprinted Grb10. Nature. 469, 534-538

Gaszner, M. and Felsenfeld, G. (2006). Insulators: exploiting transcriptional and epigenetic mechanisms. Nat Rev Genet. 7, 703-713

Gitler, A. D., Lu, M. M., Jiang, Y. Q., Epstein, J. A. and Gruber, P. J. (2003). Molecular markers of cardiac endocardial cushion development. Dev Dyn. 228, 643-650

Goll, M. G. and Bestor, T. H. (2005). Eukaryotic Cytosine Methyltransferases. Annual Review of Biochemistry. 74, 481-514

Goll, M. G. (2006). Methylation of tRNAAsp by the DNA Methyltransferase Homolog Dnmt2. Science. 311, 395-398

Grandjean, V., O'Neill, L., Sado, T., Turner, B. and Ferguson-Smith, A. (2001). Relationship between DNA methylation, histone H4 acetylation and gene expression in the mouse imprinted Igf2-H19 domain. FEBS Letters. 488, 165-169

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. Science. 329, 643-648

Gu, T.-P., Guo, F., Yang, H., Wu, H.-P., Xu, G.-F., Liu, W., Xie, Z.-G., Shi, L., He, X., Jin, S.-g., Iqbal, K., Shi, Y. G., Deng, Z., Szabó, P. E., Pfeifer, G. P., Li, J. and Xu, G.-L. (2011). The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. Nature. 477, 606-610

Gullerova, M. and Proudfoot, N. J. (2008). Cohesin complex promotes transcriptional termination between convergent genes in S. pombe. Cell. 132, 983-995

Hadjur, S., Williams, L. M., Ryan, N. K., Cobb, B. S., Sexton, T., Fraser, P., Fisher, A. G. and Merkenschlager, M. (2009). Cohesins form chromosomal cis-interactions at the developmentally regulated IFNG locus. Nature. 460, 410-413

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

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., Ancelin, K., Waldmann, T., Lacoste, N., Lange, U. C., Cesari, F., Lee, C., Almouzni, G., Schneider, R. and Surani, M. A. (2008). Chromatin dynamics during epigenetic reprogramming in the mouse germ line. Nature. 452, 877-881

Hajkova, P. (2010). Epigenetic reprogramming â€" taking a lesson from the embryo. Current opinion in cell biology. 22, 342-350

Han, L., Su, B., Li, W.-H. and Zhao, Z. (2008). CpG island density and its correlations with genomic features in mammalian genomes. Genome Biology. 9, R79

Handoko, L., Xu, H., Li, G., Ngan, C. Y., Chew, E., Schnapp, M., Lee, C. W. H., Ye, C., Ping, J. L. H., Mulawadi, F., Wong, E., Sheng, J., Zhang, Y., Poh, T., Chan, C. S., Kunarso, G., Shahab, A., Bourque, G., Cacheux-Rataboul, V., Sung, W.-K., Ruan, Y. and Wei, C.-L. (2011). CTCF-mediated functional chromatin interactome in pluripotent cells. Nature Genetics. 43, 630-638

Hark, A. T., Schoenherr, C. J., Katz, D. J., Ingram, R. S., Levorse, J. M. and Tilghman, S. M. (2000). CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus. Nature. 405, 486-489

Hayashi, M., Yamaji, Y., Kitajima, W. and Saruta, T. (1990). Aromatic L-amino acid decarboxylase activity along the rat nephron. Am J Physiol. 258, F28-33

Hikichi, T. (2003). Imprinting regulation of the murine Meg1/Grb10 and human GRB10 genes; roles of brain-specific promoters and mouse-specific CTCF-binding sites. Nucleic Acids Research. 31, 1398-1406

Hobson, D. J., Wei, W., Steinmetz, L. M. and Svejstrup, J. Q. (2012). RNA polymerase II collision interrupts convergent transcription. Mol Cell. 48, 365-374

Holliday, R. and Pugh, J. E. (1975). DNA modification mechanisms and gene activity during development. Science. 187, 226-232

Hoyer, D., Clarke, D. E., Fozard, J. R., Hartig, P. R., Martin, G. R., Mylecharane, E. J., Saxena, P. R. and Humphrey, P. P. (1994). International Union of Pharmacology classification of receptors for 5-hydroxytryptamine (Serotonin). Pharmacological reviews. 46, 157-203

Hu, J. F., Pham, J., Dey, I., Li, T., Vu, T. H. and Hoffman, A. R. (2000). Allele-specific histone acetylation accompanies genomic imprinting of the insulin-like growth factor II receptor gene. Endocrinology. 141, 4428-4435

Huang, D. W., Sherman, B. T. and Lempicki, R. A. (2008). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nature Protocols. 4, 44-57

Ichinose, H., Ohye, T., Fujita, K., Pantucek, F., Lange, K., Riederer, P. and Nagatsu, T. (1994). Quantification of mRNA of tyrosine hydroxylase and aromatic L-amino acid decarboxylase in the substantia nigra in Parkinson's disease and schizophrenia. J Neural Transm Park Dis Dement Sect. 8, 149-158

Ideraabdullah, F. Y., Vigneau, S. and Bartolomei, M. S. (2008). Genomic imprinting mechanisms in mammals. Mutation research. 647, 77-85

Iqbal, K., Jin, S.-g., Pfeifer, G. P. and Szabó, P. E. (2011). Reprogramming of the paternal genome upon fertilization involves genome-wide oxidation of 5-methylcytosine. Proceedings of the National Academy of Sciences. 108, 3642-3647

Jenni, R., Oechslin, E., Schneider, J., Attenhofer Jost, C. and Kaufmann, P. A. (2001). Echocardiographic and pathoanatomical characteristics of isolated left ventricular non-compaction: a step towards classification as a distinct cardiomyopathy. Heart. 86, 666-671

Jia, D., Jurkowska, R. Z., Zhang, X., Jeltsch, A. and Cheng, X. (2007). Structure of Dnmt3a bound to Dnmt3L suggests a model for de novo DNA methylation. Nature. 449, 248-251

Kagey, M. H., Newman, J. J., Bilodeau, S., Zhan, Y., Orlando, D. A., van Berkum, N. L., Ebmeier, C. C., Goossens, J., Rahl, P. B., Levine, S. S., Taatjes, D. J., Dekker, J. and Young, R. A. (2010). Mediator and cohesin connect gene expression and chromatin architecture. Nature. 467, 430-435

Kanduri, C., Pant, V., Loukinov, D., Pugacheva, E., Qi, C. F., Wolffe, A., Ohlsson, R. and Lobanenkov, V. V. (2000). Functional association of CTCF with the insulator upstream of the H19 gene is parent of origin-specific and methylation-sensitive. Current Biology. 10, 853-856

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

Kato, Y., Kaneda, M., Hata, K., Kumaki, K., Hisano, M., Kohara, Y., Okano, M., Li, E., Nozaki, M. and Sasaki, H. (2007). Role of the Dnmt3 family in de novo methylation of imprinted and repetitive sequences during male germ cell development in the mouse. Human Molecular Genetics. 16, 2272-2280

Kaumann, A. J. and Levy, F. O. (2006). 5-hydroxytryptamine receptors in the human cardiovascular system. Pharmacol Ther. 111, 674-706

Kelsey, G. and Bartolomei, M. S. (2012). Imprinted Genes … and the Number Is? PLoS Genetics. 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. and Curley, J. P. (2008). Epigenetics, brain evolution and behaviour. Frontiers in Neuroendocrinology. 29, 398-412

Killian, J. K., Nolan, C. M., Wylie, A. A., Li, T., Vu, T. H., Hoffman, A. R. and Jirtle, R. L. (2001). Divergent evolution in M6P/IGF2R imprinting from the Jurassic to the Quaternary. Human Molecular Genetics. 10, 1721-1728

Kim, J.-M. and Ogura, A. (2009). Changes in allele-specific association of histone modifications at the imprinting control regions during mouse preimplantation development. genesis. 47, 611-616

Kim, K. P., Thurston, A., Mummery, C., Ward-van Oostwaard, D., Priddle, H., Allegrucci, C., Denning, C. and Young, L. (2007). Gene-specific vulnerability to imprinting variability in human embryonic stem cell lines. Genome Research. 17, 1731-1742

Kim, T. H., Abdullaev, Z. K., Smith, A. D., Ching, K. A., Loukinov, D. I., Green, R. D., Zhang, M. Q., Lobanenkov, V. V. and Ren, B. (2007). Analysis of the Vertebrate Insulator Protein CTCF-Binding Sites in the Human Genome. Cell. 128, 1231-1245

Kirley, A., Hawi, Z., Daly, G., McCarron, M., Mullins, C., Millar, N., Waldman, I., Fitzgerald, M. and Gill, M. (2002). Dopaminergic system genes in ADHD: toward a biological hypothesis. Neuropsychopharmacology. 27, 607-619

Komiyama, M., Ito, K. and Shimada, Y. (1987). Origin and development of the epicardium in the mouse embryo. Anat Embryol (Berl). 176, 183-189

Kriaucionis, S. and Heintz, N. (2009). The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. Science. 324, 929-930

Kuzmin, I., Geil, L., Gibson, L., Cavinato, T., Loukinov, D., Lobanenkov, V. and Lerman, M. I. (2005). Transcriptional regulator CTCF controls human interleukin 1 receptor-associated kinase 2 promoter. J Mol Biol. 346, 411-422

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., Aumayr, K., Pasierbek, P. and Barlow, D. P. (2012). Airn transcriptional overlap, but not its lncRNA products, induces imprinted Igf2r silencing. Science. 338, 1469-1472

Lau, J. C. Y. (2004). Tissue-specific and imprinted epigenetic modifications of the human NDN gene. Nucleic Acids Research. 32, 3376-3382

Lefebvre, L., Viville, S., Barton, S. C., Ishino, F., Keverne, E. B. and Surani, M. A. (1998). Abnormal maternal behaviour and growth retardation associated with loss of the imprinted gene Mest. Nat Genet. 20, 163-169

Leighton, P. A., Saam, J. R., Ingram, R. S., Stewart, C. L. and Tilghman, S. M. (1995). An enhancer deletion affects both H19 and Igf2 expression. Genes Dev. 9, 2079-2089

Lemke, G. (1996). Neuregulins in development. Molecular and cellular neurosciences. 7, 247-262

Lewis, A., Mitsuya, K., Umlauf, D., Smith, P., Dean, W., Walter, J., Higgins, M., Feil, R. and Reik, W. (2004). Imprinting on distal chromosome 7 in the placenta involves repressive histone methylation independent of DNA methylation. Nature Genetics. 36, 1291-1295

Lewis, A. and Reik, W. (2006). How imprinting centres work. Cytogenetic and Genome Research. 113, 81-89

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

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, T. (2004). Activating and silencing histone modifications form independent allelic switch regions in the imprinted Gnas gene. Human Molecular Genetics. 13, 741-750

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. Developmental Cell. 15, 547-557

Lin, S., Ferguson-Smith, A. C., Schultz, R. M. and Bartolomei, M. S. (2011). Nonallelic Transcriptional Roles of CTCF and Cohesins at Imprinted Loci. Molecular and Cellular Biology. 31, 3094-3104

Ling, J. Q., Li, T., Hu, J. F., Vu, T. H., Chen, H. L., Qiu, X. W., Cherry, A. M. and Hoffman, A. R. (2006). CTCF mediates interchromosomal colocalization between Igf2/H19 and Wsb1/Nf1. Science. 312, 269-272

Lister, R., Pelizzola, M., Dowen, R. H., Hawkins, R. D., Hon, G., Tonti-Filippini, J., Nery, J. R., Lee, L., Ye, Z., Ngo, Q.-M., Edsall, L., Antosiewicz-Bourget, J., Stewart, R., Ruotti, V., Millar, A. H., Thomson, J. A., Ren, B. and Ecker, J. R. (2009). Human DNA methylomes at base resolution show widespread epigenomic differences. Nature. 462, 315-322

Livingstone, M. S. and Tempel, B. L. (1983). Genetic dissection of monoamine neurotransmitter synthesis in Drosophila. Nature. 303, 67-70

Lossie, A. C., Whitney, M. M., Amidon, D., Dong, H. J., Chen, P., Theriaque, D., Hutson, A., Nicholls, R. D., Zori, R. T., Williams, C. A. and Driscoll, D. J. (2001). Distinct phenotypes distinguish the molecular classes of Angelman syndrome. J Med Genet. 38, 834-845

Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F. and Richmond, T. J. (1997). Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature. 389, 251-260

Luger, K. and Richmond, T. J. (1998). The histone tails of the nucleosome. Current Opinion in Genetics & Development. 8, 140-146

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., Hahnemann, J. M. D., Kordonouri, O., Masoud, A. F., Oestergaard, E., Storr, J., Ellard, S., Hattersley, A. T., Robinson, D. O. and Temple, I. K. (2008). Hypomethylation of multiple imprinted loci in individuals with transient neonatal diabetes is associated with mutations in ZFP57. Nature Genetics. 40, 949-951

Magenis, R. E., Brown, M. G., Lacy, D. A., Budden, S. and LaFranchi, S. (1987). Is Angelman syndrome an alternate result of del(15)(q11q13)? American journal of medical genetics. 28, 829-838

Mancini-DiNardo, D. (2003). A differentially methylated region within the gene Kcnq1 functions as an imprinted promoter and silencer. Human Molecular Genetics. 12, 283-294

Mancini-DiNardo, D. (2006). Elongation of the Kcnq1ot1 transcript is required for genomic imprinting of neighboring genes. Genes & Development. 20, 1268-1282

Martin, D., Pantoja, C., Miñán, A. F., Valdes-Quezada, C., Moltó, E., Matesanz, F., Bogdanović, O., de la Calle-Mustienes, E., Domínguez, O., Taher, L., Furlan-Magaril, M., Alcina, A., Cañón, S., Fedetz, M., Blasco, M. A., Pereira, P. S., Ovcharenko, I., Recillas-Targa, F., Montoliu, L., Manzanares, M., Guigó, R., Serrano, M., Casares, F. and Gómez-Skarmeta, J. L. (2011). Genome-wide CTCF distribution in vertebrates defines equivalent sites that aid the identification of disease-associated genes. Nature Structural & Molecular Biology. 18, 708-714

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

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. Molecular and Cellular Biology. 28, 386-396

Mikkelsen, T. S., Ku, M., Jaffe, D. B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T.-K., Koche, R. P., Lee, W., Mendenhall, E., O'Donovan, A., Presser, A., Russ, C., Xie, X., Meissner, A., Wernig, M., Jaenisch, R., Nusbaum, C., Lander, E. S. and Bernstein, B. E. (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature. 448, 553-560

