UNIVERSITY OF LIVERPOOL

Epigenetic features, allelic expression and transcriptional regulatory elements at the *Peg13, Trappc9, Ago2* imprinted gene cluster in brain and neural cells

Thesis written in accordance with the requirements of the University of Liverpool for the degree of Doctor in Philosophy

By

Michael Claxton

March 2023

DECLARATION

I hereby declare that except where specific reference is made to the work of others, the contents of this thesis are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this or any other university. This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others except as specified in the text. This dissertation contains fewer than 55,000 words including bibliography, footnotes, tables and equations and has fewer than 40 Figures.

Michael Claxton

March 2023

Abstract

Genomic imprinting is a molecular mechanism that causes genes to be expressed in a parent-of-origin-specific manner due to epigenetic modifications to the genome. This results in mono-allelic or heavily biased expression of one allele, while the other remains inactive. For certain genes, imprinted allelic expression may be tissue-specific and reliant on CTCF-influenced enhancer-promoter interactions. The *Peg13* imprinted cluster, located on chromosome 15qD3 in the mouse genome, is associated with neurodevelopmental disorders such as Birk-Barel syndrome and non-syndromic intellectual disability. It consists of canonical imprinted genes that are conserved between mouse and human and brain-specific imprinted genes in the mouse. The former consists of *Peg13*, a conserved imprinted gene that exhibits a strong, non-tissue-specific, paternal expression bias, while the latter consists of *Trappc9*, *Chrac1* and *Ago2*, which display a maternal allelic expression bias of ~75% in the murine brain but exhibit biallelic expression in peripheral mouse tissues and in humans. The regulation of allele-specific expression of these genes is thought to result from a differentially methylated CpG region near the promoter of *Peg13*. To date, most imprinted expression studies have been conducted on bulk tissue data, which seeks to identify tissue-specificity for multiple imprinted clusters, novel imprinted genes and conservation of imprinting between species. However, the finding of such allelic expression biases generalised on the tissue lysate level raises the fundamental question of whether these patterns are conserved in each cell or whether there is variability and mosaicism in allelic expression between individual cells of the imprinted tissue.

To address this, I characterised the allelic expression of some of these imprinted genes in both bulk tissues and at a single-cell level using newborn C57BL/6J x cast/EiJ hybrid mice to

generate identifiable SNPs and determine allelic expression ratios. In Chapter 3, I discuss my use of three tissue samples: whole-brain lysates, cultured neural stem cells (NSCs) isolated from hippocampus tissue, and kidney, to confirm previously identified brain-specific imprinted expression for genes within the *Peg13* cluster. Using a combination of pyrosequencing and Sanger sequencing, I identified that *Trappc9*, *Ago2*, and *Chrac1* all display brain-specific imprinted expression, with varying degrees of strength, but all preferentially expressed from the maternal allele in brain tissue lysate, whereas kidney samples exhibited expected biallelic expression. Furthermore, I confirmed that both *Kcnk9* and *Peg13* do not display tissue-specific imprinting, with both whole-brain lysates and kidney samples showing preferential expression from the maternal and paternal allele, respectively. Interestingly, the data also showed that *Trappc9* and *Ago2* are not imprinted in hippocampus derived NSCs, while *Peg13* retains its strong bias of paternal allele expression in these samples. Additionally, in Chapter 3, I discuss methylation analysis of bulk tissue samples for CpG islands (CGIs) located proximal to the promoter region of each gene within the cluster. This analysis revealed that methylation patterns were consistent with current literature, displaying hypomethylation of all CGI regions within the cluster, with the exception of the CGIs located at *Trappc9* exon 2 and the *Peg13* differentially methylated region (DMR), which both exhibited high levels of methylation.

Chapter 4 focuses on answering whether single cells within a known imprinted tissue exhibit the same allele-specific expression pattern or show variability between cells. To achieve this, I cultured both neural progenitor (neurosphere) cells (NPCs) and *in vitro* differentiated neurons derived from newborn C57BL/6J x cast/EiJ hybrid mice. I used the single-cell genotyping, expression, and methylation (GEM) technique, combined with Sanger sequencing, to determine allelic expression patterns for *Peg13, Trappc9*, and *Ago2*. The

results indicated that single-cell imprinted expression was not uniform, but instead contained several variable states of allelic expression in individual NPCs and neurons. Regarding *Peg13*, while the majority of cells were in line with bulk tissue data, a small proportion of cells deviated from this expected paternal allele bias, exhibiting equal biallelic or even mono-allelic maternal expression. Furthermore, for *Trappc9* and *Ago2*, I identified a spectrum of expression states ranging from mono-allelic maternal, mono-allelic paternal, and varying levels of preferential biallelic expression. However, while the presence of multiple imprinting expression patterns within a cell population highlights an increased complexity of imprinting regulation, as a whole, the ratio of variable allele-specific expression was reflective of the bulk tissue data. In addition to analysing imprinted expression, by utilising the methylation-sensitive restriction enzyme BstUI, I also attempted to identify the presence of methylation at the promoter-proximal CGIs in single cells for genes within this cluster. The results of this analysis identified that the majority of single cells appear to be in line with the bulk tissue methylation data.

Finally, in Chapter 5, I discuss a potential molecular mechanism for the regulation of imprinted expression in this cluster. Through interrogation of the ENCODE3, UCSC and Ensembl genome browsers, I identified the location of candidate brain-specific regulatory elements relative to a CTCF site located proximal to *Peg13* in mouse that showed potential for regulating gene expression. To investigate this, I transfected promoter-reporter gene constructs into primary neurons isolated from newborn C57BL/6 mice and fibroblasts. Results showed that several of these regulatory elements exhibit either tissue-specific or general silencer activity, specific to the individual element, which may potentially contribute to the regulation of imprinted expression bias of the *Trappc9* gene. In conclusion, the expression of tissue-specific imprinted genes is likely more complex and molecularly

dynamic than initially believed within single cells at any specific time point. Single cells within an imprinted tissue appear to exhibit a wide range of allele-specific expression patterns that can deviate from the bulk-tissue standard. The mechanism behind this variation is yet to be elucidated. However, further investigation of the driving factors behind this regulation will establish a deeper understanding of imprinted single-cell transcriptomics.

Table of contents

List of Figures

List of Tables

List of Abbreviations

4C	Circular chromosome conformation capture
Ago2	Argonaute RISC Catalytic Component 2
AraC	Cytosine β-D-arabinofuranoside
Btn	Biotinylation
CGI	CpG island
ChIP	Chromatin Immunoprecipitation
Chrac1	Chromatin Accessibility Complex Subunit 1
CIP	Calf Intestinal alkaline Phosphatase
COPI	Coat Protein
CTCF	CCCTC-binding factor
Dlk1	Delta Like Non-Canonical Notch Ligand 1
DMR	Differentially Methylated Region
DNMT	DNA Methyltransferase
EGF	Epidermal Growth Factor
ENCODE	Encyclopaedia of DNA elements
ESC	Embryonic Stem Cells
ESS	Evolutionary Stable Strategy
FGF	Fibroblast Growth Factor
GEF	Guanine Exchange Factor
HAT	Histone Acetyltransferase
HDAC	Histone Deacetylase
HDM	Histone Demethylase
HEK293	Human embryonic kidney 293
HMT	Histone Methyltransferase
HP1	Heterochromatin Protein 1
HSMM	Human Skeletal Muscle Myoblast
ICR	Imprinting Control Region

Igf2	Insulin-like Growth Factor 2
Igf2r	Insulin-like Growth Factor 2 Receptor
Kcnk9	Potassium Channel Subfamily K member 9
KO	Knock out
KRAB	Krüppel Associated Box
LINE	Long Interspersed Nuclear Elements
LTR	Long Terminal Repeat
miRNA	micro RNA
ncRNA	non-coding RNA
NS	Neurospheres
NSC	Neural Stem Cell
NuRD	Nucleosome Remodelling and Histone Deacetylation
PCR	Polymerase Chain Reaction
Peg13	Paternally Expressed Gene 13
piRNA	piwi-associated RNA
QUMA	Quantification tool for Methylation Analysis
RISC	RNA-Induced Silencing Complex
RNAi	RNA Interference
RT	Reverse Transcription
SAM	S-adenosyl-L-methionine
Sc-GEM	Single cell- Genotyping, Expression, Methylation
SCRAM	Single-cell Restriction Analysis of Methylation
ScTAM-seq	Single-cell Targeted Analysis of the Methylome
SETDB1	SET Domain Bifurcated Histone Lysine Methyltransferase 1
shRNA	short-hairpin RNA
SINE	Small Interspersed Nuclear Elements
siRNA	small interfering RNA
SLC22a2	Solute Carrier Family 22 Member 2
SNP	Single Nucleotide Polymorphism
TAD	Topologically Associating Domain
TASK3	TWIK-related acid-sensitive K1 channel 3

TRAPP	Trafficking Protein Particle
Trappc9	Trafficking Protein Particle Complex 9
Trim28	Tripartite Motif Containing 28
UBE3A	Ubiquitin Protein Ligase E3A
WT	Wild Type
ZFP57	Zinc Finger Protein homolog 57

Chapter 1 – Introduction

1.1 Epigenetics

1.1.1 History and definition

The term epigenetics, derived from the Greek prefix *epi*- over/ on top of, was first introduced in 1942 by C.H. Waddington, who in turn related it to the 17th-century concept of 'epigenesis', and defined it as the complex of developmental processes between genes interacting with their environment to produce a phenotype (Deichmann, 2016; G. Zhang & Pradhan, 2014). However, since its introduction in 1942, the definition of this process has been changed and adapted to numerous concepts and fields of research, particularly within the 21st century (Bird, 2007; Felsenfeld, 2014; Haig, 2012; Morange, 2013). It was not until 2009, when a conference hosted by the Banbury Conference Centre and Cold Spring Harbor Laboratory intending to discuss chromatin-based epigenetics, that a consensus definition for the term was reached, being defined as "An epigenetic trait is a stably heritable phenotype resulting from changes in a chromosome without alterations in the DNA sequence" (Berger et al., 2009). In this case, heritable refers to both mitotic epigenetic inheritance through mitosis of somatic cells and transgenerational epigenetic inheritance through the germline such as in the case of genomic imprinting. Therefore, by this definition, epigenetics is the regulation of gene activation or repression through modification of the DNA, or its associated chromatin proteins, that does not result in a change of the DNA sequence but is inherited.

Regulation of gene expression is imperative for the correct functioning of cells and helps promote differentiation by ensuring genes related to the functions of differentiated cells are activated, while other genes are silenced. Many well-known processes have been identified

as the result of epigenetic regulation including X-chromosome inactivation, genomic imprinting and gene silencing. Additionally, the incorrect establishment of epigenetic marks, resulting in abnormal regulation of transcription has been associated with multiple disease phenotypes including cancer and chromosomal instabilities (Egger et al., 2004).

1.1.2 Epigenetic mechanisms- DNA methylation

The mechanisms by which epigenetic modifications regulate gene transcription have been extensively studied in the past few decades, with three key determinants having been identified: DNA methylation, histone modifications and non-coding RNAs (ncRNA) (Aboud et al., 2022; G. Zhang & Pradhan, 2014). Of the three factors, DNA methylation is the most well-studied epigenetic mark in the mammalian genome and is thought to be associated with the binding capability of transcription factors to the promoter regions of genes (G. Zhang & Pradhan, 2014). The process of DNA methylation is catalysed by DNA methyltransferase (DNMTs) enzymes, wherein a methyl group (CH_3) is added to the cytosine base of a CG dinucleotide (Fig 1.2). These DNMTs regulate DNA methylation by either recognising hemimethylated DNA, where they then proceed to methylate the complementary strand (DNMT1) or they establish new methylation sites that had no previous marks i.e. *de novo* methylation (DNMT3a, b) (Elhamamsy, 2016; Greenberg & Bourc'his, 2019). These CG dinucleotides are often found in an environment in which they are surrounded by other CG dinucleotides forming what is known as a CpG island (CGI), a region ~300-3000 bp long possessing a G+C content > 55% (Takai & Jones, 2002).

These CGIs are frequent targets for epigenetic regulation through DNA methylation, most notably those CGIs that are found within the promoter region of a gene. Furthermore,

previous studies have reported that ~70% of gene promoter regions overlap with a CGI indicating how prevalent DNA methylation is in regulating transcription (Aboud et al., 2022; Saxonov et al., 2006). The presence of DNA methylation at a promoter CGI region is indicative of repression of transcription as the methylated cytosines are targets of gene suppressor proteins such as the KRAB-Trim28 complex, which is associated with imprinting maintenance, and limit the interaction of DNA with transcription factors. Additionally, DNA methylation promotes the formation of heterochromatin, resulting in the tightening of DNA around the nucleosome and preventing access of transcriptional machinery to bind to the DNA (Fig 1.2) (McMahon et al., 2017; Moore et al., 2012).

Nucleosomes are repeat elements of chromatin that consist of ~150 bp of DNA wrapped around a histone octamer consisting of two copies of four core histone proteins: H2A, H2B, H3 and H4. Each nucleosome is connected through a fragment of linker DNA of variable length, usually, ~50-100 bp, which binds to the linker histone H1. The core nucleosome plus the H1 linker histone that make up the repeat subunits of the chromatin are called a chromatosome (Fig 1.1) (Mariño-Ramírez et al., 2005; Tachiwana et al., 2021). There are numerous examples of this repression occurring both in normal development, such as in X-chromosome inactivation, in tissue-specific genomic imprinting (Moore et al., 2012) as well as in disease models where the acquisition of DNA methylation in an unregulated manner results in harmful phenotypes. In fact, many cancers show hypermethylation of tumour suppressor genes and/ or hypomethylation of proto-oncogenes resulting in tumour carcinogenesis (McMahon et al., 2017).

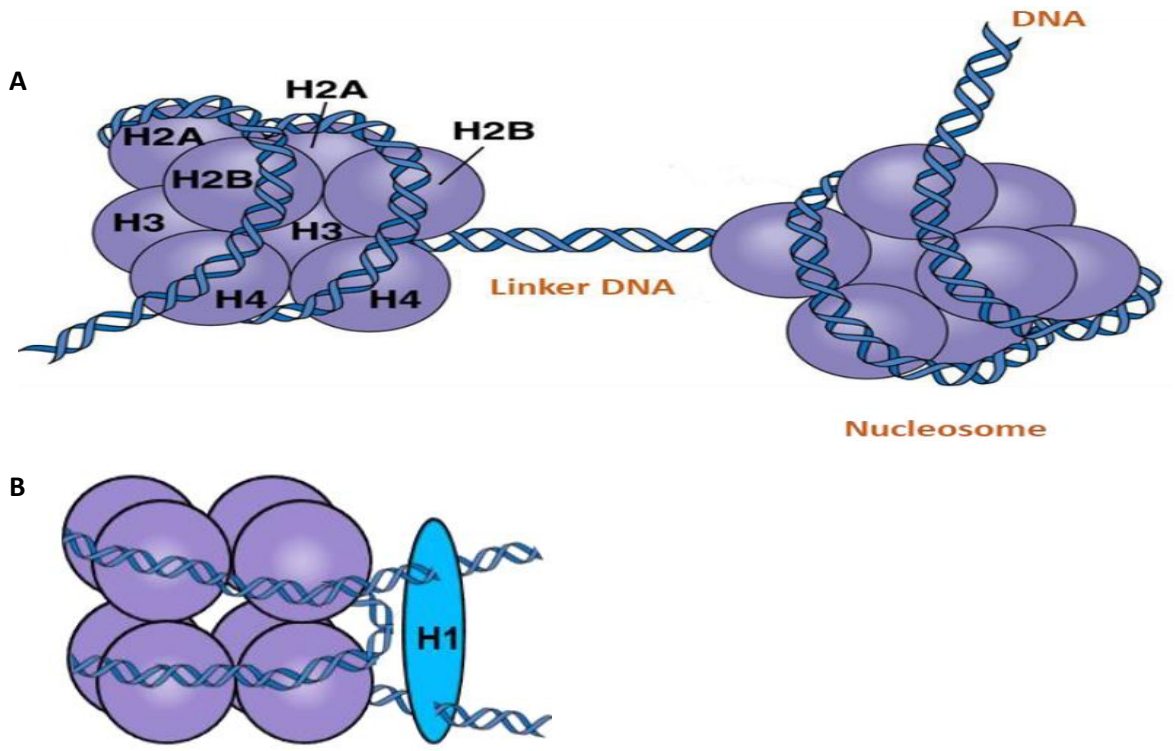


Figure 1.1: Structure of a nucleosome. A) A section of DNA ~150bp wrapped around a histone octamer comprised of two copies of four main core histone proteins: H2A, H2B, H3 and H4. B) Nucleosome units are connected by a segment of linker DNA bound to histone 1. Adapted from (Pulix, 2017)

1.1.3 Histone modification

The second form of epigenetic regulation is through post-translational modification of histone proteins, usually at their N-terminus also commonly referred to as the 'histone tail', of which there are four main categories: acetylation, methylation, phosphorylation and ubiquitination (Alaskhar Alhamwe et al., 2018; Bannister & Kouzarides, 2011). Acetylation of histones is regulated by two enzymes, histone acetyltransferases (HATs) and histone deacetylases (HDAC), that frequently target positively charged lysine residues for modification (Bannister & Kouzarides, 2011). These HAT enzymes utilise acetyl CoA to transfer an acetyl group to the ε-amino group of a lysine side chain removing the lysine's positive

charge which in turn weakens the DNA-histone interaction resulting in the loosening of the chromatin from the nucleosome. This more open form of the chromatin, dubbed euchromatin, promotes access of RNA polymerase II and other transcription factor binding which facilitate transcription (Fig 1.2). Examples of this are the acetylation of lysine 9 and lysine 27 on histone 3 (H3K9ac and H3K27ac respectively) which are both associated with activation of transcription and so are frequently found at promoter and enhancer regions. HDAC enzymes work inversely to this, removing the acetyl group and restoring the lysine residues' positive charge, thus working to repress expression (Bannister & Kouzarides, 2011; Verdone et al., 2006).

Histone methylation frequently targets lysine and arginine residues and can be associated with both activation and repression of transcription, although it is often more frequently associated with repression (Gupta et al., 2010; C. Martin & Zhang, 2005). Unlike acetylation, histone methylation does not alter the charge of the amino acid residue but instead directly influences the ability of the DNA to recruit and bind different regulatory proteins. Additionally, histone methylation has the capability for multiple methyl group binding, monomethylation (me1), dimethylation (me2) or trimethylation (me3) for lysine and mono or dimethylation for arginine (Alaskhar Alhamwe et al., 2018; Bannister & Kouzarides, 2011). The reaction is catalysed by a histone methyltransferase (HMT), which transfers up to three methyl groups from the cofactor S-adenosyl-L-methionine (SAM) to a lysine residue (replacing the hydrogen on the primary amine) or up to two methyl groups to an arginine residue (adding them to the nitrogen atom), depending on the type of HMT enzyme that is used (Blanc & phane Richard, 2017). An alternate enzyme, histone demethylase (HDM) reverses the reaction, removing methyl groups from the amino acid residue (Kaniskan et al., 2018; Morera et al., 2016). The combination of alternate amino acid targets and variation in

the number of methyl groups deposited determines whether the modification promotes or represses transcription. For example, methylation of lysine 4 on histone 3 (H3K4me) is associated with the activation of transcription while trimethylation of lysine's 9 or 27 on histone 3 (H3K9me3 and H3K27me3) correlates with the formation of heterochromatin and repression of transcription (Fig 1.2) (Aboud et al., 2022; Esteller, 2008; Nestorov et al., 2013).

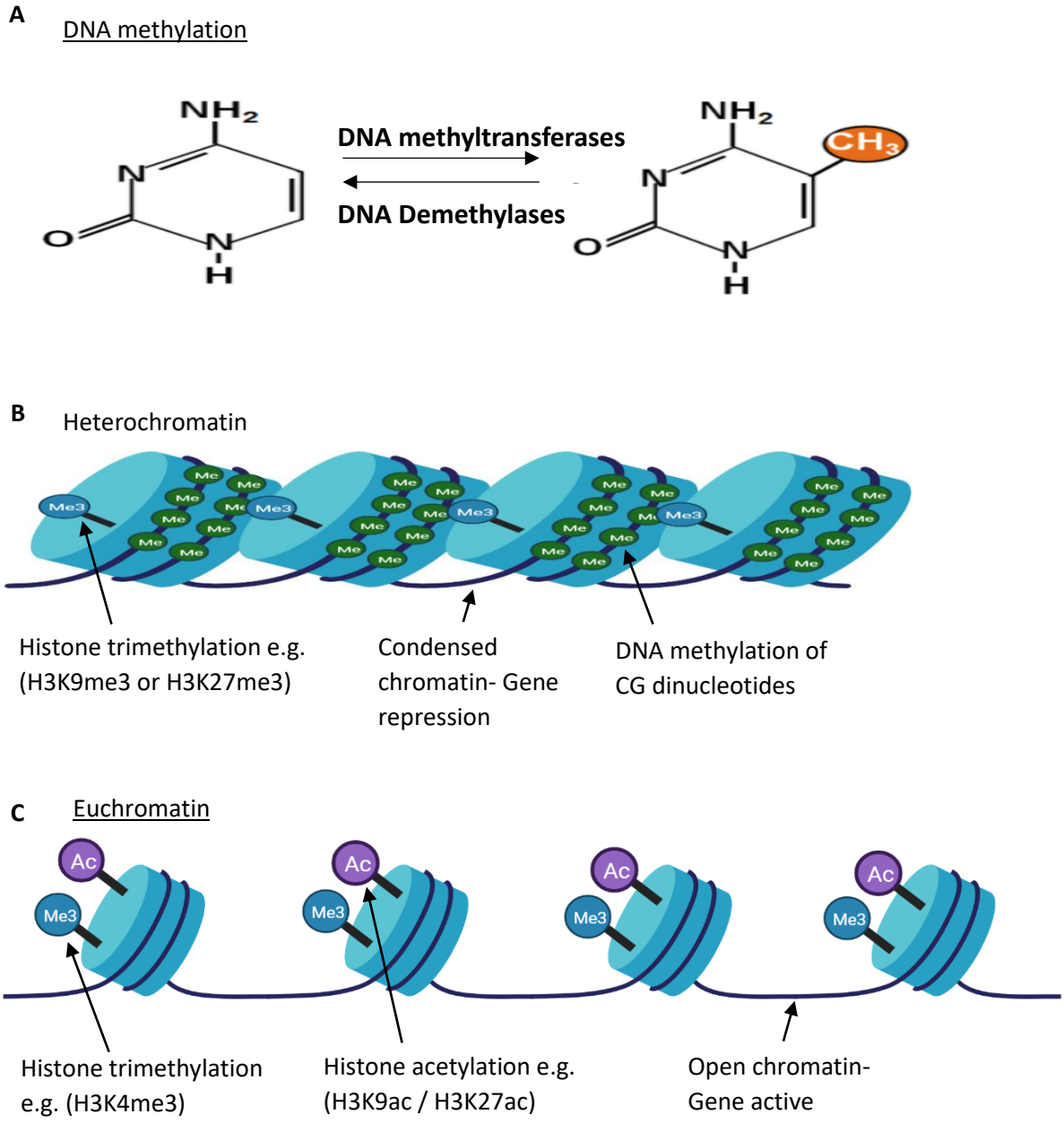


Figure 1.2: Schematic of the process of DNA methylation and formation of heterochromatin and euchromatin for regulation of transcription. A) Epigenetic modification of the DNA sequence via the addition of a methyl group (CH_3), catalysed by a DNA methyltransferase enzyme, to the 5th carbon of a cytosine base at a CG dinucleotide to form 5-methylcytosine. The reaction is reversible through a reaction catalysed by DNA demethylase. B) Formation of transcriptionally silent heterochromatin through epigenetic modification. The addition of methyl groups to the DNA and specific histone modifications such as H3K9me3 and H3K27me3 result in the chromatin tightening around the nucleosome, preventing transcriptional machinery from binding to the DNA and therefore repressing gene expression. C) Formation of euchromatin through specific histone modifications such as H3K4me3, H3K9ac and H3K27ac. These modifications result in the chromatin loosening from the nucleosome becoming open and available for transcription factor binding resulting in gene transcription.

The final two histone modifications, phosphorylation and ubiquitination, are able to work in tandem with the previous two histone modifications and can assist in the establishment of histone acetylation and methylation marks (Alaskhar Alhamwe et al., 2018). Histone phosphorylation is regulated by two types of enzymes, kinases (addition) and phosphatases (removal). This modification has been shown to regulate gene transcription, such as by phosphorylation of serine 10 on histone 3 (H3S10ph) which has been proven to influence acetylation of both H3K9 and H3K14 as well as interacting with H4K16ac (D. G. Edmondson et al., 2002; Lo et al., 2000; Zippo et al., 2009). In addition to regulating transcriptional activity, phosphorylation of histones has been associated with DNA damage repair mechanisms and control of chromatin compaction in mitosis and meiosis (Bannister & Kouzarides, 2011; Rossetto et al., 2012).

Protein ubiquitination is a post-translational modification that affects a wide range of cellular functions and signalling pathways in eukaryotes. The addition and removal of ubiquitin onto histone proteins is regulated by the enzymes histone ubiquitin ligase (addition) and ubiquitin-specific peptidase (removal), the latter also being known as deubiquitinating enzymes (DUBs), within the nucleus (Cao & Yan, 2012; Ravid & Hochstrasser, 2008; Schwertman et al., 2016). The effect of ubiquitination is dependent on the histone modified with histone 2A monoubiquitination (H2Aub) being associated with gene silencing while the same modification on histone 2B (H2Bub) is correlated with the activation of transcription. Furthermore, ubiquitination of histones, like phosphorylation, can promote alternate epigenetic marks with ubiquitination of histone 3 (H3ub) promoting acetylation of the same histone and therefore promoting transcription (Cao & Yan, 2012; Schwertman et al., 2016; Weake & Workman, 2008; X. Zhang et al., 2017).

1.1.4 Non-coding RNAs

The third epigenetic mechanism of non-coding RNA is also the most recently elucidated and least understood of the three. A ncRNA is a synthesised and functional RNA molecule that is transcribed from the DNA, but not translated into a protein. These ncRNAs can be further subdivided into two categories: small non-coding RNAs (sncRNAs) comprised of fewer than 200 nucleotides and long non-coding RNAs (lncRNAs) consisting of more than 200 nucleotides. Additionally, the sncRNAs can further be separated into small interfering RNAs (siRNAs), micro RNAs (miRNA) and Piwi-associated RNAs (Table 1.1) (piRNAs) (Aboud et al., 2022; Riedmann & Schwentner, 2010; Wei et al., 2017; P. Zhang et al., 2019). Once thought to be redundant genomic sequence, genes that code for these ncRNAs are now thought to play a pivotal role in epigenetic regulation and has been theorised to be strongly associated with gene silencing. Previous studies have found evidence of ncRNAs participating in DNA methylation and histone modification to repress transcription (Wei et al., 2017) and lncRNAs have been shown to alter chromatin conformation by promoting the formation of heterochromatin to silence gene expression (Chisholm et al., 2012; Frías-Lasserre & Villagra, 2017).

Table 1.1: Non-coding RNAs associated with epigenetic regulation of transcription. Adapted from (Wei et al., 2017).

Name	**Size**	**Source**	**Primary function(s)**
siRNA	19-24bp	Double-stranded RNA	Silence mRNA translation via degradation of mRNA
miRNA	19-24bp	Pri-miRNA	Silence mRNA translation
piRNA	26-31bp	Long single chain precursor	Transposon repression and DNA methylation
lncRNA	> 200bp	Multiple sources	Genomic imprinting X-chromosome inactivation

1.2 Genomic imprinting

1.2.1 History and definition

With the exception of certain genes found on the Y chromosome, most genes possess two copies, one found on the maternal chromosomes and the other on the paternal chromosome. The majority of these genes are expressed from both alleles equally, termed biallelic expression, in the majority of cells. However, certain genes are "marked" during gametogenesis via epigenetic modifications, such as DNA methylation, that result in them being imprinted (Bajrami & Spiroski, 2016; Ferguson-Smith & Bourc'his, 2018; Y. Li & Sasaki, 2011). The term "imprinting" was first defined by Helen Crouse in 1960 when she related it to the programmed elimination of one or two paternally derived X chromosomes in *sciarid* flies. Crouse determined that "the imprint a chromosome possesses is unrelated to the genetic constitution of the chromosome but is instead determined by the sex of the germ line through which the chromosome was inherited" (Crouse, 1960; Tucci et al., 2019).

However, it was not until the mid-1980s when experiments performed by Davor Solter and Azim Surani, in two independent studies, began to form the foundation for the study of epigenetic inheritance and gene regulation during development (Ferguson-Smith & Bourc'his, 2018; McGrath & Solter, 1984; Peters, 2014; Surani et al., 1984; Tucci et al., 2019). Solter and Surani generated mouse embryos in which each embryo was manipulated to possess either two sets of chromosomes from the mother (bi-maternal/ gynogenetic) or two sets from the father (bi-paternal/ androgenetic), rather than one from each, which when transferred into pseudo-pregnant recipient females failed to develop appropriate embryonic/ extraembryonic tissues and died *in utero*. This study deduced that while genetically equivalent, the maternal and paternal chromosomes were not functionally equivalent and

that each parental chromosome is somehow marked in a way that distinguishes itself from the other. Additionally, it established that for normal development to take place, one set of chromosomes from each parent is required (Barton et al., 1984; McGrath & Solter, 1984; Surani et al., 1984). The mechanism behind this need for both parental alleles for normal development is genomic imprinting, which is a process that acts upon gametes to mark certain genes to express in a parent-of-origin specific manner resulting in mono-allelic, or strongly biased expression, of one allele and silencing of the other post-fertilisation. After this discovery, genetic studies conducted by Cattanach & Kirk (1985) identified that imprinted genes were not evenly distributed across the genome but rather located in specific genomic regions, often referred to as imprinting clusters. In addition to this, many imprinting clusters display a high degree of conserved linkage between species, such as the human chromosome 11p15.5 region, which has previously been associated with Beckwith-Wiedmann syndrome and contains the paternally expressed *MTR1* gene, maps to an imprinted cluster found on mouse chromosome 7 highlighting a degree of evolutionary conservation (Oakey & Beechev, 2002). A potential driving force of this evolutionary conservation of genomic imprinting could be to ensure maintenance of imprinted DNA methylation patterns throughout early embryonic development and regulation of LTR-derived transcripts to ensure viability of the embryo (Hanna & Kelsey, 2021). The first imprinted genes identified were insulin-like growth factor 2 receptor (*Igf2r*)*, Igf2* and *H19* in 1991 (Barlow et al., 1991; Bartolomei et al., 1991; DeChiara et al., 1991; Ferguson-Smith et al., 1991). The *Igf2r* gene, which codes for an intracellular transport receptor, exhibits exclusive expression from the maternal allele with additional studies identifying it as being part of a cluster of imprinted genes located on chromosome 17 of the mouse genome (Barlow et al., 1991; Zwart et al., 2001). Furthermore, *Igf2*, which codes for insulin-like

growth factor II, and is involved in foetal development, was discovered to exhibit paternal-specific expression with experiments on paternal I*gf2* deletion exhibiting growth deficiencies while maternal deletion resulted in no clear deleterious phenotype (DeChiara et al., 1991). Interestingly, *H19,* which codes for a ncRNA and possesses a role in the negative regulation (or limiting) of body weight and cell proliferation (Gabory et al., 2009), is expressed from the maternal allele and exhibits a reciprocal pattern of imprinted expression with *Igf2* which is explored in section 1.2.4 – Mechanism of imprinted gene regulation.

Previous studies on genomic imprinting have not only highlighted its necessity for normal embryonic development and postnatal processes, such as regulation of the brain, behaviour and metabolism (Cleaton et al., 2014), but have also identified multiple human syndromes as a result of abnormal parent-of-origin specific expression. Approximately 100 imprinting disorders have been identified in humans such as the two neurological disorders: Prader-Willi and Angelman syndromes and the foetal overgrowth disorder Beckwith-Wiedemann Syndrome (Table 1.2) (Butler et al., 2006; Butler, 2020; Cox et al., 2002; DeBaun et al., 2003; Driscoll et al., 2017; Manipalviratn et al., 2009; Nicholls et al., 1989). The advancement of technology and new genome-wide assays have contributed considerably to the identification of these imprinted genes, our understanding of their function and the classification of imprinting syndromes. However, this advancement in technology has also brought about an increase in the complexity of imprinting. While many imprinted genes follow the initial description of genomic imprinting as the process of exclusive expression from one parental allele, many imprinted genes have now been found to exhibit preferential (> 70%) but not complete mono-allelic expression as well as multiple genes exhibiting a degree of tissue and developmental stage specificity (Andergassen et al., 2017; Babak et al., 2015; Crowley et al., 2015; Elena Martinez et al., 2014; Perez et al., 2015).

Table 1.2: List of several imprinting disorders, highlighting the chromosome affected, the aetiology of the condition and exhibited symptoms (Adapted from Eggermann et al, 2015).

Imprinting disorder	**Affected region**	**Cause**	**Symptoms**
Transient Neonatal Diabetes Mellitus (TNDM)	*6q24*: *PLAGL1*: alt-TSS	UPD of pat chromosome 6 Paternal duplications Methylation defects	IUGR, Transient diabetes, hyperglycaemia without ketoacidosis
Silver-Russell syndrome (SRS; Russell-Silver Syndrome, RSS)	*7* and *11p15 H19/IGF2*	UPD of mat chromosome 7 Hypomethylation of the *H19/IGF2*: IG DMR	IUGR, postnatal growth restriction, feeding difficulties, distinct facial features
Beckwith-Wiedemann	*11p15-H19/IGF2 and KCNQ1OT1*	UPD of pat *11q15* Abnormal methylation patterns at the H19/IGF2 and *KCNQ1OT1* DMR's Sporadic and familial point mutations	Pre- and postnatal overgrowth, increased risk of tumour formation, hemihypertrophy
Kagami-Ogata syndrome (KOS14, upd(14) pat syndrome)	*14q32-MEG3/DLK1: IG DMR*	UPD of pat chromosome 14 Maternal deletion Aberrant methylation	IUGR, polyhydramnios, abdominal and thoracal wall defects, bell-shaped thorax
Angelman syndrome	*15q11-q13*	Maternal deletion Point mutations UPD of paternal Chromosome 15 Aberrant methylation	mental retardation, microcephaly, no speech, ataxia, severe developmental delay, excessive laughing
Prader-Willi syndrome	*15q11-q13*	Paternal deletion UPD of maternal chromosome 15 Aberrant methylation	Mental retardation, neonatal hypotonia, hypogenitalism, hypopigmentation, obesity/hyperphagia

1.2.2 Evolution of genomic imprinting

Since the discovery of the first imprinted genes in 1991, approximately 150 imprinted genes in mice and 100 genes in humans have been discovered, with other species such as cows, pigs and dogs also containing variable numbers of imprinted genes (www.geneimprint.com) (Allach El Khattabi et al., 2019; Geneimprint : Genes, 2023). The majority of these imprinting clusters are conserved between human and mouse. However, the imprinting status of some genes, such as *Trappc9*, which has previously been shown to exhibit maternal biased expression in mouse brain, does not exhibit conservation and is biallelically expressed in humans (Court et al., 2014; Tucci et al., 2019). The evolutionary drive responsible for the

formation of genomic imprinting as a regulatory mechanism has been an intensely debated subject over the last few decades with multiple theories having been proposed: Haig and colleagues' kinship theory, Day and Bonduriansky's sexual antagonism theory, Wolf and Hager's maternal–offspring coadaptation theory, and the theory of imprinted genes arising from silencing of transposable elements, with each having fundamentally different perspectives on the significance of imprinting (Holman & Kokko, 2014; Patten et al., 2014; Spencer & Wolf, 2014).

Kinship theory, which was first proposed by Haig and colleagues (Haig, 2003, 2004; Haig & Westoby, 1989), suggests that genomic imprinting is a mechanism that evolved to alter gene dosage due to its differential effect on the fitness of matrilineal and patrilineal relatives. It argues that genomic imprinting is the result of a conflict of interest between the maternal and paternal alleles (Spencer & Wolf, 2014). This conflict arises when the alleles of an organism experience different levels of relatedness in a social environment i.e. an organism is more likely to encounter maternally related siblings compared to paternally related, due to multiple mating patterns, and therefore the expression of each allele has alternate consequences for their respective fitness. Additionally, the theory suggests that paternally expressed imprinted genes are usually associated with foetal growth and generate an increased demand for resources. Maternally expressed imprinted genes, however, have been correlated with a restriction on foetal growth and conservation of resources, due to a desire for equal resource allocation for all offspring during pregnancy and early postnatal life (i.e. paternal genes promote the fitness of an individual offspring vs maternal genes increasing the fitness of all offspring) (Patten et al., 2014). Due to this asymmetry in relatedness between siblings, certain variations, such as an increase in expression of a gene from the paternally inherited allele (e.g. a foetal growth factor), may increase the fitness of

itself but decrease inclusive fitness of the maternal allele to promote its own inheritance. An example of this would be the reciprocal expression of the paternally expressed *Igf2* and the maternally expressed *Igf2r*, which have been linked to opposing phenotypic effects (promotion and inhibition of prenatal growth respectively) (Haig, 2004). Upregulation of foetal growth factor expression from the paternal allele compared to the maternal allele would impact the growth of the individual and potentially its fitness while being possibly detrimental to the organism's maternal siblings through its demand for shared maternal resources. Furthermore, because of this imbalance of expression between paternal and maternal alleles, an evolutionarily stable strategy (ESS) for gene regulation is promoted. The allele that sought to improve its own fitness through variation of expression, in this case, the paternal allele through upregulation of a foetal growth factor, is selected for and remains active while the maternal allele, which expresses the same foetal growth factor to a lesser degree due to its role in the conservation of resources, is silenced to promote an appropriate level of gene expression rather than overexpression caused by active expression of both alleles (Haig, 1997; Haig & Westoby, 1989). While the originally proposed theory relied on multiple mating as the source of related asymmetry, further work has expanded the theory to cover alternate sources such as haplodiploidy and sex-based dispersal (Brandvain et al., 2011; Patten et al., 2014; Queller, 2003; Úbeda & Gardner, 2010). This theory discusses the idea that the paternal allele exploits the fact that, in mammals, the female is the only provider of resources to the foetus/ newborn. Therefore, the paternal allele ensures the survival of the offspring through greater resource demand to the detriment of the offspring's maternal siblings, while the maternal allele is generally associated with limiting resources to ensure survival of all offspring. However, this concept is only applicable to a

certain percentage of imprinted genes that are associated with development and resource allocation, while several imprinted genes have no apparent link to resource allocation.

The sexual antagonism theory proposed by Day and Bonduriansky relies on a sex-specific selection pressure where selection differs between males and females differs (Bonduriansky, 2007; Day & Bonduriansky, 2004). It supports the idea that selection favours genomic imprinting as the two alleles provide different variations of the expressed gene, with one allele being more adaptive to the selection pressure than the other. According to the theory, the non-equivalence between alleles is generated by two different scenarios: A) Selection is sexually antagonistic for a gene, meaning there is a selective advantage to an allele modifier (epigenetic marks) adjusting expression levels in a sex-specific way, therefore paternal alleles will be enriched for male benefit while maternal alleles will be enriched for female benefit. B) There is a net selective advantage to alteration of expression, independent of the sex of the offspring, provided the allele derived from the parent that experienced stronger selection in the previous generation is expressed (Patten et al., 2014). The only requirement for this theory of how genomic imprinting evolved is the idea that there is an original sex-specific selection pressure on the genes driving the formation of an imprinting bias. Furthermore, as this theory applies to any species with two sexes, the requirements of this theory are less broad than that of the other two and is more widely applicable. (Day & Bonduriansky, 2004; Holman & Kokko, 2014; Patten et al., 2014). However, despite sexual antagonism being more widely applicable than alternative theories, genomic imprinting has frequently been associated with mammals, with many species such as fish and reptiles possessing no identifiable imprinted genes. While the reasons for imprinting being more prevalent in mammals is not entirely clear, it may be related to the evolution of viviparity (live birth) in these animals, which requires a high degree of maternal-foetal interaction and

regulation of foetal growth. Additionally, the unique reproductive biology of mammals, such as the presence of a placenta and lactation, may have contributed to the selection pressure of genomic imprinting as a mechanism to regulate foetal growth and development (Barlow & Bartolomei, 2014).

The third theory behind how genomic imprinting may arise is via maternal-offspring coadaptation proposed by Wolf and Hager in 2006. Their theory suggests that there are specific maternal-offspring interactions and both the maternal and paternal alleles in an offspring code for alternate levels of fitness for these interactions, thus favouring the formation of genomic imprinting (Wolf & Hager, 2006, 2009). These interactions are predicted to be stronger in species that exhibit prolonged maternal care or where maternal traits have an impact on fitness related to traits of the offspring (Patten et al., 2014). There are two scenarios Wolf and Hager proposed to define this model: A) The single locus model wherein the same locus pleiotropically influences traits of both the mother and offspring. B) The two-locus model where alternate loci affect maternal and offspring traits. In the single-locus model, the two alleles at a diploid locus within the offspring exhibit different levels of fitness. This is predominantly due to the maternally derived allele being more likely to be present in the mother and therefore conveys an advantage to the offspring for maternal-offspring interactions. The expression of the maternal allele in the offspring is correlated to a higher likelihood of a phenotypic match between offspring and mother than the expression of the paternal allele. However, if the fitness of the interaction is better in a heightened mismatch between mother and offspring then it is more likely that the maternal allele will be silenced and preferential expression of the paternal allele will be exhibited (Patten et al., 2014; Wolf & Hager, 2006, 2009). Alternatively, in the two-locus model, past selection of the maternal-offspring interaction results in the offspring's maternal allele

expression being correlated with the mother's traits/ phenotype with which the offspring's phenotype must interact. Due to this selection, there is a degree of linkage disequilibrium between the two alleles, and so the interaction between mother and offspring is likely to have a higher fitness when expression in the offspring is limited to only the maternal allele due to the evolution of a degree of co-adaptation with the alleles present in the mother, thus promoting the formation of genomic imprinting and silencing of the paternal allele.

While these three theories explain alternative mechanisms of how genomic imprinting evolved as a result of parental allele-offspring interaction, an alternative theory for the evolution of genomic imprinting may revolve around the silencing of transposable elements via DNA methylation. Transposable elements are DNA sequences that have the ability to change their position within the genome. They are major components of eukaryotic genomes and are subdivided into multiple classes. These classes include: Class 1 elements, also known as retrotransposons, which mobilise through a 'copy-and-paste' mechanism, relying on reverse transcription, and can be subdivided into long terminal repeat (LTR) retrotransposons and non-LTR retrotransposons, which include both long and short interspersed nuclear elements (LINEs and SINEs). Class 2 elements, also known as DNA transposons, mobilise via a DNA intermediate, usually through a 'cut-and-paste' mechanism (Bourque et al., 2018; Wells & Feschotte, 2020). One theory for the evolution of genomic imprinting is that the need to silence these mobile DNA elements originated first as a means of protecting the genome from instability, which was achieved by DNA methylation and chromatin organisation. These novel insertions and neighbouring host genes may have evolved into imprinted genes that were selected for by evolutionary selection pressures, such as in the post-implantation epiblast and extra-embryonic ectoderm, where non-canonical imprints have been localised to endogenous retrovirus-K (ERVK) long terminal

repeats, acting as imprinted promoters (Hanna et al., 2019; Kaneko-Ishino & Ishino, 2015). Through previous KO experiments, it was identified that several of these retrotransposons have an essential role in the current mammalian developmental system as endogenously functional genes specific to mammals (Kaneko-Ishino & Ishino, 2015; Ono et al., 2006; Sekita et al., 2008). These genes, derived from transposable elements, have been shown to play a role in the formation, maintenance and endocrinological regulation of the placenta and a variety of brain functions. Examples of such genes include *PEG10* and *PEG11/* Retrotransposon-like 1 (*Rtl1*), which are both expressed from the paternal allele, originate from a retrotransposon, and are related to placental formation and interaction, with mutations resulting in embryonic lethality (Kaneko-Ishino & Ishino, 2015; Ono et al., 2006; Sekita et al., 2008). Previous studies identified that the *PEG10* gene is present in therian mammals such as the marsupial tammar wallaby (*Macropus eugenii*) but not in prototherian mammals, such as the egg-laying platypus (*Ornithorhynchus anatinus*), highlighting the close relationship of *PEG10* in placental formation and viviparity (Suzuki et al., 2007). Furthermore, the marsupial tammar DMR is the first example of a differentially methylated region (DMR) associated with genomic imprinting in marsupials and is limited to the 5' region of *PEG10*, unlike the eutherian DMR, which covers the promoter regions of both *PEG10* and the adjacent imprinted gene *SGCE.* This demonstrates both a common origin of the DMR imprinting mechanism in therian mammals and provides a demonstration that DMR-associated genomic imprinting can originate from the repression of exogenous DNA sequences and/or retrotransposons by DNA methylation. Additionally, mutations of DNMT3C result in deleterious effects associated with transposable element activation in germ cells, such as the post-natal activation of retrotransposons, due to a lack of methylation, at meiosis, subsequent instability of the meiotic chromatin landscape and

interruption of spermatogenesis by apoptosis (Dura et al., 2022). This theory denotes that imprinted genes evolved through the "domestication" of transposable elements that were initially silenced as a means of protecting the genome and then evolved into genes expressed in a parent-of-origin specific manner through evolutionary selection pressures.

1.2.3 Establishment and maintenance of imprinted genes

Through extensive genome studies, the majority of imprinted genes have been found in clusters of ~3-12 genes, spread over a variable length of DNA sequence ranging from 20 kb-3.7Mb (Edwards & Ferguson-Smith, 2007). The majority of these imprinted clusters are regulated by DNA methylation and are centred around a differentially methylated region (DMR), where only one allele is methylated. These DMRs are usually established in germline cells during oogenesis/ spermatogenesis, although secondary DMRs can be established during early embryogenesis, such as in the case of the imprinted *Nesp* gene. A DMR acts as an imprinting control region (ICR) regulating the expression of neighbouring genes within the cluster. The presence of methylation at the DMR is dependent on whether the allele was inherited from the mother or the father, although the mechanism behind why these individual alleles are selected for methyl group presence or absence is still debated (Edwards & Ferguson-Smith, 2007; Ferguson-Smith & Bourc'his, 2018; Maupetit-Méhouas et al., 2016).

A unique feature of these ICRs is their ability to resist the extensive genome reprogramming and the global DNA demethylation process that takes place post-fertilisation. In somatic cells, *de novo* DNA methylation of the genome occurs at the end of the preimplantation stage, according to developmental cell fates in different tissues and cell types, and are generally maintained throughout the organism's life. However, in germline cells, DNA

methylation is erased during the migration of primordial germ cells and later re-established during gametogenesis, according to the sex of the organism (Bartolomei & Ferguson-Smith, 2011; Messerschmidt, 2012). This demethylation in primordial germ cells is essential as all new paternal or maternal ICR methylation patterns must be established in sperm and oocytes respectively. This re-establishment of methylation status is imperative for the regulation of gene expression in a parent-of-origin-specific manner and serves to maintain imprinted gene expression over subsequent generations (Fig 1.3). The established ICRs, which act in an allele-specific manner, can be subdivided into ICRs that are methylated during oogenesis on the maternally inherited chromosome and ICRs that are methylated during spermatogenesis from the paternally inherited chromosome (Edwards & Ferguson-Smith, 2007; Plasschaert & Bartolomei, 2014). Interestingly, the DNA methylation of these parental-specific ICRs occurs at different stages of development, with "paternal ICR methylation occurring in the haploid (meiotic) phase of spermatogenesis whereas maternal ICR methylation occurs in the postnatal growth phase while oocytes are arrested at the diplotene stage of prophase I" (Lucifero et al., 2004; Messerschmidt, 2012). Furthermore, through genome-wide studies, it has been identified that paternal ICRs are typically associated with intergenic sequences and transposon repeats while maternal ICRs are frequently found at the promoters of genes and often overlap a promoter-proximal CGI sequence (Bartolomei & Ferguson-Smith, 2011; Veselovska et al., 2015). Within the mouse genome there is a higher frequency of maternal ICRs than that of paternal ICRs, likely due to maternal ICRs being located at promoter regions and thus exhibiting a stronger evolutionary pressure to be maintained compared to paternal ICRs located in intergenic regions (Schulz et al., 2010). The reacquisition of DNA methylation in the germline, which re-establishes sex-specific imprinting marks, is regulated by DNA methyltransferase 3A (DNMT3A) (Kaneda et al., 2004; Messerschmidt, 2012; Plasschaert

& Bartolomei, 2014). DNMT3L is also associated with the repression of transposons in male germ cells creating a potential link between genomic imprinting and silencing of repetitive elements. Additionally, a loss of DNMT3L has been shown to be correlated with a loss of maternal and paternal imprints (Hata et al., 2002; Messerschmidt, 2012).

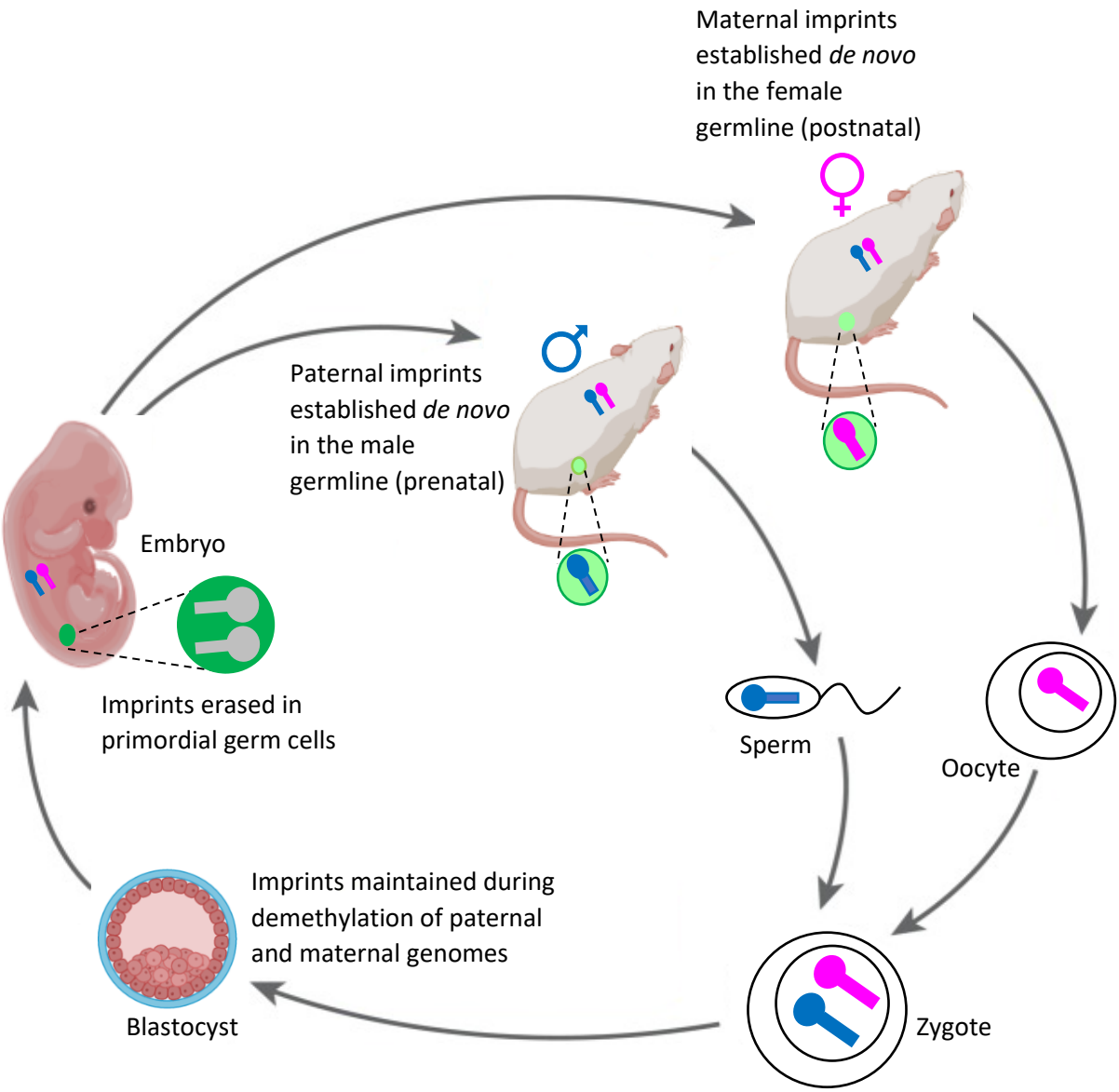


Figure 1.3: Establishment and maintenance of imprinting marks during mouse development. Genomic imprints are acquired in a sex-specific manner in developing primordial germ cells such as spermatogonia or developing oocytes (light green circles). Paternal chromosomes (shown in blue) establish imprinting marks prenatally while maternal chromosomes (shown in pink) establish these marks postnatally. These genomic imprints resist reprogramming and global changes in DNA methylation, such as the active and passive demethylation of the paternal and maternal chromosomes respectively, that occur post-fertilisation. The established imprinting marks are maintained in somatic cells. In newly formed primordial germ cells (dark green circles) however, these imprinting marks are erased (grey chromosomes) and then re-established according to the sex of the organism which will subsequently be transmitted to the next generation. Adapted from (Plasschaert & Bartolomei, 2014)

Another factor that also contributes towards the maintenance of DNA methylation at ICR in the early mouse embryo is the Tripartite motif-containing 28 (Trim28) protein. Many ICRs possess several TGCCGC motifs that, when methylated, are recognised by tissue-specific Krüppel-associated box-containing zinc-finger proteins (KRAB-ZFP), such as ZFP57, which in turn recruits Trim28 (Quenneville et al., 2011). The Trim28 (also known as KAP1 or TIF1b) protein acts as the central scaffolding component in a chromatin-modifier complex that promotes the formation of heterochromatin (X. Li et al., 2008; Messerschmidt et al., 2012; Quenneville et al., 2011). This chromatin-modifier complex, that Trim28 facilitates the binding of, consists of many factors (Fig 1.4) including the H3K9me3-catalysing histone methyltransferase SETDB1, the nucleosome remodelling and histone deacetylation (NuRD) complex, heterochromatin protein 1 (HP1), and in embryonic stem cells (ESC) the *de novo* methyltransferases DNMT3A and DNMT3B (Iyengar & Farnham, 2011; Quenneville et al., 2011; Schultz et al., 2001, 2002; Zuo et al., 2012). This heterochromatin-inducing complex remains bound to ICRs throughout preimplantation development, preserving DNA methylation at these regions, while the rest of the genome undergoes global DNA demethylation. Evidence of this complex's involvement in the maintenance of ICR was achieved through genome-wide DNA binding analysis in ESCs where 91 loci, including multiple known ICRs, are bound by the complex in a ZFP57-dependent manner (Quenneville et al., 2011). Furthermore, deletion of *ZFP57* has been shown to promote demethylation of ICRs post fertilisation leading to misregulation of imprinted expression and embryonic lethality, while deletion of *Trim28* in ESC resulted in a loss of pluripotency and arrest of growth (X. Li et al., 2008; Quenneville et al., 2011).

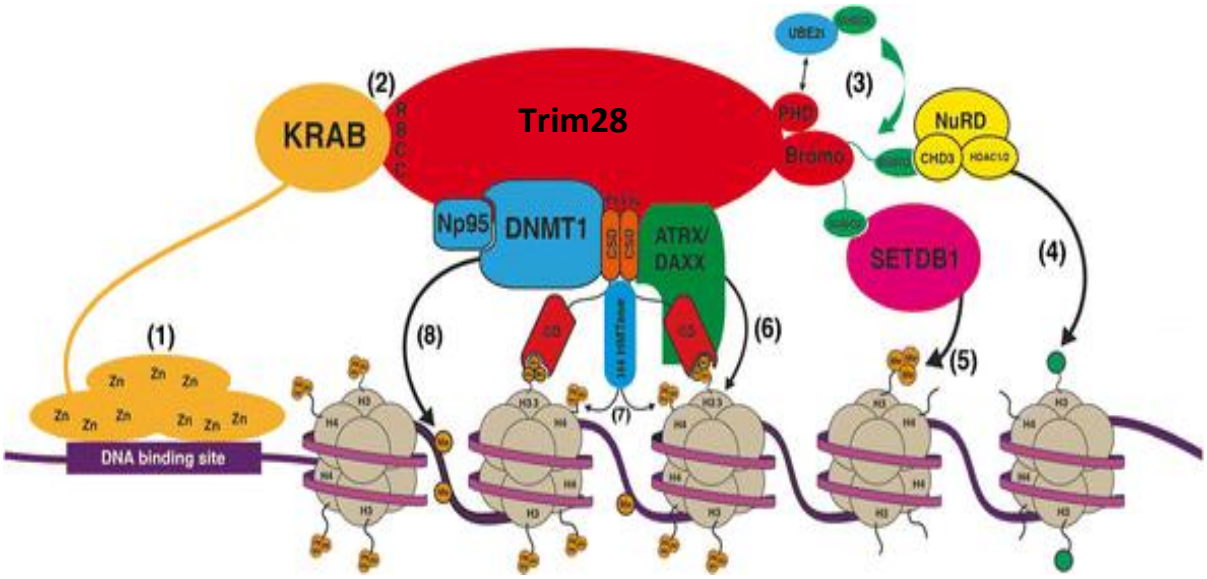


Figure 1.4: Schematic of the ZFP57/Trim28 facilitated chromatin modifier complex. 1) The KRAB-ZFP binds to specific TGCCGC motifs binding site through its Zinc finger projections. 2) The KRAB domain of the KRAB-ZFP interacts with the RBCC domain (RING domain, 2 B-box–type zinc fingers, and a coiled-coil domain). 3) The plant homeodomain (PHD) is an E3 ligase that, in co-operation with UBE2i, sumoylates the Trim28 bromodomain. 4) The now sumoylated bromodomain is bound by the nuRD complex that deacetylates nearby acetylated histones (green circle). 5) SETDB1 interacts with the sumoylated bromodomain and methylate's nearby histones to produce H3K9me3 (orange circles). 6) The ATRX/DAXX complex binds to Trim28, HP1 and H3K9me3 to incorporate the replacement histone H3.3 into the chromatin to ensure the maintenance of repressive histone modifications. 7) HP1 recruits a histone methyltransferase (HMT) that modifies nearby histones to generate H4K20me3 (orange circles) 8) DNMT1 which assists in the maintenance of DNA methylation binds to Trim28 along with Np95, a DNMT1 co-factor. This complex protects certain genomic regions such as ICR from global demethylation after fertilisation. Adapted from (P. B. Singh, 2016).

1.2.4 Mechanism of imprinted gene regulation

The establishment and maintenance of allele-specific imprinting marks through DNA methylation and histone modification is required to ensure that certain genes are expressed in a parent-of-origin-specific manner. The advancement of technology to allow for genome-wide profiling has benefitted our understanding of the mechanisms that promote this asymmetric expression immensely. As a result of these studies, two well-defined mechanisms for the regulation of imprinted gene expression have been proposed: the

insulator model and the non-coding RNA model (Fig 1.5) (Bartolomei & Ferguson-Smith, 2011; J. T. Lee & Bartolomei, 2013)**.**

The insulator model of imprinting regulation relies on the ability of the insulator protein CCCTC (CTCF) to bind to unmethylated ICRs. CTCF is a highly conserved ZFP, ubiquitously expressed in eukaryotes, and first identified as a transcriptional repressor of the chicken *c-myc* gene (Filippova et al., 1996; Klenova et al., 1993; Phillips & Corces, 2009). Previous studies have shown CTCF to bind to 55,000-65,000 sites within the mammalian genome, targeting intergenic, intragenic and promoter-proximal sites (H. Chen et al., 2012; X. Chen et al., 2008). One of the identified primary functions of CTCF is its role as an insulator. Insulator binding sites are short nucleotide sequences that act as targets for CTCF and set boundaries between genomic domains to prevent enhancer-promoter interactions and to act as a barrier against the spread of heterochromatin formation (Brasset & Vaury, 2005; Burgess-Beusse et al., 2002; Gaszner & Felsenfeld, 2006). CTCF binds to the DNA, as a result of its Zinc finger activity, where it forms a complex with cohesin, a ring-shaped multi-protein complex comprised of four subunits, Smc1, Smc3, Rad21 and SA1/2. The CTCF-cohesin complex brings together two distant regions of the genome into close proximity e.g., a promoter and enhancer, forming a stable chromatin loop (Hansen et al., 2017). Upregulation of gene transcription via enhancer-mediated activation is a prominent transcriptional regulatory mechanism in eukaryotes. These enhancers are short (100 bp-several kilobases) sequences that act as cis-regulatory elements and generally function independent of orientation to promote the transcription of genes up to 2-3 Mbp away from the enhancer region. Additional studies have also identified hundreds of thousands of enhancers within the human genome (Pennacchio et al., 2013). They are frequently marked by open chromatin-associated histone modifications such as H3K4me and H3K27ac and are able to facilitate

gene upregulation through modification of chromatin spatial organisation, resulting in chromatin looping, allowing the enhancer to be in close physical proximity with their target gene (Arnold et al., 2020; Court et al., 2014; Kulaeva et al., 2012; Pennacchio et al., 2013).

A good example of the insulator model is the regulation of gene expression at the *Igf2/H19* locus. These two genes lie 70 kb apart on chromosome 7 in the mouse genome and are reciprocally imprinted. A DMR, located 2-4 kb upstream of the transcriptional start site of *H19*, is found at this locus, where the paternal allele is hypermethylated and the maternal allele is unmethylated. The presence of hypermethylation at the paternal ICR inhibits the binding capability of the insulator complex CTCF-cohesin meaning the downstream enhancer, shared by both genes, is capable of interacting with the promoter of *Igf2*, driving transcription, while the paternal *H19* remains transcriptionally repressed. On the maternal allele, due to a lack of methyl presence at the DMR, CTCF is capable of binding to insulator sequences within the ICR, forming the CTCF-cohesin complex, and preventing the downstream enhancer from interacting with *Igf2*. However, this insulation of *Igf2* promotes enhancer interaction with *H19* resulting in upregulation of expression while *Igf2* remains inactive on the maternal allele. In this model, both *H19* and *Igf2* are competing for a shared downstream enhancer whose affinity for interaction and upregulation of expression is dependent in an allele-methylation sensitive manner resulting in reciprocal expression of the two genes (Bartolomei & Ferguson-Smith, 2011; Nordin et al., 2014; Pidsley et al., 2012; Plasschaert & Bartolomei, 2014; Sasaki et al., 2000). While this insulator model has been a long-standing theory of how imprinted expression is regulated at the *IGF2/H19* locus, a recent study conducted by Battaglia et al (2022) identified that the locus is instead potentially regulated by both a canonical and a primate-specific non-canonical enhancer, with the former promoting paternal *IGF2* expression as seen in Figure 1.5 and the latter

driving biallelic *IGF2* expression in certain proliferative tissues and cells. This non-canonical enhancer identified by Battaglia was marked by enhancer-like features in human skeletal muscle myoblasts (HSMMs) from nucleosome occupancy and methylome sequencing (NOMe–seq) data and possessed high levels of H3K27 acetylation. Additionally, in contrast to the canonical enhancer, the non-canonical enhancer region is closer to the *IGF2* gene and is accessible and unmethylated on both the maternal and paternal alleles which promotes biallelic expression in HSMMs, while ESCs, which do not possess evidence of the non-canonical enhancer, exhibited the expected paternal-specific expression of IGF2. Furthermore, while the non-canonical enhancer does not disrupt *H19* imprinting, its effect on *H19* maternal-specific expression is still unknown. The evidence of this alternate enhancer model in HSMMs, promoting biallelic expression, provides further evidence for species and tissue-specific regulation of imprinted clusters.

The other well-defined model of imprinting regulation, the ncRNA model, does not rely on enhancer-mediated activation and instead relies on the expression of a long ncRNA to repress the expression of nearby imprinted genes. As mentioned earlier in this chapter, long ncRNA molecules have the potential to promote genomic imprinting and transcriptional repression in an allele-specific manner. Examples of imprinted gene clusters that are possibly regulated using the ncRNA model include the *Igf2r* and *Kcnq1* loci (Plasschaert & Bartolomei, 2014). In this model, for *Igf2r* and *Kcnq1*, it is the paternal ICR that is unmethylated while the maternal ICR is hypermethylated. In these loci, the ICR coincides with a promoter for a long ncRNA molecule, *Airn* for the *Igf2r* loci and *Kcnq1ot1* in the *Kcnq1* loci. Expression of these long ncRNA molecules occurs in an antisense orientation and is dependent on methylation at the ICR (Fig 1.5), with the paternal unmethylated allele allowing expression of the long ncRNA while the maternal methylated allele silences its

expression. As a result of this expression of the long ncRNA from the unmethylated paternal allele, neighbouring genes such as *Igf2r, Slc22a2* and *Slc22a3* at the *Igf2r* locus are silenced (Fig 1.5). This is thought to occur due to the long ncRNA either attracting proteins that deposit repressive chromatin marks on the neighbouring genes (Nagano & Fraser, 2009) or by inhibiting RNA polymerase II recruitment at the promoter regions of the neighbouring imprinted genes (Latos et al., 2012), although these mechanisms require further study. The methylated maternal ICR however silences expression of these long ncRNA genes promoting activation of the nearby *Igf2r*, *Slc22a2* and *Slc22a3* genes resulting in them being expressed in a maternal-specific manner.