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

Morison, I. M., Ramsay, J. P. and Spencer, H. G. (2005). A census of mammalian imprinting. Trends in genetics : TIG. 21, 457-465

Mukhopadhyay, R. (2004). The Binding Sites for the Chromatin Insulator Protein CTCF Map to DNA Methylation-Free Domains Genome-Wide. Genome Research. 14, 1594-1602

Murrell, A. (2006). Genomic imprinting and cancer: from primordial germ cells to somatic cells. ScientificWorldJournal. 6, 1888-1910

Nativio, R., Wendt, K. S., Ito, Y., Huddleston, J. E., Uribe-Lewis, S., Woodfine, K., Krueger, C., Reik, W., Peters, J.-M. and Murrell, A. (2009). Cohesin Is Required for Higher-Order Chromatin Conformation at the Imprinted IGF2-H19 Locus. PLoS Genetics. 5, e1000739

Nebigil, C. G., Choi, D. S., Dierich, A., Hickel, P., Le Meur, M., Messaddeq, N., Launay, J. M. and Maroteaux, L. (2000). Serotonin 2B receptor is required for heart development. Proceedings of the National Academy of Sciences of the United States of America. 97, 9508-9513

Nebigil, C. G., Hickel, P., Messaddeq, N., Vonesch, J. L., Douchet, M. P., Monassier, L., Gyorgy, K., Matz, R., Andriantsitohaina, R., Manivet, P., Launay, J. M. and Maroteaux, L. (2001). Ablation of Serotonin 5-HT2B Receptors in Mice Leads to Abnormal Cardiac Structure and Function. Circulation. 103, 2973-2979

Nebigil, C. G. and Maroteaux, L. (2001). A novel role for serotonin in heart. Trends in cardiovascular medicine. 11, 329-335

Nix, D. A., Courdy, S. J. and Boucher, K. M. (2008). Empirical methods for controlling false positives and estimating confidence in ChIP-Seq peaks. BMC Bioinformatics. 9, 523

O'Neill, M. J., Ingram, R. S., Vrana, P. B. and Tilghman, S. M. (2000). Allelic expression of IGF2 in marsupials and birds. Development genes and evolution. 210, 18-20

Oechslin, E. and Jenni, R. (2011). Left ventricular non-compaction revisited: a distinct phenotype with genetic heterogeneity? European Heart Journal. 32, 1446-1456

Okano, M., Bell, D. W., Haber, D. A. and Li, E. (1999). DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell. 99, 247-257

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., Cheng, X. and Bestor, T. H. (2007). DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. Nature. 448, 714-717

Oswald, J., Engemann, S., Lane, N., Mayer, W., Olek, A., Fundele, R., Dean, W., Reik, W. and Walter, J. (2000). Active demethylation of the paternal genome in the mouse zygote. Current Biology. 10, 475-478

Otto, S. P. and Goldstein, D. B. (1992). Recombination and the evolution of diploidy. Genetics. 131, 745-751

Oudejans, C. (2001). Allelic IGF2R Repression Does Not Correlate with Expression of Antisense RNA in Human Extraembryonic Tissues. Genomics. 73, 331-337

Pauler, F. M., Koerner, M. V. and Barlow, D. P. (2007). Silencing by imprinted noncoding RNAs: is transcription the answer? Trends in Genetics. 23, 284-292

Pauler, F. M., Barlow, D. P. and Hudson, Q. J. (2012). Mechanisms of long range silencing by imprinted macro non-coding RNAs. Current Opinion in Genetics & Development. 22, 283-289

Perez-Pomares, J. M., Macias, D., Garcia-Garrido, L. and Munoz-Chapuli, R. (1997). Contribution of the primitive epicardium to the subepicardial mesenchyme in hamster and chick embryos. Dev Dyn. 210, 96-105

Perrot, V., Richerd, S. and Valéro, M. (1991). Transition from haploidy to diploidy. Nature. 351, 315-317

Piedrahita, J. A. (2011). The role of imprinted genes in fetal growth abnormalities. Birth Defects Research Part A: Clinical and Molecular Teratology. 91, 682-692

Plagge, A., Isles, A. R., Gordon, E., Humby, T., Dean, W., Gritsch, S., Fischer-Colbrie, R., Wilkinson, L. S. and Kelsey, G. (2005). Imprinted Nesp55 influences behavioral reactivity to novel environments. Mol Cell Biol. 25, 3019-3026

Prickett, A. R. and Oakey, R. J. (2012). A survey of tissue-specific genomic imprinting in mammals. Molecular Genetics and 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. and Trono, D. (2011). In Embryonic Stem Cells, ZFP57/KAP1 Recognize a Methylated Hexanucleotide to Affect Chromatin and DNA Methylation of Imprinting Control Regions. Molecular cell. 44, 361-372

Quinlan, A. R. and Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 26, 841-842

Recillas-Targa, F., De La Rosa-Velazquez, I. A., Soto-Reyes, E. and Benitez-Bribiesca, L. (2006). Epigenetic boundaries of tumour suppressor gene promoters: the CTCF connection and its role in carcinogenesis. J Cell Mol Med. 10, 554-568

Reik, W. and Walter, J. (2001). Genomic imprinting: parental influence on the genome. Nature Reviews Genetics. 2, 21-32

Reinartz, J., Bruyns, E., Lin, J.-Z., Burcham, T., Brenner, S., Bowen, B., Kramer, M. and Woychik, R. (2002). Massively parallel signature sequencing (MPSS) as a tool for in-depth quantitative gene expression profiling in all organisms. Briefings in functional genomics & proteomics. 1, 95-104

Ren, L., Wang, Y., Shi, M., Wang, X., Yang, Z. and Zhao, Z. (2012). CTCF mediates the cell-type specific spatial organization of the Kcnq5 locus and the local gene regulation. PLoS ONE. 7, e31416

Riesewijk, A. M., Xu, Y. Q., Schepens, M. T., Mariman, E. M., Polychronakos, C., Ropers, H. H. and Kalscheuer, V. M. (1998). Absence of an obvious molecular imprinting mechanism in a human fetus with monoallelic IGF2R expression. Biochemical and Biophysical Research Communications. 245, 272-277

Robinson, W. P. (2000). Mechanisms leading to uniparental disomy and their clinical consequences. BioEssays. 22, 452-459

Rouzaud-Laborde, C., Hanoun, N., Baysal, I., Rech, J.-S., Mias, C., Calise, D., Sicard, P., Frugier, C., Seguelas, M.-H., Parini, A. and Pizzinat, N. (2012). Role of Endothelial AADC in Cardiac Synthesis of Serotonin and Nitrates Accumulation. PLoS ONE. 7, e34893

Rugg-Gunn, P. J., Ferguson-Smith, A. C. and Pedersen, R. A. (2007). Status of genomic imprinting in human embryonic stem cells as revealed by a large cohort of independently derived and maintained lines. Human Molecular Genetics. 16, R243-R251

Sakamoto, A. (2004). Tissue-specific imprinting of the G protein Gs is associated with tissue-specific differences in histone methylation. Human Molecular Genetics. 13, 819-828

Santoro, F. and Barlow, D. P. (2011). Developmental control of imprinted expression by macro non-coding RNAs. Seminars in Cell and Developmental Biology. 22, 328-335

Sanz, L. A., Chamberlain, S., Sabourin, J.-C., Henckel, A., Magnuson, T., Hugnot, J.-P., Feil, R. and Arnaud, P. (2008). A mono-allelic bivalent chromatin domain controls tissue-specific imprinting at Grb10. The EMBO Journal. 27, 2523-2532

Sasaki, H., Jones, P. A., Chaillet, J. R., Ferguson-Smith, A. C., Barton, S. C., Reik, W. and Surani, M. A. (1992). Parental imprinting: potentially active chromatin of the repressed maternal allele of the mouse insulin-like growth factor II (Igf2) gene. Genes & Development. 6, 1843-1856

Savolainen, S. M., Foley, J. F. and Elmore, S. A. (2009). Histology atlas of the developing mouse heart with emphasis on E11.5 to E18.5. Toxicologic Pathology. 37, 395-414

Schmidt, D., Schwalie, P. C., Ross-Innes, C. S., Hurtado, A., Brown, G. D., Carroll, J. S., Flicek, P. and Odom, D. T. (2010). A CTCF-independent role for cohesin in tissue-specific transcription. Genome Research. 20, 578-588

Schulz, R., Woodfine, K., Menheniott, T. R., Bourc'his, D., Bestor, T. and Oakey, R. J. (2008). WAMIDEX: a web atlas of murine genomic imprinting and differential expression. Epigenetics : official journal of the DNA Methylation Society. 3, 89-96

Searle, A. G. and Beechey, C. V. (1978). Complementation studies with mouse translocations. Cytogenetics and cell genetics. 20, 282-303

Searle, A. G. and Beechey, C. V. (1990). Genome imprinting phenomena on mouse chromosome 7. Genet Res. 56, 237-244

Sedmera, D., Pexieder, T., Vuillemin, M., Thompson, R. P. and Anderson, R. H. (2000). Developmental patterning of the myocardium. The Anatomical record. 258, 319-337

Sexton, T., Kurukuti, S., Mitchell, J. A., Umlauf, D., Nagano, T. and Fraser, P. (2012). Sensitive detection of chromatin coassociations using enhanced chromosome conformation capture on chip. Nat Protoc. 7, 1335-1350

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. Human Molecular Genetics. 18, 1424-1438

Singh, P., Wu, X., Lee, D.-H., Li, A. X., Rauch, T. A., Pfeifer, G. P., Mann, J. R. and Szabo, P. E. (2011). Chromosome-Wide Analysis of Parental Allele-Specific Chromatin and DNA Methylation. Molecular and Cellular Biology. 31, 1757-1770

Sleutels, F., Zwart, R. and Barlow, D. P. (2002). The non-coding Air RNA is required for silencing autosomal imprinted genes. Nature. 415, 810-813

Smallwood, S. A., Tomizawa, S.-i., 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. Nature Genetics. 1-5

Stedman, W., Kang, H., Lin, S., Kissil, J. L., Bartolomei, M. S. and Lieberman, P. M. (2008). Cohesins localize with CTCF at the KSHV latency control region and at cellular c-myc and H19/Igf2 insulators. The EMBO Journal. 27, 654-666

Stöger, R., Kubicka, P., Liu, C. G., Kafri, T., Razin, A., Cedar, H. and Barlow, D. P. (1993). Maternal-specific methylation of the imprinted mouse Igf2r locus identifies the expressed locus as carrying the imprinting signal. Cell. 73, 61-71

Suetake, I. (2004). DNMT3L Stimulates the DNA Methylation Activity of Dnmt3a and Dnmt3b through a Direct Interaction. Journal of Biological Chemistry. 279, 27816-27823

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

Surani, M. A., Barton, S. C. and Norris, M. L. (1986). Nuclear transplantation in the mouse: heritable differences between parental genomes after activation of the embryonic genome. Cell. 45, 127-136

Taatjes, D. J. (2010). The human Mediator complex: a versatile, genome-wide regulator of transcription. Trends Biochem Sci. 35, 315-322

Tahiliani, M., Koh, K. P., Shen, Y., Pastor, W. A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L. M., Liu, D. R., Aravind, L. and Rao, A. (2009). Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science. 324, 930-935

Takai, D. and Jones, P. A. (2002). Comprehensive analysis of CpG islands in human chromosomes 21 and 22. Proc Natl Acad Sci U S A. 99, 3740-3745

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 & Development. 12, 3693-3702

Tremblay, K. D., Saam, J. R., Ingram, R. S., Tilghman, S. M. and Bartolomei, M. S. (1995). A paternal-specific methylation imprint marks the alleles of the mouse H19 gene. Nature Genetics. 9, 407-413

Umlauf, D., Goto, Y., Cao, R., Cerqueira, F., Wagschal, A., Zhang, Y. and Feil, R. (2004). Imprinting along the Kcnq1 domain on mouse chromosome 7 involves repressive histone methylation and recruitment of Polycomb group complexes. Nature Genetics. 36, 1296-1300

Unoki, M., Nishidate, T. and Nakamura, Y. (2004). ICBP90, an E2F-1 target, recruits HDAC1 and binds to methyl-CpG through its SRA domain. Oncogene. 23, 7601-7610

Uribe-Lewis, S., Woodfine, K., Stojic, L. and Murrell, A. (2011). Molecular mechanisms of genomic imprinting and clinical implications for cancer. Expert Reviews in Molecular Medicine. 13,

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

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, e3839

Watanabe, T., Tomizawa, S. i., Mitsuya, K., Totoki, Y., Yamamoto, Y., Kuramochi-Miyagawa, S., Iida, N., Hoki, Y., Murphy, P. J., Toyoda, A., Gotoh, K., Hiura, H., Arima, T., Fujiyama, A., Sado, T., Shibata, T., Nakano, T., Lin, H., Ichiyanagi, K., Soloway, P. D. and Sasaki, H. (2011). Role for piRNAs and Noncoding RNA in de Novo DNA Methylation of the Imprinted Mouse Rasgrf1 Locus. Science. 332, 848-852

Weber, M., Hellmann, I., Stadler, M. B., Ramos, L., Pääbo, S., Rebhan, M. and Schübeler, D. (2007). Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. Nature Genetics. 39, 457-466

Wendt, K. S., Yoshida, K., Itoh, T., Bando, M., Koch, B., Schirghuber, E., Tsutsumi, S., Nagae, G., Ishihara, K., Mishiro, T., Yahata, K., Imamoto, F., Aburatani, H., Nakao, M., Imamoto, N., Maeshima, K., Shirahige, K. and Peters, J. M. (2008). Cohesin mediates transcriptional insulation by CCCTC-binding factor. Nature. 451, 796-801

Weninger, W. J. and Mohun, T. (2002). Phenotyping transgenic embryos: a rapid 3-D screening method based on episcopic fluorescence image capturing. Nature Genetics. 30, 59-65

Wessels, A. and Sedmera, D. (2003). Developmental anatomy of the heart: a tale of mice and man. Physiological Genomics. 15, 165-176

Williams, C. A., Beaudet, A. L., Clayton-Smith, J., Knoll, J. H., Kyllerman, M., Laan, L. A., Magenis, R. E., Moncla, A., Schinzel, A. A., Summers, J. A. and Wagstaff, J. (2006). Angelman syndrome 2005: updated consensus for diagnostic criteria. Am J Med Genet A. 140, 413-418

Wood, A. J., Roberts, R. G., Monk, D., Moore, G. E., Schulz, R. and Oakey, R. J. (2007). A Screen for Retrotransposed Imprinted Genes Reveals an Association between X Chromosome Homology and Maternal Germ-Line Methylation. PLoS Genetics. 3, e20

Wossidlo, M., Nakamura, T., Lepikhov, K., Marques, C. J., Zakhartchenko, V., Boiani, M., Arand, J., Nakano, T., Reik, W. and Walter, J. o. r. (2011). 5-Hydroxymethylcytosine in the mammalian zygote is linked with epigenetic reprogramming. Nature Communications. 2, 241-248

Xie, S., Wang, Z., Okano, M., Nogami, M., Li, Y., He, W. W., Okumura, K. and Li, E. (1999). Cloning, expression and chromosome locations of the human DNMT3 gene family. Gene. 236, 87-95

Xie, W., Barr, C. L., Kim, A., Yue, F., Lee, A. Y., Eubanks, J., Dempster, E. L. and Ren, B. (2012). Base-Resolution Analyses of Sequence and Parent-of-Origin Dependent DNA Methylation in the Mouse Genome. Cell. 148, 816-831

Yalcin, B., Adams, D. J., Flint, J. and Keane, T. M. (2012). Next-generation sequencing of experimental mouse strains. Mamm Genome. 23, 490-498

Yang, Y. (2003). The Histone Code Regulating Expression of the Imprinted Mouse Igf2r Gene. Endocrinology. 144, 5658-5670

Yokomine, T., Kuroiwa, A., Tanaka, K., Tsudzuki, M., Matsuda, Y. and Sasaki, H. (2001). Sequence polymorphisms, allelic expression status and chromosome locations of the chicken IGF2 and MPR1 genes. Cytogenetics and cell genetics. 93, 109-113

Yokomine, T. (2005). Structural and functional analysis of a 0.5-Mb chicken region orthologous to the imprinted mammalian Ascl2/Mash2-Igf2-H19 region. Genome Research. 15, 154-165

Yoon, B., Herman, H., Hu, B., Park, Y. J., Lindroth, A., Bell, A., West, A. G., Chang, Y., Stablewski, A., Piel, J. C., Loukinov, D. I., Lobanenkov, V. V. and Soloway, P. D. (2005). Rasgrf1 Imprinting Is Regulated by a CTCF-Dependent Methylation-Sensitive Enhancer Blocker. Molecular and Cellular Biology. 25, 11184-11190

Yusufzai, T. M., Tagami, H., Nakatani, Y. and Felsenfeld, G. (2004). CTCF tethers an insulator to subnuclear sites, suggesting shared insulator mechanisms across species. Mol Cell. 13, 291-298

Zhang, M.-Z., Yao, B., Wang, S., Fan, X., Wu, G., Yang, H., Yin, H., Yang, S. and Harris, R. C. (2011). Intrarenal dopamine deficiency leads to hypertension and decreased longevity in mice. The Journal of clinical investigation. 121, 2845-2854

Zuo, X., Sheng, J., Lau, H. T., McDonald, C. M., Andrade, M., Cullen, D. E., Bell, F. T., Iacovino, M., Kyba, M., Xu, G. and Li, X. (2012). Zinc Finger Protein ZFP57 Requires Its Co-factor to Recruit DNA Methyltransferases and Maintains DNA Methylation Imprint in Embryonic Stem Cells via Its Transcriptional Repression Domain. Journal of Biological Chemistry. 287, 2107-2118

Appendices

Appendix 2.1 Chromatin Immunoprecipitation antibodies

Antigen	Species	Concentration	Manufacturer	Catalogue #	Lot#
IgG	Rabbit	1/250	Abcam	ab17890	674863
Rad21	Rabbit	1/200	Abcam	ab992	818494
CTCF	Rabbit	1/70	Millipore	07-729	1772428

Appendix 2.2 Taqman™ assay for illumina library quantification

Assay	Type	Sequence
Library	Primers	CCACCGAGATCTACACTCTTTCC
Quantification		GTTCGTCTTCTGCCGTATGCT
	FAM	ACACGACGCTCTTCC

Appendix 2.3 Primers for genotyping *Ddc* knockout mice

Assay	Sequence	Tm	Size (bp)
Wildtype	AAGCACCCCCTGAAAGCTGTACT	59	214
	AATACTGGTCTTGGAGGCCTTGGA		
Knockout	AAATGGCGTTACTTAAGCTAGCTTGC	59	165
	AATACTGGTCTTGGAGGCCTTGGA		

Appendix 2.4 Primers for Bisulphite PCR

Assay	Sequence	Tm	Size (bp)
CGI2 Grb10	GTGAGGTTTAAAAGATGATTAGGT	47	423
	AATAATTTAAATAATAAATAATTTTTA		
CGI3 Grb10	GGATTGATTGTATTTTAAGTATA	48	559
	ATTTCACACTCTCCTCTAAC		
Region 1	GTGGTTGAGAAATTGAGAAATTT	50	308
AK0066690	CTCTTATAAAAAAAAAACTACCA		
Region 2	TTTATATAGGGTAAGAAATTG	46	450
AK0066690	TACCATTAAATTAAACCACC		
Magel2	GTGTTTGTTGAGAGTTGTTGAGAGA	60	373
Promoter	ACCAAACAACCATAAAAACCTACAA		

Appendix 2.5 RT-PCR primers for *Grb10* transcript detection

Assay	Sequence	Tm	Size (bp)
Maternal *Grb10* (exon1A to exon 2)	CTTTGCCTTCGGTTTTTCTC CTTCTTTGTTGTGGCGAAAA	58	154
Maternal *Grb10* (exon1B to exon 2)	TCCAGAGCCCTTTTTCTGAG CTTCTTTGTTGTGGCGAAAA	58	272
Paternal *Grb10* (Exon1A to LacZ)	CTTTGCCTTCGGTTTTTCTC GCGATCTGCGTTCTTCTTCT	58	≅290
Paternal *Grb10* (Exon1B to LacZ)	TCCAGAGCCCTTTTTCTGAG GCGATCTGCGTTCTTCTTCT	58	≅290
Actin	TCGCCATGGATGACGATA TCTCCGGAGTCCATCACAAT	59	475

Appendix 2.6 RT-PCR primers for *AK0066690* transcript

Assay	Sequence	Tm	Size (bp)
AK0066690 Transcript	TTCAGCCAAGAGTGCTTAGG GCTGCTGCATGCTTATTTGT	58	184
Actin	TCGCCATGGATGACGATA TCTCCGGAGTCCATCACAAT	59	475

Appendix 2.7 RT-PCR primers for *Ddc* knockout transcript analysis

Assay	Sequence	Tm	Size (bp)
Ddc_canonical exon 1 – exon 3	AGTTGTGTCGCCACCTCCT GAGAAACCAATGCAGCCAAT	60	
Ddc_canonical exon 1 - exon 4	AGTTGTGTCGCCACCTCCT CAGCTCTTCCAGCCAAAAAG	60	
Ddc_exon1a exon 1a – exon 3	TGTCACCAAGGAGAGAGAGAGA GAGAAACCAATGCAGCCAAT	59	
Ddc_exon1a exon 1a - exon 4	TGTCACCAAGGAGAGAGAGAGA CAGCTCTTCCAGCCAAAAAG	59	
Ddc Exon1 to vector insert (1)	AGTTGTGTCGCCACCTCCT TAGCCGAATAGCCTCTCCAC	59	
Ddc Exon1 to vector insert (2)	AGTTGTGTCGCCACCTCCT TGTCTGTTGTGCCCAGTCAT	58	
Ddc Exon1a to vector insert (1)	TGTCACCAAGGAGAGAGAGAGA TAGCCGAATAGCCTCTCCAC	59	
Ddc Exon1a to vector insert (2)	TGTCACCAAGGAGAGAGAGAGA TGTCTGTTGTGCCCAGTCAT	58	
Ddc downstream	ATTGGCTGCATTGGTTTCTC TCCACCAATTAACCCAGCTC	60	
Actin	TCGCCATGGATGACGATA TCTCCGGAGTCCATCACAAT	59	

Appendix 2.8 Taqman™ assays for CTCF ChIP

Assay	Type	Sequence
cMyc	Primers	GTGAGCGGACGGTTGGA
		TCCAGAGCTGCCTTCTTAGGT
	FAM	CTGCACACACGGCTCT
DMDup	Primers	AGTGGTGTCTGTAATCTGGAGAGAA
		TGAGCCTGCATCTTATTGAAGTGAT
	FAM	CCATGTCCTATATAAGTAACACTC
Grb10-1	Primers	GGGTCCTTGTGCCCTCTATTG
		CCCCTACACAAGAATGAAGATTAAAAACC
	FAM	ACGTGTCATGCTCCCC
Grb10-2	Primers	GCCGGCTAGCACAGACTT
(DMR)		CGGGAGCTGTCCACTGG
	FAM	CCCGGTAGCGCGCACG
Grb10-4	Primers	CCAGCCTCTGAGAAAATGAAATGG
		ACAGCAAGCCCAGGGAAAA
	FAM	TCTGCCAGCACGAACC
Ddc-1	Primers	CCACAGAAGCAATTGTCTCAAGTCT
		GGCTTGAGCTAATTGAAGAGCATTT
	FAM	CCCACCAGAGGGCACC
Ddc-2	Primers	TGAAAACCTGAGGTCTAGCATCAC
		GCCTACTGCAGTTGCTCACT
	FAM	ACTGCCACCAGAGGTC

Appendix 2.9 *Ddc* expression Taqman™ assay*

Assay	Assay ID / Catalogue Number
Ddc	Mm01192099_m1
Actin	cat # 4352933E

*Primer and probe sequences and genomic location are not provided by Applied Biosystems™ however the *Ddc* probe detects both *Ddc_canonical* and *Ddc_exon1a* transcripts.

Appendix 2.10 Antibodies used in immunostaining assays

Antigen	Species	Concentration	Manufacturer	Catalogue #	Conjugate
Ddc	Rabbit	1/500	Abcam	3905	n/a
ANF	Goat	1/100	Santa Cruz	18811	n/a
MF-20	Mouse	1/50	Monoclonal from Professor Baldwin, Vanderbilt	n/a	n/a
Goat IgG	Donkey	1/300	Invitrogen	A-21432	Alexa Fluor 555
Rabbit IgG	Donkey	1/300	Invitrogen	A-31573	Alexa Fluor 647
Mouse IgG	Goat	1/400	Invitrogen	A-11001	Alexa Fluor 488
Rabbit IgG	Goat	1/400	Invitrogen	A-21429	Alexa Fluor 555

Appendix 2.11 Antibodies used in western blot analysis

Antigen	Species	Concentration	Manufacturer	Catalogue #	Conjugate
Ddc	Rabbit	1/2000	Abcam	3905	n/a
Beta-actin					n/a
Rabbit IgG			DAKO		Peroxidase

Appendix 5 Microarray data analysis

Appendix 5.1 Experimental design

Gene expression data from the four genotypes was used to construct a linear model, Figure 5.1a shows the workflow used to examine differential gene expression between *Ddc*$^{-/+}$ and *Ddc*$^{+/-}$ heart within the linear model. Initially the microarray data were quality controlled and pre-processed using the NEQC analysis package, and subsequent modelling of the data was performed using a linear model package (LIMMA), both were implemented in R. In order to assess the variation in the experiment introduced by the fact that the 12 samples were run in two batches, a principal component analysis (PCA) was performed on the first 10 000 probes from the 12 samples (Figure 5.2a). PCA revealed segregation between the two batches; therefore 'batch effect' was added as an additional component in the linear model. This meant that five variables were used in the LIMMA analysis, the four separate genotypes and the 'batch effect' (Figure 5.1a).

Appendix 5.2 Differential gene expression analysis

Within the multi-component linear model, differential gene expression was tested in heart. Comparisons were made between *Ddc*$^{+/-}$ mice that have little or no expression of *Ddc_exon1a*, because the predominately active paternal allele is knocked out, compared to mice which have levels of *Ddc_exon1a* expression comparable to the wildtype because the maternally silence copy of *Ddc* inherits the knockout allele. Differential expression analysis between *Ddc*$^{+/-}$ and *Ddc*$^{-/+}$ within the linear model revealed 33 probes with differential gene expression, following correction for multiple testing with a cutoff of $q < 0.01$. Due to the small number of genes in this list gene ontology or biochemical pathway analysis could not be conducted, therefore the gene list was extended to include all genes that showed an absolute fold change of $> \text{Log}_2(0.4)$ (representing approximately 30 % increase or decrease in gene expression) and that had