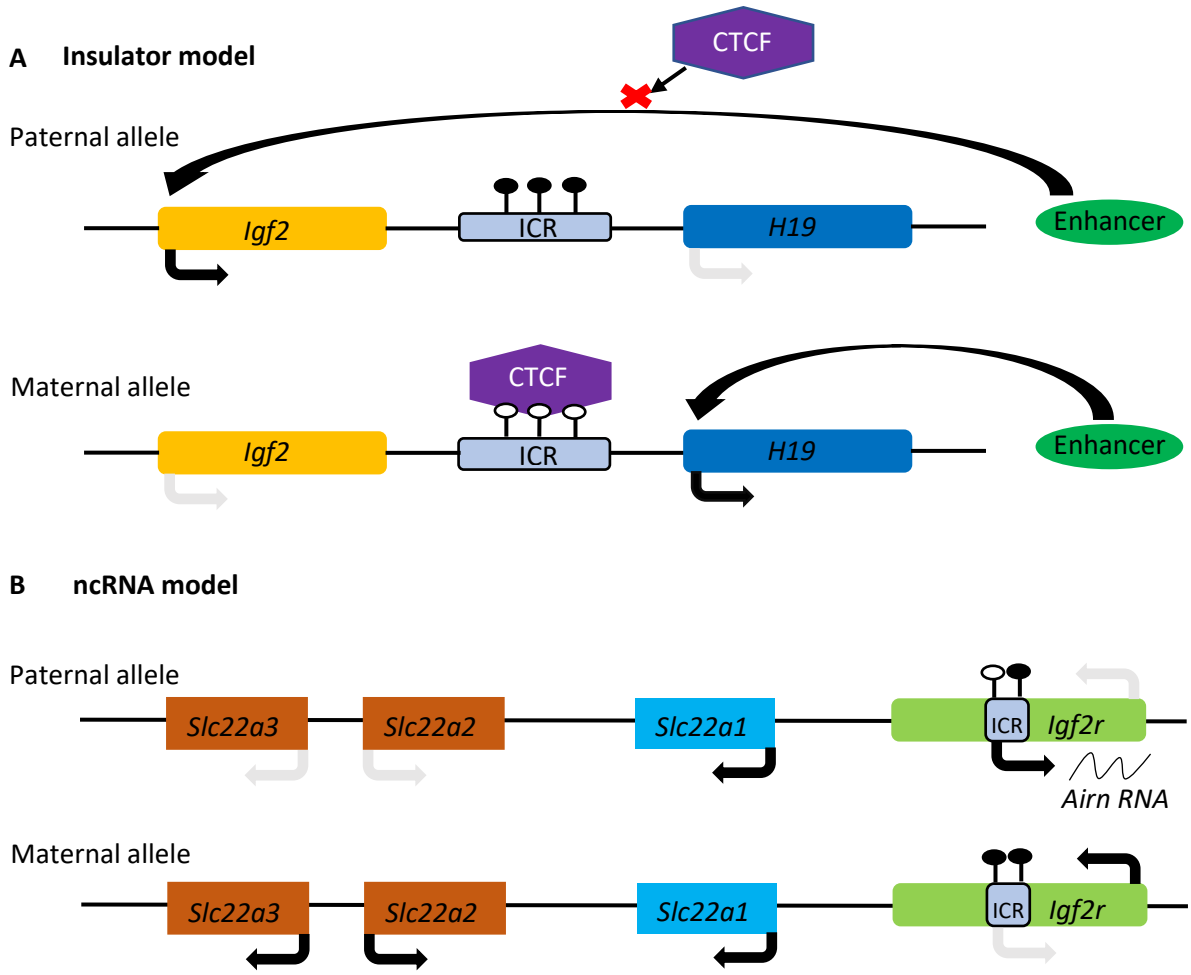


Figure 1.5: Two hypothesised mechanisms of imprinting regulation. A) The insulator model represented by the *H19/Igf2* locus (mouse Chr 7:142129262-142220553, ~91,000bp): The ICR on the paternal allele is methylated (black circles) preventing CTCF binding. The inability of CTCF to bind to the paternal ICR allows an enhancer downstream of the *H19/Igf2* locus to loop and interact with the promoter of *Igf2* activating its expression while *H19* expression remains inactive. On the maternal allele, the ICR is unmethylated (white circles) and acts as a target for CTCF binding. The CTCF-cohesin prevents interaction of the enhancer with *Igf2* and the enhancer instead interacts with the promoter of *H19* driving its expression while *Igf2* remains inactive on that allele. B) The ncRNA model represented by the *Igf2r/Slc22a* locus (mouse Chr1712637847-12988551, ~350,000bp): On the paternal allele, the ICR, located in the second intron of the *Igf2r* gene is unmethylated resulting in expression of the ncRNA *Airn* whose transcription occurs in an antisense orientation. Expression of this *Airn* corresponds with silencing of *Igf2r*, Slc22a2 and Slc22a3, possibly by blocking access of transcriptional machinery for those maternally expressed genes. The maternal allele, however, has a methylated ICR which silences the expression of *Airn* resulting in the expression of *Igf2r, Slc22a2 and Slc22a3*. The Slc22a1 gene is not imprinted and exhibits biallelic expression from both alleles. Dark arrows represent transcriptionally active genes while light grey arrows indicate repression of gene transcription. Adapted from (Marcho et al., 2015; Plasschaert & Bartolomei, 2014)

1.2.5 Imprinted genes and human disorders

Imprinted gene expression has been identified as essential for normal development, with many imprinted genes being highly expressed in prenatal development of both the embryo and extraembryonic tissues, and then subsequently downregulated post-birth, as well as in postnatal development in brain tissue (Cleaton et al., 2014; Coan et al., 2005; Wilkinson et al., 2007). Identification of multiple imprinted genes over time has provided insight into their functions, with imprinted genes expressed during development being associated with prenatal growth and development of particular lineages and postnatal homeostasis, while imprinted genes expressed in the brain are correlated with neurodevelopment, modulating metabolic axes, behaviour, learning and maternal care (Table 1.3) (Bartolomei & Ferguson-Smith, 2011). Furthermore, imprinted gene expression can exhibit a degree of tissue and cell specificity, such as *Ube3a*, which encodes the ubiquitin ligase protein UBE3A, displaying maternal-specific expression in discrete population of cells within the olfactory bulb (mitral cells), hippocampus (neurons) and cerebellum while exhibiting biallelic expression in other brain regions and peripheral tissues (Albrecht et al., 1997; Wilkinson et al., 2007). Additionally, imprinted gene expression can be dynamic and short-lasting such as in the case of the Delta Like Non-Canonical Notch Ligand 1 (*Dlk1*) gene which displays paternal-specific expression in the early embryo but loses its imprinting status and transitions to biallelic expression upon differentiation into NSCs, which is required for normal neurogenesis (Ferrón et al., 2011).

Table 1.3: A list of imprinted genes in humans that have been shown to impact numerous factors including neural and embryonic/ placental development.

Imprinted gene	Location (in human)	Function	Tissue impact
IGF2 (Insulin-like Growth Factor 2)	Chromosome 11: 2,129,112-2,158,391 (29,279 bp)	Promotes cell growth and differentiation	Affects foetal growth and development. Mutation shows impaired nutrient transport to growing foetus
H19	Chromosome 11: 1,995,129-2,001,710 (6582 bp)	Regulates cell cycle and growth	Impacts embryonic development, especially in musculoskeletal tissues and placenta
SNRPN (Small Nuclear Ribonucleoprotein N)	Chromosome 15: 24,823,637-24,978,723 (155,086 bp)	Involved in RNA splicing and processing	Affects neurodevelopment, predominantly in the brain
UBE3A (Ubiquitin Protein Ligase E3A)	Chromosome 15: 25,333,728-25,439,056 (105,329 bp)	Ubiquitin ligase involved in protein degradation and synaptic function	Impacts neuronal development and function, particularly in the brain
PEG3 (Paternally Expressed 3)	Chromosome 19: 56,810,077-56,840,728 (30,651 bp)	Involved in growth regulation	Impacts placental and foetal development. Mutations result in reduced maternal care
DLK1 (Delta Like Non-Canonical Notch Ligand 1)	Chromosome 14: 100,725,705-100,738,224 (12,519 bp)	Regulates cell differentiation and development	Affects growth and development of various tissues. Mutants have growth retardation, skeletal abnormalities, adiposity defects and abnormalities of hematopoietic lineages
GRB10 (Growth Factor Receptor-Bound Protein 10)	Chromosome 7: 50,590,063-50,793,462 (203,400 bp)	Involved in cell signalling and growth regulation	Impacts foetal and postnatal growth, as well as muscle and adipose tissue development
G_s alpha subunit (GNAS)	Chromosome 20: 58,839,718-58,911,192 (71,474 bp)	G protein regulating receptor-mediated cAMP production	Hypothalamus: controls melanocortin-mediated energy expenditure. Deletion results in glucose intolerance, insulin resistance, and obesity

Given the relationship between imprinted gene function and development, misregulation of these imprinted genes through chromosome aberrations such as deletions of imprinted regions, uniparental disomy, or mutations in proteins that regulate epigenetic regulation e.g. *Trim28*, are frequently associated with developmental disorders affecting embryonic growth and neurodevelopment (Butler, 2020; Eggermann et al., 2015; Plasschaert & Bartolomei, 2014). The most well-known imprinted disorders include Prader-Willi and Angelman syndromes, both of which are derived from mutations in the same domain on human chromosome 15 but differ in parental origin, and Beckwith-Wiedemann syndrome (Buiting, 2010; Butler, 2020). Prader-Willi syndrome, first described by Prader et al in 1956, is a neurodevelopmental disorder resulting from the absence of expression of a cluster of paternally expressed genes located at 15q11-q13 (Eggermann et al., 2015). The ICR that

regulates the expression of genes within the 15q11-q13 cluster is located within the small nuclear ribonucleoprotein-associated protein N (*SNRPN*) upstream region on the human chromosome 15 cluster (Hogart et al., 2009). Prader-Willi syndrome is characterised by two phases of symptomology occurring in infancy and early childhood. In infancy, the newborn exhibits infantile muscular hypotonia, respiratory problems and feeding difficulties, with later symptoms in early childhood including excessive eating (usually resulting in morbid obesity), variable levels of cognitive impairment, behavioural problems and growth hormone deficiency (Angulo et al., 2015; Cassidy et al., 2012). There are three known mechanisms leading to the establishment of Prader-Willi syndrome: 1) *De novo* deletion of the 15q11-q13 cluster on the paternal chromosome as a result of two proximal chromosome 15q11-q13 breakpoints (BP1 and BP2) and a distal breakpoint (BP3) which equates to ~75% of Prader-Willi cases. 2) Uniparental disomy inheritance of the maternal chromosome 15, correlating with ~20-25% of cases. 3) Failure of imprinting establishment e.g. silencing of the paternal allele through epigenetic markers, contributing to ~1% of cases (Angulo et al., 2015; Buiting, 2010; Eggermann et al., 2015). The 15q11-q13 chromosomal region contains several protein-coding genes, proposed to be involved in alternative splicing regulation, such as the paternally expressed genes *MKRN3, MAGEL2, NDN, PWRN1, C15orf2* and *SNURF-SNRPN* as well as several clusters of small nucleolar RNA (snoRNA) genes (Anderlid et al., 2014; Eggermann et al., 2015). Additionally, the region also contains the antisense paternally expressed transcript of the *UBE3A* gene, *UBE3A-ATS* thought to regulate the imprinted expression of *UBE3A* at the 3' end of the imprinted cluster (Buiting, 2010; Schaaf et al., 2013; Sonzogni et al., 2020).

Angelman syndrome is another neurodevelopmental imprinted disorder named after Harry Angelman who first identified the syndrome in 1965 after observing children with similar

abnormal features. These features included severe developmental delay, movement or balance disorder, severe limitations in speech and language and typical abnormal behaviour including a happy demeanour and excessive laughter (Angelman, 1965; Buiting et al., 2016). Clinical diagnosis of the syndrome requires fulfilment of these four major criteria alongside an exhibition of three minor criteria including postnatal microcephaly, seizures, abnormal EEG, sleep disturbance, attraction to or fascination with water, and drooling (Tan et al., 2011). The absence of expression, or the mutation, of the maternal transcript *UBE3A*, which is expressed in the human foetal brain and adult frontal cortex (Rougeulle et al., 1997; Vu & Hoffman, 1997) is the root cause of Angelman syndrome. Additionally, this lack of *UBE3A* expression has a similar foundation to the causes of Prader-Willi syndrome with the exception of chromosome alterations occurring on the maternal allele at the 15q11-q13 region rather than that of the paternal allele seen in Prader-Willi syndrome. Furthermore, the percentiles of causation for Angelman syndrome are also similar, with 70-75% of cases caused by *de novo* deletion of the maternal 15q11-q13 region, 3-7% caused by uniparental disomy of the paternal allele and 2-3% of cases caused by imprinting defects (Butler, 2020; Eggermann et al., 2015). While Prader-Willi syndrome can be attributed to these three genetic defects alone, these mechanisms correlate to 85-90% of Angelman cases. The remaining 10-15% of cases are a result of mutation of the *UBE3A* gene on the maternal allele where the sense transcript of this gene is preferentially expressed (Buiting et al., 2016; Margolis et al., 2015).

Unlike the previous two imprinted disorders, certain imprinted disorders can be associated with mutations or chromosomal aberrations of multiple loci. One of the best-known examples of this is the growth disorder Beckwith-Wiedemann syndrome, previously known as EMG syndrome due to its three main features: exomphalos, macroglossia and (neonatal)

gigantism. First reported by Wiedemann in 1964 (Wiedemann, 1964) and Beckwith in 1969, Beckwith-Wiedemann syndrome is generally a result of sporadic mutation, although autosomal dominant transmission is seen in ~10% of cases (Butler, 2020; Eggermann et al., 2015). Major features of the disorder include macrosomia, with larger than average muscle mass at birth, macroglossia, hemihyperplasia resulting in asymmetrical growth in one or more regions of the body, and occasionally (5-7% of cases) embryonal tumours, most commonly Wilms tumour (A. C. Edmondson & Kalish, 2015; Eggermann et al., 2015; Wang et al., 2020). The molecular aetiology for the disorder involves genetic and epigenetic changes on the chromosome 11p15.5 region which comprises two imprinting control regions, *H19/Igf2* and *Kcnq1/Cdkn1c*. The majority of cases (~80%) show sporadic aberrant DNA methylation at the ICRs of these imprinting clusters such as hypomethylation of the maternal allele *KCNQ10T1* DMR resulting in expression of this ncRNA from the maternal allele and downregulation of *CDKN1C*, or gain of methylation at the maternal *H19/Igf2* ICR. However, the other 20% of cases are usually a result of alternate chromosome abnormalities including maternally derived translocations and inversions of 11p15, duplications of paternal 11p15, uniparental disomy of the paternal chromosome 11 and mutation of *CDKN1C* (Butler, 2020; Wang et al., 2020). Additionally, the majority of inherited cases of the syndrome where aberrant methylation is present is frequently correlated with Microdeletions/duplications or point mutations at the ICRs while cases where methylation patterns at ICRs were normal are usually associated with mutations of *CDKN1C* (Eggermann et al., 2015).

1.3 The *Peg13/Kcnk9* cluster

1.3.1 Imprinting regulation of the *Peg13/Kcnk9* cluster

The *Peg13/Kcnk9* imprinted cluster, located on chromosome 8q24 in the human genome and chromosome 15qD3 in the mouse genome, is comprised of 5 genes: Argonaute RISC Catalytic Component 2 *(Ago2*), chromatin accessibility complex subunit 1 *(Chrac1),* Trafficking protein particle complex subunit 9 *(Trappc9),* Paternally expressed 13 *(Peg13)* and Potassium channel subfamily K member 9 *(Kcnk9)* (Fig 1.6) (Cooper et al., 2020; Court et al., 2014). The genes within the complex have been shown to play a role in neurodevelopmental functions, with mutations causing disorders in both humans and mice (Cooper et al., 2020; Delgado et al., 2014; Lessel et al., 2020; Marangi et al., 2012; Mochida et al., 2009). The ICR responsible for the regulation of imprinted expression at the *Peg13/Kcnk9* cluster has been hypothesised to be the DMR located at the *Peg13* promoter proximal CGI, due to its status as the only DMR CGI within the cluster, displaying methylation on the maternal allele, hypomethylation of the paternal allele and CTCF binding capability (Court et al., 2014; P. Singh et al., 2011; Smith et al., 2003). The genes within the locus have all been reported to be imprinted in mice, with a degree of brain-specific imprinted expression observed for *Ago2, Chrac1* and *Trappc9* with these genes displaying biallelic expression in mouse peripheral tissues. *Peg13* has been reported to exhibit imprinted expression in multiple tissues, showing little to no tissue-specificity unlike other genes within the cluster (Claxton et al., 2022; Court et al., 2014). Additionally, the imprinted expression of *Ago2, Chrac1* and *Trappc9* does not appear to be conserved, with only *PEG13* and *KCNK9* displaying a clear imprinting status in humans (Court et al., 2014).

The noncoding RNA *Peg13,* located within intron 17 of the *Trappc9* gene, has previously been shown to be expressed from the paternal allele, starting at an unmethylated CGI promoter. This paternal biased expression has been reported in multiple mouse tissue types and shows conservation in humans (Claxton et al., 2022; Court et al., 2014; Ruf et al., 2007; Smith et al., 2003). Generation of this imprinting bias is achieved through germline-derived methylation of the *Peg13* promoter on the maternal allele while the paternal allele remains unmethylated and transcriptionally active. An interesting feature of this cluster is that while *Peg13* displays multi-tissue paternal biased expression, the other imprinted genes within the cluster exhibit a preference for maternal allele expression in murine brain tissue only (Fig 1.6), with *KCNK9* exhibiting this maternal biased expression in human brain also. Additionally, while each of the four maternally expressed genes: *Ago2, Chrac1, Trappc9* and *Kcnk9* display maternal-specific expression, they do so at varying degrees with certain genes such as *Kcnk9* displaying a stronger degree of imprinting compared to that of *Trappc9* and *Ago2*. This variation in imprinting expression levels has been shown to be tissue and even cell-specific with *Chrac1* and *Ago2* displaying a similar pattern of maternal bias throughout the brain, being stronger in the cortex and weaker in the olfactory bulb, hippocampus, and cerebellum (Claxton et al., 2022; Perez et al., 2015).

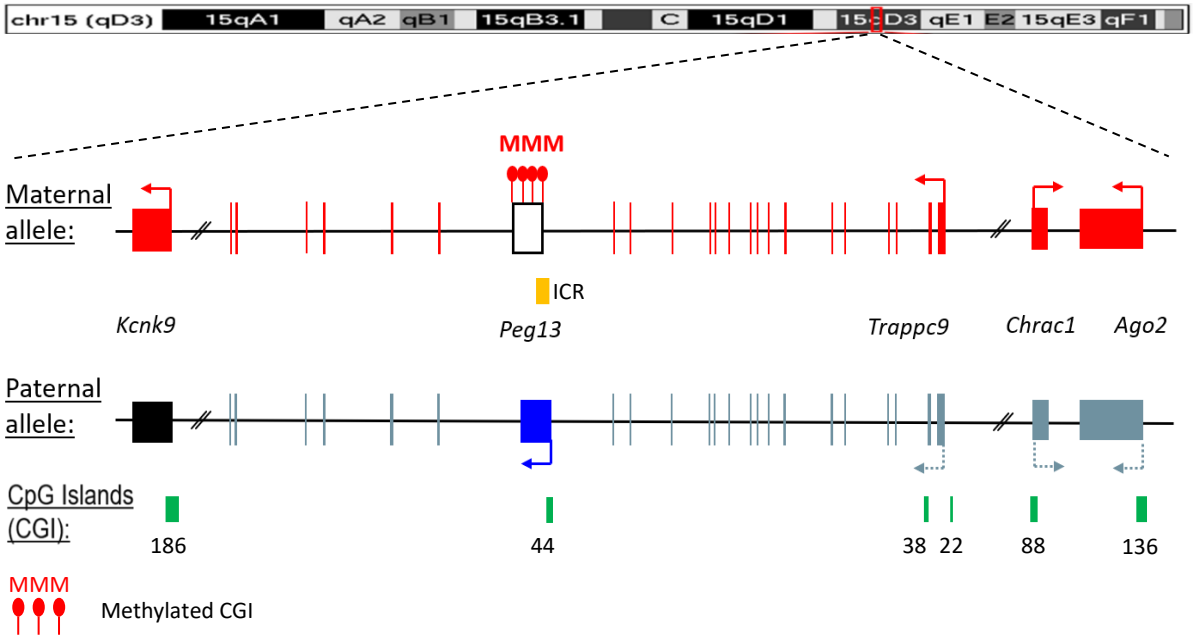


Figure 1.6: Schematic of the mouse *Peg13/Kcnk9* imprinted gene cluster, chr15:72368013-73063835 (~700,000bp). The *Peg13/Kcnk9* imprinted cluster located on chromosome 15 in the mouse genome is comprised of 5 imprinted genes. Three of the genes within the cluster: *Ago2, Chrac1* and *Trappc9* display tissue-specific preferential maternal expression (red arrows) in brain tissue and biallelic expression in peripheral tissues in mice. *Kcnk9* also displays maternal-specific expression in multiple mouse tissues and shows conservation of genomic imprinting in humans which *Ago2, Chrac1* and *Trappc9* do not. *Peg13* displays paternal-specific expression (blue arrow) in multiple tissues in both mice and humans. CGIs located proximal to the promoter region of each gene are displayed (green boxes) with the *Peg13* CGI, located between exons 17 and 18 of the *Trappc9* gene, exhibiting a differential methylation pattern: the maternal allele is methylated (red circles) resulting in silencing of *Peg13* and upregulation of the neighbouring genes while the paternal allele, which shows hypomethylation at the DMR promotes transcription of *Peg13* ncRNA and represses transcription of neighbouring genes. This DMR is thought to act as an ICR to mediate allele-specific expression of all genes within the cluster depending on tissue type. Arrows indicate the direction of transcription. Grey arrows on the paternal allele indicate repressed transcription and silencing of genes.

In addition to this, there have also been conflicting reports on the imprinting status between transcriptional variants of the *Trappc9* gene. A number of *Trappc9* transcript variants are annotated on the ENSEMBL database (https://www.ensembl.org/Mus_musculus/Info/Index) including two alternative promoters and truncated transcripts. While the *Trappc9* gene has been empirically proven to express

from the maternal allele, previous data suggest that the truncated variant (203) of *Trappc9*, which ends shortly after Peg13 in intron 17 exhibits paternal-specific expression (Gregg et al., 2010; Hsu et al., 2018). However, alternate attempts to identify variant transcript expression of *Trappc9* using RT-PCR were unable to find evidence of an alternative promoter or truncated transcript variants (Claxton et al., 2022).

1.3.2 The Peg13/Kcnk9 cluster gene functions

The functions of the five genes located within the *Peg13/Kcnk9* imprinted cluster have previously been linked to neurodevelopment, with mutations causing developmental defects in both humans and mice. *Ago2* is a highly conserved gene that is widely expressed in the mouse oocyte and during early mouse development. The gene encodes an endonuclease that functions as part of the RNA-induced silencing complex (RISC) in a process called RNA interference (RNAi) (Alisch et al., 2007; Hutvagner & Simard, 2008; J. M. Zhang et al., 2020). As a protein in humans, Ago2 is comprised of four domains: The N domain acts as a wedge to split duplexes during the RISC assembly, the middle (MID) and Piwi-Argonaute-Zwille (PAZ) domains recognise the 5' and 3' ends of the small RNA guide strand, respectively and the P-element-induced wimpy testis (PIWI) domain contains the catalytic site responsible for target cleavage. Additionally, there are two linkers, L1 and L2 between the N and PAZ domains and between the PAZ and MID domains, respectively, which contribute towards the structural stability of the RISC (Nakanishi, 2022). The *Ago2* protein displays binding capabilities with microRNAs, siRNAs and *Piwi*-interacting RNAs and facilitates their loading onto RISC, where the single strand acts as a template to recognise mRNA transcripts, allowing *Ago2* to cleave the mRNA resulting in the inhibition of

translation and degradation of transcripts (Hutvagner & Simard, 2008; Meister et al., 2004). Furthermore, *Ago2* has been proven to be essential for oogenesis and early mouse embryonic development as deletion of the gene has been correlated with the arrest of development and lethality (Kaneda et al., 2009; Lessel et al., 2020; Lykke-Andersen et al., 2008; Morita et al., 2007). The *Chrac1* gene encodes a histone-fold protein that interacts with additional histone-fold proteins to create a nucleosome-remodelling complex that, in conjunction with nucleosome remodelling factors (NURF), functions to open chromatin for transcription and replication, utilising their common ATPase ISWI catalytic subunits to increase the mobility of nucleosomes relative to DNA sequence (Guschin & Wolffe, 1999). *Kcnk9* encodes a protein called K_{2P}9.1, also known as TWIK-related acid-sensitive K1 channel 3 (TASK3), which acts as a member of the two pore-domain potassium channel (K_{2P}) subfamily. The gene has been suggested to play a role in K+-dependent apoptosis of cerebellar granule neurons in culture (Lauritzen et al., 2003), in maturation of neurons in the cerebellum during neuronal development (Zanzouri et al., 2006) and in the maintenance of background current in multiple neuronal populations (Talley et al., 2001). K_{2P} channels have also been associated with the stabilisation of resting membrane potentials of both excitable and non-excitable cells thus regulating cell activity (Cousin et al., 2022). *Peg13* is a gene that encodes a lncRNA whose function remains to be elucidated although studies have suggested it may play a role in the imprinting regulation of neighbouring genes (Court et al., 2014). Expression of *Peg13 has* been shown in multiple tissues and organs but displays the greatest level of expression in the brain in regions such as the septal and hypothalamic regions, the hippocampus and the cerebral cortex (Davies et al., 2004).

The final gene of the cluster, *Trappc9* (also known as *Trs120* in yeast) is one of the most well-studied of the five imprinted genes. It encodes a protein subunit of the transport

protein particle complex II (TRAPP II), a highly conserved trafficking molecule found in many organisms including yeast and humans (Kim et al., 2016). Expression of *Trappc9* is found predominantly in postmitotic neurons of the cerebral cortex, hippocampus and deep grey matter (Mochida et al., 2009). The majority of our understanding of the TRAPP complexes comes from studies conducted in yeast, where the complexes show a high degree of conservation. To date three TRAPP complexes (I, II and III) have been identified in *Saccharomyces cerevisiae* while two complexes have been identified in humans (II and III) (Fig 1.7) (Brunet & Sacher, 2014; Kim et al., 2016). All TRAPP complexes share a core structure of 7 polypeptides made of six subunits consisting of Trs20, 23, 31, 33, two copies of Bet3 and one Bet5 subunit (Trappc2, Trappc4, Trappc5 and Trappc6, Trappc3 and Trappc1 in mammals respectively) (Table 1.4). To this core structure that makes up the TRAPP I complex additional subunits are added, trs65 and trs120, trs130 and Tca17 (Trappc13, Trappc9, Trappc10 and Trappc2L in mammals respectively), to make the TRAPP II complex while the TRAPP III complex contains trs85 (Trappc8 in mammals) and two other subunits specific to the human TRAPP III complex, Trappc11 and Trappc12 (Fig 1.7) (Brunet & Sacher, 2014; Mbimba et al., 2018).

Table 1.4: Homology of TRAPP subunits between yeast and mammals

Yeast proteins	Mammalian proteins
Bet5	Trappc1
Trs20	Trappc2
Tca17	Trappc2L
Bet3	Trappc3
Trs23	Trappc4
Trs31	Trappc5
Trs33	Trappc6
Trs85	Trappc8
Trs120	Trappc9
Trs130	Trappc10
N/A	Trappc11
N/A	Trappc12
Trs65	Trappc13

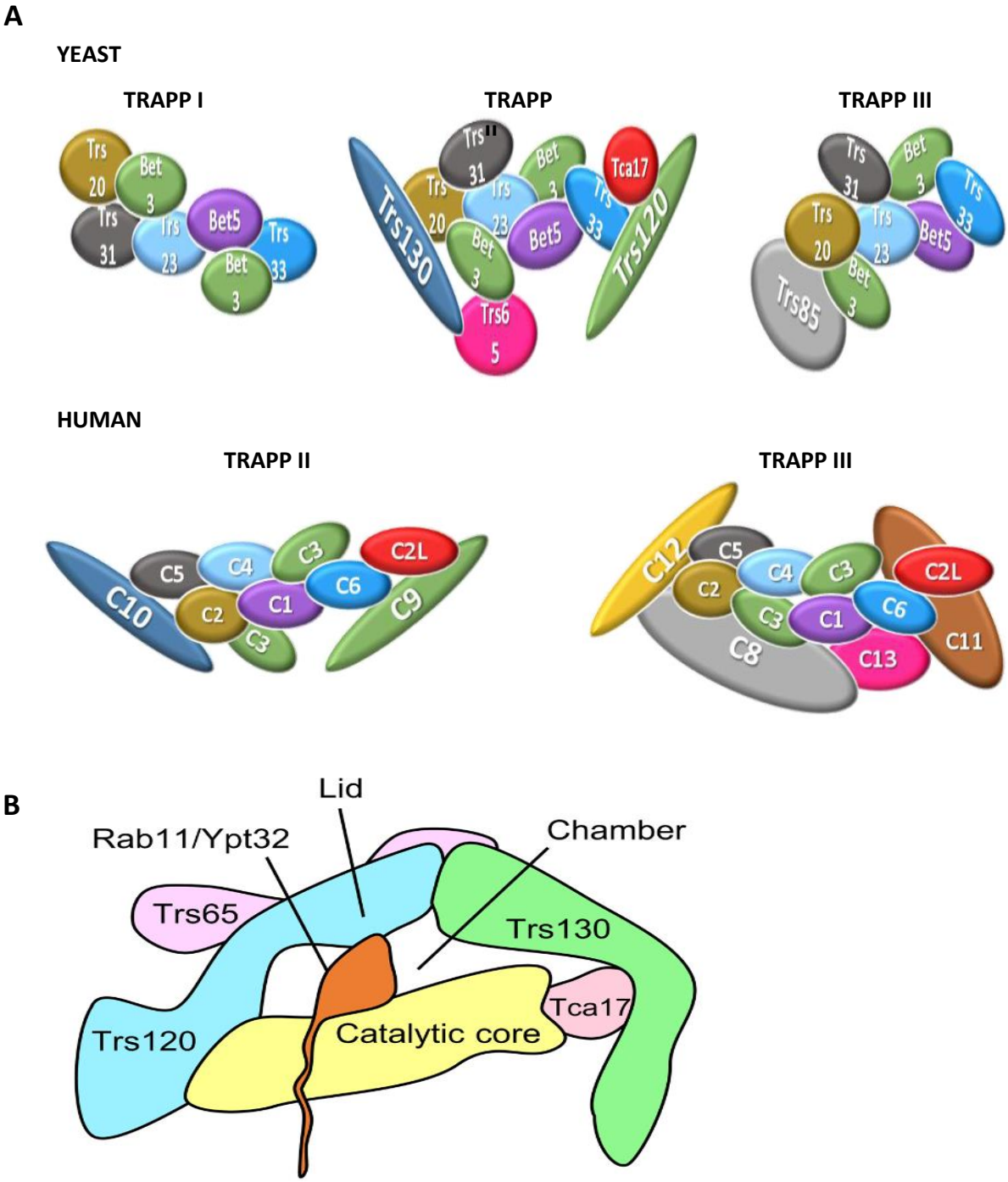


Figure 1.7: A) Structure of the TRAPP complexes in yeast and mammals. B) Cartoon schematic of the interaction of Rab11/Ypt32 with TRAPPII. Trs120/Trappc9 forms a lid to enclose Rab11 within the active site chamber Adapted from (Bagde & Christopher Fromme, 2022; and Brunet & Sacher, 2014).

The TRAPP complexes have been reported to be involved in the vesicular transport of proteins and lipids between different membrane-bound compartments, with Trapp II, and by association Trappc9, being particularly important for Golgi vesicle tethering and intra-Golgi transport while TRAPP III is associated with autophagy (Brunet & Sacher, 2014; Kim et al., 2016; Mbimba et al., 2018; Yip et al., 2010). Vesicle tethering refers to the initial long-range interaction between a vesicle and its acceptor compartment which is mediated by multi-subunit complexes such as TRAPP and long coiled-coil proteins (Barrowman et al., 2010; Yip et al., 2010). This process is then followed by membrane-vesicle fusion mediated by tethering factors that promote the organisation of SNAP REceptor (SNARE) proteins that tightly interact to form a hairpin structure and initiate membrane mixing and fusion (Cai et al., 2007; Jahn & Scheller, 2006). The process is likely aided by Rab GTPases, which act as master regulators of membrane trafficking in eukaryotic cells, with all three TRAPP complexes possessing guanine exchange factor (GEF) activity for multiple Rab proteins (Jenkins et al., 2020). The TRAPP II complex has previously been shown to act as a GEF for Rab1, Rab11 and Rab18 through binding of the Trappc9 subunit (Ke et al., 2020; C. Li et al., 2017). Activation of these Rab proteins by the TRAPP complexes induces the recruitment of effector factors involved in the fusion of the vesicle to the target membrane (Hutagalung & Novick, 2011; C. Li et al., 2017). Additionally, Trappc9 is known to mediate the interaction between the TRAPP II complex and coat protein I (COPI) serving as a tethering complex for COPI-coated vesicles specific to endoplasmic reticulum (ER)-Golgi and intra-Golgi trafficking (Cai et al., 2005; C. Li et al., 2017; Mbimba et al., 2018; A. Yamasaki et al., 2009).

1.3.3 Mutations of the Peg13/Kcnk9 cluster

Several genes of the *Peg13/Kcnk9* imprinted cluster have been associated with clinical disorders upon acquisition of specific mutations. While no clinical disorder has been characterised for *Ago2*, mutations of the gene are associated with disturbances in neurological development. Previous studies have identified the human *AGO2* gene as one of the most missense-intolerant genes within the human genome, with Exome Aggregation Consortium (ExAC) and Genome Aggregation Database (gnomAD) sequencing data producing Z scores quantified at 7.696 and 6.058 for *AGO2* respectively, substantially higher than that of other genes associated with developmental disorders (Lessel et al., 2020). Mutations of *AGO2* have been linked with impaired shRNA-mediated silencing and impaired RISC formation which have been correlated with intellectual disability, developmental delay, including delayed motor development and impaired speech development, as well as embryo lethality during early mouse development at post-implantation stages, potentially due to placental development defects in mouse *Ago2* KO's (Lessel et al., 2020; Müller et al., 2020).

Unlike *Ago2, Chrac1* and *Peg13*, all of which have no associated clinical disorders, a missense mutation, 770G→A in exon 2 of the maternal copy of *Kcnk9* has been identified as the cause of Birk-Barel syndrome (also known as *Kcnk9* imprinting syndrome (KIS)) in humans (Barel et al., 2008). The syndrome is associated with moderate to severe mental retardation, feeding difficulties in infancy, generalised hypotonia from birth, and unique dysmorphism with an elongated face. Additionally, mice lacking TASK3 (*Kcnk9*-/-) have also been shown to have impaired memory (Linden et al., 2007), sleep perturbation (D. S. J. Pang et al., 2009), possible increased likelihood of depression (Gotter et al., 2011) and vision

defects (Wen et al., 2022). While associations of phenotype to KIS have been studied, the precise molecular mechanisms that underpin the disease are yet to be identified.

From the *Peg13/Kcnk9* imprinted cluster, both function and the impact of mutations have been most extensively studied in the *Trappc9* gene. Mutations in the sequence of *TRAPPC9* have been associated with recessive autosomal non-syndromic intellectual disability (Krämer et al., 2021), autism spectrum disorder (ASD) (Hnoonual et al., 2019) and severe developmental delay (Aslanger et al., 2022). These mutations are rare, affecting ~56 individuals across the world across multiple ethnic groups. The highest proportion however is seen across the middle east, suggested to be linked to a higher percentage of consanguineous marriages in this region (Amin et al., 2022). The developmental defects caused by *TRAPPC9* mutation is characterised by multiple features that fall under the categorisation of Intellectual disability-obesity-brain malformations-facial dysmorphism syndrome (ORPHA). The severity of symptoms varies between individuals; however, the majority seem to exhibit a degree of inability or delay of speech, late milestone development e.g. walking, inability to feed themselves. With regard to morphological alterations as a result of mutation of *TRAPPC9*, symptoms can include microcephaly, cortical and cerebellar atrophy, reduction in white matter volume, thinning of the corpus callosum and dilatation of ventricles (Amin et al., 2022; Krämer et al., 2021; Marangi et al., 2012; Mochida et al., 2009). Additional features sometimes seen in *TRAPPC9* mutation patients include obesity, hypotonia, facial dysmorphism (smooth philtrum, short forehead, synophrys (unibrow), broad nasal bridge), and behavioural phenotypes such as autism and/or a happy demeanour (Krämer et al., 2021; Marangi et al., 2012). To date, all *TRAPPC9* mutations within exon sequences that result in physiological disorders, have been reported

to be loss of function mutations: nonsense, frameshift splice site or insertions/deletions (Amin et al., 2022).

To summarise, the genes of the *Peg13/Kcnk9* cluster possess many different roles associated with normal development and display varying degrees of tissue-specific imprinted expression in mice, with two genes, *KCNK9* and *PEG13,* having conserved imprinted expression in humans. Several of the genes have been associated with clinical disorders and disease phenotypes such as Birk-Barel syndrome and non-syndromic intellectual disability, that have been extensively studied.

1.4 Aims of the thesis

While the regulation of imprinted expression within the *Peg13/Kcnk9* cluster has previously been associated with the *Peg13* DMR, the molecular mechanism that underlies this process is yet to be elucidated. In a study conducted by Court et al (2014), the regulation of the human orthologous region of this cluster, located on chromosome 8q24, was identified. Using SNPs for allele identification, the study determined that only *PEG13* and *KCNK9* exhibited imprinted expression, indicating that conservation of imprinting is not observed between mouse and human for *AGO2*, *TRAPP9*, and *CHRAC1*. Moreover, the study found no sequence similarity between the mouse and human *PEG13*, despite their similar location within the genome in relation to neighbouring genes and that the maternal *PEG13* CGI was the only methylated promoter CGI within the cluster.

Utilising an *in-silico* analysis of published ChIP-sequence data, Court et al (2014) identified several strong canonical two-part CTCF (motif 1 + 2) binding sites within the *PEG13*-DMR,

highlighting that regulation of the cluster may be similar to that of the insulator model seen in the *H19/Igf2* cluster. This was confirmed using ChIP-seq datasets and analysis to identify co-localisation of CTCF and cohesin at the unmethylated *PEG13* paternal CGI. Through the binding of CTCF-cohesin, the *Peg13*-DMR exhibits enhancer-blocking activity (confirmed by Court et al using an enhancer-blocking assay in HEK293 cells) and allele-specific chromatin looping between a shared enhancer and the *KCNK9* and *PEG13* promoters (Fig 1.8). The human *KCNK9* promoter, similar to the *PEG13* promoter, exhibits strong CTCF enrichment, with both displaying evidence of enhancer chromatin signatures such as co-enrichment of H3K4me1, H3K27ac and p300 in various brain ChIP-seq datasets such as the frontal cortex and other brain regions. Additionally, lymphocytes, which express neither *KCNK9* or *PEG13*, are not associated with these histone modifications, suggesting that the enhancer responsible for regulation of these genes is brain-specific. Although the physical interaction between an enhancer region and the promoters of *KCNK9* and *PEG13* in human cerebellar samples was confirmed in this study using chromosome conformation capture (3C) (Fig 1.8), Court et al (2014) was unable to identify any informative SNPs to determine which enhancer loop interaction was specific to either allele.

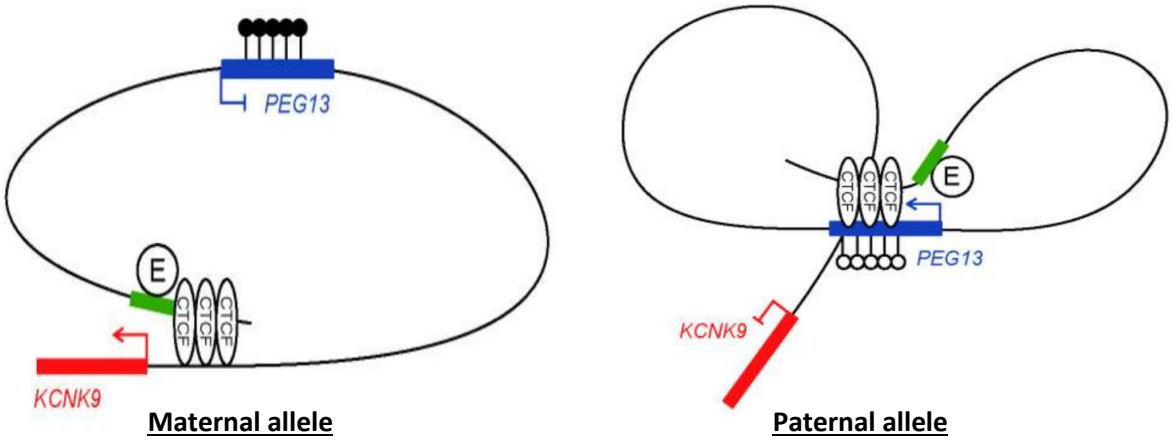


Figure 1.8: Proposed model of chromatin looping for enhancer-mediated upregulation of *PEG13* and *KCNK9* at the human 8q24 locus (Chr8 139,600,838-140,100,553, ~500,000bp). The *PEG13* DMR is methylated on the maternal allele, inhibiting the binding of CTCF insulator molecules which instead bind to an unmethylated CTCF binding site proximal to *KCNK9*. CTCF then interacts with cohesin to promote higher-order chromatin looping, bringing a distal enhancer (E) in close proximity to *KCNK9* promoting its expression while *PEG13* is transcriptionally silent. On the paternal allele, due to the absence of methylation at the *PEG13* DMR, CTCF is capable of binding and insulates *KCNK9* from enhancer activity. The enhancer instead loops into close proximity of *PEG13*, promoting its expression from the paternal allele while the paternal *KCNK9* gene is transcriptionally silent. Adapted from (Court et al., 2014)

While Court's study was able to identify distal enhancer interaction with the promoters of *PEG13* and *KCNK9* as a result of chromatin looping mediated by CTCF-cohesin interaction, there is no imprinted regulation for *AGO2, TRAPPC9* and *KCNK9* in human. However, in mice, these three genes exhibit maternal biased expression specifically in brain tissue, highlighting that the mechanism involved in regulating expression of these genes must rely on different/additional elements in mouse. The identification of a mouse brain-specific regulatory element responsible for the imprinted regulation of genes within this cluster was one of the primary motivations of this thesis research.

An additional motivation for the research undertaken in this project is that analysis of imprinted expression for this locus has previously been performed using bulk tissue samples, whereas previous studies have suggested that imprinting regulation can vary

widely on a cell-to-cell basis (Perez et al., 2016). A previous study by Bonthuis et al (2015) utilised RNAscope in situ hybridisation probes to determine whether noncanonical imprinted expression within a tissue is a result of 1) an allele expression bias in each cell, 2) skewed random mono-allelic expression effects or 3) allele silencing in a subpopulation of cells. The probes designed in this study targeted intronic regions to detect nascent RNA arising from each allele in the nucleus of cells in tissue cryosections of the arcuate nucleus (ARN) and dorsal raphe nucleus (DRN) in mouse brain. Their data concluded that noncanonical imprinted genes exhibit allele-specific expression effects in subpopulations of neurons in the brain. Given that most imprinted expression analyses of the *Peg13/Kcnk9* cluster has been conducted on bulk tissue, I aimed to determine whether the imprinting bias previously described in literature is reflective of all cells within a neuron population or whether individual subpopulations and cells exhibit alternate allele-specific expression patterns highlighting a greater degree of complexity for tissue-specific imprinting.

By addressing these questions, this thesis seeks to elevate our understanding of the imprinting regulation of the *Peg13/Kcnk9* cluster in development, focusing on imprinted expression between neural stem cell samples and differentiated brain tissue both in multi-cellular and single-cell samples. An additional aim of the thesis is to identify a brain-specific regulatory element that acts to upregulate the expression of the *Trappc9* gene, in an allele-specific manner, and identify whether the imprinted cluster is associated with enhancer-mediated regulation. In chapter 3 I discuss the expression of the genes within the Peg13/Kcnk9 cluster as well as the methylation profiles of their promoter-proximal CGI regions using bulk tissue samples for analysis. Hybrid cross new-born mice were used to generate identifiable SNPs to differentiate the maternal and paternal alleles when analysing samples to determine relative expression frequencies. Additionally, multiple tissue types

were used to isolate mRNA and genomic DNA for evaluation to confirm the tissue-specific imprinted expression and CGI methylation profiles at promoter regions that previous literature has described. In chapter 4 I discuss the variation between bulk sample imprinted expression and methylation profiles with that of single-cell samples. Utilising multiple techniques, I detail whether the relationship of allele-specific expression in a multi-cellular sample is reflected in all single cells of a similar sample, thus displaying a degree of uniformity, or whether individual cells possess variations in imprinting specificity, both in terms of allele bias and relative strength of imprinted expression across cells. Additionally, I evaluate whether the methylation of promoter CGI regions is uniform across single cells, exhibiting a similar pattern to that of the bulk sample methylation profiling conducted in chapter 3, or whether single-cell methylation status at promoter-proximal CGIs can vary across a cell population. Furthermore, I aimed to determine whether the methylation profiles exhibited in single cells alter upon differentiation, using NSC and primary neurons as my samples. Finally, in chapter 5 I investigate the presence of brain-specific regulatory elements that show enhancer like chromatin modifications and identify whether any identified elements mediate upregulation of the *Trappc9* gene using multiple unique promoter-reporter constructs.

Chapter 2- Materials and Methods

2.1 Animal Husbandry

Mouse strains C57BL/6 and Cast/EiJ were bred and maintained in the Babraham Institute Biological Support Unit. Ambient temperature was ~19-21°C and relative humidity 52%. Lighting was provided on a 12-hour light: 12-hour dark cycle including 15 min 'dawn' and 'dusk' periods of subdued lighting. After weaning, mice were transferred to individually ventilated cages with 1-5 mice per cage. Mice were fed CRM (P) VP diet (Special Diet Services) ad libitum and received seeds (e.g., sunflower, millet) at the time of cage-cleaning as part of their environmental enrichment. Breeding and maintenance of these strains were performed under licenses issued by the Home Office (UK) in accordance with the Animals (Scientific Procedures) Act 1986 and were approved by the Animal Welfare and Ethical Review Body at the Babraham Institute. Tissues were collected from newborn or adult mice and either frozen for molecular biology or processed for cell culture. Neural stem cell and primary neuron samples used in the Enhancer-reporter gene assay were obtained from C57BL/6 newborn mice obtained from the BSU facility at the University of Liverpool. These mice were kept under similar conditions to those at the Babraham institute and maintained under the Plagge project licence PPL PP0116966. Frozen tissues from C57BL/6 and Japanese Fancy Mouse (JF1) hybrid mice were kindly provided by Dr Philippe Arnaud, Université Clermont Auvergne, France.

2.2 Primers and targets

2.2.1 Primer generation

DNA primers were designed using the Primer3web V 4.1.0 (https://primer3.ut.ee/) website according to the target gene, regulatory element or CpG island they intended to amplify. Unless deviation from the standard was required, the primers were designed to meet the specifications of a CG content between 40-60%, a length of 18-23 nucleotides, a Tm of 57-62°C, and a low potential for secondary structures. Primers designed for generating amplicons used in pyrosequencing were designed using PyroMark Assay Design Software 2.0. and required an additional modification of biotinylation in one of the primer pairs highlighted as **(Btn)** for downstream applications as well as the design of a sequencing primer. Once designed, primers were ordered from Sigma Aldrich as a lyophilised powder and reconstituted into a 10µM working stock.

Table 2.1: Primers used in experiments. Chromosome positions refer to the mouse GRCm38/mm10 genome version. Primers are divided into 4 subsections: Pyrosequencing primers were designed to amplify both bisulphite converted CpG elements for methylation analysis and SNP variants located in the exons of multiple genes within the *Trappc9-Peg13* cluster. Several primers contained a biotinylation modification (**Btn**) for downstream processes of the pyrosequencing protocol. Bisulphite Sanger sequencing primers designed to amplify bisulphite converted DNA used in cloning experiments and Sanger sequencing. SC-GEM primers were used in single cell assays to determine allelic expression bias and methylation frequency. Regulatory element-luciferase primers targeting potential regulatory elements of the *Trappc9* gene and its promoter.

Primer Name	**Sequence**	**Position**
Pyrosequencing		
Ago2 Meth F1	ATTAGTATTTGAATGGGGA	Ago2 CGI
Ago2 Meth R1B	**(Btn)** CATAACTATAAAACCCAACAC	
Ago2 Meth S1	GTATTTGAATGGGGAGG	
Chrac1 Meth F1B	**(Btn)**GGAGGAGGAGGTGTGTAG	Chrac1 CGI
Chrac1 R1	CCTTAACCTTTAACTACCCTAC	
Chrac1 Meth S1	CCTTAACCTTTAACTACCC	
Kcnk9 Meth F1	GAGGTGTTTAGGATTAGATAGTT	Kcnk9 CGI
Kcnk9 Meth R1B	**(Btn)**CCTATACCAAACCTAAATCAA	
Kcnk9 Meth S1	GGTGTTTAGGATTAGATAGT	
Peg13 Meth F1	AGGTTTTGTGTGATAGTTTATTTAAG	Peg13 CGI
Peg13 Meth R1B	**(Btn)** TCTTCTATCCAACCATTTTCA	
Peg13 Meth S1	GTTGGTGTTATGTAGA	
Trappc9 CGI1 Meth F1	GGTAGAGGGATTGAGTAGTTTG	Trappc9 CGI1
Trappc9 CGI1 Meth R1B	**(Btn)** CCCCAAACAAATCTTAAACC	
Trappc9 CGI1 Meth S1	GGGATTGAGTAGTTTGGA	
Trappc9 CGI2 Meth F3	GTAGGGTGATGTGGTTGAG	Trappc9 CGI2
Trappc9 CGI2 Meth R3B	**(Btn)**ACCCTTAAACTCCAACACAAT	
Trappc9 CGI2 Meth S3	GGTGATGTGGTTGAGTA	
Ago2 snp F1	CGTTGTCATGAGGCACTTACC	Ago2 Exon 4
Ago2 snp R1B	**(Btn)** GTTGGAACAGCCTTCAGATGC	Ago2 Exon 5
Ago2 snp S1	CCATGAGGTACACCCC	Ago2 Exon 4/5
Chrac1 snp F77B	**(Btn)** AAGCCAAGAAAGCACTGAC	Chrac1 Exon 2
Chrac1 snp R77	AATATATCTGCGAGAAACTGAAG	Chrac1 Exon 3
Chrac1 snp S9	GAGAAACTGAAGTGTCTCC	Chrac1 Exon 2
Kcnk9 snp F1	AACGTGCGTACCCTGTCCT	Kcnk9 Exon 1
Kcnk9 snp R6B	**(Btn)** CTCCTCGCGCATCTCATG	Kcnk9 Exon 1
Kcnk9 snp S1	CGGTGCCGCGGTGTT	Kcnk9 Exon 1
Peg13 snp F53	GGCAAAAGGAGGCACAGAA	Peg13 Exon 1
Peg13 snp R53B	**(Btn)** GCTGCAGGGTTCTGTGCTC	Peg13 Exon 1
Peg13 snp S1	AAGCCCAGATATCTGTGT	Peg13 Exon 1
Trappc9 snp F1	CAGCGTGCCCTCTTCATCC	Trappc9 Exon 2
Trappc9 snp R3B	**(Btn)** TGCGGTGCGTCTGGAAGT	Trappc9 Exon 2
Trappc9 snp S1	TGCCCTCTTCATCCG	Trappc9 Exon 2
Trappc9_snp1 Fb	**(Btn)** CAGCAAGTACAAGAACGCCG	Trappc9 Exon 7
Trappc9_snp1 R	CTCCATGCCACGCTTCTG	Trappc9 Exon 7
Trappc9_snp1 S	CGCTTCTGAATCGCTAGG	Trappc9 Exon 7
Trappc9_snp2 Fb	**(Btn)** AAGGACTTCAGCAAAGGCACA	Trappc9 Exon 9

Trappc9_snp2 R	ATCCTGCGGGAGGCATAGA	Trappc9 Exon 9
Trappc9_snp2 S	ATAGACCAGCTCATGCA	Trappc9 Exon 9
Bisulphite Sanger sequencing analysis		
Ago2 Full CpG island F1	TTTATAGTGAAGAAGTTTGGGGAG	Ago2 CGI
Ago2 Full CpG island R1	TAACAACTAATAAAATTTAATCCTAAAC	
Chrac1 Full CpG island F1	TGAAAGGATAAATTTGTGTAGT	Chrac1 CGI
Chrac1 Full CpG island R1	AAACCCTAAACAACCTTACAAAC	
Kcnk9 Full CpG island F1	TTATTAGTTGGTTGGGGA	Kcnk9 CGI
Kcnk9 Full CpG island R1	AAAAATAAAATCATACCCCTAAAAA	
Peg13 Full CpG island F1	GGGGTTTTATTGTGTGGG	Peg13 CGI
Peg13 Full CpG island R1	CTCCATAACTCATCATTATACTACAACCA	
Trappc9 Full CGI1 F1	AGATAGAGGATTAGGTAAGTAGGGGG	Trappc9 CGI1
Trappc9 Full CGI1 R1	TAATCTTCTACTCCAAAATCTACCAA	
CpG1_02Fw	GGTGGTTTGGAGGTTTTAGGTTGTTTAG	Trappc9 CGI1
CpG1_02Rv	CCAAACAACTAACAACAAAATAACAAACTATCC	Trappc9 CGI1
Trappc9 Full CGI2 F1	TGTGAGTTTTGTAAGGTAGAGA	Trappc9 CGI2
Trappc9 Full CGI2 R1	CTTAAACTCCAACACAATAAAAA	
CpG2_01Fw	AGAGTGTGGTATGTTTTTGTTTATTAGTGTTAATG	Trappc9 CGI2
CpG2_01Rv	AACTTCTCAAAAATCTATAACCAATCTTTAAAC	Trappc9 CGI2
Sc-GEM		
B actin expression F	AAGGCCAACCGTGAAAAGAT	Exon 3
B actin expression R	GTGGTACGACCAGAGGCATAC	Exon 4
Ago2 cDNA F	CGTTACACGATGCACTTTCG	Exon 4
Ago2 cDNA Nested R	GTTGGAACAGCCTTCAGATGC	Exon 5
Ago2 cDNA Outer R	ACTGATGGAAGCCAAACCAC	Exon 5
Peg13 cDNA F	GGCAAAAGGAGGCACAGAA	Exon 1
Peg13 cDNA Nested R	AGCCTCTGTGCTAGCGTCTC	Exon 1
Peg13 cDNA R	CTCCATGGCTCATCATTGTG	Exon1
Trappc9 cDNA F	ATTGAGCTGGAAGCCTGTGT	Exon 7
Trappc9 cDNA Nested R	ATGCTGTAGCGCTGGATTTT	Exon 8
Trappc9 cDNA Outer R	CACTCGCTTGAAGAATGCTG	Exon 8
Aqp4 F	AGCAATTGGATTTTCCGTTG	Exon 2
Aqp4 Nested F	CTGGGCAAACCACTGGATAT	Exon 3
Aqp4 R	TGAGCTCCACATCAGGACAG	Exon 4
Wnt8b F	ACTCCCGAAATGGACAACTG	Exon 1
Wnt8b Nested F	GAGAGGCAATTTCCAAGCAG	Exon 2
Wnt8b R	TTACACGTGCGTTTCATGGT	Exon 3
Cdk1 F	CTGGGCAGTTCATGGATTCT	Exon 4
Cdk1 Nested F	ATCAAACTGGCTGATTTCGG	Exon 5
Cdk1 R	GATGTCAACCGGAGTGGAGT	Exon 6
Eomes F	GGCCTACCAAAACACGGATA	Exon 4
Eomes Nested F	TTCCGGGACAACTACGATTC	Exon 5
Eomes R	GTCACGTCAACTTCACAGCA	Exon 5
Mxd3 F	CCAGGGTGCATATCCAGAAG	Exon 3
Mxd3 Nested F	CTTCAGGCCTGTCCTCTGAG	Exon 4
Mxd3 R	CACATCCACCTCCAGATCCT	Exon 5
Calb2 F	TGGCGGAAGTATGACACAGA	Exon 5
Calb2 Nested F	CTCCTGAAGAAGGCCAACAG	Exon 6

Calb2 R	CCAATTTGCCATCTCCATTT	Exon 7
Igfbpl1 F	CCCTTCTGACCATGAGACCA	Exon 2
Igfbpl1 Nested F	AGGGAGTGTACCACTGCCAC	Exon 3
Igfbpl1 R	CACCCGGAACTGAGGAGTAG	Exon 3
Nestin F'	GATCGCTCAGATCCTGGAAG	Exon 2
Nestin Nested F'	TGAGAACTCTCGCTTGCAGA	Exon 3
Nestin R'	AGAGAAGGATGTTGGGCTGA	Exon 4
Trappc9 CGI1 Long F’ meth	CTGAGCAGCCTGGAACCT	Trappc9 CGI1
Trappc9 CGI1 Short F’ meth	GCTTCGGCCTAAGATCTGC	
Trappc9 CGI1 Nested R’ meth	CGCAGGCTCCAGGTCTTC	
Trappc9 CGI1 R’ meth	ATCCCTCCTCCCTCTTCG	
Trappc9 CGI2 Long F’ meth	TTGGGATAGAAGGCCACATC	Trappc9 CGI2
Trappc9 CGI2 Short F’ meth	GGAATCGTAGAGCGTGGAAC	
Trappc9 CGI2 Nested R’ meth	ACCATCACCGATTGCTTCTC	
Trappc9 CGI2 R’ meth	GTGGTAGGCCTCATCACCAT	
Peg13 CpG Long F’ meth	GGCAGTGTCGCAGGTCTT	Peg13 CGI
Peg13 CpG Short F’ meth	TCGTCTACATAGCACCAGCG	
Peg13 CpG Nested R’ meth	TTTTAGGCCTTGTGTGATAGCTC	
Peg13 CpG R’ meth	TTGAGATTTTAGGCCTTGTGTG	
Ago2 CpG Long F’ meth	CCCCAACACTTGTTTTCTCA	Ago2 CGI
Ago2 CpG Short F’ meth	GGAAATGCGTCCGTGTTTT	
Ago2 CpG Nested R’ meth	CCCCATTCAAGTGCTAATCG	
Ago2 CpG R’ meth	CCTCCCCATTCAAGTGCT	
Chrac1 CpG Long F’ meth	ACGGACGCGGAAGATG	Chrac1 CGI
Chrac1 CpG Short F’ meth	ATCATGAAGAGCTCTCCCGA	
Chrac1 CpG Nested R’ meth	CAGCGCCTCCTGGTTGAT	
Chrac1 CpG R’ meth	GCTCACCGTGGCCTTG	
Kcnk9 CpG Long F’ meth	ATGGTTGTGCGGATGGAT	Kcnk9 CGI
Kcnk9 CpG Short F’ meth	CTGTCTGGTCCTGGGCAC	
Kcnk9 CpG Nested R’ meth	CTCCCTCTGTCCCGGCTA	
Kcnk9 CpG R’ meth	CATCAGGGATGGGGAACC	
Igf2r CpG Long F’ meth	AATCTTGCGCAGGAGTGTG	Igf2r DMR
Igf2r CpG Short F’ meth	TTTGAGCTTGCCTCTCTTGC	
Igf2r CpG Nested R’ meth	TGCCATGTTACAGGAGAGATG	
Igf2r CpG R’ meth	GTTCTGTGATCAGGGCCAAC	
H19 CpG Long F’ meth	GACCATGCCCTATTCTTGGA	H19 DMR
H19 CpG Short F’ meth	GATTGCGCCAAACCTAAAGA	
H19 CpG Nested R’ meth	CCATTTGTGAATTCCAATACCAG	
H19 CpG R’ meth	ACAGCATTGCCATTTGTGAA	
Regulatory element-Luciferase assay		
Trappc9 Promoter region F1	CCACTCTGGTGTTTCTATTACT	Chr15: 72,931,663-72,934,420
Trappc9 Promoter region F2	CAGACCCTGGTGATAGCTT	
Trappc9 Promoter region R1	CCAAGCAGGAGCCGAG	

Trappc9 Promoter region R2	CTGGTGCTACTGTTTGGTATT	
Regulatory element- A F1	TAGATCAGACGCAGGACACA	chr15:72625401-72626800
Regulatory element- A R1	ACTTTGAAACTTCCCTCTTC	
Regulatory element- B F1	AGGAACTGACCCATAGGAGT	chr15:72849700-72853300
Regulatory element- B R1	CTGGGATTTGAACTCTGGA	
Regulatory element- C F1	TCACCTCTTCAAGACTCCAT	chr15:72860700-72862300
Regulatory element- C R1	CAATGGCATTCCAAACCAA	
Regulatory element- D F1	TGCACCACTCTAGCAATCTT	chr15:73019700-73023500
Regulatory element- D R1	GTCCAGGGAAGCCTTGAA	
Regulatory element- E F1	GCCCAATACCCAACAGCAG	chr15:73066701-73068500
Regulatory element- E R1	GGTTTCCTGTGTCGTTCCG	
Regulatory element- 2 F1	ACGAATGTGACCTCTCCTCC	chr15:72657300-72658700
Regulatory element- 2 R1	CCTTAGTTGTGCACACCTACA	
Regulatory element- 8 F1	GGAGGCACACAGAACCACA	chr15:72,891,601-72,892,600
Regulatory element- 8 R1	CCTCTCAAGAGCTCTAGTGAT	

2.2.2 SNP variants

SNP variants between C57BL/6 x Cast/EiJ hybrid crosses, and C57BL/6 x Japanese Fancy Mouse (JF1) in the case of *Chrac1* due to a lack of SNPs for this gene in the C57BL/6 x Cast/EiJ cross, were identified using both the ENSEMBL database https://www.ensembl.org/Mus_musculus/Info/Index and the Riken mouse database https://molossinus.brc.riken.jp/mogplus/#JF1 Primers were designed to flank these variants and span introns where applicable to ensure only reverse transcribed RNA was amplified to determine allele specific expression patterns.

Table 2.2: SNP targets for allele specific expression and methylation analysis using pyrosequencing and Sanger sequencing. SNPs indicate a variation of a single nucleotide between the C57BL/6 and Cast/EiJ mouse strains, apart from *Chrac1* which indicates a SNP between the C57BL/6 and JF1 mouse strains.

Gene	SNP ID (C57BL/6 x Cast/EiJ)	Locus (GRCm38/mm10)	Variation (Forward strand)	Exon
Expression				
Trappc9	rs31440851	73,058,335	A/G	2
	rs31443479	73,026,031	T/C	7 (include alt transcript)
Peg13	rs238259968	72,809,627	G/A	1
	rS31423566	72,809,619	A/G	1
Ago2	rs232384843	73,128,451	A/G	5
Chrac1 (C57BL/6 x JF1)	rs248258787	73,092,934	C/T	2
Kcnkk9	rs225149059	72,546,199	G/A	1
Methylation				
Trappc9 CGI1	rs31441779	73,060,948	A/G	N/A
Trappc9 CGI2	rs31440849	73,058,251	A/G	N/A
Peg13 CGI	rs31423566	72,809,619	A/G	N/A
	rs238259968	72,809,627	G/A	N/A
	rs31423567	72,809,785	G/A	N/A
	rs31423568	72,809,828	A/G	N/A
	rs242845136	72,809,926	G/A	N/A
Ago2 CGI	rs257455001	73,184,532	A/G	N/A
	rs227718844	73,184,580	G/A	N/A
Chrac1 CGI	rs249421726	73,090,306	A/T	N/A
Kcnk9 CGI	rs580519880	72,546,763	C/G	N/A
	rs587113152	72,546,762	C/T	N/A
	rs580268814	72,546,744	T/C	N/A
	rs583874465	72,546,695	T/G	N/A
	rs219245628	72,546,594	C/A	N/A
	rs259286628	72,546,593	T/G	N/A

2.3 Cell culture

All medium was warmed to 37^{o}C prior to use.