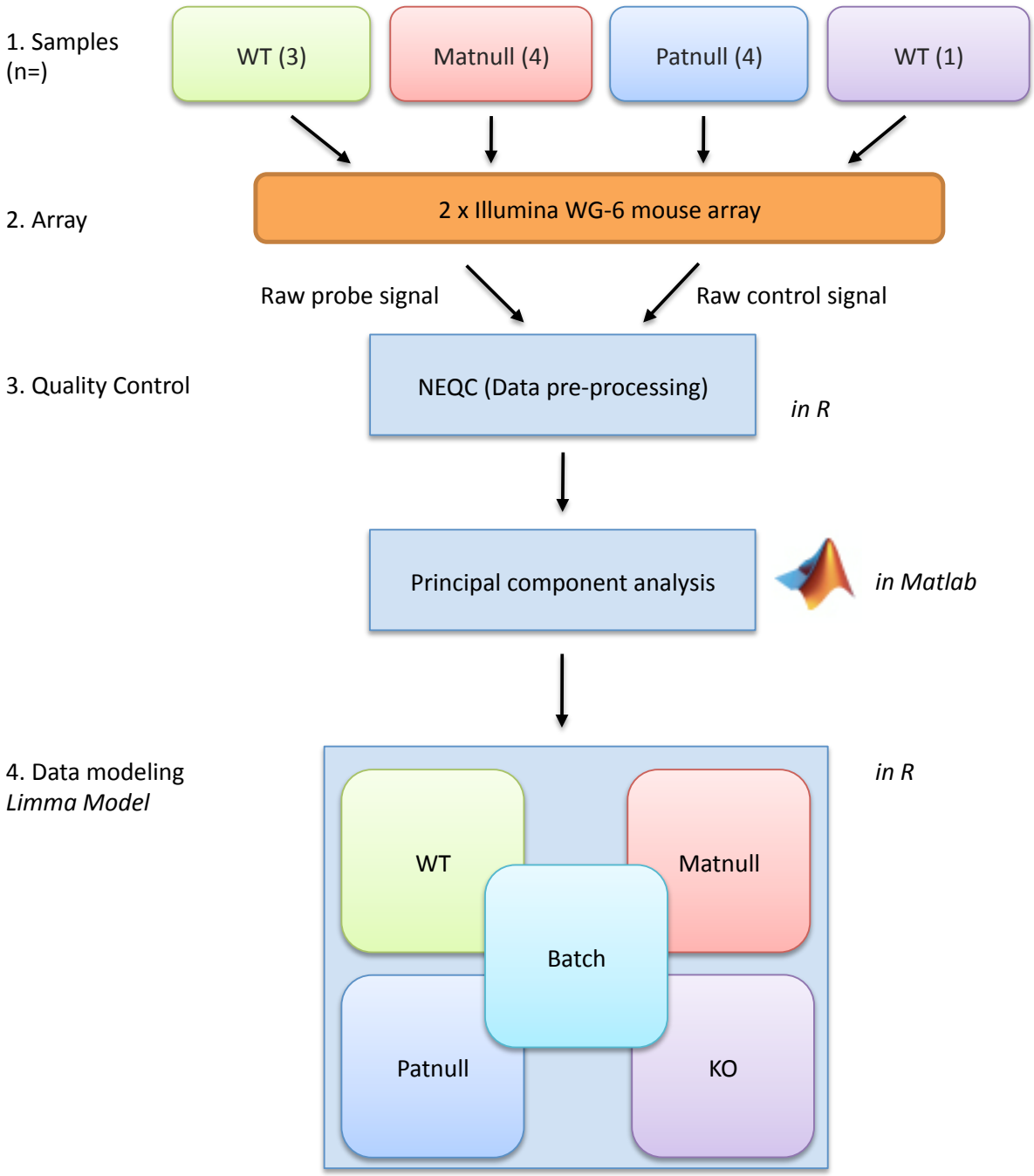


Figure 5.1a Microarray analysis workflow. Step 1 and 2: Samples are prepared and analyzed on the WG-6 mouse expression array. Step 3: Raw signals are normalized and quality controlled, batch effect is checked using principle component analysis. Step 4: Linear Model (Limma) is built using five components (four genotypes plus batch effect). Within this model comparisons between genotypes can then be made.

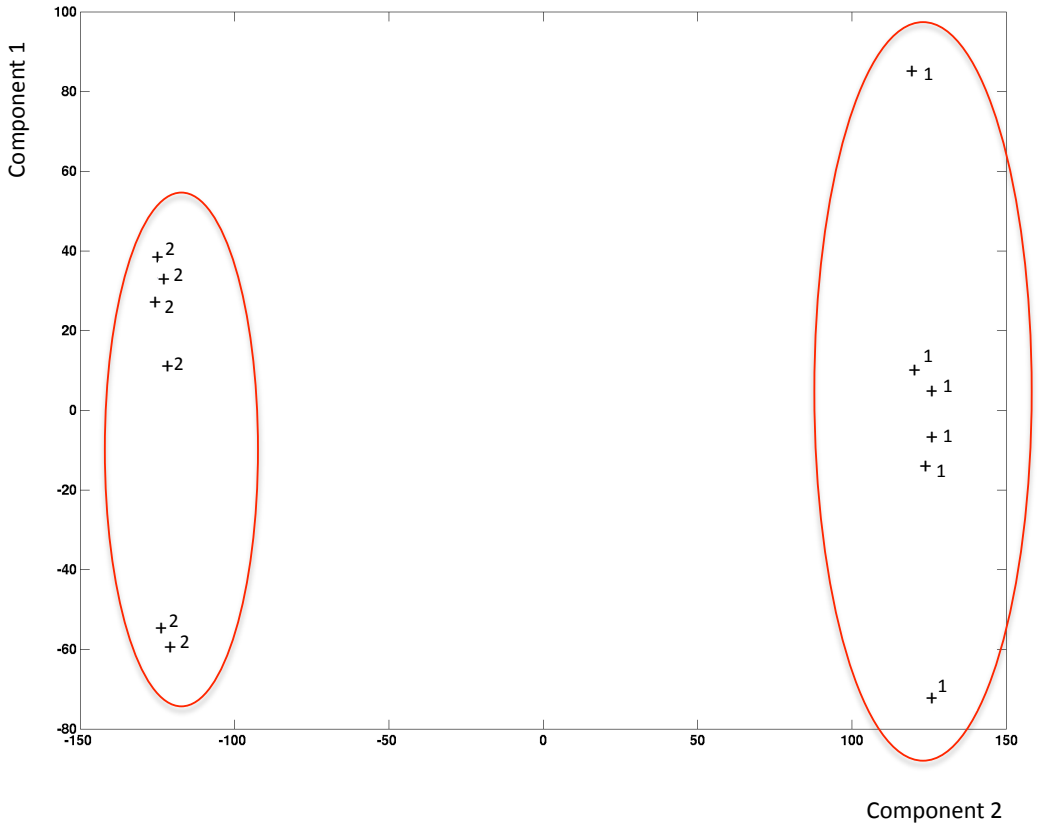


Figure 5.2a Principle component analysis (PCA) of microarray data. PCA reveals segregation based on batch. Arrays were run in two batches. Samples are labeled according to batch. All batch 1 are grouped on the right and batch 2 are grouped on the left. For this reason 'batch effect' was added as a variable in subsequent analysis.

a p-value of $p < 0.001$, before accounting for multiple testing. This stringency of differential expression yielded a list of 323 genes, which were subsequently used to analyse biological processes that may be affected by ablation of *Ddc_exon1a* expression on the paternal allele in the developing heart. The three *Ddc* probes showed the highest fold change in expression between $Ddc^{+/-}$ and $Ddc^{-/+}$ and represented the most statistically significant changes in expression, changes observed in other genes were less apparent. This is indicative of a gene that does not directly influence gene expression by acting as a transcription factor, but instead is involved in a biochemical pathway, and therefore elicits an effect via a biochemical feedback mechanism.

Appendix 5.3 Ontological analysis

The 323 genes with differential expression were analysed for overrepresented gene ontologies (GO) using three separate methods:

1) DAVID analysis
2) Ingenuity analysis
3) Mammalian phenotype browser analysis

The DAVID bioinformatic analysis tool is a bioinformatic knowledge base that utilizes a high throughput, data-mining environment to extract biological meaning from large gene lists, specifically developed for use on gene lists extracted from genome-wide studies (Huang, et al., 2008). Analysis of the 323 differentially expressed genes using DAVID revealed 22 over-represented gene ontologies with a false discovery rate (FDR) of <20 % (Table 5.1a), only one category 'morphogenesis of a branching structure' surviving multiple testing correction ($q < 0.05$). Combining these results with the analysis of *Ddc_exon1a* expression patterns in the developing mouse heart points to a potential role for Ddc in trabecular morphogenesis, as trabeculae forms a branching component of the developing heart.

Term	Count	FDR (%)
morphogenesis of a branching structure	10	0.060
striated muscle tissue development	8	2.197
actin cytoskeleton organization	9	2.312
porphyrin biosynthetic process	4	2.476
tetrapyrrole biosynthetic process	4	2.476
tube morphogenesis	9	2.891
muscle tissue development	8	3.239
actin filament-based process	9	3.454
skeletal muscle tissue development	6	3.926
skeletal muscle organ development	6	4.416
regulation of branching involved in prostate gland morphogenesis	3	5.396
erythrocyte differentiation	5	5.505
cytoskeleton organization	12	5.791
regulation of morphogenesis of a branching structure	4	6.403
erythrocyte homeostasis	5	6.815
bone development	7	7.007
porphyrin metabolic process	4	7.143
tetrapyrrole metabolic process	4	7.143
branching morphogenesis of a tube	6	11.299
morphogenesis of an epithelium	8	11.587
muscle organ development	8	12.650
positive regulation of cell division	4	16.950

Table 5.1a DAVID ontological analysis. Ontological analysis was performed on 323 genes differentially expression in *Ddc*$^{-/+}$ compared to *Ddc*$^{+/-}$ using the DAVID bioinformatic analysis tool. Count shows the number of genes in the listed ontology. Results are ranked by statistical significance based on the false discovery rate (FDR) with the most statistically significant ontology first. Genes with an FDR < 5% are shown with a white background and are the most statistically interesting.

A more in depth look at the over-represented ontologies with an FDR <20 % reveal that 16 of the 22 GO categories are associated with branching morphogenesis, muscle and organ development and cell re-organization and growth, all of which are concordant with the hypothesis that Ddc plays a role in heart muscle formation. In order to look further at these ontologies specifically, the genes involved were extracted and analysed in more detail (Table 5.2a) including whether expression increased or decreased in $Ddc^{+/-}$ hearts that lack *Ddc* expression, compared to $Ddc^{-/+}$ heart. These analyses revealed that 71 % of genes differentially expressed in these ontological groups are down-regulated in response to a reduction in the level of *Ddc* transcript, with the remaining 29 % up-regulated, suggesting Ddc may positively regulate these ontological functions (Figure 5.3a). The number of genes increased or decreased was considered for each individual ontology associated with branching morphogenesis, muscle and organ development and cell re-organization and growth, to assess whether function was down-regulated or mis-regulated for each ontology as defined by the proportion of genes in the pathway showed increased or decreased expression (Table 5.3a). These analyses suggest that Ddc functions to increase branching, muscle tissue and organ development and cell proliferation, and functions to regulate muscle tissue development and cytoskeletal organization in the developing heart.

Additional bioinformatic tools were used to extract overrepresented functions and systems that are influenced by genes mis-regulated in the $Ddc^{+/-}$ heart. Data were analyzed through the use of ingenuity pathway analysis (IPA) (Ingenuity® Systems, www.ingenuity.com), a commercially available tool that mines a repository of biological interactions and functional annotations taken from millions of individually modelled relationships between proteins, RNAs, genes, isoforms, metabolites, complexes, cells, tissues, drugs, and diseases, in a fashion similar to the DAVID tool. The ingenuity licence was used courtesy of Vanderbilt University Medical Center, TN, USA. Analyses revealed similar results to

Gene Name	Description	$Ddc^{+/-}$ vs. $Ddc^{-/+}$
Acta1	Actin, alpha 1, skeletal muscle	Down
Bmp4	Bone morphogenetic protein 4	Down
Btc	Betacellulin, epidermal growth factor family member	Up
Cacna2d2	Calcium channel, voltage-dependent, alpha 2/delta subunit 2	Down
Cnn2	Calponin 2	Down
Col2a1	Collagen, type II, alpha 1	Down
Cxcl12	Chemokine (C-X-C motif) ligand 12	Down
Epb4.1	Erythrocyte protein band 4.1	Up
Fgfr2	Fibroblast growth factor receptor 2	Down
Fhl2	Four and a half LIM domains 2	Up
Flnb	Filamin, beta	Up
Ftcd	Formiminotransferase cyclodeaminase	Up
Igfbp3	Insulin-like growth factor binding protein 3	Down
Mdk	Midkine	Down
Meox2	Mesenchyme homeobox 2	Up
Myh6	Myosin, heavy polypeptide 6, cardiac muscle, alpha	Up
Papss2	3-phosphoadenosine 5-phosphosulfate synthase 2	Down
Pdgfc	Platelet-derived growth factor, C polypeptide	Down
Pdlim3	PDZ and LIM domain 3	Down
Pitx2	Paired-like homeodomain transcription factor 2	Down
Prkcz	Protein kinase C, zeta	Down
Ptn	Pleiotrophin	Down
Rhou	Ras homolog gene family, member U	Down
Rspo3	R-spondin 3 homolog (Xenopus laevis)	Down
S100a9	S100 calcium binding protein A9 (calgranulin B)	Up
Sema3a	Sema domain, immunoglobulin domain (Ig), short basic domain, secreted, (semaphorin) 3A	Down
Sfrp1	Secreted frizzled-related protein 1	Down
Sgcg	Sarcoglycan, gamma (dystrophin-associated glycoprotein)	Up
Shroom1	Shroom family member 1	Down
Sorbs1	Sorbin and SH3 domain containing 1	Up
Tcf21	Transcription factor 21	Down
Tgif1	TGFB-induced factor homeobox 1	Down
Thsd2	Thrombospondin, type I, domain containing 2	Down
Twist1	Twist homolog 1 (Drosophila)	Down
Wnt5a	Wingless-related MMTV integration site 5A	Down

Table 5.2a Gene expression differences in *Ddc* knockout heart. Genes from overrepresented ontologies (FDR < 20%) that are associated with branching morphogenesis, muscle and organ development and cell re-organization and growth are listed. Genes down-regulated in the $Ddc^{+/-}$ when compared to $Ddc^{-/+}$ are shown in green and those up-regulated are shown in red.

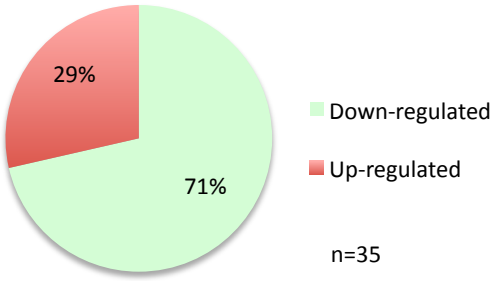


Figure 5.3a Gene expression differences in *Ddc* knockout heart. 71 % of genes associated with branching morphogenesis, muscle and organ development and cell re-organization and growth were down-regulated in hearts lacking *Ddc* expression.

GO groups with >80 % (Down-regulated in $Ddc^{+/-}$)	Number of genes in GO	Number down-regulated	Number up-regulated
Morphogenesis of a branching structure	10	10	0
tube morphogenesis	9	9	0
skeletal muscle tissue development	5	4	1
skeletal muscle organ development	5	4	1
regulation of branching involved in prostate gland morphogenesis	3	3	0
regulation of morphogenesis of a branching structure	3	3	0
bone development	7	6	1
branching morphogenesis of a tube	6	6	0
morphogenesis of an epithelium	8	8	0
positive regulation of cell division	5	4	1
GO groups with 20 < x < 80 % (Mis-regulated in $Ddc^{+/-}$)	Number of genes in GO	Number down-regulated	Number up-regulated
striated muscle tissue development	8	4	4
actin cytoskeleton organization	9	5	4
muscle tissue development	7	3	4
actin filament-based process	10	6	4
cytoskeleton organization	12	6	6
muscle organ development	8	4	4

Table 5.3a Branching morphogenesis, muscle and organ development and cell re-organization and growth ontologies. The number of genes up and down-regulated in the *$Ddc^{+/-}$* when compared to *$Ddc^{-/+}$* are shown for each ontology. Where more than 80% of genes are down-regulated Ddc is proposed to positively regulate these ontological processes. When the genes in the ontology are both up and down regulated (ratio is between 20-80%) Ddc is proposed to play a regulatory role in this ontological process.

that seen with the DAVID bioinformatic analysis, however the Ingenuity platform provides multiple outputs, the two most relevant being molecular and cellular function, and physiological systems development (Table 5.4a). The most significant overrepresented molecular functions pertained to cellular development and cellular movement. Furthermore the over-represented physiological systems development were concerned with the development and function of connective tissue, skeletal and muscular systems, tissue and the cardiovascular system providing further support for a putative role for Ddc in cardiomyocyte cell growth, proliferation and movement with in the developing myocardium.