2.3.1 Neural stem cells

Neural stem cell culture was performed as described previously by Chojnacki & Weiss, 2008; / Ferrón et al., 2007 with slight modifications. Briefly, hippocampi were dissected from newborn mouse brain in ice-cold Neurosphere growth medium (DMEM/F12 (GibcoBRL: 11320-074) supplemented with 0.6% w/v glucose (Sigma: G-7021), 0.1% $NaHCO_3$ (Sigma: S-8761), 5 mM HEPES (Sigma: H-0887), 2 mM L-Gln (GibcoBRL: 25030-024), 100 U/ml penicillin, 0.1 mg/ml streptomycin (GibcoBRL: 15240-062), 1x B27 (Gibco: 17504044), 10 ng/ml FGF-2 (Peprotech: 100-18C), 20 ng/ml EGF (Peprotech: AF-100-15), and 4 mg/ml BSA (Sigma, # B-4287)). The dissected hippocampi tissue was then transferred into 2 mL Accutase (Gibco: A11105-01) and kept at 37°C for ~10 minutes before dissociation into a single-cell suspension by gentle trituration. Trituration was performed ~20-30 times, ensuring the pipette tip was approximately 0.25 cm from the base of the falcon tube with slow discharge from the pipette tip, to minimise cell death and bubble formation. Following this, centrifugation was performed at 200g for 5 minutes. The supernatant was removed, and cells were resuspended in 1mL of fresh growth medium and triturated again to achieve a single cell suspension. Cells were counted using a 1:1 ratio of cell suspension to Trypan blue and a haemocytometer, and then plated at a density of 3,000 cells/cm^2 in non-TC-treated, non-pyrogenic suspension 60mm cell culture dishes (Corning). Neural stem cells were incubated at 37°C, 5% CO_2 for 6-8 days with intermittent medium supplementation to produce 3D Neurosphere organoids before passaging. Passaging was performed by transference of the cell media into a 15 mL falcon tube and centrifugation performed as

above. The supernatant was removed, and the cell pellet was washed and resuspended in 1 mL Ca^{2+}/Mg^{2+}-free PBS before another centrifugation as above. The PBS supernatant was removed, and the cells were treated as above following the Accutase dissociation into a single-cell suspension. Cells were re-plated at a lower density of 1,500-2,000 cells/cm^2. Alternatively, neurospheres could be stored in liquid nitrogen after centrifugation and resuspension in growth medium supplemented with 10% DMSO. For bulk and single-cell gene expression analysis, neurospheres at early passage numbers (P3-P5) were used. Alternatively, neurospheres could be differentiated into neurons at the point of passaging by seeding a single cell suspension on Poly-L-Lysine (Sigma: P4707) coated cell adherent dishes in differentiation medium (growth medium without EGF, FGF-2 and BSA, but containing 1% FBS). For single-neuron analysis, selection against replicating glial cells was started after two days of culture with 2 µM Cytosine β-D-arabinofuranoside (AraC) (Sigma: C1768) as described in the next paragraph. This differentiation process into neurons was used for the SC-GEM process outlined later.

2.3.2 Primary Neurons

Primary hippocampal neurons were cultured as described by Beaudoin et al., 2012; Ioannou et al., 2019 with modifications. Hippocampi were dissected from newborn mouse brain in ice-cold dissection medium (HBSS (Sigma: H9394) supplemented with 0.1% w/v glucose (Sigma: G7021), 10mM Hepes pH 7.4 (Sigma: H0887), and 1% Na-pyruvate (Sigma: S8636)) and the tissue was dissociated by adding an equal volume of 2x Papain stock solution (Worthington- Cat: LK003176) (generated by dissolving Papain powder in dissection media) which was then incubated at 37°C for 20 min. The supernatant was removed carefully, and the tissue was gently washed with plating medium (MEM (Gibco: 21010-026) supplemented

with 0.45% glucose, 10% FBS (Sigma: F9665), 1% Na-pyruvate, 2 mM L-Glutamine, 100 U/ml penicillin, 0.1 mg/ml streptomycin). The plating media used for washing was carefully removed and fresh plating media was added with gentle trituration performed as described above. The dissociated tissue was rinsed through a 70-µm cell strainer (Corning: 431751) and the collected cells centrifuged at 200g for 5 min. The plating media supernatant was removed, and the cells were resuspended in neuronal medium (Neurobasal medium (Gibco: 10888-022) supplemented with 2 mM L-glutamine, 100 U/ml penicillin, 0.1 mg/ml streptomycin, 1x B27 (Gibco). Cell counts were assessed as described above and the cells were then plated in Poly-L-Lysine (Sigma: P4707) coated dishes at a density of 60,000 cells/cm^2. Medium was replaced the following day, and on day two selection against replicating non-neuronal cells was started with replacing half the media with neuronal medium containing 4 µM AraC to a final dilution of 2 µM AraC to prevent glial cell growth while maintaining growth factors secreted by the growing neurons. Primary neurons were incubated at 37°C, 5% CO_2 for 5-7 days with half the medium being replaced with fresh neuronal medium every other day to dilute the AraC.

2.3.3 Fibroblasts

Mouse embryonic fibroblasts were prepared as described by Matise & Joyner, 1999 and cultured from frozen stocks in Hepes-buffered DMEM (Sigma) supplemented with 10% FBS, 2 mM L-Glutamine, 100 U/ml penicillin, 0.1 mg/ml streptomycin.

2.4 Regulatory elements, reporter gene constructs and cloning

2.4.1 Identification of potential regulatory elements

Potential brain-specific gene regulatory elements were identified from histone modification, CTCF ChIP assay, DNAse I and ATAC-seq hypersensitivity data for the newborn (P0) mouse brain in comparison to peripheral tissues. These data were extracted from the databases ENCODE3 (https://www.encodeproject.org/), UCSC Genome Browser (https://genome-euro.ucsc.edu/) and ENSEMBL (https://www.ensembl.org/Mus_musculus/Info/Index). Seven candidate brain-specific regulatory elements across the *Trappc9-Peg13* locus were found; their genomic positions (mouse GRCm38/mm10 genome version) and features are listed in (Table 5.1). To evaluate the functionality of these regulatory elements in transfected cells, promoter-reporter gene plasmids were generated.

2.4.2 Generation of reporter-gene constructs

The regulatory elements were amplified from C57BL/6 genomic DNA using Q5™ High-Fidelity DNA Polymerase (New England Biolabs: M0491S), cloned into TOPO®-plasmids (Invitrogen) and sequenced for confirmation. Similarly, four *Trappc9* promoter fragments of different lengths were cloned into TOPO®-plasmids (positions 73,061,805–73,060,204 bp, 73,061,805–73,060,975, 73,061,418–73,060,204 bp and 73,061,418–73,060,975 bp in GRC38/mm10). Topo® cloning plasmids were generated by combining ~3 µL purified target PCR product, 1µL salt solution (1.2 M NaCl, 0.06M $MgCl_2$), 1 µL Nuclease free water and 1 µL Topo vector plasmid (10 ng/µL). Samples were incubated at RT for 10 min and then placed in ice and used for chemical transformation of One Shot® competent cells. From the Topo® cloning reaction 2 µL was transferred into a vial of One Shot® chemically competent *E.coli*

and incubated on ice for 30 min. The *E.coli* Topo® clone mixture was then heat shocked in a pre-heated 42°C water bath without shaking for 30 seconds and immediately incubated in ice. Following heat shock 250 µL S.O.C cell recovery media was added to the sample which was then incubated at 37°C with horizontal shaking at 200 rpm for 1 hr. After incubation ~50 µL was spread on a prewarmed antibiotic selection plate and incubated at 37°C overnight. The next day ~10 colonies were selected and placed in separate 15mL bacterial falcon tubes containing 3mL LB broth (supplemented with ~50 µg/mL of the same antibiotic used for initial selection) and incubated at 37°C with horizontal shaking at 200 rpm overnight. Samples were then purified using the QIAprep Spin Miniprep Kit (Qiagen: 27104) and tested using restriction digests to ensure correct plasmid clone orientation and presence.

The firefly luciferase-encoding pGL4.23 [*luc2*/minP] vector (Promega) was used to generate reporter-gene constructs. First, the endogenous minimal promoter was removed from the pGL 4.23 [*Luc2*/minP] vector via a HindIII and NcoI digest. The restriction digests were performed using up to 1 µg of plasmid DNA combined with 3 µL (10x) enzyme appropriate restriction buffer, 1 µL of the restriction enzyme (*HindIII / NcoI*) and made up to 30 µL using nuclease free water. The samples were incubated at 37°C for 1 hr and purified using gel electrophoresis and a gel clean up kit. Once the minimal promoter was removed for the pGL 4.23 plasmid, the *Trappc9* promoter constructs were cut out of the respective TOPO®-plasmid they were originally cloned into using an Eco53KI and EcoRV double digest and then purified using gel electrophoresis and a clean-up kit. As the HindIII and NcoI digest used to remove the minimal promoter left sticky ends on the remaining pGL 4.23 vector, these needed to be blunted before the blunt cut *Trappc9* promoter fragment could be inserted. Blunting was achieved using T4 DNA polymerase (NEB, M0203) with all of the purified, digested pGl 4.23 vector being incubated in CutSmart® buffer (1x), supplemented with 100

µM dNTPs and 1 unit of T4 DNA Polymerase per microgram of vector DNA. The sample was then incubated at 12°C for 15 mins before stopping the reaction by adding EDTA to a final concentration of 10 mM and heating to 75°C for 20 minutes. The newly blunted pGL 4.23 vector was CIP treated to prevent self religation using 2 µL CutSmart® Buffer (10X), 1 unit of CIP (NEB, M0290), 1 pmol of DNA ends and made up to 20 µL using purified water. Samples were incubated at 37°C for 30 minutes and then purified. After CIP treatment the *Trappc9* promoter fragments were ligated into the pGL 4.23 vector using a standard blunt end ligation protocol. Individually, the four *Trappc9* promoter fragments were mixed with ~ 50 ng of the CIP treated pGL 4.23 vector using a 3x molar excess of vector to insert ratio. To this mixture, 5 µL of 5x Ligation buffer (containing ATP and PEG4000 at 25% w/v (final conc. 5%)), 1 µL of T4 DNA Ligase (NEB, M0202S) and purified water up to 15 µL was added. The sample was left at RT overnight and used for bacterial transformation the following day. Additionally, a ligation without a *Trappc9* promoter insert was used as a control to ensure the procedure worked correctly.

The four *Trappc9* promoter plasmids were assessed in a preliminary reporter gene assay for their activity. The promoter fragment 73,061,418–73,060,975 bp, which avoids an upstream dinucleotide repeat sequence stretch (comprised primarily of GT ad GA repeats) and ends before the exon 1 splice donor site, showed the highest activity and was used in further experiments in combination with the identified regulatory elements. The regulatory elements were cloned into the pGL vector at the BamHI site downstream of the *Luc2* reporter gene to reflect the same relative orientation to the *Trappc9* promoter as in the genome. Only the Reg-E element was cloned upstream of the *Trappc9* promoter as shown in Figure 2.1.

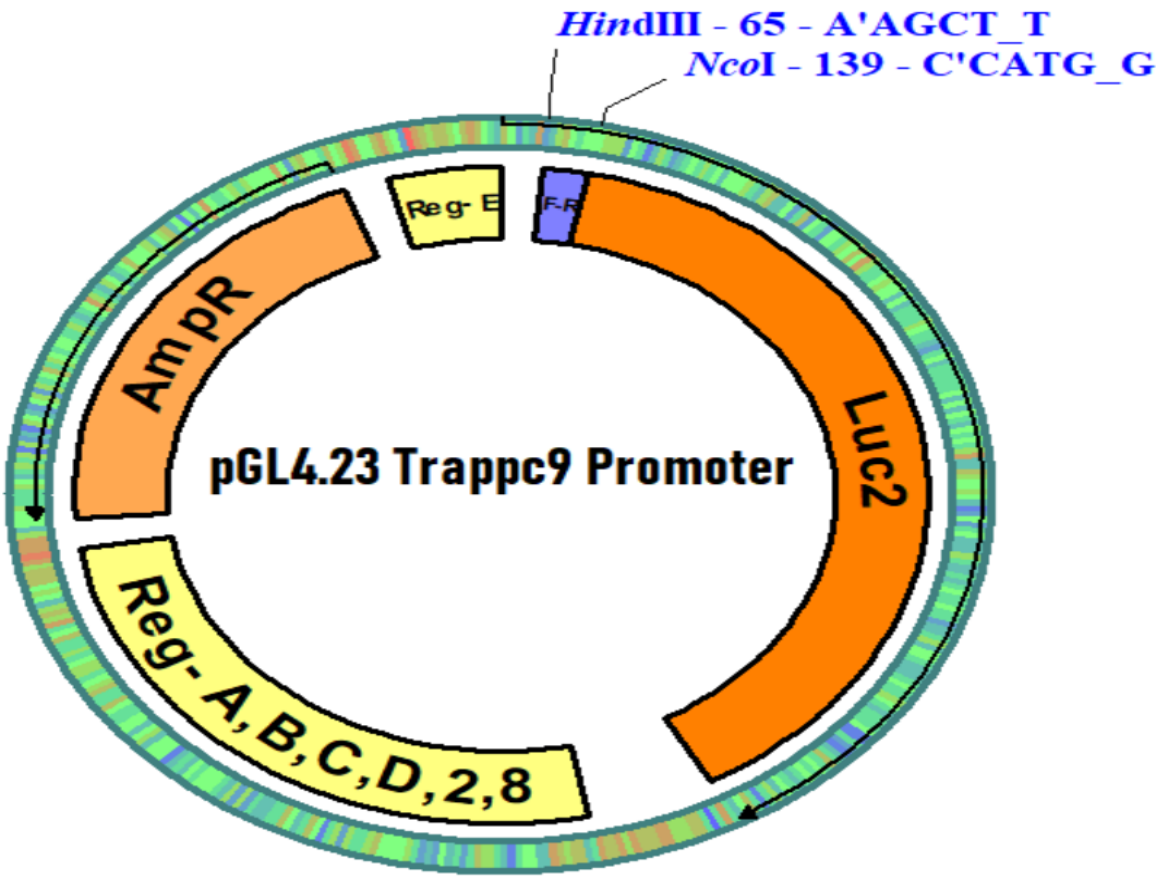


Figure 2.1: Map of the pGL4.23 plasmid with the *Trappc9* promoter and regulatory element positions shown. The minimal promoter was removed via a *HindIII* and *NcoI* restriction digest and one of the four candidate *Trappc9* promoter amplicons (shown in blue) was ligated into the plasmid to replace the minimal promoter. The regulatory elements identified using multiple genomic databases (see Chapter 5, Figure 5.1) (shown in yellow) were inserted into the plasmid in a manner that reflected their natural position in the genome. Regulatory elements A, B, C, D, 2 and 8 are all downstream of the *Trappc9* promoter in the mouse genome and so were all inserted after the *Luc2* gene to reflect this. The Reg-E candidate however is located upstream of the *Trappc9* promoter region and so was inserted upstream of the *HindIII* site to reflect its natural genomic position

2.5 Cell transfections and Reporter gene assays

The promoter-reporter gene plasmids were transfected into fibroblasts and primary hippocampal neurons once the cells reached ~70-90% confluency using the Lipofectamine 2000 reagent (Invitrogen: 11668030). Two separate mixtures were created, one containing the plasmids used in the transfection and one containing the Lipofectamine used for the transfection. The firefly luciferase-based test constructs were mixed with a Renilla luciferase control plasmid (pGL4.74, Promega) at a 100:1 ratio to normalise for transfection efficiency. This is due to previous studies determining that the transfection efficiency of fibroblasts is slightly higher (~ 15%) than that of primary neurons (~ 5%) when using lipofectamine 2000 as the transfection agent and so normalisation ensures comparability between the two cell lines (Alabdullah et al., 2019; M. Lee et al., 2017). The firefly luciferase and Renilla luciferase plasmids were mixed with Opti-MEM reduced serum media (Gibco: 31985070) to a final volume of 187.5 µL. Approximately 3.75 µg total of Firefly luciferase plasmid and 37.5 ng total Renilla luciferase plasmid (100:1 ratio) were mixed with the Opti-MEM media to achieve a final DNA concentration of 300 ng Firefly and 3 ng Renilla luciferase per well in a 48-well plate. In separate Eppendorf tubes, aliquots containing 2 µL of the Lipofectamine 2000 reagent mixed with Opti-MEM reduced serum media to a final volume of 37.5 µL were prepared alongside the plasmid mixtures.

Both the plasmid and lipofectamine mixtures were incubated at RT for ~10 min before the two mixtures were combined in a new Eppendorf tube at a 1:1 ratio of 37.5 µL each to make a total volume of 75 µL with the combined mixture incubated at RT for a further 5 minutes. Due to the plasmid vector mixture containing 187.5 µL total volume, one mixture containing both the firefly and Renilla vectors could be used to mix with 5 separate aliquots containing

the Lipofectamine at a 1:1 ratio for technical repeats if needed. From this combined mixture of lipofectamine and luciferase plasmids 30 µL was taken and added slowly to the cell media for each well containing either primary hippocampal neurons or fibroblasts. After 24 hours the media was removed from each well and replaced with fresh Neuronal or fibroblast growth media to prevent extended exposure to the lipofectamine reagent which appeared to trigger cell death upon longer exposures. Cells were lysed 48 h post transfection and luciferase activities were measured using the Dual-Luciferase® Reporter Assay System (Promega: E1910) on a Glomax Multi Detection System (Promega). Cell lysis was performed by removal of cell media and the use of ~70 µL 1x Passive Lysis Buffer per well, incubated at RT on a rocking platform for 15 mins. From the lysate 10 µL was transferred to a flat, white bottomed 96 well plate and analysed by the addition of Luciferase Assay Substrate and Stop & Glo® Substrate according to protocol instructions and measured using the Glomax Multi Detection System.

2.6 Molecular biology

2.6.1 DNA and RNA extraction

Genomic DNA and RNA was isolated from neurospheres, brain and kidney tissues using TRIzol™ reagent (Invitrogen: 15596026). TRIzol™ was added at a volume dependent on the sample used, with 1mL TRIzol™ for every 50-100mg of brain or kidney tissue and 0.75 mL TRIzol™ for every 0.25mL of cell suspension (5-10 x10^6) neural stem cells collected after centrifugation at 200 g for 5 mins. Samples were homogenised in TRIzol™ solution and incubated at RT for 5 mins. Chloroform was added at a ratio of 0.2mL for every 1 mL or TRIzol™ used and incubated at RT for 2-3 mins. Once incubated, the samples were

centrifuged for 15 mins at 12,000 × g at 4°C and the RNA containing aqueous phase was transferred into a new 1.5 mL microcentrifuge tube. Approximately 0.5 mL of isopropanol per 1 mL of TRIzol™ was added to this aqueous phase which was allowed to incubate at RT for 10 min before another centrifugation at 12,000 × g and 4°C for 10 min. The supernatant was removed leaving a white-gel like pellet containing the RNA. This pellet was resuspended in 1 mL of 75% ethanol (per 1 mL of TRIzol™ used) where it was then vortexed and centrifuged at 7500 x g at 4°C. Once again, the supernatant was removed and the pellet was left to air dry with the lid open for 5-10 min until all residual ethanol had evaporated. The pellet was then resuspended in ~30 µL of RNAse free water and incubated in a heat block set at 55–60°C for 10–15 min. Once resuspended the RNA was treated with DNAse I (New England Biolabs: M0303S) to remove any traces of genomic DNA and then inactivated by incubating the sample at 75°C for 10 minutes before cDNA was synthesised.

Genomic DNA was isolated from the remaining interphase and lower phenol-chloroform phase after removal of the RNA containing aqueous phase. To the remaining phases 0.3 mL of 100% ethanol per 1 mL TRIzol™ used was added, mixed via inversion and incubated at RT for 2-3 min. The sample was then centrifuged at 2000 x g at 4°C for 5 mins producing a small pellet of DNA. The supernatant, which contains the remaining proteins, was removed and the pellet was resuspended in 1 mL sodium citrate in 10% ethanol, pH 8.5, per 1 mL of TRIzol™ where it was incubated at RT for 30 min with occasional mixing by gentle inversion. The mixture was then centrifuged again at 2000 x g at 4°C after which the supernatant was removed. This process was repeated once again starting from the addition of the 1 mL sodium citrate. Once repeated, the pellet was then resuspended in 1.5 mL of 75% ethanol per 1 mL of TRIzol™ used and incubated for 10-20 min occasionally mixed by gentle

inversion. The sample was again centrifuged at 2000 x g at 4°C and the supernatant removed with the pellet left to air dry for 5-10 min until the residual ethanol had evaporated. Once dry, the pellet was resuspended in 0.3–0.6 mL of 8 mM NaOH and centrifuged at 12,000 × g at 4°C for 10 min to remove any insoluble material. The supernatant was transferred to a new 1.5 mL microcentrifuge tube and the pH adjusted to 7-8 using HEPES. Once purified, both DNA and RNA were measured using absorbance to measure purity and yield. We aimed to achieve as good an A260/A280 ratio as possible, with 1.8 being pure DNA and 2.0 being pure RNA.

2.6.2 DNA isolation for genotyping

DNA isolation for genotyping was performed using a modified version of the protocol stated by Laird et al., 1991. A short segment of tail or ear notch was taken from the mouse and lysed in lysis buffer (100 mM Tris, 5 mM EDTA, 200 mM NaCl, 0.2% SDS, ph8.5) containing 100µg/ml Proteinase K (Thermofisher: EO0491) at 55°C overnight. The volume of lysis buffer used was proportionate to size of tissue ~200 µl per 1 cm of tail tip, or 50 µl for an ear notch. After incubation, samples were centrifuged at max speed for 2 mins and 5 µl of the sample was transferred to a new 1.5 mL microcentrifuge tube and diluted in 95 µl of nuclease free water. The diluted sample was incubated at 95°C for 15 min to deactivate the proteinase K.

2.6.3 Reverse transcription

RNA isolated from neural stem cells, primary neurons and bulk tissues was reverse transcribed for expression analysis via qPCR, pyrosequencing, and Sanger sequencing. The

cDNA was synthesised according to the ProtoScript® II Reverse Transcriptase (New England Biolabs: M0368) protocol. In an RNAse free microcentrifuge tube ~1 µg of isolated RNA was mixed with 2 µL of random hexamer primers (60µM), 1 µL of dNTP's (10 mM) and topped up to 10 µL with nuclease-free water. The sample was incubated at 65°C to denature any RNA/primer interactions. Following denaturation, 4 µL of 5X Protoscript II buffer, 2 µL of 0.1M DTT, 1 µL Protoscript II Reverse Transcriptase (200 U/µL), 0.2 µL RNase Inhibitor (40 U/µL) and 2.8 µL Nuclease-free water were added to the sample. The sample was then incubated in a thermocycler at 25°C for 5 mins, 42°C for 1 hr and 65°C for 20 mins. After conversion of the RNA into cDNA the sample was treated with 0.5 µl RNase H (New England Biolabs: M0297S) and incubated for a further 20 mins at 37°C to remove any RNA contaminants. The sample was then diluted 5-fold for any downstream applications.

2.6.4 PCR

PCR reactions were performed using either GoTaq® Hot Start Polymerase (Promega: M5001) or Q5™ High-Fidelity DNA Polymerase (New England Biolabs: M0491S). From isolated DNA samples, 3 µl of DNA template was mixed with 15 µl nuclease free water, 6 µl 5x Flexi GoTaq buffer, 1.8 µl $MgCl_2$ (25 mM), 3 µl dNTP-mix (2 mM stock) of each nucleotide, 0.5 µl of each primer (gene dependent) from a 50 µM stock and 0.2 µl GoTaq HotStart-Polymerase. Samples were then incubated in a thermocycler under the following conditions, initial denaturation 98°C, 30 sec; 30 cycles of 98°C denaturation, 50-72°C annealing (primer dependent); 72°C, 30 sec; and final extension 72°C, 2 min. After amplification samples were stained with ~6 µl of a 5x DNA loading buffer containing Midori green and loaded onto an agarose gel (percentage of agarose powder dissolved in 1x Tris-acetate-EDTA buffer was dependent on amplicon size) for separation using gel electrophoresis. Samples were

visualised using a trans-illuminator and extracted from the gel using a scalpel and placed in a 1.5 mL microcentrifuge tube where the DNA was then purified using the Monarch® DNA Gel Extraction Kit (New England Biolabs: T1020S)

2.6.5 Bisulphite treatment and DNA methylation analysis

For DNA methylation analysis of CpG sites, bisulphite conversion of unmethylated cytosines was conducted using the EZ DNA Methylation-Gold™ kit (Zymo Research: D5001). All centrifugation steps were performed at 14,000 RCF and flow-through was discarded unless otherwise stated. Initially, the CT Conversion Reagent was prepared by adding 900 µl ddH_2O to the solid conversion reagent mixture, followed by 300 µl M-Dilution Buffer and 50 µl M-Dissolving Buffer. From the CT conversion reagent 130 µl was taken and added to a 20 µl (100 ng/µl) isolated Genomic DNA sample obtained from either neural stem cells grown in culture or newborn brain or kidney tissue samples. The samples were homogenised in the conversion reagent through gentle trituration and then briefly centrifuged before being incubated at 98 °C for 10 minutes then 64°C for 2.5 hrs using a thermocycler. After incubation, the samples were mixed with 600 µl M-Binding Buffer and mixed by inversion before being transferred to a Zymo-Spin™ IC Column. Samples were centrifuged for 60 sec, then 200 µl M-Wash Buffer was added and centrifugation repeated. 200 µl M-Desulphonation Buffer was added to columns and incubated at RT for 20 minutes, followed by centrifugation for 60 secs after which the columns were washed twice with 300 µl M-Wash Buffer, as above. The columns were then dry centrifuged for 5 mins to remove any ethanol carry-over before being placed into a new 1.5 mL DNAse free microcentrifuge tube. 50 µl M-Elution Buffer was added to the column matrix directly and the column left to stand for 10 mins before a final centrifugation for 60 sec to elute the DNA from the column matrix.

The bisulphite-treated and purified DNA was then used for PCR amplification of CGI fragments, followed either by methylation analysis via direct pyrosequencing or cloning of PCR products into TOPO®-vectors (Invitrogen) and Sanger sequencing of individually cloned plasmid samples. Sanger sequencing results were further analysed using the free online tool 'QUantification tool for Methylation Analysis' (QUMA; http://quma.cdb.riken.jp/).

2.7 Pyrosequencing

SNPs from cDNA samples as well as genomic DNA for methylation analysis after bisulphite treatment obtained from brain and kidney tissues as well as cultured neurospheres of hybrid mice, were sequenced using the PyroMark® Gold Q96 (Qiagen: 972804) protocol and read using a PyroMark® Q96 ID instrument (Qiagen). PCR and sequencing primers (Table 2.1) were designed using the PyroMark Assay Design Software 2.0. Amplicons of the target region were generated prior to the pyrosequencing reaction via PCR using Q5 Taq polymerase. These amplicons were then run on an agarose gel to confirm amplicon size and then extracted and purified using the Monarch® DNA Gel Extraction Kit. The biotinylated PCR products were immobilised on streptavidin-coated beads, denatured and then neutralised before being sequenced following the manufacturer's protocols. For expression data a threshold of ~70% expression from one allele was used to define a gene showing a strong expression bias with anything weaker being labelled either biallelic or maternal/ paternal preferential biallelic expression.

2.8 Single-Cell Genotyping, Expression and Methylation analysis

Isolation of single cells was performed using a modified protocol of Cheow et al., 2015, 2016; Lorthongpanich et al., 2013 (Fig 2.2). In the original Sc-GEM procedure, fluidigm chips were used to analyse hundreds of single-cells in an automated process. However, due to limitations of materials and funding, I was unable to perform the procedure this way and instead adapted the protocol to a standard qPCR technique. Single C57BL/6 x Cast/EiJ neural stem cells were obtained via dissociation of cultured neurosphere clusters with 1 mL of accutase and gentle trituration. The resultant cell suspension was then diluted using neural stem cell growth media and plated onto 6cm^2 dishes where they were either A) isolated for analysis, via manual isolation under a microscope utilising capillary action, developed originally from a protocol that isolated oocytes (Lorthongpanich et al., 2013; Cheow et al., 2015; Cheow et al., 2016)., or B) dissociated cells were differentiated into neurons after seven days of culture in differentiation medium, which included five days of AraC treatment. and then single neuron cells were taken to represent a single-cell reflection of bulk neurosphere and whole brain tissue lysate samples. Neurons were dissociated from culture dishes using ~3 mL Trypsin/EDTA (Sigma: T4049), diluted in differentiation medium and single cells isolated manually via capillary action under a microscope. Single cells were transferred into PCR tubes containing 5 µL of lysis buffer (CellsDirect Resuspension and Lysis Buffer, 10:1 (Invitrogen: 11739010)), with the exception of a no-DNA control tube which contained only the lysis buffer. Cells isolated and placed in lysis buffer were kept on ice during the isolation of subsequent cells to limit RNase activity and maintain transcriptome integrity. Additionally, to minimise any impact on DNA/RNA integrity from time spent in lysis buffer on ice, the time taken from the first cell being isolated to the last was capped at ~30

minutes allowing an isolation of ~20 cells in one sitting. Samples were incubated at RT for 5 min and then at 75°C for 10 min in order for the lysis to occur.

cDNA was synthesised by the addition of an equal volume of a 2x reverse transcription master mix comprised of 1 µL 10x Reverse transcriptase buffer, 0.4 µL 25x dNTP mix (100mM), 1 µL random hexamer primers, 0.5 µL Multiscribe^{TM} Reverse Transcriptase (Invitrogen: 4311235) and 2.1 µL Nuclease free water. Samples were briefly spun and incubated at 25°C for 10 min, 37°C for 2 hr and 85°C for 5 min. This was followed by protease treatment using 2 µL of a 1:1 mix of Qiagen protease (3.75 AU/mL) (Qiagen: 19157) and Nuclease free water to remove chromatin-associated proteins from genomic DNA. Samples were incubated at 50°C for 1.5 hrs then 70°C for 30 min.

Next, we perform single-cell restriction analysis of methylation (SCRAM), using BstUI (a methylation sensitive restriction enzyme targeting 5'-CG$^{\blacktriangledown}$CG-3'), to digest unmethylated CpG sites of genomic DNA. This provides the additional option of analysing DNA methylation of CGIs in the single cells. Both SpeI and MslI, which digest the *Dlk1*-Ig DMR and the *Igf2r* DMR respectively and cut independent of methylation were used as a control restriction digest in ~2-4 of the single-cell lysate samples. Additionally, two other controls were added, a no-digest control containing no enzyme (replaced with Nuclease-free water) was used as a positive control while a no DNA control was used as a separate control to check for DNA contamination. To each sample a 10 µL digestion master mix was added containing 1 µL BstuI, SpeI or MslI (10 U/µL), with the exception of the no digest control, 2 µL of the respective enzyme buffer (NEBuffer 4 or Cutsmart) and 7 µL nuclease free water, depending on whether it was an experimental or control digest. Samples were then incubated for 2 hrs at 60°C for the experimental BstuI digests and 37°C for the SpeI and MslI control digests.

Post digest enzymes were inactivated by the addition of 2 µL proteinase K (10 mg/mL) and incubation at 50°C for 2 hrs then 95°C for 10 min to inactivate the proteinase K. A multiplex PCR pre-amplification was performed using GoTaq polymerase to amplify all target genes. To each sample 30 µL of a PCR master mix containing 10 µL 5x GoTaq buffer, 5 µL $MgCl_2$ (25 mM), 1 µL dNTP's (25 mM), 5 µL of a primer pool containing primers for each target gene to be analysed using the SC-GEM method (500 nM per primer) (Table 2.1), 0.25 µL GoTaq polymerase (5 U/µL) and 8.75 µL Nuclease free water. The samples were then incubated under the following conditions: initial denaturation 95°C, 10 min; 30 cycles of 95°C, 30 sec; 60°C annealing/extension, 4 mins.

After the multiplex preamplification 5 µL of each single cell sample was transferred to a new PCR tube and 2 µL of Exonuclease I (4 U/µL) was added to remove unincorporated primers. The samples were incubated at 37°C for 30 min and then 80°C to inactivate the exonuclease. Samples were then diluted 10-fold and each sample was used to amplify multiple individual target genes via nested-primer qPCR. The qPCR mixture was comprised of 3 µL of diluted template, 2 µL of a target-specific primer pair comprised of the nested primer and the corresponding outer primer (5 µM) (Table 2.1), and 5 µL PowerUp SYBR Green 2x Master mix (Invitrogen: A25741) under the following conditions: initial denaturation 95°C, 10 min; 30 cycles of 95°C, 15 sec; 60°C annealing/extension 1 min; melt curve analysis 60-95°C ramp 5 sec/degree. Where applicable, these qPCR products were purified using MinElute PCR® purification kit (Qiagen: 28004) and Sanger-sequenced for cDNA SNP expression analysis.

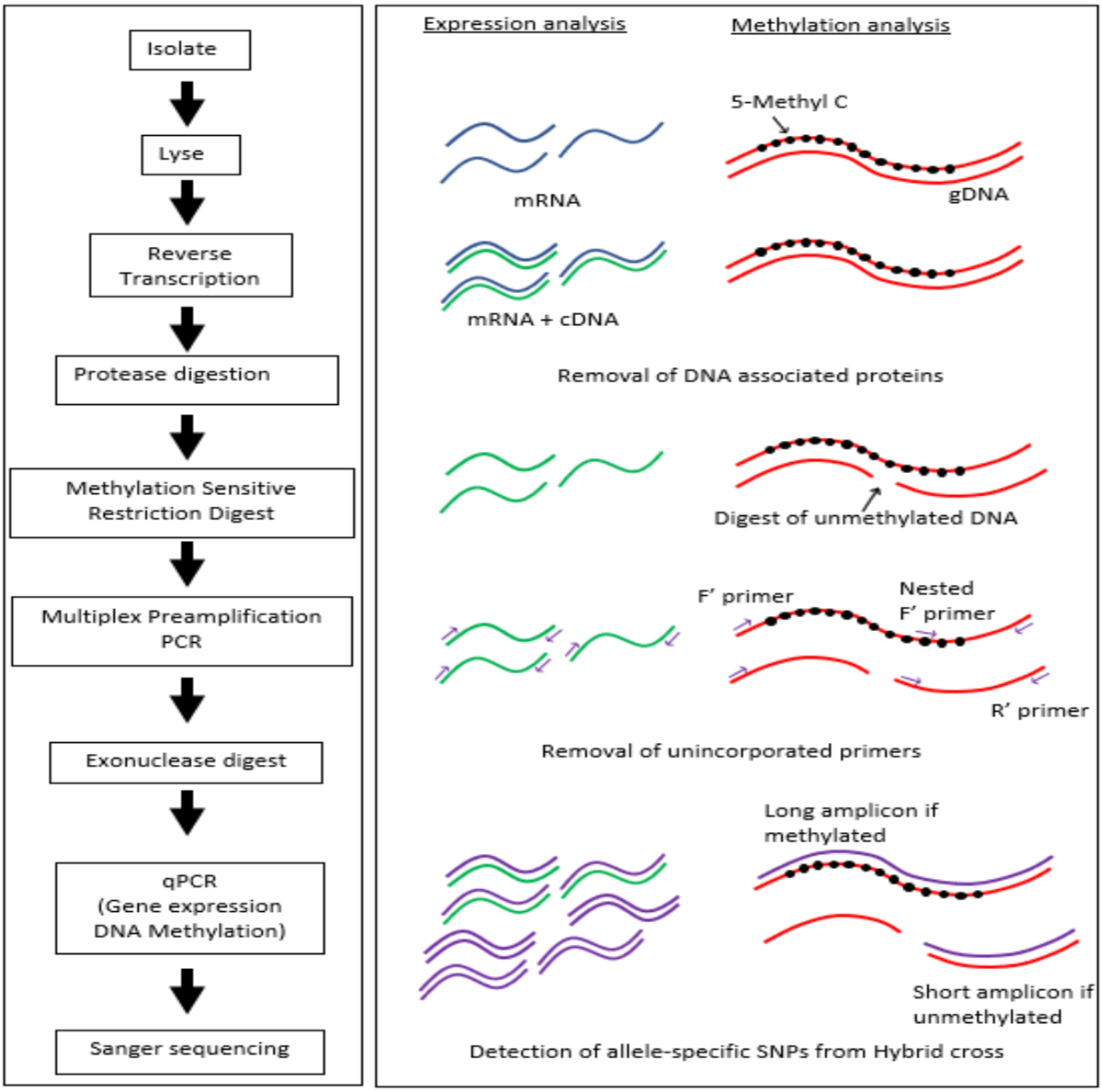


Figure 2.2: Schematic overview of the process of single-cell genotyping, expression and methylation analysis. Single-cells are isolated and then lysed to release DNA and RNA. The RNA is reverse transcribed into cDNA and then genomic DNA associated proteins are removed by protease digestion. A methylation sensitive restriction enzyme is used to digest unmethylated genomic DNA before a multiplex preamplification PCR of all genes to be analysed. After this the unincorporated primers are removed by an exonuclease enzyme and individual genes are subject to a qPCR to determine absolute gene expression levels or presence of a long or short amplicon in the case of methylation analysis. Samples are then submitted for Sanger sequencing to determine whether the expressed transcripts originated from the maternal or paternal allele based on SNPs from a hybrid mouse cross.

While the technique of identifying single-cell methylation through sc-GEM works in theory, there are multiple complications that can arise. Initially, due to working with genomic DNA of single cells, there is a low starting concentration of template for PCR reactions i.e. 0, 1 or 2 copies depending on the digest of the amplicon, leading to frequent cell dropouts due to a lack of amplification of the DNA. Additionally, when performing the BstUI restriction digest, there is no differentiation between a methylated sample producing a long amplicon and a sample where the restriction digest didn't work also resulting in the presence of a long amplicon. This was countered by performing control restriction digests. Finally, the results of the methylation analysis are binary in terms of whether the long amplicon is detected or not. There would be a 1-cycle difference expected for the Ct value theoretically if the difference in the starting material is 1 or 2 copies, based on the methylation-sensitive restriction digest, which is not reliably measurable. Due to this fact, I decided to approximate methylation data by deciding a cut-off point to decide whether the genomic DNA was considered amplified or not. I used a Ct value of 30 as the cut-off where the long and short amplicons were considered detected < 30 and undetected at Ct values > 30. If the Ct value of the short amplicon was < 30 while the long amplicon was either > 30 or did not amplify at all this was considered a successful but undetected sample. However, values just above my predetermined cut-off point while not considered amplified in my data aren't confirmed to be wrong or inaccurate. With these features in mind, I do not confirm the data I have obtained is a clear-cut picture of methylation status within the CGIs of the *Peg13-Trappc9* imprinted cluster but rather an attempt at identifying any trends that may occur between the CGIs of different genes.

2.9 Statistical analysis

The data for the promoter-reporter gene assays were analysed using GraphPad Prism v.9.3 software. Data were analysed for outliers using the ROUT method. The datasets were then analysed for normality using the Shapiro-Wilkinson test. For non-parametric datasets, Mann-Whitney U-tests were performed in comparisons of the Reg-element datasets to the basic *Trappc9* promoter dataset. For parametric datasets, unpaired *t*-tests were performed.

Chapter 3- Allelic expression biases of the imprinting cluster genes in tissues and neurospheres

3.1 Establishing an Imprinting bias

Genomic imprinting is an epigenetic phenomenon by which allelic gene expression levels are determined in a parent-of-origin specific manner (Ferguson-Smith, 2011; Patten et al., 2014; Peters, 2014; Tucci et al., 2019). The mechanism that drives this differential expression of the maternal and paternal alleles relies on chemical modifications of the DNA sequence in the form of DNA methylation. This process involves a methyl group being added to a cytosine residue at a CG dinucleotide, often found within short segments of the genome known as CpG islands which are defined as "regions of DNA with a high G + C content and a high frequency of CpG dinucleotides relative to the bulk genome", and are frequently found, but not exclusive to, promoter regions of genes (Gardiner-Garden & Frommer, 1987). Additionally, these methyl marks that facilitate a gene being imprinted are established during the development of germ cells into sperm or egg, and are maintained in somatic tissues post fertilisation by the offspring (Plasschaert & Bartolomei, 2014). Several of these CpG islands are located proximal to specific DNA regions of these imprinted genes known as DMRs promoting expression of one allele and silencing of the other (Rotondo et al., 2013). These DMR's often differ in location depending on the allele upon which they are found. Maternally methylated DMR's often, but not always, overlap with promoter regions of imprinted genes and are established prior to ovulation and are erased in the primordial germ cells of the next generation. Paternally methylated DMR's however are frequently found within intergenic regions or imprinting control centres (such as *Peg13*) where multiple imprinted genes are thought to be regulated by a single DMR. These paternal DMRs are

established very early in the male germ line and persist for the reproductive life of the organism (Ferguson-Smith, 2011). According to Bourc'his and Bestor (2006), this difference in timing of imprint establishment is likely to underlie the increasing sexual dimorphism of other aspects of imprinted gene expression (Bourc'his & Bestor, 2006). Note that CpG islands are not specific to imprinted genes as they are widespread across the genome and can be found proximal to promoters and intragenic regions of biallelically expressed genes also. These CGI's may still be methylated, but the methylation levels might not play a role in allele-specific expression as they do in imprinted genes. Instead, the methylation of these regions may be involved in other regulatory processes, such as fine-tuning the overall expression level or responding to environmental cues.

Previous studies have indicated that allele specific expression due to genomic imprinting is not uniform across an organism and in fact can show different expression profiles in a tissue-specific manner (Baran et al., 2015). One such example of this tissue-specific imprinted expression is the *IGF2* gene which shows preferential maternal expression in the brain as opposed to the canonical paternal biased expression it shows in other tissues of the body (Baran et al., 2015). Furthermore, not only can imprinted genes be tissue-specific, but they have also been shown to be dynamic in regard to their status during developmental stages and adulthood. An investigation by Babak et al, (2015), examined an atlas of imprinting in 33 mouse and 45 human developmental stages and tissues and found that nearly all imprinted genes were established in early development and either retained their parent-of-origin expression in adults or lost it completely. While the evolutionary explanation for how genomic imprinting first arose is still an intensely debated topic, it is clear that this phenomenon is crucial to normal development and impacts hundreds of genes across a multitude of species. Characterisation of these imprinted genes is an ongoing

endeavour, with their functions and mechanisms, as well as their alteration in transcriptional activity in specific tissues, being largely unknown. In this chapter I aim to analyse one such imprinted cluster, the *Peg13-Trappc9* locus located on chromosome 15 in the mouse genome, known to have imprinted genes exhibiting preferential mono-allelic expression in murine brain.

The genes within the *Trappc9* imprinting cluster, save for *Peg13*, have been characterised via RNA-seq as being tissue-specifically imprinted, exhibiting preferential maternal allele expression in mouse brain. *Peg13* which constitutes a canonical imprinted gene, defined by gene expression regulated by allelic DNA methylation as opposed to non-canonical imprinting regulated by histone modification in a DNA methylation-independent manner (Hanna & Kelsey, 2021), is the exception to this, showing a mono-allelic paternal expression pattern that is not tissue specific (Andergassen et al., 2017; Babak et al., 2015; Perez et al., 2015). We attempted to validate these findings using SNP identification from two reciprocal hybrid mouse crosses C57BL/6 x Cast/EiJ and C57BL/6 x JF1 using a combination of pyrosequencing and Sanger sequencing. Furthermore, previous literature has shown some imprinted genes may alter their expression bias, specifically in postnatal and adult NSC's, to a more biallelic pattern (Ferrón et al., 2011; Montalbán-Loro et al., 2021), therefore we decided to include primary NSC's from newborn pups as a comparison to whole brain tissue with the addition of kidney tissue as a control to test for imprinted tissue specificity.

From these results I hypothesise that I will see mono-allelic or preferential expression from the maternal allele in *Trappc9, Ago2, Chrac1* and *Kcnk9* within both whole brain lysate and neural stem cells and biallelic expression of these genes in kidney tissue. However, for *Peg13* I expect to see mono-allelic expression from the paternal allele that is not tissue

specific and so all three tissue types analysed should display this preferential paternal expression. In addition to this, current literature dictates that the *Peg13* DMR, established in germline cells, is maternally methylated resulting in silencing of the maternal allele while the paternal CGI of *Peg13*, and the promoter CGIs of the other four genes remain unmethylated (Court et al., 2014). I aimed to validate these findings and identify the DNA methylation state at the CGI's located proximal to each gene's promoter region, and to potentially identify whether DNA methylation could be involved in their tissue-specific imprinting. From these results I expect that I will see hypermethylation at the promoter CGIs of all 5 genes within the cluster, with a differential methylation pattern at the *Peg13* CGI where the paternal allele will exhibit hypomethylation and the maternal hypermethylation, characteristic of this ICR. Furthermore, I expect the 2nd CGI of *Trappc9*, located proximal to its 2nd exon, will display hypermethylation to silence any alternate transcriptional start sites and prevent the expression of isoforms.

3.2 *Ago2*

3.2.1 *Ago2* SNP pyrosequencing of cDNA for allelic expression analysis

Pyrosequencing results analysing SNP variant rs232384843 (Table 2.2) indicated that transcripts expressed from *Ago2* were predominantly transcribed from the maternal allele in brain (70%–80%) but showed equal biallelic expression in kidney (Fig 3.1) which is in line with our hypothesised results. Unexpectedly however and in contrast to brain tissue, *Ago2* did not display imprinted expression in the NSC cultures but showed equal biallelic expression (Fig 3.1).

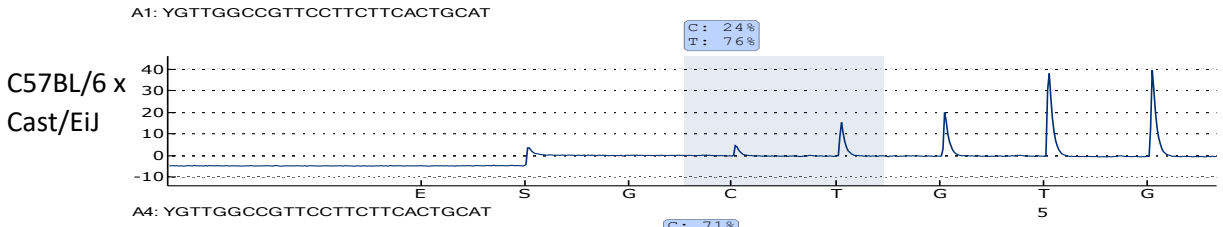
***Ago2* Exon 4-5 showing the SNP variant between C57BL/6 x Cast/EiJ**

A

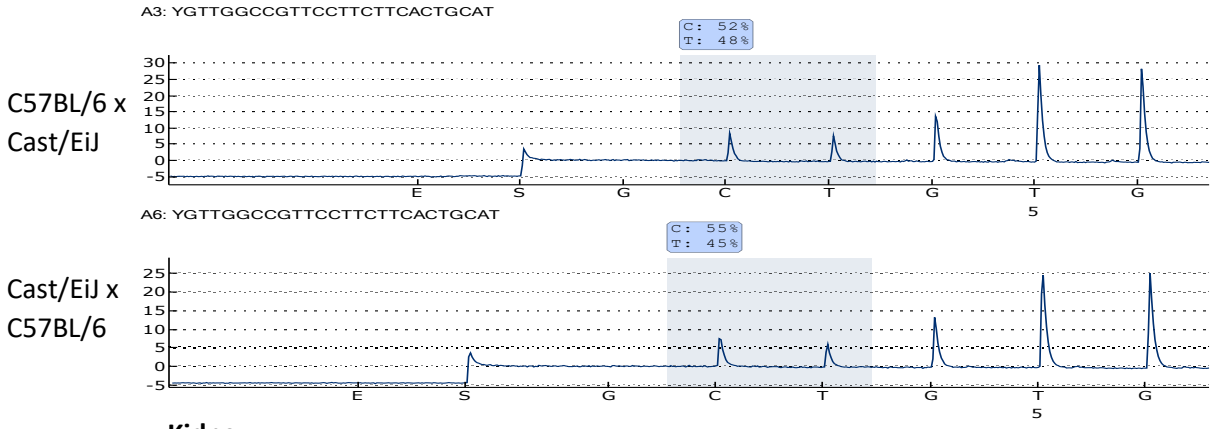
CCCTTCGAGACGATCCAGGCCCTGGACGTTGTCATGAGGCACTTACC ATCCATGAGGTACACCCCYGTTGGCCGTTCCTTCTTCACTGCATCTGA AGGCTGTTCCAACCCTCTGGGTGGGGGCAGAGA	Exon 4 Exon 5 rs232384843 A/G

B

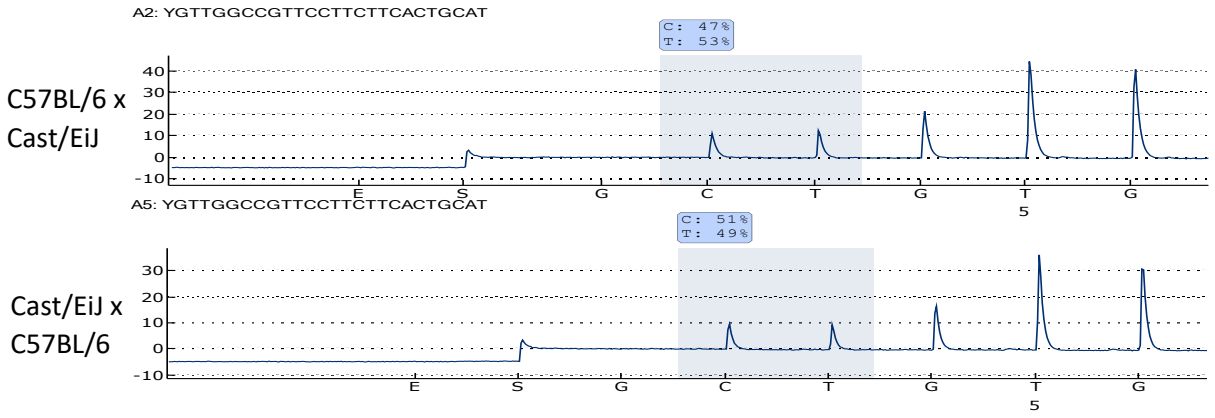
Brain



NSC



Kidney



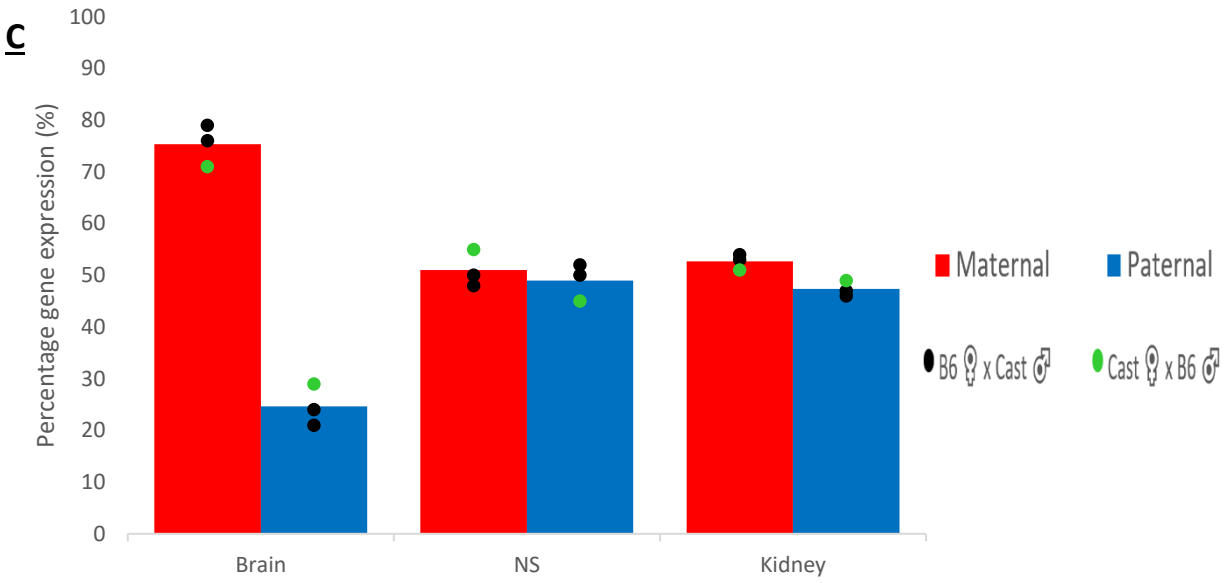
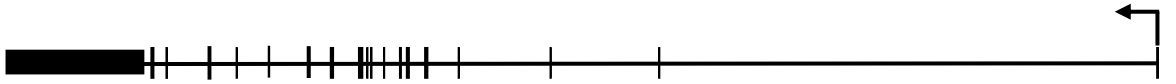


Figure 3.1: Allelic expression of *Ago2* in tissues and primary hippocampal neurospheres from C57BL/6 x Cast/EiJ hybrid mice. Parental allelic expression was quantified via SNP pyrosequencing of cDNA from tissues or cultured neurospheres (NS) obtained from newborn mice. **(A)** The SNP rs232384843 in exon 5 with flanking primers highlighted in yellow. **(B)** Representative pyrograms from Brain, NSC and Kidney samples are shown, indicating the SNP position in the sequence and the quantification of allelic expression. **(C)** Average expression (n = 3 mice) Points on the bar graphs represent individual pyrosequencing results from reciprocal crosses as shown in the legend.

3.2.2 *Ago2* pyrosequencing for DNA methylation

The *Ago2* promoter CGI located at position 73,184,092-73,185,471 on chromosome 15 in the GRCm38/mm10 sequence displayed no methylation in NSCs (Fig 3.2). While this is in line with current human brain data shown by Court et al., 2014 and confirms *Ago2*'s status as an actively transcribed gene, we do not possess methylation data for *Ago2* in brain tissue where maternal biased expression has been previously confirmed. Additionally, the use of kidney tissue as a peripheral control was inconclusive due to inaccuracy of the sequence read from the pyrosequencing experiment.



< *Ago2*-201 – ENSMUST00000044113.12 (Protein coding) – 13,800bp (19 exons) over 89.09kb

CpG island – 136 CpG

NSC

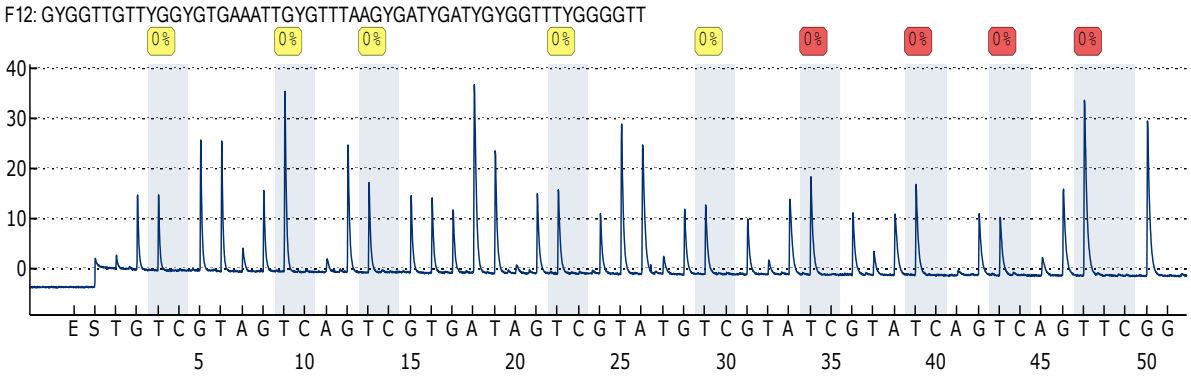


Figure 3.2: Methylation analysis of the *Ago2* promoter CGI, chr15: 73,184,092-73,185,471 using a combination of bisulphite conversion and pyrosequencing. A schematic of the *Ago2* locus showing exon/intron structure and the location of the CGI is shown. Extracted DNA underwent bisulphite conversion, and the newly converted CGI, from the forward strand, was used as a template for PCR amplification. This amplified product was then pyrosequenced. Several CG dinucleotides were assessed and the percentage as to how much of the template was converted at that region was indicated, with 0% indicating no remaining CG dinucleotides/ full conversion and 100% indicating no conversion and maintenance of CG dinucleotides due to the presence of a methyl group. Methylation frequency indicates 0% methylation in the *Ago2* CGI for NSC. Data are for neural stem cells (hippocampal neurospheres) obtained from C57BL/6 x Cast/EiJ newborn F1 hybrids.

3.2.3 *Ago2* allelic methylation analysis by Sanger sequencing

In parallel with the pyrosequencing, I also performed an independent analysis using plasmid-cloned PCR products and Sanger sequencing for each gene. The advantage of this is it allowed me to distinguish the maternal and paternal alleles using strain specific SNPs between the hybrid cross. This would allow us to see whether there was a difference in methylation status between the maternal and paternal alleles as opposed to the pyrosequencing reaction which gives an average methylation percentage but no allele

distinction. The bisulphite converted, PCR amplified, CGI sample was submitted to Sanger sequencing with the results analysed using the QUMA software. Both the maternal and paternal alleles of the *Ago2* CGI showed little to no methylation (Fig 3.3) confirming the results of our pyrosequencing data and supporting my hypothesis.

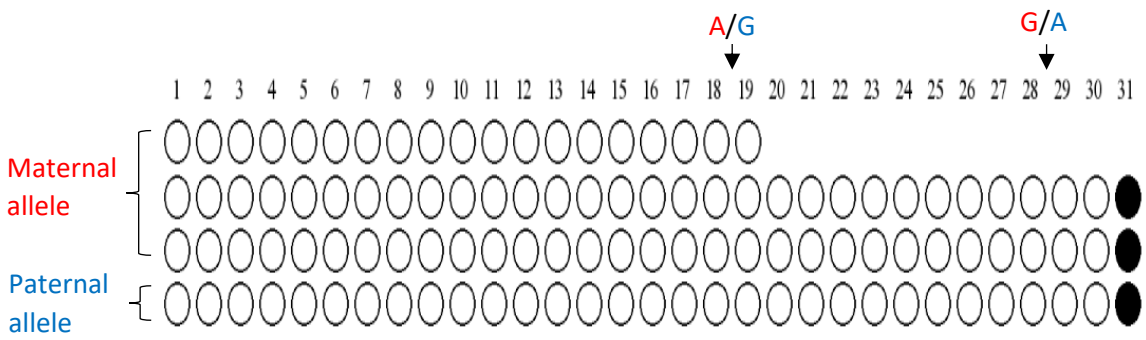


Figure 3.3: Analysis of genomic DNA methylation at the *Ago2* CGI through bisulphite-treatment followed by Sanger sequencing of cloned samples. Sequencing data were analysed using QUMA. Data are for neural stem cells (hippocampal neurospheres) obtained from C57BL/6 x Cast/EiJ newborn F1 hybrids. QUMA output summary of Sanger sequences compared to the original non bisulphite converted reference sequence. The SNPs: rs257455001 and rs227718844 were identified using Sanger sequencing and used to distinguish the maternal and paternal allele via their strain specificity. From the CGI 31 CpG dinucleotides were analysed and indicated that the promoter CGI at *Ago2* is unmethylated on both alleles in neural stem cells. Black circle = methylated CpG; white circle = unmethylated CpG).

Overall, the results of the *Ago2* expression and methylation analysis were all within expected parameters except for the anomalous result of biallelic expression within NSCs. The potential reasons for this deviation will be touched upon within the discussion of this chapter.

3.3 *Trappc9*

3.3.1 *Trappc9* SNP pyrosequencing of cDNA for allelic expression analysis

Pyrosequencing results analysing SNP variant rs31440851 (Table 2.2) indicated that *Trappc9* expression occurred predominantly from the maternal allele in brain (70%–80%) but displayed a weak maternal bias, more closely resembling equal biallelic expression, in kidney (Fig 3.4). This mirrors the expression pattern of the *Ago2* gene for both tissues which again supports our proposed hypothesis that imprinted expression of these genes is tissue specific. *Trappc9* did not however appear to be fully imprinted in the NSC cultures but did display a slight bias (< 70%) towards maternal expression (Fig 3.4). This may be reminiscent of the *Dlk1* gene which transitions from being paternally expressed to biallelic expression in NSCs. (Ferrón et al., 2011; Montalbán-Loro et al., 2021). However, despite not being significantly skewed towards a maternal bias, *Trappc9* maternal allele expression in NSC and kidney does appear be slightly stronger than the maternal expression of *Ago2* within the same tissues/cells.

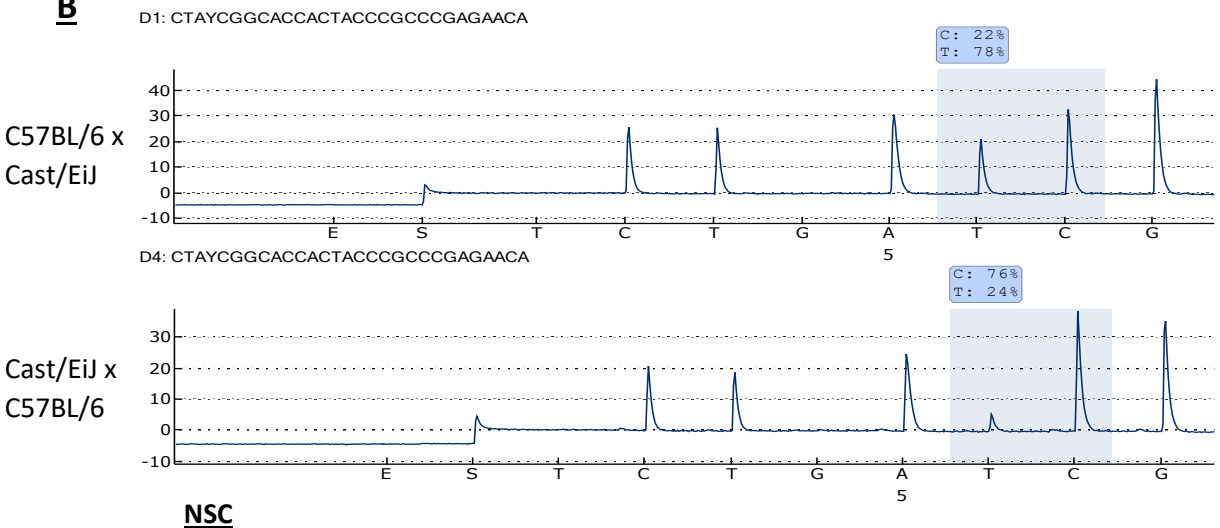
A

***Trappc9* Exon 2 showing the SNP variant between C57BL/6 x Cast/EiJ**

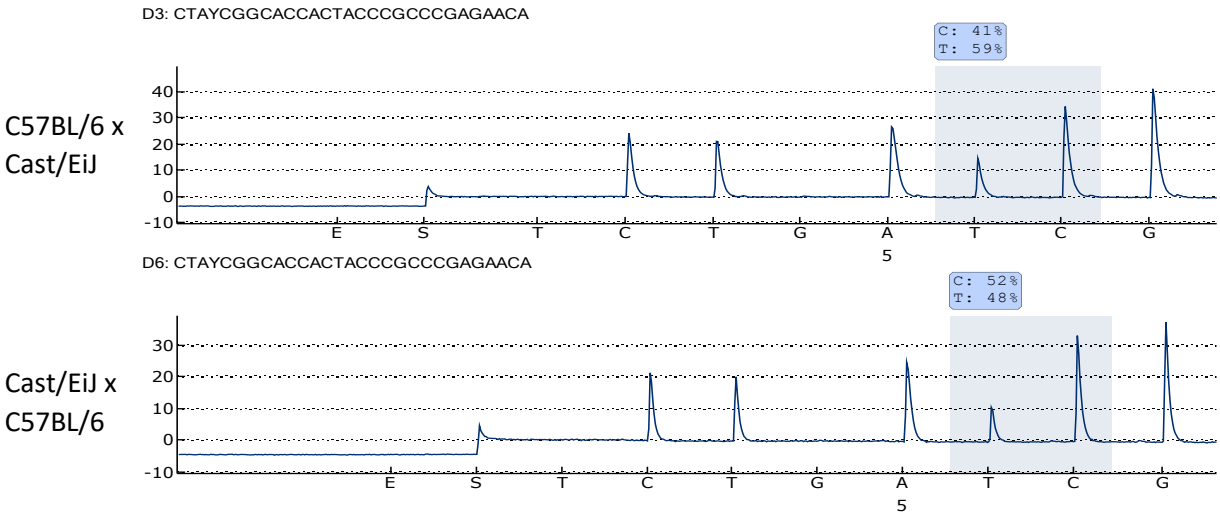
CAGCTCAGTGTGCGCGACACGCAGCGTGCCCTCTTCATCCGCTAYCG GCACCACTACCCGCCCGAGAACAACGAGTGGGGCGACTTCCAGACG CACCGCAAGGTGGTAGG	Exon 2 rs31440851 A/G

B

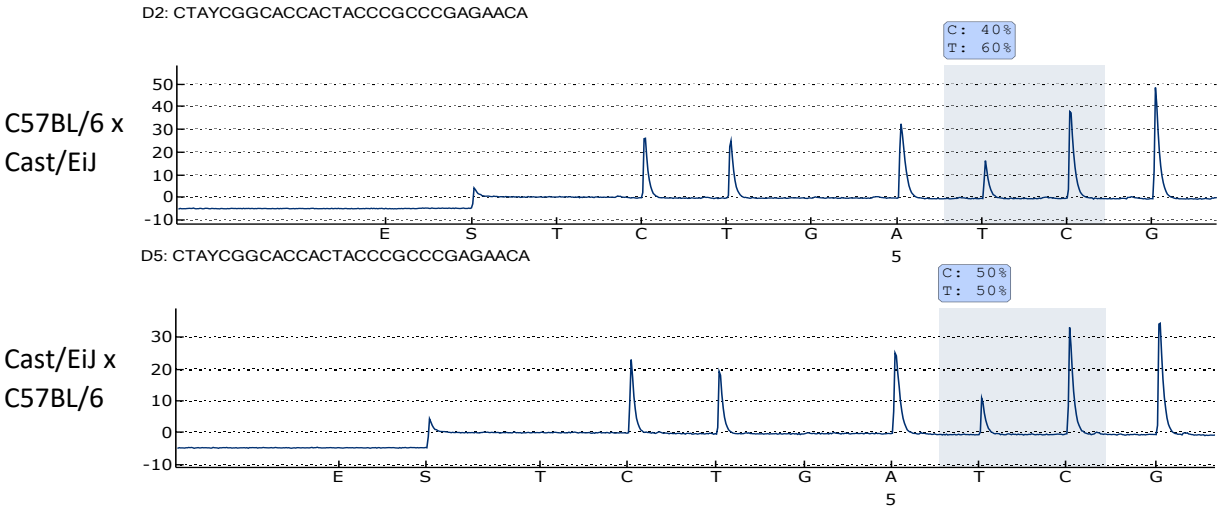
Brain



NSC



Kidney



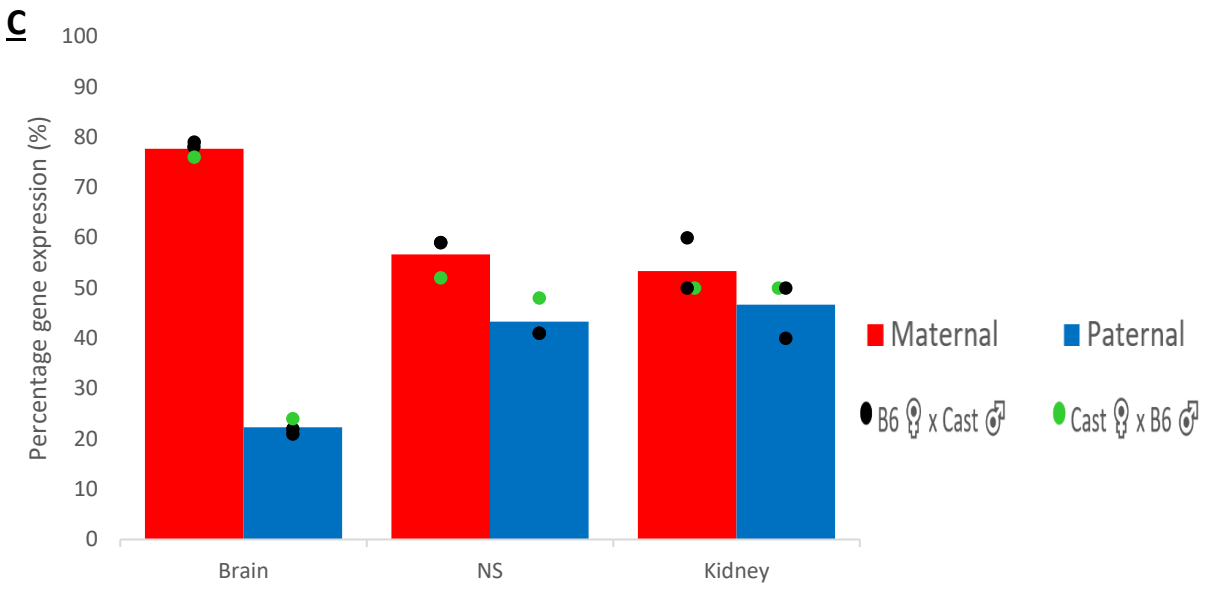
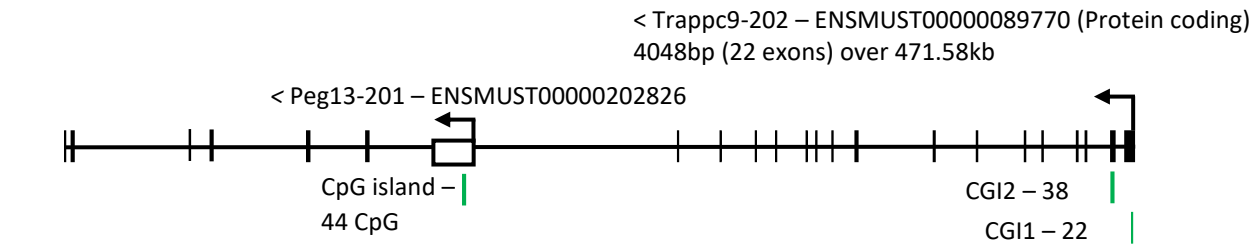


Figure 3.4: Allelic expression of *Trappc9* in brain, primary hippocampal neurospheres and kidney of C57BL/6 x Cast/EiJ hybrid mice. Parental allelic expression was quantified via SNP pyrosequencing of cDNA from tissues or cultured neurospheres (NS) obtained from newborn mice. **(A)** The SNP rs31440851 in exon 2 with flanking primers highlighted in yellow. **(B)** Representative pyrograms from Brain, NSC and kidney samples are shown, indicating the SNP position in the sequence and the quantification of allelic expression. **(C)** Average expression (n = 2–3 mice) Points on the bar graphs represent individual pyrosequencing results from reciprocal crosses as shown in the legend.