In order to consider potential phenotypes that may be affected by genes mis-regulated genes in the absence of Ddc, each of the 323 genes were cross-references with the mammalian phenotype (MP) browser, a directory of 2194 genotypes associated with abnormal heart morphology (Eppig, et al., 2012) (data retrieved February 2012). Each gene was then associated with one or multiple heart phenotypes in the following categories: ventricular, atrial, and/or valvular/septal defect, phenotypes that could not be categorised were listed as "other" and included phenotypes such as hypertension or abnormal heart patterning (Figure 5.4a). These results demonstrated that 44 genes (~14 %) mis-regulated in the *Ddc*$^{+/-}$ mutants have known published heart phenotypes. 14 genes are associated with specific defects in ventricular formation, with six further genes associating with abnormal morphology of both ventricles and atria. Nine genes have been shown to play a role specifically in valvular and septal defects and an additional six showed defects in all three categories.

Appendix 5.4 Summary of gene expression analysis

Microarray analysis and subsequent analysis of genes that are mis-regulated when *Ddc_exon1* expression is reduced in the developing heart, reveals a putative role for Ddc in heart as a positive regulator of branching formation, muscle and organ development. It also appears to play a role in

Molecular and Cellular Functions	p-value	# molecules
Cellular Development	1.30E-06 - 1.23E-02	54
Cellular Movement	1.63E-05 - 1.23E-02	46
Lipid Metabolism	3.57E-05 - 1.23E-02	28
Molecular Transport	3.57E-05 - 1.23E-02	25
Small Molecule Biochemistry	3.57E-05 - 1.23E-02	41

Physiological System Development	p-value	# molecules
Connective Tissue Development and Function	1.65E-07 - 1.23E-02	29
Skeletal & Muscular System Development and Function	1.65E-07 - 1.23E-02	54
Tissue Development	1.65E-07 - 1.23E-02	64
Cardiovascular System Development and Function	8.68E-07 - 1.23E-02	38
Embryonic Development	8.68E-07 - 1.23E-02	43

Table 5.4a Ingenuity® systems pathway analysis. 323 genes differentially expression in *Ddc*$^{-/+}$ compared to *Ddc*$^{+/-}$ analysed using the Ingenuity® bioinformatic analysis tool. Results are ranked by statistical significance based on P value with the most statistically significant ontology first.

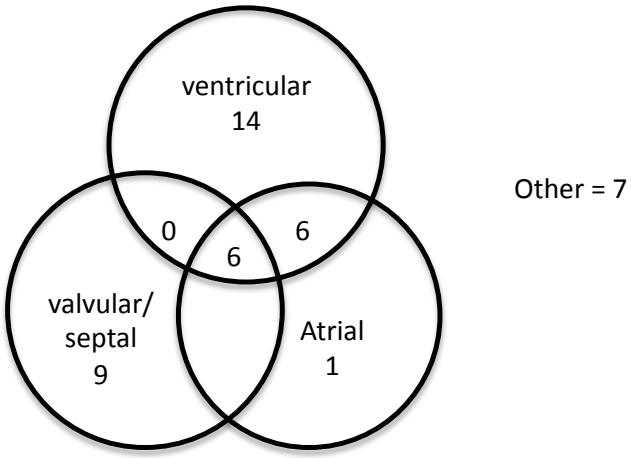


Figure 5.4a Mammalian phenotype browser analysis. Genes differentially expressed in *Ddc*$^{-/+}$ compared to *Ddc*$^{+/-}$ were cross-referenced with genes listed the mammalian phenotype browser, genes known to associated with a heart phenotype were categorised into ventricular, atrial, and/or valvular/septal defects, phenotypes which could not be assigned to one of these categories were listed as "other". Results demonstrated that 44 genes (~14 %) have known, published heart phenotypes.

correct cytoskeletal organization within muscle tissue and organ development. This would be concordant with the expression of *Ddc* in the compact layer of the developing myocardium, as the compact layer of the myocardium has high proliferative activity and serves as a source of new cells for the trabeculae and to increase thickness of the myocardium during the early proliferative phase (Sedmera, et al., 2000). Subsequent validation of differentially expressed genes has yet to be performed however this is planned (beyond the scope of this thesis).

Publications

1. **Prickett, A.R**. and Oakey, R. J. (2012) A survey of tissue specific genomic imprinting in mammals. *Mol Genet Genomics.* 287:621-630
2. **Prickett, A.R***., Barkas, N*., McCole, R.B*., Hughes, S., Amante, S.M., Schulz, R., and Oakey, R.J. Genomewide 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 research (in press)*. *Contributed equally.
3. **Prickett, A.R**., Bohm, S., Irving MD and Oakey, R.J. Imprinting in the human *Ddc* gene. Manuscript in preparation.
4. **Prickett, A.R**., Saadeh, H., Cowley, M.A., Schulz, R., Baldwin, H.S. and Oakey, R.J. A novel function for Dopa Decarboxylase in the developing mouse heart. Manuscript in preparation

Mol Genet Genomics (2012) 287:621–630
DOI 10.1007/s00438-012-0708-6

REVIEW

A survey of tissue-specific genomic imprinting in mammals

Adam R. Prickett · Rebecca J. Oakey

Received: 21 May 2012 / Accepted: 3 July 2012 / Published online: 21 July 2012

Abstract In mammals, most somatic cells contain two copies of each autosomal gene, one inherited from each parent. When a gene is expressed, both parental alleles are usually transcribed. However, a subset of genes is subject to the epigenetic silencing of one of the parental copies by genomic imprinting. In this review, we explore the evidence for variability in genomic imprinting between different tissue and cell types. We also consider why the imprinting of particular genes may be restricted to, or lost in, specific tissues and discuss the potential for high-throughput sequencing technologies in facilitating the characterisation of tissue-specific imprinting and assaying the potentially functional variations in epigenetic marks.

Keywords Imprinting · Tissue-specific · Genome-wide

Epigenetic processes are essential for mammalian differentiation and development and errors can result in developmental defects and disease. Genomic imprinting is a specialized form of gene regulation in mammals (Reik and Walter 1998) and flowering plants, whereby rather than expression of all alleles of a gene, the alleles inherited from one parent are silenced, while the other alleles remain active (Fig. 1). The functional non-equivalence of the mammalian genome was demonstrated by elegant nuclear transfer experiments in the 1980s (McGrath and Solter 1984; Surani et al. 1984) illustrating the need for a maternal and a paternal genome for development to term. Complementary studies using mice with balanced translocation chromosomes contributed to the delineation of specific regions of the genome that cause abnormalities when inherited from only a single parent (Cattanach and Kirk 1985). These mice exhibit phenotypes ranging from subtle effects on growth and development to embryonic lethality, demonstrating that many imprinted genes play a role in normal growth and development. Decades of work have revealed up to 150 likely imprinted genes in the mouse genome (Williamson et al. 2012; Schulz et al. 2008) with a significant number similarly expressed in human (Morison et al. 2001). Interestingly, the imprinted state of a gene is not necessarily conserved between mouse and human. Moreover, imprinting is not always consistent across tissues within the same organism, with some genes demonstrating imprinted expression in only a subset of the tissues in which the gene is active. It is this tissue specificity that this review seeks to explore.

In the genome, imprinted genes are frequently found in close proximity to one another, either organised into clusters (Reik and Walter 2001) or pairs (Wood et al. 2007) and occasionally as singletons (Hagiwara et al. 1997). This genomic organisation is a feature of both the mouse and human genomes. Ultimately, parent-of-origin-specific gene expression in even the largest cluster is controlled by a differentially methylated region (DMR), a small genomic region, typically with a CpG island at its heart that is methylated on one of the parental alleles, but not the other. Some of these methylation differences are established during germline development, and these regions are therefore referred to as germline DMRs (gDMRs); others are acquired during early embryonic development and are called somatic or acquired DMRs. The molecular

Communicated by J. Graw.

A. R. Prickett · R. J. Oakey (✉)
Department of Medical and Molecular Genetics,
King's College London, 8th Floor Tower Wing,
Guy's Hospital, London, SE1 9RT, UK
e-mail: rebecca.oakey@kcl.ac.uk

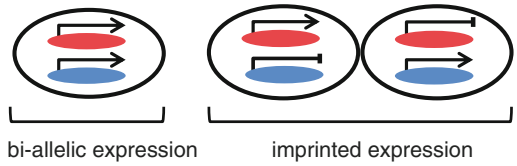


Fig. 1 Imprinted gene expression. Most genes show bi-allelic expression in mammalian cells; this means that transcription occurs roughly equally from both maternal and paternal alleles. Imprinted genes are an exception to this rule; for these genes, expression comes predominantly from one parental allele with the other allele silenced by epigenetic mechanisms. *Red bars* indicate a maternally inherited allele and *blue* indicates paternally inherited alleles (colour figure online)

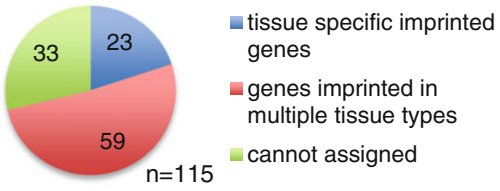


Fig. 2 Imprinted gene categories. All genes listed on WAMIDEX were stratified into three categories; those with tissue-specific imprinted expression, those imprinted in two or more tissues and those that cannot be assigned as imprinting has only been tested in one tissue type. Over a quarter of imprinted genes have only been assessed in one tissue type, mostly from screens involving a single tissue (colour figure online)

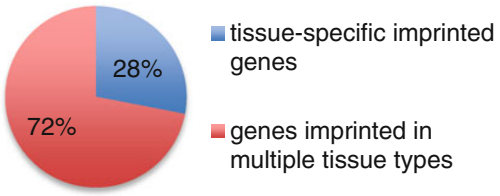


Fig. 3 Proportion of genes exhibiting tissue-specific imprinting. Of the 82 imprinted genes that have been tested for mono-allelic expression in multiple tissues, 28 % show imprinting in just one tissue type. We have considered genes tissue-specifically imprinted if parent-of-origin-specific mono-allelic expression is seen in one tissue type only and multiple other tissues have been assessed for imprinted expression (colour figure online)

mechanisms by which DMRs exert control over the expression of imprinted genes have been intensely studied (Thorvaldsen et al. 1998; Peters et al. 1999; Lin et al. 2003; Redrup et al. 2009). For example, specific histone modifications, central to chromatin organization and transcriptional regulation, are enriched in a parental allele-specific manner at DMRs (Umlauf et al. 2004) and the regulation of chromatin organization has been implicated in the control of imprinting of the well-characterised imprinted *Igf2/H19* locus via interactions mediated by DNA methylation (Hark et al. 2000; Murrell et al. 2004). One feature that has emerged is the fact that imprinting is not necessarily conserved across all tissues of a given mammal, and other factors contribute to a more complicated picture, including the use of tissue-specific promoters and epigenetic marks.

Epigenetic mechanisms are closely linked with the processes of cell differentiation and the specification of cell types and tissues that emerge during development. Here, we consider examples of tissue-specific imprinting and discuss the potential significance of these patterns of expression. We also examine the importance of next-generation sequencing technologies in elucidating the extent of tissue-specific imprinting and the associated epigenetic mechanisms.

Some genes are imprinted in a tissue-specific manner

To study examples of tissue-specific imprinting, we have summarized the collection of imprinted genes found in the Web Atlas of Murine genomic Imprinting and Differential EXpression (WAMIDEX) (Schulz et al. 2008) and stratified genes into three categories: (1) those with tissue-specific imprinted expression. We categorized genes as tissue-specific if they are imprinted in only one tissue and if multiple other tissues, or whole embryo has been tested for expression and the gene is bi-allelic; (2) those with ubiquitous imprinting, or imprinted expression in multiple different tissue types and only sporadic bi-allelic expression; or (3) those that cannot be assigned because imprinting has only been tested in one tissue type (Fig. 2). This stratification is fluid because it represents currently published data and thus genes may change categories as more data from additional tissue or cell types become available. We did not include expression data on oocytes and sperm. Of the 82 imprinted genes for which imprinted expression has been tested in multiple tissues, a surprisingly large proportion (28 %) shows tissue-specific imprinted expression in one specific tissue type (Fig. 3). These cases of tissue-specific imprinted gene expression represent a plausible and useful model for studying epigenetic variation that may be involved with the control of differential imprinting between cell types. It may also provide additional insight into the evolution of genomic imprinting in mammals.

Of the 23 tissue-specific imprinted genes (Table 1), the majority are imprinted only in extra-embryonic tissues, specifically placenta (48 %) and yolk sac (9 %), or only in the brain, including specific subsets of brain regions (39 %) (Fig. 4). We have also highlighted in Table 1 a further five genes which do not quite fit our criteria for tissue-specific imprinting, because their imprinting is not unique to one

Table 1 Details of WAMIDEX genes exhibiting imprinting in one tissue-type only

Gene	Allele	Location of tissue-specific imprinting	References
A19	p	Brain	(de la Puente et al. 2002)
Ampd3	m	Placenta	(Schulz et al. 2006)
Atp10a	m	Hippocampus and olfactory bulb	(Kayashima et al. 2003; Kashiwagi et al. 2003)
Blcap_v1a	m	Brain	(Schulz et al. 2009)
Calcr	m	Embryonic and adult Brain	(Hoshiya et al. 2003)
Commd1	m	Brain from postnatal day 14	(Wang et al. 2004)
Copg2	m	Brain	(Lee et al. 2000)
Dcn	m	Placenta	(Mizuno et al. 2002)
Ddc_exon1a	p	Embryonic and neonatal heart	(Menheniott et al. 2008)
Dhcr7	m	Placenta	(Schulz et al. 2006)
*Ins1**	p	Yolk Sac	(Giddings et al. 1994; Deltour et al. 2004)
Ins2	p	Yolk Sac	(Giddings et al. 1994; Deltour et al. 2004)
Mash2	m	Placenta	(Tanaka et al. 1999)
Nap1l4	m	Placenta	(Paulsen et al. 1998; Engemann et al. 2000)
Obph1	m	Placenta	(Umlauf et al. 2004; Higashimoto et al. 2002; Engemann et al. 2000)
Slc22a2	m	Placenta	(Zwart et al. 2001)
Slc22a3	m	Placenta	(Zwart et al. 2001)
Th	m	Placenta	(Schulz et al. 2006)
Tnfrsf23	m	Placenta	(Clark et al. 2002)
Tssc4	m	Placenta	(Paulsen et al. 2000)
Ube3a	m	Neurons	(Yamasaki et al. 2003)
Ube3aAs	p	Neurons	(Yamasaki et al. 2003)
Zfp264	p	Neonatal and adult Brain	(Kim et al. 2001)
Genes imprinted specifically in extra-embryonic tissue			
Gatm	m	Placenta and yolk sac	(Sandell et al. 2003)
Neurabin	m	Placenta and yolk sac	(Ono et al. 2003; Monk et al. 2008)
Pon2	m	Placenta and yolk sac	(Ono et al. 2003; Monk et al. 2008)
Pon3	m	Placenta and yolk sac	(Ono et al. 2003; Monk et al. 2008)
Tfpi2	m	Placenta and yolk sac	(Monk et al. 2008)

Tissue-specific imprinted genes are taken from those imprinted genes listed in the Web Atlas of Murine genomic Imprinting and Differential Expression (WAMIDEX). We categorized genes as tissue-specific if they are specifically imprinted in only one tissue type, and multiple other tissue types or whole embryo has been tested for expression and imprinting, but show bi-allelic expression. A further five genes listed are not tissue-specific by this definition, as they are imprinted in the two extra-embryonic tissues of the placenta and yolk sac; however, they have been included to illustrate extra-embryonic imprinting. "Allele" denotes the allele that is expressed where the gene is imprinted

* Denotes contradictory evidence

tissue only, but they are mono-allelically expressed only in the extra-embryonic tissues of both placenta and yolk sac.

Genomic imprinting, by the nature of epigenetically silencing one allele, presents a potential disadvantage to mammals if they are unlucky enough to be encumbered by a deleterious mutation on the active copy as this would result in complete loss of function. Explanations for why imprinting evolved have thus been energetically sought. Imprinted expression of genes specifically in placenta or the brain during the pre-weaning period in the mouse can be seen to support the "parental conflict" theory of evolution for genomic imprinting (Moore and Haig 1991). The salient points of this theory are that the maternal and paternal genomes have different interests in terms of offspring demand on maternal resource allocation, with paternally expressed imprinted genes selected to maximize demand from the mother and maternally expressed genes to act to balance demand between offspring of current and future litters (Moore and Haig 1991). If one buys into this evolutionary perspective, one might expect the imprinting status of genes to be different depending on whether the genes are being expressed in a tissue type where their function could influence the allocation of maternal resources. For genes acting in growth factor signalling pathways, such as *Igf2* or its receptor *Igf2r*, growth can be considered as a global effect, and thus tissue-specific

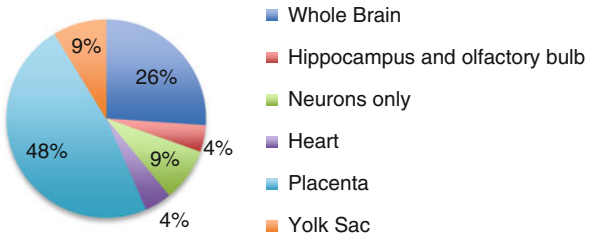


Fig. 4 Location of tissue-specific imprinting. For tissue-specific imprinted transcripts, we surveyed in which tissues these are imprinted. A large proportion is imprinted in extra-embryonic tissues of the placenta (48 %) and yolk sac (9 %). Another hot-spot of tissue-specific imprinting is whole brain (26 %) including subsets of brain regions, the hippocampus (4 %) and neurons (9 %) (colour figure online)

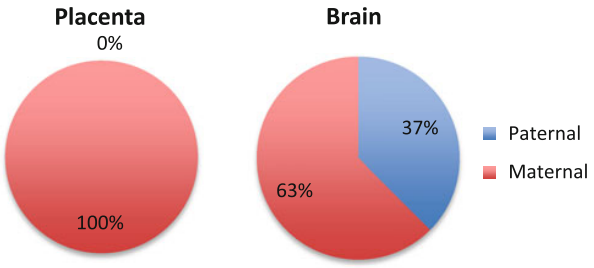


Fig. 5 Parent-of-origin of tissue-specific imprinting in the brain and the placenta. Of the genes imprinted tissue-specifically in the placenta, all are expressed from the maternal allele demonstrating maternal genome-specific contributions to placental growth and function. Genes imprinted tissue-specifically in brain are both maternally expressed (63 %) and paternally expressed (37 %) (colour figure online)

imprinting is irrelevant. However, in tissues such as the brain, which influences feeding behaviour, or placenta, which acts as a direct mediator of maternal resource allocation, pressure to evolve genomic imprinting of particular genes may be localized specifically to these tissues. Existing evidence supports a role for genomic imprinting in behaviour that is associated with demand for resources. For example, mouse models where imprinting has been disrupted due to the maternal duplication of chromosome 2 (and lack of a paternal Chr 2) die perinatally due in part to a failure to suckle (Cattanach 1986). This phenotype can be attributed, at lease in part, to genes in the imprinted GNAS cluster. This complex gene cluster is made up of multiple transcripts, each with a distinct pattern of imprinted expression. The *GnasXL* transcript in particular is paternally expressed in multiple regions of the brain associated with the control of orofacial muscles, and mice with a deletion specifically targeted to the *XL* transcript exhibit reduced ability to suckle milk, become weak and inactive and die shortly after birth (Plagge et al. 2004). The imprinted *Peg3* and *Dlk1* genes have also been shown to regulate suckling behaviour in mouse (Curley et al. 2004; da Rocha et al. 2009). Interestingly, in placenta, all 11 genes that are tissue-specifically imprinted (Table 1) are expressed from the maternal allele only, pointing to a specialized and overrepresented role for the maternal genome in placental development and function ($p < 0.005$, Chi-squared vs. mat:pat for all imprinted genes) (Fig. 5), consistent with previous findings (Schulz et al. 2010). Of the tissue-specifically imprinted genes in brain, three are paternally and five are maternally expressed, which is not significantly different from the expected ratio based on the overall numbers of maternally and paternally expressed imprinted genes (Fig. 5).

The only published example of a tissue-specific imprinted gene where imprinting is not localized to the brain or extra-embryonic tissue is Dopa decarboxylase (*Ddc*), which is imprinted solely in heart (Menheniott et al. 2008). *Ddc* demonstrates a mechanism of tightly controlled epigenetic regulation, which varies between heart and other tissue types. *Ddc* gives rise to two transcript variants termed *Ddc_canonical* and *Ddc_exon1a*, which differ only in the first untranslated exon. The *Ddc_exon1a* variant is expressed in brain and heart, and is bi-allelic in brain, but expressed from the paternal allele only in the developing heart (Fig. 6). This organization of transcripts and expression patterns indicate that epigenetic differences between brain and heart are key to controlling imprinted expression. The fine transcriptional control exerted at the *Ddc* locus, specifically in heart tissue, does not readily support the parental conflict theory of imprinting. Imprinting of *Ddc* is neither ubiquitous nor active in a tissue that is directly involved in resource provisioning from mother to offspring. We must therefore consider that there may be another evolutionary role for maternal silencing of *Ddc* in heart.

Ubiquitous and tissue-specific imprinting in a single imprinted cluster

Interestingly, the tissue-specific switch from bi-allelic gene expression to imprinted expression appears to occur independently of other imprinted genes within a single imprinting cluster. *Mash2*, *Obph1 and Tssc4* are each tissue-specifically imprinted in placental tissues, remaining bi-allelically expressed in other embryonic tissues. These genes form part of the imprinted cluster that is controlled, at least in part, by a DMR located at the promoter of *Kcnq1ot1*. This domain includes genes ubiquitously imprinted in early embryonic development namely *Kcnq1ot1, Kcnq1, Slc22a18*, *Cdkn1c* and *Phlda2* (Mancini-Dinardo et al. 2006). It has subsequently been shown that a separate genomic region downstream of the *kcnq1ot1* promoter may be responsible for

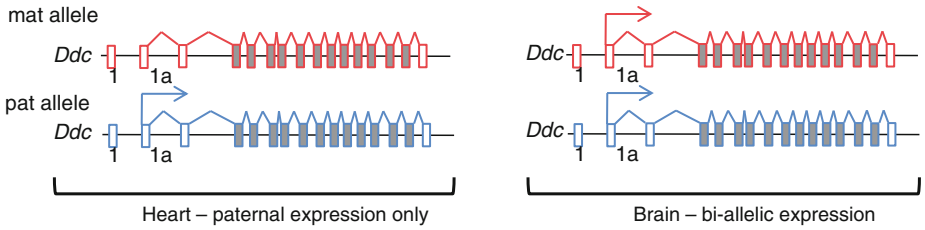


Fig. 6 Imprinted expression of *Ddc*. *Ddc* is the only example of tissue-specific imprinting in heart. The *Ddc_exon1a* transcript is expressed from both parental alleles in brain; however, in heart the maternal allele is epigenetically silenced

Table 2 Details of WAMIDEX genes exhibiting bi-allelic expression in one tissue-type only

Gene	Allele when imprinted	Location of bi-allelic expression	References
Dlk1	p	From day 7 in niche astrocytes and neural stem cells	(Ferron et al. 2011)
Frat3	p	Placenta	(Kobayashi et al. 2002)
H13 –multiple isoforms*	m	Neonatal brain	(Wood et al. 2007, 2008)
Igf2	p	Neonatal choroid plexus and Leptomeninges	(DeChiara et al. 1991; Hetts et al. 1997)
Igf2r	m	Neonatal cerebrum	(Yamasaki et al. 2005)
Inpp5f_v2	p	Adult testes	(Choi et al. 2005; Wood et al. 2007)
Mcts2	p	Adult testes	(Wood et al. 2007)
U2af1-rs1	p	Adult testes	(Wang et al. 2004; Wood et al. 2007)
Zac1	p	30-day Liver	(Piras et al. 2000)

Listed are the imprinted genes that are close to ubiquitously imprinted but which are expressed in a bi-allelic fashion in one tissue type only (data from WAMIDEX). Excluded are data from cell lines and data that are ambiguous or open to alternative interpretation

* Denotes genes that have multiple isoforms some of which are maternally expressed and other of which are paternally expressed

differential control of the ubiquitously imprinted and the tissue-specifically imprinted genes found at this locus (Mohammad et al. 2010). Further to this, the imprinting of *Kcnq1* itself is relaxed during mid-gestation in the brain and the kidney and in cardiac lineages between embryonic days 13.5 and 14.5, reverting from maternal expression to bi-allelic expression. This reversion is concurrent with changes to the chromatin structure and results in an increase in *kcnq1* transcript levels in the developing heart (Korostowski et al. 2011). Thus, evidence from this locus demonstrates how a temporal switch from imprinted to bi-allelic expression in a single tissue can be used as part of a mechanism to regulate gene dosage, how tissue-specifically imprinted genes can exist alongside genes imprinted in multiple tissue types within a single imprinting cluster and how changes to the chromatin structure may influence tissue-specific gene expression.

Imprinted genes that are bi-allelic in a specific tissue type

In the same way that a subset of imprinted genes is mono-allelic in one tissue only, another subset is almost ubiquitously imprinted and yet is expressed in a bi-allelic fashion in a single tissue (Table 2). This was first demonstrated at the *Igf2* locus, as *Igf2* reverts to bi-allelic expression in the choroid plexus and leptomeninges (DeChiara et al. 1991). In somatic tissues it has been considered that reversion to bi-allelic expression may passively result from a lack of a need to maintain the imprinted state. However, new work on the *Dlk1* gene suggests that in this case reversion to bi-allelic expression may play an important role in the regulation of gene dosage.

The switch from imprinted to bi-allelic expression of *Dlk1* in niche astrocytes and neural stem cells is achieved by the acquisition of a methylation mark, and the resulting reactivation of the maternal allele is integral for post-natal neural development (Ferron et al. 2011). So, while the bi-allelic expression of Dlk1 is important in these cells, it is also critical for imprinting of Dlk1 to be maintained in other somatic tissues to facilitate correct gene dosage. Mice harbouring a *Dlk1* BAC transgene, recapitulating a re-activation of bi-allelic expression, are born larger than wild-type littermates, but exhibit skeletal abnormities and show an increase in postnatal lethality due to reduced suckling ability (da Rocha et al. 2009). It is important therefore to consider the role of tissue-specific imprinted

expression in the context of gene dosage control, which could also drive the imprinted expression of *Ddc_exon1a* in heart.

Grb10 switches expression from the maternal to the paternal allele in a tissue-specific manner

The growth factor receptor-binding protein, *Grb10*, has to date a somewhat unique imprinted expression pattern, being expressed from the paternal allele in brain but from the maternal allele in most other peripheral tissues. During development, *Grb10* expressed from the maternal allele functions to repress growth and in adulthood mediates glucose metabolism and energy homeostasis (Smith et al. 2007; Charalambous et al. 2003, 2010). Intriguingly, in the brain, where *Grb10* is expressed from the paternal allele, it exhibits a distinct function, tempering social dominance (Garfield et al. 2011). This switch from maternal to paternal expression is associated with tissue-specific changes in epigenetic marks, namely the presence of a brain-specific somatic DMR, located upstream of the *Grb10* germline DMR, which is hypermethylated on both parental alleles in the peripheral tissues of the kidney and liver and the loss of the repressive H3K27me3 histone mark at the paternal *Grb10* promoter in the developing neural lineage (Arnaud et al. 2003; Sanz et al. 2008). *Grb10* is a unique example of a gene that exhibits a different function depending on which parental allele is expressed.