3.3.2 *Trappc9* pyrosequencing for DNA methylation

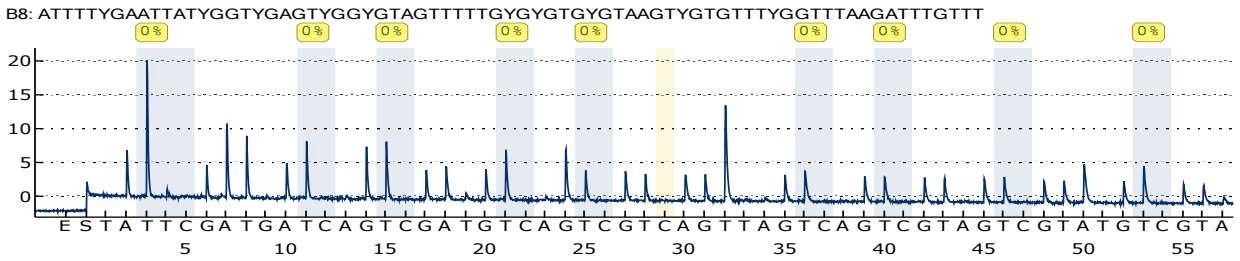
Unlike the other genes within the imprinted cluster, the *Trappc9* gene contains two CGI's, one located near the promoter/exon 1 boundary at position 73,060,920-73,061,124 labelled as CGI1 and the other located proximal to exon 2 at position 73,058,049-73,058,500 labelled CGI2 in the GRCm38/mm10 sequence (fig 3.5). Previous data has claimed that the *Trappc9* gene is associated with alternate transcriptional start sites, one being the promoter/ exon 1 boundary and the other downstream of the exon 2/ CGI2 boundary in brain tissue (Gregg et al., 2010; Hsu et al., 2018). Although we were unable to confirm this in our own experiments (Claxton et al., 2022), we aimed to identify the methylation patterns of CGIs overlapping these sequences to determine whether they showed marks of gene activation

or silencing. Previous methylation data obtained from Pulix, (2017) indicated that in adult C57BL/6 mice, CGI1 is unmethylated in both brain and peripheral tissues while CGI2 showed hypermethylation for both brain and peripheral tissues. Data obtained from the pyrosequencing reaction of *Trappc9* CGI1 was inconclusive in NSCs but showed no methylation in kidney tissue, while CGI2 showed hypermethylation (80-100%) in both NSCs and kidney, both of which are in concordance with Pulix's data (Fig 3.5).



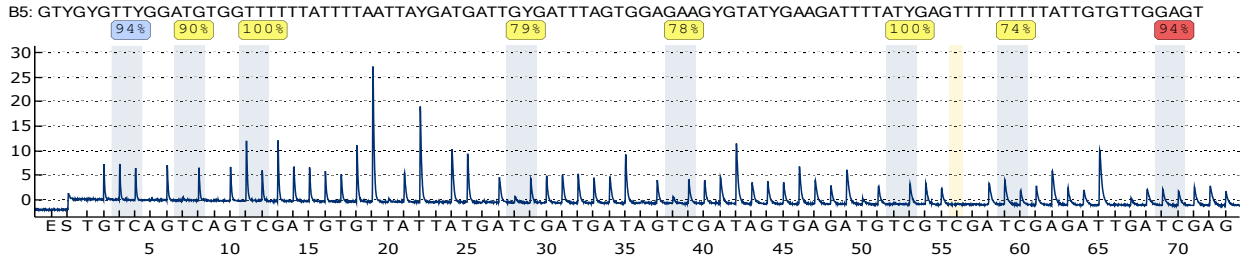
A ***Trappc9* CGI1**

Kidney



B ***Trappc9* CGI2**

NSC



Kidney

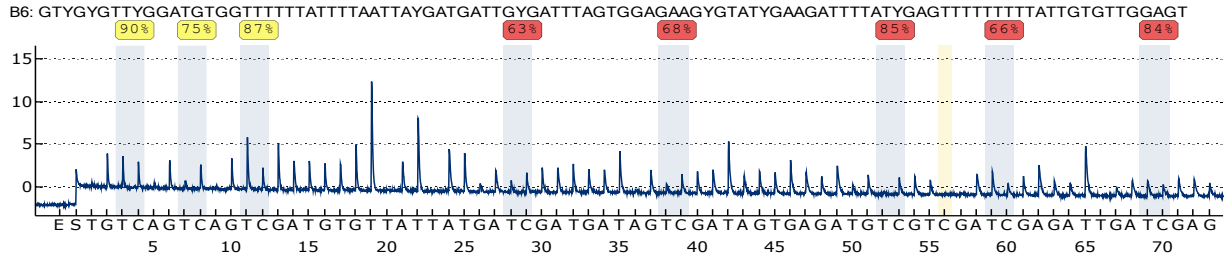


Figure 3.5: Methylation analysis of the *Trappc9* CGIs: CGI1 chr15: chr15: 73,060,920-73,061,124 located at the promoter/exon 1 boundary and CGI2 chr15: 73,058,049-73,058,500 located at exon 2 in the GRCm38/mm10 mouse genome. A schematic of the *Trappc9* locus, showing exon/intron structure and the location of CGI1 and CGI2. Extracted DNA underwent bisulphite conversion, and the newly converted CGI (from the forward strand for CGI1 and the reverse strand for CGI2) was used as a template for PCR amplification. This amplified product was then pyrosequenced. Each CG dinucleotide region was assessed and given a percentage as to how much of the template was converted at that region was indicated, with 0% indicating no remaining CG dinucleotides/ full conversion and 100% indicating no conversion and maintenance of CG dinucleotides due to the presence of a methyl group. **(A)** Methylation frequency indicates 0% methylation for *Trappc9* CGI1 in Kidney tissue. **(B)** Methylation frequency indicates ~80-100% methylation for *Trappc9* CGI2 for both NSC and Kidney tissues. Data are for neural stem cells (hippocampal neurospheres) and kidney tissue obtained from C57BL/6 x Cast/EiJ newborn F1 hybrids.

3.3.3 *Trappc9* allelic methylation analysis by Sanger sequencing

The SNP variants rs31441779 for CGI1 and rs31440849 for CGI2 (Table 2.2) were used to differentiate the maternal and paternal alleles in my Sanger sequencing analysis. Both the maternal and paternal allele showed little to no methylation at CGI1, which supports the pyrosequencing and previous data from Pulix and is in line with my expected hypothesis. CGI2 on the other hand displayed hypermethylation in both alleles further supporting the previous data and the pyrosequencing results (Fig 3.6).

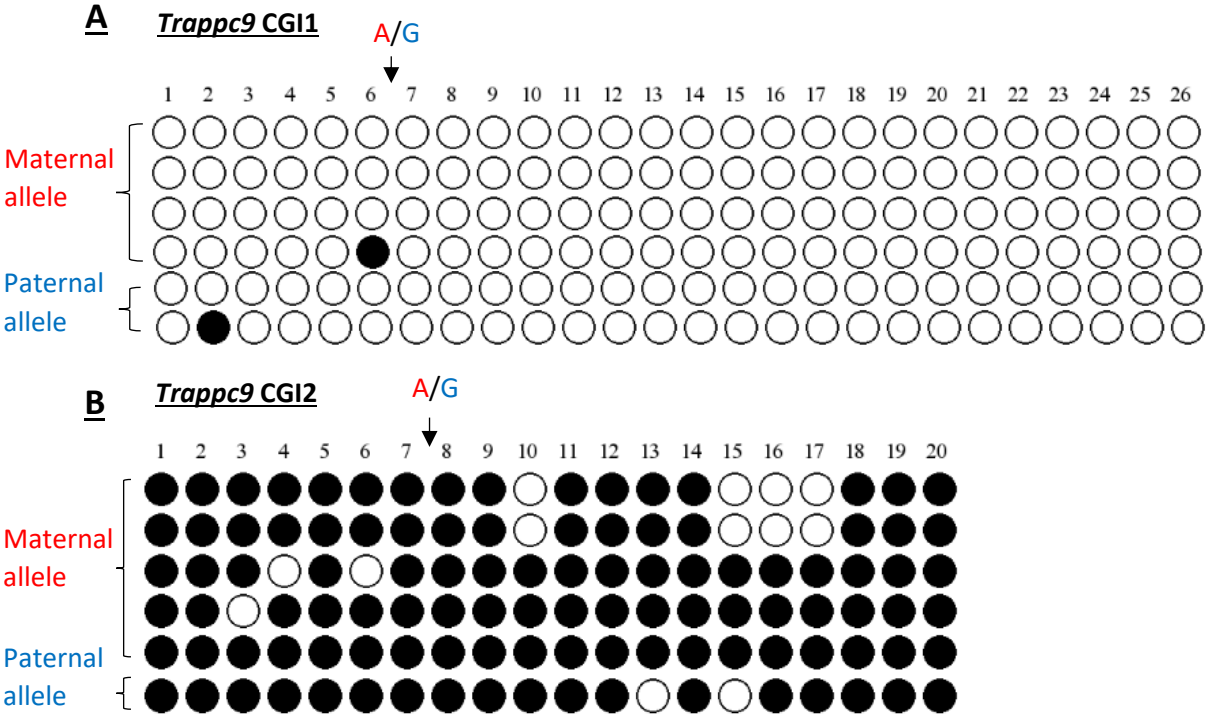


Figure 3.6: Analysis of genomic DNA methylation at *Trappc9* CGI1 and CGI2 through bisulphite-treatment followed by Sanger sequencing of cloned samples. Sequencing data were analysed using QUMA. Data are for neural stem cells (hippocampal neurospheres) obtained from C57BL/6 x Cast/EiJ newborn F1 hybrids. The SNPS rs31441779 (CGI1) and rs31440849 (CGI2) were identified using Sanger sequencing and used to distinguish the maternal and paternal allele via their strain specificity. **(A)** QUMA output summary of CGI1 targeted Sanger sequences compared to the original non bisulphite converted reference sequence. A total of 26 CpG dinucleotides were analysed highlighting that the promoter CGI1 at *Trappc9* indicated little to no methylation. **(B)** QUMA output summary of CGI2. A total of 20 CpG dinucleotides were analysed highlighting that CGI2 located proximal to the 2nd exon of *Trappc9* indicated a high degree of methylation in NSC. Black circles = methylated CpG; white circle = unmethylated CpG).

The results of the *Trappc9* expression and methylation analysis mirrored that of *Ago2*, with tissue specific imprinting favouring maternal allele expression being observed. In addition to this, the same unexpected result found in *Ago2* of biallelic expression within NSCs which were expected to show a maternal bias was evident for *Trappc9*. Furthermore, the methylation analysis of both CGIs was in line with our expectations indicating that the deviation in expression was likely not due to anomalies in the methylation state of the CGIs.

3.4 *Peg13*

3.4.1 *Peg13* SNP pyrosequencing of cDNA for allelic expression analysis

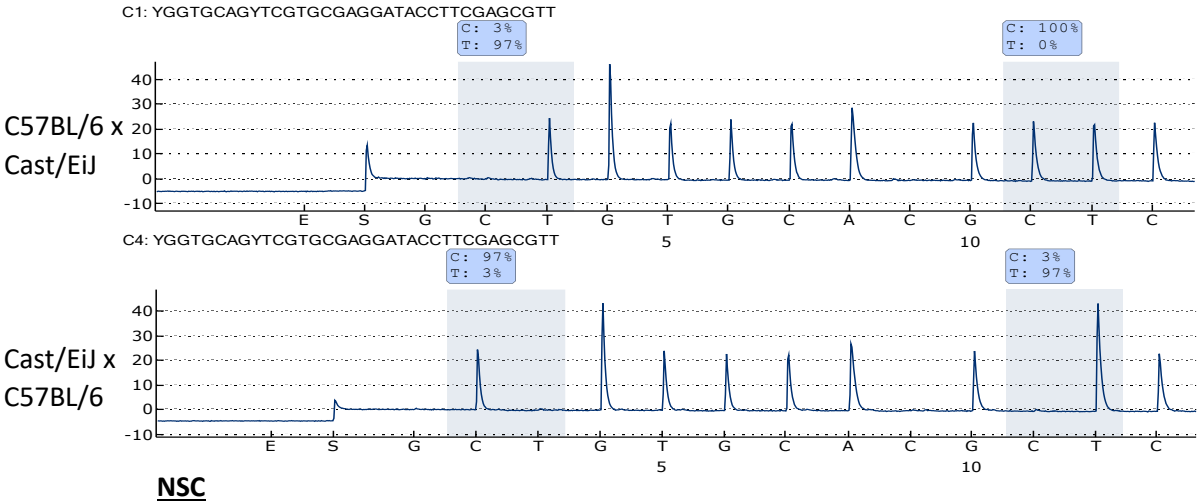
Unlike other genes within the *Trappc9-Peg13* cluster, we expected that *Peg13* would show no tissue-specific imprinting and would have a strong expression bias towards the paternal allele. Using the SNP variants rs238259968 and rs31423566 (Table 2.2) from our hybrid cross, the pyrosequencing results indicated that *Peg13* expression occurred predominantly from the expected paternal allele in brain, kidney and NSC samples showing an imprinting bias of ~80-90%, with no tissue specificity, confirming our hypothesis (Fig 3.7).

A ***Peg13* Exon 1 showing the SNP variant between C57BL/6 x Cast/EiJ**

AATGGGCAAAAGGAGGCACAGAAAAAGCCCAGATATCTGTGTYG GTGCAGYTCGTGCGAGGATACCTTCGAGCGTTGAGATACCTTCGA GCACAGAACCCTGCAGCTTGCG	Exon 1 rs238259968- G/A rS31423566- A/G

B

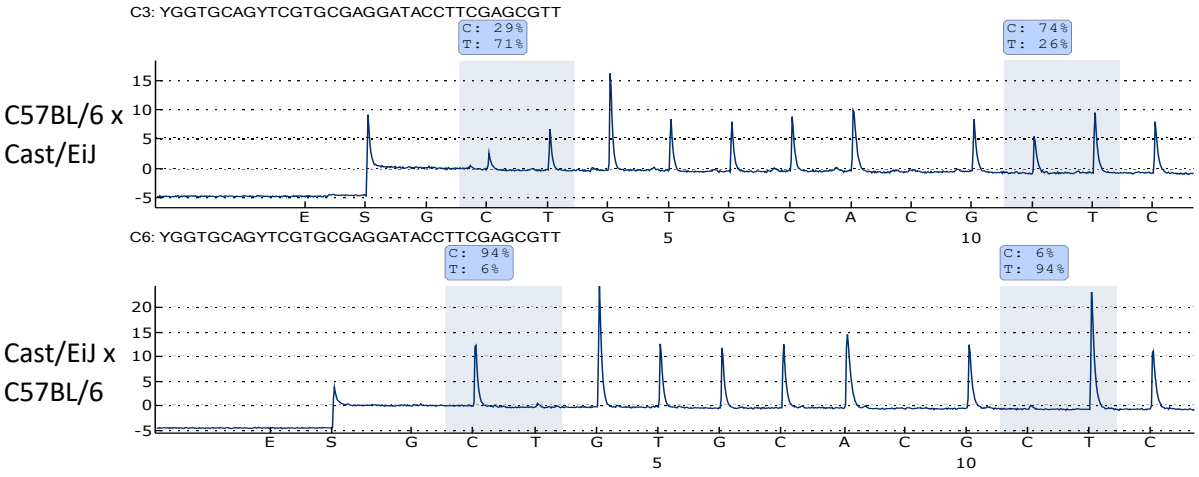
Brain



C57BL/6 x Cast/EiJ

Cast/EiJ x C57BL/6

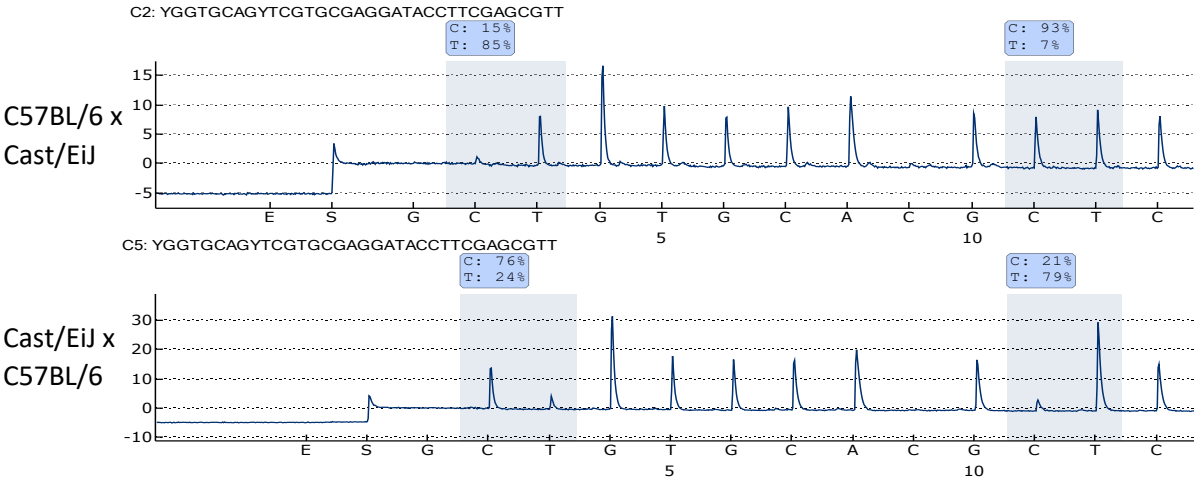
NSC



C57BL/6 x Cast/EiJ

Cast/EiJ x C57BL/6

Kidney



C57BL/6 x Cast/EiJ

Cast/EiJ x C57BL/6

C

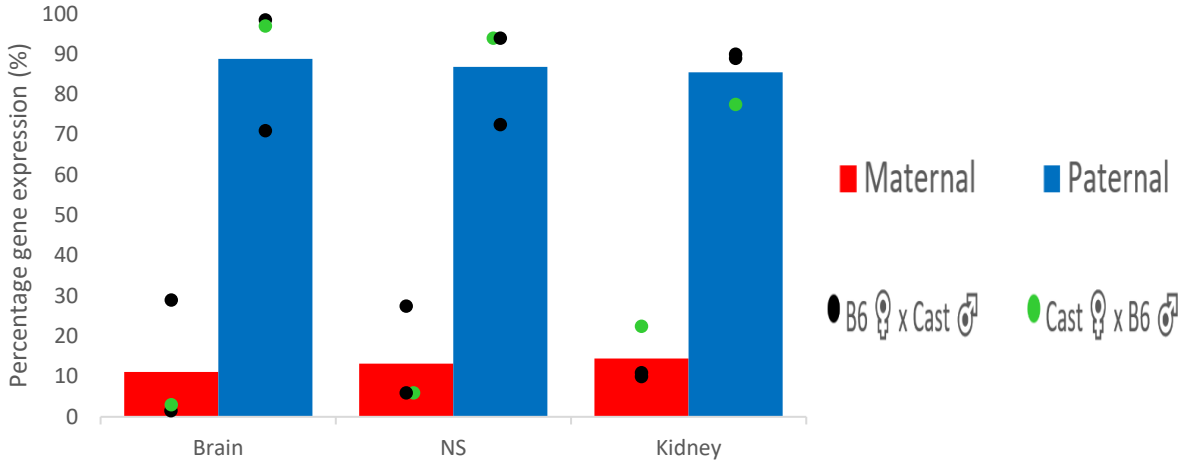


Figure 3.7: Allelic expression of *Peg13* in brain, primary hippocampal neurospheres and kidney of C57BL/6 x Cast/EiJ hybrid mice. Parental allelic expression was quantified via SNP pyrosequencing of cDNA from tissues or cultured neurospheres (NS) obtained from newborn mice. **(A)** The SNPs rs238259968 and rs31423566 in exon 1 with flanking primers highlighted in yellow. **(B)** Representative pyrograms from Brain, NSC and kidney samples are shown, indicating the SNP position in the sequence and the quantification of allelic expression. **(C)** Average expression (n = 3 mice) Points on the bar graphs represent the average of the two SNP values in the same sequence from the pyrosequencing results showing both reciprocal crosses as indicated in the legend.

3.4.2 *Peg13* pyrosequencing for DNA methylation

The imprinting regulation of this cluster of genes has been proposed to be controlled by a maternally methylated germline DMR encompassing the CGI associated with the promoter of *Peg13* (Court et al., 2014). Pyrosequencing analysis of the *Peg13* promoter CGI, located at position 72,809,538-72,810,123 in the GRCm38/mm10 sequence, indicated 35-40% methylation frequency in NSC and 40-50% in kidney samples (Fig 3.8). This percentage is in line with a differential methylation pattern, as one allele being completely methylated while the other is hypomethylated would produce a methylation frequency of 50%. Our result is slightly lower than 50% methylation which may be due to a low level of methylation on both alleles, as pyrosequencing simply indicates a percentage of total methylation frequency, but

it is more likely due to the maternal allele being hypermethylated and the paternal allele being hypomethylated, based upon previous literature surrounding this ICR.

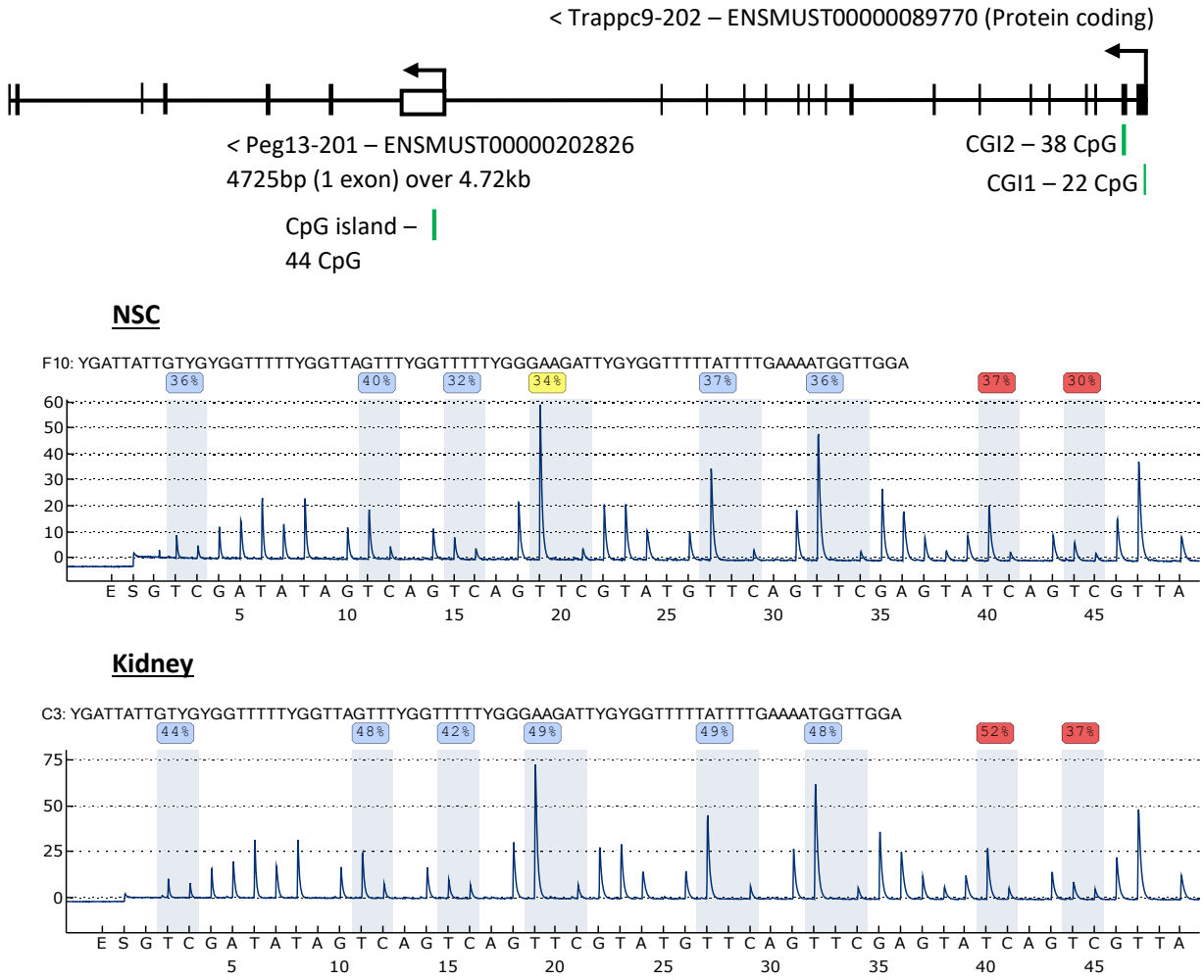


Figure 3.8: Methylation analysis of the *Peg13* promoter CGI located at position chr15: 72,809,538-72,810,123 GRCm38/mm10 mouse genome using a combination of bisulphite conversion and pyrosequencing. A schematic of the *Peg13* locus showing exon/intron structure and the location of the CGI. Extracted DNA underwent bisulphite conversion, and the newly converted CGI, from the reverse strand, was used as a template for PCR amplification. This amplified product was then pyrosequenced. Several CG dinucleotides were assessed and the percentage as to how much of the template was converted at that region was indicated, with 0% indicating no remaining CG dinucleotides/ full conversion and 100% indicating no conversion and maintenance of CG dinucleotides due to the presence of a methyl group. Methylation frequency indicates ~35-50% methylation in the *Peg13* CGI for both NSC and Kidney tissues. Data are for neural stem cells (hippocampal neurospheres) and kidney tissue obtained from C57BL/6 x Cast/EiJ newborn F1 hybrids.

3.4.3 *Peg13* allelic methylation analysis by Sanger sequencing

The SNP variants rs242845136, rs31423568, rs31423567, rs238259968 and rs31423566 which show as G/A, A/G, G/A, G/A and A/G respectively (Table 2.2) were used to distinguish between the maternal and paternal allele for *Peg13*. From the Sanger sequencing and subsequent QUMA analysis, only maternal C57BL/6 strain SNPs were identified but of these maternal samples all displayed hypermethylation (Fig 3.9) which matched our proposed expectations. While the lack of paternal allele samples was unfortunate, based on the methylation frequnecies displayed in the pyrosequencing data, coupled with the evident hypermethylation of the maternal allele, it is likely the paternal allele would display hypomethylation as we initially hypothesised.

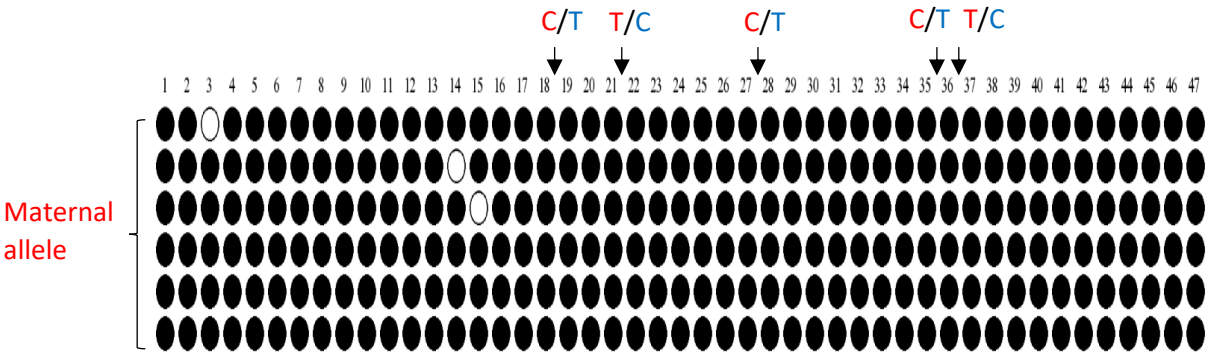


Figure 3.9: Analysis of genomic DNA methylation at the *Peg13* CGI through bisulphite-treatment followed by Sanger sequencing of cloned samples. Sequencing data were analysed using QUMA. Data are for neural stem cells (hippocampal neurospheres) obtained from C57BL/6 x Cast/EiJ newborn F1 hybrids. QUMA output summary of Sanger sequences compared to the original non bisulphite converted reference sequence. SNPs: rs242845136, rs31423568, rs31423567, rs238259968 and rs31423566 were identified using Sanger sequencing and used to distinguish the maternal and paternal allele via their strain specificity. From the 47 CpG dinucleotides analysed the promoter CGI at *Peg13* is hypermethylated on the maternal allele in NSC, paternal allele unidentified. Black circle = methylated CpG; white circle = unmethylated CpG).

The results of the *Peg13* expression and methylation analysis were completely in line with our expected hypothesis. All tissues displayed a strong, near mono-allelic paternal expression bias, indicating the gene is imprinted and displays no tissue specificity. Additionally, while no paternal alleles were obtained via the Sanger sequencing/ QUMA analysis, the maternal hypermethylation of the promoter proximal CGI identified through Sanger sequencing is supportive of previous literature highlighting a DMR at this region.

3.5 *Kcnk9*

3.5.1 *Kcnk9* SNP pyrosequencing of cDNA for allelic expression analysis

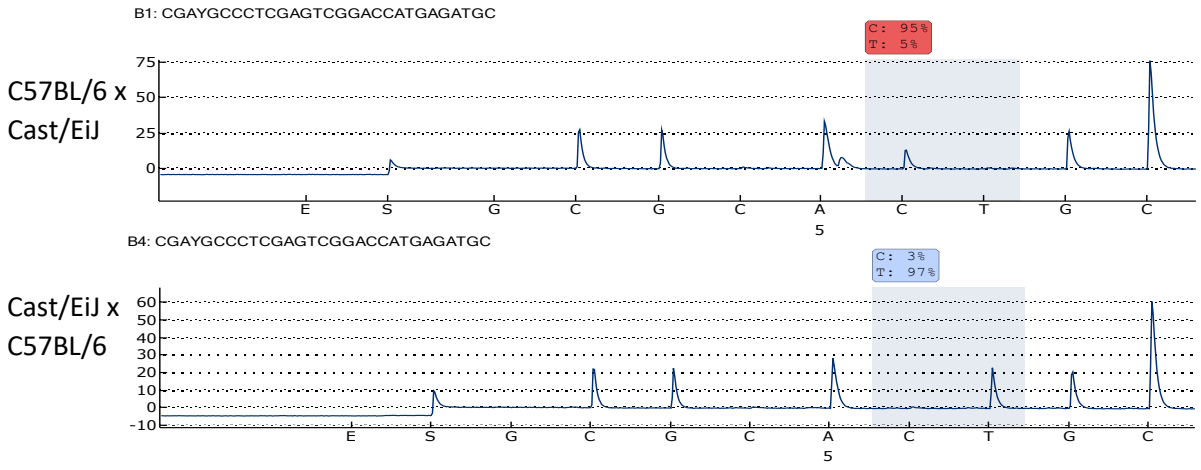
Pyrosequencing results analysing the SNP variant rs225149059 (Table 2.2) indicated that *Kcnk9* expression occurred almost exclusively from the maternal allele in brain and kidney samples (90–100%). However, in the NSC sample, allelic expression bias was inconclusive

due to the presence of a strain bias with the Cast/EiJ allele being the more highly expressed allele in both crosses (Fig 3.10). The maternal allele biased expression of *Kcnk9* in brain and kidney tissues follows the hypothesis that all genes within the *Trappc9-Peg13* save for *Peg13* display a maternal expression bias. However, it does indicate that *Kcnk9* does not display imprinting tissue specificity the way *Ago2* and *Trappc9* do.

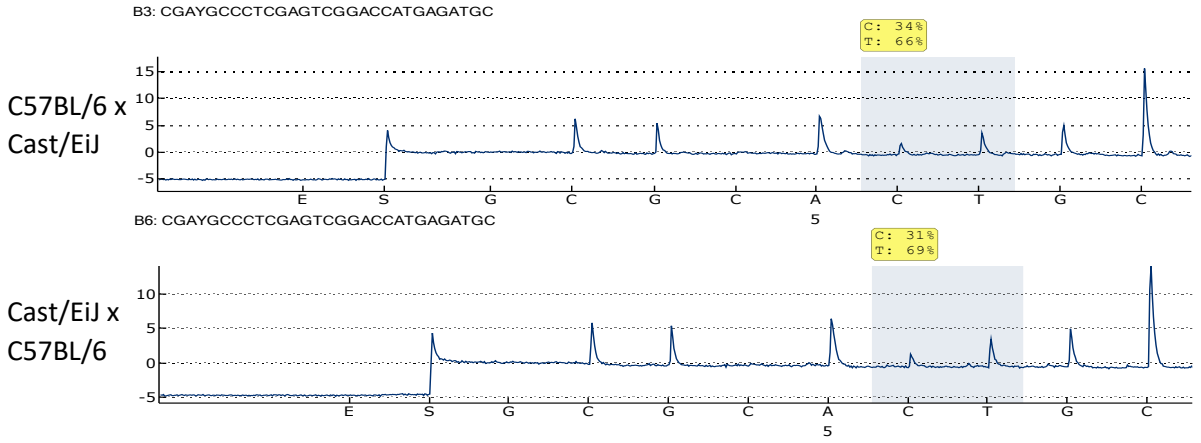
A ***Kcnk9* Exon 1 showing the SNP variant between C57BL/6 x Cast/EiJ**

GCAGAACGTGCGTACCCTGTCCTTGATCGCCTGTACCTTCACCTAC CTGCTGGTCGGTGCCGCGGTGTTCGAYGCCCTCGAGTCGGACCAT GAGATGCGCGAGGAGGAGAA	Exon 1 rs225149059 - G/A

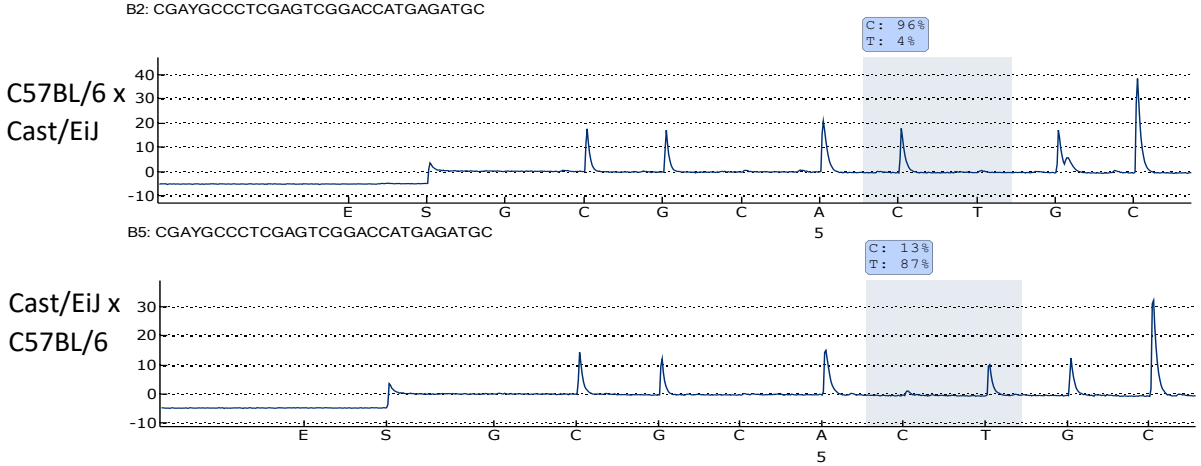
B **Brain**



NSC



Kidney



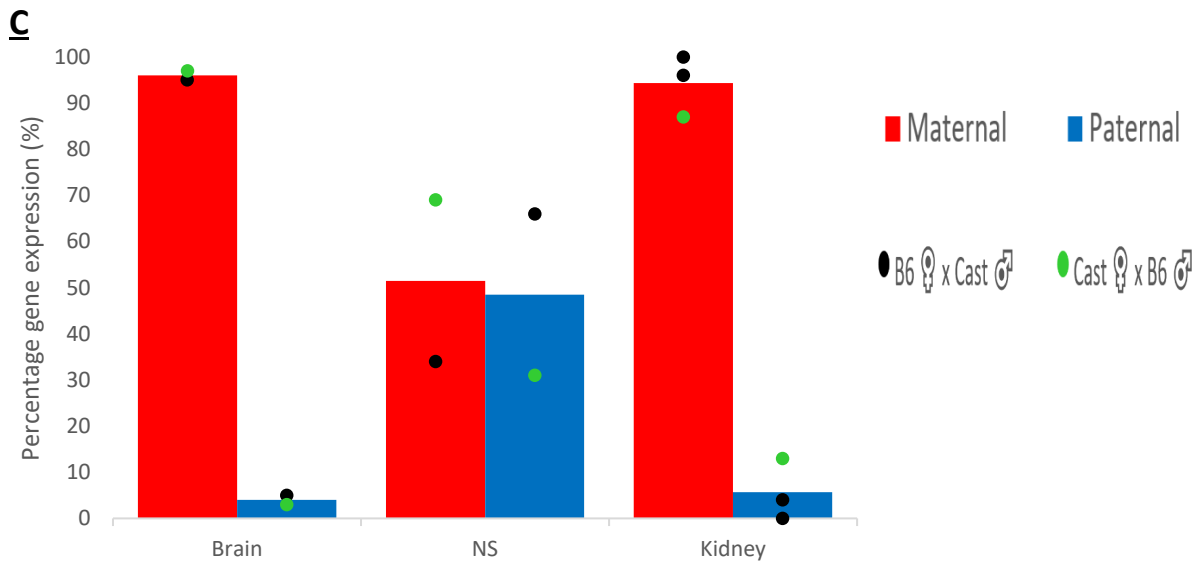


Figure 3.10: Allelic expression of *Kcnk9* in tissues and primary hippocampal neurospheres from C57BL/6 x Cast/EiJ hybrid mice. Parental allelic expression was quantified via SNP pyrosequencing of cDNA from tissues or cultured neurospheres (NS) obtained from newborn mice. **(A)** The SNP rs225149059 in exon 1 with flanking primers highlighted in yellow. **(B)** Representative pyrograms from Brain, NSC and Kidney samples are shown, indicating the SNP position in the sequence and the quantification of allelic expression. **(C)** Average expression (n = 2–3 mice) Points on the bar graphs represent individual pyrosequencing results from reciprocal crosses as shown in the legend

3.5.2 *Kcnk9* pyrosequencing for DNA methylation

For the *Kcnk9* CGI, both NSC and kidney tissue were analysed but due to sequencing errors in the pyrosequencing reaction the kidney data was inconclusive. The CGI at the promoter region of *Kcnk9* located at position chr15: 72,545,926-72,547,811 in the GRCm38/mm10 sequence showed little to no methylation (0-10%) in the NSC samples analysed (Fig 3.11).

***Kcnk9* CGI**

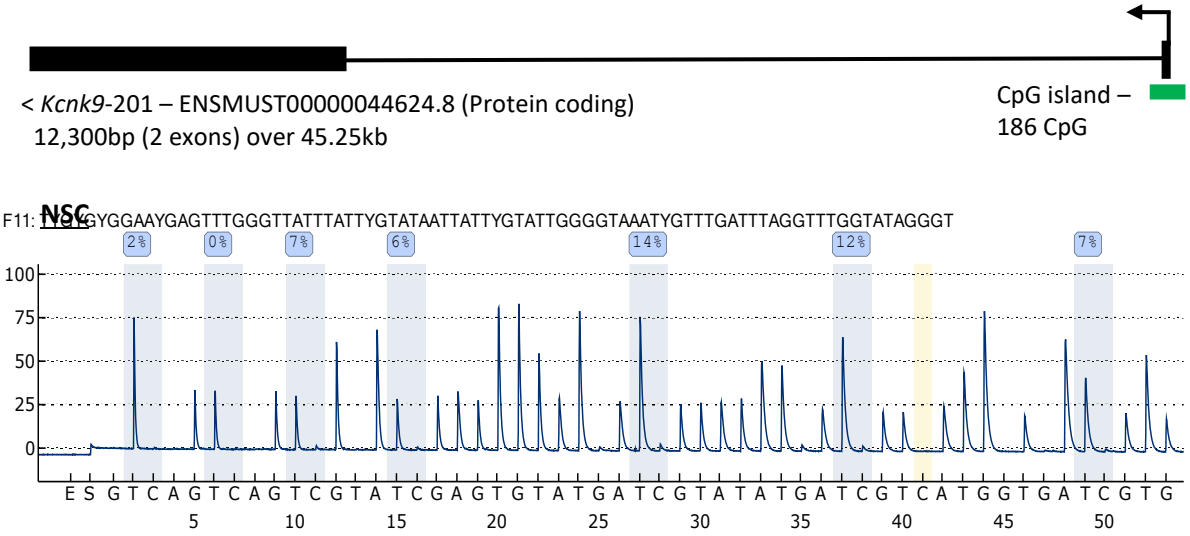


Figure 3.11: Methylation analysis of the *Kcnk9* CGI located at position 72,545,926-72,547,811 in the GRCm38/mm10 mouse genome. A schematic of the *Kcnk9* locus showing exon/intron structure and the location of the CGI. Extracted genomic DNA underwent bisulphite conversion, and the newly converted CGI, from the reverse strand, was used as a template for PCR amplification. This amplified product was then pyrosequenced. Several CG dinucleotides were assessed and the percentage as to how much of the template was converted at that region was indicated, with 0% indicating no remaining CG dinucleotides/ full conversion and 100% indicating no conversion and maintenance of CG dinucleotides due to the presence of a methyl group. Methylation frequency indicates ~0-15% methylation in the *Kcnk9* CGI for NSC. Data are for neural stem cells (hippocampal neurospheres) obtained from C57BL/6 x Cast/EiJ newborn F1 hybrids.

3.5.3 *Kcnk9* allelic methylation analysis by Sanger sequencing

The SNP variants rs583874465, rs219245628 and rs259286628 (Table 2.2) were used to distinguish between the maternal and paternal allele were in the C57BL/6 x Cast/EiJ hybrid mouse cross. From the Sanger sequencing and QUMA analysis, both maternal and paternal allele SNPs were identified with almost every CpG dinucleotide within the CGI being converted indicating a lack of methylation at KCNK9 CGI in NSCs (Fig 3.12)

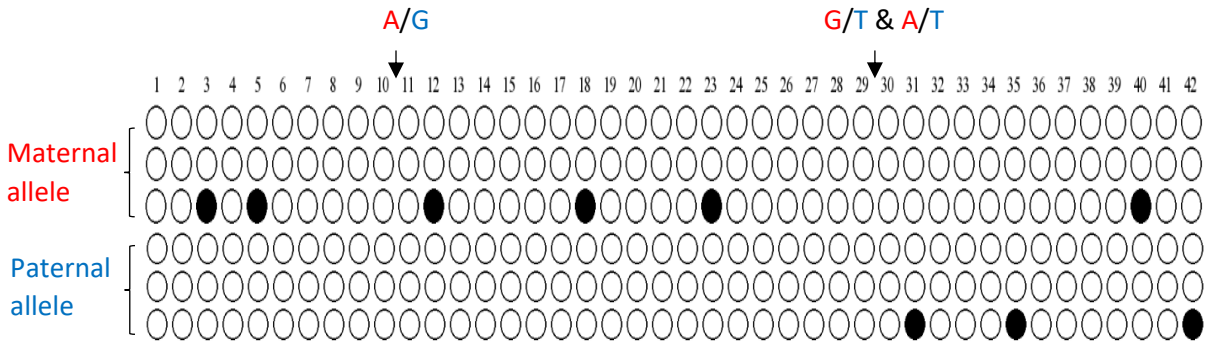


Figure 3.12: Analysis of genomic DNA methylation at the *Kcnk9* CGI through bisulphite-treatment followed by Sanger sequencing of cloned samples. Sequencing data were analysed using QUMA. Data are for neural stem cells (hippocampal neurospheres) obtained from C57BL/6 x Cast/EiJ newborn F1 hybrids. QUMA output summary of Sanger sequences compared to the original non bisulphite converted reference sequence. SNPs: rs583874465, rs219245628 and rs259286628 were identified using Sanger sequencing and used to distinguish the maternal and paternal allele via their strain specificity. From the 53 CpG dinucleotides analysed the promoter CGI at *Kcnk9* shows little to no methylation on both the paternal and maternal allele in NSC. Black circle = methylated CpG; white circle = unmethylated CpG).

The conclusion of the expression data analysed was in line with that of the expected hypothesis with *Kcnk9* displaying preferential maternal allele expression with no tissue specificity of imprinted transcription. Unfortunately, due to the strain bias where the Cast/EiJ allele was expressed more highly in both crosses, the evidence of an imprinting bias in NSCs for *Kcnk9* is inconclusive and would require further experiments before a definitive conclusion could be made. Furthermore, I observed no difference in the methylation state between the maternal and paternal alleles of *Kcnk9*, both of which are hypomethylated, in NSCs which supports previous literature.

3.6 *Chrac1*

3.6.1 *Chrac1* SNP pyrosequencing of cDNA for allelic expression analysis

The SNP variant rs248258787 (Table 2.2) in the C57BL/6 x JF1 hybrid cross was used to identify any allelic expression bias found in the *Chrac1* gene. For analysis of allele expression identification, we used a C57BL/6 x JF1 hybrid cross as opposed to the C57BL/6 x Cast/EiJ cross used for the other genes in the cluster due to the absence of an appropriate SNP between the standard hybrid cross used in our previous experiments. Pyrosequencing results of newborn brain and kidney samples indicated that *Chrac1* expression was found to have a small bias (<70%) towards the maternal allele in newborn brain and exhibited biallelic expression in kidney tissue (Fig 3.13).

***Chrac1* Exon 2/3 showing the SNP variant between C57BL/6 x JF1**

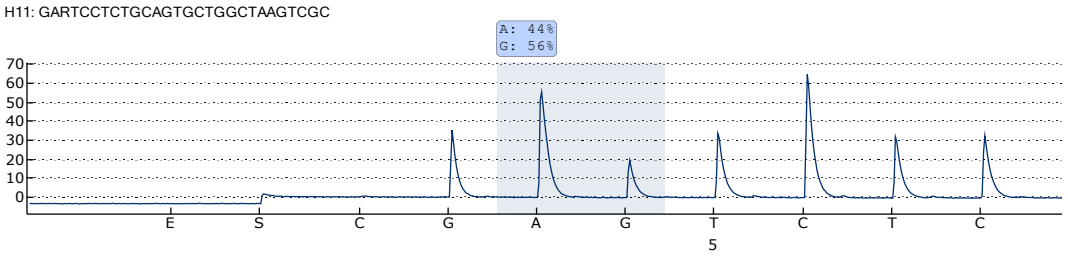
A

GGAAGTGGCAAAGCCAAGAAAGCACTGACGTACAGCGACTTAGC CAGCACTGCAGAGGAYTCGGAGACACTTCAGTTTCTCGCAGATAT ATTACCAAAGAAGATTTTAGCT	Exon 2 Exon 3 rs248258787 – G/A

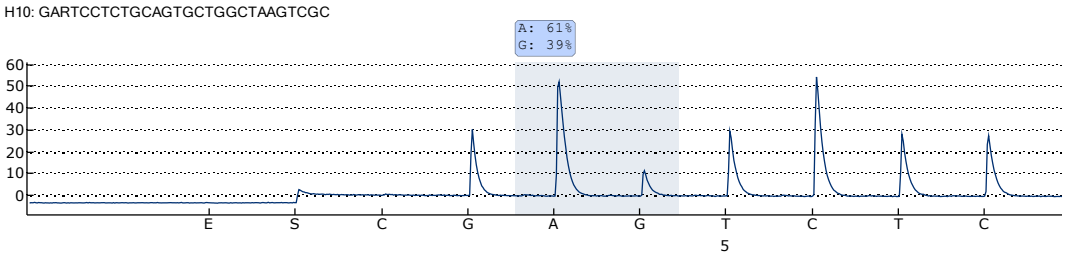
B

Brain

C57BL/6 x JF1

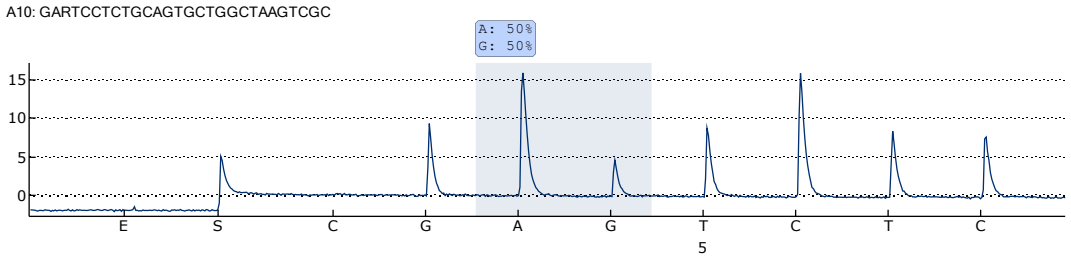


JF1 x C57BL/6



Kidney

C57BL/6 x JF1



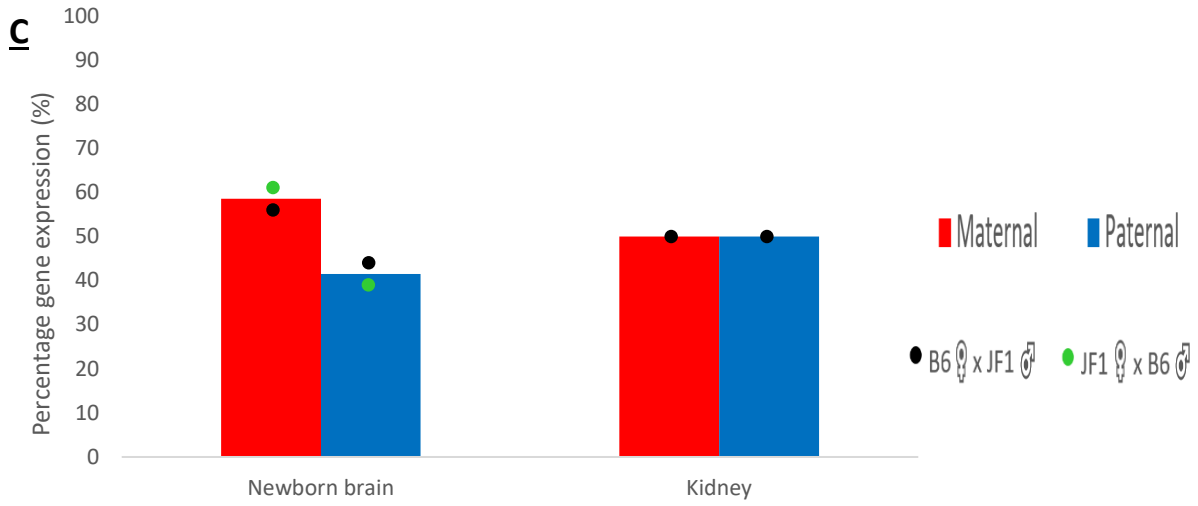
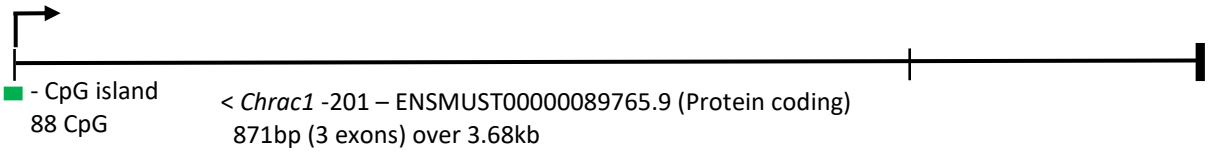


Figure 3.13: Allelic expression of *Chrac1* in newborn brain and kidney tissues from C57BL/6 x JF1 hybrid mice. Parental allelic expression was quantified via SNP pyrosequencing of cDNA from brain and kidney tissue **(A)** The SNP rs248258787 in exon 2 with flanking primers highlighted in yellow. **(B)** Representative pyrograms from newborn Brain for both crosses and Kidney for a single cross are shown, indicating the SNP position in the sequence and the percentage of each alleles expression. **(C)** Average expression (n = 1-2 mice) Points on the bar graphs represent individual pyrosequencing results from reciprocal crosses as shown in the legend

3.6.2 *Chrac1* pyrosequencing for DNA methylation

For the *Chrac1* promoter CGI both NSC and kidney tissues were analysed, however, due to sequencing errors from the pyrosequencing experiment, the kidney samples were inconclusive. The CGI at the promoter region of *Chrac1* located at position Chr15: 73,090,255-73,090,909 in the GRCm38/mm10 sequence showed little to no methylation (0-10%) in the NSC samples (Fig 3.14).

Chrac1 CGI



NSC

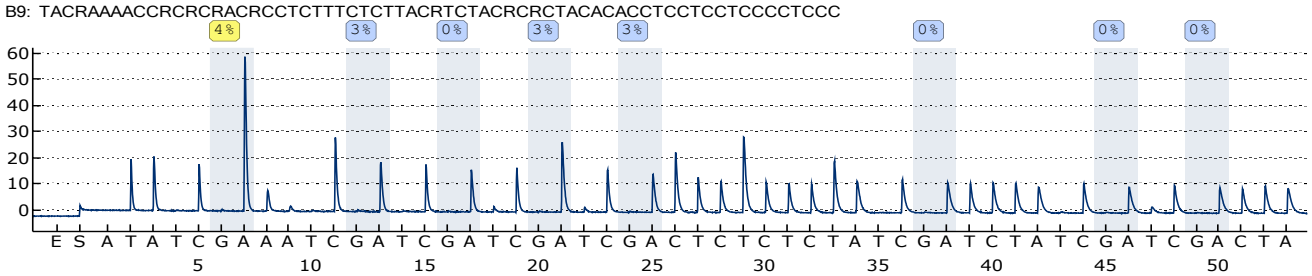


Figure 3.14: Methylation analysis of the *Chrac1* CGI located at position Chr15: 73,090,255-73,090,909 in the GRCm38/mm10 mouse genome. A schematic of the *Chrac1* locus showing exon/intron structure and the location of the CGI. Extracted genomic DNA underwent bisulphite conversion, and the newly converted CGI, from the reverse strand, was used as a template for PCR amplification. This amplified product was then pyrosequenced. Several CG dinucleotides were assessed and the percentage as to how much of the template was converted at that region was indicated, with 0% indicating no remaining CG dinucleotides/ full conversion and 100% indicating no conversion and maintenance of CG dinucleotides due to the presence of a methyl group. Methylation frequency indicates ~0-10% methylation in the *Chrac1* CGI for NSC. Data are for neural stem cells (hippocampal neurospheres) obtained from C57BL/6 x Cast/EiJ newborn F1 hybrids.

3.6.3 *Chrac1* allelic methylation analysis by Sanger sequencing

The SNP rs249421726 which shows an A/T variation in the C57BL/6 x Cast/EiJ hybrid mouse forward strand (Table 2.2) was used to distinguish between the maternal and paternal allele for *Chrac1*. From the Sanger sequencing results both the maternal and paternal allele SNPs identified almost every CpG dinucleotide was converted indicating a lack of methylation (Fig 3.15).

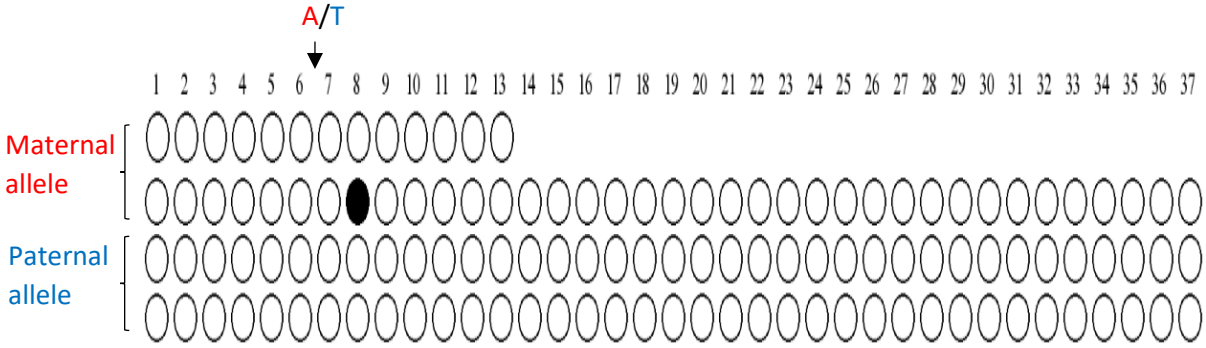


Figure 3.15: Analysis of genomic DNA methylation at the *Chrac1* CGI through bisulphite-treatment followed by Sanger sequencing of cloned samples. Sequencing data were analysed using QUMA. Data are for neural stem cells (hippocampal neurospheres) obtained from C57BL/6 x Cast/EiJ newborn F1 hybrids. QUMA output summary of Sanger sequences compared to the original non bisulphite converted reference sequence. SNP: rs249421726 was identified using Sanger sequencing and used to distinguish the maternal and paternal allele via their strain specificity. From the 37 CpG dinucleotides analysed the promoter CGI at *Chrac1* showed little to no methylation on both the maternal and paternal allele in NSC. Black circle = methylated CpG; white circle = unmethylated CpG).

3.7 Discussion

In this chapter I analysed both allele specific expression and methylation to characterise how the maternal and paternal alleles function for all genes located within the *Peg13-Trappc9* imprinting cluster. Both *Peg13* and *Kcnk9* displayed a strong (~90%) paternal and maternal expression respectively in murine brain tissue, while *Trappc9* and *Ago2* both displayed a preference for maternal expression in whole brain lysates, albeit to a more moderate extent (~75%) than that of *Kcnk9* in the mouse brain samples. All these results for allele specific expression in newborn mouse brain are in line with previous data obtained from whole transcriptome studies (Andergassen et al., 2017; Babak et al., 2015; Bonthuis et al., 2015; Bouschet et al., 2017; Crowley et al., 2015; Huang et al., 2017; Perez et al., 2015). However, *Chrac1* newborn brain maternal expression fell below the previously defined 70% threshold, indicating it is not as strongly biased towards maternal expression as *Trappc9* or

Ago2, although this has been previously observed for *Chrac1* in previous imprinting studies (Crowley et al., 2015; Perez et al., 2015). Furthermore, we observed tissue-specific imprinting of both *Trappc9* and *Ago2* (figs 3.1 & 3.4) both of which showed strong biased maternal expression in whole brain tissue but exhibited a biallelic expression pattern in kidney which is also in line with previous data and provides support that the data that I have obtained is accurate.

The most interesting and unexpected finding from this data, however, was the biallelic expression exhibited by the cultured hippocampal NSCs for *Ago2, Trappc9* given that both genes, in differentiated brain tissue, show a strong maternal preference. This is an interesting observation as studies into murine neural stem cell imprinting patterns, compared to those of more mature differentiated brain tissue at a later stage of development are limited. In addition to this, the identification that neural stem cell populations of *Trappc9* and *Ago2* appear to display an expression pattern that more closely resembles biallelic expression highlights a potential regulatory mechanism or factor that promotes an alteration of genomic imprinted expression bias upon differentiation into neurons. Although this has yet to be elucidated, some studies have already begun to identify the impact neural stem cell differentiation has on imprinting status (H. J. Lee et al., 2019). The strain bias observed in the *Kcnk9* NSC samples (fig 3.10) also highlights the importance of performing reciprocal cross analysis for these kinds of studies.

This alteration of expression state in *Ago2* and *Trappc9* NSCs compared to other neural tissues has previously been seen in the imprinted gene *Dlk1*. This imprinted *Dlk1* gene, which is widely expressed in the brain during early development, shows a preference for expression from the paternal chromosome during embryonic development (Ferrón et al.,

2011). However, previous studies have shown that the *Dlk1* gene loses it imprinting status and instead is biallelically expressed after transitioning into postnatal NSCs of the ventricular zone and hippocampal sub granular zone (Ferrón et al., 2011; Montalbán-Loro et al., 2021). This means that the mechanism controlling this shift from paternal bias to biallelic expression can override the early established imprinting mechanism generated in the germ cells and that imprinting status can be dynamic. Ferrón et al., (2011) goes on to state that this change in imprinting status is associated with postnatal gain of methylation at the DMR for *Dlk1* and is required for normal neurogenesis. However, when analysing the methylation status of the CGIs located within the *Peg13-Trappc9* cluster we found no alteration in methylation state at the *Peg-13* DMR in NSCs. Additionally, the CpG island located proximal to the *Trappc9* promoter (CGI1) remained unmethylated on both alleles while the other CpG island located near exon 2 (CGI2) retained its hypermethylation state previously identified in NSC for both alleles (figs 3.5 & 3.6). Furthermore, the CGI located near the *Ago2* promoter (figs 3.2 & 3.3) showed no methylation on either allele in NSCs. Therefore, the regulation of imprinted expression of *Trappc9* and *Ago2* in NSCs is likely to differ from how they are regulated in differentiated neurons and from how *Dlk1* imprinted expression is regulated in NSCs. Instead, regulation of *Trappc9* and *Ago2* expression and imprinting status in NSCs is more likely to be regulated by alternate regulatory features such as changes in histone modifications, transcription factor binding or as a result of other external factors such as enhancers.

Chapter 4 -Comparison of allelic expression bias and DNA methylation of CGIs of *Peg13*, *Trappc9* and *Ago2* in single neural stem cells and differentiated neurons

4.1 Introduction

Genomic imprinting, resulting in differential allelic expression, was first defined by Davor Solter and Azim Surani independently in 1984 (McGrath & Solter, 1984; Surani et al., 1984). Analysis of imprinted genes, as a result of the technology at the time, had only been possible using bulk primary cell culture or tissue lysates which have the potential to contain multiple different cell types. Although a large number of these imprinted genes show a strong allelic expression bias (>90%) and are considered mono-allelic, recent studies using RNA-seq have shown that several of these imprinted genes possess a weaker expression bias (~70%) which has previously been used as a cutoff threshold in the definition of whether a gene is considered imprinted (Andergassen et al., 2017; Babak et al., 2015; Bonthuis et al., 2015; Bouschet et al., 2017; Crowley et al., 2015; Huang et al., 2017; Perez et al., 2015). However, based on this data, when analysing imprinted gene expression and methylation frequency on a bulk tissue level it raises the question of whether the patterns exhibited are reflective of every individual cell within the lysate or merely show an average between the multitude of cells analysed (Perez et al., 2016).

The first use of single-cell expression analysis occurred in 2009 by a group attempting to compare the efficiency of single-cell sequencing versus a microarray (Tang et al., 2009; Wu et al., 2018). With this technology, individual cell expression profiles and identification of methylation states can be analysed, further advancing the field of genomic imprinting (Kelsey et al., 2017). A lower parental expression bias of ~70% may be explained by a number of single-cell expression profiles: A) All cells may exhibit the same biased biallelic

expression of 70% from one parental allele and 30% from the other. B) The bulk tissue may consist of a mixture of mono-allelic expressing cells and biallelic expressing cells where the average expression, on a bulk tissue analysis, would average a value between these two cell expression profiles e.g. 70% average expression from one allele. C) The analysed tissue could form a mosaic of different expression profiles, some mono-allelic paternal, some mono-allelic maternal and a range in-between which may form as a result of differential promoter usage in a similar manner to the *Grb10* gene in neurons versus glial cells and peripheral tissues (Claxton et al., 2022; Garfield et al., 2011; Sanz et al., 2008; Yamasaki-Ishizaki et al., 2006).

With imprinted bulk tissue expression data being explained by these different possibilities of single-cell expression behaviour, certain studies have attempted to identify which of these scenarios is most applicable to specific imprinted genes. For the *H19-Igf2* imprinted genes, SNP-FISH *in situ* hybridisation, utilising SNP-specific oligonucleotides, has been used in fixed fibroblasts and heart tissue to distinguish between mono-allelic and biallelic expression (Ginart et al., 2016). An alternative approach utilised by Bonthuis et al., (2015) was to use intronic RNAscope probes for newly generated RNA in the nucleus. This allowed them to see a mixture of cells with both mono-allelic and biallelic expression of imprinted genes in certain brain sections. While large-scale single-cell genomic imprinting analysis is still in its infancy, novel data suggest that individual cells can differ in their transcriptional status and deviate from the tissue average thus forming a mosaic (Martini et al., 2022).

4.2 Analysis of Bulk tissue versus single-cells

From the results of the pyrosequencing and Sanger sequencing data displayed in chapter 3, several conclusions can be made: I) Generally, we can see the imprinting bias that results in preferential expression of one allele over the other is more prevalent in whole brain lysate when compared to neural stem cells, except in the case of *Peg13*, where it is approximately equal in both samples showing a strong paternal bias. II) From using QUMA it is evident that there is little deviation in the methylation frequency on both alleles at the CGI regions in *Ago2, Trappc9, Kcnk9 and Chrac1* highlighting a lack of differential methylation between the maternal and paternal alleles. Furthermore, while I was unsuccessful in obtaining paternal allele data from my QUMA brain tissue analysis of *Peg13*, previous literature has shown this CGI to be differentially methylated (Court et al., 2014; Ruf et al., 2007; Smith et al., 2003) with methylation being present on the maternal allele, confirmed by my QUMA data, and a lack of methylation on the paternal allele.

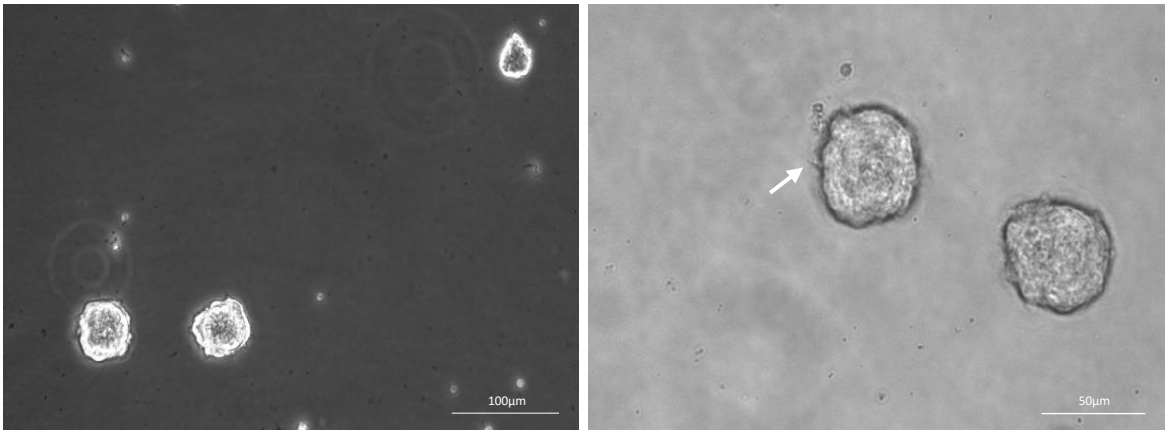
The aims of this chapter are to expand upon the findings of chapter 3 and determine whether the imprinting patterns of the genes located within the *Peg13-Trappc9* cluster are identical in both bulk-tissue and single-cells. To accomplish this, I isolated single NSCs or *in vitro* differentiated neurons from the hippocampus of newborn C57BL/6 x Cast/EiJ mice. Once isolated, the samples were then subjected to qRT-PCR using the sc-GEM (single-cell analysis of genotype, expression and methylation) technique developed by Cheow et al., 2016 and Sanger sequencing to determine allelic SNP expression bias for *Peg13*, *Trappc9* and *Ago2*.

From this analysis I expect to identify a single cell imprinted expression pattern that matched one of the three allelic expression states I described above i.e., I) All cells express

the same imprinting ratio, II) a mix of mono and biallelic expressing cells, or III) each cell expressing a variable percentage of expression from each allele with some cells being imprinted towards the maternal allele, some the paternal allele and some resembling biallelic expression. Of these three patterns, identifying all cells as displaying the same expression bias towards one allele would match the bulk tissue data but as previous studies have shown, individual cell transcriptomics can vary within a population and this would likely be the most significant find. Furthermore, with regard to the methylation analysis, I expect to see an overall majority of single cells that exhibit a similar methylation pattern to that of the bulk cells. The reason for this is that the methylation frequency of the promoter-proximal CGI's within the *Peg13-Trappc9* cluster were all in accordance with previous literature from the bulk sample analysis of whole brain lysate and NSCs. Therefore, I would expect to see little deviation from most CGI's being hypomethylated, with the exception of *Trappc9* CGI2 and the maternal allele of *Peg13*.

Finally, in addition to identifying the expression and methylation patterns of *Peg13*, *Trappc9* and *Ago2,* I also aimed to verify cell phenotype as an added confirmation to ensure the cells analysed were in fact NSC or differentiated neurons. This was achieved by verifying marker gene expression (Fig 4.1) (Hochgerner et al., 2018) for genes associated with neural stem cells and neuronal precursors.

A



B

Cell markers	Cell 1	Cell 2	Cell 3	Cell 4	Cell 5	Cell 6	Cell 7	Cell 8	Cell 9	Cell 10
Aqp4 (astrocytes)	✓	✓	×	✓	✓	✓	✓	✓	✓	✓
Eomes (nIPCs, NB1)	×	×	×	✓	✓	✓	×	✓	×	×
Mxd3 (RG, nIPCs)	×	×	✓	×	✓	✓	✓	✓	✓	×
Nestin (RG, nIPCs, vasculature)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Wnt8b (RG, nIPCs)	×	×	×	×	×	×	✓	×	×	×
Calb2 (NB1, NB2, CR)	×	×	×	×	×	×	×	×	×	×
Cdk1 (nIPCs)	✓	✓	✓	×	✓	✓	✓	✓	✓	✓
Igfbp1 (nIPCs, NB1, NB2)	×	×	✓	×	×	×	×	✓	×	×

Figure 4.1: Characterisation of neural stem cell cultures derived from hippocampi of newborn C57BL/6 x Cast/EiJ hybrid mice. (A) Images of neurospheres at day 3 of culture. Neurite-like extensions were recognisable occasionally (arrow). (B) Marker gene expression in a sample of single neurosphere cells was identified using qRT-PCR (sc-GEM) to determine cell type characteristics. Marker genes were selected according to (Hochgerner et al., 2018) and represent the following cell types: RG = Radial glia-like, nIPC = Neural intermediate progenitor cells, NB1 = Neuroblast 1, NB2 = Neuroblast 2, CR = Cajal-Retzius cell. Genes: *Aquaporin-4 (Aqp4)*, *Eomesodermin (Eomes)*, *MAX Dimerization Protein 3 (Mxd3)*, *Nestin*, *Wnt Family Member 8B (Wnt 8b)*, *Calbindin 2 (Calb2)*, *Cyclin-dependent kinase 1 (Cdk1)* and *Insulin-Like Growth Factor Binding Protein 1 (Igfbp1)*.

4.3 *Ago2*

4.3.1 *Ago2* single-cell expression

Ago2 has previously been shown to exhibit preferential maternal expression in the murine brain (Babak et al., 2015; Hsu et al., 2018; Perez et al., 2015) and our pyrosequencing expression data in chapter 3 (Fig 3.1) confirm this finding. To expand on this, I wanted to determine whether single cell imprinted expression reflects that of the bulk cell sequencing and whether all single cells express the same pattern of imprinted expression. I hypothesise that, for *Ago2*, if single-cell samples indeed follow the same pattern as bulk cell analysis then we would see, on average, equal expression from both alleles, as a result of either all cells expressing biallelically or variable expression states that average out to biallelic expression. In the single differentiated neurons however, we would find a shift in expression towards a maternal bias in single cells, similar to that of the whole brain lysate samples.

Using the SNP rs232384843 (Table 2.2) and the SC-GEM method (fig 2.2), I identified that *Ago2* allelic expression in single NSCs and differentiated neurons varied considerably between cells and there was no uniformity in cell expression. Our data indicated that NSCs display a small percentage (10%) of mono-allelic maternal expression, 16% mono-allelic paternal expression and 12% equal biallelic expression. The remaining 62% of NSCs displayed expression from both alleles but with a clear bias towards one allele (29% maternal biased, 33% paternal biased (Fig 4.2). Compared to the NSCs, the differentiated neurons possessed a greater number of cells exhibiting mono-allelic maternal (16%), equal biallelic (21%) and biased maternal (32%) expression while mono-allelic paternal (9%) and biased paternal (23%) expressing cells were reduced (Fig 4.2). (See section 4.7 for a full breakdown of single NSC and differentiated neuron expression).

A ***Ago2***

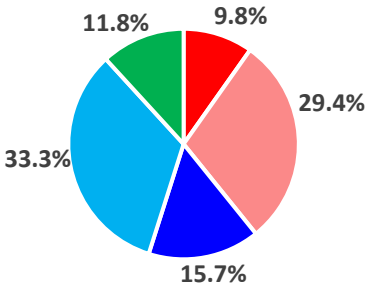
(Mat) C57BL/6 x (Pat) Cast/EiJ SNP variant: RS232384843 (T/C) Exon 4/ Exon 5

Coding strand

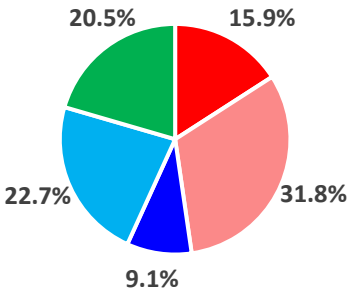
CGTTACACGATGCACTTTCGGGGCGGCTGCCCAGCGTCCCCTTCGAGACGATCCAGGCCCTGGACGTTGT
CATGAGGCACTTACCATCCATGAGGTACACCCCTGTTGGCCGTTCCTTCTTCACTGCATCTGAAGGCTGTT
CCAACCCTCTGGGTGGGGGCAGAGAAGTGTGGTTTGGCTTCCATCAGT

B

Percentage of neural stem cells with parental allele-specific expression biases (n=51)



Percentage of differentiated neurons with parental allele-specific expression biases (n=44)



C

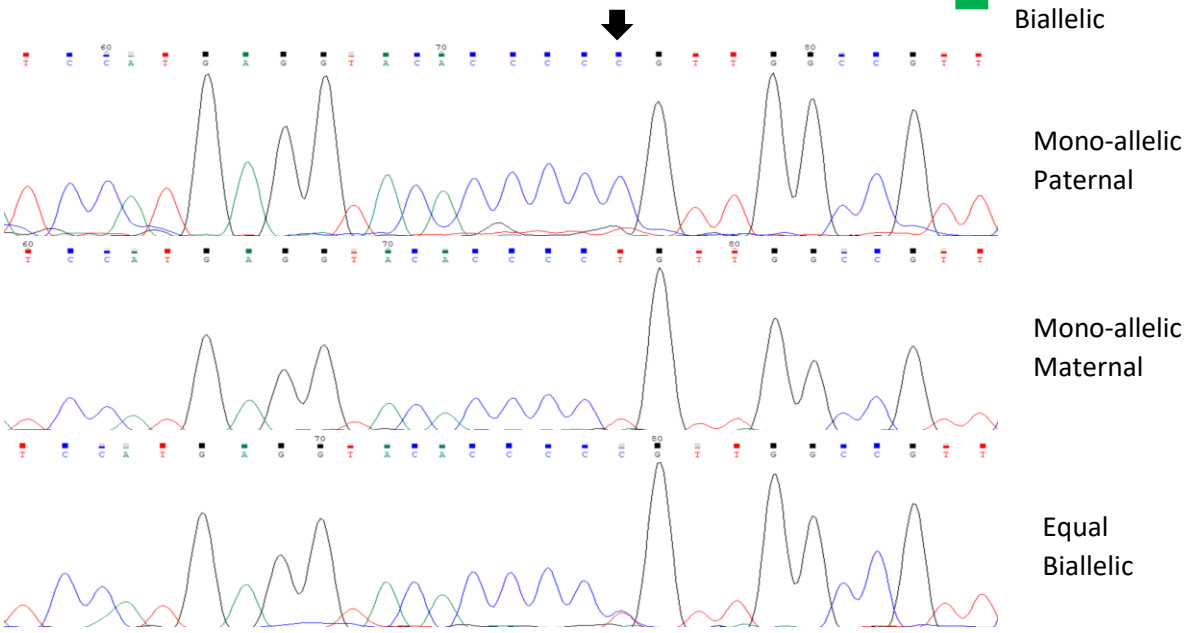


Figure 4.2: Variable single-cell allele-specific expression of the *Ago2* gene (Chr15: 72,967,693-73,056,784) analysed through Sanger sequencing of a cDNA SNP from neural stem (neurosphere) cells and their *in vitro* differentiated neurons. **(A)** The mouse cross, cDNA amplicon and SNP location in exon 5 are shown. C57BL/6 SNP variant RS232384843 is indicated in colour; F', nested R' and R' primers are underlined. **(B)** Summary data showing the proportions of NSC and differentiated neurons falling into the five categories of allelic expression. **(C)** Single-cell sequence tracks for three expression categories indicated. Alternate sequence tracks for cells with a maternal or paternal expression bias displayed a major peak for the SNP associated with the bias with a minor overlapping SNP peak for the other allele. SNP position is highlighted by an arrow.

This data indicates a surprising heterogeneity of allelic expression between individual NSCs that differs from the hypothesised biallelic expression expected in NSC indicated by the bulk neurosphere analysis.

4.3.2 *Ago2* single-cell methylation

Determination of methylation frequency within single NSCs and differentiated neurons was attempted using a methylation sensitive restriction digest as part of the sc-GEM procedure (fig 2.2). The *Ago2* gene has previously been shown to be unmethylated on the promoter proximal CGI for both alleles within mouse neurosphere NSCs which our bulk pyrosequencing and QUMA/ Sanger sequencing data confirmed. For the *Ago2* gene, our single-cell data indicate the presence of some degree of methylation with 59% (total n=22) of NSCs showing a detectable long amplicon, with a Ct value below 30 (Fig 4.3). This methylation presence at the *Ago2* CGI did not appear to be maintained in differentiated neurons with no long amplicons (n=18) being detected after BstUI digestion indicating no methylation presence.

A *Ago2* Full CGI GRCm38/mm10 chr15: 73,184,092 – 73,185,471 Forward strand

Short segment of *Ago2* CGI Long F' Short F' Nested R' **Outer R'**
L/Nes R=214bp S/NesR= 138bp BstUI digest site: 5'-CG$^{\vee}$CG-3'

AAAAAGCCTCCCCAACACTTGTTTTCTCAACCC**CGCG**AGTAAACCGCCCTCCGCTCGGCTCCAGGAGCG
GGGTGGGGAAATGCGTCCGTGTTTTTAAGTTTTCTGGAACTTTTCCAGAGAAACGGGCACGCTGGTGG
GAAAGGGGTGCCACGGCGGCCACCCGCTCCGCTAGGAAATCGCAGGGCGATT**AGCACTTGAATGGGG**
AGGGCGGCTGCCCGGCGTGAAACTGCGTCCA

B

NSC				Differentiated neuron			
Detected		Undetected		Detected		Undetected	
S' Amplicon	L' Amplicon	S' Amplicon	L' Amplicon	S' Amplicon	L' Amplicon	S' Amplicon	L' Amplicon
28.04	27.32	29.84	31.22	-	-	29.97	0.00
28.26	29.47	28.11	32.98	-	-	29.17	0.00
26.18	25.36	25.93	34.55	-	-	28.42	31.12
28.38	29.28	28.82	32.74	-	-	29.07	33.12
27.89	27.35	24.46	34.99	-	-	28.64	31.24
27.35	26.77	24.93	35.48	-	-	29.36	31.87
28.63	28.29	28.78	0.00	-	-	26.97	33.18
27.25	25.70	27.87	0.00	-	-	27.43	31.49
25.65	25.40	28.38	32.81	-	-	26.19	0.00
21.06	25.63	-	-	-	-	27.30	35.47
24.72	29.06	-	-	-	-	26.56	33.75
20.94	27.91	-	-	-	-	24.02	0.00
26.64	29.34	-	-	-	-	26.79	0.00
				-	-	26.80	34.00
				-	-	26.82	0.00
				-	-	24.13	0.00
				-	-	25.32	33.51
				-	-	26.09	33.05

C

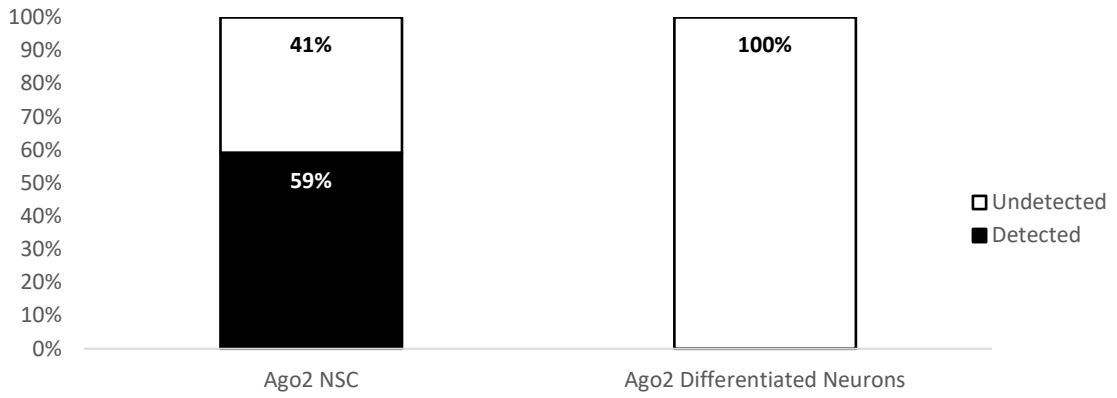


Fig 4.3: Single-cell methylation analysis of a short segment of the CGI found at the *Ago2* promoter. Full CGI located at position chr15: 73,184,092 – 73,185,471 in the GRCm38/mm10 mouse genome. **(A)** Description of the full *Ago2* CGI position and a short segment of the CGI sequence indicating the position of the upstream long amplicon F' primer (yellow), the downstream short amplicon F' primer (green), the Nested R' primer (blue) and the outer R' primer (underlined). The BstUI methylation-sensitive restriction site is highlighted in red. **(B)** Table indicating the Ct values obtained from qPCR of both the S' and L' amplicons for both NSCs and differentiated neurons after amplification of the *Ago2* CGI using the Sc-GEM method. Values > 30 were considered undetected **(C)** Percentages of single NSC (n = 22) and differentiated neurons (n = 18) showing detection of the long amplicon after BstUI digestion. In single NSC there is detection of the long amplicon indicating the presence of methylation at the CGI in 59% of cells, while in differentiated neurons no long amplicons were detected indicating complete digestion of the BstUI site highlighting a lack of methylation.

The NSC population displayed a majority of cells where the Long amplicon was detected indicating the presence of DNA methylation. This result was unexpected given the lack of methylation observed in the bulk NSC pyrosequencing and QUMA analysis and the lack of methylation observed in single differentiated neurons. Possible theories for why this deviation between single and bulk cell methylation occur will be explored in the discussion section of this chapter.

4.4 *Trappc9*

4.4.1 *Trappc9* single-cell expression

Trappc9 has also previously shown preferential maternal allele expression in the murine brain with the results obtained from the pyrosequencing expression data in chapter 3 (Fig 3.4) confirming this. However, in chapter 3, NSCs displayed a surprising biallelic expression pattern for *Trappc9* compared to the clear maternal allele preference in whole brain tissue lysate. Based on the bulk cell expression shown in chapter 3, I expect to see a biallelic expression, with a slight increase in maternal expression for single NSCs. In single differentiated neurons we would find a much more prominent bias towards the maternal allele for *Trappc9* expression, similar to that of the whole brain lysate.

Using the SNP RS31443479 and SC-GEM, I identified that *Trappc9* expression displays no uniformity across single cells as multiple alternate expression patterns were observed. In NSCs I observed mono-allelic expression in 20% of the cells (10% mono-maternal: 10% mono-paternal), while 29% displayed equal biallelic transcription. Approximately half of the NSCs displayed biased biallelic expression: 21% maternal biased, 31% paternal biased (Fig 4.4). When compared to the single NSC data, the differentiated neuron population showed a decrease in biallelic expression (10.9%) with many cells instead displaying either a maternal or paternal allele preferential expression pattern; 10.9% mono-maternal, 32.6% maternal biased biallelic expression (an increase of 11.4% compared to the NSC samples), 17.4% mono-allelic paternal expression (almost double compared to the NSC samples) and 28.3% paternal biased biallelic expression (Fig 4.4). See section 4.7 for a breakdown of imprinted expression states between NSCs and differentiated neurons.

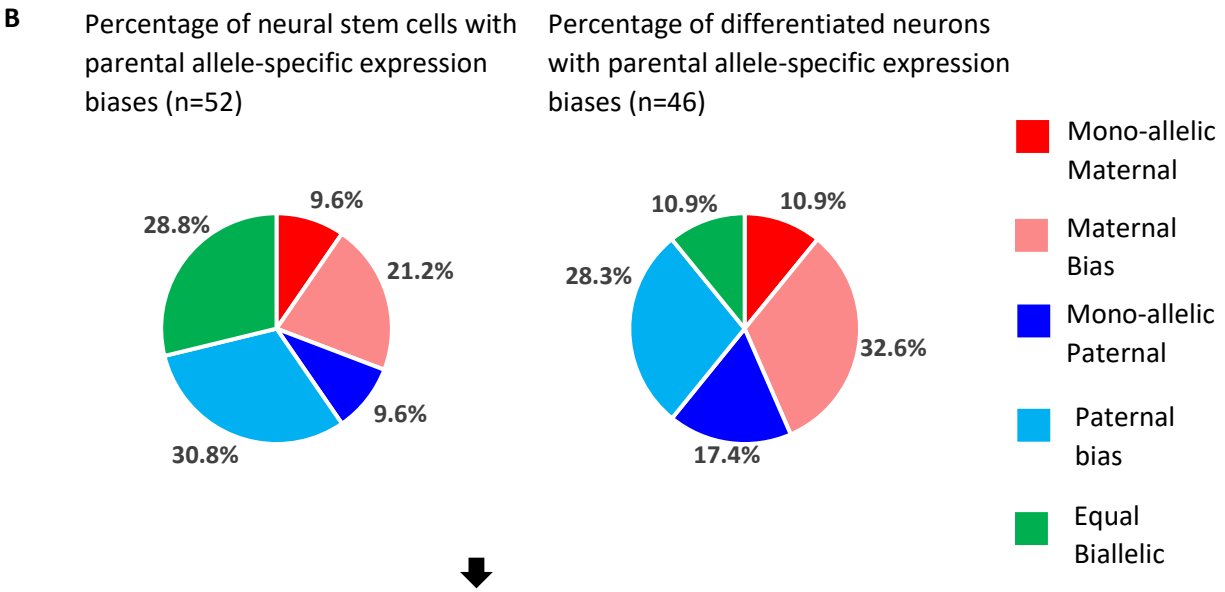
A ***Trappc9***

(Mat) C57BL/6 x (Pat) Cast/EiJ SNP variant: RS31443479 A/G Exon 7 / Exon 8

Coding strand

ATTGAGCTGGAAGCCTGTGTCAAGGCCGTGCGCGTCCT**A**GCGATTCAGAAGCGTGGCATGGAGGCTTCGGAG
TTTCTTCAGAATGCTGTGTACATCAATCTCCGGCAGCTTTCGGAAGAAGAGAAAATCCAGCGCTACAGCATCC
TGTCTGAGCTCTACGAGCTGATTGGCTTCCACCGCAAGTCAGCATTCTTCAAGCGAGTG

B



C

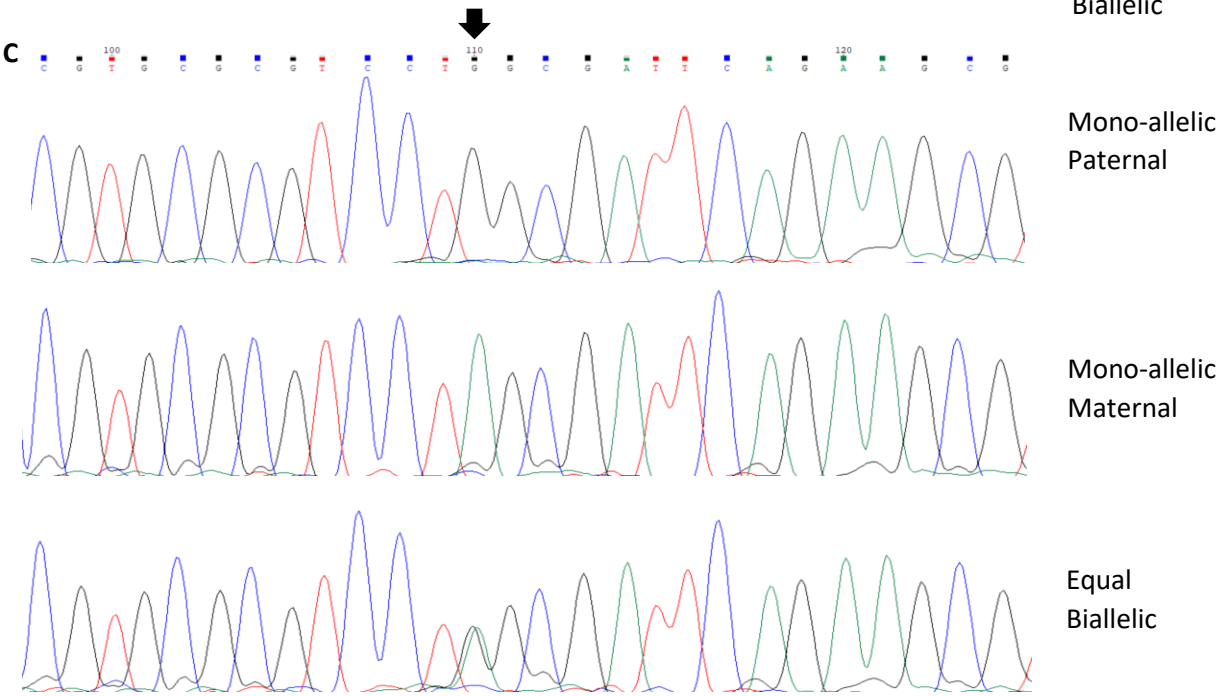


Figure 4.4: Variable single-cell allele-specific expression of *Trappc9* (Chr15: 72,461,469-72,933,053) analysed through Sanger sequencing of a cDNA SNP from neural stem (neurosphere) cells and their *in vitro* differentiated neurons. **(A)** The mouse cross, cDNA amplicon and SNP location in exon 7 are shown. C57BL/6 SNP variant, RS31443479, is indicated in colour; F', nested R' and R' primers are underlined. **(B)** Summary data showing the proportions of cells falling into the five categories of allelic expression. **(C)** Example single-cell sequence tracks for three expression categories indicated. Alternate sequence tracks for cells with maternal or paternal biased expression displayed a major SNP peak for the biased allele with a minor overlapping SNP peak for the other allele. SNP position is highlighted by an arrow.

4.4.2 Trappc9 single-cell methylation

The two CGIs located proximal to the first two exons of *Trappc9* at positions chr15: 73,060,920-73,061,124 for CGI1 and 73,058,049-73,058,500 for CGI2 were analysed using the sc-GEM procedure to identify single-cell methylation presence. Previous data suggests that CGI1 is unmethylated on both alleles while CGI2 is hypermethylated on both alleles. Based on my findings in chapter 3, I hypothesise that after BstUI digestion the long amplicon would not be detected in the CGI1 amplicon, while the methylated CGI2 BstUI site would remain intact and both long and short amplicons would be detected.

The data obtained indicated that for CGI1 the majority of both single NSCs (74%, n=34) and differentiated neurons (82%, n=34) produced no detectable long amplicon (Fig 4.5) indicating no methylation. Though the majority showed a lack of methylation, there were a small number of cells where the BstUI digestion was not successful, possibly due to the presence of methylation at the CGI1 site in some cells that the initial bulk sample pyrosequencing and Sanger sequencing did not identify or as a result of the restriction digest not working. For the CGI2 amplicons, all NSC samples showed detection of the long amplicon (100%, n = 34) while the majority (94%, n = 33) of differentiated neurons also showed detection (Fig 4.5). Both findings support our hypothesis of the *Trappc9* CGI1 being largely unmethylated and CGI2 being hypermethylated. It also shows that most single cells have similar methylation states and there is a high level of homogeneity between single cells regarding methylation frequency.

A

Trappc9 Full CGI1 GRCm38/mm10 chr15: 73,060,920-73,061,124 Forward strand

Short Segment of *Trappc9* CGI1 Long F' Short F' Nested R' **Outer R'**
L/Nes R=116bp S/NesR= 50bp BstUI digest site: 5'-CG$^{\blacktriangledown}$CG-3'

CTGGAGGGCAGCGGGACTGAGCAGCCTGGAACCTCCGAACCACCGGCCGAGCCGGCGCAGCTCCT
G**CGCG**TGCGCAAGTCGTGCTTCGGCCTAAGATCTGCCCCGGGTTTTCCCGAAGACCTGGAGCCTGC
GAGGGAGGCGG**CGAAGAGGGAGGAGGGAT**AAGAGAGAGGAGGAGGA

B

NSC				**Differentiated neuron**			
Detected		**Undetected**		**Detected**		**Undetected**	
S' Amplicon	**L' Amplicon**	**S' Amplicon**	**L' Amplicon**	**S' Amplicon**	**L' Amplicon**	**S' Amplicon**	**L' Amplicon**
19.60	26.26	27.45	0.00	26.28	29.37	28.83	0.00
23.11	27.99	26.41	0.00	26.34	28.01	27.92	0.00
21.79	25.42	22.13	32.40	26.15	28.73	27.24	0.00
20.15	26.98	25.47	0.00	25.75	27.05	27.93	0.00
20.62	27.69	22.18	0.00	26.40	28.15	26.05	0.00
19.18	23.52	24.56	0.00	25.31	27.40	26.46	0.00
20.49	21.99	22.99	0.00			26.12	0.00
20.02	25.42	24.11	0.00			28.60	0.00
19.65	25.82	20.84	34.10			27.02	0.00
		22.68	34.16			27.71	0.00
		23.44	30.24			27.98	0.00
		24.53	34.26			28.59	0.00
		21.95	33.54			28.22	33.99
		24.57	33.32			28.83	0.00
		20.45	33.76			26.41	30.35
		27.07	35.14			26.80	31.09
		19.46	30.04			28.26	0.00
		20.32	31.21			27.79	0.00
		17.83	32.79			25.46	0.00
		16.18	32.77			26.83	0.00
		26.42	33.75			25.76	31.79
		26.39	0.00			27.57	0.00
		27.35	0.00			26.36	0.00
		27.34	30.37			24.55	0.00
		26.00	33.64			26.95	0.00
						23.82	0.00
						25.17	0.00
						25.75	0.00

C

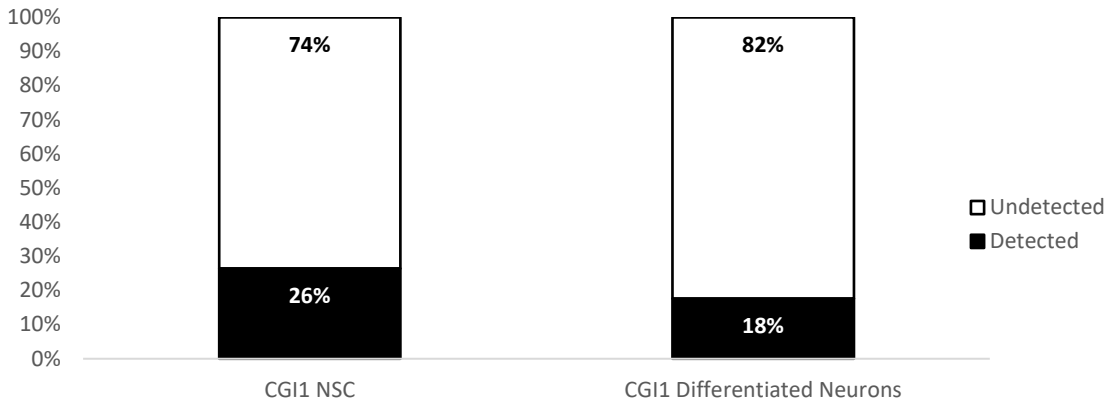


Fig 4.5: Single-cell methylation analysis of a short section of CGI1 found at exon 1 of the *Trappc9* gene, full CGI1 located at position chr15: 73,060,920-73,061,124 in the GRCm38/mm10 mouse genome for NSC and differentiated neurons. **(A)** Description of the full *Trappc9* CGI1 position and a short segment of the CGI1 sequence used for amplification indicating the position of the upstream long amplicon F' primer (yellow), the downstream short amplicon F' primer (green), the Nested R' primer (blue) and the outer R' primer (underlined). The BstUI methylation-sensitive restriction site is highlighted in red. **(B)** Table indicating the Ct values obtained from qPCR of both the Short and Long amplicons for both NSCs and differentiated neurons after amplification of the *Trappc9* CGI1 using the Sc-GEM method. Values > 30 were considered undetected. **(C)** Percentages of single NSC (n=34) and differentiated neurons (n=34) showing detection of the long amplicon after BstUI digestion. In single NSC there are only a small number of cells (26%) where the long amplicon is detected indicating a lack of methylation in most cells with the same pattern being exhibited in differentiated neurons.

A *Trappc9* Full CGI2 GRCm38/mm10 chr15: 73,058,049-73,058,500 Forward strand

Short segment of *Trappc9* CGI2 Long F' Short F' Nested R' **Outer R'**
L/Nes R=167bp S/NesR= 96bp BstUI digest site: 5'-CGᵛCG-3'

ATCGCAGTCATCGTAGTTGGGATAGAAGGCCACATCCGGGCGCGGCTGCTCAGCCACATCACC
CTGCAGCCCGAACACGAAGAGGCGGGAATCGTAGAGCGTGGAACCGTAGATCTCCTTCTGCA
CGTGGAACTTCTCGAAGGTCTGTGGCCAGTCTTTGGGCGAGAAGCAATCGGTG**ATGGTGATG**
AGGCCTACCACCTTGCGGTGCGTCTGGAAGTC

B

NSC				Differentiated neuron			
Detected		Undetected		Detected		Undetected	
S' Amplicon	L' Amplicon	S' Amplicon	L' Amplicon	S' Amplicon	L' Amplicon	S' Amplicon	L' Amplicon
22.00	27.13	-	-	24.56	25.45	25.22	0.00
21.76	28.16	-	-	24.11	25.70	23.50	0.00
14.63	15.43	-	-	25.06	26.57		
17.79	19.39	-	-	24.71	26.30		
17.91	19.52	-	-	21.82	22.92		
21.91	25.17	-	-	23.49	25.07		
19.56	20.56	-	-	22.25	24.66		
19.92	21.06	-	-	25.16	25.94		
19.09	19.96	-	-	23.73	25.31		
19.48	22.49	-	-	23.77	25.90		
17.31	23.87	-	-	26.65	27.20		
18.23	21.51	-	-	20.06	23.15		
19.54	22.76	-	-	20.17	24.64		
17.81	20.95	-	-	19.39	23.00		
15.87	18.92	-	-	20.58	23.20		
18.11	21.21	-	-	23.11	24.91		
16.28	18.18	-	-	22.82	24.42		
19.47	20.98	-	-	21.60	23.67		
17.98	19.44	-	-	21.19	23.78		
15.68	17.98	-	-	21.43	24.44		
15.65	17.77	-	-	21.40	23.86		
21.60	24.06	-	-	18.65	22.27		
12.28	13.37	-	-	22.98	27.85		
14.58	16.41	-	-	22.27	27.45		
14.39	20.03	-	-	21.78	26.43		
19.17	21.23	-	-	21.31	26.96		
13.53	15.96	-	-	20.65	25.34		
19.93	24.87	-	-	22.46	26.03		
24.95	29.13	-	-	20.39	24.34		
21.06	29.34	-	-	20.36	25.16		
19.32	24.72	-	-	21.49	25.33		
22.14	26.42	-	-				
20.19	25.20	-	-				
20.18	24.36	-	-				

C

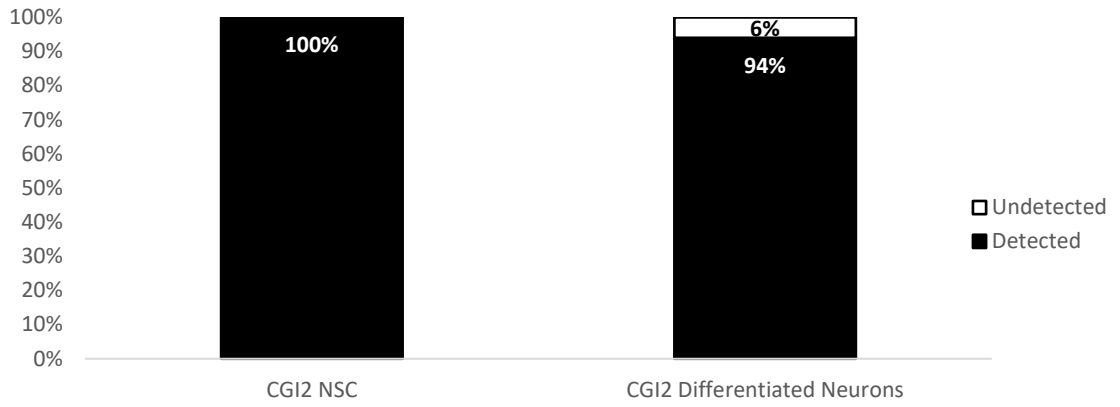


Fig 4.6: Single-cell methylation analysis of a short section of CGI2 found at exon 2 of the *Trappc9* gene, full CGI2 located at position chr15: 73,058,049-73,058,500 in the GRCm38/mm10 mouse genome in NSC and differentiated neurons. **(A)** Description of the full *Trappc9* CGI2 position and a short segment of the CGI2 sequence indicating the position of the upstream long amplicon F' primer (yellow), the downstream short amplicon F' primer (green), the Nested R' primer (blue) and the outer R' primer (underlined). The BstUI methylation-sensitive restriction site is highlighted in red. **(B)** Table indicating the Ct values obtained from qPCR of both the Short and Long amplicons for both NSCs and differentiated neurons after amplification of the *Trappc9* CGI2 using the Sc-GEM method. Values > 30 were considered undetected. **(C)** Percentages of single NSC (n=34) and *in vitro* differentiated neurons (n=33) showing detection of the long amplicon after BstUI digestion. Every NSC and most differentiated neurons (94%) produced a detectable long amplicon indicating presence of methylation in almost every cell.

4.5 *Peg13*

4.5.1 *Peg13* single-cell expression

Peg13 differs from both *Trappc9* and *Ago2* in the fact that it exhibits preferential paternal allele expression, not only in the murine brain but in additional peripheral tissues which the results obtained in chapter 3 (Fig 3.7) confirm. There appears to be no variation in the expression bias of *Peg13* across different tissue types leading to the hypothesis that in single NSCs we would find either a mono-allelic paternal or strong paternal bias which would be near identical in single differentiated neuron cells. Using the SNPs RS238259968 &

RS31423566, found in the sole exon of the *Peg13* gene, I tested NSCs and differentiated neurons for allelic expression bias via the SC-GEM method (Fig 4.7).

As expected, most NSC samples exhibited either mono-allelic paternal expression (39%) or displayed biased paternal expression (40%) showing only a weak maternal signal. Unexpectedly however, while most cells had a preference for paternal allele expression, there were a small number of NSCs that exhibited equal biallelic (2%) expression or even a preference for maternal expression with 5% showing mono-maternal expression and a further 14% showing maternal biased expression (Fig 4.7). Given the strong paternal bias observed in the bulk sample analysis and previous literature, the finding of individual cells that expressed solely from the maternal allele was a particularly interesting find. Theories as to why this may occur are touched upon in the discussion segment of this chapter.

For the *In vitro* differentiated neuron samples, most cells exhibited either mono-allelic paternal expression (29%) or preferential paternal allele expression (69%). This follows the paternal expression bias observed in the bulk whole brain lysate analysis. The heterogeneity of expression states in the differentiated neurons appeared to be reduced compared to that of the NSCs with no cells exhibiting either biallelic or mono-allelic maternal expression and only 2% of cells exhibiting preferential maternal expression (Fig 4.7). (See section 4.7 for a direct comparison of single-cell expression states between NSC's and *in vitro* differentiated neurons).

A

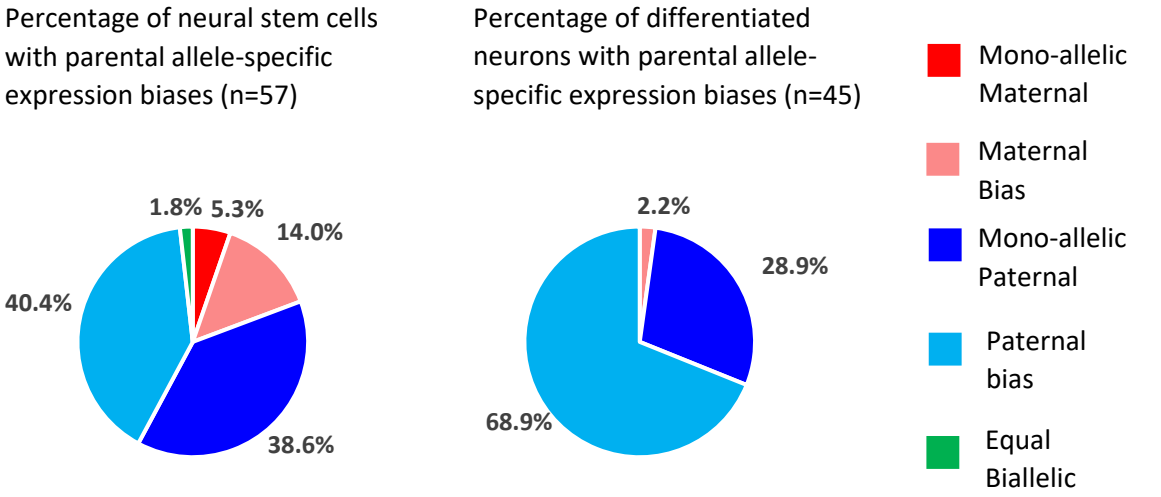
Peg13

(Mat) C57BL/6 x (Pat) Cast/EiJ SNP variants: RS238259968- (C/T) / RS31423566- (T/C) Exon 1

sense strand

GGCAAAAGGAGGCACAGAAAAAGCCCAGATATCTGTGTCGGTGCAGTTCGTGCGAGGATACCTTCGAGC
GTTGAGATACCTTCGAGCACAGAACCCTGCAGCTTGCGATGCAACGGAGAGCTTGGCCGATCTTTTTACCC
TG**GAGACGCTAGCACAGAGGCT**CGCCATCACCCTGTGGATAGGCGCATTGTGGTTGTAG**CACAATGATG
AGCCATGGAG**

B



C

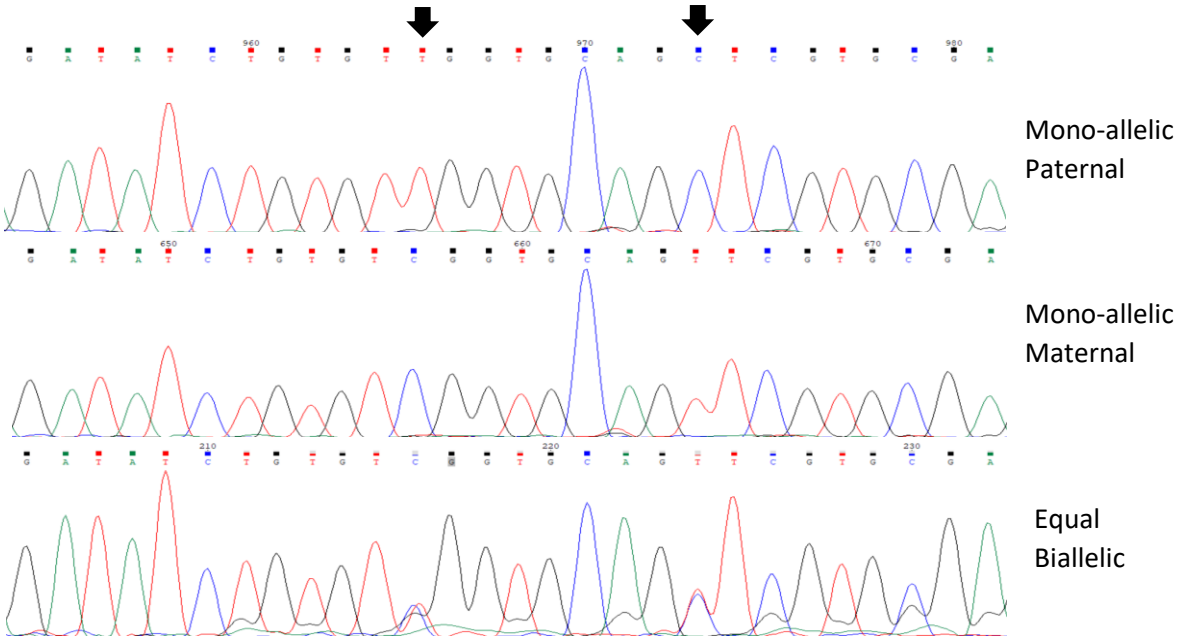


Figure 4.7: Variable single-cell allele-specific expression of *Peg13* (Chr15: 72,677,449-72,682,173) analysed through Sanger sequencing of cDNA SNPs from neural stem (neurosphere) cells and their *in vitro* differentiated neurons. (A) The mouse cross, cDNA amplicon and SNP locations in exon 1 are shown. The C57BL/6 SNP variants RS238259968 and RS31423566 are indicated in colour; F', nested R' and R' primers are underlined. (B) Summary data showing the proportions of cells falling into the five categories of allelic expression. (C) Single-cell sequence tracks for three expression categories indicated. Alternate sequence tracks for cells with maternal or paternal bias displayed a major SNP peak for the biased allele with a minor overlapping SNP peak for the other allele. SNP positions are highlighted by arrows.

4.5.2 Peg13 single-cell methylation

The *Peg13* CGI, located at position chr15: 72,809,538-72,810,123 in the GRCm38/mm10 mouse genome, has previously been shown in NSCs and whole brain tissue to be differentially methylated with the maternal allele displaying hypermethylation (confirmed by our bulk cell QUMA analysis (Chapter 3 section 3.4.3)), and the paternal allele displaying hypomethylation as indicated in previous literature. As the Sc-GEM procedure is unable to differentiate DMR's (as mentioned in section 4.2.2 of this chapter), I hypothesise that all NSC and differentiated neurons will produce a detectable long amplicon due to the expected presence of methylation on the maternal allele providing an undigested template for both forward primers to amplify regardless of the methylation state of the paternal allele.

I found that in 85% of our NSC samples (n = 34) and 89% of our differentiated neuron samples (n = 35), there was a detectable long amplicon indicating methylation in line with my hypothesis (Fig 4.8). However, 15% of NSCs and 11% of differentiated neurons did not show detection of the long amplicon as expected. This is likely due to complication with the methylation restriction digest and will be addressed in the discussion section of this chapter. The data still, however, suggest that the majority of NSCs are homogenous for methylation on at least one allele at this CGI and that this methylation pattern is maintained upon differentiation.

A *Peg13* Full CGI GRCm38/mm10 chr15: 72,809,538-72,810,123 Forward strand

Short section of the *Peg13* CGI Long F' Short F' Nested R' **Outer R'**
L/Nes R=144bp S/NesR= 50bp BstUI digest site: 5'-CG^CG-3'

GTGAAAAACTCCCGGCTCCGCGGGCAGTGTCGCAGGTCTTCTATCCAACCATTTTCAGAGTAAA
GGCCGCGGATCTTCCCGGAGAGCCGAAGCTGGCCGGAGAGCCGCGGCAATGATCGTCTACA
TAGCACCAGCGCTTGGATGAGCTATCACACAAGGCCTAAAATCTCAATAAGATGGGCTAACA
AGCCACCGGGTCCT

B

NSC				Differentiated neuron			
Detected		Undetected		Detected		Undetected	
S' Amplicon	L' Amplicon	S' Amplicon	L' Amplicon	S' Amplicon	L' Amplicon	S' Amplicon	L' Amplicon
21.51	30.00	20.37	31.89	23.38	28.74	21.41	30.14
17.20	21.65	20.62	30.70	21.79	25.59	20.87	31.33
19.14	24.66	20.19	30.57	23.12	26.28	27.52	0.00
18.72	25.43	25.58	34.57	22.52	27.85	17.52	35.06
19.83	25.39	21.96	33.12	22.12	28.47		
21.38	28.00			20.11	24.61		
19.25	25.45			21.70	25.26		
22.69	29.11			20.89	26.17		
18.57	29.07			22.24	26.01		
20.78	28.78			21.29	24.47		
22.89	27.67			21.79	25.67		
18.44	21.15			24.31	28.03		
20.49	24.50			19.31	26.78		
19.48	23.25			20.54	28.91		
21.31	25.26			20.68	27.00		
20.57	25.76			20.94	29.19		
18.33	21.71			20.74	26.19		
18.83	22.31			20.10	26.54		
12.92	12.72			21.39	26.97		
13.69	16.92			20.32	26.60		
14.08	29.64			20.78	25.75		
13.98	20.49			20.57	27.36		
11.41	13.89			17.64	22.58		
18.77	27.03			18.96	26.48		
10.11	27.66			16.67	26.88		
22.64	27.37			19.19	32.68		
23.53	26.39			17.66	27.27		
23.13	26.18			17.23	25.61		
22.49	25.25			16.81	23.70		
				17.42	26.45		
				17.43	25.98		

C

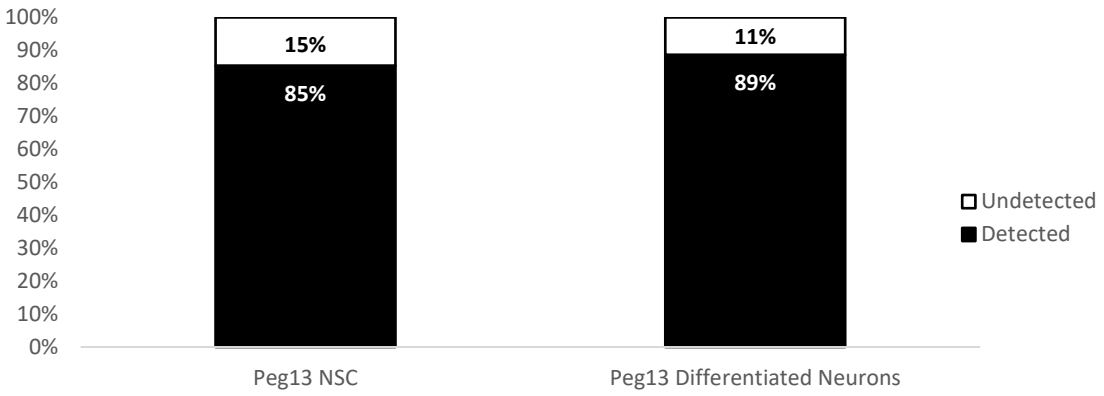


Fig 4.8: Single-cell methylation analysis of a short section of the CGI found at *Peg13*, full CGI located at position chr15: 72,809,538 -72,810,123 in the GRCm38/mm10 mouse genome. **(A)** Description of the full *Peg13* CGI position and a short segment of the CGI sequence indicating the position of the upstream long amplicon F' primer (yellow), the downstream short amplicon F' primer (green), the Nested R' primer (blue) and the outer R' primer (underlined). The BstUI methylation-sensitive restriction site is highlighted in red. **(B)** Table indicating the Ct values obtained from qPCR of both the Short and Long amplicons for both NSCs and differentiated neurons after amplification of the *Peg13* CGI using the Sc-GEM method. Values > 30 were considered undetected. **(C)** Percentages of single NSC (n = 34) and differentiated neurons (n = 35) showing detection of the long amplicon after BstUI digestion. In single NSC there is detection of the long amplicon indicating the presence of methylation at the CGI in 85% of cells, while in differentiated neurons 89% of cells had detectable long amplicons indicating a high level of methylation in both cell types.

4.6 *Kcnk9* single-cell methylation

Kcnk9 single-cell expression was not analysed due to practical and time constraints of performing multiple Sanger sequencing experiments. However, the methylation status of the *Kcnk9* CGI located at position chr15: 72,545,926-72,547,811 in the GRCm38/mm10 mouse genome was analysed using a methylation-sensitive restriction digest. As shown in chapter 3, *Kcnk9* has very little methyl presence in bulk NSC samples based on pyrosequencing and QUMA analysis. For the single-cell methylation analysis of the *Kcnk9* CGI (Fig 4.9) I hypothesise that both cell types would show very limited methylation, meaning minimal detection of the long amplicon, in concordance with their bulk cell counterparts. Our results indicated that a higher proportion of NSCs than initially predicted

showed a detectable long amplicon (40%), however, the majority still showed no methylation presence as predicted. Additionally, the differentiated neuron samples produced no detectable long amplicon indicating all cells were hypomethylated at the *Kcnk9* CGI region.

A *Kcnk9* Full CGI GRCm38/mm10 chr15: 72,545,926-72,547,811 Forward strand

Short section of *Kcnk9* CGI Long F' Short F' Nested R' **Outer R'**
L/Nes R=118bp S/NesR= 77bp BstUI digest site: 5'-CG$^{\wedge}$CG-3'

CAAGCGGTTTGCCCCAGTGCGGATGGTTGTGCGGATGGATGGCCCAAGCTCGTTC**CGCG**CG
GGCTGTCTGGTCCTGGGCACCTCGCCCCTCGAGCCTGCGGGCTCGGTTGGGGCAGGATCGG
TAGCCGGGACAGAGGGAGACCCGGGGAGG**GGTTCCCCATCCCTGATG**CGTTCCCCGGCTA
GCGGGGCCCAGCTAATCGCCTCCAGTGTGGCGGTTGCCC

B

NSC				**Differentiated neuron**			
Detected		**Undetected**		**Detected**		**Undetected**	
S' Amplicon	**L' Amplicon**	**S' Amplicon**	**L' Amplicon**	**S' Amplicon**	**L' Amplicon**	**S' Amplicon**	**L' Amplicon**
24.63	29.83	27.37	0.00	-	-	29.48	0.00
26.46	27.12	27.57	0.00	-	-	29.12	0.00
27.07	28.03	29.37	0.00	-	-	28.69	0.00
16.86	22.13	29.14	33.08	-	-	28.86	0.00
17.86	24.26	28.01	0.00	-	-	29.10	0.00
19.48	24.83	27.56	0.00	-	-	28.81	0.00
19.07	25.04	29.88	0.00	-	-	27.39	0.00
17.03	23.08	28.56	0.00	-	-	29.75	0.00
		29.61	35.38	-	-	29.18	0.00
		28.77	0.00	-	-	28.97	0.00
		23.57	30.38	-	-	27.99	0.00
		26.18	30.87	-	-	29.71	0.00
				-	-	27.49	0.00
				-	-	29.74	0.00

C

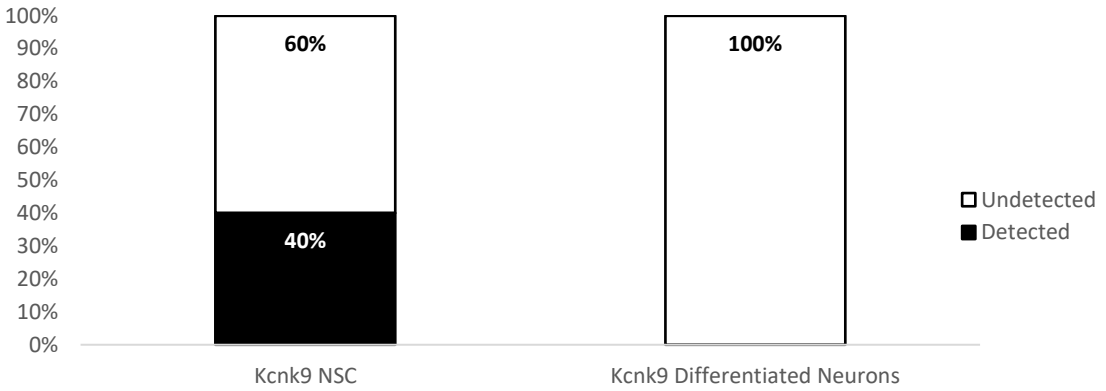


Fig 4.9: Single-cell methylation analysis of a short section of the CGI CGI found at the *Kcnk9* promoter located at position chr15: 72,545,926-72,547,811 in the GRCm38/mm10 mouse genome. **(A)** Description of the full *Kcnk9* CGI position and a short segment of the CGI sequence indicating the position of the upstream long amplicon F' primer (yellow), the downstream short amplicon F' primer (green), the Nested R' primer (blue) and the outer R' primer (underlined). The BstUI methylation-sensitive restriction site is highlighted in red. **(B)** Table indicating the Ct values obtained from qPCR of both the Short and Long amplicons for both NSCs and differentiated neurons after amplification of the *Kcnk9* CGI using the Sc-GEM method. Values > 30 were considered undetected. **(C)** Percentages of single NSC (n = 20) and differentiated neurons (n = 14) showing detection of the long amplicon after BstUI digestion. In single NSC there is detection of the long amplicon indicating the presence of methylation at the CGI in 40% of cells, while in differentiated neurons no long amplicons were detected indicating complete digestion of the BstUI site highlighting a lack of methylation.

4.7 Summary of results

Table 4.1: Summary of murine single-cell expression data for the imprinted genes *Ago2, Trappc9* and *Peg13* in both Neural stem cells and differentiated neurons. Cells were analysed using a combination of Sc-GEM and Sanger sequencing to determine the percentage of expression from each gene originating from either the maternal or paternal allele.

Single cell expression						
Gene		Mono-allelic Maternal	Biased Maternal	Mono-allelic Paternal	Biased Paternal	Equal Biallelic
Ago2	NSC	9.8%	29.4%	15.7%	33.3%	11.8%
	Neurons	15.9%	31.8%	9.1%	22.7%	20.5%
Trappc9	NSC	9.6%	21.2%	9.6%	30.8%	28.8%
	Neurons	10.9%	32.6%	17.4%	28.3%	10.9%
Peg13	NSC	5.3%	14.0%	38.6%	40.4%	1.8%
	Neurons	0%	2.2%	28.9%	68.9%	0%

For *Ago2* and *Trappc9* while the NSC and differentiated neurons displayed no uniformity between cells, the NSC expression ratios are reflective of the bulk expression data we obtained in chapter 3 when taken as an average expression of all individual cell data. For instance, when focusing on *Ago2* the combination of mono-allelic maternal and maternal bias expression indicates that ~39% of NSCs show preferential maternal expression to a varying degree. Meanwhile, the mono-allelic paternal and paternal bias expression represents 49% of NSC expression patterns observed for *Ago2*. This, in conjunction with the 11.8% of NSCs that exhibited biallelic expression, indicates NSC *Ago2* expression is diverse with a range of various expression categories and no clear allele bias. It also confirms that single-cell *Ago2* expression in NSCs, when taking all the various expression categories as an average, reflects the equal biallelic *Ago2* expression found in bulk neurosphere samples

However, with regards to the differentiated neurons, while a clear shift to a more maternal biased expression pattern and a reduction of paternal allele biased expression was evident, the single *in vitro* differentiated neurons did not reflect the same strong bias of 75% maternal biased expression of *Ago2* that were detected in the whole-brain lysate. Instead, in single *in vitro* differentiated neurons, the average expression ratio is similar to that of equal biallelic expression (47.7% maternal, 31.8% paternal and 20.5% equal biallelic). Possible explanations for this will be touched upon in the discussion segment.

The *Trappc9* single NSC data is much in line with that of *Ago2* showing a high degree of variability and roughly matching the bulk neurosphere analysis as no clear parental allele bias was observed. However, in the single differentiated neuron samples, there is an almost equal increase in both maternal biased biallelic expression and mono-allelic paternal expression when compared to the NSCs, likely as a result of the decrease in biallelic expressing cells. Despite these increases in both maternal and paternal expression, the single neuron dataset depicts no clear allelic bias in *Trappc9* expression, which contrasts with the maternal expression bias of 78% observed in whole-brain lysate from pyrosequencing (fig 3.4). These discrepancies in the data may be due to the whole brain lysate used in the bulk data analysis containing a wider range of neuronal and other cell types (e.g., astrocytes, microglia, oligodendrocytes) and a wider range of neuron types from different brain sub-regions as compared to the *in vitro* differentiated neurons, which were derived from hippocampal neurospheres and used as a single cell representative fraction of the whole brain lysate used for pyrosequencing.

For *Peg13*, the single NSC data indicates that ~79% of the NSCs analysed have a paternal bias which approximates the bulk neurosphere expression bias of 87% paternal transcripts

identified through pyrosequencing. The single differentiated neurons show ~98% bias towards paternal expression, much in line with the paternal bias of 89% identified in brain tissue. The key observation from the *Peg13* analysis was that cells display either a low level of maternal allele expression or can express solely from the maternal allele which is theorised to be transcriptionally inactive as a result of the DMR and genomic imprinting.

Overall, the majority of the single cell data was complementary to the results I obtained from the bulk cell data. However, I have now identified that the biallelic expression of NSCs is a result of variable expression states unique to each cell rather than all cells expressing in a biallelic manner, a previously theorised but unconfirmed hypothesis for this imprinted gene cluster.

Table 4.2: Summary of murine methylation at CGI's located proximal to the promoter regions of imprinted genes positioned within *Trappc9/Peg13* cluster. Both Neural stem cells and *in vitro* differentiated neurons isolated from newborn hippocampus tissue were analysed

Single cell CGI methylation			
CGI		Methylated (Detected L')	Unmethylated (undetected L')
Ago2	NSC	59%	41%
	Neurons	0%	100%
Trappc9	CGI1 NSC	26%	74%
	CGI1 Neurons	18%	82%
	CGI2 NSC	100%	0%
	CGI2 Neurons	94%	6%
Peg13	NSC	85%	15%
	Neurons	89%	11%
Kcnk9	NSC	40%	60%
	Neurons	0%	100%

The methylation analysis of the promoter proximal CGIs within the *Peg13-Trappc9* imprinted cluster were mostly reflective of the pyrosequencing and QUMA analysis results discussed in chapter 3. However, there were a few deviations in the data that were unexpected and are mentioned in the discussion segment of this chapter.

4.8 Discussion

Due to the advancement of technology moving towards single-cell transcriptome analysis and the increased sensitivity of *in situ* hybridization techniques for single-cell work, investigations into characterising imprinting gene expression on the single-cell level have become far more accurate and financially feasible (Martini et al., 2022; Varrault et al., 2020). For this project, I elected to use the Sc-GEM technique rather than single-cell RNA-seq for a more limited analysis of the genes of this imprinting cluster in single cultured NSCs and differentiated neurons (Cheow et al., 2015, 2016; Lorthongpanich et al., 2013).