Next-generation sequencing will aid the identification of novel instances of tissue-specific imprinting

Next-generation sequencing technology is beginning to provide novel insights into genomic imprinting (Cooper and Constancia 2010). To accurately define all imprinted gene expression would require the analysis of each gene transcript's expression in a pure population of each distinct developmental and terminally differentiated cell type. Although this remains a distant goal, the detection of new examples of tissue-specific imprinting and identification of the epigenetic differences that cause the parent-of-origin specific silencing in a particular cell type are facilitated by next-generation sequencing techniques applied to RNA (RNA-seq), immunoprecipitated chromatin using antibodies against DNA-binding proteins or histone modifications (ChIP-seq) and bisulphite-treated DNA (BS-seq). For example, next-generation sequencing has enabled significant progress in elucidating the establishment of the epigenetic marks that are crucial for genomic imprinting during oogenesis (Smallwood et al. 2011). Genome-wide analysis of tissue-specific imprinting can provide an additional tool for understanding how somatic changes to epigenetic marks, during growth and development, can control gene silencing by allowing direct comparisons to be made between active and silenced alleles in imprinted and non-imprinted tissues.

In principle, RNA-seq can generate quantitative measurements of gene expression for both parental alleles. One of the main problems encountered when identifying tissue-specific imprinting is the masking of mono-allelic expression in one tissue-type, by bi-allelic expression of the same gene in surrounding tissues. For example, when assaying whole brain for mono-allelic gene expression, we are actually looking at multiple brain regions, multiple cell types as well as potential contaminations that may occur during tissue dissection. The sensitivity and large dynamic range of RNA-seq can identify imprinting that has not previously been detectable. So far the use of RNA-seq for identifying novel tissue-specific imprinting has been a useful tool. Using RNA-seq, Gregg et al. (2010a, b) recently reported 1300 new examples of genes imprinted in brain including many examples of sex-specific imprinting. However, a recent re-analysis of their data suggests that this may be an overestimate and that the real number of novel imprinted genes may be much lower (Deveale et al. 2012). These studies highlight the difficulty in interpreting and modelling genome-wide expression data generated by RNA-seq and also illustrate the difficulty in distinguishing between true imprinting and either biological or technical stochastic biases in the levels of transcript from the two parental alleles (Kelsey and Bartolomei 2012). Although RNA-seq is sometimes effective in detecting imprinted expression in the mixtures of cell types found in whole tissues, to correctly pinpoint tissue-specific imprinting, one would ideally need to sequence a pure population of cells using RNA-seq. Isolating pure populations of cells will become easier as cell sorting approaches mature and the amount of input RNA required is reduced as sequencing technologies improve. Single-cell RNA-seq will ultimately allow for assessment of allelic expression in a cell-specific manner (Tang et al. 2010); however, this technology is still immature and is yet to be applied to the study of tissue-specific imprinting.

In addition to RNA-seq, further analysis of other genome-wide epigenetic marks and chromatin structure is providing a better understanding of tissue-specific epigenetic variation. This includes the study of tissue-specific changes in DNA methylation and histone modifications alongside tissue-specific differences in chromatin conformation. This will allow the characterization of the interplay between epigenetic marks, chromatin structure and imprinted gene expression in mechanistic studies. Genomic imprinting is reliant on epigenetic differences set up in the

germline and then maintained, or potentially lost or gained, during differentiation. Epigenetic differences between tissues may indicate variation in genomic imprinting in different tissues and inform further on the epigenetic mechanisms governing imprinted gene expression.

Differential methylation between parental alleles is known to control imprinted gene expression (Reik et al. 2001). Identification of variations in DMRs, particularly somatic DMRs, genome-wide and in different tissue types, are potentially indicative of tissue-specific imprinting. Advances in bisulfite-seq technology are allowing us to interrogate methylation genome-wide with base pair resolution and can be used to examine methylation in a parent-of-origin specific manner. A recently published study in mouse frontal cortex demonstrated allele-specific methylation at 32 known imprinted domains and a further 23 genomic regions (Xie et al. 2012). In the same study, Xie et al. examined the histone modifications H3K27ac and K3K4me3 using ChIP-seq as well as RNA-seq to identify parent-of-origin differences in expression. This is the first study to combine multiple genome-wide and allele-specific techniques to probe epigenetic and gene expression differences in a single in vivo tissue type. To investigate changing chromatin structure in different cell types, Hi-C has been developed as a technique for studying higher order chromatin interactions in a genome-wide context. Hi-C combines chromosome conformation capture with next-generation sequencing to profile the spatial association of different regions of DNA within the cell. Hi-C has been used to study variation in chromatin interactions between mouse embryonic stem cells and differentiated mouse cortex (Dixon et al. 2012); this type of comparison, made between tissues which display tissue-specific imprinting differences, could be used to better understand the effect of chromatin looping in governing genomic imprinting.

For the purposes of review, we have defined tissue-specific imprinting as genes that are bi-allelic in most tissues but imprinted specifically in only one tissue or cell type. It is pertinent, however, to mention *Gsα* despite the fact it does not fit our criteria for tissue-specific imprinting. *Gsα* is located in the *GNAS* cluster of imprinted genes and is a well-studied example of a gene that switches from bi-allelic to imprinted expression. The *GNAS* locus consists of multiple different imprinted transcripts; for a full review see Peters and Williamson (2007). Of these transcripts, *Gsα* is highly expressed in a bi-allelic manner in most tissues, but is expressed only from the maternal allele in a specific subset of tissues composed of renal cortex, brown and white adipose tissue (Yu et al. 1998), the anterior pituitary (Hayward et al. 2001), thyroid (Mantovani et al. 2002; Germain-Lee et al. 2002), ovary (Mantovani et al. 2002) and paraventricular nucleus of the hypothalamus (Chen et al. 2009). The *1A* DMR is known to control imprinting in these tissues (Williamson et al. 2004); however, no tissue-specific differences in methylation at this DMR have so far been found; thus tissue-specific imprinting may involve more than simple tissue-specific methylation differences and it is important to consider other epigenetic variations.

Summary

Genomic imprinting can be regarded as the 'perfect' tool for studying epigenetic silencing as both active and inactive gene copies reside in identical conditions in the same cell. By using a raft of novel experimental tools, it is now feasible to fully dissect the imprintome by studying tissue-specific differences in allelic expression. When coupled with genome-wide epigenetic analysis from the same pure cell populations, this will allow a greater understanding of epigenetic silencing at imprinted loci and will form a foundation from which we can increase our understanding of the relationship between epigenetic marks and gene silencing in a non-imprinting context.

Acknowledgments We would like to thank Dr Michael Cowley, Dr Reiner Schulz and Siobhan Hughes for their critical reading of this manuscript. We thank our funders the Wellcome Trust (RJO, Grant number 085448/Z/08/Z) and the British Heart Foundation (AP, Grant number FS/08/051/25748) for their support.

References

Arnaud P, Monk D, Hitchins M, Gordon E, Dean W, Beechey CV, Peters J, Craigen W, Preece M, Stanier P, Moore GE, Kelsey G (2003) Conserved methylation imprints in the human and mouse GRB10 genes with divergent allelic expression suggests differential reading of the same mark. Hum Mol Genet 12(9): 1005–1019

Cattanach BM (1986) Parental origin effects in mice. J Embryol Exp Morphol 97(Suppl):137–150

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

Charalambous M, Smith FM, Bennett WR, Crew TE, Mackenzie F, Ward A (2003) Disruption of the imprinted Grb10 gene leads to disproportionate overgrowth by an Igf2-independent mechanism. Proc Natl Acad Sci USA 100(14):8292–8297. doi:10.1073/pnas.1532175100153217510015321751

Charalambous M, Cowley M, Geoghegan F, Smith FM, Radford EJ, Marlow BP, Graham CF, Hurst LD, Ward A (2010) Maternally inherited Grb10 reduces placental size and efficiency. Dev Biol 337(1):1–8. doi:10.1016/j.ydbio.2009.10.011

Chen M, Wang J, Dickerson KE, Kelleher J, Xie T, Gupta D, Lai EW, Pacak K, Gavrilova O, Weinstein LS (2009) Central nervous system imprinting of the G protein G(s)alpha and its role in metabolic regulation. Cell Metab 9(6):548–555. doi:10.1016/j.cmet.2009.05.004

Choi JD, Underkoffler LA, Wood AJ, Collins JN, Williams PT, Golden JA, Schuster EF Jr, Loomes KM, Oakey RJ (2005) A novel variant of Inpp 5f is imprinted in brain, and its expression is correlated with differential methylation of an internal CpG

island. Mol Cell Biol 25(13):5514–5522. doi:10.1128/MCB.25.13.5514-5522.2005
Clark L, Wei M, Cattoretti G, Mendelsohn C, Tycko B (2002) The Tnfrh1 (Tnfrsf23) gene is weakly imprinted in several organs and expressed at the trophoblast-decidua interface. BMC Genet 3:11
Cooper WN, Constancia M (2010) How genome-wide approaches can be used to unravel the remaining secrets of the imprintome. Brief Funct Genomics 9(4):315–328. doi:10.1093/bfgp/elq018
Curley JP, Barton S, Surani A, Keverne EB (2004) Coadaptation in mother and infant regulated by a paternally expressed imprinted gene. Proc Biol Sci 271(1545):1303–1309. doi:10.1098/rspb.2004.2725
da Rocha ST, Charalambous M, Lin SP, Gutteridge I, Ito Y, Gray D, Dean W, Ferguson-Smith AC (2009) Gene dosage effects of the imprinted delta-like homologue 1 (dlk1/pref1) in development: implications for the evolution of imprinting. PLoS Genet 5(2):e1000392. doi:10.1371/journal.pgen.1000392
de la Puente A, Hall J, Wu YZ, Leone G, Peters J, Yoon BJ, Soloway P, Plass C (2002) Structural characterization of Rasgrf1 and a novel linked imprinted locus. Gene 291(1–2):287–297 pii:S0378111902006017
DeChiara TM, Robertson EJ, Efstratiadis A (1991) Parental imprinting of the mouse insulin-like growth factor II gene. Cell 64:849–859
Deltour L, Vandamme J, Jouvenot Y, Duvillie B, Kelemen K, Schaerly P, Jami J, Paldi A (2004) Differential expression and imprinting status of Ins1 and Ins2 genes in extraembryonic tissues of laboratory mice. Gene Expr Patterns 5(2):297–300. doi:10.1016/j.modgep.2004.04.013
Deveale B, van der Kooy D, Babak T (2012) Critical evaluation of imprinted gene expression by RNA-Seq: a new perspective. PLoS Genet 8(3):e1002600. doi:10.1371/journal.pgen.1002600
Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS, Ren B (2012) Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature 485(7398):376–380. doi:10.1038/nature11082
Engemann S, Strodicke M, Paulsen M, Franck O, Reinhardt R, Lane N, Reik W, Walter J (2000) Sequence and functional comparison in the Beckwith–Wiedemann region: implications for a novel imprinting centre and extended imprinting. Hum Mol Genet 9(18):2691–2706
Ferron SR, Charalambous M, Radford E, McEwen K, Wildner H, Hind E, Morante-Redolat JM, Laborda J, Guillemot F, Bauer SR, Farinas I, Ferguson-Smith AC (2011) Postnatal loss of Dlk1 imprinting in stem cells and niche astrocytes regulates neurogenesis. Nature 475(7356):381–385. doi:10.1038/nature10229
Garfield AS, Cowley M, Smith FM, Moorwood K, Stewart-Cox JE, Gilroy K, Baker S, Xia J, Dalley JW, Hurst LD, Wilkinson LS, Isles AR, Ward A (2011) Distinct physiological and behavioural functions for parental alleles of imprinted Grb10. Nature 469(7331):534–538. doi:10.1038/nature09651
Germain-Lee EL, Ding CL, Deng Z, Crane JL, Saji M, Ringel MD, Levine MA (2002) Paternal imprinting of Galpha(s) in the human thyroid as the basis of TSH resistance in pseudohypoparathyroidism type 1a. Biochem Biophys Res Commun 296(1):67–72. doi:S0006291X02008331
Giddings SJ, King CD, Harman KW, Flood JF, Carnaghi LR (1994) Allele specific inactivation of insulin 1 and 2, in the mouse yolk sac, indicates imprinting. Nat Genet 6(3):310–313. doi:10.1038/ng0394-310
Gregg C, Zhang J, Butler JE, Haig D, Dulac C (2010a) Sex-specific parent-of-origin allelic expression in the mouse brain. Science 329(5992):682–685. doi:10.1126/science.1190831
Gregg C, Zhang J, Weissbourd B, Luo S, Schroth GP, Haig D, Dulac C (2010b) High-resolution analysis of parent-of-origin allelic expression in the mouse brain. Science 329(5992):643–648. doi:10.1126/science.1190830
Hagiwara Y, Hirai M, Nishiyama K, Kanazawa I, Ueda T, Sakaki Y, 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 USA 94(17):9249–9254
Hark AT, Schoenherr CJ, Katz DJ, Ingram RS, Levorse JM, Tilghman SM (2000) CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus. Nature 405(6785):486–489. doi:10.1038/35013106
Hayward BE, Barlier A, Korbonits M, Grossman AB, Jacquet P, Enjalbert A, Bonthron DT (2001) Imprinting of the G(s)alpha gene GNAS1 in the pathogenesis of acromegaly. J Clin Invest 107(6):R31–R36. doi:10.1172/JCI11887
Hetts SW, Rosen KM, Dikkes P, Villa-Komaroff L, Mozell RL (1997) Expression and imprinting of the insulin-like growth factor II gene in neonatal mouse cerebellum. J Neurosci Res 50(6):958–966. doi:10.1002/(SICI)1097-4547(19971215)50:6<958:AID-JNR6>3.0.CO;2-C
Higashimoto K, Soejima H, Yatsuki H, Joh K, Uchiyama M, Obata Y, Ono R, Wang Y, Xin Z, Zhu X, Masuko S, Ishino F, Hatada I, Jinno Y, Iwasaka T, Katsuki T, Mukai T (2002) Characterization and imprinting status of OBPH1/Obph1 gene: implications for an extended imprinting domain in human and mouse. Genomics 80(6):575–584. doi:S0888754302970060
Hoshiya H, Meguro M, Kashiwagi A, Okita C, Oshimura M (2003) Calcr, a brain-specific imprinted mouse calcitonin receptor gene in the imprinted cluster of the proximal region of chromosome 6. J Hum Genet 48(4):208–211. doi:10.1007/s10038-003-0006-6
Kashiwagi A, Meguro M, Hoshiya H, Haruta M, Ishino F, Shibahara T, Oshimura M (2003) Predominant maternal expression of the mouse Atp10c in hippocampus and olfactory bulb. J Hum Genet 48(4):194–198. doi:10.1007/s10038-003-0009-3
Kayashima T, Yamasaki K, Joh K, Yamada T, Ohta T, Yoshiura K, Matsumoto N, Nakane Y, Mukai T, Niikawa N, Kishino T (2003) Atp10a, the mouse ortholog of the human imprinted ATP10A gene, escapes genomic imprinting. Genomics 81(6):644–647. doi:S0888754303000776
Kelsey G, Bartolomei MS (2012) Imprinted genes and the number is? PLoS Genet 8 (3):e1002601. doi:10.1371/journal.pgen.1002601
Kim J, Bergmann A, Wehri E, Lu X, Stubbs L (2001) Imprinting and evolution of two Kruppel-type zinc-finger genes, ZIM3 and ZNF264, located in the PEG3/USP29 imprinted domain. Genomics 77(1–2):91–98. doi:10.1006/geno.2001.6621
Kobayashi S, Kohda T, Ichikawa H, Ogura A, Ohki M, Kaneko-Ishino T, Ishino F (2002) Paternal expression of a novel imprinted gene, Peg12/Frat3, in the mouse 7C region homologous to the Prader–Willi syndrome region. Biochem Biophys Res Commun 290(1):403–408. doi:10.1006/bbrc.2001.6160
Korostowski L, Raval A, Breuer G, Engel N (2011) Enhancer-driven chromatin interactions during development promote escape from silencing by a long non-coding RNA. Epigenetics Chromatin 4:21. doi:1756-8935-4-2110.1186/1756-8935-4-21
Lee YJ, Park CW, Hahn Y, Park J, Lee J, Yun JH, Hyun B, Chung JH (2000) Mit1/Lb9 and Copg2, new members of mouse imprinted genes closely linked to Peg1/Mest(1). FEBS Lett 472(2–3):230–234. doi:S0014579300014617
Lin SP, Youngson N, Takada S, Seitz H, Reik W, Paulsen M, Cavaille J, Ferguson-Smith AC (2003) Asymmetric regulation of imprinting on the maternal and paternal chromosomes at the Dlk1-Gtl2 imprinted cluster on mouse chromosome 12. Nat Genet 35(1):97–102
Mancini-Dinardo D, Steele SJ, Levorse JM, Ingram RS, Tilghman SM (2006) Elongation of the Kcnq1ot1 transcript is required for