The data obtained from the Sc-GEM procedure suggest that *Trappc9* and *Ago2* showed a great deal of variability in their allelic transcriptional states in individual NSCs and neurons. All categories of allelic transcription ranging from mono-allelic, preferential allelic expression and equal biallelic expression were all represented by significant numbers of cells (figs 4.2 & 4.4). When considering all NSC data analysed for *Trappc9* and *Ago2*, it indicates no clear bias towards one allele is evident. When the variable transcription states of the NSC data were levelled out, they were in line with the bulk NSC data seen in the previous chapter where biallelic expression was observed for these two genes. Additionally, we found similar findings in the single neuron samples we differentiated from NSCs for both *Ago2* and *Trappc9* expression, with multiple transcription patterns of allele-specific expression observed. However, when comparing the single NSC samples to the

differentiated neuron samples for these two genes several small differences were identified. For *Ago2* there was a slight shift from a paternal to a maternal allelic bias, with the percentage of cells exhibiting either preferential maternal allele expression or biallelic expression increasing. Additionally, for *Trappc9*, the proportion of cells exhibiting biallelic expression was reduced in the neuron population. However, this reduction in biallelic expression did little to alter the proportion of single neuron samples exhibiting paternal or maternal biased expression, with both transcriptional states remaining relatively balanced. Even with these small differences between the neuron and NSC populations, there was no clear expression bias towards one allele in the single differentiated neurons, which contrasts with the bulk data obtained from whole brain lysates. One possible explanation for this variation between whole brain lysate and single differentiated neurons is that our single differentiated neurons may not be a true reflection of the whole brain tissue lysate. When culturing the differentiated neurons, I actively selected against the formation of dividing glial cells by adding AraC to the culture medium. Furthermore, due to the differentiated neurons being generated from the differentiation of NSCs isolated from the hippocampus *in vitro*, the culture is likely to contain only a limited range of neuronal cell types. Previous data from the Allen brain map (https://portal.brain-map.org/) *in situ* hybridization atlas and single-cell transcriptomics showed medium levels of *Ago2* and *Trappc9* expression in numerous neurons of the cortex and hippocampus, with additional lower levels of these genes in some types of neurons as well as in astrocytes and oligodendrocytes. This data is also supported by histological analysis for *Trappc9* by Ke et al (2020) as well as histological RNAscope data (Pulix, 2017).

Another potential aspect that would impact these results would be the possibility of technical issues such as allele dropout during the reverse transcription step. This would

affect the more weakly expressed allele e.g. the paternal allele of *Trappc9* or the maternal allele of *Peg13* (as indicated from the whole brain lysate expression analysis) more than the strongly expressed allele of these imprinted genes. This would lead to an increased number of cells showing mono-allelic expression for the biased allele as the weaker expressed allele would be undetected. However, allele dropout is less likely to occur in the expression analysis and more likely to occur in the methylation analysis of genomic DNA, due to their being only two template copies present in each single cell, compared to the expression analysis where there are likely hundreds of mRNA templates. Additionally, there is no indication that my single-cell data are affected by these potential technical issues and furthermore, I have found no way to test for such eventualities at extremely low numbers of RNA molecules. On the contrary, the data I have obtained from these single-cell analyses shows the opposite, with a large number of cells showing the expected expression bias but with a readily detectable expression of the weaker allele e.g. *Peg13* showing strong paternal expression but with clear weak maternal expression also occurring. Additionally, we also saw cells with predominant maternal expression of *Peg13* (fig 4.7) in complete contrast to the standard expression bias seen in previous literature. These results would not be expected if there was a significant rate of drop-out of the weakly expressed allele.

For *Peg13*, our single-cell data suggested that NSCs predominantly exhibited a paternal-expression bias which was further exaggerated in the single neuron samples, although a substantial number of these cells exhibited a paternal bias with weak maternal expression as opposed to complete mono-allelic paternal expression (fig 4.7). However, a small number of these cells, predominantly from the NSC samples, deviated from this majority to show either biallelic or maternal biased expression with a small number even exhibiting mono-allelic maternal expression. While these results were surprising, they are not

unprecedented. A previous study of imprinted gene expression using single cortical cells identified a similar variability in imprinted expression to our data and also showed occasional deviations from the expected biases (Laukoter et al., 2020). The study found that the imprinted gene *Meg3* which normally shows a strong maternal bias was found to be biallelic in a small number of cortical cells and two additional imprinted genes *Inpp5f* and *Impact*, that usually exhibit a paternal bias occasionally showed biallelic or predominantly maternal expression (Laukoter et al., 2020). The mechanisms that regulate these variable allelic expression states of imprinted genes in single cells remain to be fully identified. Several imprinted genes have been attempted to be characterised for alterations in expression states such as the gain of methylation and loss of imprinting in *Dlk1* upon differentiation into NSCs (Ferrón et al., 2011; Montalbán-Loro et al., 2021), or the use of alternative promoters on the maternal and paternal alleles for the *Grb10* gene (Garfield et al., 2011; Sanz et al., 2008; Yamasaki-Ishizaki et al., 2006). However, these models are unlikely to be responsible for our findings. An alternate model involving random mono-allelic expression (RMAE) such as transcriptional bursting (Chess, 2016; Reinius & Sandberg, 2015; Varrault et al., 2020; Xu et al., 2017) may affect the imprinted genes and be responsible for the variation we see in our single-cell samples. These RMAE effects may be stochastic and dynamic rather than produce any permanent/ longitudinal effect on transcription such as in the case of random allelic exclusion of immunoglobulin genes and may be a result of short-lived CTCF cohesion loops altering chromatin architecture (Gabriele et al., 2022).

Since the non-coding RNA *Peg13* is transcribed from the core imprinting regulatory region (DMR) of the locus, findings of *Peg13* expression other than mono-allelic paternal or paternally biased biallelic (Fig 4.7) were mostly unexpected. This raises the question of whether there is a specific correlation associated with this unexpected *Peg13* transcription

from the maternal allele coinciding with deviations from the expected maternal preferential allelic expression of the other imprinted genes of the cluster. When analysing the expression status of *Trappc9* and *Ago2* in both NSC and differentiated neurons cells that additionally showed equal biallelic, maternally biased or mono-allelic maternal expression of *Peg13*, I did not find any specific patterns or correlation of abnormal *Trappc9* or *Ago2* expression (Table 4.3). Some of these cells displayed the expected maternal bias of *Trappc9* and/or *Ago2*, but other expression patterns including mono-allelic expression states were also observed showing no clear discernible correlation between unexpected transcriptional states of the imprinted genes of the cluster.

Table 4.3: Analysing anomalous expression patterns of *Peg13* in single NSCs and differentiated neurons from Sanger sequencing results, where expression was biallelic or showed preferential expression from the maternal allele and identifying any correlation of expression for *Trappc9* and *Ago2*. For cells that showed unexpected allelic expression in *Peg13*, the allelic expression pattern for *Ago2* and *Trappc9* for the same cell was compared to see whether all genes expressed in a manner unusual to their standard transcriptional state. No clear discernible pattern could be identified as there were multiple combinations of expected and unexpected results to varying degrees of biased biallelic expression and mono-allelic expression.

Cell	*Peg13* expression	*Trappc9* expression	*Ago2* expression
Expected expression	**Mono-allelic paternal**	**Biased maternal**	**Biased maternal**
Exp 1: Cell 1	Biased maternal	Biased paternal	Biased paternal
Exp 1: Cell 2	Biased maternal	Biased paternal	equal biallelic
Exp 1: Cell 3	Biased maternal	Biased maternal	Biased paternal
Exp 1: Cell 5	Mono-allelic maternal	Biased paternal	Biased maternal
Exp 2: Cell 9	Biased maternal	N/A	N/A
Exp 3: Cell 4	Biased maternal	N/A	Biased maternal
Exp 3: Cell 20	Mono-allelic maternal	N/A	N/A
Exp 6: Cell 1	Biased maternal	Biased paternal	Biased maternal
Exp 7: Cell 3	Biased maternal	Mono-allelic paternal	Biased maternal
Exp 7: Cell 8	equal biallelic	Biased paternal	equal biallelic
Exp 7: Cell 14	Mono-allelic maternal	Biased maternal	Mono-allelic paternal
Exp 7: Cell 17	Biased maternal	Mono-allelic maternal	Biased maternal
Exp 7: Cell 4	Biased maternal	Biased paternal	Biased paternal

The CGIs located proximal to each imprinted gene promoter within the *Peg13-Trappc9* cluster were analysed for the presence of methylation using the Sc-GEM procedure (fig 2.2). Detection of both the S' and L' amplicon was indicative of methylation while detection of only the S' amplicon indicated a lack of methylation. The *Ago2* gene showed a strong presence of methylation in the NSC samples (fig 4.3) which contrasts with the bulk neurosphere NSC data obtained from pyrosequencing and QUMA analysis which indicated little to no methylation of the *Ago2* CGI in bulk NSCs. However, in the single neuron samples, differentiated from cultured NSCs, the *Ago2* promoter CGI showed no detection of the L' amplicon indicating every methylation-sensitive restriction site was digested and no methylation was present. We have previously seen alterations in methylation upon differentiation from the *Dlk1* gene, which is associated with a gain of methylation after differentiation into NSC (Ferrón et al., 2011). Due to the deviation in methylation status of the *Ago2* CGI between NSC and differentiated neuron samples, there is a possibility that DNA methylation status is not maintained post differentiation into neuron cells. However, a more likely explanation for this scenario is due to a failure of the BstUI methylation sensitive digest in the NSC samples. This would provide a template for both the long and short amplicon making the CGI appear methylated, due to the failure of the restriction enzyme. Further testing is required to determine whether methylation status is impacted by differentiation or whether using a methylation sensitive digest when analysing single-cell methylation patterns is too great a limiting factor for this type of analysis.

There are two CGIs located within the *Trappc9* gene, one located proximal to the promoter/ exon 1 (CGI1) and one located proximal to exon 2 (CGI2). The promoter CGI1 exhibited expected non-methylation patterns in a majority of cells (fig 4.5), where the L' amplicon was undetected at this region in NSCs, and an even greater percentage of cells with no

methylation in single neuron cells. This is supported by the QUMA data which determined there was no methylation on either allele at the *Trappc9* CGI1 region in bulk NSCs. The second CGI of the *Trappc9* gene located proximal to exon 2 on the other hand showed that every NSC with a detectable S' amplicon also had a detectable L', highlighting that this region is hypermethylated (fig 4.6). The single neuron samples exhibited a very similar pattern of detection of both amplicons, with only a very small percentage showing no detection of the L' amplicon, indicating hypermethylation at this CGI region. The small number of cells with an undetectable L' amplicon for *Trappc9* CGI2 in single neurons may be a result of the representative nature of the Sc-GEM technique. While we assume that the digested CG dinucleotide is representative of the CGI as a whole, as the QUMA data previously showed, there are certain CG dinucleotides whose methylation status can deviate from the surrounding environment. The small percentage of cells with an undetected L' amplicon may be a result of this, with those cells being unmethylated at a CG dinucleotide digest site, even if the surrounding CG dinucleotides were methylated, resulting in digestion by the enzyme and no detection of the L' amplicon. However, the effect of this is more limited due to the fact that the BstUI enzyme contains a CG^CG target where both CG dinucleotides need to be unmethylated for digestion to take place.

The *Peg13* CGI located at the *Peg13* promoter has previously been shown to be differentially methylated, with the maternal CGI methylated and the paternal allele unmethylated (Suzuki et al., 2011). Due to the inability of the Sc-GEM technique to differentiate between whether the maternal or paternal alleles are differentially methylated, every cell analysed using this method should theoretically show detection of both the S' and L' amplicon as every cell should contain at least one methylated copy of the genome, being the maternal allele, that would not be digested and can act as a template for

the downstream processes of the Sc-GEM procedure. Both the single NSC and differentiated neuron samples showed a majority of cells where the L' amplicon was detected indicating the presence of methylation which was expected (fig 4.8). However, a small percentage of cells in both samples showed a detectable S' amplicon but no detectable L' amplicon indicating a lack of methylation in these cells. While a small number of cells did not produce a detectable long amplicon as expected, unlike the *Ago2* and *Trappc9* samples there were multiple BstUI digest sites between the two F' amplicons and so there is a greater possibility that one of the digest sites possessed unmethylated CG dinucleotides. If even one of the CG dinucleotides is unmethylated it will result in no detection of the long amplicon due to the template for the Long amplicon being disrupted. This is one potential reason for the small percentage of cells showing no detection even though theoretically every cell should show a detectable long amplicon based on the differential methylation of the *Peg13* CGI. However, a more likely possibility is due to allelic dropout as a result of using single cell genomic DNA. When using single cell genomic DNA for analysis, only one template copy is available, unlike mRNA which usually constitutes thousands of copies per cell. If the allelic dropout affects the methylated maternal allele, then the BstUI, and resulting analysis, would be carried out on the unmethylated paternal allele which would display a short amplicon but no long amplicon due to the action of the BstUI enzyme. This would explain why *Peg13*, which I hypothesised would always produce a long amplicon due its methylated maternal allele, may show some cells as being unmethylated.

Kcnk9, which has been shown to be expressed from the maternal allele in both human and mouse brain tissue (Skaar et al., 2021), also possesses a CGI located proximal to its promoter. When I analysed this CGI using the Sc-GEM procedure, the majority of NSCs had a detectable S' amplicon but no detectable L' amplicon indicating no methyl presence (fig

4.9). However, while a minority, there were still a significant number of NSCs that produced a L' amplicon indicating the presence of methylation. This was in contrast to the single neuron samples where every cell showed a detectable S' amplicon and no detectable L' amplicon. One possible reason for the presence of detectable long amplicons in some NSCs could be due to certain cells being methylated at the BstUI digest site found within the *Kcnk9* CGI which is partially supported by the pyrosequencing and QUMA analysis of the bulk data where a small degree of methylation was present in the *Kcnk9* CGI in NSCs. Another possibility may be due to a failure of the restriction enzyme BstUI, either through reaction conditions or human error, resulting in a template for the downstream processes of the Sc-GEM procedure resulting in a detectable L' amplicon. In the differentiated neuron sample, there was no detectable long amplicon meaning that both alleles are hypomethylated as expected.

In summary, our single-cell analysis of the three imprinted genes *Ago2, Trappc9* and *Peg13* indicates a surprising heterogeneity of allelic expression states in individual cells, ranging from mono-allelic maternal to mono-allelic paternal transcription, even for the core imprinted gene of the locus, *Peg13*. Thus, our single-cell data do not support a model, in which a tissue-level imprinted gene expression status is reflected in each cell of the tissue in the same way. These findings might hint at a certain level of transcriptional noise or transient/ random bursts of transcription, which might still be able to occur at alleles that are "silenced" by genomic imprinting (Varrault et al., 2020). Furthermore, the allelic expression ratios observed in the bulk-cell expression data reflects an average of the multiple variable expression patterns of the single-cells in the NSC samples for *Ago2, Trappc9* and *Peg13* as well as the single differentiated neurons for *Peg13*. However, it does not appear to reflect the variability between *Ago2* and *Trappc9* single differentiated

neurons expression patterns which appear to show no clear discernible allelic expression bias in the same way the whole brain lysate samples do towards maternal allele expression. Another possible reason for this is due to the cells being differentiated *in vitro* and so they may not receive the same signals to silence expression of the paternal allele, or they are missing some crucial factor responsible for a shift towards preferential maternal expression.

In regard to the identification of methylation at the CGI regions of the genes within the cluster, while I cannot say for certain that our results are 100% representative of the cells as a whole due to potential limitations of the approach, the general consensus appeared to mimic that of the bulk cell data achieved via pyrosequencing and QUMA/Sanger sequencing. While *Ago2* and *Kcnk9* NSC showed a greater detection of the L' amplicon, and therefore methylation, at the promoter proximal CGI regions than initially anticipated in the NSC samples, both *Trappc9* CGI1 and CGI2 as well as the *Peg13* CGI are in line with bulk methylation data. Additionally, the *in vitro* differentiated neuron samples for every gene's promoter CGI across the *Peg13-Trappc9* cluster was in line with bulk methylation data. However, despite this single cell methylation analysis displaying promising results for methylation frequency analysis, due to its limitations, alternate single cell methylation methods may be more applicable for identifying single-cell methylation. One such technique would be single-cell Targeted Analysis of the Methylome (ScTAM-seq) which analyses thousands of cells and has a low dropout rate (Bianchi et al., 2022). This method would likely produce a more accurate representation of methylation frequency at the CGIs within the cluster, especially as the Sc-GEM technique reads a single restriction site, in this case

CG^CG from the BstUI methylation sensitive enzyme, that is supposed to represent the CGI as a whole as opposed to reading the CGI as a whole.

Chapter 5 - Analysis of potential gene regulatory regions of the *Trappc9* imprinted gene

5.1 Imprinting mechanism and regulation

Previous chapters have identified expression and methylation patterns of the genes within the *Peg13-Trappc9* imprinted cluster and the CGIs proximal to these genes' promoter regions in both bulk tissue and single-cell samples. However, the mechanism that underpins this tissue-specific imprinting and preferential maternal allele expression of *Trappc9* and *Ago2* in the mouse brain has yet to be elucidated. In addition to DNA methylation, the mechanisms involved in the regulation of imprinted gene expression comprise histone modifications, non-coding RNAs and boundary/ insulator elements that are recognised by CTCF, a methylation-sensitive DNA binding factor. The CTCF protein acts as an anchor/ boundary factory, binding at unmethylated sites, frequently within DMRs, allowing for the formation of chromatin loops via interaction with cohesin to form a CTCF-cohesin complex (Pugacheva et al., 2020). These complexes have been shown to regulate the proximity of tissue-specific enhancers, such as in the case of the imprinted *Igf2-H19* locus, to their targeted promoter regions and promote the formation of allelic topologically associated domains (TADs), thereby controlling the expression of neighbouring genes in an allele-specific manner (Bell & Felsenfeld, 2000; Llères et al., 2019).

While the imprinting status of most genes is conserved between human and mouse, some genes do not show an allelic expression bias in one or the other species. Previous studies

have shown that the genomic imprinting of *Ago2, Chrac1* and *Trappc9* observed in mice is not conserved in humans (Andergassen et al., 2017; Babak et al., 2015; Bonthuis et al., 2015; Court et al., 2014). Furthermore, while many imprinted genes show a strong allelic bias in all analysed tissues, some imprinted genes, can also be tissue specific. Examples of such tissue-specific imprinted genes include *Gnas* in defined brain regions, endocrine glands and proximal renal tubules, *Ube3a* in neurons (but not glia and peripheral tissues) and *Ago2, Chrac1, Trappc9* within the *Peg13* imprinted cluster (Claxton et al., 2022; Mabb et al., 2011; Perez et al., 2015; Peters, 2014; Weinstein et al., 2010; K. Yamasaki et al., 2003). Based on these findings there is a strong possibility that some of these non-DNA methylation regulatory features may be involved in generating the imprinting bias we see for the genes within the *Peg13-Trappc9* cluster in the murine brain.

One potential mechanism by which the tissue-specific imprinting may arise is chromatin boundaries and CTCF-regulated access to tissue-specific enhancers such as in the imprinted *Igf2-H19* locus (Bell & Felsenfeld, 2000). There is already evidence that potentially supports this theory as CTCF has been shown to bind to the unmethylated paternal allele DMR of *Peg13* in mouse fibroblasts (Prickett et al., 2013; P. Singh et al., 2011) as well as in human brain (Court et al., 2014). Evidence obtained by Court et al., (2014) showed that, in humans, CTCF regulates binding of the *Kcnk9* and *Peg13* promoters to a brain-specific enhancer, via alternate chromatin looping mechanisms, dependent on the ability of CTCF to bind to an unmethylated region proximal to the gene being regulated. The study conducted by Court was able to identify one interaction, mediated by CTCF binding, of a brain-specific enhancer with the *Kcnk9* promoter, while another interaction showed the enhancer interacting with the *Peg13* promoter. While not empirically proven, these enhancer-promoter interactions are most likely regulated by the differential methylation of the maternal and paternal alleles

at the *Peg13* DMR, in order to facilitate transcription in an imprinted manner. However, their study was unable to differentiate between the two parental alleles due to their use of human samples and no clear parental SNP associations.

These enhancers that Court *et al* mention are small regions of the genome that are major gene-regulatory elements. They are able to control cell-type-specific gene expression, most often by chromatin looping long distances so the enhancer comes in close physical proximity with the promoters of its target genes (Schoenfelder & Fraser, 2019). While studies estimate that there are hundreds of thousands of enhancer DNA regions (Pennacchio et al., 2013) it is theorised that enhancers interact with specific promoters, with which they are brought into close proximity in order to regulate the expression of that gene. A study by Beagan et al., (2020) found, in brain cortical neurons, 24,937 enhancer loops bringing specific enhancers to their target promoter. Enhancers also appear to occasionally work in tandem, often at tens or hundreds of thousands of nucleotides distant from their target genes and can coordinate with one another to control the expression of their common target gene (Schoenfelder & Fraser, 2019). Once these enhancer loops are formed, they act as a scaffold for additional co-factors to bind onto, these transcription factors then help upregulate the expression of the gene (Grossman et al., 2018). In this chapter, I discuss the process of identifying any brain-specific potential regulatory elements that may be responsible for upregulating the expression of *Trappc9*. Once identified, these features were then tested in promoter-reporter gene assays to determine which regulatory element had the greatest impact on the expression of *Trappc9*. From the analysis of these regulatory elements, I hypothesised that I would identify one or several brain specific regulatory element(s) that upregulate the expression of the *Trappc9* gene in a tissue-specific manner i.e. increased expression of *Trappc9* in brain tissue but not in control fibroblasts. However, the results I

obtained, surprisingly, showed regulatory elements that had opposite effect, a net decrease in *Trappc9* expression compared to control samples. This is likely due to a repressive/ silencer function of the element in question which is discussed later in the chapter.

5.2 Regulatory element identification

Due to the lack of conservation between the imprinting status of *Ago2, Trappc9* and *Chrac1* between mouse and human, the regulation of these genes presumably involves an enhancer element that is also not conserved. We, therefore, screened the Encode3 and Ensembl mouse genome databases for known enhancer locations, chromatin modifications and DNA accessibility (Abascal et al., 2020; Gorkin et al., 2020), across the *Trappc9-Peg13* locus. Using the Ensembl database I identified enhancer regions within the *Peg13-Trappc9* imprinted cluster and utilised previous studies submitted to the database to screen whether they were active in brain and displayed specificity i.e. were inactive in other mouse tissues, due to most of the genes within the cluster exhibiting tissue-specific imprinting, localised to the brain. Once the brain-specific enhancers located within the *Trappc9-peg13* locus were identified, I then transferred the genomic coordinates of these enhancers to the ENCODE database to search for chromatin features that denote enhancer activity e.g. histone modifications via ChIP assay data to identify H3K4 methylation, H3K27 acetylation and H3K9 acetylation histone as well as DNaseI and ATAC hypersensitive sites in newborn (P0) mouse brain. I also examined to see whether CTCF was known to bind to that region of DNA due to CTCF-ChIP interactions mentioned previously in the literature. Genomic regions that have a high frequency of these genetic modifications/ sensitivities have been shown to be highly enriched for enhancer activity (Grossman et al., 2018). If the enhancers identified using

ENCODE (and visualised in the UCSC genome browser) displayed a relatively good peak height (> 2) from the ChIP seq data ENCODE provided that feature was considered to be positive at that specific enhancer region (fig 5.1) Of all the enhancers identified, there were seven potential brain-specific regulatory elements, that we labelled A, B, C, D, E, 2 and 8, containing most if not all of these features in multiple brain tissues (Table 5.1). The rest of the enhancer regions identified from Ensembl displayed few/weak enhancer like chromatin characteristics from the ENCODE database and so were not used in this study.

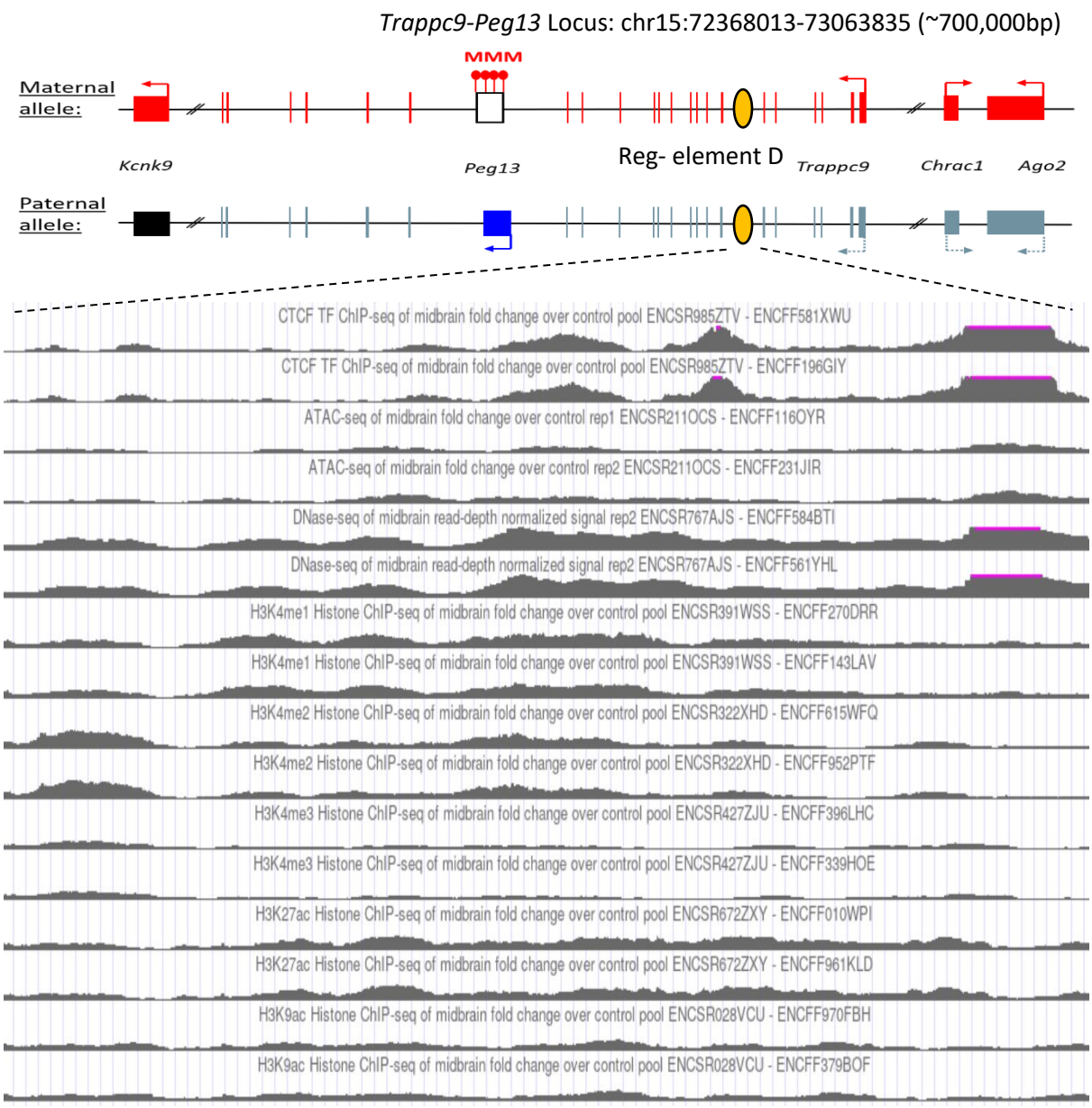


Figure 5.1: Identification of chromatin features at candidate regulatory element D (Chr15: 73,019,700-73,023,494) located within the *Trappc9* imprinted locus. Histone modifications and DNA features were identified using the ENCODE3 database and visualised in the UCSC genome browser database. Regulatory features identified include: CTCF binding, ATAC hypersensitivity, DNase I hypersensitivity, H3K4me1, H3K4me2, H3K4me3, H3K27ac and H3K9ac. The features shown are obtained from midbrain tissue samples of newborn (P0) mice at the regulatory element D position (chr15: 73,019,700-73,023,494 in the GRCm38/mm10 mouse genome). These traces show an example of how chromatin modifications were identified, further summary data of other brain regions and candidate regulatory elements are identified in a later figure.

Table 5.1: Identification of potential brain-specific regulatory elements across the *Peg13-Trappc9* locus. The candidate brain-specific regulatory regions were identified using the ENCODE3, Ensembl and UCSC Genome Browser databases to extract epigenetic and chromatin accessibility features (ChIP data for histone modifications, CTCF and RNA-polymerase II, DNAse I and ATAC hypersensitivity), which are associated with enhancer-like characteristics in the P0 mouse brain. Six different brain tissues were analysed due to the imprinting of the cluster occurring only in brain tissue. Abbreviations correspond to ENCODE3 / UCSC Genome Browser annotations. cHMM = chromHMM (chromatin hidden Markov model). cCRE = candidate cis-regulatory element. Chromosome positions refer to the mouse GRCm38/mm10 genome version.

Genomic feature	Reg - A	Reg - 2	Reg - B	Reg - C	Reg - 8	Reg - D	Reg - E
Position (Chr:15)	72,625,407-72,626,800 (1393 bp)	72,657,344-72,658,462 (1118 bp)	72,849,700-72,853,300 (3599 bp)	72,860,700-72,862,300 (1599 bp)	72,891,564-72,892,637 (1073 bp)	73,019,700-73,023,500 (3794 bp)	73,066,670-73,068,528 (1858 bp)
cHMM	En-P (4)	En-W (2) En-P (3)	En-W (2) En-S (2) En-P (3 and 4) En-Ws (3)	En-W (2) En-P (3 and 4)	En-W (4)	En-W (2, 3 and 4) En-P (3 and 4) En-S (2, 3 and 4)	En-S (2 and 4) En-P (2, 3 and 4)
cCRE	enhD	enhD	enhD	enhD	enhD	enhD	enhD
ATAC	1,2,3,4	1,2,3,4	1,2,3,4	1,2,3,4	1,2,3,4	1,2,3,4	1,2,3,4
CTCF	X	2,5	2,5	X	X	1,2,3,4,5	2,3,4
POLR2A	1	5	5	X	5	5	1,5
DNase HS	1,2,3,4	1,2,3,4	1,2,3,4	1,2,3,4	1,2,3,4	1,2,3,4	1,2,3,4
H3K4 me1	1,2,3,4,5	1,2,3,4,5	1,2,3,4,5	1,2,3,4,5	2,3,4,5	1,2,3,4,5	1,2,3,4,5
H3K4 me2	2,3,4	2,4	2,3,4	2,3,4	4	2,3,4	2,3,4
H3K4 me3	X	X	5	X	X	2,3,4,5	1,2,3,4,5
H3K27ac	2,3,5	2,3,4,5	1,2,3,4,5	2,3,4,5	2,3,4	1,2,3,4,5	1,2,3,4,5
H3K9ac	2,3,4	2,4	2,3,4	2,3,4	2,3	2,3,4	2,3,4

LEGEND		
CEREBELLUM = 1	FOREBRAIN = 2	MIDBRAIN = 3
HINDBRAIN = 4	OLFACTORY BULB = 5	FRONTAL CORTEX = 6

5.3 Testing regulatory elements

5.3.1 Identifying the promoter region

To determine the impact these regulatory elements had on the expression of *Trappc9*, we decided to use a promoter-reporter assay to identify alterations of *Trappc9* expression in the presence of different regulatory elements using cultured primary neurons as an *in vitro* model for brain tissue and fibroblasts as a control for non-imprinted peripheral tissue. The element responsible for regulating transcription of *Trappc9* would likely have specificity for the *Trappc9* promoter and so first I identified this promoter region from the UCSC genome browser, locating it around the 1st exon of *Trappc9* where we could see it showed typical promoter chromatin features. I aimed to clone the *Trappc9* promoter located at position chr15: 73,060,204-73,061,805, in the GRCm38/mm10 mouse sequence, into the firefly luciferase-encoding pGL4.23 [luc2/minP] vector, replacing the minimal promoter provided with the vector via restriction digest and ligation. The *Trappc9* promoter contained two regions of concern: an upstream dinucleotide repeat sequence, comprised of >100bp of GT and GA repeats, and a downstream splice donor site that begins at the end of *Trappc9* exon 1 and extends into the intron by ~200-300 bp. The potential issue with the former region is that it may have a repressive function, while the latter if included, has the potential to cause the transcript to splice out before going into the *Luc2* open reading frame. To determine whether these two sites might have an impact on the reporter-gene assay we designed four constructs of the *Trappc9* promoter amplified using two forward and two reverse primers labelled F1, F2, R1 and R2 to produce 4 amplicons of different lengths (positions 73,061,805–73,060,204 bp, 73,061,805–73,060,975 bp, 73,061,418–73,060,204 bp and 73,061,418–73,060,975 bp). The F2 primer avoided the upstream dinucleotide GT/ GA

repeat sequence while the R1 primer was located in exon 1 just before the splice donor site (Fig 5.2).

These candidate promoter fragments were then tested in a preliminary promoter-reporter gene assay by constructing four separate pGL4.23 Luciferase plasmids, each containing one of the four *Trappc9* promoter fragments in lieu of the minimal promoter that the original vector contained. Once the four constructs were established, they were co-transfected, alongside a Renilla vector at a 100:1 ratio, into either primary neurons, grown in culture for a week after being isolated from the hippocampus of newborn C57BL/6 mice or mouse embryonic fibroblasts as a control tissue. After 48 hours post-transfection, the samples were lysed and analysed for their luciferase activity compared to the original minimal promoter pGL4.23 plasmid. The results of this preliminary assay (Fig 5.2) showed us that all *Trappc9* promoter fragments increased the expression of luciferase and possess strong regulatory activity compared to that of the original Pgl minimal promoter (control) vector. However, of the four fragments, the F1/R2 amplicon, containing both the dinucleotide repeat sequence and the splice donor site, and the F2/R2 containing the exon 1 splice donor site had the lowest levels of luciferase activity. Of the two remaining constructs, the F1/R1 amplicon containing the dinucleotide repeat sequence showed the 2nd highest luciferase activity, although this was still lower than without either problematic region, indicating it still likely possesses a repressive function, while the F2/R1 construct containing neither of the two regions showed the highest level of luciferase activity. This data highlights that the exon 1 splice donor site appears to have the strongest impact on reducing luciferase activity while the dinucleotide repeat sequence also appears to reduce the activity, although to a lesser extent than that of the splice donor site. With this in mind, when designing the regulatory

element luciferase constructs, the F2/R1 *Trappc9* promoter construct was chosen as the promoter to replace the original minimal promote in the pGL4.23 [luc2/minP] vector.

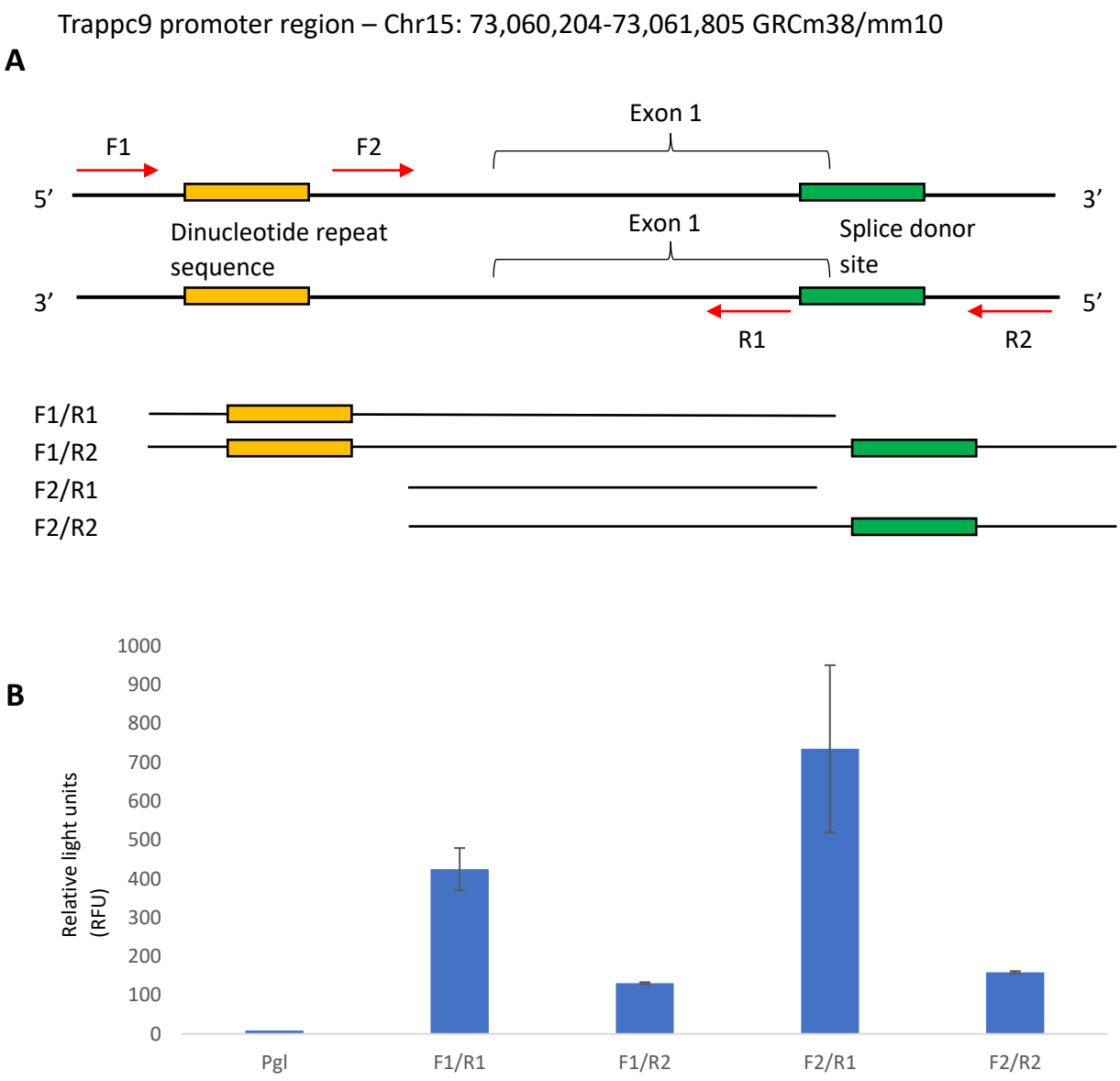


Fig 5.2: **(A)** Schematic of the four *Trappc9* promoter constructs that were generated and tested for activity. Two forward primers: F1 & F2, and two reverse primers: R1 & R2 were used in order to test the two regions of the promoter that had the potential to cause complications in the reporter gene assay: a dinucleotide repeat sequence and the exon 1 splice donor site. The four amplicons produced were of different lengths and contained one, both or neither of the two sites. **(B)** All four promoter amplicons were tested in a preliminary reporter-gene assay using a pGL4.23 vector with the minimal promoter replaced with one of the four generated *Trappc9* promoters. The original pGL4.23 [luc2/minP] was also used as a control. Experimental constructs were Co-transfected alongside a Renilla vector at a 100:1 firefly: Renilla ratio, into hippocampal primary neurons and analysed for luciferase activity 48 hours post-transfection.

5.3.2 Testing regulatory element reporter constructs

From the preliminary reporter gene assay, we decided to use the pGL 4.23 vector where the minimal promoter had been replaced with the F2/R1 *Trappc9* promoter fragment upstream of and including non-coding exon 1 consisting of 444 bp. The size of this promoter fragment is in line with standards from high throughput testing of promoter-enhancer interactions in the mouse genome (Martinez-Ara et al., 2022). The seven regulatory elements identified were amplified using PCR and initially cloned into TOPO®-plasmids. The regulatory elements were then extracted via restriction digest and cloned into the pGL 4.23 vector containing the F2/R1 *Trappc9* promoter fragment. We positioned the candidate regulatory elements depending on their relative locations within the Trappc9 locus, attempting to mimic the genomic architecture as closely as possible. Most of the candidate regulatory elements were positioned downstream of the *Luc2* reporter gene at the BamHI site to reflect the same relative orientation to the Trappc9 promoter as in the genome. Only the Reg-E element was cloned upstream of the Trappc9 promoter to maintain its original orientation as shown in Figure 2.1.

We performed these reporter gene assays in cultures of mouse primary hippocampal neurons and embryonic fibroblasts. Compared to the F2/R1 promoter-only construct, regulatory elements Reg-B and Reg-E had significant silencing effects in fibroblasts, but not neurons, indicating a tissue-/cell type-specific function (Fig 5.3). Reg-D displayed silencing activity in both neurons and fibroblasts. Unexpectedly, none of the regulatory elements showed enhancer activity in our assay conditions. All three silencing elements are located on the Trappc9-proximal side of the Peg13 DMR and CTCF-binding site. These silencer elements might contribute to the regulation of tissue-specific expression of *Trappc9, Chrac1*

and/or *Ago2 in vivo*. Reg-D might also contribute specifically to the reduced transcription of their paternal alleles in the brain, thereby generating an imprinted expression bias in this tissue, although any allele-specific mechanism remains to be elucidated.

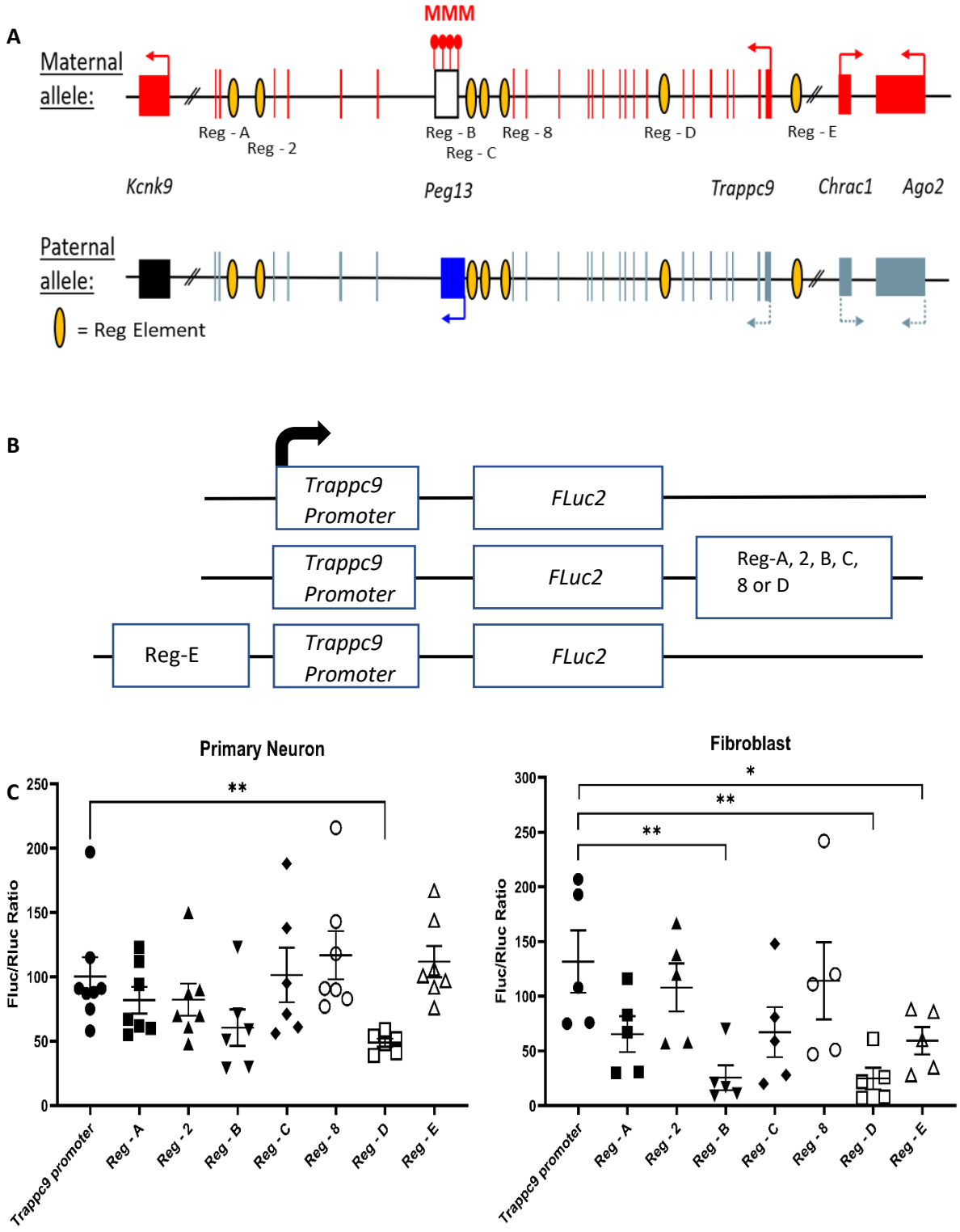


Figure 5.3: Promoter-reporter gene assays indicate silencer elements for *Trappc9*. **(A)** Scheme of the murine *Peg13 – Kcnk9 – Trappc9* cluster of imprinted genes. For *Trappc9* all introns are shown; for *Chrac1*, *Ago2* and *Kcnk9* introns have been omitted. Promoters, transcriptional activity and directions are indicated by arrows. Also indicated are the approximate locations of the candidate regulatory elements (listed in Table 5.1) across *Trappc9* introns. Not to scale. **(B)** Schematic of the promoter/enhancer constructs used for transfection of primary cells. Enhancers were positioned downstream or upstream of the promoter-reporter gene cassette depending on their relative locations within the *Trappc9* gene and in the same orientation. **(C)** Reporter gene activity of the constructs in cultures of primary hippocampal neurons from newborn mice and primary embryonic fibroblasts. Normalised activity of Firefly luciferase (FLuc) to co-transfected Renilla luciferase (RLuc) is shown. Mean values ± S.E.M. are indicated. *p ≤ 0.05; ** p ≤ 0.01.

5.4 Discussion

The presence of CTCF-cohesin binding sites, chromatin looping, and enhancer interactions has previously been shown in the human *Peg13-Kcnk9* imprinting cluster for both the *Peg13* and *Kcnk9* gene promoters by Court et al (2014) (fig 1.8). Furthermore, the binding of CTCF to the unmethylated paternal allele of the *Peg13* gDMR has been shown to be conserved in mice (Prickett et al., 2013; P. Singh et al., 2011). However, there is a possibility that tissue-specific enhancer elements and chromatin looping may differ in this species and could underly the imprinted expression of *Trappc9*, *Chrac1* and *Ago2* in murine brain tissue.

When using the ENCODE3 database to search for candidate regulatory elements potentially responsible for the regulation of imprinted expression of these genes, we considered active histone and open chromatin marks that resulted in the identification of 7 brain-specific potential regulatory regions (table 5.1 & fig 5.3A). Unexpectedly, however, these identified potential regulatory elements did not show the expected upregulation typical of enhancer function when tested in transfected primary neurons or fibroblasts. On the contrary, two of the elements (Reg-B and Reg-E) showed silencing activity specifically in fibroblasts, while a third element (Reg-D) showed silencing activity in both primary neurons and fibroblasts (fig 5.3). Active enhancers are characteristically associated with active chromatin marks H3K27ac, H3K4me1 and H3K9ac, while silencers on the other hand, are variably marked by H4K20me, H3K9me3 (typical for heterochromatin and methylated DNA) and/or H3K27me3, although chromatin at silencer regions is still expected to be open for binding of repressive transcription factors and, therefore is also associated with H3K79me2 and H3K36me3 marks (B. Pang & Snyder, 2020). While silencers are occasionally associated with these marks, there is still no widely accepted consensus on silencer-associated chromatin signatures and

previous studies have even shown that silencers may be bifunctional elements, acting through various mechanisms (Segert et al., 2021).

Furthermore, a functional study of the ENCODE3 platform performed by Martinez-Ara et al (2022) identified that the majority of annotated candidate cis-regulatory elements had no impact on transcription while comparable numbers of the remaining elements exhibited enhancer or repressor activity respectively. This was an unexpected result given that cCREs are usually predicted to be enhancers.

While the results obtained from my promoter-reporter assay showed unexpected silencing activity in several of the regulatory elements identified, further experiments are required to determine whether the silencer elements I identified within the *Trappc9* locus might function in an allele-specific way and contribute towards the brain-specific imprinted expression bias of this gene. In order to better understand the role of enhancer mediated regulation identified at this locus, future experiments, such as circular chromosome conformation capture (4c) or Hi-C, could be used to identify distal chromatin interactions with promoter regions of genes within the cluster. Previous studies have used these techniques to identify allele-specific sub-TAD organisation, correlated with CTCF insulator binding, at the *Dlk1-Dio3*, *Igf2/H19* (Llères et al., 2019) and the Growth factor receptor-bound protein 10 (*Grb10*) and dopa decarboxylase (*Ddc*) locus (Juan et al., 2022) imprinted clusters in mouse cells. These techniques would allow us to assess allele-specific TAD boundaries and distal chromatin interactions which would lead to identification of specific enhancer sequences/ locations that are responsible for regulating imprinted expression at the *Peg13/Kcnk9* cluster in mice.

Chapter 6 – Discussion and conclusions

6.1 Where we are now

In this thesis, I have investigated numerous factors related to the regulation of allele-specific expression promoted by genomic imprinting for genes located within the mouse chromosome 15 *Peg13/Kcnk9* cluster. The importance of understanding the regulation of allele-specific expression within this cluster is due to its affiliation with human disease and associated phenotypes such as that of the non-syndromic intellectual disability discussed earlier in this thesis. Developing a deeper understanding of transcriptome behaviour, and regulatory factors, in different cell types for these imprinted genes, helps us understand the mechanism behind their establishment and may lead to potential new avenues of research into possible treatment options. My first aim was to confirm the bulk tissue-specific imprinting bias previously discussed in the literature regarding *Ago2, Chrac1*, and *Trappc9* expression (Court et al., 2014; Perez et al., 2015). From this data, I identified a preference towards maternal expression of these three genes in whole brain tissue lysates, while peripheral tissues exhibited expected biallelic expression. Additionally, I observed that both *Kcnk9* and *Peg13* displayed non-tissue-specific maternal and paternal preferential expression, respectively, in whole brain and kidney tissues, with *Peg13* exhibiting an equal level of paternal-specific expression in bulk NSC samples as well. However, through my bulk-tissue analysis, I identified that NSCs do not appear to display the same imprinting expression pattern as whole brain tissue lysates (with the exception of *Peg13*), as *Ago2* and *Trappc9* both displayed biallelic/mild maternal biased expression in hippocampus-derived NSCs. This lack of imprinted expression in NSCs compared to whole brain lysate may indicate a potential role for biallelic expression in stem cell maintenance, similar to that of

the *Dlk1* gene in NSCs (Lozano-Ureña et al., 2017). Additionally, I identified that the promoter-proximal CGI regions for each gene displayed expected methylation frequencies that were in line with previous literature, identifying all CGIs as being hypomethylated, with the exceptions of *Trappc9* CGI2, located proximal to the 2nd exon of the *Trappc9* gene, and the *Peg13* maternal allele, which were both hypermethylated.

While bulk RNA sequencing has been used to identify the majority of imprinted gene expression, it has several limitations, including poor detection of lowly expressed genes, generalisation of allele-specific expression when performed on bulk-tissue samples, and data storage/processing issues. The idea of classifying specific tissues as displaying imprinted gene expression by utilising a bulk tissue sample for RNAseq is likely a generalisation and not necessarily reflective of the transcriptional patterns of the individual cells from which the tissue is made. This raises the question of how allele-specific expression is reflected in individual cells of multiple tissues and whether there is variability and mosaicism within the cell population. While single-cell imprinting analysis is a relatively new field of study that has been used to identify tissue and cell-specific imprinting patterns (Santoni et al., 2017; Varrault et al., 2020), only a limited number of these experiments have utilised hybrid mouse models to categorise allele expression ratios in single-cell populations.

To address this question of single-cell imprinting at the *Peg13/Kcnk9 cluster*, I utilised the Sc-GEM method to identify allelic expression ratios in individual NSCs and *in vitro* differentiated neurons from hybrid newborn mouse hippocampus tissue, where these genes show higher levels of expression compared to alternative brain regions (Pulix, 2017). This data highlighted that there is a high degree of variability in allelic expression for single cells, with *Ago2* and *Trappc9* exhibiting a range of mono-allelic paternal, equal biallelic, mono-

allelic maternal, and biased biallelic transcriptional states in both cell types. While *Peg13* displayed a majority of single cells with either mono-allelic or preferential paternal expression, which is in line with the bulk tissue data and previous literature, some cells did exhibit biallelic (in NSCs), preferential maternal, or even mono-allelic maternal expression patterns, indicating a higher degree of complexity for imprinted expression than initially considered. Additionally, while the single-cell expression data collected in this thesis identified variable allele-specific expression states for genes within the cluster, the methylation frequency at promoter-proximal CGIs from single cells, analysed using the methylation sensitive BstUI restriction enzyme, appeared to be mostly in line with that of bulk cell data. This is shown by the bulk cell data displaying a lack of methylation at the *Ago2* promoter-proximal CGI which correlates to the hypomethylation seen in single differentiated neuron cells for the *Ago2* CGI. Unfortunately, this does not explain the reason as to why the single NSC data exhibit a higher degree of methylation which differs from the bulk cell data. However, both *Trappc9* CGI's display the same result for both bulk and single-cell methylation data with CGI1 being hypomethylated in NSC's and differentiated neurons and CGI2 being hypermethylated. Furthermore, while the *Peg13* single-cell data is unable to differentiate between the maternal and paternal allele due to the differential methylation at the CGI, both the bulk and single-cell analysis indicated a high methylation frequency at this CGI which is in line with previous literature and provides a level of confidence in my results.

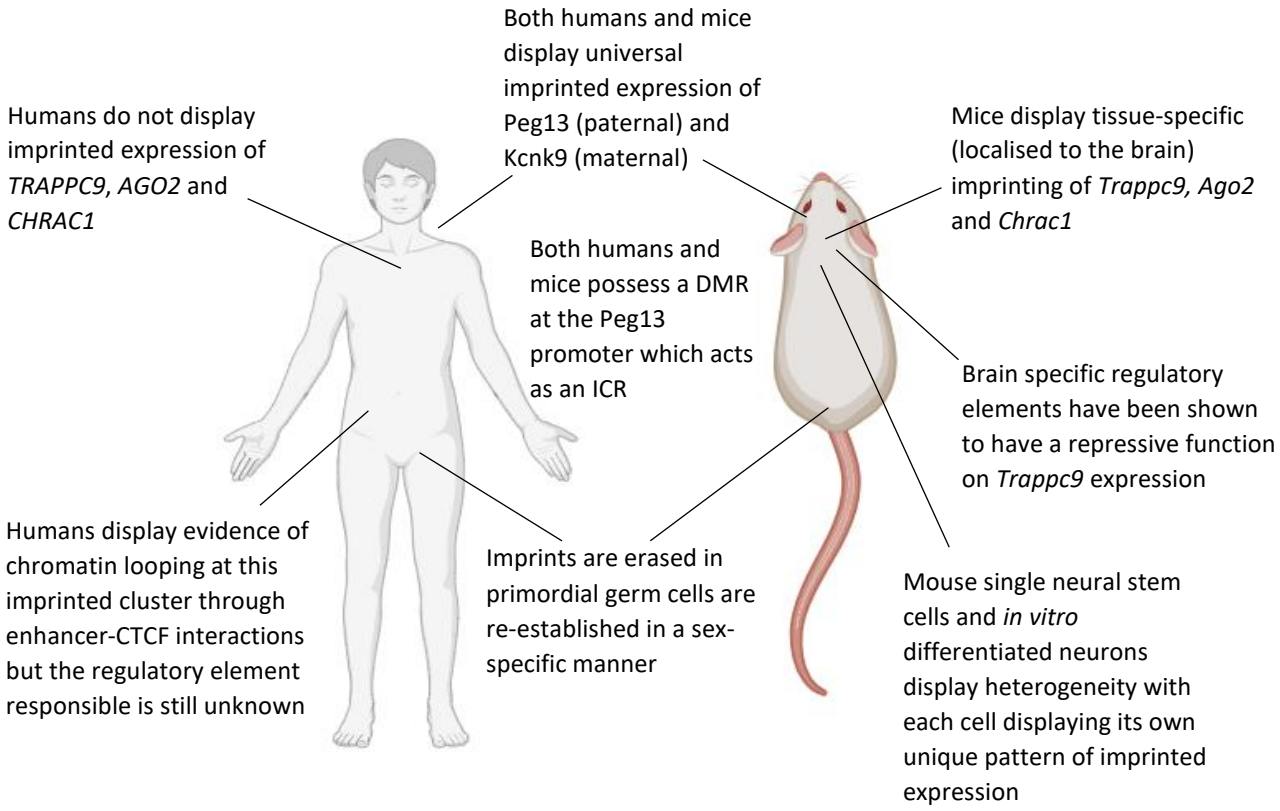


Figure 6.1: An illustration displaying some of the key similarities and differences between human and mouse regarding imprinted gene expression. In both species, imprinting patterns are established in the primordial germ cells in a sex-specific manner, both possess a DMR at the *Peg13* promoter proximal CGI with hypomethylation of the maternal allele and hypermethylation of the paternal allele. Additionally, in both, the *Peg13* and *Kcnk9* genes are paternally and maternally expressed respectively. However, in humans, *AGO2, TRAPPC9* and *CHRAC1* are not imprinted and display biallelic expression while in mice they are imprinted exclusively in brain tissue. Additionally, we have identified that single neural and stem cells and differentiated neurons display unique variable imprinted patterns within mice. Furthermore, the human *PEG13-TRAPPC9* imprinted cluster shows evidence of enhancer/chromatin looping, and while chromatin interaction experiments for this cluster have not been performed in mice, the regulatory elements surrounding the *Trappc9* gene in mice exhibit a repressive function, possibly to downregulate expression of the paternal *Trappc9*.

Although the mechanism responsible for the variations in transcriptional states has yet to be elucidated, previous studies have suggested that regulation of imprinted expression can be dynamic as a result of trans-acting factors and alterations of chromatin architecture without changes in the methylation state of the ICR (Scagliotti et al., 2021; van Ekelenburg et al.,

2023). Transient expression of the "silenced" allele generated by genomic imprinting may result from alterations of chromatin state, potentially due to dynamic binding of CTCF and establishment of temporary TAD boundaries. Live-cell imaging of ES cells has identified that CTCF-CTCF looping is relatively rare, with approximately 3-6% of loci forming loops at any given time in the cell population, with each loop lasting ~30 minutes, highlighting that these chromatin arrangements are constantly being generated and dismantled (Dehingia et al., 2022; Gabriele et al., 2022). In addition, chromatin topology is suggested to have an important role in imprinted gene expression and silencing, with the inactive X chromosome localising specifically to the nuclear periphery or perinucleolar region, which has been strongly associated with transcriptional repression and imprinted clusters seemingly undergoing changes in nuclear localisation and replication timing (Macdonald et al., 2015). Studies have suggested that this chromatin topology and nuclear architecture is dynamic and can alter upon differentiation and between cell types, generating variability in imprinted gene expression within a cell population. An example of this regulation can be found in the *H19* and *Kcnq1ot1* imprinted domains where *Kcnq1ot1* is divided into three TADs while *H19* is located in a separate TAD within mouse and human ESCs. However, in differentiated cells from the mouse brain the *Kcnq1ot1* imprinted domain and neighbouring *H19* domain become one TAD suggesting a dynamic shift in chromatin topology between these imprinted domains upon differentiation (Macdonald et al., 2015). Due to the hypothesised dynamic nature of imprinting gene regulation, CTCF binding, and chromatin topology, it is possible that the transcripts identified from the single cells I isolated and analysed were only a snapshot of certain molecular dynamics for each individual cell at the time they were lysed. This proposed hypothesis would explain the mosaic of alternate

expression ratios between the maternal and paternal alleles that I observed in cells isolated from the same brain region and grown in identical conditions.

Finally, although I was unsuccessful at identifying a regulatory element that promoted upregulation of the *Trappc9* gene, I was able to identify several elements that produced a silencing effect on *Trappc9* expression in both primary neurons, extracted from hippocampus tissue and grown *in vitro*, and in fibroblasts. Further experimentation is required to deduce whether these silencer elements have an impact on allele-specific expression and how they may contribute to the brain-specific imprinted expression bias of this gene.

6.2 Limitations

While the data obtained in this thesis provides evidence for variable imprinted expression in single-cell populations, which formulates a more complex view of imprinting regulation, it is limited by the small sample size of NSCs and differentiated neuron populations used. The original Sc-GEM method used by Cheow et al (2015) utilised C1 Fluidigm chips, which measures readouts of 1500 cells, generating large sample sizes for data analysis. However, due to multiple factors, including practical limitations, time constraints, and limited access to software and training, the single-cell isolations used for data generation in this thesis were manually isolated and underwent standard qPCR and Sanger sequencing procedures. Furthermore, the impact on the transcriptome from maintaining single cell samples in lysis buffer placed on ice was not investigated in this thesis. There are ~30 minutes that pass between the first cell being isolated to the last cell and the impact of the stress response on transcriptome behaviour for the cells isolated towards the beginning of the procedure has not been measured. Possible effects of cell stress include a reduction in RNA templates due

to RNA degradation, changes in gene expression as a response to the environmental stress, such as downregulation of non-essential genes and upregulation of stress response genes, and variability in my data e.g. perhaps the first cell isolated will have the greatest deviation in expression compared to the last cell. In the future identifying whether the genes I am analysing are impacted by the stress response and determining whether there is an impact on gene expression based on time spent in lysis buffer/ice could be investigated to determine whether it affects the results seen from the qPCR analysis.

Another factor to consider is the impact of using hybrid crosses to analyse imprinted gene expression. Previous studies have identified alterations in imprinted gene expression, such as that of Kcnq1 located on chromosome 7 when using inter sub-species hybrids such as C57BL/6 x Cast/EiJ crosses (Korostowski et al., 2012). They found that early maternal expression of *Kcnq1* is higher from the CAST/EiJ cross than from the C57BL/6J cross and it continues to have a relatively higher expression level throughout development. Additionally, they identified that this difference in expression was not contingent on differential DNA methylation. Furthermore, they identified that a developmental switch from mono- to biallelic expression was delayed in (Cast/EiJ × C57BL/6J)F_1 hybrids compared with (C57BL/6J × Cast/EiJ)F_1 hybrids (Korostowski et al., 2012). While it is unknown how impactful the hybrid cross that I used to generate data for imprinted gene expression was, it is safe to assume that further experiments, utilising additional hybrid crosses, would be beneficial for future imprinting studies to determine whether this may impact the results.

Due to the small amount of starting material used to generate the single-cell data, both expression and methylation analyses could be impacted, although it is more likely to impact the methylation analysis due to the presence of multiple copies of mRNA per cell as

opposed to the singular copy of genomic DNA. When analysing genomic DNA on a single-cell level, such as for methylation analysis of CGIs, allele dropout is a concern due to there being only two templates per cell from the maternal and paternal chromosomes. This reduces the efficiency of PCR-based techniques and may produce inaccurate results for methyl presence at specific genomic loci. Furthermore, due to the inability of the assay to distinguish between heterozygous methylated alleles and homozygous methylated alleles, it is less informative about methylation status for DMRs such as the methylated maternal *Peg13* and the unmethylated paternal *Peg13*. To address these issues, future experiments, such as single-cell Targeted Analysis of the Methylome (ScTAM-seq), which can analyse thousands of cells with a low dropout rate (~7%), may be useful for categorising DNA methylation for the *Peg13/Kcnk9* locus in future single-cell experiments (Bianchi et al., 2022). Analysing transcriptional data on a single-cell level is less likely to suffer from allele dropout due to the multiple copies of mRNA within a cell that can be targeted for reverse transcription and amplification. However, single-cell transcriptional analysis via qPCR faces its own complications such as potential PCR bias and variability due to technical factors such as the efficiency of reverse transcription and amplification. This can make it difficult to compare gene expression between different cells and is an important consideration given the Sc-GEM protocol possesses both a multiplex preamplification step after the methylation sensitive restriction digest to produce sufficient PCR product for downstream analysis and a locus-specific qPCR amplification.

An additional limitation could be the generation of allele-specific transcriptional variants from the imprinted genes being analysed. Although no transcriptional variants of Trappc9 were identified in NSCs from this project (Claxton et al., 2022), the concept of imprinted genes having transcriptional isoforms that undergo alternate imprinted regulation may

impact the single-cell transcriptional data. This imprinted isoform theory is still undergoing extensive research and further experiments may yield a better understanding of how these transcriptional variants are generated and to what degree they may impact imprinted expression data (Stelzer et al., 2015).

6.3 Future experiments

While the data obtained in this thesis helps build a foundation for understanding the increased complexity of single-cell imprinting dynamics compared to bulk tissue analysis, it is not sufficient to make definitive conclusions about imprinted transcriptional behaviour. Future experiments should focus on expanding both the sample size and cell types analysed for the *Peg13/Kcnk9* cluster and whether other imprinting clusters that display imprinted tissue specificity exhibit similar patterns of mosaicism when single cells from the imprinted issue are analysed. Additionally, the analysis of single cells directly separated out of brain tissue instead of using isolated cells that were then cultured into NSCs / *in vitro* differentiated neurons, as was performed in this thesis, would be an interesting experiment to determine whether there are any differences in expression/ methylation profiles of primary isolated cells vs those that have been cultured and passaged. Furthermore, the use of alternate single-cell RNA sequencing methods such as single-nucleus transcriptional profiling by combinatorial indexing (B. K. Martin et al., 2023) would be a key experiment in identifying any limitations of the Sc-GEM procedure and for validation that the results obtained in this thesis are consistent when using alternate profiling techniques.

Regarding the molecular mechanism that regulates imprinted expression at the *Peg13/Kcnk9* imprinted cluster, the use of circular chromosome conformation capture (4C) or Hi-C would be a key experiment in expanding our understanding of any enhancer-

mediated upregulation specific to this locus. These techniques would identify chromatin looping and organisation by cross-linking a known genomic region (the promoter of the gene being analysed) and any unknown proximal chromatin interactions. This would help to identify the sequence and location of any distal regulatory elements responsible for the regulation of gene expression, including enhancers and silencers. Recent studies have begun using Hi-C to identify these distal regulatory elements in multiple cell types (Llères et al., 2019), and new software is consistently being developed to analyse large quantities of Hi-C data to categorise thousands of promoter-enhancer interactions across the genome (Montefiori et al., 2018; Ron et al., 2017). The creation of these chromatin interaction databases would have a significant impact on identifying enhancer-mediated regulation at multiple imprinted clusters and would likely expand our understanding of imprinting regulation. Furthermore, the use of chromatin immunoprecipitation for CTCF binding may be useful in determining dynamic CTCF binding regulation of allele-specific expression in multiple cell types at different time points.

Bibliography

Abascal, F., Acosta, R., Addleman, N. J., Adrian, J., Afzal, V., Aken, B., Akiyama, J. A., Jammal, O. al, Amrhein, H., Anderson, S. M., Andrews, G. R., Antoshechkin, I., Ardlie, K. G., Armstrong, J., Astley, M., Banerjee, B., Barkal, A. A., Barnes, I. H. A., Barozzi, I., … Zimmerman, J. (2020). Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature 2020 583:7818*, *583*(7818), 699–710. https://doi.org/10.1038/s41586-020-2493-4

Aboud, N. M. al, Tupper, C., & Jialal, I. (2022). Genetics, Epigenetic Mechanism. *StatPearls*. https://www.ncbi.nlm.nih.gov/books/NBK532999/

Alabdullah, A. A., Al-Abdulaziz, B., Alsalem, H., Magrashi, A., Pulicat, S. M., Almzroua, A. A., Almohanna, F., Assiri, A. M., Al Tassan, N. A., & Al-Mubarak, B. R. (2019). Estimating transfection efficiency in differentiated and undifferentiated neural cells. *BMC Research Notes*, *12*(1), 1–7. https://doi.org/10.1186/S13104-019-4249-5/FIGURES/3

Alaskhar Alhamwe, B., Khalaila, R., Wolf, J., Bülow, V., Harb, H., Alhamdan, F., Hii, C. S., Prescott, S. L., Ferrante, A., Renz, H., Garn, H., & Potaczek, D. P. (2018). Histone modifications and their role in epigenetics of atopy and allergic diseases. *Allergy, Asthma, and Clinical Immunology : Official Journal of the Canadian Society of Allergy and Clinical Immunology*, *14*(1). https://doi.org/10.1186/S13223-018-0259-4

Albrecht, U., Sutcliffe, J. S., Cattanach, B. M., Beechey, C. v., Armstrong, D., Eichele, G., & Beaudet, A. L. (1997). Imprinted expression of the murine Angelman syndrome gene, Ube3a, in hippocampal and Purkinje neurons. *Nature Genetics*, *17*(1), 75–78. https://doi.org/10.1038/NG0997-75

Alisch, R. S., Jin, P., Epstein, M., Caspary, T., & Warren, S. T. (2007). Argonaute2 Is Essential for Mammalian Gastrulation and Proper Mesoderm Formation. *PLoS Genetics*, *3*(12), 2565–2571. https://doi.org/10.1371/JOURNAL.PGEN.0030227

Allach El Khattabi, L., Backer, S., Pinard, A., Dieudonné, M. N., Tsatsaris, V., Vaiman, D., Dandolo, L., Bloch-Gallego, E., Jammes, H., & Barbaux, S. (2019). A genome-wide search for new imprinted genes in the human placenta identifies DSCAM as the first imprinted gene on chromosome 21. *European Journal of Human Genetics*, *27*(1), 49. https://doi.org/10.1038/S41431-018-0267-3

Amin, M., Vignal, C., Eltaraifee, E., Mohammed, I. N., Hamed, A. A. A., Elseed, M. A., Babai, A., Elbadi, I., Mustafa, D., Abubaker, R., Mustafa, M., Drunat, S., Elsayed, L. E. O., Ahmed, A. E., Boespflug-Tanguy, O., & Dorboz, I. (2022). A novel homozygous mutation in TRAPPC9 gene causing autosomal recessive non-syndromic intellectual disability. *BMC Medical Genomics*, *15*(1), 1–5. https://doi.org/10.1186/S12920-022-01354-1/FIGURES/2

Andergassen, D., Dotter, C. P., Wenzel, D., Sigl, V., Bammer, P. C., Muckenhuber, M., Mayer, D., Kulinski, T. M., Theussl, H. C., Penninger, J. M., Bock, C., Barlow, D. P., Pauler, F. M., & Hudson, Q. J. (2017). Mapping the mouse Allelome reveals tissue-specific regulation of allelic expression. *ELife*, *6*. https://doi.org/10.7554/ELIFE.25125

Anderlid, B. M., Lundin, J., Malmgren, H., Lehtihet, M., & Nordgren, A. (2014). Small mosaic deletion encompassing the snoRNAs and SNURF-SNRPN results in an atypical Prader-Willi syndrome phenotype. *American Journal of Medical Genetics, Part A*, *164*(2), 425–431. https://doi.org/10.1002/AJMG.A.36307

Angelman, H. (1965). 'Puppet' Children A Report on Three Cases. *Developmental Medicine & Child Neurology*, *7*(6), 681–688. https://doi.org/10.1111/J.1469-8749.1965.TB07844.X

Angulo, M. A., Butler, M. G., & Cataletto, M. E. (2015). Prader-Willi syndrome: a review of clinical, genetic, and endocrine findings. *Journal of Endocrinological Investigation*, *38*(12), 1249–1263. https://doi.org/10.1007/S40618-015-0312-9

Arnold, P. R., Wells, A. D., & Li, X. C. (2020). Diversity and Emerging Roles of Enhancer RNA in Regulation of Gene Expression and Cell Fate. *Frontiers in Cell and Developmental Biology*, *7*, 377. https://doi.org/10.3389/FCELL.2019.00377/BIBTEX

Aslanger, A. D., Goncu, B., Duzenli, O. F., Yucesan, E., Sengenc, E., & Yesil, G. (2022). Biallelic loss of TRAPPC9 function links vesicle trafficking pathway to autosomal recessive intellectual disability. *Journal of Human Genetics*, *67*(5), 279–284. https://doi.org/10.1038/S10038-021-01007-8

Babak, T., Deveale, B., Tsang, E. K., Zhou, Y., Li, X., Smith, K. S., Kukurba, K. R., Zhang, R., Li, J. B., van der Kooy, D., Montgomery, S. B., & Fraser, H. B. (2015a). Genetic conflict reflected in tissue-specific maps of genomic imprinting in human and mouse. *Nature Genetics 2015 47:5*, *47*(5), 544–549. https://doi.org/10.1038/ng.3274

Bagde, S. R., & Christopher Fromme, J. (2022). Structure of a TRAPPII-Rab11 activation intermediate reveals GTPase substrate selection mechanisms. *Science Advances*, *8*(19), 7446. https://doi.org/10.1126/SCIADV.ABN7446/SUPPL_FILE/SCIADV.ABN7446_MOVIE_S1.ZIP

Bajrami, E., & Spiroski, M. (2016). Genomic Imprinting. *Open Access Macedonian Journal of Medical Sciences*, *4*(1), 181. https://doi.org/10.3889/OAMJMS.2016.028

Bannister, A. J., & Kouzarides, T. (2011). Regulation of chromatin by histone modifications. *Cell Research 2011 21:3*, *21*(3), 381–395. https://doi.org/10.1038/cr.2011.22

Baran, Y., Subramaniam, M., Biton, A., Tukiainen, T., Tsang, E. K., Rivas, M. A., Pirinen, M., Gutierrez-Arcelus, M., Smith, K. S., Kukurba, K. R., Zhang, R., Eng, C., Torgerson, D. G., Urbanek, C., Li, J. B., Rodriguez-Santana, J. R., Burchard, E. G., Seibold, M. A., MacArthur, D. G., … Lappalainen, T. (2015). The landscape of genomic imprinting across diverse adult human tissues. *Genome Research*, *25*(7), 927. https://doi.org/10.1101/GR.192278.115

Barel, O., Shalev, S. A., Ofir, R., Cohen, A., Zlotogora, J., Shorer, Z., Mazor, G., Finer, G., Khateeb, S., Zilberberg, N., & Birk, O. S. (2008). Maternally inherited Birk Barel mental retardation dysmorphism syndrome caused by a mutation in the genomically imprinted potassium channel KCNK9. *American Journal of Human Genetics*, *83*(2), 193–199. https://doi.org/10.1016/J.AJHG.2008.07.010

Barlow, D. P., & Bartolomei, M. S. (2014). Genomic Imprinting in Mammals. *Cold Spring Harbor Perspectives in Biology*, *6*(2). https://doi.org/10.1101/CSHPERSPECT.A018382

Barlow, D. P., Stöger, R., Herrmann, B. G., Saito, K., & Schweifer, N. (1991). The mouse insulin-like growth factor type-2 receptor is imprinted and closely linked to the Tme locus. *Nature*, *349*(6304), 84–87. https://doi.org/10.1038/349084A0

Barrowman, J., Bhandari, D., Reinisch, K., & Ferro-Novick, S. (2010). TRAPP complexes in membrane traffic: Convergence through a common Rab. In *Nature Reviews Molecular Cell Biology* (Vol. 11, Issue 11, pp. 759–763). Nature Publishing Group. https://doi.org/10.1038/nrm2999

Bartolomei, M. S., & Ferguson-Smith, A. C. (2011). Mammalian genomic imprinting. *Cold Spring Harbor Perspectives in Biology*, *3*(7). https://doi.org/10.1101/cshperspect.a002592

Bartolomei, M. S., Zemel, S., & Tilghman, S. M. (1991). Parental imprinting of the mouse H19 gene. *Nature*, *351*(6322), 153–155. https://doi.org/10.1038/351153A0

Barton, S. C., Surani, M. A. H., & Norris, M. L. (1984). Role of paternal and maternal genomes in mouse development. *Nature*, *311*(5984), 374–376. https://doi.org/10.1038/311374A0

Battaglia, S., Dong, K., Wu, J., Chen, Z., Najm, F. J., Zhang, Y., Moore, M. M., Hecht, V., Shoresh, N., & Bernstein, B. E. (2022). Long-range phasing of dynamic, tissue-specific and allele-specific regulatory elements. *Nature Genetics 2022 54:10*, *54*(10), 1504–1513. https://doi.org/10.1038/s41588-022-01188-8

Beagan, J. A., Pastuzyn, E. D., Fernandez, L. R., Guo, M. H., Feng, K., Titus, K. R., Chandrashekar, H., Shepherd, J. D., & Phillips-Cremins, J. E. (2020). Three-dimensional genome restructuring across timescales of activity-induced neuronal gene expression. *Nature Neuroscience*, *23*(6), 707–717. https://doi.org/10.1038/S41593-020-0634-6

Beaudoin, G. M. J., Lee, S. H., Singh, D., Yuan, Y., Ng, Y. G., Reichardt, L. F., & Arikkath, J. (2012). Culturing pyramidal neurons from the early postnatal mouse hippocampus and cortex. *Nature Protocols 2012 7:9*, *7*(9), 1741–1754. https://doi.org/10.1038/nprot.2012.099

Bell, A. C., & Felsenfeld, G. (2000). Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. *Nature*, *405*(6785), 482–485. https://doi.org/10.1038/35013100

Berger, S. L., Kouzarides, T., Shiekhattar, R., & Shilatifard, A. (2009). An operational definition of epigenetics. *Genes & Development*, *23*(7), 781. https://doi.org/10.1101/GAD.1787609

Bianchi, A., Scherer, M., Zaurin, R., Quililan, K., Velten, L., & Beekman, R. (2022). scTAM-seq enables targeted high-confidence analysis of DNA methylation in single cells. *Genome Biology*, *23*(1), 1–17. https://doi.org/10.1186/S13059-022-02796-7/FIGURES/3

Bird, A. (2007). Perceptions of epigenetics. *Nature*, *447*(7143), 396–398. https://doi.org/10.1038/NATURE05913

Blanc, R. S., & phane Richard, S. (2017). *Molecular Cell Review Arginine Methylation: The Coming of Age*. https://doi.org/10.1016/j.molcel.2016.11.003

Bonduriansky, R. (2007). The genetic architecture of sexual dimorphism: the potential roles of genomic imprinting and condition-dependence. *Sex, Size and Gender Roles: Evolutionary Studies of Sexual Size Dimorphism*. https://doi.org/10.1093/ACPROF:OSO/9780199208784.003.0020

Bonthuis, P. J., Huang, W. C., Stacher Hörndli, C. N., Ferris, E., Cheng, T., & Gregg, C. (2015). Noncanonical genomic imprinting effects in offspring. *Cell Reports*, *12*(6), 979–991. https://doi.org/10.1016/j.celrep.2015.07.017

Bourc'his, D., & Bestor, T. H. (2006). Origins of extreme sexual dimorphism in genomic imprinting. *Cytogenetic and Genome Research*, *113*(1–4), 36–40. https://doi.org/10.1159/000090813

Bourque, G., Burns, K. H., Gehring, M., Gorbunova, V., Seluanov, A., Hammell, M., Imbeault, M., Izsvák, Z., Levin, H. L., Macfarlan, T. S., Mager, D. L., & Feschotte, C. (2018). Ten things you

should know about transposable elements. *Genome Biology 2018 19:1*, *19*(1), 1–12. https://doi.org/10.1186/S13059-018-1577-Z

Bouschet, T., Dubois, E., Reynès, C., Kota, S. K., Rialle, S., Maupetit-Méhouas, S., Pezet, M., le Digarcher, A., Nidelet, S., Demolombe, V., Cavelier, P., Meusnier, C., Maurizy, C., Sabatier, R., Feil, R., Arnaud, P., Journot, L., & Varrault, A. (2017). In Vitro Corticogenesis from Embryonic Stem Cells Recapitulates the In Vivo Epigenetic Control of Imprinted Gene Expression. *Cerebral Cortex (New York, N.Y. : 1991)*, *27*(3), 2418–2433. https://doi.org/10.1093/CERCOR/BHW102

Brandvain, Y., van Cleve, J., Úbeda, F., & Wilkins, J. F. (2011). Demography, kinship, and the evolving theory of genomic imprinting. *Trends in Genetics : TIG*, *27*(7), 251–257. https://doi.org/10.1016/J.TIG.2011.04.005

Brasset, E., & Vaury, C. (2005). Insulators are fundamental components of the eukaryotic genomes. *Heredity 2005 94:6*, *94*(6), 571–576. https://doi.org/10.1038/sj.hdy.6800669

Brunet, S., & Sacher, M. (2014). In Sickness and in Health: The Role of TRAPP and Associated Proteins in Disease. *Traffic*, *15*(8), 803–818. https://doi.org/10.1111/TRA.12183

Buiting, K. (2010). Prader–Willi syndrome and Angelman syndrome. *American Journal of Medical Genetics Part C: Seminars in Medical Genetics*, *154C*(3), 365–376. https://doi.org/10.1002/AJMG.C.30273

Buiting, K., Williams, C., & Horsthemke, B. (2016). Angelman syndrome — insights into a rare neurogenetic disorder. *Nature Reviews Neurology 2016 12:10*, *12*(10), 584–593. https://doi.org/10.1038/nrneurol.2016.133

Burgess-Beusse, B., Farrell, C., Gaszner, M., Litt, M., Mutskov, V., Recillas-Targa, F., Simpson, M., West, A., & Felsenfeld, G. (2002). The insulation of genes from external enhancers and silencing chromatin. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(SUPPL. 4), 16433–16437. https://doi.org/10.1073/PNAS.162342499/ASSET/E97CA23B-8C5F-4463-B170-92A3E52B962D/ASSETS/GRAPHIC/PQ1623424007.JPEG

Butler, M. G. (2020). Imprinting disorders in humans: a review. *Current Opinion in Pediatrics*, *32*(6), 719. https://doi.org/10.1097/MOP.0000000000000965

Butler, M. G., Lee, P. D. K., & Whitman, B. Y. (2006). Management of prader-willi syndrome: Third edition. *Management of Prader-Willi Syndrome: Third Edition*, 1–552. https://doi.org/10.1007/978-0-387-33536-0

Cai, H., Reinisch, K., & Ferro-Novick, S. (2007). Coats, Tethers, Rabs, and SNAREs Work Together to Mediate the Intracellular Destination of a Transport Vesicle. *Developmental Cell*, *12*(5), 671–682. https://doi.org/10.1016/J.DEVCEL.2007.04.005

Cai, H., Zhang, Y., Pypaert, M., Walker, L., & Ferro-Novick, S. (2005). Mutants in trs120 disrupt traffic from the early endosome to the late Golgi. *The Journal of Cell Biology*, *171*(5), 823. https://doi.org/10.1083/JCB.200505145

Cao, J., & Yan, Q. (2012). Histone ubiquitination and deubiquitination in transcription, DNA damage response, and cancer. *Frontiers in Oncology*, *2*(MAR), 26. https://doi.org/10.3389/FONC.2012.00026/BIBTEX

Cassidy, S. B., Schwartz, S., Miller, J. L., & Driscoll, D. J. (2012). Prader-Willi syndrome. *Genetics in Medicine : Official Journal of the American College of Medical Genetics*, *14*(1), 10–26. https://doi.org/10.1038/GIM.0B013E31822BEAD0

Cattanach, B. M., & Kirk, M. (1985). Differential activity of maternally and paternally derived chromosome regions in mice. *Nature*, *315*(6019), 496–498. https://doi.org/10.1038/315496A0

Chen, H., Tian, Y., Shu, W., Bo, X., & Wang, S. (2012). Comprehensive identification and annotation of cell type-specific and ubiquitous CTCF-binding sites in the human genome. *PloS One*, *7*(7). https://doi.org/10.1371/JOURNAL.PONE.0041374

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., … Ng, H. H. (2008). Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell*, *133*(6), 1106–1117. https://doi.org/10.1016/J.CELL.2008.04.043

Cheow, L. F., Courtois, E. T., Tan, Y., Viswanathan, R., Xing, Q., Tan, R. Z., Tan, D. S. W., Robson, P., Loh, Y. H., Quake, S. R., & Burkholder, W. F. (2016). Single-cell multimodal profiling reveals cellular epigenetic heterogeneity. *Nature Methods*, *13*(10), 833–836. https://doi.org/10.1038/nmeth.3961

Cheow, L. F., Quake, S. R., Burkholder, W. F., & Messerschmidt, D. M. (2015). Multiplexed locus-specific analysis of DNA methylation in single cells. *Nature Protocols*, *10*(4), 619–631. https://doi.org/10.1038/nprot.2015.041

Chess, A. (2016). Monoallelic Gene Expression in Mammals. *Annual Review of Genetics*, *50*, 317–327. https://doi.org/10.1146/ANNUREV-GENET-120215-035120

Chisholm, K. M., Wan, Y., Li, R., Montgomery, K. D., Chang, H. Y., & West, R. B. (2012). Detection of long non-coding RNA in archival tissue: correlation with polycomb protein expression in primary and metastatic breast carcinoma. *PloS One*, *7*(10). https://doi.org/10.1371/JOURNAL.PONE.0047998

Chojnacki, A., & Weiss, S. (2008). Production of neurons, astrocytes and oligodendrocytes from mammalian CNS stem cells. *Nature Protocols*, *3*(6), 935–940. https://doi.org/10.1038/nprot.2008.55

Claxton, M., Pulix, M., Seah, M. K. Y., Bernardo, R., Zhou, P., Aljuraysi, S., Liloglou, T., Arnaud, P., Kelsey, G., Messerschmidt, D. M., & Plagge, A. (2022). Variable allelic expression of imprinted genes at the Peg13, Trappc9, Ago2 cluster in single neural cells. *Frontiers in Cell and Developmental Biology*, *10*. https://doi.org/10.3389/FCELL.2022.1022422

Cleaton, M. A. M., Edwards, C. A., & Ferguson-Smith, A. C. (2014). Phenotypic outcomes of imprinted gene models in mice: elucidation of pre- and postnatal functions of imprinted genes. *Annual Review of Genomics and Human Genetics*, *15*, 93–126. https://doi.org/10.1146/ANNUREV-GENOM-091212-153441

Coan, P. M., Burton, G. J., & Ferguson-Smith, A. C. (2005). Imprinted genes in the placenta--a review. *Placenta*, *26 Suppl A*(SUPPL.). https://doi.org/10.1016/J.PLACENTA.2004.12.009

Cooper, A., Butto, T., Hammer, N., Jagannath, S., Fend-Guella, D. L., Akhtar, J., Radyushkin, K., Lesage, F., Winter, J., Strand, S., Roeper, J., Zechner, U., & Schweiger, S. (2020). Inhibition of

histone deacetylation rescues phenotype in a mouse model of Birk-Barel intellectual disability syndrome. *Nature Communications 2020 11:1*, *11*(1), 1–14. https://doi.org/10.1038/s41467-019-13918-4

Court, F., Camprubi, C., Garcia, C. V., Guillaumet-Adkins, A., Sparago, A., Seruggia, D., Sandoval, J., Esteller, M., Martin-Trujillo, A., Riccio, A., Montoliu, L., & Monk, D. (2014). The PEG13-DMR and brain-specific enhancers dictate imprinted expression within the 8q24 intellectual disability risk locus. *Epigenetics & Chromatin*, *7*(1), 5. https://doi.org/10.1186/1756-8935-7-5

Cousin, M. A., Veale, E. L., Dsouza, N. R., Tripathi, S., Holden, R. G., Arelin, M., Beek, G., Bekheirnia, M. R., Beygo, J., Bhambhani, V., Bialer, M., Bigoni, S., Boelman, C., Carmichael, J., Courtin, T., Cogne, B., Dabaj, I., Doummar, D., Fazilleau, L., … Klee, E. W. (2022). Gain and loss of TASK3 channel function and its regulation by novel variation cause KCNK9 imprinting syndrome. *Genome Medicine*, *14*(1), 1–19. https://doi.org/10.1186/S13073-022-01064-4/FIGURES/5

Cox, G. F., Bürger, J., Lip, V., Mau, U. A., Sperling, K., Wu, B. L., & Horsthemke, B. (2002). Intracytoplasmic sperm injection may increase the risk of imprinting defects. *American Journal of Human Genetics*, *71*(1), 162–164. https://doi.org/10.1086/341096

Crouse, H. v. (1960). The Controlling Element in Sex Chromosome Behavior in Sciara. *Genetics*, *45*(10), 1429. https://doi.org/10.1093/GENETICS/45.10.1429

Crowley, J. J., Zhabotynsky, V., Sun, W., Huang, S., Pakatci, I. K., Kim, Y., Wang, J. R., Morgan, A. P., Calaway, J. D., Aylor, D. L., Yun, Z., Bell, T. A., Buus, R. J., Calaway, M. E., Didion, J. P., Gooch, T. J., Hansen, S. D., Robinson, N. N., Shaw, G. D., … de Villena, F. P. M. (2015). Analyses of allele-specific gene expression in highly divergent mouse crosses identifies pervasive allelic imbalance. *Nature Genetics*, *47*(4), 353–360. https://doi.org/10.1038/NG.3222

Davies, W., Smith, R. J., Kelsey, G., & Wilkinson, L. S. (2004). Expression patterns of the novel imprinted genes Nap1l5 and Peg13 and their non-imprinted host genes in the adult mouse brain. *Gene Expression Patterns*, *4*(6), 741–747. https://doi.org/10.1016/j.modgep.2004.03.008

Day, T., & Bonduriansky, R. (2004). Intralocus sexual conflict can drive the evolution of genomic imprinting. *Genetics*, *167*(4), 1537. https://doi.org/10.1534/GENETICS.103.026211

DeBaun, M. R., Niemitz, E. L., & Feinberg, A. P. (2003). Association of in vitro fertilization with Beckwith-Wiedemann syndrome and epigenetic alterations of LIT1 and H19. *American Journal of Human Genetics*, *72*(1), 156–160. https://doi.org/10.1086/346031

DeChiara, T. M., Robertson, E. J., & Efstratiadis, A. (1991). Parental imprinting of the mouse insulin-like growth factor II gene. *Cell*, *64*(4), 849–859. https://doi.org/10.1016/0092-8674(91)90513-X

Dehingia, B., Milewska, M., Janowski, M., & Pe Z Kowska, A. (2022). CTCF shapes chromatin structure and gene expression in health and disease. *EMBO Reports*, *23*(9), e55146. https://doi.org/10.15252/EMBR.202255146

Deichmann, U. (2016). Epigenetics: The origins and evolution of a fashionable topic. *Developmental Biology*, *416*(1), 249–254. https://doi.org/10.1016/J.YDBIO.2016.06.005

Delgado, M. S., Camprubí, C., Tümer, Z., Martínez, F., Milà, M., & Monk, D. (2014). Screening individuals with intellectual disability, autism and Tourette's syndrome for KCNK9 mutations and aberrant DNA methylation within the 8q24 imprinted cluster. *American Journal of Medical*

Genetics Part B: Neuropsychiatric Genetics, *165*(6), 472–478. https://doi.org/10.1002/AJMG.B.32250

Driscoll, D. J., Miller, J. L., Schwartz, S., & Cassidy, S. B. (2017). Prader-Willi Syndrome. *Neuropediatrics*, *29*, 124–126. https://www.ncbi.nlm.nih.gov/sites/books/NBK1330/

Dura, M., Teissandier, A., Armand, M., Barau, J., Lapoujade, C., Fouchet, P., Bonneville, L., Schulz, M., Weber, M., Baudrin, L. G., Lameiras, S., & Bourc'his, D. (2022). DNMT3A-dependent DNA methylation is required for spermatogonial stem cells to commit to spermatogenesis. *Nature Genetics 2022 54:4*, *54*(4), 469–480. https://doi.org/10.1038/s41588-022-01040-z

Edmondson, A. C., & Kalish, J. M. (2015). Genetic Advances in Intellectual Disability: Overgrowth Syndromes. *Journal of Pediatric Genetics*, *4*(3), 136. https://doi.org/10.1055/S-0035-1564440

Edmondson, D. G., Davie, J. K., Zhou, J., Mirnikjoo, B., Tatchell, K., & Dent, S. Y. R. (2002). Site-specific loss of acetylation upon phosphorylation of histone H3. *The Journal of Biological Chemistry*, *277*(33), 29496–29502. https://doi.org/10.1074/JBC.M200651200

Edwards, C. A., & Ferguson-Smith, A. C. (2007). Mechanisms regulating imprinted genes in clusters. *Current Opinion in Cell Biology*, *19*(3), 281–289. https://doi.org/10.1016/J.CEB.2007.04.013

Egger, G., Liang, G., Aparicio, A., & Jones, P. A. (2004). Epigenetics in human disease and prospects for epigenetic therapy. *Nature*, *429*(6990), 457–463. https://doi.org/10.1038/NATURE02625

Eggermann, T., Perez de Nanclares, G., Maher, E. R., Temple, I. K., Tümer, Z., Monk, D., Mackay, D. J. G., Grønskov, K., Riccio, A., Linglart, A., & Netchine, I. (2015). Imprinting disorders: a group of congenital disorders with overlapping patterns of molecular changes affecting imprinted loci. *Clinical Epigenetics*, *7*(1), 1–18. https://doi.org/10.1186/S13148-015-0143-8/TABLES/2

Elena Martinez, M., Charalambous, M., Saferali, A., Fiering, S., Naumova, A. K., Germain, D. S., Ferguson-Smith, A. C., & Hernandez, A. (2014). Genomic imprinting variations in the mouse type 3 deiodinase gene between tissues and brain regions. *Molecular Endocrinology (Baltimore, Md.)*, *28*(11), 1875–1886. https://doi.org/10.1210/ME.2014-1210

Elhamamsy, A. R. (2016). DNA methylation dynamics in plants and mammals: overview of regulation and dysregulation. *Cell Biochemistry and Function*, *34*(5), 289–298. https://doi.org/10.1002/CBF.3183

Esteller, M. (2008). Epigenetics in cancer. *The New England Journal of Medicine*, *358*(11), 1148–1159. https://doi.org/10.1056/NEJMRA072067

Felsenfeld, G. (2014). The evolution of epigenetics. *Perspectives in Biology and Medicine*, *57*(1), 132–148. https://doi.org/10.1353/PBM.2014.0004

Ferguson-Smith, A. C. (2011). Genomic imprinting: the emergence of an epigenetic paradigm. *Nature Reviews. Genetics*, *12*(8), 565–575. https://doi.org/10.1038/NRG3032

Ferguson-Smith, A. C., & Bourc'his, D. (2018). The discovery and importance of genomic imprinting. *ELife*, *7*. https://doi.org/10.7554/ELIFE.42368

Ferguson-Smith, A. C., Cattanach, B. M., Barton, S. C., Beechey, C. v., & Surani, M. A. (1991). Embryological and molecular investigations of parental imprinting on mouse chromosome 7. *Nature*, *351*(6328), 667–670. https://doi.org/10.1038/351667A0

Ferrón, S. R., Andreu-Agulló, C., Mira, H., Sánchez, P., Ángeles Marqués-Torrejón, M., & Fariñas, I. (2007). A combined ex/in vivo assay to detect effects of exogenously added factors in neural stem cells. *Nature Protocols 2007 2:4*, *2*(4), 849–859. https://doi.org/10.1038/nprot.2007.104

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., & Ferguson-Smith, A. C. (2011). Postnatal loss of Dlk1 imprinting in stem cells and niche astrocytes regulates neurogenesis. *Nature 2011 475:7356*, *475*(7356), 381–385. https://doi.org/10.1038/nature10229

Filippova, G. N., Fagerlie, S., Klenova, E. M., Myers, C., Dehner, Y., Goodwin, G., Neiman, P. E., Collins, S. J., & 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. *Molecular and Cellular Biology*, *16*(6), 2802–2813. https://doi.org/10.1128/MCB.16.6.2802

Frías-Lasserre, D., & Villagra, C. A. (2017). The importance of ncRNAs as epigenetic mechanisms in phenotypic variation and organic evolution. *Frontiers in Microbiology*, *8*(DEC), 2483. https://doi.org/10.3389/FMICB.2017.02483/BIBTEX

Gabory, A., Ripoche, M. A., le Digarcher, A., Watrin, F., Ziyyat, A., Forné, T., Jammes, H., Ainscough, J. F. X., Surani, M. A., Journot, L., & Dandolo, L. (2009). H19 acts as a trans regulator of the imprinted gene network controlling growth in mice. *Development*, *136*(20), 3413–3421. https://doi.org/10.1242/DEV.036061

Gabriele, M., Brandão, H. B., Grosse-Holz, S., Jha, A., Dailey, G. M., Cattoglio, C., Hsieh, T. H. S., Mirny, L., Zechner, C., & Hansen, A. S. (2022). Dynamics of CTCF- and cohesin-mediated chromatin looping revealed by live-cell imaging. *Science (New York, N.Y.)*, *376*(6592), 476–501. https://doi.org/10.1126/SCIENCE.ABN6583

Gardiner-Garden, M., & Frommer, M. (1987). CpG Islands in vertebrate genomes. *Journal of Molecular Biology*, *196*(2), 261–282. https://doi.org/10.1016/0022-2836(87)90689-9

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., & Ward, A. (2011). Distinct physiological and behavioural functions for parental alleles of imprinted Grb10. *Nature*, *469*(7331), 534–538. https://doi.org/10.1038/NATURE09651

Gaszner, M., & Felsenfeld, G. (2006). Insulators: exploiting transcriptional and epigenetic mechanisms. *Nature Reviews. Genetics*, *7*(9), 703–713. https://doi.org/10.1038/NRG1925

Geneimprint : Genes. (n.d.). Retrieved February 20, 2023, from https://www.geneimprint.com/site/genes-by-species

Ginart, P., Kalish, J. M., Jiang, C. L., Yu, A. C., Bartolomei, M. S., & Raj, A. (2016). Visualizing allele-specific expression in single cells reveals epigenetic mosaicism in an H19 loss-of-imprinting mutant. *Genes & Development*, *30*(5), 567–578. https://doi.org/10.1101/GAD.275958.115

Gorkin, D. U., Barozzi, I., Zhao, Y., Zhang, Y., Huang, H., Lee, A. Y., Li, B., Chiou, J., Wildberg, A., Ding, B., Zhang, B., Wang, M., Strattan, J. S., Davidson, J. M., Qiu, Y., Afzal, V., Akiyama, J. A., Plajzer-Frick, I., Novak, C. S., … Ren, B. (2020). An atlas of dynamic chromatin landscapes in mouse fetal development. *Nature*, *583*(7818), 744–751. https://doi.org/10.1038/S41586-020-2093-3

Gotter, A. L., Santarelli, V. P., Doran, S. M., Tannenbaum, P. L., Kraus, R. L., Rosahl, T. W., Meziane, H., Montial, M., Reiss, D. R., Wessner, K., McCampbell, A., Stevens, J., Brunner, J. I., Fox, S. v., Uebele, V. N., Bayliss, D. A., Winrow, C. J., & Renger, J. J. (2011). TASK-3 as a Potential Antidepressant Target. *Brain Research*, *1416C*, 69. https://doi.org/10.1016/J.BRAINRES.2011.08.021

Greenberg, M. V. C., & Bourc'his, D. (2019). The diverse roles of DNA methylation in mammalian development and disease. *Nature Reviews Molecular Cell Biology 2019 20:10*, *20*(10), 590–607. https://doi.org/10.1038/s41580-019-0159-6

Gregg, C., Zhang, J., Weissbourd, B., Luo, S., Schroth, G. P., Haig, D., & Dulac, C. (2010). High Resolution Analysis of Parent-of-Origin Allelic Expression in the Mouse Brain. *Science (New York, N.Y.)*, *329*(5992), 643. https://doi.org/10.1126/SCIENCE.1190830

Grossman, S. R., Engreitz, J., Ray, J. P., Nguyen, T. H., Hacohen, N., & Lander, E. S. (2018). Positional specificity of different transcription factor classes within enhancers. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(30), E7222–E7230. https://doi.org/10.1073/PNAS.1804663115/-/DCSUPPLEMENTAL

Gupta, S., Kim, S. Y., Artis, S., Molfese, D. L., Schumacher, A., Sweatt, J. D., Paylor, R. E., & Lubin, F. D. (2010). Histone Methylation Regulates Memory Formation. *The Journal of Neuroscience*, *30*(10), 3589. https://doi.org/10.1523/JNEUROSCI.3732-09.2010

Guschin, D., & Wolffe, A. P. (1999). Transcriptional control: SWItched-on mobility. *Current Biology*, *9*(19), R742–R746. https://doi.org/10.1016/S0960-9822(99)80473-4

Haig, D. (1997). Parental antagonism, relatedness asymmetries, and genomic imprinting. *Proceedings of the Royal Society B: Biological Sciences*, *264*(1388), 1657. https://doi.org/10.1098/RSPB.1997.0230

Haig, D. (2003). The Kinship Theory of Genomic Imprinting. *Https://Doi.Org/10.1146/Annurev.Ecolsys.31.1.9*, *31*, 9–32. https://doi.org/10.1146/ANNUREV.ECOLSYS.31.1.9

Haig, D. (2004). Genomic imprinting and kinship: how good is the evidence? *Annual Review of Genetics*, *38*, 553–585. https://doi.org/10.1146/ANNUREV.GENET.37.110801.142741

Haig, D. (2012). Commentary: The epidemiology of epigenetics. *International Journal of Epidemiology*, *41*(1), 13–16. https://doi.org/10.1093/IJE/DYR183

Haig, D., & Westoby, M. (1989). Parent-Specific Gene Expression and the Triploid Endosperm. *Https://Doi.Org/10.1086/284971*, *134*(1), 147–155. https://doi.org/10.1086/284971

Hanna, C. W., & Kelsey, G. (2021). Features and mechanisms of canonical and noncanonical genomic imprinting. *Genes and Development*, *38*(11–12), 821–834. https://doi.org/10.1101/GAD.348422.121/-/DC1

Hanna, C. W., Pérez-Palacios, R., Gahurova, L., Schubert, M., Krueger, F., Biggins, L., Andrews, S., Colomé-Tatché, M., Bourc'His, D., Dean, W., & Kelsey, G. (2019). Endogenous retroviral insertions drive non-canonical imprinting in extra-embryonic tissues. *Genome Biology*, *20*(1), 1–17. https://doi.org/10.1186/S13059-019-1833-X/FIGURES/7

Hansen, A. S., Pustova, I., Cattoglio, C., Tjian, R., & Darzacq, X. (2017). CTCF and cohesin regulate chromatin loop stability with distinct dynamics. *ELife*, *6*. https://doi.org/10.7554/ELIFE.25776.001

Hata, K., Okano, M., Lei, H., & Li, E. (2002). Dnmt3L cooperates with the Dnmt3 family of de novo DNA methyltransferases to establish maternal imprints in mice. *Development*, *129*(8), 1983–1993. https://doi.org/10.1242/DEV.129.8.1983

Hnoonual, A., Graidist, P., Kritsaneepaiboon, S., & Limprasert, P. (2019). Novel Compound Heterozygous Mutations in the TRAPPC9 Gene in Two Siblings With Autism and Intellectual Disability. *Frontiers in Genetics*, *10*. https://doi.org/10.3389/FGENE.2019.00061

Hochgerner, H., Zeisel, A., Lönnerberg, P., & Linnarsson, S. (2018). Conserved properties of dentate gyrus neurogenesis across postnatal development revealed by single-cell RNA sequencing. *Nature Neuroscience*, *21*(2), 290–299. https://doi.org/10.1038/s41593-017-0056-2

Hogart, A., Leung, K. N., Wang, N. J., Wu, D. J., Driscoll, J., Vallero, R. O., Schanen, N. C., & LaSalle, J. M. (2009). Chromosome 15q11-13 duplication syndrome brain reveals epigenetic alterations in gene expression not predicted from copy number. *Journal of Medical Genetics*, *46*(2), 86. https://doi.org/10.1136/JMG.2008.061580

Holman, L., & Kokko, H. (2014). The evolution of genomic imprinting: costs, benefits and long-term consequences. *Biological Reviews*, *89*(3), 568–587. https://doi.org/10.1111/BRV.12069

Hsu, C. L., Chou, C. H., Huang, S. C., Lin, C. Y., Lin, M. Y., Tung, C. C., Lin, C. Y., Lai, I. P., Zou, Y. F., Youngson, N. A., Lin, S. P., Yang, C. H., Chen, S. K., Gau, S. S. F., & Huang, H. S. (2018). Analysis of experience-regulated transcriptome and imprintome during critical periods of mouse visual system development reveals spatiotemporal dynamics. *Human Molecular Genetics*, *27*(6), 1039–1054. https://doi.org/10.1093/HMG/DDY023

Huang, W. C., Ferris, E., Cheng, T., Hörndli, C. S., Gleason, K., Tamminga, C., Wagner, J. D., Boucher, K. M., Christian, J. L., & Gregg, C. (2017). Diverse Non-genetic, Allele-Specific Expression Effects Shape Genetic Architecture at the Cellular Level in the Mammalian Brain. *Neuron*, *93*(5), 1094-1109.e7. https://doi.org/10.1016/J.NEURON.2017.01.033

Hutagalung, A. H., & Novick, P. J. (2011). Role of Rab GTPases in Membrane Traffic and Cell Physiology. *Physiological Reviews*, *91*(1), 119. https://doi.org/10.1152/PHYSREV.00059.2009

Hutvagner, G., & Simard, M. J. (2008). Argonaute proteins: key players in RNA silencing. *Nature Reviews Molecular Cell Biology 2008 9:1*, *9*(1), 22–32. https://doi.org/10.1038/nrm2321

Ioannou, M. S., Liu, Z., & Lippincott-Schwartz, J. (2019). A Neuron-Glia Co-culture System for Studying Intercellular Lipid Transport. *Current Protocols in Cell Biology*, *84*(1), e95. https://doi.org/10.1002/CPCB.95

Iyengar, S., & Farnham, P. J. (2011). KAP1 protein: an enigmatic master regulator of the genome. *The Journal of Biological Chemistry*, *286*(30), 26267–26276. https://doi.org/10.1074/JBC.R111.252569

Jahn, R., & Scheller, R. H. (2006). SNAREs--engines for membrane fusion. *Nature Reviews. Molecular Cell Biology*, *7*(9), 631–643. https://doi.org/10.1038/NRM2002

Jenkins, M. L., Harris, N. J., Dalwadi, U., Fleming, K. D., Ziemianowicz, D. S., Rafiei, A., Martin, E. M., Schriemer, D. C., Yip, C. K., & Burke, J. E. (2020). The substrate specificity of the human TRAPPII

complex's Rab-guanine nucleotide exchange factor activity. *Communications Biology*, *3*(1). https://doi.org/10.1038/S42003-020-01459-2

Juan, A. M., Foong, Y. H., Thorvaldsen, J. L., Lan, Y., Leu, N. A., Rurik, J. G., Li, L., Krapp, C., Rosier, C. L., Epstein, J. A., & Bartolomei, M. S. (2022). Tissue-specific Grb10/Ddc insulator drives allelic architecture for cardiac development. *Molecular Cell*, *82*(19), 3613-3631.e7. https://doi.org/10.1016/J.MOLCEL.2022.08.021

Kaneda, M., Okano, M., Hata, K., Sado, T., Tsujimoto, H., Li, E., & Sasaki, H. (2004). Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting. *Nature*, *429*(6994), 900–903. https://doi.org/10.1038/NATURE02633

Kaneda, M., Tang, F., O'Carroll, D., Lao, K., & Surani, M. A. (2009). Essential role for Argonaute2 protein in mouse oogenesis. *Epigenetics and Chromatin*, *2*(1), 1–8. https://doi.org/10.1186/1756-8935-2-9/FIGURES/4

Kaneko-Ishino, T., & Ishino, F. (2015). Mammalian-specific genomic functions: Newly acquired traits generated by genomic imprinting and LTR retrotransposon-derived genes in mammals. *Proceedings of the Japan Academy. Series B, Physical and Biological Sciences*, *91*(10), 511. https://doi.org/10.2183/PJAB.91.511

Kaniskan, H. Ü., Martini, M. L., & Jin, J. (2018). Inhibitors of Protein Methyltransferases and Demethylases. *Chemical Reviews*, *118*(3), 989–1068. https://doi.org/10.1021/ACS.CHEMREV.6B00801/ASSET/IMAGES/LARGE/CR-2016-008015_0026.JPEG

Kelsey, G., Stegle, O., & Reik, W. (2017). Single-cell epigenomics: Recording the past and predicting the future. *Science (New York, N.Y.)*, *358*(6359), 69–75. https://doi.org/10.1126/SCIENCE.AAN6826

Ke, Y., Weng, M., Chhetri, G., Usman, M., Li, Y., Yu, Q., Ding, Y., Wang, Z., Wang, X., Sultana, P., DiFiglia, M., & Li, X. (2020). Trappc9 deficiency in mice impairs learning and memory by causing imbalance of dopamine D1 and D2 neurons. *Science Advances*, *6*(47), 7781–7799. https://doi.org/10.1126/sciadv.abb7781

Kim, J. J., Lipatova, Z., & Segev, N. (2016). TRAPP Complexes in Secretion and Autophagy. *Frontiers in Cell and Developmental Biology*, *4*(MAR). https://doi.org/10.3389/FCELL.2016.00020

Klenova, E. M., Nicolas, R. H., Paterson, H. F., Carne, A. F., Heath, C. M., Goodwin, G. H., Neiman, P. E., & Lobanenkov, V. v. (1993). CTCF, a conserved nuclear factor required for optimal transcriptional activity of the chicken c-myc gene, is an 11-Zn-finger protein differentially expressed in multiple forms. *Molecular and Cellular Biology*, *13*(12), 7612–7624. https://doi.org/10.1128/MCB.13.12.7612-7624.1993

Korostowski, L., Sedlak, N., & Engel, N. (2012). The Kcnq1ot1 Long Non-Coding RNA Affects Chromatin Conformation and Expression of Kcnq1, but Does Not Regulate Its Imprinting in the Developing Heart. *PLOS Genetics*, *8*(9), e1002956. https://doi.org/10.1371/JOURNAL.PGEN.1002956

Krämer, J., Beer, M., Bode, H., & Winter, B. (2021). Two Novel Compound Heterozygous Mutations in the TRAPPC9 Gene Reveal a Connection of Non-syndromic Intellectual Disability and Autism Spectrum Disorder. *Frontiers in Genetics*, *11*, 972. https://doi.org/10.3389/FGENE.2020.00972/BIBTEX

Kulaeva, O. I., Nizovtseva, E. v., Polikanov, Y. S., Ulianov, S. v., & Studitsky, V. M. (2012). Distant activation of transcription: mechanisms of enhancer action. *Molecular and Cellular Biology*, *32*(24), 4892–4897. https://doi.org/10.1128/MCB.01127-12

Laird, P. W., Zijderveld, A., Linders, K., Rudnicki, M. A., Jaenisch, R., & Berns, A. (1991). Simplified mammalian DNA isolation procedure. *Nucleic Acids Research*, *19*(15), 4293. https://doi.org/10.1093/NAR/19.15.4293

Latos, P. A., Pauler, F. M., Koerner, M. v., Şenergin, H. B., Hudson, Q. J., Stocsits, R. R., Allhoff, W., Stricker, S. H., Klement, R. M., Warczok, K. E., Aumayr, K., Pasierbek, P., & Barlow, D. P. (2012). Airn transcriptional overlap, but not its lncRNA products, induces imprinted Igf2r silencing. *Science (New York, N.Y.)*, *338*(6113), 1469–1472. https://doi.org/10.1126/SCIENCE.1228110

Laukoter, S., Pauler, F. M., Beattie, R., Amberg, N., Hansen, A. H., Streicher, C., Penz, T., Bock, C., & Hippenmeyer, S. (2020). Cell-Type Specificity of Genomic Imprinting in Cerebral Cortex. *Neuron*, *107*(6), 1160-1179.e9. https://doi.org/10.1016/J.NEURON.2020.06.031

Lauritzen, I., Zanzouri, M., Honoré, E., Duprat, F., Ehrengruber, M. U., Lazdunski, M., & Patel, A. J. (2003). K+-dependent cerebellar granule neuron apoptosis. Role of task leak K+ channels. *The Journal of Biological Chemistry*, *278*(34), 32068–32076. https://doi.org/10.1074/JBC.M302631200

Lee, H. J., Choi, N. Y., Lee, S. W., Lee, Y., Ko, K., Kim, G. J., Hwang, H. S., & Ko, K. (2019). Alteration of Genomic Imprinting Status of Human Parthenogenetic Induced Pluripotent Stem Cells during Neural Lineage Differentiation. *International Journal of Stem Cells*, *12*(1), 31. https://doi.org/10.15283/IJSC18084

Lee, J. T., & Bartolomei, M. S. (2013). X-Inactivation, Imprinting, and Long Noncoding RNAs in Health and Disease. *Cell*, *152*(6), 1308–1323. https://doi.org/10.1016/J.CELL.2013.02.016

Lee, M., Chea, K., Pyda, R., Chua, M., & Dominguez, I. (2017). Comparative Analysis of Non-viral Transfection Methods in Mouse Embryonic Fibroblast Cells. *Journal of Biomolecular Techniques : JBT*, *28*(2), 67. https://doi.org/10.7171/JBT.17-2802-003

Lessel, D., Zeitler, D. M., Reijnders, M. R. F., Kazantsev, A., Hassani Nia, F., Bartholomäus, A., Martens, V., Bruckmann, A., Graus, V., McConkie-Rosell, A., McDonald, M., Lozic, B., Tan, E. S., Gerkes, E., Johannsen, J., Denecke, J., Telegrafi, A., Zonneveld-Huijssoon, E., Lemmink, H. H., … Kreienkamp, H. J. (2020). Germline AGO2 mutations impair RNA interference and human neurological development. *Nature Communications 2020 11:1*, *11*(1), 1–14. https://doi.org/10.1038/s41467-020-19572-5

Li, C., Luo, X., Zhao, S., Siu, G. K., Liang, Y., Chan, H. C., Satoh, A., & Yu, S. S. (2017). COPI-TRAPPII activates Rab18 and regulates its lipid droplet association. *The EMBO Journal*, *36*(4), 441–457. https://doi.org/10.15252/EMBJ.201694866

Linden, A. M., Sandu, C., Aller, M. I., Vekovischeva, O. Y., Rosenberg, P. H., Wisden, W., & Korpi, E. R. (2007). TASK-3 knockout mice exhibit exaggerated nocturnal activity, impairments in cognitive functions, and reduced sensitivity to inhalation anesthetics. *The Journal of Pharmacology and Experimental Therapeutics*, *323*(3), 924–934. https://doi.org/10.1124/JPET.107.129544

Li, X., Ito, M., Zhou, F., Youngson, N., Zuo, X., Leder, P., & Ferguson-Smith, A. C. (2008). A Maternal-Zygotic Effect Gene, Zfp57, Maintains Both Maternal and Paternal Imprints. *Developmental Cell*, *15*(4), 547–557. https://doi.org/10.1016/j.devcel.2008.08.014

Li, Y., & Sasaki, H. (2011). Genomic imprinting in mammals: its life cycle, molecular mechanisms and reprogramming. *Cell Research 2011 21:3*, *21*(3), 466–473. https://doi.org/10.1038/cr.2011.15

Llères, D., Moindrot, B., Pathak, R., Piras, V., Matelot, M., Pignard, B., Marchand, A., Poncelet, M., Perrin, A., Tellier, V., Feil, R., & Noordermeer, D. (2019). CTCF modulates allele-specific sub-TAD organization and imprinted gene activity at the mouse Dlk1-Dio3 and Igf2-H19 domains. *Genome Biology*, *20*(1), 1–17. https://doi.org/10.1186/S13059-019-1896-8/FIGURES/5

Lorthongpanich, C., Cheow, L. F., Balu, S., Quake, S. R., Knowles, B. B., Burkholder, W. F., Solter, D., & Messerschmidt, D. M. (2013). Single-cell DNA-methylation analysis reveals epigenetic chimerism in preimplantation embryos. *Science (New York, N.Y.)*, *341*(6150), 1110–1112. https://doi.org/10.1126/SCIENCE.1240617

Lo, W. S., Trievel, R. C., Rojas, J. R., Duggan, L., Hsu, J. Y., Allis, C. D., Marmorstein, R., & Berger, S. L. (2000). Phosphorylation of serine 10 in histone H3 is functionally linked in vitro and in vivo to Gcn5-mediated acetylation at lysine 14. *Molecular Cell*, *5*(6), 917–926. https://doi.org/10.1016/S1097-2765(00)80257-9

Lozano-Ureña, A., Montalbán-Loro, R., Ferguson-Smith, A. C., & Ferrón, S. R. (2017). Genomic Imprinting and the Regulation of Postnatal Neurogenesis. *Brain Plasticity*, *3*(1), 89. https://doi.org/10.3233/BPL-160041

Lucifero, D., Mann, M. R. W., Bartolomei, M. S., & Trasler, J. M. (2004). Gene-specific timing and epigenetic memory in oocyte imprinting. *Human Molecular Genetics*, *13*(8), 839–849. https://doi.org/10.1093/HMG/DDH104

Lykke-Andersen, K., Gilchrist, M. J., Grabarek, J. B., Das, P., Miska, E., & Zernicka-Goetz, M. (2008). Maternal Argonaute 2 is essential for early mouse development at the maternal-zygotic transition. *Molecular Biology of the Cell*, *19*(10), 4383–4392. https://doi.org/10.1091/MBC.E08-02-0219

Mabb, A. M., Judson, M. C., Zylka, M. J., & Philpot, B. D. (2011). Angelman syndrome: insights into genomic imprinting and neurodevelopmental phenotypes. *Trends in Neurosciences*, *34*(6), 293–303. https://doi.org/10.1016/J.TINS.2011.04.001

Macdonald, W. A., Sachani, S. S., White, C. R., & Mann, M. R. W. (2015). A role for chromatin topology in imprinted domain regulation1. *Biochemistry and Cell Biology*, *94*(1), 43–55. https://doi.org/10.1139/BCB-2015-0032

Manipalviratn, S., DeCherney, A., & Segars, J. (2009). Imprinting disorders and assisted reproductive technology. *Fertility and Sterility*, *91*(2), 305–315. https://doi.org/10.1016/J.FERTNSTERT.2009.01.002

Marangi, G., Leuzzi, V., Manti, F., Lattante, S., Orteschi, D., Pecile, V., Neri, G., & Zollino, M. (2012). TRAPPC9-related autosomal recessive intellectual disability: report of a new mutation and clinical phenotype. *European Journal of Human Genetics 2013 21:2*, *21*(2), 229–232. https://doi.org/10.1038/ejhg.2012.79

Marcho, C., Bevilacqua, A., Tremblay, K. D., & Mager, J. (2015). Tissue-specific regulation of Igf2r/Airn imprinting during gastrulation. *Epigenetics and Chromatin*, *8*(1), 1–10. https://doi.org/10.1186/S13072-015-0003-Y/FIGURES/5

Margolis, S. S., Sell, G. L., Zbinden, M. A., & Bird, L. M. (2015). Angelman Syndrome. *Neurotherapeutics*, *12*(3), 641. https://doi.org/10.1007/S13311-015-0361-Y

Mariño-Ramírez, L., Kann, M. G., Shoemaker, B. A., & Landsman, D. (2005). Histone structure and nucleosome stability. *Expert Review of Proteomics*, *2*(5), 719. https://doi.org/10.1586/14789450.2.5.719

Martin, B. K., Qiu, C., Nichols, E., Phung, M., Green-Gladden, R., Srivatsan, S., Blecher-Gonen, R., Beliveau, B. J., Trapnell, C., Cao, J., & Shendure, J. (2022). Optimized single-nucleus transcriptional profiling by combinatorial indexing. *Nature Protocols 2022 18:1*, *18*(1), 188–207. https://doi.org/10.1038/s41596-022-00752-0

Martin, C., & Zhang, Y. (2005). The diverse functions of histone lysine methylation. *Nature Reviews Molecular Cell Biology 2005 6:11*, *6*(11), 838–849. https://doi.org/10.1038/nrm1761

Martinez-Ara, M., Comoglio, F., van Arensbergen, J., & van Steensel, B. (2022). Systematic analysis of intrinsic enhancer-promoter compatibility in the mouse genome. *Molecular Cell*, *82*(13), 2519-2531.e6. https://doi.org/10.1016/J.MOLCEL.2022.04.009

Martini, P., Sales, G., Diamante, L., Perrera, V., Colantuono, C., Riccardo, S., Cacchiarelli, D., Romualdi, C., & Martello, G. (2022). BrewerIX enables allelic expression analysis of imprinted and X-linked genes from bulk and single-cell transcriptomes. *Communications Biology 2022 5:1*, *5*(1), 1–12. https://doi.org/10.1038/s42003-022-03087-4

Matise, M. P., & Joyner, A. L. (1999). Production of targeted embryonic stem cell clones. *Gene Targeting*. https://doi.org/10.1093/OSO/9780199637928.003.0007

Maupetit-Méhouas, S., Montibus, B., Nury, D., Tayama, C., Wassef, M., Kota, S. K., Fogli, A., Campos, F. C., Hata, K., Feil, R., Margueron, R., Nakabayashi, K., Court, F., & Arnaud, P. (2016). Imprinting control regions (ICRs) are marked by mono-allelic bivalent chromatin when transcriptionally inactive. *Nucleic Acids Research*, *44*(2), 621. https://doi.org/10.1093/NAR/GKV960

Mbimba, T., Hussein, N. J., Najeed, A., & Safadi, F. F. (2018). TRAPPC9: Novel insights into its trafficking and signaling pathways in health and disease (Review). *International Journal of Molecular Medicine*, *42*(6), 2991–2997. https://doi.org/10.3892/IJMM.2018.3889/HTML

McGrath, J., & Solter, D. (1984). Completion of mouse embryogenesis requires both the maternal and paternal genomes. *Cell*, *37*(1), 179–183. https://doi.org/10.1016/0092-8674(84)90313-1

McMahon, K. W., Karunasena, E., & Ahuja, N. (2017). The roles of DNA methylation in the stages of cancer. *Cancer Journal (Sudbury, Mass.)*, *23*(5), 257. https://doi.org/10.1097/PPO.0000000000000279

Meister, G., Landthaler, M., Patkaniowska, A., Dorsett, Y., Teng, G., & Tuschl, T. (2004). Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. *Molecular Cell*, *15*(2), 185–197. https://doi.org/10.1016/J.MOLCEL.2004.07.007

Messerschmidt, D. M. (2012). Should I stay or should I go: protection and maintenance of DNA methylation at imprinted genes. *Epigenetics*, *7*(9), 969–975. https://doi.org/10.4161/EPI.21337

Messerschmidt, D. M., de Vries, W., Ito, M., Solter, D., Ferguson-Smith, A., & Knowles, B. B. (2012). Trim28 is required for epigenetic stability during mouse oocyte to embryo transition. *Science (New York, N.Y.)*, *335*(6075), 1499–1502. https://doi.org/10.1126/SCIENCE.1216154

Mochida, G. H., Mahajnah, M., Hill, A. D., Basel-Vanagaite, L., Gleason, D., Hill, R. S., Bodell, A., Crosier, M., Straussberg, R., & Walsh, C. A. (2009). A Truncating Mutation of TRAPPC9 Is Associated with Autosomal-Recessive Intellectual Disability and Postnatal Microcephaly. *American Journal of Human Genetics*, *85*(6), 897. https://doi.org/10.1016/J.AJHG.2009.10.027

Montalbán-Loro, R., Lassi, G., Lozano-Ureña, A., Perez-Villalba, A., Jiménez-Villalba, E., Charalambous, M., Vallortigara, G., Horner, A. E., Saksida, L. M., Bussey, T. J., Trejo, J. L., Tucci, V., Ferguson-Smith, A. C., & Ferrón, S. R. (2021). Dlk1 dosage regulates hippocampal neurogenesis and cognition. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(11). https://doi.org/10.1073/PNAS.2015505118

Montefiori, L. E., Sobreira, D. R., Sakabe, N. J., Aneas, I., Joslin, A. C., Hansen, G. T., Bozek, G., Moskowitz, I. P., McNally, E. M., & Nóbrega, M. A. (2018). A promoter interaction map for cardiovascular disease genetics. *ELife*, *7*. https://doi.org/10.7554/ELIFE.35788

Moore, L. D., Le, T., & Fan, G. (2012). DNA Methylation and Its Basic Function. *Neuropsychopharmacology 2013 38:1*, *38*(1), 23–38. https://doi.org/10.1038/npp.2012.112

Morange, M. (2013). What history tells us XXXII. the long and tortuous history of epigenetic marks. *Journal of Biosciences*, *38*(3), 451–454. https://doi.org/10.1007/S12038-013-9354-3/METRICS

Morera, L., Lübbert, M., & Jung, M. (2016). Targeting histone methyltransferases and demethylases in clinical trials for cancer therapy. *Clinical Epigenetics 2016 8:1*, *8*(1), 1–16. https://doi.org/10.1186/S13148-016-0223-4

Morita, S., Horii, T., Kimura, M., Goto, Y., Ochiya, T., & Hatada, I. (2007). One Argonaute family member, Eif2c2 (Ago2), is essential for development and appears not to be involved in DNA methylation. *Genomics*, *89*(6), 687–696. https://doi.org/10.1016/J.YGENO.2007.01.004

Müller, M., Fazi, F., & Ciaudo, C. (2020). Argonaute Proteins: From Structure to Function in Development and Pathological Cell Fate Determination. *Frontiers in Cell and Developmental Biology*, *7*, 360. https://doi.org/10.3389/FCELL.2019.00360/BIBTEX

Nagano, T., & Fraser, P. (2009). Emerging similarities in epigenetic gene silencing by long noncoding RNAs. *Mammalian Genome : Official Journal of the International Mammalian Genome Society*, *20*(9–10), 557–562. https://doi.org/10.1007/S00335-009-9218-1

Nakanishi, K. (2022). Anatomy of four human Argonaute proteins. *Nucleic Acids Research*, *50*(12), 6618. https://doi.org/10.1093/NAR/GKAC519

Nestorov, P., Tardat, M., & Peters, A. H. F. M. (2013). H3K9/HP1 and Polycomb: two key epigenetic silencing pathways for gene regulation and embryo development. *Current Topics in Developmental Biology*, *104*, 243–291. https://doi.org/10.1016/B978-0-12-416027-9.00008-5

Nicholls, R. D., Knoll, J. H. M., Butler, M. G., Karam, S., & Lalande, M. (1989). Genetic imprinting suggested by maternal heterodisomy in nondeletion Prader-Willi syndrome. *Nature*, *342*(6247), 281–285. https://doi.org/10.1038/342281A0

Nordin, M., Bergman, D., Halje, M., Engström, W., & Ward, A. (2014). Epigenetic regulation of the Igf2/H19 gene cluster. *Cell Proliferation*, *47*(3), 189. https://doi.org/10.1111/CPR.12106

Ono, R., Nakamura, K., Inoue, K., Naruse, M., Usami, T., Wakisaka-Saito, N., Hino, T., Suzuki-Migishima, R., Ogonuki, N., Miki, H., Kohda, T., Ogura, A., Yokoyama, M., Kaneko-Ishino, T., & Ishino, F. (2006). Deletion of Peg10, an imprinted gene acquired from a retrotransposon,

causes early embryonic lethality. *Nature Genetics*, *38*(1), 101–106. https://doi.org/10.1038/NG1699

Pang, B., & Snyder, M. P. (2020). Systematic identification of silencers in human cells. *Nature Genetics 2020 52:3*, *52*(3), 254–263. https://doi.org/10.1038/s41588-020-0578-5

Pang, D. S. J., Robledo, C. J., Carr, D. R., Gent, T. C., Vyssotski, A. L., Caley, A., Zecharia, A. Y., Wisden, W., Brickley, S. G., & Franks, N. P. (2009). An unexpected role for TASK-3 potassium channels in network oscillations with implications for sleep mechanisms and anesthetic action. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(41), 17546–17551. https://doi.org/10.1073/PNAS.0907228106

Patten, M. M., Ross, L., Curley, J. P., Queller, D. C., Bonduriansky, R., & Wolf, J. B. (2014). The evolution of genomic imprinting: theories, predictions and empirical tests. *Heredity 2014 113:2*, *113*(2), 119–128. https://doi.org/10.1038/hdy.2014.29

Pennacchio, L. A., Bickmore, W., Dean, A., Nobrega, M. A., & Bejerano, G. (2013). Enhancers: five essential questions. *Nature Reviews. Genetics*, *14*(4), 288. https://doi.org/10.1038/NRG3458

Perez, J. D., Rubinstein, N. D., & Dulac, C. (2016). New Perspectives on Genomic Imprinting, an Essential and Multifaceted Mode of Epigenetic Control in the Developing and Adult Brain. *Annual Review of Neuroscience*, *39*, 347–384. https://doi.org/10.1146/ANNUREV-NEURO-061010-113708

Perez, J. D., Rubinstein, N. D., Fernandez, D. E., Santoro, S. W., Needleman, L. A., Ho-Shing, O., Choi, J. J., Zirlinger, M., Chen, S. K., Liu, J. S., & Dulac, C. (2015). Quantitative and functional interrogation of parent-of-origin allelic expression biases in the brain. *ELife*, *4*(JULY 2015), 41. https://doi.org/10.7554/ELIFE.07860

Peters, J. (2014). The role of genomic imprinting in biology and disease: an expanding view. *Nature Reviews. Genetics*, *15*(8), 517–530. https://doi.org/10.1038/NRG3766

Phillips, J. E., & Corces, V. G. (2009). CTCF: master weaver of the genome. *Cell*, *137*(7), 1194–1211. https://doi.org/10.1016/J.CELL.2009.06.001

Pidsley, R., Fernandes, C., Viana, J., Paya-Cano, J. L., Liu, L., Smith, R. G., Schalkwyk, L. C., & Mill, J. (2012). DNA methylation at the Igf2/H19 imprinting control region is associated with cerebellum mass in outbred mice. *Molecular Brain*, *5*(1), 1–9. https://doi.org/10.1186/1756-6606-5-42/TABLES/2

Plasschaert, R. N., & Bartolomei, M. S. (2014). Genomic imprinting in development, growth, behavior and stem cells. *Development (Cambridge, England)*, *141*(9), 1805. https://doi.org/10.1242/DEV.101428

Prickett, A. R., Barkas, N., McCole, R. B., Hughes, S., Amante, S. M., Schulz, R., & Oakey, R. J. (2013). Genome-wide and parental allele-specific analysis of CTCF and cohesin DNA binding in mouse brain reveals a tissue-specific binding pattern and an association with imprinted differentially methylated regions. *Genome Research*, *23*(10), 1624–1635. https://doi.org/10.1101/GR.150136.112

Pulix, M. (2017). *INVESTIGATIONS INTO THE FUNCTIONS AND REGULATION OF THE MICROCEPHALY-ASSOCIATED TRAPPC9 GENE.* https://doi.org/10.17638/03010130

Queller, D. C. (2003). Theory of genomic imprinting conflict in social insects. *BMC Evolutionary Biology*, *3*(1), 1–23. https://doi.org/10.1186/1471-2148-3-15/TABLES/10

Quenneville, S., Verde, G., Corsinotti, A., Kapopoulou, A., Jakobsson, J., Offner, S., Baglivo, I., Pedone, P. v., Grimaldi, G., Riccio, A., & 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*(3), 361–372. https://doi.org/10.1016/J.MOLCEL.2011.08.032

Ravid, T., & Hochstrasser, M. (2008). Diversity of degradation signals in the ubiquitin-proteasome system. *Nature Reviews. Molecular Cell Biology*, *9*(9), 679–689. https://doi.org/10.1038/NRM2468

Reinius, B., & Sandberg, R. (2015). Random monoallelic expression of autosomal genes: stochastic transcription and allele-level regulation. *Nature Reviews. Genetics*, *16*(11), 653–664. https://doi.org/10.1038/NRG3888

Riedmann, L. T., & Schwentner, R. (2010). miRNA, siRNA, piRNA and argonautes: news in small matters. *RNA Biology*, *7*(2), 133–139. https://doi.org/10.4161/RNA.7.2.11288

Ron, G., Globerson, Y., Moran, D., & Kaplan, T. (2017). Promoter-enhancer interactions identified from Hi-C data using probabilistic models and hierarchical topological domains. *Nature Communications 2017 8:1*, *8*(1), 1–12. https://doi.org/10.1038/s41467-017-02386-3

Rossetto, D., Avvakumov, N., & Côté, J. (2012). Histone phosphorylation: a chromatin modification involved in diverse nuclear events. *Epigenetics*, *7*(10), 1098–1108. https://doi.org/10.4161/EPI.21975

Rotondo, J. C., Selvatici, R., di Domenico, M., Marci, R., Vesce, F., Tognon, M., & Martini, F. (2013). Methylation loss at H19 imprinted gene correlates with methylenetetrahydrofolate reductase gene promoter hypermethylation in semen samples from infertile males. *Epigenetics*, *8*(9), 990–997. https://doi.org/10.4161/EPI.25798

Rougeulle, C., Glatt, H., & Lalande, M. (1997). The Angelman syndrome candidate gene, UBE3A/E6-AP, is imprinted in brain. *Nature Genetics*, *17*(1), 14–15. https://doi.org/10.1038/NG0997-14

Ruf, N., Bähring, S., Galetzka, D., Pliushch, G., Luft, F. C., Nürnberg, P., Haaf, T., Kelsey, G., & Zechner, U. (2007). Sequence-based bioinformatic prediction and QUASEP identify genomic imprinting of the KCNK9 potassium channel gene in mouse and human. *Human Molecular Genetics*, *16*(21), 2591–2599. https://doi.org/10.1093/HMG/DDM216

Santoni, F. A., Stamoulis, G., Garieri, M., Falconnet, E., Ribaux, P., Borel, C., & Antonarakis, S. E. (2017). Detection of Imprinted Genes by Single-Cell Allele-Specific Gene Expression. *American Journal of Human Genetics*, *100*(3), 444. https://doi.org/10.1016/J.AJHG.2017.01.028

Sanz, L. A., Chamberlain, S., Sabourin, J. C., Henckel, A., Magnuson, T., Hugnot, J. P., Feil, R., & Arnaud, P. (2008). A mono-allelic bivalent chromatin domain controls tissue-specific imprinting at Grb10. *EMBO Journal*, *27*(19), 2523–2532. https://doi.org/10.1038/EMBOJ.2008.142

Sasaki, H., Ishihara, K., & Kato, R. (2000). Mechanisms of Igf2/H19 Imprinting: DNA Methylation, Chromatin and Long-Distance Gene Regulation. *The Journal of Biochemistry*, *127*(5), 711–715. https://doi.org/10.1093/OXFORDJOURNALS.JBCHEM.A022661

Saxonov, S., Berg, P., & Brutlag, D. L. (2006). A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proceedings of the National*

Academy of Sciences of the United States of America, *103*(5), 1412–1417. https://doi.org/10.1073/PNAS.0510310103

Scagliotti, V., Esse, R. C. F., Willis, T. L., Howard, M., Carrus, I., Lodge, E., Andoniadou, C. L., & Charalambous, M. (2021). Dynamic expression of imprinted genes in the developing and postnatal pituitary gland. *Genes*, *12*(4). https://doi.org/10.3390/GENES12040509/S1

Schaaf, C. P., Gonzalez-Garay, M. L., Xia, F., Potocki, L., Gripp, K. W., Zhang, B., Peters, B. A., McElwain, M. A., Drmanac, R., Beaudet, A. L., Caskey, C. T., & Yang, Y. (2013). Truncating mutations of MAGEL2cause Prader-Willi phenotypes and autism. *Nature Genetics*, *45*(11), 1405. https://doi.org/10.1038/NG.2776

Schoenfelder, S., & Fraser, P. (2019). Long-range enhancer-promoter contacts in gene expression control. *Nature Reviews. Genetics*, *20*(8), 437–455. https://doi.org/10.1038/S41576-019-0128-0

Schultz, D. C., Ayyanathan, K., Negorev, D., Maul, G. G., & Rauscher, F. J. (2002). SETDB1: a novel KAP-1-associated histone H3, lysine 9-specific methyltransferase that contributes to HP1-mediated silencing of euchromatic genes by KRAB zinc-finger proteins. *Genes & Development*, *16*(8), 919–932. https://doi.org/10.1101/GAD.973302

Schultz, D. C., Friedman, J. R., & Rauscher, F. J. (2001). Targeting histone deacetylase complexes via KRAB-zinc finger proteins: the PHD and bromodomains of KAP-1 form a cooperative unit that recruits a novel isoform of the Mi-2alpha subunit of NuRD. *Genes & Development*, *15*(4), 428–443. https://doi.org/10.1101/GAD.869501

Schulz, R., Proudhon, C., Bestor, T. H., Woodfine, K., Lin, C. S., Lin, S. P., Prissette, M., Oakey, R. J., & Bourc'his, D. (2010). The Parental Non-Equivalence of Imprinting Control Regions during Mammalian Development and Evolution. *PLoS Genetics*, *6*(11), 1001214. https://doi.org/10.1371/JOURNAL.PGEN.1001214

Schwertman, P., Bekker-Jensen, S., & Mailand, N. (2016). Regulation of DNA double-strand break repair by ubiquitin and ubiquitin-like modifiers. *Nature Reviews. Molecular Cell Biology*, *17*(6), 379–394. https://doi.org/10.1038/NRM.2016.58

Segert, J. A., Gisselbrecht, S. S., & Bulyk, M. L. (2021). Transcriptional silencers: driving gene expression with the brakes on. *Trends in Genetics : TIG*, *37*(6), 514. https://doi.org/10.1016/J.TIG.2021.02.002

Sekita, Y., Wagatsuma, H., Nakamura, K., Ono, R., Kagami, M., Wakisaka, N., Hino, T., Suzuki-Migishima, R., Kohda, T., Ogura, A., Ogata, T., Yokoyama, M., Kaneko-Ishino, T., & Ishino, F. (2008). Role of retrotransposon-derived imprinted gene, Rtl1, in the feto-maternal interface of mouse placenta. *Nature Genetics*, *40*(2), 243–248. https://doi.org/10.1038/NG.2007.51

Singh, P. B. (2016). Heterochromatin and the molecular mechanisms of 'parent-of-origin' effects in animals. *Journal of Biosciences 2016 41:4*, *41*(4), 759–786. https://doi.org/10.1007/S12038-016-9650-9

Singh, P., Wu, X., Lee, D.-H., Li, A. X., Rauch, T. A., Pfeifer, G. P., Mann, J. R., & Szabó, P. E. (2011). Chromosome-Wide Analysis of Parental Allele-Specific Chromatin and DNA Methylation. *Molecular and Cellular Biology*, *31*(8), 1757. https://doi.org/10.1128/MCB.00961-10

Skaar, D. A., Dietze, E. C., Alva-Ornelas, J. A., Ann, D., Schones, D. E., Hyslop, T., Sistrunk, C., Zalles, C., Ambrose, A., Kennedy, K., Idassi, O., Carboni, G. M., Gould, M. N., Jirtle, R. L., & Seewaldt, V. L.