genomic imprinting of neighboring genes. Genes Dev 20(10): 1268–1282. doi:10.1101/gad.1416906
Mantovani G, Ballare E, Giammona E, Beck-Peccoz P, Spada A (2002) The gsalpha gene: predominant maternal origin of transcription in human thyroid gland and gonads. J Clin Endocrinol Metab 87(10):4736–4740
McGrath J, Solter D (1984) Completion of mouse embryogenesis requires both the maternal and paternal genomes. Cell 37:179–183
Menheniott TR, Woodfine K, Schulz R, Wood AJ, Monk D, Giraud AS, Baldwin HS, Moore GE, Oakey RJ (2008) Genomic imprinting of Dopa decarboxylase in heart and reciprocal allelic expression with neighboring Grb10. Mol Cell Biol 28(1):386–396. doi:10.1128/MCB.00862-07
Mizuno Y, Sotomaru Y, Katsuzawa Y, Kono T, Meguro M, Oshimura M, Kawai J, Tomaru Y, Kiyosawa H, Nikaido I, Amanuma H, Hayashizaki Y, Okazaki Y (2002) Asb4, Ata3, and Dcn are novel imprinted genes identified by high-throughput screening using RIKEN cDNA microarray. Biochem Biophys Res Commun 290(5):1499–1505. doi:10.1006/bbrc.2002.6370
Mohammad F, Mondal T, Guseva N, Pandey GK, Kanduri C (2010) Kcnq1ot1 noncoding RNA mediates transcriptional gene silencing by interacting with Dnmt1. Development 137(15):2493–2499. doi:10.1242/dev.048181
Monk D, Wagschal A, Arnaud P, Muller PS, Parker-Katiraee L, Bourc'his D, Scherer SW, Feil R, Stanier P, Moore GE (2008) Comparative analysis of human chromosome 7q21 and mouse proximal chromosome 6 reveals a placental-specific imprinted gene, TFPI2/Tfpi2, which requires EHMT2 and EED for allelic-silencing. Genome Res 18(8):1270–1281. doi:10.1101/gr.077115.108
Moore T, Haig D (1991) Genomic imprinting in mammalian development: a parental tug-of-war. Trends Genet 7:45–49
Morison IM, Paton CJ, Cleverley SD (2001) The imprinted gene and parent of origin effect database: http://igc.otago.ac.nz/home.html . Nucleic Acids Res 29:275–276
Murrell A, Heeson S, Reik W (2004) Interaction between differentially methylated regions partitions the imprinted genes Igf2 and H19 into parent-specific chromatin loops. Nat Genet 36(8):889–893. doi:10.1038/ng1402
Ono R, Shiura H, Aburatani H, Kohda T, Kaneko-Ishino T, Ishino F (2003) Identification of a large novel imprinted gene cluster on mouse proximal chromosome 6. Genome Res 13(7):1696–1705. doi:10.1101/gr.90680313/7/1696
Paulsen M, Davies KR, Bowden LM, Villar AJ, Franck O, Fuermann M, Dean WL, Moore TF, Rodrigues N, Davies KE, Hu RJ, Feinberg AP, Maher ER, Reik W, Walter J (1998) Syntenic organization of the mouse distal chromosome 7 imprinting cluster and the Beckwith–Wiedemann syndrome region in chromosome 11p15.5. Hum Mol Genet 7 (7):1149–1159. pii:ddb146
Paulsen M, El-Maarri O, Engemann S, Strodicke M, Franck O, Davies K, Reinhardt R, Reik W, Walter J (2000) Sequence conservation and variability of imprinting in the Beckwith–Wiedemann syndrome gene cluster in human and mouse. Hum Mol Genet 9(12):1829–1841
Peters J, Williamson CM (2007) Control of imprinting at the Gnas cluster. Epigenetics 2(4):207–213. pii:5380
Peters J, Wroe SF, Wells CA, Miller HJ, Bodle D, Beechey CV, Williamson CM, Kelsey G (1999) A cluster of oppositely imprinted transcripts at the *Gnas* locus in the distal imprinting region of mouse chromosome 2. Proc Nat Acad Sci 96(7):3830–3835
Piras G, El Kharroubi A, Kozlov S, Escalante-Alcalde D, Hernandez L, Copeland NG, Gilbert DJ, Jenkins NA, Stewart CL (2000) Zac1 (Lot1), a potential tumor suppressor gene, and the gene for epsilon-sarcoglycan are maternally imprinted genes: identification by a subtractive screen of novel uniparental fibroblast lines. Mol Cell Biol 20(9):3308–3315
Plagge A, Gordon E, Dean W, Boiani R, Cinti S, Peters J, Kelsey G (2004) The imprinted signaling protein XL alpha s is required for postnatal adaptation to feeding. Nat Genet 36(8):818–826. doi:10.1038/ng1397
Redrup L, Branco MR, Perdeaux ER, Krueger C, Lewis A, Santos F, Nagano T, Cobb BS, Fraser P, Reik W (2009) The long noncoding RNA Kcnq1ot1 organises a lineage-specific nuclear domain for epigenetic gene silencing. Development 136(4):525–530. doi:10.1242/dev.031328
Reik W, Walter J (1998) Imprinting mechanisms in mammals. Curr Opin Genet Dev 8(2):154–164
Reik W, Walter J (2001) Genomic imprinting: parental influence on the genome. Nat Rev Genet 2(1):21–32. doi:10.1038/35047554
Reik W, Dean W, Walter J (2001) Epigenetic reprogramming in mammalian development. Science 293(5532):1089–1093. doi:10.1126/science.1063443293/5532/1089
Sandell LL, Guan XJ, Ingram R, Tilghman SM (2003) Gatm, a creatine synthesis enzyme, is imprinted in mouse placenta. Proc Natl Acad Sci USA 100(8):4622–4627. doi:10.1073/pnas.0230424100
Sanz LA, Chamberlain S, Sabourin JC, Henckel A, Magnuson T, Hugnot JP, Feil R, Arnaud P (2008) A mono-allelic bivalent chromatin domain controls tissue-specific imprinting at Grb10. EMBO J 27(19):2523–2532. doi:10.1038/emboj.2008.142
Schulz R, Menheniott TR, Woodfine K, Wood AJ, Choi JD, Oakey RJ (2006) Chromosome-wide identification of novel imprinted genes using microarrays and uniparental disomies. Nucleic Acids Res 34(12):e88. doi:10.1093/nar/gkl461
Schulz R, Woodfine K, Menheniott TR, Bourc'his D, Bestor T, Oakey RJ (2008) WAMIDEX: a web atlas of murine genomic imprinting and differential expression. Epigenetics 3(2):89–96. pii:5900
Schulz R, McCole RB, Woodfine K, Wood AJ, Chahal M, Monk D, Moore GE, Oakey RJ (2009) Transcript- and tissue-specific imprinting of a tumour suppressor gene. Hum Mol Genet 18(1):118–127. doi:ddn32210.1093/hmg/ddn322
Schulz R, Proudhon C, Bestor TH, Woodfine K, Lin CS, Lin SP, Prissette M, Oakey RJ, Bourc'his D (2010) The parental non-equivalence of imprinting control regions during mammalian development and evolution. PLoS Genet 6(11):e1001214. doi:10.1371/journal.pgen.1001214
Smallwood SA, Tomizawa S, Krueger F, Ruf N, Carli N, Segonds-Pichon A, Sato S, Hata K, Andrews SR, Kelsey G (2011) Dynamic CpG island methylation landscape in oocytes and preimplantation embryos. Nat Genet 43(8):811–814. doi:10.1038/ng.864ng.864
Smith FM, Holt LJ, Garfield AS, Charalambous M, Koumanov F, Perry M, Bazzani R, Sheardown SA, Hegarty BD, Lyons RJ, Cooney GJ, Daly RJ, Ward A (2007) Mice with a disruption of the imprinted Grb10 gene exhibit altered body composition, glucose homeostasis, and insulin signaling during postnatal life. Mol Cell Biol 27(16):5871–5886. doi:10.1128/MCB.02087-06
Surani A, Barton SC, Norris ML (1984) Development of reconstituted mouse eggs suggests imprinting of the genome during gametogenesis. Nature 308:548–550
Tanaka M, Puchyr M, Gertsenstein M, Harpal K, Jaenisch R, Rossant J, Nagy A (1999) Parental origin-specific expression of Mash2 is established at the time of implantation with its imprinting mechanism highly resistant to genome-wide demethylation. Mech Dev 87(1–2):129–142. pii:S0925-4773(99)00158-6
Tang F, Barbacioru C, Nordman E, Li B, Xu N, Bashkirov VI, Lao K, Surani MA (2010) RNA-Seq analysis to capture the transcriptome landscape of a single cell. Nat Protoc 5(3):516–535. doi:10.1038/nprot.2009.236

Thorvaldsen J, Duran JL, Bartolomei MS (1998) Deletion of the H19 differentially methylated domain results in loss of imprinted expression of H19 and Igf2. Gene Dev 12:3693–3702

Umlauf D, Goto Y, Cao R, Cerqueira F, Wagschal A, Zhang Y, Feil R (2004) Imprinting along the Kcnq1 domain on mouse chromosome 7 involves repressive histone methylation and recruitment of Polycomb group complexes. Nat Genet 36(12):1296–1300. doi:10.1038/ng1467

Wang Y, Joh K, Masuko S, Yatsuki H, Soejima H, Nabetani A, Beechey CV, Okinami S, Mukai T (2004) The mouse Murr1 gene is imprinted in the adult brain, presumably due to transcriptional interference by the antisense-oriented U2af1-rs1 gene. Mol Cell Biol 24(1):270–279

Williamson CM, Ball ST, Nottingham WT, Skinner JA, Plagge A, Turner MD, Powles N, Hough T, Papworth D, Fraser WD, Maconochie M, Peters J (2004) A cis-acting control region is required exclusively for the tissue-specific imprinting of Gnas. Nat Genet 36(8):894–899. doi:10.1038/ng1398

Williamson CM, Blake A, Thomas S, Beechey CV, Hankcock J, Cattanach BM, Peters J (2012) Mouse imprinting data and references. http://www.har.mrc.ac.uk/research/genomic_imprinting/

Wood AJ, Roberts RG, Monk D, Moore GE, Schulz R, Oakey RJ (2007) A screen for retrotransposed imprinted genes reveals an association between x chromosome homology and maternal germ-line methylation. PLoS Genet 3 (2 e20):192–203

Wood AJ, Schulz R, Woodfine K, Koltowska K, Beechey CV, Peters J, Bourc'his D, Oakey RJ (2008) Regulation of alternative polyadenylation by genomic imprinting. Genes Dev 22(9):1141–1146. doi:10.1101/gad.473408

Xie W, Barr CL, Kim A, Yue F, Lee AY, Eubanks J, Dempster EL, Ren B (2012) Base-resolution analyses of sequence and parent-of-origin dependent DNA methylation in the mouse genome. Cell 148(4):816–831. doi:10.1016/j.cell.2011.12.035

Yamasaki K, Joh K, Ohta T, Masuzaki H, Ishimaru T, Mukai T, Niikawa N, Ogawa M, Wagstaff J, Kishino T (2003) Neurons but not glial cells show reciprocal imprinting of sense and antisense transcripts of Ube3a. Hum Mol Genet 12(8):837–847

Yamasaki Y, Kayashima T, Soejima H, Kinoshita A, Yoshiura K, Matsumoto N, Ohta T, Urano T, Masuzaki H, Ishimaru T, Mukai T, Niikawa N, Kishino T (2005) Neuron-specific relaxation of Igf2r imprinting is associated with neuron-specific histone modifications and lack of its antisense transcript Air. Hum Mol Genet 14(17):2511–2520. doi:10.1093/hmg/ddi255

Yu S, Yu D, Lee E, Eckhaus M, Lee R, Corria Z, Accili D, Westphal H, Weinstein LS (1998) Variable and tissue-specific hormone resistance in heterotrimeric Gs protein alpha-subunit (Gsalpha) knockout mice is due to tissue-specific imprinting of the gsalpha gene. Proc Natl Acad Sci USA 95(15):8715–8720

Zwart R, Sleutels F, Wutz A, Schinkel AH, Barlow DP (2001) Bidirectional action of the Igf2r imprint control element on upstream and downstream imprinted genes. Genes Dev 15(18):2361–2366. doi:10.1101/gad.206201