(2021). Epigenetic dysregulation of kcnk9 imprinting and triple-negative breast cancer. *Cancers*, *13*(23). https://doi.org/10.3390/CANCERS13236031/S1

Smith, R. J., Dean, W., Konfortova, G., & Kelsey, G. (2003a). Identification of Novel Imprinted Genes in a Genome-Wide Screen for Maternal Methylation. *Genome Research*, *13*(4), 558. https://doi.org/10.1101/GR.781503

Sonzogni, M., Zhai, P., Mientjes, E. J., van Woerden, G. M., & Elgersma, Y. (2020). Assessing the requirements of prenatal UBE3A expression for rescue of behavioral phenotypes in a mouse model for Angelman syndrome. *Molecular Autism*, *11*(1), 1–12. https://doi.org/10.1186/S13229-020-00376-9/FIGURES/5

Spencer, H. G., & Wolf, J. B. (2014). Genomic imprinting: theories and data. *Heredity*, *113*(2), 93. https://doi.org/10.1038/HDY.2014.52

Stelzer, Y., Kadener, S., & Correspondence, N. B. (2015). Differentiation of Human Parthenogenetic Pluripotent Stem Cells Reveals Multiple Tissue- and Isoform-Specific Imprinted Transcripts. *CellReports*, *11*, 308–320. https://doi.org/10.1016/j.celrep.2015.03.023

Surani, M. A. H., Barton, S. C., & Norris, M. L. (1984). Development of reconstituted mouse eggs suggests imprinting of the genome during gametogenesis. *Nature*, *308*(5959), 548–550. https://doi.org/10.1038/308548A0

Suzuki, S., Ono, R., Narita, T., Pask, A. J., Shaw, G., Wang, C., Kohda, T., Alsop, A. E., Marshall Graves, J. A., Kohara, Y., Ishino, F., Renfree, M. B., & Kaneko-Ishino, T. (2007). Retrotransposon silencing by DNA methylation can drive mammalian genomic imprinting. *PLoS Genetics*, *3*(4). https://doi.org/10.1371/JOURNAL.PGEN.0030055

Suzuki, S., Shaw, G., Kaneko-Ishino, T., Ishino, F., & Renfree, M. B. (2011). The Evolution of Mammalian Genomic Imprinting Was Accompanied by the Acquisition of Novel CpG Islands. *Genome Biology and Evolution*, *3*(1), 1276. https://doi.org/10.1093/GBE/EVR104

Tachiwana, H., Dacher, M., Maehara, K., Harada, A., Seto, Y., Katayama, R., Ohkawa, Y., Kimura, H., Kurumizaka, H., & Saitoh, N. (2021). Chromatin structure-dependent histone incorporation revealed by a genome-wide deposition assay. *ELife*, *10*. https://doi.org/10.7554/ELIFE.66290

Takai, D., & Jones, P. A. (2002). Comprehensive analysis of CpG islands in human chromosomes 21 and 22. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(6), 3740–3745. https://doi.org/10.1073/PNAS.052410099

Talley, E. M., Solórzano, G., Lei, Q., Kim, D., & Bayliss, D. A. (2001). CNS Distribution of Members of the Two-Pore-Domain (KCNK) Potassium Channel Family. *The Journal of Neuroscience*, *21*(19), 7491. https://doi.org/10.1523/JNEUROSCI.21-19-07491.2001

Tang, F., Barbacioru, C., Wang, Y., Nordman, E., Lee, C., Xu, N., Wang, X., Bodeau, J., Tuch, B. B., Siddiqui, A., Lao, K., & Surani, M. A. (2009). mRNA-Seq whole-transcriptome analysis of a single cell. *Nature Methods 2009 6:5*, *6*(5), 377–382. https://doi.org/10.1038/nmeth.1315

Tan, W. H., Bacino, C. A., Skinner, S. A., Anselm, I., Barbieri-Welge, R., Bauer-Carlin, A., Beaudet, A. L., Bichell, T. J., Gentile, J. K., Glaze, D. G., Horowitz, L. T., Kothare, S. v., Lee, H. S., Nespeca, M. P., Peters, S. U., Sahoo, T., Sarco, D., Waisbren, S. E., & Bird, L. M. (2011). Angelman syndrome: Mutations influence features in early childhood. *American Journal of Medical Genetics. Part A*, *155A*(1), 81–90. https://doi.org/10.1002/AJMG.A.33775

Tucci, V., Isles, A. R., Kelsey, G., Ferguson-Smith, A. C., Bartolomei, M. S., Benvenisty, N., Bourc'his, D., Charalambous, M., Dulac, C., Feil, R., Glaser, J., Huelsmann, L., John, R. M., McNamara, G. I., Moorwood, K., Muscatelli, F., Sasaki, H., Strassmann, B. I., Vincenz, C., & Wilkins, J. (2019). Genomic Imprinting and Physiological Processes in Mammals. *Cell*, *176*(5), 952–965. https://doi.org/10.1016/J.CELL.2019.01.043

Úbeda, F., & Gardner, A. (2010). A MODEL FOR GENOMIC IMPRINTING IN THE SOCIAL BRAIN: JUVENILES. *Evolution*, *64*(9), 2587–2600. https://doi.org/10.1111/J.1558-5646.2010.01015.X

van Ekelenburg, Y. S., Hornslien, K. S., Van Hautegem, T., Fendrych, M., Van Isterdael, G., Bjerkan, K. N., Miller, J. R., Nowack, M. K., & Grini, P. E. (2023). Spatial and temporal regulation of parent-of-origin allelic expression in the endosperm. *Plant Physiology*, *191*(2), 986–1001. https://doi.org/10.1093/PLPHYS/KIAC520

Varrault, A., Dubois, E., Le Digarcher, A., & Bouschet, T. (2020). Quantifying Genomic Imprinting at Tissue and Cell Resolution in the Brain. *Epigenomes*, *4*(3). https://doi.org/10.3390/EPIGENOMES4030021

Verdone, L., Agricola, E., Caserta, M., & di Mauro, E. (2006). Histone acetylation in gene regulation. *Briefings in Functional Genomics*, *5*(3), 209–221. https://doi.org/10.1093/BFGP/ELL028

Veselovska, L., Smallwood, S. A., Saadeh, H., Stewart, K. R., Krueger, F., Maupetit Méhouas, S., Arnaud, P., Tomizawa, S. ichi, Andrews, S., & Kelsey, G. (2015). Deep sequencing and de novo assembly of the mouse oocyte transcriptome define the contribution of transcription to the DNA methylation landscape. *Genome Biology*, *16*(1). https://doi.org/10.1186/S13059-015-0769-Z

Vu, T. H., & Hoffman, A. R. (1997). Imprinting of the Angelman syndrome gene, UBE3A, is restricted to brain. *Nature Genetics*, *17*(1), 12–13. https://doi.org/10.1038/NG0997-12

Wang, K. H., Kupa, J., Duffy, K. A., & Kalish, J. M. (2020). Diagnosis and Management of Beckwith-Wiedemann Syndrome. *Frontiers in Pediatrics*, *7*. https://doi.org/10.3389/FPED.2019.00562

Weake, V. M., & Workman, J. L. (2008). Histone ubiquitination: triggering gene activity. *Molecular Cell*, *29*(6), 653–663. https://doi.org/10.1016/J.MOLCEL.2008.02.014

Wei, J. W., Huang, K., Yang, C., & Kang, C. S. (2017a). Non-coding RNAs as regulators in epigenetics (Review). *Oncology Reports*, *37*(1), 3–9. https://doi.org/10.3892/OR.2016.5236/HTML

Weinstein, L. S., Xie, T., Qasem, A., Wang, J., & Chen, M. (2010). The role of GNAS and other imprinted genes in the development of obesity. *International Journal of Obesity (2005)*, *34*(1), 6–17. https://doi.org/10.1038/IJO.2009.222

Wen, X., Liao, P., Luo, Y., Yang, L., Yang, H., Liu, L., & Jiang, R. (2022). Tandem pore domain acid-sensitive K channel 3 (TASK-3) regulates visual sensitivity in healthy and aging retina. *Science Advances*, *8*(36), 8785. https://doi.org/10.1126/SCIADV.ABN8785/SUPPL_FILE/SCIADV.ABN8785_SM.PDF

WIEDEMANN, H. R. (1964). [FAMILIAL MALFORMATION COMPLEX WITH UMBILICAL HERNIA AND MACROGLOSSIA--A "NEW SYNDROME"?]. *Journal de Genetique Humaine*, *13*, 223–232. https://europepmc.org/article/med/14231762

Wilkinson, L. S., Davies, W., & Isles, A. R. (2007). Genomic imprinting effects on brain development and function. *Nature Reviews. Neuroscience*, *8*(11), 832–843. https://doi.org/10.1038/NRN2235

Wolf, J. B., & Hager, R. (2006). A Maternal–Offspring Coadaptation Theory for the Evolution of Genomic Imprinting. *PLOS Biology*, *4*(12), e380. https://doi.org/10.1371/JOURNAL.PBIO.0040380

Wolf, J. B., & Hager, R. (2009). Selective abortion and the evolution of genomic imprinting. *Journal of Evolutionary Biology*, *22*(12), 2519–2523. https://doi.org/10.1111/J.1420-9101.2009.01874.X

Wu, X., Yang, B., Udo-Inyang, I., Ji, S., Ozog, D., Zhou, L., & Mi, Q. S. (2018). Research Techniques Made Simple: Single-Cell RNA Sequencing and its Applications in Dermatology. *The Journal of Investigative Dermatology*, *138*(5), 1004. https://doi.org/10.1016/J.JID.2018.01.026

Xu, J., Carter, A. C., Gendrel, A. V., Attia, M., Loftus, J., Greenleaf, W. J., Tibshirani, R., Heard, E., & Chang, H. Y. (2017). Landscape of monoallelic DNA accessibility in mouse embryonic stem cells and neural progenitor cells. *Nature Genetics*, *49*(3), 377–386. https://doi.org/10.1038/NG.3769

Yamasaki, A., Menon, S., Yu, S., Barrowman, J., Meerloo, T., Oorschot, V., Klumperman, J., Satoh, A., & Ferro-Novick, S. (2009). mTrs130 is a component of a mammalian TRAPPII complex, a Rab1 GEF that binds to COPI-coated vesicles. *Molecular Biology of the Cell*, *20*(19), 4205–4215. https://doi.org/10.1091/MBC.E09-05-0387

Yamasaki-Ishizaki, Y., Kayashima, T., Mapendano, C. K., Soejima, H., Ohta, T., Masuzaki, H., Kinoshita, A., Urano, T., Yoshiura, K., Matsumoto, N., Ishimaru, T., Mukai, T., Niikawa, N., & Kishino, T. (2006). Role of DNA methylation and histone H3 lysine 27 methylation in tissue-specific imprinting of mouse Grb10. *Molecular and Cellular Biology*, *27*(2), 732–742. https://doi.org/10.1128/MCB.01329-06

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. *Human Molecular Genetics*, *12*(8), 837–847. https://doi.org/10.1093/HMG/DDG106

Yip, C. K., Berscheminski, J., & Walz, T. (2010). Molecular architecture of the TRAPPII complex and implications for vesicle tethering. *Nature Structural & Molecular Biology*, *17*(11), 1298. https://doi.org/10.1038/NSMB.1914

Zanzouri, M., Lauritzen, I., Duprat, F., Mazzuca, M., Lesage, F., Lazdunski, M., & Patel, A. (2006). Membrane potential-regulated transcription of the resting K+ conductance TASK-3 via the calcineurin pathway. *Journal of Biological Chemistry*, *281*(39), 28910–28918. https://doi.org/10.1074/jbc.M606092200

Zhang, G., & Pradhan, S. (2014). Mammalian epigenetic mechanisms. *IUBMB Life*, *66*(4), 240–256. https://doi.org/10.1002/IUB.1264

Zhang, J. M., Hou, W. B., Du, J. W., Zong, M., Zheng, K. L., Wang, W. J., Wang, J. Q., Zhang, H., Mu, Y. S., Yin, Z., Ding, C. M., Sun, Q. Y., Liu, Z. H., & Kong, Q. R. (2020). Argonaute 2 is a key regulator of maternal mRNA degradation in mouse early embryos. *Cell Death Discovery 2020 6:1*, *6*(1), 1–14. https://doi.org/10.1038/s41420-020-00368-x

Zhang, P., Wu, W., Chen, Q., & Chen, M. (2019). Non-Coding RNAs and their Integrated Networks. *Journal of Integrative Bioinformatics*, *16*(3). https://doi.org/10.1515/JIB-2019-0027

Zhang, X., Li, B., Rezaeian, A. H., Xu, X., Chou, P. C., Jin, G., Han, F., Pan, B. S., Wang, C. Y., Long, J., Zhang, A., Huang, C. Y., Tsai, F. J., Tsai, C. H., Logothetis, C., & Lin, H. K. (2017). H3 ubiquitination by NEDD4 regulates H3 acetylation and tumorigenesis. *Nature Communications*, *8*. https://doi.org/10.1038/NCOMMS14799

Zippo, A., Serafini, R., Rocchigiani, M., Pennacchini, S., Krepelova, A., & Oliviero, S. (2009). Histone crosstalk between H3S10ph and H4K16ac generates a histone code that mediates transcription elongation. *Cell*, *138*(6), 1122–1136. https://doi.org/10.1016/J.CELL.2009.07.031

Zuo, X., Sheng, J., Lau, H. T., McDonald, C. M., Andrade, M., Cullen, D. E., Bell, F. T., Iacovino, M., Kyba, M., Xu, G., & 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. *The Journal of Biological Chemistry*, *287*(3), 2107–2118. https://doi.org/10.1074/JBC.M111.322644

Zwart, R., Sleutels, F., Wutz, A., Schinkel, A. H., & Barlow, D. P. (2001). Bidirectional action of the Igf2r imprint control element on upstream and downstream imprinted genes. *Genes & Development*, *15*(18), 2361. https://doi.org/10.1101/GAD.206201

Appendix

Enhancer sequences

Enhancer sequences used in the promoter-reporter assay, coordinates represent the mouse genome. These enhancers are not conserved between mice and human.

Enhancer A:

>mm10_dna range=chr15:72,625,407-72,626,800 5'pad=200 3'pad=200 strand=+ repeatMasking=none
TGTAACTAGATCAGACGCAGGACACAGATATCATTTACCTTCTTACAAGT
CGTGATCAGGTACACTGTTATAAAACTAACGACTAACCTTGGCTTTAATT
CTTATTTTCAAGTTGAATATGTTTCTTGGGAGACTCCCTGAGACTGAGCT
GGGTCCTACATGTTAGTGAACTGGTCATGGACAGGCTCTGCCCAGGGATC
AGGAGATGCCACCCTGCCCCAGTCACCACAGGACATCTGGTCTTTGATTT
CATCTCGGATAGCTGGAGGCTGGTTCCAGAGAGCAGAGTGGGCGAGAAGT
TTTGATTTGGCCTTTTCCTCAGAGCTTCTGTCAGGAACAAATTGGCTGGG
GTGTGAGAGCTTTGAAAGGCCAAAGGCTCTTAGGGGCTAAGAGAAGACAC
CAGTCTTCCTCCTCATGATGAATAACACACAGCCAGCCAACCACCCATGG
GGCCACCCACAGTCACCTGCAATAGTGCAAAGGTTGTCAGGTTCCAATCC
ATGCTGTAGCCTGGCAAGCTGCTTACTGCTTTGAACTTGTGTCCTACTGG
TACGCAAATGATGAGGCCCAGAGCATTGATGAATACTGTGGTCAGCCCAC
AGCGGCCTTCCCAACCCCCTCCTGCCTGATGCCCCAATTCAGTCCTTCCC
TGGTAGCTGGTTTATATACTAGCATACACTGGAGGAGGGAGGGTGGTGGG
TTAATTCTAAAATTTAACTTGCTTTCCTTTTCCCACAGGGAGAAAGAGTG
GGAACAGGTTTCTAAAAATAGGCTGGCCTGTCAAAACTGTTAAGGATCAG
GCAGGGCGAAGACGCTGCTGGGGGGAGAGAGAGAGCAGTGGGGCTGCTGT
GGAGGGCTCCATGCAGGACCTTCCCCAGGCCACAGCCCTGGGGCTCGTGC
CCCTGGCACGGCCAGTCAGGACAGCCTGCTGTCCTATGGGCACAGCGACC
AGCTGGGTGATGCTGTCGGAGCGTGGTACGGGGACTGCAAATGGCTGCTG
ATGGCCACTGTCACAGCAGAAGGATGGAAGGCTGCTCTATGTCAGATGCT
GCCCTGGGTGTGCTTTTTTTCTCCTACCCCCCTCCTCACAGCTATGCAGG
GATGAAGTCACCATAGCTCCACCATCAGGTGACAAAGAAGGGCTTAGAAA
GGAGTCTCTCCAGGCATCCAGCTAGACCATGAGCAAGTCAGGCTGCGAGC
CTAGTTCCTCTAAGGCTGTCATGGAGCTCCCCAGGTACTGCTTGTGTGTG
TTGTCAGTCCCTCCTGAAGAGCCCGAGGGTTGTGTCAGACCAGGACTCTA
CTCACAACCTGTTTTCCTCTCCAAGTAAAGACGACTCTTCAGGATGGAAG
AAGAGATGTGGGGGTGGGGCAGGCCAGGGGGAAGAGGGAAGTTTCAAAGT

Enhancer 2:

>mm10_dna range=chr15:72,657,344-72,658,462 5'pad=0 3'pad=0 strand=+
repeatMasking=none
GCAACATTGTTTTGGACTAGAATTTCTATTGGGTTTTGCTGGTCACGAAT
GTGACCTCTCCTCCTTGAGCCTTGGACTATAGCACCTGTTAGAAAAGAAG
GGTGCAGAGAGGTAGAAGCCGCCTTCAAAATTTCCCCTACTTGACCAATT
CCTCTGGATTCCTCGCCTTCTTTTGGCTGACTAGAGCTGCTTTTAATTTA
TTTTATAAATTGGCTGTACAAATGCTCCATTTGGTTAATGAATTCCCAGA
GCCCGGAAGGTCCTTCCTGGCTCAGGATTCAGGACCGAGAGCGCCTGCGC
TGTCTCAGCCCCTGTGCGGCAGGCCTGGGGCCATTGTGCATGTTCTCGCT
CAGTCTACCTACCTGTCTCCACATCTGCCCAGCTGGAGCTATGCATCCTT
GGCCCTCCCCATCTATGCCTCGAAGATTCTATTCTCATAGCCAAGAGGCA
TCCTTGCTGACAGCCCTGGGGACGTGGTAGGAGGCAGAAGAAAGGCCCGA
TCTCTGCTGCTGTTGCTGTAAACCAATCCCAGATGACGCAGGGGAGACCC
GCTGCCAAGGCTGCTTGCTTTAATTAGGGGGGAGATTGATAAGAGCTGTC
AGGGCAGAATGGAGAGACTCTAAGCCTCACTGGTTTGCTGAGACATGAGA
GGGAAGGCTGAAGATTAAAAAACAGAAAACTCCCCCTGGAAAGCTCACTA
GGGGCTGAGCCGAGGAAAGAACAGCAGGGTAGGAAAGAGTTGTGGGAACT
GCCCAGTGGCGCAACCCTCCACCAAGGTGGCCTTGCCTTGTGGGTTGTTT
GGGAAGGTCACCTATGATGGCCCTTCCATATAGCCACACATCTACCACAA
TGTTTTTCACGGAAGCTGGAGACTCAGAAATGGAAGTGCCTCCTGCCTCC
TCTCCAGGCAAGAGAATGCTCATCTCTGTATCCATGCAGGGCTCCATGTT
CTTGCCTTAACTACTTGTCCCCCAAGCCTTGCCCTGCCAGAGGCCACGTA
GGTATGCAGGAGGAAGCCAGCAGAAGAGTGACAGCCTGCTGAGGCCACGT
TTAAGGATGCTGTGGCCAAGGCCTGGCTCAGGCCTGTGTAGCTCCTTTCT
GTCCCTGTGTTTTTCCCAGCTGTCTGCCATGGTATACAGATGTGTAGGTG
TGCACAACTAAGGTGAACAGATAGCTGGTGTCTCACTTGGATTCCCAACA
AAGAAGAGAAAACAGGATTTGCTCCCCAAAAGTTGTCTGCAGCATGTTCT
TGAATCCCCATGGGCAGAGAGACTAAGGCTCTGAGAAATCTTAGCTTACA
TGTCGTTTGTGATGAGAAAAAGAGACGCACTCTTTCAGGAGGCTGCTCCC
AGTTAGAGATACTGTACTTCATACACTGAAGAAAATCACGCTTTTAAAAAA

Enhancer B:

>mm10_dna range=chr15:72,849,700-72,853,300 5'pad=0 3'pad=0 strand=+
repeatMasking=none
ACTCCTATGGGTCAGTTCCTTCCCTACTACAAGTGGTATTCACCAAGATT
CTTTCTCAAACACACGTGAAAAAAGAGACTATGCATGTTAGCCATTTGTA
GCCCTTTAAGATTGCCTGAAAGCCAAGAAGCACTCAGTGGTACATACACC
CCCTGCTCACAGACCCCATGCTGGGCTGGGTTGTCTCCCTAGTCCCAATA
CTGCTAACATTGAGTCCTTGGATCCACCCCATGCCCTCTCTCCAAGCTTC
TGCCTGCTGTTGCCTGGGAAGCTAGCCTCCACCTTCCACTCCATTTCACA
CAGTTTGGTTTATAAAAAGCTGAAGTTTCTAAGTTGACATAGAAATAGCT
GAGTGAGCGTTCCTACTTAGAGGCAGTAAGCCCTGTGTGGGCATCCCAGT
GAAGACTGCAAACTCCCAGGAAGGAGGAGAGGCTGTGACAAGGCACAGAA
GCTTCAGGGTTTCTCAAATGAAACCTCCCTTAAGCCAGCGGGTAGACACA

CAATCCGCCAGTATATTAGTCCTCTTGGGAGGTGCTAAATACTTCTTAAC
TGATTTTTTTTTTAATCTATGCCAAAAGGCCATTAATTGTTTCCTGTTTT
CATCCATTTGGTCTTCATTTGTCACTAAGAGCGCTGTACATCAGCAAAGC
CCAGTCCCCATCCTCGTAGCAACTCACAGCCTGATGCTGCAGGCAGGGGA
AACCAAGGGAACTACAACCTGTGTAGGCATCCGTGTGCCCAGGCAGGGTA
GCAGCTGGGTGCCAGGTGAGGGAGGGCTGGGTTAGCTAGCAATCAGCCCC
CCTGCTGCAGAATCCCTACCCGGTAATCCTCCAGCCACTCCTCCCAACCT
GCTGCAGAATCCCTACCTGGTAATCCTCCAGCCACTCCTCCCAAGTTCTG
AAACTGTGCTTTACATCCGAGGCACTGTTCACTCCAAAGACCTATGTTAT
CCGAGGCACTGTTCACTCCAAAGACCTATGTTTGCCAATTAAGGGAGAGG
TTCTGAAGACAGAACAGATCTGTTTTGCCACTGTGGGTGGGGGCTGTATG
TCTGAGAATGCATGTGCACAGGTGTGCATGCCACCTTGTCTCCACCGAGT
GCCTGCAGTGGTGTGGCATTTAACAATGGTGTGAAAGGTTCTTTCCCAAA
TCACTGGCTGAAAGATAAAGGAGTTCAAAGACCCTCCCTACCAATTACAA
AAGGACAAAGTGTTCCCCATTCAGATTCTACACAGAAACACCTTGTATTC
TGCAATCCTCTGCCTCCAGATCCTTCAGCCCATTCACACAGCCATCTTCA
CAGAAGAAAAGTCACAGCATGGGAGATGGTAACTCCTAGCTCATGCCCTG
TAGATATTCCTTGCCACAGCCACTGTCTCTGAAGCCTGGGGGTCTCCCCT
TCCCCAAGCCATGGGCTTTTCTAGGTCCTACTTTTCTTTCACTTGACTGT
ACCAGTTAGTTATCAGTCGGTGAGGGTCGGTGCCAAATGGCTAATTGTCT
ATATCTCTGGCCAACATTTCCACTTTCTATTTATGATATTTACTTCATTC
ATCACTCTTAGCTCATCTGCCAACTATAAGCTGCTATGGGTAGAAGAACC
AGTCAGGGTTCAAACGCCATGAGTGCCTCCTACTAAGGTGTGTTGAAGGC
ACGAAGAACTGGGAATTCCATTTAGCCATGGCTCTCATCTTTCTGTGTGG
AATCCATGAGCCCTTCCAGACATCAGGTCTGTACATGCCCTAAGGAGTGC
TAGCTCAAGGTCTGCTGTGTGAATTACCAGTGAATGAATGAATATTTGGA
GGAAATACATACAATTTCTTTTCTTAATAACTTAGTAGCTGTATTGGGGA
GACCGTGATAGAGAAACAAGAACAATGGATCTTTAAGAATGCTTGTCTAT
GATTTAAGCTTTGTTAGTGATGGGCAGTAGCTACAAAGCTATCTTTAGTA
TCATCTGCTTAAAGGGCAAGAATGGAGACCATCTGTTAGCGGCAATCACT
TTTGCTGCCCATTCATTTATAAAGTATACATATTACTTAGGCATAAAATG
CACTCCTTTTTTTTTTTAATTTTTGGTTCTGAGTACATATATGTCTTCAC
ATGCCACAAAGCACCCTGGGGGGGAACAAAACATCCTATGGAGAGACTCT
TCCCAACCAAACGGGTTAATGGTAACCATGGCAACTGCCTCCAAGAATAG
CATCATCTGCAGTAGCATTCATGCTGGGACCGTCGACAGAATGGTTATAA
ATATACCTATTAGCAGCCTGAACTGAATCTTTAAACAGCACTGTTGGGAG
AAAGGGTACCTAAAAATAGCTCCCACGCTTGCATTTGAAGACGTGTCCTC
AGAAAAACCGCATCTTGTGCACATACTTTCACACACACAGGAAGATTTTT
TTTTTAATTTTCTCTTTTTCTCTAACAATGGGCACCAAAACACTGAAGAA
GCATGAGAGACAATGCAGAGCTTTATGGCCTTCTAATTTAATTGCCTCAT
TCCAAAAAACAGCCTCCACCCCCAAACAGAGCAGGCCAGACTTAGCAGTC
AGCAGGTGACAGGCAGGGAAGGGCAAGGTGGCCCCAGAGCCTGCACAGCA
GGGTGACCCACAGGCTCCACATACCATGCTCACAAGGAAGGCTGTTTCTG
CAGTGACCTCACAGGAACACTGGAGAGAGATAACAGCTACTCCCAGCAAA
GGGCCAGCCACTTCCCACATCCACATGATAGTCCTGTGCTTTGTCCACTC
TATGCAAGCAATCGTCTCCAGTGTTCCTGAGGCTTTATTCACACTGTCCC
TGAGGATGGCGAGTCTGCATACACGACAGGAAGTGAGAACCAGATGGTTA

CAGGCTTGAACTTTGGCTCAAACACATCATTACTTTTTTCGTATGAATAA
ATTAACATCTCCAGAGCTCATTTCTTTACTGCTGAAGAAGACTGGTAGAG
ACTTAGAGTACTGCTACAGGAAGTCAGGACAGAGAAATCTCAGTGCCCAG
TACCAGGCATACCCACCCTATGAGTTTCCCTCCTTGCCTTGATCCCCTTC
AACTCACAGACCCCAGCACTCTGAGAGCATGGGTCACAGTGCCTAAACAA
CAGATCTACAGAGCATGGCGCCCATCTGAAAAGCAGCTGCCAGTCAATGA
GCTACCCAGGCATAGCATGGGTGGGACAGGAGAGGGAGGAAAGGAAGCCA
CACCTCTTAAGCATAAAGGTGCATGGCTATCTGCCCAGTAGTTATGAAAG
ATAAATATTATAAAACAATATTCTGTCACTATGTGGGGCCCAGGCATTGG
TCAACCAAAGTGAAGCCAAACAGGCACAGCTTCCTGGTCACATCTTTAAC
TTGTGATCATCCTGGTCTCCACCCCGTGTGCTAGGTACCAGTACATCACA
AGCCTGATAACAGCAGGACCACTTGACTCAGTCCTATCCCTCAGTTTGTC
CTGTCTGTTCTTGGCTTCTCTTCTTATCATGTGAATTCTAATGTCAGCTC
TTCCCATTGATTTTAAAAGCCTGGTTAGGGGGCTGGTGAGATGGCTCAGT
GGGTTAGAGCACCCGACTGCTCTTCCGAAGGTCCAGAGTTCAAATCCCAGC

Enhancer C:

>mm10_dna range=chr15:72,860,700-72,862,300 5'pad=0 3'pad=0 strand=+
repeatMasking=none
TCACCTCTTCAAGACTCCATTCACTCTCTTGCTAATTGACGGGCGCAAAA
GTCAGTGTAGGCGGGACAATGGTAGGGCTTGATCTGCAGGAAGCCCTAAA
TATGTGCTGTATATATATACCCAGGGCCTCTATCATATGACTTCTCTGAT
CTCAGTGATAAACACACTAAGCCAGCAGCCCTGGCTGGGGCGCAAGGAGG
GTCAAGCACTGCATCTAACTGATGCTCAGCGGTATGCAACTTCACACCTA
CCCCTGCACAGCAAAAAGGTGCTGACACAGGCACTATAGCTGGCACAGGC
TTCAGTGCTCTCTGGAAAGAGGGCCACAAAGTGCTGCATGAGGCTTCTTC
TGTCGTCATGGCAAGCAAAATATTAACAAATTGTGTTTAATCTGCTTTCT
TTCTAGCATACAAGTAAGAGGCCTTATAGAAATCTGGCAACGGCTAAATC
TGCAGTAGAAACCCACTTTAATCCAGCTTTGCCCATGAGAAAGGAGGGCT
GAGAGGCCAACCTCAGCTTCCTCCCCATTGTCAGCAGCCTTGAGAGCAAG
AGGGTAGGGATTAGTATGAGAGTTAGAATCAGGGATGAAGGAACACAGTC
ACCAATCTGATATGGGGGCTGTCATGGCAGATCTAGAAGAGAGAGATGGC
TGAGAAAATGTAAAAATAAAAAAATCGTGAAAGCAGAGAAACACACACAC
ACACACACACACACACTGTAAAAACCCAGCTCCAAATGCAGGCATGAAAC
CAAGACTCTGCAGCAGCCTGGAGACACACTCCCTCTGCGAGTGAGCCTGC
CTGCCTTCTCCCCTGGGACAGGTGGGCAGAGGAGGGATAAGGAGCCCACC
AGGACTGATGGAGAAGTGAGCTTGGGAAGAGCACAGATGAGAAGAGCCAG
CAGGCAGGAGCACTGGGCACAGCTGCCATGTGGGGGGCAGGGCGGGGGGA
AGGGAGAGCCTGAAGGGCAGGGGTGAAGAGAGCTCCTCAAAGGAAAACCT
AATAATCAGCCACTCTCCTGTTGATTCCTCCCCACTAAATTTAATCTCCT
TCTCTCTTGTGTAAAATCCATACAGTAAGAAGCCAGCTTAGCGTCCTGCC
CCTGCACACATGCCCCTGCCGTGTTCCCTGGGCTCCCCCAGTGCATGCTG
AACAAAGGCGTTTTTCTGTCAGAGTCAGCTCACGTCCTTGTTCTGTTTTA
TGAATTGCCCTTTGTGTCTTAAGTCAGAGATATCTTTCCACACTAAAATA

CGTCTTCCTAGACCTTCATCCTTACTGGTCATGTGTCACACGGACAGATC
ATCATCAGTTAATCAAGTCGTCACTGCGGGGCATTTATGTTGCATCTACT
AACTCTGGCTTTATAAATACTACTGCGGCCGAGAACATCATGCACGCAGC
TTTGTACACCGGCCTATCTTATTTCTCAAGGTAAACTAACTGCCGAACCG
TGTGGGTCAATACATGGTGATAGGCTAAAAAAAAAAATGTCAAAGATATCT
GTGTTCCAGTCTCTAGCACCTATTAGGGTGTACCATGACACATAAGGACT
TTGCAGGAGTGACCAGTGAAGGTTATCTAGGTTGGTTTGGAATGCCATTGC

Enhancer 8:

>mm10_dna range=chr15:72,891,564-72,892,637 5'pad=0 3'pad=0 strand=+
repeatMasking=none
CCATGACCGAGAAGGAGATCGTGGACTATGTGGCCAGCCAGGTTACA
ACCGCCAAGAAGCTGCGCGGTGGTGTTGTGTTCGTGGACGAGGTGCCTAA
AGGACTGACCGGCAAGTTGGACGCCCGCAAGATCCGCGAGATTCTCATTA
AGGCCAAGAAGGGCGGCAAGATCGCCGTGTAATAATTCTAGAGTCGGGGC
GGCCGGCCGCTTCGAGCAGACATGATAAGATACATTGATGAGTTTGGACA
AACCACAACTAGAATGCAGTGAAAAAAATGCTTTATTTGTGAAATTTGTG
ATGCTATTGCTTTATTTGTAACCATTATAAGCTGCAATAAACAAGTTAAC
AACAACAATTGCATTCATTTTATGTTTCAGGTTCAGGGGGAGGTGTGGGA
GGTTTTTTAAAGCAAGTAAAACCTCTACAAATGTGGTAAAATCGATAAGG
ATCCTCTAGTGATCTAGAGTAGAGGTCCTCAGCCTTCCTCATGCTGACCC
TTTGACAGTTCCTCATGTTGTGGTGACCACCAGCCATAGTTATTTTGTTG
CTACTTCGTAGCTGTAATTTTGCTACCGTTATGAATCATAACGTAAATAT
CTGATATGCTGGATATCTGACATATGACCTCTAAAGGGGTCACAACCCAC
AGGTTGAGAACCACCAGCTTAGACAAAGCAAGGGGACGTGGACTGGGCTG
CTGAGCTTCAGGGTTGCTCAGCCCAGCTAGTGCTTTGAGAATATCAGCTA
CCAGCATTGGCGTGTTTGCAAGCCAACTGGAGATGCAGACTAGAGGGTCC
CATTGCAGGCCTCTGGAATCAGGCTCTGCATTTTAATGAGACTAGAATAT
TCAAGTCTGCGATGATGTATCCTTTGCTCTGTTGCTTTGAGCCCTGCTCA
TTGTTCTGTGTTGGCATCCCTTCCAGGAAGCCCTCCTCTCAACCCCCTGC
CTCTTCTGGCGCAGCAGTATTTTGTCTGCTGTGTTGGCAGCATCTGTCCC
CTCTGTCCTTCCCACTGTCACAACCTCCCTTTTGAAGTGCAGTGTCTGAC
ATGCTGAATGTACTCAATAAATATTTGCTGAAGAGAGCGATATCTTGGTC
CTAGGGTAGTTCCTGGGTCTGCATAGGAAGAGAAGCAGCCCTTGCCAGTA
GGGCCTGCATGTCTCTCAGTGCAGATATGCATTTACCGGAGAGACTGTTA
GGACCCTGAGTGCTGCTACCATGAATGTCCCAAGGTGACAATGCAGGCAG
CTTGAGTGGAGGCCTGACTGCCCACTGAGCCACTGAGTTTCTACCTTGCT
CCTCTGAAGAGGCCTCCAGAAGAAGAGTCGTCCTGACTTGCTAAGTGCTG
GGCTTGCACATAGCAGCAGGTAACAGCTAGTTCATCACAGGTATCCTTTG
AGCTTCCAGAAAAGACCCCTCTAAGTTCTGAAGACCACTTGCTTTGCTAT
GTCACTATCCTGCGCAGGATCCTCCTCGTCAGATCTTGGTCTGTTCCTGT
GGTTCTGTGTGCCTCC

Enhancer D:

>mm10_dna range=chr15:73,019,700-73,023,500 5'pad=0 3'pad=0 strand=+
repeatMasking=none
TGCACCACTCTAGCAATCTTCCAGAGCCTTGAAGCTCTCCCGAATAAAAA
CCAAGTCTGCTGTGCATAACAAATATGAACTAATAAAAACAATTCCTCAC
AGGAGGGCCTTGAAGTGTTAAACGCAGAAAGGTTGTGGAGCTCTTCGTGG
GACACCTTGCATCCCTACCAATGGTCACCGCTGTCATTGCTCGGAGCGGT
GCCGACACTGGAGAGAGCAGCCGGATGAATCATGCGTCAGAGATTAATAA
GTGTCTCTTCTCTGACAAGCACTGTGCCTGCACAGAGAGATGAATAAGCC
CCCATCCCTTCTCCAGAGTCTGCGTGTATTACCAGACTCGCCTACTGAAA
ACACAGTTTTAACATTCACCTCAGCTGTAACCTCCTAAGGTAAATACCCT
GGAACCTCCAAAAGCAGTGTGCCTGCGGTTCTCTCTGTGTCTGCAGTGCG
CCTGCGCTACTCTCCGTGCCTGCATCTGTTTACATTCTGCCTTGTTCTAA
AGAAGAATGTAAAATCAATTTTTTAAACACACACAATTTAAGAGCAGTAA
CAGCCAAAATATAAAAAAAAAAAAATCAAGGTGAAAAAAATGAGATTATAG
TTAAATAAAACCAGAGGAAGGTTCGATCAGAAAAATGCCTGCCAGGAAGT
GGGTGTAATTATTAAGGTATCTGTAGACTTGATGCTAAGGTTCCTAACTG
CTAAAGCGGAGAGGGAAAGCCAAGATTCCCCAAGCCCATGAGACCAAAGT
GACTTGCAAGTCTCCAAAGAGGAATTCCCAAGTATCCCTCTCACCCCTCC
TTAAAAATTACAGCTGCCTGCAGGGCACTGAGGTCTGGCTTCTTTTATCA
CCTTTTCTGGTTAGAGTGGAGATTTGAAACTCAGCTCAATGGGCAGTGCC
TTTCCAAGTTGGGAGGGGGAGGGGAGGGGGTAGAGCAGAGCTGGCCCCCA
AGGGGTCTCTGCAGACCCCTGGCTCTGGAATGACTGAGCCCCCCTCACTG
TCCTCACAGACTTCCACTCTCCTGGACTCTTAGTCAAGCCACACACCAGA
ATCCTGTTAGACCAGAAATTTCTCAGTTCTCTCTCTCTCTCCTCTCAATT
CCAGAAAGGAATCTAGCAGGTGACCTTCCTTCTCCACCCCCCCCCATTCA
ATGGTTTCCTTGTAAGAGCTGTGGATCCCCCCACGGTTCCCCCACAGTTC
CCCCACGGACTGACTGGCCCTGGAAGCCATTGGCCAGTTGGCCAGCTTTG
ACCTGTCACAGCATTACCTCAGCCCCTTCCCCAGATGACACCCTTCCCCA
CTCTCCCACCCTTCTGACTCTGAGCTCAACCTGACAGGAGCACTAGGGAT
GCAGTCACTCTGGTGCCTACCAGGCTCTTGCAGGGTTTCCTTAGCAACAG
CTTCATTAAGGAAACCCATCAAAAAAGAAATCCGAGGAGGAGGAGGCTGT
TCTTGCATCTGAACAGAGTAGCCTGCATAAAATCTGAATGGCTTTGCAAA
GAAAATTAACCAAAGACAAAGAGGTTGAAAGTCCCCTCTAAAGTCTCATA
AAGGAGAATGTTTTAAAGGGCACCAATTGTTCTGGAACTTAGAGATGTGA
GGAGGCGGAGGACGGGGCCACGCTGCGACCCTCAGAAGCTGGTCTAGTTT
TCCAGACACTTCCTCAGTAGCACTTTCTCGTAAGGTCACCTGCCATTTTC
TGTTGAAAACAAAAATGTATTCAGTCCCTCAGGCAGCCTAGGGTGGGGGA
GGGGAGCTGCCACCGGGTTGGGCTTTTGGAATCACACAAGCGAGGGAGCG
AGGGGGACTGAAAGAACAATGGGAGACTTTGTGAGCTGAGCTGGGTCTCA
CCCCTCAGTTTCCCTGTCTTTAAAACTGGCTGCAAGTGGTGAGGAAAGGA
ACTGACGAGACCTTTCTCAGGGTCTGCTCAGCACCAGATGGAGCCGCCAC
AAGCCCCACCAGCCAGCTACAGGGGAGACAGTCATCGCAGCCTGTGCATC
TTTCTTTGTTGCACCTTGCTCTACATTGCAACAGTGCTTATGCTCCTGGG
ATGAAGGCCTCCCGAGCTGCTCCTCCGCAGAAGGTATCCAATGGGAGCGT
GGTCTTTGTTTCTGTAGATCTGACTCACTTGGGAATTGTTTGCCTCCATG
TCTTTTGTTTATCAGCCGGTTTCCCCTCCTCAGCGCCGGCCAGTGTGGAC

CATGACTGTCTGCAGCTCACTGGACTTCAGGAGCAGGCTAGCAGGCAGGC
AGACAGAAACACTTTGCCTAGAGTCACTCCTGCCCATCCCAGAGGCAAGG
GAAGGGAAGGGAAGGAACACAAAGGGTGCATGAACACGTGTCCCAGGGAC
CCCTCACCTCTAAGCCACCATCTTCTTTGGTGCCTGAATCATTGTTCCTT
TTCCTAACCTCCCTCTGCAGATAGGGGTGGCTACATTTAGACCACAGAAT
CTCAAGGAACAGATTCCTACTTCAAGAAGTCTGTGAATGGATGCCAGTGA
GGTTTGCAGCCTCCTTGCCACCCTGCTGTTCCCAGACGGTGTGCAGCGGC
ACTAAGTAAAGGAGAGAGGTGGCTGCTTGCTCTCCACAGAAGACAAGCTG
AGAGGGTGCTGCATGGCTCCTCTATTCACAAACAGGGGAAAATGATGACT
CAAGTCCCAGCCCCTCTGTTCTCATATGAACCTTGTCATTTGCCTGTGAA
ACCCTCATGACCATTGCACAGATCTGTACACAGCCAGCTATTTGTTGCCA
TTTGGGAGTGCCAGATAATTACCAGAAATCACTCGCCTAAACCAGCCAAC
ACACGCAGGTCACCCTGTGCCTGCACAGGGAAGAGGGACTCTGCATCTAT
TTTGCAGAGCACCGATAAAGACAGTTTTATCCTCCAGTTCTGAGGTTTTG
GATTTTTGAGACAGAGTTTCTCTGTGTAATATCCCTGACTGTCCTGGACT
CCCTTTGTAGACCAGGCTGGCCTGAAACTCAAGAAATCTGCCTGCCTCTG
CCTCCTGAGTGCTGGGATTCAAAGTGTGTGGCACCACCAGCTCATGCAGG
GATTCTGAAAGACAGATGGAACTTGCACTAGTCACATCTCCATTCCTTCC
CAAATGTCCGCCATTTGCCGCCGAAGTCCTTCGTTACCATCAGCTCCACA
ATGGATCACACTTGATTAAGAACCAAAAGCTTGTTCTATTTCAGAAAAAC
TCCAGGCCCTTCTCTTTACTGTGATTATAAAATCCCACCAGAACTCATTA
TTTTGCTGCCTTGTCAACTAGCAGACAAACATCAGGATGAGACAAGGACC
GCAAAGTGCCTTCCTCCTCCGGCAGAATTCAAAGGCTGGTAACCGCCCTG
TGCCGCAGGCAGGGCCACAATTGCCACAGAGCATTGCTGCCCTCTAGTGT
TTGCCTGACGCCGAGCAAGATGATAGCAGAGCTGATCCAAACTACAGGGC
CCCACTGTGAATAGATAGCAGGTTGCAGTAAGTCCAGATCTCCTCCCTTT
AATCTGCAAAGCGGACTTCAGTAACTGCAGAGAGTGAGGGCCCACAGGTC
TGTAGAAGAGGAGCAGGGCTGGGAAGCCTGGCAGCCACCCACTACTCTAG
CACCAGCTAAGCCCAACAGACTCCACTCCCCAGACCTTTCCGGTCTGCTT
GGAGCCTCTTCTGTGTTCAAAGCTGTCTCCTGAACCGCCTGGGCCTGATA
ACAGCAGTAGCCCACTGCAGAGCTGTGGAAGGACACAGGACCATCTAACC
TAAATCCAGCCTCCCAGCTACAAATAATTCAAGGCTTCCCTGGACCTTTCC

Enhancer E:

>mm10_dna range=chr15:73,066,670-73,068,528 5'pad=0 3'pad=0 strand=+
repeatMasking=none
TGTAAGCACAGCTATAGAGGAGCTAACCCGCTAACTTCTGCAGCACCAGC
ACTCAATACAAACATACTGGTATAAAGACAATAGTCAATGACTACGAGGG
ATAATTGGGGGCAGGGCTATCACACAGCCCAAGCTGGCTTGAAACTAAGT
AATCCAGGATGACCTTGATCTTTTGAGGCTTTTGCCTCCACCTCCCAAAT
TCAGAATAATAGGGATGTGGGAACTGAGCTATATGCATGCTAACCAAGTA
GTCTACCAGCTGAGGCCCACCTTCAGCCAAACTCCTGGAATTTAACCCCA
ACGTTAGTGGACACAAACTCATTTCAAACCTCTCAGTGCCTCTGAGAGAC
AACTCTACTGCCCAGGTGAATGTAAAGCGAAAGAAAGGGGCGCAGTAGTA

AATTCTCAGCAAGCTGGGATGCTCTGCAGATCTGGGGTTGTGCTACACCT
GATTTCAGAGGTGGAGCCACCTGCAAATGTCCCTGGAAGAAAGAAGAATG
TACCCCCTTTTGCAAAACACTTATCACACCAGACTTGAGAGGAGAGGGCG
ACACTGACGTGTCAACCAAAGCATGGGCTCCTGGGGGATGGGGAGTGGGG
GACATGGTCGAAGTCACAAAACCTGGTATCACTCAATCCATTCCTTGAAT
GAATAGAATGAATGAGTGAATGAATGAAAGGCTGTGTGCTACGAATCCCT
AAACCAGGCACCAGTCGACCAATCAACGCCTTCTCCTGCAGCTCCGCCCC
CGAGGCCACACACCCCAGCCACTGTCAGAGTTCCCTCAGACTGCTTGCAA
ACTCCGGAGTAGTGGATTACAATCCTTTCTGAGTCCTGCGCATGCTCTGC
GGTGACTAGGGGGGACAAATTTAAAAGCGATTGGCTGGGGGGAAGACAAG
CTCCGCCCGACTCCGCCCACCCAGGCTCAGGACCAATCCACAGCCGAAGG
CTCGCCTTTCCCAGCACCGCCCCGCCGCCCGCCGCGCCCTGGGGCGGAAC
CTATCGTGAATCAGGTTGAGAAAAGAGAAGGGAGGGGCGAACGGGAAGTG
ACGCACGAGCGCCAACGGAAGGCGGGCGGCCGGAGTTTGTTGCCCTGGGA
ACTGAGCTCTCGACTACGTCAATCTTGCGTCATAGGTCTTGTATATTTTA
CGCATGCGCGGCCACATAGCAGAGCGCGGGAAGAACACTAATTTAGGCCA
AAAATGGGTTCAGGGTCGCCTGTGAAAGTGGCCTGGATTGGGTGGAAAAG
GGTGAGGTGTCCGGCCTGAGGCCCAGAAGGTGGGATAGGCGTCTCTATGA
AGATTCACACATTTTTTTTTCCTTTTGGGACAGAAGCCCCAGTAGCCCA
GGCTGGCCTTGAACTCCCTAATTAGCTAAGGGTAACCTTGAACTCGTGAT
CCCCAGCTTCCACCAGCGTGCCGAGACCGCAGGCTTAGATCATCACACTT
GGTTTATACCGTGCTGAGGGTCAAACCCAGGACCTCACAAGCCAGTCCCT
CTTATCCACAGAGCTACAGTTTCGTCCCTTCCTGAATTTGTATTCAAAAC
TCATCTGTGCCAGAACACGTCATTCCGGGTAACTTGAATGAATTTGTGTT
CTTGGCCATTCACCCCTGTTTGGCTCAGAAAAAAAAAAAAAAAACTAAAC
TAAAAAAAACCAATCATGCCTTTTTTTATTATTCTTTGTAACCACACCTT
AAAGCCGGGTGTGGTGGTACAAGCCTTTAATTCCAGCACTCTGGAGGCAG
AGGCAGATGATTCTCTGAGTTCAAGACCAGTCCGGTTTACAGAAAGATTC

Trappc9 mouse promoter sequence- No sequence homology with human

CCACTCTGGTGTTTCTATTACTATTGTTGTTGTTTCCAGGC
AGGTTTTCTCTATGTAATAGGTTAGCCTCGAACTCAGAGATTCTCTTCCTCTCCTGGGAT
TAAAGGCGTGTGCCGCCACCATCACTACATTTTAATGCATTTTTAATTGAAAAGAAATGC
ATACATATGGTAGAAAAACATCAATCAGAATACTATAAGCACATAAGGGAGAGTCTCCCC
TTCTGTCTTTTTGTTTCTGGAGCCACTGTTATGTATGCACACTGTAACCTCTATACATTA
CCGTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGAGAGAGAGAGA
GAGAGAGAGAGAGAGAGAGAGACAGACAGACAGATAGACAGATAGACAGACCCTGGTGAT
AGCTTGGGGTCTTATCTAAGAGTCAACAGGGAAATGTTAATCCAGCCTGGAGTCGGTGAC
TGGCGGTTCCCACGTAGGCACCTCCCCTAGTAAAGGTTTCTGCGCTCTTCGCTCCTCCTC
CGGCTCCTCCCATCTCCTGTCCCTTCCTGCTCCCTGCTCCCTCCCCGGATCCTCCACTTC
CTTCCCTTCACACTAGTCTTCTACTCCAGAGTCTGCCAGTACCACCCCTTCTTCCTCCTC
CTCCTCTCTCTTATCCCTCCTCCCTCTTCGCCGCCTCCCTCGCAGGCTCCAGGTCTTCGG
GAAAACCCGGGGCAGATCTTAGGCCGAAGCACGACTTGCGCACGCGCAGGAGCTGCGCCG

GCTCGGCCGGTGGTTCGGAGGTTCCAGGCTGCTCAGTCCCGCTGCCCTCCAGGCCTCGGC
TCCTGCTTGGGGCCGCTGCGCTGTGTAAGTGCCCGGTTAGGGCCGCGCGTTGGGGTTCCT
GGCCGGAGGGGCTGGCTGGGGGCCTAGCCCGGTCTGCACCTGCCCGGGTCCCTTGTCAAC
CCATGTCCCCCCCTGCTTGCCTGGTCCTCTGTCTGTCAGGTCCCCTTGTCTGCCCAGGTC
CCCTGCCCAGCCGGGGACAGCCTGCTACCCTGTTGCCAGTCGTTTGGCGCCCCGCCTACA
CCTGGGGTGCGGTATCCGCACCTAAAGGATCTGCGTACACCTGTCCGGGATGGGCTACAC
CTGCACCTCTTGGCGGGCCAGGCCGTCTGGCCCTAAGTCCCTCCTCTCTTCCGGTTCTGG
GACCTGGAATTGGGCTACACTGCACGCCGTGCGCCACGCCCCTCGCCCACAATTCTAGGC
CAGGAACAGGGTGCACTATCTCAAGCGGCCCAGTGCCTCAAACCCAGCCCAATGTGCAGT
TAACCTGGAATGAGATCTTTGCGCTGTAAACTCCAGTTTATCTCCAGTAACAGCTTGGGA
GAAGTGTCACTTTGCACTGGCCTTTGATCTCACCTCGCTGTTATATCCCCTCTTGGTACA
CTGTTAAGCAGAGCTTTCTGGTGTTAGGGGCCTTTAAGGGAGCTGAAACCAGGCTGCGTT
GGATTCAGTGAAAATGTATGAGGCCGACACACCCCCTTTAAAAGCAGTGACTTTGCAGAG
TTAAGGAGCTACAGTTGCTTCCTCATCTGCCTTGTTTGTTAATACCAAACAGTAGCACCA
GGGGTATGACATTGGAAACCTAGTAATATTCCGGAAGATTTCCCGGTGAGGTTTGGTTTT
CTTTGTAAATTGTTTTAGATGGTAGACTTATTAGTGGGATGGTGGTGGTGGGGAGATAAT
TGTCAAGAAACTTAATCTAAAAAAAAAAAAAAAAAAAAGAAAGAAAGAAAACAGTCCAGC
ATGAGACTCCTCCTGCCCGGGAGGATTGCCCGTTCAAGGCCAGCTTGGGCAAGGAGACTC
AGAGTCAAAAGAGGGGTGGGAGCAAGGGCTCATTGGGAGAGCACTTACTAAACACGTGCA
AGATCCTGTTCTGTCCCTAGCACCTTGTACAGTCCAGGGGCTCAACTATTGTATTCCAGG
ATCATTTCAGAACCATGCCTTTAGCTGTCGCTCTTCATCGTTGATCATTGCGGGGGTGGG
GGTGGGGGTGGGGATGGAGATGGAGAGTTGGTTGTTAGTAGTAGAGAGTTTCAGAACAAT
AAGATGGAAAGTTCTGGAAATGGATGATGGTAACATTGTACAACACCGAATACTCATGTT
TTTGGAACTGCATGCTTAAAAGTGGCAAATCTTAGGTTCATGTATTTTTTAAATATTTAT
TTATTTATTTTGTTTGAGACAGGGCTTCTCTGTGTAGCCCTGGCTGTCCTTGAACTCAGA
AATCCACCTACCTCTGCCTCCCAAGTGCTGGGATTAAAGGTGTGTGCCACTACTGCCCAG
CTGATTTATTTATTTTTATTTATACGAGTACACTGTAGCTGTCTTCAGACACACACCAGA
ACAGGGCATTGGATCCCATACAGATGGTTGTGAGCCACCATGTGGTTACTGGGAATTGAA
CTCAGGACTTCTGGAAGAGCAGTCAGTGCTCTTAACTGCTGAGCCATCTCTCCAGCTCCC
AGCATTGTTATTCGTAACTACTGCACCATCTCGCCAACTCTATTTATTTATTTATTTATT
TTATTTTATTTTTTAACAAATACTCTTTTTGAGATGGGGAACTCACTATGCAGCTCTGGC
TGGCTTGAAACTTACTTGGTAGACTAGGCTAGATTTGAACTCATAGAGACTTGGCAGCCT
CTGCCTTCTAAGAGGCTGGCATTAAAGGCACACTTTTTTTTTTTTTTTAAAGATAAGCAT
GACTGTGTAGTCCAGGCTCATCAGGAACTTATGGCCCAACACTCTGGCCTCCTGAGAGCC
TTGCATTGATCATATTTCGTGACTCTGTAGAGAGACCACATCCTCATTAGCAGTACTACC
TGGTCTTGCTTTCCTGAGGAATACATAAGCTCACCGTGGTAGCACATCAGGCGATTCTTA
GCTACTCTTGCCTATAAATAGCTTGTTTCACTAGTTAGATGCCTGGTAGGACTTCTAACC
AGAAGTGTTTTGAGTTTACATATTAAACTGAGGGGTCAGACAGAGTTGGGATTAGTGATA
CTCCTGTTCATGAAGAGACCCCCAATGCTGGTCTATTTGGTCCATGGTTAGTTTCAACTC
CAGCTTCTCATAGAAGGGGTAGGTAGCCTTGTGGGGTACTGTGTGAATCACTTGTGGCCA
CCCAGAGTGATAGCCAGGGTTGCCAGTGAGGATGAAGAGTGTGGCATGTCTCTGTCCACC
AGTGCTAATGTGTTATTGTGAGTTCTGCAAGGCAGAGATGTCTGTCATAGACTCTTTTCC
CTTTGGACAG

Acknowledgements

The completion of the thesis has been the result of countless hours of writing, time spent in the lab, and analysing results until far too late an hour. However, while I am personally responsible for writing this dissertation, I am fully aware that I would have been unable to do so without the plethora of people who have supported me along the way, both in my work and personal life, that have enabled me to reach where I am today.

There are so many people I have to thank for supporting me on this journey but the first and most important thank you has to be to my supervisor Toni who has guided me from a young, inexperienced graduate and taught me what it means to be a good scientist. Despite the mistakes I made along the way and the countless questions I have asked, his door has always been open to me, and the teachings and experience he has provided me will always be at the core of what makes me a scientist. An additional source of guidance that helped shape me into the researcher I am today and who I must thank is my secondary supervisor, Dr Daniel Messerschmidt, who helped me feel at ease during my two-year placement at the Agency for Science, Technology and Research in Singapore and provided essential support, especially during the troubling time of the COVID-19 pandemic when I was so far from my friends and family. His experience, sense of humour and critical thinking were a massive help and I will always be grateful for his support. Additional thanks need to be given to our collaborators, both for this thesis and the publication that I attained during this time. The pyrosequencing data would not have been possible without the assistance of Dr Lakis Liloglou who helped design the primers and allowed us to use his equipment, and Dr Gavin Kelsey and Dr Philippe Arnaud provided us with hybrid tissue samples that allowed a large portion of this research to take place.

The next people to thank are the amazing people I had the chance to work beside in the lab on a day to day basis both in Liverpool and Singapore. Sultan Aljuraysi guided me on the use of statistical software and assisted with dissections and cell cultures which was incredibly helpful, while Megan Smith was always there to vent to whenever an experiment went wrong and is one of the reasons I survived the stress of completing a PhD. The master's students, Ralph Bernado and Peng Zhou, were some of the best students I have ever seen, being an immense help in maintaining cell cultures, assisting with experiments and countless other things along with becoming great friends outside of the lab. A huge thank you also goes to other members of the Liverpool lab: Bettina, Thomas (Wilm), Trish, Chris, Arthur, Katherine, Francesco, Ana, Alejandra, Claudia, Niamh, Adams, present and previous MRes and undergraduate students, it has been an immense pleasure getting to know you all and seeing the amazing work you're doing. In Singapore, the whole Messerschmidt lab has helped me at one point or another but the two people who deserve a special thank you are Dr Michelle Seah, who taught me how to manually isolate single cells and hosted many game and dinner parties, and Dr Hui Mun Loh who was responsible for anything else I needed and has the patience of a saint for always showing me where reagents and equipment were in the lab even after I had asked a dozen times before.

The last people to thank are some of the most important people in my life, who have always encouraged me and helped me through so much and without whom this thesis wouldn't have been possible. The first are my parents who have supported me throughout my entire university experience and laid the foundation for what I have achieved today. I love you, you have made me the man I am today and I will always be grateful for everything you've done for me. Next are my brothers and their partners who have always celebrated my achievements with me and have been there to support me when things have been difficult.

And finally, Natashaa, who has been a pillar of emotional support throughout these last four and a half years, has seen me at my lowest and my best and has always loved me unconditionally throughout, I love you and thank you for everything!